Insect Systematics and Diversity, (2018) 2(2): 1; 1–9
doi: 10.1093/isd/ixy002
Research

Evolution

Phylogeography of Walnut-Infesting *Rhagoletis suavis* (Diptera: Tephritidae) Flies

Mary M. Glover,[1,5] Scott P. Egan,[2] Glen R. Hood,[2] Juan Rull,[3] Martin Aluja,[4] and Jeffrey L. Feder[1]

[1]Department of Biological Sciences, University of Notre Dame, Galvin Life Science Bld., Notre Dame, IN 46530 [2]Department of BioSciences, Rice University, Houston, TX 77005, [3]PROIMI Biotecnología-CONICET, LIEMEN-División Control Biológico de Plagas, Avenue Belgrano y Pje. Caseros, T4001MVB San Miguel de Tucumán, Tucumán, Argentina, [4]Instituto de Ecología A.C, Carretera Antigua a Coatepec no. 351, Col. El Haya, Xalapa, Veracruz 91070, México, and [5]Corresponding author, e-mail: mglover@nd.edu

Subject Editor: Patrick Mardulyn

Received 22 September 2017; Editorial decision 25 January 2018

Abstract

Discerning the biogeography and historical ranges of organisms is important to understanding the processes causing population divergence and speciation. Mountainous regions in North America have contributed to widespread divergence within animals and plants as species become geographically isolated and diverge. Here, we investigate patterns of divergence for six species of walnut-infesting flies in the *Rhagoletis suavis* species group (Diptera: Tephritidae) in the United States and Mexico based on sequence analysis of mitochondrial DNA (mtDNA) encoded Cytochrome Oxidase I and II genes (COI and COII). We resolved the relationship of the newly described species, *Rhagoletis ramosae*, found in the highlands of Mexico, within the *R. suavis* species group to test alternate hypotheses of migration and divergence. Bayesian phylogenetic analysis supported a clade in which *R. ramosae* was most closely related to *Rhagoletis zoqui/Rhagoletis completa*, found in eastern Mexico. This implies that the Sierra Madre Oriental and not the Sierra Madre Occidental have been a major conduit of migration, isolation, and speciation for walnut flies between Mexico and the United States. Comparisons of mtDNA divergence for *R. suavis* group flies with the *Rhagoletis pomonella* and *Rhagoletis cingulata* species groups suggested that despite current similarities in geographic distributions, these taxa do not share a common biogeographic history, diverging in different regions at different times in the past. Patterns displayed by *Rhagoletis* flies can be compared to patterns seen in other organisms through the Southwestern and Eastern United States, and Mexico to develop a fuller understanding of the biogeography of these regions of North America.

Key words: speciation, phylogenetics, biogeography, mtDNA

Allopatry, the geographic separation of populations (Feder 2017), is often an important factor facilitating population divergence leading to speciation in many animals and plants (Mayr 1963, Coyne and Orr 2004, Nosil 2012). When populations reside in different areas with little or no gene flow between them, they can accumulate genetic differences independent from one another. Given enough time, the isolated populations can eventually accrue enough differences to become reproductively isolated and diverge into separate species. Understanding the geographic distribution of species through time and space (biogeography) is therefore a key element to discerning the processes giving rise to new biodiversity (Klicka and Zink 1997, Avise and Walker 1998, Avise et al. 1998). Knowing the historic distributions of taxa can also be important for discerning how climate and ecology interact with spatial isolation and local adaptation to affect dispersal and population divergence (Lieberman 2003).

The genus *Rhagoletis* (Diptera: Tephritidae) contains a number of different species groups native to North America that have contributed to our understanding of the roles that geography and ecology can play in the speciation process (Bush 1966, Bush and Smith 1998, Feder et al. 2003). The *Rhagoletis suavis* species group includes six species (*Rhagoletis suavis* (Loew), *Rhagoletis juglandis* (Cresson), *Rhagoletis completa* (Cresson), *Rhagoletis boycei* (Cresson), *Rhagoletis zoqui* (Bush), and *Rhagoletis ramosae* (Hernández-Ortiz)) (Fig. 1A), which are currently largely geographically isolated from one another in the United States and Mexico (Fig. 1B). *R. suavis* flies exploit their host plants by infesting the husks of walnuts (*Juglans* spp. (Fagales: Juglandaceae)) (Fig. 1C) and are thought to have undergone allopatric speciation (Bush 1966), which is different from most *Rhagoletis* species groups that overlap geographically. Walnut flies also differ from other *Rhagoletis* species groups in that the species are morphologically

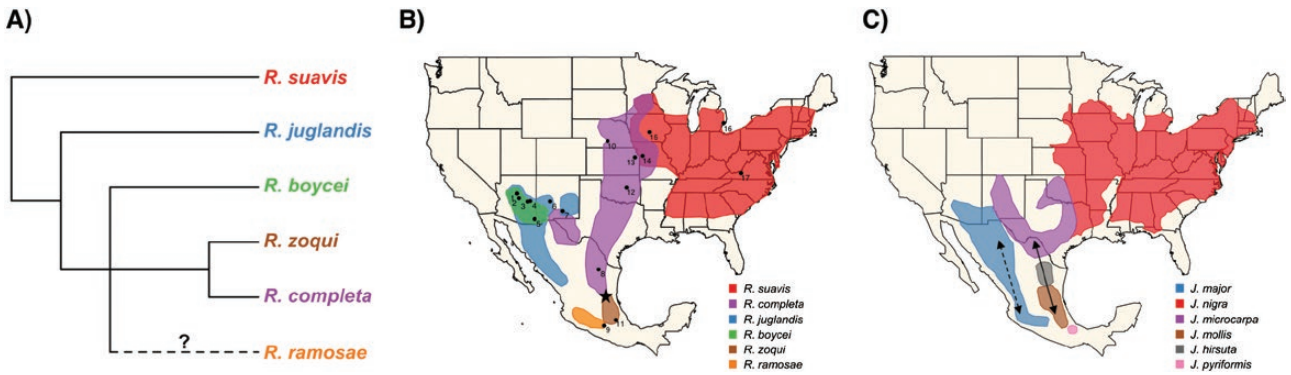

Fig. 1. Phylogenetic relationships of *R. suavis* group flies, their geographic distributions, and the ranges of their walnut host trees in North America. (A) mtDNA cladogram of the *R. suavis* group showing relationships resolved from previous studies (Smith and Bush 1997, Bush and Smith 1998). The relationship of *R. ramosae* in the earlier studies is depicted as unknown by the dotted line and question mark (B) Species distributions of walnut-infesting *Rhagoletis* in the United States and Mexico. Site locations for flies sequenced for COI and COII are distributed across distributions of each species, and numerical designations for sites analyzed in the study are shown (Supplementary Table S2). The hybrid zone between *R. completa* and *R. zoqui* found in Acahuales, Tamaulipas, Mexico is indicated by a star; (C) Species distributions of walnuts (*Juglans* spp.). We used mtDNA sequencing to determine the primary connectivity of walnut flies with alternative hypotheses that divergence occurred to the east along the Sierra Madre Oriental Mountains (solid arrow = SMOr) or in the west through the Sierra Madre Occidental Mountains (dotted arrow = SMOc).

