Identifying Diagnostic Genetic Markers for a Cryptic Invasive Agricultural Pest: A Test Case Using the Apple Maggot Fly (Diptera: Tephritidae)

Authors: Doellman, Meredith M., Hood, Glen R., Gersfeld, Jacob, Driscoe, Amanda, Xu, Charles C. Y., et al.

Source: Annals of the Entomological Society of America, 113(4) : 246-256

Published By: Entomological Society of America

URL: https://doi.org/10.1093/aesa/saz069

Annals of the Entomological Society of America, 113(4), 2020, 246–256
doi: 10.1093/aesa/saz069
Advance Access Publication Date: 16 March 2020
Special Collection: Advanced Genetic Analysis of Invasive Arthropods Special Collection

Identifying Diagnostic Genetic Markers for a Cryptic Invasive Agricultural Pest: A Test Case Using the Apple Maggot Fly (Diptera: Tephritidae)

Meredith M. Doellman,[1,5,*,] Glen R. Hood,[2,*] Jacob Gersfeld,[1] Amanda Driscoe,[1] Charles C. Y. Xu,[3] Ryan N. Sheehy,[1] Noah Holmes,[1] Wee L. Yee,[4] and Jeffrey L. Feder[1]

[1]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, [2]Department of Biological Sciences, Wayne State University, Detroit, MI 48202, [3]Redpath Museum & Department of Biology, McGill University, Montreal, QC H3A 0C4, Canada, [4]USDAARS, Temperate Tree Fruit & Vegetable Research Unit, Yakima Agricultural Research Laboratory, Wapato, WA 98951 and [5]Corresponding author, e-mail: mdoellma@nd.edu

*Denotes equal author contribution.

Subject Editor: Jeffrey Morisette

Received 13 May 2019; Editorial decision 22 November 2019

Abstract

Insect pests destroy ~15% of all U.S. crops, resulting in losses of $15 billion annually. Thus, developing cheap, quick, and reliable methods for detecting harmful species is critical to curtail insect damage and lessen economic impact. The apple maggot fly, *Rhagoletis pomonella*, is a major invasive pest threatening the multibillion-dollar apple industry in the Pacific Northwest United States. The fly is also sympatric with a benign but morphologically similar and genetically closely related species, *R. zephyria*, which attacks noncommercial snowberry. Unambiguous species identification is essential due to a zero-infestation policy of apple maggot for fruit export. Mistaking *R. zephyria* for *R. pomonella* triggers unnecessary and costly quarantines, diverting valuable control resources. Here we develop and apply a relatively simple and cost-effective diagnostic approach using Illumina sequencing of double-digest restriction site-associated DNA markers. We identified five informative single-nucleotide polymorphisms (SNPs) and designed a diagnostic test based on agarose gel electrophoresis of restriction enzyme-digested polymerase chain reaction amplification products (RFLPs) to distinguish fly species. We demonstrated the utility of this approach for immediate, 1-d species identification by scoring apple- and snowberry-infesting flies from known hosts, reared from fruit collected at 11 sites throughout Washington. However, if immediate diagnosis is not required, or hundreds to thousands of specimens must be assessed, then a direct Illumina-based sequencing strategy, similar to that used here for diagnostic SNP identification, can be powerful and cost-effective. The genomic strategy we present is effective for *R. pomonella* and also transferable to many cryptic pests.

Key words: apple maggot, ddRADseq, RFLP, species diagnostic

It is becoming increasingly apparent that many insects are composed of difficult to distinguish cryptic host races and species specialized on different plants, some attacking economically important crops and others causing no agricultural harm (Walter 2005, Shu-sheng et al. 2012, Hendrichs et al. 2015). Moreover, foreign invasive pests are often difficult to detect and identify at points of entry (Carruthers 2003, Stouthamer et al. 2017). These problems call for the development of new strategies to rapidly and accurately identify populations and specimens of ecological and agricultural concern in a cost-effective manner to enact immediate control measures to prevent their detrimental establishment and spread.

The apple maggot fly, *Rhagoletis pomonella* (Walsh), is a recently introduced and invasive economic pest of domesticated apples in the Pacific Northwest (PNW) region of the United States (AliNiazee and Penrose 1981, AliNiazee and Westcott 1986, Brunner 1987, Tracewski et al. 1987, Dowell 1988, Hood et al. 2013, Sim et al. 2017). These flies are an ideal test case for DNA diagnostics because they are a textbook example of a group of insect taxa composed of numerous cryptic host races and sibling species specialized to feed on the fruit of different host plants, some of economic importance and others of no commercial value (Bush 1966, Feder et al. 1988, McPheron et al. 1988, Berlocher et al. 1993, Berlocher and Feder 2002). A problem faced by apple growers in the PNW is distinguishing the quarantined

and introduced *R. pomonella* from its harmless and native sibling species, *R. zephyria* Snow (Diptera: Tephritidae), that infests the fruit of the nonagriculturally important snowberry shrub (*Symphoricarpos* spp.) (Yee et al. 2009, 2011). In the PNW, snowberry-attacking flies are common and extensively co-occur with *R. pomonella*. When captured on field monitoring traps as adults, it is extremely difficult to visually distinguish between *R. zephyria* and *R. pomonella* and larvae of the two species are morphologically identical (Yee et al. 2009, 2011). A rapid and reliable genetic test is therefore required to unambiguously identify *R. pomonella* and determine where to target resources to control the fly.

The conventional history of *R. pomonella*'s presence in the PNW posits that the fly was introduced to the region via larval-infested apples about 40 years ago from the eastern United States (AliNiazee and Penrose 1981, AliNiazee and Westcott 1986, Brunner 1987, Tracewski et al. 1987, Dowell 1988, Hood et al. 2013, Sim et al. 2017). In the east, *R. pomonella* is a model for host shifts to economically important fruit (Bush 1966, Feder et al. 1988, McPheron et al. 1988). In the mid-1800s, *R. pomonella* shifted from its native, ancestral host downy hawthorn, *Crataegus mollis* Scheele (Rosales: Rosaceae), to establish populations on domesticated apple, *Malus pumila* Mil. (Rosales: Rosaceae), becoming a major pest of the crop as a result. It has been hypothesized that the fly was first introduced to the PNW in Portland, OR, where the first report of apple infestation was made in 1979 (AliNiazee and Penrose 1981). The introduction hypothesis is complicated, however, by the presence of *R. pomonella* infesting native black hawthorn, *C. douglasii* Lindley (Rosales: Rosaceae), and introduced ornamental hawthorn, *C. monogyna* Jacquin (Rosales: Rosaceae), in the PNW (Yee 2008, Yee and Goughnour 2008, Yee et al. 2012, Hood et al. 2013). It is therefore possible that *R. pomonella* is native on black hawthorn in the PNW and shifted onto apple after the plant was introduced by settlers in the early 19th century. However, geographic surveys of hawthorns throughout the PNW, and microsatellite genotyping of flies support the recent introduction hypothesis for *R. pomonella*'s presence in apple fruit (Yee and Goughnour 2008, Yee et al. 2012, Hood et al. 2013, Sim et al. 2017). Regardless, the presence of *R. pomonella* infesting black hawthorns and ornamental hawthorns requires that these two host plants also be considered in *R. pomonella* pest control strategies.

A genetic diagnostic test to distinguish *R. zephyria* from *R. pomonella* is now imperative as the apple-infesting fly has spread from the hypothesized area of initial introduction near Portland, OR, north into Washington on the western side of the Cascade Mountains, and into the Columbia River Gorge and other passages. Currently, the fly is encroaching on the commercial apple-growing region of central Washington, where it threatens a $2.25 billion-a-year industry (Fig. 1; Yee and Goughnour 2008, Yee et al. 2012, Mertz et al. 2013). In central Washington, there is a zero-infestation policy for apple export to foreign markets and for domestic consumption (WSDA 2001, Yee et al. 2012). If apple maggots are trapped within one-half-mile of an orchard, the surrounding area is considered 'threatened' and requires field inspection. If larvae are found in the vicinity, the surrounding area is placed under quarantine, affecting growers' access to export markets as apples cannot be shipped fresh but instead must undergo a prolonged cold treatment. To complicate the issue, positive visual identification of adult *R. pomonella* trapped in or near orchards is difficult because of the abundance of morphologically cryptic *R. zephyria* in central Washington (Yee et al. 2009, 2011). False positives are costly, resulting in unwarranted inspection and quarantine that misdirect control efforts from critical host spots, resulting in spraying and plant removal from non-apple maggot

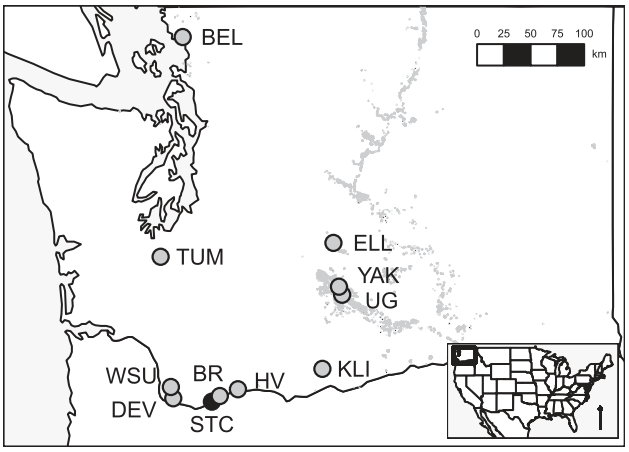

Fig. 1. Map of collecting sites in Washington State, United States. The black point denotes St. Cloud, Skamania County, WA, where flies included in the ddRADseq study were collected. Ten additional sites indicated in gray were combined with St. Cloud in the RFLP survey. The apple-growing region of central Washington is designated with light gray shading (Agricultural Land Use Data; WSDA 2017). See Supp Table 1 (online only) for a description of sites and additional details concerning sample sizes and collection dates.

infested areas. Computer simulations indicate that implementing more effective control practices that slow the fly's spread by just 10% can save 8 million dollars a year. Without improvement, however, all apple-producing areas in Washington are predicted to be infested in less than 30 yr (Zhao et al. 2007).

