ABSTRACT

Title of Document: GEOGRAPHY AND GENETICS OF ECOLOGICAL SPECIATION IN PEA APHIDS

Joan A. West, Ph.D., 2008

Directed By: Associate Professor David J. Hawthorne, Department of Entomology

During ecological speciation, divergent natural selection drives evolution of ecological specialization and genetic differentiation of populations on alternate environments. Populations diverging onto the same alternate environments may be geographically widespread, so that divergence may occur at an array of locations simultaneously. Spatial variation in the process of divergence may produce a pattern of differences in divergence among locations called the Geographic Mosaic of Divergence. Diverging populations may vary in their degree of genetic differentiation and ecological specialization among locations. My dissertation examines the pattern and evolutionary processes of divergence in pea aphids (*Acyrthosiphon pisum*) on alfalfa (*Medicago sativa*) and clover (*Trifolium pretense*).

In Chapter One, I examined differences among North American aphid populations in genetic differentiation at nuclear, sequence-based markers and in ecological specialization, measured as aphid fecundity on each host plant. In the East, aphids showed high host-plant associated ecological specialization and high

genetic differentiation. In the West, aphids from clover were genetically indistinguishable from aphids on alfalfa, and aphids from clover were less specialized. Thus, the pattern of divergence differed among locations, suggesting a Geographic Mosaic of Divergence.

In Chapter Two, I examined genomic heterogeneity in divergence in aphids on alfalfa and clover across North America using amplified fragment length polymorphisms (AFLPs). The degree of genetic differentiation varied greatly among markers, suggesting that divergent natural selection drives aphid divergence in all geographic locations. Three of the same genetic markers were identified as evolving under divergent selection in the eastern and western regions, and additional divergent markers were identified in the East.

In Chapter Three, I investigated population structure of aphids in North America, France, and Sweden using AFLPs. Aphids on the same host plant were genetically similar across many parts of their range, so the evolution of host plant specialization does not appear to have occurred independently in every location. While aphids on alfalfa and clover were genetically differentiated in most locations, aphids from alfalfa and clover were genetically similar in both western North America and Sweden. High gene flow from alfalfa onto clover may constrain divergence in these locations.

GEOGRAPHY AND GENETICS OF ECOLOGICAL SPECIATION IN PEA APHIDS

By

Joan A. West

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2008

Advisory Committee:
Associate Professor David Hawthorne, Chair
Professor Charles Mitter
Adjunct Professor Sonja Scheffer
Professor Barbara Thorne
Professor Gerald Wilkinson

Acknowledgements

I thank my dissertation advisor, Dave Hawthorne, for valuable advice, help on all aspects of the dissertation, and insightful discussions that greatly improved this research. Thank you to my committee members, Charlie Mitter, Barbara Thorne, Sonja Scheffer, and Jerry Wilkinson for thoughtful feedback on my research and the dissertation, and for challenging my views on sympatric speciation. I thank Matt Hare for answering my many questions about data analyses and for suggesting that gene swamping could explain the pattern in the West.

I thank Sara Via for her contributions. She also came up with the idea of the Geographic Mosaic, and the work with live aphids was done in her lab.

I owe many thanks to all the past and present (and adjunct) members of the Hawthorne and Via labs for assistance in development of the research ideas, aphid collections, marker development, data collection, and for their friendship, inspiration, and camaraderie: Renee Godinez, Gwen Shlichta, Andreanna Welch, Gina Conte, Christy Olson, Melanie Delion, Megan McCarthy, Stacey Bealmear, Diane Kang, and Rob Ahern. Andrea Bagley collected part of the genetic data used in Chapters 2 and 3, and I am thankful to her and to Dave Hawthorne for the opportunity to analyze this data as part of my dissertation.

I am very appreciative of Maria Murray, Jaime Grace, and Sky Lesnick for insightful discussions and suggestions for the data analysis and interpretation.

I also thank Julia Ferrari for the opportunity to collect aphids in England, for data collection, and for thoughtful discussions.

I am tremendously grateful to all my dear friends for their encouragement, for providing an incredible support network, and for always reminding me to keep things in perspective. Thank you especially to Laurie Alexander for her exuberance and wisdom, to Cindy Darcy for keeping me on track, to Peter Soukup for everything, to my roommates on Marlbrough Way for their friendship, and to Julie Byrd for insightful comments on all aspects of this project, and making me think deeply about host plant specialization.

Finally, I thank my sister, Gail Brooks, and my parents, Neal and Nancy West, for always being there for me.

Funding for this research comes from an NSF Biology of Small Populations RTG, and grants and the Darwin Fellowship from the BEES program at the University of Maryland.

Table of Contents

List of Tables

List of Figures

Chapter 1: The geographic mosaic of divergence: geographic differences in genetic divergence and ecological specialization of pea aphids on alfalfa and clover in North America

Chapter 2: Ecological divergence across the genome and across geographic locations in pea aphids

Chapter 3: Population structure of pea aphids on alfalfa and clover in North America and Europe and the origins of host plant specialization

Introduction

A major controversy in evolutionary biology regards the relative importance of allopatric and sympatric speciation, and whether speciation can occur without geographic isolation of populations (Futuyma and Mayer 1980, Via 2001, Berlocher and Feder 2002, Coyne and Orr 2004). During allopatric speciation, reproductive isolation evolves as a byproduct of divergent evolution in geographically isolated populations (Mayr 1963). In contrast, during sympatric speciation, divergence occurs in populations within range of one another that can still exchange genes (Bush 1969, Bolnick and Fitzpatrick 2007). From the Modern Synthesis until the 1980's, allopatric speciation was thought to be the primary mode of speciation, explaining the vast majority of biological diversity (Coyne and Orr 2004). In the last twenty years, theoretical and empirical research has caused a resurgence in support for sympatric speciation (reviewed in Via 2001). Recently, a pluralistic view of speciation has emerged (Schilthuizen 2000, Jiggins and Bridle 2004, Rundle and Nosil 2005), and the focus of speciation research may be shifting from determining the geography of speciation to more tractable questions about the genetic basis and evolutionary mechanisms of speciation (Via 2001, Schluter 2001).

Here, I describe the historical and conceptual context of my dissertation research on the geography and genetics of divergence in pea aphids on alfalfa and clover. I briefly describe the allopatric/sympatric speciation controversy and several limitations to the allopatric/sympatric dichotomy for classifying the mode of speciation. Then, I highlight recent shifts in the focus and approaches of speciation research.

Allopatric/sympatric speciation controversy

The view that allopatric speciation is the primary mode of speciation has dominated the field of evolutionary biology since the 1930's (Futuyma and Mayer 1980, Coyne and Orr 2004). Mayr (1963) found strong evidence to support allopatric speciation in a wide range of animal groups and argued forcefully for the primacy of allopatric speciation in explaining the majority of biological diversity. Dobzhansky and Muller showed how the evolution of hybrid sterility and inviability could evolve due to fixation of incompatible alleles in geographically isolated populations (Orr 1995, Turelli *et al*. 2001). These models provide a genetic mechanism to explain allopatric speciation, and suggested that because the genome acts as a coadapted unit, any gene flow would prevent speciation. This assumption has had a lasting impact on speciation research but has recently been challenged by genetic evidence for divergence-with-gene-flow (Emelianov *et al*. 2004, Savolainen *et al*. 2006).

Sympatric speciation occurs when restricted gene flow between diverging populations is due not to geographic isolation but rather the biology of the organisms (Futuyma and Mayer 1980). Sympatric speciation can happen when adaptation to alternate environments leads to the evolution of reproductive isolation. The idea that ecology and natural selection are important for speciation can be traced back to Darwin (1859). One of the early proponents of sympatric speciation, Bush (1969), showed that sympatric speciation may be especially common in phytophagous insects due to host shifts and specialization onto the new host plant. Since then, support for sympatric speciation has come from theoretical research and empirical studies of sister taxa and incipient species (Bush 1994, Via 2001).

Recent theory has contributed to a resurgence of interest in sympatric speciation (Via 2001, Berlocher and Feder 2002, Bolnick and Fitzpatrick 2007). Theoretical models show the conditions under which sympatric speciation are likely to occur (Dieckmann and Doebeli 1999, Kondrashov and Kondrashov 1999, reviewed in Turelli *et al*. 2001), and these conditions may be quite common in nature (Kondrashov and Mina 1986, Bolnick and Fitzpatrick 2007).

Phylogeographic comparative studies have also investigated the relative importance of allopatric and sympatric speciation in already diverged species (Coyne and Orr 2004). Because speciation is a gradual process that occurs over long periods of time and cannot be directly observed in nature, sister species, as recent products of speciation, can be compared to infer the biogeography of speciation (Barraclough and Vogler 2000, Noor and Feder 2004). Analyses of the phylogenies and geographic ranges of closely related species have been cited as providing support for both allopatric speciation (Coyne and Orr 2004) and for sympatric speciation (Via 2001). However, the utility of this approach for determining the relative likelihood of sympatric and allopatric speciation has been challenged (Losos and Glor 2003, Coyne and Orr 2004 but see Mallet 2005). The current geographic ranges of the diverged species may differ from their distribution at the time of speciation (Losos and Glor 2003), and populations that are currently sympatric may have diverged allopatrically (Grant and Grant 1997, Rundle and Nosil 2005). Thus, the most reliable evidence for sympatric speciation comes from cases where additional evidence suggests a sympatric origin such as for adaptive radiations in isolated locations or where historical information is available (i.e. Schliewen *et al*. 1994, Wilson *et al.* 2000, Savolainen *et al*. 2006).

Strong evidence for sympatric speciation also comes from case studies of probable incipient species (Via 2001, Berlocher and Feder 2002, Bolnick and Fitzpatrick 2007). Highly divergent populations including host races, host associated populations, and ecotypes may be at intermediate stages in the divergence process (Jaenike 1981, Dres and Mallet 2002). By studying natural populations at different stages of speciation, a general sequence of important steps in speciation can be reconstructed (Dres and Mallet 2002). Inferences about the process of speciation from studies of incipient species are often criticized because populations at intermediate stages of divergence may never evolve into full species, and could, in fact, fuse back into a single population (i.e. Taylor *et al*. 2006). For this reason, studies of incipient species are more accurately characterized as studies of the process of divergence, because speciation may or may not proceed to completion. Nevertheless, studies of incipient species provide some of the best information about the genetics and mechanisms of divergence because the factors that contribute to divergence can be directly studied.

Studies of incipient species often take population genetic and ecological genetic approaches to study the evolution of the traits contributing to reproductive isolation and divergence (Via 2001). Host races in a variety of systems have been investigated using these approaches: larch budmoths (Emelianov *et al.* 2004), soapberry bug (Carroll *et al*. 1997), the apple maggot and flowering dogwood flies (Feder and Bush 1991), goldenrod gall makers (Waring *et al.* 1990, Itani *et al.* 1998), walking sticks (Nosil *et al*. 2006), and holly leafminers (Scheffer and Hawthorne 2007). The shift of *Rhagoletis pomonella* from hawthorn to cultivated apple provides the most highly cited and historically well-documented example of sympatric host race formation (Bush 1994, Via 2001, Coyne and

Orr 2004), though new genetic studies provide an extra complication to this claim (see below). Detailed genetic and ecological research also makes pea aphids on alfalfa and clover a particularly rich model system for studying divergence (Brisson and Stern 2006), and in this dissertation, I investigate the process of divergence in pea aphids on alfalfa and clover.

Recent shifts in speciation research

It is finally being recognized that determining the relative likelihoods of allopatric and sympatric speciation may be an intractable goal due to theoretical and logistical issues. Neither strictly allopatric nor sympatric speciation may explain the diversity of most organisms. Partially as a consequence of the complexities in describing the geographic context of speciation, new classifications of modes of speciation and important new questions in speciation research have emerged.

Determining the geographic context of divergence is difficult for several reasons (Losos and Glor 2003, Coyne and Orr 2004). Most critically, the geographic ranges of diverging populations may change during the process of divergence so that speciation involves both sympatric and allopatric phases (Grant and Grant 1997, Losos and Glor 2003, Rundle and Nosil 2005). For example, divergence may occur as a byproduct of divergent selection in allopatry, and then speciation can be completed following secondary contact due to sexual selection or reinforcement (Rundle and Nosil 2005). New terms describing a mixed speciation mode have been proposed such as "allo-parapatric" and "allo-sympatric" speciation (Coyne and Orr 2004). There are considerable practical limitations to such a classification (Mallet 2005, Xie *et al*. 2007),

but it does reflect the need for a more complex description of the geography of speciation.

Another problem with the allopatric/sympatric dichotomy is highlighted by studies that show that the geographic origin of the genetic variation contributing to divergence may differ from the geographic context of the divergence of the populations. In the apple maggot fly, the genetic variation contributing to divergence may have arisen in allopatry, even if the divergence occurred in sympatry (Feder *et al.* 2003, Michel *et al*. 2007).

Finally, the geography of speciation may also be variable across the range of diverging taxa at any one point in time. This means the diverging populations may be sympatric in some parts of their range, and allopatric in other parts. The consequence of this geographic complexity to the process and outcome of speciation has received little theoretical and empirical attention, and this provides the motivation behind the Geographic Mosaic of Divergence (see Chapter 1). For all these reasons, the strict allopatric/sympatric dichotomy is too simple to describe the complexity of the process of speciation observed in nature (Schilthuizen 2000, Jiggins and Bridle 2004, Michel *et al*. 2007).

Recently, speciation models have shifted from the allopatric/sympatric dichotomy, and now modes of speciation are most often defined in terms of the evolutionary mechanism and forces leading to speciation (Schluter 2001, Via 2001). This does not mean that the geography of speciation is irrelevant because the type of evolutionary force is still dependent on the geographic context. The focus is now on understanding these forces and the genetics of speciation in their geographic context. For

instance, genetic drift gives rise to genetic incompatibilities in allopatry, while reinforcement only occurs in sympatry. Ecological speciation can occur in both allopatry and sympatry (Schluter 2001) and the challenge is to understand the genetic, behavioral, and ecological factors that influence the process of speciation and how the geographic context of divergence influences the process.

One of the ongoing controversies in ecological speciation is about whether divergence can occur with ongoing gene flow. A wealth of genetic data and new types of analysis has provided several clear instances of divergence-with-gene-flow (Noor and Feder 2006), and show how gene flow may not inhibit the evolution of reproductive isolation (Wu 2001, Wu and Ting 2004). Several lines of evidence have informed this controversy. Coalescent models can be used to infer past gene flow during speciation for already diverged taxa (Nielson and Wakeley 2001, Hey and Nielsen 2004), and provide strong genetic evidence that speciation occurred in the face of gene flow (Machado *et al.* 2002). Next, studies of host races show how populations can maintain major ecologically-based differences and still exchange genes at moderate rates. Ongoing gene flow contributes to introgression only at neutral loci while the loci involved in divergence differentiate in response to selection (Barton and Bengtsson 1986). Recent methods based on assignment methods provide strong evidence for and quantify rates of ongoing gene flow between host races (Manel *et al*. 2005). Evidence for different rates of gene flow among loci can then be inferred from F_{ST} outlier analysis (Beaumont and Nichols 1996, Storz 2005, Beaumont 2005). This provides compelling evidence for divergence-with-gene-flow (Wilding *et al.* 2001, Campbell and Bernatchez 2004, Emelianov *et al.* 2004, Savolainen *et al.* 2006, Scheffer and Hawthorne 2007). In my dissertation, I

investigate the genetics of divergence by quantify rates of ongoing gene flow and identify F_{ST} outliers to understand how pea aphids maintain major differences in host plant specialization that contribute to partial reproductive isolation in the face of ongoing gene flow.

Chapter 1: The geographic mosaic of divergence: geographic differences in genetic divergence and ecological specialization of pea aphids on alfalfa and clover in North America

Abstract

Ecological divergence may occur differently across geographically widespread sets of populations. Differences in natural selection among ecologically variable geographic locations may result in a range of specialization and patterns of divergence among populations in those locations, which may be called a Geographic Mosaic of Divergence. The interaction of local adaptation and gene flow among these geographic locations has the potential to either accelerate or retard the evolution of reproductive isolation, or prevent speciation from proceeding altogether. Important geographical differences were found across North America between divergent populations of pea aphids both in patterns of host plant specialization and genetic differentiation at molecular loci. In eastern North America, pea aphids (*Acyrthosiphon pisum*) from alfalfa (*Medicago sativa*) and clover (*Trifolium pretense*) were highly genetically differentiated and ecologically specialized on their host plants. In contrast, pea aphids on alfalfa and clover from several western locations were genetically indistinguishable from each other or from eastern aphids collected from alfalfa. Aphids from western clover showed some variation in ability to use clover. Clover specialization may be re-evolving in the West, or high directional migration from alfalfa onto clover could be homogenizing aphid populations in the West.

Introduction

Adaptation of organisms to different resources can lead to phenotypic and genetic divergence as populations become ecologically specialized in alternative environments (Futuyma and Moreno 1988, Jaenike 1990, Schluter 2001, Funk *et al*. 2002, Fry 2003). This ecological divergence can result in ecological speciation when adaptation to different environments causes assortative mating either directly, by influencing mate choice, or indirectly when habitat choice leads to mate choice (Schluter 2001 Turelli *et al.* 2001, Via 2001, Via 2002, Rundle and Nosil 2005). Although the dynamics of divergence are typically examined at a single location, host-associated populations may be geographically widespread and exhibit genetic population structure (Avise 2000, Thompson 2005). Then, the processes of divergence may occur at an array of locations simultaneously. Spatial variation in the process of divergence, including differences in natural selection, may cause specialization and genetic divergence to occur at different rates or to have different characteristics across geographic populations (Itami *et al.* 1998, Lu and Bernatchez 1999, Scriber 2002, Fernandez *et al.* 2005, Nosil *et al.* 2006), thus creating a mosaic of populations each responding to a common set of ecological challenges but with varying outcomes. The geographic heterogeneity of ecological divergence may inform observations of the process of specialization and ecological speciation in ways that could not be predicted without consideration of population structure or by empirical studies of single populations.

The outcome of ecological divergence for geographically widespread pairs of populations is dependent upon the interaction of three evolutionary processes acting simultaneously (Figure 1). First, genetic variation available for phenotypic evolution

may vary among locations because of differing population history and demography (Avise 2000, Thompson 2005). Differences in genetic variation among locations could cause populations under the same selective regime to evolve along different evolutionary trajectories. Second, because of local ecological differences, the targets and intensity of selection may also differ among locations, influencing the evolution of traits associated with ecological specialization and reproductive isolation (Thompson 2005, Schemske and Bierzychudek 2001). The strength of divergent natural selection for resource use may differ due to resource abundance or quality and the presence of alternate resources (How *et al.* 1993, Janz and Nylin 1997, Stireman and Singer 2003, Strauss and Karban 1998, Bernays 2001, Bernays *et al.* 2004, Nosil *et al.* 2006). Other selective forces may differ altogether such as those due to differences in climate (Scriber 2002) or ecological communities (Itami *et al.* 1998, Campbell 2003). These differing patterns of natural selection may not only lead to differences in the responses to selection in each location, but they may also cause population-specific changes in the genetic variation, influencing the first evolutionary process (Figure 2).

Third, the amount of connectivity among populations determines the degree to which these populations evolve independently or collectively (Rieseberg and Burke 2001, Morjan and Rieseberg 2004, Figure 2). Gene flow among geographically separated populations may differ, both in magnitude and reciprocity, due to population size or other ecological or environmental factors (Denno *et al*. 1996, Hanski and Simberloff 1997, Sork *et al.* 1999). Consequently, some populations concurrently adapting to a common resource may freely exchange alleles at key loci, increasing the possibility of similar patterns of adaptation, while others may be relatively isolated, possibly evolving

independently perhaps through different allelic substitutions or changes at different loci (Hoekstra and Nachman 2003, Colosimo *et al.* 2005, Kronforst *et al.* 2006).

These features of genetic variation, selection, and gene flow may interact such that geographically distributed sets of populations undergoing ecological divergence or speciation will vary in their degree of reproductive isolation and ecological specialization (Itami *et al.* 1998, Lu and Bernatchez 1999, Scriber 2002, Fernandez *et al.* 2005, Nosil *et al.* 2006). Thus, divergent populations in different geographic locations may occupy different positions along the divergence continuum, from polymorphic populations to locally adapted races to species (Jiggins and Mallet 2000, Dres and Mallet 2001). Collections of more or less independent populations may be distributed in a spatial patchwork, which I refer to as a "geographic mosaic of ecological divergence", which, if contributing to the evolution of reproductive isolation, may become a “geographic mosaic of speciation”.

While the pattern and process of the geographic mosaic have not previously been described, important geographic differences have been found in degree of divergence and local patterns of selection in several examples of ecological divergence and speciation. These examples include goldenrod gall makers (Waring *et al.* 1990, Itani *et al.* 1998), whitefish ecotypes (Lu and Bernatchez 1999), intertidal snail morphs (Cruz *et al.* 2004, Fernandez *et al.* 2005), walking sticks (Nosil *et al*. 2005) and the flowering dogwood fly and apple maggot fly (Feder and Bush 1991). Differences in gene flow between hybridizing species among geographic locations are also documented in mosaic hybrid zones of two *Ipomopsis* species (Campbell 2003, Aldridge 2005) and in tiger salamander species (Fitzpatrick and Shaffer 2004). The ecological mechanisms behind geographic

heterogeneity in divergence are as variable as the locations. To list only a few, there may be variability in resource abundance or quality and the presence of alternate resources (How *et al.* 1993, Fitzpatrick and Shaffer 2004), climatic variability (Itami *et al.* 1998, Scriber 2002), and the presence of parasitoids and predators (Itami *et al.* 1998). For example, the presence of an alternate resource niche in one location can serve as a mating site for both types of specialists, increasing the rate of hybridization (Fitzpatrick and Shaffer 2004).

The geographic mosaic of divergence has a theoretical basis in classical population genetics, ecology and more recent evolutionary biology (Hanski and Simberloff 1997, Wade and Goodnight 1998, Thrall and Burdon 2002, Thompson 2005). Wright described how evolutionary forces interact simultaneously to produce evolutionary change in subdivided populations, and how gene flow allows the joint evolution of interconnected populations (Wright 1969, Wade and Goodnight 1998). Metapopulation models show the importance of population size, the spatial arrangement of populations, and colonization and extinction of patches for shaping gene flow among locations (Hanski and Simberloff 1997, Thrall and Burdon 2002). Also, while the process of divergence described here does not involve coevolution, the theory of how interacting populations evolve across variable landscapes described by Thompson's "geographic mosaic of coevolution" (Thompson 1994, 1999, 2005) is directly relevant to how divergence may occur in spatially distributed sets of diverging populations. Specifically, the concept of "geographic selection mosaics" describes how divergent selection varies among locations (Thompson 2005). "Trait remixing" due to gene flow across locations, genetic drift, mutation, and population extinction and recolonization

continually shapes the genetic variation available for natural selection (Thompson 2005). Further investigations of how the three evolutionary processes of variation, selection, and gene flow interact to give rise to the genetic mosaic would greatly enhance the understanding of the process of divergence (Figure 1).

Pea aphids (*Acyrthosiphon pisum*) have become a model system for the study of ecological divergence onto different host plants (Via 2000, Coyne and Orr 2004, Brisson and Stern 2006), and they have recently expanded their geographic range (Sanderson 1900, Davis 1915). I investigate heterogeneity in genetic divergence and ecological specialization among pea aphid populations using alfalfa and clover in an array of locations across North America. I ask, are there differences in levels of genetic diversity among populations? Moreover, are there differences in gene flow between populations on the same resource across geographic locations? Next, I look for evidence of the geographic mosaic pattern of divergence by asking, are there differences in levels of genetic differentiation and ecological specialization among geographically widespread pairs of populations?

To assess differences among locations in genetic variation, I used five nuclear, non-coding, sequence-based (STS) markers and two allozymes. Next, I assessed population connectivity across locations for aphids on the same host plant to assess the genetic independence of geographic populations. In addition, I investigated population structure of aphids on each host plant and in each location. To look for evidence of the geographic mosaic of ecological divergence in pea aphids, at several geographic locations, I examined levels of genetic differentiation between aphids on each host plant and I measured aphid fecundity on each host plant as a measure of ecological

specialization. My approach combines the use of molecular genetic tools and the genetic analysis of traits associated with host use to estimate the extent of genetic and phenotypic divergence among geographically separated population pairs using alfalfa and clover as host plants.

Study system

Life history

Pea aphids are non-host alternating and cyclically parthenogenic, reproducing clonally during the summer and undergoing sexual reproduction in the fall (Eastop 1973, Blackman 1987). They overwinter in cold climates as diapausing eggs, and fundatrices hatch in the spring that give rise to parthenogenic lineages (Blackman 1987). Pea aphid clones can produce either wingless (apterous) or winged (alate) progeny. Apterous individuals achieve higher population growth rates than alates, and during the course of the summer there is clonal selection for host plant related performance, primarily on apterous individuals (Sandstrom 1996). Thus, it is important to measure fecundity of the apterous individuals when assessing pea aphid host plant specialization. In response to environmental stress and crowded conditions, alates are produced which are capable of dispersal (Weisser and Braendle 2001, Caillaud *et al*. 2002). This high mobility may allow significant gene flow among geographically isolated populations (McVean ***et al.*** 1999).

Pea aphid host-associated populations

The pea aphid complex includes three subspecies, and one of these subspecies, *Acyrthosiphon pisum pisum* (Harris), includes populations specialized on alfalfa (*Medicago sativa*) and red clover (*Trifolium pretense*; Eastop 1973, Blackman and Eastop

2000, McVean and Dixon 2002). Pea aphids and their host plants are thought to be native to Eurasia (Small 1996, Muller *et al.* 2003), though they have expanded their range and are now distributed worldwide in temperate climates. Pea aphids on alfalfa and red clover are ecologically specialized and/or genetically structured in parts of their range (in France; Frantz *et al.* 2006, in Sweden; Sandstrom 1996, in England; Ferarri *et al.* 2006, in New York; Via 1999). A different pattern was found by Leonardo and Muiru (2003) in California, suggesting differences among locations in specialization on alfalfa and clover (see below).

Introduction into North America

Pea aphids were introduced into North America at least 150 years ago, perhaps during the introduction of their host plants (Eastop 1973, Sanderson 1900, Petit 1905, Folsom 1909, Davis 1915). Red clover was introduced by European colonists to the eastern United States by 1663, and was the predominant forage and cover crop in the temperate East and Midwest until about 1950, though its use has subsequently declined (Taylor and Quesenberry 1996). In contrast, alfalfa was introduced to California around 1854 (Putnam 1997) and rapidly gained popularity as a forage crop in the western states (Westgate 1908), and later, following the development of regional plant varieties, into the Midwest and East (Folsom 1909). Presently, alfalfa and clover are often cultivated in mixtures or patchworks of adjacent fields in dairy production areas of the East and Midwest (Barnes *et al.* 1995). In the West and Great Plains, alfalfa is grown much more commonly than clover (Taylor and Quesenberry 1996).