distinct from one another in wing pattern and body coloration and display different courtship behaviors, which is thought to be associated with sexual selection (Bush 1966, Bush and Smith 1998). Some walnut fly species currently overlap with one another in a portion of their geographic ranges, likely the result of secondary contact (Bush 1966, Bush and Smith 1998). In the southwestern United States, *R. juglandis* co-occurs in part of its distribution with *R. boycei*, while in the east it co-exists with *R. completa* (Fig. 1B). In the midwestern United States, *R. completa* and *R. suavis* partially overlap, while in Mexico *R. completa* and *R. zoqui* co-occur in a narrow region in the area of Acahuales. A hybrid zone has been shown to exist between two species, *R. completa* and *R. zoqui*, in Acahuales, Mexico (see star in Fig. 1B; Rull et al. 2012). No other hybrid zone has yet been found between any other pair of overlapping *R. suavis* species in nature.

Except for *R. ramosae*, the phylogenetic relationships of *R. suavis* group flies in the United States and Mexico have been determined based on mitochondrial DNA (mtDNA) (Fig. 1A) (Smith and Bush 1997, Rull et al. 2013). *R. ramosae* is native to the Eje Volcánico Trans Mexicano (the central Altiplano or trans-Mexican volcanic belt; abbreviated EVTM hereafter) (Rull et al. 2013). The EVTM in the highlands of Mexico is a hotspot of biodiversity because of the mountainous terrain and volcanic regions that separate populations (Knowles 2001, Bryson et al. 2011a , Domínguez-Domínguez et al. 2011). Pleistocene glacial–interglacial climatic cycles, and the resultant expansions and contractions of the Mexican pine–oak forest, have driven widespread divergence within plants (Luna Vega et al. 1999, Jaramillo-Correa et al. 2008), scarabs (Lobo and Halffter 2000), reptiles, amphibians (Bryson et al. 2011b), and mammals (Ceballos et al. 2010).

From previous mtDNA phylogenetic analysis, two scenarios were proposed by Rull et al. (2013) regarding *R. ramosae* (Fig. 1C). The first hypothesis is that *R. ramosae* is most closely related to *R. boycei* or *R. juglandis* in the southwestern United States. If this hypothesis is true, then an important axis of past gene flow and current geographic isolation for walnut flies occurred in is the Sierra Madre Occidental Mountains (SMOc) in western Mexico, resulting in the divergence of *R. ramosae* in the south and *R. boycei* or *R. juglandis* to the north. The second hypothesis is that *R. ramosae* is most closely related to *R. completa* or *R. zoqui*. In this case, the divergence of *R. ramosae* could be attributed to an ~90 km gap in suitable *Rhagoletis* habitat that exists between the EVTM and Sierra Madre Oriental Mountains (SMOr) in east-central Mexico, potentially corresponding to the location and time of divergence in other species groups.

Resolving the relationship of *R. ramosae* to other *R. suavis* taxa can allow us to compare the speciation history of the *R. suavis* species group to other *Rhagoletis* flies in the highlands of Mexico. For example, host races of the species *Rhagoletis pomonella* (Walsh) are also present throughout the United States and Mexico where they attack the fruit of several different species of hawthorns (*Crataegus* spp. (Rosales: Rosaceae)), as well as domesticated apple (*Malus domestica* (Miller) (Rosales: Rosaceae) (Bush 1969). Unlike walnut flies, *R. pomonella* has been proposed to speciate in the absence of complete geographic isolation in the process of shifting and ecologically adapting to new host plants (Bush 1969, Berlocher and Feder 2002). An isolated population of hawthorn-infesting *R. pomonella* exists in the EVTM and is estimated to have become separated from flies in the SMOr over one million years ago (Feder et al. 2003, 2005; Michel et al. 2007). It has been hypothesized that periodic gene flow from the EVTM into the SMOr and more northern United States populations of hawthorn flies played a role in the radiation of the *R. pomonella* group onto a number of novel hosts with differing fruiting times (Feder et al. 2003, 2005; Xie et al. 2007). By comparing the estimates of divergence time for *R. ramosae* with hawthorn-infesting *R. pomonella* in the EVTM, we can infer if the two flies became isolated by an historically shared or different geographic isolating event(s).

The objective of the current study was to resolve the biogeography of the *R. suavis* species group through DNA sequence analysis of a portion of the mitochondrial genome of these flies. We sequenced the Cytochrome Oxidase subunit I and II (COI and COII) mtDNA genes from 33 individuals from the *R. suavis* species group and the outgroup taxon, *Rhagoletis cingulata* (Loew) (Diptera:Tephritidae), to determine the phylogenetic relationship of *R. ramosae* to the five other members of the *R. suavis* group and test the two hypotheses of migration proposed by Rull et al. (2013). We also compared the divergence pattern of the *R. suavis* flies to other species groups of *Rhagoletis* in North America to determine whether any similarities that exist may be due to a shared biogeographic history of isolation.

Methods

Species Collections

Walnuts infested with *Rhagoletis* fly larvae were collected from 1983 to 2012 from a total of 16 sites across the United States and Mexico (Table 1, Fig. 1B) and taken back to the laboratory and held on wire

mesh racks over plastic collecting trays in a greenhouse experiencing ambient environmental conditions. Trays were monitored on a daily basis, and after larvae fed within the husks of walnuts, left the fruit, and formed puparia in the trays, they were collected and stored in Petri dishes containing moist vermiculite. With the exception of samples collected in 2012, which were immediately frozen at –20°C as pupae, the Petri dishes containing pupae were stored at 4°C for 4–6 mo to simulate winter, after which time they were kept at 22°C in a constant temperature room and 14:10 h light:dark cycle until flies eclosed as adult and were frozen at –20°C for later genetic analysis.

