Spring 2012

Rapid Molecular Species Identification of a Morphologically Cryptic Apple Pest.

Emily Green
Western Washington University

WESTERN
WASHINGTON UNIVERSITY

Honors Program
516 High Street
Bellingham, Washington 98225-9112
(360) 650-3034
www.wwu.edu/depts/honors

HONORS THESIS

In presenting this Honors paper in partial requirements for a bachelor's degree at Western Washington University, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this thesis is allowable only for scholarly purposes. It is understood that any publication of this thesis for commercial purposes or for financial gain shall not be allowed without my written permission.

Signature _____

Rapid molecular species identification of a morphologically cryptic apple pest.

Senior Honors Project

Emily Green
Department of Biology
Western Washington University
Bellingham, WA

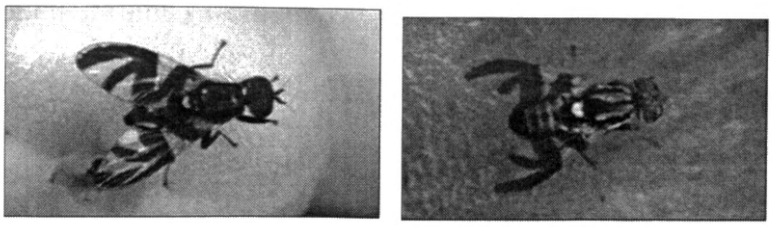
Advisor: Dr. Dietmar Schwarz
Spring 2012

Abstract

Effective management of agricultural pests depends upon accurate identification of those pests. In Washington State, identification of the apple maggot, *Rhagoletis pomonella*, is hindered by the presence of *Rhagoletis zephyria*, a morphologically almost identical species. Using real-time polymerase chain reaction (qPCR), I identified three markers which differ in allele frequency between the two populations. Using these markers and the software NewHybrids, I was able to identify pure *R. pomonella*, *R. zephyria*, and hybrid flies collected from apple, snowberry, and blackhaw fruit in Bellingham, WA, and flies collected from blackhaw in central Washington. Modeling reveals that NewHybrids is able to distinguish 97.7% of pure *R. pomonella* flies, 100% pure *R. zephyria* flies, and 95% first-generation hybrid (F1) flies. Additional markers can be used to identify backcrossed (BC) and second- generation hybrid (F2) flies. These markers provide additional tools for agricultural monitors to check the spread of this persistent agricultural pest and for researchers studying the unique evolutionary history of these sister taxa.

Introduction

The development of rapid and relatively inexpensive DNA technology has provided an additional tool for distinguishing between species, especially in the case of cryptic species (Hébert, 2004). Cryptic species are almost morphologically indistinguishable, but are genetically dissimilar from each other. The presence of cryptic species in a population has implications on measurements of diversity, conservation, and management (Hébert, 2004). In Washington, identification of the agricultural pest *Rhagoletis pomonella* (the apple maggot) is complicated by the presence of the cryptic *Rhagoletis zephyria* (the snowberry maggot). Currently, the two species are identified using morphological measurements. However, because of variation in body size, these measurements are not always conclusive (Yee, 2009). Additionally, morphological analysis can only be applied to adult individuals, while it is the larval ("maggot") stages which cause damage to the fruit. The goal of this study is to identify genetic markers which can be used to distinguish *R. pomonella* from *R. zephyria* in Washington. These markers will provide a novel tool for both agricultural managers and scientific investigators—applicable for flies of all developmental stages, as well as damaged adult specimens.



Rhagoletis zephyria (left) and *Rhagoletis pomonella* (right) (Schwarz Lab, WWU)