The precise details about the number of introductions and source of the introductions of the pea aphid in North America are unknown. The earliest reports of the

pea aphid in the US are from around 1880 in the Midwest (Sanderson 1900), though pea aphids were documented across most of the northern tier of North America by the early 1900's (Davis 1915). Beginning in 1899, devastating outbreaks of pea aphids occurred in the Midwest and East on pea and red clover, though not on alfalfa (Folsom 1909, Davis 1915), suggesting clover aphids were already specialized at that point. Pea aphids show low mtDNA diversity in both New York (Barrette *et al.* 1994) and in Europe (Birkle and Douglass 1999) and no other study has compared variation between these locations, so the genetic consequences of the introduction are unknown. Because of the possibility of multiple introductions and range expansions during the invasion of the pea aphid, it is possible that there are differences in genetic variation and diversity across North America.

Pea aphid host races

Pea aphid populations from New York and Iowa have strong preferences for their host plant (Via 1999; Caillaud and Via 2000), and because pea aphids mate on their host plants, assortative mating tends to occur among aphid clones found together on the same host plant (Via 1999, Via *et al.* 2000). Pea aphids in Iowa and New York also show highly specialized performance on alfalfa and clover that is not substantially altered by experience (Via 1991). Migrants to the alternate host plant are therefore strongly disadvantaged which may contribute to increased assortative mating, further limiting the likelihood of cross-host plant gene flow (Via *et al.* 2000). Similarly, hybrids, which are viable but rare in nature, have lower fitness on both parental host plants due mostly to impaired resource use rather than to intrinsic incompatibilities (Via *et al.* 2000). In a sample from New York, host plant specialization of pea aphids on alfalfa and clover was

influenced by several complexes of nuclear genes, located on all four of the pea aphid linkage groups (Hawthorne and Via 2001). In the only other analysis of host plant specialization of pea aphids in North America, Leonardo and Muiru (2003) reported that some pea aphids collected from white clover in California were more specialized on that crop whereas aphids collected from other plants were more generalized. Because the pattern of specialization in California was so different than that seen in New York and Iowa, this suggests that there could be variability in ecological specialization across locations, and thus the pattern of the geographic mosaic of ecological divergence in pea aphids.

Methods

Aphid collections

Pea aphids were collected from ten locations across North America between 1997 and 2003 during mid-summer (Table 1). At most locations, aphids were collected from two or more closely adjacent fields of alfalfa (*Medicago sativa*) and red clover (*Trifolium pratense*). In California, aphids were collected from white clover (*Trifolium repens*), and in Cortez, Colorado, where no red clover was found, aphids were collected from alfalfa. Pea aphids were sampled in widely dispersed locations across fields to avoid sampling the same clone. Aphids from five locations (Maryland, New York, Iowa, Washington, and British Columbia) representative of major alfalfa and/or clover growing regions (hereafter the "subset" locations) were maintained as clonal lineages on their host plant of origin for genetic analysis. Samples of each clone were also frozen (-80 °C) for DNA analysis. Pea aphids collected from the five other geographic locations (California,

Oregon, Northern Colorado, Southern Colorado, and Michigan) were preserved in ethanol immediately. Genomic DNA was extracted from aphids from all 10 geographic locations using the DNeasy Tissue Kit (QIAGEN).

STS markers

Genetic variation in pea aphids on alfalfa and clover across North America was assessed using nuclear, codominant sequence markers. Ten aphids from each host plant at each location (Table 1) were genotyped for five codominant non-coding sequence tagged site (STS) markers (Ic380, IIIc598, IIIc650, IIc870, and Ic901) (Table 2). Each marker was amplified in 25 uL PCRs containing 1 uL DNA, 1X PCR buffer, a marker-specific $MgCl_2$ concentration (Table 2), 0.1 mM dNTPs, 0.1 mM of each primer, and 1 unit of Taq DNA Polymerase (New England Biolabs). Amplifications were performed on a MJ Research thermocycler: 2 minutes at 94 °C, followed by 35 cycles of 20 seconds at 94 °C, 20 seconds at the annealing temperature (Table 2), and 30 seconds at 72 °C, with a final 2 minutes at 72 °C.

A combination of direct sequencing, single strand conformation polymorphism (SSCP) analysis, and sequencing of SSCP bands was used to obtain sequence data and to resolve haplotype phase. Genotyping strategy differed among markers (Table 2). One marker, IIIc598, was directly sequenced and haplotype phase was resolved using FASTPHASE (Scheet and Stephens 2006). This program uses information from known haplotypes and population information to reconstruct haplotypes from unphased sequence data. For two markers, Ic380 and IIc870, most haplotypes were resolved using SSCP analysis. For these two markers, amplification products were evaluated on 35 x 45 cm vertical gels, with 0.4X MDE Gel as a matrix (SSCP Analysis protocol, BioWhittaker).

Gels were run at 4 Watts and 4 °C for 20-72 hours, silver stained (Silver Sequence, Promega), and scored manually. Multiple examples of each SSCP haplotype from each gel were sequenced to confirm haplotype identity. Briefly, bands taken from the gel with a sterile pipette tip and placed into a PCR cocktail. The amplification product was purified (ExoSAP-IT, USB Corp.) and sequenced (BigDye protocols, ABI) in both directions using the same primers and conditions used in the original amplification. Several haplotypes could not be differentiated by SSCP analysis, so individuals with these SSCP haplotypes were directly sequenced.

Finally, for the other two markers, IIIc650 and Ic901, all samples were directly sequenced, and cis-trans phase was resolved for ambiguous genotypes using three methods. First, haplotype phase was resolved by inference (Clark 1990). Then, for haplotypes found by inference but not also found in homozygotes, bands were individually excised from SSCP gels and sequenced to confirm haplotype identity. Third, for individuals heterozygous for indels or an internal and polymorphic microsatellite, haplotypes were determined using the methods described by Flot *et al.* (2006). Briefly, diploid sequencing yields double peaks due to the superposition of two sequences of different lengths. The two haplotypes can be deduced by manually lining up the forward and reverse sequence traces.

Allozymes

To increase the number of markers for Bayesian multilocus analysis of population structure, aphids from the five "subset" geographic locations were also genotyped using two allozyme markers. Pea aphids were genotyped for *Pep-GL* (EC 3.4.13.11) and *Pep-LGG* (EC 3.4.11) using cellulose-acetate gel electrophoresis (Herbert and Beaton 1989).

Previous surveys using these markers in New York showed that several alleles are strongly host associated, and both allozymes show high genetic differentiation between aphids on alfalfa and clover (Via 1999).

Genetic diversity

Four measures of genetic diversity at the five STS markers were compared across locations for aphids on alfalfa and on clover using the full dataset. Because diversity measures are affected by sample size, comparisons were made for the two major regions (i.e. the "East" and the "West", see F_{ST} and STRUCTURE results for justification, below). Allelic richness is the average number of alleles per locus in a sample, and is more sensitive for detecting the effects of short, severe bottlenecks than heterozygosity. Allelic richness was estimated in FSTAT (Goudet 2001) using ElMousadik and Petit's (1996) method to account for sample size. Observed heterozygosity (H_o), the observed proportion of heterozygotes, and Nei's gene diversity or expected heterozygosity (H_s), the probability that two alleles at a random locus differ in a sample, was each calculated in FSTAT. For aphids on alfalfa and on clover separately, I tested for differences between the populations from the East and the West for these three diversity measures in FSTAT using 1000 permutations. Finally, nucleotide diversity (π), the average number of nucleotide differences among sequences in a sample, was estimated in SITES (Hey and Wakeley 1997) for each marker, and then the mean and standard error of nucleotide diversity across markers were calculated.

F_{ST} and exact tests

To measure pea aphid population genetic structure on host plants and across geographic locations at several scales, I calculated F_{ST} with five STS markers using the methods of Weir and Cockerham (1984) in the program Genepop on the Web (Raymond and Rousset 1995). To measure the genetic independence of populations of aphids on the same host plant across locations, F_{ST} was calculated across geographic locations for the two host plants separately, which reflect genetic structuring across geography. Next for each host plant separately, pairwise F_{ST} was calculated between each pair of geographic populations. To determine if host plant-based divergence differs among geographic locations, F_{ST} between aphids on alfalfa and clover was calculated at each geographic location. These F_{ST}'s between host plants may reflect divergence due to resource use. Standard errors of F_{ST} were estimated by jackknifing over loci. Differences between estimates were tested based upon 95% confidence intervals generated by bootstrapping over loci.

Because exact tests of population differentiation are more sensitive for detecting genetic differentiation than F_{ST} for small sample size, exact tests were performed to determine if there were the same allelic distributions at each location (Raymond and Rousset 1995). Analyses were done in FSTAT between each geographic population using 1000 permutations and a Bonferroni adjustment for multiple tests.

STRUCTURE

To further examine genetic structure of aphids from alfalfa and clover in each geographic location, Bayesian analyses were used to assign individuals based upon their multilocus genotype into genetically distinct groups (STRUCTURE 2.0, Pritchard *et al.*

2000). To determine the number of populations represented by the data (called "K"), I used the admixture model without population of origin information. Under this model, the allele frequencies of the parental populations (called "clusters") are determined by Bayesian analysis using only individuals' multilocus genotypes. Clusters are constructed to maintain Hardy-Weinberg equilibrium and linkage equilibrium within genetic populations. Individuals inherit a fraction of their genome from each inferred cluster, and have a proportion of ancestry derived from each cluster (q), which ranges from 0 to 1. Individuals with mixed ancestry have intermediate values of q. Because multilocus analysis is more effective with more markers, I analyzed the subset dataset with 5 markers plus two allozymes for the five representative locations. The model was run for K = 1-6 with 5 independent runs for each K. Models were run for a burn-in period of 400,000 MCMC iterations and a data collection period of 1,000,000 MCMC iterations using the admixture model, correlated allele frequencies and no prior information (Falush *et al.* 2003). Two methods to evaluate K were used: the maximal log probability of the data, Pr (X|K) (Pritchard *et al.* 2000) and the modal value of ΔK, which is based on the rate of change in the log probability of the data in consecutive runs for each K (Evanno *et al.* 2005).

STRUCTURE assumes that each inferred cluster is in Hardy-Weinberg Equilibrium (HWE), and that markers are in linkage equilibrium. Possible deviations from HWE were determined using probability tests for each cluster and markers at α = 0.05 (following Bonferroni correction for multiple tests) using Genepop on the Web (Raymond and Rousset 1995). Linkage disequilibrium within each cluster defined by STRUCTURE was also tested for each pair of loci in Genepop on the Web.

Ecological specialization of aphids on alfalfa and clover

Ecological specialization of individual aphid clones from four of the five "subset" locations (New York, Maryland, Washington, and British Columbia) was assessed by measuring pea aphid fecundity on both alfalfa and clover using a reciprocal transplant design. Aphid fecundity on each host plant is correlated with fitness (Via 1999). Aphid performance was measured using the protocol described in Via and Hawthorne (2002). The number of offspring produced in an apterous aphid's first nine days of adult life was measured, which provides a good estimate of aphid fecundity (Via *et al.* 2001). For each geographic location, 10-48 pea aphid clonal genotypes were tested in replicate (Table 1). All experiments used the medium red variety of clover and Oneida VR variety of alfalfa, which are the common varieties grown in New York. Experiments were conducted in controlled environmental chambers at 21.5 °C/ 15 °C and a light-dark cycle of 16L/8D to replicate summer conditions.

Aphid fecundity was analyzed using the MIXED procedure in SAS (Littel *et al.* 1996, SAS Institute 2001), with "home plant" and "test plant" as fixed effects, and "aphid clone" as a random effect. Population means are reported as least-squares means, while clone means were estimated as a best linear unbiased predictor (BLUP). Each geographic population was tested separately without a common control set of clones in each population. For this reason, some care should be taken when populations are compared.

Results

Genetic diversity

STS markers contained several types of variation including SNPs, indels, and microsatellites (Table 3). While the shortest sequence (Ic901) only had 2-4 variable sites and 3 to 4 haplotypes in each population, other markers were more variable. For instance, IIc870, the longest sequence, had 10-20 variable sites and 5-9 haplotypes in each population. Genetic diversity of aphids was consistently higher in the West than in the East for samples from both alfalfa and clover (Table 4). For aphids on alfalfa, there were significant differences according to permutation testing in FSTAT between the eastern and western populations in allelic richness (AR, East = 2.937; AR, West = 3.468; $p = 0.001$) and gene diversity (H_s, East = 0.494; H_s, West = 0.595; $p = 0.001$). For aphids on clover, there were also significant differences according to permutation testing in FSTAT between the eastern and western population in allelic richness (AR, East = 2.858; AR, West = 3.412; $p = 0.005$), observed heterozygosity (H_o, East = 0.467; H_o, West = 0.661; $p = 0.012$), and gene diversity (H_s, East = 0.517; H_s, West = 0.647; $p = 0.005$). Nucleotide diversity was also higher in the West compared to the East for most markers (Table 3 and 4).

F_{ST} and exact tests

At the regional scale across North America, pea aphids showed significant genetic structuring of allele frequencies between eastern and western clover (F_{ST} across locations = 0.14), and to a much lesser extent for alfalfa (F_{ST} across locations = 0.03). This indicates significant genetic structuring across geographic locations for populations using

the same host plant ($p<0.01$). For aphids on both alfalfa and clover, pairwise F_{ST} between locations within each region were low (i.e. within the East and the West), and none of the locations were significantly differentiated according to Exact Tests of population differentiation (Table 5). Within each region, the pairwise F_{ST} between locations ranges from -0.023 to 0.035 for aphids on alfalfa and from -0.022 to 0.195 for aphids on clover (Table 5). The highest pairwise F_{ST} within a region was between Michigan and Maryland aphids on clover, two of the more distantly located populations within either region. Pairwise F_{ST}s for those comparisons between locations in the East and the West tended to be higher than within-region comparisons, and were generally higher for aphids on clover than for alfalfa (Table 5). Within both alfalfa and clover, several populations between the East and the West were significantly genetically differentiated according to exact tests (Table 5).

Molecular divergence measured by F_{ST} between host plants varied across geographic locations. Genetic differentiation was higher in the eastern populations compared to the western ones (Figure 3). There was also variation in levels of differentiation within each region. In the East, Maryland showed the highest F_{ST} between aphids on alfalfa and clover, and F_{ST} in Michigan was moderate.

STRUCTURE

Analysis of these molecular genetic data using a Bayesian method (STRUCTURE) was consistent with the F_{ST} analysis. Both methods for determining the number of genetic populations, or clusters, indicated that the data represent two clusters (Figure 4). The highest L(K) and the mode of ΔK were both for K=2 indicating two genetic groups. Under the admixture model without population of origin information,

there was a correspondence between the inferred genetic clusters and host plant of origin in the three "eastern" locations Iowa, New York, and Maryland (Figure 5). In the East, aphids from alfalfa belong to one cluster, while aphids from clover belong to the other (Figure 5). In the West, aphids from alfalfa and clover belong to the same cluster, which was the same as the eastern alfalfa cluster. Thus, there was genetic differentiation between clover aphids from the East and West, but there were not any differences across locations for aphids on alfalfa.

Within each of the clusters defined by STRUCTURE, there were significant departures from HWE for two markers (IIIc598 and Pep-LGG), and in the Alfalfa-Western clover cluster for another marker (IIc870, $p < 0.004$). In addition, there were significant departures from linkage equilibrium for three pairs of loci (IIc870-IIIc650, Ic380-IIIc650, and IIIc650-Pep-LGG). The first two pairs of loci are not located on the same linkage groups, and Pep-LGG has not been located on the linkage map (Hawthorne and Via 2001, West, unpublished data). However, STRUCTURE is robust to some deviations from assumptions.

Ecological specialization of aphids on alfalfa and clover

In the East, aphid populations on their home plant had higher fecundity than aphids from the other host plant (Figure 6A). The interaction between host plant of origin and test plant was significant in both New York ($F = 164.34$, $p < 0.0001$) and in Maryland ($F = 626.55$, $p < 0.0001$). Aphids from alfalfa had higher fecundity on alfalfa than they did on clover in New York ($F = 79.69$, $p < 0.001$) and in Maryland ($F = 360.98$, $p < 0.001$), and aphids from clover had higher fecundity on clover than they did on alfalfa in New York ($F = 87.73$, $p < 0.001$) and in Maryland ($F = 260.63$, $p < 0.001$).

In the West, aphid populations from alfalfa showed similar patterns of host-plant based fecundity as those in the East. The interaction between host plant of origin and test plant was significant in British Columbia ($F = 78.57$, $p < 0.0001$) and in Washington ($F = 18.84$, $p < 0.0001$), and aphids from alfalfa had higher fecundity on alfalfa than they did on clover in British Columbia ($F = 337.88$, $p < 0.001$) and in Washington ($F = 269.91$, $p < 0.001$). However, aphids from western clover had as high or higher fecundity on alfalfa as they did on clover (Figure 6A). In British Columbia, aphids from clover showed no significant difference in their fecundity on alfalfa and clover ($F = 0.29$, $p = 0.59$), and in Washington, aphids from clover had higher fecundity on alfalfa ($F = 10.45$, $p = 0.004$). Furthermore, western aphids from clover showed a range of fecundities on alfalfa, and the mean fecundities on each plant for each aphid clone overlapped with the fecundities of aphids from alfalfa (Figure 6B). Thus, western aphids from alfalfa and clover were phenotypically more similar and less ecologically differentiated from one another. In addition, the aphids from clover in the West tended to be less specialized than eastern aphids from clover. The degree of ecological specialization between aphids collected from the two host plants thus varied among geographic locations.

Discussion

Diverging populations may be geographically dispersed when incipient species experience a range expansion or already widespread populations similarly encounter alternate resource types. Geographic differences among diverging populations in genetic variation and local features of natural selection coupled with some genetic isolation results in a geographic mosaic of divergence, and possibly also a geographic mosaic of speciation (Figure 1). This variation in the degree of divergence among the geographic

sets of divergent populations indicates that populations across locations are at different stages of the processes of divergence and speciation (Jiggins and Mallet 2000, Dres and Mallet 2002).

The first goal of the study was to determine if there are differences in genetic diversity among different geographic populations using the same resource, which could contribute to a geographic mosaic. Surprisingly large levels of diversity for an introduced species were found, especially since there was little mtDNA variation found in previous studies (Barrette *et al.* 1994, Birkle and Douglass 1999). I found that aphids on alfalfa and clover showed significantly more genetic diversity in the western than in the eastern populations. Differences in genetic variation available for natural selection among populations may contribute to the geographic mosaic, constraining specialization or divergence or allowing evolution to proceed in different directions in different locations (Futuyma *et al.* 1995). Measuring genetic variation at markers linked to genomic regions involved in divergence or at the genes themselves, if they can be identified, could further detail the genetic changes that occur during divergence and their geographic context (Feder *et al.* 2003, Colosimo *et al.* 2005).

Differences between the East and West in selection or colonization history could explain these differences in genetic diversity. Introductions may have occurred from genetically distinct source populations and there may have been multiple introductions. If aphids on alfalfa were introduced to the West first, then a reduction of diversity in the East could be the result of range expansion. Pea aphids on clover were reported early in the Midwest and East (Folsom 1909, Davis 1915), so clover specialists may have a different colonization history. Comparisons of North American and European

populations could reveal the detailed history of the introductions, which could help explain the observed difference in genetic diversity.

The second goal of this study was to determine if there were differences in gene flow among aphid populations on the same resource among geographic locations. Within the regional sampling locations, there was no significant population structure for aphid populations on the same host plant, indicating that either gene flow is high or shared ancestry is recent, and therefore collective evolution is possible and likely has occurred within a regional scale. Between regions, there was significant genetic structure across North America between the East and West. Genetic differentiation of pea aphids from the same host plant across their range was greater for aphids on clover than for aphids on alfalfa (Table 5). Aphids from alfalfa were genetically similar across North America, while aphids from clover were genetically distinct between the East and the West (Figure 4). This could be explained by the differences in the ecology of the two host plants. Alfalfa is more abundant than clover especially in regions of the arid West (Barnes and Sheaffer 1995). Higher densities of alfalfa across North America can sustain larger aphid populations and yield more potential migrants, and larger patch sizes allows higher gene flow among locations (Kareiva 1983, McCauley 1991). Aphid clones also vary substantially in their tendency to produce alates (Lamb and MacKay 1979, Bommarco and Ekbom 1996, Weisser and Braendle 2001), so there could be genetically-based differences among aphid populations in migratory tendency. If genetic similarity of alfalfa populations is due to gene flow and not shared ancestry, the high gene flow across locations for aphids on alfalfa could allow important variation for alfalfa specialization to spread easily among populations across their North American range.

Finally, I investigated the pattern of the geographic mosaic by measuring the degree of genetic divergence and ecological specialization of aphids on alfalfa and clover across North America. I found that pea aphids have a pattern of divergence consistent with a geographic mosaic of divergence. Ecological divergence in pea aphids involves genetic differentiation due to the accumulation of genetic differences between populations (due to divergent selection, drift, and/or reduced gene flow), and increased specialization, specifically the ability of aphids to use their host plants. I found that in North America, pea aphids on alfalfa and clover are geographically variable in both genetic differentiation at molecular loci and demographic measures of host plant specialization. In the East, pea aphids from alfalfa and clover are highly genetically differentiated and ecologically specialized on their host plants, similar to results found in previous studies (Via 1991, Via 1999, Via *et al.* 2000). In contrast, aphids on alfalfa and clover in the West were genetically indistinguishable, with the aphids from clover more similar to the aphids from alfalfa in host plant specialization and in genetic structure.

The geographic mosaic described here suggests that pea aphid populations in different locations have different histories and are at different stages of ecological divergence. Populations in the East may be much farther along the process of divergence, while those in the West are at a much earlier stage or are subject to higher rates of introgression. In the West, greater phenotypic than genetic divergence is consistent with very recent or seasonal divergence. Investigation of rates of hybridization between aphids on alfalfa and clover in each location are required to determine if populations represent host-associated populations or host races (Dres and Mallet 2002). Nevertheless, the difference between the molecular genetic and phenotypic patterns is

characteristic of populations in early stages of ecological speciation because divergence occurs first at the loci contributing to variation in the traits under divergent selection and is measurable first in markers linked to those loci (Charlesworth *et al.* 1997, McKay and Latta 2002). Only when there is further reproductive isolation and sufficient time for drift-mutation processes to act will there be similar divergence in the remainder of the genome (Wu 2001).

Possible causes of the geographic mosaic of divergence in pea aphids

There are several, non-mutually exclusive hypotheses to explain the geographic mosaic pattern in pea aphids that involve (1) differences in local ecological conditions and natural selection or (2) nonuniform genetic variation due to aphid colonization history:

Western clover aphids may be less specialized and more genetically similar to aphids on alfalfa because of reduced natural selection for clover specialization and host plant abundance. Natural selection for clover specialization may be stronger in the East than the West because of the greater abundance of clover grown in the East (Barnes and Sheaffer 1995). Furthermore, selection could be altered by differences in host plant traits in different regional clover varieties (Taylor and Smith 1995). Also, host plant persistence influences insect dispersal (Peterson and Denno 1998), and higher gene flow can occur among insect populations where host plants are not available year-round (Denno *et al.* 1996). In some regions of the West, clover is grown predominantly for seed, while alfalfa is grown for both seed and forage. When clover is grown for seed, plants dry out in late summer before being harvested and become unsuitable for aphids to mate on in the fall. In regions where clover is grown for seed, clover specialists may die

out, or disperse to other host plants and there may be selection for use of alternate host plants. Any of these could decrease selection for clover specialization. Detailed ecological study of the pea aphids and their host plants could reveal differences between regions in agricultural practices such as irrigation, crop rotations, harvesting, plant varieties, or alternate host plants that could affect host plant use.

Not only may selection for clover specialization differ, but high directional gene flow from alfalfa onto clover could lead to gene swamping, where alleles for clover specialization are lost (Lenormand 2002). Host plant patch size influences migration rates of insect specialists (Kareiva 1983, McCauley 1991). The abundance of alfalfa in the West and large population sizes of alfalfa specialists may cause migration rates from alfalfa to clover to be very high due to demographic causes. Directional migration rates may vary because aphids on alfalfa and clover could differ in their migratory tendency. There is genetic variation among aphids in how readily the production of winged forms is induced in response to crowded conditions (Lamb and MacKay 1979, Bommarco and Ekbom 1996, Weisser and Braendle 2001). If aphids on alfalfa have greater migratory tendency than aphids on clover, there could be higher gene flow from alfalfa onto clover. Thus, even if selection favors clover specialization in the West, gene flow may limit the evolution of clover specialization.

Differences in genetic variation due to the introduction of the aphids into North America could constrain the evolution of clover specialization in the West but not the East. If only alfalfa specialists colonized the West, western aphids on both alfalfa and clover may be the descendents of these colonists. The aphids seen on clover in the West may be the result of a secondary colonization of clover by aphids from alfalfa.

Populations of alfalfa specialists from all geographic locations are likely to have some genetic variation for clover use due to ongoing hybridization, and the current levels of clover specialization could be due to only a few generations of selection (Frazer 1972). Thus, the genetic similarity of aphids in the West on alfalfa and clover could be because western clover aphids are very recent colonists from alfalfa. This could be tested experimentally: western alfalfa specialists could be introduced onto clover for several generations to determine if a similar level of clover specialization could be achieved after only a few generations.

Selection on variation present in natural populations can repeatedly give rise to ecological specialization in different geographic locations. In another model system for ecological speciation, parallel evolution of the benthic and limnetic forms of sticklebacks has been shown to be due to selection on standing genetic variation present in natural population of the "ancestral" populations resulting in similar yet independent evolution of ecologically important phenotypes (Colosimo *et al.* 2005). Thus, parallel evolution of specialization can arise rapidly because it can involve sorting of ancestral variation in new locations.

Determining the cause of lower genetic and phenotypic differentiation of pea aphids in the West compared to the East requires additional study of the genetics and ecology of host plant specialization. However, it seems unlikely that clover specialists were unable to colonize the West, given their success in colonizing new continents, so differences between regions in the strength of divergent selection and hybridization seem the most likely cause.

Whatever the reason for the different degrees of divergence in pea aphids in North America, whether the geographic mosaic of divergence becomes a geographic mosaic of speciation in aphids will likely depend upon the interaction of gene flow between locations and local processes within geographic locations. It is possible that aphids on alfalfa and clover may continue to diverge and become separate species. There could also, however, be a long-term maintenance of variation among locations in divergence, or perhaps an erosion of differentiation. If divergence in the West is constrained by a lack of genetic variation for clover specialization, additional immigration and gene flow from eastern clover types could accelerate divergence in the West. In this case, geographically distributed populations using clover would become more similar to each other and the geographic mosaic of speciation would be transient. On the other hand, if regional ecological differences continue to affect divergence in pea aphids, then divergence and speciation might continue to proceed differently in the East and West. Measuring rates of gene flow across locations and studying the ecology of the pea aphids and their host plants in the West may reveal which scenarios are most likely for the future of the pea aphids.

Consideration of the genetic and ecological causes and consequences of a geographic mosaic significantly enriches our understanding of the process of adaptive divergence in contemporary populations. Analysis of geographic mosaics of divergence should motivate both the analysis of geographically isolated populations and the processes that determine the degree to which they evolve independently or collectively.