Mitochondrial Sequencing

Total genomic and mtDNA was extracted from individual flies using the Puregene DNA purification kit (Qiagen, Inc. Hilden, Germany). The polymerase chain reaction (PCR) was performed on purified DNA isolated from individual flies to amplify a portion of the mtDNA COI and COII genes. COI was amplified using the primer pairs Lep1F (5′-ATTCAACCAAT CATAAAGATATTGG-3′) and Lep1R (5′-TAAACTTCTGGATGTCCAAAAAATCA-3′) (Hebert et al. 2004). COII was amplified using either the primer pairs C1-J-2792 (*George:* 5′-ATACCTCGACGTTATTCAGA-3′) and TK-N-3772 (*Eva*: 5′-GAGACCAT TACTTGCTTTCAGTCATCT-3′) (Simons et al. 1994) or the newly designed *R. suavis* group specific forward (5′-CAACTTACATTCTTTCATGATC-3′) and reverse (5′-GAATTTACTCTATTTGTAATTC-3′) primers, which are internal to *George* and *Eva* primers. *George* and *Eva* primers did not amplify universally across the *R. suavis* species group, potentially due to a mutation in the primer site; thus, the analysis was only done on a fragment internal to these primers. PCR was performed with an initial denature step of 94°C for 5 min, followed by 35 cycles of 94°C for 20 s, 50°C/56°C for COI and COII, respectively, for 20 s, and 72°C for 30 s, and a final elongation step of 72°C for 10 min. Following SAP/exo cleanup, the PCR products were sequenced via Sanger sequencing at the University of Arizona Genomics Core (Tuscon, AZ).

Phylogenetic Analysis

Mitochondrial DNA sequences generated for a total of 33 flies from the 16 walnut collecting sites and *R. cingulata*, a closely related cherry-infesting fly used as an outgroup, were cleaned and trimmed using SEQUENCHER (GeneCorp, Ann Arbor, MI). The sequences are available on GenBank (accession numbers: MF977833–MF977898). The resulting sequences were then aligned against one another using MUSCLE (Edgar 2004) and the alignment refined, truncating the ends of the sequences to only include sites with information at each base pair (bp) using GBlocks (Castresana 2000) to generate reads with a total length of 541 bp for COI and 409 bp for COII for phylogenetic analysis. jModelTest2 (Guindon and Gascuel

Table 1. Site locations, GPS coordinates, and collection information for walnut-infesting *R. suavis* species group flies and the outgroup, *R. cingulata,* sequenced for mtDNA

ID	Location	Latitude	Longitude	Collector
boycei_1	Chiricahua Mountains, AZ	31.9344	–109.2844	J.L. Feder
boycei_2	Ft. Apache, AZ	34.0645	–109.9135	J.L. Feder
boycei_3	Sedona, AZ	34.9126	–111.7268	D. Howard
boycei_4	Sedona, AZ	34.9126	–111.7268	D. Howard
boycei_5	Sedona, AZ	34.9126	–111.7268	D. Howard
boycei_6	Chiricahua Mountains, AZ	31.9344	–109.2844	J.L. Feder
boycei_7	Chiricahua Mountains, AZ	31.9344	–109.2844	J.L. Feder
boycei_8	Sedona, AZ	34.9126	–111.7268	D. Howard
completa_1	Sacramento Mountains, NM	32.9149	–105.3547	J.L. Feder
completa_2	Sacramento Mountains, NM	32.9149	–105.3547	J.L. Feder
completa_3	Edmond, OK	35.6227	–96.4424	M.M. Glover
completa_4	Lawrence, KS	38.9747	–95.2467	R. Cook
completa_5	Lawrence, KS	38.9747	–95.2467	R. Cook
completa_6	Kearney, NE	40.6936	–99.0897	M.M. Glover
completa_7	Lawrence, KS	38.9747	–95.2467	R. Cook
completa_8	Edmond, OK	35.6227	–96.4424	M.M. Glover
juglandis_1	Cibola, AZ	34.0242	–107.1303	J.L. Feder
juglandis_2	Carrizo, AZ	34.0007	–110.2639	J.L. Feder
juglandis_3	Pine, AZ	34.3846	–111.4547	J.L. Feder
juglandis_4	Chiricahua Mountains, AZ	31.9344	–109.2844	J.L. Feder
juglandis_5	Chiricahua Mountains, AZ	31.9344	–109.2844	J.L. Feder
ramosae_1	Taxco, Guerrero, Mexico	18.5616	–99.6048	J. Rull
ramosae_2	Taxco, Guerrero, Mexico	18.5616	–99.6048	J. Rull
ramosae_3	Taxco, Guerrero, Mexico	18.5616	–99.6048	J. Rull
suavis_1	Mt. Clemens, MI	42.5861	–82.8684	R. Cook
suavis_2	Blacksburg, VA	37.2255	–80.4259	D. Cochran
suavis_3	Buckner, MO	39.1342	–94.2011	R. Cook
suavis_4	Colfax, IA	41.6751	–93.2385	M.M. Glover
suavis_5	Mt. Clemens, MI	42.5861	–82.8684	R. Cook
suavis_6	Mt. Clemens, MI	42.5861	–82.8684	R. Cook
suavis_7	Blacksburg, VA	37.2255	–80.4259	D. Cochran
zoqui	Huamantla, Tlaxcala, Mexico	19.3261	–97.9254	J. Rull
cingulata	Coahuila, Nueva León, Mexico	25.8136	–100.3673	S.P. Egan

ID = species identification.

2003, Darriba et al. 2012) identified the Hasegawa, Kishino and Yano (1985) model with invariable sites and rate variation among sites (HKY + I + G) as the nucleotide substitution model having the highest likelihood for explaining the walnut fly mtDNA data based on the Akaike information criterion (AIC). The HKY + I + G model was subsequently employed in a Bayesian phylogenetic analysis with COI, COII, and COI and COII concatenated sequences in MrBayes v.3.2.6 (Huelsenbeck and Ronquist 2001, Ronquist et al. 2012). Bayesian analyzes were conducted with two independent runs each with four Markov chains (three heated and one cold) for one million generations, with a 25% burn in. Default uninformative prior settings were used. Branch lengths for the Bayesian consensus mtDNA trees were estimated with maximum likelihood in PhyML 3.0 based on the HKY + I + G substitution model (Guindon and Gascuel 2003, Guindon et al. 2010). The resulting mtDNA trees were visualized with *R. cingulata* rooted as the outgroup with the APE package in R v3.3.3 (Paradis et al. 2004, R Core Team, 2017).

