Development of microsatellite markers for an outbreaking species of oak gall wasp, *Zapatella davisae* (Hymenoptera: Cynipidae), in the northeastern United States

J.C. Andersen¹*, C.P. Camp¹†, M.J. Davis¹, N.P. Havill² and J.S. Elkinton¹

¹Department of Environmental Conservation, University of Amherst
Massachusetts, 160 Holdsworth Way, Amherst, MA 01003, USA; ²United States Department of Agriculture Forest Service Northern Research Station, Hamden, CT 06514, USA

**Abstract**

Populations of the recently described black oak gall wasp, *Zapatella davisae* Buffington (Hymenoptera: Cynipidae), have been identified as the cause of extensive tree damage and mortality to black oaks, *Quercus velutina* Lamarck (Fagales: Fagaceae), in the northeastern United States. Relatively little is known, however, about the distribution, phylogenetic placement, and lifecycle of this important tree pest. Therefore, we conducted next-generation sequencing using the Ion Torrent™ PGM (ThermoFisher Scientific, Inc.) platform to develop genomic resources for the study of *Z. davisae* and for other closely related species of oak gall wasps. Individual sequence reads were aligned, assembled into unique contigs, and the contigs were then utilized for the *in silico* isolation and development of microsatellite markers. In total, we screened 36 candidate microsatellite loci, of which 23 amplified consistently (five polymorphic and 18 monomorphic). We then examined whether the polymorphic loci could be used to infer whether populations of *Z. davisae* from Cape Cod and Nantucket are sexual or asexual by calculating several metrics of genetic diversity that might indicate the mode of reproduction. These included testing for statistical deviations from Hardy–Weinberg equilibrium (HWE) and for linkage disequilibrium (LD), observations for the presence of the Meselson effect, and by calculating the probability that clonal individuals are more prevalent than would be expected in a randomly mating population. While we found significant deviations from HWE and more clonal individuals than expected, our estimates of the Meselson effect were inconclusive due to limited sampling, and we found no evidence of LD. Therefore, the sexual/asexual status of *Z. davisae* populations remains uncertain.

**Keywords:** next-generation sequencing, ion torrent, population genetics, reproductive mode, parthenogenesis

(Accepted 16 October 2018; First published online 29 November 2018)
Development of microsatellite markers for an outbreaking species of oak gall wasp, Zapatella davisae

Introduction

Wasps in the family Cynipidae (Hymenoptera) form intimate associations with their plant hosts through the formation of galls – plant induced structures that provide protection and enhanced nutrition for developing larvae (Ronquist & Liljeblad, 2001; Stone et al., 2002). Morphological and phylogenetic revisions support the division of the Cynipidae into several distinct tribes, within which the oak gall wasps (Cynipini) form a monophyletic assemblage (Liljeblad & Ronquist, 1998; Rokas et al., 2002, 2003; Ronquist et al., 2015). However, within the Cynipini, new genera and species are frequently being described (Pujade-Villar et al., 2010; Medianero, Nieves-Aldrey & Melika, 2011; Ide et al., 2012; Pujade-Villar et al., 2012; Melika et al., 2013; Tang et al., 2016) and existing species and genera are often re-assigned making inferences into the biological characteristics of members of this group difficult (Drown & Brown, 1998; Melika & Abrahamson, 2007). While the vast majority of the oak gall wasps persist at low population level densities, due in part to regulation by parasitoids and inquilines (Washburn & Cornell, 1981; Moriya et al., 1989; Stone et al., 2002), occasionally species are known to outbreak and cause host plant damage and even mortality (Eliason & Potter, 2000; Cooper & Rieske, 2007; Pujade-Villar et al., 2014).