Tables

Table 1. Pea aphids were collected from alfalfa and clover in ten locations across North America. At each location, aphids were collected from 0-4 fields of each host plant. Geographic locations from which aphids were collected live and used for allozymes and/or genetic analysis of fecundity (the "subset" locations) are indicated with *. 10 aphids were collected from each host plant in each location for use in molecular genetic analysis. Additional aphids were collected for measuring fecundity.

Location name	Coll-ection Date	Latitude	Longitude	No. alfalfa fields sampled	No. clover fields sampled	No. aphids from alfalfa for fecundity	No. aphids from clover for fecundity	No. replicates per clone for fecundity
Middleton, MD*	2001	39° 28' 42" N	77° 31' 21" W	3	3	48	46	3
Ithaca, NY*	1999	42° 26' 38" N	76° 27' 57" W	1	2	10	11	3
Michigan	2002	42° 24' 00" N	85° 24' 00" W	1	1	-	-	-
Iowa City, IA*	2002	41° 44' 25" N	91° 29' 01" W	3	3	-	-	-
N. Colorado	1997	40° 01' 37" N	105° 15' 04" W	4	2	-	-	-
Cortez, CO	2003	37° 20' 13" N	108° 48' 02" W	2	0	-	-	-
Ontario, OR	2002	43° 45' 48" N	117° 02' 09" W	2	3	-	-	-
E. Washington*	1999	46° 58' 04" N	119° 02' 49" W	2	2	30	12	2
Summerland, BC*	1998	49° 01' 59" N	119° 27' 00" W	3	3	26	20	4
California	2003	39° 44' 44" N	122° 11' 14" W	3	3	-	-	-

Table 2. Primer sequences, PCR conditions, mapping information, and genotyping methods for pea aphid STS markers.

Marker*	Primer sequence (5'-3')	Sequence length (bp)	[MgCl2] (mM)	Ta (°C)	Genotyping method
Ic380	L: ACTTACAAGTCTAATTTTGAAG	354	3.7	48	SSCP
	R: GTAATGTCACTATTAGAAG				
IIIc598	L: TGTGTACCTACACGGCAAAG	328	3.0	58	Sequence
	R: ATGGCAGCGGTGAGTGGCGATG				
IIIc650	L: GAGGTTTTCATCATTTTTGCCTA	568	1.5	56	SSCP & Sequence
	R: GCGTACATGCAGCAGTATCA				
IIc870	L: GTACTCTGGTACTCAATGAAACC	751	1.5	52	SSCP
	R: GCATTGTGAAACGACGCAAACG				
Ic901	L: ACGGACAGCTACTCAATCGTTAG	250	1.5	60	SSCP & Sequence
	R: CATTCCACGAGACTTCTAGGTCGT				

*Marker naming convention is linkage group, marker type (c=codominant), and size of original AFLP band.

Table 3. Summary of nucleotide variation and genetic polymorphism of the 5 STS markers for North American pea aphids on alfalfa and clover. Aphids in each region and on the two host plant had different numbers of segregating sites, and contained several types of variation including SNPs, indels, and microsatellites. Nucleotide diversity (π) is estimated for aphids from alfalfa and clover in each region in SITES (Hey and Wakeley 1997).

Locus	Host plant	Region	No. of sequences	Sequence length (bp)	Segregating sites (all)	SNPs	Indels	Microsats	No. haplotypes	Nucleotide diversity (π)
Ic380	Alfalfa	East	72	354	8	6	2	0	5	0.00608
		West	113	354	9	7	2	0	6	0.00638
	Clover	East	76	354	5	5	0	0	3	0.00526
		West	81	354	8	6	2	0	5	0.00595
IIIc598	Alfalfa	East	72	328	11	11	0	0	11	0.00345
		West	114	328	7	7	0	0	14	0.00406
	Clover	East	78	328	8	8	0	0	13	0.00489
		West	82	328	8	8	0	0	15	0.00470
IIIc650	Alfalfa	East	71	568	11	10	0	1	6	0.00598
		West	116	568	11	10	0	1	5	0.00803
	Clover	East	78	568	11	10	0	1	3	0.00075
		West	84	568	12	11	0	1	6	0.00835
IIc870	Alfalfa	East	71	751	10	8	2	0	5	0.00234
		West	114	751	16	14	2	0	8	0.00338
	Clover	East	71	751	20	18	2	0	9	0.00628
		West	78	751	16	14	2	0	8	0.00350
Ic901	Alfalfa	East	72	250	4	4	0	0	3	0.00129
		West	107	250	2	2	0	0	3	0.00155
	Clover	East	78	250	3	3	0	0	4	0.00492
		West	78	250	2	2	0	0	3	0.00247

Table 4. Genetic variation and polymorphism of North American pea aphids on alfalfa and clover. Allelic richness, observed heterozygosity (H_o), and gene diversity or expected heterozygosity (H_s) were calculated in FSTAT using the five STS markers. Nucleotide diversity (π) is estimated for aphids from alfalfa and clover in each region using SITES, and the mean and standard error is shown across five STS markers.

Collection host plant	Collection region	Allelic richness	Observed heterozygosity (Ho)	Gene diversity (Hs)	Nucleotide diversity (π)
Alfalfa	East	2.937	0.536	0.494	0.00383 ± 0.00096
	West	3.468	0.625	0.595	0.00468 ± 0.00114
Clover	East	2.858	0.467	0.517	0.00442 ± 0.00095
	West	3.412	0.661	0.647	0.00499 ± 0.00132

Table 5. Genetic differentiation was measured across geographic locations for pea aphids collected from alfalfa (A) and from clover (B). Pairwise F_{ST} between populations of pea aphids are in the bottom diagonal (Weir and Cockerham (1984) estimates from GENEPOP ON THE WEB). Exact tests of population differentiation were calculated in FSTAT using 1000 permutations. Significant differentiation is indicated by a "*" in the top diagonal using a Bonferroni adjustment for multiple tests. All analyses used 5 STS markers.

A. Alfalfa

		West						East			
		CA	BC	WA	OR	CO	SCO	IA	MI	NY	MD
West	CA		NS	NS	NS	NS	NS	NS	NS	*	NS
	BC	0.021		NS	NS	NS	NS	NS	NS	*	NS
	WA	0.023	-0.004		NS	NS	NS	*	NS	*	NS
	OR	0.009	-0.018	0.014		NS	NS	*	NS	*	NS
	CO	0.031	0.023	0.035	0.032		NS	NS	NS	NS	NS
	SCO	-0.011	-0.023	0.008	-0.022	0.030		NS	NS	*	NS
East	IA	0.077	0.057	0.094	0.063	0.009	0.053		NS	NS	NS
	MI	0.004	0.010	0.001	0.030	-0.003	-0.011	0.033		NS	NS
	NY	0.103	0.064	0.063	0.075	0.013	0.075	0.023	0.017		NS
	MD	0.087	0.052	0.081	0.060	0.025	0.033	0.017	0.013	0.010	

B. Clover

		West					East			
		CA	BC	WA	OR	CO	IA	MI	NY	MD
West	CA		NS	NS	NS	NS	*	*	*	*
	BC	0.057		NS	NS	NS	*	NS	*	*
	WA	0.042	0.040		NS	NS	*	NS	*	*
	OR	0.010	0.014	0.001		NS	NS	*	NS	*
	CO	-0.001	0.028	0.025	-0.022		NS	NS	NS	NS
East	IA	0.291	0.188	0.266	0.223	0.215		NS	NS	NS
	MI	0.219	0.061	0.214	0.151	0.120	0.031		NS	NS
	NY	0.219	0.080	0.189	0.141	0.160	-0.012	0.003		NS
	MD	0.544	0.434	0.546	0.448	0.493	-0.002	0.195	0.076	

Figures

Figure 1. Summary of the three evolutionary processes that interact concurrently to give rise to the pattern of the geographic mosaic of divergence

Three evolutionary processes:

1. Genetic variation for traits important to ecological specialization and reproductive isolation differs among locations.

2. The targets and intensity of selection may differ among locations.

3. Gene flow among geographically separated populations may differ in magnitude and reciprocity.

The Pattern:

Geographic sets of divergent populations vary in degree of ecological specialization and genetic differentiation across their range.

Figure 2. Populations diverging onto different resources that are spatially widespread may form a geographic mosaic of divergence. At each geographic location, the diverging populations, those sympatric populations undergoing divergent selection for specialization on alternate resources, may experience different rates of gene flow. Gene flow among the geographic populations, those spatially and genetically isolated populations on the same resource, may vary across their range. Some locations may be more isolated, and evolving largely independently (Location 1), while populations in other locations may be connect by high gene flow, and be evolving collectively (Location 2 and 3).

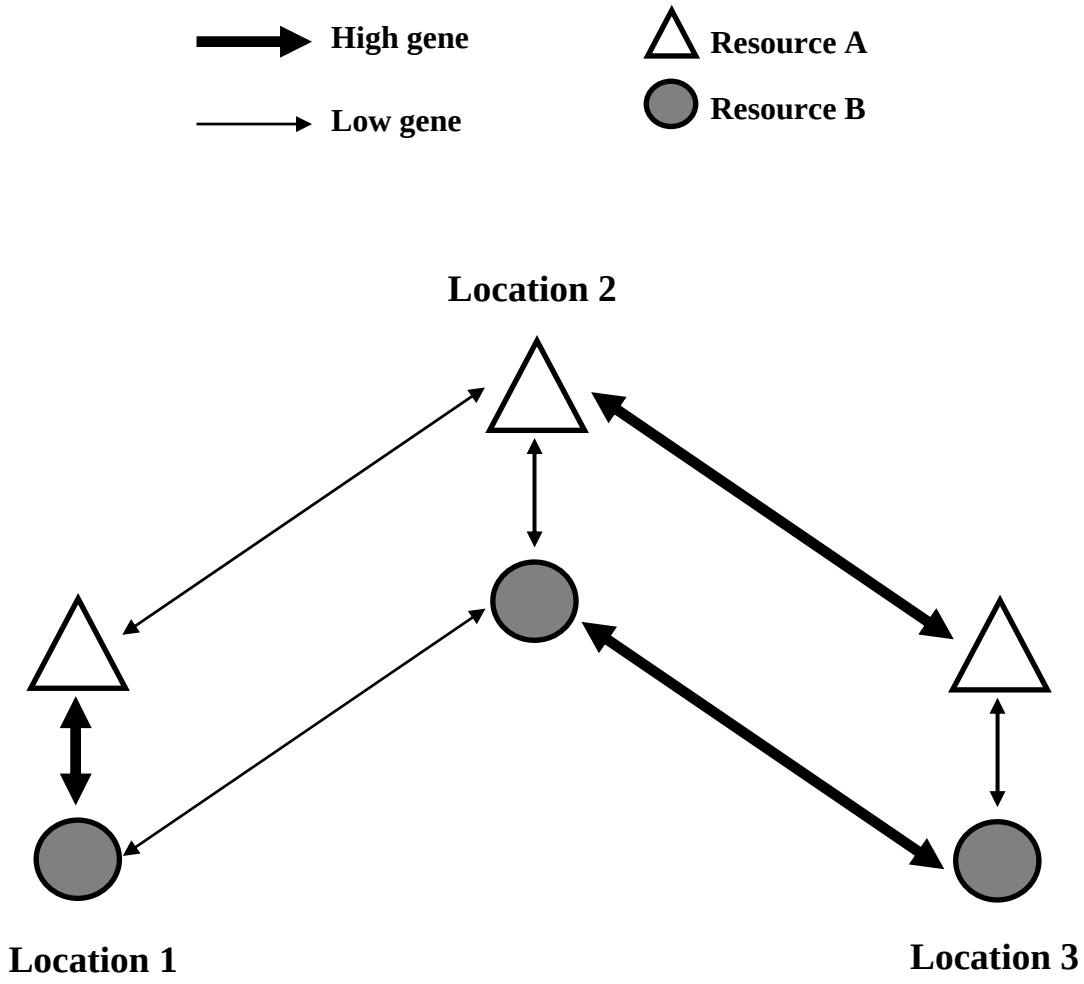


Figure 3. F_{ST} between aphids on alfalfa and clover varied across geographic locations using 5 STS markers. F_{ST} was estimated as Weir and Cockerham's (1984) theta in FSTAT, and standard errors were estimated by jackknifing over loci. There were large differences in host-associated divergence between the populations in the eastern regions (grey) and western regions (white).

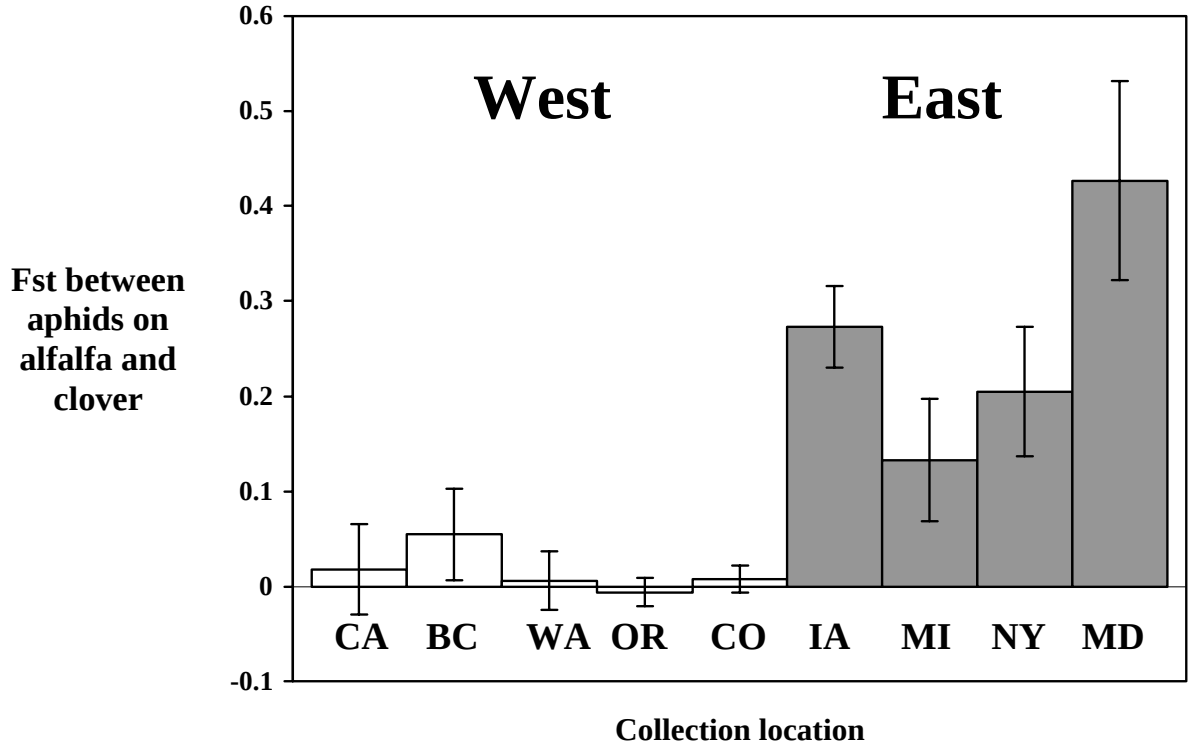
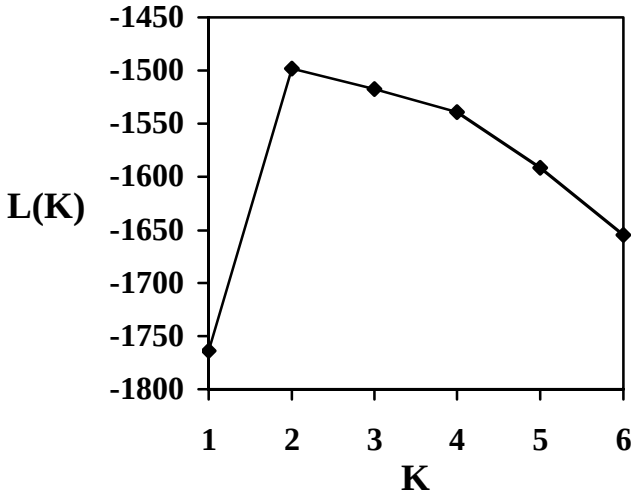


Figure 4. Model selection for STRUCTURE was determined two ways from five independent runs of K = 1 - 6. The maximal value of the log-likelihood probability for the data for a given K [ln(Pr X|K) or L(K)] is suggested by Prichard *et al.* 2000). Evanno *et al.* (2005)'s method uses the peak of ΔK, which is based on the rate of change in the log probability of the data in consecutive runs of K, to determine the number of populations represented by the data (B).

A. Pritchard's method



B. Evanno's method

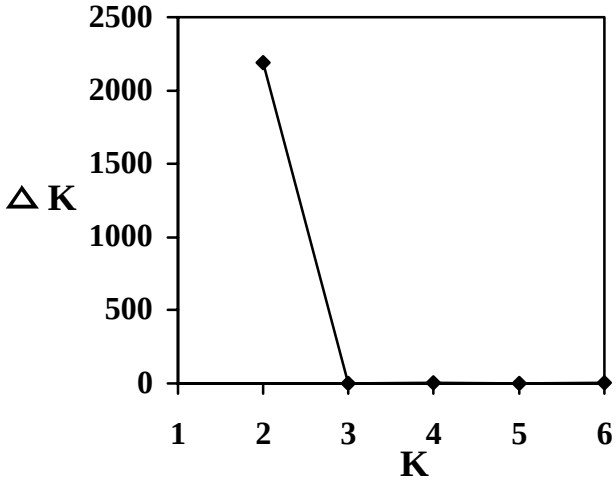


Figure 5. Genetic structure of pea aphids on alfalfa and clover across North America inferred with STRUCTURE using 5 STS and 2 allozymes. Each individual aphid is represented by a bar. Aphids were sorted following assignment by host plant and collection location (i.e. British Columbia, BC; Washington, WA; Iowa, IA; New York, NY; and Maryland, MD). Using the admixture model, aphids were assigned proportionally into the two inferred clusters, and the proportion of each individual's ancestry in the two inferred clusters is shown.

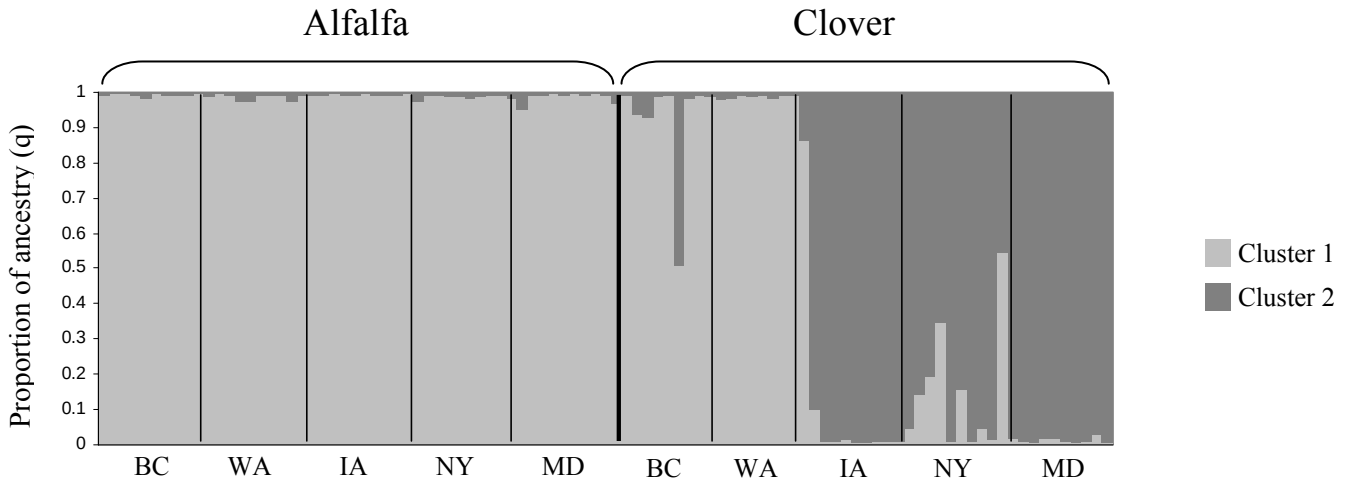
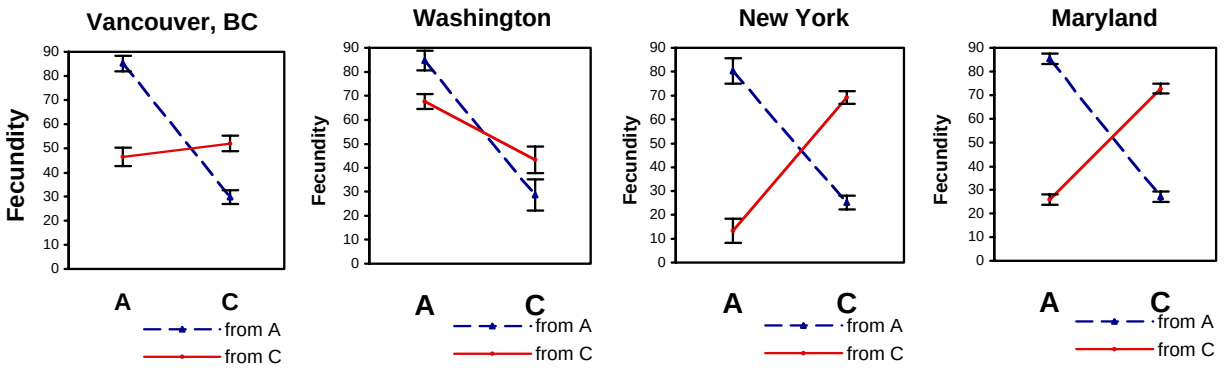
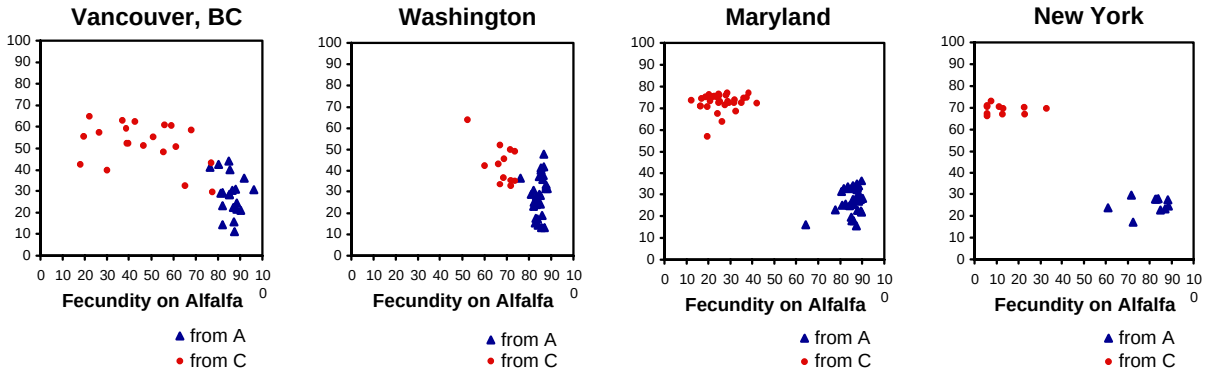


Figure 6. Comparison among geographic locations of the quantitative genetic variation in pea aphid host plant specialization. The number of offspring each aphid produces during its first nine days of adult life (fecundity) provides a measure of host plant specialization. (A) Plots of population means and 95% CI. (B) Scatterplots of means for each aphid clone (BLUPs) for fecundity on each host plant

A.



B.



Chapter 2: Ecological divergence across the genome and across geographic locations in pea aphids

Abstract

Determining how the genetic basis of adaptation and divergence may vary across geographic locations may reveal how evolutionary forces interact to generate biological diversity. During ecological divergence, genetic differentiation may occur first at regions of the genome responsible for traits involved in ecological specialization, while the remainder of the genome remains similar due to ongoing gene flow. When ecological divergence onto the same alternate resources occurs in an array of locations simultaneously, genetic differentiation may involve similar or different genomic regions, genes, and alleles. I examined patterns of genetic differentiation between sympatric pairs of pea aphids on alfalfa and clover in multiple geographic locations in North America using amplified fragment length polymorphisms (AFLPs). A common pattern of genetic heterogeneity in divergence was found in all locations across North America, providing genetic evidence that natural selection has been important in the pea aphid divergence. However, the proportion of markers identified as F_{ST} outliers were not statistically different among locations, and the markers identified as outliers in the West were a subset of those identified in the East. In the East, migrants and hybrids on both host plants were identified using assignment tests, suggesting pea aphids on alfalfa and clover represent host races, and that ecological divergence is maintained in the face of ongoing hybridization.

Introduction

During the process of ecological divergence, populations evolve specialization to alternate environments (Dieckmann and Doebil 1999, Schluter 2001, Coyne and Orr 2004, Rundle and Nosil 2005). Divergent natural selection can drive the evolution of reproductive isolation between resource specialists and results in decreased gene flow between populations. Genetic changes happen first at the genes and genomic regions that contribute to phenotypic differentiation, causing genetic changes to occur at different rates across the genome (Harrison 1991, Charlesworth *et al*. 1997, Barton and Whitlock 1997, Wu 2001, Latta 2004). When diverging populations are widely spatially distributed, genetic changes may also occur at different genes or involve different alleles across geographic locations (Rieseberg and Burke 2001, Nachman 2005). Despite recent advances in understanding the genetics of divergence (Vasemagi and Primmer 2005, Noor and Feder 2006), surprisingly little is known about how genetic changes occur both across the genome and across geographic space (Colosimo *et al.* 2005, Michel *et al.* 2007).

Because the process of speciation cannot easily be observed in nature, one good approach is to study the genetics of speciation in populations at intermediate stages of divergence (Campbell and Bernatchez 2004, Savolainen *et al.* 2006). This allows the evolutionary processes and genetic changes contributing to restricted gene flow and phenotypic differentiation to be directly investigated. The study of divergent populations can be used to make inferences about the process of speciation since populations at intermediate stages of divergence may eventually evolve into distinct species (Dres and Mallet 2002). Populations at intermediate stages of divergence represent different stages

along a continuum of genetic and phenotypic differentiation and reproductive isolation, ranging from locally adapted populations, to host races, and species (Dres and Mallet 2002). Host races are at an important and well-studied stage on the divergence continuum because they experience ongoing (though possibly reduced) gene flow, yet are highly phenotypically divergent and somewhat reproductively isolated (Jaenike 1981, Dres and Mallet 2002, Coyne and Orr 2004).

Classical population genetic theory (Wright 1940, Hartl and Clark 1997) describes how natural selection and gene flow interact to produce genetic changes during the process of divergence at a single locus under divergent selection in a single location. As two sympatric populations evolve in response to local divergent selection for alternate environments, they continue to experience gene flow. Allele frequencies in each population are initially similar at all loci in both populations. As gene flow decreases between the diverging populations, allele frequencies begin to differ between populations, and genetic divergence between populations increases. If gene flow ceases between populations, alternate alleles may eventually become fixed between populations.

