MSU
LIBRARIES

RETURNING MATERIALS:
Place in book drop to remove this checkout from your record. FINES will be charged if book is returned after the date stamped below.

9-25-96

[illegible] 12
JUN 2 0 2012

[illegible] 1997

THE BIOCHEMICAL GENETICS OF HOST RACE FORMATION AND SYMPATRIC SPECIATION IN *RHAGOLETIS POMONELLA* (DIPTERA: TEPHRITIDAE)

By

Jeffrey Lee Feder

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Zoology

1989

5454633

ABSTRACT

THE BIOCHEMICAL GENETICS OF HOST RACE FORMATION AND SPECIATION FOR *RHAGOLETIS POMONELLA*

By

Jeffrey Lee Feder

The allopatric separation of populations is widely thought to be a prerequisite for speciation in animals. However, true fruit flies in the *Rhagoletis pomonella* (Walsh) sibling species group have been proposed to speciate sympatrically (that is, in the absence of geographic isolation) as a consequence of shifting and adapting to new host plants. The recent shift of the species *R. pomonella* from its native host hawthorn (*Crataegus mollis* Scheele) to introduced domestic apple (*Malus pumila* Mill) provides a test of whether host specialization is sufficient to differentiate fly populations without geographic barriers to gene flow.

A test of the sympatric speciation hypothesis was conducted by examining hawthorn and apple infesting populations of *R. pomonella* for evidence of genetic differentiation using the technique of starch gel electrophoresis. Six allozyme loci (*Malic enzyme*, *Aconitase-2*, *Mannose phosphate isomerase*, *NADH-diaphorase-2*, *Aspartate aminotransferase-2* and *Hydroxyacid dehyrodenase*) consistently showed significant allele frequency differences at paired apple and hawthorn sites across the eastern United States. These six loci mapped to three different regions of the genome and linkage disequilibrium was observed between non-allelic genes within each region. Allele frequencies for five of the six loci displaying host associated differences also co-varied significantly with latitude. Inter-host genetic differentiation is therefore superimposed on clinal patterns of variation within the host races such that the magnitude of host associated divergence is a function of latitude. In addition, the allozyme studies suggested that ambient temperature is an important environmental parameter influencing host-parasite interactions for *R. pomonella* as allele frequencies for *Malic enzyme* and *Aconitase-2* for hawthorn flies correlated closely with the configuration of irregular iso-thermal clines in Michigan and Wisconsin. Further support for the temperature/development hypothesis came from the fact that frequency clines for both host races were steepest across three regions in the Midwest that corresponded with major ecological transition zones. The results from this Dissertation therefore confirm that partially reproductively isolated "host races" can rapidly evolve in sympatry and implicate differences in host recognition and host associated development (related to ambient temperature) as key traits restricting gene flow between hawthorn and apple populations.

Dedication

This Dissertation is dedicated to Benjamin Walsh, the father of sympatric speciation for phytophagous insects. Also to my family and to Emma Wang for their long-standing support, encouragement and love, without which I would not have been able to complete this study.

Acknowledgments

I would like to thank my major professor Dr. Guy L. Bush for stimulating my interest in evolutionary biology and for involving me in research on *Rhagoletis* fruit flies. Special mention must also be made for my Doctoral committee consisting of Dr. Donald O. Straney, Dr. Thomas Friedman, Dr. Wes Brown and Dr. J. Mark Scriber for having the patience to allow me to mature and develop as a research scientist and person. Charley Chilcote and Dan Howard played key roles in perfecting the electrophoretic techniques and discussing the experimental strategies which were used in this Dissertation. I would also like to thank Scot M. Williams, Stewart H. Berlocher and Bruce A. McPheron for reading and making useful comments on an earlier draft of this Dissertation. Jill Millett, Bruce RJ. Angus MacMillan, Tim A. "Slim" Hunt, Miss Anne Wertheimer, Kristy Placido, Jo Ann Palma, Sue Jamrog, Angel Lake, James Edward Zablotny ("Z"), Todd Bierbaum, Scot "Phineas" Williams, Matthew D.B. Liebold, Syd Hess, Sue Bourbeau, and Don Swiderski assisted in numerous aspects of the study including fly collecting, fly rearing and gel running. Ron Prokopy and Sylvia S. Cooley kindly supplied me with flies from Massachusetts, Stewart H. Berlocher, Bruce A. McPheron and D. Courtney Smith with flies from Illinois, Ann L. Averill and Harvey Reissig with flies from New York and Willis Neilson with flies from Canada. Bruce S. Weir provided me with useful information on F-statistics and linkage analysis. Fred Nurnberger, Doug Clark, Wayne M. Wendland and Ken Scheeringa were gracious in giving me access to weather station data from the states of Michigan, Wisconsin, Illinois and Indiana, respectively. I would also like to thank Ed and Jim Hansen and Mr. & Mrs. Melvin Turner for the use of their land. Suzanne Tolin, Angela Sorrells and Lynda Belisle lent much appreciated secretarial help. I am also grateful to Karen Cebra for drawing Figures 1, 2, 3, 5, and 9. Dorie Bush played a critical part in assuring that I never lacked for nutritional fortitude or friendly banter. Have I missed anybody? If so, thanks. Finally, a small but special honorable mention must also go to incipient Dr. Ronald W. DeBry for finishing his dissertation on schedule with mine.

Table of Contents

List of Tables

List of Figures

Introduction

"In dim outline evolution is evident enough. But that particular and essential bit of the theory of evolution which is concerned with the origin and nature of species remains utterly mysterious."

William Bateson 1922
Evolutionary faith and modern doubts
Science 55:55-61.

Modes of Speciation

Although written almost 70 years ago Bateson's statement is still largely valid today. Empirically demonstrating the mechanistic basis for speciation (the cladistic splitting of evolutionary lineages in sexually reproducing organisms) remains one of the most difficult tasks facing evolutionary biologists. As enumerated by M.J.D. White (1978), a fully documented history of speciation must include:

1.) A precise map of the present geographic distributions of taxa.

2.) Geological and climatological evidence suggesting past distributions.

3.) A detailed morphological description of the taxa.

4.) Ecological data on preferred habitats and the life history of the organisms including information on any daily, seasonal or yearly cycles.

5.) Behavioral information regarding any ethological isolating mechanisms or habitat selection for the taxa.

6.) A detailed genetic description of the taxa based on cytological, biochemical and molecular data.

7.) An extensive study of morphological and genetic geographic variation.

8.) Results of experimental hybridization studies in the laboratory.

9.) Information on the frequency of hybridization in nature, provided of course, that the taxa co-occur in at least some portion of their respective ranges.

Allopatric Models of Speciation

Needless to say there is no example of speciation for which all of this information is available. As a result, much attention has been focused on the theoretical basis for speciation. The traditional view of speciation first advanced by Wagner (1868, 1889) and later more clearly defined by Jordan (1896) requires the complete geographic separation (allopatry) of conspecific populations as a pre-requisite for divergence. This classic "dumbbell" model of speciation (Mayr, 1942) involves three different stages. 1.) The range expansion of a population into a new area. 2.) The development of a geographic barrier between the two areas. 3.) The evolution of genetic modifications within populations in the two areas such that when the populations again come into contact they will be kept apart by genetic isolating mechanisms. The emphasis on the need for the complete geographic isolation of populations stems from the view that the genome is a highly integrated genetic system maintained at the status quo by a balance between epistatic interactions among genes and inter-demic gene flow. As stated by Mayr (1963, pp. 518, 521):

> "The real problem of speciation is not how to produce difference but rather how to escape from the cohesion of the gene complex. No one will comprehend how formidable this problem is who does not understand the power of the cohesive forces that are responsible for the coadapted harmony of the gene pool......The importance of complete isolation becomes evident as soon as the extensive epistatic effects of genes are properly taken into consideration. As a consequence, a population cannot change drastically as long as it is exposed to the normalizing effects of gene flow."

As a consequence, Mayr (1963, pg. 525) believes that "genetic differences between populations are built up not by the independent changes of gene frequencies at different loci, but by the harmonious restructuring of the total genotype." Speciation is therefore the process whereby reproductive isolation develops between populations as a by product of the breakdown and restructuring of the genetic system within geographically isolated demes experiencing different selection pressures.

> "As soon as they are separated, two populations will drift apart in their genetic contents for a number of reasons. The probability is nil that the same mutations will occur in the two populations in the same sequence. Each incorporated mutation changes the genetic background of the population and thus affects the selective value of all subsequent mutations. Furthermore, recombination will produce different genotypes in the two gene pools, and, since the same gene may have different selective values in different genotypes, this will likewise lead to a gradual shifting of gene frequencies......The selective agents to which the two separated populations are exposed are also not the same, since there are no two places on the face of the earth where even the physical

> environment is quite identical. Where a population is completely isolated , the composition of the biota is invariably different and this shift of the biotic environment results in an additional powerful difference in selection pressure. This is true for competition, predation, and whatever other ecological interactions may exist.
>
> These local conditions exert selection pressures leading to a steady change of gene contents and to the development of numerous new adjustments. Having to become adapted to a new environment may lead to a sufficiently drastic genetic reconstruction to affect the nature of isolating mechanisms, including potential sterility barriers with sister populations."
>
> E. Mayr (1963, pp. 527,528)

An important aspect of the classic "dumbbell" model is that population sizes are assumed to be large during all stages of speciation. Consequently, the evolution of genetic isolating mechanisms between allopatric populations is a slow and gradual process. However, range expansions of populations during stage 1 of allopatric speciation may involve only a few colonists. This consideration and earlier theoretical work by Wright (1940a,b), led Mayr (1942) to propose an auxillary form of allopatric speciation based on the "founder principle" in which a new isolate formed by only a few individuals could rapidly undergo the genetic turnover necessary for speciation. As discussed by Mayr (1942, 1954, 1963):

> "The founder population differs from the parental population not only in the drastic reduction of the diversity of its gene pool, but also by being exposed to a totally new constellation of environmental factors, biotic as well as genetic ones. The most important of these is the sudden conversion from an open to a closed-and at that a small closed-population......After conversion of the population into a closed one, homozygosity increases inevitably......and as a consequence of their increased frequency in the founder population, homozygotes will be much more exposed to selection and those genes will be favored which are specially

> viable in homozygous condition. Thus the soloist is now the favorite, rather than the good mixer. We come thus to the important conclusion that the mere change of the genetic environment may change the selective value of a gene very considerably (and particularly the change from an open to a closed population). This change, in fact, is the most drastic genetic change (except for polyploidy and hybridization) which may occur in a natural population, since it may affect all loci at once. Indeed, it may have the character of a veritable genetic revolution. Furthermore, this genetic revolution, released by the isolation of the founder population, may well have the character of a chain reaction. Changes in any locus will in turn affect the selective value at many other loci, until finally the system has reached a new state of equilibrium......Therefore, during a genetic revolution, the population will pass from one well integrated and stable condition through a highly unstable period to another period of balanced integration......I use the term genetic revolution advisedly, because the genetic reconstitution was in all these cases far too drastic to be described in terms of a change of gene frequencies. Surely, a change of gene frequencies also occurred, but it was associated with a sufficiently drastic replacement of supergenes that it resulted in the acquisition of a new epigenotype."

Carson (1959, 1965, 1968, 1971) has further developed Mayr's idea by suggesting that a series of sudden population increases ("flushes") and decreases ("crashes") greatly enhances the probability of speciation by the founder principle (although in retrospect, Carson's ideas on the founder-flush hypothesis can also be directly traced back to Wright, 1940a,b).

Sympatric Models of Speciation

Although there are few doubts about the reality of allopatric speciation, questions have been raised over the universality of geographic isolation as a pre-requisite for speciation in sexually reproducing animals and about the genetic basis underlying the speciation process (Fisher, 1930; Lewis, 1962, 1966; Bazykin, 1965, 1969; Maynard-Smith, 1966; Bush, 1966, 1969a,b, 1975a; White, 1968, 1978; Murray, 1972; Endler, 1977; Templeton, 1981). One non-allopatric mode of speciation which has gained considerable attention is termed "sympatric speciation" and is based on the premise that populations can diverge as a consequence of adapting to different ecological niches or habitats which co-occur within the same geographic area. Sympatric models of speciation owe their origin to Benjamin Walsh (1864, 1867) who, five years after Darwin (1859) published "*On the Origins of Species*", postulated that reproductive isolation among certain species of phytophagous insects was the result of these insects sympatrically shifting and adapting to new host plants. Mathematical models of sympatric speciation can be traced back to Maynard Smith's (1966) application of the work of Levene (1953) on the effects of selection in heterogeneous environments to the question of population divergence.

The possibility of sympatric speciation has been criticized by Mayr (1963, pp. 451,452, 527) who has stated that:

> "There is no evidence to support the notion of ecological races as a category distinct from that of geographical races. A theory of ecological speciation would have to prove that a population of a species established itself in a new kind of habitat within the cruising range of individuals of the parental populations and merely by conditioning and habitat selection segregated itself in the new habitat until it had reached species level......Since there are normally very numerous ecological niches within the dispersal area of a deme, niche specialization is impossible without continued new pollution in every generation by migrants.....Contiguous populations of a species are therefore held together by such a close tie of genetic cohesion that it is impossible to divide this essentially single gene pool into two. There is no mechanism (except for instantaneous speciation through polyploidy) by which the cohesion might be loosened to the extent that it would permit the sympatric development of two independent gene complexes which would give

reproductive isolation and ecological compatibility."

Nevertheless, during the century following publication of Walsh's theory, several putative cases of host shifts supporting a sympatric mode of divergence have appeared in the literature (Brues, 1924; Thorpe, 1930; Smith, 1941; Bush, 1966; see review by Diehl and Bush, 1984). Perhaps the most well known example is that for certain species groups of true fruit flies in the genus *Rhagoletis* (Diptera: Tephritidae). Because *Rhagoletis* larvae are highly specific, internal frugivorous parasites of their host plants and mate selection in these flies is directly coupled to host selection, Bush (1966, 1969a,b, 1975a) has argued that variation for traits related to host recognition and host associated survivorship could act as premating barriers to gene flow and result in the sympatric divergence of fly populations specialized on different host plants.

Bush (1969a,b; 1975a) contends that Mayr has overlooked some critical points with regard to sympatric speciation and the genetics of speciation. First, Mayr neglects to distinguish between migration and effective gene flow. Individuals occupying different habitats are likely to be under different selection pressures. Therefore, although migration may occur between con-specific populations occupying different habitats, effective levels of gene flow may still be quite low due to selection acting against immigrants entering a niche to which they are not well adapted. Second, mating in many phytophagous insects and parasites occurs only within preferred habitats. Such organisms therefore have mating systems which are inherently biased toward the development of positive assortative mating following a habitat or host shift. Consequently, any trait conferring differential habitat usage simultaneously produces a pre-mating isolating barrier which further reduces inter-demic gene flow. Third, genomes are not necessarily highly "co-adapted gene complexes" and speciation may not require or be initiated by a "genetic revolution". For instance, although traits related to host recognition are undoubtedly polygenetic (Jaenike, 1987) changes in just a few genes or even a single locus could have major effects on proper host identification (Huettle and Bush, 1972; Carson and Ohta, 1981; Thompson 1988). Populations in which mating is confined to specific habitats could therefore diverge rapidly due to fixation of alternative allelic variants at only a limited number of key polymorphic loci controlling habitat choice.

Summary of Similarities and Differences between Sympatric and Allopatric Modes of Speciation

Perhaps it would be good to digress for a moment to summarize and further clarify the debate concerning the geographic context of speciation. As written by Mayr (1963, pg. 451,453):

> "Theories of sympatric and geographic speciation agree in their emphasis on the importance of ecological factors in speciation. They differ in the sequence in which the steps of the speciation process follow each other. The theory of geographic speciation lets an extrinsic event separate the single into several gene pools, with the ecological factors playing their major role after the populations have become geographically separated. According to the theory of sympatric speciation the splitting of the gene pool itself is caused by ecological factors, and whatever spatial isolation of the thus formed populations ensues is a secondary, later phenomenon......The question to be solved, then, is whether there are any natural phenomena known that can be interpreted as cases of incipient species in a process of sympatric speciation. The complex of phenomena usually labeled biological races has indeed been cited as evidence for such sympatric speciation and must therefore be

studied carefully".

Or as succinctly stated by Maynard Smith (1966):

> "The crucial step in sympatric speciation is the establishment of a stable polymorphism in a heterogeneous environment. Whether this paper is regarded as an argument for or against sympatric speciation will depend on how likely such a polymorphism is thought to be, and this in turn depends on whether a single gene difference can produce selective coefficients large enough to satisfy the necessary conditions."

Rhagoletis pomonella as a Model System of Host Race Formation and Sympatric Speciation

As mentioned before, the biology and natural history of true fruit flies in the genus *Rhagoletis* are ideally suited for studies of population divergence and speciation. In particular, members of the *R. pomonella* sibling species group offer a unique opportunity to directly test various tenants of sympatric speciation theory. Members of the *R. pomonella* complex are sympatric or broadly parapatric in their geographic distributions across North America (Bush, 1966). The four described and one undescribed species comprising the group are all indigenous to North America and each species attacks a different, non-overlapping set of host plants (Bush, 1966). Although several pairs of *R. pomonella* group species can be crossed in the laboratory, these species are genetically distinct and rarely hybridize in nature (Bush, 1966; Berlocher and Bush, 1982; Feder *et al.*, in press; Feder and Bush, submitted; Williams *et al.*, manuscript in prep.). Most importantly, the species *R. pomonella* is actively generating new, sympatric host associated populations (Ward, 1866; Walsh, 1867; Herrick, 1920; Shervis *et al.*, 1970; Prokopy and Bush, 1972; Prokopy and

Berlocher, 1980). The most famous and best documented host shift for *R. pomonella* is that from its native host hawthorn (*Crataegus* spp.) to introduced, domestic apple (*Malus pumila*). The known historical timeframe (~150 years ago; see Walsh, 1867; Illingworth, 1912; O'Kane, 1914; Porter, 1928; Bush, 1966) and probable location of the shift (Hudson River Valley, New York) allow a direct analysis of the host specialization process from a context that excludes the possibility of prior geographic isolation between apple and hawthorn infesting populations. Therefore, if behavioral and/or genetic differences exist between hawthorn and apple populations at sympatric field sites, then this would strongly support a sympatric mode of divergence for these flies.

The objective of this dissertation is to examine hawthorn and apple infesting forms of *R. pomonella* for evidence of genetic differentiation. Three general questions will be addressed. First, do genetic differences exist between hawthorn and apple flies and, if so, do these differences indicate that gene flow is restricted between the two host associated populations? In other words, are hawthorn and apple flies partially reproductively isolated "host races" as predicted by sympatric speciation theory? Second, what is the nature of any observed genetic variation between sympatric apple and hawthorn populations. Is genetic differentiation widespread throughout the genome, as indicative of the "genetic revolution" of Mayr, or are host related differences restricted to allele frequency changes at only a limited number of loci located in specific regions of the genome? Third, what is the geographic pattern of genetic variation for *R. pomonella*, and does this pattern reveal anything about possible factors differentiating apple and hawthorn populations? Is selection a possibility or is the pattern of intra- and inter-host divergence suggestive of stochastic processes such as genetic drift?

Starch Gel Electrophoresis of Soluble Proteins

Starch gel electrophoresis of soluble proteins forms the basis for the survey of genetic polymorphism in *R. pomonella* populations. The technique was first used for analyzing genetic variation in natural populations of *Drosophila melanogaster* by Lewontin and Hubby (1966) and in humans by Harris (1966), and has since become a standard biochemical tool in the study of evolution. In protein gel electrophoresis, an individual is ground in a stabilizing buffer solution and the homogenate is subjected to an electric field in a starch, cellulose acetate or acrylamide gel medium. Protein variants can be distinguished because amino acid substitutions (due to changes in the DNA sequence encoding the protein) can cause net charge or conformational changes in the enzyme which alter its rate of migration in the electric field. Specific histochemical stains are used to visualize the mobilities of a protein or class of proteins in the gel. Mobility differences among or within individuals for the same enzyme system are usually indicative of allelic variation at the locus coding for the protein. Unfortunately, not all genetic variation is detected by starch gel electrophoresis of soluble proteins (Lewontin, 1974). However, studies suggest that the technique does give a fair approximation of the extent of genetic polymorphism in populations (Ramshaw *et al.*, 1979). In this dissertation, the term allozyme refers to enzyme loci resolved by different histochemical stains and isozyme refers to differentially migrating enzyme systems which were detected using the same histochemical stain.

Life History of *R. pomonella*

Before continuing further, I would like to present some additional remarks on the life history of *R. pomonella* and on F-statistics which I hope will clarify the subsequent presentation of the data. First, *R. pomonella* generally has a univoltine life history cycle across its range (Bush, 1966; Dean and Chapman, 1973). Adult females lay their eggs directly into the flesh of the host fruit while the fruit is still on the tree. Eggs hatch within two days and immatures develop through three larval instars over a period of 3-5 weeks. A larvae feeds and completes development only within the single host fruit in which it hatched. After the fruit abscises from the host plant and falls to the ground, late third instar larvae leave the fruit and burrow a few inches into the soil where they pupate after 3-4 days. Flies therefore overwinter as pupae in the soil beneath the tree which they infested as larvae during the preceding summer. Most flies terminate diapause and eclose as adults the next summer, although some pupae require two winters of chilling in the field before completing development (Phipps and Dirks, 1933b).

F-statistics for the Analysis of Population Structure

F-statistics were invented by Sewall Wright as a method for quantifying population structure with respect to the amount of inbreeding within populations and subpopulations (Wright, 1943, 1951). Wright developed separate inbreeding coefficients for an individual relative to its subpopulation (F_{IS}), for an individual with respect to the total population (F_{IT}), and for the correlation between randomly chosen gametes within subpopulations relative to the total population (F_{ST}). The three inbreeding coefficients are related algebraically by the formula:

$$F_{ST} = (F_{IT} - F_{IS}) / (1 - F_{IS})$$

An alternative concept of F-statistics has been developed by Nei (1977) based on characterizing the degree of genetic differentiation among existing populations rather than on inbreeding or coancestry within populations. Nei's method has certain advantages over Wright's in that it does not make explicit assumptions about the nature of factors influencing gene and genotype frequencies. In contrast, mutation, selection and migration are assumed to have minimal effects on the genetic structure of populations under Wright's calculations. However, because of problems related to sampling theory, direct comparisons of Nei's F_{ST} values between populations is not always advisable. Since in many instances in this dissertation I use F_{ST} values for comparative purposes, I calculated F-statistics using the formulae of Weir and Cockerham (1984); a methodology which retains Wright's original conceptual definition of F-statistics as inbreeding coefficients.

Organization of the Body of the Thesis

The body of the thesis is organized into six chapters. The first five chapters are presented in scientific format, with each Chapter being subdivided into an introduction, materials and methods, results, and discussion sections. Chapter 6 is a concluding summary. For ease of reference, tables containing large amounts of allele frequency data appear at the end of Chapters 3, 4 and 5 following the discussion sections.

Chapter 1 establishes a Mendelian mode of inheritance as well as linkage group associations and relative map positions for 12 different polymorphic allozymes in *R. pomonella.* Chapter 2 reports the observation of significant allele frequency heterogeneity for allozymes between hawthorn and apple infesting populations of *R. pomonella* at a sympatric field site in western Michigan. Chapter 3 addresses the question of geographic variation for apple and hawthorn populations through a survey of allozyme variation across a major portion of *R. pomonella*'s range in eastern North America. Chapters 4 and 5 present a more fine grained analysis of spatial and temporal genetic differentiation between hawthorn and apple host races. Chapter 4 explores the regional pattern of divergence on a county by county basis in the midwestern United States by analyzing populations along five latitudinal transects which run from Michigan through Indiana and from Wisconsin through Illinois. Chapter 5 dissects the pattern of local and microgeographic genetic variation within a 60 km^2 area surrounding the original study site near the town of Grant, in western Michigan. Chapter 5 therefore assesses whether patterns observed on a broad geographic scale across eastern North America are spatially and temporally consistent on a very fine level down to individual host trees, over years, within an old field. Chapter 6 is a concluding summary which includes some brief comments on my future research plans for the *R. pomonella* group.

Chapter 1

Inheritance and Linkage Relationships for Allozymes in *Rhagoletis pomonella*

Introduction

The apple maggot fly, *Rhagoletis pomonella* (Walsh) (Diptera: Tephritidae) is of economic importance because it is a major pest of domesticated apples throughout North America. Apples (*Malus pumila*) are not, however, the native host of *R. pomonella*. Apple populations of the fly are thought to have originated within the past 150 years via one or more host shift(s) from hawthorns (*Crataegus* spp.; Bush, 1966, 1969a,b, 1975a). The historical context of the shift(s) makes *R. pomonella* interesting to evolutionary biologists as well as to applied entomologists as a model system for the study of sympatric host race formation and speciation.

Our understanding of the population genetics and phylogenetic relationships of *R. pomonella* and related species has primarily come from starch gel electrophoresis of soluble proteins (Berlocher, 1976; Berlocher and Bush, 1982). Allozymes have also provided genetic markers toward the construction of a linkage map for the apple maggot (Berlocher and Smith, 1983) which is crucial for future research on the genetics of host selection in these flies. The scope of electrophoretic studies has been limited, however, by the number of enzyme loci which can be stained and reliably scored in *R. pomonella*.

The objective of this Chapter is to resolve and characterize new allozyme systems in the apple maggot. Previous studies on a laboratory stock of *R. pomonella* established a Mendelian mode of inheritance for allozymes at 12 electrophoretic loci (Berlocher, 1980; Berlocher and Smith, 1983). Ten of these loci were subsequently mapped to four different linkage groups. Here, I expand on the original results by verifying a genetic basis for six additional allozymes and determining map positions for five of these loci in the same laboratory fly stock. Linkage disequilibrium is also examined between pairs of non-allelic genes for the laboratory strain and compared to values observed in natural *R. pomonella* populations.

Materials and Methods

Crossing of *R. pomonella*

In nature, *R. pomonella* has a univoltine life cycle, spending the winter months in a facultative pupal diapause. For this reason, crosses were performed on a non-diapausing laboratory strain of the apple maggot that has a generation time of approximately three months. The strain originated from an apple population collected near Geneva, New York in the mid 1970's and has gone through approximately 40 generations in the laboratory.

Fifty single-pair crosses were started by placing pairs of virgin flies into cages containing Red Delicious apples. The apples were removed from the cages every other day and replaced with fresh apples. After six weeks the crosses were terminated and surviving parents frozen at -80°C until they could be analyzed electrophoretically. F_1 progeny were reared to adulthood from the infested apples and also stored at -80°C. In 15 of the original 50 matings parents were frozen alive and crosses produced sufficient progeny (>20) to warrant genetic analysis. Eighteen progeny were initially run from each of these 15 crosses with additional progeny being examined from crosses involving parents with especially informative genotypes.

Electrophoretic Methods

Standard starch gel electrophoretic techniques were used (Berlocher and Bush, 1982). Individual flies were sexed and homogenized with a glass rod in 20 µl of a 0.1 M Tris, 1 mM EDTA, 0.13 mM NADP, 0.15 mM NAD, 0.3 M beta-MeEtOH (pH 7.0) grinding buffer. When possible, parents and their progeny were examined on the same gel to facilitate the correct scoring of allozymes. Gels were made from 11% electrostarch (O. Hiller and Co.) using one of three different buffer systems: *system 1*, 3.7 mM Citric acid, 0.1% v/v N-(30 amino-propylmorpholine)(pH 6.1); *system 2*, 8 mM Tris, 3 mM Citric acid (pH 6.7); *system 3*, 76 mM Tris, 5 mM Citric acid (pH 8.7). Electrode running buffer for *system 1* was a 10x concentrated form of gel buffer 1, for *system 2* was 0.22 M Tris, 86 mM Citric acid (pH 6.3), and for *system 3* was 0.3 M Borate (pH 8.2). Gels were sliced and stained for the 11 different enzymes listed in Table 1 along with their abbreviations, enzyme commission numbers and gel buffer systems.

Table 1. Enzyme systems resolved for the laboratory strain of *R. pomonella*. Isozymes that migrated nearest to the cathode were designated 1, the second nearest 2, etc.

Enzyme	Abbreviations for isozymes resolved	Enzyme Commission number	Gel buffer system
Aconitase	Acon 1* & 2	EC 4.2.1.3	2
Aminoacylase	Acy	EC 3.5.1.14	3
Aspartate amino-transferase	Aat 1 & 2	EC 2.6.1.1	1 & 2
NADH-diaphorase	Dia 1*, 2 & 3*	EC 1.6.2.2	2
beta-Hydroxyacid dehydrogenase	Had	EC 1.1.1.30	3
Isocitrate dehydrogenase	Idh	EC 1.1.1.42	1
Malic enzyme	Me	EC 1.1.1.40	3
Mannose phosphate isomerase	Mpi	EC 5.3.1.8	2
Peptidase	Pep 1* & 2	EC 3.4.13.9	3
Phosphoglucose isomerase	Pgi	EC 5.3.1.9	2
Phosphoglucosemutase	Pgm	EC 2.7.5.1	1

* Indicates a monomorphic locus

Genetic Nomenclature

Sixteen different allozymes were resolved in the study, 12 of which were polymorphic (Table 1). Isozymes that migrated the nearest to the cathode were designated 1, the second nearest 2, etc. Six of the polymorphic loci resolved (*Pgm*, *Had*, *Idh*, *Acon-2*, *Aat-2* and *Pgi*) have been previously characterized in *R. pomonella* (Berlocher, 1980; Berlocher and Smith, 1983) and their original allele designations were retained whenever possible. Alleles for six previously unresolved polymorphic loci (*Me*, *Mpi*, *Acy*, *Pep-2*, *Dia-2*, *Aat-1*) were named in the conventional manner with the most common electromorph for a locus designated 100 and all other alleles assigned according to their relative anodal mobility to the common type. The most common allele at each of the new loci was determined from a parallel electrophoretic study of 11 natural *R. pomonella* populations distributed across eastern North America from Door Co., Wisconsin to Nova Scotia, Canada (see Chapter 3).

Calculation of Linkage Disequilibrium

Linkage disequilibrium refers to deviations from non-randomness in the associations among non-allelic genes within gametes and in the union of these gametes in forming zygotes. When coupling and repulsion double heterozygotes can be distinguished it is possible to partition linkage disequilibrium into within gamete (D_w) and between united gametes (D_b) components (Cockerham and Weir, 1977; Weir, 1979). If f_{11} denotes the frequency which two non-allelic electromophs A and B are found together on the same gamete within the same individual and g_{11} is the frequency which A and B are found on different gametes within the same individual, then:

$$D_w = f_{11} - g_{11} \quad \text{and}$$

$$D_b = g_{11} - pq,$$

where p and q are the respective frequencies of alleles A and B in the population (Note: $D_w + D_b$ is equal to the conventional measure, D, of linkage disequilibrium.). A series of four different null hypotheses concerning D_w and D_b can be tested by Chi-square goodness of fit statistics and are explained in Table 2.

When coupling and repulsion heterozygotes cannot be distinguished or if loci are unlinked, a composite measure of linkage disequilibrium (Δ^{AB}) proposed by Peter Burrows (Cockerham and Weir, 1977; Weir, 1979) can still be calculated where,

$$\Delta^{AB} = D_w + 2D_b = f_{11} + g_{11} - 2pq$$

A test statistic of the hypothesis $\Delta^{AB} = 0$ (H4; Table 2) is provided by,

$$X^2(1\ \mathrm{df}) = N(\Delta^{AB})^2 / [p(1-p)+D^{AA}]\,[q(1-q)+D^{BB}]$$

where N is the sample size and D^{AA} and D^{BB} measure deviations of observed from expected frequencies of individuals homozygous for the non-allelic genes A and B, respectively. Direct comparisons of Δ^{AB} between populations are complicated, however, by the fact that Δ^{AB} is gene frequency dependent. Therefore, for inter-populational comparisons we calculated correlation coefficients as suggested by Laurie-Ahlberg and Weir (1979) in which,

$$r^{AB} = \Delta^{AB} / [p(1-p) + D^{AA}]\,[q(1-q) + D^{BB}].$$

Table 2. Null hypotheses (H1-H4) concerning tests of linkage disequilibrium within gametes (D_w) and between united gametes (D_b).

Test	Null hypothesis	Explanation of Test
H1	Db=0	Test of random association of non-allelic genes between uniting gametes.
H2	Db+Dw=0	Test of random association of non-allelic genes within gametes. Db+Dw is equal to the conventional measure of linkage disequilibrium (D). H2 is, therefore, the standard test for gametic disequilibrium.
H3	Dw=0	Test of non-allelic genes being associated more strongly within gametes than between uniting gametes.
H4	Dw+2Db=0	Test of whether composite measure of linkage disequilibrium involving non-allelic gene associations within and between uniting gametes is significantly different than 0.

Results and Discussion

Segregation Studies

Alleles for all 12 polymorphic loci were found to have a genetic basis and all F_1 progeny had genotypes that could be transmitted from their test cross parents (see Table 3 for segregation results; data for *Acon-2*, *Aat-2*, *Pgm*, *Idh*, *Pgi* and *Had* are omitted because Mendelian segregation for alleles at these loci has already been demonstrated by Berlocher and Smith, 1983). Only one locus, *Aat-1*, consistently failed to fit Mendelian expectations for an autosomal locus. Heterozygote males were not found for *Aat-1*. This suggests that *Aat-1* is sex-linked and that males are the heterogametic (XY) sex in *R. pomonella*. Confirmation of sex-linkage for *Aat-1* came from crosses between heterozygous females and single banded males (males which were presumed to be hemizygous). Female progeny from such crosses were either homozygous for the allele possessed by their fathers or heterozygous. Male offspring, however, were always single banded for one of the *Aat-1* alleles possessed by their mothers and were, therefore, hemizygous (Table 3).

Three "cross types" showed significant deviations from Mendelian expectations (Table 3; Note: G-contingency tests indicated that genotypic frequencies were homogeneous among test crosses pooled to form all of the cross types analyzed in the study.). An excess of $Dia\text{-}2^{100}/Dia\text{-}2^{90}$ progeny was observed in cross type 6 for *Dia-2* while cross types 6 and 7 for *Pep-2* produced an excess of $Pep\text{-}2^{100}/Pep\text{-}2^{106}$ heterozygotes. Although the three significant deviations out of a total of 62 tests could be due to any number of factors including selection, they are most easily explained by random type I error.

Table 3. Segregation results from test crosses for six newly resolved allozymes for the apple maggot. Data for *Aat-1* are given separately according to sex (F=female, M=male). The number of test crosses pooled for each cross type is given in parentheses following the numerical designation for the cross type.

Locus	Cross Type	Parental Genotypes	F1 Genotypes and Progeny Number				X^2	(d.f.)
Me	1 (6)	80/80 x 80/80	80/80: 108				----	
	2 (8)	80/80 x 100/80	80/80: 84	100/80: 85			0.01	(1)
	3 (1)	100/80 x 100/80	80/80: 2	100/80: 11	100/100: 5		2.06	(2)
Mpi	1 (11)	100/100 x 100/100	100/100: 197				----	
	2 (4)	100/100 x 100/70	100/100: 52	100/70: 46			0.37	(1)
Acy	1 (13)	100/100 x 100/100	100/100: 186				----	
	2 (2)	100/100 x 100/89	100/100: 10	100/89: 18			2.29	(1)
Dia-2	1 (3)	100/100 x 100/100	100/100: 80				----	
	2 (1)	85/85 x 85/90	85/85: 10	85/90: 8			0.22	(1)
	3 (4)	100/85 x 100/100	100/100: 46	85/100: 62			2.37	(1)
	4 (2)	100/85 x 85/85	85/85: 16	85/100: 20			0.44	(1)
	5 (3)	100/85 x 100/85	100/100: 13	85/100: 32	85/85: 9		3.77	(2)
	6 (2)	100/90 x 100/85	100/100: 6	85/100: 10	100/90: 16	85/90: 4	9.33*	(3)
Pep-2	1 (3)	100/100 x 100/100	100/100: 54				----	
	2 (1)	106/106 x 100/100	100/106: 44				----	
	3 (3)	100/100 x 100/110	100/100: 22	100/110: 32			1.85	(1)
	4 (3)	100/106 x 100/100	100/100: 27	100/106: 27			0.00	(1)
	5 (3)	100/106 x 100/106	100/100: 9	100/106: 28	106/106: 13		1.36	(2)
	6 (1)	100/106 x 100/94	100/100: 11	100/106: 23	100/94: 2	106/94: 13	18.80***	(3)
	7 (1)	100/106 x 110/106	106/106: 3	100/106: 9	110/106: 1	100/110: 4	8.17*	(3)

Table 3. (cont'd.)

Locus	Cross Type	Parental Genotypes	F1 Genotypes and Progeny Number			X^2	(d.f.)
Aat-1	1 (7)	F -100/-100 x -100 M	F	-100/-100: 63		----	
			M	-100: 63		----	
	2 (1)	F -100/-64 x -64 M	F	-64/-64: 2	-64/-100: 7	2.77	(1)
			M	-100: 11	-64: 6	1.47	(1)
	3 (4)	F -100/-64 x -100 M	F	-100/-100: 9	-64/-100: 9	0.00	(1)
			M	-100: 10	-64: 8	0.22	(1)
	4 (2)	F -100/-110 x -100 M	F	-100/-100: 17	-100/-110: 19	0.11	(1)
			M	-100: 23	-110: 13	2.77	(1)
	5 (1)	F -100/-110 x -110 M	F	-110/-110: 3	-100/-110: 6	1.00	(1)
			M	-100: 14	-110: 7	2.32	(1)

---

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ Significant deviations from genotypic expectations as determined by Chi-square tests. Calculations for *Aat-1* were done separately for male and female progeny.

Confirmation of a genetic basis of inheritance for *Me*, *Dia-2*, *Acy*, *Mpi*, *Pep-2* and *Aat-1* brings to 18 the number of polymorphic allozymes characterized in *R. pomonella.* Extensive biochemical surveys of natural apple maggot populations are now possible on a scale approaching those of other study organisms. The combination of genetic, behavioral and ecological data now available for the apple maggot makes the animal a unique organismal system for the study of evolutionary processes within a historical timeframe.

Linkage Results

Restricted crossing over in *Rhagoletis* males (Berlocher and Smith, 1983) makes it possible to determine whether two loci are linked based on the pattern of assortment observed in crosses between double heterozygous males and homozygous females. Our results confirmed that exchange is limited in males in that only one recombinant genotype was observed out of a total of 660 cases in which crossing over in males could be detected. We therefore used the simple method of following the assortment pattern in double heterozygote male crosses to determine linkage groups. Linkage assignments for *Acon-2*, *Idh*, *Aat-2*, *Had*, *Pgi* and *Pgm* were in agreement with the relationships previously reported for the apple maggot (Figure 1; see Berlocher and Smith, 1983). In addition *Dia-2*, *Me*, *Mpi* and *Pep-2* were mapped to linkage groups I, II, II, and III, respectively. *Aat-1* was placed in linkage group V which was designated by Berlocher and Smith (1983) as containing the sex-determining factor(s). *Acy* was not mapped to any specific linkage group. However, I can rule out its association with loci on groups I and II due to its random assortment with *Idh*, *Aat-2*, *Dia-2* and *Acon-2* in test crosses involving double heterozygote males. Also, *Acy* did

Figure 1. Linkage group assignments and map distances (in centimorgans) between loci for *R. pomonella*. Assignments and distances given in parentheses are based on data from Berlocher and Smith (1983).

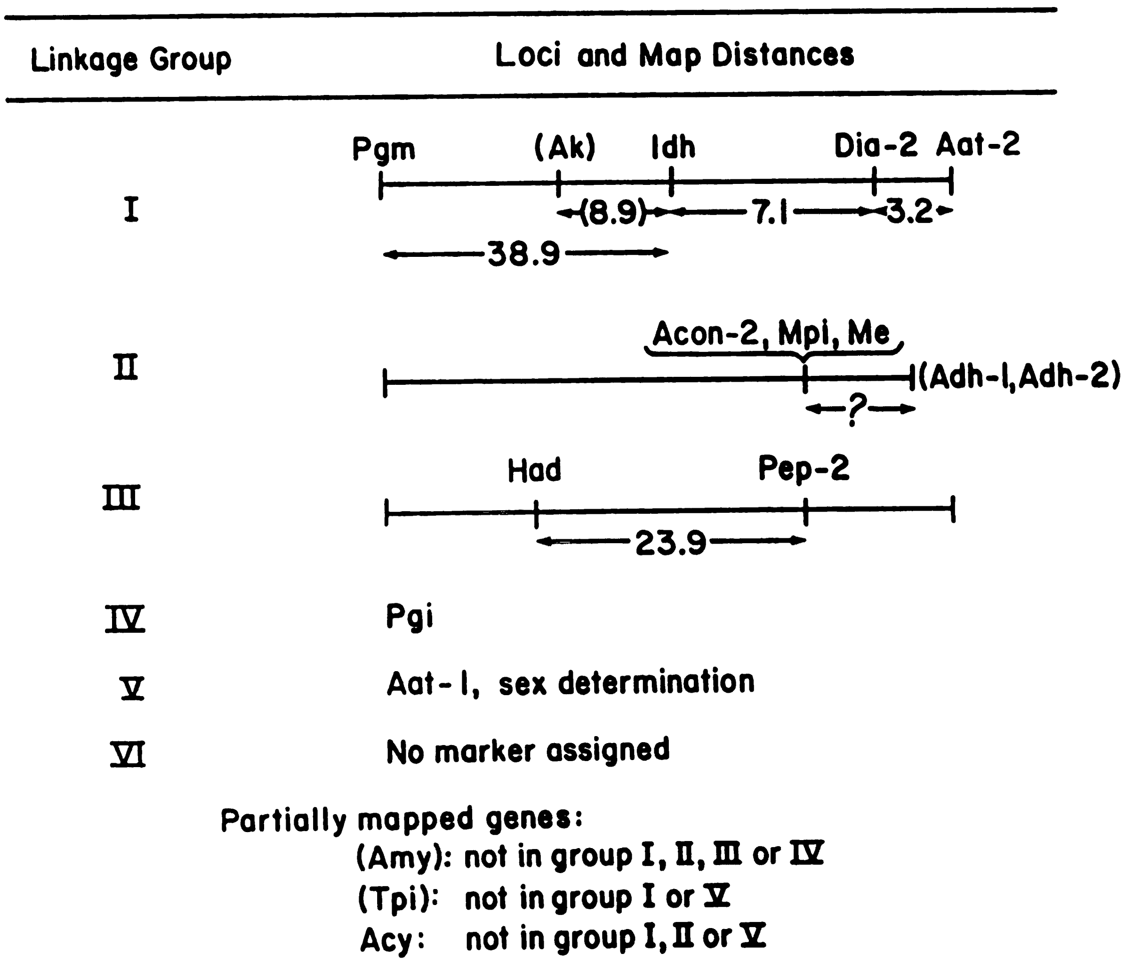

Linkage Group
Loci and Map Distances
I
Pgm
(Ak)
Idh
Dia-2
Aat-2
(8.9)
7.1
3.2
38.9
II
Acon-2, Mpi, Me
(Adh-1,Adh-2)
?
III
Had
Pep-2
23.9
IV
Pgi
V
Aat-1, sex determination
VI
No marker assigned
Partially mapped genes:
(Amy): not in group I, II, III or IV
(Tpi): not in group I or V
Acy: not in group I, II or V

Figure 1.

not show sex linkage so it is not in linkage group V.

Map distances and the relative gene order of linked loci can be determined from crosses involving female parents that are heterozygous for multiple loci. The results for linkage group I differed slightly from those of Berlocher and Smith (1983) in that in our study the crossover distance between *Idh* and *Aat-2* was greatly reduced (10.3 centimorgans (cM) compared to 30.3; Table 4). Recombination rates between *Idh* and *Aat-2* were heterogeneous, however, among the test crosses of Berlocher and Smith (1983). We also found cross to cross variation in map distances between *Aat-2* and *Idh* ranging from 0 to 22.2 cM (Table 4). The lower map distance observed in this study could therefore be the result of calculating exchange rates from crosses involving a higher percentage of inversion heterozygote females than in the previous work. Support for this interpretation comes from the finding of significant gametic disequilibrium between pairs of non-allelic genes encoded at *Aat-2* and *Idh* (see discussion in next section on Linkage Disequilibrium and Genetic Variation in the Laboratory Stock). Unfortunately, extensive ectopic pairing of polytene chromosomes in *Rhagoletis* (Bush,1966) prevents us from directly verifying the existence of inversion polymorphism.

Map distances were also determined between *Had* and *Pep-2* in linkage group III and *Acon-2*, *Mpi* and *Me* in linkage group II (Table 4, Figure 1). Surprisingly, not a single recombinant genotype was observed among *Acon-2*, *Mpi* and *Me* in 162 progeny scored.

Five allozyme linkage groups are now known for *R. pomonella* which presumably correspond to 5 of the 6 chromosomes comprising the haploid karyotype of these flies (Bush, 1966). Any biochemical, morphological or behavioral trait can therefore be mapped to a specific

Table 4. Map distances (in centimorgans) calculated between loci from single pair crosses.

Loci	Cross #					
Linkage group I	16	20	23	25	37	Ave.
Dia-2/Aat-2	0.0	3.7	11.1	0.0	0.0	3.2
Dia-2/Aat-2	11.1	7.4	11.1	5.6	0.0	7.1
Aat-2/Idh	11.1	11.1	22.2	5.6	0.0	10.3
Idh/Pgm	--	--	--	--	38.9	38.9
Linkage group III	28	32	Ave.			
Pep-2/Had	22.4	27.8	23.9			

linkage group. Unfortunately, we are currently unable to associate specific linkage groups with their corresponding chromosomes due to the poor quality of polytene chromosome preparations in *R. pomonella*. However, we assume, like previous investigators (Berlocher and Smith, 1983), that the unmarked sixth linkage group represents the small dot chromosome observed in metaphase preparations (Bush, 1966).