A recent episode demonstrating the need for a diagnostic test and its significance for agriculture occurred during the 2011 apple maggot survey. Seven unidentified fly larvae were discovered infesting the fruit of a Chinese crabapple tree, *Malus spectabilis* (Aiton) Borkhausen (Rosales: Rosaceae), in a homeowner's yard in Kennewick, WA near an active orchard (Yee and Klaus 2013). If these larvae were identified as *R. pomonella*, then a quarantine on apples would have been imposed on the surrounding counties of Benton, Franklin, and Walla Walla. It was therefore imperative to rapidly determine whether the larvae were *R. pomonella* as opposed to another species of *Rhagoletis*. To do so, we genetically scored five of the flies and found that the unknown mystery maggots were western cherry fruit flies, *R. indifferens* Curran (Diptera: Tephritidae), and not *R. pomonella* (Saint-Jean et al. 2013). The *R. indifferens* identification avoided quarantine of the three counties, saving the Washington State Department of Agriculture (WSDA), United States Department of Agriculture (USDA), and local pest control agencies an estimated one-half million dollars in administrative, inspection, and control costs.

The identification of the crabapple larvae was possible because previously developed microsatellites that amplify in both *R. pomonella* and *R. indifferens* have alleles unique to each species (Maxwell et al. 2009). Distinguishing *R. pomonella* and *R. zephyria* is more challenging, however, as populations of these two species share many alleles in common (Berlocher et al. 1993, Feder et al. 1999, Green et al. 2013). In previous work, however, one cDNA (P2956) and one allozyme (*Had*) out of 11 markers scored (Green et al. 2013) showed substantial allele frequency differences between *R. pomonella* and *R. zephyria* at several sites outside the central apple growing area of Washington. In addition, 13 of 19 microsatellites surveyed at nine 'sympatric' field sites where *R. pomonella* and *R. zephyria* co-occur across Washington revealed pronounced and statistically significant allele frequency differences of >0.30 between the species (Arcella et al. 2015). Moreover, most of the microsatellites analyzed contained private alleles present in either snowberry- or apple-infesting flies at

frequencies ranging from only 0.03 to 0.15. Thus, detectable genetic differences exist between *R. pomonella* and *R. zephyria*. However, a more comprehensive DNA sequencing approach, for example, generating genome-wide single-nucleotide polymorphisms (SNPs), is needed to systematically identify whether loci displaying potentially greater differentiation than P2956 and *Had* exist that can be combined with methods to develop a cost-effective, efficient, rapid, and accurate diagnostic genetic test.

Here, we apply next-generation genomic DNA sequencing to develop a rapid and genetic diagnostic tool to distinguish *R. pomonella* from *R. zephyria* as a means to eliminate the potential for false-positive identification of apple maggot flies. The strategy involves Illumina sequencing of double-digest restriction site-associated DNA markers (ddRADseq) of flies reared from infested host fruit collected at a site where *R. pomonella* and *R. zephyria* co-occur at St. Cloud recreation area near Skamania, WA (Figs. 1 and 2). Double-digest RADseq was used to identify SNPs displaying exceptional allele frequency differences between apple- and snowberry-origin flies. We then focused on five of the most informative SNPs identified in the population survey to develop a species-specific diagnostic test via agarose gel electrophoresis resolution of restriction enzyme-digested DNA fragment length polymorphisms (RFLPs) of polymerase chain reaction (PCR) amplification products. Finally, we demonstrated the application and feasibility of the genetic test for species diagnostic screening by scoring an additional 526 flies from 11 sites throughout Washington. This tool, while still in the developmental stages awaiting proper validation, can potentially soon be implemented in the WSDA annual apple maggot fly survey. More generally, our work highlights how population genomic tools can be developed and used to diagnose and distinguish closely related species of concern.

Methods

Sample Collection and ddRAD Sequencing

Larval-infested host fruit were collected from apple, black hawthorn, and ornamental hawthorn trees, and snowberry bushes, at St. Cloud, Skamania County, WA in late summer and early fall of 2011 (Fig. 1). Maggots were reared directly out of the fruit, and a subset was allowed to overwinter and then develop to adulthood, following standard *Rhagoletis* husbandry methods (Neilson and McAllan 1965; Feder et al. 1989, 1990, 1993). Both larvae and adults (which were sexed upon emergence) were preserved at −80°C for later genetic analysis. Roughly equal proportions of adults and larvae were used for DNA sequencing to generate total sample sizes for ddRADseq of 48 *R. zephyria* and 674 *R. pomonella* specimens (n = 342 apple-; n = 188 black hawthorn-; and n = 144 ornamental hawthorn-infesting flies). Whole-body DNA was extracted separately for each fly according to the Qiagen Puregene protocol (Qiagen Inc., Valencia, CA).

Double-digest Restriction Amplified DNA libraries were prepared for *Rhagoletis* as described in Egan et al. (2015) and Ragland et al. (2017). Briefly, DNA samples isolated from each fly were digested with the restriction enzymes EcoRI and MseI. Illumina sequencing adapters and 8–10 bp unique barcodes were ligated to the sticky ends of restriction fragments and the ligation products for each sample amplified separately in its own PCR reaction. Amplification products were then pooled and purified with a 0.8× volume of Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN). Size selection for fragments in the range of 300–500 bp was next performed using a BluePippin (Sage Science, Beverly, MA). The resulting libraries were sequenced (100 bp paired-end reads) in two lanes on an Illumina HiSeq 2000 platform.

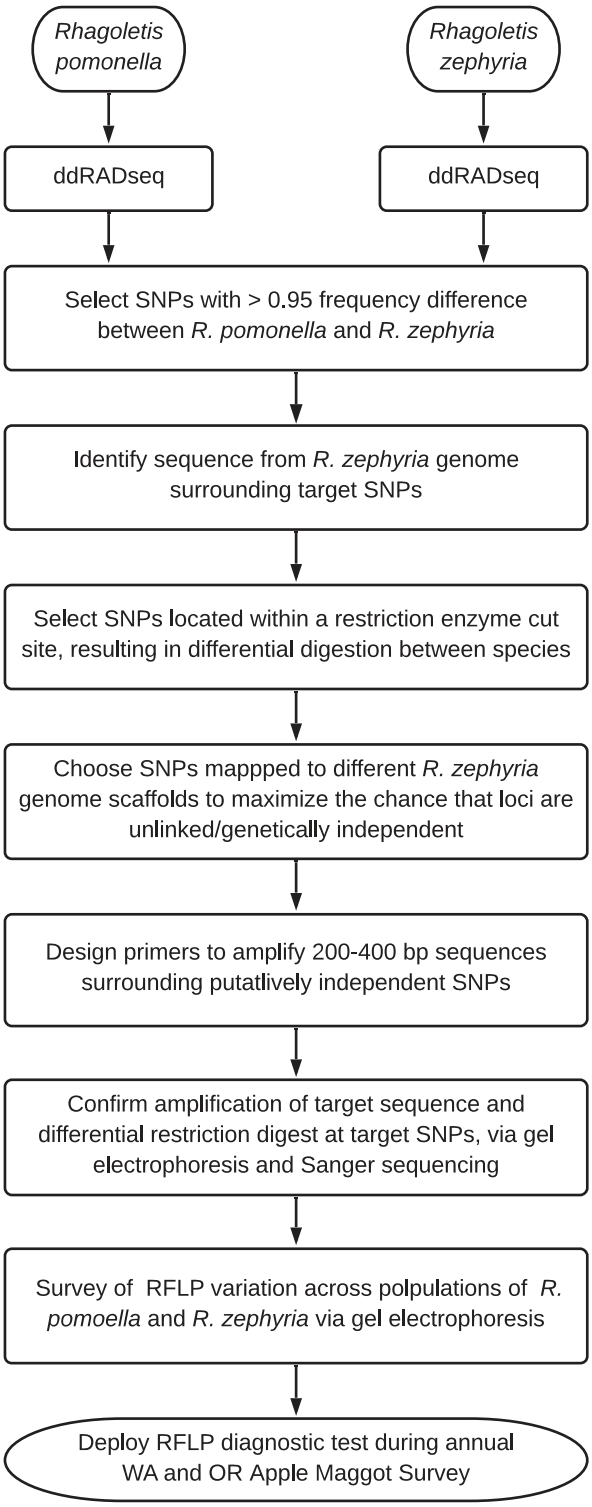

Fig. 2. A schematic of the strategy deployed for development of species diagnostic markers for discriminating *R. pomonella* and *R. zephyria* using next-generation sequencing data. This approach is amendable for the detection of any cryptic species of concern from a genetically closely related congener.

