THE GENETICS OF SPECIATION AND THE ORIGIN OF GENOMIC DIVERGENCE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Erik B. Dopman

August 2005

THE GENETICS OF SPECIATION AND THE ORIGIN OF GENOMIC DIVERGENCE

Erik B. Dopman, Ph. D.

Cornell University 2005

Speciation involves the origin of trait differences that limit or prevent gene exchange and ultimately results in daughter populations that form monophyletic or exclusive genetic groups. However, for recently diverged populations or species, between which reproductive isolation is often incomplete, gene genealogies will be discordant and most regions of the genome will display nonexclusive genealogical patterns. In these situations, genome regions for which one or both species are exclusive groups may mark the footprint of recent selective sweeps. Alternatively, such regions may include or be closely linked to "speciation genes," genes involved in reproductive isolation. Therefore, comparisons of gene genealogies allow inferences about the genetic architectures of both reproductive isolation and adaptation. Contrasting genealogical relationships in sexually isolated Z and E pheromone strains of the European corn borer moth (ECB) demonstrate the relevance of this approach.

Genealogies for five gene regions in ECB are discordant, and for only one molecular marker, the sex-linked gene *Tpi*, are the two pheromone strains exclusive groups. A genetic linkage map provides the context for understanding genealogical discordance. The factors responsible for sexual isolation, male behavioral response (*Resp*) and female pheromone production (*Pher*), and the factor causing temporal isolation, post-diapause development (*Pdd*), were placed on a linkage map that also contained the mapping positions of the gene genealogies. *Pher* maps to an autosome,

whereas *Resp* and *Pdd* are sex-linked. The exclusive gene, *Tpi*, maps to a position on the sex chromosome that is indistinguishable from *Pdd*; *Resp* maps 20-30 cM away. Neutral demography involving population expansion and population substructure can explain most genetic patterns among loci; however, *Tpi* shows evidence for non-neutral evolution. Because *Tpi* is tightly linked to *Pdd*, recent evolution of this reproductive barrier may be responsible for the evolution of genetic divergence between these incipient species.

BIOGRAPHICAL SKETCH

Erik Bryant Dopman, a Taurus, was born in Clearwater, Florida on May 18, 1976 to William Conrad Dopman and Mary Therese McCarty. Before he was one, Erik moved with his mother, father and older brother, Jim, to Freeport, Illinois. In the small industrial town of Freeport, Erik managed to pick up the basics in class through his junior year. Alas, he was denied the distinction of acquiring Freeport Senior High's letter jacket, which prominently displayed the school's mascot, the salted pretzel. Erik moved from Illinois and finished secondary school in Garland, Texas, where it was more common for students to be "panthers," or "rams," rather than a scrumptious bar snack.

Following high school, Erik joined 49,999 of his peers at the University of Texas at Austin. At UT, Erik welcomed anonymity and really did not mind becoming student number AE613…42. Erik's focused temperament mixed quite well with the independent can-do attitude drilled by the Texas educational system. These qualities served Erik at UT by helping him navigate the many opportunities that were available to the initiated and motivated. Erik's experiences at UT help to solidify his interests in biology, and he became involved in independent research under the mentorship of Greg Sword and David Hillis. It was at a scientific meeting while presenting his undergraduate research that he met his future graduate school mentor, Rick Harrison. Erik discovered that he and Rick shared many interests and they both decided that collaboration would be fruitful, but more importantly, fun!

ACKNOWLEDGMENTS

Special thanks go to my major advisor Rick Harrison for helping me sharpen my scientific thinking, for always being available to talk, and for encouraging me to make room for life's many experiences. I hope to advise my students using him as an example and I am looking forward to our continued friendship and collaboration.

Charles Aquadro and Amy McCune rounded out my committee and offered insight from different perspectives. For their advice, I am very grateful. Of course, this work could not have been accomplished without the support from my lab mates. Steve Bogdanowicz offered guidance and patience early on, and he became a good friend as the years passed. I am especially indebted to Steve, Sean Mullen, Luana Maroja, and Jose Andres for helping me find humor in every day. I would also like to thank all of my good friends in Ithaca who helped make graduate school a wonderful experience.

I would like to thank Wendell Roelofs, Paul Robbins, Charles Linn, Kathy Poole, Paul Feeny, Jim Walgenbach, George Kennedy, Denis Bourguet, and Gabor Szocs for assistance and advice with ECB. I also want to thank Peter Andolfatto, Shankar Iyer, and Brielle Fischman for their assistance with the molecular genetic work. Kevin Thornton was especially helpful with the modeling.

My research was supported by the Graduate School at Cornell University, the Department of Ecology and Evolutionary Biology at Cornell University, a U.S. Environmental Protection Agency Science to Achieve Results fellowship for graduate environmental study (U-94589501-1), and by two grants to Rick Harrison from the U.S. Department of Agriculture (2001-35302-11123) and from the National Science Foundation (DEB-0415343).

Finally, I want to thank my dearest Leila. I never would have thought that companionship for my heart, mind, and body was so near for so many years. She truly is an amazing woman.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Why study speciation?

If asked to make a collection of leaves that dot the path during a forest walk, a person of just about any age will be forced to appreciate the abundance of variation in leaf shape and size. What becomes immediately apparent, even for young children, is the obvious way in which variation among leaves is partitioned. Clear boundaries exist that divide patterns in nature; leaves of similar shape, such as the needles gathered from multiple white pine trees, display fewer differences than leaves of different shape, such as between pine needles and the flat, broad leaves from white maple trees. Thus, biological variation in nature is discontinuous. The discrete clusters that are produced from these discontinuities, namely species, form the fundamental unit of biological organization in nature.

In 1859, observations of morphological, behavioral, and physiological variation provided the impetus for Darwin to outline the most compelling explanation for the origin of biological variation – gradual evolution by natural selection (DARWIN 1859). This work provided an explanation for variation within species. However, neither Darwin nor his contemporaries were able account for how the gradual, continuous process of evolution could produce the discontinuities that were ubiquitous in nature.

During the "modern" evolutionary synthesis of Mendelian genetics and Darwinian evolution, Dobzhansky and Mayr focused attention on the breeding patterns of co-occurring populations as providing a critical solution to the "species problem" that had plagued biologists for over a century (DOBZHANSKY 1937; MAYR

1942). Specifically, the splitting of an ancestral species into two or more daughter species, and the persistence of species where sympatric, could be explained by the evolution of reproductive isolation between populations. The evolution of traits that impede interbreeding results in populations that, as a consequence of isolation, can evolve unique differences that accrue over evolutionary time. In other words, it is the origin of reproductive barriers that permits the production and/or persistence of discontinuities in biological variation. Without these barriers, recently diverged populations that exhibit even minimal differences (at more than one gene) would not be able to retain their distinctiveness; fusion would eventually occur. This notion forms the basis for the broadly accepted and widely applied biological species concept, codified by Mayr in 1942.

Reproductive barriers represent a special class of traits in sexually reproducing species because they permit discontinuities at all other traits to persist when populations co-occur. Since I am broadly interested in understanding the cause of biological discontinuities in nature, reproductive barriers necessarily form the heart of the problem. It follows that the origin of reproductive barriers during the speciation process provides the appropriate context. Answers to many long-standing questions are within reach once genetic changes that affect reproductive isolation are identified. Thanks to recent innovations in statistical analysis of DNA data, once the genetic determinants for a given trait have been characterized, the evolutionary mechanism(s) responsible for shaping patterns of variation can be discerned. For example, an evolutionary genetic analysis of loci for reproductive barriers brings us closer to understanding the causes of speciation. These same statistical innovations also make it possible to turn from a strictly descriptive to an analytical approach for inferring population and species history (e.g., DOPMAN *et al.* 2002).

Dissertation Research Goals

My dissertation research focuses on the genetics of reproductive barriers that act before mating. Pre-mating barriers such as behavioral, temporal or ecological differences are an understudied form of isolation. Most work has concentrated on post-mating barriers that reduce hybrid fitness, although pre-mating barriers also play an important role in this capacity for species that hybridize (COYNE and ORR 1998, 2004). The proximate goal for my research on the genetic basis of speciation is to define the genetic architecture of the relevant phenotypic traits; that is, to define the number, chromosomal location, and phenotypic effects of genes involved in reproductive isolation (VIA and HAWTHORNE 1998).

Speciation involves the origin of trait differences that limit or prevent gene exchange, and ultimately results in daughter populations that form monophyletic or exclusive genetic groups. However, for recently diverged populations or species, between which reproductive isolation is often incomplete, gene genealogies will be discordant and most regions of the genome will display non-exclusive genealogical patterns. In these situations, genome regions for which one or both species are exclusive groups may be closely linked to "speciation genes," genes involved in reproductive isolation. Alternatively, such regions may reflect local selective sweeps at genes for species-specific adaptations (MAYNARD SMITH and HAIGH 1974). Therefore, another research goal is to make comparisons of gene genealogies to infer the genetic architecture of reproductive isolation and adaptation. In turn, the genetic architecture for divergent traits can be used to confirm patterns of genealogical descent between recently diverged lineages.

Genetic differences are often used to reconstruct the evolutionary history of species, to characterize recent demographic processes, and to identify molecular

evolutionary forces. Unfortunately, the evolution of species, populations, and genes all contribute to genetic differentiation, and it may often be impossible to disentangle the effects at any single scale in order to discern a correct interpretation at another. However, for those systems in which abundant natural history information is available, it may be possible to assess the impact and contribution of known demographic processes on genetic variation. In this manner, the potentially overwhelming array of explanations for the evolution of genetic variation can be limited, or at the very least, bounded, by ecologically plausible scenarios. My final dissertation research goal is to use field data and knowledge of recent natural history to investigate the contribution of reproductive isolation and other ecological processes on local and genome-wide patterns of genetic variation.

The Study System

Traits that currently act to prevent gene exchange between species may not have historically been important, and they may have evolved subsequent to complete reproductive isolation. Moreover, often many genes affect reproductive isolation and if each contributes to complete isolation independently, it can be impossible to discern the relative importance of independent barriers that led to speciation (e.g., COYNE and ORR 2004). Therefore, populations in the incipient stages of speciation, such as closely related species, subspecies, races or strains, provide the clearest window into the speciation process. It is for these reasons that I chose to study the European corn borer moth, *Ostrinia nubilalis*.

The European corn borer (ECB) is native to Europe, North Africa, and Western Asia (MUTUURA and MONROE 1970). Early in the 20th century, ECB was introduced on multiple occasions to North America in shipments of broom corn from

Italy and Hungary (CAFFREY and WORTHLEY 1927; SMITH 1920). Since its introduction, ECB has quickly spread throughout the United States and Canada east of the Rocky Mountains.

Across most of its range, ECB consists of two behaviorally isolated pheromone strains that differ in the sex pheromone produced by females and the response elicited in males (KLUN and COOPERATORS 1975). In the Z strain, females produce and males respond to a 3:97 mixture of E and Z-Δ11-14:OAc, whereas in the E strain females produce and males respond to a 99:1 (E)/(Z) blend (KLUN *et al.* 1973). Partial sexual isolation stems from differences in male response to these alternative pheromone blends (GLOVER *et al.* 1990; ROELOFS *et al.* 1987). Cross attraction rarely occurs in the laboratory (GLOVER *et al.* 1990; LINN *et al.* 1997). In nature, hybrid genotypes exist in low frequency (GLOVER *et al.* 1991; KLUN and HUETTEL 1988).

The genetic factors responsible for major differences in female pheromone blend production, *Pher*, and male behavioral response to those blends, *Resp*, exhibit simple Mendelian inheritance (DOPMAN *et al.* 2004; GLOVER *et al.* 1990; KLUN and MAINI 1979; ROELOFS *et al.* 1987). *Pher* is autosomal, whereas *Resp* is on the Z chromosome and is sex linked. Like other Lepidoptera, female ECB are the heterogametic sex.

In addition to differences in sexual communication, ECB also exhibit variation in life cycle. In New York State, populations are either bivoltine with an early first generation and a late second generation, or univoltine with a single generation in the middle of the season (ROELOFS *et al.* 1985). In southern regions of North America, ECB can exhibit a "multi-voltine" life cycle with up to four generations per season. Based on voltinism and sex pheromone blend, New York populations consist of three distinct races: univoltine Z (UZ), bivoltine E (BE), and bivoltine Z (BZ) (GLOVER *et al.* 1991; KLUN and HUETTEL 1988). Adults from univoltine populations are partially

temporally isolated from those of the bivoltine populations (ECKENRODE *et al.* 1983; ROELOFS *et al.* 1985).

Bivoltine and univoltine life cycle patterns in New York reflect genetically determined differences in post-diapause development time (PDD), the time to pupation for over-wintering larvae under temperature and photoperiod conditions conducive to breaking diapause (GLOVER *et al.* 1992). Less is known about the determinants for multi-voltinism, but a stronger environmental component seems likely. The genetic factor responsible for major differences in development time, *Pdd*, is controlled by a simple Mendelian factor that shows sex linkage (GLOVER *et al.* 1992).

Other than pheromone communication and life cycle differences, strains or races of ECB in North America are difficult to distinguish (LIEBHERR 1974; LIEBHERR and ROELOFS 1975). Multiple genetic surveys have revealed extensive shared genetic variation, but little differentiation (CIANCHI *et al.* 1980; HARRISON and VAWTER 1977; MARCON *et al.* 1999; PORNKULWAT *et al.* 1998; WILLETT and HARRISON 1999). However, significant allele-frequency differences occur at a sex-linked gene that encodes the enzyme triose phosphate isomerase (TPI). At this locus, BE populations in New York are fixed for the *Tpi-1* allele and UZ and BZ populations are segregating for both *Tpi-1* and *Tpi-2*, with *Tpi-2* being the more common allele (GLOVER *et al.* 1991).

Major Research Findings: A Preface

Genetics of reproductive barriers: One of the most important aspects of speciation in animals is the evolution of pre-mating reproductive barriers, but most studies on the genetics of speciation have focused on post-mating barriers, primarily in *Drosophila*

(COYNE and ORR 2004). In my research, I have elucidated the genetics of behavioral and temporal isolation that are leading to population divergence in E and Z pheromone strains of the European corn borer moth. I constructed a complete ECB genetic linkage map that covers the entire genome. The map consists of 213 AFLP markers, 45 microsatellite markers, and a gene that shows frequency differences between pheromone strains (*Tpi*). On this map I identified gene regions responsible for differences in pheromone blend production (*Pher*) and male behavioral response (*Resp*), components of the pheromone communication system known to be important for behavioral isolation. I also mapped a major factor responsible for differences in development time (*Pdd*) that causes temporal isolation between ECB populations.

Consequences of reproductive barriers for genealogical discordance: Contrasting genealogical relationships can facilitate inference into the genetic architecture of reproductive isolation and adaptation. I demonstrate the relevance of this approach using sexually isolated pheromone strains of ECB. Gene genealogies for five gene regions are discordant, and for only one molecular marker, the sex-linked gene *Tpi*, are the Z and E pheromone strains exclusive groups. Independent of life cycle differences, the gene genealogy for *Tpi* reveals North American pheromone strains as exclusive genetic groups, even where they co-occur. In contrast, at four other loci strains share identical DNA sequences. *Tpi* maps to a position on the sex chromosome that is indistinguishable from a major factor affecting differences in post-diapause development time (*Pdd*). The major gene determining male behavioral response to pheromone (*Resp*) is also sex-linked, but the gene maps far away on the Z chromosome. Exclusivity at *Tpi* may be a consequence of these linkage relationships, because evidence from phenotypic variation in natural populations implicates both

Pdd and *Resp* as candidates for genes involved in recent sweeps and/or reproductive isolation between strains.

Disentangling genome-wide and locus-specific effects of reproductive isolation: Here, knowledge of the demographic history for sexually isolated pheromone strains of ECB are used to investigate the contribution of reproductive isolation on local and genome-wide patterns of genetic variation. Patterns of genetic variation for a multi-locus data set are consistent with the action of both genome-wide stochastic effects and locus-specific deterministic effects. Specifically, neutral demographic processes involving population substructure and population expansion can account for much of the observed variation among genetic markers, but one locus appears to have been influenced by non-neutral evolution. Variation at *Tpi* is consistent with recent selection and reveals pheromone strains of ECB as exclusive genetic groups, but only in North America. Patterns of genealogical descent and genetic variation at *Tpi*, and the indistinguishable mapping positions of *Tpi* and *Pdd*, suggest that ecological adaptation is responsible for the origin and possibly the maintenance of this locus-specific pattern. As differences in development time incidentally generate temporal isolation, this reproductive barrier represents a byproduct of adaptation to different environments.

Broader Impacts

The original theory for descent and diversification as proposed by Darwin and the contributors to the Modern Synthesis fits the Z and E pheromone strains of ECB well. Gradual evolution within ECB pheromone strains at *Pdd*, *Pher*, and *Resp* has generated behavioral and temporal reproductive isolation between populations that is

responsible for the origin and/or persistence of discontinuities observed in nature at *Tpi*, even in sympatry. By utilizing computational and molecular genetic tools, my work advances earlier speciation theory by making an explicit link between the origin of genetic discontinuities, an attribute that ultimately characterizes species as unique entities, and the evolution of reproductive isolation between populations. A subtle, but far-reaching implication from my work is that for populations in the incipient stages of speciation, such as closely related species, subspecies, races or strains, discontinuities may only exist at or near genes for reproductive isolation – the only locus that reveals Z and E pheromone strains of ECB to be exclusive (*Tpi*) maps to a genomic position that is indistinguishable from a divergent trait (*Pdd*).

REFERENCES

CAFFREY, D. J., and L. H. WORTHLEY, 1927 A progress report on the investigations of the European corn borer. U. S. Department of Agriculture Bulletin 1476: 155pp.

CIANCHI, R., S. MAINI and L. BULLINI, 1980 Genetic distance between pheromone strains of the European corn borer *Ostrinia nubilalis* different contribution of variable substrate regulatory and nonregulatory enzymes. Heredity 45: 383-388.

COYNE, J. A., and H. A. ORR, 1998 The evolutionary genetics of speciation. Phil. Trans. R. Soc. B. Biol. Sci. 353: 287-305.

COYNE, J. A., and H. A. ORR, 2004 Speciation. Sinauer Associates, Sunderland, Massachusetts.

DARWIN, C., 1859 The origin of species. John Murray, London.

DOBZHANSKY, T., 1937 Genetics and the Origin of Species. Columbia Univ. Press, New York, NY.

DOPMAN, E. B., S. M. BOGDANOWICZ and R. G. HARRISON, 2004 Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). Genetics 167: 301-309.

DOPMAN, E. B., G. A. SWORD and D. M. HILLIS, 2002 The importance of the ontogenetic niche in resource-associated divergence: Evidence from a generalist grasshopper. Evolution 56: 731-740.

ECKENRODE, C. J., P. S. ROBBINS and J. T. ANDALORO, 1983 Variations in flight patterns of European corn borer (Lepidoptera: Pyralidae) in New York. Environ. Entomol. 12: 393-396.

GLOVER, T., M. CAMPBELL, P. ROBBINS and W. ROELOFS, 1990 Sex-linked control of sex pheromone behavioral responses in European corn borer moths (*Ostrinia nubilalis*) confirmed with TPI marker gene. Arch. Insect Biochem. Phys. 15: 67-77.

GLOVER, T. J., J. J. KNODEL, P. S. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1991 Gene flow among three races of European corn borers (Lepidoptera:Pyralidae) in New York State. Environ Entomol 20: 1356-1362.

GLOVER, T. J., P. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1992 Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. Arch. Insect Biochem. Phys. 20: 107-117.

HARRISON, R. G., and A. T. VAWTER, 1977 Allozyme differentiation between pheromone strains of the European corn borer, *Ostrinia nubilalis*. Ann. Entomol. Soc. Am. 70: 717-720.

KLUN, J. A., O. L. CHAPMAN, K. C. MATTES, WOJTKOWS.PW, M. BEROZA *et al.*, 1973 Insect sex pheromones - minor amount of opposite geometrical isomer critical to attraction. Science 181: 661-663.

KLUN, J. A., and COOPERATORS, 1975 Insect sex pheromones: Intraspecific pheromonal variability of *Ostrinia nubilalis* in North America and Europe. Environmental Entomology 4: 894-894.

KLUN, J. A., and M. D. HUETTEL, 1988 Genetic regulation of sex pheromone production and response: Interaction of sympatric pheromonal types of European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). J. Chem. Ecol. 14: 2047-2061.

KLUN, J. A., and S. MAINI, 1979 Genetic basis of an insect chemical communication system: The European corn borer. Environ. Entomol. 8: 423-426.

LIEBHERR, J., 1974 Studies on two strains of European corn borer, *Ostrinia nubilalis* (Hübner), pp. 1-70 in Masters Thesis. thesis, Cornell University, Ithaca, NY.

LIEBHERR, J., and W. L. ROELOFS, 1975 Laboratory hybridization and mating period studies using two pheromone strains of *Ostrinia nubilalis*. Ann. Entomol. Soc. Am. 68: 305-309.

LINN, C., E. JR., M. S. YOUNG, M. GENDLE, T. J. GLOVER and W. L. ROELOFS, 1997 Sex pheromone blend discrimination in two races and hybrids of the European corn borer moth, *Ostrinia nubilalis*. Physiol. Entomol. 22: 212-223.

MARCON, P., D. B. TAYLOR, C. E. MASON, R. L. HELLMICH and B. D. SIEGFRIED, 1999 Genetic similarity among pheromone and voltinism races of *Ostrinia nubilalis* (Hubner) (Lepidoptera: Crambidae). Insect Mol. Biol. 8: 213-221.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

MAYR, E., 1942 Systematics and the Origin of Species. Columbia University Press, New York.

MUTUURA, A., and E. MONROE, 1970 Taxonomy and distribution of the European corn borer and allied species: Genus *Ostrinia* (Lepidoptera: Pyralidae). Mem. Entomol. Soc. Canada 71: 1-112.

PORNKULWAT, S., S. R. SKODA, G. D. THOMAS and J. E. FOSTER, 1998 Random amplified polymorphic DNA used to identify genetic variation in ecotypes of the European corn borer (Lepidoptera: Pyralidae). Ann. Entomol. Soc. Am 91: 719-725.

ROELOFS, W., T. GLOVER, X. H. TANG, I. SRENG, P. ROBBINS *et al.*, 1987 Sex-pheromone production and perception in European Corn Borer moths is determined by both autosomal and sex-linked genes. Proc. Natl. Acad. Sci. U.S.A. 84: 7585-7589.

ROELOFS, W. L., J. W. DU, X. H. TANG, P. ROBBINS and C. ECKENRODE, 1985 Three European corn borer populations in New York based on sex pheromones and voltinism. J. Chem. Ecol. 11: 829-836.

SMITH, H. E., 1920 Broom corn, the probable host in which *Pyrausta nubilalis* Hubn. reached America. J. Econ. Entomol. 13: 425-430.

VIA, S., and D. J. HAWTHORNE, 1998 The genetics of speciation: promises and prospects of quantitative trait locus mapping, pp. 352-366 in Endless Forms, edited by D. J. HOWARD and S. H. BERLOCHER. Oxford Univ. Press, New York, NY.

WILLETT, C. S., and R. G. HARRISON, 1999 Insights into genome differentiation: Pheromone-binding protein variation and population history in the European corn borer (*Ostrinia nubilalis*). Genetics 153: 1743-1751.

CHAPTER 2

GENETIC MAPPING OF SEXUAL ISOLATION IN THE EUROPEAN CORN BORER *

ABSTRACT

The E and Z pheromone strains of the European corn borer (ECB) provide an exceptional model system for examining the genetic basis of sexual isolation. Differences at two major genes account for variation in female pheromone production and male behavioral response, components of the pheromone communication system known to be important for mate recognition and mate choice. Strains of ECB are morphologically indistinguishable, and surveys of allozyme and DNA sequence variation have revealed significant allele frequency differences only at a single sex-linked locus, *Tpi*. Here we present a detailed genetic linkage map of ECB using AFLP and microsatellite markers and map the factors responsible for pheromone production (*Pher*) and male response (*Resp*). Our map covers 1697 cM and identifies all 31 linkage groups in ECB. Both *Resp* and *Tpi* map to the Z (sex) chromosome, but the distance between these markers (> 20 cM) argues against the hypothesis that patterns of variation at *Tpi* are explained by tight linkage to this "speciation gene." However, we show, through analysis of marker density, that *Tpi* is located in a region of low recombination and suggest that a second Z-linked reproductive barrier could be responsible for the origin and/or persistence of differentiation at *Tpi*.

* DOPMAN, E.B., S. M. BOGDANOWICZ AND R. G. HARRISON, 2004 Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*) Genetics 167: 301-309.

INTRODUCTION

Although evolutionary biologists and systematists often do not agree on the relative merits of competing species concepts, most students of the speciation *process* adopt the biological species concept (BSC), in which the origin of new species involves the evolution of intrinsic barriers to gene exchange between diverging populations (COYNE and ORR 1998; HARRISON 1998). As a consequence, research on the genetic basis for speciation has focused on identification and characterization of the genetic elements that contribute to reproductive barriers. The proximate goal is to define the genetic architecture of the relevant phenotypic traits, that is to define the number, chromosomal location, and phenotypic effects of genes involved in reproductive isolation (VIA and HAWTHORNE 1998; ORR 2001).

