

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

J.C. Andersen^{1*} , C.P. Camp^{1†} , 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)

*Author for correspondence

Phone: 413-545-2665

Fax: 413-545-4358

E-mail: jcandersen@umass.edu

†Present address: Biomedical Engineering Graduate Interdisciplinary Program, University of Arizona, 1127 E. James E. Rogers Way, Tucson, AZ 85721, USA.

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 Massachusetts on Cape Cod and Martha's Vineyard (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 (Schaefer *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 aid 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 ($\text{ng } \mu\text{l}^{-1}$) 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'-TCCCAGTCACGACGT-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 manufacture'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 GoTaq 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 GoTaq (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_s), 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_{IS}) 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/GenoDiveManual.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 SRR5570816) as part of BioProject PRJNA386803.

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 $11x \pm 15x$. 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 average numbers of alleles ($n = 2$), as well as similar average heterozygosity ($H_o = 0.177$ for Nantucket and $H_o = 0.198$ for

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

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

Count, the number of candidate loci for each motif; average repeats, the average length of repeat units; Max repeats and Min repeats, the maximum and minimum lengths of repeat units, respectively.

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.

Locus name	Repeat type	Repeat motif and range	<i>n</i>	<i>Num</i>	<i>Eff_num</i>	<i>H_o</i>	<i>H_s</i>	<i>H_t</i>	<i>H'_t</i>	<i>G_{IS}</i>	HWE
0781	Tetra-	CAGA ^{2–10}	31	2	1.155	0.094	0.140	0.139	0.138	0.329	0.160
0887	Tri-	AAT ^{10–12}	32	3	1.215	0.000	0.190	0.200	0.211	1.000	0.000
1333	Tetra-	TGTC ^{13–17}	29	2	1.175	0.000	0.161	0.171	0.182	1.000	0.007
1643	Tri-	ATT ^{6–14}	31	6	1.414	0.216	0.306	0.303	0.300	0.295	0.068
2121	Di-	AG ^{16–23}	32	2	2.000	1.000	0.500	0.500	0.500	–1.000	1.000

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*.

Population	<i>n</i>	<i>Num</i>	<i>Eff_num</i>	<i>H_o</i>	<i>H_s</i>	<i>H_t</i>	<i>H'_t</i>	<i>G_{IS}</i>	HWE
Nantucket	13	1.857	1.288	0.177	0.181	0.181	–	0.023	0.000
Dennis	19	2.000	1.303	0.198	0.190	0.190	–	–0.041	0.000

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.

	Locus 781	Locus 887	Locus 1333	Locus 1643	Locus 2121
CV	3.106	na	na	2.799	0
Number of alleles	2	3	2	6	2

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).

of clarity as to the overall number of species and genera that make up the Cynipini – likely due to the vast number of species yet to be found and the taxonomic validity of existing species (see references in Melika & Buss, 2002). While several recent morphological and phylogenetic studies have clarified some of the higher level relationships among the tribes in the Cynipidae (e.g., Liljeblad & Ronquist, 1998; Rokas *et al.*, 2002; Rokas *et al.*, 2003; Ronquist *et al.*, 2015), relationships within the Cynipini remain in flux as evident by the numerous revisions and re-assignments of species and genera, and the frequent identification of novel species (e.g., Buffington & Morita 2009; Pujade-Villar *et al.*, 2010; Medianero, Nieves-Aldrey & Melika 2011; Ide, Wachi & Abe 2012; Pujade-Villar *et al.*, 2012; Melika *et al.*, 2013; Tang *et al.*, 2016).

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

I suggests that individuals of *Z. davisae* in the northeastern United States have multiple close matches (>97% similarity) with individuals from Florida (BOLD: BBHYH546-10), Nova Scotia (GenBank: KR804678.1, GenBank: KR782782.1), Ontario (BOLD: CNROF073-13, BOLD: CNTIF3344-15, BOLD: CNRGD596-15, GenBank: KR409819.1, GenBank: KR414957.1), and Tennessee (BOLD: GMGSM442-13). 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). In addition we hope to use these microsatellite markers to determine the relationship between *Z. davisae* and *Callirhytis ceropteroides* Bassett, a species previously reported to have caused black oak damage and mortality in Connecticut and New York State (Bassett, 1900; Pike *et al.*, 2006), as understanding the biology and natural enemies of *C. ceropteroides* may help inform the control of *Z. davisae* as well.