Genetic Divergence

We estimated the divergence times of the COII gene from *R. suavis* species and *R. pomonella*, as COI sequences for other species groups are not available, using a Bayesian Markov Chain Monte Carlo (MCMC) approach. COII sequences for the *R. pomonella* species in the eastern United States (n = 4), SMOr (n = 2), and EVTM (n = 2) were retrieved from Xie et al. (2008) through GenBank (Table 2). The sequences were aligned with the walnut fly COII sequences and truncated to the same 409 bp length fragment. Bayesian MCMC was run independently five times with 10 million generations each using Beast 2.4.7 (Bouckaert et al. 2014, Drummond et al. 2006). We used a gamma site model with four category counts and an HKY DNA substitution model. We used a relaxed lognormal clock with a rate of 3.54% divergence, a common estimate for the mtDNA molecular clock in insects (Papadopoulou et al. 2010). The tree was estimated with a Yule model and priors on the *R. suavis* and *R. pomonella* groups to constrain for monophyletic clades for the species group. Independent runs were combined using LogCombiner and summary tree produced using TreeAnnotator with at 10% burn-in (Drummond and Rambaut 2007).

Results

Most clades in the *R. suavis* group were well supported in the concatenated COI and COII tree (Fig. 2). For example, *R. suavis, R. juglandis, R. boycei*, and the clade containing, *R. completa*, *R. zoqui*, and *R. ramosae* each formed monophyletic groups with Bayesian posterior probabilities (PP) ≥0.99 (PP = 1.0, 1.0, 1.0, and 0.99, respectively). However, the phylogenetic relationships among *R. completa*, *R. zoqui*, and *R. ramosae* remained unresolved. *R. ramosae* possessed two unique bp differences not found in any other walnut fly species, one in COI (bp position 354) and one in COII (bp position 158). These two autapomorphies characterized *R. ramosae* as a unique clade. However, *R. completa* and *R. zoqui* could not be clearly distinguished, making the branching order within the clade containing *R. completa*, *R. zoqui*, and *R. ramosae* uncertain and the group polyphyletic based on mtDNA haplotypes (Fig. 2). Regardless, the COI and COII gene tree indicated that *R. ramosae* is more closely related to *R. zoqui*/*R. completa* and not to *R. suavis, R. juglandis,* or *R. boycei.*

The mtDNA Bayesian phylogeny for concatenated COI and COII sequences (Fig. 2, individual trees are shown in Supplementary Fig. S1) was congruent with previous analyses of the *R. suavis* group (Smith and Bush 1997, Bush and Smith 1998). The concatenated COI and COII gene tree had *R. suavis* as the first species to branch from the common ancestor of walnut flies, followed by R. *juglandis* speciating from a clade containing all the remaining taxa found in the southwestern United States and Mexico (Fig. 2). *Rhagoletis boycei* was the next species to become isolated and diverged in southwestern New Mexico and the southeastern and central portions of Arizona from the common ancestor of *R. completa*, *R. zoqui*, and *R. ramosae* (Fig. 2). Here, *R. boycei* is now found at higher elevations, while the co-occurring *R. juglandis* is present at lower elevations (Bush 1966). Finally, the clade containing *R. completa*, *R. zoqui*, and *R. ramosae* diversified from north to south, respectively, along an area including the SMOr and EVTM.

The Bayesian tree estimates of divergence time followed the same pattern of speciation inferred for the walnut flies (Fig. 3). Estimated divergence times (and 95% highest posterior density, HPD) for the *R. suavis, R. juglandis*, and *R. boycei* clades were 0.749 (0.458–1.068), 0.635 (0.362–0.869), and 0.419 (0.240–0.619) MYA, respectively. We were unable to estimate divergence times among *R. completa*, *R. zoqui*, and *R. ramosae* due to lack of mtDNA resolution for this clade. However, given their low level of mtDNA differentiation, these taxa were inferred to form recently.

The hawthorn-infesting population of *R. pomonella* in the EVTM formed a separate clade from the rest of the *R. pomonella* group with a divergence time of 0.395 (0.180–0.6322) MYA. The rest of the populations, including *R. pomonella* from the SMOr and apple and hawthorn-infesting populations in the eastern United States, did not form a monophyletic clade and exhibited recent divergence.

Discussion

The mtDNA phylogeny for COI and COII genes resolved the relationship of *R. ramosae* to the other members of the *R. suavis* group, indicating that *R. ramosae* is most closely related to *R. completa* and *R. zoqui*, differing only by two substitutions, one each in the COI and COII genes. These results suggest that *R. ramosae* in the EVTM diverged relatively recently from *R. completa* and *R. zoqui* in the

Table 2. Accession numbers and sampling information for mtDNA COII sequences for *R. pomonella* used for estimate divergence time

ID	GenBank accession	Species	Population	Host plant
pom_EVTM1	EU109153	*R. pomonella*	EVTM	Hawthorn
pom_EVTM2	EU109154	*R. pomonella*	EVTM	Hawthorn
pom_East_haw1	EU109156	*R. pomonella*	US-East	Hawthorn
pom_East_haw2	EU109157	*R. pomonella*	US-East	Apple
pom_East_haw3	EU109158	*R. pomonella*	US-East	Hawthorn
pom_East_apple	EU109159	*R. pomonella*	US-East	Hawthorn
pom_SMOr1	EU109160	*R. pomonella*	SMOr	Hawthorn
pom_SMOr2	EU109161	*R. pomonella*	SMOr	Hawthorn