Genetic changes across the genome

Allele frequencies may change at many loci in the genome during divergence (Harrison 1991, Black *et al*. 2001, Luikart *et al*. 2003). This can be studied by simultaneously sampling many variable loci across the genome. Loci evolving under evolutionary forces that have locus-specific effects (including selection, mutation, and recombination) can be distinguished from those evolving under neutral forces (such as genetic drift and gene flow), which should influence all loci the same way (Black *et al.* 2001, Luikart *et al.* 2003). I will refer to "loci" as the particular genomic regions, while

"genes" are the loci contributing to phenotypic traits. "Markers" are those genetic tools that are used to investigate variation at particular loci. To the extent that recombination breaks up the genome into independently evolving segments, allele frequencies at individual loci across the genome change at different rates under different evolutionary forces (Lewontin and Krakauer 1973, Harrison 1991). Specifically, allele frequencies at the genes contributing to phenotypes under selection will change more rapidly than allele frequencies at all but the most closely linked neutral loci (Lewontin and Krakauer 1973, Barton and Bengtsson 1986). Markers linked to genes under selection can be identified from their extreme patterns of divergence compared with neutral loci, thereby identifying genomic regions evolving under locus-specific forces such as natural selection (Harrison 1991).

To identify genomic regions involved in ecological divergence, F_{ST} between divergent populations is measured at many markers (Storz 2005, Beaumont 2005). Wright's F_{ST} quantifies the amount of genetic differentiation among populations, and is calculated as the standardized variance in allele frequencies among populations. Markers with extremely high values of F_{ST} ("F_{ST} outliers") are inferred to evolve under divergent selection. Because the genes contributing to phenotypic differences between populations and loci linked to those genes diverge first, F_{ST} outliers mark genomic regions involved in ecological specialization or reproductive isolation (Beaumont and Nichols 1996, Vitalis *et al*. 2001). The process of ecological divergence produces a pattern of genomic heterogeneity in divergence where the genome is a patchwork of highly differentiated and undifferentiated regions (Wu 2001, Emelianov *et al.* 2004, Turner *et al.* 2005). In contrast, under speciation by drift (including some forms of allopatric speciation),

heterogeneity in divergence is not expected since all loci evolve under neutral evolutionary forces. While the population genomics approach has been applied to several cases of ecological divergence (Wilding *et al.* 2001, Campbell and Bernatchez 2004, Emelianov *et al.* 2004, Scheffer and Hawthorne 2007), it is unclear how much of the genome is typically involved during ecological divergence and how quickly changes accumulate across the genome (Noor and Feder 2006).

Genomic heterogeneity in divergence is expected to be a changing, transient feature of the genome. Early in the divergence process, the few highly divergent genomic regions involved in speciation are surrounded by large, undifferentiated genomic regions (Wu 2001, Turner *et al*. 2005). As populations diverge, genetic changes accumulate due to selection and drift, and divergence in the rest of the genome comes into concordance with the initial "speciation" loci (Wu and Ting 2004). Thus, populations at different stages in the process of divergence are expected to vary in their patterns of genetic differentiation, both in overall genetic differentiation (overall F_{ST}, Giles and Goudet 1997) and the proportion of their genome under divergent selection. Host-associated populations are expected to have small overall F_{ST} and few outliers, while host races should have both higher overall F_{ST} and increased numbers of outlier F_{ST} loci. This population genomic approach can be coupled with analysis of introgression rates and detailed ecological study to determine where each population falls on the divergence continuum. Populations at different stages of divergence can be studied not only by looking across different systems, but also by looking across geographic locations when sets of diverging populations are widespread.

Genetic changes across geography

Pairs of sympatric populations may diverge simultaneously in response to the same alternate environments in different geographic locations, creating a mosaic of populations each evolving under divergent selection but with varying outcomes (Chapter 1). Three evolutionary processes interact simultaneously to influence divergence and adaptation in widespread populations: (1) Genetic variation for traits important for ecological divergence may vary among locations, which can occur because of differing population history. (2) The targets and intensity of selection may differ among locations (Itami *et al*. 1998, Schemske and Bierzychudek 2001), influencing the evolution of traits associated with ecological specialization and reproductive isolation (Lu and Bernatchez 1999, Scriber 2002, Fernandez *et al.* 2005, Nosil *et al* 2006). Selection can be constrained by lack of genetic variation, and can also shape genetic variation for important traits. (3) Gene flow among geographically separated populations may differ, both in magnitude and reciprocity, such that some locations freely exchange genetic variation while others evolve more independently (Sork *et al*. 1999, Rieseberg and Burke 2001). Selection and gene flow may interact such that the level of ecological specialization and reproductive isolation varies across locations (Itami *et al.* 1998, Lu and Bernatchez 1999, Scriber 2002, Fernandez *et al.* 2005, Nosil *et al*. 2006) producing a pattern that is called the "geographic mosaic of divergence" (Chapter 1). Diverging populations in different locations may represent different stages along the divergence continuum or different outcomes of the divergence process (Jiggins and Mallet 2000, Dres and Mallet 2002).

The genetic changes contributing to ecological specialization and divergence may differ across the geographic mosaic. Genetic changes across the geographic sets of diverging populations can involve different genes (and thus genomic regions) or the same genes but different alleles. Recent studies suggest that divergence in different locations may involve the same genes (Colosimo *et al.* 2005, Kronforst *et al.* 2006), while different genes are involved in other cases (Hoekstra and Nachman 2003). However, it is not clear how genetic variation differs across locations that vary in degree of divergence. Studying the genetic changes that occur during different stages of divergence could reveal the sequence of genetic changes that occur during the process of divergence.

Insect specialists on agricultural crops provide some of the most well studied examples of ecological-based divergence (Bush 1975, Craig *et al.* 1997, Dres and Mallet 2002, Berlocher and Feder 2002). Pea aphids on alfalfa and clover are highly ecologically specialized, genetically differentiated, and partially reproductively isolated populations that are adapted to alternate host plants (Via 1991, 1999, Via *et al.* 2000). Sympatric populations of pea aphids on alfalfa and clover differ in their extent of ecological specialization between the eastern and western regions of North America, and therefore may be at different stages in the process of divergence (Chapter 1). Aphids on alfalfa and clover in the eastern region represent likely host races, but rates of ongoing gene flow have not been measured. In this study, I ask:

(1) Is there a common pattern of genomic heterogeneity in divergence between aphids on alfalfa and clover across geographic locations?

(2) How does the genetic basis of pea aphid specialization and divergence vary across geographic locations?

(3) Is there hybridization between aphids on alfalfa and clover?

Here, I investigate genetic divergence in pea aphids at three spatial scales using different datasets to address these three questions: (a) between the eastern and western regions of North America (the "Regional dataset") (b) among three populations in the East (the "Eastern dataset"), and (c) with additional genetic markers for one location of the East (the "New York dataset"). First, I determine if there is a common pattern of genomic heterogeneity in divergence in pea aphids on alfalfa and clover across their North American range. This is expected for ecologically-based divergence driven by natural selection (Wilding *et al*. 2001, Emelianov *et al.* 2003). I use F_{ST} outlier analysis to identify regions of the genome potentially linked to genes for ecological specialization and reproductive isolation. Second, I compare the F_{ST} outlier markers in locations across North America to determine if the genetic basis of ecological divergence differs across geographic locations. Then, I examine migration and hybridization between pea aphids on alfalfa and clover in locations in North American where they are most specialized. If aphids on alfalfa and clover experience hybridization, this would define these populations as host races, an important stage along the divergence continuum.

Methods

Aphids on alfalfa and clover

Pea aphids (*Acyrthosiphon pisum* Harris) comprise host-associated populations on herbaceous legumes including alfalfa (*Medicago sativa*) and clover (*Trifolium pratense*) (Via 1999, Via *et al.* 2000). In eastern North America, pea aphids show highly specialized performance on their host plants (Via 1991), and there is selection against

migrants to the alternate host plant and hybrids (Via *et al*. 2000). Pea aphids also have strong preferences for alfalfa and clover (Via 1999; Caillaud and Via 2000), and because pea aphids mate on their host plants, assortative mating tends to occur among aphid clones found together on the same host plant, which contributes to reproductive isolation. Aphid specialization may be influenced by several complexes of nuclear genes, located on all four of the pea aphid linkage groups (Hawthorne and Via 2001) that correspond to the four pea aphid chromosomes (Sun and Robinson 1966). Thus, several genomic regions may be under divergent selection.

Aphid collections

Aphid collections were conducted at three spatial scales. At the largest scale, aphids were collected between 1996-1999 from two regions of North America where aphids are evolving somewhat independently (Chapter 1): the "East" (Iowa, Maryland, and New York) and the "West" (Washington and Oregon) (Regional dataset, Table 1A). Aphids from different locations within regions were pooled to achieve greater sample sizes at the regional scale. At a smaller spatial scale, intensive sampling within the eastern region was conducted in 2001 and 2002 to assess variation among genetically and phenotypically similar locations where aphids on alfalfa and clover are highly ecologically specialized. Aphids were sampled from Iowa, Maryland, and New York (Eastern dataset, Table 1B). The New York dataset includes the same aphids sampled from New York for the Eastern dataset but genotyped using 52 additional markers (83 total).

At each location, aphids were collected from adjacent fields of alfalfa (*Medicago sativa*) and red clover (*Trifolium pratense*). Aphids were sampled in widely dispersed locations across fields to avoid sampling the same clone.

AFLP analysis

Genomic DNA from single aphid adults was extracted using the DNeasy Tissue Kit (QIAGEN), eluted into 200 uL of supplied buffer, and stored at -20°C. Aphids were genotyped for amplified fragment length polymorphisms (AFLPs; Vos *et al*. 1995), which are nuclear, multilocus genetic markers, which survey variation across the genome (Luikart *et al* 2003). Genotyping followed the method used previously for pea aphids (Hawthorne and Via 2001), but with different primer combinations for each dataset (Table 2). For the regional dataset, two primer combinations generated 44 polymorphic markers. For the Eastern dataset, fourteen different primer combinations produced 31 polymorphic markers. For the New York dataset, four additional primer combinations were added to those used in the Eastern dataset, and additional markers were scored, to generate a total of 83 polymorphic markers.

Digestion of genomic DNA used two six-base recognition restriction enzymes (*Pst*I and *Eco*RI) to reduce the number of fragments generated. Digestion of genomic DNA and ligation of adaptors to the ends of the restriction fragments was performed in 50 uL reactions containing 25 uL of genomic DNA, New England Biolabs (Beverly, MA) #4 restriction enzyme buffer, 1.8 mM DTT, 2 mM ATP, 100 ng/uL bovine serum albumen, 20 units *Pst*I, 20 units *Eco*RI, 6 units T4 DNA ligase, and 5 pM of each double-stranded adaptor (Table 2). Reactions were incubated at 37°C for 3-5 hours in a shaker oven to generate the AFLP construct.

Two rounds of amplification generate the AFLP fragments. The first (preselective) amplification used primers complementary to the adapter sequences (Table 2). Each sample was amplified in 50 uL PCRs containing 5 uL of the AFLP construct as template, 1X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 5 pM of each core primer (Table 2), and 1 unit of Taq DNA Polymerase (New England Biolabs). Amplifications were performed on a MJ Research thermocycler: 1 minute at 95 °C, followed by 21 cycles of 10 seconds at 95 °C, 30 seconds at 52°C, and 1 minute and 30 seconds at 72 °C, with a final 5 minutes at 75 °C. Preamplification products were diluted 1:2 with 10mM Tris.

For the second, selective round of amplification, two primer combinations were used based on a pair of core primer sequences differing in their selective extensions (Table 2). Each sample was amplified in 50 uL PCRs containing 5 uL of the diluted preamplification product as template, 1X PCR buffer, 3 mM $MgCl_2$, 0.2 mM dNTPs, 5 pM of each selective primer, and 1 unit of Taq DNA Polymerase (New England Biolabs). The selective amplification used a touchdown PCR which began with 30 seconds at 95 °C, followed by 12 cycles of 10 seconds at 95 °C, 40 seconds at the annealing temperature, and 1 minute and 30 seconds at 72 °C, where the annealing temperature started at 65°C but then decreased by 0.7°C each cycle. This was followed by 35 cycles of 11 seconds at 95 °C, 40 seconds at 56°C, and 1 minute and 30 seconds at 72 °C, with a final 5 minutes at 75°C. The selective amplification product was diluted 2:1 with loading dye (98% formamide, 10mM EDTA, 0.025% xylene cynol, 0.025% bromophenol blue). Samples were denatured at 94°C for 2 minutes, and 2 uL of sample were evaluated on 35 x 45 cm vertical, 4.3% denaturing polyacrylamide gels (National Diagnostics, Atlanta,

GA). Gels were run at 60 Watts for 4 hours, and silver stained (Silver Sequence, Promega). AFLP bands were visualized on a light box, and individuals were scored for the presence and absence of bands at each polymorphic locus. Because AFLPs are dominant markers, band-present heterozygotes and band-present homozygotes cannot be distinguished. Unbanded individuals are band-absent homozygotes.

Detecting F_{ST} outliers

To identify markers linked to genomic regions potentially under divergent selection, I used F_{ST} outlier analysis in DFDIST (Beaumont and Nichols 1996). F_{ST} between aphids on alfalfa and clover for each marker was compared to the expected neutral distribution, modeled using computer simulations, to find markers with high F_{ST}. Separate outlier F_{ST} analyses were done within each region (Regional dataset) and location (Eastern dataset). In DFDIST, allele frequencies were estimated from dominant markers using the Bayesian method developed by Zhivotovsky (1999). Outlier loci were detected using multiple rounds of simulation as suggested by Beaumont and Nichols (1996). First, coalescent simulations of a neutral model of migration and drift generated the null distribution of F_{ST} around a weighted mean F_{ST} calculated from the empirical distribution of the data. The weighted mean F_{ST} is calculated with the upper and lower 30% of markers removed, so is relatively insensitive to the presence of outliers. The 0.95 quantile of the neutral simulated distribution was calculated. The probability that each marker was outside the range expected under neutrality was calculated. Those markers with probabilities greater than 0.95 were considered outliers. Additional rounds of simulations were performed with an expected F_{ST} calculated with the outlier markers removed, and additional outliers were detected based upon this lower mean F_{ST}. This

was repeated until no additional outliers were detected. Only markers where the frequency of the rare allele was greater than 2% were included. All simulations used two populations modeled to represent the two host-associated populations with beta of 0.25 and 50,000 iterations.

To determine whether the proportion of markers identified as F_{ST} outliers differed between regions, I used Fisher's Exact test to test for an association of the proportion of outliers detected and geographic location.

Identification of migrants and hybrids

Two methods were used to investigate migration and hybridization between aphids on alfalfa and clover and to determine if the most differentiated eastern populations have low rates of introgression, and therefore represent host races. Bayesian analysis in STRUCTURE was used to identify individual migrants and hybrids to estimate rates of migration. The distribution of hybrid index scores for aphids on each host plant shows the proportion of the population with individuals of mixed ancestry in each location. Separate analyses were conducted for the three locations of the east (Eastern dataset) that show similar levels of ecological specialization and reproductive isolation to determine if they also show similar patterns of migration and hybridization.

STRUCTURE

To examine genetic differentiation of aphids, identify migrants and hybrids, and measure rates of hybridization between aphids on alfalfa and clover, I used Bayesian assignment-test analysis of multilocus AFLP markers implemented in STRUCTURE 2.2 (Falush *et al.* 2007). I first used the admixture model without population of origin information to determine whether populations cluster by host plant of origin (Pritchard *et*

al. 2000). Under this model, the allele frequencies of the inferred genetic populations (called "clusters") are determined by Bayesian analysis using only individuals' multilocus genotypes. Clusters are constructed to maximize Hardy-Weinberg equilibrium and linkage disequilibrium within genetic populations. In the model, individuals can inherit a fraction of their genome from each inferred cluster, and have a proportion of ancestry derived from each cluster (q), which ranges from 0 to 1. Individuals with mixed ancestry have intermediate values of q (i.e. an F1 hybrid has q = 0.5). The proportion of membership of each predefined population (i.e. Iowa alfalfa, New York clover, etc.) into each of the inferred clusters reflects the correspondence of location and host plant to the clusters.

I next used the admixture model with the USEPOPINFO option in STRUCTURE 2.2 to assign individuals into hybrid categories, and obtain statistical support of the classification (Pritchard *et al.* 2000, Falush *et al.* 2007). Under this model, most individuals are residents, with full ancestry in the cluster of origin, some are migrants, with full ancestry in the other cluster, and some are hybrids, with past admixture. I estimated the probability of admixture in the past three generations (GENSBACK = 3, MIGRPRIOR = 0.05), which models hybrids as F1, backcross, or second-generation backcrosses. Individuals were classified as "residents" if they had a posterior probability greater than 0.5 of having full ancestry on the host plant of origin, and as "migrants" if they had the greatest probability of being from the alternate host plant. "Hybrids" were considered those individuals with less than 0.5 posterior probability of having pure ancestry from either plant.

For all STRUCTURE analyses, I fixed the number of populations at K = 2, because these samples (Regional dataset) represent two host-associated genetic populations (see Chapter 3), and because our questions are concerned with gene flow between aphids on alfalfa and clover. Models were run assuming correlated allele frequencies between populations (Falush *et al.* 2003) for a burn in period of 100,000 Markov Chain Monte Carlo iterations and a data collection period of 1,000,000 Markov Chain Monte Carlo iterations.

Hybrid index

To characterize hybridization between aphids on alfalfa and clover, I examined the distribution of hybrid index scores for aphids collected from alfalfa and clover in the three eastern locations. The distribution of hybrid index scores shows the relative number of individuals with intermediate and parental-type genotypes in a population. The hybrid index score was calculated as the proportion of markers for each aphid individual that show the alfalfa phenotype. Aphids from alfalfa are expected to have a hybrid index close to 1, aphids from clover are expected to have a hybrid index of 0, and F_1 hybrids should be 0.5. I calculated a hybrid index using the AFLP markers with above average F_{ST}, (the top 15 calculated using DFDIST in each location, Beaumont and Nichols 1996). Using high F_{ST} markers maximizes the resolving power of the method for populations with markers that do not show fixed differences. This included all the F_{ST} outliers identified in each location in addition to non-outlier markers. In comparison to hybrid index analyses using markers with fixed differences between populations, using lower F_{ST} markers may overestimate the number of individuals with intermediate hybrid indices, and thus provides a conservative estimate of the degree of host-associated

differentiation. I defined the "alfalfa phenotype" as the phenotype seen most commonly in alfalfa (either banded or unbanded) for each marker, and then counted the number of markers showing the alfalfa phenotype for each individual. The "hybrid index score" for each individual was calculated as the number of markers showing the alfalfa phenotype divided by the number of markers (15 total).

Analyses of New York samples using additional markers (New York dataset)

Sampling many markers across the genome improves accuracy of multilocus analyses (Storz 2005, Vaha and Primmer 2006). To determine if my results would be different if additional markers were used, I repeated the F_{ST} outlier analysis in DFDIST and identification of migrants and hybrids in STRUCTURE with the same samples from New York using 53 additional AFLPs for a total of 83 polymorphic AFLPS (New York dataset). I compared the proportion of markers identified as outliers, and the number of hybrids and migrants identified in the New York and Eastern datasets.

Results

Detecting F_{ST} outliers

Only polymorphic AFLP markers were used in the analysis. The Regional dataset had 45 polymorphic AFLPs in the East and 34 polymorphic AFLPs in the West. In the Eastern dataset, all 31 AFLPs were polymorphic in all locations. Neutral simulations in DFDIST were based on the weighted mean F_{ST} for all polymorphic markers, which was higher in the East than the West (Table 3) but similar among locations within the East (Table 3). In all locations, variability among markers in F_{ST} was high (Figure 1 and 2).

In all populations, the mean F_{ST} was much higher than the median, indicating a skewed distribution of F_{ST} (Table 3).

Multiple F_{ST} outliers were detected in all locations (Figure 1 and 2). The proportion of markers identified as outliers ranged from 0.06 to 0.18 (Table 3). There were a greater proportion of F_{ST} outlier markers found in the East compared to the West. However, the proportion of significant F_{ST} outliers was not statistically different between the two regions (two-tailed Fisher's Exact test, $p = 0.35$). The proportion of markers identified as outliers also varied among the three locations within the eastern region. The six significant F_{ST} outliers identified in the Eastern dataset are distributed on three of the four linkage groups (Hawthorne and Via 2001, West unpublished data). None of the outliers are located within 10 cM of one another. The AFLP markers in the Regional dataset have not been mapped.

Detection of the same F_{ST} outlier in different locations suggests a similar genetic basis of ecological divergence across geographic locations. In the East and West (the Regional dataset), three F_{ST} outlier markers were the same in both locations, and five were restricted to the East (Table 4A). Two of the shared F_{ST} outliers (Pat-Etc 528 and Pct-Etg 950) also had the highest F_{ST} values in both locations, and were both detected in the first round of F_{ST} simulation in DFDIST. For locations within the East (the Eastern dataset), one outlier was common to all locations, two were in two locations, and three were restricted to a single location (Table 4B). Different markers were used in the Regional and Western datasets, so comparisons could not be made across these spatial scales.

Identification of migrants and hybrids

STRUCTURE

Under the admixture model without population of origin information, there was a correspondence between the inferred genetic clusters and the pre-defined populations of origin, indicating genetic differentiation among host-associated populations (Figure 3A). The proportion of membership of each pre-defined population into the inferred cluster was high for both aphids from alfalfa (0.93, 0.94, and 0.92 for Iowa, New York, and Maryland respectively) and for aphids from clover (0.94, 0.95, and 0.87 for Iowa, New York, and Maryland respectively).

Using prior information about aphid host plant of origin, pea aphid migrants and hybrids were identified on both alfalfa and clover in all three locations in the East (Eastern dataset, Figure 3B, Table 5). Seven aphid individuals in Iowa, Maryland, and New York had probabilities greater than 75% of having both parents from the alternate ecotype (indicated by a * in Figure 3B). All these "migrant" individuals had less than 1% probability of being a resident. Eight individuals identified as "hybrids" had less than a 50% posterior probability of having full ancestry on the other host plant (shown by a "**" in Figure 3B). However, there was uncertainty about the number of generations back admixture occurred, which means the type of hybrid (i.e. F1, backcross, or later generation backcross) could not be identified. Only one individual had $p > 0.5$ in any hybrid category; an aphid collected from alfalfa in Iowa had $p = 0.51$ of having one parent from clover (i.e. F1).

Hybrid Index

The distribution of individual hybrid index scores for aphids from alfalfa and clover in Iowa, New York, and Maryland forms a bimodal distribution in all three eastern locations (Figure 5). However, aphid populations on alfalfa and clover show a range of hybrid index scores, and the distributions were overlapping. Most individuals showed similar hybrid index scores as other individuals from the same host plant. Fewer aphids (10.5 %, 9.4%, and 5.7% in Iowa, Maryland, and New York respectively) had intermediate hybrid indices (between 0.3 and 0.5), suggesting they are of mixed ancestry. Individual aphids identified as hybrids in STRUCTURE had intermediate hybrid indices (indicated by "**" in Figure 5), though not all individuals with intermediate hybrid indices were identified as hybrids in STRUCTURE. This discrepancy could be because STRUCTURE uses data from all the markers, while the hybrid index analysis uses only the most divergent markers. Several aphids showed genotypes characteristic of the aphids from the other crop, and are likely migrants. Most of these individuals were identified as migrants using STRUCTURE (indicated by "*" in Figure 5).

Analyses of New York samples using additional markers (New York dataset)

F_{ST} outlier analysis and identification of migrants and hybrids was repeated for the aphids on alfalfa and clover from New York using an additional 52 markers. All 83 AFLPs were polymorphic between alfalfa and clover, and so could be subject to F_{ST} outlier analysis. Mean F_{ST} between aphids on alfalfa and clover was similar between datasets (0.21 and 0.19 for the Eastern and New York datasets respectively). All outliers in the Eastern dataset were also outliers in the New York dataset, and additional outliers were detected (Figure 4D). Estimates of the proportion of outlier markers in New York

were similar between analyses as well (0.12 and 0.11 for the Eastern and New York datasets respectively, Table 3).

STRUCTURE results were consistent when more markers were included, yielding only slightly higher estimates of the proportion of membership of each pre-defined population into each of the inferred clusters (0.98 for alfalfa and 0.96 for clover) (Figure 4A). The number of inferred migrants was identical in analyses using 31 and 83 AFLPs, but one individual classified as a resident using 31 markers was classified as a hybrid using 83 markers (Figure 4B), which is expected because the power to detect hybrids is increased with additional markers (Evanno *et al.* 2005, Vaha and Primmer 2006).

Discussion

Genetic divergence between aphids on alfalfa and clover varied across the genome and across geographic locations in North America. Pea aphids from all locations showed a pattern of genomic heterogeneity in divergence, providing strong genetic evidence that natural selection drives pea aphid divergence. The genetic basis of divergence showed some similarity across locations, but there were also important differences in the F_{ST} outliers identified. This suggests differences and similarities in the genes for divergence across locations, and/or that populations represent different stages on the divergence continuum. Finally, low rates of hybridization were measured between pea aphids on alfalfa and clover in eastern North America, which should now be defined as host races, an intermediate stage in the process of divergence.

Geographic variation in genomic heterogeneity of divergence

Pea aphids on alfalfa and clover showed a common pattern of heterogeneity in divergence across geographic locations, implicating natural selection as a cause of divergence (Emelianov *et al.* 2003). All locations in North America showed a pattern of high variability in F_{ST} among markers, a skewed distribution of F_{ST}, and multiple, significant F_{ST} outliers, suggesting these genomic regions have a level of divergence outside the range expected if they were evolving under neutral processes. This is consistent with detailed previous research in this system showing that divergent natural selection can be a rapid and effective force driving ecological divergence in sympatry (Via 1991, 1999, Via *et al.* 2000). F_{ST} outliers could all be linked to a single divergent region or they may indicate many regions, depending on their distribution across the genome. In the Western dataset where the markers are located on the linkage map (Hawthorne and Via 2001, Via and West, unpublished), the F_{ST} outliers are distributed on three of the four aphid linkage groups, and most are not closely linked. Therefore, identification of multiple F_{ST} outliers suggests that multiple genomic regions are differentiated between aphids on alfalfa and clover in each location.

Aphids on alfalfa and clover in different regions and locations showed differences in the pattern of genomic heterogeneity in divergence, suggesting differences among locations in the genetic basis of divergence and that they are at different places in the divergence continuum. Genetic divergence between aphids on alfalfa and clover, measured by overall F_{ST} of AFLP markers, was greater in the East than the West (Table 3). This confirms the results from nuclear STS markers found with a different aphid sample (Chapter 1) and provides further evidence for the pattern of a geographic mosaic

of divergence in pea aphids. Furthermore, a higher proportion of F_{ST} outliers were found in the East than the West, though the difference was not statistically significant. However, these results are suggestive, and it could be that larger or more genomic regions show the effect of divergent natural selection in the East. Natural selection or genetic variation could vary among locations and aphids on alfalfa and clover in the West may be at an earlier stage of divergence, so may not have diverged at as many loci. Because populations at different stages of divergence are rarely considered, differences in sets of diverging populations in patterns of differentiation across the genome have not been documented before. Comparing populations on the same resources at early and intermediate stages of divergence may allow us to dissect the sequence of genetic changes that occur during divergence.

