ENTOMOLOGICAL
SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Molecular Entomology

Partial Ribosomal Nontranscribed Spacer Sequences Distinguish *Rhagoletis zephyria* (Diptera: Tephritidae) From the Apple Maggot, *R. pomonella*

J. J. Smith,[1,2,4,] **P. Brzezinski,**[2] **J. Dziedziula,**[2] **E. Rosenthal,**[2] **and M. Klaus**[3]

[1]Department of Entomology, Michigan State University, 244 Farm Lane, Room 243, Michigan State University, East Lansing, MI 48825-1115, USA, [2]Lyman Briggs College, Michigan State University, 919 E. Shaw Lane, Room E-35, East Lansing, MI 48825-3804, USA, [3]Plant Protection Division, Washington State Department of Agriculture, 21 North 1st Avenue Suite 103, Yakima, WA 98902, USA, and [4]Corresponding author, e-mail: jimsmith@msu.edu

Subject Editor: Kent Shelby

Abstract

The apple maggot, *Rhagoletis pomonella* (Walsh), was introduced into the apple-growing regions of the Pacific Northwest in the U.S.A. during the past 60–100 yr. Apple maggot (larvae, puparia, and adults) is difficult to distinguish from its morphologically similar sister species, *Rhagoletis zephyria* Snow, which is native and abundant in the Pacific Northwest. While morphological identifications are common practice, a simple, inexpensive assay based on genetic differences would be very useful when morphological traits are unclear. Here we report nucleotide substitution and insertion–deletion mutations in the nontranscribed spacer (NTS) of the ribosomal RNA gene cistron of *R. pomonella* and *R. zephyria* that appear to be diagnostic for these two fly species. Insertion–deletion variation is substantial and results in a 49 base-pair difference in PCR amplicon size between *R. zephyria* and *R. pomonella* that can be scored using agarose gel electrophoresis. PCR amplification and DNA sequencing of 766 bp of the NTS region from 38 *R. pomonella* individuals and 35 *R. zephyria* individuals from across their geographic ranges led to the expected PCR fragments of approx. 840 bp and 790 bp, respectively, as did amplification and sequencing of a smaller set of 26 *R. pomonella* and 16 *R. zephyria* flies from a sympatric site in Washington State. Conversely, 633 bp mitochondrial COI barcode sequences from this set of flies were polyphyletic with respect to *R. pomonella* and *R. zephyria*. Thus, differences in NTS PCR products on agarose gels potentially provide a simple way to distinguish between *R. pomonella* and *R. zephyria*.

Key words: Apples, Fruit flies, molecular diagnostics, rDNA, snowberries

Tephritid flies in the *Rhagoletis pomonella* species group (Diptera: Tephritidae) have been of great interest and utility for studying fundamental evolutionary patterns and processes (Bush 1969; Feder et al. 1988; Powell et al. 2014; Doellman et al. 2018, 2019). However, these flies are also important in apple-growing regions in eastern North America as insect pests of apple (*Malus domestica* Bork.). In the Pacific Northwest of the U.S.A., apple maggot (*Rhagoletis pomonella* (Walsh)) is an introduced, major quarantine pest of apple (Dowell et al. 1990) that was first reported in Oregon in 1979 (AliNiazee and Penrose 1981), with the initial infestations there arising from its introduction from eastern North America (McPheron 1990b, Arcella et al. 2015). In the apple-growing areas of Oregon, Washington, and British Columbia, quarantine zones are in effect, economic costs of

potential infestation are high, and there is zero tolerance for the presence of apple maggot in apple shipments destined for export (WSDA 2001, Bush et al. 2002, CFIA 2017, Galinato et al. 2018).

Detection of apple maggot in and near apple orchards in the Pacific Northwest is complicated by the presence of *Rhagoletis zephyria* Snow, which is native to the Pacific Northwest (Gavrilovic et al. 2007), and common and abundant in apple-growing areas. *Rhagoletis zephyria* is a close relative of *R. pomonella* within the *pomonella* species group (Gavrilovic et al. 2007, Xie et al. 2007), and infests fruit of snowberries, *Symphoricarpos* spp. (Caprifoliaceae). *Symphoricarpos albus* var. *laevigatus* (Fern.), which is both common and abundant throughout the Pacific Northwest, is the primary host for *R. zephyria* in apple-growing regions.

Rhagoletis mendax Curran is a third species that is very closely related genetically and morphologically similar to *R. pomonella* and *R. zephyria* (Bush 1966). However, *R. mendax* is known only from *Vaccinium* spp. and *Gaylussacia* spp. hosts, and its geographic range is limited to the eastern half of North America (Payne and Berlocher 1995, Smith et al. 2001). Thus, *R. mendax* does not cause a threat to apples or create a problem for the apple industry in the Pacific Northwest.

Because *R. zephyria* is virtually identical morphologically to *R. pomonella* (Bush 1966), this creates problems for morphological identifications in areas where the ranges of these species overlap. Despite these similarities, Westcott (1982) was able to draw upon the work of Curran (1924), Wasbauer (1963), and Bush (1966, 1969) to establish morphological criteria for distinguishing between adults of the two species. Males of the two species can be distinguished from each other based on a difference in the surstylus shape, while females differ in mean ovipositor length. These characters have been used extensively as the basis for species diagnosis of field-caught adult flies in annual apple maggot surveys in Washington State. Ovipositor length can be used to distinguish between females of the two species, but large *R. zephyria* females can be difficult to distinguish from small *R. pomonella* females; males of the two species appear to be reliably distinguished by surstylus shape (Westcott 1982). Additionally, the introduction of *R. pomonella* into the Pacific Northwest over the past 50 yr has created problems for species identification, as introgression appears to have been ongoing (McPheron 1990a), presumably between the relatively abundant *R. zephyria* males and the introduced *R. pomonella* females (Arcella et al. 2015).

A simple, inexpensive diagnostic tool to distinguish between *R. pomonella* and *R. zephyria* would be very useful when morphological characters, especially in females, are unclear (i.e., ovipositor lengths are intermediate between typical specimens). DNA sequencing of the barcode or some other region of the mitochondrial DNA (Hebert et al. 2004) would appear to be the logical choice for this task, but published work to date (Smith and Bush 1997, Schwarz et al. 2005, Michel et al. 2007) has not revealed unique, diagnostic haplotypes for *R. pomonella* and *R. zephyria* based on mitochondrial DNA sequences. It is unclear whether the inability of mtDNA to distinguish between *R. pomonella* and *R. zephyria* is due to insufficient mitochondrial DNA sequence data (more nucleotides need to be sequenced per individual) or if the history and structure of *R. pomonella* and *R. zephyria* populations is such that their mitochondrial DNA haplotypes are shared because of ongoing introgression and/or incomplete lineage sorting.

To date, the use of multilocus genotypes to identify unknown field-caught flies as either *R. zephyria* or *R. pomonella* has had the highest potential for success (Schwarz et al. 2005, Michel et al. 2007). Green et al. (2013) described a quantitative real-time PCR approach for species diagnosis in *R. zephyria* and *R. pomonella* based on four single nucleotide polymorphisms (SNPs) derived from expressed sequence tags (ESTs). While the assay developed by Green et al. (2013) was useful in most cases for distinguishing whether a fly originated from apple, black hawthorn (*Crataegus douglasii* Lindl.), or snowberry (S. *albus* var. *laevigatus*), their assay was complicated by the presence of *R. zephyria* alleles that presumably had introgressed into the *R. pomonella* gene pool in areas where the two species were in close proximity. More recently, Doellman et al. (2020) developed a diagnostic test based on PCR amplification of alleles at five nuclear loci followed by restriction enzyme digestion to reveal differences between field-caught *R. pomonella* and *R. zephyria*, that appears to be diagnostic but is somewhat technically demanding.

Here we describe genetic variation in the rRNA gene cistron of *R. pomonella* and *R. zephyria* (Fig. 1) that provides an additional useful tool for distinguishing between individuals in these two fly species, either as field-caught adults or larvae. As a part of a larger study in our lab on species relationships in the genus *Rhagoletis* (Hulbert 2018), we examined molecular variation in a 625-nucleotide fragment of the D2 expansion segment region of the 28S rRNA gene. This study revealed a single nucleotide insertion mutation in the 28S D2 expansion region of *R. zephyria* individuals that appeared to distinguish these flies from *R. pomonella*. Subsequent assemblies of larger portions of the rDNA cistron of *R. pomonella* and *R. zephyria* using Illumina NextGen sequences (Smith et al. in preparation) indicated that there may be several DNA mutations, both single nucleotide substitutions and larger scale insertion–deletion (indel) mutations, upstream of the 18S structural rRNA gene, in both the external transcribed spacer (ETS) and nontranscribed spacer (NTS) regions of the rDNA cistron.

We report here genetic variation in a portion of the ribosomal NTS region that appears to be useful for distinguishing between *R. pomonella* and *R. zephyria*, with indel mutations that potentially provide the basis for a simple diagnostic test for these two species based on a 49 bp size difference in their PCR amplicons. The ribosomal region that we examined (the 3' end of the NTS region; Fig. 1) contains both DNA substitutions and indel variation that appear to be fixed between *R. pomonella* and *R. zephyria*. On the other hand, *R. pomonella* and *R. mendax*, the third named species in the *pomonella* group sibling species complex, appear to be polyphyletic with respect to each other based on the NTS gene region, while *R. zephyria* remains monophyletic with respect to both *R. pomonella* and *R. mendax*.

Materials and Methods

Taxon Sample

The taxon sample for this study consists of both a geographic sample of 101 flies collected from across the northern US, and a sample collected mainly from Beacon Rock State Park near the Columbia River in Washington State. The geographic sample consists of 101 individual flies (38 *R. pomonella*, 35 *R. zephyria*, and 26 *R. mendax*)
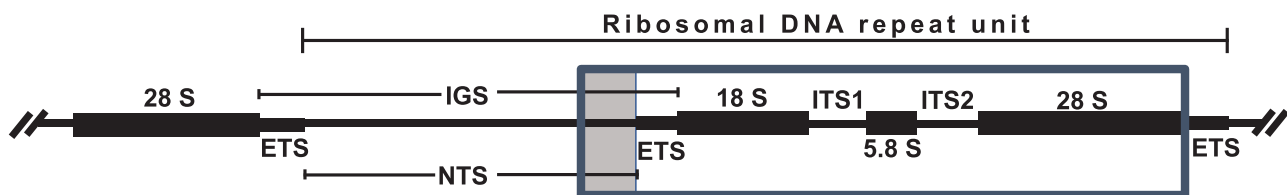

Fig. 1. **Diagram of the eukaryotic ribosomal DNA repeat unit (cistron), showing the location of the NTS area sequenced in this study.** The complete repeating unit of rDNA is shown, showing the relative locations and sizes of the different components. This basic structure, with minor variation, is conserved across the eukaryotes. Our overall rDNA alignment is 9,221 nucleotide positions, which comprise the area outlined with the bold box. We determined sequences (approx. 800 bp) from the 3' end of the nontranscribed spacer (NTS; shaded area within bold box) within the IGS region for 141 *R. pomonella*, *R. mendax*, and *R. zephyria* flies. Figure adapted from Hwang and Kim (1999).