Until very recently, this research agenda was pursued successfully only in a few model organisms, and indeed most of our knowledge about the genetics of speciation came from detailed studies of hybrid sterility and inviability in *Drosophila* (COYNE and ORR 1998). A recent review of the evolutionary genetics of speciation (COYNE and ORR 1998) summarized case studies that provided "rigorous" genetic analysis of traits known to cause reproductive isolation. Approximately 75% of the studies used *Drosophila* as a model system, and 60-70% of those studies focused on post-zygotic isolation. Although the genetics of postmating isolation in *Drosophila* continues to be a major focus for speciation research (*e.g.*, PRESGRAVES *et al.* 2003), premating barriers (*i.e.*, those that derive from behavioral, temporal, or ecological differences) may be the most important components of reproductive isolation in many taxa (COYNE and ORR 1989). Genetic analysis of premating barriers brings added complexity because they are often composed of many context-dependent traits for which individual effects on reproductive isolation are difficult to quantify (BUTLIN and

RITCHIE 2001). As a result, studies of the genetics of premating barriers have lagged behind those of postmating barriers. However, the availability of molecular tools facilitates linkage mapping studies in animals and plants of particular interest to speciation researchers and allows identification of genetic factors for ecological, behavioral, or morphological differences associated with (perhaps responsible for) reproductive isolation between closely related species or races (BRADSHAW *et al*. 1995; RIESEBERG *et al*. 1999; HAWTHORNE and VIA 2001; PEICHEL *et al*. 2001; PARSONS and SHAW 2002).

One of the non-*Drosophila* model systems mentioned by COYNE and ORR (1998) is the European corn borer (ECB) moth, *Ostrinia nubilalis* (Crambidae). Like many moths, ECB uses a chemical communication system for long-distance mate attraction. Stationary female moths emit a particular pheromone blend that is used by males with the "appropriate" behavioral response for orientation and navigation toward the signaling female. Although in most moths this pheromone system provides species specificity and is invariant within species (LÖFSTEDT 1993), in ECB two pheromone strains exist: one in which females produce and males respond to a 3:97 ratio of E/Z-11-tetradecenyl acetate (Δ11-14:OAc) (Z strain), and another in which females produce and males respond to a 99:1 E/Z blend (E strain) (KLUN *et al*. 1973; KOCHANSKY *et al*. 1975). Both Z and E borers are found in Europe and were introduced into eastern North America early in the 20th century, presumably on shipments of broom corn (CAFFREY and WORTHLEY 1927).

Laboratory analyses of E and Z borers suggest that sexual isolation stems from stereotypic differences in male response to pheromone blend composition (KLUN and COOPERATORS 1975; LIEBHERR and ROELOFS 1975; CARDÉ *et al*. 1978; GLOVER *et al*. 1991; LINN *et al*. 1997; E.B. Dopman, unpublished results). Z-strain males fly to the 3:97 E/Z pheromone, but not to any other blends (ROELOFS et al. 1987; GLOVER et al.

1990). In contrast, although most E males respond to the 99:1 E/Z pheromone produced by E females, a substantial proportion fly to intermediate blends and an occasional male is attracted to the Z blend. Thus, in spite of the use of alternative sexual communication systems, cross attraction between moths occurs at a low frequency. Field and laboratory data indicate that hybridization occurs when E males court and mate with Z females (LIEBHERR and ROELOFS 1975; GLOVER *et al.* 1991). Of 150 females that were analyzed for pheromone production at localities where Z and E populations were sympatric, thirteen (8.7%) produced the characteristic 65:35 E/Z hybrid blend (GLOVER *et al.* 1991). When hybrid males are produced they respond over a broad range of pheromone blends, but rarely to the E blend (ROELOFS *et al.* 1987; GLOVER *et al.* 1990). Hybrid males and females exhibit no obvious reduction in fitness in the laboratory (LIEBHERR and ROELOFS 1975).

Segregation patterns assessed from F1, F2, and backcross progeny produced using Z and E strain parents indicate that pheromone production and male behavioral response are each determined by single major genes. Pheromone production exhibits autosomal inheritance, whereas male behavioral response is sex-linked (KLUN and MAINI 1979; ROELOFS *et al.* 1987; GLOVER *et al.* 1990). In Lepidoptera, females are the heterogametic sex and are ZW, whereas males are ZZ.

Although differentiated with respect to pheromone communication, the Z and E strains of ECB are otherwise difficult to distinguish (LIEBHERR 1974), and the taxonomic status of these strains has remained uncertain (CARDÉ *et al.* 1978). Surveys of allozymes (HARRISON and VAWTER 1977; CIANCHI *et al.* 1980), mitochondrial DNA (MARCON *et al.* 1999), RAPDs (PORNKULWAT *et al.* 1998), and a nuclear gene (WILLETT and HARRISON 1999) have revealed significant allele frequency differences at only one locus, the gene encoding the enzyme triose phosphate isomerase (TPI). In upstate New York, E-strain populations are fixed for the *Tpi-1* allele, whereas Z-strain

populations are segregating for both *Tpi-1* and *Tpi-2* at intermediate frequencies (GLOVER *et al*. 1991). Even where they occur together in the same fields, E and Z borers remain differentiated for *Tpi*, suggesting that gene flow is limited or absent for this locus. Because *Tpi* and male pheromone response are both Z-linked (GLOVER *et al*. 1990), the two loci may map close to each other on the Z chromosome. Close physical linkage and/or low recombination between loci could explain the apparent reduction of introgression in spite of ongoing hybridization. Furthermore, a recent selective sweep at the locus for male response could have initially driven differentiation at the *Tpi* locus via genetic hitchhiking.

Here we describe the first genetic linkage map for ECB based on a combination of dominant AFLP and co-dominant microsatellite molecular markers. Because crossing-over during oogenesis does not occur in female Lepidoptera, families can be generated in which all markers on the same chromosome co-segregate as a single unit. Conversely, crossing-over during spermatogenesis in male Lepidoptera allows families to be generated in which recombinational distance affects the co-segregation of markers on the same chromosome. This "biphasic" nature of crossing-over facilitates mapping of major loci underlying phenotypes by using a sequential analysis of female and male-informative crosses (HECKEL *et al*. 1999). We used a female-informative backcross to establish linkage groups for segregating markers and traits (BC1F) and a male-informative backcross to order markers and traits within linkage groups (BC1M). Within the context of this linkage map, we locate the *Tpi* gene and the factors responsible for pheromone blend production and male behavioral response. In addition, marker densities across the Z chromosome are used to estimate changes in local recombination rates.

MATERIALS AND METHODS

Insect populations: Two cultures of ECB, maintained by Wendell Roelofs and colleagues at the New York State Agricultural Experiment Station (NYSAES) in Geneva, NY, were used as sources for initiating mapping families. The first culture consisted of Z-strain insects that were isolated from field-collected larvae, pupae, and adults in corn stubble from Bouckville, New York in April 1994. The second culture consisted of E-strain insects isolated from corn stubble near Geneva, New York in May 1996. We set up 12 F1 crosses between E females and Z males and chose one cross that was segregating at the *Tpi* locus to begin the backcross generation. Backcross families were started using an F1 female crossed to an E male for BC1F and an F1 male crossed to an E female for BC1M (Figure 2.1). All insects were reared under a 16:8 L:D photoperiod at 25-30° and 50-60% relative humidity.

Marker development: Genomic DNA was extracted from adult moths using a DNeasy Tissue Kit (QIAGEN, Valencia, CA). AFLP markers (VOS *et al.* 1995) were developed with the restriction enzymes *Eco*RI and *Mse*I (NEB, Beverly, MA) using the AFLP plant-mapping kit for small genomes (50-500 Mb) according to the manufacturers' protocol (Applied Biosystems, Foster City, CA). We developed markers for ECB using 12 primer pairs for BC1F, nine of which were used for BC1M plus one additional primer pair not used in BC1F. For the selective amplification, the *Eco*R1 primer was fluorescently labeled, allowing detection of fragments on an ABI 377 automated sequencer (Applied Biosystems, Foster City, CA). Primers generated between 50-100 AFLP bands per primer pair, and detected fragments ranged in size from ≈ 50 to ≈ 450 base pairs. AFLP fragments were sized in GENESCAN (v 3.1, ABI) and segregating markers were visualized with GENOTYPER (v 2.1, ABI).

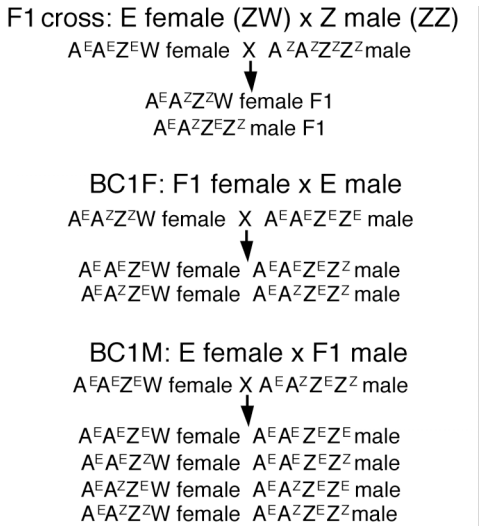


Figure 2.1 "Biphasic" mating design (HECKEL *et al.* 1999) that establishes linkage group relationships and orders markers within linkage groups using female-informative (BC1F) and male-informative (BC1M) backcrosses, respectively. E and Z superscripts indicate origin of autosomes (A) and sex chromosomes (Z, W) from grandparents in F1 cross.

Sequences containing microsatellites were isolated as described by HAMILTON *et al*. (1999), with some modifications. Genomic DNA digestion and linker ligation occurred simultaneously, and these fragments were enriched for microsatellites by hybridization to biotinylated dimeric and tetrameric oligonucleotides. After magnetic capture, double-stranded DNA was recovered by PCR and ligated to pUC 19. Aliquots of this ligation were used to transform *E. coli* (DH5-α) cells. Colonies were plated on Luria agar supplemented with ampicillin, replicated onto nylon membranes, and probed with the same oligonucleotides (now radiolabeled) that were used in the enrichment. Sequences from positive colonies were obtained with universal M13 primers that flank the cloning site. Primers were designed using PRIMERSELECT (DNASTAR Inc., Madison, WI), and unlabeled and 5' fluorescently labeled primers (HEX, 6-FAM) were ordered from Integrated DNA Technologies (Coralville, IA).

PCR reactions for genotyping individuals contained 1 μl (approximately 10 ng) genomic DNA in a 10 μl total volume that included 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2 mM $MgCl_2$, 0.2 mM each dNTP, 1.5 pmole each primer, and 0.5 units Platinum Taq DNA polymerase (Invitrogen, Carlsbad, CA). Thermal cycling consisted of a "touchdown" procedure that began at 95° for 50 sec, 67° for 1 min, 72° for 1 min. The annealing temperature was dropped 2° per cycle for the first seven cycles (range of temperatures 67° to 55°), followed by 28 cycles with an annealing temperature of 52°. Up to six loci were multiplexed per lane on an ABI 377 automated sequencer according to the manufacturers' protocols. Fragments were analyzed identically to AFLPs.

Genotyping moths for *Tpi*: Moths were genotyped for *Tpi* either by characterizing protein mobility on cellulose acetate plates or by use of a diagnostic restriction site difference. For protein electrophoresis, three legs from adult ECB in BC1M were removed prior to genomic DNA extraction, and homogenized leg tissue

was applied to cellulose acetate plates for allozyme analysis. Samples were electrophoresed for 13 min at 200 volts in Tris/Glycine buffer. Agar-overlay stain components (Sigma, St. Louis, MO) for single 76 x 76 mm acetate plates included 3.5 mg dihydroxyacetone phosphate as a substrate together with 2 ml NAD (2 mg/ml), 10 mg Na_2HAsO_4, 130 μl MTT (10 mg/ml), 130 μl PMS (2 mg/ml), 20 μl G_3PDH, and 3 ml agar (16 mg/ml). From *Tpi* DNA sequence data (E.B. Dopman, L. Perez, S.M. Bogdanowicz, R.G. Harrison, unpublished results) we identified a diagnostic *Bsr*I restriction site between the *Tpi-1* and *Tpi-2* alleles. Following amplification of genomic DNA using primers that flank the recognition site, the PCR product was digested with *Bsr*I (NEB, Beverly, MA) following the manufacturers' protocol.

Female pheromone blend and male behavioral response assays: Adult 1-2 day old females in the BC1F and BC1M mapping populations were prepared for the pheromone blend assay by being placed in 10 ml plastic cups covered with aluminum foil. Following an 18-24 hour period in the dark, abdominal tips were exposed and clipped into 1 drop of HLPC-grade hexane (Sigma). Tips were removed after 1 hour and solutions were stored at –80°. Pheromone samples were concentrated to 1 μl under an N_2 stream before being injected into a HP 5890 GC (for BC1F) or a Shimadzu GC-17A (for BC1M) gas chromatograph, both of which were equipped with 30 m polar EC-Wax columns (Alltech, Deerfield, IL). The running program for the HP 5890 began at 80° for 2 min followed by a ramping rate of 15°/min to 190°, 4°/min to 220°, and ending at 200° for 4 min, whereas for the Shimadzu, running conditions were 40° for 2 min, increased at 15°/min to 250° and held for 10 min. Under these conditions, females that are E/Z at the production locus will exhibit a 65:35 ratio of E/Z Δ11-14:OAc and females that are E/E at this locus will exhibit a 99:1 ratio of E/Z Δ11-14:OAc (ROELOFS *et al*. 1987).

Male behavioral response for BC1M males was characterized by flying 1-2 day old males to synthetic pheromone blends in a sustained-flight tunnel according to the conditions and protocol of LINN *et al*. (1997). Flight tunnel conditions were 19°, 50-60% relative humidity, 0.5 m/sec air speed, and 11 lux incandescent red-light intensity. Males were taken to the flight tunnel room 1 hour before lights off and placed individually in 12 x 5 cm screen cages for acclimatization. Moths were first assayed with a 99:1 E/Z blend followed by 3:97 blend of E/Z. Pheromones (30 μg) were impregnated on rubber septa placed upwind of males and a behavioral sequence was recorded that included activation, taking flight, orientation in the odor plume, upwind flight at least 10 cm from release point, upwind flight to the mid-point in the tunnel, upwind flight at least 10 cm from the pheromone source, and contact with the pheromone source. Positive responses were those for which the male exhibited upwind flight to within 10 cm of the pheromone source. Under these conditions, all responding control males with the E/E genotype for male response flew to the 99:1 E/Z source, but not to 3:97 E/Z. Conversely, all responding control hybrid males with the E/Z genotype for male response flew only to the 3:97 E/Z blend, or to both blends. Thus, test males flying only to the 99:1 E/Z source were E/E at the response locus and test males flying to both or to only the 3:97 E/Z source were E/Z at the response locus.

BC1F - Linkage group construction and test of effect on pheromone production: In BC1F, informative markers were those present in the F1 mother and absent in the recurrent E male parent (Figure 2.1). Using the labeling scheme developed for use in MAPMAKER/EXP 3.0 (LANDER *et al*. 1987), the presence of alleles inherited from the paternal Z grandparent was scored as “H” and absence as “A”, whereas the presence and absence of alleles were scored as “A” and “H,” respectively, for alleles inherited from the maternal E grandparent. Since pheromone production and male response exhibit simple Mendelian inheritance, traits were

similarly coded in BC1F and BC1M. Markers were sorted into linkage groups with a LOD score of 3.0 and a maximum recombinational distance of 40 cM. Because of the absence of crossing-over in the F1 mother, 31 groups of co-segregating markers were expected, each group segregating in a 1:1 ratio. These groups represent different chromosomes inherited from the Z and E grandparents through the F1 mother, and the contribution of each chromosome to pheromone production was tested using the interaction χ^2 statistic at $\alpha = 0.05/31$ where the counts of BC1F females' pheromone blend (99:1 E/Z or 65:35 E/Z) were compared to chromosomal origin (Z or E grandparent).

BC1M - Mapping markers and traits within linkage groups: In BC1M, informative markers were those present in the F1 father and absent in the recurrent E female parent (Figure 2.1). Segregation ratios (1:1) were tested using a χ^2 test and homologous markers from BC1F were noted to identify groups of markers originating from the same chromosome. Linkage was established under an initial score of LOD ≥ 3.0 with a maximum distance of 40 cM. For linkage groups possessing less than 9 markers, the "compare" command was used to order markers and traits within linkage groups and distances were estimated using the Kosambi mapping function. The "suggest subset" command was used to identify equally spaced markers for linkage groups with 10 or more markers and the "build" or "try" commands were used to place additional markers at a threshold of 2.0. Finally, marker order was tested using the "ripple" command. Markers that were not automatically placed because of a low number of recombinants were added to the map under their most likely positions as long as their addition did not result in substantial map extension.

Estimation of recombination rate: For the majority of species in which a physical map is unknown, information about genome-wide recombination patterns can be obtained indirectly. LYON (1976) suggested that genomic variation in chiasmata

would create a non-random distribution of marker densities on a genetic map, provided that markers exhibit a random physical distribution. Areas of high recombination would exhibit low marker clustering, whereas in areas of low recombination markers would cluster. Thus, the inverse of marker density on the genetic map provides a proportional estimate of local recombination rate. Using a microsatellite-based genetic map, LYON'S indirect approach correctly identified regions of low recombination in *Drosophila melanogaster* (NACHMAN and CHURCHILL 1996), and this method has been applied to *Mus musculus* and sea beet (*Beta vulgaris* subsp. *maritima*) using microsatellites and RFLP markers, respectively (NACHMAN and CHURCHILL 1996; KRAFT *et al*. 1998).

We compared the number of markers per linkage group with those expected under a Poisson distribution to test for a random physical distribution of AFLP and microsatellite markers among ECB chromosomes. Within each linkage group, the number of markers in a 10 cM bandwidth on the genetic map was compared to that expected under a Poisson distribution to detect significant marker density variation. Density estimates could be inflated due to co-dominant AFLP markers mapping to the same position; however, co-dominance would only be possible if markers exhibited complimentary segregation and if markers mapping to the same position were derived from the same primers. We checked for these allelic effects and markers that exhibited this pattern were removed from the analysis. Local recombination rates were estimated for the Z chromosome by taking the inverse of the cosine-kernel density function with an 8 cM bandwidth using the R environment (v 1.7-18, ISBN 3-901167-51-X). A reflecting boundary was employed to minimize biased estimates near edges (see NACHMAN and CHURCHILL 1996), and rates were scaled to those expected with equal sized chromosomes and a genome size of 500 Mb (the size estimated for the pyralid moth *Galleria melonella*, a close relative of ECB) (GREGORY 2001).

RESULTS

BC1F - Linkage group construction and test of effect on pheromone production: Fifty-seven female BC1F progeny produced sufficient pheromone for gas chromatographic analysis. Of these, 12 females producing 99:1 E/Z and 13 females producing 65:35 E/Z were chosen as material for genetic analysis. Six BC1F males were also analyzed to identify markers linked to the sex chromosomes. A total of 216 AFLPs and 32 microsatellites segregated in BC1F, and as expected, the majority of markers originated from the paternal Z grandparent because the recurrent parent was from the E strain (Figure 2.1). Linkage analysis produced groups of co-segregating markers indicating a lack of recombination in female ECB. The number of markers for the 30 groups identified ranged in size from 2 to 26, and the single unlinked AFLP marker was included as group 31 because n = 31 for ECB (GUTHRIE *et al*. 1965). Only linkage group 12 showed a significant interaction between pheromone blend and chromosomal origin after correcting for multiple comparisons ($\alpha = 0.05/31$) (Figure 2.2). Females producing the 99:1 E/Z pheromone derived their chromosome 12 from the E grandparent and females producing the 65:35 E/Z pheromone derived their chromosome 12 from the Z grandparent.

BC1M - Mapping markers and traits within linkage groups: In the BC1M mapping family, 81 males showed unambiguous flight behaviors and 38 females produced measurable pheromone. Seventy-nine BC1M progeny were genotyped for AFLP markers (41 males and 38 females), and 38 progeny were genotyped for microsatellites. All males and females were genotyped for *Tpi*. A total of 213 AFLP and 45 microsatellite markers segregated in BC1M. By noting which markers in BC1M showed homology with those used in BC1F, apparently independent linkage groups could be located on the same chromosome, even when they did not sort

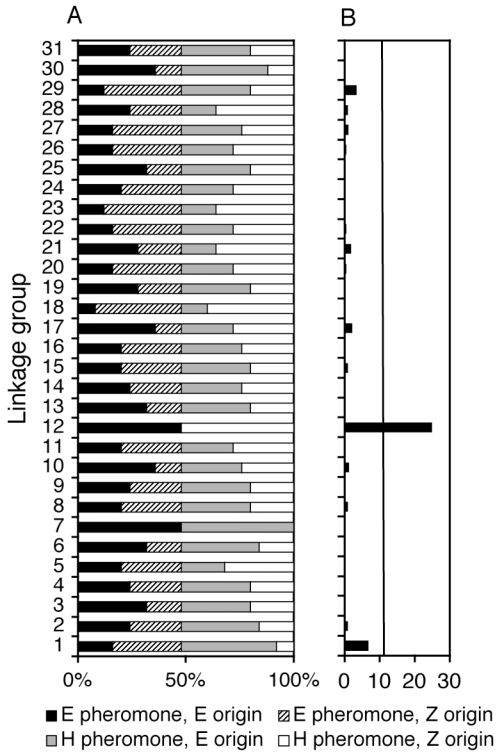


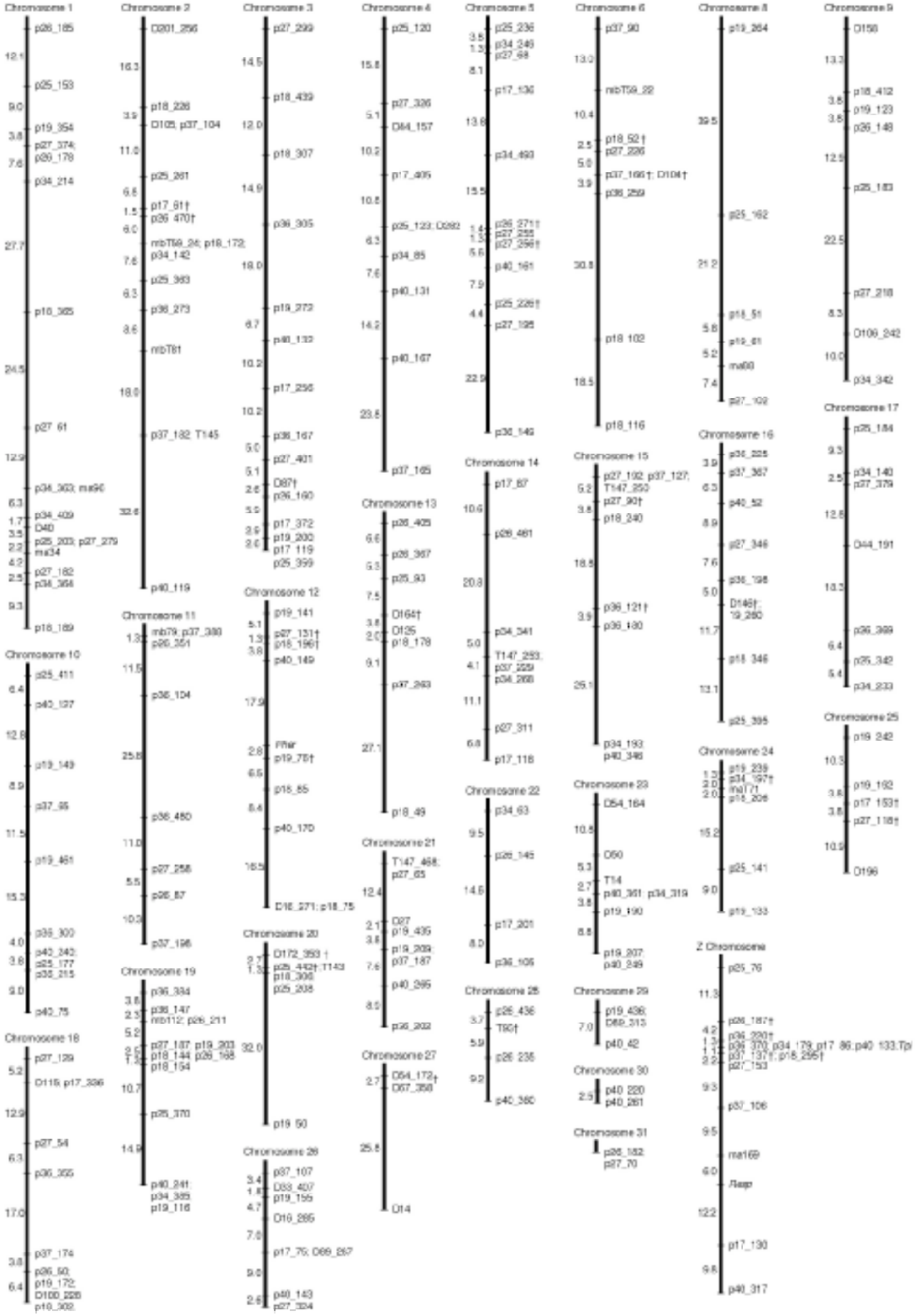
Figure 2.2 Genomic scan for the factor responsible for pheromone blend production in female ECB. (A) For each linkage group, the proportion of females in BC1F with E (E) or hybrid (H) pheromone blend and chromosome origin. (B) Chi-square tests for the interaction of linkage group origin and pheromone blend. The vertical line indicates the corrected threshold for significance ($\alpha = 0.05/31$).

together under the LOD threshold of 3.0. These fragmented groups coalesced under a higher LOD score of 4.0, producing a total of 31 linkage groups with 2-18 markers per group. The combined AFLP and microsatellite map covered 1697.3 cM with linkage groups ranging from 2.5 cM (group 30) to 127.3 cM (group 1) in length. The average interval distance between mapped markers was 8.8 cM.