Bioinformatic Analysis of ddRADseq Data

Raw sequences were first edited for adapter contamination with Trimmomatic v0.32 (Bolger et al. 2014), after which barcodes were removed, individuals demultiplexed with a custom python script (Assour 2014, 2015), and the quality of reads assessed

with FASTQC v0.11.2 (Andrews 2010). Sequences were then aligned to the *R. zephyria* draft genome (NCBI accession number: GCA_001687245.1) using the *bwa mem* algorithm, removing reads having a mapping quality score below 30. Single-nucleotide polymorphisms were called using the GATK Unified Genotyper v3.3 (McKenna et al. 2010). We then used VCFtools v0.1.14 (Danecek et al. 2011) to remove indels and filter the data set for biallelic SNPs with a minimum quality score of 30. Individual SNPs were retained for analysis if 1) they were present in at least 50% of the flies sequenced; 2) the rarer allele had a minimum estimated frequency of ≥0.05 across the combined *R. zephyria* and *R. pomonella* data set; 3) allele counts did not significantly deviate at $P \leq 0.05$ from the expected binomial distribution; 4) the genotype distribution did not deviate significantly at $P \leq 0.01$ from Hardy–Weinberg equilibrium within each of the four host-associated populations; and 5) the mean depth of coverage (number of reads) per individual was ≤15. This resulted in a total of 26,951 biallelic SNPs retained for analysis in adjusted sample sizes of $N = 45$ for *R. zephyria*, and $N = 275$ for apple-, $N = 81$ for black hawthorn-, and $N = 115$ for ornamental hawthorn-infesting populations of *R. pomonella* at St. Cloud. Probabilities of single locus genotypes and allele frequencies for the 26,951 SNPs were calculated following McKenna et al. (2010). Tests for population differentiation and genotypic clustering of *R. pomonella* and *R. zephyria* were conducted using fastSTRUCTURE (Raj et al. 2014) and by performing principal component analysis (PCA) using *adegenet* in R v3.4.3 (Jombart 2008, Jombart and Ahmed 2011, R Core Team 2018).

Designing Diagnostic Makers

Single-nucleotide polymorphisms were identified as candidates for PCR primer design and possible inclusion in the RFLP species diagnostic test if they showed an estimated allele frequency difference of ≥0.95 between *R. zephyria* and the apple-infesting population of *R. pomonella* at St. Cloud. For each of these candidate markers, the 5 bp sequence immediately flanking either side of the SNP was examined to determine whether it represented a site that could result in diagnostic digestion by a restriction enzyme between *R. zephyria* and *R. pomonella* (i.e., a site where an enzyme will cut the target sequence in one species, but not the other). The mean read length of unique fly DNA resulting from Illumina HiSeq 2000 sequencing was only 85 bp. As a result, individual contigs containing variable restriction enzyme cut sites were not of sufficient length to design primers that would generate RFLPs of appropriately different sizes to distinguish by standard and low-priced 1% agarose gel electrophoresis. Consequently, contigs containing variable cut sites were aligned to larger scaffolds of the *R. zephyria* draft genome, allowing primers to be designed that approximately centered the restriction site in an amplicon having a total length of ~400 bp. Flanking sequences that contained additional cut sites for the restriction enzyme of interest were eliminated from further marker development and testing. The criteria for development was that each of the primer pairs 1) had to be ~20 bp in length; 2) possess 40–60% GC content; 3) have similar annealing temperatures; and 4) represent unique sequences mapping to different scaffolds in the *R. zephyria* genome to maximize the number of physically unlinked markers included in the diagnostic test (see below). In some cases, primer sites had to be moved closer to the target SNP to accommodate the above criteria, reducing the total amplicon length below 400 bp. In one case (zeph10), the best position for the reverse primer incorporated the target SNP at the 3′ end, which results in the presence (*R. zephyria*) or absence (*R. pomonella*) of a diagnostic PCR band (see below).

Linkage Relationships and Genetic Independence of Diagnostic Markers

An important criterion for why SNPs were selected for primer development was because they mapped to different major scaffolds in the *R. zephyria* genome, suggesting that these loci are not closely physically linked to one another on chromosomes. To explicitly test for genetic independence and estimate the number of independent genomic regions of high divergence, pairwise correlation coefficients (r) were calculated among genotypes for all SNPs showing exceptionally high frequency differences between *R. pomonella* and *R. zephyria* as determined by ddRADseq, across the apple-, black hawthorn-, and ornamental hawthorn-infesting populations of *R. pomonella* at St. Cloud. In these analyses, genotypes were encoded as 1 = homozygote for common allele, 2 = heterozygote, 3 = homozygote for alternate allele and correlation analyses were performed in R v3.4.3 (R Core Team 2017).

Population Survey Evaluating Diagnostic Markers

To demonstrate the utility of the developed RFLP markers, we conducted a population survey by scoring 211 *R. pomonella* and 315 *R. zephyria* sampled from 11 sites across western and central Washington. For this analysis, flies obtained by rearing larval *R. pomonella* and *R. zephyria* from infested host fruit collected from 2009 to 2016 (Fig. 1; Supp Table 1 [online only]) were allowed to overwinter and then develop to adulthood. Adult flies were sexed and then preserved at −80°C for later genetic analysis. Three of the 11 sites in the survey (Ellensburg, Yakima, and Union Gap) are also located in central Washington near the apple-growing region of the state, where *R. pomonella* has not yet invaded; therefore, we sampled only *R. zephyria* from these sites (Fig. 1; Supp Table 1 [online only]). Additionally, only *R. zephyria* from Klickitat and *R. pomonella* from Tumwater were available, although both fly species have previously been sampled near these sites (Arcella et al. 2015). The remaining six sites, where both *R. pomonella* and *R. zephyria* are common, were sampled for both species, including an additional 23 apple- and 37 snowberry-infesting flies collected from St. Cloud in 2016, which were not genotyped in the ddRADseq analysis.

We extracted DNA from individual whole-body flies for PCR amplification using the primer pairs developed for diagnostic RFLP testing. Reaction conditions for PCR amplification included 1 µl of template DNA, 1.5 µl of 10× PCR buffer, 1.2 µl of 25 mM $MgCl_2$, 0.3 µl each of 10 µM forward and reverse primers, 0.6 µl of 25 µM of each dNTPs, 0.08 µl of 5 U/µl Taq polymerase, and 10 of µl water. After an initial denaturation step at 94°C for 2 min, DNA was amplified for a total of 35 cycles (one cycle = 94°C for 30 s, reannealing for 45 s at the temperature specified for each primer pair in Table 1, and extension at 72°C for 1 min, with a final extension step for 15 min). Restriction digestion reactions were performed on 10 µl of PCR products from zeph2, zeph5, zeph6, and zeph7 using the enzymes NcoI, HpyCH4V, AluI, or PstI, respectively, at 37°C, per the manufacturer's instructions (Table 2) (New England Biolabs, Ipswich, MA). The RFLP fragments were electrophoretically resolved on 1% agarose gels and individual genotypes for the diagnostic SNPs were assigned based on the resulting fragment sizes (Table 2).

Sanger Sequencing

Sanger sequencing was performed on a small subset of surveyed individuals for each primer pair chosen for the diagnostic test to verify that the PCR amplicons represented the intended target sequences (GenBank accession numbers MN812839–MN812863).

Table 1. Primer sequences and PCR annealing temperatures for the five species diagnostic markers distinguishing *R. pomonella* and *R. zephyria*. Note that for zeph10, the bolded 'G' nucleotide at the 3′ end of the reverse primer. This polymorphism is complementary to the 'C' allele in *R. zephyria*, but not the 'A' allele in *R. pomonella* (Table 2); therefore, the zeph10 primer pair produces a PCR amplicon in *R. zephyria* but not *R. pomonella*.

Locus	Forward primer (5′-3′)	Reverse primer (5′-3′)	Annealing temperature
zeph2	TGTAAGTCGGTCGGTCACTC	CGCTTCGTAACTGATTATGCG	63.5°C
zeph5	AATAGTTCAGCTGTGTACGA	ATGCGTCCAATAACCGAAAA	59.5°C
zeph6	TTTAACTTGGGCACGAAAAA	GACTCACTAAAATGTGGAAT	57.5°C
zeph7	AATTAGAGAACAGCGTGCGC	CCTGTATCTGTTCAGTCAGC	60.8°C
zeph10	ATGCGTAATTGCGCGCGATG	AACTGAAATGCGCAGTATTC**G**	68.4°C

Table 2. Amplicon lengths, SNP alleles, restriction enzymes generating cut site differences, digestion temperatures, and lengths of restriction fragments for each of five diagnostic markers distinguishing *R. zephyria* (*Rz*) and *R. pomonella* (*Rp*). Note that for zeph10, *R. zephyria* and *R. pomonella* are distinguished by presence (Rz) or absence (Rp) of the 221 bp PCR amplicon; therefore, zeph10 is diagnostic without a restriction enzyme digest step.