Table 1. *Rhagoletis pomonella, R. zephyria,* and *R. mendax* analyzed in this study, including collection information (locality, host plant, date)

Taxon	Designation	State/Province	Locality (County)	n	Host	Collector	Date	Lat	Long
R. mendax	MNJ	New Jersey	Rutgers Expt Stn (Burlington)	2*	Vaccinium corymbosum	OL	7/VIII/2000	39.72	−74.51
R. mendax	MEL	Michigan	MSU Clinical Center (Ingham)	10	Vaccinium corymbosum	JEC/BF	5/VIII/1999	42.72	−84.46
R. mendax	MJP	Indiana	Jasper-Pulaski WR (Jasper)	9	Gaylussacia dumosa	JS	1/VIII/1999	41.15	−86.97
R. mendax	MWB	Ontario	Wainfleet Bog (Niagara)	3	Vaccinium corymbosum	DD/JS	1/VIII/2001	42.91	−79.30
R. pomonella	PIL	Illinois	Riverwoods (Lake)	5*	Crataegus mollis	VG/MJ	12/IX/1999	42.19	−87.88
R. pomonella	PMA	Massachusetts	Amherst (Hampshire)	9*	Crataegus mollis	AR/GW/VG	24/IX/2000	42.39	−72.53
R. pomonella	PMN	Minnesota	Staples (Todd)	1*	Crataegus mollis	GLB/DB	5/IX/1994	46.40	−94.93
R. pomonella	PNE	Nebraska	I-80, Exit 285 (Buffalo)	1*	Crataegus mollis	VG	3/IX/1999	40.70	−98.84
R. pomonella	PUT	Utah	Wellsville (Cache)	4	Crataegus douglasii	DA	23/VIII/2000	41.63	−111.93
R. pomonella	PCO	Colorado	Boulder (Larimer)	4	Crataegus mollis	VG/BB	5/IX/1999	40.00	−105.29
R. pomonella	PNY	New York	Geneva (Seneca)	7	Crataegus mollis	AR/GW/VG	24/IX/2000	42.87	−77.01
R. pomonella	PPA	Pennsylvania	Biglerville (Centre)	4	Malus domestica	MC	18/VIII/1999	39.93	−77.25
R. pomonella	PWA	Washington	St. Cloud Ranch (Skamania)	1*	Crataegus douglasii	VG	12/IX/2000	45.60	−122.11
R. pomonella	MK_haw	Washington	Beacon Rock St. Pk. (Skamania)	26	Crataegus douglasii	MK	IX/2000	45.62	−122.03
R. zephyria	ZEL	Michigan	East Lansing (Ingham)	2*	Symphoricarpos albus var. laevigatus	JS/AR	10/VIII/2000	42.73	−84.48
R. zephyria	ZMN	Minnesota	Hawley (Clay)	2*	Symphoricarpos occidentalis	GLB/DB	7/IX/1994	46.87	−96.47
R. zephyria	ZMT	Montana	Melstone (Musselshell)	2*	Symphoricarpos occidentalis	GLB/DB	9/IX/1994	46.59	−107.89
R. zephyria	ZND	North Dakota	Mandan (Burleigh)	1*	Symphoricarpos occidentalis	GLB/DB	9/IX/1994	46.69	−100.90
R. zephyria	ZPA	Pennsylvania	State College (Centre)	5*	Symphoricarpos albus var. laevigatus	VG	IX/1995	40.79	−77.86
R. zephyria	ZSD	South Dakota	Custer (Pennington)	1*	Symphoricarpos occidentalis	VG	9/IX/1999	43.75	−103.36
R. zephyria	ZNY	New York	Geneva (Seneca)	8	Symphoricarpos albus var. laevigatus	JS	18/VIII/1997	42.87	−77.01
R. zephyria	ZWI	Wisconsin	Waukesha (Waukesha)	4	Symphoricarpos occidentalis	VG	12/IX/1999	43.00	−88.27
R. zephyria	ZMA	Massachusetts	Amherst (Hampshire)	4	Symphoricarpos albus var. laevigatus	RJP	IX/1998	42.39	−72.53
R. zephyria	ZCO	Colorado	Boulder (Larimer)	4	Symphoricarpos albus var. laevigatus	VG/BB	5/IX/1999	40.01	−105.27
R. zephyria	ZWA	Washington	Dixie (Walla Walla)	2	Symphoricarpos albus var. laevigatus	GLB/DB	31/VIII/1997	46.14	−118.17
R. zephyria	MK_snow	Washington	Beacon Rock St. Pk. (Skamania)	10	Symphoricarpos albus var. laevigatus	MK	IX/2000	45.62	−122.03
R. zephyria	MK_snow	Washington	St. Cloud Ranch (Skamania)	6	Symphoricarpos albus var. laevigatus	MK	IX/2000	45.60	−122.11
R. pomonella Mex.	R pom Mex	Mexico	San Diego, Mex. Mex.	2*	Crataegus mexicana	JG	X/1995	19.26	−99.10

One individual from each of sample with an asterisk (*) was sequenced using Illumina NextGen (Smith et al. in preparation), except *R. pomonella* Mex., in which both flies were sequenced.

Collectors: DA, Diane Alston; BB, Branislav Blagojevic; DB, Dorie Bush; GLB, Guy L. Bush; MC, Marty Condon; JEC, Joe Crossno; DD, Danny Ducat; BF, Ben Fernandez; VG, Vesna Gavrilovic; JG, Jorge Graciano; MJ, Matt Jaycox; MK, Mike Klaus; OL, Oscar Liburd; RJP, Ron J. Prokopy; AR, Angela Roles; JS, Jim Smith; GW, George Weiblen.

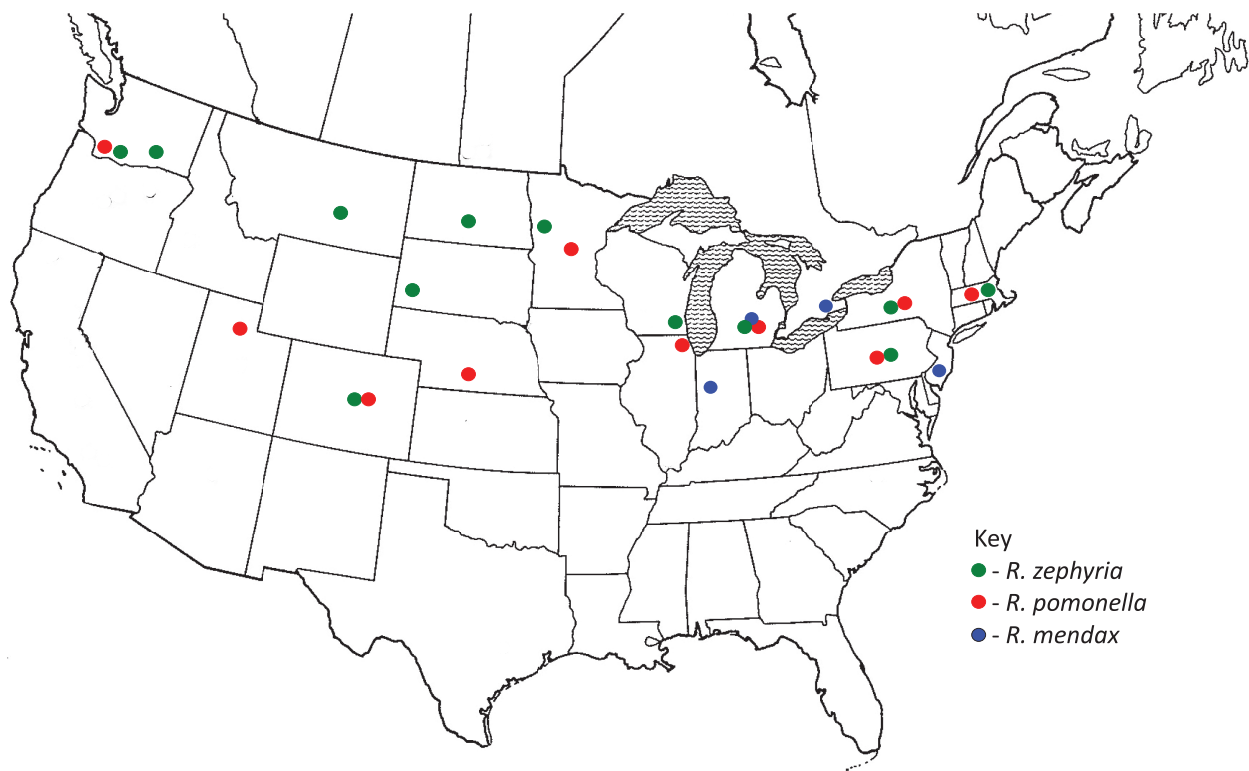Latitudes and Longitudes in *italics* are approximations.

Fig. 2. Map showing collection locations of fly samples used in this study. Locations where fly samples were collected from hawthorns or apples (*R. pomonella*), from snowberries (*R. zephyria*), and from blueberries or huckleberries (*R. mendax*) are shown. The location where flies were collected from hawthorn in Mexico (*R. pomonella* Mex. EVTM) is not shown. Further information on each collection locality can be found in Table 1.

from populations across their respective geographic ranges (Table 1; Fig. 2), plus two *R. pomonella* Mex. individuals from the Eje Volcanico Trans Mexicano (EVTM; Michel et al. 2007). Following Green et al. (2013), species names were assigned to individuals collected from apple or hawthorn as *R. pomonella*, individuals collected from snowberry as *R. zephyria*, and individuals collected from blueberry as *R. mendax*. This set of 101 flies includes 14 flies from which we have obtained approx. 9,000 bp sequences of the ribosomal DNA cistron (partial NTS, ETS, 18S, ITS1, 5.8S, ITS2, and 28S; GenBank Accession #'s MN507539–MN507552) and 14,000 bp of mitochondrial DNA (all 37 genes; GenBank Accession #'s MN443932–MN443945) via Illumina NextGen sequencing (Table 1; Smith et al. in preparation). The Beacon Rock sample includes 26 individual *R. pomonella* flies collected from *C. douglasii* and 10 individual *R. zephyria* flies collected from *S.albus* at Beacon Rock State Park (WA) in 2000, and also includes six individual *R. zephyria* flies collected nearby from *S. albus* at St. Cloud Ranch (WA) in 2000. Beacon Rock St. Park and St. Cloud Ranch correspond to sites 17 and 18, respectively, in Mattson et al. (2022).