Linkage groups appear to be surprisingly well conserved between *Rhagoletis* and its distant confamilial relative the medfly, *Ceratitis capitata*. For example, the loci *Pgm*, *Idh* and *Aat* are all linked in the medfly as in *R. pomonella* (see Figure 1 and Gasperi *et al.*, 1986). Also, *Mpi* and *Adh* map to the same chromosome in the medfly while *Had*, *Pgi*, *Mpi-Adh*, and *Aat-Idh-Pgm* constitute four different linkage groups (Gasperi *et al.*, 1986). Genetic studies of additional allozyme loci and Tephritid species will therefore provide valuable insight into the evolution of specific gene clusters.

Aat-1 represents the first sex-linked marker mapped in *R. pomonella*. The segregation pattern for *Aat-1* is the first evidence that males are the heterogametic (XY) sex in the species. Hemizygosity for *Aat-1* in males implies that the X and Y chromosomes are not completely homologous. However, morphologically distinct sex chromosomes have not been found for *R. pomonella* (Bush, 1966). It is possible that an undetected translocation between an autosome and the sex determining chromosome (linkage group V) was responsible for generating sex-linkage for *Aat-1*. XY: autosome translocations have been found for *R. tabellaria* (McPheron and Berlocher, 1985) and in the *R. suavis* group (Berlocher, 1984a), however, translocated loci were not hemizygous in these cases. Determination of linkage relationships for *Aat-1* in additional *Rhagoletis* species will help resolve this point.

Linkage Disequilibrium and Genetic Variation in the Laboratory Stock

Progeny testing of double heterozygote parents allows complete identification of coupling and repulsion two locus genotypes. Therefore, D_W (linkage disequilibrium within gametes) and D_b (disequilibrium between united gametes) could be calculated between pairs of linked, non-allelic genes for the *R. pomonella* laboratory strain (Table 5). Hypothesis 1 ($D_b = 0$; see Table 2 for explanation of hypotheses 1-4) was true for 35 of 37 pairwise tests. Only $Me^{100}/Acon\text{-}2^{89}$ ($X^2 = 3.9$, $\underline{P} \leq 0.05$, 1 df) and $Mpi^{100}/Acon\text{-}2^{100}$ ($X^2 = 6.71$, $\underline{P} \leq 0.01$, 1 df) showed significant deviations from non-random union of gametes. These two significant tests may be explained by random type I error.

When gametes unite at random and survivorship is equal for coupling and repulsion genotypes, D_W is a direct measure of linkage disequilibrium within gametes and H2 ($D_W + D_b = 0$) becomes equivalent to H3 ($D_W = 0$; Cockerham and Weir, 1977). Of the 35 tests in which H1 ($D_b = 0$) was true, 23 were significant for either H2 or H3 (Table 5). Non-random associations were found between non-allelic electromorphs for *Me/Acon-2*, *Idh/Aat-2*, *Idh/Dia-2*, *Aat-2/Dia-2* and *Had/Pep-2*. Linked genes, therefore, showed substantial amounts of genetic covariation in the non-diapause line.

Burrows disequilibrium values (Δ^{AB}) were calculated between pairwise combinations of unlinked genes in the laboratory stock. Only 6 of 140 tests of the hypothesis $\Delta^{AB} = 0$ were significant and no pattern was discernable among loci displaying significant tests (data not presented). The general lack of disequilibrium between alleles at unlinked loci indicates

Table 5. Linkage disequilibrium between pairs of non-allelic genes in the laboratory stock of *R. pomonella*. Results are given for disequilibrium within gametes (D_w), between united gametes (D_b) and for a composite measure of D_w and D_b (Δ, Burrows value). See Table 6 for explanations of test hypotheses H1-H4.

Loci/Alleles	Dw	D_b	Δ^{AB}	Significant Hypotheses
Me^{100}/Acon^{100}	0.1000	−0.0305	0.0389	H2** H3*
Me^{100}/Acon^{95}	−0.0330	−0.0111	−0.0555	H4*
Me^{100}/Acon^{89}	−0.0667	0.0416	0.0167	H1* H3*
Mpi^{100}/Acon^{100}	0.0667	−0.0411	−0.0150	H1** H3*
Idh^{100}/Dia-2^{85}	−0.1670	0.0225	−0.1220	H2*** H3*** H4**
Idh^{100}/Dia-2^{100}	0.1670	−0.0333	0.1000	H2*** H3** H4*
Idh^{118}/Dia-2^{85}	0.1830	−0.0186	0.1460	H2*** H3*** H4***
Idh^{118}/Dia-2^{100}	−0.1830	0.0277	−0.1280	H2*** H3*** H4**
Idh^{104}/Dia-2^{100}	0.0167	0.0055	0.0277	H2* H4*
Idh/^{100}/Aat-2^{21}	0.1167	−0.0216	0.0733	H2*** H3**
Idh^{100}/Aat-2^{75}	−0.0833	0.0366	−0.0100	H3*
Idh^{100}/Aat-2^{123}	−0.0833	0.0366	−0.0100	H2** H3* H4*
Idh^{118}/Aat-2^{50}	−0.0667	0.0144	−0.0377	H2*
Idh^{118}/Aat-2^{75}	−0.1167	0.0128	−0.0911	H2*** H3** H4*
Idh^{118}/Aat-2^{75}	0.1000	−0.0400	0.0200	H3*
Idh^{118}/Aat-2^{123}	0.0677	0.0058	0.0783	H2** H3* H4*
Dia-2^{85}/Aat-2^{50}	−0.0667	0.0011	−0.0644	H2** H4**
Dia-2^{85}/Aat-2^{21}	−0.1833	0.0394	−0.1044	H2*** H3*** H4***
Dia-2^{85}/Aat-2^{75}	0.1500	−0.0300	0.0900	H2*** H3** H4*
Dia-2^{85}/Aat-2^{123}	0.0667	−0.0092	0.0483	H2** H3*
Dia-2^{100}/Aat-2^{50}	0.0833	−0.0111	0.0611	H2*** H4*
Dia-2^{100}/Aat-2^{21}	0.1500	−0.0277	0.0944	H2*** H3** H4**
Dia-2^{100}/Aat-2^{75}	−0.1500	0.0330	−0.0833	H2*** H3** H4*
Dia-2^{100}/Aat-2^{123}	−0.0500	0.0000	−0.0500	H2* H4*
Pep-2^{100}/Had^{100}	−0.0667	−0.0064	−0.0790	H2** H4*
Pep-2^{100}/Had^{122}	0.0500	−0.0006	0.489	H2*

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ Chi-square test (1 df).

that genetic interactions across chromosomes are not strong in the non-diapause strain.

The non-diapausing line has a surprisingly high level of heterozygosity (Table 6). Ten of the 12 loci analyzed showed a trend for an excess of heterozygosity over expectation. Several loci in the non-diapausing strain also displayed significantly higher levels of heterozygosity than were observed in 11 natural populations of *R. pomonella* (Table 6). Small population sizes associated with laboratory rearing should theoretically lead to a loss of neutral allozyme variability according to the equation:

$$H_t/H_o = (1-(1/2N_e))^t ,$$

where H_t is the heterozygosity of a population after t generations in the laboratory, H_o is the heterozygosity of the founding population and N_e is the effective size of the laboratory population (Wright, 1931). If N_e=150 and t=40 (reasonable approximations for the laboratory stock), the ratio of H_t to H_o should theoretically be 0.875. Although we do not know H_o for the original New York population, an estimate of 0.406 can be made based on the average level of heterozygosity observed in 11 natural populations (Table 6). The average H_t for the laboratory population is 0.439 so that the actual value of H_t/H_o is 1.08. The laboratory line therefore has a slightly higher, rather than a lower, overall level of heterozygosity compared to natural populations.

Heterozygosity is not, however, the only aspect of genetic variation which can be affected by a sudden reduction in population size. Allelic diversity is also a sensitive indicator of whether a population experienced a

Table 6. Observed and expected heterozygosity and the presumed number of alleles lost in the laboratory stock compared to natural apple maggot populations.

Locus	Laboratory Population			Ave 11 natural populations
	H obs	H exp.	Alleles lost	H obs ± SD
Me	0.333	0.278	0	0.472 ± 0.066
Mpi	0.133	0.125	4	0.247 ± 0.076
Acon-2	0.767	0.654	4	0.561 ± 0.164
Aat-2	0.800	0.738	7	0.737 ± 0.035
Dia-2	0.633	0.506	0	0.369 ± 0.050
Idh	0.467	0.529	3	0.422 ± 0.033
Had	0.433	0.406	2	0.373 ± 0.089
Pgi	0.400	0.444	5	0.182 ± 0.052
Aat-1	0.533	0.424	2	0.299 ± 0.075
Acy	0.066	0.064	2	0.348 ± 0.092
Pep-2	0.533	0.475	0	0.503 ± 0.160
Pgm	0.167	0.156	6	0.344 ± 0.091
All loci	0.439	0.400	2.9	0.406

recent bottleneck (Nei *et al.*, 1975; Maruyama and Fuerst, 1985). This point certainly holds true for *R. pomonella* as the laboratory strain appears to have lost almost half the number of electromorphs normally found in wild apple populations (Table 6). However, we can only assume that alleles missing from the laboratory stock were lost because we have no information concerning the genetic composition of the original apple population from which the non-diapausing line was derived. Nevertheless, our rough estimate of a 40% decline in allelic diversity for the laboratory line agrees reasonably well with the colony having gone through a bottleneck of around 150 individuals (see Table 3 of Nei *et al.*, 1975). The slightly higher level of heterozygosity in the non-diapausing strain is therefore surprising because of the obvious effect that the population bottleneck had on allelic diversity.

What factors, then, are responsible for maintaining gametic disequilibrium and the higher than expected levels of heterozygosity in the laboratory population? Founder effects and drift related to the smaller size of the laboratory population are undoubtedly involved in generating gametic disequilibrium. However, several of the loci displaying disequilibrium are separated by a number of map units. For example, *Idh* and *Aat-2* are 10.3 cM apart and *Had* and *Pep-2* are separated by 23.9 cM. Recombination should be common among these loci and result in the rapid decay of linkage disequilibrium. Consequently, founder effects and drift cannot be the only factors responsible for gametic disequilibrium in the laboratory line. The heterogeneity of recombination rates observed among allozyme markers in linkage group I suggests that inversion polymorphisms exist in the laboratory population. It is therefore possible that heterosis associated with inversion heterozygote genotypes is maintaining gametic disequilibrium and genetic heterozygosity in the laboratory stock.

Analysis of Disequilibrium between Laboratory and Natural Populations

Comparisons of correlation coefficients (r^{AB}) between the laboratory and a natural apple maggot population from western Michigan indicate a drastically different pattern of linkage disequilibrium for *Me/Acon-2* , *Pep-2/Had* and *Dia-2/Aat-2* (Table 7). For instance, *Me*^{100}*/Acon-2*^{95}, *Pep-2*^{100}*/Had*^{100} and *Dia-2*^{85}*/Aat-2*^{21} are highly positively correlated in nature while being highly negatively correlated in the laboratory strain. Besides *Me/Acon-2*, *Had/Pep-2*, *Dia-2/Aat-2* and *Me/Mpi*, however, no other pair of loci show a consistent pattern of linkage disequilibrium in the natural population.

Why is the pattern of linkage disequilibrium between non-allelic genes different in nature than in the laboratory? Geographic differentiation between populations from Geneva, New York (the original source of the laboratory stock) and Michigan could account for the differences. However, analysis of other natural *R. pomonella* populations from across the northeastern United States (Chapter 3) has revealed the same pattern of disequilibrium as seen in the Michigan population ruling out geographic variation as a possible explanation.

Differences in the breeding structure of the laboratory and natural populations could also account for the observed differences in linkage disequilibrium. Unfortunately, the composite nature of r^{AB} does not allow us to determine the extent to which the non-random association of genes within gametes (D_w) and the non-random union of gametes (D_b) affect linkage disequilibrium in the natural population. If D_w is the same in

Table 7. Correlation coefficients calculated between pairs of non-allelic genes for the laboratory stock and a natural apple fly population collected from Grant, Michigan.

Alleles tested	r^{AB} (laboratory)	r^{AB} (nature)
$Me^{100}/Acon^{95}$	-0.374*	0.379***
Me^{100}/Mpi^{100}	0.069	-0.214*
$Mpi^{100}/Acon^{100}$	-0.010	-0.242*
$Idh^{100}/Dia\text{-}2^{85}$	-0.530**	0.044
$Idh^{100}/Dia\text{-}2^{100}$	0.499*	-0.055
$Idh^{118}/Dia\text{-}2^{85}$	0.631***	-0.018
$Idh^{118}/Dia\text{-}2^{100}$	-0.570**	0.018
$Idh^{104}/Dia\text{-}2^{100}$	0.373*	0.098
$Idh^{100}/Aat\text{-}2^{123}$	-0.399*	-0.036
$Idh^{118}/Aat\text{-}2^{21}$	-0.431*	0.131
$Idh^{118}/Aat\text{-}2^{123}$	0.454*	0.058
$Dia\text{-}2^{85}/Aat\text{-}2^{50}$	-0.473**	0.203
$Dia\text{-}2^{85}/Aat\text{-}2^{21}$	-0.604***	0.250**
$Dia\text{-}2^{85}/Aat\text{-}2^{100}$	-0.218	-0.333***
$Dia\text{-}2^{100}/Aat\text{-}2^{50}$	0.463*	-0.307***
$Dia\text{-}2^{100}/Aat\text{-}2^{21}$	0.564**	-0.456***
$Dia\text{-}2^{100}/Aat\text{-}2^{100}$	-0.163	0.380***
$Dia\text{-}2^{100}/Aat\text{-}2^{75}$	-0.428*	0.254**
$Dia\text{-}2^{100}/Aat\text{-}2^{123}$	-0.366*	0.031
$Pep\text{-}2^{100}/Had^{100}$	-0.457*	0.486***

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ Chi-square test of Burrow's values.

nature as it is in the laboratory then selection, differential fecundity or fertility, deviations from non-random mating and/or gene frequency differences between sexes, could all be significant in the natural population. Almost all single locus Hardy-Weinberg tests, however, were in equilibrium for the Michigan population (see site 5 in Tables 21-32). This suggests that mating is random and single locus selection is not strong within the wild apple maggot fly population. Also, gene frequencies were not significantly different between males and females in the western Michigan population (data not shown). These findings all suggest that it is D_w, rather than D_b, which differs between laboratory and natural *R. pomonella* populations.

Although we cannot completely discount non-random union of gametes in nature, two observations make it tempting to speculate that selection for high fecundity and/or rapid development has favored alternative combinations of non-allelic genes for *Me/Acon-2*, *Pep-2/Had* and *Dia-2/Aat-2* in the laboratory population. First, test crosses between laboratory strain flies generally produce about twice as many F1 offspring as those between wild type parents (J.L. Feder, personal observation). Second, allele frequencies for *Me*, *Acon-2* and *Had* correlate with the timing of diapause termination and adult eclosion in natural *R. pomonella* populations (Feder *et al.*, manuscript in preparation).

Regardless of its causal basis, the results from this study indicate that the non-diapausing line has an altered genetic constitution. Genetic, physiological and behavioral changes associated with laboratory rearing are now known for a number of insect species including *Drosophila melanogaster* (Laurie-Ahlberg and Weir, 1979); the screw worm, *Cochliomyia hominivorax* (Bush and Neck, 1976; Bush *et al.*, 1976; Singh, 1984); the codling moth, *Laspeyresia pomonella* (Bush, 1975b); the olive fruit fly, *Dacus oleae* (Prokopy *et al.*, 1975; Bush and Kitto, 1979; Economopoulous, 1980; Economopolous and Zervas, 1982; Loukas *et al.*, 1985); and the med fly, *Ceratitis capitata* (Saul, 1986; for a more extensive review of the insect literature see Menken and Ulenberg, 1987). Although we are not aware of any studies directly comparing the behavior of laboratory and wild apple maggot flies, the genetic changes observed for the non-diapausing line give a warning that the colony may also be behaviorally modified.

Chapter 2

Genetic Differentiation between Sympatric Host Races of the Apple Maggot Fly, *Rhagoletis pomonella*

Introduction

True fruit flies in the *Rhagoletis pomonella* species group have been proposed to speciate sympatrically (that is, in the absence of geographic isolation) as a consequence of shifts to previously unexploited host plants (Bush, 1966, 1969a,b, Bush, 1975a). Because *Rhagoletis* larvae are host specific fruit parasites and mate selection in these flies is directly coupled to host plant recognition (Prokopy *et al.*, 1971,1972), variation for larval survivorship and host preference traits can act as genetically based barriers to gene flow and subsequently result in the sympatric divergence of fly populations adapted to alternative hosts. The shift of *R. pomonella* from its native host hawthorn (*Crataegus* spp.) to domestic apples (*Malus pumila*) in eastern North America within the last 150 years (Walsh, 1867) provides an opportunity to determine whether host specialization is sufficient to differentiate populations without prior periods of geographic isolation. In this Chapter, I report finding significant genetic differentiation between co-occurring hawthorn and apple populations of *R. pomonella* at a field site near Grant, Michigan. The result confirms that hawthorn and apple flies represent partially reproductively isolated "host races" and is consistent with a sympatric mode of divergence for these flies.

Materials and Methods

R. pomonella flies were collected from an old field located near the town of Grant (Newaygo Co.) in western Michigan (Figure 2). In 1984 and 1985 flies were sampled as larvae from infested fruit and reared to adulthood in the laboratory. In contrast, adults were captured directly off of host fruits by sweep net in 1986. Standard horizontal starch gel electrophoretic techniques were used (see Chapter 1 for details and also Berlocher and Bush, 1982) and a total of 29 different allozyme systems were resolved (Table 8). Alleles were numbered according to their relative anodal mobilities, with the most common allele for a locus assigned a value of 100 and used as a standard. Isozyme systems that migrated the nearest to the cathode were designated 1, the second closest 2, etc. Thirteen loci were polymorphic (frequency of the common allele less than 0.95; Table 8) and, excluding *Aat-1* which is sex-linked, only 15 significant deviations from Hardy-Weinberg equilibrium were observed for these loci in 290 tests (data not shown). The significant deviations displayed no regular pattern and their number does not differ from random expectation due to type I error.

Results

Six allozyme loci, *Aconitase-2* (*Acon-2*), *Malic enzyme* (*Me*), *Mannose phosphate isomerase* (*Mpi*), *Aspartate amino-transferase* (*Aat-2*), *NADH-Diaphorase-2* (*Dia-2*) and *beta-Hydroxy acid dehydrogenase* (*Had*),

Figure 2. Scale diagram of study site located near the town of Grant, in western Michigan. The site is an old field which has remained fallow since at least 1922 (E. Hansen, pers. comm.). Host tree designations (A=apple, H=hawthorn) are given along with collecting dates and inter-tree distances (in meters).

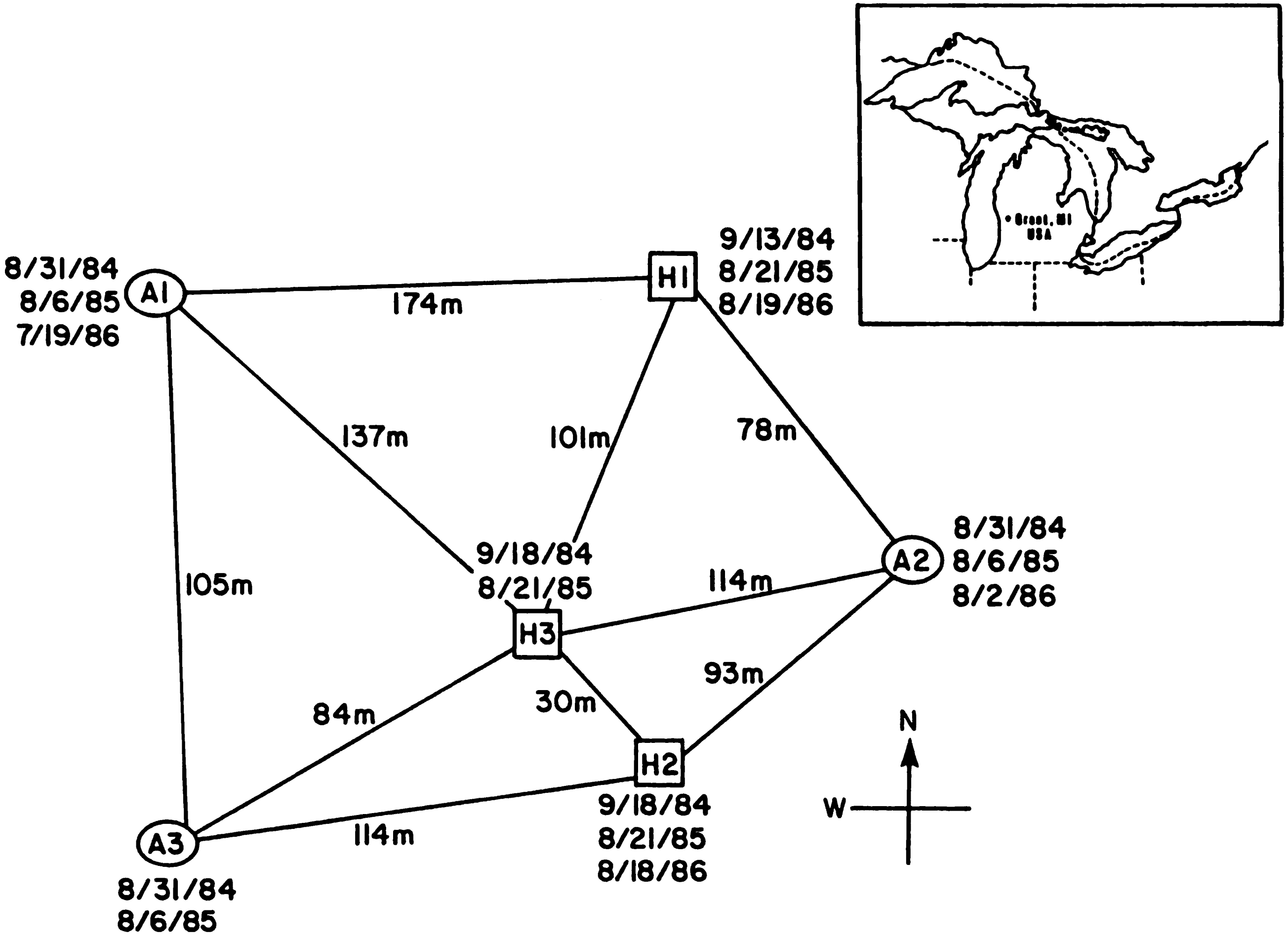

Grant, MI
USA
A1
8/31/84
8/6/85
7/19/86
H1
9/13/84
8/21/85
8/19/86
A2
8/31/84
8/6/85
8/2/86
H3
9/18/84
8/21/85
H2
9/18/84
8/21/85
8/18/86
A3
8/31/84
8/6/85
174m
137m
101m
78m
105m
114m
93m
30m
84m
114m
N
W

Figure 2.

Table 8. Allozymes resolved for the host races of *R. pomonella*. Enzyme abbreviations are given in parentheses. Isozymes that migrated the nearest to the cathode are designated system 1, the second nearest 2, etc.

Monomorphic or Essentially Fixed Enzymes		Polymorphic Enzymes	
Napthyl-acid phosphatase	(Acph)	Aminoacylase	(Acy)
Aconitase-1	(Acon-1)	Aconitase-2	(Acon-2)^{t}
Alcohol dehydrogenase-1 & 2	(Adh-1 & 2)*	Adenylate kinase	(Ak)
Aldolase	(Aldo)	Aspartate amino transferase 1 & 2	(Aat-1 & 2^{t})
NADH-diaphorase-1 & 3	(Dia-1 & 3)	NADH-diaphorase-2	(Dia-2)^{t}
Fumarase	(Fum)	ß-Hydroxyacid dehydrogenase	(Had)^{t}
Hexokinase	(Hk)	Isocitrate dehydrogenase	(Idh)
Malate dehydrogenase-1 & 2	(Mdh-1 & 2)*	Malic enzyme	(Me)^{t}
Peptidase-1 & 3	(Pep 1 & 3)	Mannose phosphate isomerase	(Mpi)^{t}
Superoxide dismutase	(Sod)	Peptidase-2	(Pep-2)
Trehalase	(Tre)	Phosphoglucose isomerase	(Pgi)
Triose phosphate isomerase	(Tpi)	Phosphoglucomutase	(Pgm)

---

* Indicates an essentially fixed locus with the frequency of the common allele > 0.95

t Indicates a locus showing significant inter-host allele frequency heterogeneity at the Grant, Michigan site.

showed significant allele frequency differences between flies collected from sympatric hawthorn and apple trees at the Grant, Michigan site (Tables 9, 10). Allele frequencies for *Acon-2* and *Me* were significantly different between hawthorn and apple flies within each of the three years sampled and did not vary significantly among either apple or hawthorn trees within any year (Tables 9, 10). *Aat-2*, *Dia-2* and *Mpi* also showed host related differences. However, gene frequencies varied among flies from different apple trees for *Aat-2* in 1985 and for *Dia-2* in 1984 and 1985 as well as among hawthorn trees for *Mpi* in 1985 (Tables 9, 10). Allele frequencies for *Had* differed significantly between hawthorn and apple flies in 1985 but not in 1984 or 1986. *Had* also varied significantly among apple trees in 1984. None of the other 7 polymorphic loci resolved in this study showed significant inter- or intra-host frequency differences.

Allele frequencies were reasonably constant between 1984 and 1985. Of the 13 polymorphic loci, only apple tree #1 for *Me* (G=9.67, $\underline{P} \leq 0.01$, 1 df) and hawthorn tree #1 for *Dia-2* (G=4.66, $\underline{P} \leq 0.05$, 1 df) had significant frequency differences between 1984 and 1985 (data for *Acon-2*, *Me*, *Mpi*, *Aat-2*, *Dia-2* and *Had* are presented in Table 10). However, eight out of a total of 20 tests for *Me*, *Acon-2*, *Mpi*, *Aat-2* and *Dia-2* showed significant differences between 1985 and 1986. Allele frequencies for both apple and hawthorn flies displayed a general trend to become more "apple-like" in 1986 compared to previous years (Table 9). The shift in 1986 allele frequencies could be explained by selection either in the field or during laboratory rearing as flies collected as larvae in 1985 and adults in 1986 represent different life history stages of the same generation. However, frequency shifts between 1985 and 1986 were in the same direction and of

Table 9. Allele frequencies for *Me*, *Acon-2*, *Mpi*, *Aat-2*, *Dia-2* and *Had* for flies collected from hawthorn and apple trees at the Grant,Michigan study site. n = sample size.

YEAR	HOST	Me		Acon-2			Mpi			Aat-2			Dia-2		Had	
(Stage)		n	80	n	100	95	n	100	37	n	100	21	n	100	n	100
1984	Apple-1	99	0.707	99	0.545	0.152	99	0.899	0.056	99	0.279	0.187	96	0.688	99	0.803
(Larvae)	Apple-2	84	0.631	86	0.512	0.221	86	0.884	0.046	86	0.326	0.157	86	0.767	86	0.907
1985	Apple-1	25	0.480	25	0.480	0.240	25	0.900	0.060	25	0.300	0.120	24	0.729	25	0.740
(Larvae)	Apple-2	52	0.587	62	0.484	0.185	65	0.869	0.085	52	0.393	0.183	63	0.802	65	0.838
	Apple-3	51	0.618	50	0.511	0.220	50	0.880	0.060	51	0.294	0.157	50	0.650	51	0.726
1986	Apple-1	46	0.663	46	0.598	0.109	46	0.946	0.000	46	0.272	0.130	46	0.669	46	0.837
(Adults)	Apple-2	39	0.603	39	0.584	0.167	39	0.897	0.000	39	0.269	0.141	39	0.628	39	0.859
1984 (Larvae)	Haw-1	101	0.272	101	0.297	0.490	101	0.703	0.178	101	0.465	0.104	100	0.895	101	0.876
1985	Haw-1	52	0.269	64	0.250	0.500	65	0.785	0.085	51	0.372	0.088	64	0.797	65	0.885
(Larvae)	Haw-2	53	0.274	53	0.273	0.500	53	0.717	0.216	53	0.415	0.066	53	0.877	53	0.877
	Haw-3	51	0.225	52	0.250	0.645	52	0.683	0.183	50	0.410	0.030	52	0.865	52	0.875
1986	Haw-1	61	0.385	61	0.443	0.410	61	0.795	0.148	61	0.402	0.090	61	0.820	61	0.813
(Adults)	Haw-2	63	0.444	63	0.365	0.365	63	0.787	0.131	63	0.460	0.048	63	0.802	63	0.841

Table 2 is abbreviated and contains the subset of alleles showing the highest levels of inter-host differentiation.

Table 10. G-contingency tests for allele frequency differences for *Me*, *Acon-2*, *Mpi*, *Aat-2*, *Dia-2* and *Had* at the Grant, Michigan site. Degrees of freedom are given in parentheses.

Year/(Stage)	Test	Me	Acon-2	Mpi	Aat-2	Dia-2	Had
1986	Among Apple trees	0.67 (1)	1.44 (3)	1.39 (2)	5.72 (4)	0.86 (1)	0.16 (1)
(Adults)	Among Haw trees	0.90 (1)	6.71 (3)	0.65 (2)	2.34 (4)	0.13 (1)	0.61 (1)
	Between Apple-Haw	19.70 (1)***	34.60 (3)***	38.10 (2)***	14.79 (4)**	11.32 (1)***	0.44 (1)
1985	Among Apple trees	3.07 (2)	3.71 (6)	0.96 (4)	16.45 (8)*	6.55 (2)*	4.90 (2)
(Larvae)	Among Haw Trees	0.78 (2)	10.67 (6)	11.84 (4)*	10.98 (8)	3.33 (2)	0.06 (2)
	Between Apple-Haw	60.70 (1)***	44.10 (3)***	-	-	-	10.97 (1)***
	Apples-Haw Tree 1	-	-	7.10 (2)*	-	-	-
	Apples-Haw Tree 2	-	-	15.50 (2)***	-	-	-
	Apples-Haw Tree 3	-	-	18.50 (2)***	-	-	-
	Apple Tree 1-Haws	-	-	-	5.43 (4)	3.47 (1)	-
	Apple Tree 2-Haws	-	-	-	21.22 (4)***	1.11 (1)	-
	Apple Tree 3-Haws	-	-	-	16.93 (4)**	16.36 (1)***	-
1984	Among Apple Trees	2.39 (1)	6.32 (3)	1.13 (2)	2.17 (4)	4.14 (1)*	8.12 (1)**
(Larvae)	Between Apple-Haw	85.70 (1)***	59.30 (3)***	32.40 (2)***	31.10 (4)***	-	-
	Apple Tree 1-Haws	-	-	-	-	26.42 (1)***	4.01 (1)*
	Apple Tree 2-Haws	-	-	-	-	9.11 (1)**	0.91 (1)
Between	Apple Tree 1	5.01 (1)*	6.30 (3)	6.47 (2)*	2.93 (4)	0.17 (1)	1.87 (1)
1985/1986	Apple Tree 2	0.05 (1)	3.19 (3)	12.60 (2)**	17.98 (4)**	7.30 (1)**	0.16 (1)
(Larvae/	Haw Tree 1	3.43 (1)	15.50 (3)**	5.25 (2)	0.88 (4)	0.21 (1)	3.20 (1)
Adults)	Haw Tree 2	7.33 (1)**	9.39 (3)*	3.00 (2)	6.52 (4)	2.45 (1)	0.62 (1)
Between	Apple Tree 1	9.67 (1)**	5.60 (3)	0.04 (2)	3.75 (4)	0.62 (1)	0.92 (1)
1984/1985	Apple Tree 2	1.45 (1)	2.72 (3)	1.22 (2)	5.52 (4)	0.50 (1)	3.19 (1)
(Larvae)	Haw Tree 1	0.16 (1)	4.85 (3)	5.97 (2)	4.12 (4)	4.66 (1)*	0.05 (1)

---

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-contingency test. Gene frequencies for non-significant intra-host tests were pooled across host trees and these pooled totals were used for inter-host tests (e.g., Between Apple-Haw). When significant intra-host heterogeneity was observed for a locus, inter-host tests were conducted on a tree by tree basis against the pooled total for the other host (e.g., Apples-Haw tree 1).

similar magnitude for both hawthorn and apple flies (Table 9). Therefore, if selection did occur, its effects were uniform across host plants.

Allele frequencies for *Acon-2*, *Me*, *Mpi*, *Aat-2*, *Dia-2* and *Had* may not be evolving independently. Hitchhiking effects for linked loci and epistatic selection can coordinately change allele frequencies at a number of loci. To examine these possibilities, correlation coefficients were calculated between pairs of non-allelic genes based upon Burrows disequilibrium values (Cockerham and Weir, 1977; Laurie-Ahlberg and Weir, 1979). Significant disequilibrium was found between *Me/Acon-2*, *Me/Mpi*, *Acon-2/Mpi* and *Aat-2/Dia-2* (Table 11). These results agree with the known genetic map of *R. pomonella* as *Aat-2* and *Dia-2* are 3.2 centimorgans apart on linkage group I while *Acon-2*, *Me* and *Mpi* are very tightly linked on linkage group II (see Chapter 1). Loci showing genetic differentiation between hawthorn and apple flies therefore map to three different regions of the genome, and hitchhiking effects are likely among non-allelic genes within at least two of these regions.

Interactions between unlinked loci were not readily apparent from the correlation coefficients. Twenty-one significant tests of gametic disequilibrium were observed in 314 pairwise comparisons between *Mpi*, *Acon-2* or *Me* and either *Aat-2* or *Dia-2* (data not shown). *Had*, which has been mapped to linkage group III (see Chapter 1 and Berlocher and Smith, 1983), was in gametic equilibrium with either *Acon-2*, *Me*, *Mpi*, *Aat-2* or *Dia-2* in 153 of 160 tests. Although inter-chromosomal levels of gametic disequilibrium cannot be completely explained by random type I errors, no consistent pattern was observed in either the loci or alleles involved in the significant tests. Allele frequencies for unlinked loci, therefore, appear to be evolving independently in both hawthorn and apple populations.

Table 11. Correlation coefficients (r^{AB}) calculated between pairs of non-allelic genes based upon disequilibrium values.

Year/ (Stage)	Host	Me^{100}/Acon-2^{95}	Me^{80}/Acon-2^{100}	Me^{100}/Mpi^{100}	Mpi^{100}/Acon-2^{95}	Aat-2^{21}/Dia-2^{85}
1984	Apple 1	0.3031***	0.2417**	−0.2823*	−0.3137***	0.5045***
(Larvae)	Apple 2	0.4272***	0.2529*	−0.1276	−0.1755	0.4752***
	Haw 1	0.2770**	0.2303*	−0.1980*	−0.3039**	0.6692***
1985	Apple 1	0.2327	−0.0033	−0.3770*	−0.4354***	0.6991**
(Larvae)	Apple 2	0.3314*	0.1132	−0.1997	−0.1782	0.2935*
	Apple 3	0.5097***	0.2806*	−0.1833	−0.2660	0.6337**
	Haw 1	0.4634**	0.2837*	−0.2780*	−0.2285	0.5082***
	Haw 2	0.3723**	0.4356***	−0.2731	−0.1681	0.5153***
	Haw 3	0.3951*	0.4481**	−0.3420*	−0.2488	0.5357***
1986	Apple 1	0.3311**	0.1605	−0.1679	−0.3311*	0.5513***
(Adults)	Apple 2	0.3229**	0.1231	−0.2724	−0.0362	0.4481***
	Haw 1	0.4551**	0.3552**	−0.3445**	−0.3017*	0.4990***
	Haw 2	0.5882***	0.3173*	−0.1853	−0.4217***	0.5267***

---

* $P \leq .05$ ** $P \leq .01$ *** $P \leq .001$ Significance determined by Chi-squared goodness of fit tests of Burrows value.

Discussion

Several factors, alone or in combination, could be responsible for the observed allele frequency differences between hawthorn and apple flies. Among the possibilities are: (1.) Post mating reproductive isolation between hawthorn and apple flies; (2.) A genetic bottleneck associated with the founding of the apple race; (3.) Differential larval survivorship associated with the host fruit environment; (4.) Differential host recognition by adult flies; and (5.) Temporal differences in the timing of adult emergence.

It is very unlikely that post mating reproductive isolation is involved in differentiating host populations. Hawthorn and apple flies readily mate in the laboratory and produce viable F_1 progeny (R. Prokopy, pers. comm.). Percent egg hatch is similar for apple x apple, hawthorn x hawthorn and hawthorn x apple test crosses (Reissig and Smith, 1978). Although data from F_2 and backcrosses are needed to completely rule out postmating isolation, the likelihood of reproductive incompatibility between these flies is remote.

A genetic bottleneck during the colonization(s) of apple by *R. pomonella* may have initially caused gene frequency differences between the host races due to drift. However, for neutral genetic differences to persist requires restricted gene flow between sympatric apple and hawthorn populations. Variation for intrinsic biological factors (i.e., differential larval survivorship, host recognition and/or adult emergence) are therefore necessary to account for the continued maintenance of host race differentiation regardless of its original cause.

Although differential selection on larval populations infesting hawthorn and apple fruits could account for genetic divergence, larval selection alone cannot explain the significant frequency differences observed between adults captured directly from hawthorn and apple fruits in 1986 (Table 9, 10). Random dispersal of flies between host plants would homogenize adult gene frequencies even if larval selection was intense (Felsenstein, 1981). Furthermore, adult hawthorn and apple flies show the same pattern and magnitude of host related differentiation as larvae. Gene flow, therefore, must be restricted between hawthorn and apple populations indicating that the host races are not randomly mating in nature.

Traits involved in host recognition and the timing of adult eclosion are the most likely candidates responsible for reducing gene flow between hawthorn and apple races. Host acceptance behaviors do differ between hawthorn and apple flies (Prokopy *et al.*, 1988) and prior adult experience can affect host preference in *R. pomonella* (Prokopy *et al.*, 1982). Field studies have shown that hawthorn and apple flies differ in their mean adult emergence times by approximately a week and a half in the field (Diehl, 1984; Feder *et al.*, manuscript in preparation) and by even more under laboratory conditions (Smith, 1986, 1988; Feder *et al.*, manuscript in preparation). Also, the distribution of adults at the Grant site closely follows the fruiting phenology of their host plants as early in the summer flies are abundant on ripe apples but scarce on immature hawthorns while the reverse is true 3-4 weeks later in the season when hawthorns ripen (see Figure 2 for adult collecting dates). Differences in diapause termination and host preference, together with adult conditioning, could all contribute to the establishment of an assortative mating system with early emerging adults tending to reproduce on apples and later emerging flies on hawthorns.

The existence of genetic differentiation between sympatric hawthorn and apple populations of *R. pomonella* at the Grant, Michigan site confirms the status of these flies as non-randomly mating host races. The recent formation of the apple race indicates that these flies are diverging sympatrically in the absence of geographic isolation. Additional research is needed, however, to clarify which biological factors are most responsible for creating and maintaining host related frequency differences at the Grant, Michigan site.

The geographic pattern of genetic differentiation between apple and hawthorn populations of *R. pomonella* appears to be complex, however. McPheron *et al.* (1988a) have also documented significant inter-host frequency differences at a sympatric site in Urbana, Illinois. Their results differ slightly from mine, however, in the pattern of loci and frequencies of alleles displaying inter-host differentiation. In addition, genetic analysis of geographic variation across the eastern United States (see Chapters 3, 4) indicates that north-south allele frequency clines exist for both apple and hawthorn flies across their ranges for five of the six loci showing host associated heterogeneity. Inter-host differences are therefore superimposed on latitudinal patterns of variation within the hawthorn and apple races and these clines help to explain the genetic differences between sites in Michigan and Illinois. The results from the geographic survey raise the question of whether sympatric mechanisms alone account for the development of complete reproductive isolation between *R. pomonella* populations. We consequently feel that it is inappropriate to definitively state that hawthorn and apple races represent "incipient" species. However, the *R. pomonella* complex does contain a number of sympatrically distributed sibling species specialized on different host plants which are inter-fertile in laboratory crosses, yet remain distinct in nature (Bush, 1966; Berlocher and Bush, 1982; Feder and Bush, submitted; Feder *et al.*, submitted).

Chapter 3

The Geographic Pattern of Genetic Differentiation between Host Races of *Rhagoletis pomonella*

Introduction

True fruit flies belonging to the *Rhagoletis pomonella* sibling species group are at the center of a long standing controversy concerning modes of speciation. Speciation in animals has traditionally been assumed to require the complete geographic separation of populations in order to eliminate the homogenizing effects of gene flow (Mayr, 1963). However, Bush (1966, 1969a,b, 1975a) has proposed that certain species groups in the genus *Rhagoletis* diverge sympatrically via shifts to new host plants. Because *Rhagoletis* larvae are host specific fruit parasites and mate selection in these flies is directly coupled to host plant recognition (Prokopy *et al.*, 1971, 1972), variation for traits related to host preference and larval survivorship can produce genetically based barriers to gene flow. This reduction in gene flow can result in the sympatric formation of partially reproductively isolated "host races" and eventually species adapted to different host plants.

The recent sympatric shift of the species *R. pomonella* from its native host hawthorn (*Crataegus* spp.) to domestic apple (*Malus pumila*) provides an opportunity to determine whether host specialization can differentiate fly populations in the absence of prior periods of geographic isolation.

Rhagoletis pomonella was first reported attacking domestic apples in the Hudson Valley, New York in the 1860's (Ward, 1866; Walsh, 1867). Agricultural records document the subsequent spread of the "apple fly" from New York throughout eastern North America over the last 120 years (Illingworth, 1912; O'Kane, 1914; Porter, 1928; Bush, 1969a; see Figure 3). The implication that apple flies originated from just a single or relatively few colonization events may be misleading. The spread of the apple fly from New York could also have resulted from several independent host shifts corresponding to expansions in the range of domestic apple cultivation. It is clear, however, that hawthorns are the original, endemic host plant of *R. pomonella*. Although several native species of crabapple exist in North America, none support populations of the fly (Illingworth, 1912; O'Kane, 1914; Porter, 1928; Bush, 1966). The reason for this has not been adequately studied but it has been suggested that fruits of native crabapple species ripen too late for *R. pomonella* to complete development before the onset of winter (O'Kane, 1914). The historical timeframe of the host shift is also reliable as domestic apples were introduced into North America from Europe only within the last 350 years (Chapman and Lienk, 1971). Therefore, at no time during the formation of the "apple race" was it geographically separated from hawthorn infesting populations.

In Chapter 2, I reported the finding of genetic differentiation between sympatric apple and hawthorn populations of *R. pomonella* at a field site in western Michigan. Six allozyme loci displayed significant host related frequency differences for both adult and larval life history stages of the fly (Tables 9, 10). These six loci mapped to three different regions of the genome (Chapter 1) and significant disequilibrium was observed

Figure 3. Collecting sites for *R. pomonella* across the eastern United States and Canada (see Table 12 for descriptions of study sites). Dotted portion of map designates the region where flies were first observed attacking apples. Dotted and shaded areas encompass the distribution of apple flies when the first published report of the race was made by Ward (1866). Years are given for the first observation of the apple race in a region. The current distributions of the apple and hawthorn races are indicated by the solid and dashed lines, respectively. Figure 3 is a modification of Figure 3 from Bush (1969a).

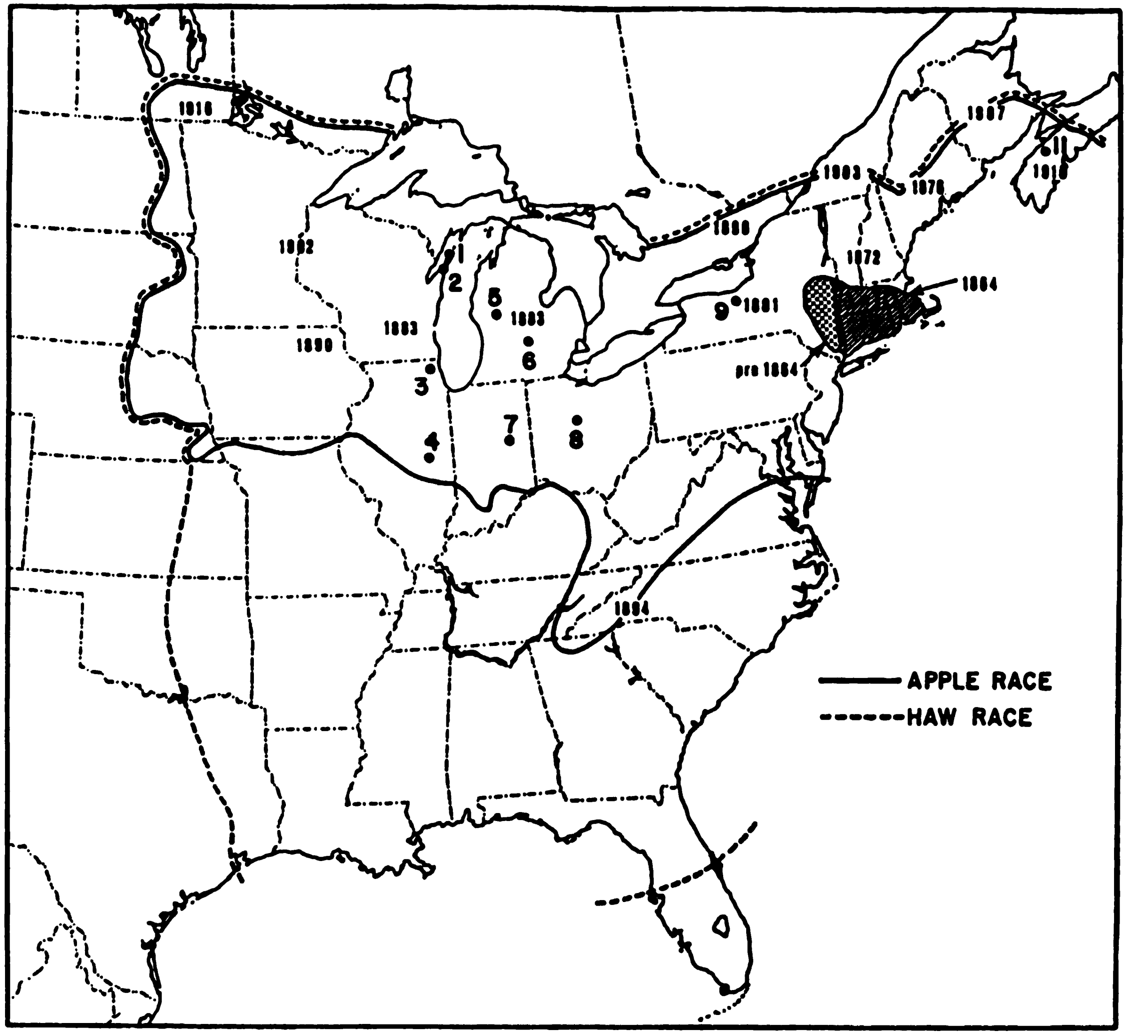

1916
1902
1883
1890
1883
1
2
3
4
5
6
7
8
9
1881
1880
1903
1876
1897
11
1916
1872
1864
pre 1864
1894
APPLE RACE
HAW RACE

Figure 3.

between non-allelic genes within, but not between, two of these regions (Table 11). Hawthorn and apple populations at the Michigan site therefore represent partially reproductively isolated "host races", a result supporting a sympatric mode of divergence for these flies.