The locus for pheromone blend production (*Pher*) mapped to linkage group 12, 9.3 ± 5.5 cM from AFLP p18_85 and 17.9 ± 6.0 cM from AFLP p40_149 (Figure 2.3). Six additional AFLPs and one microsatellite were placed on linkage group 12, producing a 62.3 cM map. The gene for male response (*Resp*) mapped to the Z chromosome (linkage group 7), 12.2. ± 4.3 cM away from AFLP p17_130 and 6.0 ± 2.5 cM away from microsatellite ma169. Interestingly, *Tpi* mapped 28.1 ± 4.1 cM from *Resp*. The Z chromosome map consisted of 12 additional AFLP markers for a total map length of 66.9 cM. A cluster of markers on the Z chromosome grouped in the region surrounding *Tpi*, including four AFLP markers mapping to the same position as *Tpi* and two AFLPs mapping together at a second site.

Estimation of recombination rate: The distributions of both microsatellite ($\chi^2 = 1.51$, d.f. = 3, $P = 0.59$) and AFLP ($\chi^2 = 2.58$, d.f. = 3, $P = 0.83$) markers among linkage groups were not significantly different from random expectations. In contrast, after correcting for multiple comparisons ($\alpha = 0.05/28$) six linkage groups showed non-random marker distributions, including the Z chromosome. Linkage groups 3 ($\chi^2 = 20.72$, d.f. = 4, $P = 0.0004$), 10 ($\chi^2 = 27.89$, d.f. = 4, $P = 7\text{x}10^{-6}$), and 14 ($\chi^2 = 20.9$, d.f. = 4, $P = 0.0009$) showed a more even distribution of markers than expected, whereas linkage groups 19 ($\chi^2 = 25.1$, d.f. = 5, $P = 2\text{x}10^{-5}$), 20 ($\chi^2 = 12.92$, d.f. = 2, $P = 0.0016$), and the sex chromosome ($\chi^2 = 37.45$, d.f. = 5, $P = 6.9\text{x}10^{-7}$), showed a distribution of markers with a significant excess of marker clustering. Other linkage groups may exhibit a random distribution of markers, but a lack of power due to low

Figure 2.3 Genetic linkage map for ECB showing position of factors responsible for male behavioral response (*Resp*), female pheromone blend (*Pher*), and *Tpi*. AFLP markers are labeled pXX_XX indicating primer combination and allele size, respectively. All other markers are microsatellites. † denotes positions ordered at LOD < 2.0 (see Methods); all other positions ordered at LOD ≥ 2.0.


Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 8
Chromosome 9
Chromosome 10
Chromosome 11
Chromosome 12
Chromosome 13
Chromosome 14
Chromosome 15
Chromosome 16
Chromosome 17
Chromosome 18
Chromosome 19
Chromosome 20
Chromosome 21
Chromosome 22
Chromosome 23
Chromosome 24
Chromosome 25
Chromosome 26
Chromosome 27
Chromosome 28
Chromosome 29
Chromosome 30
Chromosome 31
Z Chromosome

marker density is likely for some linkage groups. Using the indirect method of LYON (1976), the average recombination rate for the Z chromosome was estimated at 4.2 cM/Mb (2.1 cM/Mb sex-averaged rate) and based on the kernel density estimator, recombination rates were lower near the middle of the chromosome with rates ranging from 2.3 to 5.9 cM/Mb (Figure 2.4). *Tpi* mapped to the area of reduced recombination, whereas male response mapped to the opposite end of the Z chromosome in an area of higher recombination.

DISCUSSION

One fundamental goal for the study of speciation is to characterize the genetic basis for trait differences that contribute to reproductive isolation among natural populations. A necessary prerequisite is knowledge of the underlying genetic architecture, but unfortunately, for all but a few species this information is lacking (see PARSONS and SHAW 2002). However, although the availability of a dense genetic map can assist in gene isolation, many reproductive barriers have a complex, polygenic basis (COYNE and ORR 1998; RITCHIE and PHILLIPS 1998). Furthermore, unless allele frequencies are known in natural populations, it is unclear whether the phenotypic variance components explained by multiple QTL factors in a mapping population reflect an important *biological* contribution to reproductive isolation in natural populations. Because traits that play a central role in reproductive isolation in ECB are determined by single major genetic factors, phenotypic frequencies necessarily reflect genotypic frequencies in natural populations, making this insect an excellent model for studying both the genetics and functional significance of divergence in insect communication. Here we have used crosses between the two naturally occurring pheromone strains to generate the first linkage map for ECB and to locate the genetic

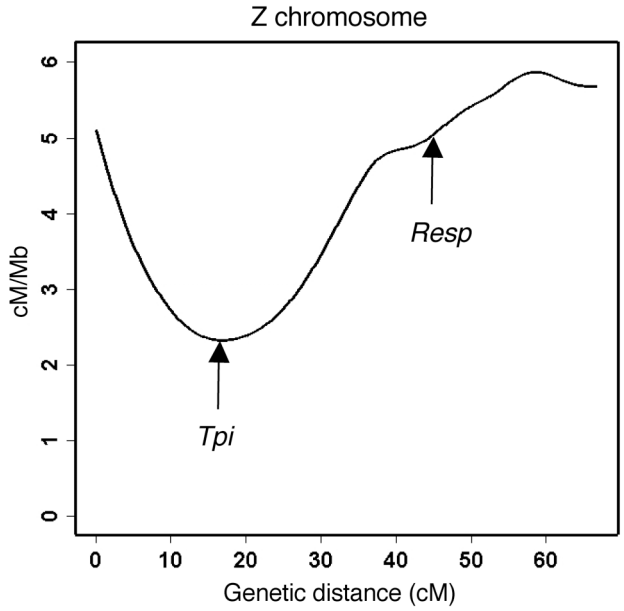


Figure 2.4 Local recombination rates on the ECB Z chromosome estimated by taking the inverse of the cosine-kernel density function of marker density in a 8 cM window using a reflecting boundary. The map positions of *Tpi* and *Resp* are indicated.

factors that account for variation in female pheromone production and male behavioral response, components of the communication system which are clearly involved in mate recognition and mate choice.

Our linkage map for ECB covers 1697.3 cM across 31 linkage groups with an average interval distance of 8.8 cM (Figure 2.3). Genome sizes in moths are generally in the 400-1000 Mb range (GREGORY 2001), but the closest relative to ECB for which genome size has been determined has a genome size of 500 Mb. Assuming a similar genome size for ECB, the relationship between physical and recombinational map distances would be ≈ 290 kb/cM, at the lower end of the range seen in other insects (HUNT and PAGE 1995; GREGORY 2001; PARSONS and SHAW 2002). This figure should decrease, however, as additional markers will extend the map length on those linkage groups with few markers. That we identify the same number of linkage groups as chromosome can be attributed in part to our sequential female and male-informative backcross mapping design (Figure 2.1). This strategy allowed us to recognize homologies between markers assorting as non-recombining chromosomes in BC1F and segregating markers on recombining chromosomes in BC1M. By establishing homology, we were able to accurately sort markers in BC1M, in which high levels of recombination combined with the large number of chromosomes in ECB might have led us to identify many more linkage groups than there are chromosomes. Given the low to moderate marker density common for non-model organisms, this can be a problematic even when chromosome numbers are not large (PARSONS and SHAW 2002). Thus, our ability to identify 31 linkage groups in an insect with a large haploid number underscores the utility of the biphasic approach (HECKEL *et al*. 1999) for accurate linkage map construction and argues for use of this approach in Lepidoptera and other taxa with differences in crossing-over between the sexes.

From the female-informative backcross, we detected one autosomal linkage group (chromosome 12) that contributes significantly to variation in E/Z pheromone blend between the E and Z strains of ECB (Figure 2.2). Through the male-informative backcross, we were able to map the factor responsible for this variation and to locate AFLP markers within 10 cM (Figure 2.3). Previous research has shown that the Z (sex) chromosome has a significant effect on variation in male behavioral response between E and Z strains (GLOVER *et al.* 1990), and as expected, we placed the factor that accounts for differences in male response on the Z chromosome between a microsatellite and an AFLP marker. We assign the names *Pher* and *Resp*, respectively, to the factors responsible for pheromone production and male behavioral response.

Differences in the pheromone blend produced by female ECB are likely due to changes in the specificity of the reactions in which Δ11-14-carbon-precursor acids are reduced and acetylated to produce the E and Z acetates, which are the pheromone components. Both hybrids and "pure" strain ECB females produce a ≈ 70:30 mixture of E/Z precursor acids. Because the acetylation of the alcohol precursors is not selective (ZHU *et al.* 1996) the 99:1 E/Z, 65:35 E/Z, and 3:97 E/Z pheromone blends are likely generated by the differential specificity of alleles at a locus encoding a reductase (ROELOFS *et al.* 1987; ROELOFS and WOLF 1988). Thus, a reductase represents a candidate gene for observed variation in pheromone blend production. A candidate gene for male response is less clear, but considering that males of both strains have the peripheral sensory physiology to detect all three pheromone blends (LINN *et al.* 1999), it seems reasonable to invoke a Z-linked gene encoding a protein involved in central nervous system processing of alternative pheromone input.

The *Tpi* locus is the only marker assayed thus far that exhibits significant differences in allele frequencies between the pheromone strains, and these differences persist even where E and Z borers are sympatric (GLOVER *et al.* 1991). Initial

divergence between the E and Z strains at *Tpi* might have been the result of a selective sweep at the locus itself or at a closely linked locus, but persistence of allele frequency differences in sympatry requires linkage to a factor or factors causing reduced hybrid fitness and/or assortative mating. Selective sweeps eliminate shared ancestral polymorphism and selection against hybrids eliminates introgressed genetic variation upon secondary contact (TING *et al*. 2000). In both cases, the region of reduced genetic variation is expected to be larger in regions of low recombination (*e.g*., AQUADRO 1997). Indeed, local recombination rates have been found to positively correlate with levels of genetic variation in *D. melanogaster* and in other species (AQUADRO 1997; KRAFT *et al*. 1998). In accord with these findings, our indirect method of estimating recombination rate using marker densities showed a 45% reduction from the average recombination rate on the Z chromosome in the area adjacent to *Tpi* (Figure 2.4). Although limited by its assumption of a random physical distribution of markers, this indirect approach provides a means to compare the relative differences in local recombination rates across the genome (NACHMAN and CHURCHILL 1996). Indeed, *Tpi* maps to a region that has the lowest estimated recombination rate (highest marker clustering) for the entire ECB genetic map, and therefore, may be more susceptible to selection across a broader genomic region.

Prior to this study, selection at the male response locus seemed a plausible explanation for the origin and maintenance of differentiation at *Tpi* because both are Z linked. Scenarios could be proposed involving either a selective sweep of alleles for male response or incompatibilities due to hybrid males having reduced success in obtaining mates. However, the map distance of 28.1 ± 4.1 cM between *Resp* and *Tpi* (Figure 2.3) makes it unlikely that *Tpi* allele frequencies have been influenced via hitchhiking associated with selection at the *Resp* locus. Furthermore, the persistence of linkage disequilibrium between *Resp* and *Tpi* in sympatric populations of E and Z

strains becomes more difficult to explain. One possibility is that *Resp* does have an influence on *Tpi*, but through its effects on assortative mating. Limited mating between E and Z borers would cause any initial linkage disequilibrium to decay slowly between *Resp* and *Tpi*. A second explanation is that strong premating isolation and/or substantially reduced F1 hybrid fitness impedes all gene flow between pheromone strains; such a strong barrier would maintain linkage disequilibrium indefinitely. Yet another possibility is that a second reproductive barrier on the Z chromosome is limiting gene flow at *Tpi*, possibly in addition to the effects of *Resp*.

Post-diapause development (PDD) describes the time it takes for overwintering larvae to pupate under diapause-breaking conditions. In upstate New York, populations of ECB exhibit either one or two generations per year as a consequence of individual differences in PDD (GLOVER *et al*. 1992). The shorter PDD exhibited by bivoltine moths translates into an early first generation and a late second generation. Univoltine moths tend to fly in the time window between the flight periods of these two generations. At the Geneva, NY site, where gene flow has been monitored, E strain moths are bivoltine and Z strain moths are univoltine. Thus at this site, differences in PDD result in temporal isolation of the two pheromone strains. Furthermore, PDD has been shown through its association with *Tpi* genotype to be linked to *Tpi* and therefore must be determined by a major gene (or cluster of genes) on the Z chromosome (GLOVER *et al*. 1992). If selection on PDD has promoted differentiation at *Tpi* (*e.g.*, through a selective sweep or reduced hybrid fitness) and/or is responsible for the maintenance of differentiation (*e.g.*, through temporal isolation or hybrid incompatibilities), then this factor should map much closer to *Tpi* than *Resp*. Moreover, if patterns of genetic differentiation in ECB are truly a function of linkage to potential "speciation" genes like *PDD*, *Resp*, and *Pher*, then markers that are further away on the Z chromosome should exhibit less differentiation.

Beyond the understanding of the genetics of mating signals and responses provided by comparisons of the ECB strains, broader comparisons between ECB and other *Ostrinia* species should be informative about the evolution of sexual communication in the genus. The adzuki bean borer (*Ostrinia scapulalis*), a congener of ECB, expresses striking similarity in its pheromone blend variation to ECB strains (HUANG *et al*. 2002) and the genes responsible for pheromone blend production and male response in the adzuki bean borer may map to the same location as *Resp* and *Pher* in ECB. This would be a significant finding because although parallel speciation, defined as the independent evolution of the "same" species, has been suggested (SCHLUTER and NAGEL 1995), cases of parallel speciation are strengthened if there is evidence for equivalent change in the same "speciation" genes. If comparisons of the ECB linkage map and a corresponding map for adzuki bean borer confirm such a relationship, support would be provided for the evolution of parallel reproductive barriers between incipient species (pheromone strains).

REFERENCES

AQUADRO, C. F., 1997 Insights into the evolutionary process from patterns of DNA sequence variability. Curr. Opin. Genet. Dev. 7: 835-840.

BRADSHAW, H. D., S. M. WILBERT, K. G. OTTO and D. W. SCHEMSKE, 1995 Genetic mapping of floral traits associated with reproductive isolation in Monkeyflowers (*Mimulus*). Nature 376: 762-765.

BUTLIN, R., and M. G. RITCHIE, 2001 Evolutionary biology: Searching for speciation genes. Nature 412: 31-33.

CAFFREY, D. J., and L. H. WORTHLEY, 1927 A progress report on the investigations of the European corn borer. U.S.D.A., Dept. Bull. No. 1476, Government Printing Office, Washington, D. C.

CARDÉ, R. T., W. L. ROELOFS, R. G. HARRISON, A. T. VAWTER, P. F. BRUSSARD *et al.*, 1978 European corn borer: pheromone polymorphism or sibling species. Science 199: 555-556.

CIANCHI, R., S. MAINI and L. BULLINI, 1980 Genetic distance between pheromone strains of the European corn borer, *Ostrinia nubilalis*: different contribution of variable substrate, regulatory and non-regulatory enzymes. Heredity 45: 383-388.

COYNE, J. A., and H. A. ORR, 1989 Patterns of speciation in *Drosophila*. Evolution 43: 362-381.

COYNE, J. A., and H. A. ORR, 1998 The evolutionary genetics of speciation. Phil. Trans. R. Soc. Lond. B 353: 287-305.

GLOVER, T., M. CAMPBELL, P. ROBBINS and W. ROELOFS, 1990 Sex-linked control of sex-pheromone behavioral responses in European corn borer moths (*Ostrinia*

nubilalis) confirmed with TPI marker gene. Arch. Insect Biochem. Phys. 15: 67-77.

GLOVER, T. J., J. J. KNODEL, P. S. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1991 Gene flow among three races of European corn borers (Lepidoptera: Pyralidae) in New York state. Environ. Entomol. 20: 1356-1362.

GLOVER, T. J., P. S. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1992 Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. Arch. Insect Biochem. Phys. 20: 107-117.

GREGORY, T. R., 2001 Animal genome size database. http://www.genomesize.com.

GUTHRIE, W. D., E. J. DOLLINGER and J. F. STETSON, 1965 Chromosome studies of the European corn borer, smartweed borer, and lotus borer (Pyralidae). Ann. Entomol. Soc. Am. 58: 100-105.

HAMILTON, M. B., E. L. PINCUS, A. DI FIORE and R. C. FLEISCHER, 1999 Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. Biotechniques 27: 500-507.

HARRISON, R. G., 1998 Linking evolutionary pattern and process, pp. 19-31 in *Endless Forms*, edited by D. J. HOWARD and S. H. BERLOCHER. Oxford Univ. Press, New York, NY.

HARRISON, R. G., and A. T. VAWTER, 1977 Allozyme differentiation between pheromone strains of European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). Ann. Entomol. Soc. Am. 70: 717-720.

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

HECKEL, D. G., L. J. GAHAN, Y. B. LIU and B. E. TABASHNIK, 1999 Genetic mapping of resistance to *Bacillus thuringiensis* toxins in diamondback moth using biphasic linkage analysis. Proc. Natl. Acad. Sci. USA 96: 8373-8377.

HUANG, Y. P., T. TAKANASHI, S. HOSHIZAKI, S. TATSUKI and Y. ISHIKAWA, 2002 Female sex pheromone polymorphism in adzuki bean borer, *Ostrinia scapulalis*, is similar to that in European corn borer, *O. nubilalis*. J. Chem. Ecol. 28: 533-539.

HUNT, G. J., and R. E. PAGE, 1995 Linkage map of the Honey Bee, *Apis mellifera*, based on RAPD markers. Genetics 139: 1371-1382.

KLUN, J. A., O. L. CHAPMAN, K. C. MATTES, WOJTKOWS.P. W., M. BEROZA *et al.*, 1973 Insect sex pheromones: minor amount of opposite geometrical isomer critical to attraction. Science 181: 661-663.

KLUN, J. A., and S. MAINI, 1979 Genetic basis of an insect chemical communication system: European corn borer. Environ. Entomol. 8: 423-426.

KLUN, J. A. and COOPERATORS, 1975 Insect sex pheromones - intraspecific pheromonal variability of *Ostrinia nubilalis* in North America and Europe. Environ. Entomol. 4: 891-894.

KOCHANSKY, J., R. T. CARDÉ, J. LIEBHERR and W. L. ROELOFS, 1975 Sex pheromones of the European corn borer in New York. J. Chem. Ecol. 1: 225-231.

KRAFT, T., T. SALL, I. MAGNUSSON-RADING, N. O. NILSSON and C. HALLDEN, 1998 Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). Genetics 150: 1239-1244.

LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALY *et al.*, 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

LIEBHERR, J., 1974 Studies on two strains of European corn borer, *Ostrinia nubilalis* (Hübner). Masters Thesis. Cornell University, Ithaca, NY.

LIEBHERR, J., and W. ROELOFS, 1975 Laboratory hybridization and mating period studies using two pheromone strains of *Ostrinia nubilalis*. Ann. Entomol. Soc. Am. 68: 305-309.

LINN, C., K. POOLE, A. ZHANG and W. ROELOFS, 1999 Pheromone-blend discrimination by European corn borer moths with inter-race and inter-sex antennal transplants. J. Compar. Phys. A 184: 273-278.

LINN, C. E., M. S. YOUNG, M. GENDLE, T. J. GLOVER and W. L. ROELOFS, 1997 Sex pheromone blend discrimination in two races and hybrids of the European corn borer moth, *Ostrinia nubilalis*. Phys. Entomol. 22: 212-223.

LÖFSTEDT, C., 1993 Moth pheromone genetics and evolution. Phil. Trans. R. Soc. Lond. B 340: 167-177.

LYON, M. F., 1976 Distribution of crossing-over in mouse chromosomes. Genet. Res. 28: 291-299.

MARCON, P., D. B. TAYLOR, C. E. MASON, R. L. HELLMICH and B. D. SIEGFRIED, 1999 Genetic similarity among pheromone and voltinism races of *Ostrinia nubilalis* (Hübner) (Lepidoptera: Crambidae). Insect Mol. Biol. 8: 213-221.

NACHMAN, M. W., and G. A. CHURCHILL, 1996 Heterogeneity in rates of recombination across the mouse genome. Genetics 142: 537-548.

ORR, H. A., 2001 The genetics of species differences. Trends in Ecology & Evolution 16: 343-350.

PARSONS, Y. M., and K. L. SHAW, 2002 Mapping unexplored genomes: A genetic linkage map of the Hawaiian cricket *Laupala*. Genetics 162: 1275-1282.

PEICHEL, C. L., K. S. NERENG, K. A. OHGI, B. L. E. COLE, P. F. COLOSIMO *et al.*, 2001 The genetic architecture of divergence between threespine stickleback species. Nature 414: 901-905.

PORNKULWAT, S., S. R. SKODA, G. D. THOMAS and J. E. FOSTER, 1998 Random amplified polymorphic DNA used to identify genetic variation in ecotypes of the European corn borer (Lepidoptera: Pyralidae). Ann. Entomol. Soc. Am. 91: 719-725.

PRESGRAVES, D. C., L. BALAGOPALAN, S. M. ABMAYR and H. A. ORR, 2003 Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature 423: 715-719.

RIESEBERG, L. H., J. WHITTON and K. GARDNER, 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152: 713-727.

RITCHIE, M. G., and S. D. F. PHILLIPS, 1998 The genetics of sexual isolation, pp. 291-308 in *Endless Forms*, edited by D. J. HOWARD and S. H. BERLOCHER. Oxford Univ. Press, New York, NY.

ROELOFS, W., T. GLOVER, X. H. TANG, I. SRENG, P. ROBBINS *et al.*, 1987 Sex-pheromone production and perception in European corn borer moths is determined by both autosomal and sex-linked genes. Proc. Natl. Acad. Sci. USA 84: 7585-7589.

ROELOFS, W. L., and W. A. WOLF, 1988 Pheromone biosynthesis in Lepidoptera. J. Chem. Ecol. 14: 2019-2031.

SCHLUTER, D., and L. M. NAGEL, 1995 Parallel speciation by natural selection. Am. Nat. 146: 292-301.

TING, C. T., S. C. TSAUR and C. I. WU, 2000 The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. Proc. Natl. Acad. Sci. USA 97: 5313-5316.

VIA, S., and D. J. HAWTHORNE, 1998 The genetics of speciation: promises and prospects of quantitative trait locus mapping, pp. 352-366 in *Endless Forms*,

edited by D. J. HOWARD and S. H. BERLOCHER. Oxford Univ. Press, New York, NY.

VOS, P., R. HOGERS, M. BLEEKER, M. REIJANS, T. VANDELEE *et al.*, 1995 AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407-4414.

WILLETT, C. S., and R. G. HARRISON, 1999 Insights into genome differentiation: Pheromone-binding protein variation and population history in the European corn borer (*Ostrinia nubilalis*). Genetics 153: 1743-1751.

ZHU, J. W., C. H. ZHAO, F. LU, M. BENGTSSON and C. LOFSTEDT, 1996 Reductase specificity and the ratio regulation of E/Z isomers in pheromone biosynthesis of the European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). Insect Biochem. Mol. Biol. 26: 171-176.

CHAPTER 3

CONSEQUENCES OF REPRODUCTIVE BARRIERS FOR GENEALOGICAL DISCORDANCE IN THE EUROPEAN CORN BORER

ABSTRACT

Speciation involves the origin of trait differences that limit or prevent gene exchange and ultimately results in daughter populations that form monophyletic or exclusive genetic groups. However, for recently diverged populations or species, between which reproductive isolation is often incomplete, gene genealogies will be discordant and most regions of the genome will display nonexclusive genealogical patterns. In these situations, genome regions for which one or both species are exclusive groups may mark the footprint of recent selective sweeps. Alternatively, such regions may include or be closely linked to "speciation genes," genes involved in reproductive isolation. Therefore, comparisons of gene genealogies allow inferences about the genetic architectures of both reproductive isolation and adaptation. Contrasting genealogical relationships in sexually isolated pheromone strains of the European corn borer moth (ECB) demonstrate the relevance of this approach. Genealogies for five gene regions are discordant, and for only one molecular marker, the sex-linked gene *Tpi*, are the two pheromone strains exclusive groups. *Tpi* maps to a position on the sex chromosome that is indistinguishable from a major factor affecting differences in post-diapause development time (*Pdd*). The major gene determining male behavioral response to pheromone (*Resp*) is also sex-linked, but maps 20-30 cM away. Exclusivity at *Tpi* may be a consequence of these linkage relationships, because evidence from phenotypic variation in natural populations

implicates both *Pdd* and *Resp* as candidates for genes involved in recent sweeps and/or reproductive isolation between strains.