One such outbreaking species in the northeastern United States is the newly described black oak gall wasp, Zapatella davisae Buffington et al. (Buffington et al., 2016). This species has caused extensive tree damage and mortality to black oaks, Quercus velutina Lamarck (Fagales: Fagaceae), primarily in the coastal regions of Connecticut, Massachusetts, New York and Rhode Island, with defoliation greatest in the coastal regions of Connecticut, Massachusetts, and New York (Davis, 2017). In these regions efforts to control outbreaking populations include injections of systemic pesticides into black oak trees including imidacloprid or emamectin benzoate (Davis, 2017; Davis & Elkinton, 2018). Unfortunately, basic information about the life history of Z. davisae, such as its geographic distribution are currently unknown (Davis, 2017). It is also unclear whether Z. davisae reproduces sexually, asexually, or has strictly alternating sexual and asexual generations at different times of the year – a form of cyclical parthenogenesis that is unique to the Cynipidae (Atkinson et al., 2002; Stone et al., 2002).

While Z. davisae was described as asexual based on the lack of observed males in sampled populations (Buffington et al., 2016), definitively identifying the mode of reproduction for a species can be a difficult process for several reasons (as reviewed in Schurko et al., 2009). These include, but are not limited to (1) that the presence of males does not guarantee that sexual reproduction occurs (Schafer et al., 2006); (2) that field observations of sex can be misleading because females in species that engage in intercourse do not necessarily utilize the sperm (Neiman & Lively, 2005) and (3) that many species are described as asexual due to the lack of males, but that cryptic sex can be common (Stone et al., 2008). Due to these difficulties, molecular surveys can be used to provide additional insights into a species’ mode of reproduction (Schurko et al., 2009). These methods include examinations for statistical deviations from the Hardy–Weinberg equilibrium (HWE; i.e., deviations from expected allele-frequencies for a randomly mating sexual population; Hardy 1908), and the detection of strong linkage disequilibrium (LD; i.e., the lack of statistical independence for alleles from different loci; Lewontin & Kojima 1960; Slatkin 2008). Unfortunately, both measures of HWE and LD are known to show deviations from expectations for reasons other than lack of sexual reproduction (e.g., gene-flow, selection, expanding/contracting population sizes, etc.). Therefore, additional molecular measures have been proposed. One measure, termed the Meselson effect, seeks to quantify allelic differentiation at a locus under the assumption that for sexual lineages recombination will constrain the differentiation of alleles whereas in asexual lineages the lack of recombination will result in greater differentiation of alleles (Birky, 1996). A second approach seeks to calculate statistical estimations for clonal population structure based on the genetic distance between individuals while accounting for possible scoring errors during molecular genotyping (Meirmans & Van Tienderen, 2004). However, as with deviations from HWE and the presence of LD, these molecular methods are not always definitive, and combinations of field surveys and molecular surveys are therefore desirable to provide a more accurate assessment of reproductive mode than either approach independently.

Therefore, in an effort to aide future research of this important pest species, we utilized next-generation sequencing to develop genomic resources for Z. davisae with the specific goals of: (1) generating sequence data that can be applied to the studies of Z. davisae and other oak gall wasps, (2) generating microsatellite loci that can be used in future studies of the population genetics of Z. davisae, and (3) utilizing the polymorphic microsatellite loci to examine the mode of reproduction for populations of Z. davisae on Cape Cod and Nantucket.

Materials and methods

Ion-torrent library preparation and sequencing

Whole genomic DNA was extracted from 20 individual wasps collected on Cape Cod using the Qiagen PureGene Tissue kit (Qiagen Corp.) following the manufacturer’s recommendations. DNA concentration and purity for each extract were then estimated using a NanoDrop 2000c spectrophotometer (ThermoFisher Scientific, Inc.). The two extracts with the highest concentration of DNA (ng μl⁻¹) and the highest ratios of DNA to contaminants (i.e., 260/280 and 260/230 ratios) were then selected for next-generation sequencing library preparation. These included one individual each from Dennis, MA (C04) and West Harwich, MA (C12). Libraries were individually prepared for sequencing on the Ion Torrent™ PGM (ThermoFisher Scientific, Inc.) platform, using the NEXTFlex™ DNA-seq kit (Bioo Scientific®, Inc.) with individual barcoding adaptors for each library (A5 for C04, and A7 for C12) by the Functional Genomics Laboratory at the University of California Berkeley. Library concentrations were then quantified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.), and pooled prior to sequencing at the DNA Analysis Facility at Science Hill, at Yale University on an Ion Torrent™ PGM 400 using the 318 v2 Chip kit.