PCR Primers

We designed PCR primers and developed PCR conditions that reliably amplify an approximately 800 bp portion of the NTS gene region of the rDNA cistron in *R. pomonella*, *R. zephyria*, and *R. mendax* individuals (Fig. 1). In preliminary work, examination of approx. 9,000 bp of the rDNA cistron from 14 individual *Rhagoletis* individuals (six *R. zephyria*, five *R. pomonella*, one *R. mendax*, and two *R. pomonella* Mex. flies; GenBank Accession #'s MN507539–MN 507552) revealed marked indel variation between the *R. pomonella* and *R. zephyria* individuals in the ETS and NTS regions. We examined this further by designing PCR primers

that could amplify the DNA segments in which these indels occurred. Two primer pairs were designed using Primer-BLAST as implemented by the NCBI website (Ye et al. 2012). The first pair (PMZrDNA1F/PMZrDNA1R; Table 2) was designed to amplify approx. 800 bp of DNA located at the 3' end of the nontranscribed spacer (NTS) of the rDNA cistron. The second pair (PMZrDNA2F/PMZrDNA2R) was designed to amplify approx. 800 bp of DNA located at the 3' end of the external transcribed spacer (ETS) and the 5' end of the 18S rDNA structural gene. Examination of variation in these two rDNA regions across the 14 individual *Rhagoletis* individuals indicated that the NTS region amplified using PMZrDNA1F/PMZrDNA1R was more likely to provide a marked size difference in fragments between *R. pomonella* and *R. zephyria* individuals. All the work presented here is thus based on amplifications of the NTS region using the PMZrDNA1F/PMZrDNA1R primer pair.

The mitochondrial COI barcode primers of Hebert et al. (2004), LEP F1 and LEP R1 (Table 2), were used to amplify the mitochondrial COI region from the same set of flies from which we amplified the NTS and ETS regions. As observed by Barr et al. (2021) in European cherry fruit fly *Rhagoletis cerasi* (L.), the quality of some COI sequence reads from *Rhagoletis* flies in our sample was low, leading to 'dirty' sequences. In two cases, PCR of whole fly *Rhagoletis* extracts using LEP F1 and LEP R1 resulted in the amplification of COI fragments that were very clean upon sequencing, but yielded a DNA sequence that matched exactly the COI sequence from *Wolbachia pipientis* (GenBank Accession CP 042445.1). This phenomenon has been observed before (Smith et al. 2012), and we thus followed the recommendation of Barr et al. (2021) and used the primer pair TY-J-1460 and C1-N-2191 (Simon et al. 1994) for a number of subsequent COI amplifications, modifying the TY-J-1460

Table 2. DNA sequences of PCR primers used and/or discussed in this study

a) Nuclear Ribosomal NTS and ETS primers		
PMZrDNA1F	5′-TTRCTTCCGTGAGGTGTGTG-3′	This study
PMZrDNA1R	5′-AAACTAACTGTGGTAAGACCAGGA-3′	This study
PMZrDNA2F	5′-TGATGTATGCACAAATGCATTGAAA-3′	This study
PMZrDNA2R	5′-TGTGGTAGCCGTTTCTCAGG-3′	This study
b) Mitochondrial COI primers		
LCO-1490	5′-GGTCAACAAATCATAAAGATATTGG-3′	Folmer et al. (1994)
LEP F1 Hebert	5′-ATTCAACCAATCATAAAGATAT-3′	Hebert et al. (2004)
LEP F1 Smith	5′-ATTCAACCAATCATAAAGATATTGG-3′	Smith et al. (2007)
HCO-2198	5′-TAAACTTCGGGGTGACCAAAAAATCA-3′	Folmer et al. (1994)
LEP R1 Hebert	5′-TAAACTTCTGGATGTCCAAAAA-3′	Hebert et al. (2004)
LEP R1 Smith	5′-TAAACTTCTGGATGTCCAAAAAATCA-3′	Smith et al. (2007)
C1-N-2191	5′-CCCGGTAAAATTAAAATATAAACTTC-3′	Simon et al. (1994)
TY-J-1460	5′-TACAATTTATCGCCTAAACTTCAGCC-3′	Simon et al. (1994)
TY-J-1460_Rhag	5′-TACAGTCTATCGCCTAAATTCAGCC-3′	This study

Nucleotide symbols match IUPAC convention.

primer to match the known *Rhagoletis* sequence of the tRNA$_{Tyr}$ gene (TY-J-1460_Rhag; Table 2).

PCR

All PCR amplification was carried out using a BioRad T100 Thermocycler in 25 µl reactions using GotaqFlexi (Promega, Madison, WI) with the following reagents (and concentrations): reaction buffer (1X), MgCl$_2$ (8 mM), dNTP (0.5 mM each), forward and reverse primers (0.5 mM each), DNA polymerase (2.5 u), and DNA template (~18–92 ng). The forward and reverse primers used in the PCR are listed in Table 2. Thermocycler conditions used to amplify the approx. 800 bp ribosomal NTS fragment using the PMZrDNA1F/PMZrDNA1R primer pair were as follows: initial denaturation was at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 30 s, primer annealing at 58°C for 30 s, and product elongation at 72°C for 1 min (×35), followed by a final product elongation at 72°C for 5 min. To amplify the approx. 650 bp mitochondrial COI fragment using the LEP F1/LEP R1 primer pair, the conditions used were as described above except the annealing temperature was set at 46°C and the product elongation step was 2 min. These latter conditions were also used to amplify COI using the TY-J-1460_Rhag/C1-N-2191 primer pair.

Agarose Gel Electrophoresis

Agarose gel electrophoresis for determining if PCR was successful in preparation for DNA sequencing was carried out with a MiniOne gel apparatus using 1.5–2.0% (w/v) agarose in a gel volume of 15–20 ml. MiniOne gels were run in 1X TBE buffer for 15–20 min. For the higher resolution of PCR products in the 700–900 bp range, larger 2% agarose gels were used and run at 80 V for 4 hr. DNA bands were visualized using SYBR Green (Invitrogen S33102; Waltham, MA) with a UV transilluminator and photographed using a mobile phone. A 100 bp ladder (Gold Biotechnology D001-500; St. Louis, MO) was used to estimate band sizes.

DNA Sequencing and Sequence Alignment

PCR products for DNA sequencing were purified using QIAquick PCR Purification Kits (Qiagen; Hilden, Germany) according to the manufacturer's recommendations. Sanger sequencing was performed at the Michigan State University Research Technology Support Facility via BigDye Terminator Sequencing on an Applied Biosystems 3730xl DNA Analyzer (Foster City, CA, USA) using the PCR primers as sequencing primers. Base calls were made within MEGA (vers. 7.2; Kumar et al. 2016) based on the sequences of both the forward and reverse strands. DNA nucleotide substitutions observed within the data sets were verified by visual examination of .ab1 file tracings within MEGA. DNA sequence alignment of the NTS amplicons was carried out using a TCS alignment (Chang et al. 2015) of 14 complete ribosomal DNA cistron sequences from *R. pomonella*, *R. zephyria*, and *R. mendax* (described above) as a scaffold. All of the NTS sequences generated for this study via Sanger sequencing readily and unambiguously aligned to this scaffold. Mitochondrial COI sequences obtained for this study readily aligned to a set of published COI sequences from *R. pomonella* and *R. mendax* (Frey et al. 2013).

Phylogenetic Analysis and Network Construction

Phylogenetic analysis of the aligned NTS sequences was carried out via the neighbor-joining algorithm of Saitou and Nei (1987) as implemented in PAUP*4.0 Version 4.0a (build 167) for Macintosh (X86) (Swofford 2000). The minimum evolution criterion was used as this has been shown to be among the most efficient methods when sequences are closely related (Nei 1991). The NTS sequences from two *R. pomonella* Mex. EVTM were used as the outgroup. Neighbor-joining topologies were examined in Parsimony mode in PAUP to identify unique alleles for generating input files for minimum-spanning network construction. Character changes on branches were identified in PAUP using 'Show Reconstructions'. Minimum spanning networks (Bandelt et al. 1999) were generated using the computer program PopArt (http://popart.otago.ac.nz) with default settings.

Results

Preliminary PCR of the NTS/ETS region of the ribosomal DNA with both primer pairs PMZrDNA1F/PMZrDNA1R and PMZrDNA2F/PMZrDNA2R using a selected number of *R. pomonella* and *R. zephyria* individuals to provide template DNA led to amplification of the predicted NTS and ETS regions of the rDNA cistron sequences that had been obtained by Illumina NextGen sequencing (GenBank Accession #'s MN507539–MN507552). Because indel variation in the NTS sequences appeared to provide an opportunity to distinguish between *R. pomonella* and *R. zephyria* based on PCR product length differences, we chose to focus our work on the NTS sequences obtainable using the primer pair PMZrDNA1F/PMZrDNA1R.

NTS Sequences From the Geographic Sample

PCR-amplification using the primer pair PMZrDNA1F/PMZrDNA1R led to a PCR product of approx. 840 bp in the *R. pomonella* and *R. mendax* individuals tested from the geographic sample (*n* = 64), and approx. 790 bp in the *R. zephyria* individuals from the geographic sample (*n* = 35), based on visualization of fragments via agarose gel electrophoresis (Fig. 3) and the size of the sequenced PCR amplicons (see below).

We obtained DNA sequences from the amplified NTS region from 38 *R. pomonella*, 35 *R. zephyria*, and 26 *R. mendax* individuals in the geographic sample. DNA sequencing of each PCR product indicated that these PCR products were 839 bp in *R. pomonella* and *R. mendax*, and 792 bp in *R. zephyria*, respectively, with minor size variation within species due to the presence of short 1–2 bp indels in some sequences. Base calling and alignment of the sequences were straightforward and unambiguous, resulting in 766 aligned nucleotide positions, which is included in .fas format in Supp File 1a (online only).

The PCR products obtained with primers PMZrDNA1F and PMZrDNA1R from all flies studied here yielded DNA sequences that were very clean. There was no evidence of PCR product size heterogeneity, as would be evidenced by sequences that become unreadable at a particular point. This would appear to indicate that the NTS indel structure did not vary within individuals for this PCR amplicon, and that the ribosomal DNA cistrons are well homogenized within the three species examined here. On the other hand, we did observe what could be interpreted as evidence of DNA substitution heterozygosity within *R. mendax*: some changes observed within the *R. mendax* individuals appeared to involve heterozygous states (double peaks; state assigned as the higher of the two peaks). However, no double peaks were observed in the NTS sequences from *R. pomonella* and *R. zephyria* individuals; all of the base calls were clean and we had no difficulty assigning the character states within these sequences.

The overall size difference of the NTS fragment between *R. pomonella* and *R. mendax* individuals compared to *R. zephyria* individuals is accounted for by the presence of a characteristic set of indels in the NTS alignment, each of which appears to have been uniquely derived (relative to the *R. pomonella* Mex. EVTM outgroup) in either all *R. zephyria* individuals, or in all *R. pomonella* and *R. mendax* individuals (Table 3). For example, all *R. zephyria* individuals examined contained a single deletion of 15 nucleotides from positions 318–332 in the NTS alignment relative to *R. pomonella* and *R. mendax* (Fig. 4). Conversely,

all *R. pomonella* and *R. mendax* individuals contain a 20-nucleotide insertion from positions 352–371. Consideration of all of the indels in the PCR amplicon corresponding to the NTS region examined accounts for the size difference of 49 nucleotides between the ribosomal NTS PCR products obtained from *R. pomonella* and *R. mendax* individuals, compared to those obtained from *R. zephyria* individuals.