INTRODUCTION

Describing and interpreting historical patterns of descent and diversification are the principal goals of evolutionary biology. These patterns are commonly inferred from DNA sequence data, using genealogies or phylogenies based on single gene regions. However, the reconstructed relationships between recently diverged lineages often fail to reveal species as monophyletic or exclusive groups (groups in which all members are more closely related to each other than to individuals outside the group). For an extended period following speciation, patterns of descent inferred from DNA sequence data will vary across the genome by chance alone, and shared ancestral polymorphism will be common (HUDSON and COYNE 2002; NEIGEL and AVISE 1986; TAJIMA 1983). As a consequence, multiple gene genealogies for the same set of closely related species produce discordant trees (BELTRAN *et al.* 2002; BROUGHTON and HARRISON 2003; KLIMAN *et al.* 2000; MACHADO and HEY 2003; TING *et al.* 2000; WANG *et al.* 1997). However, genetic variation linked to advantageous alleles within daughter species can "hitchhike" to fixation along with the selected allele, resulting in the purging of ancestral polymorphism and a more rapid approach to exclusivity or monophyly (MAYNARD SMITH and HAIGH 1974). The size of the chromosomal region affected by these "selective sweeps" is determined by the strength of selection and the rate of recombination.

Once a pattern of exclusivity becomes established in geographically isolated populations, it will persist, but hybridization and gene flow may cause the pattern to erode when populations come into secondary contact. At "speciation genes," loci for

which trait differences result in hybrid unfitness or positive assortative mating, gene exchange is impeded or prevented between races or species, effectively maintaining exclusive patterns (BARTON and HEWITT 1981; HARRISON 1990; WU 2001). Elsewhere in the genome, introgression may be extensive, leading to shared alleles and an erosion of exclusive relationships. The result is a semi-permeable species boundary, with permeability depending on the genetic marker (BARTON and HEWITT 1981; HARRISON 1990; WU 2001). Thus, the genetic architectures of both reproductive isolation and selective sweeps have important consequences for genealogical patterns. Specifically, markers closely linked to speciation genes or to loci that have experienced a recent selective sweep are more likely to exhibit an exclusive relationship between pairs of recently diverged species (or subspecies, races, and strains).

In this study, we focus on reproductively isolated strains of the European corn borer moth (ECB), *Ostrinia nubilalis*. ECB is native to Europe, North Africa, and parts of western Asia, but it was introduced into North America on broom corn from Italy and Hungary early in the 20th century (CAFFREY and WORTHLEY 1927). In both Europe and North America, ECB consists of two behaviorally isolated strains that differ in the sex pheromone produced by females and the response elicited in males. In the Z strain, females produce and males respond to a 3:97 mixture of E and Z-Δ11-14:OAc, whereas in the E strain females produce and males respond to a 99:1 (E)/(Z) blend (KLUN *et al.* 1973). Sexual isolation stems from a stereotypic male response to these alternative pheromone blends (GLOVER *et al.* 1990; ROELOFS *et al.* 1987). Although there are clearly two distinct sexual communication systems, with cross attraction rarely occurring in the laboratory, hybrid genotypes can be found in nature at North American sites (GLOVER *et al.* 1991; KLUN and HUETTEL 1988). The genetic factors responsible for major differences in female pheromone blend production, *Pher*,

and male behavioral response to those blends, *Resp*, exhibit simple Mendelian inheritance (DOPMAN *et al.* 2004; GLOVER *et al.* 1990; KLUN and MAINI 1979; ROELOFS *et al.* 1987). *Pher* is autosomal, whereas *Resp* is sex linked.

ECB populations are also characterized by variation in numbers of generations per year (voltinism). In New York State, populations are either bivoltine with an early first generation and a late second generation, or univoltine with a single generation in the middle of the season (ROELOFS *et al.* 1985). Adults from the univoltine population are temporally isolated from those of the bivoltine population. Voltinism patterns reflect differences in post-diapause development time (PDD), the time to pupation for over-wintering larvae under temperature and photoperiod conditions conducive to breaking diapause. PDD is controlled by a major factor on the sex chromosome (GLOVER *et al.* 1992). Based on voltinism and sex pheromone blend, New York ECB populations consist of three distinct races: univoltine Z (UZ), bivoltine E (BE), and bivoltine Z (BZ) (GLOVER *et al.* 1991; KLUN and HUETTEL 1988).

Although diverged with respect to sexual communication and life history, strains or races of ECB are otherwise difficult to distinguish (LIEBHERR 1974; LIEBHERR and ROELOFS 1975). Genetic surveys of allozymes (CIANCHI *et al.* 1980; HARRISON and VAWTER 1977), mitochondrial DNA (MARCON *et al.* 1999), RAPDs (PORNKULWAT *et al.* 1998), and a nuclear gene (WILLETT and HARRISON 1999) have revealed extensive shared genetic variation, but little differentiation. However, significant allele-frequency differences occur at a sex-linked gene that encodes the enzyme triose phosphate isomerase (TPI). At this locus, BE populations in New York are fixed for the *Tpi-1* allele and UZ and BZ populations are segregating for both *Tpi-1* and *Tpi-2*, with *Tpi-2* being the more common allele (GLOVER *et al.* 1991). Allele frequency differences between pheromone strains for the sex-linked *Tpi* locus might be maintained by tight linkage to the gene for male response, but using inter-strain

crosses Dopman et al. mapped *Resp* > 20 cM away from *Tpi* on the sex chromosome, and argued that the origin of patterns of variation at *Tpi* cannot be explained by linkage to this factor.

Here, we use the ECB genetic linkage map as a framework for a comparative genealogical analysis of sequence data that includes *Tpi* and four additional molecular markers. We infer genealogical relationships for these five markers (three sex-linked, one autosomal, and mtDNA) for each strain and an outgroup species. We then show that genealogies are discordant and that only *Tpi* reveals evidence of exclusivity between Z and E moths. Finally, we address the relationship between genealogical patterns and the genetic architecture of reproductive barriers by mapping our markers relative to the positions of *Resp* and *Pdd*.

MATERIALS AND METHODS

Insect Populations

Insect mapping families: Two cultures of ECB, maintained by Wendell Roelofs and colleagues at the New York State Agricultural Experiment Station (NYSAES) in Geneva, NY, were used as sources for initiating mapping families. The first culture consisted of UZ-strain insects that were derived from field-collected larvae, pupae, and adults found in corn stubble from Bouckville, New York in April 1994. The second culture consisted of BE-strain insects derived from corn stubble near Geneva, New York in May 1996. Details of mass rearing of insects can be found in (ROELOFS *et al.* 1985).

Field collected insects: Both pheromone strains occur in New York, and sympatric populations have been documented at a number of localities (GLOVER *et al.*

1991; ROELOFS *et al.* 1985). The Geneva, NY population has been monitored for ECB and consists of BE and UZ moths [e.g., (ROELOFS *et al.* 1985)]. Thus, female larvae and pupae collected from corn stubble in May of 2000 and 2004 that were reared under diapause-breaking conditions to assess post-diapause development (PDD) time could be scored as BE or UZ. Fourteen BE and six UZ females were used for genetic analysis from this locality. Eight UZ females were collected for analysis from a UZ population in Madison, NY in May 2000, and four BZ females were collected from a BZ population in Eden, NY in October 2000.

Both pheromone strains also occur in North Carolina, with Z borers found in the western part of the state and E borers more common in the east. A zone of overlap exists where males are found in pheromone traps baited with either the Z or E blends (SORENSON *et al.* 1992). For genealogical analysis we used three female ECB collected in May 2002 from potato plants in Weeksville, Pasquotank Co., N.C. (E strain) and two females collected from corn in Fletcher, Henderson Co., N.C. (Z strain). An Asian corn borer female (*Ostrinia furnicalis*) was used as an outgroup species.

Marker Development

AFLP and microsatellite markers: Development of AFLP and microsatellite markers is described in (DOPMAN *et al.* 2004).

Triose phosphate isomerase: A *Tpi* fragment was initially amplified with degenerate primers designed using *Tpi* sequences from *Heliothis virescens* (GenBank accession U23080), *Spodoptera littoralis* (L39011) *Drosophila melanogaster* (AE003772), and *Anopheles merus* (U82707). ECB-specific primers were then

designed and used for PCR amplification of most of the *Tpi* sequence, with the exception of the furthest 3' exon. The 3' end of the gene was obtained by PCR amplification from cDNA. mRNA was isolated using Oligotex extraction (Qiagen, Valencia, CA), reverse transcribed, and PCR amplified with the ECB and polyT linker primers. Thermal cycling used the profile: 94° for 45 sec, 50° for 45 sec, and 72° for 1.3 min. The ~1.6 kb genomic fragment that includes the entire coding region plus introns was amplified using primers ECBtpi_for1A (5' AGATGTCAAAATTCAACTCAG) and ECBtpi_rev5 (5' AGCACCCTTCGGCACTT), and sequenced using this primer pair and two internal primers (ECBtpi_for5A: 5' AGGCAGACCAAGGCACTCTTGCC, ECBtpi_rev3A: 5' TTCGGTACCGATGGCCCATACAG). *Tpi* was placed on the Z linkage map as described in Dopman et al. (DOPMAN *et al.* 2004). We assessed electrophoretic mobility for TPI in a subset of field-collected insects to compare mobility class with *Tpi* sequence in an effort to identify the amino acid substitution responsible for electrophoretic mobility variation. Because female Lepidoptera are heterogametic, female ECB possess a single copy of either the *Tpi-1* or the *Tpi-2* allele.

Kettin: Degenerate primers provided by P. Andolfatto were used for initial *Kettin* (*Ket*) amplification, and ECB-specific primers amplified a ~1.3 kb fragment (ECBketF: 5' TGAAATCCCGGAACCAGTAACA, ECBketR: 5' TTGAGGTGAGTAGTGAAAATAGGAG) under the amplification profile: 94° for 45 sec, 53.5° for 45 sec, and 72° for 1.3 min. Sequencing primers were ECBket308F: 5' CTAGGTGAAGCAGTAACGACAGC and ECBket348R: 5' ATCCAAAGTAACGAATCCGAAATC. *Ket* was mapped either by a diagnostic restriction site (*Rsa*I) or by amplification using a primer that extends over a polymorphic indel (ECBket_RSA3F: 5'

TATGAATCAGTTACCTACATAACTAGGTAC). Restriction enzyme digestion followed manufacturer's protocols (NEB, Beverly, MA), and PCR followed previous *Ket* cycling conditions with a change in annealing temperature to 50.7°.

Lactate dehydrogenase: A BLASTN search of the *Bombyx mori* EST database (http://www.ab.a.u-tokyo.ac.jp/silkbase/) with a *Papilio glaucus Lactate dehydrogenase* (*Ldh*) sequence (P. Andolfatto, unpublished data) yielded a significant alignment (1e-52; EST sequence number wdS30992). Primers were designed from *B. mori* sequence that was conserved with respect to the *P. glaucus* sequence (Ldhbm_65F: 5' ATCGCCAGTAACCCCGTGG, Ldhbm_376R: 5' CGATAGCCCAGGAAGTGTATCCCTTC), and ECB-specific internal primers were developed (ECB_ldhF: 5' GGCTCCGGCACCAACCTGGACTC, ECB_ldhR: 5' CGTAGGCGCTCTTCACCACCATCTCA). Other than a 55° annealing temperature for the *B. mori*-derived primers and a 58° annealing temperature for the ECB-derived primers, cycling conditions for *Ldh* were identical to those used for *Ket*. PCR products were ~600 and ~500 bp, respectively, for the two primer pairs. Like *Ket*, *Ldh* was mapped using either a diagnostic restriction site (*Alu*I) or by amplification using a primer that extends over a polymorphic indel (ECBldh_E84R: 5' GAATATCAGAACAAACAAAGGTC). Digestion followed manufacturers' protocols (NEB, Beverly, MA), and PCR followed previous cycling conditions with a change in annealing temperature to 64.2°.

Pheromone binding protein: A ~1.6 kb fragment of DNA that included *Pheromone binding protein* (*Pbp*) was amplified using primers ECEP5 and ECPA as described by Willett and Harrison (WILLETT and HARRISON 1999). Amplified fragments were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) and

sequence data were obtained for the 5' end of the gene (WILLETT and HARRISON 1999). ACB sequence was AF133630 from GenBank.

Cytochrome oxidase I: A ~1.2 kb fragment of *COI* was amplified using primers Ron (5' GGATCACCTGATATAGCATTCCC) and Pat (5' TCCAATGCACTAATCTGCCATATTA) at 50° annealing, but otherwise, thermal cycling was the same as above.

PCR and sequencing: All PCR reactions and sequencing with an ABI PRISM 377 (Applied Biosystems, Foster City, CA) followed protocols described in (DOPMAN *et al.* 2004). Sequence data were manipulated with the DNASTAR programs (Madison, WI) using default parameters.

Genealogical analysis

Tests of exclusivity between E and Z moths were conducted by comparing ML trees with those obtained from heuristic searches with the constraint tree: (ACB, ((E), (Z))) or ((E), (Z)). When ACB was not specified in the constraint, it could be attached on any branch and the resulting tree would still be compatible with the constraint. Exclusivity could therefore be defined in terms of the monophyly of E and paraphyly of Z, monophyly of Z and paraphyly of E, or reciprocal monophyly. Shimodaira-Hasegawa tests [SH test; (SHIMODAIRA and HASEGAWA 1999)] in PAUP* (v. 4.0b10) (SWOFFORD 2000) were used to compare the ML tree with the ML tree compatible with the constraint using 10,000 RELL bootstrap replicates. Optimal substitution models under ML for each data set were identified using Modeltest (v.3.6) (POSADA and CRANDALL 1998), but a full-parameterized GTR+Γ+I model was used for SH tests

in order to maximize the ML scores and to minimize biases due to differences in the number of free parameters between trees.

Other genealogical analyses were conducted in PAUP* under the maximum parsimony (MP) optimality criterion. All MP analyses used heuristic searches that assumed unordered, equal-weight characters with gaps treated as a "fifth base." Multiple-base insertions/deletions (indels) were down-weighted by reducing to single-base indels under the assumption that such characters represent single evolutionary changes. Starting trees were obtained using stepwise addition, the addition sequence was random using 10 replicates, the branch-swapping algorithm was TBR, and MULTREES was in effect. Bootstraps values were obtained using a heuristic search with stepwise addition, random-addition sequence, TBR branch-swapping, and 1,000 replicates. Finally, the congruence of characters from each marker was assessed by performing a partition-homogeneity test with 1,000 replicates.

Z chromosome genetic map

Details of crosses, linkage map construction, and phenotype assessment of male behavioral response can be found in (DOPMAN *et al.* 2004) and references therein. The reported Z linkage map includes data from 116 BC1 progeny (78 males and 38 females), all of which were genotyped for *Ket* (this study), *Ldh* (this study), and *Tpi* (DOPMAN *et al.* 2004). Microsatellite marker ma169 was genotyped for all male progeny, which were also phenotyped for male behavioral response. Finally, seventy-nine progeny (41 males and 38 females) were genotyped for eight AFLP markers.

PDD has been shown through its association with TPI electrophoretic mobility to be linked to the *Tpi* gene and therefore must map to the Z chromosome . A BC1

family was generated to map PDD differences on the Z chromosome using a three-point test cross with *Tpi* and *Ldh* as markers. After hatching BC1 progeny, a 12:12 L:D photoperiod was used to induce maximum diapause . After 35 days, borers in diapause were reared under diapause-breaking conditions of 16:8 L:D and the time to pupation was noted every two days. Under these conditions, the average PDD for parental stocks is 14.50 (SE: ± 0.55) for bivoltine females and 43.94 (SE: ± 1.64) for univoltine females .

RESULTS

Genealogical analysis

Aligned sequence data for *COI*, *Ket*, *Ldh*, and *Tpi* consisted of 17 E and 20 Z-strain ECB, plus the ACB outgroup (Table 3.1). Only 12 of the 17 E-strain insects were sequenced for *Pbp*. After reduction of multiple-base indels, the aligned data sets ranged in size from 1449 characters for *Tpi* to 291 characters for *Ldh* (*Ldh* had 167 intronic characters removed that were difficult to align) (Table 3.1). Although the length of the *Ldh* data set was smaller than those for the other markers, it contained the second largest proportion of variable to total characters (*Tpi*: 55/1449; *Ket*: 37/690; *Ldh*: 31/291; *Pbp*: 59/477; *COI*: 35/1195). Loci showed a wide range of ingroup nucleotide diversity and average sequence divergence to the outgroup (Table 3.1). The average pairwise-sequence divergence within ECB ranged from 0.1% for COI to 2.6% for *Ldh*. Average sequence divergence to the outgroup ranged from 1% to > 3.5%.

Characters from the five markers were significantly incongruent, as indicated by the partition-homogeneity test ($P = 0.001$). Non-parametric bootstrap support

Table 3.1. Summary information for sampled markers after reducing multiple-base indels.

Locus	L [a]	C [b]	$\bar{d}_{within}$ [c]	d_{within} [d]	$\bar{d}_{between}$ [e]
COI	1195	8	0.11	0-0.34	2.29
Ket	690	23	0.62	0-1.85	1.1
Ldh	291	23	2.59	0-7.25	3.56
Pbp	477	32	2.12	0-4.65	3.14
Tpi	1449	16	0.39	0.41-0.76	1.98

[a] Total number of characters.
[b] No. of parsimony-informative characters, with indels.
[c] Average divergence (%) within ECB.
[d] Range of average divergence (%) between Z and E ECB.
[e] Average divergence (%) between ECB and ACB.

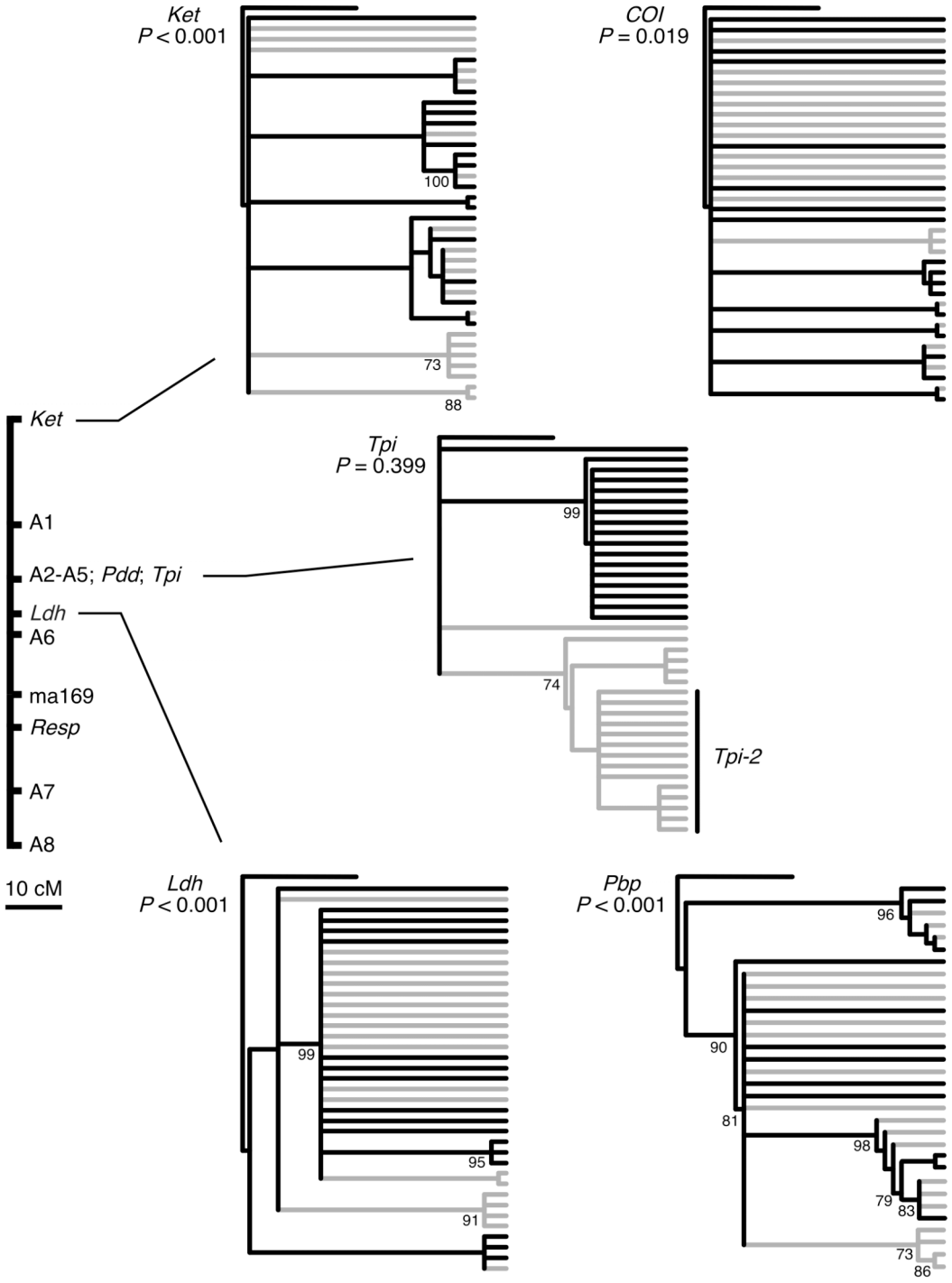
for clades varied by locus, with *COI* showing little overall clade support and the other loci each containing at least two nodes with ≥ 70% support (Figure 3.1). Although strict reciprocal monophyly between E and Z moths was not present in any gene tree, the *Tpi* genealogy came close. Only two clades had bootstrap support ≥ 70% in *Tpi*; one which contained 16 of the 17 E-strain borers (99% support) and the other that contained 19 of the 20 Z-strain borers (74% support). Additionally, there were no shared haplotypes between E and Z moths at *Tpi* (Table 3.1). Each of the other markers had at least one group that contained multiple Z and E moths with identical haplotypes. Groups ranged in size from one E and two Z moths for *Ket* to seven E and eight Z moths for *COI*.

For each locus, we tested the hypothesis that Z and E strains of ECB form exclusive genealogical groups (Figure 3.1). We compared the likelihood of the ML unconstrained tree with the likelihood of the tree that enforced a pattern of exclusivity for the Z and E strains. The exclusivity hypothesis, ((E), (Z)), was rejected for *COI*, *Ket*, *Ldh*, and *Pbp,* but not for *Tpi* (SH test, *COI*: $P = 0.019$; *Ket*: $P < 0.001$; *Ldh*: $P < 0.001$; *Pbp*: $P < 0.001$; *Tpi*: $P = 0.399$) (Figure 3.1). The ML tree for *Tpi* also did not differ significantly from one in which a reciprocal monophyletic relationship exists between Z and E pheromone strains [i.e., (ACB, ((E), (Z))); $P = 0.399$].

The choice of phylogenetic analysis had no significant effect on tree topology. For each locus, the ML tree did not significantly differ from the MP tree(s) (gaps included, Wilcoxon sign-ranks test, *COI*: $P = 1.0$; *Ket*: $P \geq 0.18$; *Ldh*: $P = 1.0$; *Pbp*: $P \geq 0.21$; *Tpi*: $P \geq 0.26$). Gene genealogies shown in Figure 3.1 represent the consensus of reconstructions for the MP analysis.

Character-state optimization on an MP tree for moths characterized electrophoretically for TPI located a single character-state change on the branch separating females that are hemizygous for the *Tpi-1* allele from females that have the

Figure 3.1 Genealogies and mapping positions for molecular markers and traits divergent between ECB pheromone strains. A genetic linkage map of the Z (sex) chromosome shows the mapping positions for *Ket*, *Tpi*, *Ldh*, AFLPs (A1-A8), microsatellite ma169, male behavioral response (*Resp*), and post-diapause development (*Pdd*). Also shown are consensus MP gene genealogies for the sex-linked markers *Ket*, *Tpi*, and *Ldh*, the mtDNA marker, *COI*, and the autosomal locus, *Pbp.* Genealogies are derived from Z and E-strain females collected in New York and North Carolina. E strain borers are in black, whereas Z strain borers are in grey. The *p*-value associated with each genealogy indicates whether the observed ML tree differs significantly from a ML tree that forces an exclusive relationship between Z and E borers. Values near branches represent bootstrap support. ECB exhibiting the *Tpi-2* electrophoretic form of the enzyme TPI are indicated on the *Tpi* genealogy.


Ket
P < 0.001
100
73
88
COI
P = 0.019
Ket
A1
A2-A5; Pdd; Tpi
Ldh
A6
ma169
Resp
A7
A8
10 cM
Tpi
P = 0.399
99
74
Tpi-2
Ldh
P < 0.001
99
95
91
Pbp
P < 0.001
96
90
81
98
79
83
73
86

Tpi-2 allele. This single character-state change is a non-synonymous substitution at site 1194 in the reduced data set. Moths that carry the *Tpi-1* allele have an asparagine at amino acid residue 189, whereas moths that carry the *Tpi-2* allele have a lysine at this position. The substitution of a lysine for an asparagine is the only character-state change that maps to the branch separating females with different TPI alleles (Figure 3.1). It is also one of only two non-synonymous substitutions in the data set, the other being an autapomorphic arginine to threonine substitution. In both ACB and the moth *Heliothis virescens*, *Tpi* sequences have an asparagine at residue 189, suggesting that the *Tpi-2* allele is derived within the ECB lineage. The substitution of a positively charged lysine for a neutral amino acid accounts for the slower anodal migration of the *Tpi-2* allele in gel electrophoresis.