Locus	Amplicon length (*Rz*)	Amplicon length (*Rp*)	*Rz* allele	*Rp* allele	Restriction enzyme	Recognition site	*Rz* bands	*Rp* bands
zeph2	401 bp	461 bp	C	G	NcoI	CCATGG	200 bp; 201 bp	461 bp
zeph5	263 bp	263 bp	T	C	HpyCH4V	TGCA	202 bp; 61 bp	263 bp
zeph6	401 bp	394–402 bp	A	T	AluI	AGCT	202 bp; 199 bp	394–402 bp
zeph7	156 bp	156 bp	T	C	PstI	CTGCAG	156 bp	94 bp; 62 bp
zeph10	221 bp	—	C	A	—	—	221 bp	—

Nondigested PCR amplicons were sequenced for three *R. pomonella* and three *R. zephyria* adults, one each from the Bellingham, Devine, and St. Cloud, on a 96-capillary 3730xl DNA Analyzer (Applied Biosystems, Waltham, MA). Chromatograms were visually inspected and sequences were trimmed for quality and aligned in Sequencher 5.1 (GeneCorp, Ann Arbor, MI) and *blast* (Altschul et al. 1990) was used to confirm alignment to the target regions of the *R. zephyria* draft genome. We did not obtain sequences for *R. pomonella* for primer pair zeph10, as the 3′ mismatch in the reverse primer prevented PCR amplification of this marker.

MiSeq Study Resolving Ambiguous Individuals

To resolve the identity of putative hybrid individuals identified in the RFLP survey and demonstrate the utility and efficiency of our proposed ddRADseq approach, 10 additional individuals were sequenced on an Illumina MiSeq. Ambiguous samples included a putative F1 hybrid collected from black hawthorn at Home Valley, as well as individuals collected from apple at Washington State University (WSU) (n = 1) and snowberry at Klickitat (n = 1) and Yakima (n = 1), which were heterozygous for one to two RFLP markers (Supp Table 3 [online only]). We chose three unambiguously identified *R. zephyria*, from St. Cloud (n = 1) and WSU (n = 2), and three unambiguously identified *R. pomonella*, from St. Cloud, WSU, and Home Valley (n = 1 each), as representative parental samples. Sequencing libraries were prepared according to the ddRADseq protocol outlined above and sequenced on an Illumina MiSeq platform. The initial bioinformatic analyses through SNP identification with GATK were also conducted as described above (McKenna et al. 2010). From raw MiSeq SNP set, we selected all 26,951 SNPs that were retained for analysis in the ddRADseq data set above to maximize overlap between the two studies. In addition, only SNPs that were biallelic and present in at least 50% of the MiSeq samples were retained, resulting in 18,727 SNPs.

Two analyses were conducted to evaluate the ancestry of these 10 samples. Both approaches were based on a set of 35 SNPs with estimated allele frequency differences ≥ 0.95 between *R. pomonella* and *R. zephyria* in the ddRADseq study, which were also present in the MiSeq data set. First, Introgress was used to visualize the genotypic composition at these loci (homozygous or heterozygous) of potential hybrids compared to parental populations (Gompert and Buerkle 2010). Second, we used NewHybrids to calculate the posterior probability of individual assignment to each of six categories: *R. pomonella*, *R. zephyria*, F1, F2, F1 backcrossed to *R. pomonella*, and F1 backcrossed to *R. zephyria* (Anderson and Thompson 2002). We ran the analysis for 1,000,000 sampling sweeps, following 100,000 burn-in sweeps and report the average posterior probabilities from 10 independent replicate runs. Although results are presented for only the 10 MiSeq samples, the NewHybrids analysis included all individuals from the ddRADseq data set for robust estimation of allele frequencies.

Results

ddRADseq Divergence Between Apple- and Snowberry-Infesting Flies

Double-digest RADseq resulted in the identification of a total of 26,951 SNPs segregating in the adjusted samples of 45 *R. zephyria* and 471 *R. pomonella* from St. Cloud. Analyses via PCA and fastSTRUCTURE based on these SNPs distinguished *R. zephyria* and *R. pomonella* as distinct genotypic clusters. Additionally, no fly was identified as being a F1 hybrid between *R. zephyria* and *R. pomonella* at St. Cloud (Fig. 3A and B).

Despite forming distinct genotypic clusters, none of the 26,951 SNPs genotyped was diagnostically fixed for alternate alleles between *R. zephyria* and *R. pomonella* (Fig. 3C). Several SNPs showed substantial allele frequency differences between the two species, however. For example, there was a noticeable minor peak of allele frequency divergence centered at a difference of ~0.85 (Fig. 3C). Most importantly, for diagnostic testing, 40 SNPs displayed exceptionally high frequency differences between the species of ≥0.95. For 29 of the SNPs showing exceptional divergence, one allele was extremely common in *R. pomonella* (frequency ≥ 0.95), but was absent from *R. zephyria* (i.e., the frequency of the alternate SNP allele was 1.0 in *R. zephyria*, and ≤0.05 in *R. pomonella*). In the remaining 11 cases, the frequency of the common allele in one taxon was reciprocally rare in the other (≤0.03). Finally, in no single case was an allele

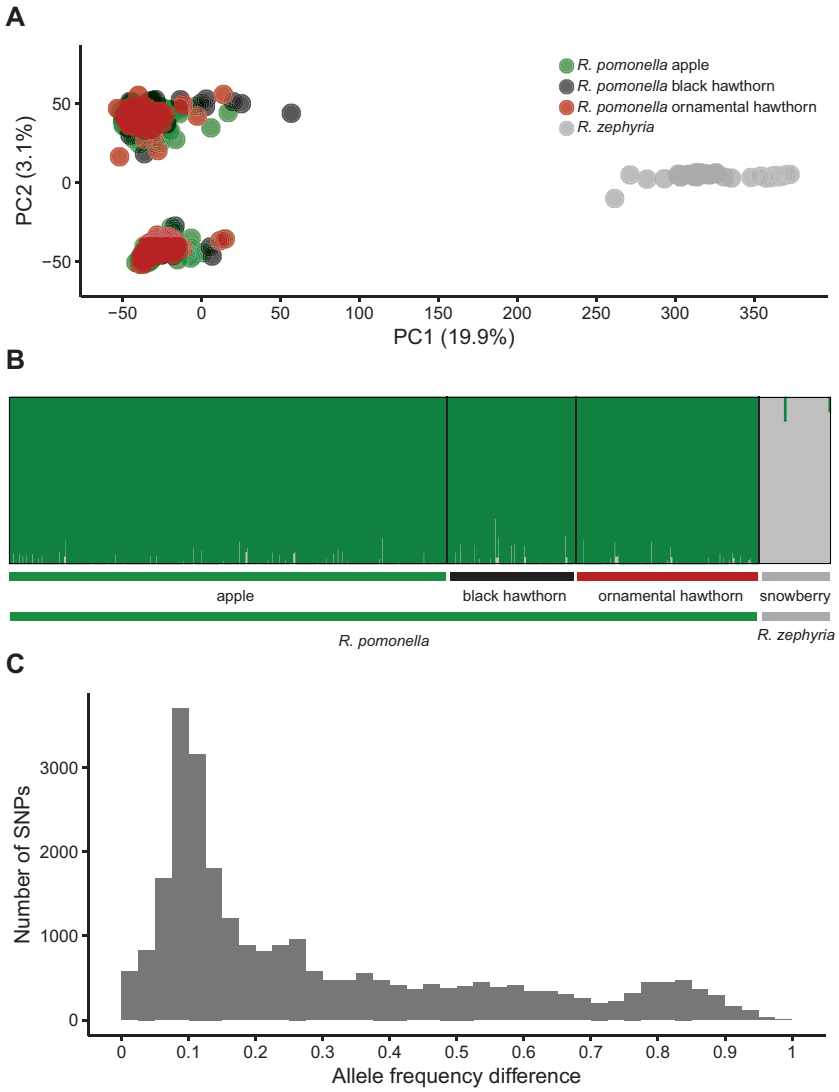

Fig. 3. (A) Principal component analysis (PCA) for ddRADseq SNPs genotyped for *R. pomonella* collected as larvae in infested apples (n = 275; green circles), black hawthorns (n = 81; black circles), and ornamental hawthorns (n = 115; red circles); and *R. zephyria* sampled from snowberries (n = 45; gray circles) at St. Cloud. Principle component 1 (PC1), plotted along the x-axis, separates the two species, while PC2 plotted along the y-axis separates male and female *R. pomonella*. (B) fastSTRUCTURE plot for *R. pomonella* (green) and *R. zephyria* (gray) at St. Cloud derived for K = 2 subpopulations, as indicated by the *chooseK.py* function implemented in fastSTRUCTURE (Raj et al. 2014). (C) The distribution of SNP allele frequency differences between *R. pomonella* (apple-, black hawthorn-, and ornamental hawthorn-infesting flies combined) versus *R. zephyria* at St. Cloud.

fixed (1.0) or exceptionally high in frequency in *R. pomonella* and also rare (<0.05) in *R. zephyria*.