Background

The genus *Rhagoletis* is a model for sympatric speciation (speciation without geographic isolation) (Bush, 1969). *Rhagoletis pomonella* (Walsh, 1967) is native to eastern North America, where it infests hawthorns (*Crataegus* sp). In the 19th century *R. pomonella* expanded its range to include apples (*Malus* sp.) introduced from Europe (Bush, 1969). Similar host shifts have occurred to rose (Prokopy, 1980), cherry, plum, and pear (Yee, 2006). Observation of these shifts and the resulting reproductive isolation between populations lead to the formulation of a model for sympatric speciation via host shift (Bush, 1969). Because mating and egg laying occurs on the host fruit, *Rhagoletis* populations are usually sexually isolated by their host fidelity (Feder, 1999). Host races may also become temporally isolated emergence times alter in response to a new host's fruiting cycle (Smith, 1988). However, reproductive isolation is incomplete in many cases, as illustrated by the production of viable hybrids in the lab (Prokopy, 1980; Yee, 2010) and the identification of genetic flow between several species in the wild—notably *R. zephyria* and *R. mendax* in the eastern United States (Schwarz, 2005) and *R. pomonella* and *R. zephyria* in the west (Feder, 1999). The maintenance of host specific populations despite gene flow, the effect of hybrids on wild populations, and the changes necessary to shift hosts are all sources for scientific investigation.

Study of *Rhagoletis* flies also has economic implications. Many of the fruits infested by *Rhagoletis* are economically important crops. In Washington, the infestation of apples by *R. pomonella* is of particular concern. The United State is the second largest producer of apples, and Washington State accounts for 60% of this production (Agriculture, 2011). Because of the ability of *R. pomonella* to switch to new hosts and

rapidly invade new regions, fresh apples must be certified to be maggot-free before export. The Washington Department of Agriculture rigorously monitors apple maggots using sticky traps to catch adult flies (Apple, 1993). Accurate identification is essential to support economic and ecological goals: misidentification of *R. zephyria* results in economic loss for orchards as well as premature expansion of the quarantine region, while misidentification of *R. pomonella* delays implementation of preventative measures to check the spread of the pest into new regions.

Signs found along I-5 in Washington alert travelers to the presence of apple maggot quarantine (WSDA)

Currently, individuals are identified as either *R. pomonella* or *R. zephyria* based on morphological differences. Females are distinguished using the length of the ovipositor, while males are identified based on the orientation of their surstyli (Yee, 2009). However, there is overlap in these measurements between the species, compounding the possibility of misidentification (Yee, 2009). Genetic markers would provide an additional tool for distinguishing these two species. Members of the *Rhagoletis* species complex have proved extremely difficult to separate using genetic markers: no diagnostically fixed allelic differences have been found between *R. pomonella* and *R. zephyria* (McPheron et al, 1988; Feder et al., 1999; Berlocher 2000).

However, researchers have identified differences in allele frequencies between populations of *R. pomonella* infesting apples and hawthorns (Xie, 2008), as well as different allelic frequencies between *R. pomonella* and *R. zephyria* (Feder, 1999). The goal of this study is to identify markers with varying allele frequencies between the populations, providing an alternative to diagnostic markers.

Methods

Sample Collection

Pupae were collected from fallen fruit from four populations of *Rhagoletis*: *R. pomonella* from apples (*Malus* sp.) and blackhaw (*prunifolium* spp.) in Bellingham, WA; *R. zephyria* from snowberry (*Symphoricarpos spp.*) in Bellingham WA; and *R. pomonella* from blackhaw sites in Eastern Washington (Kittitas, Klickitat, Walla Walla, and Yakima counties). DNA extracts were made using the DNAeasy tissue kit (Qiagen, Hilden, Germany), according to the manufacturer's recommendations.

Genotyping

Five potentially diagnostic alleles were identified using GeneBank and screened using *R. pomonella* and *R. zephyria* samples from Bellingham apple and snowberry. Of these five, three exhibited different allele frequencies between the two populations and were selected for further analysis. Individuals were genotyped using Real-Time PCR (qPCR) reactions, which uses fluorescently labeled primers to indicate the presence of specific alleles. qPCR reactions were performed twice for each individual. Additional reactions were performed when necessary to clarify conflicting results.

Data Analysis

The genetic data of the four populations was analyzed using the software NewHybrids v. 1.0 (Anderson, 2001), which uses a Bayesian model to calculate the posterior probability (pp) that an individual belongs to one of six different genotype frequency classes, based on their genotypes. I used uniform priors and 1,000,000 steps

for sampling. Each run was repeated three times and pp values were averaged for each category and individual. I assigned individuals to a specific class when $pp \geq 0.95$. Individuals with $0.95 > pp \geq 0.5$ were tentatively assigned to a class. Individuals with $pp < 0.5$ were classified as generic hybrid if the sum of the pp of the first-generation hybrid (F1), second-generation hybrid (F2), and backcross (BC) categories was ≥ 0.5. There were no flies which did not fit one of these criteria. This analysis was repeated with the inclusion of eight additional microsatellite and allozyme markers.