Several critical questions concerning the geographic pattern of genetic variation for apple and hawthorn flies still need to be answered, however. The first issue is whether host associated differentiation is widespread across the range of *R. pomonella*. In this regard, McPheron *et al.* (1988a) have also found significant inter-host allele frequency differences between apple and hawthorn flies at a sympatric site in Urbana, Illinois. Are these two sites representative of host related divergence for *R. pomonella* across eastern North America or do apple and hawthorn flies differ on just a local or regional scale?

The second question is whether any trends exist in the partitioning of genetic variation between and among apple and hawthorn populations and, if so, whether these patterns point to factors responsible for differentiating the host races. For instance, do the same loci consistently display inter-host heterogeneity between apple and hawthorn flies across sites or are different genes involved at every site? Differences do exist in the loci and alleles showing host associated differentiation between *R. pomonella* populations from Michigan and Illinois (see Chapter 2 and McPheron *et al.*, 1988a). For example, *Diaphorase-2* displayed higher levels of inter-host heterogeneity in Michigan than in Illinois populations, while the reverse was true for *Hydroxyacid dehydrogenase*. In addition, *Aconitase-2* ^{95} allele frequencies were reversed between apple and hawthorn populations in Michigan and Illinois. Are these the patterns of variation expected if differential selection were occurring between the host races or is genetic drift a more plausible explanation? Significant linkage disequilibrium was also observed among loci displaying inter-host differentiation at the Michigan site. How common is linkage disequilibrium in *R. pomonella* populations? If certain non-allelic genes are regularly in disequilibrium does this mean that host related selection is causing specific regions of the genome to co-evolve?

Finally, the apple race reportedly arose in the Hudson Valley, New York and subsequently spread across the rest of eastern North America. Is there any genetic evidence for this? For example, is the apple race less variable across its range than the hawthorn race as would be expected if the apple race owed its origin to just a single, or relatively few, founder events? Do peripheral apple populations possess alleles characteristic of hawthorn populations near the supposed home of the apple race in New York or are they genetically more similar to local hawthorn populations, as would be predicted by the multiple shift hypothesis? Answers to all of the above questions bear directly on the evolutionary significance of sympatric host race formation in the *R. pomonella* group.

In this Chapter, I report the results of an electrophoretic study of allozyme variation for hawthorn and apple populations of *R. pomonella* from across the eastern United States and Canada. The findings indicate that significant genetic differentiation exists between hawthorn and apple flies across most of the range of *R. pomonella*. The same set of six loci are consistently responsible for host related differences, and five of these six loci show a pattern of geographic variation indicative of latitudinal allele frequency clines in the two host races.

Materials and Methods

Flies were collected as larvae from infested fruit at ten paired feral apple (*Malus pumila*) and hawthorn (*Crataegus mollis*) sites across the eastern United States as well as from an apple site in Kentville, Nova Scotia (see Figure 3 and Table 12 for locations and descriptions of study sites). Sympatric site 5 (Grant, Michigan) represents the same apple and hawthorn populations I analyzed in Chapter 2. Likewise, flies surveyed from site 4 (Urbana, Illinois) were from the same populations, but were not the same individuals, that McPheron *et al.* (1988a) studied. Multiple year samples were collected for six of the ten paired sites (Table 12). At site 5 in 1987 and site 7 (Gas City, Indiana), larvae were dissected from infested fruit and were immediately frozen in liquid nitrogen for later electrophoretic analysis. Otherwise, we brought infested fruit back into the laboratory and allowed larvae to pupate in moist vermiculite. Pupae were sifted from the vermiculite and subsequently stored at 4° C to simulate winter diapause conditions. After five months, we removed the pupae from the cold and held them at 22° C, under a 15:9 hr. light:dark cycle to terminate diapause. Eclosing adults were collected and stored at -80° C until they were analyzed electrophoretically.

Standard horizontal starch gel electrophoretic techniques were used and are described fully in Chapter 1 and in Berlocher and Bush (1982). I stained for a total of 21 different enzyme systems and was able to resolve 29 allozymes (see Table 8 for a list of resolved loci and their abbreviations). Isozymes which migrated the nearest to the cathode were designated 1, the second closest 2, etc. Alleles at each locus were numbered according to their relative anodal mobilities, with the most common allele at a locus

Table 12. Collecting sites for *R. pomonella* across the eastern United States and Canada. A=apple, H=hawthorn, sym = paired site where hawthorn and apple trees sampled were within 100 meters or less of one another.

Site #	Location	Years Collected	Fruit Collected	Distance (km) Seperating Host Plants	Latitude
1.	Hogan farm, Ephraim, Wisc.	84-86	A,H	Sym	45°07'
2.	Kuehn farm, Carlsville, Wisc.	84,85	A,H	Sym	44°56'
3.	Hyw. 41, near Waukegan, Ill.	84,86	A,H	3.3	42°20'
4.	Univ. Ill. campus, Urbana, Ill.	86	A,H	Sym	40°05'
5.	112th St., near Grant, Mich.	84-87	A,H	Sym	43°21'
6.	Mich. St. Univ. campus, E. Lansing, Mich.	84,85	A,H	1.5	42°41'
7.	Hyw. 69, near Gas City, Ind.	87	A,H	6.0	40°31'
8.	Hyw. 31, near Kenton, Ohio	86	A,H	5.0	40°39'
9.	N.Y.S. Ag. Exp. Station, Geneva, N.Y.	85	A,H	Sym	42°53'
10.	Univ. Mass. campus, Amherst, Mass.	85,86	A,H	1.0	42°24'
11.	Ag. Canada Res. Station, Kentville, Nova Scotia, Canada	85	A	—	44°57'

assigned a value of 100 and used as a standard (Note: The relative mobilities of electrophoretic alleles resolved for the 13 polymorphic loci are given in Tables 21-32 which appear after the Discussion section on pages 91-102).

Significant deviations from Hardy-Weinberg genotypic expectation were determined by G-tests. Alleles were pooled, if required, so that the expected frequency of all genotypic classes were >1. The Levene correction (Spiess, 1977) was applied if sample sizes were <100 or if genotypic classes had expected numbers <1. G-contingency tests were used to test for allele frequency heterogeneity between host populations and across sites and years. Alleles for a locus were pooled, when necessary, to ensure that each cell in the G-contingency test had an observed number ≥ 5. F-statistics were calculated by the method of Weir and Cockerham (1984) with variances estimated by jackknifing over populations, loci or years. Corrections for unequal sample sizes in Weir and Cockerham's method can result in negative F_{ST} values. In these instances we report F_{ST} as 0. Correlation coefficients (r^{AB}) were calculated between pairs of non-allelic genes based upon Burrow's disequilibrium values (Cockerham and Weir, 1977). To test the null hypothesis $r^{AB}=0$, the correlation coefficients were z transformed to approximate a normal distribution with variance:

$$\sigma_z^2 = 1 / n-3.$$

Arcsine transformed allele frequencies for apple and hawthorn populations were linearly regressed against latitude for the ten paired sites (Note: At sites with multiple year samples I transformed the mean allele frequency for the population averaged across years). I tested the null hypothesis that the regression coefficient (i.e., slope of the linear regression equation) = 0 by ANOVA using F-tests. Curvilinear regressions were also performed based upon second order polynomials but F-tests revealed that they did not significantly improve the fit of regression lines compared to first order equations.

Results

Thirteen of the 29 allozyme loci resolved were polymorphic (frequency of the common allele $\leq$ 0.95, Table 8). A Mendelian mode of inheritance has been established in *R. pomonella* for 12 of these polymorphic loci with *Aat-1* being sex-linked (see Chapter 1; Berlocher and Smith, 1983). Excluding *Aat-1*, 30 significant deviations from Hardy-Weinberg equilibrium were observed in 670 tests (Tables 21-32). The significant deviations displayed no regular pattern and were fewer than the number expected due to random type I error (30 compared to 33.5). Genotypic frequencies for autosomal loci were, therefore, generally in Hardy-Weinberg equilibrium within both apple and hawthorn populations. This was not true for the sex-linked locus *Aat-1* in that 27 out of 33 populations showed significant departures from equilibrium proportions. However, when females (which are the homogametic sex in *R. pomonella*; see Chapter 1) were considered separately from males (which are hemizygous and therefore should not be included in calculations of Hardy-Weinberg equilibrium), females were generally in Hardy-Weinberg equilibrium for *Aat-1* (2 significant deviations out of 33 total tests).

R. pomonella flies infesting apples were genetically different from those infesting hawthorns. Although no fixed allele difference was found for any of the 29 loci resolved, significant allele frequency differences were observed between apple and hawthorn populations at nine of the ten paired collecting sites (Tables 13, 21-32; Note: Paired sites 1-10 are organized by latitude from north to south in these three Tables). Only site 8 near Kenton, Ohio did not differ significantly for at least 1 of the 13 polymorphic loci analyzed.

Host associated genetic variation displayed a distinct dichotomy. Six loci (*Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* and *Had*) consistently showed significant host related differences (60 out of a total of 113 G-contingency tests were significant for these six loci; Table 13). In contrast, the remaining seven polymorphic loci analyzed in the study had much lower levels of inter-host variation. Only 10 out of 113 tests were significant for these seven loci and there was no trend evident in either the loci or alleles involved in the significant differences.

Inter-host differentiation was stable over time for the six sites where multiple year collections were made. At site 5 (Grant, Michigan) for instance, *Me*, *Acon-2*, *Mpi*, *Dia-2* and *Aat-2* all showed significant allele frequency heterogeneity between apple and hawthorn flies over a 4 year period from 1984 to 1987 (Table 13). The only other locus which varied significantly between host populations at this site was *Had* in 1985.

Allele frequencies within apple or hawthorn populations fluctuated across years, however (Table 14). This was especially true for hawthorn flies in which 16 out of 83 G-contingency tests indicated significant heterogeneity over time (Table 14). Intra-host temporal variation did not overlap in frequency or compare in magnitude to inter-host differences, however. F-statistics calculated across years for *Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* and *Had* at site 1 and site 5 document this point (Tables 15, 16). Overall F_{ST} values for these six loci across years between host populations

Table 13. G-contingency tests for allele frequency heterogeneity at paired sites between hawthorn and apple infesting populations of *R. pomonella*. Degrees of freedom for tests are given in parentheses.

Site	Year	Me(1)	Acon-2(2)	Mpi(2)	Dia-2(1)	Aat-2(4)	Had(1)	7 other polymorphic loci
1	84	**	*					
	85	***	**	*				Aat-1(1)*
	86	***	***	***	**			——
2	84	*					***	
	85	***	***	***	***	***	**	
5	84	***	***	***	***	**		
	85	***	***	***	***	***	**	
	86	***	***	***	***	***		——
	87	***	***	***	***	***		——
9	85	***	**		***	**	**	Aat-1(1)*, Acy(1)**
6	84		***	**				Aat-1(1)*
	85	*	***		**			
10	85	***			*			Pep-2(2)**
	86						*	
3	84	**			***	*	*	Pgm(1)*
	86	***				*		Idh(1)**
8	86				—			
7	87						**	Pgi(1)**, Acy(1)*
4	86		*				**	Ak(1)*

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G- contingency test

Table 14. G-contingency tests for intra-host allele frequency heterogeneity across years and sites. Degrees of freedom for tests are given in parentheses.

Across Years		Loci tested						
Site	Host	Me	Acon-2	Mpi	Dia-2	Aat-2	Had	7 other polymorphic loci
1	A							Pgm(2)*
	H		(4)*					Acy(1)**
2	A						(1)*	———
	H			(2)*	(1)*	(4)*		
5	A					(12)***		
	H	(3)***	(6)***	(6)**	(3)**		(3)*	Ak(1)*
6	A							
	H	(1)*	(2)*					Pep-2(3)**
10	A							Pep-2(2)***
	H							
3	A							Aat-1(1)**
	H							Aat-1(1)***, Idh(1)**
Across Sites								
Apple		(9)***	(18)***	(9)***	(8)**	(27)***	(9)***	Pep-2(18)***, Acy (9)**, Aat-1(9)***
Hawthorn		(9)***	(18)***	(9)***	(8)***	(27)***	(9)***	Pep-2(18)*, Aat-1(9)**

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G- contingency test

Table 15. F_{ST} values across years for apple and hawthorn populations at site 1 (Ephraim, Wisconsin).

	Site #1 (1984-86)		
Locus	**Apple x Y**	**Hawthorn x Y**	**Hosts x Y**
Me	0.0000 ± 0.0016[a]	0.0114 ± 0.0208	0.0796 ± 0.0241
Acon-2	0.0042 ± 0.0122	0.0074 ± 0.0078	0.0541 ± 0.0143
Mpi	0.0000 ± 0.0077	0.0000 ± 0.0015	0.0000 ± 0.0065
Dia-2	0.0000 ± 0.0057	0.0000 ± 0.0007	0.0072 ± 0.0081
Aat-2	0.0000 ± 0.0027	0.0072 ± 0.0107	0.0057 ± 0.0038
Had	0.0053 ± 0.0151	0.0080 ± 0.0133	0.0039 ± 0.0042
Total	0.0000 ± 0.0027[b]	0.0054 ± 0.0020	0.0284 ± 0.0163

---

[a] Standard deviation - Jackknife estimate over years.

[b] Standard deviation - Jackknife estimate over loci.

[c] Hosts x Years (Y) are F_{ST} values when both hawthorn and apple populations are included in the calculations. Hosts x Y is therefore representative of the interaction between host associated and temporal genetic variation.

Table 16. F_{ST} values across years for apple and hawthorn populations at site 5 (Grant, Michigan).

	Site #5 (1984-87)		
Locus	**Apple x Y**	**Hawthorn x Y**	**Hosts x Y**
Me	0.0019 ± 0.0038	0.0259 ± 0.0270	0.0960 ± 0.0235
Acon-2	0.0015 ± 0.0025	0.0099 ± 0.0095	0.0604 ± 0.0136
Mpi	0.0000 ± 0.0015	0.0133 ± 0.0082	0.0322 ± 0.0125
Dia-2	0.0025 ± 0.0036	0.0070 ± 0.0068	0.0247 ± 0.0074
Aat-2	0.0018 ± 0.0020	0.0014 ± 0.0020	0.0080 ± 0.0020
Had	0.0022 ± 0.0042	0.0032 ± 0.0034	0.0046 ± 0.0026
Total	0.0017 ± 0.0002	0.0084 ± 0.0040	0.0377 ± 0.0188

[a] Standard deviation - Jackknife estimate over years.

b Standard deviation - Jackknife estimate over loci.

[c] Hosts x Years (Y) are F_{ST} values when both hawthorn and apple populations are included in the calculations. Hosts x Y is therefore representative of the interaction between host associated and temporal genetic variation.

were over 4 x greater at both sites 1 and 5 than values calculated for hawthorn populations alone and 20 x greater than those for apple flies (Tables 15, 16).

Geographic variation was observed among both hawthorn and apple populations across the eastern United States and Nova Scotia. G-contingency tests indicated significant geographic variation for *Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat 1 & 2*, *Had* and *Pep-2* for both host races (Table 14). *Acy* also varied significantly among apple populations. F-statistics were calculated to quantify the degree of intra-host geographic variation (Table 17). Only data from a single year for hawthorn and apple populations from each of the ten paired collecting sites were used in the calculations (1986 being chosen whenever possible, with 1985 results used for sites 2 and 9 and 1987 data for site 7). Overall F_{ST} values for all 13 polymorphic loci were significantly greater than zero for both apple (0.0117 ± 0.0054) and hawthorn flies (0.0450 ± 0.0208) as indicated by standard deviations derived by jackknifing over loci. Geographic differentiation was much greater among hawthorn than apple populations, however, as evidenced by the nearly 4 fold larger overall F_{ST} value for hawthorn flies (see above).

Substantial heterogeneity existed among loci in their F_{ST} values for populations across the eastern United States (Table 17). For example, *Me*, *Acon-2* and *Had* all had a F_{ST} of over 0.100 for hawthorn flies while *Pgm*, *Acy*, *Ak*, *Idh* and *Aat-1* all had a F_{ST} of under 0.005 among the same hawthorn populations. In fact, the same six loci (*Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* and *Had*) which consistently showed host related differences also displayed the highest levels of geographic differentiation. If just these six loci are considered, apple populations had a F_{ST} of 0.0166 ± 0.0107 while hawthorn populations had a F_{ST} of 0.0821 ± 0.0383 (Table 17). In

Table 17. F_{ST} values across apple and hawthorn populations calculated by the method of Weir and Cockerham (1984). Negative F_{ST} values are reported as 0.0000.

Locus	Apple populations	Hawthorn populations
Dia-2	0.0042 ± 0.0049[a]	0.0676 ± 0.0404
Aat-2	0.0020 ± 0.0021	0.0122 ± 0.0067
Me	0.0173 ± 0.0094	0.1916 ± 0.0557
Acon-2	0.0378 ± 0.0172	0.1225 ± 0.0388
Mpi	0.0140 ± 0.0063	0.0426 ± 0.0185
Had	0.0208 ± 0.0066	0.1098 ± 0.0421
Pep-2	0.0137 ± 0.0068	0.0063 ± 0.0043
Idh	0.0000 ± 0.0019	0.0033 ± 0.0069
Ak	0.0000 ± 0.0034	0.0016 ± 0.0008
Pgm	0.0000 ± 0.0027	0.0024 ± 0.0031
Pgi	0.0062 ± 0.0070	0.0076 ± 0.0078
Aat-1	0.0197 ± 0.0150	0.0006 ± 0.0082
Acy	0.0024 ± 0.0029	0.0013 ± 0.0025
All 13 loci	0.0117 ± 0.0054[b]	0.0450 ± 0.0208
First 6 loci	0.0166 ± 0.0107	0.0821 ± 0.0383
Last 7 loci	0.0061 ± 0.0042	0.0036 ± 0.0019

---

[a] Standard deviation - jackknife estimate over populations

[b] Standard deviation - jackknife estimate over loci

contrast, F_{ST} values for the remaining seven polymorphic loci were 0.0061 ± 0.0042 for apple flies and just 0.0036 ± 0.0019 for hawthorn flies.

Despite the observed spatial diversity among *R. pomonella* populations, no obvious relationship existed between the geographic location of a population and the number of electrophoretic alleles scored for the population. Of the 11 sites we analyzed, site 10 (Amherst, Massachusetts) was located the closest to the Hudson Valley, New York, the supposed source of the apple race. Apple infesting flies were, in fact, already present in western Massachusetts at the time of the first published account of the new race by Ward in 1866 (Figure 3). Fifty-five electromorphs were observed in the apple population on the campus of the University of Massachusetts at Amherst for the 13 polymorphic loci resolved in this study. In comparison, apple populations near the northern and southern ends of the flies distribution in Door County, Wisconsin (site 1), and Urbana, Illinois (site 4), had 59 and 50 alleles, respectively, not appreciably different from the number scored at site 10. In fact, the number of alleles detected for apple populations 1, 10 and 4 was proportional to the sample size at each site (Tables 21-32). Furthermore, apple populations at all three of these sites contained either more or the same number of alleles as sympatric hawthorn populations (55, 53 and 50 for hawthorn populations 1, 10 and 4, respectively). The apple population at Kentville, Nova Scotia (site 11), did show a reduced amount of genetic variation (39 electromorphs). However, *Acy* was not scored for site 11. *Acy* had an average of four alleles in apple populations so the total number of alleles for Kentville should be adjusted to approximately 43. In addition, the sample size at site 11 (39 individuals) was smaller than that of any other population analyzed, which reduced the likelihood of detecting rare alleles at the Kentville site. I also do not have a paired hawthorn population for site 11 to verify whether the fewer number of alleles present in the Kentville apple population is due to a general reduction in genetic diversity for *R. pomonella* in this portion of its range. It therefore appears that if a population bottleneck did occur during the formation of the apple race, then either it was not severe enough to greatly reduce genetic variation in apple populations or subsequent gene flow from hawthorn populations has restored the original level of allelic diversity to the apple race.

Genetic variation was not randomly distributed among either apple or hawthorn populations across the eastern United States, however. Linear regression analysis revealed significant relationships between latitude and allele frequencies for *Me*^{100}, *Acon-2*^{95}, *Mpi*^{100}, *Had*^{100}and *Pep-2*^{100} across both host populations and for *Dia-2*^{100} across hawthorn populations (Table 18). Except for *Pep-2*^{100}, the absolute value of the slopes of the significant regression equations were always greater for hawthorn than apple flies (Table 18). For *Pep-2* ^{100} the slopes were essentially the same between the host races (Table 18). Hawthorn flies, therefore, displayed steeper north-south frequency clines for *Me*^{100}, *Acon-2*^{95}, *Mpi*^{100}, *Dia-2* ^{100} and *Had* ^{100} than apple flies; a result indicative of the higher levels of geographic differentiation observed in the hawthorn race.

The geographic pattern of inter-host differentiation is due in large part to the differences in the slopes of latitudinal allele frequency clines between hawthorn and apple populations. Figure 4 graphically depicts the relationship between latitude and the average arcsine transformed allele frequency for *Me*^{100}, *Acon-2*^{95}, *Mpi*^{37}, *Dia-2*^{100} and *Had*^{100} for 9 of the 10 paired sites in the eastern United States (Note: Site 8 is not included in Figure 4 because data for *Dia-2*^{100} are missing at this site). Inter-host

Table 18. Results of linear regression analysis of arcsin transformed allele frequencies and latitude across the 10 paired apple and hawthorn sites. r^2=explained genetic variation. Asterisk(s) denote significant deviations of regression coefficients (i.e., slopes of regression lines) from 0.

Locus	Allele	Apple populations		Hawthorn populations	
		r^2	Slope ± Std. Err.	r^2	Slope ± Std. Err.
Dia-2	100	0.245	0.687 ± 0.456	0.551*	3.400 ± 1.200
Aat-2	100	0.082	-0.424 ± 0.499	0.136	-0.713 ± 0.635
Idh	100	0.000	-0.003 ± 0.504	0.398	-0.614 ± 0.311
Ak	100	0.217	0.526 ± 0.353	0.040	-0.320 ± 0.558
Pgm	100	0.084	-0.332 ± 0.386	0.172	0.427 ± 0.331
Me	100	0.811***	2.500 ± 0.428	0.735**	6.500 ± 1.400
Acon-2	95	0.531*	3.300 ± 1.100	0.830***	6.700 ± 1.100
Mpi	100	0.569*	-1.700 ± 0.514	0.662**	-2.80 ± 0.718
Pep-2	100	0.615**	-1.700 ± 0.489	0.569*	-1.400 ± 0.444
Had	100	0.558*	1.900 ± 0.597	0.603**	4.300 ± 1.200
Pgi	100	0.176	0.855 ± 0.655	0.001	-0.007 ± 0.872
Aat-1	-100	0.086	-0.884 ± 0.662	0.284	1.200 ± 1.000
Acy	100	0.006	-0.171 ± 0.737	0.051	0.246 ± 0.377

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ ANOVA F-test

Figure 4. Average arcsin transformed allele frequency for Me^{100}, $Acon\text{-}2^{95}$, Mpi^{37}, $Dia\text{-}2^{100}$ and Had^{100} for paired apple and hawthorn sites plotted against latitude. Site 8 (Kenton, Ohio) is not included because data for Dia-2^{100} are missing for this site. The linear regression equation is given along with the amount of explained variation (r^2) and probability levels (P) for the ANOVA F-tests of the null hypothesis that the slope of the regression line = 0.

ARCSIN TRANSFORMED GENE FREQUENCY (ME, ACON-2, MPI, DIA-2, HAD)
vs.
LATITUDE

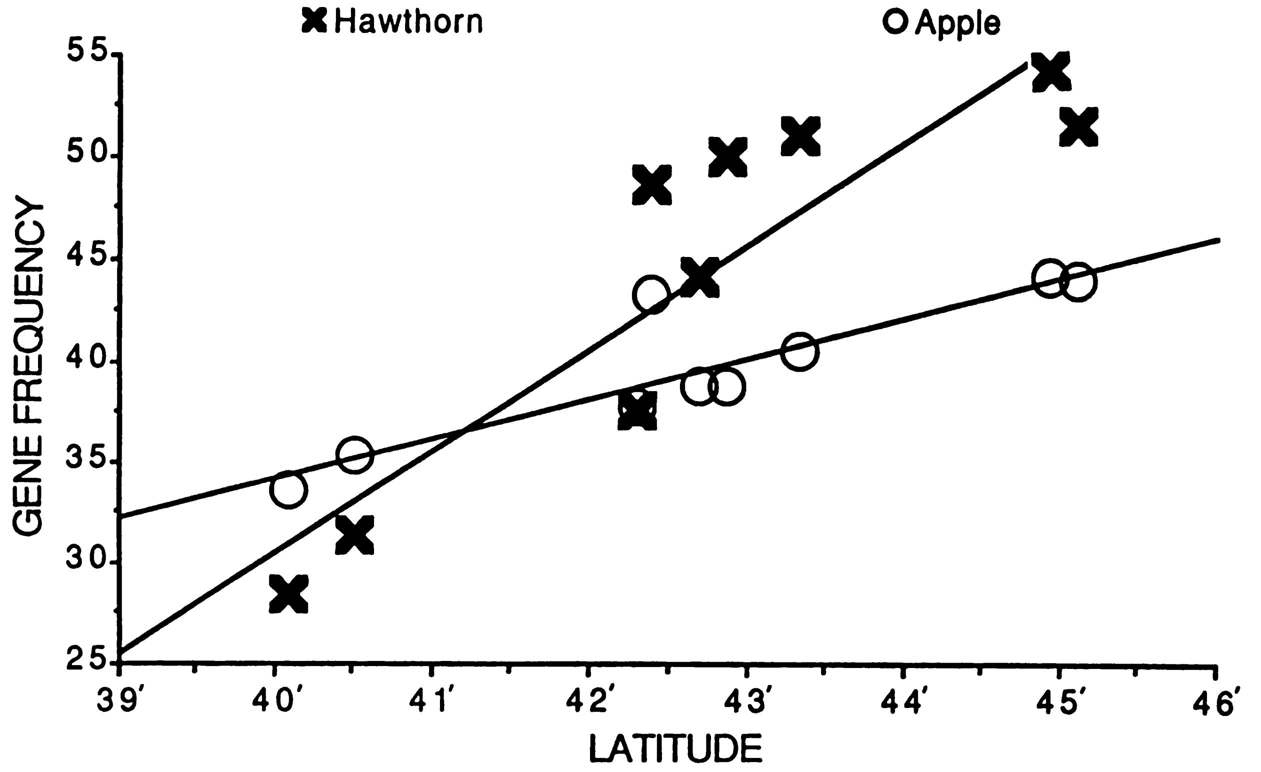


Figure 4.

differentiation is greatest at northern sites where allele frequencies for *Me*^{100}, *Acon-2*^{95}, *Mpi*^{37}, *Dia-2*^{100} and *Had*^{100} are generally higher in hawthorn than apple populations (Tables 21-25). At centrally located sites, frequency clines for the two host races converge and apple and hawthorn flies become less genetically distinct until at site 8 (latitude 40° 39') no locus shows significant inter-host heterogeneity (Table 13). Host associated allele frequency differences appear again at southern sites. However, in the south, frequencies for *Me*^{100}, *Acon-2*^{95}, *Dia-2*^{100}and *Had*^{100} tend to be lower in hawthorn than apple populations, the reverse of what was found for northern populations. The latitude of a site is, therefore, an accurate predictor of the magnitude and pattern of inter-host differentiation.

The fact that almost all of the loci showing host associated differences also co-vary significantly with latitude suggests that allele frequencies for these loci may not be evolving independently in fly populations. *Me*, *Acon-2*, *Mpi*, *Aat-2*, *Dia-2*, *Had* and *Pep-2* do, in fact, map to only three different regions of the genome (Chapter 1; Berlocher & Smith 1983). *Dia-2* and *Aat-2* are located 3.2 centimorgans apart on linkage group I while *Me*, *Acon-2* and *Mpi* are very tightly linked on linkage group II (no recombination was observed among *Me*, *Acon-2* and *Mpi* for 162 test cross progeny scored in Chapter 1). Finally, *Had* is on linkage group III at a distance of 23.9 centimorgans from *Pep-2*.

The linkage relationships for *Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2*, *Had* and *Pep-2* suggest that hitchhiking effects and/or epistatic genetic interactions are occurring among loci. To further examine this possibility, I calculated correlation coefficients between all pairs of non-allelic genes for these seven loci based on the method of Cockerham and Weir (1977).

Significant disequilibrium was found between *Me/Acon-2*, *Me/Mpi*, *Acon-2/Mpi*, *Aat-2/Dia-2* and *Had/Pep-2* for most apple and hawthorn populations (Tables 19 and 20). The patterns of disequilibrium between non-allelic genes were the same both within and between the host races throughout the eastern United States and Nova Scotia. Allele frequencies for linked loci within each of the three genomic regions showing inter-host differentiation are, therefore, co-evolving.

Linkage disequilibrium appears to exist only within chromosomal regions displaying host associated variation, however. *Idh* maps 7.1 centimorgans away from *Dia-2* and 10.3 centimorgans from *Aat-2* on linkage group I (Chapter 1). Nevertheless, *Idh* was in equilibrium with *Dia-2* in all 19 tests conducted. In addition, allele frequencies for *Idh* did not co-vary significantly with those for either *Ak* or *Pgm* which are located 8.9 and 38.7 centimorgans, respectively, from *Idh* on the opposite side of linkage group I as *Dia-2* and *Aat-2* (Chapter 1; Berlocher and Smith, 1983).

Interactions between unlinked loci were also not readily apparent from the correlation coefficients. Only 24 out of 627 correlation coefficients were significantly different from zero between pairs of unlinked loci for apple flies while 28 of 590 tests were significant for hawthorn flies (data not shown). No pattern was observed for either the loci or alleles involved in the significant tests. Allele frequencies for unlinked loci, therefore, appear to be evolving independently in both hawthorn and apple populations.

Table 19. Correlation coefficients between non-allelic genes (Cockerham and Weir, 1977) for apple populations across the eastern United States and Canada.

Linkage Group	Loci	Alleles	Apple Sites (Year) 11(85)	1(85)	2(85)	5(85)	9(86)	6(85)
I.	Dia-2/Aat-2	100/21	-0.61***	-0.56***	-0.46***	-0.59***	-0.44***	-0.63***
		100/50	-0.51***	-0.24	-0.32***	-0.35***	-0.43***	-0.41***
		100/75	0.34*	0.18	0.35***	0.34***	0.39***	0.41***
		100/100	0.42*	0.47***	0.30***	0.49***	0.42***	0.44***
II.	Me/Acon-2	100/95	0.44**	0.54***	0.49***	0.36***	0.20	0.17*
		100/100	-0.24	-0.54***	-0.41***	-0.15	-0.07	-0.09
		100/106	0.20	0.03	-0.01	0.17	0.03	0.15
	Me/Mpi	100/100	-0.41*	-0.44**	-0.17	-0.21*	0.13	-0.23**
	Mpi/Acon-2	100/95	-0.24	-0.40**	-0.16	-0.22*	-0.11	-0.31***
		100/100	-0.06	-0.33*	0.07	0.09	-0.01	0.15
		37/95	0.03	0.34*	0.23**	0.20*	0.26	0.24**
III.	Had/Pep-2	100/100	0.30	0.43**	0.44***	0.49***	0.51***	0.36***

* $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$ Significance determined by t-test of z transformed correlation coefficients.

Table 19. (cont'd.)

Linkage Group	Loci	Alleles	Apple Sites (Year)				
			10(85)	3(86)	8(86)	7(87)	4(86)
I.	Dia-2/Aat-2	100/21	-0.49**	-0.45***	---	-0.28	-0.55***
		100/50	-0.37*	-0.40**	---	-0.36*	-0.49***
		100/75	0.54***	0.33**	---	0.16	0.51***
		100/100	0.24	0.33**	---	0.32*	0.54***
II.	Me/Acon-2	100/95	-0.05	0.19	-0.38	0.01	-0.09
		100/100	-0.08	-0.19	0.34	-0.44**	-0.08
		100/106	0.12	0.15	0.58*	0.59***	0.26
	Me/Mpi	100/100	0.34*	-0.02	-0.09	-0.15	-0.03
	Mpi/Acon-2	100/95	0.26	-0.04	-0.11	-0.12	-0.16
		100/100	-0.25	-0.02	0.00	0.04	0.19
		37/95	0.00	---	---	---	---
III.	Had/Pep-2	100/100	0.61***	0.47***	0.57***	0.28	0.15

* $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$ Significance determined by t-test of z transformed correlation coefficients.

Table 20. Correlation coefficients between non-allelic genes (Cockerham and Weir, 1977) for Hawthorn populations across the eastern United States and Canada.

Linkage Group	Loci	Alleles	Hawthorn Sites (Year) 1(85)	2(85)	5(85)	9(86)	6(85)
I.	Dia-2/Aat-2	100/21	-0.29*	-0.35***	-0.36***	-0.50***	-0.37***
		100/50	-0.78***	-0.62***	-0.55***	-0.47***	-0.47***
		100/75	0.33*	0.23**	0.33	0.39**	0.33***
		100/100	0.51***	0.28***	0.21*	0.21	0.39***
II.	Me/Acon-2	100/95	0.42**	0.35***	0.41***	0.28*	0.30***
		100/100	-0.28	-0.39***	-0.38***	-0.39***	-0.19*
		100/106	0.50***	0.02	-0.07	-0.07	0.07
	Me/Mpi	100/100	0.08	-0.20*	-0.28***	-0.16	-0.15
	Mpi/Acon-2	100/95	-0.08	-0.13	-0.22**	-0.22	-0.27**
		100/100	0.12	0.10	0.22**	0.22*	0.22*
		37/95	0.17	0.24**	0.13	0.05	0.25**
III.	Had/Pep-2	100/100	0.26	0.36***	0.33***	0.35*	0.56***

* $p \le 0.05$ ** $p \le 0.01$ *** $p \le 0.001$ Significance determined by t-test of z transformed correlation coefficients

Table 20. (cont'd.)

Linkage Group	Loci	Alleles	Hawthorn Sites (Year) 10(85)	3(86)	8(86)	7(87)	4(86)
I.	Dia-2/Aat-2	100/21	-0.09	-0.45***	---	-0.34*	-0.58***
		100/50	-0.76***	-0.52***	---	-0.55***	-0.57***
		100/75	0.26	0.35**	---	0.30*	0.50***
		100/100	0.23	0.41***	---	0.54***	0.70***
II.	Me/Acon-2	100/95	0.32*	0.11	0.26	0.11	0.05
		100/100	-0.34*	-0.26*	-0.41*	-0.09	-0.15
		100/106	0.02	0.16	0.24	-0.10	-0.17
	Me/Mpi	100/100	-0.19	0.00	-0.26	-0.24	-0.11
	Mpi/Acon-2	100/95	-0.02	0.12	0.07	0.04	-0.55***
		100/100	-0.04	-0.05	0.09	0.25	0.34*
		37/95	-0.05	---	---	---	---
III.	Had/Pep-2	100/100	0.52***	0.43**	0.13	0.20	0.42***

* $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$ Significance determined by t-test of z transformed correlation coefficients

Discussion

We can now answer many of the questions about geographic variation posed in the introduction to this Chapter. First, is genetic divergence widespread between apple and hawthorn infesting populations of *R. pomonella*? The answer to this question is yes. Significant allele frequency differences were found at 9 out of 10 paired apple and hawthorn sites examined across the eastern United States. Host related differentiation is, therefore, not just a local or regional phenomenon in *R. pomonella*, but is common across eastern North America.

Second, do genetic differences between apple and hawthorn populations display any patterns or trends? Yes, the same set of six loci (*Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* and *Had*) are primarily responsible for inter-host differentiation. In addition, five of these six loci show latitudinal allele frequency clines for the host races. Furthermore, the six loci displaying host associated variation are not randomly distributed throughout the genome but, instead, map to three specific chromosomal regions. Significant linkage disequilibrium exists in natural fly populations between linked, non-allelic genes within each of these three regions (*Had*, which maps to a separate chromosome from the other five loci, was tested against *Pep-2* which is located 23.9 centimorgans away from *Had* on linkage group III), but genetic interactions between chromosomes are not readily apparent from the data. It therefore seems that only a limited number of genetic changes at specific locations in the genome are involved in the initial host divergence process. Although I have probably not identified all of the regions responsible for differentiating host populations, it appears that a "genetic revolution" involving a restructuring of interactions across the genome has not accompanied the shift(s) of *R. pomonella* to apples.

Does the pattern of geographic variation for *R. pomonella* point to possible factors differentiating the host races? Theoretical models of sympatric speciation (Maynard-Smith, 1966; Bush, 1969a, 1975a; Diehl and Bush, 1989) stress that habitat specific mating is a key trait for the formation of new host races. In Chapter 2, I showed that adult flies do not randomly migrate between different species of host plants but, instead, tend to attack the same species of host plant that they previously infested as larvae (larval conditioning has not been found for *R. pomonella,* although adult conditioning has been documented; Prokopy *et al*., 1982). Because courtship and mating in *R. pomonella* occur almost exclusively on or near the fruits of the host plant (Prokopy *et al*., 1971, 1972), these differences in host preference produce genetically based, pre-mating barriers to gene flow. There is, consequently, little doubt that differential host utilization by hawthorn and apple flies results in a system of positive assortative mating which helps to maintain genetic variation between the host races.

Two observations make it unlikely that any of the six loci displaying inter-host frequency differences are directly involved in the host recognition process, however. First, five of these six loci also show latitudinal allele frequency clines within the host races. This is not what would be expected for genes encoding host preference behaviors, traits which should be relatively invariant across the range of a host race. Second, the allozymes we resolved in this study catalyze steps in general metabolic pathways. *Me* and *Acon*, for instance, are part of the TCA cycle while *Aat* converts glutamate into oxaloacetate which is used to replenish TCA intermediates. *Mpi* facilitates the entry of mannose into glycolysis and

Had and *Dia* are both involved in fatty acid biochemistry. *Had* transforms acetoacetate to D-beta-hydroxybutyrate during ketosis while *Dia* is a member of a non-phosphorylating electron-transport chain which is active during the desaturation of fatty acids. It is improbable that general metabolic enzymes such as *Me*, *Acon*, *Mpi*, *Dia*, *Aat* and *Had* directly affect behaviors associated with proper host recognition.

There is no a priori reason to presume, however, that any of the allozymes showing inter-host variation are directly under selection, however. Perhaps a more reasonable initial assumption is that these loci are linked to genes being differentially selected for in the host races and that selection is strong enough to produce hitchhiking effects for linked genes. Inversions could play a contributing role in the co-evolution of such linked arrays of genes by reducing recombination among loci sufficiently for selection to cause genetic hitchhiking. Unfortunately, the poor quality of polytene chromosome preparations in *R. pomonella* prevent us from directly testing the inversion hypothesis.

Regardless of the form or nature of selection, its effects do not appear to be uniform across the range of either apple or hawthorn flies. Except for *Aat-2*, allele frequencies for all loci showing host related differences also co-vary significantly with latitude. This result suggests that the environmental factor(s) responsible for selection is in some way correlated with latitude. Furthermore, because allele frequency clines are steeper among hawthorn than apple populations, apple flies must either be buffered from or constrained by environmental conditions to a greater extent than hawthorn flies. As a consequence, geographic variation is reduced among apple populations compared to hawthorn populations.

Natural selection is not, of course, the only possible factor influencing the pattern of geographic variation in the host races. A genetic bottleneck during the founding of the apple race could also help explain why apple flies are less genetically variable than hawthorn flies. However, two points argue that natural selection is more important than stochastic processes in shaping the overall genetic structure of *R. pomonella* populations.

First, not all polymorphic loci show reduced amounts of geographic variation among apple compared to hawthorn populations. The overall F_{ST} value for *Idh*, *Ak*, *Aat-1*, *Pgm*, *Pgi*, *Pep-2* and *Acy* across the eastern United States is actually higher for apple (0.0061 ± 0.0042; Table 17) than hawthorn flies (0.0036 ± 0.0019). It is, therefore, only those loci which show host associated differentiation which are less variable in the apple race. Allele frequencies for the six loci displaying inter-host variation are not homogeneous among apple populations, however (Table 14). In fact, *Me*, *Acon-2*, *Mpi*, *Aat-2*, *Dia-2* and *Had* are among the most geographically heterogeneous loci for the apple race (Table 17). Stochastic processes such as genetic drift and/or founder events are expected to have roughly equivalent effects on neutral genes. The diversity of geographic patterns observed among loci for apple and hawthorn flies is, therefore, not predicted by neutral theory and is strongly suggestive of selection.

Second, allele frequencies for *Acon-2* indicate that at least some gene flow may occur between the host races. $Acon^{75}$ is a relatively rare allele usually found at frequencies of 0.05 or less in fly populations (Table 22). However, at site 6 (E. Lansing, Michigan) and to a certain extent at site 5 (Grant, Michigan), $Acon^{75}$ frequencies are elevated in both hawthorn and apple populations. Even low levels of gene flow between local apple and hawthorn populations would homogenize gene frequencies for neutral alleles. This may account for the limited amount of geographic variation observed for polymorphic loci not showing host related differentiation and for why apple and hawthorn populations share even the rarest alleles in common (Tables 21-32). However, inter-host gene flow does not explain the existence of or differences in latitudinal frequency clines observed for apple and hawthorn flies. If the frequency clines in the apple race were only the result of a genetic bottleneck followed by local gene flow from hawthorn populations, then allele frequencies for apple populations should converge with those of the hawthorn race over time. However, inter-host differentiation was consistent across years at sites where multi-year collections were made (Table 13). This result argues that differential selection is strong enough to offset local levels of inter-host migration and that selection is largely responsible for both the existence of, and differences in, allele frequency clines for apple and hawthorn flies. But what environmental factor is responsible for selection in *R. pomonella*?

Several lines of evidence suggest that ambient temperature is an important environmental factor differentially affecting apple and hawthorn flies. Because *R. pomonella* is univoltine and life expectancy of adults is only 3-6 weeks in the field (J. L. Feder, pers. obs.; see also review by Boller and Prokopy, 1976), flies must eclose as adults at times closely matching the fruiting phenologies of their host plants. Fruits of domestic apple varieties favored by *R. pomonella* generally ripen and abscise 3-4 weeks earlier than those of hawthorns. "Temporal windows" for infesting apples and hawthorns are, therefore, different. Ambient temperature is, of course, correlated with latitude and has a direct effect on the timing of fruit maturation. Temperature has also been shown to be a key determinant in initiating, maintaining and terminating diapause for *R. pomonella* (Prokopy, 1968; Reissig *et al.*, 1979). Apple and hawthorn flies could, therefore, be differentially adapted to temperature as an indicator of the phenologies of their respective host plants. The distribution of adults and larvae does, in fact, follow the fruiting phenologies of apples and hawthorns. Early in the season flies are abundant on ripe apples but scarce on immature hawthorns while the reverse is true later in the season when hawthorns ripen. Apple and hawthorn flies also show genetically based differences in the timing of adult eclosion with apple flies emerging significantly earlier than hawthorn flies under controlled laboratory conditions and in the field (Diehl, 1984; Smith, 1986, 1988; Feder *et al.*, manuscript in preparation).

The temperature/development hypothesis also provides a possible explanation for latitudinal variation in *R. pomonella*. If larvae develop too early in the summer (a potential problem for populations in southern latitudes), then at least part of the population will not diapause. Instead, these flies will emerge as a second generation of "non-diapausing" adults in the fall when host fruit is no longer available. Almost all larvae do, in fact, fail to diapause when held at a constant temperature above 28° C and a light:dark cycle of 17:7 hr. (Prokopy, 1968) and second generations of flies have been reported in field studies (Caesar and Ross, 1919; Porter, 1928; Phipps and Dirks, 1933b). Alternatively, if larvae develop too slowly then they are exposed to predators and parasites for a longer period of time and run the risk of freezing to death during the first frost. Freezing is a major problem for flies in the north as we have commonly observed frozen larvae in fruit toward the end of the field season. Clinal variation can, therefore, be due to the premium placed on rapid larval development in northern populations compared to slower development and/or constitutive diapause in the south. Latitudinal variation for developmental traits have, in fact, been reported for a number of other insect species (Istock and Weisberg, 1987; Scriber and Hainze, 1987). Egg transplant experiments performed in the laboratory under controlled nutritional and environmental conditions would provide a direct test of whether southern *R. pomonella* have slower developmental rates or are more prone to diapause than northern flies.

A puzzling observation is still unaccounted for by the temperature/development hypothesis. Why do hawthorn infesting flies show greater north-south genetic variation than apple flies? One possibility is that some important physiological, chemical or phenological characteristic varies more with latitude for hawthorn than apple trees. For example, hawthorn trees may show greater north-south diversity in the period between or timing of fruit maturation and abscission than apples. Apples are, after all, a commercially cultivated fruit crop introduced from Europe. Consequently, apples should be less genetically and phenotypically variable across their range compared to native hawthorn species. The relative nutritional value of hawthorns and apples may also be an important consideration. Evidence indicates that hawthorns provide a better host-fruit environment for *R. pomonella* larvae than apples, as larvae of both apple and hawthorn origin exhibited greater survivorship in hawthorn fruits in egg transplant experiments (Prokopy *et al.*, 1988). These results are not surprising since hawthorns are the native host of *R. pomonella*. Larval development may, therefore, be inherently faster in hawthorn fruits. As a result, the failure to diapause could be as much, or more, of a problem for hawthorn flies than apple flies at southern latitudes despite the later maturation time for hawthorn fruits. Reciprocal larval transplant experiments performed in the field between different hosts using apple and hawthorn origin flies from northern and southern latitudes would help resolve why allele frequency clines are steeper in the hawthorn than in the apple race.