At a smaller scale, the three more ecologically specialized eastern populations all showed high F_{ST} between aphids on alfalfa and clover, indicating high genetic differentiation. However, the variation in the percentage of outlier markers was larger than expected. Iowa had only two F_{ST} outliers while New York and Maryland had four. This is not consistent with what is expected if the percentage of outliers is correlated to the stage of divergence, and suggests further studies are needed (see below).

The large proportion of outlier markers (6-18%) in all locations suggests portions of the genome are evolving due to divergent selection for host plant specialization, despite hybridization between aphids on alfalfa and clover. While some examples of ecologically based divergence show similar percentages of outlier markers, 11-15% in leaf miners (Scheffer and Hawthorne 2007), fewer outliers have been identified in other systems: 1.4-3.2 % in fish (Campbell and Bernatchez 2004), 5% in snails (Wilding *et al.*

2001), and 1.5% in island palms (Savolainen *et al.* 2006). The large proportion of F_{ST} outlier markers may be explained by a combination of factors including the polygenic basis of host plant specialization in aphids (Hawthorne and Via 2001), strong selection, and the population history of pea aphids into North America. Models of divergence that account for population structure indicate strong selection can produce high F_{ST} at neutral loci quite distant from the locus under selection due to hitchhiking and population structure (Charlesworth *et al.* 1997). Because pea aphids have a cyclically parthenogenic life cycle, clonal selection during the summer, combined with rapid population growth rates, mean that selection for host plant performance can be very strong (Halkett *et al.* 2005). Furthermore, selection is often especially strong in agricultural systems because host plants, especially alfalfa, are grown in large monocultures. Furthermore, the population history of pea aphids may influence the number of genomic regions that show differentiation. If aphids from alfalfa and aphids from clover were introduced into North America from different source populations, some of the differentiation may reflect genetic differences that were accumulated allopatrically, and thus may not mark genomic regions under host plant specialization. Future studies are needed to put these genetic differences in proper historical and biogeographic context.

Geographic variation in the genetic basis of specialization and divergence

The second objective was to determine if the same genomic regions are involved in pea aphid alfalfa-clover ecological specialization and divergence, as indicated by F_{ST} outlier analysis, across geographic locations. Both similar and different F_{ST} outliers were identified in two relatively isolated parts of their introduced range, eastern and western North America, and in three locations within the eastern region. A similar genetic basis

among locations for genetic variation for alfalfa and clover specialization is not surprising given the recent introduction of these populations into North America and evidence for gene flow among locations (Chapter 1). If aphids from alfalfa in different locations share a common history (Chapter 1), they would be expected to share key variation for host plant use. Even low rates of gene flow can spread genetic variation under common selective pressure, especially for alleles with large effects on the traits, allowing populations within species to evolve collectively (Rieseberg and Burke 2001, Morjan and Rieseberg 2004).

The presence of different F_{ST} outliers suggests that the genetic basis of specialization onto alfalfa and clover may differ somewhat between geographic locations, especially between the eastern and western regions of North America. The same selection pressure can produce parallel phenotypic changes using different genes, such as in parallel evolution, when gene flow between populations is restricted (Hoekstra and Nachman 2003, Colosimo *et al.* 2004). Alternatively, populations could be at different stages in divergence, if local selection varies geographically. Alfalfa and clover are not uniform resources, differing in abundance, quality, and composition across North America (Barnes *et al.* 1995). Populations could also differ for selection on other ecologically important traits with strong host plant association such as associated endosymbionts (Simon *et al.* 2003, Tsuchida *et al.* 2004), and resistance to parasitoids (Henter and Via 1995, Ferrari *et al.* 2001) and fungi (Ferrari and Godfray 2003). Determining the cause of the difference in genetic differentiation between regions will ultimately require several approaches (Vasemagi and Primmer 2005). Construction of a QTL map for aphids from alfalfa and clover populations in the West would allow

comparison of the genetic basis among regions, and would complement the F_{ST} outlier approach to understanding the genetics of divergence (LeCorre and Kremer 2003).

Hybridization between aphids on alfalfa and clover

Aphids on alfalfa and clover showed low rates of migration and hybridization, yet still maintain high phenotypic differentiation and genetic differentiation at genes and linked genomic regions. Assignment-based tests using genetic markers estimated actual migration rates of about 0.02 between pea aphid host-associated populations. Furthermore, identification of hybrids with a range of admixture percentages provides evidence that migrants to the alternate host plant occasionally survive to reproduce, and their offspring, with a combination of genetic elements from both host-associated populations, contribute genetic material to future generations. Hybrid index analysis reveals a range of hybrid types present at low frequencies on both host plants. The bimodal distribution of individual hybrid index scores, with a greater frequency of later-generation hybrids and fewer intermediates, implies that populations have strong, but not complete, barriers to gene flow. This is indicative of a later stage in the process of divergence, such as host races (Jiggins and Mallet 2000).

Pea aphids are frequently cited as an example of insect host races undergoing ecological speciation in sympatry (Coyne and Orr 2004). However, the status of pea aphids as host races has been questioned (Dres and Mallet 2002) because it is not known if they undergo hybridization and gene flow. Host races represent a point along the divergence continuum, undergoing actual gene flow and hybridization at an appreciable rate ($m > 1\%$ per generation, Dres and Mallet 2002), while sibling species hybridize only rarely. While hybridization can constrain divergence, it also allows populations to

potentially exchange variation for traits under positive selection in both environments (Arnold *et al.* 1999, Rieseberg and Burke 2001, Morjan and Rieseberg 2004). This is the first study to directly show hybridization and gene flow between aphids on alfalfa and clover, and it suggests pea aphids represent host races (Dres and Mallet 2002). This is consistent with indirect estimates of migration using F_{ST} with allozymes (Via 1999) and laboratory trials of host preference (Via *et al.* 2001), which provided suggestive evidence for continuing gene flow between aphid on alfalfa and clover (Dres and Mallet 2002, Coyne and Orr 2004). Together with extensive ecological, behavioral, and genetic research on pea aphids (summarized in Dres and Mallet 2002), the identification of hybrids in this study places the pea aphids in eastern North America in the category of host races, and in the context of other studies of ecological speciation.

Conclusions

This study provides insights into the mechanisms by which evolutionary forces contribute to the geographic mosaic of divergence in pea aphids, and highlights the value of taking a geographical perspective to the study of the genetics of ecological speciation. I confirmed the pattern of the geographic mosaic of divergence found in Chapter 1 using different markers and aphid samples: the eastern region of North America showed greater genetic differentiation between aphids on alfalfa and clover than the western region. Results from this study provided important insights into the three evolutionary processes by which this pattern arises and is maintained: (1) Genetic variation for important traits may differ among locations. While some genetic variation was similar across locations, the presence of different F_{ST} outliers among locations suggests important differences as well. This could cause the genetic basis of traits for ecological specialization or

divergence to differ, or even the outcome of divergence to vary across locations. (2) Selection may also differ among locations. The common pattern of genomic heterogeneity in divergence across geographic locations suggests that selection drives divergence between aphids on alfalfa and clover. This confirms previous ecological genetic research implicating divergent natural selection as the primary cause of pea aphid divergence (Via 1999, Via *et al*. 2000). The finding that there were fewer F_{ST} outliers in the west compared to the east suggests selection varies across locations. Fewer genomic regions may be under selection in the west if some genes for specialization are not important in the west, or divergent selection could be weaker, perhaps because of host plant conditions. (3) Gene flow among geographically separated populations on the same resource may differ. While it remains to be seen what the rates of ongoing migration among locations are, shared genetic variation for traits involved in specialization and divergence, as suggested by shared outliers, suggests specialists on the same resource are evolving somewhat collectively. This implies that important variation for specialization could be spread by gene flow across locations, even if there are some differences in selection and genetic variation across locations.

Taking a geographic perspective provides important insights to the genetics of ecological speciation, and shows how populations at different stages in divergence vary in their pattern of differentiation across the genome. Sampling across the geographic range of host-associated populations allows sampling different parts of the divergence continuum- specifically, eastern populations of pea aphids may represent host races, as suggested by their rates of ongoing migration and hybridization and their bimodal hybrid index pattern. In contrast, the western populations may represent an earlier stage of

divergence, such as host-associated populations, though future research will be needed to characterize hybridization rates in this region. Comparison of the genetics of divergence between these host races in the East with the host-associated populations in the West suggest more of the genome is under selection and more of the genome shows a pattern of divergence at later stages of divergence.

Tables

Table 1. Pea aphids were collected from alfalfa and clover across North America (A). Additional aphids were collected in the three eastern locations (B).

A. Regional dataset

Region	Location	No. aphids from alfalfa in location	No. aphids from clover in location	No. aphids from alfalfa in region	No. aphids from clover in region
West	E. Washington	29	12	41	16
	Oregon	12	4		
East	Iowa City, Iowa	10	10	29	28
	Beltsville, Maryland	9	8		
	Ithaca, New York	10	10		

B. Eastern and New York datasets*

Location	No. aphids from alfalfa in location	No. aphids from clover in location
Iowa City, Iowa	48	47
Middleton, Maryland	94	66
Ithaca, New York	85	89

* Only the New York samples are used in the New York dataset.

Table 2. Sequences of adaptors and primers used for AFLP protocols. Sequences are provided 5' to 3'. A variable number of polymorphic AFLP markers were generated for each primer combination for each aphid collection to generate three datasets.

PstI adaptors	EcoRI adaptors
TGTACGCAGTCTTAC	AATTGGTACGCAGTC
CTCGTAGACTGCGTACATGCA	CTCGTAGACTGCGTACC

The core sequence of primers for EcoRI and PstI amplicons

Pst1 core:	GACTGCGTACATGCAG
EcoRI core:	GACTGCGTACCAATTC

Selective Primers

PstI+AG(Pag)	GACTGCGTACATGCAGAG
PstI+AT (Pat)	GACTGCGTACATGCAGAT
PstI+CA (Pca)	GACTGCGTACATGCAGCA
PstI+CT (Pct)	GACTGCGTACATGCAGCT
EcoRI+AG (Eag)	GACTGCGTACCAATTCAG
EcoRI+CA (Eca)	GACTGCGTACCAATTCCA
EcoRI+CG (Ecg)	GACTGCGTACCAATTCCG
EcoRI+CT (Ect)	GACTGCGTACCAATTCCT
EcoRI+GC (Egc)	GACTGCGTACCAATTCGC
EcoRI+TC (Etc)	GACTGCGTACCAATTCTC
EcoRI+TG (Etg)	GACTGCGTACCAATTCTG

Number of Polymorphic AFLP markers generated from each primer combination for each aphid collection

Primer combination	Regional dataset (East + West)	Eastern dataset (Iowa,, Maryland, New York)	New York dataset (New York only)
Pct-Etg	29	-	8
Pca-Egc	-	3	4
Pat-Ect	-	3	4
Pct-Etc	-	3	7
Pat-Egc	-	4	10
Pca-Etc	-	3	4
Pca-Eag	-	1	2
Pca-Ecg	-	2	3
Pct-Eca	-	3	5
Pct-Eag	-	4	5
Pag-Eag	-	1	2
Pct-Ect	-	1	2
Pca-Ect	-	2	5
Pct-Ecg	-	1	4
Pct-Egc	-	1	7
Pag-Etc	-	-	4
Pat-Eag	-	-	1
Pat-Eca	-	-	6
Pct-Eat	15	-	-
Total No. Markers:	44	31	83

Table 3. The number of F_{ST} outlier markers identified using DFDIST is shown out of the total number of markers examined in each location for the three datasets. The overall mean F_{ST} and median F_{ST} of all markers is calculated for each location. The proportion of markers within each dataset identified as outliers between pea aphids on alfalfa and clover is provided for each location.

Dataset	Population	Mean F_{ST}	Median F_{ST}	No. F_{ST} Outliers	Proportion of Markers
Regional	East	0.17	0.05	8 / 45	0.18
	West	0.02	-0.006	3 / 34	0.09
Eastern	Iowa	0.22	0.17	2 / 31	0.06
	Maryland	0.21	0.13	4 / 31	0.12
	New York	0.19	0.12	4 / 31	0.12
New York	New York	0.19	0.12	9 / 83	0.11

Table 4. Genetic differentiation between aphids on alfalfa and clover for each outlier AFLP marker are shown for each geographic region. Marker names are composed of the name of the primer combination from which that fragment was generated and the approximate size of that fragment. The size of the fragment corresponds to the marker name in Hawthorne and Via (2001) for the Eastern dataset. Markers were identified as F_{ST} outliers in DFDIST in one, two, or all locations.

A. Eastern and Western North America (Regional dataset)

Marker Name	F_{ST} Outlier	F_{ST} A-C (East)	F_{ST} A-C (West)
Pat-Etc 528	Both	0.84	0.14
Pct-Etg 950	Both	0.55	0.20
Pct-Etg 363	Both	0.37	0.09
Pat-Etc 790	East	0.20	-0.01
Pat-Etc 264	East	0.37	-0.01
Pat-Etc 484	East	0.32	0.01
Pct-Etg 292	East	0.32	-0.02
Pat-Etc 465	East	0.47	-0.01

B. Locations within Eastern North America (Eastern dataset)

Marker Name	F_{ST} Outlier	F_{ST} A-C (Iowa)	F_{ST} A-C (New York)	F_{ST} A-C (Maryland)
Pct-Etc 420	All	0.77	0.72	0.51
Pct-Etc 740	IA,NY	0.69	0.57	0.25
Pat-Ect 373	NY,MD	0.60	0.58	0.54
Pca-Etc 396	MD	0.34	0.41	0.66
Pct-Eag 425	NY	0.03	0.68	0.01
Pca-Ecg 499	MD	0.61	0.34	0.53

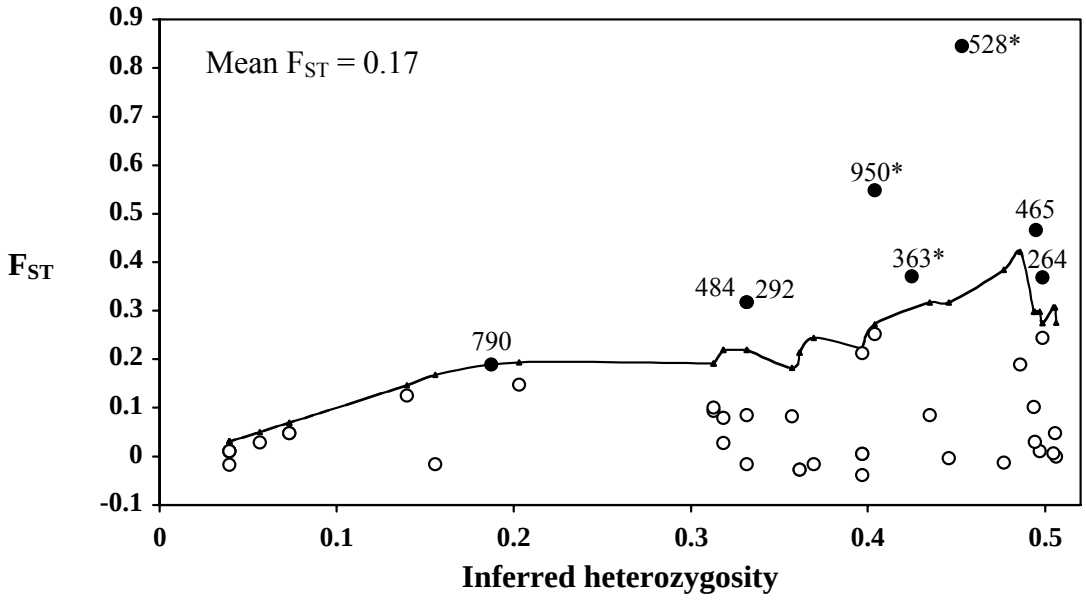
Table 5. Aphid migrants and hybrids on alfalfa and clover were identified using STRUCTURE in three locations in eastern North America. To obtain a more accurate estimate of migration in New York, an additional 52 AFLP markers were analyzed (New York dataset). The migration rate from the other host plant was calculated as the proportion of migrants from the other ecotype present on that host plant.

Dataset	Location	Host plant	Residents	Migrants	Hybrids	Migration Rate
Eastern	Iowa	Alfalfa	46	1	1	0.02
		Clover	46	0	1	-
	Maryland	Alfalfa	89	3	2	0.03
		Clover	62	1	3	0.02
	New York	Alfalfa	85	0	0	-
		Clover	86	2	1	0.02
New York	New York	Alfalfa	85	0	0	-
		Clover	85	2	2	0.02

Figures

Figure 1. F_{ST} between aphids on alfalfa and clover estimated from variable AFLP loci are shown against heterozygosity estimates in the two North American regions (Regional dataset). A solid line represents the 0.95 quantile of the neutral simulated distribution. The mean F_{ST} over all markers is indicated for each region. Solid points represent significant F_{ST} outlier loci, and are labeled with a number reflecting the size of the AFLP band. F_{ST} outliers present in both regions are starred.

A. Eastern North America



B. Western North America

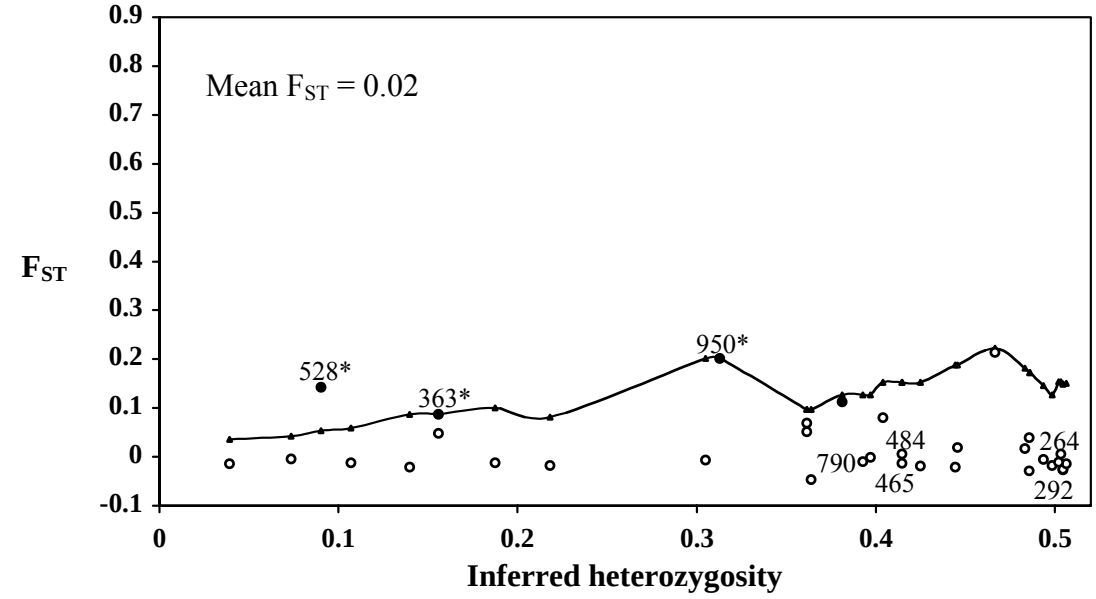
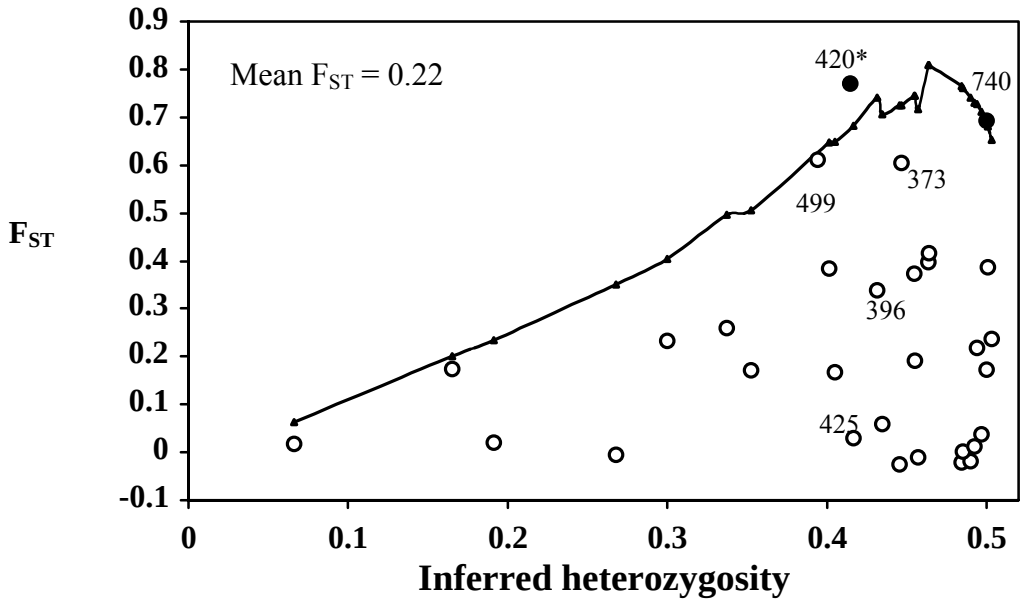
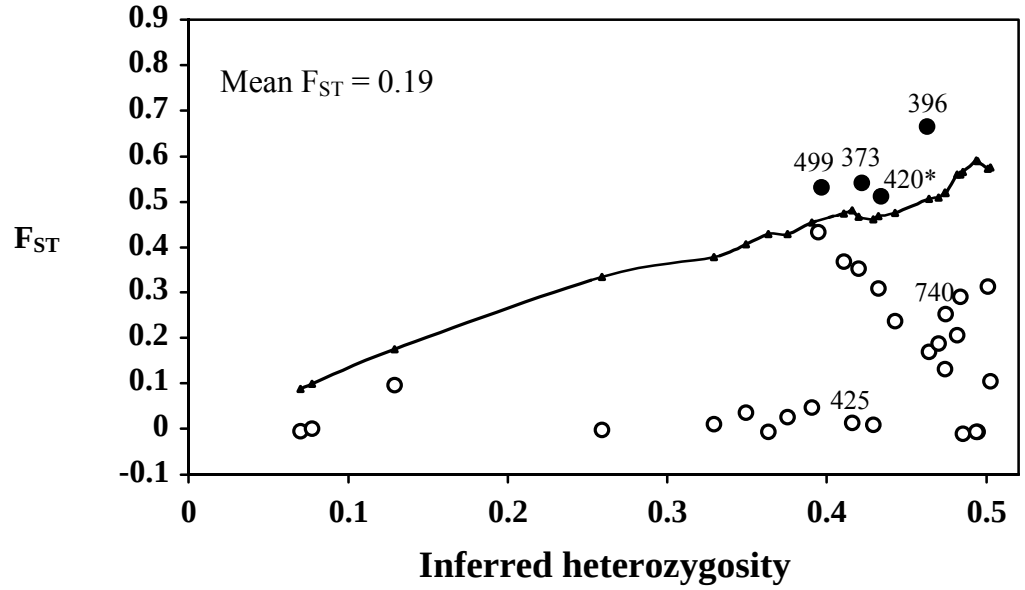


Figure 2. F_{ST} between aphids on alfalfa and clover estimated from variable AFLP loci are shown against heterozygosity estimates in three locations in Eastern North America, Iowa (A), Maryland (C), and New York (C) using 31 markers (Eastern dataset) and for New York using 83 markers (D). A solid line represents the 0.95 quantile of the neutral simulated distribution. The mean F_{ST} over all markers is indicated for each region. Solid points represent significant F_{ST} outlier loci, and are labeled with a number reflecting the size of the AFLP band. The F_{ST} outlier present in all three regions is starred.

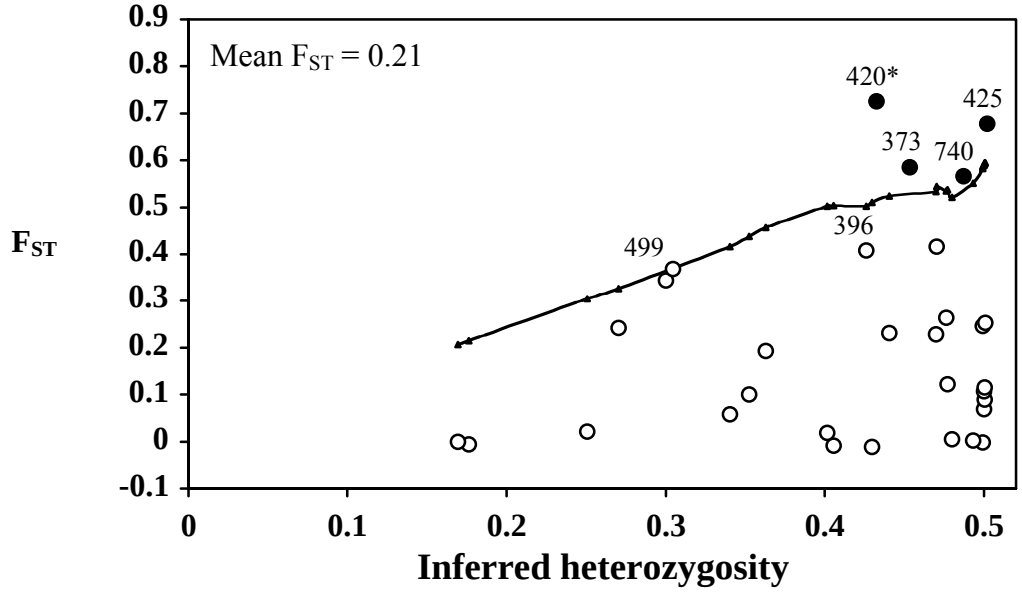
A. Iowa



B. Maryland



C. New York (31 loci)



D. New York (83 loci)

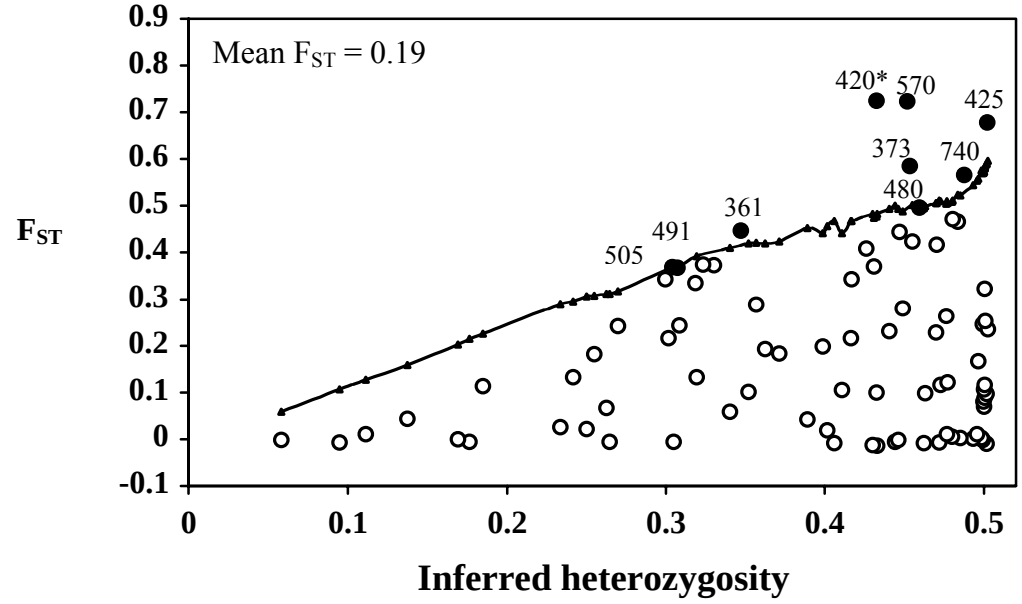


Figure 3. Genetic structure of pea aphids on alfalfa and clover in Eastern North America inferred with STRUCTURE using 31 AFLPs (Eastern Dataset). Each individual aphid is represented by a bar, and were sorted following assignment. (A) Using the admixture model, aphids were assigned proportionally into the two inferred clusters. (B) Using the population of origin information, the probability of admixture in the last three generations is calculated. Migrants (*) and hybrids (**) that were identified are indicated.