We also observed single-nucleotide substitution variation within the *R. pomonella*, *R. mendax*, and *R. zephyria* individuals of the geographic sample at 12 nucleotide positions in the aligned data, with 11 of these being phylogenetically informative. A minimum evolution neighbor-joining tree based only on DNA substitution mutations shows the relationships of the NTS sequences obtained from the 38 *R. pomonella*, 35 *R. zephyria*, and 26 *R. mendax* individuals in the geographic sample (Fig. 5). Even though DNA substitution variation was slight, the NTS sequences from the 35 *R. zephyria* individuals formed a monophyletic group supported by a bootstrap value of 96. Three of the observed DNA substitutions (position 4, A → T; position 41, G → C; position 113, A → G) are uniquely derived in all of the *R. zephyria* individuals, and only in *R. zephyria* individuals, providing what appear to be diagnostic nucleotide substitutions (character states) for *R. zephyria*, above and beyond the set of indel mutations described above. Within the *R. zephyria* cluster, there is also a group of 9 sequences, also with strong bootstrap support, that is mainly composed of *R. zephyria* individuals from the eastern U.S. and differs from the other *R. zephyria* sequences by a single nucleotide change at position 696 in the alignment. While *R. zephyria* was monophyletic in the NTS tree, *R. pomonella* and *R. mendax* were not, appearing polyphyletic with respect to each other in the neighbor-joining tree (Fig. 5). Three *R. mendax* individuals (MJP5, MEL8, and MEL20) share the sequence that is most common for *R. pomonella*, while one *R. pomonella* individual (PNY7) shares its sequence with four *R. mendax* individuals.

Using the neighbor-joining tree as a basis, we defined 15 alleles (genotypes) within the NTS data set. Construction of a minimum-spanning network (Fig. 6) provides a visual representation of the relationships of these 15 alleles and their relative abundance within each of the three ingroup species. Allelic diversity at the NTS locus was higher in *R. mendax* than it was in either *R. zephyria* or *R. mendax*. While *R. zephyria* and *R. pomonella* had three alleles each, *R. mendax* had nine, despite the fact that there were fewer *R. mendax* individuals in the sample (*n* = 26) than *R. zephyria* (*n* = 35) or *R. pomonella* (*n* = 38). All of the *R. zephyria* individuals
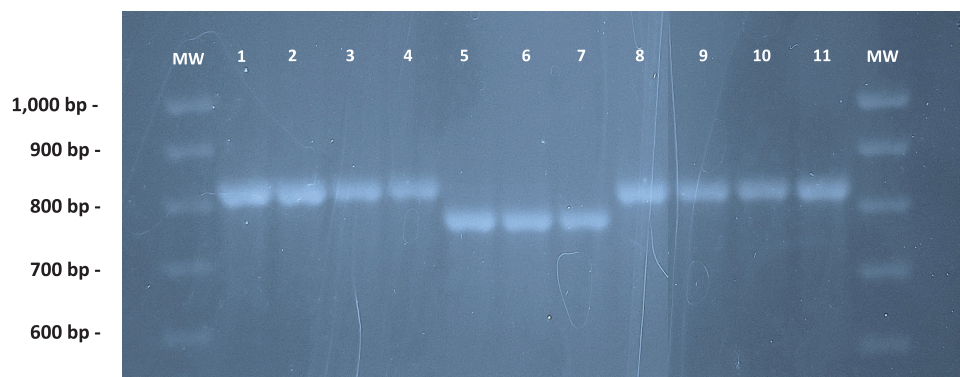

Fig. 3. Agarose gel showing the NTS size difference between the PCR products obtained from *R. pomonella*, *R. mendax*, and *R. zephyria* individuals. PCR products were obtained using primer pair PMZrDNA1F and PMZrDNA1R. Lanes: MW, 100 bp ladder; Lanes 1–4, PNY 12–15; Lanes 5–7, ZWI 12, 15, and 16; Lanes 8–11, MEL 11, 16–18. Agarose gel (2%) run for 4 hr at 80 V. Stained with SYBR Green.

had one of three alleles (Hap11, Hap12, Hap13), and each of these is found in *R. zephyria* individuals only. *R. pomonella* individuals also had one of three alleles, with *R. pomonella* and *R. mendax* sharing two of these alleles (Hap1 and Hap5). Hap1 was the most frequent allele in the data set, being found in 36 *R. pomonella* individuals and three *R. mendax* individuals.

Mitochondrial COI Sequences From the Geographic Sample

A different picture emerged when the geographic sample was analyzed using the mitochondrial COI barcode region. Mitochondrial COI sequences were obtained from the same set of individuals in the Geographic sample from which NTS sequences were obtained. Base calling and alignment of the sequences again were straightforward, resulting in 633 aligned nucleotide positions corresponding to the BOLD barcode region (Ratnasingham and Hebert 2007). The complete set of COI sequences is included in .fas format in Supp File 1b (online only). As was observed in the NTS data set, DNA sequence variation was slight within the COI data set, with observed

variation at 14 nucleotide positions within the ingroup, and 8 of these being phylogenetically informative. We included in the COI data set six *R. mendax* and *R. pomonella* COI sequences from the literature (Frey et al. 2013) and we observed COI sequences in the geographic sample that matched each of these. *Rhagoletis zephyria* is polyphyletic in the minimum evolution neighbor-joining COI tree with respect to both *R. pomonella* and *R. mendax* (Fig. 7). Three *R. zephyria* individuals (ZWI10, ZPA2, and ZPA10) appear in a cluster that contains predominantly *R. mendax* and *R. pomonella*, while two *R. pomonella* individuals (PNY5 and PWA19) are grouped with the majority of the *R. zephyria* individuals in the Geographic sample. While the neighbor-joining tree shows some apparent structure in the *R. pomonella*, *R. mendax*, and *R. zephyria* individuals included in the geographic sample, bootstrap support overall was very weak, with values >50 obtained for only two clusters, one containing three *R. mendax* individuals and the other containing two *R. pomonella* individuals.

Using the neighbor-joining tree as a basis, we defined 17 COI haplotypes. Construction of a minimum-spanning network (Fig. 8) shows extensive COI haplotype sharing between *R. pomonella*, *R. mendax*, and *R. zephyria*, with none of the three ingroup species appearing as a monophyletic group. While most of the *R. zephyria* individuals have 'private' haplotypes (Haplotypes 8, 11, or 12), some *R. zephyria* individuals share haplotypes with *R. pomonella* (Haplotype 7), *R. mendax* (Haplotype 2), or both (Haplotype 1). Unlike the NTS locus, where allele diversity was higher in *R. mendax* than it was in *R. zephyria* or *R. mendax.*, mitochondrial COI haplotype diversity is highest in *R. pomonella*, which has 10 haplotypes while *R. mendax* has six and *R. zephyria* has five. At the NTS locus, *R. pomonella* individuals only had one of three closely related alleles, while at the COI locus the ten haplotypes found in *R. pomonella* individuals are spread across the haplotype network.

Table 3. Insertion/deletion mutations in the NTS region that distinguish between *R. pomonella* and *R. zephyria*

Position(s)	Insertion/Deletion	# bp	Taxa
2–3	Deletion	2 bp	all *R. zephyria*
7	Deletion	1 bp	all *R. zephyria*
318–332	Deletion	15 bp	all *R. zephyria*
345–351	Deletion	7 bp	all *R. zephyria*
352–371	Insertion	20 bp	all *R. pomonella*
431–434	Insertion	4 bp	all *R. pomonella*

Position(s) refer to the 766 bp NTS alignment; Deletion or insertion is relative to the NTS sequences from *R. pomonella* Mexico EVTM.
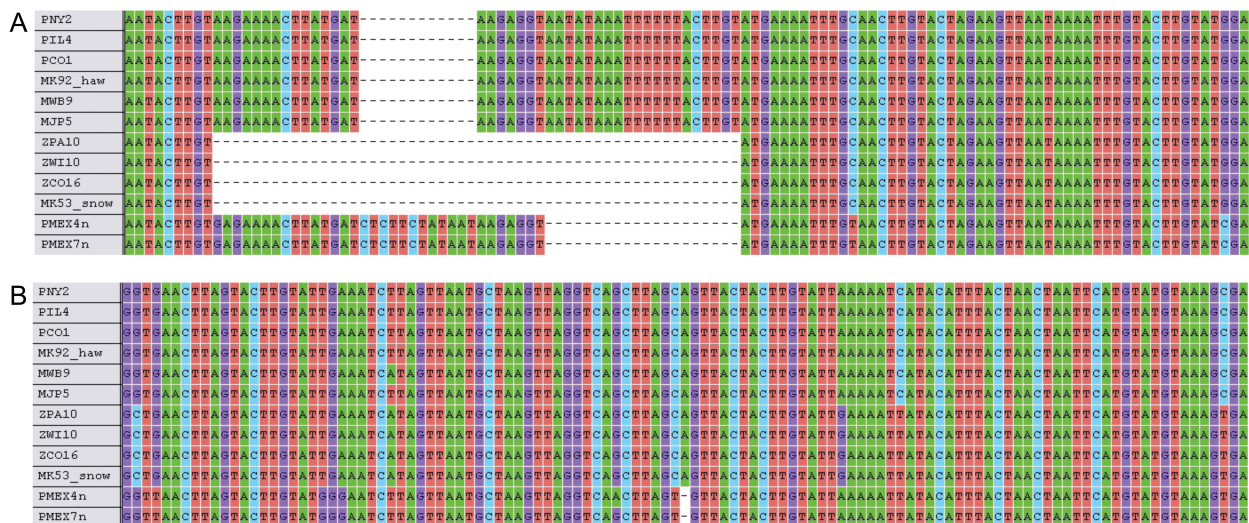