The validity of the temperature/development hypothesis hinges, of course, on a relationship between developmental rates and the allozyme loci showing latitudinal variation. Preliminary data indicate that such a relationship does exist as allele frequencies for *Me*, *Acon-2* and *Had* correlate with the timing of adult emergence for hawthorn flies (Feder *et al*., manuscript in preparation). Further classical and quantitative genetic experiments are needed, however, to clarify the exact relationship *Me*, *Acon-2* and *Had* (or linked genes) have with development in *R. pomonella*.

Secondary contact could also account for allele frequency clines in *R. pomonella*. Northern hawthorn populations may have become isolated and genetically differentiated from southern populations in refugia during pleistocene glaciations. However, the secondary contact hypothesis has no supporting data and it is much more likely that *R. pomonella* and their *Crataegus* hosts recolonized northern latitudes from the south following the retreat of ice sheets 10,000-20,000 years ago. In addition, secondary contact does not account for the frequency clines in the apple race and why these clines are more gradual than those existing for hawthorn flies. As we discussed earlier, our results strongly suggest that differential selection is occurring both within and between the host races and is largely responsible for the observed pattern of allele frequency clines. Finally, a genetic survey of hawthorn populations across latitudinal transects in the Midwest (see Chapter 4) revealed that allele frequencies for *Me* and *Acon-2* co-vary with irregularities in local thermal conditions. This result agrees with the temperature/development hypothesis and is not readily explained by secondary contact.

Another question I posed in the introduction is whether the geographic pattern of genetic variation is indicative of the origin(s) of the apple race? Unfortunately, the results of this study are ambiguous with regard to whether the apple race formed from a single or multiple host shifts. The reduced amount of geographic variation for apple flies is superficially consistent with a genetic bottleneck occurring during a single or relatively few colonizations of apple by *R. pomonella*. However, the clinal pattern of geographic variation indicates that differential selection occurs across the range of the apple fly. Selection is a complicating factor which makes it difficult to pinpoint a "genetic origin" for the apple race. In addition, gene frequencies for *Acon-2* indicate that at least some gene flow may be occurring on a local scale between the host races. Gene flow between apple and hawthorn populations would further obscure any traces of the genetic origin(s) of the apple race and could account for why even peripheral apple infesting populations contain roughly the same number of electromorphs as hawthorn or centrally located apple populations.

A possible analytical approach to investigating the origin(s) of the apple race would be to construct "phylogenetic networks" among apple and hawthorn populations based upon pairwise genetic distance measures between populations (for a review of this subject see Felsenstein, 1988). If the apple race formed from a single host shift then apple populations should cluster together on the network. Alternatively, if the multiple shift hypothesis is true then local apple and hawthorn populations should generally be closer together on the network than geographically distant populations of either race. Unfortunately, evolutionary models forming the underlying theoretical basis for network building algorithms are based on the selective neutrality of genetic characters (Felsenstein, 1988), an assumption which is clearly false for the host races. Furthermore, several loci analyzed in this study were shown to be tightly linked (Chapter 1) and co-evolving (Chapter 2; Tables 19, 20), which violates the requirement for the independence of characters used to derive distance estimates (Felsenstein, 1988). One possible solution to the above problems would be to use only loci such as *Pgm*, *Pgi*, *Ak*, *Idh*, *Aat-2* and *Acy* which do not appear to be under differential directional selection in *R. pomonella* (although, of course, this does not rule out the possibility of stabilizing selection) and are in linkage equilibrium with one another. However, allele frequencies for these six loci are phylogenetically uninformative in the sense that they display very little variation both between and among apple and hawthorn populations (Tables 13, 17), presumably due to the homogenizing effects of gene flow among populations. Genetic data therefore do not exclusively support or reject either the single or multiple shift hypotheses. In retrospect, it may be impossible to resolve this issue for the apple race probably may very well have originated and spread from the Hudson River Valley, New York, but, in the process of expanding its range into new areas, became subjected to repeated episodes of gene flow from local hawthorn populations.

A critical issue not fully explored in the introduction is whether apple and hawthorn populations represent incipient species. The results of this study demonstrate that genetically differentiated host races can arise in sympatry. In addition, the *R. pomonella* group contains a number of sympatrically distributed sibling species specialized on different hosts, many of which are inter-fertile in laboratory crosses, yet remain distinct in nature (Bush, 1966; Berlocher and Bush, 1982; Feder *et al.*, in press; Feder and Bush, submitted). All of these observations are consistent with host shifts representing the initial step in the speciation process. However, significant geographic variation exists within both hawthorn and apple races. The pattern of geographic variation suggests that host races are adapted to local environmental conditions and do not evolve uniformly as single panmictic units across their ranges. Whether or not divergent evolutionary processes occur within host races is not central to the question of sympatric speciation, however. The important consideration is whether host races are distinct evolutionary entities. That is, does the shift of *R. pomonella* to apples ensure that apple and hawthorn races are on separate evolutionary pathways. If the processes occurring within one race have little or no effect on evolution in other races, then host shifts will generate potential species.

To determine whether host races are distinct evolutionary units requires characterizing levels of effective gene flow between hawthorn and apple populations. Although not an easy task, mark-recapture and larval transplant experiments are possible. A series of mark-recapture studies at sympatric sites along a latitudinal transect would determine whether host fidelity is constant for apple and hawthorn flies throughout their respective ranges. Larval transplant experiments at the same sympatric sites as the mark-recapture studies would establish survivorship levels for the offspring of inter-host migrants. Taken together, the results would provide a quantitative estimate of effective gene flow between apple and hawthorn races.

Site 8 (Kenton, Ohio) documents the need for additional studies quantifying inter-host gene flow. No genetic differentiation was found between hawthorn and apple flies at this site. Migration of adults between apple and hawthorn trees may, therefore, be sufficient to counter balance the effects of host associated selection. If true, then apple and hawthorn flies could represent a mosaic of both evolutionary distinct host races and freely inter-breeding populations across eastern North America. Alternatively, site 8 might be located in an area where selection pressures on the allozymes are similar for apple and hawthorn infesting flies. Host recognition differences may, therefore, still exist between hawthorn and apple flies at the site and inter-host migration be minimal. However, because none of the allozymes resolved in this study are likely to be involved in the host recognition process, host related polymorphism would not be detected. Further studies measuring inter-host gene flow are, therefore, critical for evaluating the fidelity of host utilization and the evolutionary significance of genetic differentiation between hawthorn and apple flies.

Even in the absence of direct experiments on gene flow, the allozyme results, plus recently obtained behavioral and emergence data, indicate that apple and hawthorn flies are partially reproductively isolated across a majority of their distributions. Prokopy *et al.* (1988) have demonstrated that differences exist in the host acceptance behaviors of apple and hawthorn flies. Apple flies also terminate diapause and emerge as adults significantly earlier than hawthorn flies under controlled conditions and in the field (Diehl, 1984; Smith, 1986, 1988; Feder *et al.*, manuscript in preparation). I have also been able to qualitatively document that adult flies do not randomly migrate between apple and hawthorn trees but, instead, tend to attack the same species of host plant they infested as larvae (Chapter 2). Finally, the results of the current electrophoretic study imply that selection is acting differentially on flies infesting hawthorns and apples. All available evidence therefore indicates that hawthorn and apple flies are diverging sympatrically on separate evolutionary pathways.

Table 21. Allele frequencies for *Me* and *Had* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

Me										Had									
Site	Year	Host	N	80	100	Host	N	80	100	Host	N	97	100	122	Host	N	97	100	122
11	85	H	--	--	--	A	39	0.398	0.602	H	--	--	--	--	A	39		0.795	0.205
1	84	H	52	0.308	0.692	A	52	0.510	0.490	H	52		0.856	0.144	A	52		0.750	0.250
1	85	H	51	0.196	0.804	A	52	0.539	0.461	H	51		0.824	0.176	A	50		0.850	0.150
1	86	H	110	0.309	0.691	A	106	0.542	0.458	H	109	0.014	0.766	0.220	A	106	0.005	0.787	0.208
2	84	H	92	0.370	0.630	A	9	0.562	0.438	H	92		0.891	0.109	A	9		0.500	0.500
2	85	H	156	0.304	0.696*	A	133	0.553	0.447	H	157	0.003	0.863	0.134	A	133	0.004	0.774	0.222
5	84	H	101	0.272	0.728	A	183	0.671	0.329	H	101	0.005	0.876*	0.119	A	185	0.003	0.851	0.146
5	85	H	156	0.256	0.774	A	128	0.578	0.422	H	170		0.879	0.121	A	141	0.004	0.780	0.216
5	86	H	296	0.441	0.559	A	293	0.621	0.379	H	298		0.826	0.174	A	301		0.804	0.196
5	87	H	467	0.345	0.655	A	706	0.620	0.380	H	464		0.830	0.170	A	688		0.797	0.203
9	85	H	110	0.382	0.618	A	111	0.653	0.347*	H	103	0.005	0.864	0.131	A	82		0.762	0.238
6	84	H	128	0.664	0.336	A	130	0.704	0.296	H	130	0.004	0.673	0.323	A	130		0.669	0.331
6	85	H	144	0.559	0.441	A	163	0.641	0.359	H	142	0.004	0.662	0.334	A	158	0.003	0.678	0.319
10	85	H	54	0.417	0.583	A	48	0.656	0.344	H	54		0.796*	0.204	A	50		0.770	0.230
10	86	H	39	0.449	0.551	A	50	0.580	0.420	H	39		0.854	0.146	A	51		0.735	0.265
3	84	H	52	0.837	0.163	A	52	0.654	0.346	H	52		0.817	0.183	A	52		0.683	0.317
3	86	H	84	0.780	0.220	A	82	0.579	0.421	H	84		0.643	0.357	A	83		0.633	0.367
8	86	H	47	0.681	0.319	A	40	0.688	0.312	H	47		0.745	0.255	A	50		0.630	0.370
7	87	H	40	0.800	0.200*	A	52	0.789	0.211	H	40		0.525	0.475	A	52	0.019	0.702	0.279
4	86	H	60	0.825	0.175	A	60	0.742	0.258	H	60	0.008	0.425	0.567	A	60		0.617	0.383

* $\underline{P} \le 0.05$ ** $\underline{P} \le 0.01$ *** $\underline{P} \le 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 22. Allele frequencies for *Acon-2* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

							Acon-2											
Site	Year	Host	N	73	75	89	95	100	106	114	Host	N	73	75	89	95	100	106	114
11	85	H	–	–	–	–	–	–	–	–	A	39			0.090	0.460	0.372**	0.077	
1	84	H	52			0.067	0.481	0.346	0.087	0.019	A	51			0.039	0.284	0.520	0.147	0.010
1	85	H	51		0.010	0.059	0.608	0.284	0.039		A	49		0.010	0.010	0.367	0.531	0.082	
1	86	H	111		0.054	0.122	0.504	0.266	0.045	0.009	A	106	0.005	0.014	0.024	0.259	0.514	0.175	0.009
2	84	H	92	0.005	0.005	0.111	0.511*	0.289***	0.067	0.011	A	9		0.056	0.056	0.279	0.444	0.167	
2	85	H	158		0.009	0.076	0.591	0.256	0.067		A	132	0.004		0.045	0.348	0.436	0.163	0.004
5	84	H	101		0.035	0.059	0.490	0.297	0.114	0.005	A	185		0.084	0.030	0.184	0.530	0.159	0.013
5	85	H	169	0.003	0.068	0.062	0.544	0.257	0.059	0.006	A	137	0.004	0.113	0.058	0.208	0.493***	0.120	0.004
5	86	H	297		0.062	0.081	0.392	0.389	0.072	0.003	A	295	0.002	0.122	0.040	0.168	0.531	0.136	0.002
5	87	H	429		0.026	0.062	0.479	0.346	0.078	0.009	A	665		0.064	0.029	0.194	0.560	0.148	0.005
9	85	H	85		0.006	0.118	0.365	0.381	0.124	0.006	A	82			0.091	0.183	0.428	0.280	0.018
6	84	H	130		0.154	0.061	0.342	0.308	0.131	0.004	A	129		0.182	0.039	0.194	0.434	0.139	0.012
6	85	H	124		0.129	0.008	0.443	0.319	0.097	0.004	A	151	0.007	0.153	0.023	0.218	0.460	0.119	0.020
10	85	H	54		0.009	0.074	0.491	0.333	0.093		A	44			0.170	0.375	0.409	0.023	0.023
10	86	H	40			0.050	0.375	0.413	0.100	0.062	A	52		0.010	0.058	0.433	0.365	0.115	0.019
3	84	H	58		0.017	0.034	0.155	0.638	0.138	0.017	A	58		0.009	0.052	0.103	0.560	0.259	0.017
3	86	H	82			0.061	0.134	0.683	0.116	0.006	A	76	0.013	0.020	0.033	0.151	0.586	0.171	0.026
8	86	H	43			0.081	0.128	0.651	0.140		A	42			0.071	0.120	0.619	0.190	
7	87	H	51			0.030	0.078	0.794	0.078	0.020	A	52		0.010		0.067	0.788	0.125	0.010
4	86	H	60	0.017	0.017	0.125	0.083	0.642	0.092	0.025	A	60	0.092	0.008	0.108	0.117	0.475	0.183	0.017

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 23. Allele frequencies for *Mpi* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

						Mpi											
Site	Year	Host	N	16	31	37	70	100	125	Host	N	16	31	37	70	100	125
11	85	H	--	--	--	--	--	--	--	A	39		0.026	0.026	0.013	0.871	0.064
1	84	H	52		0.019	0.125	0.048	0.789	0.019	A	52			0.048	0.096	0.808	0.048
1	85	H	51			0.166	0.049	0.765	0.020	A	50		0.010	0.060	0.080	0.800	0.050
1	86	H	104		0.024	0.149	0.034	0.788	0.005	A	99		0.020	0.040	0.076	0.854	0.010
2	84	H	92			0.120	0.076	0.771	0.033	A	9				0.056	0.944	
2	85	H	158			0.199	0.092	0.668	0.041	A	132		0.004	0.091	0.080	0.788	0.038
5	84	H	101		0.005	0.178	0.074	0.703	0.040	A	185		0.005	0.051	0.046	0.893	0.005
5	85	H	170	0.003	0.006	0.156	0.065	0.732	0.038	A	140		0.004	0.071	0.039	0.879	0.007
5	86	H	287		0.026	0.106	0.057	0.808*	0.003	A	302		0.010	0.038	0.041	0.911	
5	87	H	417		0.035	0.096	0.047	0.817***	0.005	A	660		0.013	0.049	0.046	0.892	
9	85	H	85	0.012	0.012	0.059	0.047	0.846	0.024	A	52	0.010	0.048	0.019	0.038	0.875	0.010
6	84	H	130	0.008	0.038	0.065	0.131	0.746	0.012	A	130		0.015	0.046	0.085	0.854	
6	85	H	130		0.004	0.085	0.115	0.742	0.054	A	153		0.016	0.072	0.075	0.817	0.020
10	85	H	54		0.019	0.120	0.037	0.815	0.009	A	47		0.010	0.043	0.032	0.872	0.043
10	86	H	36		0.042	0.042	0.028	0.874	0.014	A	51		0.069	0.049	0.010	0.862	0.010
3	84	H	52		0.010	0.019	0.019	0.942		A	51		0.010		0.078	0.902	0.010
3	86	H	83		0.042	0.030	0.048	0.880		A	79		0.025	0.006	0.070	0.899	
8	86	H	47		0.032	0.085	0.032	0.851		A	49		0.020	0.020	0.031	0.929	
7	87	H	40		0.025		0.050	0.925		A	50		0.060	0.030	0.050	0.860	
4	86	H	58		0.017	0.017	0.033	0.933		A	45		0.022		0.044	0.933	

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 24. Allele frequencies for *Aat-2* and *Dia-2* for hawthorn (H) populations of *R. pomonella* across the eastern United States and Canada.

		Aat-2													Dia-2			
Site	Year	Host	N	01	21	32	40	50	59	75	84	100	113	123	130	Host	N	85	100
11	85	H	--	--	--	--	--	--	--	--	--	--	--	--	--	H	--	--	--
1	84	H	52		0.096	0.038		0.125		0.260		0.471		0.010		H	52	0.216	0.784
1	85	H	51		0.176		0.010	0.108		0.216		0.441		0.039	0.010	H	51	0.235	0.765
1	86	H	92	0.005	0.077	0.005		0.141		0.354		0.386		0.027	0.005	H	108	0.218	0.782
2	84	H	92		0.141	0.005		0.141	0.005	0.337		0.321		0.049		H	91	0.214	0.786
2	85	H	156		0.083	0.003		0.080	0.003	0.352		0.420	0.003	0.048	0.007	H	155	0.132	0.868
5	84	H	101	0.005	0.104	0.005		0.089	0.005	0.292		0.465		0.035		H	100	0.115	0.885
5	85	H	154		0.058	0.010		0.146	0.007	0.344	0.003	0.399		0.026	0.007	H	169	0.157	0.843
5	86	H	287	0.002	0.084			0.150	0.003	0.298		0.437		0.023	0.003	H	296	0.208	0.792
5	87	H	414	0.001	0.091	0.002		0.127		0.271		0.461	0.001	0.040	0.006	H	407	0.206	0.794
9	85	H	70		0.043			0.143		0.386		0.407		0.021		H	83	0.127	0.873
6	84	H	129		0.139		0.008	0.186	0.012	0.217*		0.403		0.035		H	127	0.311	0.689
6	85	H	135	0.004	0.104	0.022	0.007	0.133	0.007	0.274	0.007	0.408		0.033		H	124	0.250	0.750
10	85	H	54		0.102			0.120		0.352		0.380		0.046		H	53	0.160	0.840
10	86	H	40		0.100			0.163	0.025	0.400		0.312				H	39	0.218	0.782
3	84	H	52		0.058	0.038	0.019	0.067		0.240		0.510		0.048	0.019	H	51	0.098	0.902
3	86	H	80		0.106			0.156	0.006	0.213		0.475		0.044		H	77	0.260	0.740
8	86	H	30		0.100			0.167		0.300		0.433				H	--	--	--
7	87	H	40		0.200			0.200	0.075	0.175		0.375			0.025	H	40	0.388	0.612
4	86	H	60	0.017	0.142			0.317		0.200		0.283		0.025	0.017	H	58	0.500	0.500

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 25. Allele frequencies for *Aat-2* and *Dia-2* for apple (A) populations of *R. pomonella* across the eastern United States and Canada.

				Aat-2												Dia-2			
Site	Year	Host	N	01	21	32	40	50	59	75	84	100	113	123	130	Host	N	85	100
11	85	A	39		0.103	0.013		0.205		0.422		0.244		0.013		A	39	0.269	0.731
1	84	A	52		0.125	0.019		0.163		0.356		0.327		0.010		A	52	0.269	0.731
1	85	A	50	0.020	0.160	0.010	0.020	0.110		0.350		0.300		0.020	0.010	A	50	0.300	0.700
1	86	A	104		0.149	0.010		0.168		0.332		0.327		0.010	0.005	A	105	0.329	0.671
2	84	A	9		0.056			0.222		0.167		0.500			0.056	A	9	0.278	0.722
2	85	A	133	0.004	0.180	0.008		0.173	0.041	0.274		0.286		0.026	0.008	A	132	0.326	0.674
5	84	A	185	0.003	0.173	0.003	0.008	0.127	0.005	0.338	0.005	0.300		0.022	0.016	A	182	0.275	0.725
5	85	A	128	0.004	0.160	0.012	0.004	0.145	0.019	0.262	0.012	0.336	0.004	0.027	0.016	A	137	0.266	0.734
5	86	A	298	0.008	0.156	0.003	0.003	0.164	0.005	0.267*		0.361		0.027	0.005	A	299	0.341	0.659
5	87	A	662	0.002	0.150	0.002	0.001	0.186*	0.001	0.285**		0.344*		0.026	0.002	A	672	0.310	0.690
9	85	A	78		0.154			0.250		0.327		0.256		0.013		A	82	0.390	0.610
6	84	A	130		0.204	0.008	0.008	0.188	0.035	0.215	0.004	0.308**		0.023	0.008	A	126	0.377	0.623
6	85	A	156	0.003	0.157	0.032	0.003	0.173	0.022	0.231	0.003	0.343	0.010	0.019	0.003	A	149	0.369	0.631
10	85	A	49		0.184	0.010	0.020	0.173	0.010	0.296		0.306				A	44	0.284	0.716
10	86	A	51	0.010	0.078			0.265	0.010	0.314		0.294		0.029		A	40	0.313	0.687
3	84	A	51		0.167	0.029	0.010	0.118	0.029	0.225	0.020	0.343	0.010	0.039	0.010	A	52	0.279	0.721
3	86	A	71		0.190	0.007		0.176		0.296		0.289	0.014	0.028		A	74	0.331	0.669*
8	86	A	25		0.100			0.120		0.360		0.420				A	--	--	--
7	87	A	50		0.090			0.220	0.010	0.300	0.010	0.330		0.040		A	41	0.341	0.659
4	86	A	60		0.167			0.233		0.258		0.317		0.008	0.017	A	60	0.400	0.600

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 26. Allele frequencies for *Pep-2* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

							Pep-2										
Site	Year	Host	N	84	94	100	106	110	120	Host	N	84	94	100	106	110	120
11	85	H	--	--	--	--	--	--	--	A	39			0.769	0.077	0.154	
1	84	H	51		0.069	0.510	0.265	0.157		A	51	0.010	0.029	0.539**	0.235	0.186	
1	84	H	51		0.049	0.451	0.265	0.216*	0.020	A	50		0.060	0.420	0.330	0.150	0.040
1	86	H	--	--	--	--	--	--	--	A	--	--	--	--	--	--	--
2	84	H	92	0.005	0.060	0.424	0.283	0.228		A	8			0.625	0.375		
2	85	H	157	0.006	0.057	0.475	0.229	0.232		A	132		0.030	0.538	0.205	0.223	0.004
5	84	H	99		0.056	0.495	0.242	0.207***	--	A	183		0.071	0.486	0.189	0.254	
5	85	H	169	0.006	0.041	0.527	0.254	0.172*		A	136	0.004	0.088	0.474	0.228	0.206	
5	86	H	--	--	--	--	--	--	--	A	--	--	--	--	--	--	--
5	87	H	--	--	--	--	--	--	--	A	--	--	--	--	--	--	--
9	85	H	43			0.488	0.174	0.337		A	52		0.010	0.481	0.211	0.298	
6	84	H	116		0.039	0.543	0.198	0.220		A	129	0.004	0.039	0.516**	0.232	0.209	
6	85	H	131	0.011	0.034	0.576*	0.260	0.111	0.008	A	151	0.003	0.056	0.536	0.245	0.159	
10	85	H	54		0.019	0.528	0.278	0.176		A	46		0.065	0.370	0.239**	0.326	
10	86	H	33		0.045	0.652	0.197	0.106		A	50		0.030	0.640	0.240	0.090	
3	84	H	52		0.019	0.519	0.240	0.221		A	52	0.019	0.067	0.519	0.260	0.135	
3	86	H	--	--	--	--	--	--	--	A	--	--	--	--	--	--	
8	86	H	32			0.563	0.140	0.297		A	31		0.016	0.597	0.097	0.290	
7	87	H	39			0.654*	0.179	0.167		A	42		0.012	0.631	0.238	0.119	
4	86	H	60		0.017	0.542	0.267	0.175		A	60		0.008	0.667	0.175	0.150	

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 27. Allele frequencies for *Aat-1* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

							Aat-1								
Site	Year	Host	N	-48	-64	-88	-100	-110	Host	N	-48	-64	-88	-100	-110
11	85	H	--	--	--	--	--	--	A	39		0.115		0.885	
1	84	H	52		0.134		0.837***	0.029	A	52		0.173		0.788***	0.039
1	85	H	50		0.100		0.900*		A	50	0.020	0.230		0.700***	0.050
2	84	H	92		0.179		0.799***	0.021	A	9		0.111		0.833	0.056
2	85	H	155	0.042	0.116	0.006	0.810***	0.025	A	131	0.004	0.137	0.011	0.802***	0.046
5	84	H	101	0.005	0.149		0.811***	0.035	A	185		0.173	0.003	0.784***	0.040
5	85	H	155	0.023	0.171	0.006	0.774***	0.026	A	127		0.205	0.004	0.756***	0.035
9	85	H	39		0.141		0.821	0.038	A	42		0.048		0.929	0.024
6	84	H	129	0.016	0.244		0.709***	0.031	A	130	0.008	0.154		0.792***	0.046
6	85	H	130	0.008	0.181	0.008	0.781***	0.023	A	154	0.006	0.162		0.782***	0.049
10	85	H	54		0.176		0.787***	0.037	A	46		0.120		0.870	0.010
10	86	H	39		0.115		0.795*	0.090	A	52		0.077		0.865**	0.058
3	84	H	51		0.245		0.706***	0.049	A	51		0.206		0.774***	0.020
3	86	H	49		0.031		0.959*	0.010	A	48		0.021	0.021	0.917***	0.042
8	86	H	31		0.177		0.774*	0.048	A	37		0.068		0.892***	0.040
7	87	H	40	0.025	0.300	0.012	0.663***		A	51		0.226		0.725	0.049
4	86	H	60		0.158	0.008	0.817**	0.017	A	60		0.100	0.008	0.867***	0.025

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 28. Allele frequencies for *Pgi* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

Pgi																	
Site	Year	Host	N	53	80	100	117	130	145	Host	N	53	80	100	117	130	145
11	85	H	--	--	--	--	--	--	--	A	39			0.975			0.025
1	84	H	52			0.923			0.071	A	52			0.932		0.010	0.058
1	85	H	51			0.911		0.010	0.079	A	51			0.911		0.020	0.069
2	84	H	92	0.011	0.033	0.880			0.076	A	9			0.944			0.056
2	85	H	158		0.025	0.896		0.013	0.066	A	133		0.008	0.883	0.008	0.026	0.075
5	84	H	101		0.025	0.881	0.005	0.005	0.084	A	185		0.005	0.906	0.005	0.005	0.079
5	85	H	157	0.003	0.003	0.914			0.080	A	126		0.016	0.881		0.016	0.087
9	85	H	54	0.019	0.019	0.870		0.009	0.083	A	51	0.020	0.010	0.814		0.029	0.127
6	84	H	130		0.012	0.892	0.004		0.092	A	130		0.004	0.885		0.027	0.081
6	85	H	128	0.016	0.004	0.910		0.008	0.063	A	154		0.013	0.903		0.006	0.075
10	85	H	54		0.009	0.944			0.046	A	46		0.033	0.891		0.022	0.054
10	86	H	41			0.890	0.024		0.085	A	52	0.029		0.904			0.067
3	84	H	52	0.010		0.904			0.086	A	51	0.010	0.029	0.833		0.020	0.108
3	86	H	48		0.010	0.927		0.010	0.052	A	45	0.011		0.900		0.011	0.078
8	86	H	41	0.024		0.793			0.183	A	50	0.010		0.890		0.030	0.070
7	87	H	40			0.838			0.162	A	52			0.971			0.029
4	86	H	60			0.917*			0.083	A	60		0.017	0.850			0.133

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 29. Allele frequencies for *Pgm* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

		Pgm															
Site	Year	Host	N	80	92	95	100	111	118	Host	N	80	92	95	100	111	118
11	85	H	--	--	--	--	--	--	--	A	39	0.026		0.064	0.846		0.064
1	84	H	52	0.058		0.048	0.797	0.087	0.010	A	52	0.058	0.010	0.125	0.702	0.096	0.010
1	85	H	51	0.039		0.059	0.873	0.029		A	51	0.030		0.049	0.833	0.088	
2	84	H	92	0.022	0.016	0.033	0.858	0.060	0.011	A	9		0.056		0.889		0.056
2	85	H	157	0.022	0.010	0.064	0.796	0.092	0.016	A	133	0.030	0.007	0.064	0.842	0.053	0.004
5	84	H	101	0.005	0.010	0.045	0.851	0.079	0.010	A	185	0.011		0.070	0.816	0.095	0.008
5	85	H	156	0.010	0.016	0.074	0.821	0.064	0.016	A	126	0.012	0.012	0.056	0.774	0.107	0.040
9	85	H	54	0.009	0.027	0.056	0.815	0.028	0.065	A	52	0.038	0.010	0.058	0.846	0.029	0.019
6	84	H	129	0.016	0.004	0.050	0.791	0.132	0.008	A	130	0.023	0.004	0.069	0.812	0.077	0.015
6	85	H	135	0.015		0.056	0.830	0.067	0.033	A	154	0.010	0.010	0.091	0.808	0.062	0.019
10	85	H	54	0.019	0.009	0.102	0.722	0.139	0.009	A	46	0.011	0.022	0.087	0.826	0.054	
10	86	H	41	0.073			0.829	0.061	0.037	A	52	0.058	0.010	0.019	0.856	0.048	0.010
3	84	H	52	0.029		0.038	0.885	0.019	0.029	A	52	0.038		0.058	0.769	0.135	
3	86	H	48	0.010	0.021	0.052	0.803	0.104	0.010	A	46	0.011	0.011	0.109	0.804	0.054	0.011
8	86	H	48	0.021		0.042	0.823	0.063	0.052	A	49		0.010	0.092	0.816	0.051	0.031
7	87	H	40	0.013	0.013	0.112	0.813	0.013	0.037	A	52	0.038	0.019	0.038	0.827	0.048	0.029
4	86	H	58	0.034		0.043	0.776	0.129	0.017	A	58	0.034	0.017	0.034	0.837	0.060	0.017

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 30. Allele frequencies for *Acy* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

						Acy							
Site	Year	Host	N	78	89	100	110	Host	N	78	89	100	110
11	85	H	--	--	--	--	--	A	--	--	--	--	--
1	84	H	52		0.298	0.635	0.067	A	51	0.020	0.196	0.725	0.059
1	85	H	51		0.167	0.804	0.029	A	50	0.010	0.130	0.810	0.050
2	84	H	92	0.027	0.179	0.778	0.016	A	8		0.125	0.750	0.125
2	85	H	156	0.006	0.218	0.774	0.032	A	133	0.019	0.135*	0.793	0.053
5	84	H	65		0.231	0.738	0.031	A	--	--	--	--	--
5	85	H	166	0.006	0.208	0.750	0.036	A	138	0.022	0.163	0.743	0.072
9	85	H	50	0.010	0.290	0.670	0.030	A	52	0.010	0.115	0.837	0.038
6	84	H	--	--	--	--	--	A	--	--	--	--	--
6	85	H	133	0.007	0.158	0.752	0.083	A	141	0.018	0.184	0.713	0.085
10	85	H	54	0.009	0.204	0.722	0.065	A	47		0.149	0.830	0.021
10	86	H	38		0.171	0.816	0.013	A	46		0.065	0.913	0.022
3	84	H	52	0.019	0.183*	0.721*	0.077	A	52	0.010	0.183	0.740	0.067
3	86	H	--	--	--	--	--	A	--	--	--	--	--
8	86	H	37		0.203	0.730	0.067	A	40	0.012	0.200	0.750	0.038
7	87	H	40		0.200	0.688	0.112	A	51		0.137	0.833	0.029
4	86	H	60		0.142	0.733	0.125	A	60	0.080	0.175	0.767	0.050

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 31. Allele frequencies for *Idh* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

Idh															
Site	Year	Host	N	82	100	104	118	133	Host	N	82	100	104	118	133
11	85	H	--	--	--	--	--	--	A	39		0.679	0.013	0.295	0.013
1	84	H	52	0.019	0.596	0.019	0.356	0.010	A	52	0.010	0.586		0.394	0.010
1	85	H	48	0.010	0.542		0.406	0.042	A	49		0.622		0.337	0.041
2	84	H	92	0.011	0.669	0.016	0.299	0.005	A	9		0.611		0.389	
2	85	H	157	0.016	0.618	0.032	0.322	0.013	A	133	0.026	0.586	0.011	0.357	0.019
5	84	H	88	0.011	0.665	0.006	0.295	0.023	A	172	0.006	0.578	0.015	0.395	0.006
5	85	H	156	0.006	0.631	0.022	0.324	0.016	A	127	0.012	0.626	0.012	0.350	
9	85	H	53	0.009	0.651		0.340		A	52	0.010	0.692	0.038	0.260	
6	84	H	130	0.008	0.661		0.323	0.008	A	130	0.012	0.635	0.015	0.331	0.008
6	85	H	138	0.014	0.623	0.004	0.359		A	153	0.013	0.644	0.033	0.297	0.013
10	85	H	54	0.019	0.648*	0.028	0.306		A	49		0.704	0.010	0.286	
10	85	H	40	0.012	0.625	0.063	0.238	0.063	A	51		0.686	0.020	0.294	
3	84	H	52		0.577	0.019	0.385	0.019	A	52	0.029	0.702	0.038	0.211	0.019
3	86	H	48		0.781	0.021	0.198		A	48		0.604	0.042	0.344	0.010
8	86	H	48	0.010	0.667	0.031	0.271	0.021	A	48		0.584	0.052	0.333	0.031
7	87	H	40		0.638		0.325	0.037	A	52	0.019	0.625	0.019	0.327	0.010
4	86	H	60		0.658	0.017	0.308	0.017	A	60		0.600	0.017	0.350	0.033

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 32. Allele frequencies for *Ak* for hawthorn (H) and apple (A) populations of *R. pomonella* across the eastern United States and Canada.

		Ak													
Site	Year	Host	N	68	87	100	111	118	Host	N	68	87	100	111	118
11	85	H	--	--	--	--	--	--	A	39		0.115	0.872		0.013
1	84	H	52		0.067	0.923		0.010	A	52		0.077	0.913		0.010
1	85	H	50	0.010	0.090	0.840		0.060	A	50		0.090	0.840	0.050	0.020
2	84	H	92		0.103	0.853	0.010	0.033	A	9		0.111	0.889		
2	85	H	156	0.003	0.106	0.865	0.016	0.016	A	130	0.012	0.062	0.881	0.030	0.015
5	84	H	101	0.015	0.040	0.926		0.020	A	185	0.027	0.051	0.900		0.022
5	85	H	157	0.003	0.108	0.869	0.003	0.016	A	125		0.076	0.884	0.008	0.032
9	85	H	28		0.071	0.911		0.018	A	40	0.013	0.087	0.900		
6	84	H	130	0.004	0.100	0.873	0.004	0.019	A	130	0.004	0.108	0.865	0.015	0.008
6	85	H	124	0.004	0.073	0.867	0.016	0.040	A	152	0.007	0.033	0.898	0.050	0.013
10	85	H	54		0.083	0.907		0.009	A	45	0.011	0.122	0.856		0.011
10	86	H	39	0.013	0.141	0.820		0.026	A	52		0.048	0.913	0.010	0.029
3	84	H	52	0.019	0.087	0.856		0.038	A	51	0.010	0.049	0.902		0.039
3	86	H	48	0.010	0.042	0.927		0.021	A	53		0.093	0.895		0.012
8	86	H	29		0.172	0.810		0.018	A	29		0.086	0.879	0.017	0.017
7	87	H	40	0.050	0.025	0.913		0.012	A	52	0.010	0.087	0.875	0.029	
4	86	H	60	0.008	0.075	0.917			A	60		0.167	0.817	0.017	

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Chapter 4

Gene Frequency Clines for *Rhagoletis pomonella* Host Races in the mid-Western United States

Introduction

Speciation, the rendering of one closed genetic system into two, is a fundamental process responsible for the evolution of life on earth. Unfortunately, speciation is a rare, time consuming event from the perspective of the biologist. Consequently, much of our understanding of the process is based on population genetic models and indirect analyses of closely related species. Mathematical models of speciation tell us what is theoretically possible but do not constitute proof that taxa diverged by a particular mode of speciation. Post hoc studies of species differences are also plagued by a major problem: It is extremely difficult to distinguish differences which were actively responsible for restricting inter-populational gene flow and initiating the divergence process from differences which arose subsequent to speciation. We therefore have very few examples directly documenting how populations become reproductively isolated from one another. Consequently, questions have been raised as to whether speciation occurs by "macro-evolutionary" mechanisms which qualitatively differ from "micro-evolutionary" processes responsible for gene frequency changes within populations (Gould and Eldredge, 1977).

The apple maggot fly, *Rhagoletis pomonella*, has a biology and natural history which makes it ideal for the study of speciation. *R. pomonella* is part of a sibling species complex of true fruit flies whose members are broadly sympatric in their geographic distributions across North America (Bush, 1966). *Rhagoletis* larvae are obligate internal parasites in the fruit of their hosts, with each species in the group being monophagous or oligiophagous for a different set of host plants (Bush, 1966). Adult females lay their eggs directly into the host fruit which they identify by specific visual, tactile and olfactory cues (Bush 1969a,b). Males are attracted by the same cues and courtship and mating occur almost exclusively on or near the fruits of the host plant (Prokopy *et al.*, 1971, 1972). Because host recognition and mate selection are directly coupled in *Rhagoletis*, variation for host preference and host associated survivorship traits can act as genetically based, pre-mating barriers to gene flow. These considerations led Bush (1966, 1969a,b, 1975) to propose that speciation in the *R. pomonella* group occurs sympatrically and is initiated when flies colonize and adapt to new host plants.

Verifying that sympatric speciation occurs in the *R. pomonella* group is a two step process. First I must show that traits adapting fly populations to different host plants are also responsible for reproductively isolating host associated populations. Second, I must demonstrate that these "host specific" traits can evolve within geographically continuous populations. In this regard, the recent shift of the species *R. pomonella* from its native host hawthorn (*Crataegus* spp.) to introduced, domestic apple (*Malus pumila*; Ward, 1866; Walsh, 1867) provides a unique ecological timeframe to determine whether host specialization is sufficient to differentiate populations in the absence of geographic barriers to gene flow. Agricultural records suggest a probable origin for the apple infesting "race" in the Hudson Valley, New York, and document the apple fly's subsequent spread across eastern North America over the last 120 years (see Illingworth, 1912; O' Kane, 1914; Bush, 1969a). Because domestic apples were introduced into the United States by European settlers (Chapman and Lienk, 1971), we have an upper limit of approximately 350 years for the age of the apple infesting form of *R. pomonella*. Of course, apple populations could conceivably have originally shifted from a host other than hawthorns. However, the only other reasonable native hosts in North America are five species of endemic crab apple (*Pyrus angustifolia*, *P. glabrata*, *P. ioensis*, *P. rivularis* and *P. coronaria*), none of which support populations of *R. pomonella* (O'Kane, 1914; Porter, 1928; Bush, 1966). The reason for this is unclear but it has been suggested that the fruits of native crabs are too acidic and ripen too late for *R. pomonella* larvae to complete development before the onset of winter (O'Kane, 1914). It is therefore highly likely that hawthorns were the original host for the apple race.

In the preceding Chapters, I have shown that sympatric apple and hawthorn populations of *R. pomonella* are genetically differentiated. Consistently significant allele frequency differences exist for six allozyme loci between sympatric populations of the two host races from across the eastern United States (Chapter 3). These six loci map to three different regions of the genome (Chapter 1) and linkage disequilibrium occurs in natural populations between non-allelic genes within each of these three regions (Chapters 2, 3). Allele frequencies for five of the six loci displaying inter-host differentiation also co-vary significantly with latitude among both apple and hawthorn populations (Chapter 3). This suggests that north-south allele frequency clines exist for *R. pomonella*. However, the geographic analysis of eastern North America involved widely separated populations from a variety of different longitudes. Verification of clines therefore requires more detailed examination of fly populations collected along latitudinal transects.

The purpose of the current study is two-fold. First, I want to describe the regional pattern of allozyme variation for *R. pomonella* in the midwestern United States. For instance, how extensive is geographic variation for *R. pomonella* in the Midwest? What levels of intra- and inter-host variation exist on a county by county basis within a geographic region? Do apple and hawthorn flies continue to show consistent genetic differences when examined on a much finer geographic scale than previous studies? Do allele frequency clines exist for *R. pomonella*? If so what is the nature of these clines? Do all loci co-vary similarly with latitude or do loci differ in the shapes and configurations of their frequency clines? In short, the first set of questions addresses how genetic variation is partitioned within and between *R. pomonella* host races.

The second aim of this Chapter is to explore the causal basis for any observed geographic patterns of variation. In other words, what factors are responsible for differentiating apple and hawthorn populations? Does natural selection act differentially between the two races? Do flies migrate randomly between different species of host plants? Do the same processes act uniformly across the Midwest or do local conditions dictate the extent and pattern of genetic differentiation? Ambient temperature has been implicated as an important environmental factor influencing host-parasite interactions for *R. pomonella* (Prokopy, 1968; Reissig *et al.*, 1979). Is there any indication of this in the regional pattern of genetic variation? Finally, if gene frequency clines exist are they the result of differential selection (perhaps related to ambient temperature) or is secondary contact a possibility? The second set of questions therefore focuses on whether reproductive isolation in *R pomonella* is the direct outcome of ecological and behavioral adaptations resulting from host shifts.

Materials and Methods

Flies were collected from 34 different sites across the midwestern United States from 1985 to 1987 (Figure 5, Table 33). In general, the sites were organized into five north-south transects which ran from the states of Wisconsin through Illinois or from Michigan through Indiana (see Figure 5 and Table 34 for a map and breakdown of sites comprising each of the five latitudinal transects). At 28 of the sites, both hawthorn (*Crataegus mollis*) and feral apple (*Malus pumila*) infesting populations were sampled. Host plants were sympatric (that is, separated by a distance of less than 100 meters) at 13 of the 28 paired locations. In only five instances were apple and hawthorn trees situated more than 4 km apart with a maximum distance of 8 km at site 8 near Parkerdale, Indiana. Six of the 28 paired sites (listed as 9, 15, 19, 20, 26 and 28 in Table 33) were the same apple and hawthorn populations that were analyzed in the earlier allozyme study of the eastern United States (these sites were designated 7, 5, 1, 2, 3 and 4, respectively, in Chapter 3).

Three different strategies were used to collect flies. At a majority of sites larvae were dissected from fallen fruit in the field and immediately

Figure 5. Collecting sites for *R. pomonella* in the midwestern United States (see Table 33 for detailed descriptions of study sites). Solid lines indicate isothermal clines for growing degree days base 50^{o}F (GDD). GDD values for Illinois and Indiana sites are given in parentheses. Dashed lines connect sites comprising each of the five latitudinal transects collected in the study. ⊙ = site where the frequency of *Acon-2*^{75} was > 0.100 for either apple or hawthorn flies. ◉ = site where the frequency of *Acon-2*^{75} was > 0.100 for both apple and hawthorn flies. Δ = site where the frequency of *Acon-2*^{89} was > 0.200 for hawthorn flies and/or > 0.100 for apple flies.

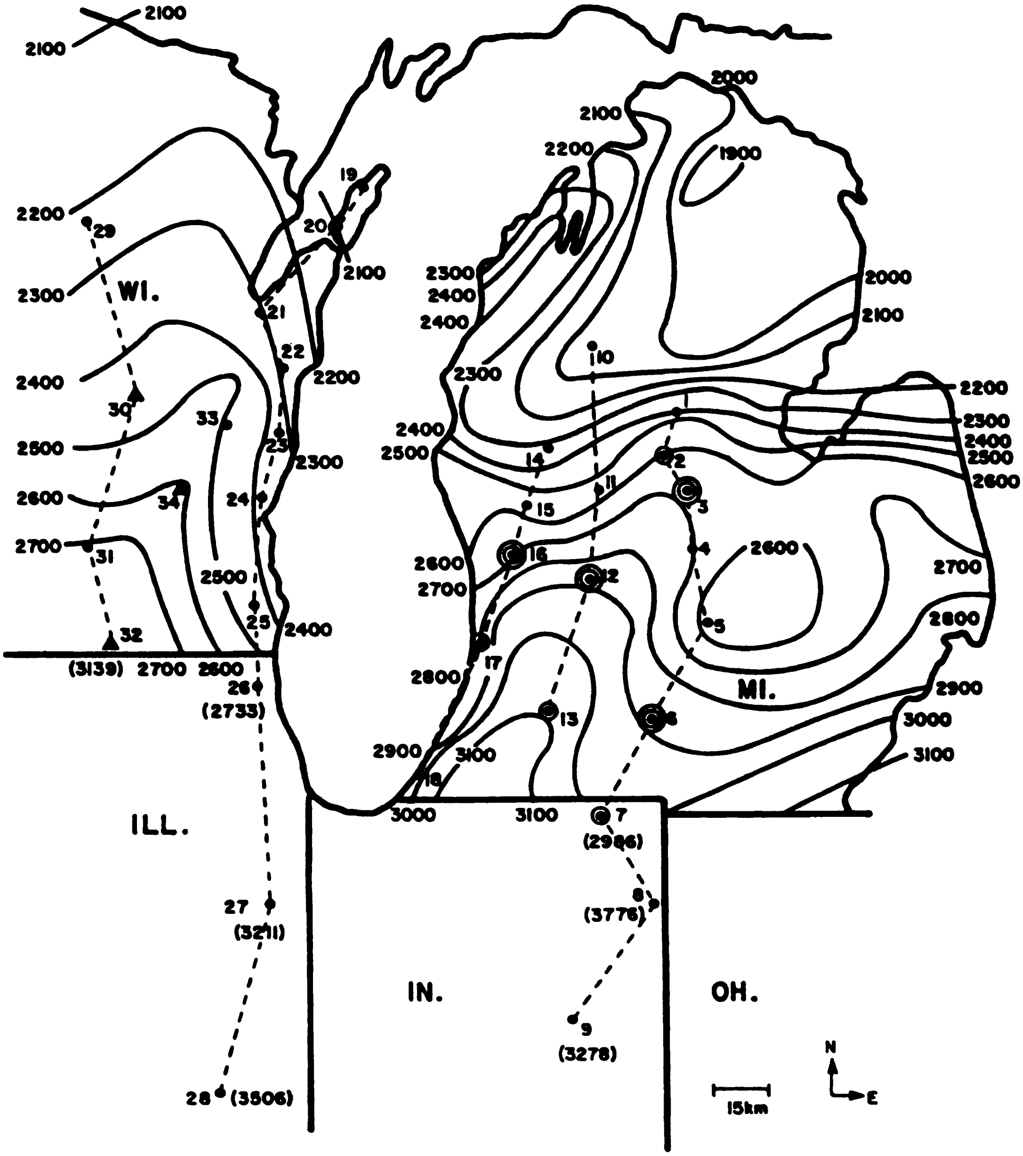

2100
2100
2200
29
2300
WI.
2400
30
33
2500
2600
34
24
2700
31
2500
25
2400
32
(3139)
2700
2600
26
(2733)
19
20
2100
21
22
2200
23
2300
ILL.
27
(3211)
28
(3506)
2000
2100
2200
1900
2300
2400
2400
2300
2400
2500
10
2000
2100
2200
2300
2400
2500
2600
1
2
3
14
11
15
16
12
4
2600
2600
2700
2700
5
2800
17
2800
MI.
2900
3000
3100
13
6
2900
3100
18
3000
3100
7
(2986)
8
(3776)
IN.
OH.
9
(3278)
15km
N
E

Figure 5.