The diagnostic power of each marker used in the study was calculated by estimating the allele frequency differential between the two species, δ (Shriver et al., 1997). δ is calculated $\frac{1}{2} \Sigma \left| (p_x - p_y) \right|$, where $p_x - p_y$ is the difference between the allele frequencies of the two populations at a specific locus.

To test the ability of NewHybrids to correctly assign known flies of known genetic background, I used the program HybridLab to create F1, F2, and BC individuals from model parental populations with the same allele frequencies as the pure individuals ($p > 0.95$). The resulting model populations were analyzed using NewHybrids and were scored to assess the accuracy of NewHybrid's classification.

Results

Population structure and diagnostic loci

NewHybrids grouped the individiuals into two distinct groupings; one group containing all apple and blackhaw origin flies, the other made up of snowberry origin flies. Analysis using the three selected markers classified all Bellingham blackhaw and apple origin flies as pure or putative pure *R. pomonella* (Figure 1). All Bellingham snowberry flies were classified as pure *R. zephyria*. In the eastern samples, I found 3 possible BC *R. pomonella* flies and two possible pure *R. zephyria* flies (Figure 1). The remaining eastern flies were all pure *R. pomonella*.

Inclusion of the eight additonal makers in the analysis classifies more individuals as hybrid, though the populations still retain the disticnt grouping (Figure 2). Using the eleven markers, I identified 1 BC *R. pomonella* and 1 undetermined hybrid in the Bellingham apple flies; 2 BC *R. pomonella* and 1 undetermined hybrid in the Bellingham blackhaw flies; and 10 BC *R. pomonella*, 1 BC *R. zephyria*, and 1 undetermined hybrid in the central Washington flies (Figure 1). All the Bellingham snowberry samples were pure or putative *R. zephyria*.

The majority of markers used in this study have high diagnostic power, with an allele frequency differential ranging from $\delta=1$ to $\delta=0.5$, though two markers, P22 and Idh, had allele frequecncy differentials of $\delta=0.49$ and $\delta=0.06$, respectivly (Figure 3). P2956 appears to be fully diagnostic.

Accuracy of NewHybrids

Using the three loci, NH correctly identified 97.7% of all pure *R. pomonella*, 100% of all pure *R. zephyria*, 95% of F1s, and 16.7% of F2s (Figure 4). The program classified the remaining 2.3% of pure *R. pomonella* individuals as general hybrids and the remaining 5% of F1 as general hybrids. 30.3% of F2 flies were classified as F1, 33.7% were general hybrids, and the remaining were all pure *R. pomonella* or v *R. zephyria*. NH was unable to correctly identify any BC individuals; ~50% of the time these individuals were considered to be pure *R. pomonella* or *R. zephyria* individuals, while the remainder were considered undetermined hybrids.

Using all eleven loci, NH correctly identified 93.5% of pure *R. pomonella* flies, 99.5% of pure *R. zephyria*, 77.5% of *R. pomonella* BC, 93.5% of F1, 69% of F2, and 77% of *R. zephyria* BC (figure 4). The remaining 6.5% of pure *R. pomonella* flies were identified as either BC *R. pomonella* (6%) or undetermined hybrids (0.5%). Incorrectly assigned pure *R. zephyria* were all assigned as BC *R. zephyria*. 6% of BC *R. pomonella* were assigned to the F2 category, 3.5% were considered F1, 7.5% were *R. pomonella*, and 5.5% were undetermined hybrids. The incorrectly assigned BC *R. zephyria* were assigned as F1 (4.5%), F2 (6%), *R. zephyria* (6.5%), and undetermined hybrid (6%). 1.5% of F1 flies were assigned as BC *R. pomonella*, 0.5% as BC *R. zephyria*, and 4.5% as undetermined hybrid. The remaining F2 flies were assigned as F1 (5%), BC *R. pomonella* (16.5%), BC *R. zephyria* (8%), and undetermined hybrid (2.5%).