Fig. 4. Selected aligned NTS DNA sequences from the *R. pomonella*/ *R. mendax*/ *R. zephyria* taxon sets showing consistent differences between *R. pomonella* and *R. zephyria*. Shown are three NTS sequences from *R. pomonella* (PNY2, PIL4, PCO1), three from *R. zephyria*, (ZPA10, ZWI10, ZCO16), and two from *R. mendax* (MWB9, MJP5) from the Geographic sample, plus one sequence from hawthorn (MK92_haw) and one from snowberry (MK53_snow) from the Beacon Rock sample, and two sequences (PMEX4n, PMEX7n) from the *R. pom.* Mex. outgroup. A) Positions 309–423 in the 766 bp NTS alignment. This segment of the alignment shows a set of indels that appear to be diagnostic between *R. pomonella* and *R. zephyria* (*R. mendax* shares the *R. pomonella* indels in all instances). B) Positions 40–154 in the alignment. This part of the alignment has four single nucleotide substitution mutations (41, 52, 119, and 152) that appear to be derived (based on outgroup comparison) in *R. zephyria* (positions 41 and 113) or *R. pomonella* and *R. mendax* (positions 119 and 152). MEL 25 and MJP2 share the ancestral state at position 119, and MEL25, MJP2, MEL19, and MEL29 share the ancestral state at position 152, with all *R. zephyria*. Variation at position 67 is homoplasious.
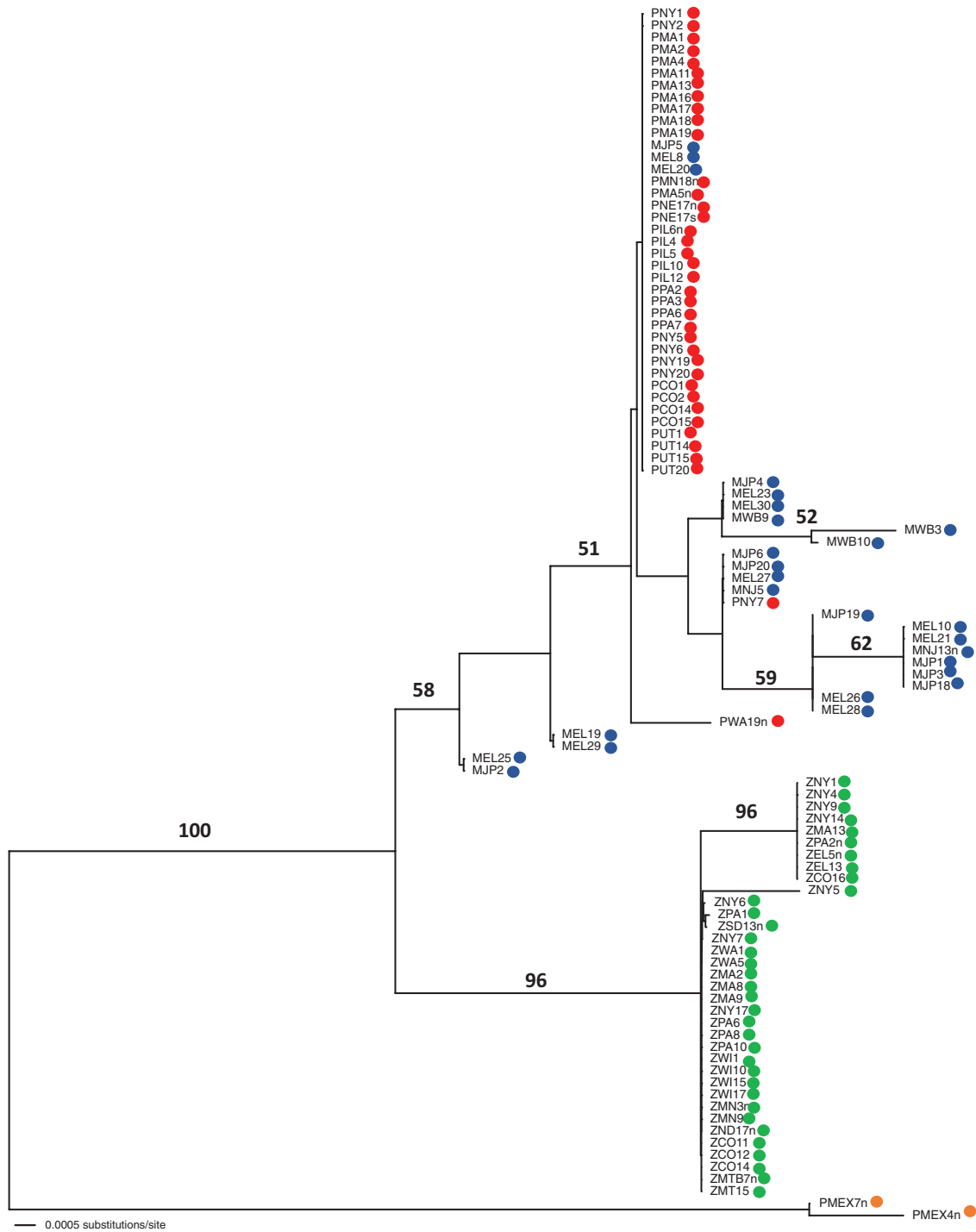
Fig. 5. Neighbor joining tree showing relationships of 766 bp NTS sequences obtained from *R. pomonella*, *R. zephyria*, and *R. mendax* individuals in the Geographic Sample. Sequences drawn from *R. pomonella*, "P"; *R. mendax*, "M"; *R. zephyria*, "Z"; and *R. pomonella* Mex. EVTM (PMex; outgroup), are shown. Numbers on branches are bootstrap values (1,000 replicates). Scale bar = *p*-distance.

NTS and COI Sequences From the Beacon Rock Sample

Incorporation into the analysis of a set of 27 individual *Rhagoletis* flies reared from black hawthorn and 17 individuals reared from snowberry from Beacon Rock State Park and St. Cloud Ranch (the Beacon Rock sample; Supp Table S1 [online only]) served as a preliminary 'field-test' of the ability of the NTS PCR product differences to serve as a diagnostic for identification purposes in areas where

R. pomonella and *R. zephyria* co-exist. All of the flies in the Beacon Rock sample were collected as larvae within infested fruit in 2000 by one of the authors (MK), sent to another of the authors (JJS), overwintered as pupae and reared to adults in the lab (with emerged adults randomly labeled to mask host of origin) and then sent back to the Washington State Department of Agriculture (WSDA) for species identification using morphological characteristics. Locality information and morphological scoring data for these flies is shown in
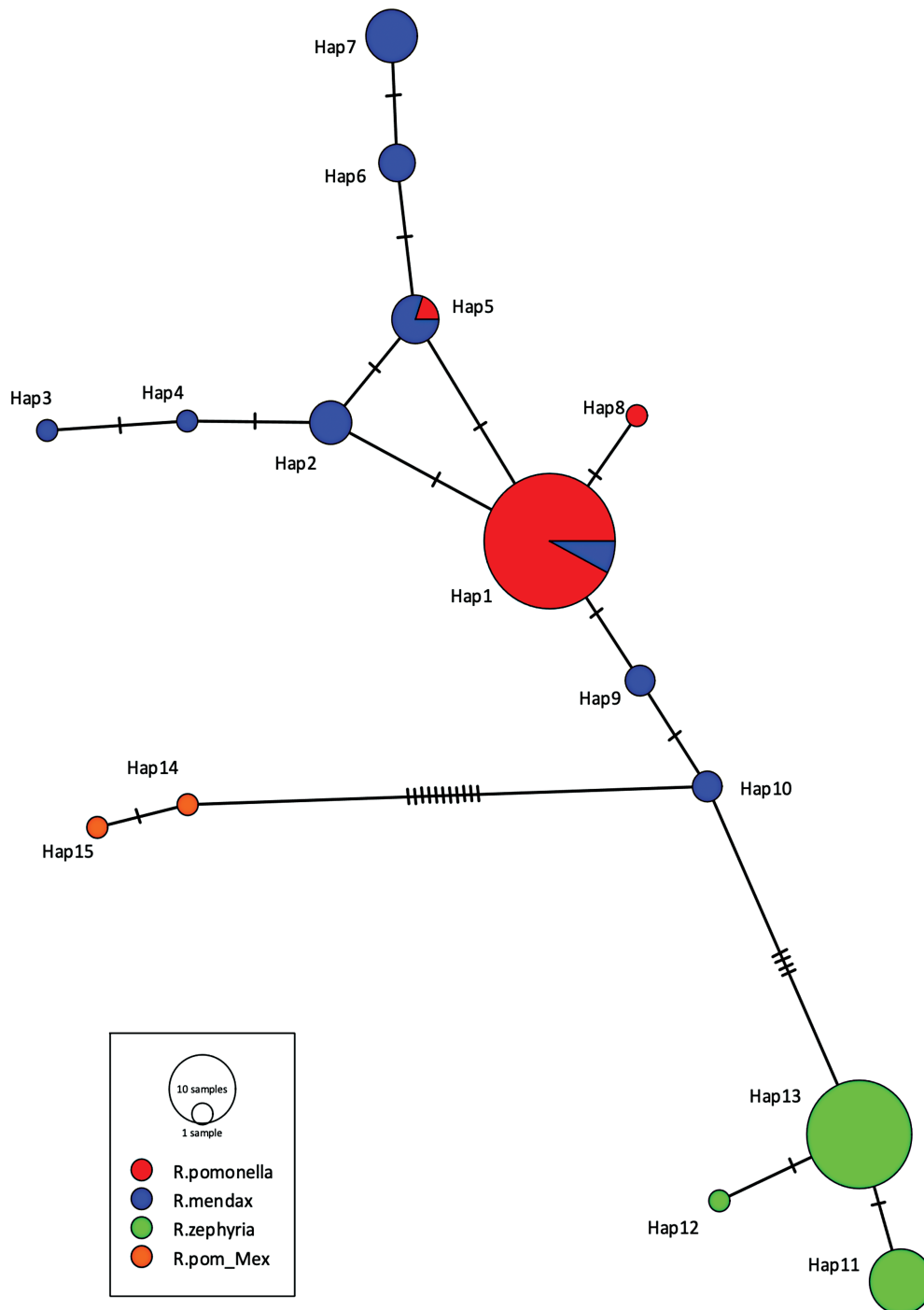
Fig. 6. Minimum spanning network showing relationships of rDNA NTS sequences drawn from *R. pomonella*, *R. mendax*, and *R. zephyria*. The snowberry fly, *R. zephyria*, is distinguished from *R. mendax* and *R. pomonella* by four nucleotide substitutions, while *R. pomonella* and *R. mendax* share alleles (1 and 5).

Supp Table S1 (online only). All 12 of the males reared from black hawthorn were identified as *R. pomonella* based on morphological criteria, while 6 of the 15 females reared from black hawthorn were identified as *R. pomonella* and the other 9 were identified as R. *zephyria* (based on ovipositor length). On the other hand, all 17 of the flies (9 females and 8 males) reared from snowberry in the Beacon Rock sample were identified as *R. zephyria* based on morphological criteria.

When we sequenced the PMZrDNA1F/PMZrDNA1R PCR product from these flies, all 26 NTS sequences obtained from