Mode of reproduction

In addition to causing damage to black oaks, *Z. davisae* has been described as asexual 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 (MJD), the Arborjet Corp, Woburn, MA and the Woodbourne Arboretum, Melville, NY (MJD 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

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.
- Andersen, J.C. & Mills, N.J. (2014) iMSAT: a novel approach to the development of microsatellite loci using barcoded Illumina libraries. *BMC Genomics* **15**, 1–9.
- Askew, R.R., Melika, G., Pujade-Villar, J., Schonrogge, K., Stone, G.N. & Nieves-Aldrey, J.L. (2013) Catalogue of parasitoids and inquiline in cynipid oak galls in the West Palaearctic. *Zootaxa* **3643**, 1–133.
- Atkinson, R.J., McVean, G.A.T. & Stone, G.N. (2002) Use of population genetic data to infer oviposition behaviour: species-specific patterns in four oak gallwasps (Hymenoptera: Cynipidae). *Proceedings of the Royal Society B-Biological Sciences* **269**, 383–390.
- Bassett, H.F.B. (1900) *Callirhytis ceropteroides*, Connecticut. *Transactions of the American Entomological Society* **26**, 334.
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2013) Genbank. *Nucleic Acids Research* **41**, D36–D42.
- Bhargava, A. & Fuentes, F.F. (2010) Mutational dynamics of microsatellites. *Molecular Biotechnology* **44**, 250–266.
- Birky Jr, C.W. (1996) Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. *Genetics* **144**, 427–437.
- Buffington, M.L. & Morita, S.I. (2009) Not all oak gall wasps gall oaks: the description of *Dryocosmus rileypokei*, a new, apostate species of Cynipini from California. *Proceedings of the Entomological Society of Washington* **111**, 244–253.
- Buffington, M.L., Melika, G., Davis, M. & Elkinton, J.S. (2016) The description of *Zapatella davisae*, new species, (Hymenoptera: Cynipidae) a pest gall wasp of black oak (*Quercus velutina*) in New England, USA. *Proceedings of the Entomological Society of Washington* **118**, 14–26.
- Cook, J.M., Rokas, A., Pagel, M. & Stone, G.N. (2002) Evolutionary shifts between host oak sections and host-plant organs in *Andricus* gallwasps. *Evolution* **56**, 1821–1830.