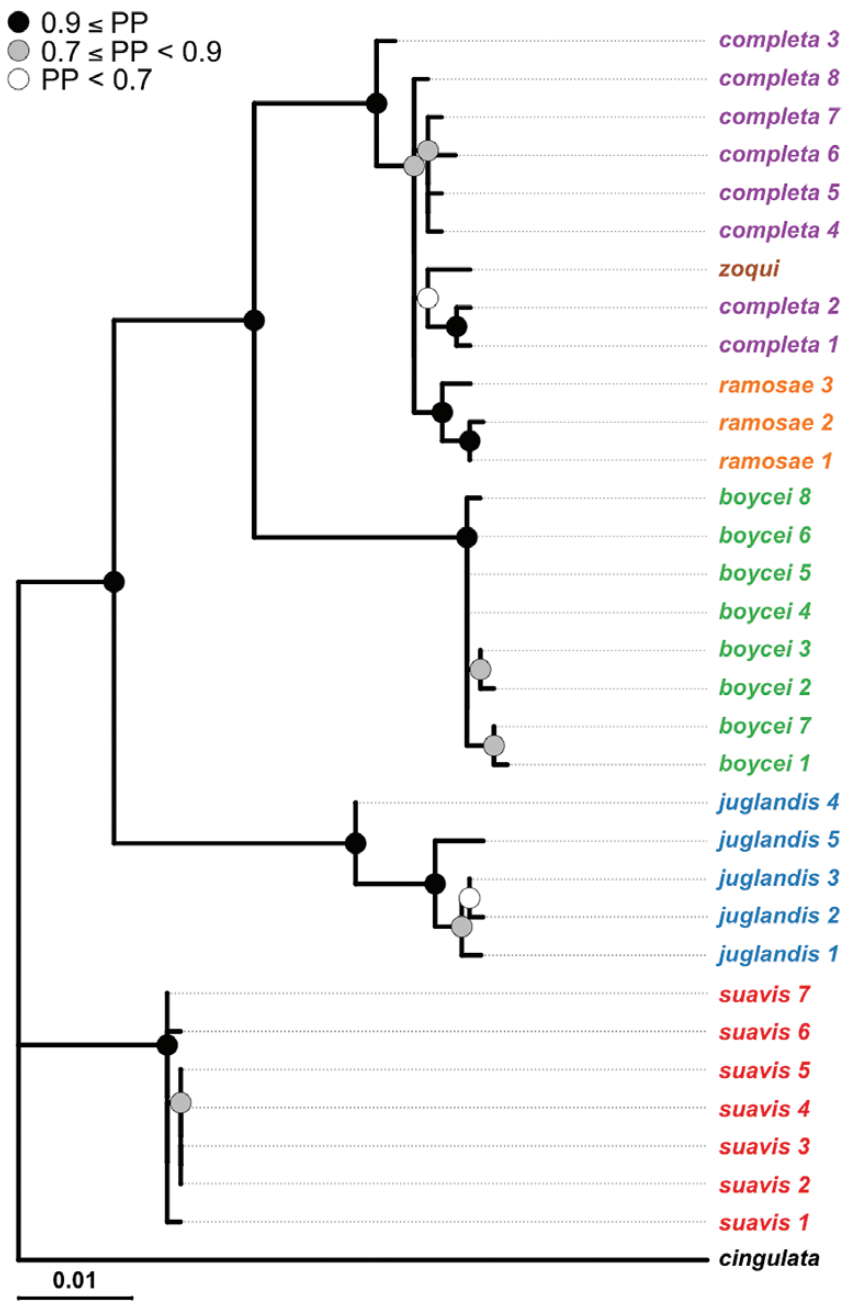

Fig. 2. Bayesian consensus mtDNA tree for combined COI and COII, in the *R. suavis* species group. Gene tree topologies were estimated for the concatenated COI (541 bp) and COII (409) gene fragments using a Bayesian phylogenetic analysis, with the posterior probabilities (PP) shown for each node (black: PP ≤ 0.9, grey: 0.7 ≤ PP < 0.9, white: PP < 0.7). The branch lengths were estimated through maximum likelihood, using PhyML, with *R. cingulata* used as the outgroup. *R. ramosae* is most closely related to *R. completa/R. zoqui*, with the three species forming a monophyletic clade. The position of *R. suavis* and *R. juglandis* is reversed in the COI versus COII gene tree, with *R. juglandis* branching first, followed by *R. suavis*. Scale for legend = 0.01 substitutions/bp.

SMOr. This is supported by the evidence of hybridization between *R. ramosae* and both *R. completa* and *R. zoqui* in a laboratory setting (Tadeo 2014). The fact that *R. completa* does not form a monophyletic clade could be due to gene flow and/or incomplete lineage sorting between the species in the SMOr. A hybrid zone between *R. completa* and *R. zoqui* in Acahuales, Mexico (Rull et al. 2012) supports the hypothesis that ongoing introgression may be contributing to the lack of mtDNA phylogenetic resolution for these two taxa. In addition, the finding that *R. ramosae* is most closely related to species found in the SMOr implies that the major geographic axis for differentiation of walnut flies in Mexico was along the SMOr and EVTM (solid arrow in Fig. 1C), and not the SMOc (dotted arrow in Fig. 1C). As discussed in Rull et al. (2013), these results suggest that divergence of these taxa likely occurred in conjunction with the separation of *Juglans major* (Heller) and *J. mollis* (Engelm.)/*J. pyriformis* (Liebmann) that exists in the SMOr. Thus, the SMOr Mountains may have been a corridor for past migration and isolation between Mexico and the southwestern and eastern United States for walnut fly populations.

Mitochondrial DNA sequencing has its limitations for inferring phylogenetic relationships and in estimating divergence times between taxa (Rubinoff and Holland 2005). The current study only