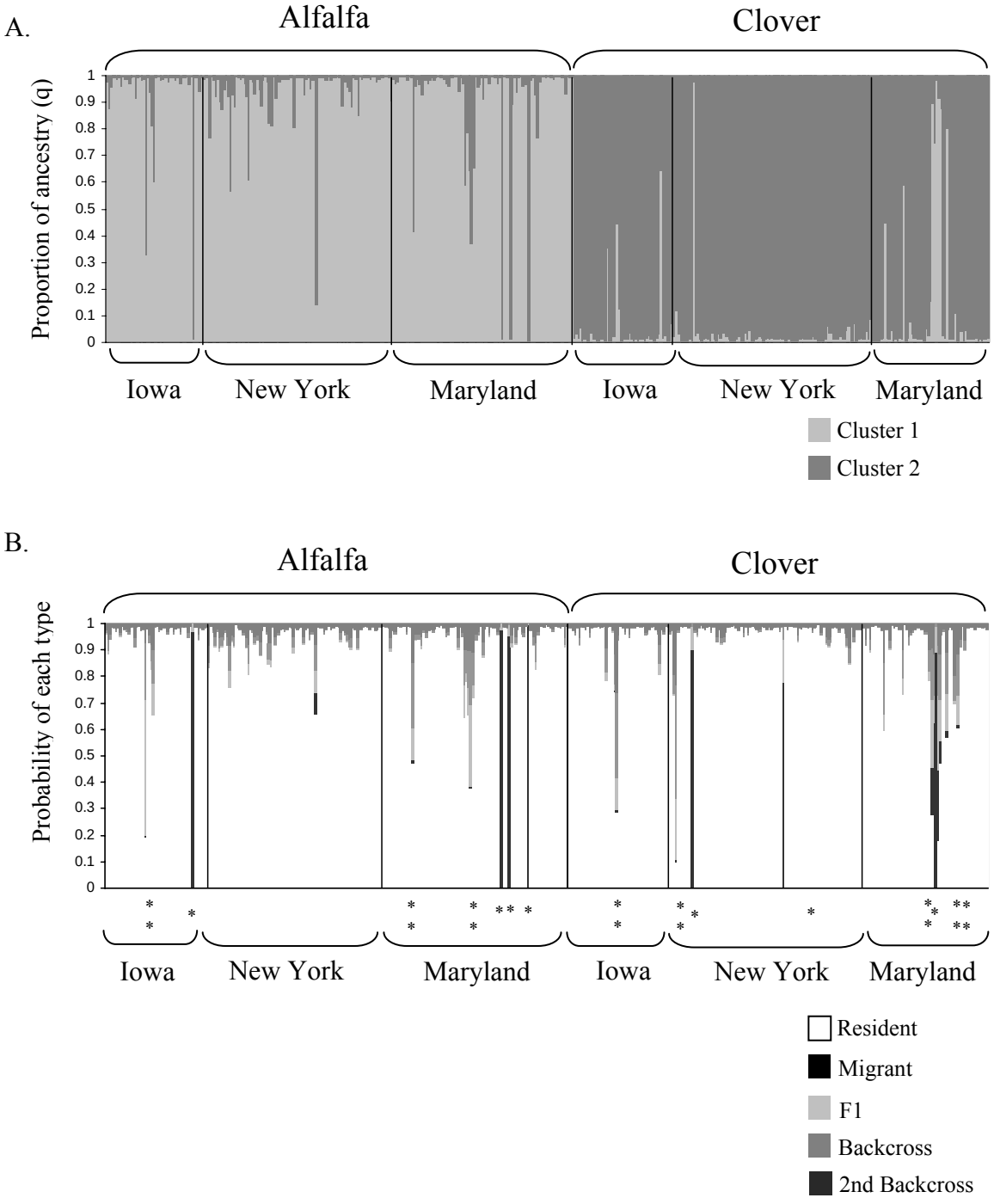


Figure 4. Genetic structure of pea aphids on alfalfa and clover in New York inferred with STRUCTURE using 83 AFLPs (New York Dataset). Each individual aphid is represented by a bar, and were sorted following assignment. (A) Using the admixture model, aphids were assigned proportionally into the two inferred clusters. (B) Using the population of origin information, the probability of admixture in the last three generations is calculated. Migrants (*) and hybrids (**) that were identified are indicated.

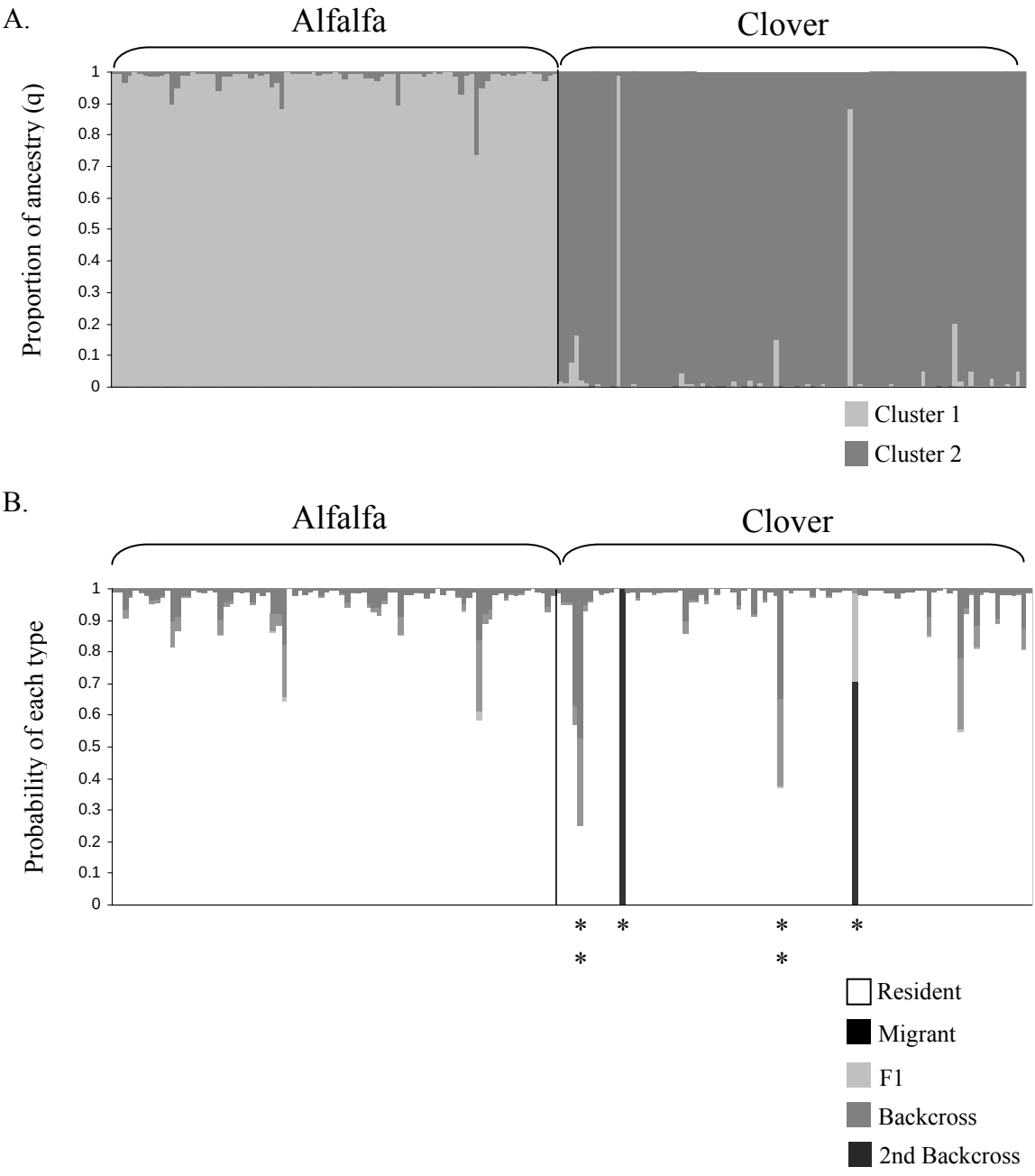
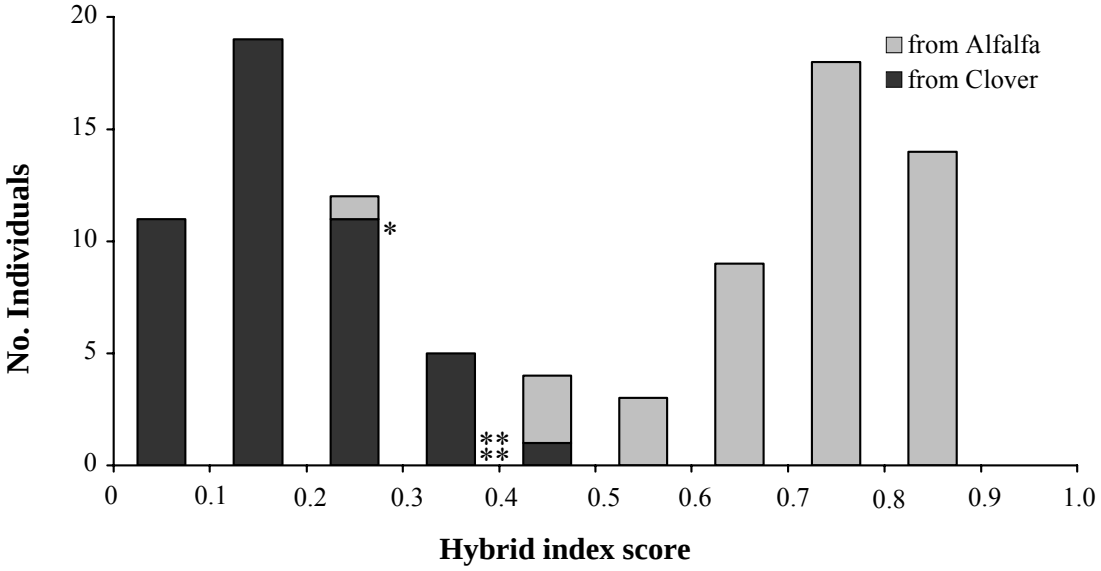
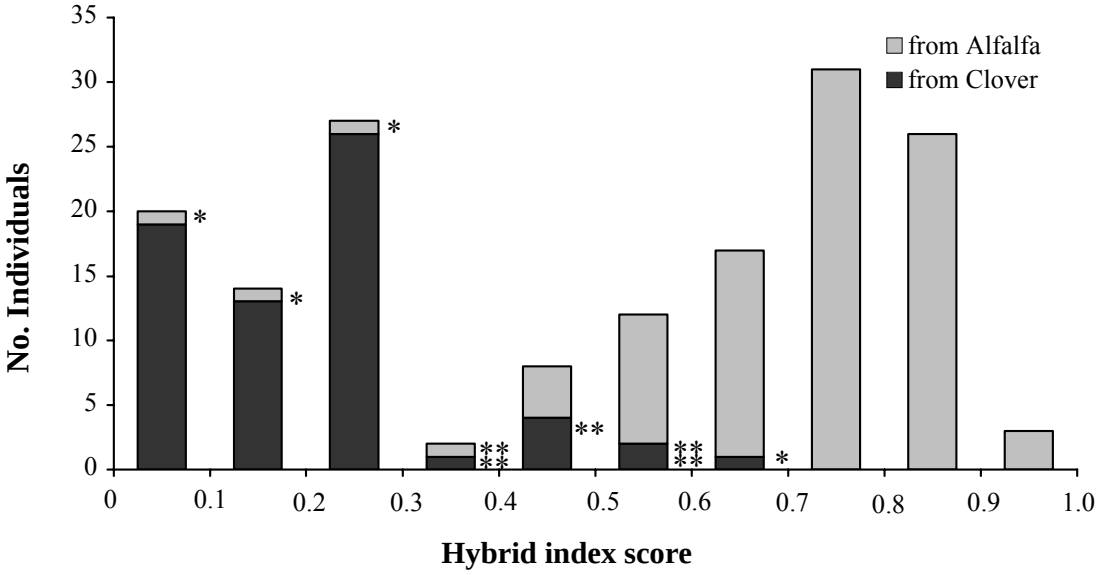


Figure 5. The distribution of individual hybrid index scores for aphids from alfalfa (light grey) and clover (dark grey) in Iowa (A), Maryland (B), and New York (C) (Eastern dataset). The hybrid index score for each individual is calculated as the proportion of highly differentiated markers showing the alfalfa phenotype. Migrants (*) and hybrids (**) identified using STRUCTURE are indicated to the right for comparison.

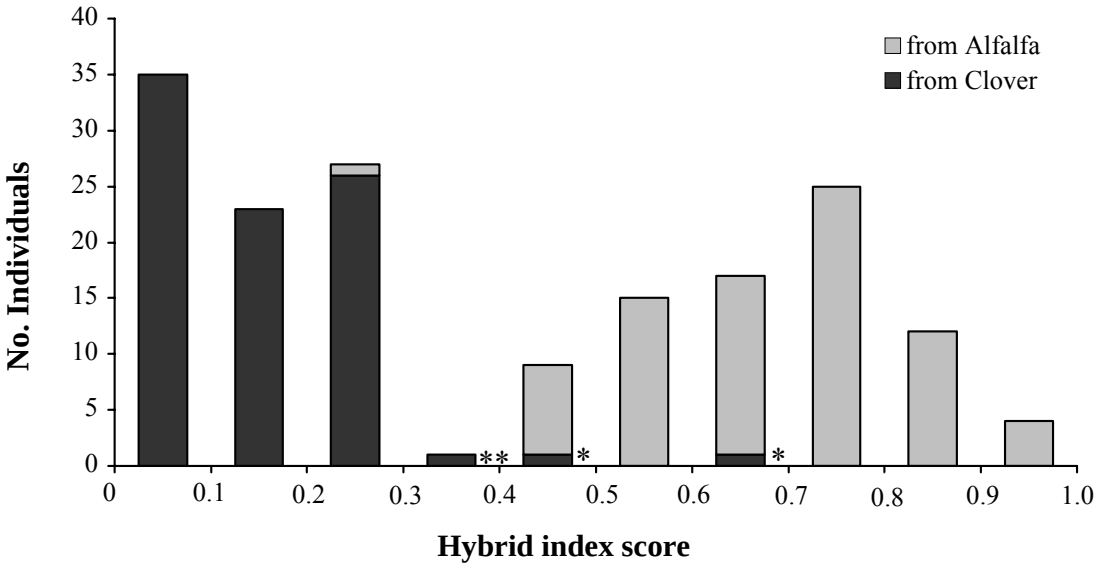
A. Iowa



B. Maryland



C. New York


40
35
30
25
20
15
10
5
0
No. Individuals
from Alfalfa
from Clover
**
*
*
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Hybrid index score

Chapter 3: Population structure of pea aphids on alfalfa and clover in North America and Europe and the origins of host plant specialization

Abstract

Two scenarios may explain the origin of sympatric insect host races onto geographically widespread alternate host plants. A single host shift may be followed by range expansion of already specialized populations. Alternately, an ancestral lineage could have colonized multiple locations and at each there were independent host shifts. Pea aphid host races on alfalfa and clover have recently expanded their range across Europe and into North America. The population structure of pea aphids was investigated using amplified-fragment length polymorphisms (AFLPs). Neighbor-joining and Bayesian analysis of population structure both provide evidence of a main divergence of aphids on alfalfa and clover. However, aphid populations group by location in both western North America and Sweden, suggesting either independent host shifts or ongoing gene flow at these locations.

Introduction

Divergent natural selection for resource specialization may lead to genetic divergence of populations and even speciation (Schluter 2001). One approach to the study of ecological speciation is to examine populations at early stages divergence, such as host-associated populations or host races, to understand the evolutionary processes and genetic changes contributing to speciation (Dres and Mallet 2002). Populations at early stages of divergence may be geographically widespread, and may differ in their degrees of ecological specialization and reproductive isolation across their range (Chapter 1). The process of divergence in each location may be influenced by population history and structure, which can shape genetic variation for specialization and reproductive isolation (Jiggins and Bridle 2004). Understanding the geographic context of the origin of ecological specialization may reveal why the rate or outcome of divergence can vary across geographic locations.

Two scenarios may explain the origin of specialized populations or host races onto geographically widespread alternate resources (Johannesson 2001). A single divergence onto two resources, either in sympatry or allopatry, may be followed by range expansions of already specialized populations across geographic locations (called "single divergence"). If the initial divergence did not result in complete reproductive isolation, secondary contact may result in hybridization or even homogenization of populations (Coyne and Orr 2004). Under other conditions, reinforcement could eventually drive evolution of reproductive isolation at each location (Rice and Hostert 1993, Howard 1993, Coyne and Orr 2004).

Alternately, a single ancestral lineage may expand its range into multiple regions with the same two resources. At some locations independently, there may be host shifts involving changes in the ability to recognize and use the new resource, and subsequent evolution of specialization onto the same two resources (called "parallel evolution of specialization"). Parallel evolution of specialization results in the convergent evolution of phenotypically similar pairs of host-associated populations at each location (Stanhope *et al.* 1993, McPeek and Wellborn 1998, Schluter 2001, Nosil *et al.* 2002, Rolan-Alvarez *et al.* 2004). Specialization and reproductive isolation can involve different genetic loci and alleles in each location (Hoekstra and Nachman 2003), though this is not always the case (Colosimo *et al.* 2005, Kronforst *et al.* 2006). When the adaptation to alternate resources directly or indirectly leads to reproductive isolation, parallel evolution of specialization can lead to parallel speciation (Schluter and Nagel 1995, Johannesson 2001). Determining whether the origin of a widely distributed set of host-associated populations involved a single divergence or parallel evolution of specialization may help to reveal the factors that shape the evolution of specialization and reproductive isolation across their range.

Phytophagous insects are often ecologically specialized on their host plants, and this specialization may be an important factor promoting their diversification (Ehrlich and Raven 1964; Eastop 1971, Mitter *et al*. 1988, Janz *et al.* 2006). Because many phytophagous insects mate on their host plants, host plant shifts and subsequent evolution of specialization directly lead to shifts in mate choice, causing assortative mating (Via 2001). Pea aphids (*Acyrthosiphon pisum pisum* Harris) on alfalfa (*Medicago sativa*) and red clover (*Trifolium pratense*) provide a model system in which to study the evolution of

host plant specialization and ecological-based divergence (Brisson and Stern 2006). Pea aphids and their host plants are thought to be native to Eurasia (Small 1996, Muller *et al*. 2003), though they have expanded their range and are now distributed worldwide in temperate climates (Eastop 1971). France is considered within the ancestral range of the pea aphid (Simon *et al.* 2003) while range expansion into Sweden (and England) is thought of have occurred in the last 1000 years. Pea aphids were first documented in North America around 1880 in the Midwest (Sanderson 1900, Folsom 1909). The number of introductions and sources of the introductions of the pea aphid in North America are unknown.

Pea aphids comprise host-associated populations on several leguminous host plants (Ferrari *et al.* 2006), but pea aphid populations on alfalfa and clover are more closely related to one another than to populations on pea (*Pisum sativum*), *Lotus*, *Cytisus scoparium*, and *Ononis repens* (West, unpublished data). Pea aphids on alfalfa and red clover are ecologically specialized and/or genetically structured in North America, Sweden, France, and England (Via 1991,Via 1999, Sandstrom 1996, Simon *et al.* 2003, Bournoville *et al.* 2004, Ferarri *et al.* 2006), though no study has directly compared ecological specialization or genetic differentiation among locations. In France, pea aphids are genetically differentiated on alfalfa and clover at allozymes and microsatellite markers (Simon *et al*. 2003), and there is genetic evidence for ongoing gene flow among host-associated populations (Frantz *et al.* 2006). In Sweden, aphids collected from both alfalfa and clover were each specialized on their host plant, and the performance of aphids from alfalfa on alfalfa increased during the course of the summer season, showing the effect of selection for host plant specialization (Sandstrom 1996). In North America,

aphid populations in the East and West of North America differ in levels of ecological specialization and genetic differentiation (Via 1991, 1999, Leonardo and Muri 2003, Chapter 1 and 2). In the East, pea aphids are highly specialized and genetically differentiated at nuclear markers (Via 1991, 1999, Via *et al*. 2000, Caillaud and Via 2000, Chapter 1). While eastern and western aphids from alfalfa were similar, aphids from clover in the west were less specialized than those from the east and genetically indistinguishable from aphids from alfalfa (Chapter 1). A different pattern was found by Leonardo and Muiru (2003) in California. They found clover specialists on white clover, and more generalized aphids on both alfalfa and clover that performed well on either plant.

Pea aphid population genetic structure has not been investigated at a broad geographic scale, so it is unclear if specialization arose once or several times (Simon *et al.* 2003). While mitochondrial sequence variation in pea aphids is low both in New York (Barrette *et al.* 1994) and in Europe (Birkle and Douglass 1999), nuclear genetic markers are highly variable in aphids (allozymes, Via 1999; STS markers, Chapter 1; microsatellites, Simon *et al.* 1999). Here, I investigate population structure of pea aphids using another nuclear genetic marker, amplified fragment length polymorphisms (AFLPs).

The goal of this research is to investigate the geographical context of the evolution of ecological specialization in pea aphids on alfalfa and clover across North America and in several European locations, and the genetic consequences of their introductions into North America. Specifically, I wanted to determine whether aphids on alfalfa and clover are the result of a single divergence onto alfalfa and clover followed by

multiple invasions of already specialized populations, or parallel evolution of specialization in some locations. Inferring population history from genetic patterns is notoriously difficult because the same genetic pattern can have several historical and demographic causes (Hare 2001, Knowles and Maddison 2002). Specifically, genetic similarity may be due to population history or ongoing gene flow. I examined the genetic structure of aphid populations at multiple, presumably neutral, genetic loci as a first step in understanding the origin of ecological specialization. Genetic similarity of populations at these loci is due to the combined effects of population history and neutral processes including ongoing gene flow. Grouping of populations by only host plant provides strong support for a single divergence, though there could be structure within the host associated clades if locations are somewhat isolated. Even a single grouping by location suggests either parallel evolution of specialization at that location or high ongoing gene flow (and distinguishing these alternatives will require further study). I used nuclear genetic markers to address the following questions:

1. Do levels of genetic variation differ between aphid populations on alfalfa and clover in each location, or between populations on each host plant across locations?

2. Are aphid populations more genetically similar to sympatric populations on alternate host plants, or allopatric populations on the same host across their range?

Methods

Aphid collections

Aphids were collected from alfalfa and clover in France, Sweden, and five locations in North America between 1996 and 1999 (Table 1). 6-10 aphids from each

location on each host plant were genotyped. The aphids from North America are the same as those used in the regional dataset in Chapter 2. Genomic DNA was extracted from aphids using the DNeasy Tissue Kit (QIAGEN). Single aphid adults were extracted individually, eluted into 200 uL of supplied buffer, and stored at -20°C.

AFLP genotyping

Aphids were genotyped for amplified fragment length polymorphisms (AFLPs; Vos *et al*. 1995), following the protocol described in Chapter 2, which is a modified procedure developed by Vos *et al.* (1995). Briefly, genomic DNA was digested with restriction enzymes, and adaptors were ligated to the ends. Two combinations of selective primers, based on a pair of core primer sequences (Pst1 and EcoR1), were used to amplify digested DNA (Table 1). Amplification products were analyzed on 4.3 % denaturing polyacrylamide gels (National Diagnostics), silver stained (Silver Sequence, Promega), and scored for band presence or absence manually.

Genetic diversity

Levels of genetic diversity were compared between aphids on alfalfa and aphids on clover and among geographic populations. Genetic diversity at AFLPs was measured using three statistics. The number of polymorphic sites and average number of pairwise differences were calculated in Arlequin (ver. 3.01, Excoffier *et al.* 2005). Nei's gene diversity within populations was estimated using the Bayesian method of allele frequency estimation of Zhivotvsky (1999) in AFLPsurv (Vekemans 2002). This estimation does assume Hardy-Weinberg Equilibrium (HWE), but previous analysis of codominant

markers (both sequence-based and allozymes) show these North American populations not deviating largely from HWE (Via 1999, Chapter 1).

Genetic structure

The genetic structure of pea aphids on alfalfa and clover across locations was analyzed using STRUCTURE 2.2 for AFLPs (Falush *et al*. 2007). I first determined the number of genetic populations (K) represented by the data by running the model for K = 1-10 with 5 independent runs for each K. I used the admixture model, correlated allele frequencies and no prior information because this is most sensitive for detecting subtle population structure (Falush *et al.* 2003). Models were run for a burn in period of 10^5 MCMC iterations and a data collection period of 10^6 MCMC iterations. Two methods were used to evaluate K. First, the maximal log probability of the data, Pr (X|K) was used to determine K (Pritchard *et al.* 2000). The modal value of ΔK was also used, which is based on the rate of change in the log probability of the data in consecutive runs for each K (Evanno *et al.* 2005). Individuals were assigned probabilistically into genetic populations defined by STRUCTURE for the most likely value of K. The proportion of ancestry of each predefined population (i.e. New York alfalfa, Maryland clover) in each inferred cluster was calculated. If there had been a single divergence and specialization onto these host plants, populations would group into clusters by host plant, whereas even a single grouping by geography suggests multiple host shifts or high ongoing gene flow. To examine differences among individuals within predefined populations and to identify possible migrants or hybrids, I calculated the admixture proportion for each individual (q), which is the proportion of ancestry each individual derived from each cluster.

Analysis of molecular variance (AMOVA) in Arlequin (Schneider *et al*. 2000) was used to determine if there was a significant partitioning of variation by host plant or geography, and to quantify the variation due to these factors. I used a hierarchical model for genotypic data and partitioned the genetic variation among host plants, among locations, and within populations.

Neighbor-joining population trees were constructed to visualize genetic similarities among aphid populations from alfalfa and clover in North America, France and Sweden in PHYLIP (version 3.6, Felsenstein 1989). Distances between populations were calculated as Nei's genetic distance, and support was assessed by bootstrapping over loci.

Results

Genetic diversity

Two AFLP primer combinations generated a total of 59 polymorphic markers over all populations. The number of polymorphic markers within each population varied between 20 and 36 (Table 2). Aphids on alfalfa from New York and Maryland had the highest number of polymorphic markers. Moderate levels of genetic diversity were observed within populations on both alfalfa and clover, with the average number of pairwise differences within populations varying from 6.6 to 14.7 and the Nei's gene diversity varied from 0.11 to 0.21. Aphids from alfalfa tended to have higher diversity than aphids from clover but not consistently and levels of variation were similar across locations (Table 2).