- Cooper, W.R. & Rieske, L.K. (2007) Community associates of an exotic gallmaker, *Dryocosmus kuriphilus* (Hymenoptera: Cynipidae), in Eastern North America. *Annals of the Entomological Society of America* **100**, 236–244.
- Corral, J.M., Piwczynski, M. & Sharbel, T.F. (2009) Allelic sequence divergence in the apomictic *Boechera holboellii* complex. pp 495–516 in Schön, I., Martens, K. & Dijk, P. (Eds) *Lost Sex*. Dordrecht, Springer.
- Davis, M.J. (2017) *Biology, molecular systematics, population dynamics and control of a stem gall wasp, Zaptella davisae* (Hymenoptera: Cynipidae). PhD. Dissertation, University of Massachusetts, Amherst.
- Davis, M.J. & Elkinton, J.S. (2018) Efficacy of two systemic insecticides with stem gall wasp, *Zaptella davisae* (Hymenoptera: Cynipidae) on Black Oak. *Journal of Economic Entomology* **111**, 986–988 (doi: <https://doi.org/10.1093/jee/tox358>).
- Drown, D.M. & Brown, J.M. (1998) Molecular phylogeny of North American oak-galling cynipini (Hymenoptera: Cynipidae) supports need for generic revision. *Biology of Gall-Inducing Arthropods* **199**, 241–246.
- Eliason, E.A. & Potter, D.A. (2000) Impact of whole-canopy and systemic insecticidal treatments on *Callirhytis cornigera* (Hymenoptera: Cynipidae) and associated parasitoids on pin oak. *Journal of Economic Entomology* **93**, 165–171.
- Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* **5**, 435–445.
- Faircloth, B.C. (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92–94.
- Hardy, G. H. (1908). Mendelian proportions in a mixed oopulation. *Science* **28**, 49–50.
- Ide, T., Wachi, N. & Abe, Y. (2012) Three new species and a new record of *Cycloneuroterus* (hymenoptera: Cynipidae: Cynipini) inducing galls on *Cyclobalanopsis* in Japan. *Annals of the Entomological Society of America* **105**, 539–549.
- Jurka, J. (2000) Repbase update: a database and an electronic journal of repetitive elements. *Trends in Genetics* **9**, 418–420.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P. & Drummond, A. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)* **28**, 1647–1649.
- Lewontin, R.C. & Kojima, K. (1960) The evolutionary dynamics of complex polymorphisms. *Evolution* **14**, 458–472.
- Liljeblad, J. & Ronquist, F. (1998) A phylogenetic analysis of higher-level gall wasp relationships (Hymenoptera: Cynipidae). *Systematic Entomology* **23**, 229–252.
- Luo, R.B., Liu, B.H., Xie, Y.L., Li, Z.Y., Huang, W.H., Yuan, J.Y., He, G.Z., Chen, Y.X., Pan, Q., Liu, Y.J., Tang, J.B., Wu, G.X., Zhang, H., Shi, Y.J., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C. L., Cheung, D.W., Yiu, S.M., Peng, S.L., Zhu, X.Q., Liu, G.M., Liao, X.K., Li, Y.R., Yang, H.M., Wang, J. & Lam, T.W. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18.
- MacDonald, C., Brookes, C.P., Edwards, K.J., Baker, A., Lockton, S. & Loxdale, H.D. (2003) Microsatellite isolation and characterization in the beneficial parasitoid wasp *Diaeretiella rapae* (M'Intosh) (Hymenoptera: Braconidae: Aphidinae). *Molecular Ecology Notes* **3**, 601–603.
- Mayer, C. (2010) Phobos v. 3.3.11. Available online at http://www.rub.de/spezzoo/cm/cm_phobos.htm.
- Medianero, E., Nieves-Aldrey, J.L. & Melika, G. (2011) Two new Neotropical species of oak gall wasps of the genus *Loxaulus* mayr (Hymenoptera: Cynipidae: Cynipini) from Panama. *Zootaxa* **2811**, 37–46.
- Meglec, E., Costedoat, C., Dubut, V., Gilles, A., Malausa, T., Pech, N. & Martin, J.F. (2010) QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics (Oxford, England)* **26**, 403–404.
- Meglec, E., Pech, N., Gilles, A., Dubut, V., Hingamp, P., Trilles, A., Grenier, R. & Martin, J.F. (2014) QDD version 3.1: a user-friendly computer program for microsatellite selection and primer design revisited: experimental validation of variables determining genotyping success rate. *Molecular Ecology Resources* **14**, 1302–1313.
- Meirmans, P.G. & Van Tienderen, P.H. (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792–794.
- Melika, G. & Abrahamson, W.G. (2007) Review of the Nearctic gallwasp species of the genus *Bassetia* ashmead, 1887, with description of new species (Hymenoptera: Cynipidae: Cynipini). *Acta Zoologica Academiae Scientiarum Hungaricae* **53**, 131–148.
- Melika, G. & Buss, E.A. (2002) Description of the sexual generation of *Callirhytis quercuscornigera* and a new inquiline (Hymenoptera: Cynipidae). *Florida Entomologist* **85**, 625–631.
- Melika, G., Tang, C.T., Sinclair, F., Yang, M.M., Lohse, K., Hearn, J., Nicholls, J.A. & Stone, G.N. (2013) A new genus of oak gallwasp, *Cyclocynips* melika, Tang & Sinclair (Hymenoptera: Cynipidae: Cynipini), with descriptions of two new species from Taiwan. *Zootaxa* **3630**, 534–548.
- Moriya, S., Inoue, K., Otake, A., Shiga, M. & Mabuchi, M. (1989) Decline of the chestnut gall wasp population, *Dryocosmus kuriphilus* Yasumatsu (Hymenoptera, Cynipidae) after the establishment of *Torymus-sinensis* Kamijo (Hymenoptera, Torymidae). *Applied Entomology and Zoology* **24**, 231–233.
- Neiman, M. & Lively, C.M. (2005) Male New Zealand mud snails (*Potamopyrgus antipodarum*) persist in copulating with asexual and parasitically castrated females. *American Midland Naturalist* **154**, 88–96.
- Pike, C.C., Robison, D.J. & Abrahamson, L.P. (2006) Cynipid gall wasps in declining black oak in New York: relationships with prior tree history and crown dieback. pp 123–132 in Ozaki, K., Yukawa, J., Ohgushi, T., & Price, P.W. (Eds) *Galling Arthropods and Their Associates*. Tokyo, Springer.
- Primmer, C.R., Moller, A.P. & Ellegren, H. (1996) A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology* **5**, 365–378.
- Pujade-Villar, J., Romero-Rangel, S., Chagoyan-Garcia, C., Equihua-Martinez, A., Estrada-Venegas, E.G. & Melika, G. (2010) A new genus of oak gallwasps, *Kinseyella* Pujade-Villar & Melika, with a description of a new species from Mexico (Hymenoptera: Cynipidae: Cynipini). *Zootaxa* **2335**, 16–28.
- Pujade-Villar, J., Hanson, P., Medina, C.A., Torres, M. & Melika, G. (2012) A new genus of oak gallwasps, *Zaptella* Pujade-Villar & Melika, gen. n., with a description of two new species from the Neotropics (Hymenoptera, Cynipidae, Cynipini). *Zookeys* **210**, 75–104.
- Pujade-Villar, J., Cibrian-Tovar, D., Barrera-Ruiz, U.M. & Melika, G. (2014) *Andricus breviramuli* Pujade-Villar n. sp. (Hymenoptera: Cynipidae: Cynipini): new pest of oaks in Mexico. *Southwestern Entomologist* **39**, 97–106.