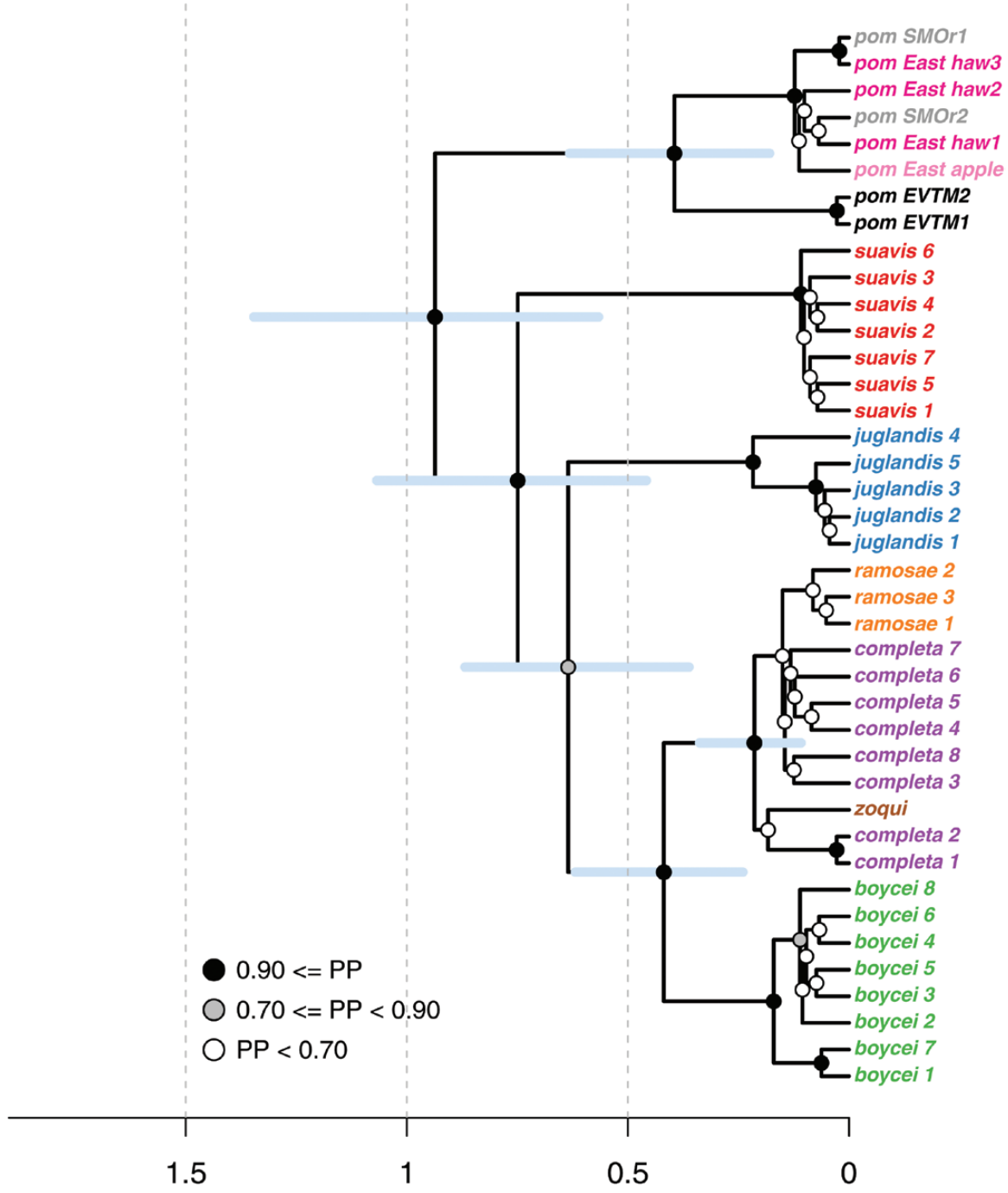

Fig. 3. Bayesian estimates of divergence time for *R. suavis* and *R. pomonella* species. Levels of divergence were estimated for a 409 bp fragment of the COII gene for *R. suavis* species and *R. pomonella* in the eastern United States, SMOr, and the EVTM. Posterior probabilities for the tree topology are colored on the nodes (black: PP ≤ 0.9, grey: 0.7 ≤ PP < 0.9, white: PP < 0.7), with light blue bars indicating the 95% highest priority density for node ages.

considered two mitochondrial genes, which may not give a complete picture of the genetics of speciation and its geographic context for the *R. suavis* group. Including nuclear genes in the study and increasing the scope of the genomic analysis through next-generation sequencing could provide greater resolution to the history of divergence for walnut flies. For example, mtDNA analysis showed low resolution in the clade containing *R. completa*, *R. zoqui*, and *R. ramosae*, but the phylogeny could be improved by increasing the number of loci. However, for the main objective of the current study, the mtDNA COI and COII genes were clearly sufficient in showing that *R. ramosae* is closely related to *R. completa* and *R. zoqui* and much more distantly related to the other walnut-infesting fly species in the group. Given the maternal inheritance of mtDNA, it is always possible that the close affinity of *R. ramosae* to *R. completa* and *R. zoqui* reflects a recent sweep of a mitochondrial haplotype through *R. ramosae*, potentially due to a cytoplasmic factor or associated endosymbiont, and not reflect the evolutionary relationship of the species to *R. completa* and *R. zoqui*. However, if this were the case, then the mtDNA haplotype and endosymbiont would have likely been transferred together to *R. ramosae* from one of the

other taxa, as the acquisition of a cytoplasmic factor or endosymbiont alone would not result in *R. ramosae* displaying a close affinity to *R. completa* or *R. zoqui*. This would require hybridization between these species, which would support their close phylogenetic relationship.

We found that *R. suavis* and *R. pomonella*, despite displaying broad similarities in their current geographic distributions through the eastern United States, EVTM, and SMOr, do not appear to share a common biogeographic history. First, both species groups are found in the EVTM, including the walnut fly, *R. ramosae*, and hawthorn-infesting *R. pomonella* (Rull et al. 2006, 2013). *R. ramosae* is relatively closely related to the other walnut-infesting flies *R. completa* and *R. zoqui* in the SMOr, differing by only one unique substitution in the COII gene (0.24% sequence divergence; Fig. 3). In comparison, hawthorn-infesting populations of *R. pomonella* in the EVTM are much more diverged from *R. pomonella* in the SMOr and eastern United States (Fig. 3), as also found in previous studies which show that *R. pomonella* in the EVTM are geographically isolated and genetically diverged from the *R. pomonella* group (Feder et al. 2003). *R. pomonella* in the EVTM are also post-zygotically isolated from flies in the EVTM, congruent with strong differentiation between the two populations (Rull et al. 2010). We estimated here that the population of *R. pomonella* in the EVTM diverged around 0.395 (95% HPD: 0.180–0.6322) MYA based on the COII gene (Fig. 3)

Second, the *R. suavis* and *R. pomonella* groups both have populations in the SMOr and eastern United States but show different levels of divergence between the regions. For walnut flies, *R. boycei, R. juglandis*, and *R. suavis* show deep splits between the flies in the SMOr and EVTM, with estimated divergence times 0.419 (0.240–0.619), 0.635 (0.362–0.869), and 0.749 (0.458–1.068) MYA. In comparison, hawthorn-infesting populations of *R. pomonella* in the eastern United States and SMOr display no fixed difference in COII in the current or in a previous study, suggesting a close evolutionary relationship (Fig. 3; Xie et al. 2007).

The cherry-infesting species *R. cingulata*, belonging to the *R. cingulata group*, also overlaps in its geographic distribution with *R. suavis* and *R. pomonella* species in North America. In particular, *R. cingulata* co-occurs with *R. pomonella* and *R. suavis* in the eastern United States and SMOr. However, cherry-infesting flies have not been found in the EVTM (Rull et al. 2011). Similar to *R. pomonella, R. cingulata* displays limited mtDNA divergence between the SMOr and the eastern United States, with approximately 0.34% divergence (H. Schuler, personal communication). Like walnut flies, populations of *R. cingulata* also exist in the southwestern United States, where *R. pomonella* is not native (Bush 1966). In contrast to the walnut fly species *R. boycei* and *R. juglandis* in the southwestern United States, which show deep divergence in COII relative to flies in Mexico, cherry-infesting *R. cingulata* displays no mtDNA divergence between the southwestern United States and SMOr (H. Schuler, personal Communication), possibly due to either gene flow or relatively recent isolation (Fig. 3).