Genetic structure

Bayesian analysis of population structure using AFLP revealed genetic differentiation of aphid populations by both host plant and location. The number of genetic populations was not conclusive by Pritchard's method, because L(K) increased with increasing K, though the rate of increase declined at higher K (Figure 1A). However, the mode of ΔK was found to be three, indicating three genetic clusters (Figure 1B). Most predefined aphid populations showed high proportion of membership into one of the three inferred genetic clusters (Table 3, Figure 2), especially aphids from clover in the eastern North America and aphids from alfalfa in Sweden and France. Other populations showed evidence of mixed ancestry including aphids from clover in France and aphids from alfalfa in New York and Maryland.

The three inferred genetic clusters each included multiple predefined populations (Figure 2). Cluster 1 included aphids from alfalfa in North America and aphids from Washington clover. Cluster 2 included Swedish aphids from alfalfa and clover and French aphids from alfalfa. Cluster 3 included aphids on clover from Washington and aphids from clover from France. Aphids on alfalfa in North America were different from those in France and Sweden, while aphids on clover in eastern North America were similar to those on clover in Sweden. Finally, two populations grouped by location (Washington and Sweden). Thus, aphids group by both host plant and location.

While most individual aphids from the same host plant and location were similar, some populations had individuals from different clusters (Figure 2). For example, while most aphids in Washington had high proportion of ancestry in Cluster 1 (Figure 2), some

aphids in Washington had high proportion of membership in Cluster 3, which includes aphids from eastern clover (Figure 2).

AMOVA showed that most of the genetic variation was found within each population, and there was significant partitioning of the variation both within host plants and among locations ($p < 0.00001$, Table 4).

Neighbor-joining tree based population analysis using AFLPs were used to investigate genetic similarity of populations of pea aphids on alfalfa and clover across locations. Grouping of aphid populations from different locations onto the same host plant is expected if specialization involved a single, well-sorted divergence. Groupings by location suggest either parallel evolution of specialization or ongoing gene flow. I found that populations grouped by both host plant and geography, as found in STRUCTURE (Figure 3). Aphids from clover in eastern North America (New York, Maryland, and Iowa) form a clade with high bootstrap support (96%). In addition, aphids from Washington alfalfa and clover group with one another. This clade was identified in both neighbor-joining and STRUCTURE analyses.

Discussion

This study of aphid population structure in North America, France, and Sweden provides several important insights into the geographic context of pea aphid specialization on alfalfa and clover. Previous research has shown that the aphid specialization and divergence on alfalfa and clover described in North America (Via 1991, 1999, Via *et al.* 2001), is also present in France (Simon *et al.* 2003, Frantz *et al.* 2006) and Sweden (Sanderson 1996). However, it was unclear if specialization arose once in Europe and specialists insects were introduced to North America, or if

specialization evolved independently in North America and in Europe (Simon *et al.* 2003, Coyne and Orr 2004, Frantz *et al.* 2006). In this study, I found that levels of genetic diversity at AFLPs were similar in aphid populations on different host plants and between populations in the native and introduced range, suggesting that severe bottlenecks have not resulted from the recent colonization of North America. I then showed that aphid populations group mainly by host plant, suggesting a main single divergence. Aphids grouped by location in Sweden and Washington, suggesting either parallel evolution of specialization or ongoing gene flow in these locations. A combination of factors may influence aphid population structure, and pea aphid population history may involve a single divergence with high rates of hybridization in some locations.

Genetic diversity

Demographic changes, such as bottlenecks, and sampling during an introduction may alter genetic variation available for further evolution and divergence in the introduced range (Nei *et al.* 1975, Muller-Scharer *et al.* 2004). I found that genetic diversity of pea aphids on alfalfa and clover measured with AFLPs was similar across geographic locations. In addition, aphids from North America, part of the introduced range of the pea aphid, did not have decreased diversity compared to the European populations. While invasive and introduced species many times show reduced variation associated with founder events (Hufbauer *et al.* 2004, Hawley *et al.* 2006), they also may have similar or even elevated genetic variation compared to the native range due to multiple introductions, which can contribute to their success in the new range (Kolbe *et al.* 2004, Bossdorf *et al.* 2005). Multiple introductions and gene flow among previously independent lineages may provide broader genetic variation in introduced species (Levin

2003). In contrast to previous comparisons of diversity between eastern and western North America using sequence-based markers (Chapter 1), AFLPs did not show higher diversity in western North America, though AFLPs provide less resolution because they are dominant and only di-allelic. Future studies should compare genetic diversity of aphid populations across these locations using sequence-based markers.

Genetic structure

Population genetic analysis of aphids from North America, France and Sweden using AFLPs show that aphid populations grouped by host plant in most locations, though aphids from Sweden and Washington grouped by location. Therefore, the origin of pea aphid specialization cannot be entirely explained by either a single split onto alfalfa and clover with invasions of each host-associated populations ("single divergence") nor only independent host shifts and parallel evolution of specialization in each location ("parallel evolution"**)**. Instead, pea aphids on alfalfa and clover appear to be the result of a combination of these processes, or ongoing gene flow (Figure 4).

In France, part of the native pea aphid range, aphids on alfalfa and clover were genetically differentiated, as has been shown previously using different nuclear markers (Simon *et al*. 2003, Frantz *et al*. 2006). Initial divergence onto alfalfa and clover may have occurred in the native range (including France) and provided a source of specialists into several locations. Aphids on alfalfa in Sweden are genetically similar to those in France (Figure 2), suggesting Swedish alfalfa could have been colonized by specialized alfalfa migrants that had already diverged in France or a genetically similar source population. Clearly, not all possible source populations have been sampled in this study, so other genetically similar source populations still need to be considered.

Ecological specialization on clover in eastern North America did not arise locally. Aphids on clover in France and North America were genetically similar in both neighbor-joining tree analysis and in STRUCTURE, which is consistent with previous results using another type of nuclear marker, allozymes (Simon *et al*. 2003). This suggests the aphids on clover in the eastern North America were derived from already-specialized clover lineages, such as those in France.

Alfalfa specialization in eastern North America does not appear to have arisen after the introduction either. Neighbor-joining analysis showed that aphids on alfalfa in eastern North America were similar to those in France, but there was poor resolution in that part of the tree. Analysis in STRUCTURE showed that aphids from alfalfa in North America differed genetically from those on alfalfa in Europe. Could aphids from alfalfa in North America be from a different, unsampled source population? Historical information suggests that alfalfa was introduced from many locations, with much coming with the Spanish colonizers (Russelle 2001). Assuming that aphids were spread along with their host plants, a western introduction from Mexico or South America is possible (Russelle 2001). Alternatively, random sampling and demographic changes due to the introduction could have produced a shift in allele frequencies, resulting in the detection of another genetic cluster due to this bottleneck, even if the aphids came from France. Aphids on alfalfa in the east with ancestry in cluster 2 provide good evidence that these populations have a European origin, though they are probably not recent migrants. Future studies should sample aphids from a broader geographic range to distinguish these alternate hypotheses. Nevertheless, there is no evidence to suggest that specialization on alfalfa arose independently in North America.

Ongoing gene flow or parallel evolution of clover specialization?

Aphids from clover in both western North America and Sweden were genetically similar to aphids in those same locations on alfalfa. These results using AFLPs confirm results from a previous study of population structure in North America using sequence-based markers (Chapter 1). In that study, western populations on clover were also shown to be much less specialized on their host plants. This is the first time genetic differentiation between aphids on alfalfa and clover has been measured in Sweden, though pea aphids on alfalfa and clover in Sweden are specialized on their host plants (Sandstrom 1996).

Two factors may contribute to the genetic similarity of aphids from alfalfa and clover in Washington and Sweden: (1) parallel evolution of clover specialization (2) directional gene flow from aphids from alfalfa onto clover. Parallel evolution of clover specialization from introduced alfalfa populations could explain the genetic similarity of aphids on alfalfa and clover in western North America and Sweden. Aphid specialists on alfalfa but not clover may have been introduced to western North America and Sweden. Alfalfa populations may have genetic variation for clover use present at low frequency, which may be continually replenished due to introgression between host-associated populations in other parts of their range (Chapter 2). Selection acting on standing genetic variation for clover specialization could give rise to specialist clover populations in these locations. Clover populations would thus be genetically similar to alfalfa populations.

Directional migration and gene flow from alfalfa onto clover may constrain evolution despite strong selection, and even cause gene swamping (Sandoval 1994, Lenormand 2002). In western North America, alfalfa is more numerically abundant

(Barnes and Sheaffer 1995), the population size of alfalfa specialists could be greater, and the sheer numerical difference between alfalfa and clover specialists cause greater migration of aphids from alfalfa onto clover, and explain the genetic similarity of aphids on alfalfa and clover.

Directional gene flow could also be caused by variation among aphids in migratory tendency. In Sweden, Sandstrom (1996) suggested that differences among aphids on alfalfa and clover in host plant performance over the course of the season could be caused by differential migratory tendencies of aphids, and colonization of clover by less specialized aphids. He suggested that perhaps aphids on clover tended to stay on their host plants, while aphids from alfalfa migrated more frequently. Migratory tendency in aphids is due to the production of winged morphs (alates) which is induced by crowded conditions or poor host plant quality (Eastop 1971). Aphid clones vary substantially in their tendency to produce alates (Lamb and MacKay 1979, Bommarco and Ekbom 1996, Weisser and Braendle 2001). If migratory tendency is correlated with host plant use, the genetic similarity of aphid from clover in Sweden to the European alfalfa clade could be explained by high directional migration and introgression. Environmental or ecological changes (or in this case, changes in population distributions) can cause collapse of barriers to gene flow between previously divergent populations into a single hybrid swarm (Rhymer and Simberloff 1996, Grant and Grant 2002, Taylor *et al.* 2006). Both alfalfa and clover specialists may have been introduced to Sweden and western North America, but clover specialization and/ or genetic differentiation may be limited by hybridization with alfalfa migrants, so speciation may be going "in reverse" in these parts of the range (Taylor *et al.* 2006).

Sympatric speciation?

Pea aphids are frequently cited as an example of sympatric speciation (Coyne and Orr 2004). It is often pointed out that, while pea aphids are sympatric and highly divergent across much of their range, it is unclear if the origin of pea aphid specialization and reproductive isolation evolved in sympatry or allopatry (Coyne and Orr 2004). In addition, it has been unclear if specialization has evolved once or several times (Simon *et al*. 2003, Frantz *et al*. 2006). Much of the research on the mechanisms of pea aphid specialization and divergence has focused on populations in the introduced range (i.e. in eastern North America, Via 1991, 1999, Via *et al*. 2000, Hawthorne and Via 2001). These results shed light on these controversies. Pea aphids in eastern North America did not diverge from one another sympatrically in their current location (eastern North America). In fact, populations from alfalfa and clover may have been introduced from allopatric source populations (i.e. French clover and an unknown alfalfa population). However, the geographic context of the origin of pea aphid specialization is distinct from the geographic context of the maintenance of that specialization, and understanding the latter is critical for understanding the evolution of currently sympatric host-associated populations. Whether diverging populations continue to diverge and eventually form completely reproductively isolated species depends critically on current evolutionary processes and ecological factors. Focusing on the evolutionary forces that currently drive speciation rather than the geographic context of the origin of divergence reflects a current trend towards understanding the process of speciation (Via 2001, Schluter 2001). This research shows that even if the origin of specialist populations is the same (as it could be

for the East and West), the outcome of divergence can be very different across the geographic distribution of host-associated populations.

Conclusions

Several conclusions can be made about the origin of pea aphid populations across North America and parts of Europe from consideration of their population genetic structure. Aphids on the same host plant were genetically similar across many parts of their ranges, suggesting that populations in many locations have not evolved host use independently. While specialization and divergence of pea aphids onto alfalfa and clover could have occurred only once, the pattern of divergence is not uniform across their range, suggesting a geographic mosaic of ecological divergence that extends beyond Western North America (Chapter 1). This means that the pattern of ecological specialization and genetic differentiation varies across locations, and the pattern of genetic divergence can be strikingly different across the range of introduced taxa. Local processes are critical for the maintenance and the further evolution of pea aphid specialization. For example, hybridization may be causing high genetic similarity between aphids on alfalfa and clover in Sweden and western North America, and it could be enough to derail the process of divergence in these location. The outcome of specialization and divergence across the range may be determined by the interaction of local processes, because evolution can occur somewhat independently in each geographic location, and gene flow among geographic locations.

Tables

Table 1. Sequences of adaptors and primers used for AFLP protocols. Sequences are provided 5' to 3'.

EcoRI adaptors
AATTGGTACGCAGTC
CTCGTAGACTGCGTACC

PstI adaptors
TGTACGCAGTCTTAC
CTCGTAGACTGCGTACATGCA

The core sequence of primers for EcoRI and PstI amplicons
EcoRI core: GACTGCGTACCAATTC
Pst1 core: GACTGCGTACATGCAG

Selective Primer Combination 1
Pst1 overhang: CT
EcoRI overhang: TG

Selective Primer Combination 2
Pst1 overhang: AT
EcoRI overhang: TC

Table 2. Genetic diversity at AFLPs of aphids on alfalfa and clover within North American and European pea aphid populations. The number of aphids sampled (N) is shown.

Host plant	Population	N	No. poly-morphic markers	Ave. No. Pairwise differences within Pops.	Gene diversity ± S.E. (H_e)
Alfalfa	E. Washington	10	23	9.2	0.146 ± 0.179
	Iowa City, IA	10	36	13	0.201 ± 0.021
	Ithaca, NY	10	36	11.5	0.185 ± 0.020
	Beltsville, MD	9	33	12.5	0.192 ± 0.020
	Sweden	9	26	10.2	0.160 ±0.019
	France	6	33	14.7	0.218 ± 0.021
Clover	E. Washington	10	24	9.3	0.143 ± 0.018
	Iowa City, IA	10	20	6.6	0.116 ± 0.017
	Ithaca, NY	10	22	8.2	0.138 ± 0.019
	Beltsville, MD	8	23	8.2	0.143 ± 0.019
	Sweden	9	34	13.3	0.189 ± 0.019
	France	8	22	7.1	0.120 ± 0.016

Table 3. Proportion of membership of predefined aphid populations from each host plant and location in each of the three genetic clusters inferred in STRUCTURE using AFLPs. Proportions greater than 0.5 are in bold.

Predefined population		Inferred genetic cluster		
Host plant	Location	1	2	3
Alfalfa	E. Washington	**0.852**	0.018	0.131
	Iowa City, IA	**0.868**	0.022	0.110
	Ithaca, NY	**0.608**	0.326	0.066
	Beltsville, MD	**0.738**	0.190	0.072
	Sweden	0.015	**0.966**	0.019
	France	0.019	**0.971**	0.010
Clover	E. Washington	**0.771**	0.015	0.215
	Iowa City, IA	0.009	0.006	**0.985**
	Ithaca, NY	0.032	0.025	**0.943**
	Beltsville, MD	0.117	0.022	**0.861**
	Sweden	0.091	0.**815**	0.093
	France	0.012	0.265	**0.722**

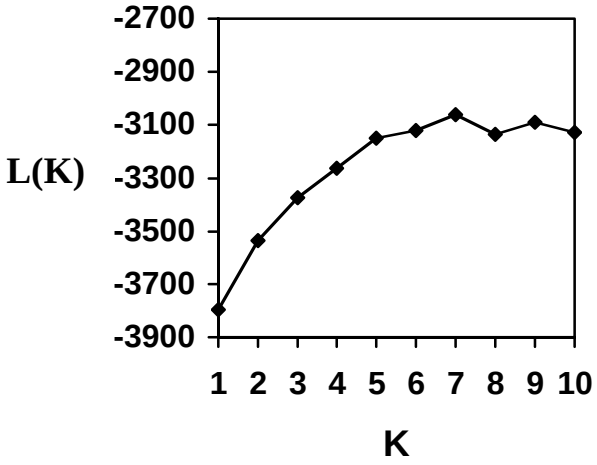
Table 4. AMOVA results for pea aphids on alfalfa and clover across North America, France and Sweden using AFLPs.

Source	d.f.	SS	Var.	% Var.	P
Among host plants	1	53.0	0.571	7.62	<0.00001
Within host plants	107	710.5	6.921	92.38	
Among locations	5	143.5	0.427	5.88	<0.00001
Within locations	103	620.1	6.845	94.12	

Figures

Figure 1. Model selection for STRUCTURE was determined two ways from five independent runs of K = 1 - 10 using AFLPs to determine the number of populations represented by the data. (A) The maximal value of the log-likelihood probability for the data for a given K [ln(Pr X|K) or L(K)] is suggested by Prichard *et al.* 2000). (B) Evanno *et al.* (2005)'s method uses the peak of ΔK, which is based on the rate of change in the log probability of the data in consecutive runs of K.

A. Pritchard's Method



B. Evanno's Method

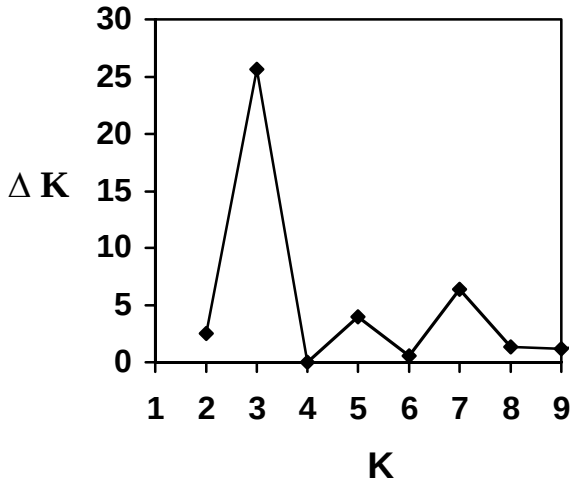


Figure 2. Genetic structure of pea aphids on alfalfa and clover in parts of their native and introduced ranges inferred with STRUCTURE using AFLPs. Each individual aphid is represented by a bar, and is assigned proportionally into the three inferred clusters. Aphids were sorted into collection location and host plant of origin following assignment.

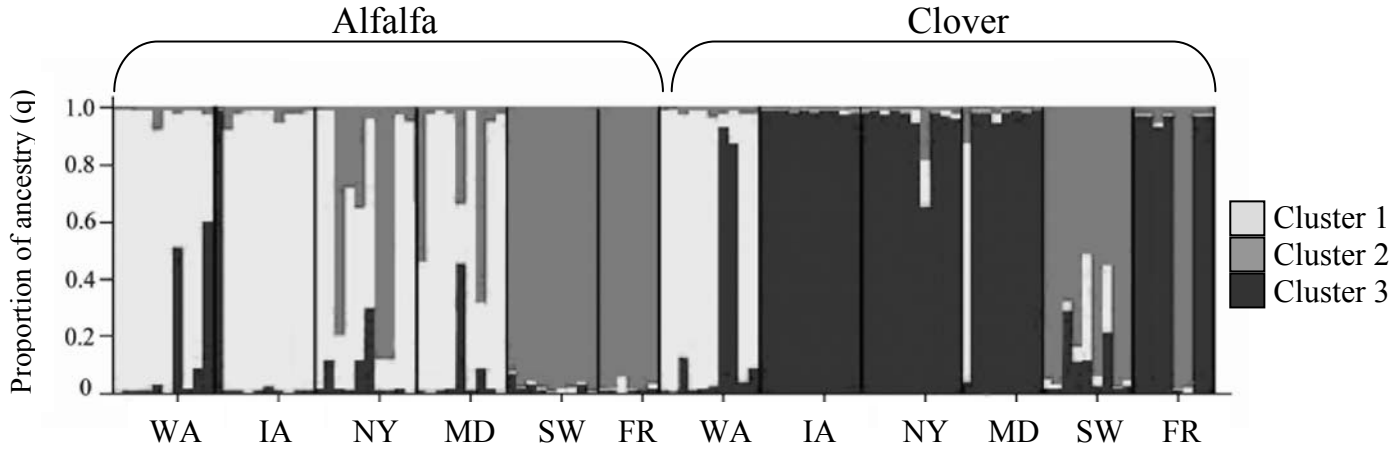


Figure 3. Neighbor-joining tree for aphid populations on alfalfa and clover using all 59 polymorphic AFLPs. Bootstrap values above 75% are shown. Two letter abbreviations are for each location. Populations from alfalfa are indicated by a blue triangle, and populations from clover are indicated by a red circle.

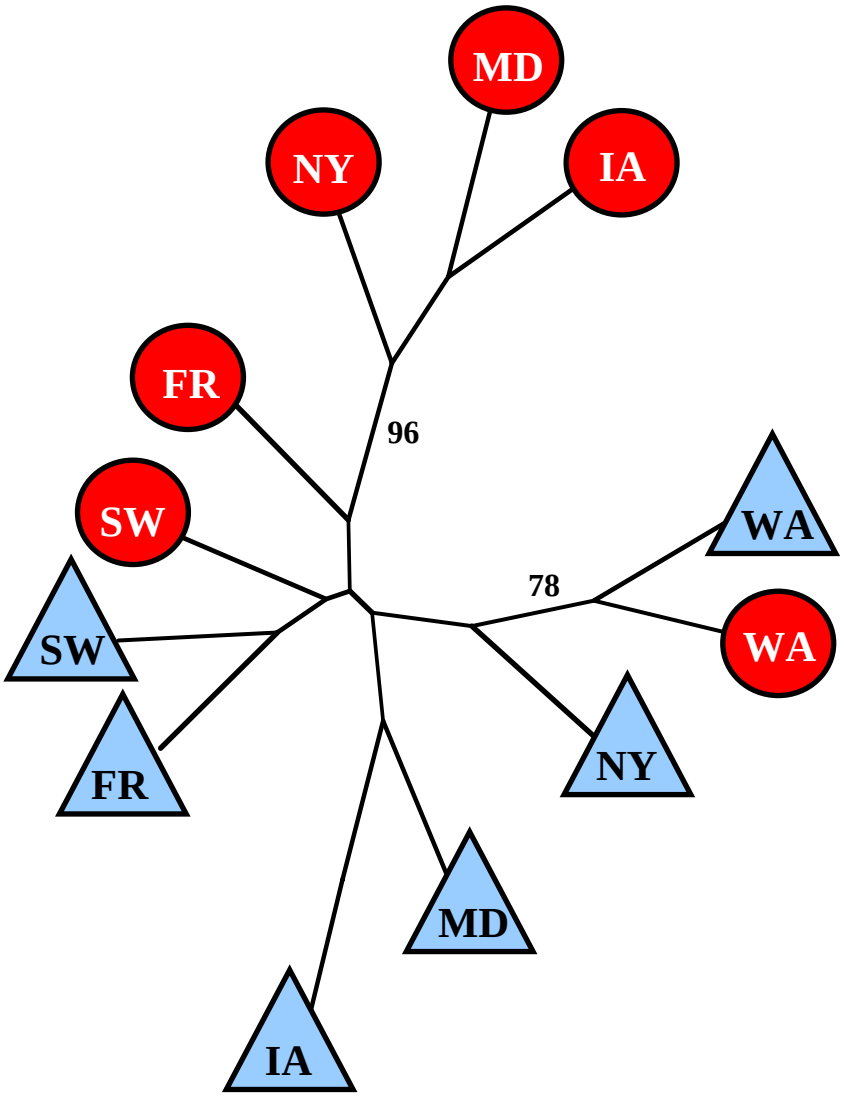
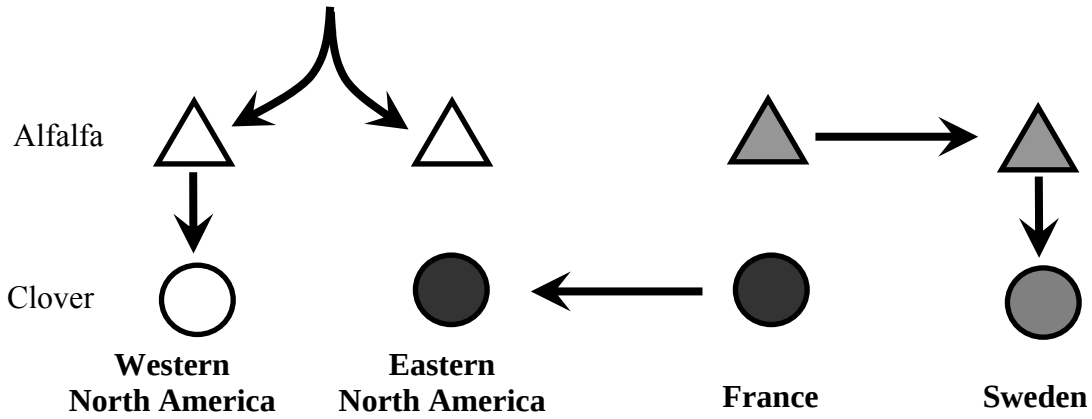


Figure 4. Hypothesized origin of host-associated population of pea aphids on alfalfa and clover. Genetically similar populations are shaded with the same color corresponding to inferred clusters in STRUCTURE. Arrows represent hypothesized introductions or ongoing gene flow.



Literature Cited

Aldridge, G. 2005. Variation in frequency of hybrids and spatial structure among *Ipomopsis* (Polemoniaceae) contact sites. New Phyotologist 167: 279-288.

Arnold, M. L., Bulger, M. R., Burke, J. M., Hempel, A. L., and J. H. Williams. 1999. Natural hybridization- how low can you go? (and still be important). Ecology 80: 371– 381.

Avise, J. C. 2000. Phylogeography. Harvard University Press, Cambridge, Mass.

Barnes, D. K., and C. C. Sheaffer. 1995. Alfalfa. P.p. 205-216. *in* R. F. Barnes, D. A. Miller, and C. J. Nelson. Forages. Iowa State University Press, Ames, Iowa.

Barraclough, T. G. and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. American Naturalist 155: 419-434.

Barrette, R.J., Crease, T.J., Herbet, P. D. N., and S. Via. 1994. Mitochondrial DNA diversity in the pea aphid *Acyrthosiphon pisum*. Genome 37: 858-865.

Barton, N. and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridizing populations. Heredity 57: 357-376.

Barton, N. and M. C. Whitlock. 1997. The evolution of metapopulations. P.p. 183-210 *in* I. Hanski and M. E. Gilpin. Metapopulation Biology. Academic Press, San Diego, California.

Beaumont, M. A. 2005. Adaptation and speciation: what can F_{ST} tell us? Trends in Ecology and Evolution 20:435-440.

Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. Proc. R. Soc. Lond. B 263: 1619-1626.

Berlocher S.H. and J. L. Feder. 2002. Sympatric speciation in phytophagous insects: Moving beyond controversy? Annual Review of Entomology 47: 773-815.

Bernays, E.A. 2001. Neural limitations in phytophagous insects: Implications for diet breadth and evolution of host affiliation. Annual Review of Entomology. 46: 703-727.

Bernays, E.A., M.S. Singer, and D. Rodrigues. 2004. Foraging in nature: Foraging efficiency and attentiveness in caterpillars with different diet breadths. Ecological Entomology. 29: 389-397.