Development of Markers for Species Diagnosis

To develop a RFLP-based diagnostic test, we concentrated on the 29 SNPs showing exceptional allele frequency differences between the species at St. Cloud that were invariant (fixed) in *R. zephyria*. For five of these 29 SNPs designated zeph2, zeph5, zeph6, zeph7, and zeph10, the variable nucleotide was contained within a restriction enzyme cut site that could potentially distinguish *R. pomonella* from *R. zephyria* (Table 2). Alignment of the five contigs containing these SNPs with the *R. zephyria* draft genome showed that each sequence mapped to a different major scaffold in snowberry-infesting flies. Each of the five scaffolds to which the markers mapped was between 17 and 337 kb in length, and the map position of each contig within its respective scaffold suggested that the five markers were separated by at least 60 kb. Assignment of the contigs to scaffolds allowed for the regions of DNA flanking the variable cut sites to be expanded outward beyond the contigs to permit PCR primer pairs to be designed to specifically amplify each locus (Table 1). For zeph2, zeph5, zeph6, and zeph7, the sizes of the amplicons generated by the designed primers were predicted to range from 156 to 401 bp (Table 2). For zeph10, the amplicon was expected to be 221 bp long in *R. zephyria* and be absent in *R. pomonella* due to the variable site differentiating *R. pomonella* being contained within the 3′ end of the reverse primer (Tables 1 and 2).

Sanger sequencing of PCR products generated for three representative sequences from both *R. zephyria* and *R. pomonella* verified that in all six cases the primers designed for zeph2, zeph5, zeph6, zeph7, and zeph10 amplified the intended target sequences (GenBank accession numbers MN812839–MN812863). However, inspection of the sequencing data revealed that *R. pomonella* is polymorphic for several small indels (1–7 bp) for zeph6 that slightly altered the amplicon length, as well as a large insertion of 60 bp in zeph2, which

correspondingly increased the amplicon size in *R. pomonella* only (Table 2). In neither case did the presence of indels alter the diagnostic utility of the marker. For each target locus, Sanger sequencing also confirmed the presence of the species characteristic SNP and the respective restriction site polymorphism, predicted from the ddRADseq data. The resulting RFLP fragment sizes for each marker and species are listed in Table 2.

Linkage Relationships and Genetic Independence of Diagnostic Markers

Correlation analyses showed that the five exceptionally diverged SNPs between *R. pomonella* and *R. zephyria* were associated with three independently evolving genomic regions: 1) zeph2, zeph6, and zeph10; 2) zeph5; 3) zeph7. Despite mapping to three different scaffolds and being separated by at least 60 kb, zeph2, zeph6, and zeph10 displayed highly significant inter-locus correlations in *R. pomonella* at St. Cloud ($r > 0.9$, $P < 0.000001$, >200 df for each of the three pairwise comparisons, including $r = 1.0$ between zeph2 and zeph10). Markers zeph5 and zeph7 were both genetically independent from each other and from zeph2, zeph6, and zeph10.

RFLP-Based Population Surveys Testing the Diagnostic Markers

To assess the utility of the five markers developed for distinguishing *R. pomonella* and *R. zephyria*, we performed a RFLP population survey of a total of 526 apple-, black hawthorn-, and snowberry-infesting flies collected from 11 sites across Washington (Fig. 1; Supp Table 1 [online only]). Due to the relatively high cost of the restriction enzyme HpyCH4V, only a subset of 48 *R. pomonella* and 48 *R. zephyria* were scored for zeph5, while all 526 flies were assessed for the remaining four markers. Across all 11 sites in the survey, all but one fly could be unambiguously diagnosed as either *R. pomonella* or *R. zephyria* based on PCA of multilocus RFLP results, with the first principal component accounting for over 99% of the genetic variation (Fig. 4A and B; Supp Table 3 [online only]). The one ambiguous individual was collected from black hawthorn at Home Valley, and had a multilocus genotype implying that it was heterozygous at all five RFLP loci and most likely an F1 hybrid between *R. pomonella* and *R. zephyria* (Fig. 4A and B). Two flies reared from snowberries (one each from Klickitat and Yakima) were heterozygous for the PstI enzyme cut site for zeph7, variation that was not present for *R. zephyria* at St. Cloud and presumed to be absent in the species (Table 2; Supp Tables 2 and 3 [online only]). Thus, it is possible that analysis of large sample sizes of snowberry flies in the future will reveal that *R. zephyria* shares all SNPs with *R. pomonella*, albeit at very low frequencies undetectable from the St. Cloud data set.

MiSeq Study Resolving Ambiguous Individuals

Additional ddRADseq analysis for individuals of ambiguous identity following the RFLP survey clarified their ancestry. Introgress analysis provided further support for the F1 hybrid status of the ambiguous individual collected from black hawthorn at Home Valley (Fig. 5). The estimated proportion of *R. zephyria* ancestry (hybrid index) for this individual (HV.BH.009) was 0.4998 (0.3784–0.6211, 95% confidence interval; Fig. 5A) having heterozygous genotypes

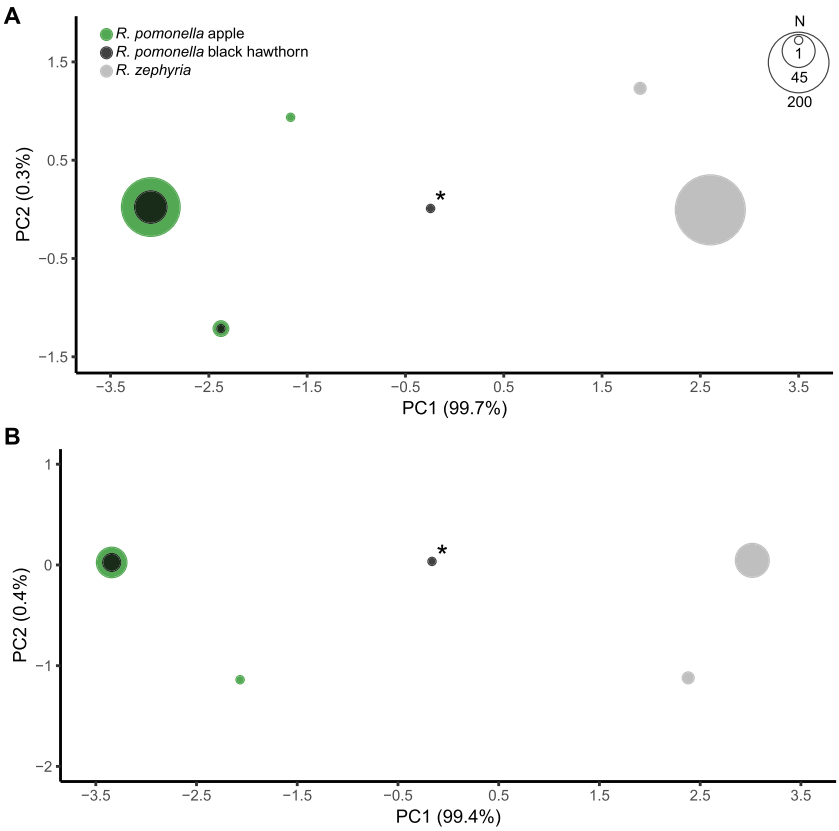

Fig. 4. (A) Principal component analysis (PCA) for RFLP survey data for *R. pomonella* ($n = 238$) and *R. zephyria* ($n = 280$) collected from 11 sites across Washington based on the four RFLP markers (zeph2, zeph6, zeph7, and zeph10); (B) PCA for *R. pomonella* ($n = 77$) and *R. zephyria* ($n = 51$) for the subset of individuals scored for all five RFLP markers (zeph2, zeph5, zeph6, zeph7, and zeph10). The size of the circles reflect the number of flies (n) having the same scores for PC1 and PC2. The single individual black hawthorn-infesting fly from the Home Valley (HV) with a high probability of being an F1 hybrid is denoted in both figures with an asterisk.