Thus, despite their broadly shared geographic distributions, walnut, hawthorn, and cherry-infesting flies have different patterns of genetic differentiation through the southwestern United States, eastern United States, and Mexico implying that they differ in the times of separation across North America. Except in the EVTM, the *R. suavis* group shows greater geographic mtDNA differentiation, and by inference past spatial isolation, than *R. pomonella* and *R. cingulata* flies across their common shared range. For example, the separation of walnut flies in the SMOr and in the eastern United States appears to have occurred historically earlier than that for hawthorn- and cherry-infesting flies from these two regions. It may be that the species *R. suavis* in the eastern United States represents a comparatively old range expansion from Mexico and the southwestern United States into the eastern United States, while the center of endemism for hawthorn- and cherry-infested flies may be the eastern United States itself. However, at the current time, we cannot rule out the alternate possibility that the walnut flies may have been originally native to the eastern United States and became isolated in the southwestern United States and Mexico after migrating from the east. Future nuclear DNA sequencing may resolve these issues, for example, by indicating where genomic diversity (allelic polymorphism) is greatest and where it is reduced, potentially allowing identification of the region of origin for walnut flies in North America.

The general pattern of increased mtDNA differentiation among walnut flies is consistent with the hypothesis that members of the *R. suavis* group have speciated allopatrically in the absence of gene flow (Bush 1966). Although *R. suavis* group flies tend to attack different species of walnut as their major native hosts (Supplementary Table S1, Rull et al. 2013), host specificity of *R. suavis* flies does not appear to have contributed to their divergence. Each species in the *R. suavis* group is capable of attacking any walnut present in its range (Rull et al. 2013). However, the geographic ranges of the different walnuts do not overlap greatly in North America, and the ranges of flies closely follow those of their primary walnut host, generating geographic isolation (Fig. 1). We, therefore, argue that it is the largely disjunct nature of the distribution of walnuts rather than any biological characteristic of the walnut hosts themselves that contributed to the radiation of the *R. suavis* group. The timing of walnut speciation, however, greatly predates that of *R. suavis* husk flies (Rull et al. 2013). The origin of the walnut genus, *Juglans,* in North America (*c.* 60 MYA), and of many present-day walnut species (40–45 MYA), is much older than that for *R. suavis* group flies (Fjellstrom and Parfitt 1995). As a result, host walnuts imposed a populations structure of geographic isolation on flies, as colonists moved into new regions and became largely associated with a new walnut host, possibly related to range expansions and contractions occurring during glacial and inter-glacial periods.

In contrast to walnut flies, speciation in the *R. pomonella* group and for certain *R. cingulata* flies has been hypothesized to occur via host shifting without complete geographic isolation. Thus, *R. pomonella* and *R. cingulata* populations infesting hawthorns and black cherries, respectively, may have experienced higher levels of gene flow when they diverged compared to walnut flies. In addition, it appears that *R. pomonella* populations in the SMOr and *R. cingulata* populations in the SMOr and southwestern United States may have been more recently contiguous with populations in the eastern United States than is the case with walnut flies, the major exception is the hawthorn-infesting population of *R. pomonella* in the EVTM (Rull et al. 2006). Thus, *R. pomonella* and *R. cingulata* generally display reduced mtDNA differentiation than walnut-infesting flies across North America.

Rhagoletis flies are not the only organisms to be affected by the mountainous geography of the highlands of Mexico experiencing geographic isolation and subsequent divergence (Knowles 2001, McCormack et al. 2008, Ruiz-Sanchez and Specht 2013). The different mountain ranges in central Mexico, including the SMOr, SMOc, and EVTM, have been shown to structure populations with between mountain ranges or even in some cases along a mountain range, as is seen in Mexican stoneroller fish in the SMOc (Domínguez-Domínguez et al. 2011). Similar to *R. pomonella* which shows strong differentiation between the EVTM and SMOr, many taxa exhibit distinct breaks in populations between the different mountain

ranges in Mexico. McCormack et al. (2008) reported high mtDNA sequence divergence between Mexican Jay populations inhabiting the SMOc, SMOr, and EVTM with speciation occurring 0.7 MYA. Gene flow is also reduced due to geographic isolation in the EVTM in the Mexican shrub, *Nolina parviflora*, resulting in strong differentiation and population structure in the different mountain ranges (Ruiz-Sanchez and Specht 2013). Further work is needed to determine how congruent the patterns and timing of diversification of these organisms are with *Rhagoletis*. Our results suggest that there is no universal biogeographical process that affected all *Rhagoletis* species groups the same and in unison. It will be interesting to determine how the differences displayed by *Rhagoletis* flies match patterns seen in other organisms in the southwestern and eastern United States and Mexico, to develop a fuller and deeper understanding of the biogeography of these regions of North America.

Supplementary Data

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

Acknowledgments

We thank M.M. Doellman, C. Tait, H. Schuler, P. Meyers, E. Hernandéz, E. Tadeo, E. Acosta, L.A. Miller, C.C. Xu, G. McLain, M. DiGaetano, and Q.D. Read for help with fly collections and rearing in this study. This work was supported by the University of Notre Dame Advanced Diagnostics & Therapeutics Initiative (S.P.E., J.L.F.) and Environmental Change Initiative (S.P.E., J.L.F.), the Instituto de Ecología, A.C. (INECOL) and the Campaña Nacional Contra las Moscas de la Fruta (Convenios INECOL-SENASICA-IICA/CONACOFI) (M.A.), a Rice University Huxley Faculty Fellowship to S.P.E., and NSF and USDA grants to S.P.E. and J.L.F.

References Cited

Avise, J. C., and D. E. Walker. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proc. R. Soc. B Biol. Sci. 265: 457–463.

Avise, J. C., D. Walker, and G. C. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. R. Soc. B Biol. Sci. 265: 1707–1712.

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

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and Drummond, A. J. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10: e1003537.

Bryson, R. W., R. W. Murphy, M. R. Graham, A. Lathrop, and D. Lazcano. 2011a. Ephemeral Pleistocene woodlands connect the dots for highland rattlesnakes of the *Crotalus intermedius* group. J. Biogeogr. 38: 2299–2310.