Genomic assembly

Raw sequence reads were assembled into contigs using the de novo assembly program SOAP de novo 2 (Luo et al., 2012) with the default settings running on the University of Massachusetts Green High Performance Computing Cluster. Assembled contigs were compared with the NCBI GenBank.
‘nt’ sequence database (downloaded 11 July 2016) using the ‘blastn’ search algorithm (Altschul et al., 1990; Benson et al., 2013), with one best match sequence retained for each contig.

Microsatellite development

Contigs from the SOAP de novo 2 assembly were then used as reference sequences for microsatellite locus identification using the Phobos v. 3.3.11 (Mayer, 2010) plugin in Geneious v. 8.1.6 (Kearse et al., 2012), iMSAT (Andersen & Mills, 2014), and MSATCOMMANDER (Faircloth, 2008). For the analyses using Phobos, candidate microsatellite loci were filtered to only include exact repeats, a minimum unit length of two repeats, a maximum unit length of ten repeats, and to have no more than two successive ‘N’ calls in each contig. For the analyses using iMSAT, candidate microsatellite loci were filtered so that they would not include additional polymorphic sites within 150 bp of the target repeat. For the analyses using MSATCOMMANDER, candidate microsatellite loci were filtered to include only those loci with five or more repeat units. In addition, the raw sequence reads were assembled and used to identify microsatellite loci with the program QDD v. 3.1 (Meglecz et al., 2010, 2014) using the default parameters and sequences with successfully designed primers were then compared with the RepBase database of transposable elements (Jurka, 2000) to avoid loci that might be duplicated in the genome.

Primers for candidate microsatellite loci were then modified so that the forward primer would include a 5′ M13 tail (5′-TCCCAAGTCAGCTACG-3′) and the reverse primer would include a 5′ poly-T tail (5′-GTTT-3′) allowing for the use of a single 6-FAM fluorescently labeled M13 primer during primer testing, following the method presented by Schuelke (2000). Candidate loci were then filtered to only include those from which primers could be designed with an optimal $T_m$ of 55–60°C, and priority for screening was given to those loci with the largest number of repeat units.

Microsatellite marker testing

To verify the utility of candidate microsatellite loci for Z. davisae population genetics research, primer pairs were tested using field collected-individuals from two populations in Massachusetts (Nantucket and Dennis). Individuals were collected by dissecting Q. velutina twigs with visible galls, and specimens identified as Z. davisae were then placed in 95% ethanol and stored at −20°C prior to molecular analyses. DNA was extracted from 13 field-collected individuals from Nantucket and 19 field-collected individuals from Dennis using the Qiagen DNeasy kit following the manufacturer’s recommendations. For each candidate locus, polymerase chain reactions (PCR) were performed using the following 25 µl reaction conditions: 18 µl water, 5 µl Promega GoTag 5× buffer (Promega, Corp.), 0.5 µl dNTP (Promega), 0.1 µl 6-FAM M13 primer (100 µM), 0.05 µl candidate microsatellite forward primer (10 µM), 0.5 µl candidate microsatellite reverse primer (10 µM) and 0.2 µl Promega GoTag (Promega). For all primer pairs, a touchdown protocol was used that cycled through annealing temperatures of 57–50°C (Andersen & Mills, 2014). PCR products were visualized via electrophoresis using 1.5% agarose gels, and for reactions where >50% of samples had visible bands, PCR products were genotyped at the DNA Analysis Facility on Science Hill at Yale University on a 3730xl 96-Capillary Genetics Analyzer (Applied Biosystems, Inc.). Genotype scores were determined in comparison to the LIZ-500 size standard through the microsatellite plugin in Geneious v. 8.1.6. Candidate loci from which all genotyped samples had the same fragment length were recorded as ‘monomorphic’, and candidate loci from which more than one fragment length was observed among genotyped samples (though with no more than two alleles for any one individual) were recorded as ‘polymorphic’.