Z chromosome genetic map

The expanded Z chromosome map has a total length of 76 cM, and is bounded on one end by *Ket* and by an AFLP marker on the other (Figure 3.1). *Tpi* and four AFLP markers map 29 cM from *Ket*, *Ldh* is 6 cM further along the sex chromosome, and *Resp* maps 20 cM beyond *Ldh*. In the three-point test cross used to map the major factor for PDD, hemizygous females from the BC1 mapping family showed a clear bimodal distribution for female PDD time (Figure 3.2). A 12-day period starting on day 25 and ending on day 36 divides females into bivoltine (fast PDD) and univoltine (slow PDD) groups that correspond with parental PDD. Out of 41 fast-developing bivoltine females, 37 expressed the *Tpi-1*/*Ldh-1* genotype and four expressed the *Tpi-1*/*Ldh-2* genotype. Out of 30 slow-developing univoltine females, 27 expressed the *Tpi-2*/*Ldh-2* genotype and three expressed the *Tpi-2*/*Ldh-1* genotype. Thus, we

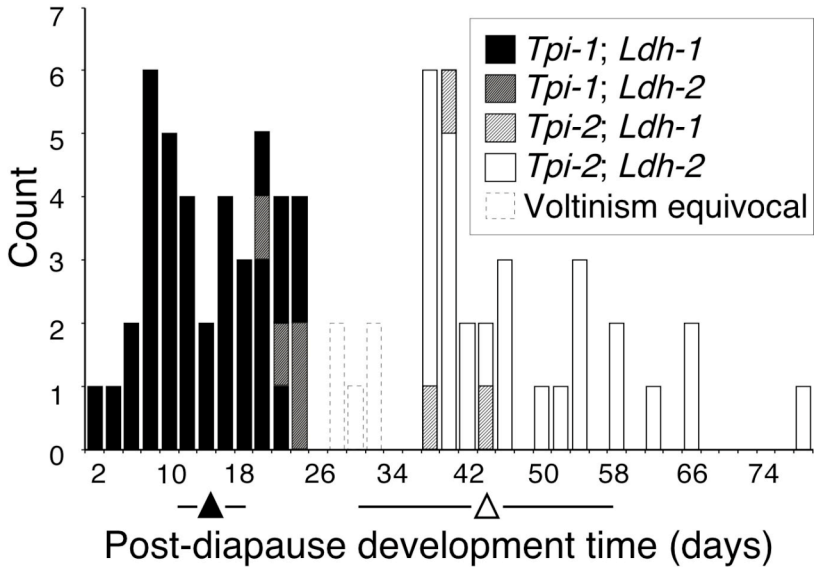


Figure 3.2 Histogram showing bimodal post-diapause development (PDD) time for female BC1 offspring genotyped for *Tpi* and *Ldh*. The 12-day period starting on day 25 and ending on day 36 divides females into bivoltine (fast PDD, N = 41) and univoltine (slow PDD, N = 30) groups that correspond with parental phenotypes. Symbols and bars below the X-axis indicate the average and standard deviation of PDD for parental bivoltine (▲, N = 40) and univoltine (△, N = 70) genotypes (GLOVER *et al.* 1992).

observed no recombinants between PDD time and *Tpi*, and seven recombinants between *Tpi* and *Ldh* [which did not differ from (DOPMAN *et al.* 2004); $\chi^2 = 0.823$; $P = 0.364$]. The factor for PDD (*Pdd*) therefore maps to the same position as *Tpi* (Figure 3.1).

DISCUSSION

Although speciation ultimately results in divergence across the entire genome, random lineage sorting and introgression will cause shared polymorphism to persist among recently diverged populations and closely related species. Indeed, multi-locus studies of closely related species frequently report discordant genealogical patterns that fail to support species boundaries based on morphological, behavioral, and ecological characters (BELTRAN *et al.* 2002; BROUGHTON and HARRISON 2003; MACHADO and HEY 2003). In this study, we focus on the behaviorally isolated Z and E pheromone strains of European corn borer moth. We assess the genealogical discordance for five independent molecular markers, and test for evidence of an exclusive genetic relationship that agrees with the history of reproductive isolation between strains.

In ECB, gene genealogies based on *COI*, *Ket*, *Ldh*, *Pbp*, and *Tpi* exhibit significant discordance ($P = 0.001$; Figure 3.1), and exclusivity of Z and E strains is rejected for four out of five markers. The genealogies for *COI*, *Ket*, *Ldh*, and *Pbp* each depict an interdigitated pattern between pheromone strains that is significantly more likely than a genealogy that enforces strain monophyly (Figure 3.1). *COI* is less variable than the other markers (Table 3.1), and mtDNA does not clearly support any haplotype grouping. *Ket*, *Ldh*, and *Pbp*, however, each exhibit high nonparametric bootstrap support for at least two haplotype groups, but none show support for distinct

Z and E clades (Figure 3.1). In contrast to the genealogical patterns at the other four markers, the genealogy for *Tpi* supports the existence of two distinct haplotype groups composed of Z and E-strain moths. Moreover, the *Tpi* tree is not significantly different from one in which strains form reciprocally monophyletic groups ($P = 0.339$; Figure 3.1). The test has sufficient power to reject the monophyly hypothesis, given that *COI* rejects the null hypothesis at the 5% level but harbors only 50% of the variation at *Tpi* (Table 3.1; Figure 3.1).

Although allozyme electrophoresis might suggest that the two pheromone strains share the *Tpi-1* "allele," genealogical analysis demonstrates that E strain and Z strain *Tpi-1* alleles fall into two distinct haplotype groups (Figure 3.1). Therefore, sharing of alleles with the same electrophoretic mobility does not reflect similarity due to recent common ancestry. Furthermore, the *Tpi-2* allele, found only in the Z strain, is apparently recently derived from Z-strain *Tpi-1* via a single non-synonymous substitution.

Possible explanations for exclusivity at *Tpi* and extensive haplotype sharing at the other four markers include random lineage extinction, selective sweeps, and differential introgression. With genetically isolated daughter populations of equal and constant size, the probability of exclusivity due to random lineage extinction for neutral markers depends largely on the time since divergence and the effective copy number of the marker. For example, because mtDNA is haploid and maternally inherited, monophyly is expected with high probability well before monophyly at even a small number of diploid autosomal loci (HUDSON and COYNE 2002). Although random lineage sorting may explain some genealogical variation among markers sampled within ECB, the observation that mtDNA does not show any evidence of exclusivity between strains argues against random lineage extinction as the only explanation for pattern differences between *Tpi* and the other markers.

Selective sweeps influence the sorting process by reducing the time to achieve reciprocal monophyly between diverged populations (MAYNARD SMITH and HAIGH 1974). Whether selective sweeps not associated with lineage-specific adaptations become restricted to only one daughter lineage, and are therefore able to generate rapid exclusivity, depends on the amount of gene flow. Appreciable gene exchange will allow positively selected genomic regions to spread in both daughter lineages, whereas limited gene flow can isolate sweeps to a single daughter population.

Gene flow and selection can interact in another way to produce discordant patterns. When diverged populations hybridize, alleles at genes encoding species-specific adaptations or reproductive barriers (and closely linked regions) will introgress less than neutral or positively selected markers (BARTON and HEWITT 1989; MACHADO and HEY 2003; RIESEBERG *et al.* 1999). Therefore, genes contributing to isolation will continue to exhibit patterns of exclusivity while other gene regions become homogenized by introgression. This process would be expected to generate a positive association between the genomic location of genes contributing to isolation and markers showing genetic divergence, a pattern that has been found among closely related *Drosophila* species (MACHADO and HEY 2003; NOOR *et al.* 2001).

We have extended a previous genetic map for the ECB sex chromosome (DOPMAN *et al.* 2004) to obtain perspective about the genomic distribution of the discordant Z-linked genealogies relative to each other, and to the Z-linked genes for male behavioral response (*Resp*) and post-diapause development (*Pdd*), genetic factors with divergent phenotypic effects in ECB (Figure 3.1). We found that *Ket* mapped to one end of the linkage group, and that *Ldh* mapped between *Tpi* and *Resp. Pdd* mapped to a position indistinguishable from *Tpi*.

When combined with genealogical information and field data, the Z chromosome linkage map reveals several interesting features. The first is that apparent

disequilibrium exists between *Tpi* and *Resp* in sympatric localities (GLOVER *et al.* 1991), in spite of > 20 cM distance separating these two loci (Figure 3.1). A fixed inversion difference could explain such a pattern (NOOR *et al.* 2001), but our mapping data (based on E x Z crosses) do not support this possibility (DOPMAN *et al.* 2004). A second feature of the map is that *Ldh* maps closer to *Resp* than *Tpi*, and yet unlike *Tpi*, *Ldh* shows a high degree of haplotype sharing between pheromone strains and rejects exclusivity ($P < 0.001$). Perhaps the most striking feature of the map, however, is the indistinguishable mapping positions of *Tpi* and *Pdd*. That is, the only locus that shows any evidence for exclusivity maps close to a factor affecting differences in development time.

One explanation for the observed pattern of locus-specific exclusivity is that limited introgression (and persistent disequilibrium) characterizes much of the sex chromosome, but disequilibrium and monophyly have become discernible only in those regions that have experienced a recent selective sweep (e.g., the *Pdd*/*Tpi* region). For many gene regions between *Resp* and *Pdd* (e.g., *Ldh*), ancestral polymorphisms have persisted and the Z and E strains are not exclusive groups. Sharing of haplotypes and absence of discernible disequilibrium where Z and E moths are sympatric would then reflect persistent ancestral polymorphism and not contemporary gene exchange. In this scenario, it remains possible that *Resp* does play an important role in slowing the decay of linkage disequilibrium between *Tpi* and *Resp* by limiting or preventing introgression between the Z and E strains.

If contemporary introgression (rather than shared ancestral polymorphism) between Z and E strains is responsible for shared haplotypes at *Ket*, *Ldh*, *Pbp*, and *COI*, then *Tpi* presumably marks the location of a speciation gene that renders the surrounding chromosomal region immune to the homogenizing effects of introgression. F1 hybrids between Z and E moths do exist in nature (GLOVER *et al.*

1991; KLUN and HUETTEL 1988), but the reproductive fate of these moths remains unknown. If F1 hybrids backcross, the factor linked to *Tpi* would have to be working in concert with *Resp* through epistatic selection for disequilibrium between *Tpi* and *Resp* to persist. Indeed, it may be difficult for recombinant moths in sympatric localities (e.g., UE moths) to find mating partners if they are sexually active at the "wrong" time (i.e., when the opposite strain is active).

Under the scenario involving contemporary introgression, the observation that *Tpi* remains differentiated with respect to pheromone strain argues against differences persisting simply as a result of temporal isolation from *Pdd*. *Tpi-2* alleles, found at intermediate frequency in BZ populations, have never been found in the BE race (GLOVER *et al.* 1991). Therefore, gene flow between pheromone strains is absent in this gene region even when the strains have the same life cycle.

The observation that the only marker which reveals strains of ECB to be monophyletic (*Tpi*) also maps to the same position as a major factor for a divergent phenotype (*Pdd*) strongly suggests a role for natural selection. This observation lends support to a model of speciation in which species become differentiated at some gene regions through natural selection, while at other loci shared variation persists because of recent or historical gene flow. The Drosophila species group *D. pseudoobscura*, *D. p. bogotana*, and *D. persimilis* have also been cited as examples of the importance of selection in speciation because reciprocal monophyly only occurs at markers linked to regions containing hybrid sterility genes (MACHADO and HEY 2003; NOOR *et al.* 2001). In ECB, if natural selection has had a role in shaping patterns of genetic variation at *Tpi*, either through a recent selective sweep or by eliminating incompatibilities in hybrid offspring, then *Pdd*, which shows tight linkage with *Tpi*, provides a candidate target of selection. Confirming that natural selection, as opposed

to non-equilibrium demography, has shaped patterns of genetic variation among markers represents the next step in understanding the genetics of speciation in ECB.

The closely related species, subspecies, races, or strains that form the foundation for speciation research often exhibit discordant phylogenies or gene genealogies for different molecular markers. The origin of monophyly, often associated with selective sweeps in these groups, may provide the diagnostic markers that allow us to identify the genomic location of reproductive barriers or adaptations. Genealogies from these (and other) gene regions help to clarify patterns of descent and diversification, and when combined with a genetic map containing the locations of barriers and sweeps, reveal a more complete picture of the speciation process.

REFERENCES

BARTON, N. H., and G. M. HEWITT, 1981 Hybrid zones and speciation, pp. 109-145 in Evolution and Speciation, edited by W. R. ATCHLEY and D. S. WOODRUFF. Cambridge Univ. Press, Cambridge, UK.

BARTON, N. H., and G. M. HEWITT, 1989 Adaptation, speciation and hybrid zones. Nature 341: 497-503.

BELTRAN, M., C. D. JIGGINS, V. BULL, M. LINARES, J. MALLET *et al.*, 2002 Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating Heliconius butterflies. Mol Biol Evol 19: 2176-2190.

BROUGHTON, R. E., and R. G. HARRISON, 2003 Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics 163: 1389-1401.

CAFFREY, D. J., and L. H. WORTHLEY, 1927 A progress report on the investigations of the European corn borer. U. S. Department of Agriculture Bulletin 1476: 155pp.

CIANCHI, R., S. MAINI and L. BULLINI, 1980 Genetic distance between pheromone strains of the European corn borer *Ostrinia nubilalis* different contribution of variable substrate regulatory and nonregulatory enzymes. Heredity 45: 383-388.

DOPMAN, E. B., S. M. BOGDANOWICZ and R. G. HARRISON, 2004 Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). Genetics 167: 301-309.

GLOVER, T., M. CAMPBELL, P. ROBBINS and W. ROELOFS, 1990 Sex-linked control of sex pheromone behavioral responses in European corn borer moths (*Ostrinia*

nubilalis) confirmed with TPI marker gene. Arch. Insect Biochem. Phys. 15: 67-77.

GLOVER, T. J., J. J. KNODEL, P. S. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1991 Gene flow among three races of European corn borers (Lepidoptera:Pyralidae) in New York State. Environ Entomol 20: 1356-1362.

GLOVER, T. J., P. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1992 Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. Arch. Insect Biochem. Phys. 20: 107-117.

HARRISON, R. G., 1990 Hybrid zones: Windows on evolutionary process, pp. 69-128 in Oxford Surveys in Evolutionary Biology, edited by D. FUTUYMA and J. ANTONOVICS. Oxford Univ. Press, Oxford, UK.

HARRISON, R. G., and A. T. VAWTER, 1977 Allozyme differentiation between pheromone strains of the European corn borer, *Ostrinia nubilalis*. Ann. Entomol. Soc. Am. 70: 717-720.

HUDSON, R. R., and J. A. COYNE, 2002 Mathematical consequences of the genealogical species concept. Evolution 56: 1557-1565.

KLIMAN, R. M., P. ANDOLFATTO, J. A. COYNE, F. DEPAULIS, M. KREITMAN *et al.*, 2000 The Population Genetics of the Origin and Divergence of the Drosophila simulans Complex Species. Genetics 156: 1913-1888.

KLUN, J. A., O. L. CHAPMAN, J. C. MATTES, P. W. WOJTOWSKI, M. BEROZA *et al.*, 1973 Insect sex pheromones: Minor amount of opposite geometrical isomer critical to attraction. Science 181: 661-663.

KLUN, J. A., and M. D. HUETTEL, 1988 Genetic regulation of sex pheromone production and response: Interaction of sympatric pheromonal types of European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). J. Chem. Ecol. 14: 2047-2061.

KLUN, J. A., and S. MAINI, 1979 Genetic basis of an insect chemical communication system: The European corn borer. Environ. Entomol. 8: 423-426.

LIEBHERR, J., 1974 Studies on two strains of European corn borer, *Ostrinia nubilalis* (Hübner), pp. 1-70 in Masters Thesis. thesis, Cornell University, Ithaca, NY.

LIEBHERR, J., and W. L. ROELOFS, 1975 Laboratory hybridization and mating period studies using two pheromone strains of *Ostrinia nubilalis*. Ann. Entomol. Soc. Am. 68: 305-309.

MACHADO, C. A., and J. HEY, 2003 The causes of phylogenetic conflict in a classic Drosophila species group. P. Roy. Soc. Lond. B. Bio. 270: 1193-1202.

MARCON, P., D. B. TAYLOR, C. E. MASON, R. L. HELLMICH and B. D. SIEGFRIED, 1999 Genetic similarity among pheromone and voltinism races of *Ostrinia nubilalis* (Hubner) (Lepidoptera: Crambidae). Insect Mol. Biol. 8: 213-221.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

NEIGEL, J. E., and J. C. AVISE, 1986 Phylogenetic Relationships Of Mitochondrial Dna Under Various Demographic Models Of Speciation. Evolutionary Processes and Theory; Workshop 1985: 515-534.

NOOR, M. A., K. L. GRAMS, L. A. BERTUCCI, Y. ALMENDAREZ, J. REILAND *et al.*, 2001 The genetics of reproductive isolation and the potential for gene exchange between Drosophila pseudoobscura and D. persimilis via backcross hybrid males. Evolution 55: 512-521.

PORNKULWAT, S., S. R. SKODA, G. D. THOMAS and J. E. FOSTER, 1998 Random amplified polymorphic DNA used to identify genetic variation in ecotypes of the European corn borer (Lepidoptera: Pyralidae). Ann. Entomol. Soc. Am 91: 719-725.

POSADA, D., and D. A. CRANDALL, 1998 Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817-818.

RIESEBERG, L. H., J. WHITTON and K. GARDNER, 1999 Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152: 713-727.

ROELOFS, W., T. GLOVER, X. H. TANG, I. SRENG, P. ROBBINS *et al.*, 1987 Sex-pheromone production and perception in European Corn Borer moths is determined by both autosomal and sex-linked genes. Proc. Natl. Acad. Sci. U.S.A. 84: 7585-7589.

ROELOFS, W. L., J.-W. DU, X.-H. TANG, P. ROBBINS and C. J. ECKENRODE, 1985 Three European corn borer populations in New York based on sex pheromones and voltinism. J. Chem. Ecol. 11: 829-936.

SHIMODAIRA, H., and M. HASEGAWA, 1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Molecular Biology and Evolution 16: 1114-1116.

SORENSON, C. E., G. C. KENNEDY, W. VAN DUYN, J. R. BRADLEY, JR. and J. F. WALGENBACH, 1992 Geographical variation in pheromone response of the European corn borer, *Ostrinia nubilalis*, in North Carolina. Entomol. exp. appl. 64: 177-185.

SWOFFORD, D. L., 2000 PAUP*. Phylogenetic analysis using parisomy (*and other methods). pp. Sinauer, Sunderland, MA.

TAJIMA, F., 1983 Evolutionary Relationship of DNA-Sequences in Finite Populations. Genetics 105: 437-460.

TING, C. T., S. C. TSAUR and C. I. WU, 2000 The phylogeny of closely related species as revealed by the genealogy of a speciation gene, Odysseus. Proc. Natl. Acad. Sci. U.S.A. 97: 5313-5316.

WANG, R. L., J. WAKELEY and J. HEY, 1997 Gene flow and natural selection in the origin of Drosophila pseudoobscura and close relatives. Genetics 147: 1091-1106.

WILLETT, C. S., and R. G. HARRISON, 1999 Insights into genome differentiation: Pheromone-binding protein variation and population history in the European corn borer (*Ostrinia nubilalis*). Genetics 153: 1743-1751.

WU, C., 2001 The genic view of the process of speciation. J. Evol. Biol 14: 851-865.

CHAPTER 4

USING NATURAL HISTORY TO DISTINGUISH BETWEEN NEUTRAL AND NON-NEUTRAL MOLECULAR GENETIC EVOLUTION: A CASE STUDY USING INCIPIENT SPECIES

ABSTRACT

Genetic differences are used to reconstruct the evolutionary history of species, to characterize recent demographic processes, and to identify molecular evolutionary forces. Unfortunately, the evolution of species, populations, and genes all contribute to genetic differentiation, and it may often be impossible to disentangle the effects at any single scale in order to discern a correct interpretation at another. However, for those systems in which abundant natural history information is available, it may be possible to assess the impact and contribution of known demographic processes on genetic variation. In this manner, the potentially overwhelming array of explanations for the evolution of genetic variation can be limited, or at the very least, bounded, by ecologically plausible scenarios. We employ this method to distinguish between neutral and non-neutral evolution in a multi-locus data set for populations in the incipient stages of speciation. We use our knowledge of the demographic history for sexually isolated pheromone strains of European corn borer moth (ECB) to investigate the contribution of reproductive isolation on patterns of genetic variation. We find that neutral demography involving population expansion and population substructure can explain most genetic patterns; however, one locus shows evidence of non-neutral evolution. Variation at *Tpi* is consistent with recent selection and the genealogy for the locus reveals Z and E pheromone strains of ECB as exclusive genetic groups, but only

in North America. *Tpi* is tightly linked to *Pdd*, a divergent trait that causes reproductive isolation between North American populations, suggesting that recent evolution of this trait has resulted in the evolution of genomic differences between these incipient species. Our work indicates that integrating natural history into evolutionary genetic analyses brings additional confidence to interpretations of genetic differences between species, populations, and molecules.

INTRODUCTION

Speciation describes the evolution of reproductive isolating barriers that limit gene flow between diverged populations (Dobzhansky 1937). A consequence of isolating barriers is fragmentation of populations that may eventually form monophyletic or exclusive genetic groups for many genomic regions. However, for closely related species much of the genome may be shared because of recent isolation and/or gene flow. As a result, most genomic regions will not reveal the demographic process of speciation - the population history of reproductive isolation (population divergence), as indicated by subsequent genetic divergence leading to genetic "exclusivity." In these situations, gene regions that show genetic divergence by grouping closely related species into discrete clusters can have important implications for understanding the history of population divergence.

If diverged species come into secondary contact and have not evolved complete reproductive isolation, then hybridization and gene exchange may occur. Gene flow will eventually homogenize any genetic divergence that has been attained prior to secondary contact; however, divergence can persist at "speciation genes," or genes for reproductive barriers (BARTON and HEWITT 1981, 1985; HARRISON 1990; WU 2001; WU and TING 2004). For hybridizing species, speciation genes interact to

cause hybrid incompatibilities (e.g., PRESGRAVES *et al.* 2003; e.g., WITTBRODT *et al.* 1989). Because the effect of a speciation gene is to reduce hybrid fitness, gene flow will be limited at or near these loci. At gene regions that are neutral in both species, differentiation will ultimately be eroded by gene flow. The rate of erosion for these neutral regions is a function of the rate of introgression between diverged species and the rate of recombination with any linked speciation gene. A semi-permeable genome between species is expected under these conditions, with permeability depending on linkage to speciation genes.

In addition to the important role of selection in maintaining locus-specific genetic divergence, selection may be the primary force responsible for the origin of genetic differentiation between closely related species by driving the evolution of reproductive barriers. Data from field studies (e.g., Boughman 2001; Jiggins et al. 2001) and from molecular analyses of genes involved in reproductive isolation (e.g., BARBASH *et al.* 2004; COLOSIMO *et al.* 2005; e.g., PRESGRAVES *et al.* 2003) have corroborated early inference of the importance of selection in speciation (Darwin 1859). Thus, divergence at speciation genes may often arise by natural or sexual selection, evolving either as incidental byproducts of local adaptation to different environments (e.g., Boughman 2001; Jiggins et al. 2001) or as a consequence of the reinforcement of reproductive isolation (Noor 1999). Selection on species-specific adaptations or speciation genes has a clear impact on population divergence of species, and theoretical predictions can be made for the locus-specific effect of selection on speciation genes.

Neutrality may characterize most genetic variation (Kimura 1983), but linkage to advantageous mutations can have predictable effects on neutral genetic patterns (but see Gillespie 2000). Directional selection on advantageous mutations can sweep tightly linked neutral alleles to fixation, thereby eliminating other neutral variants

(Maynard Smith and Haigh 1974). Among the patterns predicted by these selective sweeps is an excess of low frequency mutations relative to neutral expectations (BRAVERMAN *et al.* 1995; KAPLAN *et al.* 1989). The "genetic hitchhiking" of neutral variants with selected mutations also influences the genealogical structure of genetic variation between species. Directional selection on genes for adaptations and reproductive barriers can quickly generate a genealogical pattern of genetic divergence that matches the history of reproductive isolation between daughter lineages. By eliminating linked neutral genetic variation, selective sweeps purge shared ancestral polymorphism and reveal population divergence between reproductively isolated groups.

Through the maintenance and/or origin of genetic divergence at genes for reproductive barriers and adaptations, selection can promote locus-specific genetic patterns that reveal the history of reduced gene flow and reproductive isolation between species. However, the reproductive isolation of populations creates genome-wide patterns that resemble those stemming from selection. Specifically, populations exhibiting substructure due to reproductive isolation will randomly produce neutral loci showing genetic exclusivity (Takahata and Slatkin 1990; Wakeley 2000) as well as hallmarks of selection (Nielsen 2001; Przeworski 2002). Population substructure can therefore spuriously generate patterns consistent with the presence of a speciation gene or a gene for adaptation. Other neutral demographic processes, such as population expansion, can have similar stochastic, genome-wide effects that can be misleading. Thus, in general, neutral demographic explanations involving stochastic genome-wide effects must be explored before convincing arguments for selection can be accepted. The crucial issue for closely related species is whether a single neutral demographic scenario involving substructure (or other demographic processes) can explain patterns of variation across all loci examined. This is an important issue if

genetic patterns are used to infer the presence of genes for reproductive barriers or adaptations, both of which are critical in the diversification of species.