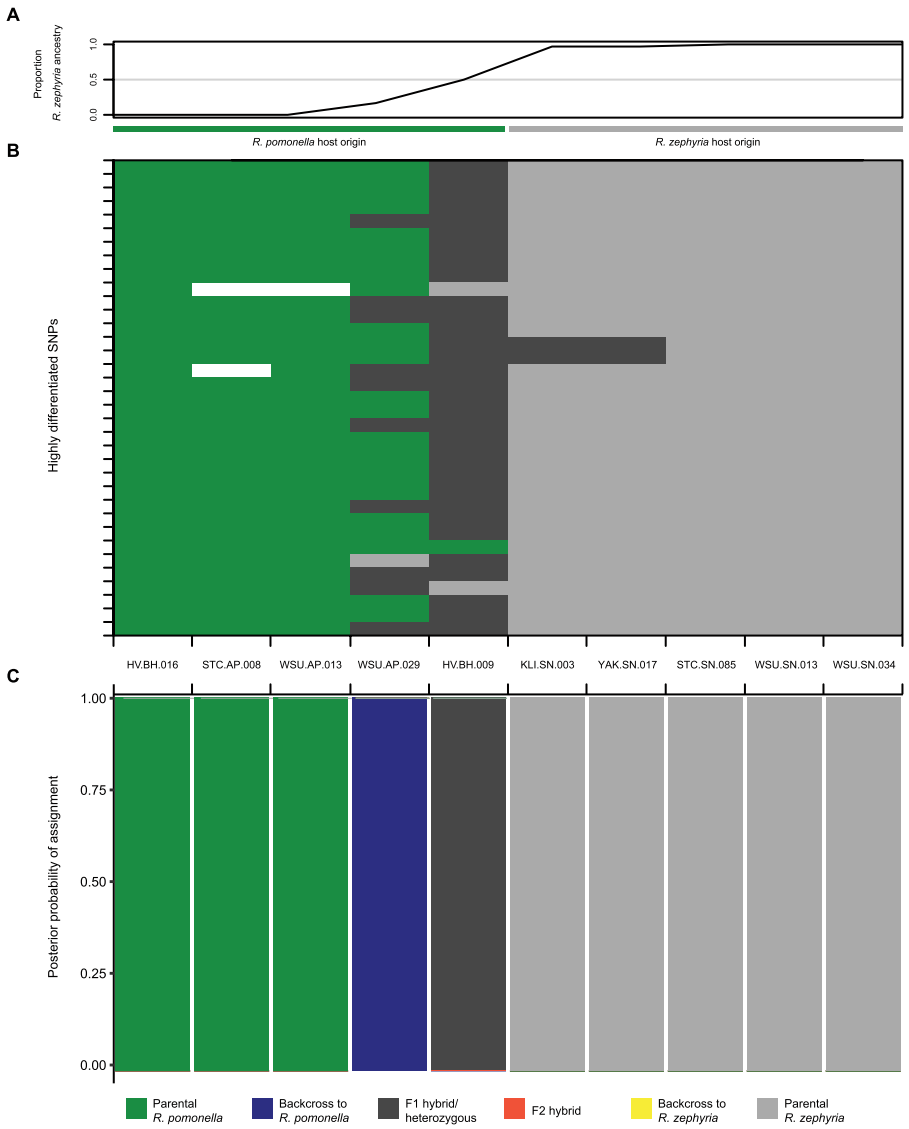

Fig. 5. Results of the MiSeq study to resolve ambiguous individuals in the RFLP survey, based on the 35 SNPs with estimated allele frequency differences ≥0.95 between *R. pomonella* and *R. zephyria* in the large ddRADseq study, which were also present in the MiSeq data set. (A) Calculated in Introgress, the estimated hybrid index (proportion of *R. zephyria* ancestry) for the putative F1 hybrid collected from black hawthorn at Home Valley (HV.BH.009) was 0.4998 (0.3784–0.6211, 95% CI). (B) Introgress analysis indicated that HV.BH.009 was heterozygous at 32 of 35 highly differentiated SNPs, while an additional individual collected from apple at WSU (WSU.AP.029) was heterozygous for 10 of 35 SNPs. The 35 highly differentiated SNPs are arrayed along the y-axis and individuals along the x-axis; green shading—homozygosity for the *R. pomonella* allele, light gray—homozygosity for the *R. zephyria* allele, dark gray—heterozygosity, and white—missing data. (C) Analysis with NewHybrids suggested a high probability (>0.99) that HV.BH.009 was indeed an F1 offspring of a *R. pomonella* × *R. zephyria* cross (indicated in dark gray). This analysis also identified WSU.AP.029 as the probable (>0.999) offspring of a backcross to *R. pomonella*. All other individuals were unambiguously assigned to the correct species. Individuals included in the MiSeq study are listed by site abbreviation (Fig. 1; Supp Table 1 [online only]), host (BH = black hawthorn, AP = apple, SN = snowberry), and individual ID number (Supp Table 3 [online only]).

(one *R. pomonella* and one *R. zephyria* allele) at 32 of 35 highly differentiated SNPs (Fig. 5B). Given that none of these 35 SNPs is differentially fixed between *R. pomonella* and *R. zephyria*, heterozygosity across all 35 SNPs would not be expected for true F1 hybrids. NewHybrids also identified HV.BH.009 as the F1 offspring of a *R. pomonella* × *R. zephyria* cross with a very high probability (>0.99; Fig. 5C). In addition, both analyses suggest that an additional individual collected from apple at WSU (WSU.AP.029) is likely the offspring of an F1 backcross to *R. pomonella* (posterior probability > 0.999; Fig. 5C). This individual, which was heterozygous at zeph2 and zeph10 (Supp Table 3 [online only]), was heterozygous at an additional 8 (total of 10) of 35 highly differentiated SNPs and had an estimated hybrid index of 0.1667, with a 95% confidence interval including 0.25, consistent with expectations for an F1 backcross (0.0890–0.2707; Fig. 5A and B). Although the two snowberry flies from Yakima (YAK.SN.017) and Klickitat (KLI.SN.003) were heterozygous for zeph7, they were not heterozygous for any other highly differentiated SNP and were unambiguously identified as *R. zephyria* by NewHybrids (Fig. 5B and C).

Discussion

Insect pests destroy ~15% of all U.S. crops, resulting in a loss of $15 billion annually (USDA 2014). Thus, developing quick, cost-effective, and reliable methods for detecting harmful species is critical to curtail insect damage. This task can be difficult, however, when pest

species are cryptic and co-occur with morphologically indistinguishable and harmless congeners. The situation can be further complicated when the pest species is invasive and has recently colonized an area where accurate information on the distribution of the insect across an entire region is needed to assess and control its spread. Moreover, agricultural crops may not represent the primary or ancestral host of a pest species but rather be a secondary or derived plant attacked by the insect (Walter 2005, Shu-sheng et al. 2012, Hendrichs et al. 2015). Thus, there is the added necessity of having to monitor a pest insect on both economic and noncrop hosts to determine the risks these latter populations may pose for attacking economically valuable plants. All of the above considerations apply to the introduction of the apple maggot fly, *R. pomonella*, into the PNW, where it has threatened the annual multibillion-dollar apple industry of the region for many years.

One way to address these problems is through the use of next-generation genomic sequencing techniques to devise simple genetic diagnostic tests to distinguish pest from nonpest taxa and assess their distributions through field surveys. We adopted this strategy here, using the apple maggot fly as a test case, to develop a five-marker RFLP genetic diagnostic tool to differentiate the invasive pest *R. pomonella* from its harmless snowberry-infesting sibling species, *R. zephyria*. We then demonstrated the utility of the test through a population survey of 526 flies sampled across 11 sites in Washington (Supp Table 1 [online only]). The *R. pomonella* RFLP diagnostic panel has several attractive features. First, it is rapid. Starting with DNA isolation from specimens and concluding with gel electrophoresis and genotyping of RFLPs, the species identity of an individual can be determined in a single work day. In addition, the RFLP survey can be easily combined with portable DNA diagnostic devices (i.e., 'mobile' or 'portable' labs) to help identify pests at the site of infestation to further decrease the time it takes to make informed management decisions (Zhang et al. 2016). To further reduce sample processing times, our approach can be combined with assays such as loop-mediated isothermal amplification (LAMP) reaction that proceeds at a constant temperature, precluding the use of traditional thermocyclers during amplification in field-based settings (Notomi et al. 2000).

Second, the approach is readily applicable to small sample sizes and can be scaled to screen moderate numbers of individuals easily without automation. Most importantly, the assay is well-suited for use during annual apple maggot surveys to determine the identity of one or a few individuals requiring immediate attention (e.g., Saint-Jean et al. 2013), with the general approach adaptable to use for any monitoring program of a pest or invasive species.

Third, the error rate for mistyping individuals is fairly low. In the case of our RFLP diagnostic panel, when conservatively considering independent markers zeph5, zeph6, and zeph7 only, for example, the chance of misidentifying a field trapped *R. pomonella* as *R. zephyria* is ~1.63×10^{-6}. This estimate assumes 1) assignment of a sample to the species for which it has a higher proportion of characteristic alleles (i.e., ≥4 of 6), and 2) the SNP allele frequencies estimated for *R. pomonella* at St. Cloud. Of course, additional RFLP markers could be designed to further increase the accuracy (i.e., decrease the false-positive rate) of species discrimination and facilitate the identification of individuals of hybrid ancestry. However, adding many additional markers would begin to negate the benefit of the rapidity of the RFLP test. We therefore advocate the use of ddRADseq in rare cases of ambiguous species diagnosis, via a reasonable number of RFLP markers. This is especially true for individual(s) captured in sensitive areas, where the fate of an orchard or implementation of a quarantine with significant financial consequences hinges on the test results. In these time-sensitive cases, involving relatively few samples ($N < 25$ to ensure adequate coverage), we recommend sequencing on the less expensive and relatively quick Illumina MiSeq, as we have demonstrated here; full processing of samples from DNA extraction and ddRADseq library preparation through sequencing and bioinformatic analysis could be completed in less than 1 wk, if services are available in-house. Larger sample sizes (hundreds to thousands) could be sequenced on higher-throughput Illumina platforms (e.g., HiSeq, NextSeq, NovaSeq) and would take a minimum of several weeks to process.