Population genetic surveys for Z. davisae

For each polymorphic locus, the following standard population genetic statistics were recorded: the number of alleles observed among populations (Num), the effective number of alleles (Eff_num), the observed heterozygosity (H_o), the expected heterozygosity within populations (H_e), the expected total heterozygosity (H_t), the total heterozygosity corrected for sampling from a limited number of populations (H'_t), and the inbreeding coefficient (G_o) using the program GENODIVE V 2.0b27 (Meirmans & Van Tienderen, 2004). These statistics were also calculated for each of the two surveyed populations. For each locus and each population, HWE and evidence LD were tested using GENEPOP (Raymond & Rousset, 1995; Rousset, 2008). To determine whether a specific microsatellite locus showed possible evidence of the Meselson effect (i.e., that two alleles from the same locus show a high level of differentiation), we compared the fragment length of alleles with each locus and calculated ‘coefficients of variation’ (CV) following Corral et al. (2009). Briefly, for each locus the standard deviation of the number of repeat motifs (σ) is divided by the mean number of repeats (μ) to estimate the CV. Lastly, to test whether populations of Z. davisae exhibit clonal population structure (i.e., the presence of more clones than would be expected under random-mating), we used the ‘Assign Clones’ algorithm implemented in GENODIVE with the stepwise-mutation model. Following the recommendations provided in the GENODIVE manual (available at: http://www.bentleydrummer.nl/software/GenoDive Manual.pdf), to account for scoring errors and/or possible mutations between clones, analyses were run using multiple genetic distances using the stepwise mutation model, of 0, 1, 10 and 20. For analyses at each distance, missing data were scored as ‘not counted’, and the presence of clonal diversity was assessed using the ‘Corrected Nei’s diversity index’. Significance was determined based on 999 bootstrap permutations, with alleles randomized over individuals within populations (i.e., the default settings for GENODIVE).

Results

Ion-torrent library preparation and sequencing

Sequencing of the individual barcoded libraries resulted in the generation of 5,032,726 sequence reads. Only 14 sequence reads could be assigned to sample C12, with the vast majority of sequence reads (4,136,285) being assigned to sample C04, while an additional 896,427 sequence reads were unassigned to either sample. Mean read lengths were 204, 244, and 243 bp to library C12, C04, and unassigned, respectively. After removal of low quality (<Q20) reads, 1,017,672,661 bp of Z. davisae associated genomic data was retained. Raw sequence reads for Z. davisae are available at the NCBI Short Read Archive (accession number SRR570816) as part of BioProject PRJNA386803.
Development of microsatellite markers for an outbreaking species of oak gall wasp, *Zapatella davisae* 513

Genomic assembly

De novo alignment of raw sequence reads resulted in the assembly of 832 unique contigs. The average length of contigs was 380 ± 263 bp, with an average coverage of 11x ± 15x. Sixty-three contigs were of sufficient length (≥250 bp) to conduct ‘blastn’ searches. Of those 63 contigs, the vast majority (n = 52) were matched with published sequences from insects, and the majority of those (n = 44) matched published hymenopteran sequences. The full search results using ‘blastn’ including the taxonomic distributions for each sequence are presented in Supplementary table S1.

Microsatellite development

In total, 1252 candidate microsatellite loci were identified across the four utilized software programs. PHOBOS identified the most candidate loci (n = 671), followed by QDD (n = 550). Both iMSAT (n = 16) and MSATCOMMANDER (n = 15) identified far fewer candidate loci. The number of candidate loci identified by each program for different repeat motifs, as well as the average, maximum and minimum numbers of repeats are presented in table 1.