Figure 1: Proportion of individuals allocated to each genotype classe. Individuals came from Bellingham Apple (n=23), Bellingham blackhaw (n= 16), Bellingham snowberry, (n=23), or Eastern blackahw (n= 20) fruit.

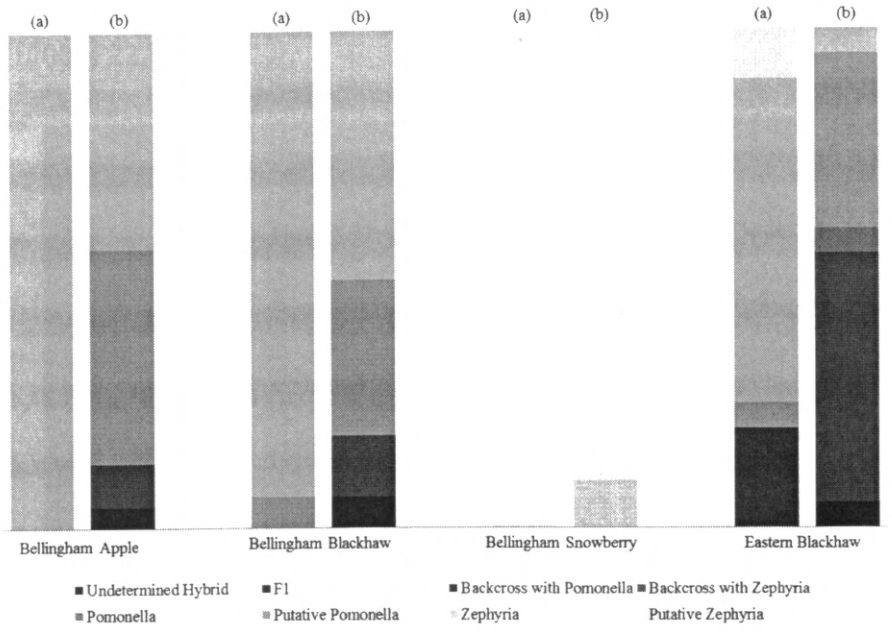
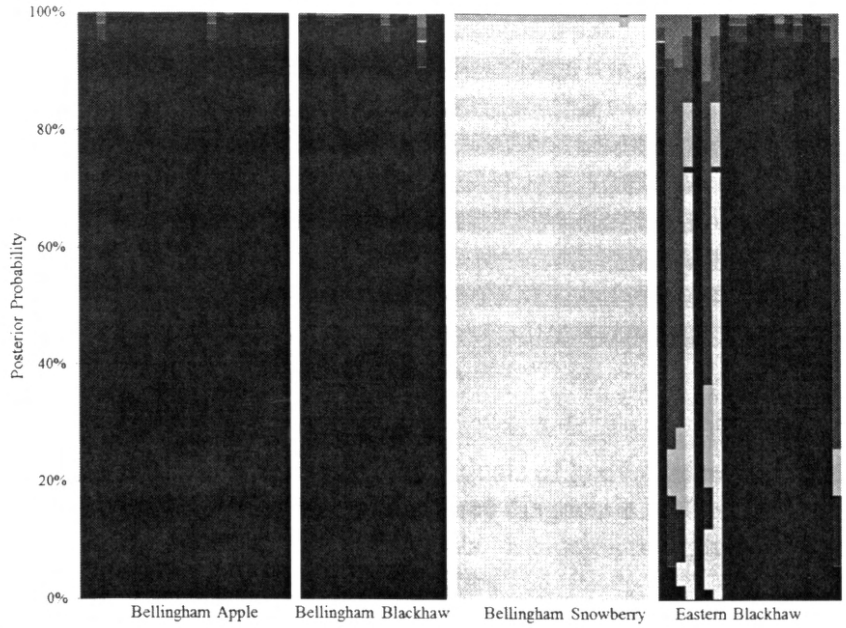
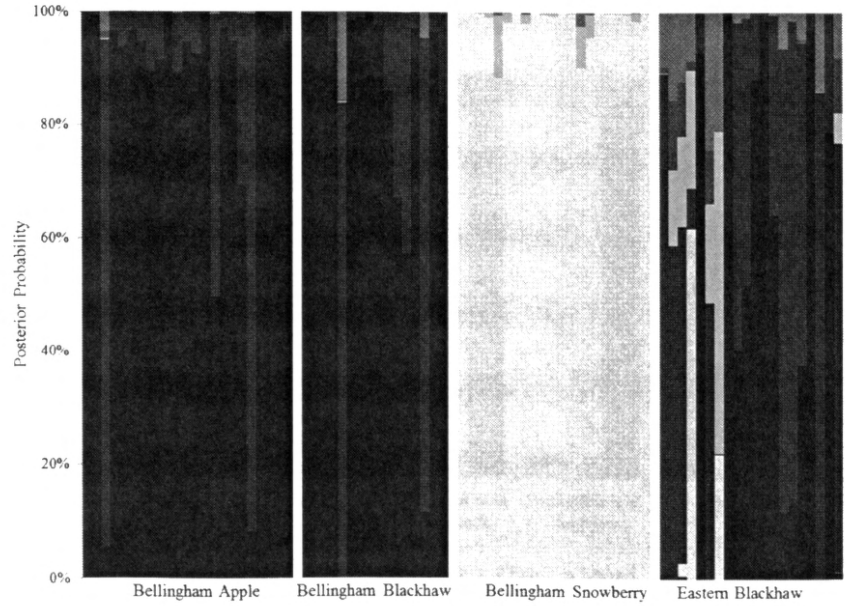