Lastly, the RFLP approach is generally inexpensive and should cost less than $10.00 per individual. However, the restriction enzyme HpyCH4V we used to score marker zeph5 was comparatively expensive, costing $3.00 per digestion reaction. The use of this enzyme stemmed, in part, from the construction of the original ddRADseq libraries, via digestion of genomic DNA samples with EcoRI and MseI. Given the low costs of these enzymes and generally high frequency of recognition site occurrence in the genome, it would have been advantageous, if possible, to have been able to include EcoRI and MseI cut sites in developing the RFLP diagnostic test. Given these considerations, if we were starting the study anew, we would explore the possibility of using alternate restriction enzymes or random shearing procedures for library construction. Alternatively, it is also possible to compare ddRADseq loci, generated using EcoRI and MseI, between species to identify 'null' alleles. If many reads of a particular sequence are observed for one taxon but not the other (i.e., 'null' or missing in this taxon), then this could indicate the absence of a cut site (i.e., a SNP in the cut site) for EcoRI or MseI in one of the target species. This pattern could be used in combination with an organism's genome sequence to design primers to amplify the region around the variable EcoRI or MseI cut site and allow for the inclusion of the cheaper EcoRI and MseI enzymes in the RFLP diagnostic test.

Our study is not without its limitations, however. For example, the diagnostic SNPs were developed from only a single location in western Washington, where *R. pomonella* infesting apple and ornamental hawthorn and black hawthorn and *R. zephyria* infesting snowberry are sympatric. Given the potential for gene flow between the two fly species (Green et al. 2013, Arcella et al. 2015), we chose this site for marker development to ensure that our diagnostic SNPs had a high probability of detecting low levels of introgression and hybridization. As we refine these diagnostic tools, we should include additional populations of *R. pomonella* and *R. zephyria* to account for potential geographic differences in SNP frequencies that may exist across Washington and the remainder of their ranges. Additionally, to truly validate the diagnostic SNPs for use in annual WSDA fruit inspections, the markers need to be formally tested in additional taxa closely related to *R. pomonella* and *R. zephyria*. *Rhagoletis pomonella* and *R. zephyria* belong to the *R. pomonella* sibling species complex, a group of at least six taxa that are morphologically and genetically closely related and whose ranges are broadly sympatric across the eastern United States (Berlocher 2000, Xie et al. 2008). However, we note *R. pomonella* and *R. zephyria* are the only two members of this group with ranges that extend into the northwestern United States. Regardless, we cannot rule out the possibility that other taxa in the *R. pomonella* species complex could be introduced into the PNW, which could complicate our diagnostic approach. In the future, we will validate these SNPs in additional taxa to safeguard against potential future introductions of *R. pomonella* group taxa into the PNW.

An important result from the current survey was the detection of one individual infesting black hawthorn at Home Valley that likely represents an F1 hybrid between *R. pomonella* and *R. zephyria* (Figs. 4 and 5). Based on our genotyping of a total of 526 individuals in the RFLP population survey and an additional 516 flies by ddRADseq at St. Cloud, this would translate into an estimated hybridization

rate of ~0.1% (=1/1,042) between the two species, approximately an order of magnitude less than what was reported in a population survey using microsatellites by Arcella et al. (2015). Of course, much larger sample sizes are needed to place a high degree of confidence in this estimate. Nevertheless, low level hybridization and introgression could help account for why none of the 26,951 SNPs we genotyped in the study was completely diagnostically fixed for alternate alleles between *R. pomonella* and *R. zephyria*.

The issue of hybridization is critical for pest management of the apple maggot fly for several reasons. First, hybrids between the two species could pose a threat to apple crops and, consequently, it may become advisable to clear snowberry bushes in the area of apple orchards, which would be a daunting and expensive control measure. Second, it has been hypothesized that introgression of genes from snowberry- into apple-infesting fly populations might be introducing alleles conferring greater resistance to desiccation in *R. pomonella*.

The apple-growing regions of central Washington are hotter and more arid than areas west of the Cascade Mountains, where *R. pomonella* was originally introduced and is commonly found in back yard and feral apples (Arguez et al. 2010, Siler et al. 2013). Thus, any factor increasing the potential for *R. pomonella* pupae to survive in drier soil conditions during overwinter diapause is problematic. Manipulative laboratory experiments have indicated that pupal mortality for *R. pomonella* originating from sites west of the Cascades in the PNW is indeed much greater compared to *R. zephyria* under lower humidity conditions (Hill 2016). Thus, evidence for low levels of hybridization found herein is worrisome if *R. zephyria* alleles conferring increased resistance to desiccation are capable of introgressing into *R. pomonella* and facilitating the fly's spread into the more arid, apple-growing region of central Washington. The possibility that *R. pomonella*'s advance into central Washington may be aided by adaptive gene flow from *R. zephyria* must be taken seriously and further investigated (Green et al. 2013, Arcella et al. 2015).

In conclusion, we developed and applied a relatively simple and cost-effective genomics approach to distinguish *R. pomonella* from *R. zephyria*. The RFLP diagnostic tests we developed is most useful when an immediate answer is required within a day for a modest number of specimens and is amendable to detecting any cryptic species of concern from a genetically closely related congener. However, as expenses for high-throughput DNA sequencing decrease, it may become feasible to conduct genome-based sequencing tests for large numbers of specimens, especially for taxa having relatively small genome sizes. The next step in our program is to validate our results with further testing, and then fully integrate our genetic test into the annual survey for the apple maggot fly in Washington and Oregon, as needed. Regardless, the genomic strategy we present for *R. pomonella* is transferable to potentially any invasive or pest organism at a modest initial investment of perhaps $5,000–10,000 and has the potential to save farmers and stakeholders in the apple industry from costly measures of quarantine control.

Supplementary Data

Supplementary data are available at *Annals of the Entomological Society of America* online.

Data Availability Statement

The raw sequence and filtered SNP data from this study are available from the Dryad Digital Repository (Doellman et al. 2019): https://doi.org/10.5061/dryad.x0k6djhfj.

Acknowledgments

We thank the Clark County Washington 78th Street Heritage Farm, the Washington State University Research and Extension Unit, Vancouver, WA, the Notre Dame Genomics and Bioinformatics Core Facility, and Jacqueline Lopez-Erickson, Michael Pfrender, Blair Wolfley, Doug Stienbarger, Terry Porter, Robert Goughnour, and Kathleen Rogers for their support and assistance. J.G. was funded by National Science Foundation Research Experience for Undergraduates Fellowship (IIS-1560363). This work was supported, in part, by grants to J.L.F. and G.R.H. from the National Science Foundation (DEB-1310850) and to J.L.F. from the United States Department of Agriculture (NIFA-2015-67013-23289), and to W.L.Y. by the Washington Tree Fruit Research Commission, Washington State Commission on Pesticide Registration, and United States Department of Agriculture Foreign Agriculture Service.

References Cited

AliNiazee, M. T., and R. L. Penrose. 1981. Apple maggot in Oregon: a possible new threat to the Northwest apple industry. Bull. Entomol. Soc. Am. 27: 245–246.

AliNiazee, M. T., and R. L. Westcott. 1986. Distribution of the apple maggot *Rhagoletis pomonella* (Diptera: Tephritidae) in Oregon. J. Entomol. Soc. Br. Columbia. 83: 54–56.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Anderson, E. C., and E. A. Thompson. 2002. A model-based method for identifying species hybrids using multilocus genetic data. Genetics. 160: 1217–1229.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arcella, T., G. R. Hood, T. H. Powell, S. B. Sim, W. L. Yee, D. Schwarz, S. P. Egan, R. B. Goughnour, J. J. Smith, and J. L. Feder. 2015. Hybridization and the spread of the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae), in the northwestern United States. Evol. Appl. 8: 834–846.

Arguez, A., I. Durre, S. Applequist, M. Squires, R. Vose, X. Yin, and R. Bilotta. 2010. NOAA's U.S. climate normals (1981–2010) [monthly normals]. Natl. Centers Environ. Information. doi:10.7289/V5PN93JP [22 July 2019].

Assour, L. 2014. Trimmer. https://github.com/UTbioinf/Hot-Rad.

Assour, L. A. 2015. Species divergence: computational methods to explore and resolve biological questions related to the evolution of species. Ph.D dissertation, University of Notre Dame, Notre Dame, IN.

Berlocher, S. H. 2000. Radiation and divergence in the *Rhagoletis pomonella* species group: inferences from allozymes. Evolution. 54: 543–557.

Berlocher, S. H., and J. L. Feder. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? Annu. Rev. Entomol. 47: 773–815.

Berlocher, S. H., B. A. McPheron, J. L. Feder, and G. L. Bush. 1993. Genetic differentiation at allozyme loci in the *Rhagoletis pomonella* (Diptera: Tephritidae) species complex. Ann. Entomol. Soc. Am. 86: 716–727.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30: 2114–2120.

Brunner, J. F. 1987. Apple maggot in Washington state: a review with special reference to its status in other western states. Melanderia. 45: 33–51.

Bush, G. L. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera, Tephritidae). Bull. Museum Comp. Zool. Harvard Univ. 134: 431–562.

Carruthers, R. I. 2003. Invasive species research in the United States Department of Agriculture-agricultural research service. Pest Manag. Sci. 59: 827–834.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, et al.; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27: 2156–2158.