After filtering candidate microsatellite loci to those that had optimal primer sites, 120 candidate microsatellite loci were retained, from which 36 pairs of candidate loci were tested. These included 11 loci identified by PHOBOS, 12 loci identified by iMSAT, ten loci identified by MSATCOMMANDER, and four loci from QDD (note: one locus, 971, was identified by both iMSAT and MSATCOMMANDER). Results from QDD indicated the presence of many additional optimal candidate loci, however, because this program is computationally more intensive than the others examined, results were not available until later stages of primer testing. A complete list of all candidate microsatellite loci, including their repeat motifs, repeat lengths, fasta formatted primer sequences (in 5’ to 3’ orientation), fasta formatted sequence for the microsatellite repeat region (including flanking regions), and the program from which they were identified are provided in Supplementary table S2.

Of the 36 loci tested, 18 were monomorphic (five from PHOBOS, eight from iMSAT, five from MSATCOMMANDER), five were polymorphic (two from PHOBOS, two from iMSAT, one from QDD), 13 loci (three from PHOBOS, two from iMSAT, five from MSATCOMMANDER, three from QDD) either failed to amplify or had bands indicating the amplification of PCR products of the target fragment length but resulted in fragment lengths that could not be scored (generally the result of stutter or nonspecific amplification). Based on these results, iMSAT had the highest percentage of candidate loci that resulted in scoreable fragment lengths (83%), while QDD had the highest percentage of candidate loci that were polymorphic (25%).

Population genetic surveys for *Z. davisae*

Population genetic statistics were generated for the five polymorphic loci. Among individuals collected on Nantucket and in Dennis, three loci (0781, 1333, and 2121) were bi-allelic, one locus (0887) was tri-allelic, and one locus (1643) was hexa-allelic. A summary of the genetic diversity for each of the polymorphic microsatellite loci is presented in table 2. Populations from Nantucket and Dennis had similar genetic diversity for each of the polymorphic microsatellite loci (table 3). These values ranged from 0 for locus 2121, to ~3 for locus 781 and locus 1643.

### Table 1. Summary of candidate microsatellite loci identified by different repeat motifs using PHOBOS, iMSAT, MSATCOMMANDER, and QDD.

<table>
<thead>
<tr>
<th>Motif</th>
<th>Count</th>
<th>Average repeats</th>
<th>Max repeats</th>
<th>Min repeats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Di</td>
<td>62</td>
<td>4.427</td>
<td>8.5</td>
<td>4</td>
</tr>
<tr>
<td>Tri</td>
<td>112</td>
<td>3.265</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Tetra</td>
<td>139</td>
<td>3.119</td>
<td>32.75</td>
<td>2.5</td>
</tr>
<tr>
<td>Penta</td>
<td>111</td>
<td>2.283</td>
<td>2.8</td>
<td>2.2</td>
</tr>
<tr>
<td>Hexa</td>
<td>126</td>
<td>2.064</td>
<td>2.833</td>
<td>2</td>
</tr>
<tr>
<td>Septa</td>
<td>77</td>
<td>2.165</td>
<td>2.857</td>
<td>2</td>
</tr>
<tr>
<td>Octa</td>
<td>23</td>
<td>2.076</td>
<td>2.375</td>
<td>2</td>
</tr>
<tr>
<td>Nona</td>
<td>11</td>
<td>2.121</td>
<td>2.667</td>
<td>2</td>
</tr>
<tr>
<td>Deca</td>
<td>10</td>
<td>2.31</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>SUM</td>
<td>671</td>
<td>4.071</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Di</td>
<td>14</td>
<td>3.5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Tri</td>
<td>2</td>
<td>2.125</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>SUM</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Di</td>
<td>8</td>
<td>6.125</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Tri</td>
<td>3</td>
<td>5.333</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Tetra</td>
<td>4</td>
<td>12.25</td>
<td>32</td>
<td>5</td>
</tr>
<tr>
<td>SUM</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Di</td>
<td>62</td>
<td>4.548</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>Tri</td>
<td>112</td>
<td>3.509</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Tetra</td>
<td>139</td>
<td>3.525</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>Penta</td>
<td>111</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Hexa</td>
<td>126</td>
<td>2.286</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>SUM</td>
<td>550</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dennis. A summary of genetic diversities for both populations is presented in table 3. Both populations displayed highly significant departures from the HWE (P < 0.0001 for both) while there was no evidence of LD within or across populations (P > 0.05 for all comparisons). Following Corral et al. (2009), estimates of CV were calculated for three of the five polymorphic microsatellite loci (table 4). These values ranged from 0 for locus 2121, to ~3 for locus 781 and locus 1643.