Bryson, R. W., R. W. Murphy, A. Lathrop, and D. Lazcano-Villareal. 2011b. Evolutionary drivers of phylogeographical diversity in the highlands of Mexico: a case study of the *Crotalus triseriatus* species group of montane rattlesnakes. J. Biogeogr. 38: 697–710.

Bush, G. L. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera, Tephritidae). Bull. Museum 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 (N. Y). 23: 237–251.

Bush, G. L., and J. J. Smith. 1998. The genetics and ecology of sympatric speciation: a case study. Res. Popul. Ecol. (Kyoto). 40: 175–187.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17: 540–552.

Ceballos G., J. Arroyo-Cabrales, and E. Ponce. 2010. Effects of Pleistocene environmental changes on the distribution and community structure of the mammalian fauna of Mexico. Quat. Res. 73: 464–473.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods. 9: 772.

Domínguez-Domínguez, O., M. Vila, R. Pérez-Rodríguez, N. Remón, and I. Doadrio. 2011. Complex evolutionary history of the Mexican stoneroller *Campostoma ornatum* Girard, 1856 (Actinopterygii: Cyprinidae). BMC Evol. Biol. 11: 153.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7: 214.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4: e88

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Feder, J. L. 2017. Allopatric. *In* B.D. Roitberg (ed.) Ref. Modul. Life Sci. Elsevier Press, Amsterdam, Netherlands.

Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith, W. L. Perry, V. Gavrilovic, K. E. Filchak, J. Rull, and M. Aluja. 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. Proc. Natl. Acad. Sci. U. S. A. 100: 10314–10319.

Feder, J. L., X. Xie, J. Rull, S. Velez, A. Forbes, B. Leung, H. Dambroski, K. E. Filchak, and M. Aluja. 2005. Mayr, Dobzhansky, and Bush and the complexities of sympatric speciation in *Rhagoletis*. Proc. Natl. Acad. Sci. U. S. A. 102 (Suppl): 6573–6580.

Fjellstrom, R. G., and D. E. Parfitt. 1995. Phylogenetic analysis and evolution of the genus *Juglans* (Juglandaceae) as determined from nuclear genome RFLPs. Plant Syst. Evol. 197: 19–32.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52: 696–704.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59: 307–321.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Ecol. 22: 160–174.

Hebert, P. D. N., 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.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17: 754–755.

Jaramillo-Correa, J. P., E. Aguierre-Planter, D. P. Khasa, L. E. Eguiarte, D. Piñero, G. R. Furnier, and J. Bousquet. 2008. Ancestry and divergence of subtropical montane forest isolates: molecular biogeography of the genus *Abies* (Pinaceae) in southern México and Guatemala. Mol. Ecol. 17: 2476–2490.

Klicka, J., and R. M. Zink. 1997. The importance of recent ice ages in speciation: a failed paradigm. Science. 277: 1666–1669.

Knowles, L. L. 2001. Genealogical portraits of speciation in montane grasshoppers (genus *Melanoplus*) from the sky islands of the Rocky Mountains. Proc. R. Soc. B Biol. Sci. 268: 319–324.

Lieberman, B. S. 2003. Paleobiogeography: the relevance of fossils to biogeography. Annu. Rev. Ecol. Evol. Syst. 34: 51–69.

Lobo, J.M., and G. Hallffter. 2000. Biogeographical and Ecological Factors Affecting the Altitudinal Variation of Mountainous Communities of Coprophagous Beetles (Coleoptera: Scarabaeoidea): a Comparative Study. Ann. Entomol. Soc. Am. 93: 115–126.

Luna Vega, I., A. O. Alcántara, O. D. Espinosa, and J. J. Morrone. 1999. Historical relationships of the Mexican cloud forests: a preliminary vicariance model applying par- simony analysis of endemicity to vascular plant taxa. J Biogeogr. 26: 1299–1305.

Mayr, E. 1963. Animal Species and Evolution. Harvard University Press, Harvard, MA.

McCormack, J. E., A. T. Peterson, E. Bonaccorso, and T. B. Smith. 2008. Speciation in the highlands of Mexico: genetic and phenotypic divergence in the Mexican jay (*Aphelocoma ultramarina*). Mol. Ecol. 17: 2505–2521.

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.

Nosil, P. 2012. Ecological Speciation. Oxford University Press, Oxford, UK.

Papadopoulou, A., I. Anastasiou, and A. P. Vogler. 2010. Revisiting the insect mitochondrial molecular clock: the Mid-Aegean trench calibration. Mol. Biol. Evol. 27: 1659–1672.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20: 289–290.

R Core Development Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Compution, Vienna, Austria. https://www.R-project.org

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61: 539–542.

Rubinoff, D. and B. S. Holland. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54: 952–961.

Ruiz-Sanchez, E., and C. D. Specht. 2013. Influence of the geological history of the Trans-Mexican volcanic belt on the diversification of *Nolina parviflora* (Asparagaceae : Nolinoideae). J. Biogeogr. 40: 1336–1347.

Rull, J., M. Aluja, J. L. Feder, and S. Berlocher. 2006. Distribution and host range of hawthorn-infesting *Rhagoletis* (Diptera:Tephritidae) in Mexico. Ann. Entomol. Soc. Am. 99: 662–672.

Rull, J., M. Aluja, and J. L. Feder. 2010. Evolution of intrinsic reproductive isolation among four North American populations of *Rhagoletis pomonella* (Diptera: Tephritidae). Biol. J. Linn. Soc. 100: 213–223.

Rull, J., M. Aluja, and J. L. Feder. 2011. Distribution and basic biology of black cherry-infesting *Rhagoletis* (Diptera: Tephritidae) in México. Ann. Entomol. Soc. Am. 104: 202–211.

Rull, J., E. Tadeo, M. Aluja, L. Guillen, S. P. Egan, and J. L. Feder. 2012. Hybridization and sequential components of reproductive isolation between parapatric walnut-infesting sister species *Rhagoletis completa* and *Rhagoletis zoqui*. Biol. J. Linn. Soc. 107: 886–898.

Rull, J., M. Aluja, E. Tadeo, L. Guillen, S. Egan, M. Glover, and J. L. Feder. 2013. Distribution, host plant affiliation, phenology, and phylogeny of walnut-infesting *Rhagoletis* flies (Diptera: Tephritidae) in Mexico. Biol. J. Linn. Soc. 110: 765–779.

Simons, 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.

Tadeo, E. 2014. Diferenciación y aislamiento reproductivo entre especies de moscas del grupo *suavis* en México. PhD Thesis, Universidad Veracruzana.

Xie, X., J. Rull, A. P. Michel, S. 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.

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.