Table 33. Numerical designations and descriptions of field sites. Sites were generally organized into five latidudinal transects (see Figure 5 and Table 34). Latitude and growing degree days base 50°F are also given for sites.

Site	Year	Location	Stage[a]	Host[b]	Dist[c] (km)	Latit.	GDD
1	87	Rodgers Ave, Clare, MI.	L(L)	A,H	3.2	43°40'	2380
2	87	MI. 20, Mt. Pleasant, MI.	L(L)	A,H	3.0	43°35'	2660
3	87	Alger Rd., Alma, MI.	L(L)	A,H	sym	43°25'	2650
4	87	MI. 27, Eureka, MI.	L(Ad,L)	A,H	sym	43°06'	2720
5	86	Jolly Rd., Okemos, MI.	R(R)	A,H	1.0	42°41'	2570
6	87	F Dr., Eckford, MI.	L(Ad)	A,H	sym	42°09'	2880
7	87	IN. 20, Lagrange, IN.	(L)	H	-	41°38'	2986
8	87	Interstate 24, Parkdale, IN.	L(L)	A,H	8.0	41°11'	3376
9	87	Interstate 69, Gas City, IN.	L(L)	A,H	7.1	40°50'	3278
10	87	No. 45 Rd., Cadillac, MI.	L(L)	A,H	1.0	44°11'	2060
11	87	MI. 46, Lakeview, MI.	L	A	-	43°25'	2550
12	87	Grand River Ave, Saranac, MI.	L(Ad)	A,H	4.9	42°52'	2870
13	87	Sprinkle Rd., Portage, MI.	L(L)	A,H	4.2	43°14'	3083
14	87	MI. 20, Woodville, MI.	L(Ad)	A,H	sym	43°40'	2320
15	87	112th St., Grant, MI.	Ad,L(Ad,L)	A,H	sym	43°21'	2550
16	87	24th St., Herrington, MI.	L(Ad)	A,H	2.0	43°04'	2690
17	87	MI. 89, Fennville, MI.	L(Ad,L)	A,H	1.5	42°34'	2800
18	87	East Rd., Three Oaks, MI.	Ad(Ad)	A,H	sym	41°46'	2940
19	86	Hogan Farm, Ephraim, WI.	R(R)	A,H	sym	45°07'	2060
20	85	Kuehn Farm, Carlsville, WI.	R(R)	A,H	sym	44°56'	2100
21	87	County MM Rd., Green Bay, WI.	L(L)	A,H	sym	41°28'	2310
22	87	Gass Lake Rd., Manitowoc, WI.	L(Ad)	A,H	sym	44°02'	2290
23	87	County EE Rd., Sheboygan, WI.	L(Ad)	A,H	sym	43°43'	2340
24	87	County C. Rd., Port Wash.,WI.	L(Ad)	A,H	sym	43°22'	2380
25	87	50th. St., N. Cape, WI.	L(Ad)	A,H	0.5	42°46'	2450
26	86	ILL. 132, Waukegan, ILL.	R(R)	A,H	3.3	42°20'	2733
27	87	River Park, Kankakee, ILL.	(L)	H	-	41°09'	3211
28	86	Univ.ILL. Campus, Urbana, ILL.	R(R)	A,H	sym	40°05'	3506
29	87	WI. 29, Hatley, WI.	L(L)	A,H	1.6	44°52'	2230
30	87	WI. 49, Berlin, WI.	L(L)	A,H	6.3	43°54'	2450
31	87	I 90 and WI. 12, Madison, WI.	(L)	H	-	43°00'	2710
32	87	WI. 81, Beloit, WI.	L(L)	A,H	2.5	42°30'	3139
33	87	WI. 23, Greenbush, WI.	(L)	H	-	43°46'	2490
34	87	WI. 67, Iron Ridge, WI.	(L)	H	-	43°22'	2600

---

[a] Life history stage in which flies were collected. L=larvae, Ad=adult, R=flies sampled as larvae and reared to adulthood in laboratory. Stage designation given in parentheses is for hawthorn flies, without parentheses for apple flies.

[b] Host plants from which flies were collected at a site. A=apple, H=hawthorn.

[c] Distance (km) separating apple and hawthorn trees sampled at paired sites. sym=sympatric site where host trees were within 100 m.

frozen in liquid nitrogen (these sites are listed as L in Table 33). Only late second and third instar larvae were electrophoretically analyzed from these collections. At other locations adult flies were captured directly from off of their respective host plants using a vacuum aspiration system and were subsequently frozen in liquid nitrogen (these sites are designated Ad in Table 33). For three hawthorn sites (4, 15, 17) and one apple site (15) both larvae and adults were collected from host trees to allow comparisons of the two life history stages which represent sequentially different fly generations (i.e., parents and their immediate offspring). At sympatric site 15, adults and larvae were collected from both apple and hawthorn host plants to determine whether inter-host genetic differentiation was consistent across different life history stages and to qualitatively gauge the extent of inter-host migration. Finally, fruits were collected from the ground beneath host trees at the five sites designated R in Table 33 and the larvae infesting these fruits were reared to adulthood in the laboratory. Laboratory rearing included chilling pupae at 4°C in a refrigerator for five months to simulate winter diapause conditions.

Standard horizontal starch gel electrophoretic techniques were used and are described fully in Chapter 1 and in Berlocher and Bush (1982). We resolved and scored the six allozyme loci which previously displayed the greatest amounts of inter-host and/or latitudinal variation in the earlier survey of the eastern United States (these six allozymes being: *Malic enzyme* [*Me*], *Aconitase-2* [*Acon-2*], *Mannose phosphate isomerase* [*Mpi*], *NADH-diaphorase-2* [*Dia-2*], *Aspartate amino-transferase-2* [*Aat-2*] and *Hydroxyacid dehydrogenase* [*Had*]). A Mendelian mode of inheritance has been established for all six of these allozymes (Chapter 1). Isozymes that migrated the nearest to the cathode were designated system 1, the second nearest system 2, etc. Alleles were numbered according to their relative anodal mobilities with the most common allele for each locus designated 100 and used as a standard. Electromorphs resolved for each of the six loci are given in Tables 39-43, which appear after the Discussion section on pages 141-150. Only two alleles were scored for the loci *Dia-2* and *Me*. Consequently, only the frequencies for the rarer of the two alleles for these loci appear in Table 40. Three alleles were scored for *Had*, but one of these alleles (Had^{97}) was very rare and so for brevity only frequencies for the second most common allele (Had^{122}) are given in Table 40. Electrophoretic alleles migrated identically in adult and larval samples.

Genotypic frequencies were examined for deviations from Hardy-Weinberg expectation by G-tests. Alleles were pooled, as required, so that all genotypic classes had expected numbers >1. The Levene correction (Spiess, 1977) was applied when sample sizes were <100 or when alleles could not be pooled to make expected numbers >1. G-contingency tests were used to test for allele frequency heterogeneity. Alleles were pooled, when necessary, to ensure that each cell in the G-contingency test had an observed number ≥5. F-statistics were calculated by the method of Weir and Cockerham (1984) with standard deviations estimated by jackknifing over loci or populations. Corrections for unequal sample sizes in the method of Weir and Cockerham can result in negative F_{ST} values. In these instances we report F_{ST} as 0.

Linear and curvilinear regression analyses were performed between arcsine transformed allele frequencies and either latitude or growing degree days (GDD) base 50°F. GDD is calculated by determining the number of degrees that the average temperature for a day was above 50°F and summing these daily totals throughout the course of the year (negative daily values are considered to be 0). Fifty degrees Fahrenheit is not an arbitrarily chosen temperature for calculating GDD but, rather, represents the approximate lower threshold condition under which post-pupal development occurs for *R. pomonella* (Reissig *et al.*, 1979). Fahrenheit was used as the unit of temperature in this study because weather stations in the midwestern United States do not provide GDD values in centigrade base. Unfortunately, converting GDD between the two temperature scales is not exact without daily readings from stations, a task which imposed insurmountable logistical and computational hurdles. GDD values from Wisconsin were the yearly averages from 1950 to 1974, in Illinois from 1951-1980, in Michigan from 1931 to 1960, and in Indiana from 1975 to 1987. Regressions were performed separately for apple and hawthorn populations from Michigan/Indiana and Wisconsin/Illinois transects. At sites where both adults and larvae were collected, the allele frequencies used in the regressions were the mean between the two life history stages.

Results

Consistently significant allele frequency differences were observed for *Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2*, and *Had* between apple and hawthorn fly populations from across the midwestern United States (Tables 34, 39-43). Of the 28 paired apple and hawthorn sites, only site 4 (Eureka, Michigan) and site 32 (Beloit, Wisconsin) did not show a significant difference for at least one of the six loci analyzed. *Me* and *Acon-2* displayed the greatest amount of inter-host variation being significantly different at 14 and 21 paired sites, respectively. *Me* and *Acon-2* are, in fact, tightly linked on chromosome II along with *Mpi* (Chapter 1) and significant

Table 34. G-contingency tests for allele frequency heterogeneity between paired apple and hawthorn populations. Degrees of freedom for tests involving each locus are given in parentheses following locus abbreviations. At site 15 both adults (Ad) and larvae (L) were tested for inter-host variation.

Site	Me(1)	Acon-2(2)	Mpi(1)	Dia-2(1)	Aat-2(3)	Had(1)
Transect 1						
1	***	***	***			
2	*	*	*	*		
3	**	***				
4						
5	*	**				
6		*				
8		*				
9						**
Transect 2						
10	***	***		*		***
12	***	***				
13						*
Transect 3						
14	***	***	***	*	***	
15(Ad)	***	***	***	*	*	***
15(L)	***	***	***	***	***	*
16		***		***	***	
17		***				
18	**					
Transect 4						
19	***	***		**		
20	***	***	**	***	***	**
21	*	**				
22		*				*
23					**	
24		**		*	**	
25		***				
26	***				**	
28		**				**
Transect 5						
29		**				
30	**	***				
32						

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-contingency test

linkage disequilibrium has been found between non-allelic genes for *Me* and *Acon-2* in natural apple and hawthorn populations (Chapters 2, 3). Genetic hitchhiking due to selection at a linked locus could therefore account for the observed correlation in allele frequencies between *Me* and *Acon-2*. The same may also be true for *Dia-2* and *Aat-2*, which are separated by a map distance of 3.2 centimorgans on chromosome I (Chapter 1). In this study, 5 of the 7 sites which showed significant frequency differences between apple and hawthorn flies for *Aat-2* were also significant for *Dia-2* (Table 34). High levels of linkage disequilibrium have also been observed between *Dia-2* and *Aat-2* in natural fly populations (Chapters 2, 3).

Allele frequencies for *Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* and *Had* were similar in intra-host comparisons between adults and larvae collected from hawthorns at sites 4, 15, and 17 and from apples at site 15. Only *Me* for hawthorn flies at site 15 (Grant, Michigan) showed a significant frequency difference between adult and larval life history stages out of a total of 24 tests (Me^{100} adults = 0.593, n = 345; Me^{100} larvae = 0.655, n = 467; G-contingency test = 6.61, $\underline{P} \leq 0.05$). Larvae dissected from host fruits represent the offspring of adults collected from off of apple and hawthorn trees. The near constancy of gene frequencies observed between the two generations suggests that intra-host selection is not intense between the two life-history stages. The fact that hawthorn and apple populations were generally in Hardy-Weinberg equilibrium lends further support for this view (Tables 39-43). The 16 significant deviations out of 330 total tests for apple populations and the 24 significant deviations out of 407 tests for hawthorn populations do not differ appreciably from the number of significant tests expected due to random type I error (16.5 and 20.3, respectively, for apple and hawthorn populations).

The pattern and magnitude of inter-host differentiation between hawthorn and apple populations was similar for both larval and adult life history stages at site 15 (Tables 34, 39-43). Gene frequencies were significantly different between apple and hawthorn populations for both larvae and field captured adults for *Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2*, and *Had* (Table 34). As mentioned before, only *Me* for hawthorns displayed significant intra-host variation at site 15 between adults and larvae. Adults collected directly from apple trees are, therefore, as genetically distinct from hawthorn adults as apple larvae are from hawthorn larvae. This finding indicates that adult flies do not migrate freely between apple and hawthorn trees but instead tend to utilize the same species of host plant which they infested as larvae. Because courtship and mating in *R. pomonella* occur almost exclusively on or near the fruits of their host plants, differences in host utilization between hawthorn and apple flies at site 15 produce pre-mating barriers to gene flow which help maintain host associated polymorphism.

Geographic variation was apparent among both apple and hawthorn populations across the midwestern United States. Substantial spatial heterogeneity was observed across each of the five north-south transects (Table 35) and the magnitude of geographic variation differed between the two host races. Hawthorn flies displayed greater inter-populational differentiation than apple flies as evidenced by the 5 x higher overall F_{ST} value among hawthorn compared to apple populations (0.0626 and 0.0125, respectively; Table 36). In addition, individual loci and alleles showed variable levels of geographic variation. For instance, *Me*, *Acon*, and *Had* had F_{ST} values of 0.1403, 0.0984 and 0.806, respectively, among hawthorn populations and 0.218, 0.212 and 0.183 among apple populations. In

Table 35. G-contingency tests for allele frequency heterogeneity among apple and hawthorn populations along five latitudinal transects in the midwestern United States.

Host	Transect	Me	Acon-2	Mpi	Dia-2	Aat-2	Had
Apple	1		***	*			
	2	***	***				
	3	**	***		***	***	
	4	***	***	**			***
	5	**			**		***
Hawthorn	1	***	***	**			***
	2	***	***	**	***	*	***
	3	***	***	*	***	*	***
	4	***	***	***	***	***	***
	5	*	***	***			***

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-contingency test

Table 36. F_{ST} values (Weir and Cockerham, 1984) across the 28 paired apple and hawthorn populations analyzed in the study. R refers to the F_{ST} value calculated by pooling all alleles remaining for a locus besides those listed in the table.

Locus	Allele	Apple	Hawthorn
Me	100	0.0218 ± 0.0092[a]	0.1403 ± 0.0314
Acon-2	100	0.0214 ± 0.0120	0.1033 ± 0.0305
	95	0.0312 ± 0.0175	0.1346 ± 0.0377
	R	0.0132 ± 0.0057	0.0380 ± 0.0170
	Total	0.0212 ± 0.0099	0.0984 ± 0.0257
Mpi	100	0.0123 ± 0.0060	0.0341 ± 0.0170
Dia-2	100	0.0059 ± 0.0049	0.0379 ± 0.0165
Aat-2	100	0.0023 ± 0.0033	0.0025 ± 0.0037
	75	0.0050 ± 0.0053	0.0126 ± 0.0064
	50	0.0020 ± 0.0024	0.0266 ± 0.0121
	R	0.0031 ± 0.0029	0.0061 ± 0.0036
	Total	0.0032 ± 0.0025	0.0108 ± 0.0042
Had	100	0.0183 ± 0.0063	0.0806 ± 0.0229
All loci		0.0125 ± 0.0058[b]	0.0626 ± 0.0293

Jackknife estimates of standard deviation calculated over populations[a] or over loci[b].

comparison, *Aat-2* had a F_{ST} of 0.0108 for hawthorn flies and 0.0032 for apple flies across the same set of populations. Furthermore, F_{ST} among hawthorn populations for the alleles *Acon-2*^{100} and *Acon-2*^{95} (0.1033 and 0.1346, respectively; Table 36) were greater than the value for the remaining pooled alleles at this locus (0.0380). A similar pattern existed for allelic variants of *Aat-2* among hawthorn populations (Table 36). The observed heteogeneity of F_{ST} values suggests that differential selection is occurring among alleles and loci, and that the effects of this selection are more pronounced for hawthorn than apple flies.

Patterns of geographic variation were not random across the Midwest. Linear and curvilinear regressions performed between arcsine transformed allele frequencies and either latitude or growing degree days (GDD base 50^{o}F) indicated that frequency clines exist for both apple and hawthorn flies. In only 6 of 52 cases did second order polynomials significantly improve the fits of the regression lines compared to first order linear equations (Tables 37, 38). Consequently, I will confine my discussion to the results of the first order analysis. Apple flies did not display nearly as many significant linear regressions for individual loci with either GDD or latitude as hawthorn flies (8 compared to 22). Also, r^2 values (the amount of genetic variation accounted for by the regression) were always lower among apple than hawthorn populations (Tables 37, 38). Apple populations therefore have more unexplained geographic variation than hawthorn populations; a result which suggests that stochastic processes (i.e., genetic drift or founder events) may play a more prominent role in shaping the genetic architecture of local apple than hawthorn populations.

The absolute values of the slopes of the regression lines were always greater for hawthorn than apple populations (Tables 37, 38). This result is

Table 37. First order linear regression analyses between arcsin transformed allele frequencies and either latitude or growing degree days base 50°F (GDD) for apple populations along either Michigan/Indiana or Wisconsin/Illinois transects. r^2 = amount of genetic variation explained by the linear regression.

			Apple Populations			
			Latitude		GDD[a]	
Transect	Locus	Allele	r^2	Slope ± Std.Err.	r^2	Slope ± Std.Err.
Mi./In.	Me	100	0.22	2.954 ± 1.439	0.37**	-1.100 ± 0.379
	Acon-2	95	0.39**[t]	3.998 ± 1.304	0.37**	-1.100 ± 0.389
	Mpi	100	0.00	-0.074 ± 1.308	0.01	0.156 ± 0.382
	Dia-2	100	0.04	0.904 ± 1.119	0.02	-0.163 ± 0.333
	Aat-2	100	0.00	0.289 ± 1.066	0.01	-0.096 ± 0.313
	Had	100	0.15	1.655 ± 1.006	0.26*	-0.638 ± 0.276
	All loci[b]		0.37**	1.646 ± 0.548	0.49**	-0.555 ± 0.145
Wi./Ill.	Me	100	0.19	1.455 ± 0.940	0.14	-0.406 ± 0.325
	Acon-2	95	0.27	2.022 ± 1.048	0.15	-0.509 ± 0.377
	Mpi	100	0.42*	-1.827 ± 0.680	0.46*	0.641 ± 0.218
	Dia-2	100	0.32	1.696 ± 0.785	0.23	-0.485 ± 0.278
	Aat-2	100	0.01	0.231 ± 0.638	0.01	-0.030 ± 0.214
	Had	100	0.52**	2.892 ± 0.880	0.47*	-0.919 ± 0.308
	All loci[b]		0.59**	1.697 ± 0.446	0.44*	-0.488 ± 0.173

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ Regression coefficient (slope of first order linear regression) > 0, ANOVA F-test.

[a] Slopes and standard errors for GDD regressions are in units of 10^{-2}.

[b] *Mpi*^{37} allele frequencies were used instead of those for *Mpi*^{100} in the All loci regressions.

[t] Second order polynomial significantly improved fit of regression line compared to first order equation (ANOVA).

Table 38. First order linear regression analyses between arcsin transformed allele frequencies and either latitude or growing degree days base 50°F (GDD) for hawthorn populations along either Michigan/Indiana or Wisconsin/Illinois transects. r^2 = amount of genetic variation explained by the linear regression.

			Hawthorn Populations							
			Latitude				GDD[a]			
Transect	Locus	Allele	r^2	Slope	±	Std.Err.	r^2	Slope	±	Std.Err.
Mi./In.	Me	100	0.72***[t]	10.639	±	1.717	0.83**[t]	-3.400	±	0.395
	Acon-2	95	0.82***	11.842	±	1.454	0.86***	-3.600	±	0.374
	Mpi	100	0.72***[t]	-5.436	±	0.871	0.70***	1.600	±	0.271
	Dia-2	100	0.61***	4.203	±	0.877	0.68***	-1.300	±	0.237
	Aat-2	100	0.52**	1.890	±	0.898	0.63***	-0.622	±	0.122
	Had	100	0.77***	7.658	±	1.081	0.88***	-2.500	±	0.232
	All loci[b]		0.83***	6.945	±	0.813	0.91***	-2.200	±	0.174
Wi./Ill.	Me	100	0.51**[t]	5.891	±	1.535	0.50**[t]	-1.900	±	0.518
	Acon-2	95	0.58***	5.700	±	1.285	0.65***	-2.000	±	0.396
	Mpi	100	0.50**	3.443	±	0.918	0.60***	1.300	±	0.273
	Dia-2	100	0.46**	-3.350	±	0.980	0.29*	-0.897	±	0.372
	Aat-2	100	0.01	0.283	±	0.807	0.01	-0.078	±	0.269
	Had	100	0.80***	5.572	±	0.746	0.81***	-1.900	±	0.244
	All loci[b]		0.81***	4.040	±	0.521	0.80***	-1.300	±	0.178

---

* *P* ≤ 0.05 ** *P* ≤ 0.01 *** *P* ≤ 0.001 Regression coefficient (slope of first order linear regression) > 0, ANOVA F-test.

[a] Slopes and standard errors for GDD regressions are in units of 10^{-2}.

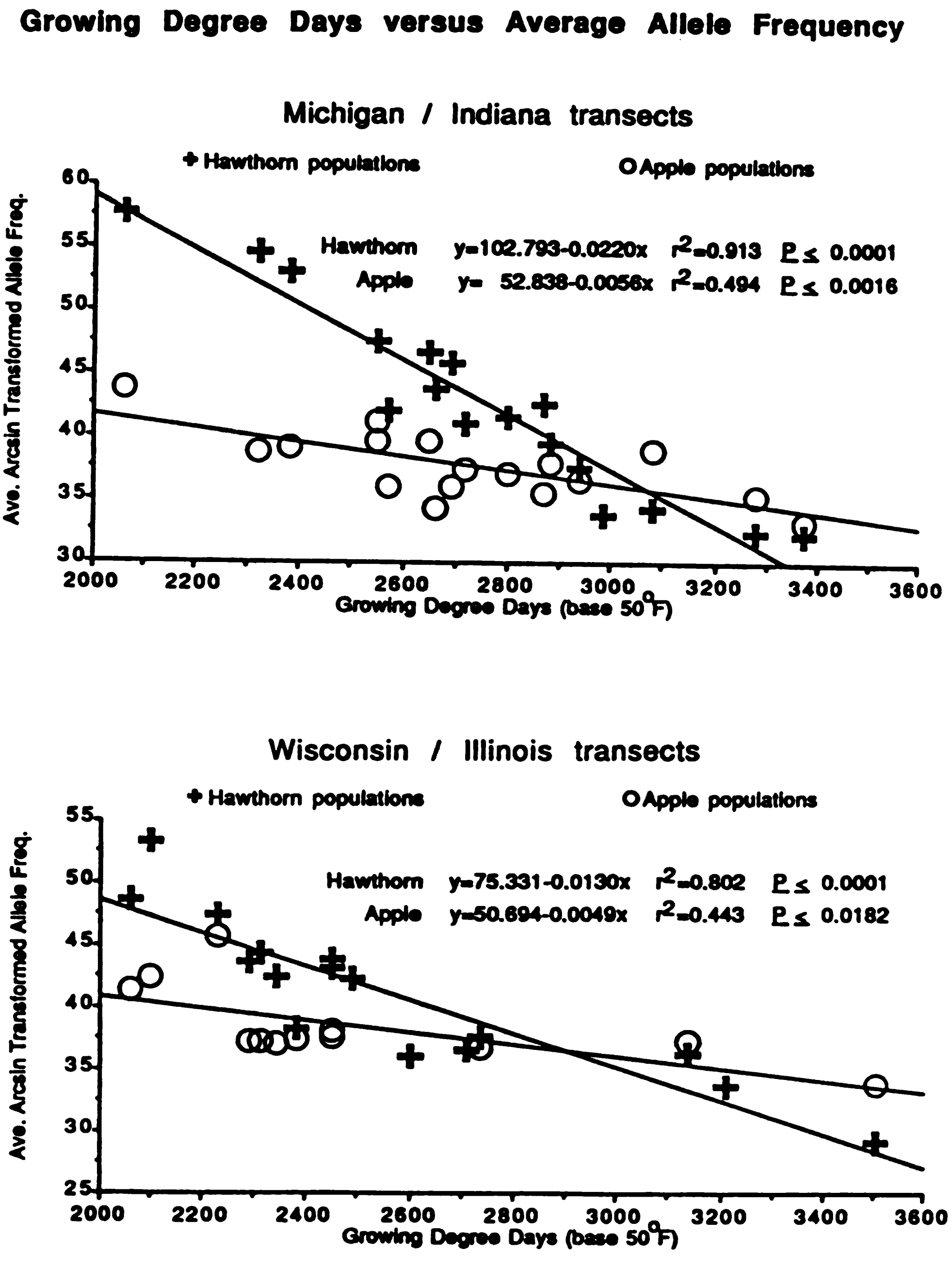
[b] *Mpi*^{37} allele frequencies were used instead of those for *Mpi*^{100} in the All loci regressions.

[t] Second order polynomial significantly improved fit of regression line compared to first order equation (ANOVA).

consistent with the higher levels of genetic variation found among hawthorn populations and also helps to account for the observed pattern of inter-host differentiation. Figure 6 graphically depicts the relationship between GDD and the average arcsine transformed allele frequency for *Me*^{100}, *Acon-2*^{95}, *Mpi*^{37}, *Dia-2*^{100}, *Aat-2*^{100} and *Had*^{100} for sites along Michigan/Indiana and Wisconsin/Illinois transects (Note: Plotting allele frequencies versus latitude gave roughly the same result as graphing allele frequencies against GDD). Inter-host differentiation was greatest at the colder, northern sites and was primarily due to the contributions of *Me*, *Acon-2* and *Mpi* (Tables 34, 39-41). At centrally located sites with more moderate temperatures, hawthorn and apple populations were less genetically distinct, until at site 4 near Eureka, Michigan, and site 34 in Beloit, Wisconsin, none of the six loci analyzed in the study showed a significant inter-host frequency difference (Table 34). Hotter sites further to the south once again began to show increased inter-host heterogeneity. However, unlike the northern sites, *Had* displayed the greatest amount of inter-host differentiation in Illinois and Indiana (Tables 34, 40). Genetic differentiation between apple and hawthorn flies is, therefore, a function of local thermal conditions (or latitude, as the two factors are correlated), with inter-host divergence resulting from differences in the slopes of allele frequency clines between the two races.

Me, *Acon-2*, *Mpi*, *Aat-2*, *Dia-2* and *Had* showed both similarities and differences in their clinal patterns of variation across the Midwest. Because allele frequency clines were more clearly defined for the hawthorn race, my discussion will initially focus on hawthorn populations along transect 1 (Michigan/Indiana) and transect 4 (Wisconsin/ Indiana), the two most intensively surveyed transects in the study (Note: Site 10 was considered to

Figure 6. Average arcsin transformed allele frequency for *Me*^{100}, *Acon-2*^{95}, *Mpi*^{37}, *Aat-2*^{100}, *Dia-2*^{100} and *Had*^{100} for apple and hawthorn populations along Michigan/Indiana and Wisconsin/Illinois transects plotted against growing degree days base 50°F. The first order linear regression equations are given for apple and hawthorn flies along with r^2 values (the amount of genetic variation explained by the regression equation) and the probability level (P) for the ANOVA F-test of the null hypothesis that the slope of the regression line = 0.



Growing Degree Days versus Average Allele Frequency

Michigan / Indiana transects

Wisconsin / Illinois transects

Figure 6.

be the northern most site along transect 1 in this discussion). For transect 4, allele frequencies for Me^{100}, $Acon\text{-}2^{95}$ and Mpi^{37} displayed sharp drops between hawthorn populations 20 and 21 north of Green Bay in eastern Wisconsin, leveled off and even showed somewhat of an increase between populations 21 through 25, and then fell for a second time south of site 25 near the boarder between Wisconsin and Illinois (Figure 7, Tables 39-41). The linked loci *Me*, *Acon-2* and *Mpi* therefore behaved in a similar fashion among transect 4 hawthorn populations. In contrast, Had^{100} frequencies did not change appreciably between hawthorn sites 20 and 21 but did fall along with Me^{100}, $Acon\text{-}2^{95}$ and Mpi^{37} south of population 25 (Figure 8, Table 40). $Dia\text{-}2^{100}$ frequencies decreased steadily from hawthorn populations 20 to 24 (Figure 8, Table 40). However, sites 25 and 26 formed a plateau of elevated $Dia\text{-}2^{100}$ frequencies that was more pronounced and occurred slightly further south, than similar peaks for Me^{100}, $Acon\text{-}2^{95}$, Mpi^{37} and Had^{100}. Allele frequencies for $Dia\text{-}2^{100}$, like those for Me^{100}, $Acon\text{-}2^{95}$, Mpi^{37}, and Had^{100}, also declined sharply toward the southern end of transect 4. $Aat\text{-}2^{100}$ displayed a marked drop in allele frequency at hawthorn site 24 which was also seen for $Dia\text{-}2^{100}$ and to a lesser extent for Had^{100}, but was not observed for $Acon\text{-}2^{95}$, Me^{100} or Mpi^{37} (Figures 7, 8). In general, $Aat\text{-}2^{100}$ showed a very similar pattern of geographic variation along transect 4 as $Dia\text{-}2^{100}$ (Figure 8, Tables 40, 42) as would be expected due to the close proximity of *Aat-2* and *Dia-2* on linkage group I (Chapter 1).

Comparisons of allele frequency clines for hawthorn flies between transects 1 and 4 revealed both similarities and differences. As was the case for transect 4, Me^{100}, $Acon\text{-}2^{\,95}$ and Mpi^{37} all displayed large frequency decreases at the northern end of transect 1 between sites 10 and 4 (Figure 7,

Figure 7. Allele frequencies for *Me*^{100}, *Acon-2*^{95} and *Mpi*^{37} for apple and hawthorn populations along transect 1 (Michigan/Indiana) and transect 4 (Wisconsin/Illinois). Sites are arranged by decreasing latitude from left to right along the x-axis. Site 10 is included as the northern most site for transect 1.

Allele Frequencies for Me^{100}, Acon-2^{95} and Mpi^{37} Along Transect 1 (Mi./Ind.) and Transect 4 (Wi./Ill.)

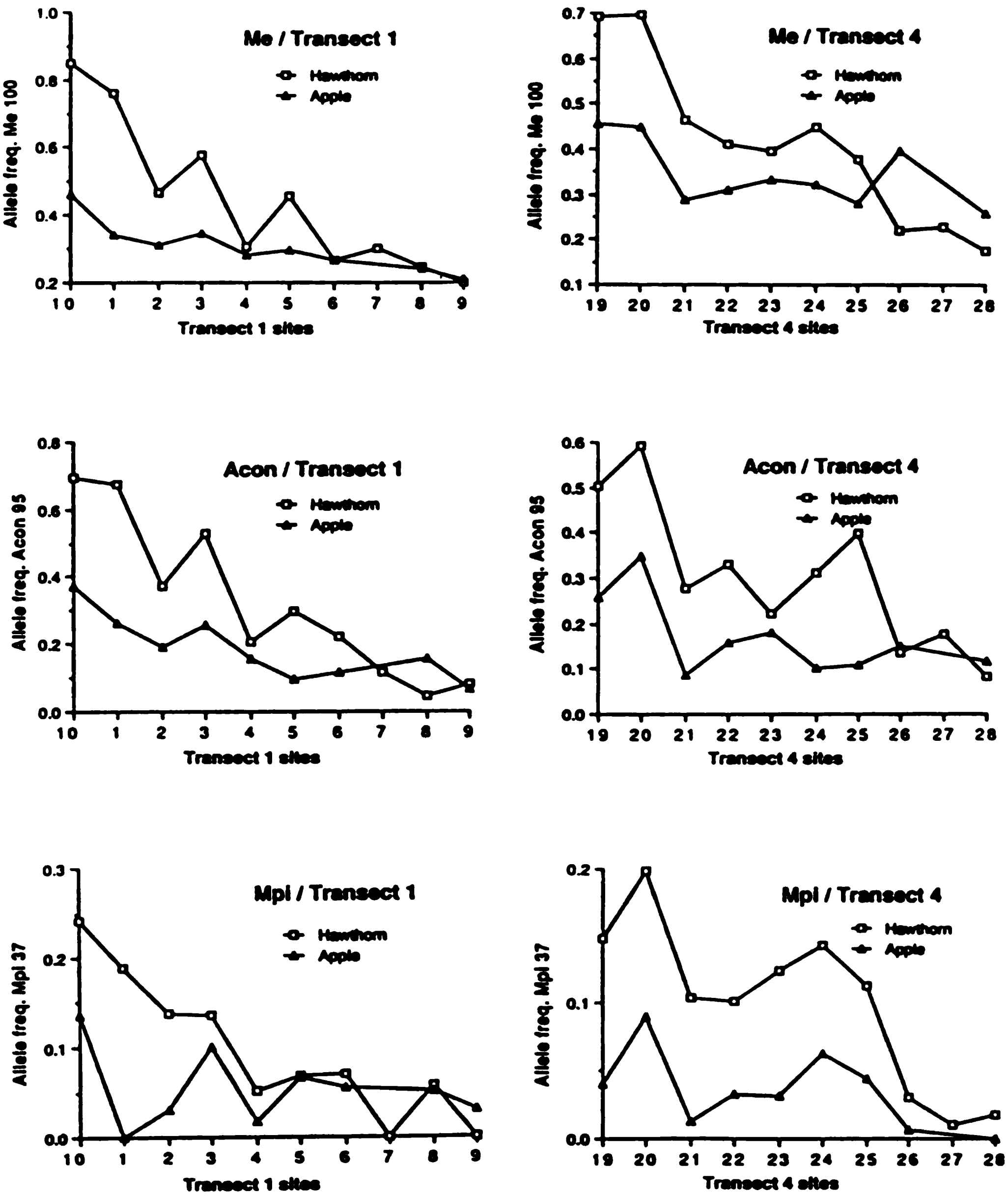


Figure 7.

Figure 8. Allele frequencies for *Aat-2*^{100}, *Dia-2*^{100} and *Had*^{100} for apple and hawthorn populations along transect 1 (Michigan/ Indiana) and transect 4 (Wisconsin/Illinois). Sites are arranged by decreasing latitude from left to right along the x-axis. Site10 is included as the northern most site for transect 1.

Allele Frequencies for Aat-2^{100}, Dia-2^{100} and Had^{100} Along Transect 1 (Mi./Ind.) and Transect 4 (Wi./Ill.)

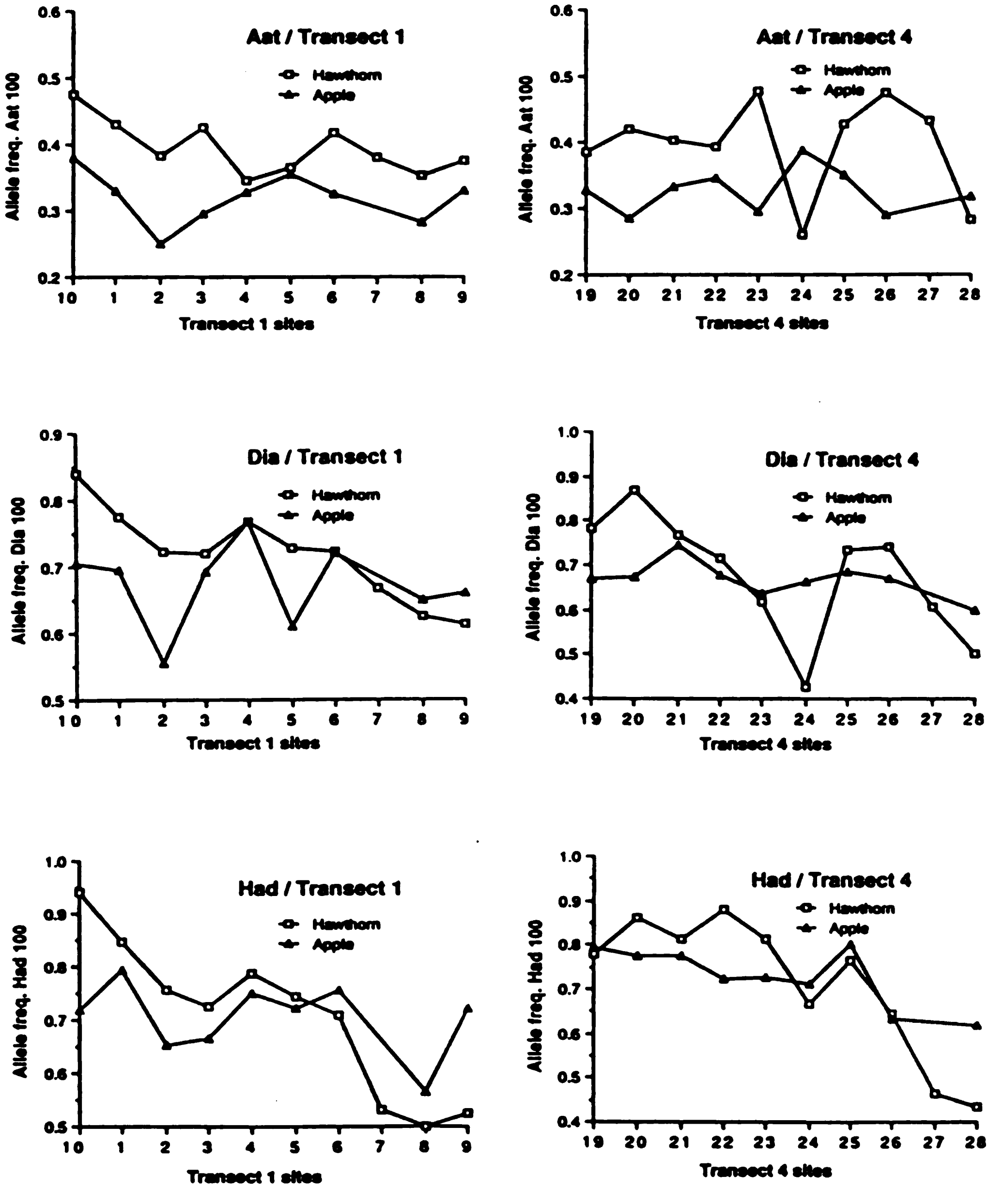


Figure 8.

Tables 39-41). However, allele frequencies for *Me*^{100} and *Acon-2*^{95} fluctuated more sharply among centrally located populations in Michigan than in Wisconsin (a possible cause for the frequency peaks observed for hawthorn flies at site 5 will be discussed later). *Had*^{100} frequencies changed roughly equivalently between the two transects (Figure 8, Table 40) but *Dia-2*^{100} and *Aat-2*^{100} clines were different, as both *Dia-2*^{100} and *Aat-2*^{100} displayed much more even and gradual reductions in allele frequencies along transect 1 than transect 4 (Figure 8, Tables 40, 42).

Allele frequency clines for apple infesting populations, while not as distinct or as statistically significant as those for hawthorn flies, were also apparent. For instance, *Me*^{100} and *Acon-2*^{95} frequencies both dropped between apple populations 20-21 and 1-4 just as they did for the corresponding paired hawthorn populations (Figure 7, Tables 39, 40). In addition, *Mpi*^{37} allele frequencies fell between apple sites 20 and 21 and sites 10 and 1 (Figure 7, Table 41). *Aat-2*^{100}, *Dia-2*^{100} and *Had*^{100} allele frequencies also displayed a general decline in allele frequencies between apple sites 10 and 2 along transect 1 just as they did for hawthorn flies (Figure 8, Tables 42, 43). However, none of these three loci showed an obvious decrease in frequency among apple populations at the northern end of transect 4. *Had*^{100} did vary significantly with both latitude and GDD for apple sites in Wisconsin and Illinois (Table 37). However, *Had*^{100} frequencies changed more gradually and uniformly among apple populations along transect 4 than for hawthorn populations (Figure 8, Table 40).

Allele frequencies for *Acon-2* suggest that at least some gene flow may be occurring between local apple and hawthorn populations. Frequencies for *Acon-2*^{75} were higher for both hawthorn and apple populations in central and western Michigan than they were in any other region of the Midwest (Figure 5, Table 39). The same phenomena also occurred for *Acon-2*^{89} at sites 30 and 32 in Wisconsin and Illinois (Figure 5, Table 39). In addition, the rare allele *Acon-2*^{73} was present only in apple and hawthorn populations along transect 4 (Table 39). Selection could be responsible for the elevated frequencies of certain *Acon-2* alleles only at specific paired apple and hawthorn sites. However, *Acon-2* allele frequencies were usually significantly different between the two host races at paired sites (Table 34); a result suggesting that selection pressures differ between sympatric populations of hawthorn and apple flies. It is also unlikely that stochastic processes alone could account for the independent increases of rare *Acon-2* alleles in apple and hawthorn populations at only certain paired sites. Inter-host gene flow must therefore be at least partly responsible for the increased frequencies of *Acon-2*^{75} and *Acon-2*^{89} in both host races in central Michigan and Wisconsin, respectively, and for the occurrence of the rare allele *Acon-2*^{73} only along transect 4. Gene flow may also account for the observation in Chapter 3 that polymorphic allozymes which do not display host associated differentiation at sympatric sites also show low levels of intra-host geographic variation across eastern North America.

Although latitude and temperature are inter-related environmental factors, their relationship is not perfect. Several irregularities exist in the pattern of isothermal clines through Michigan and Wisconsin (Figure 5) which are informative for determining whether latitude or GDD is the most accurate predictor of gene frequencies for *R. pomonella* populations. For Michigan/Indiana transects GDD proved to be the more reliable determinant and explained approximately 10% more of the genetic variation among both apple and hawthorn populations than latitude (see r^2 values for all loci in Tables 37, 38). GDD and latitude accounted equally well for genetic variation among hawthorn populations in Wisconsin and Illinois (Table 38) but latitude was a better predictor among apple populations (r^2 latitude = 0.59, r^2 GDD = 0.44; Table 37). Unfortunately, four hawthorn sites in Wisconsin and Illinois (27, 31, 33, 34) did not have paired apple collections. Three of these four unpaired sites (31, 33, 34) were located within thermal warm areas in Wisconsin (Figure 5). The addition of apple populations from these three sites would more than likely make the r^2 values for latitude and GDD among Wisconsin/Illinois apple populations very similar to one another. If one also considers that GDD estimates for sites are less accurate than latitude (GDD values being based on multiple year averages obtained from the closest weather station to a collecting site in years when temperature data were available), then ambient temperature would appear to be a slightly better indicator of the genetic composition of local *R. pomonella* populations, at the very least in Michigan and Indiana, than latitude.

Hawthorn populations displayed several allele frequency shifts for Me^{100} and $Acon\text{-}2^{95}$ that were, in fact, correlated with irregularities in local thermal conditions. First, site 5 (Okemos, Michigan) is located within a pocket of colder GDD in central Michigan (Figure 5). Gene frequencies for Me^{100} and $Acon\text{-}2^{95}$ were significantly higher in hawthorn than in apple populations at this site (Me^{100} hawthorn = 0.455 and Me^{100} apple = 0.298; $Acon\text{-}2^{95}$ hawthorn= 0.295 and $Acon\text{-}2^{95}$ apple = 0.098; Tables 34, 39, 40, Figure 7). Site 4 (Eureka, Michigan) is located approximately 20 km north of site 5 but is warmer. Me^{100} and $Acon\text{-}2^{95}$ did not show inter-host differentiation at site 4 (Table 34) and were present at respective frequencies of 0.283 and 0.214 in hawthorn larvae and 0.333 and 0.200 in hawthorn adults. Warmer temperature sites south of Okemos, Michigan (sites 6-9; see Figure 5) showed the same trends for *Me* and *Acon-2* allele frequencies that were observed at site 4 (Figure 7, Tables 39, 40). A "blip" therefore occurred for *Me*^{100} and *Acon-2*^{95} frequencies that corresponded with a cold bubble of GDD in central Michigan.

A second example involves fingers of higher GDD which extend up into Michigan around the Kalamazoo area (see site 13 in Figure 5) and in Wisconsin northeast of the city of Madison (see site 31). Gene frequencies for *Me*^{100} and *Acon-2*^{95} were almost always lower for hawthorn populations within these warm areas (site 13, Portage, Michigan, and sites 33 and 34 near Greenbush and Iron Ridge, Wisconsin) compared to sites of equivalent latitude in parallel transects (sites 6, 17 and 18 in the case of site 13, sites 23 and 30 for site 33, and site 24 for site 34; see Tables 39 and 40 for *Acon-2* and *Me* allele frequencies, respectively). The only exception to this general trend was for *Acon-2*^{95} in comparisons between sites 23 and 33, for which the frequency of the allele was lower at the colder site 23 (0.221) than at the warmer site 33 (0.306).

Finally, in a related study of hawthorn populations along a transect running from north of Milwaukee, Wisconsin to south of Chicago, Illinois (Feder *et al.*, manuscript in prep.), gene frequencies for *Me*^{100} and *Acon-2*^{95} dropped dramatically around Milwaukee corresponding to an increase in GDD presumably due to a heating effect of the city.

Discussion

The results from this study strongly suggest that partially reproductively isolated host races can evolve in sympatry. However, to prove that the apple race originated via a sympatric host shift I must still show that 1.) the apple infesting form of *R. pomonella* was not reproductively isolated from hawthorn flies prior to its colonization of apples and 2.) that the apple race could not have evolved in allopatry from the hawthorn race and then been secondarily introduced into the northeastern United States.