Significant presence (P < 0.001) of clonal population structure was observed at genetic distances of 0, 1, and 10, but not at a distance of 20. Individuals could be assigned to one of seven unique clones with a distance threshold of 0, to five unique clones with a distance threshold of 4, three unique clones with a distance threshold of 6, two unique clones with a distance threshold of 10, and all individuals were assigned as the same clone with a distance threshold of 28.

### Discussion

Oak gall wasps (tribe Cynipini) have fascinated researchers for generations with their capacity to manipulate their hosts into producing a diversity of plant-induced structures (Ronquist & Liljeblad, 2001; Cook et al., 2002; Stone et al., 2002; Askew et al., 2013). Many of these species can even produce morphologically distinct galls on different plant tissues, and/or on the same host tissue but at different times in the year (Cook et al., 2002). However, the study of these complex plant–insect interactions has been hampered by an overall lack
Table 2. Summary of allelic diversity for polymorphic microsatellite loci.

<table>
<thead>
<tr>
<th>Locus name</th>
<th>Repeat type</th>
<th>Repeat motif and range</th>
<th>n</th>
<th>Num</th>
<th>Eff_num</th>
<th>H_o</th>
<th>H_s</th>
<th>H_t</th>
<th>H′_t</th>
<th>G-IS</th>
<th>HWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0781</td>
<td>Tetra-</td>
<td>CAGA&lt;sup&gt;2-10&lt;/sup&gt;</td>
<td>31</td>
<td>2</td>
<td>1.155</td>
<td>0.094</td>
<td>0.140</td>
<td>0.139</td>
<td>0.138</td>
<td>0.329</td>
<td>0.160</td>
</tr>
<tr>
<td>0887</td>
<td>Tri-</td>
<td>AAT&lt;sup&gt;10-12&lt;/sup&gt;</td>
<td>32</td>
<td>3</td>
<td>1.215</td>
<td>0.000</td>
<td>0.190</td>
<td>0.200</td>
<td>0.211</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1333</td>
<td>Tetra-</td>
<td>TGTC&lt;sup&gt;13-17&lt;/sup&gt;</td>
<td>29</td>
<td>2</td>
<td>1.175</td>
<td>0.000</td>
<td>0.161</td>
<td>0.171</td>
<td>0.182</td>
<td>1.000</td>
<td>0.007</td>
</tr>
<tr>
<td>1643</td>
<td>Tri-</td>
<td>ATT&lt;sup&gt;14&lt;/sup&gt;</td>
<td>31</td>
<td>6</td>
<td>1.414</td>
<td>0.216</td>
<td>0.306</td>
<td>0.303</td>
<td>0.300</td>
<td>0.295</td>
<td>0.066</td>
</tr>
<tr>
<td>2121</td>
<td>Di-</td>
<td>AG&lt;sup&gt;16-23&lt;/sup&gt;</td>
<td>32</td>
<td>2</td>
<td>2.000</td>
<td>1.000</td>
<td>0.500</td>
<td>0.500</td>
<td>0.500</td>
<td>−1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

n, number of genotyped individuals; Num, number of alleles observed among populations; Eff_num, effective number of alleles; H_o observed heterozygosity, H_s heterozygosity within populations; H_t total heterozygosity; H′_t, corrected total heterozygosity; G-IS, inbreeding coefficient; HWE, probability of deviation from the Hardy–Weinberg equilibrium.