Figure 2: Population structure as determined by NewHybrids based on three diagnostic loci, P1700, P22, and P2956 (a) and eight additional markers (b). Each column represents one individual. Groups of columns represent distinct populations, designated by host plant and location.

(a) P1700, P22, P2956



(b) All markers



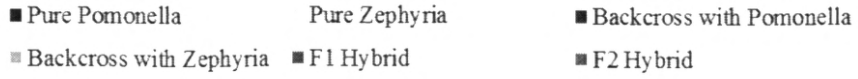


Figure 3: Ranking of loci based on δ, a measure of the diagnostic power of a locus. δ is calculated as $\frac{1}{2} \Sigma \left| P_x - P_y \right|$, where Px and Py are the frequencies of a specific allele in population X and Y.[5] P1700, P2956, and P22 (boxed) were used as diagnostic loci to distinguish *R. pomonella* and *R. zephyria*.

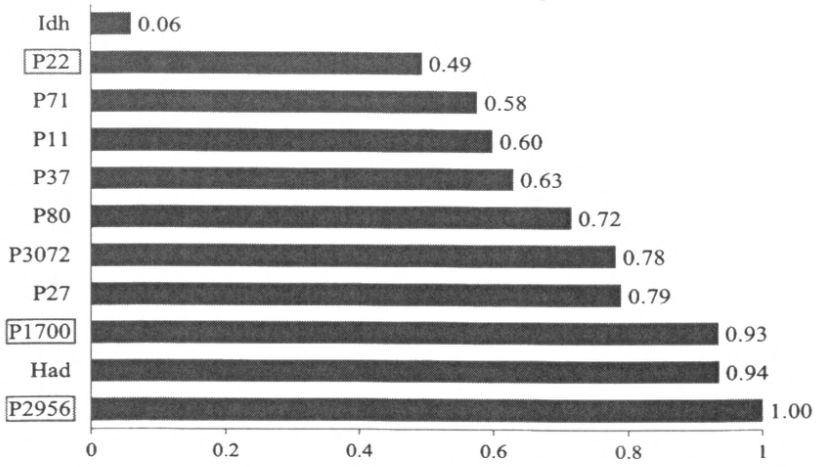
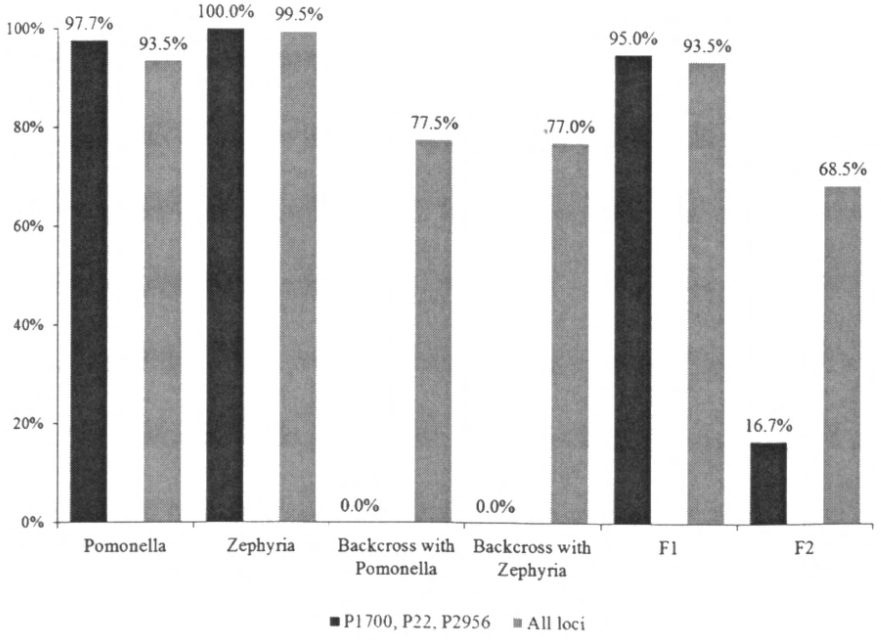