Here we use ecological information from reproductively isolated pheromone strains of European corn borer (ECB) to make a link between genes and ecology. We determine that patterns of genetic variation for a multi-locus data set for ECB are consistent with the action of both genome-wide stochastic effects and locus-specific deterministic effects. Specifically, neutral demographic processes involving population substructure and population expansion can account for much of the observed variation among genetic markers, but one locus appears to have been influenced by selection. Genealogical patterns of descent at this locus, the sex-linked gene *Tpi*, reveal strains as exclusive genetic groups, whereas strains share genetic variation at the other markers (Dopman et al. submitted). We suggest that ecological adaptation is responsible for the origin and possibly the maintenance of this locus-specific pattern because the major genetic factor explaining differences in development time (*Pdd*) is tightly linked to *Tpi* (Dopman et al. submitted). Differences in development time incidentally generate temporal isolation; therefore, this reproductive barrier may represent a byproduct of adaptation to different environments. This system allows for an important connection to be made between selection, reproductive isolation, and the origin of genetic exclusivity, an attribute that ultimately characterizes species as discrete, unique entities.

Background on the European Corn Borer

The European corn borer is native to Europe, North Africa, and Western Asia (Mutuura and Monroe 1970). Early in the 20th century, ECB was introduced on multiple occasions to North America in shipments of broom corn from Italy and

Hungary (CAFFREY and WORTHLEY 1927; SMITH 1920). Since its introduction, ECB has quickly spread throughout the United States and Canada east of the Rocky Mountains.

Across most of its range, ECB consists of two behaviorally isolated pheromone strains that differ in the sex pheromone produced by females and the response elicited in males (e.g., Klun and Cooperators 1975). In the Z strain, females produce and males respond to a 3:97 mixture of E and Z-Δ11-14:OAc, whereas in the E strain females produce and males respond to a 99:1 (E)/(Z) blend (Klun et al. 1973). Partial sexual isolation stems from differences in male response to these alternative pheromone blends (GLOVER *et al.* 1990; ROELOFS *et al.* 1987). Cross attraction rarely occurs in the laboratory (Glover et al. 1990; Linn et al. 1997). In nature, hybrid genotypes exist in low frequency (GLOVER *et al.* 1991; KLUN and HUETTEL 1988).

The genetic factors responsible for major differences in female pheromone blend production, *Pher*, and male behavioral response to those blends, *Resp*, exhibit simple Mendelian inheritance (DOPMAN *et al.* 2004; GLOVER *et al.* 1990; KLUN and MAINI 1979; ROELOFS *et al.* 1987). *Pher* is autosomal, whereas *Resp* is on the Z chromosome and is sex linked. Like other Lepidoptera, female ECB are the heterogametic sex.

In addition to differences in sexual communication, ECB also exhibits variation in life cycle. In New York State, populations are either bivoltine with an early first generation and a late second generation, or univoltine with a single generation in the middle of the season (Roelofs et al. 1985). In southern regions of North America, ECB can exhibit a “multi-voltine” life cycle with up to four generations per season. Based on voltinism and sex pheromone blend, New York populations consist of three distinct races: univoltine Z (UZ), bivoltine E (BE), and bivoltine Z (BZ) (GLOVER *et al.* 1991; KLUN and HUETTEL 1988). Adults from

univoltine populations are partially temporally isolated from those of the bivoltine populations (ECKENRODE *et al.* 1983; ROELOFS *et al.* 1985).

Bivoltine and univoltine life cycle patterns in New York reflect genetically determined differences in post-diapause development time (PDD), the time to pupation for over-wintering larvae under temperature and photoperiod conditions conducive to breaking diapause (Glover et al. 1992). Less is known about the determinants for multi-voltinism, but a stronger environmental component seems likely. The genetic factor responsible for major differences in development time, *Pdd*, is controlled by a simple Mendelian factor that shows sex linkage (Glover et al. 1992; Dopman et al. submitted). *Pdd* maps to a position on the Z chromosome that is > 20 cM away from *Resp* (Dopman et al. submitted), in a region of the sex chromosome that shows reduced levels of recombination rate (DOPMAN *et al.* 2004).

Beyond the documented differences in pheromone communication and life cycle, strains or races of ECB in North America are difficult to distinguish (Liebherr 1974; Liebherr and Roelofs 1975). Multiple genetic surveys have revealed extensive shared genetic variation, but little differentiation (CIANCHI *et al.* 1980; HARRISON and VAWTER 1977; MARCON *et al.* 1999; PORNKULWAT *et al.* 1998; WILLETT and HARRISON 1999). However, significant allele-frequency differences occur at a sex-linked gene that encodes the enzyme triose phosphate isomerase (TPI). At this locus, BE populations in New York are fixed for the *Tpi-1* allele and UZ and BZ populations are segregating for both *Tpi-1* and *Tpi-2*, with *Tpi-2* being the more common allele (Glover et al. 1991). The gene genealogy for *Tpi* reveals North American pheromone strains as exclusive genetic groups, independent of life cycle differences, while at four other loci strains share identical DNA sequences (Dopman et al. submitted). *Pdd* maps to a position on the Z chromosome that is indistinguishable from *Tpi*, suggesting that the genetic divergence at *Tpi* may have been affected by tight linkage to this gene.

MATERIALS AND METHODS

Field-Collected Insects and Sequence Data

Sympatric populations of Z and E borers have been documented at a number of localities in New York (GLOVER *et al.* 1991; ROELOFS *et al.* 1985). The Geneva, NY (Gen) population has been monitored for ECB and consists of BE and UZ moths (e.g., Roelofs et al. 1985). Thus, female larvae and pupae collected from corn stubble in May of 2000 and 2004 that were reared under diapause-breaking conditions to assess post-diapause development (PDD) time could be scored as BE or UZ. Fourteen Gen BE and six Gen UZ females were used for genetic analysis from this locality (Table 4.1). Eight females were collected for analysis from a UZ population in Madison, NY in May 2000 (Mad UZ), and four females were collected from a BZ population in Eden, NY in October 2000 (Eden BZ) (Table 4.1).

Both pheromone strains also occur in North Carolina, with Z borers found in the western part of the state and E borers more common in the east. A zone of overlap exists where males are found in pheromone traps baited with either the Z or E blends (Sorenson et al. 1992). For population genetic analysis we used three female ECB collected in May 2002 from potato plants in Weeksville, Pasquotank Co., N.C. (E strain) and two females collected from corn in Fletcher, Henderson Co., N.C. (Z strain) (Table 4.1).

We collected new sequences from Italy and Hungary ECB because North American insects are believed to be derived from these regions (Caffrey and Worthley1927). Female moths collected from maize in Europe were used for genetic analysis. We used five E-strain insects from Piacenza, Italy (Ita E) in September 2002, and we used five Z-strain females from Kety, Hungary (Hun Z) that were part of a

Table 4.1. Sample information.

Sample Name	Locality	Strain	Voltinism
Gen BE g1	Geneva, NY, USA	E	Bivoltine
Gen BE g2	Geneva, NY, USA	E	Bivoltine
Gen BE g3	Geneva, NY, USA	E	Bivoltine
Gen BE g5	Geneva, NY, USA	E	Bivoltine
Gen BE g6	Geneva, NY, USA	E	Bivoltine
Gen BE 7	Geneva, NY, USA	E	Bivoltine
Gen BE 8	Geneva, NY, USA	E	Bivoltine
Gen BE 9	Geneva, NY, USA	E	Bivoltine
Gen BE 15	Geneva, NY, USA	E	Bivoltine
Gen BE 16	Geneva, NY, USA	E	Bivoltine
Gen BE 17	Geneva, NY, USA	E	Bivoltine
Gen BE 18	Geneva, NY, USA	E	Bivoltine
Gen BE 19	Geneva, NY, USA	E	Bivoltine
Gen BE 20	Geneva, NY, USA	E	Bivoltine
NC E 17	Weeksville, NC, USA	E	Unknown
NC E 18	Weeksville, NC, USA	E	Unknown
NC E 19	Weeksville, NC, USA	E	Unknown
Ita E n1	Piacenza, Italy	E	Unknown
Ita E n12	Piacenza, Italy	E	Unknown
Ita E n15	Piacenza, Italy	E	Unknown
Ita E n34	Piacenza, Italy	E	Unknown
Ita E n45	Piacenza, Italy	E	Unknown
Gen UZ 83	Geneva, NY, USA	Z	Univoltine
Gen UZ 127	Geneva, NY, USA	Z	Univoltine
Gen UZ 138	Geneva, NY, USA	Z	Univoltine
Gen UZ 136	Geneva, NY, USA	Z	Univoltine
Gen UZ 69 T2	Geneva, NY, USA	Z	Univoltine
Gen UZ 44 T2	Geneva, NY, USA	Z	Univoltine
Mad UZ 128	Madison, NY, USA	Z	Univoltine
Mad UZ 155 T2	Madison, NY, USA	Z	Univoltine
Mad UZ 125 T2	Madison, NY, USA	Z	Univoltine
Mad UZ 84 T2	Madison, NY, USA	Z	Univoltine
Mad UZ 46 T2	Madison, NY, USA	Z	Univoltine
Mad UZ 124 T2	Madison, NY, USA	Z	Univoltine
Mad UZ 49 T2	Madison, NY, USA	Z	Univoltine
Mad UZ 137 T2	Madison, NY, USA	Z	Univoltine
Eden BZ 278	Eden, NY, USA	Z	Bivoltine
Eden BZ 282 T2	Eden, NY, USA	Z	Bivoltine
Eden BZ 279 T2	Eden, NY, USA	Z	Bivoltine
Eden BZ 280 T2	Eden, NY, USA	Z	Bivoltine
NC Z 4	Fletcher, NC, USA	Z	Unknown

Table 4.1 (Continued)

NC Z 5	Fletcher, NC, USA	Z	Unknown
Hun Z 1	Kety, Hungary	Z	Unknown
Hun Z 9	Kety, Hungary	Z	Unknown
Hun Z 8	Kety, Hungary	Z	Unknown
Hun Z 2	Kety, Hungary	Z	Unknown
Hun Z 10	Kety, Hungary	Z	Unknown
ACB	Hengshui Co., China	N/A	N/A

colony started in August 2001 (Table 4.1). An Asian corn borer (ACB) female (*Ostrinia furnicalis*) was used as an outgroup species.

Alleles from three sex-linked genes, *Triose phosphate isomerase* (*Tpi*), *Kettin* (*Ket*), and *Lactate dehydrogenase* (*Ldh*), an autosomal locus, *Pheromone binding protein* (*Pbp*), and mitochondrial *cytochrome oxidase I* (*COI*) were obtained from 17 E-strain and 20 Z-strain ECB females from North America following the protocol described in Dopman et al. (submitted). New sequences were collected from the ten European moths for *Tpi*, *Ldh*, and *COI* using oligonucleotide primers, PCR conditions, and sequencing protocols described by Dopman et al. (submitted).

Statistical Analyses

Alignment gaps were excluded from all statistical analyses. For polymorphic sites, we assumed that the ancestral state was that possessed by ACB and that the probability of back mutation was negligible. Summary statistics were calculated using DnaSP (v. 4.0) (Rozas et al. 2003) and included estimators of polymorphism, Θ (Watterson 1975) and π (Nei 1987); an estimate of divergence, K (Nei 1987); summaries of the allele frequency spectrum, Tajima's D (Tajima 1989b), Fu and Li's D^* (Fu 1993), Fu and Li's F^* (Fu 1993), and Fay and Wu's H (Fay and Wu 2000); a measure of linkage disequilibrium, Z_nS (Kelly 1997); haplotype diversity, H_d (Depaulis and Veuille 1998); and recombination, C (Hudson 1987). Significance was determined through 10,000 standard-neutral coalescent simulations under the conservative assumption of no recombination (but see below).

Multi-locus comparisons of polymorphism and divergence (HUDSON *et al.* 1987) compared ECB to ACB and were performed using the program HKA (distributed by Jody Hey at http://lifesci.rutgers.edu/heylab). Ten thousand standard-

neutral coalescent simulations were performed to assess significance. Several hierarchical analyses evaluated population genetic structure and were performed using Arlequin (v. 2.001) (SCHNEIDER *et al.* 2000). Significance was determined by permutation (10,000 repetitions).

We constructed gene genealogies for each marker using Neighbor-Joining and Tamura-Nei distances (Tamura and Nei 1993) in PAUP* (v. 4.0b10) (Swofford 2000). For North American ECB, we also constructed gene trees compatible with a constraint enforcing strain exclusivity [i.e., ((E), (Z))]. The difference in tree length score, as assessed under parsimony, between the unconstrained tree and the tree enforcing exclusivity represents a new statistic describing the disagreement of a marker with exclusivity of Z and E pheromone strains. Tree-length difference (*TLD*) ranges between zero (a tree compatible with exclusivity) and the maximum number of polymorphic characters for a locus.

Demographic Modeling

We determined the statistical significance of three summary statistics using coalescent simulation under non-equilibrium demographic scenarios likely to have been experienced by ECB pheromone strains. A symmetric-island model of migration was examined using simulated data sets generated from "ms" (Hudson 2002). Tajima's *D* and Fay and Wu's *H* were chosen as summary statistics for the allele frequency spectrum and *TLD* summarized information about genealogical exclusivity between Z and E pheromone strains. *D* and *H* were calculated from the simulated data sets with a program that uses "libsequence" (Thornton 2003), and genealogical reconstruction was performed in PAUP*. Simulations were based on the observed number of segregating sites at each locus, the number of sequences from each strain

(for North American ECB, 17 E strain and 20 Z strain), and the empirically derived per-gene recombination rate, *C*. For each locus, 10,000 coalescent simulations were generated under six population migration rates, $M = 4N_0m$ (where N_0 is the effective population size at time, $t = 0$, and m is the fraction of each subpopulation consisting of migrant individuals per generation). These rates were 0.01, 0.1, 0.5, 1.0, 1.5, 4, and 20. A critical threshold value of 0.05 was used to determine significance.

Population size change was explored in addition to population substructure. The current effective population size, N_0, was changed to $x^* N_0$ at time t^*4N_0 generations ago. The extent and time of population size change was chosen in part from information on the recent population history of ECB. Population migration rates explored under these models varied from M = 0.1 to M = 100.

RESULTS

Patterns of Genetic Structure

Dopman et al. (submitted) evaluated gene genealogies for North American ECB and found that four genetic markers did not reveal a genealogical pattern of exclusivity between Z and E pheromone strains. In contrast, *Tpi* depicted Z and E strains as exclusive genetic groups. Here, coalescent simulations under the standard neutral model indicate that *Tpi* is the only locus that has a *TLD* that differs significantly from neutral expectations (Figure 4.1, Table 4.2). With panmixia assumed by the standard neutral model, the high genetic divergence between strains and low *TLD* at *Tpi* has a low probability ($P = 0.0002$), whereas the lack of genetic divergence and positive *TLD* for the other markers has a high probability (i.e., $P \geq 0.5$). The unconstrained gene genealogy for *Tpi* was perfectly compatible with a

Figure 4.1. Gene genealogy for *Tpi*. Different symbols represent moths from different populations. Unshaded symbols are E-strain moths and shaded symbols are Z-strain moths. The tree is rooted with ACB.

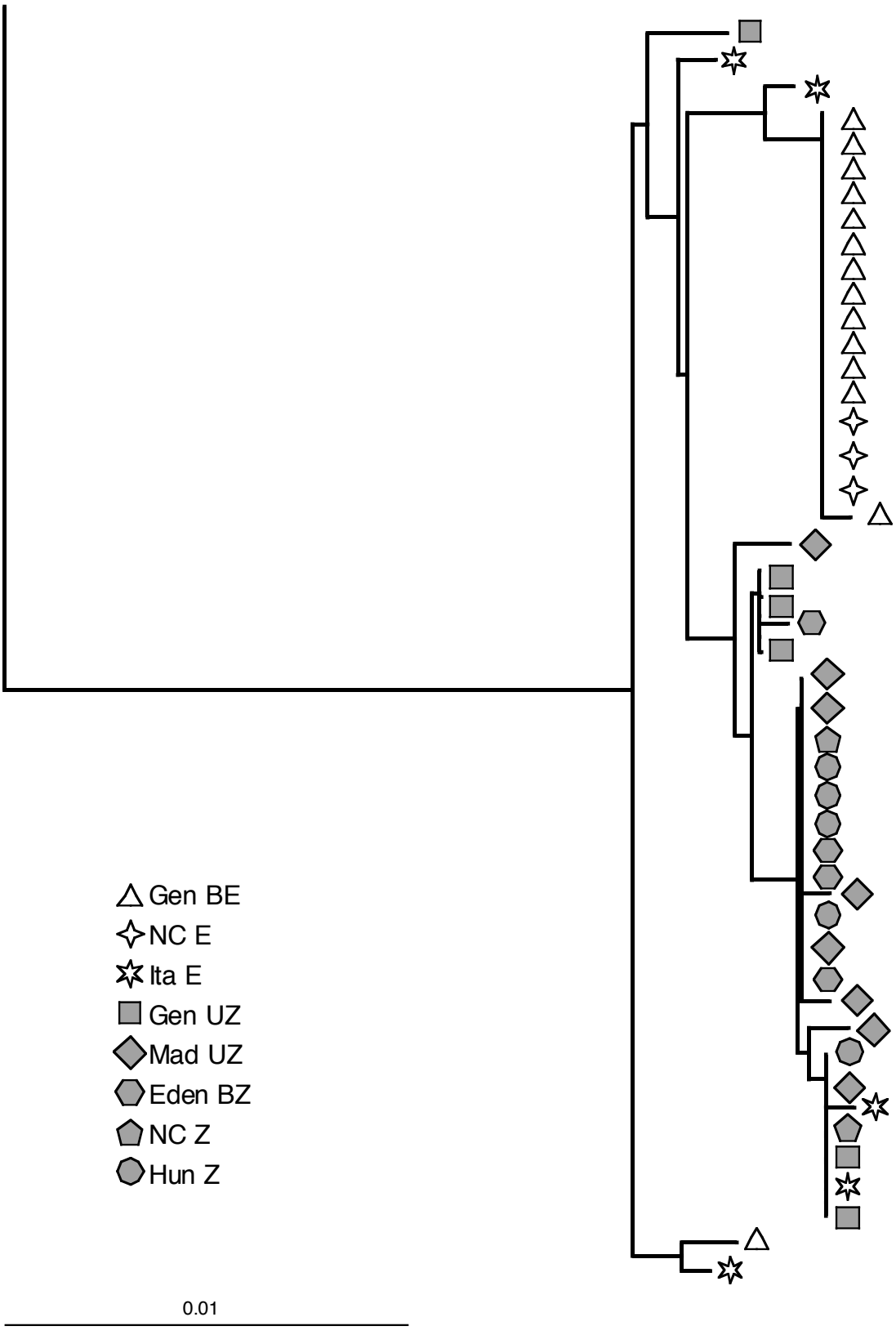

Gen BE
NC E
Ita E
Gen UZ
Mad UZ
Eden BZ
NC Z
Hun Z
0.01

Table 4.2. Tree-length difference (*TLD*) between the unconstrained and constrained (exclusive) NJ gene genealogies

Locus	TLD	*P*-value [a]
COI	4	0.936
Ket	15	0.573
Ldh	6	0.779
Pbp	22	0.837
Tpi	0	**2.00E-04**

[a] Probability of observing a tree-length difference (*TLD*) that is equal to or more extreme than the observed *TLD* between the unconstrained tree and the tree enforcing exclusivity between pheromone strains, ((E), (Z)). For *Tpi,* probability of observing *TLD* of 0. Results are based on 10,000 coalescent simulations under the conservative assumption of no recombination.

reciprocally monophyletic tree (*TLD* = 0), whereas additional steps were required to generate a monophyletic tree for the other genetic markers.

Hierarchical analyses of genetic variation (AMOVA) confirm genealogical patterns of variation between North American pheromone strains. Most covariance components show no significant genetic structure, regardless of hierarchy (where V_A = among strain; V_B = among populations within strain; V_C = within populations; Table 4.3); however, *COI*, *Ket*, *Ldh*, and *Pbp* each show minimal differentiation between strains (*Ket*, *Ldh*, and *Pbp*: $V_A \leq 6\%$; $\Phi_{CT} \leq 0.06$; $P \geq 0.14$ and *COI*: $V_A = 10.28\%$; $\Phi_{CT} = 0.10$; $P = 0.07$). In contrast, *Tpi* harbors between strain variation that is close to eight times more than the other markers and indicates very high divergence (*Tpi*: $V_A = 79.03\%$; $\Phi_{CT} = 0.79$; $P = 0.07$). Only *Pbp* was genetically structured between populations within strains ($V_B = 16.6\%$; $\Phi_{SC} = 0.15$; $P = 0.02$) due to the four haplotypes from Eden BZ (removing Eden BZ makes this component no longer significant, $V_B = 9.05\%$; $\Phi_{SC} = 0.08$; $P = 0.14$). Finally, only *Tpi* exhibits significant and substantial genetic structure between populations relative to the entire sample ($V_C = 21.2\%$; $\Phi_{SC} = 0.79$; $P < 0.0001$).

Although the AMOVA analysis did not show significant strain genetic structure for *Tpi* in North American populations, population pairwise Φ_{ST} reveals significant and strong genetic differentiation in comparisons involving North American E insects (Table 4.4). Φ_{ST} is above 0.6 in all comparisons with North American Z, and remains high in comparisons with both European samples (0.39-0.98). Indeed, Φ_{ST} between the North American Geneva BE sample and the Italian E sample is almost as great as the values between Z and E strains ($\Phi_{ST} = 0.54$; $P = 0.0005$). In contrast to the strong genetic divergence exhibited between North American E and Italian E, ECB from Italy show low (and non-significant) genetic divergence with several Z-strain populations.

Table 4.3. AMOVA and hierarchical analyses of North American ECB pheromone strains

Source of variation (%)	*COI*	*Ket*	*Ldh*	*Pbp*	*Tpi*
Among strains (V_A)	10.48	5.69	1.72	-10.01	79.03
Among populations within strain (V_B)	-10.53	-5.09	1.35	**16.6**	-0.23
Within populations (V_C)	100.05	99.39	96.93	93.41	**21.2**
Fixation indices					
Φ_{CT} (strains/total)	0.105	0.057	0.017	-0.100	0.790
Φ_{SC} (populations/strain)	-0.118	-0.054	0.014	**0.151**	-0.011
Φ_{ST} (population/total)	0.000	0.006	0.031	0.066	**0.788**

Negative variance components and fixation indices indicate an absence of structure. Bold indicates significant at 0.05 level (10,000 permutations).

Table 4.4. *Tpi* population pairwise Φ_{ST}.

	Gen BE	NC E	Eden BZ	NC Z	Gen UZ	Mad UZ	Ita E	Hun Z
Gen BE	--							
NC E	-0.18	--						
Eden BZ	**0.85**	**0.91**	--					
NC Z	**0.86**	0.98	-0.10	--				
Gen UZ	**0.72**	**0.63**	-0.03	-0.17	--			
Mad UZ	**0.83**	**0.85**	-0.03	-0.30	0.06	--		
Ita E	**0.54**	0.39	0.21	0.04	0.04	**0.26**	--	
Hun Z	**0.88**	**0.98**	0.04	-0.21	0.09	-0.09	0.30	--

Negative fixation indices indicate an absence of structure. Bold indicates significant at 0.05 level (10,000 permutations). Gen BE = BE moths from Geneva, NY, NC E = E from Weeksville, NC, Eden BZ = BZ from Eden, NY, NC Z = Z from Fletcher, NC, Gen UZ = UZ from Geneva, NY, Mad UZ = UZ from Madison, NY, Ita E = E from Piacenza, Italy, Hun Z = Z from Kety, Hungary.

The *Tpi* gene genealogy (Figure 4.1) and a hierarchical analysis combining North American and European populations (Table 4.5) confirm that the E strain population from Italy differs from North American E in patterns of polymorphism. As shown in figure 1, some Italian E haplotypes (and all Hungarian Z haplotypes) fall into the Z-strain clade. Correspondingly, when Ita E are included in an AMOVA, a significant among population within strain component exists ($V_B = 8.82\%$; $\Phi_{SC} = 0.24$; $P = 0.004$), but when Italy E is removed, genetic divergence disappears ($V_A = -0.08\%$; $\Phi_{SC} = -0.004$; $P = 0.6$). Genealogies for *COI* and *Ldh* also show Italian E and Hungarian Z sequences falling throughout the tree (Figure 4.2 and Figure 4.3). *Ldh* shows no evidence of genetic differentiation between strains, whereas *COI* shows significant (but minimal) divergence (Table 4.5).

Standard Neutrality Tests

We calculated various summary statistics to examine whether patterns of genetic variation among loci were expected under the standard neutral model. Samples from North America and Europe were analyzed separately because *Tpi* only shows deviant patterns relative to other genetic markers for North American ECB (Table 4.2, 4.4, 4.5 and Figure 4.1) (Dopman et al. submitted). For each marker, North American haplotypes from the same strain were combined for calculations.