Table 3. Summary of allelic diversity for populations of Z. davisae.

<table>
<thead>
<tr>
<th>Population</th>
<th>n</th>
<th>Num</th>
<th>Eff_num</th>
<th>H_o</th>
<th>H_s</th>
<th>H_t</th>
<th>H′_t</th>
<th>G-IS</th>
<th>HWE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nantucket</td>
<td>13</td>
<td>1.857</td>
<td>1.288</td>
<td>0.177</td>
<td>0.181</td>
<td>0.181</td>
<td>−</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>Dennis</td>
<td>19</td>
<td>2.000</td>
<td>1.303</td>
<td>0.198</td>
<td>0.190</td>
<td>0.190</td>
<td>−</td>
<td>−0.041</td>
<td>0.000</td>
</tr>
</tbody>
</table>

n, number of genotyped individuals from each population; Num, number of alleles observed among populations; Eff_num, effective number of alleles; H_o observed heterozygosity, H_s heterozygosity within populations; H_t total heterozygosity; H′_t, corrected total heterozygosity; G-IS, inbreeding coefficient; HWE, probability of deviation from the Hardy–Weinberg equilibrium.

Table 4. CV for polymorphic microsatellite loci.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Locus</th>
<th>Locus</th>
<th>Locus</th>
<th>Locus</th>
<th>CV</th>
<th>Number of alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>781</td>
<td>887</td>
<td>1333</td>
<td>1643</td>
<td>2121</td>
<td>3.106</td>
<td>na</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2.799</td>
<td>0</td>
</tr>
</tbody>
</table>

CV, coefficient of variation. Calculated by dividing the standard deviation of the number of repeat motifs for a locus (σ) by the mean number of repeats (μ).

I suggests that individuals of Z. davisae in the northeastern United States have multiple close matches (>97% similarity) with individuals from Florida (BOLD: BBHY/H546-10), Nova Scotia (GenBank: KR084678.1, GenBank: KR782782.1), Ontario (BOLD: CNROFO73-13, BOLD: CNTIF3344-15, BOLD: CNRGD596-15, GenBank: KR409819.1, GenBank: KR414957.1), and Tennessee (BOLD: GMGSNM3442-13). However, due to the limited number of available sequences in public databases, and the lack of resolution provided by a single mitochondrial locus, it is unclear whether populations in the northeastern United States are recently introduced from the southeastern United States (Davis, 2017) reports that the highest % similarity matches were to individuals in Florida and Tennessee), or whether Z. davisae has a broad geographic range that includes much of eastern North America. Therefore, we hope to use the microsatellites developed in this study to sample populations of Z. davisae from throughout the eastern United States to determine whether outbreaks are the result of enemy release (which would be predicted if populations in the northeastern United States were recently introduced), or due to some other factor (if it has a broad geographic range that includes much of the eastern United States).

Development of microsatellite loci

The lack of taxonomic clarity that currently defines species in the Cynipini presents challenges for the study and control of the recently described black oak gall wasp, Z. davisae, that has been causing extensive damage and tree mortality of black oaks, Q. velutina, on Cape Cod, Long Island, Martha’s Vineyard, and Nantucket Island (Buffington et al., 2016). For example, little is known about its phylogenetic placement in the Cynipini, and/or its geographic distribution in the northeastern United States (Davis, 2017). Recent DNA sequencing by Davis (2017) of the mitochondrial locus cytochrome oxidase