Birkle, L. M. and A. E Douglas. 1999. Low genetic diversity among pea aphid (*Acyrthosiphon pisum*) biotypes of different plant affiliation. Heredity 82: 605-612.

Black, W. C., Baer, C. F., Antolin, M. F., and N. M. DuTeau. 2001. Population genomics: Genome-wide sampling of insect populations. Annual Review of Entomology 46: 441-469.

Blackman, R. L. 1987. Aphids: Their biology, natural enemies and control. Elsevier, Amsterdam.

Blackman, R.L. and V.F. Eastop. 2000. Aphids on the World's Crops: An Identification and Information Guide. 2nd ed. John Wiley and Son's, Ltd., Chichester, England.

Bolnick, D. I. and B. M. Firzpatrick. 2007. Sympatric speciation: Models and empirical evidence. Annual Review of Ecology and Evolution 38: 459-487.

Bommarco, R. and B. Ekbom. 1996. Variation in pea aphid population development in three different habitats. Ecological Entomology. 21: 235-240.

Bossdorf, O., Auge, H., Lafuma, L., Rogers, W. E., Siemann, E., and D. Prati. 2005. Phenotypic and genetic differentiation between native and introduced plant populations. Oecologia 144: 1-11.

Bournoville, R., Carré, S., Badenhausser, I., Simon, J.C., Hennis, C. and C. Greze. 2004. Host-race of the pea aphid, *Acyrthosiphon pisum*: biological criteria and feeding behaviour of clones originating from legumes. P. p. 413-419 J. C. Simon, C. A. Dedryver, C. Rispe and M. Hullé, eds. Aphids in a New Millennium. Proceedings of the Sixth International Symposium on Aphids.

Brisson J. A., and D. L. Stern. 2006. The pea aphid, *Acyrthosiphon pisum*: an emerging genomic model system for ecological, developmental and evolutionary studies. Bioessays 28: 747-755.

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

Bush, G. L. 1975. Modes of animal speciation. Annual Review of Ecology and Systematics 6: 339-364.

Bush, G. L. 1994. Sympatric speciation in animals: new wine in old bottles. Trends in Ecology and Evolution 9: 285-288.

Caillaud, M. C., Boutin, M., Braendle, C., and J. C. Simon. 2002. A sex-linked locus controls wing polymorphism in males of the pea aphid, *Acyrthosiphon pisum* (Harris). Heredity 89: 346-352.

Caillaud, M. C., and S. Via. 2000. Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. Am. Nat. 156: 606–621.

Campbell, D. 2003. Natural selection in *Ipomopsis* hybrid zones: implications for ecological speciation. New Phytologist 161: 83-90.

Campbell, D., and L. Bernatchez. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. Molecular Biology and Evolution 21: 945-956.

Carroll, S. P., Dingle, H. and S. P. Klassen. 1997. Genetic differentiation of fitness-associated traits among rapidly evolving populations of the soapberry bug. Evolution 51: 1182-1188.

Charlesworth, B., Nordborg, M., and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genetical Research 70: 155-174.

Clark, A. G. 1990. Inference of haplotypes from PCR-amplified samples of diploid populations. Molecular Biology and Evolution 7: 111-122.

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 307: 1928–1933.

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

Craig, T. P., Horner, J. D., and J. K. Itami. 1997 Hybridization studies on the host races of *Eurosta solidaginis*: Implications for sympatric speciation. Evolution 51: 1552-1560.

Cruz, R., Vilas, C., Mosquera, J., and C. Garcia. 2004. Relative contribution of dispersal and natural selection to the maintenance of a hybrid zone in *Littorina*. Evolution 58: 2734-2746.

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.

Davis, J. J. 1915. The pea aphid with relation to forage crops. Bulletin of the United States Department of Agriculture No. 276.

Denno, R. E., Roderick, G. K., Peterson, M. A., Huberty, A. F., Dobel, H. G., Eubanks, M. D., Losey, J. E., and G. A. Langellotto. 1996. Habitat persistence underlies intraspecific variation in the dispersal strategies of planthoppers. Ecological Monographs 66: 389-408.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400: 354-357.

Drès, M., and J. Mallet. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. Phil. Trans. R. Soc. B 357: 471–492 .

Eastop, V.F. 1971. Keys for the identification of *Acyrthosiphon* (Hemiptera: Aphididae). Bulletin of the British Museum (Natural History) Entomology. 26: 1–115.

Ehrlich, P. R. and P. H. Raven. 1964. Butterflies and Plants: A Study in Coevolution. Evolution. 18: 586-608.

ElMousadik, A., and R. J. Petit 1996. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L) Skeels] endemic to Morocco. Theoretical and Applied Genetics 92: 832-839.

Emelianov, I., Simpson, F., Narang, P., and J. Mallet. 2003. Host choice promotes reproductive isolation between host races of the larch budmoth *Zeiraphera diniana*. Journal of Evolutionary Biology 16: 208-218.

Emelianov, I., Marec, F., and J. Mallet. 2004. Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proceedings of the Royal Society of London B 271: 97-105.

Evanno, G. S., Regnaut, G. S., and J. Goudet. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14: 2611-2620.

Excoffier, L., Laval, G., Schneider, S. 2005. Arlequin, Version 3.0: An integrated software package for population genetics data analysis. Available at: http://cmpg.unibe.ch/software/arlequin3.

Falush, D., Stephens, M. and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-1587.

Falush, D., Stephens, M., and J. K. Pritchard. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7: 574-578.

Feder, J. L., Berlocher, S. H., Roethele, J. B., Dambroski, H., Smith, J. J., Perry, W. L., Gavrilovic, V., Filchak, K. E., Rull, J., and M. Aluja. 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in *Rhagoletis*. Proceedings of the National Academy of Sciences of the United States of America 100: 10314-10319.

Feder, J. L., and G. L. Bush. 1991. Genetic variation among apple and hawthorne host races of *Rhagoletis-Pomonella* across an ecological transition zone in the midwestern United States. Entomologia Experimentalis et Applicata 59: 249-265.

Felsenstein, J. 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics 5:164–166.

Fernandez, J., Galindo, J., Fernandez, B., Perez-Figueroa, A., Caballero, A., and R. Rolan-Alvarez. 2005. Genetic differentiation and estimation of effective population size and migration rates in two sympatric ecotypes of the marine snail *Littorina saxatilis*. Journal of Heredity 96: 460-464.

Ferrari, J., and H. C. J. Godfray. 2003. Resistance to a fungal pathogen and host plant specialization in the pea aphid. Ecology Letters 6: 111-118.

Ferrari, J., Godfray, H. C. J., Faulconbridge, A. S., Prior, K., A, and S. Via. 2006. Population differentiation and genetic variation in host choice among pea aphids from eight host plant genera. Evolution 60: 1574–1584.

Ferrari, J., Muller, C. B., Kraaijeveld, A. R., and H. C. J. Godfray. 2001. Clonal variation and covariation in aphid resistance to parasitoids and a pathogen. Evolution 55: 1805-1814.

Frazer, B. D. 1972. Population dynamics and recognition of biotypes in pea aphid (Homoptera- Aphididae). Canadian Entomologist 104: 1729-1733.

Fitzpatrick, B. M., and H. B. Shaffer. 2004. Environment-dependent admixture dynamics in a tiger salamander hybrid zone. Evolution 58: 1282-1293.

Flot, J. F., Tillier, A., Samadi, S., and S. Tillier. 2006. Phase determination from direct sequencing of length-variable DNA regions. Molecular Ecology Notes 6: 627-630.

Frantz, A., Plantegenest, M., Mieuzet, L., and J. C. Simon. 2006. Ecological specialization correlates with genotypic differentiation in sympatric host-populations of the pea aphid. Journal of Evolutionary Biology 19: 392-401.

Folsom, J. W. 1909. The insect pests of clover and alfalfa. Illinois Agricultural Experiment Station, Bulletin no. 134:111-197.

Fry, J.D. 2003. Detecting ecological trade-offs using selection experiments. Ecology. 84: 1672-1678.

Funk, D.J., K.E. Filchak, and J.L. Feder. 2002. Herbivorous insects: model systems for the comparative study of speciation ecology. Genetica. 116: 251-267.

Futuyma, D. J., Keese, M. C., and D. J. Funk. 1995. Genetic constraints on macroevolution- the evolution of host affiliation in the leaf beetle genus Ophraella. Evolution 49: 797-809.

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

Futuyma, D. J., and G. Moreno. 1988. The evolution of ecological specialization. Annual Review of Ecology and Systematics 19: 207-233.

Giles, B. E., and J. Goudet. 1997. Genetic differentiation in *Silene dioica* metapopulations: Estimation of spatiotemporal effects in a successional plant species. American Naturalist 149: 507-526.

Goudet J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3) Available at http://www.unil.ch/izea/softwares/fstat.html.

Grant, P. R., and B. R. Grant. 1997. Genetics and the origin of bird species. Proc Natl Acad Sci 94: 7768–7775.

Grant, P. R., and B. R. Grant. 2002. Unpredictable Evolution in a 30-Year Study of Darwin's Finches. Science 296: 707 – 711.

Halkett, F., Simon, J.C., and F. Balloux. 2005. Tackling the population genetics of clonal and partially clonal organisms. Trends in Ecology and Evolution 20: 194-201.

Hanski, I. and D Simberloff. 1997. The metapopulation approach. Pp. 5-26 *in* Hanski, I. and M. Gilpin, eds. Metapopulation biology, ecology, genetics and evolution. Academic Press.

Hare, M. P. 2001. Prospects for nuclear phylogeography. Trends Ecol. Evol. 16:700-706.

Harrison, R. G. 1991. Molecular changes at speciation. Annual Review of Ecology and Systematics 22: 281-308.

Hartl, D. L., and A. G. Clark. 1997. Principles of Population Genetics. Sinauer, Sunderland, Mass.

Hawley, D. M., Hanley, D., Dhondt, A. A., and I. J. Lovette. 2006. Molecular evidence for a founder effect in invasive house finch (*Carpodacus mexicanus*) populations experiencing an emergent disease epidemic. Molecular Ecology 15:263–275

Hawthorne, D. J., and S. Via. 2001. Genetic linkage of ecological specialization and reproductive isolation in pea aphids. Nature 412: 904-907.

Henter, H. J. and S. Via. 1995. The potential for coevolution in a host-parasitoid system. I. Genetic variation within an aphid population in susceptibility to a parasitic wasp. Evolution 49: 427-438.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167: 747-760.

Hey, J., and J. Wakeley. 1997. A coalescent estimator of the population recombination rate. Genetics 145: 833-846.

Herbert, P. D. N. and M. J. Beaton. 1989. Methodologies for allozyme analysis using cellulose acetate electrophoresis. Helena Laboratories. Beaumont, Texas.

Hoekstra, H. E., and M. W. Nachman. 2003. Different genes underlie adaptive melanism in different populations of rock pocket mice. Mol. Ecol. 12: 1185–1194.

How, S. T., Abrahamson, W. G., and T. P. Craig. 1993. Role of host-plant phenology in host use by *Eurosta-Solidaginis* (Diptera, Tephritidae) on *Solidago* (Compositae). Environmental Entomology 22: 388-396.

Howard, D. J. 1993. Reinforcement: the origin, dynamics and fate of an evolutionary hypothesis. In. Hybrid zones and the evolutionary process. Ed. R. G. Harrison. P. 118-142. Oxford University Press.

Hufbauer, R. A., Bogdanowicz, S. M., and R. G. Harrison. 2004. The population genetics of a biological control introduction: mitochondrial DNA and microsatellite variation in native and introduced populations of *Aphidus ervi*, a parasitoid wasp. Molecular Ecology 13: 337–348.

Itami, J. K., Craig, T. P., and J. D. Horner. 1998. Factors affecting gene flow between the host races of *Eurosta solidaginis*. Pp. 375-407. In: Genetic Structure and Local Adaptation in Natural Insect Populations. Eds. Mopper, S. and S. Y. Strauss. Chapman and Hall, New York.

Jaenike, J. 1981. Criteria for ascertaining the existence of host races. American Naturalist 117: 830-834.

Jaenike, J. 1990. Host Specialization in Phytophagous Insects. Annual Review of Ecology and Systematics. 21: 243-273.

Janz, N. and S. Nylin. 1997. The role of female search behaviour in determining host plant range in plant feeding insects: A test of the information processing hypothesis. Proceedings of the Royal Society of London Series B-Biological Sciences. 264: 701-707.

Janz, N., Nylin, S., and N. Wahlberg. 2006. Diversity begets diversity: host expansions and the diversification of plant-feeding insects. BMC Evolutionary Biology 6:4.

Jiggins, C. D. and J. R. Bridle. 2004. Speciation in the apple maggot fly: a blend of vintages? Trends in Ecology and Evolution 19: 111-114.

Jiggins, C. D., and J. Mallet. 2000. Bimodal hybrid zones and speciation. Trends in Ecology and Evolution 15: 250-255.

Johannesson, K. 2001. Parallel speciation: a key to sympatric divergence. Trends in Ecology and Evolution 16: 148–153.

Kareiva, P. M. 1983. Local movement in herbivorous insects- applying a passive diffusion model to mark-recapture field experiments. Oecologia 57: 322-327.

Knowles, L. L., and W. P. Maddison. 2002. Statistical phylogeography. Mol. Ecol. 11: 2623–2635.

Kolbe, J. J., Glor, R. E., Rodriguez-Schettino, L., Chamizo-Lara, A., Larson, A., and J. B. Losos. 2004. Genetic variation increases during biological invasion by a Cuban lizard. Nature 431: 177–181.

Kondrashov, A. S. and F. A. Kondrashov. 1999. Interactions among quantitative traits in the course of sympatric speciation. Nature 400: 351-354.

Kondrashov, A. S. and M. V. Mina. 1986. Sympatric speciation: when is it possible? Biological Journal of the Linnean Society 27: 201-223.

Kronforst, M. R., Young, L. G., Blume, L. M., and L. E. Gilbert. 2006. Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. Evolution 60: 1254–1268.

Lamb, R. J., and P. A. MacKay. 1979. Variation in migratory tendency within and among natural populations of the pea aphid, *Acyrthosiphon pisum*. Oecologia 39:289-299.

Latta, R. G. 2004. Gene flow, adaptive population divergence and comparative population structure across loci. New Phytologist 161: 51-58.

Le Corre, V., and A. Kremer. 2003. Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. Genetics 164: 1205-1219.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends in Ecology and Evolution 17: 183–189.

Leonardo, T. E. and Muiru, G. T. 2003 Facultative symbionts are associated with host plant specialization in pea aphid populations. Proceedings of the Royal Society of London. 270: S209-S212.

Levin, D. A. 2003. Ecological speciation: lessons from invasive species. Systematic Botany 28: 643– 650.

Lewontin, R. C., and J. Krakauer. 1973. Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. Genetics 74: 175-195.

Littel, R.C., Milliken, G. A., Stroup, W. W., and R.D. Wolfinger. 1996. SAS systems for mixed models. SAS Inst., Cary, NC.

Losos, J. B. and R. E. Glor. 2003. Phylogenetic comparative methods and the geography of speciation. Trends in Ecology and Evolution 18: 220-227.

Lu, G. Q., and L. Bernatchez. 1999. Correlated trophic specialization and genetic divergence in sympatric lake whitefish ecotypes (*Coregonus clupeaformis*): Support for the ecological speciation hypothesis. Evolution 53: 1491-1505.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., and P. Taberlet. 2003. The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics 4: 981-994.

Mallet, J. 2005. Speciation in the 21st Century. Heredity 95: 105-109.

Manel, S., Gaggiotti, O. E. and R. S. Waples. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology and Evolution 20: 136-142.

Mayr, E. 1963. Animal Species and Evolution. Belknap Press.

McCauley, D. E. 1991. The effect of host plant patch size variation on the population structure of a specialist herbivore insect, Tetraopes-Tetraophthalmus. Evolution 45: 1675-1684.

McKay, J. K., and R. G. Latta. 2002. Adaptive population divergence: markers, QTL and traits. Trends in Ecology and Evolution 17: 285-291.

McPeek, M. A., and G. A. Wellborn. 1998. Genetic variation and reproductive isolation among phenotypically divergent amphipod populations. Limnology and Oceanography 43: 1162-1169.

McVean, R. I. K., Dixon, A. F. G., and R. Harrington. 1999. Causes of regional and yearly variation in pea aphid numbers in eastern England. Journal of Applied Entomology 123: 495-502.

McVean, R.I.K. and A.F.G. Dixon. 2002. The host plant range of the pea aphid subspecies *Acyrthosiphon pisum* spp. *destructor* (Johnson) (Hom., Aphididae). Journal of Applied Entomology. 126: 281-286.

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

Mitter, C., Farrell, B. and B. Wiegmann. 1988. The Phylogenetic Study of Adaptive Zones: Has Phytophagy Promoted Insect Diversification? American Naturalist 132: 107-128.

Morjan, C. L. and L. H. Rieseberg. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Molecular Ecology 13: 1341-1356.

Muller, M. H., Prosperi, J. M., Santoni, S., and J. Ronfort. 2003. Inferences from mitochondrial DNA patterns on the domestication history of alfalfa (*Medicago sativa*). Molecular Ecology 12: 2187-2199.

Muller-Scharer, H., Schaffner, U., and T. Steinger. 2004. Evolution in invasive plants: implications for biological control. Trends Ecol. Evol. 19: 417–422.

Nachman, M. W. 2005. The genetic basis of adaptation: lessons from concealing coloration in pocket mice. Genetica 123: 125-136.

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

Nielson, R. and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158: 885-896.

Noor, M. A. F., and J. L. Feder. 2006. Speciation genetics: evolving approaches. Nature Reviews Genetics 7: 851-861.

Nosil, P., Crespi, B. J., and C. P. Sandoval. 2002. Host–plant adaptation drives the parallel evolution of reproductive isolation. Nature 417: 440–443.

Nosil, P., Sandoval, C. P., and B. J. Crespi. 2006. The evolution of host preferences in allopatric vs. parapatric populations of *Timema cristinae* walking-sticks. Journal of Evolutionary Biology 19: 929-942.

Orr, H. A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics 139: 1805-1813.

Peterson, M. A., and R. F. Denno. 1998. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. American Naturalist 152: 428-446.

Pettit, R.H. 1905. Insects of the garden. Michigan State Agricultural College Experiment Station Bulletin. 233: 41-41.

Pritchard, J. K., Stephens, M., and P. Donnelly. 2000 Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Putnam, D. 1997. History, importance, and production dynamics of alfalfa in California. Proceedings, 27th National Alfalfa Symposium and 26th Annual California Alfalfa Symposium, December 9-10, Sand Diego, CA.

Raymond, M. and F. Rousset. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity, 86:248-249.

Rhymer, J. M., and D. S. Simberloff. 1996. Extinction by hybridization and introgression. Annual Review of Ecology and Systematics 27: 83–109.

Rice, W. R., and E. E. Hostert. 1993. Laboratory Experiments on Speciation: What Have We Learned in 40 Years? Evolution 47:1637-1653.

Rieseberg, L. H., and J. M. Burke. 2001. The biological reality of species: gene flow, selection, and collective evolution. Taxon 50: 47-67.

Rolan-Alvarez, E., Caballo, M., and J. Galindo. 2004. Nonallopatric and parallel origin of local reproductive barriers between two snail ecotypes. Molecular Ecology 13: 3415-3424.

Rundle, H. D. and P. Nosil. 2005. Ecological speciation. Ecology Letters 8: 336-352.

Russelle, M. 2001. Alfalfa. Am. Sci. 89: 252-259.

Sanderson, E. D. 1900. The destructive pea louse in Delaware. Delaware Agricultural Experiment Station Bulletin 49: 14-24.

Sandoval, C. P. 1994. The effects of the relative geographic scales of gene flow and selection on morph frequencies in the walking-stick *Timema cristinae*. Evolution 48: 1866-1879.

Sandström, J. 1996. Temporal changes in host adaptation in the pea aphid, *Acyrthosiphon pisum*. Ecol. Entomol. 21:56–62.

Savolainen, V., Anstett, M. C., Lexer, C., Hutton, I., Clarkson, J. J., Norup, M. V., Powell, M. P., Springate, D., Salamin, N., and W. J. Baker. 2006. Sympatric speciation in palms on an oceanic island. Nature 441: 210-213.

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. American Journal of Human Genetics 78: 629-644.

Scheffer, S. J., and D. J. Hawthorne. 2007. Molecular evidence of host-associated genetic divergence in the holly leafminer *Phytomyza glabricola* (Diptera : Agromyzidae): apparent discordance among marker systems. Molecular Ecology 16: 2627-2637.

Schemske, D. W., and P. Bierzychudek. 2001. Perspective: Evolution of flower color in the desert annual *Linanthus parryae*: Wright revisited. Evolution 55: 1269-1282.

Schilthuizen, M. 2000. Dualism and conflicts in understanding speciation. Bioessays 22: 1134-1141.

Schilthuizen, M. 2001. Frogs, Flies, and Dandelions. Oxford University Press, Oxford, U. K.

Schliewen, U. K., Diethard Tautz, D., and S. Pääbo. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368, 629 - 632

Schluter, D. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372–380.

Schluter, D., and L. Nagel. 1995. Parallel speciation by natural selection. Am. Nat. 146: 292-301.

Schneider, S., Roessli, D. and L. Excoffier. 2000. Arlequin: a software for population genetic analysis. User manual ver. 2.000. Genetics and Biometry Lab. Dept. of Anthropology. University of Geneva, Geneva.

Scriber, J. M. 2002. Latitudinal and local geographic mosaics in host plant preferences as shaped by thermal units and voltinism in *Papilio* spp. (Lepidoptera). European Journal of Entomology 99: 225-239.

Simon, J.-C., S. Baumann, P. Sunnucks, P. D. N. Hebert, J.-S. Pierre, J.-F. Le Gallic, and C.-A. Dedryver. 1999. Reproductive mode and population genetic structure of the

cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. Mol. Ecol. 8: 531–545.

Simon, J.-C., S. Carré, M. Boutin, N. Prunier-Leterme, B. Sabater-Muñoz, A. Latorre, and R. Bournoville. 2003. Host-based divergence in populations of the pea aphid: insights from nuclear markers and the prevalence of facultative symbionts. Proc. R. Soc. Lond. B. 270:1703–1712.

Small, E. 1996. Adaptations to herbivory in alfalfa (*Medicago sativa*). Can. J. Bot. 74: 807-822.

Sork, V. L., Nason, J., Campbell, D. R., and J. F. Fernandez. 1999. Landscape approaches to historical and contemporary gene flow in plants. Trends in Ecology and Evolution 14: 219-224.

Stanhope, M. J., B. Hartwick, and D. Baillie. 1993. Molecular phylogeographic evidence for multiple shifts in habitat preference in the diversification of an amphipod species. Mol. Ecol. 2::99–112.

Strauss, S. Y. and R. Karban. 1998. The strength of selection: Intraspecific variation in the host-plant quality and the fitness of herbivores. Pp. 156-177 *in* S. Mopper and S. Y. Strauss, eds. Genetic Structure and Local Adaptation in Natural Insect Populations, Chapman and Hall, New York.

Stireman, J.O. and M.S. Singer. 2003. What determines host range in parasitoids? An analysis of a tachinid parasitoid community. Oecologia. 135: 629-638.

Storz, J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology 14: 671-688.

Sun, R. Y. and A. G. Robinson. 1966. Chromosome studies on 50 species of aphids. Canadian Journal of Zoology 44: 649.

Taylor, E. B., Boughman, J. W., and M. Groenenboom. 2006. Speciation in reverse: morphological and genetic evidence of the collapse of a three-spined stickleback (*Gasterosteus aculeatus*) species pair. Molecular Ecology 15: 343–355.

Taylor, N. L. and K. H. Quesenberry. 1996. Red Clover Science. Kluwer Academic Publishers, Boston.

Thompson, J. N. 1994. The Coevolutionary Process. The University of Chicago Press, Chicago.

Thompson, J. N. 1999. Specific hypotheses on the geographic mosaic of coevolution. American Naturalist 153: S1-S14.

Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. The University of Chicago Press, Chicago.

Thrall, P. H., and J. J. Burdon. 2002. Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. Plant Pathology 51: 169-184.

Tsuchida, T., Koga, R., and T. Fukatsu. 2004. Host plant specialization governed by facultative symbiont. Science 303: 1989.

Turelli, M., N.H. Barton, and J.A. Coyne. 2001. Theory and speciation. Trends in Ecology and Evolution. 16: 330-343.

Turner, T. L., Hahn, M. W., and S. V. Nuzhdin. 2005. Genomic islands of speciation in *Anopheles gambiae*. PLOS BIOLOGY 3: 1572-1578.

Vaha, J. P., and C. R. Primmer. 2006. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology 15: 63-72.

Vasemagi, A., and C. R. Primmer. 2005. Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Molecular Ecology 14: 3623-3642.

Vekemans X. 2002. AFLP-SURV version 1.0. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Belgium.

Via, S. 1991. The genetic structure of host plant adaptation in a spatial patchwork: demographic variability among reciprocally transplanted pea aphid clones. Evolution 45: 827-852.

Via, S. 1999. Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. Evolution 53: 1446-1457.

Via, S., Bouck, A. C., and S. Skillman. 2000. Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. Evolution. 54:1626–1637.

Via, S. 2001. Sympatric speciation in animals: the ugly duckling grows up. Trends in Ecology and Evolution 16: 381-390

Via, S. 2002. The ecological genetics of speciation. American Naturalist 159: S1-S7.

Vitalis, R., Dawson, K., and P. Boursot. 2001. Interpretation of variation across marker loci as evidence of selection. Genetics 158: 1811-1823.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van der Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995. AFLP: A new technique for DNA fingerprinting. Nucleic Acids Res. 23:4407-4414.

Wade, M. J. and C. J. Goodnight. 1998. Perspective: The theories of Fisher and Wright in the context of metapopulations: When nature does many small experiments. Evolution 52: 1537-1553.

Waring, G. L., Abrahamson, W. G., and D. J. Howard. 1990. Genetic differentiation among host-associated populations of the gallmaker Eurosta-Solidaginis (Diptera, Tephritidae). Evolution 44: 1648-1655.

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

Weisser, W.W., and C. Braendle. 2001. Body colour and genetic variation in winged morph production in the pea aphid. Entomol Exp Appl 99:217–223

Westgate, J. M. 1908. Alfalfa. USDA Farmers' Bulletin 339.

Wilding, C. S., Butlin, R. K., and J. Grahame. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. Journal of Evolutionary Biology 14: 611-619.

Wilson, A. B., Noack-Kunnmann, K., and A. Meyer. 2000. Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological differentiation. Proc. R. Soc. Lond B. 267: 2133-2141.

Wright, S. 1940. Breeding structure of populations in relation to speciation. American Naturalist 74: 232-248.

Wright, S. 1969. Evolution and the genetics of populations. University of Chicago Press. Chicago.

Wu, C.I. 2001. The genic view of the process of speciation. Journal of Evolutionary Biology 14: 851-865.

Wu, C. I., and C. T. Ting. 2004. Genes and speciation. Nature Reviews Genetics 5: 114-122.

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

Zhivotovsky, L. A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology 8: 907-913.