individuals from Beacon Rock that were reared from black hawthorn had the unique set of NTS indels and DNA substitutions (described above) that were associated with the *R. pomonella* flies in the geographic sample, while all 16 NTS sequences of individuals from Beacon Rock and St. Cloud Ranch that were reared from snowberry had the unique set of NTS indels and DNA substitutions that were associated with all of the *R. zephyria* flies in the geographic sample (Supp Figs. S1 and S2 [online only]). Thus, the 49 bp PCR-amplicon length difference that was observed between *R. pomonella* and *R. zephyria* individuals in the geographic sample was also observed
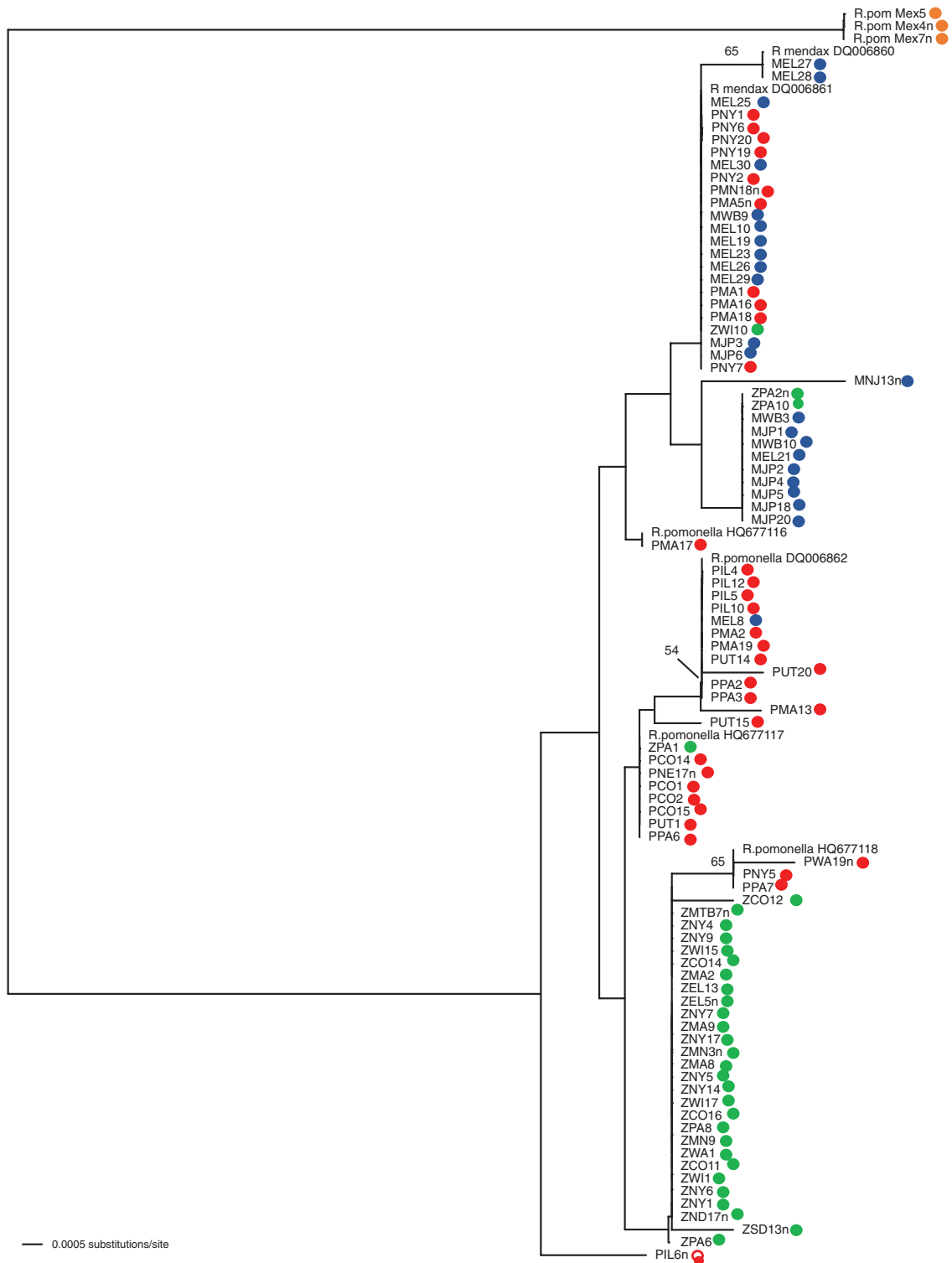
Fig. 7. Neighbor-joining tree showing relationships of 633 bp mitochondrial COI sequences obtained from *R. pomonella*, *R. zephyria*, and *R. mendax* individuals in the Geographic Sample. Sequences drawn from *R. pomonella* individuals, "P"; *R. mendax*, "M"; *R. zephyria*, "Z". *R. pomonella* Mex. EVTM is shown as the outgroup (*R. pom* Mex.). Numbers on branches are bootstrap values (500 replicates). Scale bar = *p*-distance.

in the flies reared from black hawthorns (*R. pomonella*) and flies reared from snowberries (*R. zephyria*) in the Beacon Rock sample.

A neighbor-joining tree with the combined Geographic sample and Beacon Rock sample is shown as Supp Fig. S1 (online only), and the NTS allele network including flies from both the Geographic sample and the Beacon Rock sample is shown as Supp Fig. S2 (online only). All of the individuals reared from snowberry in the Beacon Rock sample clustered with *R. zephyria* in the NTS NJ tree, while all of the individuals reared from black hawthorn in the Beacon Rock sample clustered with *R. pomonella* and *R. mendax* (Supp Fig. 1 [online only]). There were no unique NTS sequences (alleles) observed in the Beacon Rock sample (Supp Fig. 2 [online only]). Twenty-five
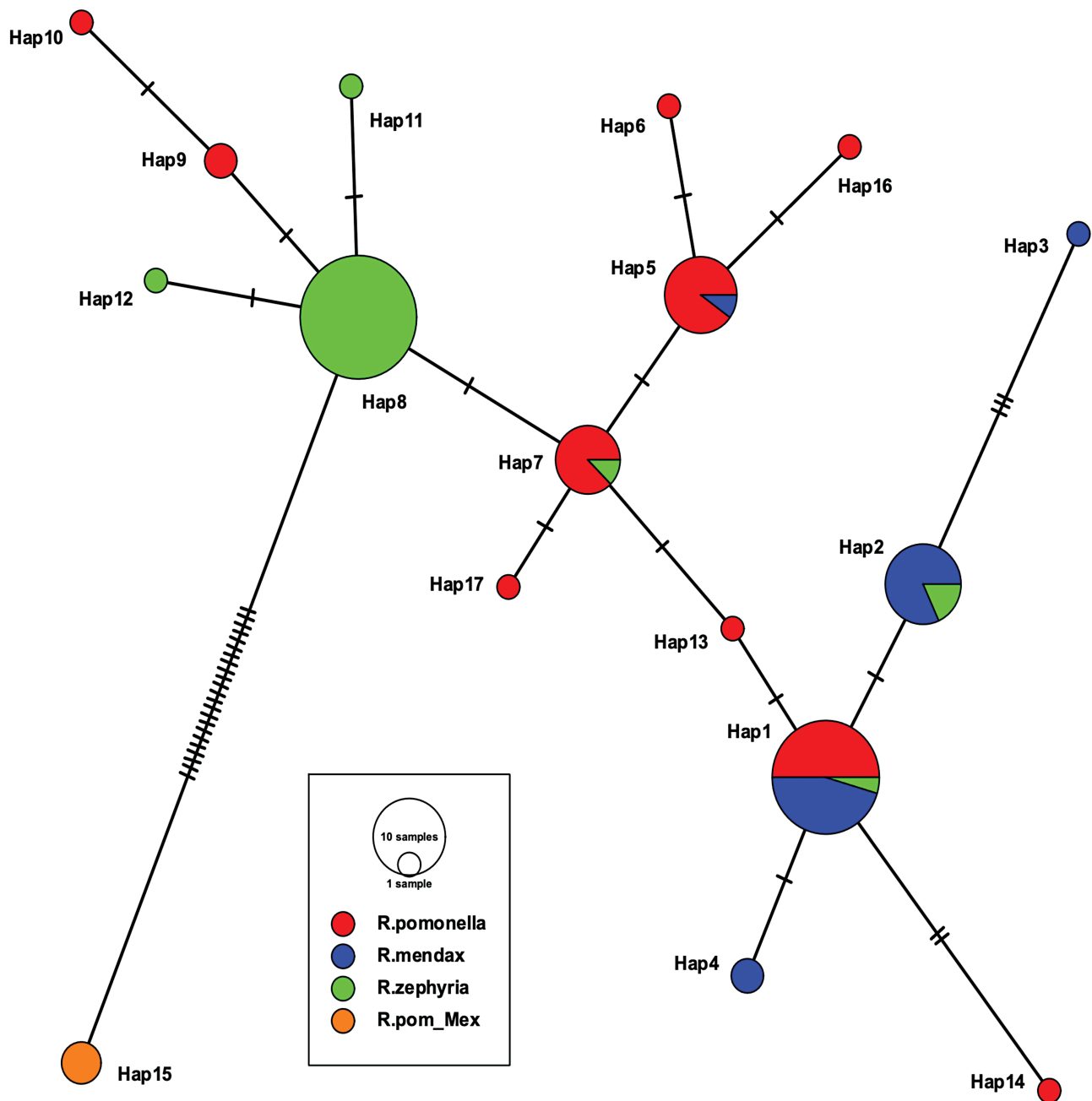
Fig. 8. Haplotype network showing relationships of mitochondrial COI sequences drawn from *R. pomonella, R. mendax,* and *R. zephyria* in the Geographic sample. Tick marks represent individual DNA substitution mutations between haplotypes. While most *R. zephyria* flies have Haplotype 8, which is private, other *R. zephyria* flies share haplotypes with both *R. pomonella* and *R. mendax* individuals elsewhere in the network.

of the 26 individuals reared from black hawthorn in the Beacon Rock sample carried NTS allele 1, while one individual (MK34) carried allele 9, which is shared with two *R. mendax* individuals (MEL19 and MEL29). Conversely, all of the 16 individuals reared from snowberry from Beacon Rock State Park and St. Cloud Ranch carried NTS allele 13, which was the main allele exclusively associated with *R. zephyria* in the geographic sample.

The COI neighbor-joining tree based on the combined Geographic and Beacon Rock samples is shown as Supp Fig. S3 (online only), and the COI haplotype network including flies from both the Geographic sample and the Beacon Rock sample is shown as Supp Fig. S4 (online only). No new COI sequences were observed when the Beacon

Rock sample was added to the analysis, and the neighbor-joining tree shows that the COI sequences from the flies reared from black hawthorn and snowberry were not strictly associated with COI sequences drawn from *R. pomonella* and *R. zephyria*, respectively (Supp Fig. S3 [online only]). While most Beacon Rock sample flies from snowberry (12/16) carried COI haplotype 8, MK97, which was reared from snowberry and identified as *R. zephyria* based on morphology, carried COI haplotype 1, which was predominantly associated with *R. pomonella* and *R. mendax* (ZWI10 also had COI haplotype 1). Similarly, Beacon Rock sample flies reared from black hawthorn carried four haplotypes. While most of these flies carried COI haplotype 1 (16/26), others showed COI haplotypes that were

more closely related to those carried mainly by *R. zephyria* flies in the geographic sample (e.g., haplotype 10; Supp Figs. S3 and S4 [online only]).

Discussion

NTS Sequences Distinguish *R. pomonella* From *R. zephyria*

All of the *R. zephyria* individuals analyzed in this study formed a monophyletic group based on DNA sequences of an 800 bp PCR product amplified from the nuclear rDNA NTS region (Supp Fig. 1 [online only]). This monophyletic group was supported unambiguously by three DNA substitutions in the 766 bp NTS alignment (positions 4, 41, and 113). In addition, *R. zephyria* was supported as a monophyletic group by a set of four unique insertion–deletion mutations within the alignment (Table 3), which serve as synapomorphies defining this clade, using *R. pomonella* Mex. EVTM as the outgroup, and appear to unambiguously distinguish *R. zephyria* from *R. pomonella* and *R. mendax*.

The indel structure occurring between positions 318 and 371 in the alignment contributes the bulk of the size difference of 49 nucleotides in the PCR products obtained from *R. zephyria* individuals compared to the PCR products obtained from *R. pomonella* and *R. mendax* individuals (Fig. 4). This DNA size fragment difference between *R. zephyria*, *R. pomonella*, and *R. mendax* individuals is readily apparent on agarose gels of the PCR products from these three different fly species (Fig. 2), and provides the potential for the development of a rapid, simple, inexpensive agarose gel-based test to use in efforts to distinguish between *R. zephyria* and *R. pomonella* in the apple-growing regions of the Pacific Northwest. A method based on these NTS differences would allow discrimination between *R. pomonella* and *R. zephyria* simply by scoring size differences between amplified PCR products via agarose gel electrophoresis.