Figure 4: Percentage of simulated individuals of known genetic origin correctly classified by New Hybrids. Use of the three diagnostic loci resolved the majority of *R. pomonella*, *R. zephyria*, and F1 individuals. Inclusion of eight additional loci successfully identified most BC and F2 flies.



Discussion

The goal of this study was to identify genetic markers capable of distinguishing *R. pomonella* from *R. zephyria*. The three markers I selected consistently grouped apple and blackhaw origin, distinct from snowberry origin flies This division is consistent with the evolutionary history and ecological preferences of *R. pomonella* and *R. zephyria*. The most informative marker, P2956, appears at this time to be fixed in the Bellingham *R. zephyria* population—in our screen of 23 individuals, I found no heterozygous flies. Its apparent fixation in the Bellingham *R. zephyria* population could be the result of selection, acting either on P2956 or on nearby genes. However, further sampling, including *R. zephyria* from central Washington, is needed before P2956 can be considered a fixed diagnostic difference.

Using the three selected loci, NewHybrids was able to correctly assign the majority of pure and F1simulated flies. The program was unable to identify any BC individuals. However, for agricultural analysis confusion of a BC *R. pomonella* as pure *R. pomonella* is trivial, as both will infest apples. The same is true for BC *R. zephyria* flies. Difficulty arises when identifying F2 flies, however use of additional markers can increase the accuracy of the identification.

The results of this study further indicate the presence of gene flow between *R. pomonella* and *R. zephyria*. This gene flow appears to be asymmetric, with more *R. zephyria* alleles found in *R. pomonella* populations that vice versa, results consistent with earlier findings (Schwarz, unpublished data). Additionally, more evidence of introgression is found in central Washington flies than in Bellingham flies. This perhaps reflects the later arrival of these flies in central Washington. *Rhagoletis pomonella* was

first detected in Washington in 1980 and quickly spread throughout western Washington. In contrast, it was not until 2004 that Kittitas and Yakima counties in central Washington were quarantined. Given the more recent introduction of *R. pomonella* into these areas, one explanation for the greater degree of admixture in the population could be that there has not been enough time for barriers to gene flow to arise in central Washington, as they have in western Washington. Another possibility is that the small size of founder populations increased the probability of hybridizing with larger, well established *R. zephyria* populations. Further information about the population structure of *R. zephyria* in eastern and central Washington is needed to understand the degree of introgression between the two species.

The ever expanding ease and accessibility of genetic technology has led to the tantalizing possibility of using genetic markers to identify species, in lieu of morphological measurements and ecological descriptions. However, in the case of sibling species such as *R. pomonella* and *R. zephyria*, which are separated by host fidelity, such fixed diagnostic markers may be difficult or impossible to find. Some researchers suggest that *Rhagoletis* taxa are best described as 'quantitative genetic' species, species distinguished by differences in allele frequencies, rather than fixed genetic differences (Mallet, 1995). While this presents a challenge for species diagnosis, this study demonstrates that genes which differ in allele frequencies between two *Rhagoletis* species can be used in place of diagnostic markers to distinguish individuals. Such markers can provide an alternative method for agricultural monitoring, helping to check the spread of this persistent, cryptic agricultural pest.