Mode of reproduction

In addition to causing damage to black oaks, Z. davisae has been described as asexually based on the lack of observed males. Recent molecular work for oak gall wasps in the genus Andricus, however, found evidence for widespread cryptic sex among previously assumed asexual species (Stone et al., 2008). Therefore, we were curious as to whether or not we could utilize molecular methods to determine the reproductive mode of Z. davisae. While these methods have their advantage...
of overcoming many of the biases inherent in field observations, definitive determination of a species’ sexual status can still be difficult (Ellegren, 2004; Selkoe & Toonen, 2006; Bhargava & Fuentes, 2010). Unfortunately, our efforts for combining field observations with comparisons to multiple genetic signatures (deviations from the HWE, presence of LD, presence of the Meselson effect, and statistical tests for clonal population structure) were inconclusive. Our field surveys collected only females, and both of our sampled populations showed highly significant ($P < 0.0001$) deviations from HWE and a significant presence of clonal population structure was observed at several genetic distances. However, deviations from HWE and the presence of clonal population structure could be the result of other factors, such as genetic bottlenecks following founder events and are therefore not necessarily indicative of asexual reproduction. In addition, our efforts to follow the approach of Corral et al. (2009) to determine the presence of the Meselson effect were inconclusive as we cannot compare measures with our polymorphic loci in reference to a known sexual population, as done by Corral et al. (2009). Lastly, we found no evidence for the presence of LD within or among populations of *Z. davisae*. While the lack of LD among Cynipini asexual populations is similar to that found by Stone et al. (2008), in total, our results are inconclusive as to the sexual/asexual status of *Z. davisae*. However, we hope the developed microsatellites will be used in future studies that include populations of *Z. davisae* from throughout its geographic distribution and/or include sexually reproducing congeners to further explore the mode of reproduction for *Z. davisae*.

**Conclusions**

Here we provide a description of recently developed genomic resources that can be applied to the study of the biology and the control of *Z. davisae*, an important pest species of *Q. velutina* (black oaks) in the northeastern United States. These resources include over one billion base pairs of sequence data, over 900 assembled contigs, and 120 candidate microsatellite loci. We tested the utility of 36 of the microsatellite loci for *Z. davisae* population genetics research and found that half of them resulted in monomorphic genotype scores, while only five were polymorphic. Monomorphic loci are not normally useful for population genetic studies (see Ellegren, 2004; Selkoe & Toonen, 2006), however; here we provide the primers and thermocycler conditions for these loci as supplemental information because our sampling of *Z. davisae* comes from two geographically proximal locations that may both be recent introductions and it is likely that when populations from greater geographic distances are observed that some of these microsatellite loci may also be found to be polymorphic. In addition, while the mutation dynamics of microsatellite loci can make their amplification difficult among closely related species (Primmer et al., 1996; Scribner & Pearce, 2000; MacDonald et al., 2003; Selkoe & Toonen, 2006), numerous studies have successfully cross-amplified microsatellite loci, and thus some of the monomorphic loci presented here may be polymorphic in other species of gall wasps. Finally, while our genetic tests to determine the reproductive mode of *Z. davisae* were inconclusive, we hope that these microsatellite loci, the raw sequence reads, and the assembled contigs can be used for future studies of *Z. davisae* populations from across its distribution, in addition to studies of the population genetics of Cynipini species more broadly, and can contribute to the study of the evolution of gall formation and of reproductive modes in this fascinating group.

**Supplementary material**

The supplementary material for this article can be found at https://doi.org/10.1017/S0007485318000858

**Acknowledgements**

The authors thank H. Broadley for her valuable feedback and for kindly ordering supplies, J. Boettner for building a functional molecular lab on a shoestring budget and for collecting last-minute samples, E. Mooshain for field assistance and laboratory identification work, D. Newman, M. Labbé, B. Normark, and two anonymous reviewers for their constructive comments and suggestions on an earlier version of this manuscript. Funding was provided by the University of Massachusetts Amherst Department of Environmental Conservation (MJF), the Arborjet Corp, Woburn, MA and the Woodbourne Arboretum, Melville, NY (MJF and JSE), the U.S. Forest Service Northern Research Station to NPH. A portion of this research represents part of Honors Thesis project of CPC, as part of the University of Massachusetts Amherst Commonwealth College Honors Program.

**References**