Table 4.6a and 4.6b present the summary statistics for North American E and Z pheromone strains. Under the standard neutral model with the conservative assumption of no recombination, *Tpi* within the E strain is the only locus that shows consistently significant values (Table 4.6a). E strain *Tpi* has an excess of low-frequency polymorphisms ($D = -2.235$), an excess of singletons and external-branch

Table 4.5. AMOVA and hierarchical analyses of North American and European ECB pheromone strains.

Source of variation (%)	*COI*	*Ldh*	*Tpi* with Italy E	*Tpi* without Italy E
Among strains (V_A)	**7.35**	1.13	**63.67**	**81.02**
Among populations within strain (V_B)	-2.6	-0.04	**8.82**	-0.08
Within populations (V_C)	95.25	98.92	**27.51**	**19.06**
Fixation indices				
Φ_{CT} (strains/total)	**0.073**	0.011	**0.637**	**0.810**
Φ_{SC} (populations/strain)	-0.028	0.000	**0.243**	-0.004
Φ_{ST} (population/total)	0.048	0.011	**0.725**	**0.809**

Negative variance components and fixation indices indicate an absence of structure. Bold indicates significant at 0.05 level (10,000 permutations).

Figure 4.2. Gene genealogy for *COI*. Different symbols represent moths from different populations. Unshaded symbols are E-strain moths and shaded symbols are Z-strain moths. The tree is rooted with ACB.

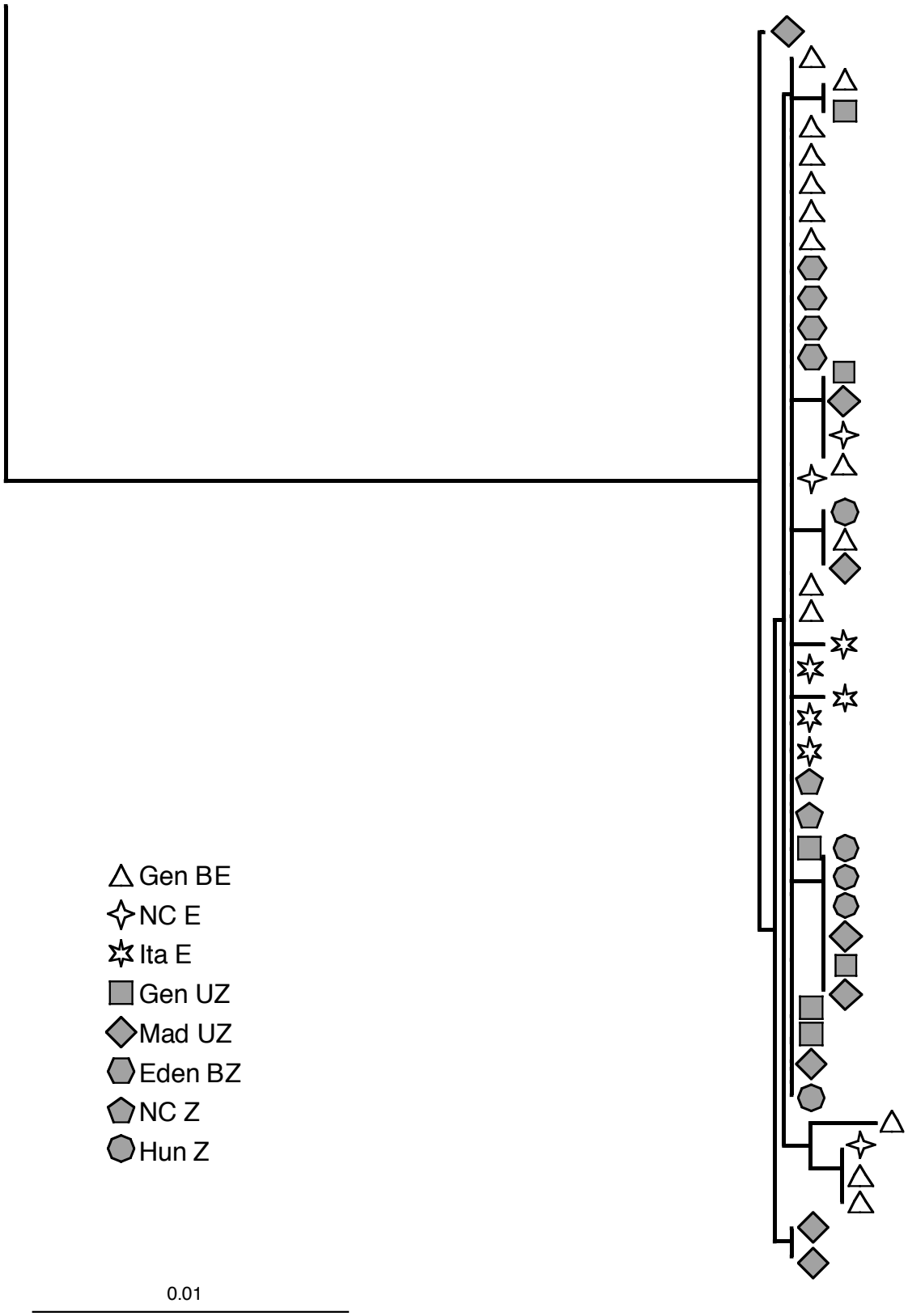

Gen BE
NC E
Ita E
Gen UZ
Mad UZ
Eden BZ
NC Z
Hun Z
0.01

Figure 4.3. Gene genealogy for *Ldh*. Different symbols represent moths from different populations. Unshaded symbols are E-strain moths and shaded symbols are Z-strain moths. The tree is rooted with ACB.

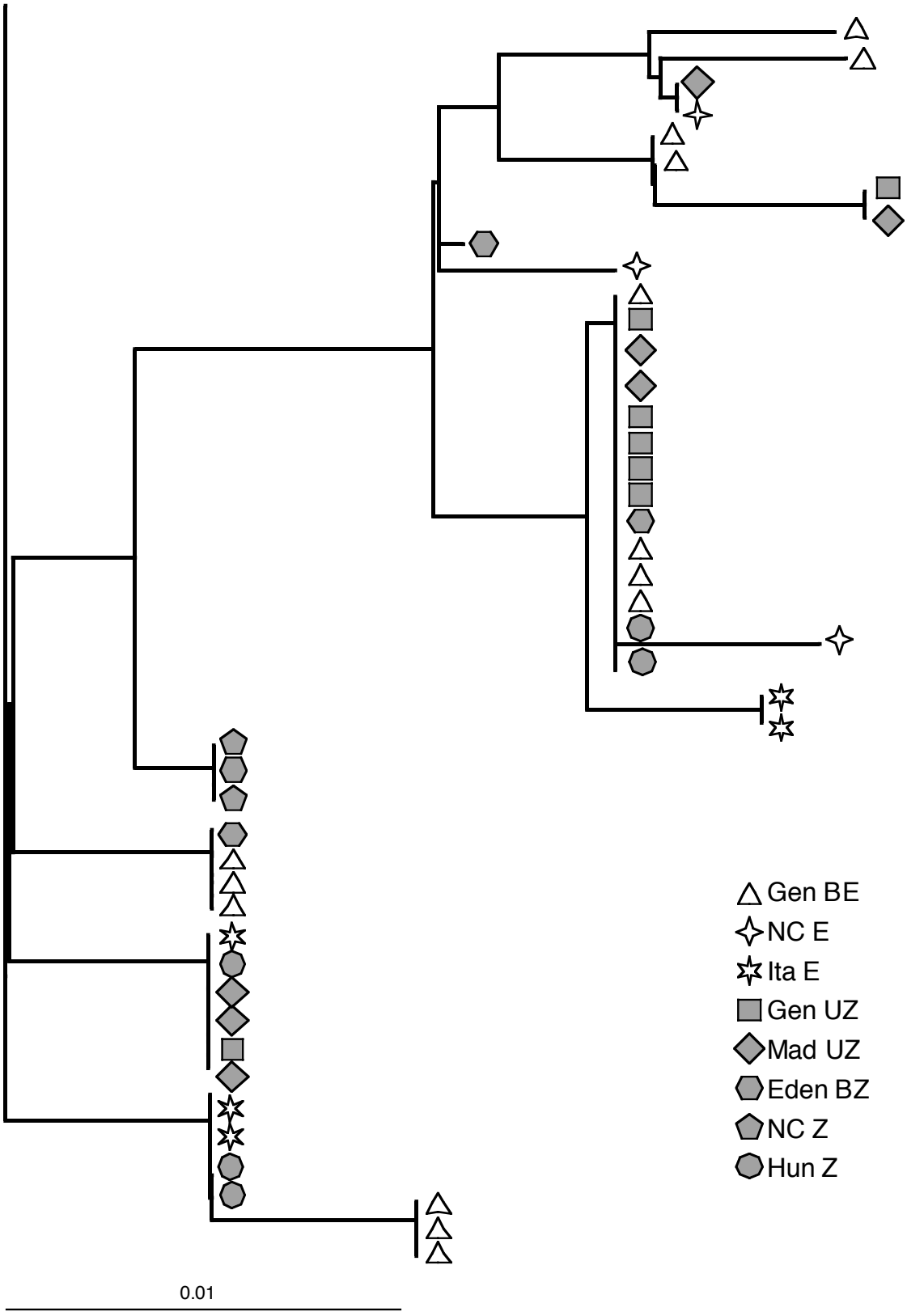

Gen BE
NC E
Ita E
Gen UZ
Mad UZ
Eden BZ
NC Z
Hun Z
0.01

Table 4.6a. Summary statistics for loci in North American E strain ECB.

Locus	N [a]	L [b]	S [c]	K [d]	Θ [e]	π [f]	D [g]	D^* [h]	F^* [i]	H [j]	Z_nS [k]	H_d [l]
COI	17	1195	8	0.023	0.0020	0.0013	-1.295	-1.204	-1.502	-0.125	0.085	0.765
Ket	17	704	17	0.011	0.0071	0.0069	-0.159	0.961	0.745	3.772	0.367	0.912
Ldh	17	192	10	0.014	0.0154	0.0170	0.380	0.514	0.549	0.904	0.140	0.904
Pbp	12	474	28	0.030	0.0203	0.0213	0.235	0.377	0.387	0.242	0.254	0.955
Tpi	17	1454	11	0.021	0.0022	0.0009	**-2.235**	**-3.138**	**-3.327**	**-11.912**	**0.819**	**0.228**

Bold values indicate significance at 0.05 level under the standard neutral model with no recombination (n = 10,000 simulations, two-sided).

[a] Sample size.

[b] Sequence length, not including alignment gaps.

[c] Number of polymorphic sites excluding indels.

[d] Average percent sequence divergence between ECB and ACB (Nei 1987).

[e] Heterozygosity, $4N_e\mu$, using the number of polymorphic sites (4Neμ for Pbp, 2Nefμ for mtDNA, and 3Neμ for Z-linked loci) (Watterson 1975).

[f] Heterozygosity, $4N_e\mu$, using the average number of nucleotide differences per site ($4N_e\mu$ for Pbp, $2N_ef\mu$ for mtDNA, and $3N_e\mu$ for Z-linked loci) (Nei 1987).

[g] Tajima's *D* (1989).

[h] Fu and Li's *D** (1993).

[i] Fu and Li's *F** (1993).

[j] Fay and Wu's *H* (2000).

[k] Linkage disequilibrium (Kelly 1997).

[l] Haplotype diversity (Nei 1987).

Table 4.6b. Summary statistics for loci in North American Z strain ECB.

Locus	N [a]	L [b]	S [c]	K [d]	Θ [e]	π [f]	D [g]	D^* [h]	F^* [i]	H [j]	Z_nS [k]	H_d [l]
COI	20	1195	7	0.023	0.0016	0.0009	-1.543	-1.204	-1.502	-2.642	**0.007**	0.784
Ket	20	703	21	0.011	0.0084	0.0055	-1.302	-1.986	-2.075	2.853	0.253	0.905
Ldh	20	192	8	0.012	0.0117	0.0127	0.288	0.211	0.270	0.158	0.152	0.800
Pbp	20	469	40	0.031	0.0246	0.0204	-0.681	-0.815	-0.903	-1.621	0.124	**0.989**
Tpi	20	1444	14	0.019	0.0027	0.0018	-1.220	-1.029	-1.258	**-6.737**	0.168	0.842

Bold values indicate significance at 0.05 level under the standard neutral model with no recombination (n = 10,000 simulations, two-sided).

[a] Sample size.

[b] Sequence length, not including alignment gaps.

[c] Number of polymorphic sites excluding indels.

[d] Average percent sequence divergence between ECB and ACB (Nei 1987).

[e] Heterozygosity, $4N_e\mu$, using the number of polymorphic sites ($4Ne\mu$ for Pbp, $2Nef\mu$ for mtDNA, and $3Ne\mu$ for Z-linked loci) (Watterson 1975).

[f] Heterozygosity, $4N_e\mu$, using the average number of nucleotide differences per site ($4N_e\mu$ for Pbp, $2N_ef\mu$ for mtDNA, and $3N_e\mu$ for Z-linked loci) (Nei 1987).

[g] Tajima's D (1989).

[h] Fu and Li's D^* (1993).

[i] Fu and Li's F^* (1993).

[j] Fay and Wu's H (2000).

[k] Linkage disequilibrium (Kelly 1997).

[l] Haplotype diversity (Nei 1987).

polymorphisms (D^* = -3.138 and F^* = -3.327), and an excess of high frequency derived polymorphisms (H = -11.912). In addition to a negative skew in the site frequency spectrum, *Tpi* exhibits significant linkage disequilibrium (Z_nS = 0.819) and significantly reduced haplotype diversity (H_d = 0.228). *COI* displays a skew in the allele frequency spectrum in the same direction as *Tpi*, but the values are within the 95% CI under the neutral model. The statistics for the other loci do not appear atypical for neutral expectations.

A more complex pattern of variability exists for North American Z strain ECB, but only two statistics are significant following coalescent simulations (Table 4.6b). Haplotype diversity for *Pbp* is greater than neutral expectations (H_d = 0.989), and H for *Tpi* indicates an excess of derived variants at high frequency (H = -6.737). Although these are the only significant statistics, a feature of the Z strain data deserves further comment; three of five Z strain loci show a marked negative skew in the site frequency spectrum (e.g., $D \leq$ -1.20), consistent with recent population expansion. Finally, summary statistics for *COI*, *Ldh*, and *Tpi* for Italian (E strain) and Hungarian (Z strain) ECB all are within expectations of the neutral model (Table 4.7a, 4.7b).

Polymorphism and divergence levels across loci were not significantly different from neutral expectations. HKA tests that compared polymorphism when pheromone strains were combined (North America E/Z and Europe E/Z) and separate (North America E versus North America Z) did not reveal a significant pattern (combined: χ^2 = 4.06, P = 0.19; North America E strain: χ^2 = 7.17, P = 0.07; North America Z strain: χ^2 = 7.1, P = 0.08). McDonald-Kreitman tests could not be performed because of a lack of sufficient variation.

Table 4.7a. Summary statistics for loci in Italian E strain ECB.

Locus	N [a]	L [b]	S [c]	K [d]	Θ [e]	π [f]	D [g]	D^* [h]	F^* [i]	H [j]	Z_nS [k]	H_d [l]
COI	5	1077	2	0.024	0.0009	0.0007	-0.973	-0.973	-0.954	0.600	0.063	0.700
Ldh	5	196	6	0.011	0.0147	0.0174	1.241	1.241	1.286	1.300	0.574	0.800
Tpi	5	1452	15	0.018	0.0050	0.0052	0.406	0.000	0.087	0.800	0.307	1.000

Bold values indicate significance at 0.05 level under the standard neutral model with no recombination (n = 10,000 simulations, two-sided).

[a] Sample size.

[b] Sequence length, not including alignment gaps.

[c] Number of polymorphic sites excluding indels.

[d] Average percent sequence divergence between ECB and ACB (Nei 1987).

[e] Heterozygosity, $4N_e\mu$, using the number of polymorphic sites ($4Ne\mu$ for Pbp, $2Nef\mu$ for mtDNA, and $3Ne\mu$ for Z-linked loci) (Watterson 1975).

[f] Heterozygosity, $4N_e\mu$, using the average number of nucleotide differences per site ($4N_e\mu$ for Pbp, $2N_ef\mu$ for mtDNA, and $3N_e\mu$ for Z-linked loci) (Nei 1987).

[g] Tajima's *D* (1989).

[h] Fu and Li's D^* (1993).

[i] Fu and Li's F^* (1993).

[j] Fay and Wu's *H* (2000).

[k] Linkage disequilibrium (Kelly 1997).

[l] Haplotype diversity (Nei 1987).

Table 4.7b. Summary statistics for loci in Hungarian Z strain ECB.

Locus	N [a]	L [b]	S [c]	K [d]	Θ [e]	π [f]	D [g]	D^* [h]	F^* [i]	H [j]	Z_nS [k]	H_d [l]
COI	5	1077	2	0.024	0.0009	0.0009	0.241	0.241	0.239	0.000	0.375	0.700
Ldh	5	196	5	0.009	0.0122	0.0143	1.124	1.124	1.156	1.100	0.500	0.800
Tpi	5	1452	1	0.019	0.0003	0.0003	-0.817	-0.817	-0.772	-1.200	-	0.400

Bold values indicate significance at 0.05 level under the standard neutral model with no recombination (n = 10,000 simulations, two-sided).

[a] Sample size.

[b] Sequence length, not including alignment gaps.

[c] Number of polymorphic sites excluding indels.

[d] Average percent sequence divergence between ECB and ACB (Nei 1987).

[e] Heterozygosity, $4N_e\mu$, using the number of polymorphic sites ($4Ne\mu$ for Pbp, $2Nef\mu$ for mtDNA, and $3Ne\mu$ for Z-linked loci) (Watterson 1975).

[f] Heterozygosity, $4N_e\mu$, using the average number of nucleotide differences per site ($4N_e\mu$ for Pbp, $2N_ef\mu$ for mtDNA, and $3N_e\mu$ for Z-linked loci) (Nei 1987).

[g] Tajima's *D* (1989).

[h] Fu and Li's *D** (1993).

[i] Fu and Li's *F** (1993).

[j] Fay and Wu's *H* (2000).

[k] Linkage disequilibrium (Kelly 1997).

[l] Haplotype diversity (Nei 1987).

Testing Non-Equilibrium Demography

Population Substructure: Recent simulation results demonstrate that neutral demographic effects such as substructure and population size changes can have stochastic genome-wide effects that cause empirically determined statistics to depart from standard neutral expectations (Nielsen 2001; Przeworski 2002). Therefore, it is possible that a single non-equilibrium demographic history can account for both the extreme statistics observed at *Tpi* (e.g., *TLD* = 0 and *D* = -2.235 in the E strain) and the more moderate statistics observed at the other genetic markers for North American pheromone strains (Table 4.2, 4.4, 4.5, 4.6a, and 4.6b). We limit our analysis to the recent population history for North American ECB strains because *Tpi* does not exhibit genetic divergence in European strains (Figure 4.1, Table 4.5) and exhibits patterns of polymorphism that appear consistent with the neutral model (Table 4.7a, 4.7b).

For each North American strain we compared the observed *TLD*, Tajima's *D*, and Fay and Wu's *H* to values obtained through coalescent simulation under non-equilibrium demographic conditions likely to have been experienced by North American ECB pheromone strains. Simulations assumed the empirical recombination rate for each locus, the observed number of segregating sites, and the number of sampled sequences (Table 4.8). Because Z and E pheromone strains exhibit (at least) partial reproductive isolation throughout the species range, an island model of substructure involving two populations was explored.

Simulations involving low migration rates ($4N_em$ = 0.01, 0.1, 0.5) indicated that the probability of the observed *TLD* for *COI*, *Ket*, *Ldh*, and *Pbp* was low (Table 4.9). This result was expected given that the unconstrained genealogies for these loci differed from an exclusive tree by at least four changes (Table 4.2). *TLD* for *Tpi*,

Table 4.8. Input for coalescent simulation.

Locus	N [a]	S[b]	*C*[c]
COI	37	11	0
Ket	37	23	1.4
Ldh	37	11	70.4
Pbp	32	47	13.4
Tpi	37	22	4.9

[a] Sample size.

[b] Number of polymorphic sites, excluding indels.

[c] Per-gene recombination parameter inferred from the empirical data (Hudson 1987).

which was compatible with an exclusive tree (*TLD* = 0), had a high probability under these same conditions. The pattern for the data reversed as migration rates increased. With increasing migration rate, simulations were able to recover the observed *TLD* at *COI*, *Ket*, *Ldh*, and *Pbp*, but not at *Tpi* (Table 4.9).

H was significant for *Tpi* under portions of migration parameter space, but otherwise, values for the data could be explained by non-equilibrium demography (table 6a, b, table 9). This was expected as *H* has been shown to be sensitive to population substructure (Przeworski 2002). *P*-values for *H* were below the 0.05 critical threshold for *Tpi* in both pheromone strains when migration rates were high ($4N_em$ = 4, 20). Under more moderate migration rates ($4N_em$ = 1, 1.5), *H* remained significant for the E strain only. In general, *p*-values for *H* increased with lower migration rate (table 9).

In contrast to the pattern observed for *H*, for several loci, *p*-values for *D* increased and then decreased with increasing migration rate (Table 4.9). Within the Z strain, *D* for *COI* was significant when migration was both low ($4N_em$ = 0.01, *P* = 0.031) and high ($4N_em$ = 1.5-20, $P \leq 0.049$). *P*-values were below 0.05 for *Tpi* in the Z strain when migration rates were high ($4N_em$ = 4, 20). Within the E strain, *Tpi* was the only locus that had a significantly negative *D* value. *P*-values for E strain *Tpi* decreased with increasing migration rate, but were below 0.05 across all models (*p*-values ranged from *P* = 0.008 when $4N_em$ = 0.01 to $P < 0.0001$ when $4N_em$ = 20). Because simulations involving substructure alone could not account for the empirically observed values of our statistics, we added a component to the model that describes a change in population size.

Table 4.9. Probability of observing summary statistics under various population migration rates

Locus	M [a]	*D* [b] E strain	*D* [b] Z strain	*TLD* [c]	*H* [d] E strain	*H* [d] Z strain
COI	0.01	0.074	**0.031**	**0.001**	0.987	0.955
Ket		0.467	0.069	**0.0001**	1.000	1.000
Ldh		0.595	0.548	**<0.0001**	1.000	0.999
Pbp		0.602	0.271	**<0.0001**	1.000	1.000
Tpi		**0.008**	0.063	0.995	0.935	0.990
COI	0.1	0.126	0.075	0.054	0.881	0.730
Ket		0.457	0.129	**0.021**	1.000	0.997
Ldh		0.616	0.578	**0.003**	0.996	0.985
Pbp		0.581	0.221	**0.009**	0.995	0.999
Tpi		**0.008**	0.126	0.810	0.475	0.779
COI	0.5	0.121	0.073	0.321	0.639	0.341
Ket		0.417	0.109	0.053	1.000	0.981
Ldh		0.611	0.558	**0.016**	0.942	0.855
Pbp		0.545	0.158	**0.040**	0.933	0.926
Tpi		**0.002**	0.097	0.255	0.077	0.253
COI	1	0.101	0.057	0.511	0.522	0.214
Ket		0.397	0.084	0.083	0.998	0.967
Ldh		0.599	0.545	0.056	0.884	0.773
Pbp		0.545	0.121	0.080	0.849	0.812
Tpi		**0.001**	0.065	0.074	**0.028**	0.112
COI	1.5	0.094	**0.049**	0.593	0.463	0.166
Ket		0.380	0.067	0.110	0.997	0.960
Ldh		0.600	0.549	0.102	0.859	0.720
Pbp		0.530	0.093	0.125	0.786	0.722
Tpi		**4.00E-04**	0.057	**0.027**	**0.017**	0.070
COI	4	0.079	**0.043**	0.779	0.390	0.114
Ket		0.392	0.053	0.244	0.993	0.941
Ldh		0.636	0.581	0.302	0.787	0.620
Pbp		0.571	0.078	0.360	0.688	0.562
Tpi		**<0.0001**	**0.041**	**0.002**	**0.005**	**0.037**
COI	20	0.083	**0.041**	0.906	0.351	0.095
Ket		0.431	0.059	0.483	0.990	0.921
Ldh		0.691	0.637	0.630	0.737	0.552
Pbp		0.635	0.092	0.741	0.604	0.446

Table 4.9 (Continued)

Tpi	**<0.0001**	**0.039**	**<0.0001**	**0.002**	**0.026**

Bold values indicate significance at 0.05 level with empirically-derived recombination (Hudson 1987). Results are based on 10,000 simulations.
[a] Population migration rate, $4N_em$.
[b] Tajima's *D* (1989).
[c] Probability of observing a tree-length difference (*TLD*) that is equal to or more extreme than the observed *TLD* between the unconstrained tree and the tree enforcing exclusivity between pheromone strains, ((E), (Z)). For *Tpi*, probability of observing *TLD* of 0.
[j] Fay and Wu's *H* (2000).