All available evidence indicates that the apple race is directly descendant from a hawthorn infesting form of *R. pomonella* and was not derived from an unrecognized sibling species. For instance, experimental crosses give no indication of any postmating or ethological premating reproductive isolation between hawthorn and apple flies (Reissig and Smith, 1978; Smith, 1988; R. Prokopy, pers. comm.). Also, no single morphological or genetic character exists which diagnostically distinguishes apple from hawthorn flies, as the host races are currently known to differ only in allele frequencies for six specific allozyme loci. In fact, because allele frequency clines are steeper among hawthorn than apple populations, hawthorn populations at southern sites in the Midwest are actually more genetically similar to northern apple populations than they are to northern hawthorn populations. Consequently, different apple populations from across the Midwest do not cluster together genetically as a discrete subdivision from hawthorn populations as might be expected if the apple race was formed from a sister taxon to the hawthorn race. Furthermore, the non-random pattern of $Acon\text{-}2^{73}$, $Acon\text{-}2^{75}$ and

Acon-2^{89} allele frequencies in Michigan and Wisconsin (Figure 1) and the similar pattern of linkage disequilibrium in apple and hawthorn populations (Chapters 2, 3) suggest that at least some gene flow is occurring between local apple and hawthorn populations. It is therefore very unlikely that apple and hawthorn populations of *R. pomonella* are completely reproductively isolated.

It is still possible, of course, that different races of *R. pomonella* exist on different species of hawthorns and that the apple race was derived from just one of these "hawthorn races". As many as 1,000 species of *Crataegus* comprising 19 different species groups may, in fact, be endemic to North America (Fernald, 1950; Correll and Johnston, 1970). However, the taxonomic status of a majority of these *Crataegus* species is questionable as hybridization, polyploidy and apomixis is common in the genus (Phipps, pers. comm.). Many of the trees that we identified as *C. mollis* in this study may, therefore, have been of hybrid origin. Nevertheless, *R. pomonella* attacks only a restricted set of *Crataegus* "species", with infestations confirmed for only 14 endemic hawthorns (Bush, 1966; Wasbauer, 1972; Berlocher, 1976; McPheron et al., 1988). In the northeastern United States, only *C. mollis* or hawthorn species which are phenotypically similar to *C. mollis* in the *Brainerdianae*, *Coccinnea* , *Macracantha*, *Pruinosa*, *Punctata* and *Tenuifoliae* series have been reported to be parasitized by the fly (O'Kane, 1914; Wasbauer, 1952; Bush, 1966). It therefore appears that *R. pomonella* utilizes only a small number of potential hawthorn hosts in the Northeast, attacking only those species that share a similar fruiting phenology and fruit characteristics; A trend not conducive to the formation of different hawthorn races. In addition, electrophoretic analysis of *R. pomonella* populations infesting *C. punctata*, *C. brachyacantha*, *C. douglasii* and *C. monogyna* from the southern and western United States has given no evidence for genetically differentiated hawthorn races (Berlocher, 1976; McPheron, 1987). Although we cannot completely discount the possibility of host specific hawthorn races, all available data suggest that hawthorn populations of *R. pomonella* represent a single race which displays latitudinal variation.

Several points also make it unlikely that the apple race formed in allopatry from the hawthorn race and was subsequently introduced into the Hudson Valley region of New York. First, genetic bottlenecks due to reductions in population size frequently occur when a species colonizes a new area. Reduced amounts of genetic variation have, in fact, been observed for newly introduced populations of *R. pomonella* and *R. completa* in the western United States (McPheron *et al.*, 1988c; Berlocher, 1984b). However, apple populations in the eastern United States have as much or more allelic diversity as eastern hawthorn populations (Chapter 3). There is, therefore, no genetic evidence indicating that the apple race was introduced into the New England area.

Second, the *R. pomonella* species group is endemic to North America (Bush, 1966), thereby ruling out the possibility that the apple race originated in Europe and was brought to the United States along with cultivated apples. Therefore, if the apple race formed in allopatry, it must have originated somewhere in North America. The range of domestic apples is, however, contained entirely within that of hawthorns in North America. Consequently, even if *R. pomonella* did not originally shift onto apples in the Hudson Valley, the apple race could not have been geographically isolated from hawthorn infesting populations at the time of its inception.

Third, B.D. Walsh (1867) noted at the time of the initial discovery of the apple race that although *R. pomonella* "exists both in the East and in the West, it attacks the cultivated apple only in a certain limited region, even in the East, for this new and formidable enemy of the apple is found in the Hudson-river valley, but has not yet reached New Jersey". Therefore, up until 1867, no apple infesting population of *R. pomonella* had been reported outside of the New England area. We now know, however, that an isolated population of *R. pomonella* does infest both hawthorns and domestic apples in the highlands of Mexico (Bush, 1966). But Mexican flies are unique in that they possess a hyaline spot at the base of the apical band on their wings which makes them morphologically distinguishable from all other taxa in the *R. pomonella* group (Bush, 1966). The apple race in the United States could not, therefore, have been derived from the Mexican population.

Although the Mexican population does not clarify the issue of the geographic origin of the apple race, it does lend further indirect support for hawthorns as the original host for the apple race. In Mexico, *R. pomonella* has independently shifted from hawthorns to domestic apples within historical times (apples were introduced into Mexico from Spain in 1522; Standley, 1922). However, Mexico does not have any endemic species of crab apple (Standley, 1922; J. Beaman, pers. comm.). Therefore, in Mexico at least, *R. pomonella* shifted directly from hawthorns to domestic apples.

Two factors are primarily responsible for the genetic differences observed between hawthorn and apple populations of *R. pomonella*. First, adult flies do not migrate randomly between hawthorn and apple trees. This was seen from the fact that adult flies collected off of hawthorn and apple trees at site 15 showed the same pattern and magnitude of genetic divergence as their larval offspring. Host fidelity therefore exists for these flies with adults tending to utilize the same species of host plant which they infested as larvae.

Second, differential selection occurs for flies infesting apples and hawthorns and appears to be related in some way with ambient temperature. The latitudinal allele frequency clines among both apple and hawthorn populations suggest that selection pressures are not constant across the ranges of both host races, however. Allele frequency clines are also steeper for hawthorn flies than for apple flies and intersect between the two host races in the central portion of *R. pomonella*'s range (Figures 6-8). This accounts for the greater amount of geographic differentiation in the hawthorn race as well as for why northern and southern sites show higher levels of inter-host divergence than centrally located areas in the Midwest.

Further research is needed in three areas to clarify specific details of the host specialization process. These include: 1.) establishing the extent of and mechanisms responsible for host fidelity and host related selection; 2.) determining how temperature acts as a differential selective factor and 3.) estimating effective levels of gene flow among and between apple and hawthorn populations.

Mark-recapture and controlled rearing experiments would help address questions concerning host fidelity and gene flow. Collecting larvae from infested fruit and marking and releasing the resulting "naive" adults the following summer would provide estimates of intra- and inter-host migration. Hardwired behavioral differences in host recognition and conditioning in adults are two likely mechanisms responsible for host fidelity (larval conditioning has never been found for *Rhagoletis* flies but adult conditioning has; Prokopy *et al.*, 1982). Releasing naive apple flies under hawthorn trees and vice-versa would establish the relative importance of genetically based differences in host preference. Adult flies could also be exposed to alternative host fruits in the laboratory prior to their release. Analyses of the subsequent host acceptance behavior of these "trained" flies under field conditions would indicate the extent to which adult conditioning affects host choice. Furthermore, by combining the results of mark-recapture studies with those from reciprocal egg transplant experiments measuring survivorship levels of progeny derived from hawthorn x hawthorn, hawthorn x apple and apple x apple crosses, I will be able to directly estimate effective gene flow between the host races.

Ambient temperature has always been regarded as an important environmental factor influencing host plant-parasite interactions for *Rhagoletis*. *Rhagoletis pomonella* is, in general, univoltine across its range and adult life expectancy has been estimated at from 3 to 6 weeks in the field (see review by Boller and Prokopy, 1976). Adults must therefore eclose at times paralleling the fruiting phenologies of their host plants. "Temporal windows" for infesting apples and hawthorns are different with the fruit of domestic apple varieties favored by *R. pomonella* generally ripening and falling one month earlier than fruit of native hawthorn species. In addition, flies must also develop as larvae and pupae at rates closely matching local environmental conditions. If development proceeds too quickly flies may not diapause and emerge as a second generation in the fall when suitable host fruit is no longer available. Alternatively, if development occurs too slowly flies are exposed to potential predators and parasites for a longer period of time and also run the risk of freezing to death during the first frost. Ambient temperature does, of course, correlate with latitude and is also a major determinant of the phenology of fruit maturation.

Temperature may therefore have a profound effect on development rates for hawthorn and apple flies across their respective ranges. Consequently, latitudinal variation in temperature could account for the allele frequency clines in the two races. Furthermore, differences in the fruiting phenologies of apple and hawthorn trees could explain why allele frequency clines differ between the two host races.

The temperature hypothesis is contingent, of course, on *Me*, *Acon-2*, *Mpi*, *Dia-2* and *Had* being related in some way with developmental timing in *R. pomonella*. Elucidating the genetic basis for diapause related traits in insects has proven to be a difficult task, however. Although diapause characters do show latitudinal variation associated with temperature and photoperiod (see review by Tauber and Tauber, 1976), few examples exist documenting the genetic basis for this phenotypic variation. One possible exception may be the pitcher plant mosquito, *Wyeomyia smithii*, in which a weak association has been found between *Phosphoglucomutase* and *Phenylalanyl-leucine amino-peptidase* (two loci showing latitudinal frequency clines within the species) and phenotypic variation for diapause and developmental timing (Istock and Weisberg, 1987). Genetic and/or phenotypic markers associated with diapause characters have also been found for the green lacewings, *Chrysopa carnea* and *C. downesi* (Tauber *et al.*, 1977); the European corn borer, *Ostrinea nubilalis* (Shower, 1981); and the eastern tiger swallowtail butterfly, *Papilio glaucus* (Rockey *et al.*, 1987). *R. pomonella* appears to represent another example as preliminary studies indicate that allele frequencies for *Me*, *Acon-2* and *Had* correlate with the timing of diapause termination and adult eclosion for hawthorn flies (Feder *et al.*, manuscript in preparation)

A potentially interesting test of the temperature/development hypothesis for *R. pomonella* would be to sample hawthorn and apple fly populations along altitudinal transects in the Appalachian mountains. The prediction would be that if temperature is important and migration rates are low or moderate, allele frequency clines similar to those observed in the Midwest should also be found for the two host races along the slopes of these mountains. Larval transplant experiments performed under controlled rearing conditions would also provide valuable information on how temperature influences host plant interactions for *R. pomonella.*

It is not uncommon for allozyme surveys to uncover one or more loci whose alleles display a clinal pattern of variation correlated with some aspect of the physical environment (Johnson, 1971; Clegg and Allard, 1972; Hamrick and Allard, 1972; Christiansen and Frydenberg, 1974; Koehn *et al.*, 1976; see review by Hedrick *et al.*, 1976). Latitudinal, seasonal and altitudinal variation in temperature have been implicated as causal factors for a number of allozyme clines including those for the freshwater fish, *Catostomus Clarkii* (Koehn, 1969); the marine fish, *Fundulus heteroclitus* (Mitton and Koehn, 1975); the fathead minnow, *Pimephales promelas* (Merritt, 1972); the fruit fly, *Drosophila melanogaster* (Berger, 1971; Johnson and Schaffer, 1973; Miller, Pearcey and Berger, 1975); and the butterfly, *Colias meadii* (Johnson, 1976). What is unusual about *R. pomonella* is that ambient temperature apparently has different consequences for conspecific populations infesting different host plants. Host plant associations therefore add an additional level of environmental heterogeneity for *R. pomonella* which helps maintain increased amounts of intraspecific genetic variation; a result consistent with predictions of niche theory (Levene, 1953; Maynard-Smith, 1966, 1970; Levins, 1968, Christiansen and Feldman, 1975; Felsenstein, 1976).

Allele frequency clines for hawthorn and apple flies are not completely discordant, however, and regions where the two races show similar patterns of genetic variation have important evolutionary implications. The three areas where the largest allele frequency changes occur for apple and/or hawthorn populations are located in or near major ecological transition zones in the Midwest. Sites 1-4 in Michigan and 20-21 in Wisconsin are situated at the boundary between deciduous eastern forest and arboreal lake forest habitats. Likewise, sites 25-26 near the Illinois/Wisconsin border lie where the central prairie meets deciduous eastern forest. The ecological consequences of these zones are significant and mark cutoff points in the distribution of numerous plant species. Changes in voltinism and/or host plant associations occur across these regions for several different phytophagous insects including the European corn-borer, (Showers, 1981) and the swallowtail butterfly, *Papilio glaucus* (for review see Scriber and Hainze, 1987). The finding of a similar pattern for *R. pomonella* supports the hypothesis that these zones have a general impact on the evolution of life history traits for phytophagous insects.

The observed frequency clines in *R. pomonella* could, of course, be due to secondary contact rather than selection. However, there is no historical or biogeographic evidence indicating that geographically isolated northern and southern populations of *R. pomonella* ever existed. It is therefore much more likely that *R. pomonella* recolonized northern latitudes along with hawthorns following the retreat of ice sheets in North America some 10,000 to 20,000 years ago than for populations of the fly to have expanded from pleistocene refugia. Also, although apple and hawthorn flies both display allele frequency clines, the configurations of these clines do show some differences between the two host races and among different loci within the races. Furthermore, *Me*^{100} and *Acon-2*^{95} frequencies for hawthorn populations co-vary with irregularities in local thermal conditions. These latter two observations are not consistent with the secondary contact hypothesis and argue for differential selection as the causal basis for allele frequency clines in *R. pomonella.*

Another important question is whether apple and hawthorn host races are incipient species. A host race is a parasitic population which, by adapting to a preferred host, has become partially reproductively isolated from other conspecific populations specialized on alternative hosts (Diehl and Bush, 1984). Host races may therefore be at various stages of divergence ranging from populations which almost freely interbreed to those that rarely exchange genes. Based on available data, it is difficult to judge whether apple and hawthorn flies have reached or will ever reach the evolutionary point of no return when they should be considered distinct species. In essence, the results of this and the preceding chapters have shown that host specific traits can evolve within geographically continuous populations. The key question then is whether these host associated adaptations are sufficient, in and of themselves, to eventually cause the complete reproductive isolation of populations. Indirect support that they are has come from additional population genetic studies involving sibling species in the *R. pomonella* complex. For instance, *R. mendax* is a closely related sibling species to *R. pomonella* which infests blueberries (*Vaccinium* sp.) and huckleberries (*Gaylussacia* sp.) in the eastern United States and Canada (Bush, 1966). Although field captured *R. pomonella* and *R. mendax* adults will hybridize in the laboratory and produce fertile and viable F1 progeny (J. Frey pers. comm.; Feder and Bush, submitted), sympatric populations of the sibling species are genetically distinct in nature and to not interbreed (Berlocher and Bush, 1982; Feder *et al.*, submitted). Genetic analysis of reproductively mature adults collected from interdigitated blueberry bushes and apple trees indicate that *R. mendax* and *R. pomonella* maintain 100% complete host fidelity in the field (Feder and Bush, submitted). Differential host plant recognition is therefore a very effective pre-mating isolating barrier between at least two sympatrically distributed sibling species in the *R. pomonella* group.

The allele frequency clines indicate that both hawthorn and apple races are not evolving uniformly across the Midwest, however. Primary frequency clines have been implicated in speciation (Fisher, 1930; Murray, 1972; Endler, 1977) and it is possible that *R. pomonella* populations are "isolated by distance" as a consequence of being differentially adapted to local thermal conditions. Both hawthorn and apple races did display sharp allele frequency "steps" across environmental transition zones in the Midwest (Figures 7, 8). In addition, hawthorn populations at the northern and southern ends of *R. pomonella's* range are essentially fixed for alternative alleles at the loci *Had*, *Dia-2* and *Me* (McPheron, 1987). *R. pomonella* populations at different latitudes may therefore be "genetically incompatible" with one another even though they are inter-connected by a series of populations exchanging genes. It will be interesting to see whether allele frequency clines for *R. pomonella* can become steep enough to result in "hybrid zones" given the fly's apparent ability to disperse over relatively long distances.

Whether or not *R. pomonella* host races are "cohesive evolutionary units" is not central to the issue of sympatric speciation, however. The important question is whether host recognition behaviors are constant across the range of a race. For example, if flies infesting hawthorns continue to prefer hawthorns when experimentally transplanted into different portions of their distribution, then host shifts ensure the divergence of populations regardless of whether selection pressures vary from locality to locality within races. The results from site 15 confirm that some host fidelity exists for *R. pomonella* in at least one portion of the fly's range. However, apple and hawthorn populations at sites 4 and 32 showed no significant genetic differentiation (Table 34). Is the lack of genetic divergence between apple and hawthorn populations at these two sites due to high levels of inter-host gene flow? Or, alternatively, is host fidelity still strong at sites 4 and 32 but selection pressures on *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* and *Had* the same for apple and hawthorn flies? The shapes of the allele frequency clines among hawthorn and apple populations suggest that the latter possibility is true. However, detailed studies measuring inter-host gene flow across the range of *R. pomonella* are needed to resolve this point and establish the extent to which apple and hawthorn races are discrete populations on separate evolutionary pathways.

In conclusion, *R. pomonella* is rapidly becoming a model organism for the study of speciation. Genetic differences exist between recently formed host races of the species and the geographic pattern of allozyme variation suggests that differences in host plant utilization and host associated development (in some way related to ambient temperature) are responsible for reducing gene flow between apple and hawthorn races. My next task is to characterize exactly how these factors operate at the population level and to translate their mode of action to the speciation process. In this regard, the ability to hybridize several different *R.*

pomonella species in the laboratory will help us elucidate the genetic basis for host associated traits, such as differential host recognition, which appear to play a major role in reproductively isolating taxa in the *R. pomonella* group.

Table 39. Allele frequencies for *Acon-2* for hawthorn (H) and apple (A) populations of *R. pomonella* across the mid-western United States. At sites 4, 15 and 17 where both larvae and adults were collected, (L) refers to larvae and (Ad) to adults. Sites are arranged within the five transects according to latitude from north to south.

Acon-2

Site	Host	N	73	75	89	95	100	106	114	Host	N	73	75	89	95	100	106	114
Transect 1 (Mi./In.)																		
1	H	51		0.029	0.059	0.676	0.157	0.059	0.020	A	50		0.020	0.050	0.260	0.510	0.160	
2	H	47		0.011	0.011	0.372	0.479	0.128		A	39		0.102	0.026	0.192	0.538	0.128	0.013
3	H	38		0.132	0.026	0.526	0.276	0.040		A	47		0.245	0.053	0.255	0.372	0.074	
4(Ad)	H	30			0.017	0.200	0.650	0.133		A	-	-	-	-	-	-	-	-
4(L)	H	28				0.214**	0.696*	0.072	0.018	A	29		0.017		0.155	0.638	0.190	
5	H	44		0.079	0.023	0.295	0.500	0.079	0.023	A	41		0.024	0.012	0.098	0.646	0.220	
6	H	36		0.361	0.069	0.222*	0.278	0.069		A	47		0.223	0.032	0.117	0.511	0.117	
7	H	30		0.133	0.050	0.117	0.617	0.050	0.033	A	-	-	-	-	-	-	-	-
8	H	35		0.014	0.029	0.043	0.743	0.171		A	29				0.155	0.552	0.293	
9	H	51			0.029	0.078	0.794	0.078	0.020	A	52		0.010		0.067	0.788	0.125	0.010
Transect 2 (Mi.)																		
10	H	39			0.043	0.696	0.250	0.011		A	39		0.064	0.089	0.372	0.372	0.103	
11	H	-	-	-	-	-	-	-	-	A	41		0.024	0.049	0.256	0.524	0.146	
12	H	40		0.138	0.037	0.312	0.425	0.088		A	45		0.189	0.089	0.089	0.511	0.122	
13	H	48		0.135	0.115	0.167	0.552	0.021	0.010	A	34		0.088	0.044	0.147	0.544	0.162	0.015
Transect 3 (Mi.)																		
14	H	42		0.012	0.107	0.619	0.250	0.012		A	50		0.100	0.060	0.160**	0.500	0.180	
15(Ad)	H	317		0.047	0.062	0.435	0.377	0.077	0.002	A	357		0.083	0.040	0.164	0.542	0.165	0.006
15(L)	H	429		0.026	0.062	0.479	0.346	0.078	0.009	A	665		0.064	0.029	0.194	0.560	0.148	0.005
16	H	32		0.172	0.016	0.391	0.344	0.078		A	36		0.157	0.043	0.100	0.543	0.157	
17(Ad)	H	36		0.153	0.028	0.361	0.333	0.111	0.014	A	-	-	-	-	-	-	-	-
17(L)	H	32		0.047	0.016	0.406	0.469	0.062		A	30		0.054	0.036	0.054	0.820	0.036	
18	H	32		0.094	0.062	0.250	0.532	0.062		A	31		0.097	0.048	0.177*	0.436	0.242	

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg equilibrium.

Table 39. (cont'd.)

Acon-2

Site	Host	N	73	75	89	95	100	106	114	Host	N	73	75	89	95	100	106	114
Transect 4 (Wi./Ill.)																		
19	H	111		0.054	0.122	0.504	0.266	0.045	0.009	A	106	0.005	0.014	0.024	0.259	0.514	0.175	0.009
20	H	158	0.003	0.009	0.070	0.592	0.256	0.066	0.003	A	132	0.004		0.045	0.348	0.436	0.163	0.004
21	H	43		0.023	0.081	0.279*	0.511	0.105		A	40	0.025	0.025	0.075	0.088	0.562	0.213	0.012
22	H	38		0.026	0.053	0.329	0.487	0.079	0.026	A	45		0.022	0.067	0.156	0.589	0.144	0.022
23	H	43		0.058	0.046	0.221*	0.523	0.151		A	50		0.030	0.070	0.180	0.610	0.110	
24	H	29		0.018	0.086	0.310	0.500	0.086		A	49		0.010	0.061	0.102	0.673	0.154	
25	H	49		0.010	0.143	0.398	0.408	0.041		A	79		0.032	0.070	0.108	0.595	0.190	0.006
26	H	82			0.061	0.134	0.683	0.116	0.006	A	76	0.013	0.020	0.033	0.151	0.586	0.171	0.026
27	H	51		0.029	0.167	0.176*	0.490	0.127	0.010	A	-	-	-	-	-	-	-	-
28	H	60	0.017	0.017	0.125	0.083	0.642	0.092	0.025	A	60	0.092	0.017	0.099	0.117	0.475	0.183	0.017
Transect 5 (Wi.)																		
29	H	32			0.125	0.484	0.344	0.047		A	42			0.062	0.238	0.538	0.150	0.012
30	H	44		0.068	0.239	0.420	0.193	0.011		A	44		0.093	0.116	0.140	0.593	0.058	
31	H	29		0.017	0.103	0.345	0.448	0.086		A	-	-	-	-	-	-	-	-
32	H	45		0.011	0.200	0.089	0.489	0.211		A	44		0.013	0.105	0.184	0.539	0.158	
Misc. (Wi.)																		
33	H	18			0.110	0.306	0.528	0.056		A	-	-	-	-	-	-	-	-
34	H	36			0.236	0.125	0.528	0.111		A	-	-	-	-	-	-	-	-

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg equilibrium.

Table 40. Allele frequencies for *Me*, *Had* and *Dia-2* for hawthorn (H) and apple (A) populations of *R. pomonella* across the mid-western United States. At sites 4, 15 and 17 where both adults and larvae were collected, (L) refers to larvae and (Ad) to adults. Sites are arranged within transects according to latitude from north to south.

	Me						**Had**						**Dia-2**					
Site	Host	N	80	Host	N	80	Host	N	122	Host	N	122	Host	N	85	Host	N	85
Transect 1 (Mi./In.)																		
1	H	46	0.239	A	50	0.660	H	49	0.153	A	51	0.206	H	49	0.224*	A	51	0.304
2	H	47	0.532	A	40	0.687	H	47	0.245	A	39	0.346	H	47	0.277**	A	28	0.446**
3	H	38	0.421	A	48	0.656	H	38	0.276	A	48	0.333	H	32	0.281**	A	44	0.307
4(Ad)	H	30	0.667	A	-	-	H	29	0.241	A	-	-	H	30	0.233*	A	-	-
4(L)	H	30	0.717	A	30	0.717	H	30	0.183	A	30	0.250**	H	28	0.232	A	30	0.233
5	H	44	0.545	A	45	0.702	H	45	0.256	A	45	0.278	H	44	0.273	A	45	0.389
6	H	36	0.736	A	49	0.735	H	36	0.292	A	49	0.245	H	36	0.278	A	48	0.281
7	H	30	0.700	A	-	-	H	31	0.468	A	-	-	H	27	0.333	A	-	-
8	H	35	0.757	A	30	0.760	H	35	0.500	A	30	0.433	H	36	0.375	A	30	0.350
9	H	40	0.800*	A	52	0.789	H	40	0.475	A	52	0.279	H	40	0.388*	A	41	0.341
Transect 2 (Mi.)																		
10	H	50	0.150	A	39	0.538	H	50	0.060	A	39	0.282	H	47	0.160	A	39	0.295
11	H	-	-	A	41	0.537	H	-	-	A	42	0.238	H	-	-	A	38	0.329
12	H	40	0.587	A	45	0.837	H	40	0.250	A	45	0.300	H	40	0.275	A	45	0.267
13	H	48	0.771	A	34	0.706	H	48	0.458*	A	33	0.288	H	47	0.404	A	31	0.274
Transect 3 (Mi.)																		
14	H	42	0.226	A	50	0.480	H	42	0.107	A	50	0.180	H	31	0.145*	A	47	0.319
15(Ad)	H	345	0.407	A	357	0.620	H	340	0.154	A	356	0.226	H	313	0.244	A	307	0.298
15(L)	H	467	0.345	A	706	0.633	H	464	0.170	A	688	0.203	H	407	0.206	A	672	0.310
16	H	33	0.561	A	36	0.708	H	33	0.303	A	36	0.292	H	33	0.136	A	35	0.386
17(Ad)	H	25	0.660	A	-	-	H	29	0.362	A	-	-	H	38	0.302	A	-	-
17(L)	H	33	0.637	A	30	0.634	H	33	0.318	A	30	0.183	H	29	0.327	A	30	0.467
18	H	32	0.750	A	31	0.500	H	32	0.313	A	31	0.322	H	32	0.359	A	31	0.484

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg equilibrium.

Table 40. (cont'd.)

	Me						Had						Dia-2					
Site	Host	N	80	Host	N	80	Host	N	122	Host	N	122	Host	N	85	Host	N	85
Transect 4 (Wi./Ill.)																		
19	H	110	0.309	A	106	0.542	H	109	0.220	A	106	0.208	H	108	0.218	A	105	0.329
20	H	156	0.304*	A	133	0.553	H	157	0.137	A	133	0.226	H	155	0.132	A	132	0.326
21	H	43	0.535	A	42	0.714	H	43	0.186	A	42	0.226	H	41	0.232	A	41	0.256
22	H	39	0.590	A	45	0.689	H	25	0.120	A	45	0.278	H	37	0.284	A	45	0.322
23	H	43	0.605	A	51	0.666	H	43	0.186	A	51	0.275	H	43	0.384	A	51	0.363
24	H	29	0.552	A	50	0.680**	H	15	0.333	A	47	0.287	H	29	0.575	A	49	0.337
25	H	49	0.623	A	80	0.719*	H	34	0.235	A	65	0.200	H	34	0.265	A	65	0.315
26	H	84	0.780	A	82	0.604	H	84	0.357	A	83	0.367	H	77	0.260	A	74	0.331**
27	H	53	0.774	A	-	-	H	52	0.538	A	-	-	H	43	0.395	A	-	-
28	H	60	0.825	A	60	0.742	H	60	0.567	A	60	0.383	H	58	0.500	A	60	0.400
Transect 5 (Wi.)																		
29	H	32	0.470	A	42	0.595	H	32	0.125	A	42	0.083	H	32	0.156	A	42	0.143
30	H	44	0.568	A	44	0.795	H	44	0.295*	A	44	0.250	H	39	0.295	A	42	0.321
31	H	29	0.930	A	-	-	H	29	0.345	A	-	-	H	27	0.296	A	-	-
32	H	45	0.667	A	44	0.636	H	43	0.372	A	44	0.341	H	38	0.250	A	37	0.365
Misc. (Wi.)																		
33	H	20	0.750	A	-	-	H	20	0.250	A	-	-	H	20	0.225	A	-	-
34	H	37	0.815	A	-	-	H	36	0.333	A	-	-	H	35	0.314*	A	-	-

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg equilibrium.

Table 41. Allele frequencies for *Mpi* for hawthorn (H) and apple (A) populations of *R. pomonella* across the mid-western United States. At sites 4, 15 and 17 where both adults and larvae were collected, (L) refers to larvae and (Ad) to adults. Sites are arranged within transects according to latitude from north to south.

Mpi

Site	Host	N	31	37	70	100	125	Host	N	31	37	70	100	125
Transect 1 (Mi./In.)														
1	H	37	0.013	0.189	0.095	0.703		A	34	0.029		0.015	0.941	0.015
2	H	47		0.138	0.043	0.819		A	33	0.030	0.030		0.940	
3	H	37	0.040	0.136	0.040	0.784		A	35	0.086	0.100	0.029	0.786	
4(Ad)	H	30		0.050	0.067	0.883		A	-	-	-	-	-	-
4(L)	H	30		0.050	0.067	0.883		A	30		0.017	0.033	0.950	
5	H	45		0.067	0.044	0.867	0.022	A	38	0.013	0.066	0.026	0.895	
6	H	36	0.028	0.069	0.042	0.861		A	28	0.107	0.054		0.839	
7	H	31			0.016	0.968	0.016	A	-	-	-	-	-	-
8	H	35	0.014	0.057	0.029	0.900		A	30		0.050		0.950	
9	H	40	0.025		0.050	0.925		A	50	0.060	0.030	0.050	0.860	
Transect 2 (Mi.)														
10	H	29	0.035	0.241	0.035	0.672	0.017	A	33	0.015	0.136	0.091	0.758	
11	H	-	-	-	-	-	-	A	42	0.012	0.095		0.893	
12	H	40	0.025	0.050	0.088	0.838		A	45	0.022	0.067	0.056	0.833	0.022
13	H	30		0.050	0.033	0.900	0.017	A	34	0.015	0.089	0.059	0.823	0.015
Transect 3 (Mi.)														
14	H	42	0.012	0.214	0.048	0.714	0.012	A	47	0.032	0.042	0.011	0.915	
15(Ad)	H	308	0.013	0.112	0.047	0.818	0.010	A	353	0.013	0.045	0.054	0.885	0.004
15(L)	H	417	0.035	0.096	0.047	0.818***	0.005	A	660	0.011	0.049	0.046	0.892	
16	H	28	0.036	0.143	0.036	0.786		A	34		0.059	0.074	0.867	
17(Ad)	H	31	0.039	0.013	0.066	0.882		A	-	-	-	-	-	-
17(L)	H	38	0.016	0.064	0.081	0.839		A	27	0.019	0.037	0.074	0.870	
18	H	32	0.016	0.078	0.016	0.890		A	31		0.065	0.080	0.855	

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-test for significant deviation from Hardy-Weinberg equilibrium.

Table 41. (cont'd.)

Mpi

Site	Host	N	31	37	70	100	125	Host	N	31	37	70	100	125
Transect 4 (Wi./Ill.)														
19	H	104	0.024	0.149	0.034	0.788	0.005	A	99	0.020	0.040	0.076	0.854	0.010
20	H	158		0.199	0.092	0.668	0.041	A	132	0.004	0.091	0.079	0.789	0.038
21	H	43	0.012	0.105	0.047	0.802	0.023	A	38	0.013	0.013	0.066	0.908	
22	H	39		0.102	0.026	0.872		A	45	0.022	0.033	0.033	0.912	
23	H	40		0.125	0.087	0.788		A	31	0.032	0.032	0.032	0.903	
24	H	21		0.143	0.023	0.833		A	40	0.012	0.063	0.050	0.875	
25	H	40	0.025	0.113		0.862		A	80	0.044	0.044	0.025	0.888	
26	H	83	0.042	0.030	0.048	0.880		A	79	0.025	0.006	0.070	0.899	
27	H	53	0.028	0.010	0.019	0.943		A	-	-	-	-	-	-
28	H	60	0.017	0.017	0.033	0.933		A	45	0.022		0.044	0.933	
Transect 5 (Wi.)														
29	H	32		0.062	0.062	0.860	0.016	A	41	0.037	0.037	0.037	0.853	0.037
30	H	42	0.012	0.083	0.107	0.762	0.036	A	40	0.025	0.050	0.113	0.812	
31	H	29			0.034	0.966		A	-	-	-	-	-	-
32	H	45			0.044	0.956		A	34	0.029		0.029	0.942	
Misc. (Wi.)														
33	H	20	0.075	0.025	0.050	0.825	0.025	A	-	-	-	-	-	-
34	H	33	0.076	0.030	0.015	0.879		A	-	-	-	-	-	-

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg equilibrium.

Table 42. Allele frequencies for *Aat-2* for hawthorn (H) populations of *R. pomonella* across the mid-western United States. At sites 4, 15 and 17 where both larvae and adults were collected, (L) refers to larvae and (Ad) to adults. Sites are arranged within the five transects according to latitude from north to south.

Aat-2

Site	Host	N	01	21	32	40	50	59	75	84	100	112	123	130
Transect 1 (Mi./In.)														
1	H	43		0.070			0.151		0.337		0.430		0.012	
2	H	47		0.074			0.223		0.245		0.383		0.064	
3	H	34		0.044			0.235*		0.250		0.426		0.044	
4(Ad)	H	30		0.100			0.134	0.017	0.433		0.266		0.050	
4(L)	H	27	0.018	0.111			0.167		0.259		0.426		0.018	
5	H	37		0.148			0.230		0.230		0.365		0.027	
6	H	36	0.028	0.069	0.028		0.222	0.042	0.139		0.417		0.042	0.014
7	H	25		0.140			0.140	0.020	0.280		0.380		0.040	
8	H	34		0.206			0.235*		0.176		0.353		0.029	
9	H	40		0.200			0.200	0.038	0.175		0.375			0.013
Transect 2 (Mi.)														
10	H	38		0.066			0.118		0.276		0.513		0.026	
11	H	-	-	-	-	-	-	-	-	-	-	-	-	-
12	H	34		0.118			0.162		0.265		0.412		0.044	
13	H	47	0.011	0.106			0.309	0.011	0.213		0.351			
Transect 3 (Mi.)														
14	H	42		0.059			0.167		0.274		0.476		0.024	
15(Ad)	H	271	0.006	0.103	0.006		0.138	0.013	0.275		0.415		0.039	0.006
15(L)	H	414	0.001	0.091	0.002	0.001	0.127		0.271		0.461		0.040	0.006
16	H	33		0.061			0.076		0.303		0.545		0.015	
17(Ad)	H	36		0.153	0.014		0.167		0.181		0.389		0.097	
17(L)	H	27		0.167	0.037		0.148		0.241		0.389		0.018	
18	H	32		0.140			0.234		0.250		0.359		0.016	

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for deviation from Hardy-Weinberg equilibrium.

Table 42. (cont'd.)

Aat-2

Site	Host	N	01	21	32	40	50	59	75	84	100	112	123	130
Transect 4 (Wi./Ill.)														
19	H	92	0.005	0.077	0.005		0.141		0.353		0.386		0.027	0.005
20	H	156		0.083	0.003		0.080	0.003	0.353		0.420	0.003	0.048	0.006
21	H	41		0.098	0.012		0.134	0.012	0.280	0.012	0.402		0.049	
22	H	37		0.176			0.176		0.216*		0.392		0.027	0.013
23	H	43	0.023	0.128			0.221	0.012	0.116		0.477		0.012	0.012
24	H	23		0.130			0.456		0.152		0.261*			
25	H	48	0.010	0.135			0.198		0.198*		0.427		0.031	
26	H	80		0.106			0.156	0.006	0.213		0.475		0.044	
27	H	52		0.183			0.183*		0.192		0.433		0.009	
28	H	60	0.017	0.142			0.317		0.200		0.283		0.025	0.017
Transect 5 (Wi.)														
29	H	28		0.089			0.161	0.018	0.339		0.339		0.054	
30	H	44	0.011	0.136			0.136	0.011	0.307		0.375		0.023	
31	H	18		0.167			0.167		0.083		0.528		0.056	
32	H	43		0.105			0.221		0.232		0.407		0.035	
Misc. (Wi.)														
33	H	20		0.075			0.175		0.225		0.475		0.050	
34	H	33		0.121			0.152		0.242		0.394*		0.091	

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for deviation from Hardy-Weinberg equilibrium.

Table 43. Allele frequencies for *Aat-2* for apple (A) populations of *R. pomonella* across the mid-western United States. At sites 4, 15 and 17 where both larvae and adults were collected, (L) refers to larvae and (Ad) to adults. Sites are arranged within transects according to latitude from north to south.

Aat-2

Site	Host	N	01	21	32	40	50	59	75	84	100	113	123	130
Transect 1 (Mi./In.)														
1	A	50		0.140			0.160		0.330		0.330		0.040	
2	A	34		0.191*			0.236*		0.279	0.015	0.250		0.029	
3	A	44		0.170		0.011	0.273		0.227		0.295*		0.023	
4(Ad)	A	-	-	-	-	-	-	-	-	-	-	-	-	-
4(L)	A	29		0.069			0.259		0.310		0.328		0.017	0.017
5	A	41		0.220			0.134		0.268		0.354		0.024	
6	A	46		0.130			0.196	0.011	0.272**	0.011	0.326		0.054	
7	A	-	-	-	-	-	-	-	-	-	-	-	-	-
8	A	30		0.150			0.283		0.283		0.283			
9	A	50		0.090			0.220	0.010	0.300	0.010	0.330		0.040	
Transect 2 (Mi.)														
10	A	38		0.132			0.237*	0.013	0.263		0.381		0.013	
11	A	36	0.014	0.097			0.153		0.292		0.403		0.042	
12	A	42		0.119			0.190		0.381		0.274		0.036	
13	A	34		0.118			0.190		0.221		0.368		0.103	
Transect 3 (Mi.)														
14	A	47		0.181	0.032		0.149	0.011	0.447		0.181			
15(Ad)	A	297	0.002	0.163	0.003		0.175		0.300		0.337		0.020	
15(L)	A	662	0.002	0.150	0.002	0.001	0.186*	0.001	0.285**		0.344*		0.026	0.002
16	A	33		0.288			0.242		0.121		0.319			
17(Ad)	A	-	-	-	-	-	-	-	-	-	-	-	-	-
17(L)	A	25		0.160			0.300		0.140		0.400			
18	A	31	0.016	0.161			0.306		0.242		0.194		0.065	0.016

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for deviation from Hardy-Weinberg equilibrium.

Table 43. (cont'd.)

Aat-2

Site	Host	N	01	21	32	40	50	59	75	84	100	113	123	130
Transect 4 (Wi./Ill.)														
19	A	104		0.149	0.010		0.168		0.332		0.327		0.010	0.005
20	A	133	0.004	0.180	0.008		0.173	0.041	0.274		0.286		0.026	0.008
21	A	39		0.115	0.026		0.141	0.013	0.372		0.333			
22	A	45		0.167			0.200*		0.278		0.344		0.011	
23	A	51		0.176			0.167		0.304		0.294		0.038	0.020
24	A	49		0.153	0.020		0.173	0.010	0.245		0.388		0.010	
25	A	76		0.184			0.151		0.270		0.349		0.046	
26	A	71		0.190	0.007		0.176		0.296		0.289	0.014	0.028	
27	A	-	-	-	-	-	-	-	-	-	-	-	-	-
28	A	60		0.167			0.233		0.258		0.317		0.008	0.017
Transect 5 (Wi.)														
29	A	41		0.085			0.122	0.016	0.293		0.439		0.024	
30	A	42		0.119			0.202	0.012	0.286	0.012	0.369			
31	A	-	-	-	-	-	-	-	-	-	-	-	-	-
32	A	33		0.167	0.015		0.182		0.197		0.409		0.021	
Misc. (Wi.)														
33	A	-	-	-	-	-	-	-	-	-	-	-	-	-
34	A	-	-	-	-	-	-	-	-	-	-	-	-	-

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-test for deviation from Hardy-Weinberg equilibrium.

Chapter 5

Local and Microgeographic Genetic Variation between Host Races of *Rhagoletis pomonella*

Introduction

The niche-variation hypothesis (Levene, 1953; Van Valen, 1965; Levins, 1968) predicts that genetic polymorphism is correlated with the ability of a population to utilize different environmental resources. If true, then differential selection in heterogeneous environments should be an important factor maintaining genetic variation (Levene, 1953; Levins, 1968; Maynard-Smith, 1970) and promoting speciation (Maynard-Smith, 1966) in natural populations.

Niche theory has found wide application in the study of phytophagous insects due in part to the existence of numerous pest "biotypes" which attack economically important plant cultivars (see review by Diehl and Bush, 1984). Differential habitat use is also central to a long-standing debate concerning modes of speciation. Speciation in animals is generally thought to require the geographic isolation of populations to eliminate the homogenizing effects of gene flow (Mayr,1963). However, Bush (1966, 1969a,b, 1975) has proposed that speciation in certain phytophagous and parasitic insects occurs sympatrically and is initiated when insects shift and become affiliated with new hosts.

Although a number of examples demonstrate associations between particular phenotypic or genotypic attributes of a population and some aspect of the environment (see reviews by Hedrick *et al.*, 1976 and McDonald, 1983), documenting the existence of host associated polymorphism in sexually reproducing insects has been difficult (Futuyma and Mayer, 1980; Futuyma and Peterson, 1985). First, it must be established that populations utilizing different hosts are conspecific and not, in fact, distinct species. Second, characters varying between habitats must have a genetic basis and not be environmentally induced modifications (this is not a particularly strict requirement for allozyme or molecular genetic studies). Third, resource related polymorphism has to be consistent on both fine and broad spatial scales, as well as being temporally stable, to distinguish it from variation due to stochastic processes such as founder effects and genetic drift. Examples of resource associated variation do exist for sympatric populations (for reviews see Diehl and Bush, 1984 and Futuyma and Peterson, 1985) but, in almost all of these cases, the evolutionary and/or geographic origin of the host-insect relationship are not known.

True fruit flies in the *Rhagoletis pomonella* sibling species complex have a biology and natural history which make them ideal candidates for studies of host associated polymorphism and speciation. *Rhagoletis* larvae are obligate internal parasites in the fruit of their host plants and each member of the *R. pomonella* group is monophagous or oligiophagous for a different set of host plants (Bush, 1966). Females lay their eggs directly into the fruit of the host plant which they identify by specific visual, tactile and olfactory cues (Bush 1969a,b). Males are attracted by the same cues, and courtship and mating occur almost exclusively on or near the fruits of the host plant (Prokopy *et al.*, 1971, 1972). Because host recognition and mate selection are directly coupled in *Rhagoletis*, variation for host preference traits can produce genetically based, pre-mating barriers to gene flow. This reduction in gene flow can result in the evolution of partially reproductively isolated host races, and eventually species, adapted to different host plants (Bush, 1966, 1969a,b, 1975).

R. pomonella group species and their host plants overlap broadly with one another in their distributions across eastern North America (Bush, 1966) and at no time during their evolutionary history was it likely for geographically isolated fly populations to have existed. This point is most evident in the recent shift of the species *R. pomonella* from its native host hawthorn (*Crataegus* spp.) to introduced domestic apple (*Malus pumila*). Agricultural records indicate a probable origin for the apple infesting fly in the Hudson Valley, New York and document the subsequent spread of the apple fly across the eastern United States and Canada over the last 120 years (Ward, 1866; Walsh, 1867; Bush, 1969a). Hawthorn and apple races of *R. pomonella* therefore offer a unique opportunity to study the initial stages of the divergence process. The historical timeframe of *R. pomonella*'s shift(s) to apples provides a direct test of whether partially reproductively isolated "host races" can form sympatrically in the absence of prior periods of geographic isolation.

The previous studies in this dissertation have shown that sympatric apple and hawthorn populations of *R. pomonella* are genetically differentiated. Significant gene frequency differences exist for six allozyme loci between sympatric apple and hawthorn populations from across the eastern United States (Chapters 2, 3, 4). These six loci map to three different regions of the genome (Chapter 1) and linkage disequilibrium occurs in natural populations between non-allelic genes within each of these three regions (Chapter 2, 3). The pattern of linkage disequilibrium is similar for apple and hawthorn flies suggesting that inversions are present in different frequencies in the two races (Chapters 2, 3). However, gene frequencies for five of the six loci displaying inter-host differentiation also show significant linear relationship with latitude (Chapters 3, 4). The slopes of these latitudinal allele frequency clines are different, however, between hawthorn and apple flies implying that differential selection occurs both within and between host races across their ranges (Chapters 3, 4).

Gene frequency clines among hawthorn and apple populations suggest that host races may not evolve uniformly across their respective ranges. Differential selection caused by vagaries in environmental conditions could dramatically shape the local pattern of intra- and inter-host differentiation. In addition, hawthorn and apple fruits are not necessarily constant resources in either time or space. Host trees can mast and it is not unusual for entire sites to be virtually devoid of fruit in certain years. Mark and recapture studies indicate that adult flies will migrate over distances as great as 2 km when host fruit is not readily available (Phipps and Dirks, 1933a, Bourne *et al.*, 1934; Maxwell, 1968). As a result, founder events and differential migration could drastically affect patterns of host associated polymorphism for *R. pomonella*. McPheron *et al.*, (1988b) have, in fact, found significant microgeographic allozyme variation among fly populations infesting different hawthorn trees within 210 meters of one another at a field site in Urbana, Illinois. It is, therefore, imperative to characterize local and microgeographic levels of differentiation among host populations as a baseline for assessing processes occurring on broader geographic scales.

The current study addresses the issue of local and microgeographic genetic differentiation by examining allozyme variation for apple and hawthorn populations within a 60 km^2 area near the town of Grant, in western Michigan (Newaygo Co.). The results indicate that genetic differentiation among apple and hawthorn populations are a microcosm of patterns observed across the eastern United States with inter-host divergence being a consequence of differences in north-south allele frequency clines between the two races. Genetic variation is, therefore, finely subdivided in *R. pomonella* according to host plant utilization and to local environmental conditions, a result supporting predictions of the niche variation and sympatric speciation hypotheses.