The availability of the rDNA NTS region adjacent to the ETS as a molecular marker will complement and extend other tools available to distinguish *R. pomonella* from *R. zephyria* for quarantine purposes. Traditionally, adults obtained via trap catches in apple maggot surveys have been analyzed for morphological differences in ovipositor length and surstylus shape (Westcott 1982, Yee et al. 2011). The PCR assay described here based on size differences in the NTS PCR amplicon would be particularly useful for individuals in the 'grey area' in these analyses. The *R. pomonella* individuals and *R. zephyria* individuals from Beacon Rock State Park and St. Cloud Ranch analyzed in this paper were scored by morphology in the WSDA lab (Supp Table 1 [online only]), which resulted in scoring nine *R. pomonella* females as *R. zephyria* (ovipositor length below threshold) as judged by the NTS sequence data. While the advanced morphometric techniques now available based on wing shape and aculeus shape (Yee et al. 2009, 2011) would likely correctly identify these flies as *R. pomonella*, all of the flies reared from black hawthorn in the Beacon Rock sample carry the set of diagnostic DNA nucleotide substitutions and indel mutations characteristic of *R. pomonella* and not *R. zephyria*.

Green et al. (2013), and more recently Doellman et al. (2020), described SNP-based diagnostic tools that would also identify all individuals in this study to their correct species. The method described by Doellman et al. (2020) is based on PCR-amplification and restriction digestion of PCR-amplified products at five nuclear loci. While accurate, the method is potentially more time and labor intensive than the assay described here. Agarose gel separation of DNA fragments following PCR amplifications is standard, and the possibility

of errors associated with a restriction digestion step would be avoided in an NTS-based assay. No DNA sequencing would be required for distinguishing between *R. pomonella* and *R. zephyria*, and the known sequence differences between the two species provide opportunity for the development of a qPCR assay.

While successful PCR-amplification is required for the NTS assay to distinguish between *R. zephyria*, *R. pomonella*, and *R. mendax*, no further processing is necessary beyond running and visualizing the agarose gel. The primers designed for use in the current study (PMZrDNA1F and PMZrDNA1R), and the PCR conditions used, appear to be specific with respect to amplification of NTS only from *R. zephyria*, *R. pomonella*, and *R. mendax*. However, one shortcoming of the current study is that the specificity of these primers beyond *R. zephyria*, *R. pomonella*, and *R. mendax* has not been determined. There is one mismatch between the sequence of the forward primer PMZrDNA1F and the corresponding sequence in *R. pomonella* Mex. EVTM, and either four mismatches (*R. tabellaria* (Fitch) and *R. electromorpha* Berlocher) or five (*R. persimilis* Bush) in PMZrDNA1F in *Rhagoletis tabellaria* species group flies (Supp Table 2 [online only]), which is a sister group to the *R. pomonella* species group within the genus *Rhagoletis* (Hulbert 2018). Sequences corresponding to the PMZrDNA1F primer have not yet been found in either *R. cornivora* Bush or *R. bushi* Hulbert and Smith. Whether the PMZrDNA1F and PMZrDNA1R primer pair can be used to amplify the NTS region in *R. pomonella* Mex. EVTM, *R. cornivora*, and the four *R. tabellaria* group species, and under what conditions, remain to be empirically determined.

NTS as a Molecular Marker

One significant aspect of the current study is the possible use of the NTS region of the ribosomal DNA just upstream from the ETS to explore systematics and evolutionary questions in other insect groups. The NTS region may prove to be particularly useful in circumstances like *R. pomonella* and *R. zephyria*, where closely related and/or cryptic species present taxonomic problems, and positive species ID is necessary for pest regulatory purposes. However, the usefulness of NTS as a marker may be limited by the history of gene flow between the species examined. While NTS appears to have fixed differences between *R. zephyria* and both *R. pomonella* and *R. mendax*, the NTS region is not diagnostic for these latter two sister species, based either on indel variation or DNA substitutions. Until the introduction of *R. pomonella* into western North America, *R. zephyria* had been geographically separated from *R. pomonella* except for along a narrow band of overlap from Minnesota into the eastern U.S. (Gavrilovic et al. 2007, Arcella et al. 2015). This historical lack of gene flow may have been necessary for the NTS divergence observed in this study to have occurred. However, there has been some gene flow between *R. mendax* and *R. zephyria* where they overlap in eastern North America (Schwarz et al. 2005), and NTS remained diagnostic in eastern North American samples of those species examined here. On the other hand, *R. mendax* and *R. pomonella* are broadly sympatric in eastern North America, with ample opportunities for gene flow between them, and the NTS region is not diagnostic for these two sister species, either based on indel variation or DNA substitutions.

Nonetheless, analysis of NTS regions may be useful for addressing questions in instances where there are shared mitochondrial DNA haplotypes across species boundaries (e.g., the mealybugs *Planococcus citri* Russo (Hemiptera: Pseudococcidae) and *P. minor* (Maskell), Rung et al. 2009; several species of *Neodiprion* sawflies (Hymenoptera: Diprionidae), Linnen and Farrell 2007; the blow

flies *Lucilia coeruleiviridis* Macquart (Diptera: Calliphoridae) and *Lucilia Mexicana* Macquart, Debry et al. 2013), and in insect taxa where two or more divergent mitochondrial lineages exist within a single species (e.g., the tree weta *Hemideina crassidens* (Blanchard) (Orthoptera: Anostostomatidae), Morgan-Richards et al. 2017).

While the 18S and 28S ribosomal RNA structural genes, and the internal transcribed spacers (ITS1 and ITS2), of the ribosomal DNA (rDNA) cistron have been used extensively for phylogenetics and population-level studies (the review of Hillis and Dixon (1991) has been cited over 2000 times), the external transcribed spacer (ETS) and nontranscribed spacer (NTS) regions, which together comprise the ribosomal intergenic spacer (IGS), have been relatively underutilized in systematics and species identification in insects, and exploration of these uses may prove fruitful. Shaikevich et al. (2013) described IGS sequences from mosquitoes (*Culex* spp.; Diptera: Culicidae) and indicated that these might prove useful for population genetics and phylogenetic analyses, but this study has been cited only once in the literature since publication (Web of Science; accessed 22 June 2021). On the other hand, the NTS and ETS gene regions seem to have found some use in phylogenetics studies, mainly in plants, with most of these studies employing the ETS region (e.g., Linder et al. 2000, Valtuena et al. 2017).

For reasons that are not entirely clear, the work of Shaikevich et al. (2013), who showed that the IGS regions of *Culex* species provided promising molecular markers for population genetics studies and phylogenetic analysis, has not seen broad applicability. This may be, in part, because the ribosomal IGS regions are not as easy to use as the 18S, 28S, ITS1, and ITS2 sequences, which are readily accessible and easy to amplify via PCR, and have proven to be successful in a range of phylogenetics studies (e.g., Si et al. 2020). In addition, because they are not transcribed, the NTS region will not be included in results obtained using RNASeq protocols. Our work allowed us to find the NTS regions by virtue of the use of whole genome NextGen sequencing, originally targeting high coverage sequences to assemble nearly complete mtDNA sequences (Smith et al. in preparation). The IGS regions also appear to be fast evolving (Stage and Eickbush 2007), so they are not readily obtained via homology searches. The DNA sequences of the NTS regions analyzed in this study from *R. pomonella*, *R. mendax*, and *R. zephyria* are apparently unique: BLAST searches of GenBank using the NTS PCR product sequences did not return any matches. Nonetheless, we were able to align our NTS sequences, at least partially, to the NTS sequence from *Bactrocera tryoni* (Froggatt) (Diptera: Tephritidae) (Gilchrist et al. 2014) and used this as the basis for hypothesizing the position of the ribosomal binding site.

The data presented here also illustrate why mtDNA has not proven useful for species diagnostics in *R. pomonella* and *R. zephyria*, as there are shared COI mtDNA haplotypes between these two species (Fig. 8). One possible explanation for this observation is that we simply have not sequenced enough mtDNA characters; the dataset here includes only 633 of the mitochondrion's approximately 15,000 bp. However, NextGen sequences of complete coding regions of the mtDNA (37 genes, 14,942 bp) from five *R. pomonella*, six *R. zephyria*, and one *R. mendax* individuals included in this study (GenBank Accession #'s MN443934–MN443945) did not provide species-level resolution (Smith et al. in preparation); instead, we observed what appears to be cytonuclear discordance between *R. zephyria* and *R. mendax* in the eastern U.S. This discordance presumably arises via introgression of *R. mendax* mtDNA into *R. zephyria*, which was described by Schwarz et al. (2005) in their report of the 'speciation by hybridization' of an (undescribed) *Rhagoletis* species, the *Lonicera* fly. The extent to which these introgressions have occurred (and are occurring) in *R. pomonella*,

R. mendax, and *R. zephyria* is unclear and further work remains to be done to provide a clearer picture of these dynamics.

Future Work

Another shortcoming of the current study is that we have not determined the extent of fixation of the IGS differences in a large sample such as was employed by Doellman et al. (2020). In our study, we analyzed 38 *R. pomonella*, 26 *R. mendax*, and 35 *R. zephyria* in the geographic sample, and an additional 27 *R. pomonella* and 17 *R. zephyria* in the Beacon Rock sample. We also have not yet analyzed F1 hybrids between *R. pomonella* and *R. zephyria* (Yee and Goughnour 2011, Yee et al. 2013), or backcross individuals. Work to expand the sample size and experiments to determine the outcomes given hybrid and introgressed individuals remains to be carried out. Results of such studies may inform questions about the inheritance of rDNA cistrons and the homogenization of the different regions within them.

We also did not take into consideration or try to determine the nature of the variation leading to the dual NTS sequence signals observed at some nucleotide positions in some *R. mendax* sequences. It is unclear whether these dual signals occurred because of variation within the multiple copies of the ribosomal DNA cistron within the same haploid genome or if there are differences at some positions due to heterozygosity. In either case, we did not observe any differences in the indel patterns within any individual sequences, which would have been observable as sequences that fell out of register from each other at a certain point in a particular read. Because *R. mendax* was not a major focus of this study, and because the sequences in question did not impact the finding that *R. zephyria* is diagnosable based on its indel pattern, we did not pursue these lines of inquiry here and leave them to a future study.

Supplementary Data

Supplementary data are available at *Journal of Economic Entomology* online.

Acknowledgments

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.

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.

Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16: 37–48.

Barr, N. B., D. Garza, L. A. Ledezma, and D. A. Salinas. 2021. Using the rDNA internal transcribed spacer 1 to identify the invasive pest *Rhagoletis cerasi* (Diptera: Tephritidae) in North America. J. Econ. Entomol. 114: 360–370.

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

Bush, G. L. 1969. Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). Evolution. 23: 237–251.

Bush, M. R., M. Klaus, A. Antonelli, and C. Daniels. 2002. Protecting backyard apple trees from apple Maggot. Washington State University Extension Bulletin #1928, Pullman, Washington.