Acknowledgements

Special thanks to Dr. Dietmar Schwarz for his support and guidance. Thanks also to Harlan Gough, who performed the allozyme genotyping, Kristen Amskar, who did most of the P1700 and P3072 genotyping, and to Sheina Sims, who contributed the microsatellite data.

Literature Cited

Anderson, EC & EA Tompson. 2001. A model-based method for identifying species hybrids using multilocus genetic data. Genetics. 160: 1217-1229.

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

Tree fruit research and extension center: Orchard pest management. Available from http://jenny.tfrec.wsu.edu/opm/displaySpecies.php?pn=140 (accessed 16 January 2011).

Burgarella, C., Z Lorenzo, R Jabbour-Zahab, R Lumaret, E Guichoux, RJ Petit, A Soto, & L Gil. 2009. Detection of hybrids in nature: application to oaks (*Quercus suber Q. ilex*). Heredity. 102:442-452.

Bush, G. L. 1969. Sympatric host race formation and speciation in frugivourous flies of the genus *Rhagoletis* (Diptera, Tephritidae). Evolution. 23(2):237-251.

Feder, JL, SM Williams, SH Berlocher, BA McPheron, & GL Bush. 1999. The population genetics of the apple maggot fly, *Rhagoletis pomonella*, and the snowberry maggot, *R. zephyria*. Implications for models of sympatric speciation. Entomologia Experimentalis et Applicata. 90:9-24.

Hébert, PD, EH Penton, JM Burns, DH Janzen, & W Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulterator*. Proceedings of the National Academy of Sciences of the United States of America. 101:14812-14817

Mallet, J. 1995. A species definition for the modern synthesis. Trends in Ecology and Evolution. 10(7):294-299.

McPheron, BA, DC Smith, & SH Berlocher. 1988. Genetic differences between host races of *Rhagoletis pomonella*. Nature. 336:64-66.

Prokopy, RJ & SH Berlocher. 1980. Establishment of *Rhagoletis pomonella* (Diptera: Tephritidae) on rose hips in southern New England. The Canadian Entomologist. 112(12):1319-1320.

Schwarz, D, BM Matta, NL Shakir-Botteri, & BA McPheron. 2005. Host shift to an invasive plant triggers rapid animal hybrid speciation. Nature. 436:546-549.

Shriver, MD, MW Smith, L Jin, A Marcini, JM Akey, R Deka, & RE Ferrell. 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. American Journal of Human Genetics. 60:957-964.

Smith. DC. 1988. Heritable divergence of *Rhagoletis pomonella* host races by seasonal asynchrony. Nature. 336:66-67

Agriculture: a cornerstone of Washington's economy. Available from http://agr.wa.gov/AgInWa/ (accessed 18 January 2011).

Xie, X, AP Michel, D Schwarz, J Rull, S Velez, AA Forbes, M Aluja, & JL Feder. 2008. Radiation and divergence in the *Rhagoletis pomonella* species group: inferences from DNA sequence data. Journal of Evolutionary Biology. 21:900-913

Yee, W & R Goughnour. 2006. New host records for the Apple Maggot, *Rhagoletis pomonella* (Diptera Tephritidae) in Washington State. Pan-Pacific Entomologist. 82(1):54-60.

Yee, WL, PS Chapman, HD Sheets, & TR Unruh. 2009, Analysis of body measurements and wing shape to discriminate *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae) in Washington State. Annalys of the Entomological Society of America. 102(6):1013-1028.

Yee, W & RB Goughnour. 2010. Mating frequencies and production of hybrids by *Rhagoletis pomonella* and *Rhagoletis zephyria* (Diptera: Tephritidae) in the laboratory. The Canadian Entomologist. 143:82-90.

Zhao, Z, T Wahl & TL Marsh. 2007. Economic Effects of Mitigating Apple Maggot Spread. Canadian Journal of Agricultural Economics 55:501-516.