Doellman, M. M., G. R. Hood, J. Gersfeld, A. Driscoe, C. C. Y. Xu, R. N. Sheehy, N. Holmes, W. L. Yee, and J. L. Feder 2019. Data from: Identifying diagnostic genetic markers for a cryptic invasive agricultural pest: a test case using the apple maggot fly (Diptera: Tephritidae). Dryad Digital Repository. https://doi.org/10.5061/dryad.x0k6djhfj.

Dowell, R. V. 1988. History of apple maggot in the western United States, p. 95. *In* R. V. Dowell, L. T. Wilson, and V. P. Jones (eds.), Apple maggot west. University of California Division of Agriculture and Natural Resources, Oakland, CA.

Egan, S. P., G. J. Ragland, L. Assour, T. H. Q. Powell, G. R. Hood, S. Emrich, P. Nosil, and J. L. Feder. 2015. Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. Ecol. Lett. 18: 817–825.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1988. Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. Nature. 336: 61–64.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1989. Inheritance and linkage relationships of allozymes in the apple maggot fly. J. Hered. 80: 277–283.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1990. Regional, local and microgeographic allele frequency variation between apple and hawthorn populations of *Rhagoletis pomonella* in western Michigan. Evolution. 44: 595–608.

Feder, J., T. Hunt, and G. Bush. 1993. The effects of climate, host-plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. Entomol. Exp. Appl. 69: 117–135.

Feder, J. L., S. M. Williams, S. H. Berlocher, B. A. McPheron, and G. L. Bush. 1999. The population genetics of the apple maggot fly, *Rhagoletis pomonella* and the snowberry maggot, *R. zephyria*: implications for models of sympatric speciation. Entomol. Exp. Appl. 90: 9–24.

Gompert, Z., and C. Alex Buerkle. 2010. introgress: a software package for mapping components of isolation in hybrids. Mol. Ecol. Resour. 10: 378–384.

Green, E., K. Almskaar, S. B. Sim, T. Arcella, W. L. Yee, J. L. Feder, and D. Schwarz. 2013. Molecular species identification of cryptic apple and snowberry maggots (Diptera: Tephritidae) in western and central Washington. Environ. Entomol. 42: 1100–1109.

Hendrichs, J., M. T. Vera, M. De Meyer, and A. R. Clarke. 2015. Resolving cryptic species complexes of major tephritid pests. Zookeys. 540: 5–39.

Hill, J. 2016. Adaptive variation in desiccation resistance in *Rhagoletis*. M.S. thesis, Western Washington University, Bellingham, WA.

Hood, G. R., W. Yee, R. B. Goughnour, S. B. Sim, S. P. Egan, T. Arcella, G. Saint-Jean, T. H. Q. Powell, C. C. Y. Xu, and J. L. Feder. 2013. The geographic distribution of *Rhagoletis pomonella* (Diptera: Tephritidae) in the western United States: introduced species or native population? Ann. Entomol. Soc. Am. 106: 59–65.

Jombart, T. 2008. *Adegenet*: a R package for the multivariate analysis of genetic markers. Bioinformatics. 24: 1403–1405.

Jombart, T., and I. Ahmed. 2011. *Adegenet 1.3-1*: new tools for the analysis of genome-wide SNP data. Bioinformatics. 27: 3070–3071.

Maxwell, S. A., G. Rasic, and N. Keyghobadi. 2009. Characterization of microsatellite loci for the western cherry fruit fly, *Rhagoletis indifferens* (Diptera: Tephritidae). Mol. Ecol. Resour. 9: 1025–1028.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

McPheron, B. A., D. C. Smith, and S. H. Berlocher. 1988. Genetic differences between host races of *Rhagoletis pomonella*. Nature. 336: 64–66.

Mertz, C., D. Koong, and S. Anderson. 2013. Washington Annual Agriculture Bulletin. United States Dep. Agric., Natl. Agric. Stat. Serv., Northwest Reg. F. Off., Olympia, WA.

Neilson, W. T. A., and J. W. McAllan. 1965. Effects of mating on fecundity of the apple maggot, *Rhagoletis pomonella* (Walsh). Can. Entomol. 97: 276–279.

Notomi, T., H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino, and T. Hase. 2000. Loop-mediated isothermal amplification of DNA. Nucleic Acids Res. 28: E63.

R Core Team. 2018. R: a language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Ragland, G. J., M. M. Doellman, P. J. Meyers, G. R. Hood, S. P. Egan, T. H. Q. Powell, D. A. Hahn, P. Nosil, and J. L. Feder. 2017. A test of genomic modularity among life-history adaptations promoting speciation with gene flow. Mol. Ecol. 26: 3926–3942.

Raj, A., M. Stephens, and J. K. Pritchard. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics. 197: 573–589.

Shu-sheng, L., J. Colvin, and P. J. De Barro. 2012. Species concepts as applied to the whitefly *Bemisia tabaci* systematics: how many species are there? J. Integr. Agric. 11: 176–186.

Siler, N., G. Roe, D. Durran, N. Siler, G. Roe, and D. Durran. 2013. On the dynamical causes of variability in the rain-shadow effect: a case study of the Washington Cascades. J. Hydrometeorol. 14: 122–139.

Sim, S. B., M. M. Doellman, G. R. Hood, W. L. Yee, T. H. Q. Powell, D. Schwarz, R. B. Goughnour, S. P. Egan, G. S. Jean, J. J. Smith, et al. 2017. Genetic evidence for the introduction of *Rhagoletis pomonella* (Diptera: Tephritidae) into the Northwestern United States. J. Econ. Entomol. 110: 2599–2608.

Saint-Jean, G., S. P. Egan, W. L. Yee, and J. L. Feder. 2013. Genetic identification of an unknown *Rhagoletis* fruit fly (Diptera: Tephritidae) infesting Chinese crabapple: implications for apple pest management. J. Econ. Entomol. 106: 1511–1515.

Stouthamer, R., P. Rugman-Jones, P. Q. Thu, A. Eskalen, T. Thibault, J. Hulcr, L.-J. Wang, B. H. Jordal, C.-Y. Chen, M. Cooperband, et al. 2017. Tracing the origin of a cryptic invader: phylogeography of the *Euwallacea fornicatus* (Coleoptera: Curculionidae: Scolytinae) species complex. Agric. For. Entomol. 19: 366–375.

Tracewski, K. T., J. F. Brunner, S. C. Hoyt, and S. R. Dewey. 1987. Occurrence of *Rhagoletis pomonella* (Walsh) in hawthorns, *Crataegus*, of the Pacific Northwest. Melanderia. 45: 19–25.

USDA. 2014. National agriculture statistics service. Field Office, Olympia, WA. www.nass.usda.gov.

Walter, G. 2005. Insect pest management and ecological research. Cambridge University Press, Cambridge, United Kingdom.

WSDA. 2001. Washington administrative code 16-470-108. Distribution of infested or damaged fruit is prohibited. Washington State Department of Agriculture. http://lawfilesext.leg.wa.gov/law/wsr/2001/11/01-11-145.htm.

WSDA. 2017. Agricultural land use. https://agr.wa.gov/PestFert/natresources/AgLandUse.aspx.

Xie, X., A. P. Michel, D. Schwarz, J. Rull, S. Velez, A. A. Forbes, M. Aluja, and J. L. Feder. 2008. Radiation and divergence in the *Rhagoletis pomonella* species complex: inferences from DNA sequence data. J. Evol. Biol. 21: 900–913.

Yee, W. L. 2008. Host plant use by apple maggot, western cherry fruit fly, and other *Rhagoletis* species (Diptera: Tephritidae) in central Washington state. Pan-Pac. Entomol. 84: 163–178.

Yee, W. L., and R. B. Goughnour. 2008. Host plant use by and new host records of apple maggot, western cherry fruit fly, and other *Rhagoletis* species (Diptera: Tephritidae) in western Washington state. Pan-Pac. Entomol. 84: 179–193.

Yee, W. L., and M. W. Klaus. 2013. Development of *Rhagoletis indifferens* Curran (Diptera: Tephritidae) in crabapple. Pan-Pac. Entomol. 89: 18–26.

Yee, W. L., P. S. Chapman, H. D. Sheets, and T. R. Unruh. 2009. Analysis of body measurements and wing shape to discriminate *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae) in Washington state. Ann. Entomol. Soc. Am. 102: 1013–1028.

Yee, W. L., H. D. Sheets, and P. S. Chapman. 2011. Analysis of surstylus and aculeus shape and size using geometric morphometrics to discriminate *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 104: 105–114.

Yee, W. L., M. W. Klaus, D. H. Cha, C. E. Linn, R. B. Goughnour, J. L. Feder, and H. Hagedorn. 2012. Abundance of apple maggot, *Rhagoletis pomonella*, across different areas in central Washington, with special reference to black-fruited hawthorns. J. Insect Sci. 12: 1–14.

Zhang, W., S. Guo, W. S. P. Carvalho, Y. Jiang, and M. J. Serpe. 2016. Portable point-of-care diagnostic devices. Anal. Methods. 8: 7847–7867.

Zhao, Z., T. Wahl, and T. Marsh. 2007. Economic effects of mitigating apple maggot spread. Can. J. Agric. Econ. Can. d'agroeconomie. 55: 499–514.