CFIA. 2017. RMD-16-01: pest risk management document—determining the regulatory status of *Rhagoletis pomonella* (apple maggot). Government of Canada, Canadian Food Inspection Agency, Ottawa, ON, Canada.

Chang, J. M., P. Di Tommaso, V. Lefort, O. Gascuel, and C. Notredame. 2015. TCS: a web server for multiple sequence alignment evaluation and phylogenetic reconstruction. Nucleic Acids Res. 43: W3–W6.

Curran, C. H. 1924. *Rhagoletis symphoricarpi*, a new trypaneid from British Columbia (Dipt.). Can. Ent. 61: 62–63.

DeBry, R. W., A. Timm, E. S. Wong, T. Stamper, C. Cookman, and G. A. Dahlem. 2013. DNA-based identification of forensically important *Lucilia* (Diptera: Calliphoridae) in the continental United States. J. Forensic Sci. 58: 73–78.

Doellman, M. M., G. J. Ragland, G. R. Hood, P. J. Meyers, S. P. Egan, T. H. Q. Powell, P. Lazorchak, M. M. Glover, C. Tait, H. Schuler, et al. 2018. Genomic differentiation during speciation-with-gene-flow: comparing geographic and host-related variation in divergent life history adaptation in *Rhagoletis pomonella*. Genes. 9: 262.

Doellman, M. M., S. P. Egan, G. J. Ragland, P. J. Meyers, G. R. Hood, T. H. Q. Powell, P. Lazorchak, D. A. Hahn, S. H. Berlocher, P. Nosil, et al. 2019. Standing geographic variation in eclosion time and the genomics of host race formation in *Rhagoletis pomonella* fruit flies. Ecol. Evol. 9: 393–409.

Doellman, M. M., G. R. Hood, J. Gersfeld, A. Driscoe, C. C. Y. Xu, R. N. Sheehy, N. Holmes, W. L. Yee, J. L. Feder. 2020. Identifying diagnostic genetic markers for a cryptic invasive agricultural pest: a test case using the apple maggot fly (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 113: 246–256. https://doi-org.proxy2.cl.msu.edu/10.1093/aesa/saz069.

Dowell, R. N., L. T. Wilson, and V. P. Jones, (eds). 1990. Apple Maggot in the west: history, biology, and control. University of California (System). Division of Agriculture and Natural Resources, Publ. 3341.

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.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial *cytochrome c oxidase* I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3: 294–299.

Frey, J. E., L. Guillén, B. Frey, J. Samietz, J. Rull, and M. Aluja. 2013. Developing diagnostic SNP panels for the identification of true fruit flies (Diptera: Tephritidae) within the limits of COI-based species delimitation. BMC Evol. Biol. 13: 106.

Galinato, S. P., R. K. Gallardo, D. M. Granatstein, and M. Willett. 2018. Economic impact of a potential expansion of pest infestation: Apple Maggot in Washington State. HortTechnology. 28: 651–659.

Gavrilovic, V., G. L. Bush, D. Schwarz, and J. J. Smith. 2007. *Rhagoletis zephyria* Snow (Diptera: Tephritidae) in the Great Lakes basin: a native insect on native hosts? Ann. Entomol. Soc. Am. 100: 474–482.

Gilchrist, A. S., D. C. Shearman, M. Frommer, K. A. Raphael, N. P. Deshpande, M. R. Wilkins, W. B. Sherwin, and J. A. Sved. 2014. The draft genome of the pest tephritid fruit fly *Bactrocera tryoni*: resources for the genomic analysis of hybridising species. BMC Genomics. 15: 1153.

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.

Hebert, P. D., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. U. S. A. 101: 14812–14817.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. 66: 411–453.

Hulbert, D. L. 2018. *Systematics of the genus* Rhagoletis *(Diptera: Tephritidae): New species, phylogeny, and justifications*. Unpublished Ph. D. Dissertation, Michigan State University, East Lansing, MI.

Hulbert, D., M. D. Jackson, G. R. Hood, and J. J. Smith. 2018. Description of a new *Rhagoletis* (Diptera: Tephritidae) species in the *tabellaria* species group. Insect Syst. Divers. 2:1–14.

Hwang, U. W., and W. Kim. 1999. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. Korean J. Parasitol. 37: 215–228.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger data sets. Mol. Biol. Evol. 33: 1870–1874.

Linder, C. R., L. R. Goertzen, B. V. Heuvel, J. Francisco-Ortega, and R. K. Jansen. 2000. The complete external transcribed spacer of 18S-26S rDNA: amplification and phylogenetic utility at low taxonomic levels in asteraceae and closely allied families. Mol. Phylogenet. Evol. 14: 285–303.

Linnen, C. R., and B. D. Farrell. 2007. Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. Evolution. 61: 1417–1438.

Mattson, M., G. R. Hood, W. L. Yee, M. M. Doellman, D. J. Bruzzese, R. B. Goughnour, A. L. Driscoe, S. Van Dexter, C. Tait, M. M. Glover, et al. 2022. Recursive adaptation in action: allochronic isolation and divergence of host-associated populations of Rhagoletis pomonella following its recent introduction to western USA. Entomologia Experimentalis et Applicata 170: 48–63.

McPheron, B. A. 1990a. Genetic structure of apple maggot fly (Diptera: Tephritidae) populations. Ann. Entomol. Soc. Am. 83: 568–577.

McPheron, B. A. 1990b. Implications of genetic variation in western apple maggots for understanding population biology. Ch. 3 In R. N. Dowell, L. T. Wilson, V. P. Jones (eds.), Apple Maggot in the west: history, biology, and control, University of California (System). Division of Agriculture and Natural Resources Publ. 3341. University of California. 95 pp.

Michel, A. P., J. Rull, M. Aluja, and J. L. Feder. 2007. The genetic structure of hawthorn-infesting *Rhagoletis pomonella* populations in Mexico: implications for sympatric host race formation. Mol. Ecol. 16: 2867–2878.

Morgan-Richards, M., M. Bulgarella, L. Sivyer, E. J. Dowle, M. Hale, N. E. McKean, and S. A. Trewick. 2017. Explaining large mitochondrial sequence differences within a population sample. R. Soc. Open Sci. 4: 170730.

Nei, M. 1991. Relative efficiencies of different tree-making methods for molecular data, pp. 90–128 In M. M. Miyamoto, and J. Cracraft (eds). Phylogenetic analysis of DNA sequences, Oxford University Press, New York, NY.

Payne, J. A., and S. H. Berlocher. 1995. Distribution and host plants of the blueberry maggot, *Rhagoletis mendax* (Diptera: Tephritidae) in southeastern North America. J. Kans. Entomol. Soc. 68: 133–142.

Powell, T. H., A. A. Forbes, G. R. Hood, and J. L. Feder. 2014. Ecological adaptation and reproductive isolation in sympatry: genetic and phenotypic evidence for native host races of *Rhagoletis pomonella*. Mol. Ecol. 23: 688–704.

Ratnasingham, S., and P. D. Hebert. 2007. bold: the barcode of life data system (http://www.barcodinglife.org). Mol. Ecol. Notes. 7: 355–364.

Rung, A., D. R. Miller, and S. J. Scheffer. 2009. Polymerase chain reaction-restriction fragment length polymorphism method to distinguish three mealybug groups within the *Planococcus citri–P. minor* species complex (Hemiptera: Coccoidea: Pseudococcidae). J. Econ. Entomol. 102: 8–12.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425.

Schwarz, D., B. M. Matta, N. L. Shakir-Botteri, and B. A. McPheron. 2005. Host shift to an invasive plant triggers rapid animal hybrid speciation. Nature. 436: 546–549.

Shaikevich, E. V., M. V. Zagoskin, and D. V. Mukha. 2013. Comparative characteristics of the intergenic spacer of the ribosomal RNA gene cluster in mosquitoes of the genus *Culex* (Diptera: Culicidae). Mol. Biol. 47: 364–372.

Si, C., K. Chen, R. Tao, C. Su, J. Ma, L. Li, J. Hao, and Q. Yang. 2020. Genetic differentiation and divergence time of Chinese Parnassius (Lepidoptera: Papilionidae) species based on nuclear internal transcribed spacer (ITS) sequence data. J. Entomol. Sci. 55: 520–546.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87: 651–701.

Smith, J. J., and G. L. Bush. 1997. Phylogeny of the genus *Rhagoletis* (Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase II. Mol. Phylogenet. Evol. 7: 33–43.

Smith, J. J., V. Gavrilovic, and D. R. Smitley. 2001. Native *Vaccinium* spp. infested by *Rhagoletis mendax* (Blueberry maggot; Diptera: Tephritidae) in the Great Lakes region: a potential source of inoculum for infestation of cultivated blueberries. J. Econ. Entomol. 94: 1378–1385.

Smith, M. A., D. M. Wood, D. H. Janzen, W. Hallwachs, and P. D. Hebert. 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proc. Natl. Acad. Sci. U. S. A. 104: 4967–4972.

Smith, M. A., C. Bertrand, K. Crosby, E. S. Eveleigh, J. Fernandez-Triana, B. L. Fisher, J. Gibbs, M. Hajibabaei, W. Hallwachs, *et al.* 2012. Wolbachia and DNA barcoding insects: patterns, potential, and problems. PLoS One. 7: e36514.

Stage, D. E., and T. H. Eickbush. 2007. Sequence variation within the rRNA gene loci of 12 Drosophila species. Gen. Res. 17: 1888–1897.

Swofford, D. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.

Valtueña, F. J., T. Rodriguez-Riaño, J. Lopez, C. Mayo, and A. Ortega-Olivencia. 2017. Peripatric speciation in an endemic Macaronesian plant after recent divergence from a widespread relative. PLoS One. 12: e0178459.

Wasbauer, M. S. 1963. Taxonomic discrimination of *Rhagoletis pomonella* from the eastern United States and *Rhagoletis zephyria* from California. Calif. Dept. Agric. T-9: 1–6.

Westcott, R. L. 1982. Differentiating adults of apple maggot, *Rhagoletis pomonella* (Walsh) from snowberry maggot, *Rhagoletis zephyria* Snow (Diptera: Tephritidae) in Oregon. Pan-Pac. Entomol. 58: 25–30.

WSDA. 2001. Washington Administrative Code 16-470-108. Distribution of infested or damaged fruit is prohibited. Washington State Department of Agriculture. Available from: https://apps.leg.wa.gov/WAC/default.aspx?cite=16-470-108. Accessed 29 December 2021.

Xie, X., J. Rull, A. P. Michel, A. Velez, A. A. Forbes, N. F. Lobo, M. Aluja, and J. L. Feder. 2007. Hawthorn-infesting populations of Rhagoletis pomonella in Mexico and speciation mode plurality. Evolution (N. Y) 61: 1091–1105.

Yee, W. L., and R. B. Goughnour. 2011. Mating frequencies and production of hybrids by *Rhagoletis pomonella* and *R. zephyria* (Diptera: Tephritidae) in the laboratory. Can. Ent. 143: 82–90.

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.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. L. Madden. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinform. 13: 134.

Yee, W. L., P. S. Chapman, and H. D. Sheets. 2013. Comparative body size and shape analyses of F1 hybrid *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae). Ann. Entomol. Soc. Am. 106: 410–423.