Materials and Methods

All flies analyzed in this study were collected as larvae from infested ground fruit. At three sites (A, E and H; see Figure 9a; Note: Site E is the original sympatric site analyzed in Chapter 2) in 1987, larvae were dissected from infested fruit and immediately frozen in liquid nitrogen for later electrophoretic analysis. In all other cases, infested fruit was transported back to the laboratory and placed in trays of moist vermiculite. Fruit from individual trees and from different sites were kept in separate trays and covered with fine mosquito netting to prevent larvae from moving between trays. All trays were stored in a greenhouse under constant temperature (21° C) and light (15:9 hr., light:dark cycle) conditions. After two months pupae were sifted from the vermiculite and

Figure 9. Collecting sites for *R. pomonella* in western Michigan. Panel b is a diagram of local populations sampled within a 60 km^2 area near the town of Grant, Michigan. o = apple tree, Δ = hawthorn tree. Panel c gives the distribution of apple and hawthorn trees at site E. A=apple H=hawthorn.

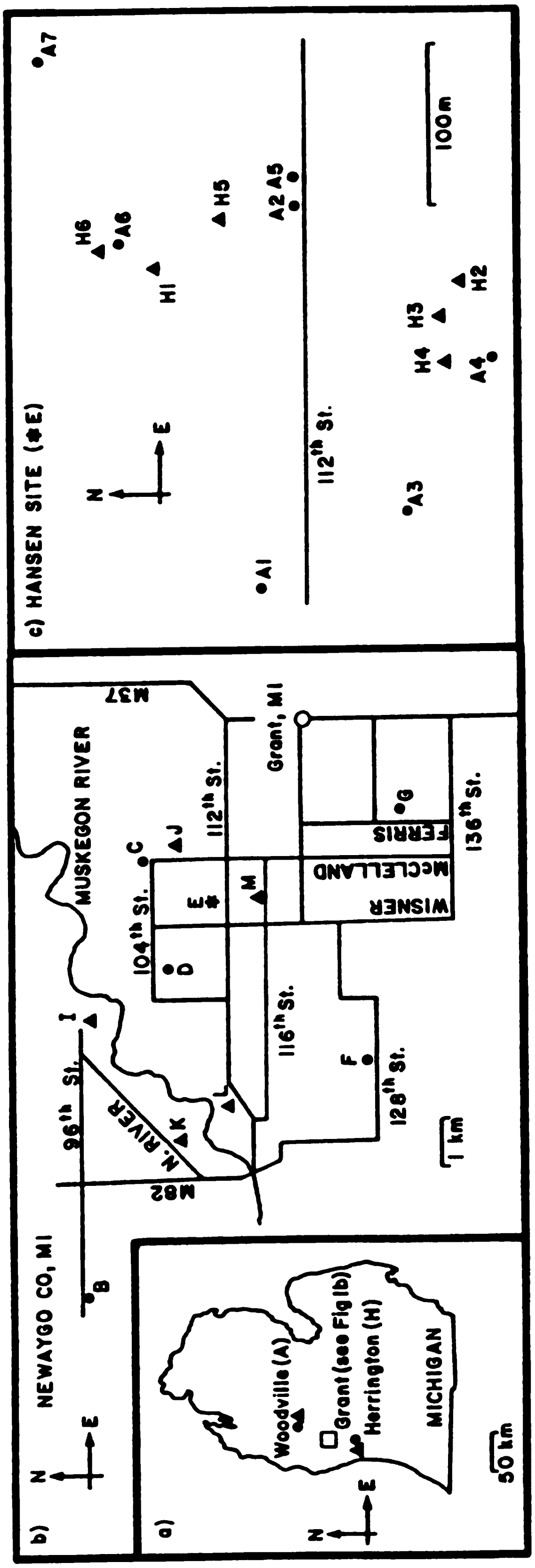

a)
N
E
Woodville (A)
Grant (see Fig 1b)
Herrington (H)
MICHIGAN
50 km
b)
N
E
NEWAYGO CO, MI
B
M82
96th St.
N. RIVER
K
I
MUSKEGON RIVER
104th St.
D
C
E
J
112th St.
M
116th St.
F
128th St.
1 km
M37
Grant, MI
WISNER
McCLELLAND
FERRIS
G
136th St.
c) HANSEN SITE (#E)
N
E
A1
A3
112th St.
A4
H4
H3
H2
H1
H6
A6
H5
A2 A5
A7
100m

Figure 9.

kept at 4° C for five months to simulate winter diapause conditions. Pupae were removed from the cold and held at 21° C, 15:9 L:D to terminate diapause. Adults were collected when they emerged after 1-3 months of heating and were frozen at -80° C until they could be analyzed electrophoretically.

Sympatric hawthorn and apple trees (sympatric being defined as host plants separated by a distance of less than 0.4 km) were individually sampled over a four year period at a centrally located site designated E (Figure 9b,c). Two apple trees and one hawthorn tree at site E were monitored across all four years of the study. Collections were made from other trees in years when enough fruit was available to ensure against population extinction. The year 1987 was exceptional with virtually every tree fruiting in large quantity at site E, except for hawthorn trees 2 and 3. Analysis of microgeographic variation was therefore most extensive in 1987.

Local geographic variation was assessed in 1986. Flies from five hawthorn and five apple trees were sampled from within a 60 km^2 area surrounding site E (Figure 9a,b). Local hawthorn and apple sites were not sympatric but were distributed as uniformly as possible in all directions around site E. In addition, sympatric hawthorn and apple populations were collected in 1987 from sites A and H located approximately 30 km north and south, respectively, of site E (Figure 9a).

Standard horizontal starch gel electrophoretic techniques were used and are described fully in Chapter 1 and in Berlocher and Bush (1982). We resolved and scored six different allozyme systems: *Malic enzyme* (*Me*), *Aconitase-2* (*Acon-2*), *Mannose phosphate isomerase* (*Mpi*), *NADH-diaphorase-2* (*Dia-2*), *Aspartate amino-transferase-2* (*Aat-2*) and

Hydroxyacid dehydrogenase (*Had*). These were the six allozymes which showed the highest levels of inter-host differentiation in the previous geographic surveys of genetic variation across eastern North America (Chapters 3, 4). Isozymes that migrated the nearest to the cathode were designated 1, the second nearest 2, etc. Alleles were numbered according to their relative anodal mobility to the common allele (100) at a locus. Tables 50-55, which follow the Discussion section on pages 183-188, list a majority of the electrophoretic alleles which were resolved in this study. Only two alleles were scored for the loci *Me* and *Dia-2* and so only the relative mobilities and frequencies for the rarer of the two alleles for these allozymes are given in Tables 50-55. Three alleles were scored for the locus *Had*, but one of these alleles (*Had*^{97}) was very rare and so only frequencies for the second most common allele (*Had*^{122}) are presented in Tables 52, 53 and 55. A Mendelian mode of inheritance has been established for all six of the loci analyzed in the study (Chapter 1) and electromorphs migrated identically in larval and adult samples.

Genotype frequencies were examined for deviations from Hardy-Weinberg expectation by G-tests. Alleles were pooled, if required, so that all genotypic classes had expected frequencies >1. The Levene correction (Spiess, 1977) was applied when sample sizes were < 100 or when it was not possible to pool alleles to make the expected frequency for all genotypic class >1. Allele frequencies were analyzed for significant heterogeneity between host plants and across time and space using G-contingency tests. Allele frequencies were pooled, when necessary, so that each cell in the G-contingency test had a frequency ≥ 5. F-statistics were calculated by the method of Weir and Cockerham (1984) with standard deviations estimated by jackknifing over populations, years or loci. Corrections for unequal sample sizes in the method of Weir and Cockerham can result in negative F_{ST} values. In these instances we report F_{ST} as 0. One of the F-statistics, f (the correlation of genes within individuals within populations $\simeq F_{IS}$ of Wright, 1951) was determined independently for each population from allele frequency data (see Barker *et al.*, 1986). Multi-allele f values were tested against the null hypothesis of f=0 by a Chi-square, single-degree of freedom test (Barker *et al.*, 1986).

Results

Results are presented sequentially for either one locus or two loci at a time. We first discuss the microgeographic pattern of differentiation for a gene(s) on a tree by tree basis at site E and then follow with a description of genetic variation among local *R. pomonella* populations in the surrounding area.

Acon-2

Significant allele frequency differences between apple and hawthorn flies were observed for *Acon-2* on both local and microgeographic scales. At site E, *Acon-2* displayed significant inter-host differentiation in each of the four years of the study but showed no significant intra-host variation within any year (Tables 44, 50, 51). Gene frequencies for *Acon-2* were heterogeneous for hawthorn flies across years at site E (Table 45). However, intra-host temporal variation neither overlapped nor was of similar magnitude as inter-host differentiation (Tables 50, 51). F-statistics quantify this point. The F_{ST} value for *Acon-2* calculated for both host populations over years at site E (0.0604 ± 0.0136, jackknife estimate of standard deviation calculated over years; Table 46) was 6 x greater than the

Table 44. G-contingency tests for intra- and inter-host allele frequency heterogeneity at site E.

Year	Test	Me	Acon-2	Mpi	Dia-2	Aat-2	Had
1984	Among Apple Trees	2.4(1)	6.3(3)	1.1(2)	2.2(4)	4.1(1)*	8.1(1)**
	Between Apple-Haw	85.7(1)***	59.3(3)***	32.4(2)***	31.1(4)***	--	--
	Apple Tree 1-Haws	--	--	--	--	26.4(1)***	4.0(1)*
	Apple Tree 2-Haws	--	--	--	--	9.1(1)**	0.9(1)
1985	Among Apple Trees	3.1(2)	3.7(6)	1.0(4)	16.5(8)*	6.6(2)*	4.9(2)
	Among Haw Trees	0.8(2)	10.6(6)	11.8(4)*	10.9(8)	3.3(2)	0.1(2)
	Between Apple-Haw	60.7(1)***	44.1(3)***	--	--	--	11.0(1)***
	Apples-Haw Tree 1	--	--	7.1(2)*	--	--	--
	Apples-Haw Tree 2	--	--	15.5(2)***	--	--	--
	Apples-Haw Tree 3	--	--	18.5(2)***	--	--	--
	Apple Tree 1-Haws	--	--	--	5.4(4)	3.5(1)	--
	Apple Tree 2-Haws	--	--	--	21.2(4)***	1.1(1)	--
	Apple Tree 3-Haws	--	--	--	16.9(4)**	16.4(1)***	--
1986	Among Apple Trees	8.4(1)**	0.4(3)	4.1(1)*	1.1(1)	0.6(3)	3.0(1)
	Among Haw Trees	0.6(1)	0.6(3)	0.1(1)	0.1(1)	4.9(3)	0.0(1)
	Between Apple-Haw	--	80.0(3)***	--	26.8(1)***	21.8(3)***	0.9(1)
	Apple Tree 1-Haws	46.4(1)***	--	29.6(1)***	--	--	--
	Apple Tree 2-Haws	8.7(1)**	--	7.3(1)**	--	--	--

Table 44. (cont'd.)

Year	Test	Me	Acon-2	Mpi	Dia-2	Aat-2	Had
1987	Among Apple Trees	33.8(6)***	24.0(18)	10.7(6)	27.8(6)***	27.5(18)	23.0(6)***
	Among Haw Trees	3.7(3)	6.1(9)	3.0(3)	5.3(3)	10.1(9)	3.6(3)
	Between Apple-Haw	--	209.1(3)***	24.2(1)***	--	35.0(3)***	--
	Apple Tree 1-Haws	39.6(1)***	--	--	16.5(1)***	--	10.0(1)**
	Apple Tree 2-Haws	43.1(1)***	--	--	1.2(1)	--	0.4(1)
	Apple Tree 3-Haws	44.4(1)***	--	--	7.5(1)**	--	3.3(1)
	Apple Tree 4-Haws	95.3(1)***	--	--	0.0(1)	--	1.9(1)
	Apple Tree 5-Haws	33.1(1)***	--	--	0.5(1)	--	2.7(1)
	Apple Tree 6-Haws	48.7(1)***	--	--	0.1(1)	--	0.2(1)
	Apple Tree 7-Haws	163.1(1)***	--	--	0.0(1)	--	2.4(1)

---

* $\underline{P} \leq 0.05$ ** $\underline{P} \leq 0.01$ *** $\underline{P} \leq 0.001$ G-contingency test . Degrees of freedom for tests are given in parentheses. Allele frequencies for non-significant intra-host tests were pooled across host trees and these pooled totals were used for inter-host tests (i.e., between apple-haw). When significant intra-host heterogeneity was observed for a locus, inter-host tests were conducted on a tree by tree basis against the pooled total for the other host (i.e. apples-haw tree 1).

Table 45. G-contingency tests for intra-host allele frequency heterogeneity across years at site E and among local apple (A) and hawthorn (H) populations. Degrees of freedom for tests are given in parentheses.

Test	Host	Me	Acon-2	Mpi	Dia-2	Aat-2	Had
Across years-site E	A	6.0(3)	10.8(6)	5.1(6)	7.3(3)	33.3(12)***	7.5(3)
	H	39.1(3)***	28.6(6)***	21.1(6)***	13.4(3)***	19.1(12)	8.6(3)*
Among local sites	A	16.4(5)**	25.1(10)**	8.0(5)	9.4(5)	17.9(15)	3.7(5)
	H	4.4(5)	23.0(10)*	7.5(5)	5.0(5)	30.5 (15)*	11.3(5)*

* $P \leq 0.05$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-contingency test

Table 46. F_{ST} values calculated by the method of Weir and Cockerham (1984) across years for apple and hawthorn trees at site E.

Locus	Apple Tree 1	Apple Tree 2	All Apples	Hawthorn Tree 1	All Hawthorns	Both Hosts[c]
Me	0.0311 ± 0.0143[a]	0.0011 ± 0.0017	0.0019 ± 0.0038	0.0197 ± 0.0177	0.0259 ± 0.0270	0.0960 ± 0.0235
Acon-2	0.0011 ± 0.0017	0.0013 ± 0.0044	0.0015 ± 0.0025	0.0043 ± 0.0046	0.0099 ± 0.0095	0.0604 ± 0.0136
Mpi	0.0000 ± 0.0027	0.0000 ± 0.0006	0.0000 ± 0.0025	0.0127 ± 0.0147	0.0133 ± 0.0082	0.0322 ± 0.0125
Dia-2	0.0060 ± 0.0096	0.0086 ± 0.0128	0.0025 ± 0.0036	0.0086 ± 0.0078	0.0070 ± 0.0068	0.0247 ± 0.0074
Aat-2	0.0057 ± 0.0032	0.0014 ± 0.0021	0.0018 ± 0.0020	0.0015 ± 0.0015	0.0014 ± 0.0020	0.0080 ± 0.0020
Had	0.0000 ± 0.0038	0.0067 ± 0.0120	0.0022 ± 0.0042	0.0021 ± 0.0020	0.0032 ± 0.0034	0.0046 ± 0.0026
All loci	0.0064 ± 0.0036[b]	0.0018 ± 0.0010	0.0017 ± 0.0002	0.0060 ± 0.0028	0.0084 ± 0.0040	0.0377 ± 0.0188

---

[a] Standard deviation - jackknife estimate over years.

[b] Standard deviation - jackknife estimate over loci.

[c] Both hosts are the F_{ST} values when both apple and hawthorn trees are included in the calculations. Both host values are therefore representative of the interaction between host associated and temporal genetic variation.

value for hawthorn flies alone (0.0099 ± 0.0095) and over 50 x greater than the F_{ST} for apple flies (0.0015 ± 0.0025).

The allele frequency differences found between apple and hawthorn flies at site E suggest that differential selection may be occurring between the host populations. However, if selection occurs for *Acon-2*, its effects were not reflected in genotypic deviations from Hardy-Weinberg expectation within host populations. Only 3 out of a total of 96 tests showed significant deviations for *Acon-2* when individual host trees were analyzed within years (Tables 50, 51). In addition, individual tree f values at site E showed no trend for either an excess or deficiency of heterozygotes (Table 47).

Inter-host differentiation for *Acon-2* was also consistent across the 12 local apple and hawthorn populations analyzed in 1986. The highest frequency for the allele $Acon\text{-}2^{95}$ in any apple population was 0.268 at site D (Table 54). In comparison, the lowest frequency for $Acon\text{-}2^{95}$ for a hawthorn population was 0.392 at site E (0.392 represents the pooled $Acon\text{-}2^{95}$ frequency among hawthorn trees at site E in 1986 and also happens to be the lowest frequency observed for hawthorn flies at site E for all four years of the study, the next lowest frequency being 0.479 in 1987; Table 54). Nevertheless, allele frequencies for *Acon-2* were still significantly different in 1986 between apple flies at site D and hawthorn flies at site E (G-contingency test = 6.3, $\underline{P} \leq 0.05$, 2 df).

$Acon\text{-}2^{95}$ allele frequencies among local hawthorn populations fit within a broader geographic pattern of latitudinal clines observed across the Midwest (Chapter 4). Local hawthorn populations I through M had $Acon\text{-}2^{95}$ frequencies that were bracketed within the range of values found for hawthorn flies at site A (0.619) and site H (0.392) located

Table 47. Distribution of f values ($\simeq F_{IS}$ of Wright, 1951) for individual populations calculated by the method of Barker, East and Weir (1986). The number of f values significantly greater or less than 0 are given in parentheses.

	Site E		Local Populations		All Sites	
Locus	$f < 0$	$f > 0$	$f < 0$	$f > 0$	$f < 0$	$f > 0$
Me	16	8	8(3)	4	24(3)	12
Acon-2	10(1)	14(1)	8	4	18(1)	18(1)
Mpi	19(2)	5	9(1)	3	28(3)	8
Dia-2	12(1)	12	9	3	21(1)	15
Aat-2	9(1)	15(2)	3	9(1)	12(1)	24(3)
Had	10	14(2)	8	4(1)	18	18(3)
All loci	76(5)	68(5)	45(4)	27(2)	120(9)	96(7)

approximately 30 km north and south, respectively, of site E (Table 54, Figure 9). However, *Acon-2*^{95} allele frequencies in 1986 still fluctuated between 0.571 at site L and 0.392 at site E. Latitudinal clines for *Acon-2*^{95} were, therefore, not perfect across local hawthorn populations.

Acon-2 also showed significant intra-host variation among local apple and hawthorn populations (Table 45). For example, the frequency of *Acon-2*^{75} was 0.250 in hawthorn population K compared to 0.038 at hawthorn site J (Table 54). The same was also true for local apple populations where *Acon-2*^{75} varied in frequency from 0.205 to 0.026 (Table 54). Despite the significant amount of local variation for *Acon-2*, only *Acon-2*^{95} at site J showed a significant deviation from Hardy-Weinberg equilibrium for this locus out of a total of 88 tests (Table 54). Also, local intra-host heterogeneity for *Acon-2* did not overshadow inter-host differentiation. The F_{ST} values among local apple (0.0103 ± 0.0074; Table 48) and hawthorn populations (0.0127 ± 0.0098) were approximately 4 x lower than the value across all host populations (0.0472 ± 0.0184).

Me and *Mpi*

Me and *Mpi* also displayed consistent microgeographic differentiation between apple and hawthorn flies at site E (Table 44). However, unlike *Acon-2*, significant intra-host variation was observed for *Me* and *Mpi* within years. *Me* varied significantly among apple trees at site E in both 1986 and 1987 (Table 44). However, inter-host allele frequency differences for *Me*^{80} were never less than 0.100 and were usually greater than 0.200 between any chosen pair of hawthorn and apple trees within a given year (Tables 50, 51). F_{ST} values for *Me* of 0.1233 ± 0.0313 across

Table 48. F_{ST} values for local apple and hawthorn populations calculated by the method of Weir and Cockerham (1984).

Locus	Apple	Hawthorn	Both Hosts[c]
Me	0.0185 ± 0.0159[a]	0.0000 ± 0.0043	0.0406 ± 0.0141
Acon-2	0.0103 ± 0.0074	0.0127 ± 0.0098	0.0472 ± 0.0184
Mpi	0.0093 ± 0.0086	0.0000 ± 0.0018	0.0214 ± 0.0169
Dia-2	0.0084 ± 0.0067	0.0202 ± 0.0180	0.0384 ± 0.0177
Aat-2	0.0018 ± 0.0060	0.0801 ± 0.0598	0.0427 ± 0.0354
Had	0.0000 ± 0.0036	0.0092 ± 0.0147	0.0038 ± 0.0049
All loci	0.0071 ± 0.0032[b]	0.0321 ± 0.0230	0.0395 ± 0.0046

---

[a] Standard deviation- jackknife estimate over sites.

[b] Standard deviation- jackknife estimate over loci.

[c] Both hosts are F_{ST} values when both local apple and hawthorn populations are included in the calculations. Both host values therefore represent the interaction between host associated and spatial genetic variation.

Table 49. F_{ST} values for hawthorn and apple trees at site E in 1985 and 1987 calculated by the method of Weir and Cockerham (1984).

	1985			1987		
Locus	Apples[c]	Hawthorns[c]	Both Hosts[d]	Apples[c]	Hawthorns[c]	Both Hosts[d]
Me	0.0092 ± 0.0240[a]	0.0000 ± 0.0039	0.1233 ± 0.0313	0.0075 ± 0.0066	0.0007 ± 0.0058	0.0844 ± 0.0269
Acon-2	0.0000 ± 0.0036	0.0087 ± 0.0167	0.0748 ± 0.0229	0.0018 ± 0.0024	0.0000 ± 0.0028	0.0543 ± 0.0122
Mpi	0.0000 ± 0.0031	0.0064 ± 0.0146	0.0390 ± 0.0171	0.0043 ± 0.0039	0.0030 ± 0.0054	0.0145 ± 0.0089
Dia-2	0.0252 ± 0.0335	0.0059 ± 0.0139	0.0325 ± 0.0289	0.0203 ± 0.0128	0.0041 ± 0.0078	0.0275 ± 0.0167
Aat-2	0.0096 ± 0.0169	0.0000 ± 0.0058	0.0124 ± 0.0077	0.0029 ± 0.0043	0.0028 ± 0.0064	0.0091 ± 0.0054
Had	0.0159 ± 0.0259	0.0000 ± 0.0018	0.0234 ± 0.0177	0.0200 ± 0.0129	0.0000 ± 0.0041	0.0169 ± 0.0107
All loci	0.0062 ± 0.0051[b]	0.0015 ± 0.0043	0.0500 ± 0.0225	0.0064 ± 0.0031	0.0013 ± 0.0013	0.0350 ± 0.0159

[a] Standard deviation - jackknife estimate over years.

[b] Standard deviation - jackknife estimate over loci.

[c] Apple and hawthorn F_{ST} values quantify the level of intra-host microgeographic genetic variation within years.

[d] Both hosts provide relative estimates of inter-host genetic differentiation at site E within years.

all host trees at site E in 1985 (Table 49) and 0.0844 ± 0.0269 in 1987 document the substantial amount of inter-host differentiation for this locus.

Mpi displayed significant heterogeneity among apple trees at site E in 1986 and among hawthorn trees in 1985 (Table 44). However, unlike *Acon-2* and *Me*, apple and hawthorn populations overlapped slightly in their gene frequencies for *Mpi* in 1987. Hawthorn tree #4 had a frequency for Mpi^{100} of 0.860 in 1987 while apple tree #4 had a frequency of 0.858 (Tables 50, 51). Within no other year was the frequency of Mpi^{100} greater for a hawthorn tree than for an apple tree. In fact, the pooled Mpi^{100} gene frequency for hawthorn trees at site E was still significantly different from that for pooled apple trees even in 1987 (G-test = 24.2, $\underline{P} \leq 0.001$, 1 df; Table 44). However, the magnitude of inter-host variation was much less for *Mpi* than for either *Acon-2* or *Me* as evidenced by the lower F_{ST} values calculated for *Mpi* across both hosts in 1985 and 1987 (0.0390 ± 0.0171 and 0.0145 ± 0.0089, respectively; Table 49).

Gene frequencies for *Me* and *Mpi* at site E varied significantly across years for hawthorn flies but not for apple flies (Table 45). The result was indicative of a general trend in the partitioning of intra-host differentiation at site E. Hawthorn flies showed greater temporal heterogeneity than apple flies. The overall F_{ST} value across years for all six loci was almost 5 x higher for hawthorn populations (0.0084 ± 0.0040; Table 46) than apple populations (0.0017 ± 0.0002). In contrast, apple flies displayed higher levels of spatial differentiation within years at site E. This was exemplified by the 5 x larger F_{ST} values for all loci among apple trees in 1985 and 1987 (0.0062 ± 0.0051 and 0.0064 ± 0.0031, respectively; Table 49) compared to hawthorn trees (0.0015 ± 0.0043 and 0.0013 ± 0.0013).

Local differentiation for *Me* and *Mpi* was similar to the pattern of microgeographic variation observed for these loci at site E. Hawthorn populations I through M showed limited spatial heterogeneity for *Mpi* and *Me* in 1986 ($F_{ST} = 0.0000$ for both loci across hawthorn sites; Table 48). In comparison, apple populations displayed higher levels of local spatial diversity (F_{ST} *Me* = 0.0185 ± 0.0159, F_{ST} *Mpi* = 0.0093 ± 0.0086). In fact, significant allele frequency heterogeneity existed among local apple populations for *Me* (G-contingency test = 16.4, $\underline{P} \leq 0.01$, 5 df; Table 45).

Inter-host differentiation among local host populations, while not as great as that observed for *Acon-2*, was also apparent for both *Me* and *Mpi*. F_{ST} values for *Me* and *Mpi* across all 12 local host populations were over 2 x greater than the values calculated for apple populations alone (Table 48). *Mpi* allele frequencies were homogeneous among both local apple and hawthorn populations (Table 45), while pooled hawthorn and apple *Mpi* frequencies were significantly different from each other (G = 33.4, $\underline{P} \leq 0.001$, 1 df). Local apple and hawthorn populations overlapped in Mpi^{100} allele frequencies, however. Apple sites B and C both had Mpi^{100} frequencies below that of the highest hawthorn population at site I (Table 54). Mpi^{100} frequencies for local sites therefore formed a continuum, with apple populations tending to have higher frequencies than hawthorn populations. A similar pattern was also observed for Me^{80}. However, unlike *Mpi*, local differentiation for *Me* can be partly explained by gene frequency clines. All apple populations north of site E had Me^{80} allele frequencies ≤ 0.560 while all sites south of E had Me^{80} frequencies ≥ 0.630 (Table 54). Apple site E fluctuated in frequency between 0.578 and 0.671 from 1984 to 1987. Local hawthorn sites displayed similar clinal tendencies with northern populations generally having lower Me^{80}

frequencies than southern populations (Table 54). Inter-host differences for *Me* were therefore superimposed on clinal patterns of variation, with apple flies having higher Me^{80} allele frequencies than hawthorn flies at every site where sympatric populations were sampled (sites A, E and H; Table 54).

Genotypic proportions for *Me* and *Mpi* showed general trends for an excess of heterozygotes at both the local and microgeographic levels. Twenty four and 28 out of a total of 36 f values were negative (indicating an excess of heterozygotes) for *Me* and *Mpi* , respectively (Table 47). Three f values out of 36 were significantly less than 0 for both *Me* and *Mpi.*

Aat-2 and *Dia-2*

Aat-2 and *Dia-2* had roughly equivalent patterns of genetic differentiation as *Mpi.* Both loci showed significant microgeographic variation among apple trees at site E within 2 of the 4 years surveyed (Table 44). For the two years when intra-host gene frequencies were homogeneous for *Dia-2* or *Aat-2*, significant inter-host differences were observed (Table 44). However, frequencies for $Dia\text{-}2^{85}$ and $Aat\text{-}2^{100}$ overlapped between apple and hawthorn trees in both 1985 and 1987 (Tables 52, 53). A spectrum of gene frequencies was therefore also seen for *Dia-2* and *Aat-2* at site E. Hawthorn trees tended to reside on one end of the frequency distribution and apple trees on the other with certain trees (apple trees # 2, 7 and hawthorn tree # 1, 5) overlapping in the middle of the spectrum (Tables 52, 53). *Aat-2* was unusual in that it was the only locus which varied significantly across years for apple flies at site E (Table 45). *Dia-2* displayed the normal temporal pattern showing heterogeneity across years for hawthorn but not for apple populations (Table 45).

Dia-2 and *Aat-2* had the highest F_{ST} among the six local hawthorn populations (0.0202 ± 0.0180 and 0.0801 ± 0.0598 respectively, Table 48). Intra-host variation for *Aat-2* among local hawthorn populations was actually greater than inter-host levels observed for *Me* and *Acon-2* (Table 48). For example, $Aat\text{-}2^{50}$ varied in frequency from 0.256 to 0.083 between hawthorn sites I and M while $Aat\text{-}2^{21}$ changed from 0.0191 to 0.050 between sites L and M (Table 55).

Inter-host differentiation still existed for both *Aat-2* and *Dia-2*, however, as $Aat\text{-}2^{100}$ and $Dia\text{-}2^{100}$ gene frequencies were generally higher in local hawthorn than apple populations (Table 55). In fact, pooled *Dia-2* allele frequencies across hawthorn and apple sites were significantly different from one another in 1986 (G-contingency test = 23.4, $\underline{P} \leq 0.001$, 1 df). However, just like Mpi^{100}, inter-host frequency differences for $Aat\text{-}2^{100}$ and $Dia\text{-}2^{100}$ were not discreet and overlap was found among local hawthorn and apple populations. $Dia\text{-}2^{100}$ did not display a local pattern of variation within either host race that was indicative of a latitudinal frequency cline. $Aat\text{-}2^{75}$ did show a hint of a frequency cline among local apple populations (Table 55), yet the allele did not co-vary significantly with latitude in a broader geographic survey of apple populations from Michigan through Indiana (Chapter 4).

Aat-2 was unusual in that it was the only locus which displayed a trend for an excess of homozygotes (positive f values; Table 47). Four f values for *Aat-2* were significantly different than 0 out of a total of 36 tests (Table 47). Genotype frequencies for *Aat-2* therefore deviated from equilibrium proportions more often than expected by chance.

Had

Had showed intra-host variation at both the local and microgeographic levels but only the slightest trace of inter-host differentiation. *Had* varied significantly among apple trees at site E in 1984 and 1987 as well as between hosts in 1985 (Table 44). Allele frequencies for the locus were heterogeneous for hawthorn flies across years at site E as well as among local hawthorn populations in 1986 (Table 45). However, *Had* displayed lower levels of inter-host differentiation between local hawthorn and apple populations than any of the other five loci analyzed in the study (F_{ST} *Had* = 0.0038 ± 0.0046; Table 48). In addition, no geographic pattern of variation was discernable for *Had* allele frequencies. *Had* f values were equally divided between positive and negative estimates (Table 47).

Discussion

The results of this and the preceding studies confirm that partially reproductively isolated host races of phytophagous insects can evolve in sympatry by shifting and adapting to new host plants. *R. pomonella* populations are genetically finely subdivided according to the species of host plant they infest; a result which agrees with predictions of the niche variation hypothesis. Even for different species of host trees within 100 meters of one another at site E, Me^{100} and $Acon\text{-}2^{95}$ allele frequencies are significantly higher for hawthorn than apple flies. Inter-host differentiation is also stable over time. Although allele frequencies for Me^{100} and $Acon\text{-}2^{95}$ did vary significantly for hawthorn flies at site E over a four year period, intra-host variation did not compare in magnitude or overlap in frequency with inter-host differences. *Me* and *Acon-2* are in fact, closely linked on chromosome II in *R. pomonella* (Chapter 1) and disequilibrium has been found between non-allelic genes for these two loci in natural fly populations (Chapters 2, 3). The similar pattern of microgeographic variation for *Me* and *Acon-2* is therefore not surprising and suggests that selection in this region of the genome is causing genetic hitchhiking effects among linked loci.

Genetic studies of *R. pomonella* suggest that differences in host recognition and host associated survivorship are two important factors differentiating fly populations infesting hawthorns and apples. Mark-recapture studies and larval transplant experiments would help determine the extent to which these factors reduce gene flow between the two host races. The fine level of inter-host divergence revealed in this Chapter indicates that reproductive isolating mechanisms can be reliably examined even for local *R. pomonella* populations. Evidence already exists indicating that adult flies do not randomly migrate between different species of host plants. Adults eclosing from under and collected directly from off of hawthorn and apple trees at site E showed the same pattern and magnitude of inter-host divergence as larvae sampled from ground fruit (Chapters 2, 4). *R. pomonella* therefore tend to utilize the same species of host plant as adults which they infested as larvae. We do not know whether inherent, genetically based differences in host recognition and/or conditioning in adults (Prokopy *et al.*, 1982) cause this bias in host choice but mark-recapture and larval transplant experiments could be designed to specifically address this issue.

The association between mate and habitat selection for *R. pomonella* host races has important implications for the modeling of non-allopatric speciation. The theoretical work of Felsenstein (1981) has often been cited to support the view that sympatric speciation is a rare phenomenon in nature (Futuyma and Mayer, 1980; Butlin, 1987). However, in Felsenstein's model individuals randomly select habitats, an assumption which is not valid for *R. pomonella* group flies and potentially many other phytophagous and parasitic insects (Price, 1980; Diehl and Bush, 1984). More recent theoretical studies (Bush and Diehl, 1982; Rice, 1984; Diehl and Bush, 1989) indicate that sympatric populations can differentiate rapidly when mating is confined to specific habitats chosen by the animals. These models suggest that non-allopatric divergence is as much, or more dependent, on the fidelity of habitat related assortative mating than on the sheer intensity of differential selection within habitats. This consideration has not been adequately appreciated in our understanding of how heterogeneous environments maintain genetic diversity and promote speciation.

An interesting implication of my work is that host races may not evolve uniformly across their ranges. Gene frequency clines, although not perfect, were observed for Me^{100} and $Acon\text{-}2^{95}$ even across the limited geographic area surveyed in this study. Inter-host differentiation for *Me* and *Acon-2* is, therefore, superimposed on north-south clinal patterns of intra-host variation. Besides *Me* and *Acon-2*, however, none of the other four loci analyzed, with the possible exception of *Aat-2* for apple flies, showed a distinct pattern of clinal variation among local populations. However, *Mpi*, *Dia-2* and *Had* do display latitudinal allele frequency clines when analyzed on a broader geographic scale across the midwestern United States (Chapter 4). For a number of reasons discussed in Chapters 3 and 4, secondary contact is an unlikely explanation for the observed frequency clines. Instead, environmental factors affecting *R. pomonella* appear to be experienced in a graded fashion from north to south, with the two races being differentially affected at a given latitude. Environmental gradients are rarely exact, however, and vagaries in local environmental conditions can obscure broader geographic patterns of genetic variation. This could account for why allele frequencies for *Mpi*, *Dia-2* and *Had* were not correlated with latitude among local populations despite showing clinal patterns of variation across the Midwest.

Differences in micro-environmental conditions could also explain the pattern of intra-host variation at site E. Although apple flies at site E had higher levels of intra-host variation within years than hawthorn flies, hawthorn populations were more variable across years. This suggests that differences exist in the environmental "buffering capacities" or "selective constraints" imposed by the two species of host plants, with flies attacking hawthorns being more acutely affected by environmental fluctuations than flies infesting apples. Buffering capacities may vary more among individual apple trees than hawthorn trees, however, accounting for the higher levels of intra-host variation for apples within years. Indirect support for this hypothesis comes from the fact that apple tree #1 at site E, the most variable apple tree across years, also deviated the most in allele frequencies from the other apple trees within years (Tables 50-53). In addition, egg to pupal survivorship is known to fluctuate widely among different varieties of apples (Phipps and Dirks, 1933b; Reissig, 1979), with flies surviving better on earlier, sweeter varieties. Reciprocal egg transplant experiments, among different apple and hawthorn trees at site E, are needed to directly test the environmental buffering hypothesis.

The discussion has thus far concentrated on the importance of deterministic factors in influencing the genetic structure of fly populations. However, stochastic processes undoubtedly interact with selection to shape local and microgeographic patterns of genetic differentiation. *Aat-2* represents the best example of this. High levels of intra-host variation exist for *Aat-2*^{21}, *Aat-2*^{50} and *Aat-2*^{75} among local hawthorn populations. The pattern of differentiation for these alleles appears to be random among sites which suggests the action of genetic drift.

Aat-2 gene frequencies are not totally dictated by chance, however. Allele frequencies for *Aat-2*^{100} are generally higher in hawthorn than apple populations (Tables 52, 53, 55). In addition, genotypic frequencies for *Aat-2* reveal a significant trend for an excess of homozygotes (Table 47) and deviate from Hardy-Weinberg equilibrium more often than would be expected by random chance (Table 47). The excess of homozygotes for *Aat-2* could be due to selection, errors in enzyme scoring caused by a silent allele(s), positive assortative mating within host populations or a Wahlund effect resulting from population subdivision on a fine scale. The latter possibility seems unlikely because none of the other five loci analyzed had a surplus of homozygotes (Table 47). Population sub-structuring should affect neutral genes in an equivalent manner. Consequently, the lack of a significant excess of homozygotes at loci besides *Aat-2* argues against a Wahlund effect.

Positive assortative mating within apple and hawthorn populations based on *Aat-2* genotypes is also unlikely. *Dia-2* is closely linked to *Aat-2* (Chapter 1) and non-allelic genes for these two loci show substantial amounts of linkage disequilibrium in natural populations (Chapters 2, 3). If non-random mating were occurring relative to *Aat-2*, its effects should also be apparent for *Dia-2*. However, *Dia-2* had an excess of heterozygotes thereby making assortative mating an improbable scenario.

Little information is currently available on the frequencies of null alleles in *Rhagoletis* populations. Null alleles appear to be relatively rare in natural populations, however (Langley *et al.*, 1981). In addition, no individual was observed to be homozygous for a null allele out of a total of 2,729 flies scored in the study. Although we cannot completely rule out the possibility of null alleles, selection at or near the *Aat-2* locus is the most probable explanation for the excess of homozygotes.

The loci analyzed in this study therefore appear to represent a continuum in terms of the inter-action of evolutionary processes. *Me* and *Acon-2* form one side of the spectrum with allele frequencies for these two loci primarily determined by selection. In contrast, genetic drift may play a larger role in influencing local patterns of variation for *Aat-2* and *Had*. There is no reason to assume, however, that selection acts directly on *Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* or *Had*. The map distances of these six loci to genes actually under differential selection must, therefore, also be considered when interpreting the relative importance of stochastic versus deterministic processes.

To quantify the interplay between natural selection and genetic drift requires detailed analyses of the distribution and movement of adult flies. First, estimates of effective population sizes are needed for local hawthorn and apple populations. Second, we must determine whether adults confine their reproductive activities to a single tree (perhaps the same tree they infested as larvae) or several different trees at a site. Finally, neighborhood sizes must be established for *R. pomonella* which require estimates of migration rates among local populations. Previous work indicates that *R.*

pomonella adults can travel over distances of up to 2 km (Phipps and Dirks, 1933b; Bourne *et al.*, 1934; Maxwell, 1968) but that most flies disperse less than 60 m from release plants (Neilson, 1971; Buriff, 1973; Reissig, 1977). Maxwell (1968) and Reissig (1977) both report the proportion of apple flies remaining on a release tree as approximately 10%. This implies that adult flies distribute their reproductive capacities among a number of different apple or hawthorn trees. If true, then either differential selection on larvae or differential fecundity among adults must be responsible for the observed microgeographic genetic variation among apple trees at site E.

In addition to being spatially patchy, host resources are also temporally variable. Partial or complete fruit failure can dramatically affect population structure. For example, in mast years entire local sites can be virtually devoid of host fruit. On a microgeographic scale, hawthorn trees #2 and #3 at site E usually had large fruit loads. However, these two trees bore no fruit in 1987. Adult flies emerging from under hawthorn trees #2 and #3 in 1987 were, therefore, forced to disperse in search of suitable host fruit. Such forced migrations would greatly increase mixing within and among local fly populations and could account for the general trend for an excess of heterozygotes for loci in this study (Table 47). In addition, adult flies may not randomly recolonize hawthorn trees #2 and #3 in subsequent years. Founder events could therefore play an important role in affecting the pattern of genetic differentiation at site E.

The diapause characteristics of *R. pomonella* may ameliorate some of the effects of mast years, however. In both the field and the laboratory, a certain proportion of pupae (ca. 10%) require exposure to more than one cycle of chilling and heating before eclosing as adults (Phipps and Dirks, 1933b). Adults emerging after two winters would tend to homogenize gene frequencies over years for trees experiencing fruit failure provided, of course, that the loci analyzed are not involved either directly or indirectly with diapause termination. This later assumption may not be true, however, as preliminary studies indicate a relationship between allele frequencies for *Me*, *Acon-2* and *Had* and the timing of adult eclosion for the hawthorn race (Feder *et al.*, manuscript in preparation). Continued monitoring of hawthorn trees #2 and #3 would help tease apart how diapause and founder effects influence gene frequency distributions within and among fly populations.

The results of this study indicate that local and microgeographic genetic variation for hawthorn and apple flies are a microcosm of the patterns observed across eastern North America. Detailed examination of local demes will therefore provide useful information concerning the processes involved in host specialization and population divergence. Further studies are needed to discern exactly how selection, drift and migration interact to produce the fine level of genetic subdivision observed for *R. pomonella*. Perhaps foremost, we must identify the nature and mode of action of selective factors differentially affecting apple and hawthorn populations.

Previous studies indicate that ambient temperature has an important influence on the phenology of host-parasite interactions for *R. pomonella* (Prokopy, 1968; Reissig *et al.*, 1979; also see Chapter 4). Temperature is a key factor in the initiation, maintenance and termination of diapause in these flies (Prokopy, 1968; Reissig *et al.*, 1979) and, of course, has a direct effect on the phenology of fruit maturation for host plants. Apple and hawthorn flies do, in fact, differ in the timing of adult eclosion with apple flies emerging consistently earlier than hawthorn flies in both the laboratory and the field (Diehl, 1984; Smith, 1986, 1988; Feder *et al.*, manuscript in preparation). Temperature also correlates with latitude, which could account for the observed north-south frequency clines in *R. pomonella*. The temperature/development hypothesis therefore requires serious consideration and further empirical testing.

Positive correlations between genetic diversity and environmental variation have been documented for populations both in the laboratory (Powell, 1971; McDonald and Ayala, 1974; Powell and Wistrand, 1978) and in the field (Clarke, 1968; Berger, 1971; Dobzhansky and Ayala, 1973; Kettlewell, 1973; Taylor and Powell, 1976; Saul *et al.*, 1978; Mully *et al.*, 1979; Steiner,1979; Franklin, 1981; Barker *et al.*, 1986; Lynch, 1987). The finding of significant allozyme variation between host associated races of *R. pomonella* indicates that in addition to environmental heterogeneity in time and space, biotic considerations related to host-parasite interactions can also promote and maintain genetic polymorphism in populations. Furthermore, because *R. pomonella* is an economically important pest species, much is known about the fly's ecology, behavior and natural history (see review by Boller and Prokopy, 1976). This information, combined with the recent formation of several new host races in the species (Herrick, 1920; Bush, 1966; Shervis *et al.*, 1970; Prokopy and Bush, 1972; Prokopy and Berlocher, 1980) makes *R. pomonella* a unique organismal system for understanding evolutionary processes within a real time framework. In addition, many of the sibling species in the *R. pomonella* group, although distinct in nature, can be hybridized in the laboratory to produce viable and fertile F1 progeny. Experiments can therefore be performed to elucidate the genetic basis of traits responsible for restricting gene flow among *R. pomonella* flies. Comparisons can then be made to determine whether factors affecting gene frequencies within *R. pomonella* host races are directly involved in the development of reproductive isolation between *R. pomonella* group species.

Table 50. Allele frequencies for *Me*, *Acon-2* and *Mpi* for individual apple (A) trees collected at site E from 1984 to 1987. See Figure 9c for a diagram of site E which includes apple tree designations. N = sample size.

		Me		**Acon-2**								**Mpi**						
Yr	Tree	N	80	N	73	75	89	95	100	106	114	N	16	31	37	70	100	125
84	A1	99	0.707	99		0.096	0.005	0.152	0.545	0.192	0.010	99		0.005	0.056	0.035	0.899	0.005
	A2	84	0.631	86		0.070	0.058	0.221	0.512	0.122	0.017	86		0.006	0.046	0.058	0.884	0.006
85	A1	25	0.480	25		0.160	0.020	0.240	0.480	0.100		25			0.060	0.040	0.900	
	A2	52	0.587	62		0.113	0.097	0.185	0.484**	0.113	0.008	65			0.085	0.038	0.869	0.008
	A3	51	0.618	50	0.010	0.090	0.030	0.220	0.510	0.140		50		0.010	0.060	0.040	0.880	0.010
86	A1	170	0.671	172		0.137	0.026	0.163	0.529	0.142	0.003	179		0.008	0.034	0.028	0.930	
	A2	123	0.556	123	0.004	0.102	0.061	0.175	0.532	0.126		123		0.012	0.045	0.061	0.882	
87	A1	154	0.549	155		0.065	0.029	0.210	0.516*	0.177	0.003	141		0.007	0.035	0.032	0.926	
	A2	120	0.579	117		0.038	0.017	0.209	0.573	0.154	0.008	111		0.023	0.054	0.041	0.883	
	A3	70	0.643	67		0.067	0.030	0.157	0.612	0.134		70		0.007	0.050	0.043	0.900	
	A4	134	0.679	126		0.071	0.044	0.218	0.520	0.143	0.004	127		0.020	0.039	0.079	0.858	0.004
	A5	40	0.675	36		0.097	0.028	0.097	0.681	0.097		40			0.037	0.025	0.938	
	A6	71	0.655	55		0.082	0.036	0.164	0.536	0.155	0.027	56			0.062	0.027	0.911	
	A7	117	0.637	109		0.060	0.018	0.197	0.596	0.128		111		0.017	0.070	0.048	0.865*	

---

*$P \leq 0.05$ **$P \leq 0.01$ ***$P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 51. Allele frequencies for *Me*, *Acon-2* and *Mpi* for individual hawthorn (H) trees collected at site E from 1984 to 1987. See Figure 9c for a diagram of site E which includes hawthorn tree designations. N = sample size.