- Raymond, M. & Rousset, F. (1995) GENEPOP (VERSION-1.2) – Population-genetics software for exact tests and ecumenicism. *Journal of Heredity* **86**, 248–249.
- Rokas, A., Nylander, J.A.A., Ronquist, F. & Stone, G.N. (2002) A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. *Molecular Phylogenetics and Evolution* **22**, 206–219.
- Rokas, A., Melika, G., Abe, Y., Nieves-Aldrey, J.L., Cook, J.M. & Stone, G.N. (2003) Lifecycle closure, lineage sorting, and hybridization revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. *Molecular Phylogenetics and Evolution* **26**, 36–45.
- Ronquist, F. & Liljeblad, J. (2001) Evolution of the gall wasp-host plant association. *Evolution* **55**, 2503–2522.
- Ronquist, F., Nieves-Aldrey, J.L., Buffington, M.L., Liu, Z.W., Liljeblad, J. & Nylander, J.A.A. (2015) Phylogeny, evolution and classification of gall wasps: the plot thickens. *PLoS One* **10**, e0123301.
- Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–106.
- Schaefer, I., Domes, K., Heethoff, M., Schneider, K., Schön, I., Norton, R. A., Scheu, S. & Maraun, M. (2006) No evidence for the 'Meselson effect' in parthenogenetic oribatid mites (Oribatida, Acari). *Journal of Evolutionary Biology* **19**, 184–193.
- Schuelke, M. (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18**, 233–234.
- Schurko, A.M., Neiman, M. & Logsdon, J.M. (2009) Signs of sex: what we know and how we know it. *Trends in Ecology & Evolution* **24**, 208–217.
- Scribner, K. & Pearce, J. (2000) Microsatellites: evolutionary and methodological background and empirical applications at individual, population, and phylogenetic levels. pp. 235–271 in Baker, A. (Ed.) *Molecular Methods in Ecology*. London, England, Blackwell Science Limited.
- Selkoe, K.A. & Toonen, R.J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615–629.
- Slatkin, M. (2008) Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* **9**, 477–485.
- Stone, G.N., Schonrogge, K., Atkinson, R.J., Bellido, D. & Pujade-Villar, J. (2002) The population biology of oak gall wasps (Hymenoptera: Cynipidae). *Annual Review of Entomology* **47**, 633–668.
- Stone, G.N., Atkinson, R.J., Rokas, A., Nieves-Aldrey, J.L., Melika, G., Acs, Z., Csoka, G., Hayward, A., Bailey, R., Buckee, C. & McVean, G.A.T. (2008) Evidence for widespread cryptic sexual generations in apparently purely asexual *Andricus* gallwasps. *Molecular Ecology* **17**, 652–665.
- Tang, C.T., Sinclair, F., Hearn, J., Yang, M.M., Stone, G.N., Nicholls, J.A., Schweger, S. & Melika, G. (2016) Eight new species of *Cycloneuroterus melika* & Tang gallwasps from Taiwan and mainland China (Hymenoptera: Cynipidae: Cynipini). *Zootaxa* **4088**, 451–488.
- Washburn, J.O. & Cornell, H.V. (1981) Parasitoids, patches, and phenology – their possible role in the local extinction of a Cynipid gall wasp population. *Ecology* **62**, 1597–1607.