		Me		**Acon-2**								**Mpi**						
Yr	Tree	N	80	N	73	75	89	95	100	106	114	N	16	31	37	70	100	125
84	H1	101	0.272	101		0.035	0.059	0.490	0.297	0.114	0.005	101		0.005	0.178	0.074	0.703	0.040
85	H1	52	0.269	64		0.094	0.078	0.500	0.250*	0.078		65	0.008	0.015	0.085	0.069	0.785	0.038
	H2	53	0.274	53	0.010	0.094	0.085	0.500	0.273	0.028	0.010	53			0.216	0.057	0.717	0.010
	H3	51	0.225	52		0.010	0.019	0.645	0.250	0.067	0.010	52			0.183	0.067	0.683	0.067
86	H1	147	0.425	147		0.058	0.085	0.401	0.378	0.078		146		0.025	0.122	0.047	0.802	0.004
	H2	149	0.456	150		0.067	0.077	0.383	0.400	0.067	0.007	150		0.027	0.091	0.068	0.811	0.003
87	H1	140	0.364	136		0.037	0.051	0.467	0.353	0.085	0.007	125		0.024	0.104	0.040	0.824**	0.008
	H4	57	0.368	51		0.010	0.118	0.510	0.275	0.069	0.019	57		0.035	0.061	0.044	0.860	
	H5	88	0.375	69		0.014	0.058	0.471	0.384	0.072		80		0.025	0.081	0.056	0.838	
	H6	182	0.308	173		0.029	0.055	0.483	0.347	0.078	0.009	155		0.048	0.110	0.048	0.787*	0.006

*$P \leq 0.05$ **$P \leq 0.01$ ***$P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 52. Allele frequencies for *Dia-2*, *Aat-2* and *Had* for individual apple (A) trees collected at site E from 1984 to 1987. See Figure 9c for a diagram of site E which includes apple tree designations. N = sample size.

		Dia-2		**Aat-2**													**Had**	
Yr	Tree	N	85	N	01	21	32	40	50	59	75	84	100	113	123	130	N	122
84	A1	96	0.312	99	0.005	0.187		0.010	0.140	0.010	0.389	0.005	0.279		0.025		99	0.197
	A2	86	0.233	86		0.157	0.006	0.006	0.110		0.337	0.006	0.326		0.017	0.035	86	0.093
85	A1	24	0.271	25		0.120			0.240		0.300		0.300		0.020	0.020	25	0.260
	A2	63	0.192	52		0.183	0.010		0.058	0.038	0.260	0.019	0.393	0.010	0.019	0.010	65	0.162
	A3	50	0.350	51	0.010	0.157	0.020	0.010	0.185	0.010	0.245	0.010	0.294		0.039	0.020	51	0.274
86	A1	176	0.358	175	0.009	0.157			0.166	0.006	0.263**		0.371		0.029		178	0.219
	A2	123	0.317	123	0.008	0.154	0.008	0.008	0.163	0.008	0.272		0.346		0.024	0.008	123	0.163
87	A1	148	0.409	146		0.178	0.003		0.247		0.288		0.267		0.017		144	0.257
	A2	122	0.238	110	0.009	0.141		0.005	0.132		0.318***		0.377		0.014	0.005	112	0.188
	A3	67	0.396	67		0.187			0.201		0.299		0.299		0.015		70	0.236
	A4	133	0.267	134	0.004	0.154			0.173	0.004	0.271**		0.365*		0.038		135	0.207
	A5	37	0.311	35		0.071			0.214		0.329		0.329		0.043	0.014	39	0.103
	A6	54	0.278	56		0.107	0.009		0.205		0.277		0.357		0.045		71	0.155*
	A7	111	0.270	114		0.145	0.004		0.149		0.254		0.417		0.031		117	0.192

*$\underline{P} \leq 0.05$ **$\underline{P} \leq 0.01$ ***$\underline{P} \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 53. Allele frequencies for *Dia-2*, *Aat-2* and *Had* for individual hawthorn (H) trees collected at site E from 1984 to 1987. See Figure 9c for a diagram of site E which includes hawthorn tree designations. N = sample size.

		Dia-2		**Aat-2**													**Had**	
Yr	Tree	N	85	N	01	21	32	40	50	59	75	84	100	113	123	130	N	122
84	H1	100	0.105	101	0.005	0.104	0.005		0.089	0.005	0.292		0.465		0.035		101	0.124
85	H1	64	0.203	51		0.088	0.010		0.137	0.020	0.324		0.372		0.039	0.010	65	0.115
	H2	53	0.123	53		0.066			0.151		0.358		0.415		0.010		53	0.123
	H3	52	0.135	50		0.030	0.020		0.150		0.340	0.010	0.410		0.030	0.010	52	0.125
86	H1	146	0.212	143		0.105	0.003		0.168		0.304		0.399		0.021		149	0.174
	H2	150	0.203	143		0.063	0.014		0.126		0.294		0.479		0.024		149	0.174
87	H1	139	0.169	131		0.084	0.004		0.115		0.267		0.473	0.004	0.042	0.011	140	0.171
	H4	37	0.189	51		0.098			0.118		0.284*		0.441*		0.058		57	0.132
	H5	69	0.261	65		0.108	0.008		0.115		0.362*		0.377		0.031		85	0.212
	H6	162	0.219	167	0.003	0.087			0.144		0.234		0.491		0.036	0.006	182	0.162

*$P \leq 0.05$ **$P \leq 0.01$ ***$P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 54. Allele frequencies for *Me*, *Acon-2* and *Mpi* for local apple (A) and hawthorn (H) populations. See Figure 9a,b for study site designations. N = sample size.

			Me		**Acon-2**								**Mpi**					
Site	Host	Year	N	80	N	73	75	89	95	100	106	114	N	31	37	70	100	125
A	A	87	50	0.480	50		0.100	0.060	0.160	0.500	0.180		47	0.032	0.042	0.011	0.915	
B	A	86	57	0.512	41		0.085	0.012	0.256	0.451	0.195		35	0.043	0.114	0.014	0.829	
C	A	86	29	0.448**	19		0.026	0.052	0.237	0.474	0.184	0.026	9		0.112		0.888	
D	A	86	42	0.560	41	0.012	0.061	0.024	0.268	0.439	0.183	0.012	42	0.012	0.107	0.036	0.845	
E	A	84	183	0.671	185		0.084	0.030	0.184	0.530	0.159	0.013	185	0.005	0.051	0.046	0.893	0.005
E	A	85	128	0.578	137	0.004	0.113	0.058	0.208	0.493	0.120	0.004	140	0.004	0.071	0.039	0.879	0.007
E	A	86	293	0.621	295	0.002	0.122	0.041	0.168	0.531	0.136	0.002	302	0.010	0.038	0.041	0.911*	
E	A	87	706	0.620	665		0.064	0.029	0.194	0.560	0.148	0.005	600	0.013	0.049	0.046	0.892	
F	A	86	42	0.630	41		0.085	0.012	0.232	0.536	0.110	0.024	12		0.042		0.958	
G	A	86	42	0.726	39		0.205	0.051	0.128	0.359	0.256		36	0.010	0.038	0.041	0.911	
H	A	87	36	0.708	35		0.157	0.043	0.100	0.543	0.157		27		0.065	0.080	0.855	
A	H	87	42	0.226	42		0.012	0.107	0.619	0.250	0.012		42	0.012	0.214	0.048	0.714	0.012
I	H	86	40	0.425	40		0.087	0.062	0.450	0.375	0.025		41	0.012	0.122	0.012	0.853	
J	H	86	42	0.389*	39		0.038	0.115	0.500*	0.320	0.026		40	0.012	0.162	0.025	0.800	
K	H	86	40	0.388	34		0.250	0.103	0.441	0.206			38	0.013	0.184	0.039	0.763	
E	H	84	101	0.272	101		0.035	0.059	0.490	0.297	0.114	0.005	101	0.005	0.178	0.074	0.703	0.040
E	H	85	156	0.256	169	0.003	0.068	0.062	0.544	0.257	0.059	0.006	170	0.009	0.156	0.065	0.732	0.038
E	H	86	296	0.441	297		0.062	0.081	0.392	0.389	0.072	0.003	287	0.026	0.106	0.057	0.808*	0.003
E	H	87	467	0.345	429		0.026	0.062	0.479	0.346	0.078	0.009	417	0.035	0.096	0.047	0.817	0.005
L	H	86	42	0.405	35		0.043	0.043	0.571	0.300	0.043		25	0.020	0.120	0.100	0.760	
M	H	86	64	0.500	51		0.059	0.059	0.451	0.353	0.078		54	0.019	0.120	0.028	0.833	
H	H	87	33	0.561	32		0.172	0.016	0.392	0.344	0.077		28	0.036	0.143	0.036	0.786	

* $P \leq 0.005$ ** $P \leq 0.01$ *** $P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Table 55. Allele frequencies for *Dia-2*, *Aat-2* and *Had* for local apple (A) and hawthorn (H) populations. See Figure 9a,b for study site designations. N = sample size.

			Dia-2		**Aat-2**													**Had**	
Site	Host	Yr	N	85	N	01	21	32	40	50	59	75	84	100	113	123	130	N	122
A	A	87	47	0.319	47		0.181	0.032		0.149	0.011	0.447		0.181				50	0.180
B	A	86	41	0.463	37		0.216*			0.216*		0.311		0.230		0.027		36	0.222*
C	A	86	18	0.361	9		0.167			0.167		0.444		0.222				24	0.208
D	A	86	41	0.293	41	0.012	0.146			0.195		0.366		0.268		0.012		42	0.155
E	A	84	182	0.275	185	0.003	0.173	0.003	0.008	0.127	0.005	0.338	0.005	0.300		0.022	0.016	185	0.146
E	A	85	137	0.266	128	0.004	0.160	0.012	0.004	0.145	0.019	0.262	0.012	0.336	0.004	0.027	0.016	141	0.216
E	A	86	299	0.341	298	0.008	0.156	0.003	0.003	0.164	0.005	0.267*		0.361		0.027	0.005	301	0.196
E	A	87	672	0.310	662	0.002	0.150	0.002	0.001	0.186	0.001	0.285		0.344		0.026	0.002	688	0.203
F	A	86	37	0.378	42		0.274			0.095		0.309*		0.321				42	0.226*
G	A	86	42	0.452	41		0.207	0.012		0.195		0.293		0.293				42	0.262
H	A	87	35	0.386	33		0.288			0.242		0.121		0.319				36	0.292
A	H	87	31	0.145	42		0.059			0.167		0.274		0.476		0.024		42	0.107
I	H	86	40	0.325	39		0.128			0.256	0.013	0.179		0.410		0.013		42	0.155
J	H	86	40	0.087	41		0.061			0.110	0.012	0.366*		0.427		0.024		41	0.280
K	H	86	40	0.312	31		0.177			0.177		0.307		0.307		0.032		40	0.087
E	H	84	100	0.115	101	0.005	0.104	0.005		0.089	0.005	0.292		0.465		0.035		101	0.119
E	H	85	169	0.157	154		0.058	0.010		0.146	0.007	0.344	0.003	0.399		0.026	0.007	170	0.121
E	H	86	296	0.208	287	0.002	0.084			0.150	0.003	0.298		0.437		0.023	0.003	298	0.174
E	H	87	407	0.206	414	0.001	0.091	0.002		0.127		0.271		0.461	0.001	0.040	0.006	464	0.170
L	H	86	40	0.250	34		0.191			0.191		0.250		0.338		0.029		42	0.190
M	H	86	63	0.238	30		0.050			0.083		0.367*		0.483		0.017		64	0.156
H	H	87	33	0.136	33		0.061			0.076		0.303		0.545		0.015		33	0.303

*$P \leq 0.05$ **$P \leq 0.01$ ***$P \leq 0.001$ G-test for significant deviation from Hardy-Weinberg expectation.

Chapter 6

Concluding Summary and Remarks

> "One would think that it should no longer be necessary to devote much time to this topic, but past experience permits one to predict with confidence that the issue will be raised again at regular intervals. Sympatric speciation is like the Lernaean Hydra which grew two new heads whenever one of its old heads was cut off. There is only one way in which final agreement can be reached and that is to clarify the whole relevant complex of questions to such an extent that disagreement is no longer possible".
>
> E. Mayr (1963, pg. 451)

In this dissertation, I have shown that apple and hawthorn populations of *R. pomonella* are genetically differentiated host races which arose in sympatry. I first reported the observation of significant inter-host allele frequency variation for six allozymes (*Me*, *Acon-2*, *Mpi*, *Dia-2*, *Aat-2* and *Had*) at a field site near the town of Grant, in western Michigan. A number of alternative, but not mutually exclusive, causal explanations were present which could account for the genetic difference. Most importantly, however, adults captured directly from off of host trees showed a similar pattern of inter-host variation as larvae reared from host fruits. Adult flies are therefore not randomly migrating between different species of host plants and, as a consequence, hawthorn and apple flies are not randomly mating in nature. Differential host utilization is therefore a key factor restricting gene flow between the host races; a finding consistent with predictions of sympatric speciation models in which host race formation represents the initial stage of the divergence process.

I next verified a Mendelian mode of inheritance and determined map positions for the six allozymes displaying allele frequency differences between the host races. *Dia-2* and *Aat-2* were shown to be separated by a distance of 3.2 centimorgans on linkage group I, while *Me*, *Acon-2*, and *Mpi* were found to be tightly linked on chromosome II. *Had* was localized on linkage group III. The six loci showing differentiation between apple and hawthorn populations therefore mapped to only three specific regions of the genome. A genetic basis of inheritance and linkage associations were also determined for three other previously unresolved allozymes. One of these three loci, *Aspartate amino-transferase-1* (*Aat-1*), was found to be sex linked and indicated for the first time that males are the heterogametic (XY) sex in *R. pomonella*. In addition, significant gametic disequilibrium was observed between pairs of linked, non-allelic genes in the laboratory strain used for the genetic mapping experiments. The pattern of linkage disequilibrium for non-diapausing flies was different from that seen in natural populations, however, suggesting that the laboratory line has become genetically modified during culture.

A geographic study of allozyme variation was then conducted for hawthorn and apple populations from across the eastern United States and Canada. The results showed that host associated genetic divergence is widespread across the range of *R. pomonella* and is primarily due to the same six loci which varied at the Grant, Michigan site. Although genetic variation was much less among apple than among hawthorn populations across eastern North America (consistent with the apple race having gone through a recent bottleneck), the pattern of geographic differentiation gave no genetic clue as to the original source of the apple race. In addition, high levels of linkage disequilibrium were observed between loci within, but not between, each of the three genomic regions displaying host related differences. The patterns of disequilibrium among loci were similar, however, between apple and hawthorn flies. Therefore, although at least three specific regions of the genome of *R. pomonella* are co-evolving , the relationships among genes within these regions are the same for the two host races. Consequently, it is just the frequencies of certain gene or chromosomal "blocks" which differs between apple and hawthorn populations; a result which suggests that inversion polymorphisms exist for these regions and are promoting hitchhiking effects by reducing recombination among linked loci. Furthermore, epistatic interactions were not found between the different chromosomal regions displaying racial variation. Taken together, these data indicate that host race formation and, by implication, speciation, in the *R. pomonella* group involves genetic changes at only a few key loci, rather than a "revolution" involving a restructuring of genetic interactions throughout the genome.

Latitudinal allele frequency clines were also found among both apple and hawthorn populations for a majority of the allozymes displaying inter-host variation. The slopes of the allele frequency clines were different for apple and hawthorn races, however, being much steeper among hawthorn than apple populations. Inter-host genetic differentiation is therefore superimposed on clinal patterns of variation within the host races such that the magnitude of host associated divergence is a function of latitude. The allele frequency clines suggest that the host races are not experiencing uniform selection pressures across their ranges but instead may be strongly influenced by local environmental conditions. The configurations of the allele frequency clines do, in fact, implicate ambient temperature as an important environmental parameter affecting host-parasite interactions for *R. pomonella*. Allele frequencies for two loci (*Aconitase-2* and *Malic enzyme*) in the hawthorn race correlated with irregularities in isothermal clines for growing degree days (base 50° F) in the states of Wisconsin and Michigan. Furthermore, allele frequencies changed most dramatically for both apple and hawthorn populations across three areas in the Midwest that mark major ecological transition zones. Variation in voltinism and host plant associations are now known for a number of different phytophagous insects across these three regions indicating that environmental changes within these areas have a general impact on the evolution of life history traits for insects.

Finally, I examined local and microgeographic genetic variation among and between apple and hawthorn populations within a 60 km^2 area surrounding the original field site near Grant, Michigan. The results indicated that local and microgeographic genetic differentiation for these flies are a microcosm of the patterns seen across eastern North America. Even for trees separated by distances of less than 100 meters at the Grant site, *Malic Enzyme*^{100} and *Aconitase-2*^{95} allele frequencies were always significantly higher for hawthorn than apple flies every year over a four year period from 1984 to 1987. The two loci also showed a similar pattern of inter-host differentiation across 12 local populations analyzed in 1986. However, the local pattern of genetic variation was not identical to that seen at the Grant site as allele frequencies for *Me* and *Acon-2* showed a general correlation with latitude even across the limited area surveyed in the study. Although genetic drift and founder effects undoubtedly shape the fine scale spatial and temporal pattern of genetic variation for *R. pomonella*, differences in host recognition and host related survivorship had to be invoked to account for the maintenance of genetic differentiation between apple and hawthorn races. The data therefore agreed with the niche variation hypothesis that differential selection in heterogeneous environments promotes genetic polymorphism and lent further support to a sympatric mode of divergence for *R. pomonella*.

Several important questions still need to be answered concerning speciation in the *R. pomonella* group, however. First, we need to have a better understanding of the factors involved in differentiating *R. pomonella* populations. This requires establishing the extent of and mechanisms responsible for host plant fidelity and host associated survivorship, determining exactly how temperature acts as a differential selective factor, and estimating levels of effective gene flow among and between apple and hawthorn populations across the range of *R. pomonella* in North America. Second, we must more accurately quantify the extent and pattern of genetic divergence between host races and sibling species in the group. At least three genomic regions show host associated allele frequency differences. Are these the only regions of the genome containing genes responsible for the formation of the apple race or are more, as yet unidentified, regions involved? Do sibling species in the group show greater amounts of genetic variation in the same regions of the genome which display racial differences or are species specific mutations randomly scattered throughout the genome? What is the nature of genetic change within genomic regions differing among the host races and sibling species (do base pair substitutions, additions or deletions predominate?) and what phenotypic traits do these regions encode? High levels of linkage disequilibrium were found between non-allelic genes within the three regions of the genome displaying allele frequency differences between apple and hawthorn flies. Is linkage disequilibrium common throughout the genome of these flies or is disequilibrium confined only to those specific regions diverging between the host races and sibling species? Are inversion polymorphisms more prevalent in those regions of the genome displaying linkage disequilibrium? Third and finally, we have to demonstrate that host associated traits, such as differential host recognition, are directly responsible for the development of reproductive isolation among higher taxa in group. In other words, we must show that host races are, in fact, incipient species which may have reached the evolutionary point of no return. An important point mentioned in this regard is whether host fidelity behaviors are constant across the distribution of a host race. If they are and barring extinction, then host races should be considered to be potential species.

As a post-doctoral fellow at Princeton University and later as a faculty member at a major research University, I plan to continue work on the *R. pomonella* group through a series of experiments addressing the issues raised above. Field experiments will be performed to measure levels of intra- and inter-host gene flow and to identify and characterize specific phenotypes related to host recognition and host associated development that are restricting gene flow between apple and hawthorn populations. Molecular genetic studies are currently underway with Dr. Martin Kreitman of Princeton University to construct an extensive restriction site polymorphism linkage map for *R. pomonella*. The development of such a map will let us comprehensively screen the genome of these flies for genetic differentiation. In addition, the linkage map will provide a large number of genetic markers evenly distributed across the genome thereby making it possible to resolve potentially polygenic traits related to host preference and development, into discrete and mappable Mendelian factors. The feasibility and power of this strategy has recently been demonstrated by Paterson *et al.* (1988) working on the domestic tomato, *Lycopersicon esculentum*. Dr. Eric Lander of the Whitehead Institute and Harvard University, a co-researcher on the tomato project, has agreed to lend his expertise and computer programs in a collaborative effort with Dr. Kreitman and myself to accomplish similar phenotypic mapping experiments for *Rhagoletis*. What makes the *R. pomonella* group particularly amenable to genetic studies of speciation is that several pairs of sibling species in the complex, although genetically distinct in nature, can be hybridized in the laboratory to produce viable and fertile F1 progeny which can be backcrossed to either parental type. It will therefore be possible to determine whether genetic changes affecting phenotypes responsible for host shifts also form the underlying basis for reproductive isolation among higher taxa in the group.

By combining a variety of different experimental strategies, *Rhagoletis* workers will eventually reach their long term goal of understanding the ecological, behavioral and genetic basis for host race formation and speciation in the *R. pomonella* group. The unique biology and natural history of the *R. pomonella* group therefore make the flies an ideal model system for bridging the supposed gap between "micro-evolutionary" changes occurring at the population level and the "macro-evolutionary" process of speciation.

List of References

List of References

Barker, J.S.F., P.D. East and B.S. Weir. 1986. Temporal and microgeographic variation in allozyme frequencies in a natural population of *Drosophila buzzatii.* Genetics 112:577-611.

Bateson, W. 1922. Evolutionary faith and modern doubts. Science 55:55-61.

Bazykin, A.D. 1965. On the possibility of sympatric species formation. Bull. Moscow Soc. Nat. Biol. Div. 70:161-165.

Bazykin, A.D. 1969. Hypothetical mechanisms of speciation. Evolution 23:685-687.

Berger, E.M. 1971. A temporal survey of allelic variation in natural and laboratory populations of *Drosophila melanogaster*. Genetics 67:121-136.

Berlocher, S. H. 1976. The genetics of speciation in *Rhagoletis* (Diptera: Tephritidae). Ph. D. Thesis. University of Texas at Austin.

Berlocher, S. H. 1980. Genetics of allozymes of the apple maggot. J. Heredity 71: 63-67.

Berlocher, S. H. 1984a. Segregation and linkage of allozymes of the walnut husk fly. J. Heredity 75:392-396.

Berlocher, S.H. 1984b. Genetic changes coinciding with the colonization of California by the walnut husk fly, *Rhagoletis completa*. Evolution 38: 906-918.

Berlocher, S.H. and G.L. Bush. 1982. An electrophoretic analysis of (Diptera:Tephritidae) phylogeny. Syst. Zool. 31:136-155.

Berlocher, S.H. and D.C. Smith. 1983. Segregation and mapping of allozymes of the apple maggot fly. J. Heredity 74:337-340.

Bourne, A.I., W.H. Thies and F.R. Shaw. 1934. Some observations on long distance dispersal of apple maggot flies. J. Econ. Ent. 27:352-355.

Boller, E.F. and R.J. Prokopy. 1976. Bionomics and management of *Rhagoletis*. Ann. Rev. Ent. 21:223-246.

Brues, C.T. 1924. The specificity of food plants in the evolution of phytophagous insects. Am. Nat. 58:127-144.

Buriff, C.R. 1973. Recapture of released adult apple maggot flies on sticky-board traps. Environ. Entomol. 2:757-758.

Bush, G.L. 1966. The taxonomy, cytology and evolution of the genus *Rhagoletis* in North America (Diptera: Tephritidae). MCZ, Cambridge, Mass.

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

Bush, G.L. 1969b. Mating behavior, host specificity, and the ecological significance of sibling species in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). Am. Nat. 103:669-672.

Bush, G.L. 1975a. Modes of animal speciation. Ann. Rev. Ecol. Syst. 6:339-364.

Bush, G.L. 1975b. Genetic variation in natural insect populations and its bearing on mass-rearing programmes. pp.9-18. In: *Controlling Fruit Flies by the Sterile-Insect Technique*. International Atomic Energy Agency. Vienna, Austria.

Bush, G.L. and R.W. Neck, Jr. 1976. Ecological genetics of the screwworm fly, *Cochliomyia hominivorax* (Diptera:Calliphoridae) and its bearing on the quality control of mass reared insects. Environ. Entomol. 5:821-826.

Bush, G.L., R.W. Neck Jr., and G.B. Kitto. 1976. Screwworm eradication: Inadvertent selection for noncompetitive ecotypes during mass rearing. Science 193:491-493.

Bush, G.L. and G.B. Kitto. 1979. Research on the genetic structure of wild and laboratory strains of olive fly. In: *F.A.O. Report, Development of Pest Management Systems for Olive Culture Programs*. F.A.O. of the United Nations. Rome. 27pp.

Bush, G.L. and S.R. Diehl. 1982. Host shifts, genetic models of sympatric speciation and the origin of parasitic insect species. pp. 297-306. In: *Insect-Plant Relationships*. Proceedings of the 5th International symposium. (J.H. Visser and A.K. Minks, eds.) Pudoc, Wageningen, Holland.

Butlin, R. 1987. A new approach to sympatric speciation. Tree 2:310-311.

Caesar, L. and W.A. Ross. 1919. The apple maggot. Ontario Dept. Agr. Bull. 275.

Carson, H. 1959. Genetic conditions which promote or retard the formation of species. Cold Spring Harbor Symp. Quant. Biol. 24:87-105.

Carson, H. 1965. Chromosomal morphism in geographically widespread species of *Drosophila*. pp. 503-531. In: *Genetics of Colonizing species*. (H.G. Baker and G.L. Stebbins, eds.) Academic Press, New York.

Carson, H. 1968. The population flush and its genetic consequences. pp. 123-137. In: *Population Biology and Evolution.* (R.C. Lewontin, ed.) Syracuse University Press, Syracuse, New York.

Carson, H. 1971. Speciation and the founder principle. Stadler Genetics Symposia 3:51-70.

Carson, H. and A.T. Ohta. 1981. Origin of the genetic basis of colonizing ability. pp. 365-370. In: *Evolution Today, The Proceedings of the Second International Congress of Systematic and Evolutionary Biology.* (G.G.E. Scudder and J.L. Reveal, eds.). Hunt Institut. for Botanical Documentation, Carnegie-Mellon Univ., Pittsburgh, Pa.

Chapman, P.J. and S.E. Lienk. 1971. *Tortricid Fauna of Apple in New York.* Cornell Univ., Ithaca, New York. 122 pp.

Christiansen, F.B. and O. Frydenberg. 1974. Geographic patterns of four polymorphisms in *Zoarces viviporus* as evidence of selection. Genetics 77:765-770.

Christiansen, F.B. and M.W. Feldman. 1975. Subdivided populations: A review of the one- and two-locus deterministic theory. Theor. Pop. Biol. 7:13-38.

Clarke, B. 1968. Balanced polymorphism and regional differentiation in land snails. In: *Evolution and Environment.* (E.T. Drake, ed.) Yale Univ. Press, New Haven, Conn.

Clegg, M.T. and R.W. Allard. 1972. Patterns of genetic differentiation in the slender wild oat species, *Avena barbata.* PNAS 69:1820-1824.

Cockerham, C.C. and B.S. Weir. 1977. Digenic descent measures for finite populations. Genetical Res. 30:12-147.

Correll, D.S. and M.C. Johnston. 1970. Manual of the Vascular Plants of Texas. 1881 pp. Texas Research Foundation, Renner, Texas.

Darwin, C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* John Murray, London, England.

Dean, R.W. and P.J. Chapman. 1973. Bionomics of the apple maggot in eastern New York. Search Agric. Entomol. Geneva No. 3.

Diehl, S.R. 1984. The role of host shifts in the ecology and speciation of *Rhagoletis* flies (Diptera:Tephritidae). Ph. D. Thesis. Univ. of Texas at Austin.

Diehl, S.R. and G.L. Bush. 1984. An evolutionary and applied perspective of insect biotypes. Ann. Rev. Entomol. 29:471-504.

Diehl, S.R. and G.L. Bush. 1989. The role of habitat preference in adaptation and speciation. In: *Speciation and Its Consequences.* (D. Otte and J. Endler, eds.) Sinauer Assoc.

Dobzhansky, Th. and F.J. Ayala. 1973. Temporal frequency changes of enzyme and chromosomal polymorphisms in natural populations of *Drosophila.* PNAS 70:680-683.

Economopoulous, A.P. 1980. SIRM against the olive fruit fly: Differences between wild and lab-reared (normal or sterilized) insects. pp. 17-26. In: *Proceedings of XVI. Int. Congress of Entomology, Symp. on Fruit Fly Problems.* Kyoto and Naha, Japan.

Economopolous, A.P. and G.A. Zervas. The quality problem in olive flies produced for SIT experiments. pp. 357-368. In: *Proceedings of Int. Symp. on Fruit Flies of Economic Importance.* Athens, Greece. 1982.

Endler, J.A. 1977. *Geographic variation, speciation, and clines.* Princeton University Press, Princeton, New Jersey.

Feder, J.L. and G.L. Bush. Differential host utilization is an effective premating barrier between sibling species in the *Rhagoletis pomonella* group. Evolution (submitted).

Feder, J.L., C.A. Chilcote and G.L. Bush. Are the apple maggot, *Rhagoletis pomonella* and the blueberry maggot, *R. mendax*, distinct species? Implications for sympatric speciation. Entomol. exp. appl. (In press).

Felsenstein, J. 1976. The theoretical population genetics of variable selection and migration. Ann. Rev. Genet. 10:253-280.

Felsenstein, J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals. Evolution 35:124-138.

Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. Ann. Rev. Genet. 22:521-565.

Fernald, M.L. 1950. *Gray's Manual of Botony*. Eigth Edition. 1632 pp. American Book Company, New York.

Fisher, R.A.F. 1930. *The Genetical Theory of Natural Selection.* Clarendon, Press. Oxford, England.

Franklin, I.R. 1981. An analysis of temporal variation at isozyme loci in *Drosophila melanogaster*. pp. 217-236. In: *Genetic Studies of Drosophila Populations. Proceedings of 1979 Kioloa Conference.* (J.B. Gibson and J.G. Oakeshott, eds.). Australian Natl. Univ., Canberra.

Futuyma, D.J. and G.C. Mayer. 1980. Non-allopatric speciation in animals. Syst. Zool. 29:254-271.

Futuyma, D.J. and S.C. Peterson. 1985. Genetic variation in the use of resources by insects. Ann. Rev. Entomol. 30:217-238.

Gasperi, G., A. Malacrida, M. Tosetti and R. Milani. 1986. Enzyme variability: A tool for investigating the genome organization and the population structure of *Ceratitus capitata*. pp. 153-161. In: *Fruit Flies of Economic Importance 84*. (R. Cavalloro, ed.) A.A. Balkema, Rotterdam, Netherlands.

Gould, S.J. and N. Eldredge. 1977. Punctuated equilibrium: The tempo and mode of evolution reconsidered. Paleobiology 3:115-151.

Hamrick, J.L. and R.W. Allard. 1972. Microgeographic variation in allozyme frequencies in *Avena barbata*. PNAS 69:2100-2104.

Harris, H. 1966. Enzyme polymorphisms in man. Proc. Roy. Soc. Ser. B 164:298-310.

Hedrick, P.W., M.E. Ginevan and E.P. Ewing. 1976. Genetic polymorphism in heterogeneous environments. Ann. Rev. Ecol. Syst. 7:1-32.

Herrick, G.W. 1920. The apple maggot in New York. Cornell Univ. Agr. Exp. Sta. Bull. 402:89-101.

Huettel, M.D. and G.L. Bush. 1972. The genetics of host selection and its bearing on sympatric speciation in *Procecidochares* (Diptera: Tephritidae). Entomol. Exp. Appl. 15:465-480.

Illingworth, J.F. 1912. A study of the biology of the apple maggot (*Rhagoletis pomonella*), together with an investigation of methods of control. Cornell Univ. Agr. Expt. Stat. Bull. 324:129-187.

Istock, C.A. and W.G. Weisberg. 1987. Strong habitat selection and the development of population structure in a mosquito. Evol. Ecol. 1: 348-362.

Jaenike, J. 1987. Genetics of oviposition-site preference in *Drosophila tripunctata*. Heredity 59:363-369.

Johnson, F.M. and H.E. Schaffer. 1973. Isozyme variability in species of the genus *Drosophila*. VII. Genotype-environmental relationships in populations of *Drosophila melanogaster* from the eastern United States. Biochem. Genet. 10:149-163.

Johnson, G.B. 1976. Polymorphism and predictability of the alpha-glycerophosphate dehydrogenase locus in *Colias* butterflies. Biochem. Genet. 14:403-425.

Johnson, M.S. 1971. Adaptive lactate dehydrogenase variation in the crested blenny, *Anoplorchus*. Heredity 27:205-226.

Jordan, K. 1896. On mechanical selection and other problems. Novit. Zool. 3:426-525.

Kettlewell, H.B.D. 1973. *The Evolution of Melanism*. Clarendon, Oxford, England.

Koehn, R.K. 1969. Esterase heterogeneity: Dynamics of a polymorphism. Science 163:943-944.

Koehn, R.K., R. Milkman and J.B. Mitton. 1976. Population genetics of marine pelecypods. IV. Selection, migration and genetic differentiation in the blue mussel, *Mytilus edulis*. Evolution 30:2-32.

Langley, C.H., R.A. Voelker, A.J. Leigh Brown, S. Ohnishi, B. Dickson and E. Montgomery. 1981. Null allele frequencies at allozyme loci in natural populations of *Drosophila melanogaster*. Genetics 99:151-156.

Laurie-Ahlberg, C. C. and B. S. Weir. 1979. Allozymic variation and linkage disequilibrium in some laboratory populations of *Drosophila melanogaster*. Genetics 92: 1295-1314.

Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. Am. Nat. 87:331-333.

Levins, R. 1968. *Evolution in changing environments*. Princeton Univ. Press, Princeton, N.J.

Lewis, H. 1962. Catastrophic selection as a factor in speciation. Evolution 16:257-271.

Lewis, H. 1966. Speciation in flowering plants. Science 152:167-172.

Lewontin, R.C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York, New York. pp. 346.

Lewontin, R.C. and Hubby. 1966. A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila psuedoobscura*. Genetics 54:595-609.

Loukas, M.G., A.P. Economopoulos, E. Zouros and Y. Verini. 1985. Genetic changes in artificially reared colonies of the Olive Fruit Fly. Ann. Ent. Soc. Amer. 78:159-165. 1985.

Lynch, M. 1987. The consequences of fluctuating selection for isozyme polymorphisms in *Daphnia*. Genetics 115:657-669.

Maruyama, T. and P.A. Fuerst. 1985. Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics 111:675-689.

Maxwell, C.W. 1968. Apple maggot adult dispersion in a New Brunswick apple orchard. J. Econ. Ent. 61:103-106.

Maynard Smith, J. 1966. Sympatric speciation. Am. Nat. 100:637-650.

Maynard Smith, J. 1970. Genetic polymorphism in a varied environment. Am. Nat. 104:487-490.

Mayr, E. 1942. *Systematics and the origin of species.* Columbia University Press, New York.

Mayr, E. 1954. Change of genetic environment and evolution. pp. 157-180. In: *Evolution as a Process.* (J. Huxley, A.C. Hardy and E.B. Ford, eds.) Allen and Unwin, London, England.

Mayr, E. 1963. *Animal Species and Evolution.* Harvard Univ. Press, Cambridge, Mass. 797 pp.

McDonald, J.F. 1983. The molecular basis of adaptation: A critical review of relevant ideas and observations. Ann. Rev. Ecol. Syst. 14:77-102.

McDonald, J.F. and F.J. Ayala. 1974. Genetic response to environmental heterogeneity. Nature 250:572-574.

McPheron, B.A. 1987. The population genetics of the colonization of the western United States by the apple maggot, *Rhagoletis pomonella* (Walsh) (Diptera: Tephritidae). Ph. D. Dissertation, Univ. of Illinois, Urbana, Illinois.

McPheron, B. A. and S. H. Berlocher. 1985. Segregation and linkage of allozymes of *Rhagoletis tabellaria.* J. Heredity 76: 218-219.

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

McPheron, B.A., D.C. Smith and S.H. Berlocher. 1988b. Microgeographic genetic variation in the apple maggot, *Rhagoletis pomonella.* Genetics 119:445-451.

McPheron, B.A., C.D. Jorgensen and S.H. Berlocher. 1988c. Low genetic variability in a Utah cherry-infesting population of the apple maggot, Rhagoletis pomonella. Entomol. Exp. Appl. 46:155-160.

Menken, S.B.J. and S.A. Ulenberg. 1987. Biochemical characters in agricultural entomology. pp. 305-360. In: *Agricultural Zoological Reviews. Vol. 2.* Intercept, Ponteland, Newcastle upon Tyne.

Merritt, R.B. 1972. Geographic distribution and enzymatic properties of lactate dehydrogenase allozymes in the fat head minnow, *Pimephales promelas*. Am. Nat. 196:173-184.

Miller, S., R.W. Pearcy and E. Berger. 1975. Polymorphism at the alpha-glycerophosphate dehydrogenase locus in *Drosophila melanogaster*. I. Properties of adult allozymes. Biochem. Genet. 13:175-188.

Mitton, J.B. and R.K. Koehn. 1975. Genetic organization and adaptive response of allozymes to ecological variables in *Fundulus heteroclitus*. Genetics 79:97-111.

Mully, J.C., J.W. James and J.S.F. Barker. 1979. Allozyme genotype-environment relationships in natural populations of *Drosophila buzzatii*. Biochem. Genet. 17:105-126.

Murray, J. 1972. *Genetic Diversity and Natural Selection*. Oliver and Boyd, Edinburg, Scotland.

Nei, M. 1977. F-statistics and analysis of gene diversity in subdivided populations. Ann. Hum. Genet. 41:225-233.

Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. Evolution 29:1-10.

Neilson, W.T.A. 1971. Dispersal studies of a natural population of apple maggot adults. J. Econ. Ent. 64(3):648-653.

O'Kane, W.C. 1914. The apple maggot. New Hampshire Expt. Sta. Bull. No. 171., 120 pp.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726.

Phipps, C.R. and C.O. Dirks. 1933a. Dispersal of the apple maggot fly-1932 studies. J. Econ. Ent. 5:576-582.

Phipps, C.R. and C.O. Dirks. 1933b. Notes on the biology of the apple maggot fly. J. Econ. Ent. 26:349-358.

Porter, B.A. 1928. The Apple Maggot. U.S. Dept. Agr. Tech. Bull. 66., 48 pp.

Powell, J.R. 1971. Genetic polymorphism in varied environments. Science 174:1035-1036.

Powell, J.R. and H. Wistrand. 1978. The effect of heterogeneous environments and a competitor on genetic variation in *Drosophila*. Am. Nat. 112:935-947.

Price, P.W. 1980. *Evolutionary biology of parasites*. Princeton Univ. Press, Princeton, N.J.

Prokopy, R.J. 1968. Influence of photoperiod, temperature and food on initiation of diapause in the apple maggot. Canad. Ent. 100:318-329.

Prokopy, R.J., E.W. Bennett and G.L. Bush. 1971. Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). I. Site of assembly. Canad. Ent. 103:1405-1409.

Prokopy, R.J., E.W. Bennett and G.L. Bush. 1972. Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). II. Temporal organization. Canad. Ent. 104:97-104.

Prokopy, R.J. and G.L. Bush. 1972. Apple maggot infestation of pear. J. Econ. Ent. 65:597.

Prokopy, R.J., G.E. Haniotakis and A.P. Economopoulos. 1975. Comparative behavior of lab-cultured and wild-type *Dacus oleae* flies in the field. pp. 101-108. In: *Controlling Insects by the Sterile Insect Technique*. International Atomic Energy Agency,Vienna.

Prokopy, R.J. and S.H. Berlocher. 1980. Establishment of *Rhagoletis pomonella* (Diptera: Tephritidae) on rose hips in southern New England. Can. Ent. 112:1319-1320.

Prokopy, R.J., A.L. Averill, S.S. Cooley and C.A. Roitberg. 1982. Associative learning in egglaying site selection by apple maggot flies. Science 218:76-77.

Prokopy, R.J., S.R. Diehl, and S.S. Cooley. 1988. Behavioral evidence for host races in *Rhagoletis pomonella* flies. Oecologia 76: 138-147.

Ramshaw, J.A.M., J.A. Coyne and R.C. Lewontin. 1979. The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics 93:1019-1037.

Reissig, W.H. 1977. Capture of released apple maggot flies, *Rhagoletis pomonella* (Diptera:Tephritidae), on sticky panels. Can. Ent. 109:631-634.

Reissig, W.H. 1979. Survival of apple maggot larvae, *Rhagoletis pomonella* (Diptera: Tephritidae), in picked and unpicked apple. Can. Ent. 111:181-187.

Reissig, W. H. and D.C. Smith. 1978. Bionomics of *Rhagoletis pomonella* in *Crataegus*. Ann. Ent. Soc. Amer. 71: 155-159.

Reissig, W.H., J. Barnard, R.W. Weires, E.H. Glass and R.W. Dean. 1979. Prediction of apple maggot fly emergence from thermal unit accumulation. Environ. Entomol. 8:51-54.

Rice, W.R. 1984. Disruptive selection on habitat preference and the evolution of reproductive isolation: a simulation study. Evolution 38:1251-1260.

Rockey, S.J., J.H. Hainze and J. M. Scriber. 1987. Evidence of a sex-linked diapause response in *Papilio glaucus* subspecies and their hybrids. Physiol. Entomol. 12:181-184.

Saul, S.H. 1986. Genetics of the Mediterranean fruit fly (*Ceritatis capitata*) (Wiedemann). pp. 73-108. In: *Agricultural Zoological Reviews. Vol 1*. Intercept, Ponteland, Newcastle upon Tyne.

Saul, S.H., M.J. Sinsko, P.R. Grimstad and G.B. Craig, Jr. 1978. Population genetics of the mosquito *Aedes triseriatus*: Genetic-ecological correlations at the *esterase* locus. Am. Nat. 112:333-339.

Scriber, J.M. and J.H. Hainze. 1987. Geographic invasion and abundance as facilitated by differential host plant utilization abilities. pp. 433-468. In: *Insect Outbreaks: Ecology and Evolutionary Processes*. (P. Barbarosa and J.C. Shultz, eds.) Academic Press, N.Y.

Shervis, L.J., G.M. Boush and C.F. Koval. 1970. Infestation of sour cherries by the apple maggot: Confirmation of a previously uncertain host status. J. Econ. Ent. 63:294-295.

Showers, W.B. 1981. Geographic variation of the diapause response in the European corn borer. In: *Insect Life History Patterns* (R.F. Denno and H. Dingle, eds.) pp. 97-111. Springer-Verlag, Berlin and New York.

Singh, P. 1984. Insect diets: Historical developments, recent advances and future prospects. pp. 32-44. In: *Advances and Challenges in Insect Rearing*. (E.C. King and N.C. Leppla, eds.). Agricultural Research Series. USDA. Washington, D.C.

Smith, D.C. 1986. *Genetics and Reproductive Isolation of Rhagoletis flies*. Ph. D. dissertation. Univ. of Illinois, Urbana, Illinois.

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

Smith, H.S. 1941. Racial segregation in insect populations and its significance in applied entomology. J. Entomol. 34:1-12.

Spiess, E.B.1977.*Genes in Populations*. Wiley and Sons, New York. 780pp.

Standley, P.C. 1922. *Trees and Shrubs of Mexico*. 1721 pp. Smithsonian Press, Washington, D.C.

Steiner, W.M. 1979. Genetic variation in Hawaiian *Drosophila*. VI. Seasonally dependent gene changes in *Drosophila mimica*. Evolution 33:543-563.

Tauber, C.A., M.J. Tauber and J.R. Nechols. 1977. Two genes control seasonal isolation in sibling species. Science: 197: 592-593.

Tauber, M.J. and C.A. Tauber. 1976. Insect seasonality: Diapause maintenance, termination, and postdiapause development. Ann. Rev. Entomol. 21:81-107.

Taylor, C.E. and J.R. Powell. 1976. Microgeographic differentiation of chromosomal and enzyme polymorphisms in *Drosophila pseudoobscura*. Genetics 85:681-695.

Templeton, A.R. 1981. Mechanisms of speciation- A population genetic approach. Ann. Rev. Ecol. Syst. 12:23-48.

Thompson, J. 1988. Evolutionary genetics of oviposition preference in Swallowtail Butterflies. Evolution 42:1223-1234.

Thorpe, W.H. 1930. Biological races in insects and allied groups. Biol. Rev. 5:177-212.

Van Valen, L. 1965. Morphological variation and the width of the niche. Am. Nat. 99:377-390.

Wagner, M. 1868. *Die Darwin'sche Theorie und das Migrationsgesetz der Organismen.* Duncker and Humblot, Leipzig, Austria.

Wagner, M. 1889. Die Entstehung der Arten durch raumliche Sonderung Benno Schwalbe, Basel, Switzerland.

Walsh, B.D. 1864. On phytophagic varieties and phytophagic species. Proc. Ent. Soc. of Philadelphia 3:403-430.

Walsh, B.D. 1867. The apple-worm and the apple maggot. J. Hort. 2:338-343.

Ward, C. 1866. Answers to correspondents. Practical Entomologist 2:20-21.

Wasbauer, M.S. 1972. An annotated host catalog of the fruit flies of America north of Mexico (Diptera: Tephritidae). Bureau of Entomol., Dept. of Agric., Sacarmento , Calif. Occasional papers no. 19.

Weir, B. S. 1979. Inferences about linkage disequilibrium. Biometrics 35: 235-254.

Weir, B.S. and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38:1358-1370.

White, M.J.D. 1968. Models of speciation. Science 159:1065-1070.

White, M.J.D. 1978. *Modes of Speciation.* W.H. Freeman and Co., San Francisco. pp. 455.

Williams, S.M., J.L. Feder, B.A. McPheron, S.H. Berlocher and G.L. Bush. The effects of recent colonization on hybridization and introgression between two sibling species in the Rhagoletis pomonella group (Diptera: Tephritidae). (Manuscript in preparation).

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16: 97-159.

Wright, S. 1940a. Breeding structure of populations in relation to speciation. Am Nat. 74:232-248.

Wright, S. 1940b. The statistical consequences of Mendelian heredity in relation to speciation. pp. 161-183. In: The New Systematics. (J. Huxley, ed.) Oxford Univ. Press, Oxford, England.

Wright, S. 1943. Isolation by distance. Genetics 28:114-138.

Wright, S. 1951. The genetical structure of populations. Ann. Eugen. (Lond.) 15:323-354.