Population Substructure and Population Expansion: If a new migrant population is started from a small number of pioneer individuals, a loss of genetic variation can occur that results in positive values for Tajima's *D* because mutations of low frequency are lost through genetic drift (e.g., (FAY and WU 1999; HAMBLIN and VEUILLE 1999; TAJIMA 1989a; WALL *et al.* 2002). In contrast, for a migrant population that does not experience a bottleneck and rapidly expands, variability may be less affected and *D* would tend to be negative. Overall, derived North American populations do not show an obvious reduction in variability (Table 4.6a, 4.6b and Table 4.7a, 4.7b), and average *D* is negative for North American samples when compared to European ECB (*COI*, *Ldh*, and *Tpi*) (North America E $\bar{D}$ = -1.05 vs. Europe E $\bar{D}$ = 0.23 and North America Z $\bar{D}$ = -0.83 vs. Europe Z $\bar{D}$ = 0.18). Average *D* is also negative when considering all markers for North American strains (North America E $\bar{D}$ = -0.62 with *Tpi* and *D* = -0.21 without *Tpi*; North America Z $\bar{D}$ = -0.89). Therefore, the North American genetic data are not consistent with a founder effect from a small number of migrants, but instead suggest the effects of rapid population expansion. Because we are particularly interested in explaining genetic divergence between North American strains at *Tpi* and the marked negative value for *D* at *Tpi* within the E strain (Table 4.6a), we chose to model the combined effects of recent population expansion and substructure.

Expansion was step-wise, instantaneous, and because *Tpi* shows deviant genetic patterns relative to the other loci only for North American populations, we chose to model expansion of ECB into North America from Europe. Therefore, we assumed that expansion was great in magnitude (going backward in time, from N_o to $0.1N_o$ [model (a)] or $0.01N_o$ [model (b)], where N_o is the effective population size at time, t = 0) and occurred very recently (at $3.75 \times 10^{-4} * 4N_o$ [model (a)] or $3.75 \times 10^{-5} 4N_o$ [model (b)] generations ago). Assuming that $N_o = 10^5$ for model (a) and that N_o =

10^6 for model (b), then expansion in each model took place 150 generations ago. This value is consistent with our knowledge of recent population expansion following introduction of ECB to North America 100 years ago (Caffrey and Worthley 1927), and our parameters under each model represent the average number of generations for bivoltine and univoltine populations during this time.

The range of migration rates that were explored under each model were partly determined by the statistic probabilities. That is, we attempted to explore parameter space that could explain the observed statistics with high probability. Table 10a and 10b present the probabilities for model (a) and model (b), respectively. As with the substructure models that assumed an equilibrium population size (table 9), both expansion models could not account for the observed *TLD* at *COI*, *Ket*, *Ldh*, and *Pbp* when population migration rate was low (Table 4.10a, 4.10b). Marginal probabilities for *TLD* increased with increasing migration rate for these loci, whereas *p*-values for *Tpi* decreased. *H* showed similar patterns for both expansion models, and qualitatively resembled the substructure models (Table 4.9) in that *p*-values decreased with increasing M. For both expansion models, *D* for *COI* in the Z strain was significant at low migration rates. *D* was also significant for *Ket* in the Z strain under low migration. Across all models involving both substructure and expansion, *D* for *Tpi* in the E strain could not be explained ($P \leq 0.013$) (Table 4.10a, 4.10b).

DISCUSSION

Genetic variation is affected by both neutral and non-neutral processes that have acted at various periods during the evolutionary time course of a particular gene region. In addition to being subject to forces that have opposing or synergistic effects on genetic variation, in all likelihood, these processes often overlap. As a result,

Table 4.10a. Probability of observing summary statistics under various population migration rates with 10-fold population expansion.

Locus	M [a]	*D* [b] E strain	*D* [b] Z strain	*TLD* [c]	*H* [d] E strain	*H* [d] Z strain
COI	0.1	0.072	**0.032**	**0.001**	0.983	0.955
Ket		0.491	0.075	**0.001**	1.000	1.000
Ldh		0.616	0.595	**0.001**	1.000	0.998
Pbp		0.627	0.306	**0.001**	0.998	0.999
Tpi		**0.011**	0.086	0.990	0.896	0.967
COI	1	0.125	0.072	0.055	0.873	0.723
Ket		0.485	0.138	**0.028**	1.000	0.995
Ldh		0.637	0.606	**0.019**	0.987	0.969
Pbp		0.624	0.314	**0.033**	0.965	0.981
Tpi		**0.016**	0.156	0.758	0.443	0.698
COI	4	0.127	0.080	0.282	0.668	0.382
Ket		0.454	0.141	0.064	0.998	0.982
Ldh		0.620	0.580	0.070	0.936	0.863
Pbp		0.580	0.289	0.137	0.837	0.873
Tpi		**0.008**	0.150	0.307	0.122	0.298
COI	10	0.107	0.061	0.391	0.516	0.212
Ket		0.417	0.106	0.080	0.996	0.964
Ldh		0.603	0.562	0.172	0.859	0.739
Pbp		0.559	0.235	0.259	0.698	0.688
Tpi		**0.004**	0.112	0.065	**0.037**	0.119
COI	15	0.095	0.051	0.600	0.472	0.162
Ket		0.410	0.093	0.144	0.995	0.952
Ldh		0.616	0.569	0.237	0.820	0.686
Pbp		0.548	0.209	0.333	0.633	0.603
Tpi		**0.002**	0.101	**0.028**	**0.027**	0.085
COI	20	0.096	0.051	0.668	0.435	0.147
Ket		0.410	0.078	0.176	0.993	0.940
Ldh		0.609	0.571	0.300	0.797	0.656
Pbp		0.556	0.204	0.412	0.594	0.553
Tpi		**0.002**	0.091	**0.013**	**0.021**	0.065

Bold values indicate significance at 0.05 level with empirically-derived recombination (Hudson 1987). Results are based on 10,000 simulations.

[a] Population migration rate, $4N_em$.

[b] Tajima's *D* (1989).

Table 4.10a (Continued)

[c] Probability of observing a tree-length difference (*TLD*) that is equal to or more extreme than the observed *TLD* between the unconstrained tree and the tree enforcing exclusivity between pheromone strains, ((E), (Z)). For *Tpi*, probability of observing *TLD* of 0.
[j] Fay and Wu's *H* (2000).

Table 4.10b. Probability of observing summary statistics under various population migration rates with 100-fold population expansion.

Locus	M [a]	*D* [b] E strain	*D* [b] Z strain	*TLD* [c]	*H* [d] E strain	*H* [d] Z strain
COI	0.1	0.050	**0.013**	**<0.0001**	0.998	0.987
Ket		0.470	**0.040**	**<0.0001**	1.000	1.000
Ldh		0.636	0.586	**<0.0001**	1.000	1.000
Pbp		0.609	0.313	**<0.0001**	1.000	1.000
Tpi		**0.007**	**0.045**	1.000	0.980	0.996
COI	1	0.079	**0.030**	**0.001**	0.985	0.956
Ket		0.482	0.069	**0.001**	1.000	1.000
Ldh		0.643	0.617	**0.001**	0.998	0.995
Pbp		0.631	0.308	**0.001**	0.998	0.999
Tpi		**0.009**	0.082	0.990	0.881	0.963
COI	4	0.103	0.056	**0.012**	0.946	0.864
Ket		0.488	0.109	**0.008**	1.000	0.999
Ldh		0.656	0.623	**0.010**	0.993	0.979
Pbp		0.633	0.309	**0.009**	0.982	0.991
Tpi		**0.013**	0.128	0.923	0.666	0.861
COI	20	0.143	0.081	0.137	0.793	0.565
Ket		0.481	0.141	**0.046**	1.000	0.989
Ldh		0.632	0.601	0.076	0.950	0.893
Pbp		0.609	0.324	0.093	0.892	0.925
Tpi		**0.013**	0.165	0.545	0.245	0.490
COI	50	0.122	0.077	0.325	0.633	0.334
Ket		0.448	0.132	0.079	0.998	0.978
Ldh		0.622	0.594	0.160	0.886	0.794
Pbp		0.578	0.285	0.204	0.751	0.786
Tpi		**0.011**	0.145	0.220	0.098	0.234
COI	75	0.118	0.068	0.440	0.571	0.261
Ket		0.440	0.113	0.095	0.997	0.969
Ldh		0.607	0.577	0.207	0.853	0.736
Pbp		0.570	0.273	0.262	0.688	0.706
Tpi		**0.007**	0.133	0.119	0.059	0.156
COI	100	0.106	0.060	0.507	0.523	0.218
Ket		0.431	0.102	0.118	0.997	0.961

Table 4.10b (Continued)

Ldh	0.611	0.581	0.259	0.825	0.703
Pbp	0.563	0.261	0.310	0.640	0.645
Tpi	**0.005**	0.125	0.073	**0.047**	0.120

Bold values indicate significance at 0.05 level with empirically-derived recombination (Hudson 1987). Results are based on 10,000 simulations.

[a] Population migration rate, $4N_em$.

[b] Tajima's *D* (1989).

[c] Probability of observing a tree-length difference (TLD) that is equal to or more extreme than the observed TLD between the unconstrained tree and the tree enforcing exclusivity between pheromone strains, ((E), (Z)). For *Tpi*, probability of observing TLD of 0.

[j] Fay and Wu's *H* (2000).

explanations for genetic variation among populations, species or genes can be incorrect. Thus, making correct interpretations for genetic variation represents one of the most important and fundamental challenges for biology. In this study, we use natural history information for the European corn borer moth to disentangle the effects of neutral demography on genetic variation from explanations involving selection.

Z and E pheromone strains of ECB exhibit behavioral isolation stemming from differences in sexual communication, and some North American populations exhibit temporal isolation stemming from differences in development time (ECKENRODE *et al.* 1983; KLUN *et al.* 1973; LIEBHERR and ROELOFS 1975; LINN *et al.* 1997; ROELOFS *et al.* 1985). In spite of these obvious reproductive barriers, Dopman et al. (submitted) discovered that only a single sex-linked locus, *Tpi*, reveals evidence for genealogical exclusivity between pheromone strains in North America. The four other loci that were sampled indicate that strains share substantial genetic variation. We added to this data set European moths from Hungary and Italy, two countries that have been identified as being likely sources for the North American introduction early in the 20th century (Smith 1920; Caffrey and Worthley 1927). We found that *Tpi* only reveals genealogical exclusivity of pheromone strains for North American populations of ECB (Figure 4.1). Italian E-strain moths are more evenly distributed throughout the gene genealogy, whereas Hungarian Z-strain moths fall within a large Z strain clade. The genealogical pattern at *Tpi* is validated by hierarchical analyses, which indicate that North American E-strain insects are genetically diverged from both Z-strain and European E-strain moths (Table 4.4, Table 4.5). This finding indicates that genetic divergence at *Tpi* has evolved recently within North American E strain ECB.

Positive Selection and Genetic Hitchhiking

As natural selection adapts populations to local environmental conditions, genetic hitchhiking with positively selected mutations purges shared ancestral variation (Maynard Smith and Haigh 1974). In this manner, selective sweeps can generate the geographic and genomic localization of genealogical exclusivity observed for North American ECB pheromone strains. Genetic hitchhiking also produces reduced levels of polymorphism, an excess of low frequency mutations, and high linkage disequilibrium (Braverman et al. 1995; Kelly 1997; Depaulis and Veuille 1998). In our data set, *Tpi* shows evidence for a significant skew in the site frequency spectrum towards low frequency variants, although it does so for most summary statistics only for E strain samples of ECB from North America (Table 4.6a). The other loci for both strains in North America and Europe show patterns that are within standard neutral expectations (Table 4.6a, 4.6b and Table 4.7a, 4.7b). North American E insects also show evidence for significant linkage disequilibrium and reduced haplotype diversity at *Tpi*, two other patterns that are consistent with a recent selective sweep near *Tpi* within the E strain.

The localized geographic structure, reduced diversity, excess of low frequency mutations, and high linkage disequilibrium at *Tpi*, and the general lack of such patterns at other loci, suggest recent positive selection near *Tpi* and genetic hitchhiking in the E strain of North American ECB. These results are compatible with the recent history of ECB in North America and the genomic position of divergent traits between pheromone strains.

That populations of ECB encountered environmental differences when they were introduced to North America 100 years ago seems almost certain. Equally likely are differences in the length of time available during the year to complete

development. Early range maps place ECB around the Great Lakes, eastern New York, and New England (Caffrey and Worthley 1927). If presented with a shorter or longer growing season in these regions, then for larvae to successfully reach the overwintering fifth-instar stage before winter, a life history step(s) would have to be shortened or lengthened. An important life history stage for ECB is post-diapause development time, the time to pupation for fifth-instar larvae following winter. A shorter or longer post-diapause development for overwintering larvae in the beginning of the season will also influence the developmental stage achieved by larvae in the final generation. Indeed, the major genetic factor for post-diapause development, *Pdd*, largely controls the number of generations per year in New York (bivoltine or univoltine). Thus, *Pdd* indirectly affects the likelihood that the generation immediately prior to winter successfully survives to the following year by influencing the developmental stage achieved by insects before the onset of winter.

Tpi maps to the same genomic region as *Pdd* in an area that exhibits the lowest estimated recombination rate for the entire genome (Dopman et al. 2004; Dopman et al. submitted). Therefore, selection on *Pdd* to alter diapause-development and linkage between *Tpi* and *Pdd* provides a plausible scenario for localized genetic patterns between strains in North America. That genetic divergence and evidence for hitchhiking at *Tpi* is between pheromone strains (Z vs. E) and not voltinism races (bivoline vs. univoltine), suggests convergence of post-diapause development between bivoltine Z (BZ) and bivoltine E (BE) races in New York.

Purifying Selection and Background Selection

Although genetic hitchhiking and the positive selection of advantageous mutations may commonly characterize the *origin* of speciation genes (e.g., Presgraves

et al. 2003; Barbash et al. 2004), the maintenance of genetic divergence demands an alternative explanation. For hybridizing species, it has been proposed that genetic divergence at speciation genes or species-specific adaptations will be maintained by purifying selection acting to remove deleterious mutations introduced by introgression from a diverged species, whereas in neutral gene regions, interspecific gene exchange will be unimpeded (Barton and Hewitt 1981; 1985; Harrison 1990). The genomic pattern of differential introgression produced by deleterious and neutral (or advantageous) loci has long been of interest to hybrid zone researchers for characterizing the genetic architecture and evolution of adaptations and speciation genes (Barton and Hewitt 1981; 1985; Harrison 1990).

Gene genealogies and coalescent theory have also provided insight about the effects of purifying selection on genetic variation. Originally proposed to describe the removal of strongly deleterious recurrent mutations within species, the "background selection" model of evolution may provide testable predictions for genetic variation between hybridizing species (CHARLESWORTH *et al.* 1993; CHARLESWORTH *et al.* 1995; HUDSON and KAPLAN 1994, 1995). If we assume that deleterious alleles within species are derived from introgression at speciation genes, rather than recurrent mutation, then the expectations for genetic variation between hybridizing species matches those under the background selection model with a mutation rate on the order of the interspecific migration rate. With strong selection and a high mutation (= migration) rate, background selection will generate an excess of low frequency alleles and a reduction in levels of genetic variation in regions containing speciation genes (Hudson and Kaplan 1994; Charlesworth et al. 1995). Therefore, the deviant patterns of variation at *Tpi* may represent the effects of tight linkage to *Pdd*, with *Pdd* acting as a speciation gene between pheromone strains in North America. Indeed, selection might be quite strong against offspring with intermediate phenotypes or with the

diapause development time that makes them active at a time when they are unable to attract or find mates (e.g., a univoltine E moth at an univoltine Z/BE locality). Genetic divergence between strains and evidence for selection only in E strain ECB suggests that *Pdd* is not acting as a speciation gene between univoltine Z and bivoltine Z populations.

Stochastic Non-equilibrium Demography

Theoretical and empirical results suggest that genes for reproductive barriers, or speciation genes, are often targets of selection and are more likely to reveal genetic divergence or exclusivity between closely related species. One might conceivably use these signatures to localize the genomic positions for traits involved in the biological process of speciation; however, reductions in gene flow will randomly produce gene regions displaying both exclusivity and hallmarks of selection (Tahakata and Slatkin 1990; Depaulis and Veuille 1998; Wakeley 2000; Nielsen 2001; Przeworski 2002). Therefore, it is important when using this approach to assess whether neutral demography can fully explain patterns of genetic variation and genealogical descent between closely related species. Here, we use our knowledge of the recent population history of North American ECB to assess whether patterns of variation at *Tpi* and the other loci could have a simple neutral explanation.

We explored a coalescent island model with varying population migration rates (M) to assess whether population substructure, as influenced by sexual isolation, could account for North American genetic patterns. In particular, the direction of the significant values for the summary statistics at *Tpi* in the E strain are expected under some models of substructure involving a small number of sequences from an unsampled population that have been introduced to the sampled population via

introgression (Table 4.6a) (Kelly 1997; Depaulis and Veuille 1998; Nielsen 2001; Przeworski 2002). On close inspection, the gene genealogy for *Tpi* appears to support this scenario because a single E strain sequence falls far outside the other E strain sequences (Figure 4.1). Can substructure explain both the pattern at *Tpi* and the other markers?

We chose to explore population substructure using three statistics: Tajima's *D*, Fay and Wu's *H*, and tree length difference from a parsimony tree (*TLD*). *D* and *H* each provides different summaries of the site frequency spectrum, while *TLD* summarizes the extent that a given genealogy differs from an exclusive tree. Together these statistics offer a complimentary summary for patterns of descent and genetic variation, both of which are affected by demography and selection.

An intermediate level of population migration (M = 1.0) explains most of the calculated statistics for most loci, as indicated by the fewest number of significant marginal probabilities (two) (Table 4.9). In this model, the only empirically derived values that remain significantly different from neutral expectations are *D* and *H* at *Tpi*, but only within the E strain. This suggests that E strain *Tpi* represents an outlier relative to the other loci even when considering substructure, a pattern expected if *Tpi* were influenced by selection. Indeed, as indicated by significant *p*-values for *D* within the E strain at *Tpi* across all models, which range from M = 0.01 to M = 20, substructure alone is an insufficient explanation for the data.

We added population expansion to the substructure model to assess whether this addition provided a better fit to the observed statistics. We explored various population migration rates under two alternative expansion models, one which characterized a 10-fold population increase (Table 4.10a) and another a 100-fold increase (Table 4.10b). Both models assumed that population expansion occurred recently to simulate expansion of ECB populations upon introduction to North

America 100 years ago (Caffrey and Worthley 1927). For each expansion model there existed at least one population migration rate that could explain all but one of the statistics. Like the best substructure model (Table 4.8), both of the best expansion models showed significant *p*-values for *D* at *Tpi* in the E strain (Table 4.10a, 4.10b). This occurred at M = 4 for the 10-fold model and at M = 50 or M = 75 for the 100-fold model. Also like the substructure model, Tajima's *D* in E strain samples for *Tpi* could not be explained under any combination of parameters. Thus, even when simple neutral demographic processes are taken into account, E strain *Tpi* represents an outlier.

Conclusions

Two independent lines of evidence support selection as the cause of genetic divergence at *Tpi*: patterns of genealogical descent and genetic variation (this study) and mapping data indicating tight linkage with *Pdd* (Dopman et al. submitted). However, genetic hitchhiking and background selection, which correspond to positive or purifying selection on speciation genes, can create similar patterns of genealogical descent, genetic variation, and genetic architecture (Maynard Smith and Haigh 1974; Barton and Hewitt 1981, 1985; Kaplan et al. 1989; Braverman et al. 1995; Harrison 1990). For closely related species, positive selection and genetic hitchhiking may be responsible for the origin of genetic divergence in gene regions containing speciation genes, whereas purifying selection and background selection may be responsible for the maintenance of divergence in these same regions for species that hybridize. Thus, for ECB pheromone strains it is unclear whether selection at *Tpi* reveals the contemporary demographic history of reproductive isolation between populations, or whether it reveals the contemporary locus-specific history of reproductive isolation

between alleles at a speciation gene. That is, *Tpi* may reveal genetic divergence because of a recent selective sweep in one of two reproductively isolated pheromone strains or because selection is acting to maintain genetic divergence at this locus in the face of contemporary introgression. Identical DNA sequences shared by both strains at all of the loci but *Tpi* might help discriminate between these possibilities (Figure 4.2 and Figure 4.3) (Dopman et al. submitted), but shared variation can represent either shared ancestral polymorphism or contemporary gene exchange. In order to discern whether genetic hitchhiking or background selection explains divergence at *Tpi*, we will focus on rapidly evolving neutral markers that may show allele frequency differences if reproductive isolation characterizes ECB pheromone strains where they occur together.

Patterns of genetic variation and genealogical descent are commonly used to infer the history of species, populations, and genes. Unfortunately, both genome-wide and locus-specific forces conspire to produce genetic patterns that confuse neutral and deterministic explanations. The conflation of neutral demography with selection presents specific challenges when interpreting the evolutionary history for closely related species, subspecies, races or strains, groups that arguably offer the clearest window into the speciation process. This is because the evolution of reproductive isolation between populations is a demographic process that has *both* stochastic, genome-wide and deterministic, locus-specific effects on genetic variation. The resolution to these challenges may only be evident when patterns of genetic variation and genealogical descent are combined with natural history and genetic mapping of adaptations and reproductive barriers. If so, then Z and E pheromone strains of the European corn borer moth provide an excellent opportunity for investigating how the divergence of populations during speciation causes the divergence of genomes.

REFERENCES

BARBASH, D. A., P. AWADALLA and A. M. TARONE, 2004 Functional divergence caused by ancient positive selection of a Drosophila hybrid incompatibility locus. PLoS Biol 2: e142.

BARTON, N. H., and G. M. HEWITT, 1981 Hybrid zones and speciation, pp. 109-145 in Evolution and Speciation, edited by W. R. ATCHLEY and D. S. WOODRUFF. Cambridge Univ. Press, Cambridge, UK.

BARTON, N. H., and G. M. HEWITT, 1985 Analysis Of Hybrid Zones. Annual Review Of Ecology and Systematics 16: 113-148.

BOUGHMAN, J. W., 2001 Divergent sexual selection enhances reproductive isolation in sticklebacks. Nature 411: 944-948.

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The Hitchhiking Effect On the Site Frequency-Spectrum Of Dna Polymorphisms. Genetics 140: 783-796.

CAFFREY, D. J., and L. H. WORTHLEY, 1927 A progress report on the investigations of the European corn borer. U. S. Department of Agriculture Bulletin 1476: 155pp.

CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics. 134: 1289-1303.

CHARLESWORTH, D., B. CHARLESWORTH and M. T. MORGAN, 1995 The Pattern Of Neutral Molecular Variation Under the Background Selection Model. Genetics 141: 1619-1632.

CIANCHI, R., S. MAINI and L. BULLINI, 1980 Genetic distance between pheromone strains of the European corn borer *Ostrinia nubilalis* different contribution of

variable substrate regulatory and nonregulatory enzymes. Heredity 45: 383-388.

COLOSIMO, P. F., K. E. HOSEMANN, S. BALABHADRA, G. VILLARREAL, JR., M. DICKSON *et al.*, 2005 Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. Science 307: 1928-1933.

DARWIN, C., 1859 The origin of species. John Murray, London.

DEPAULIS, F., and M. VEUILLE, 1998 Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol Biol Evol 15: 1788-1790.

DOBZHANSKY, T., 1937 Genetics and the Origin of Species. Columbia Univ. Press, New York, NY.

DOPMAN, E. B., S. M. BOGDANOWICZ and R. G. HARRISON, 2004 Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (*Ostrinia nubilalis*). Genetics 167: 301-309.

ECKENRODE, C. J., P. S. ROBBINS and J. T. ANDALORO, 1983 Variations in flight patterns of European corn borer (Lepidoptera: Pyralidae) in New York. Environ. Entomol. 12: 393-396.

FAY, J. C., and C. I. WU, 1999 A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol 16: 1003-1005.

FAY, J. C., and C. I. WU, 2000 Hitchhiking under positive Darwinian selection. Genetics 155: 1405-1413.

FU, Y.-X. A. W.-H. L., 1993 Statistical tests of neutrality of mutations. Genetics 133: 693-709.

GILLESPIE, J. H., 2000 Genetic drift in an infinite population. The pseudohitchhiking model. Genetics 155: 909-919.

GLOVER, T., M. CAMPBELL, P. ROBBINS and W. ROELOFS, 1990 Sex-linked control of sex pheromone behavioral responses in European corn borer moths (*Ostrinia nubilalis*) confirmed with TPI marker gene. Arch. Insect Biochem. Phys. 15: 67-77.

GLOVER, T. J., J. J. KNODEL, P. S. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1991 Gene flow among three races of European corn borers (Lepidoptera:Pyralidae) in New York State. Environ Entomol 20: 1356-1362.

GLOVER, T. J., P. ROBBINS, C. J. ECKENRODE and W. L. ROELOFS, 1992 Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. Arch. Insect Biochem. Phys. 20: 107-117.

HAMBLIN, M. T., and M. VEUILLE, 1999 Population structure among African and derived populations of Drosophila simulans: evidence for ancient subdivision and recent admixture. Genetics 153: 305-317.

HARRISON, R. G., 1990 Hybrid zones: Windows on evolutionary process, pp. 69-128 in Oxford Surveys in Evolutionary Biology, edited by D. FUTUYMA and J. ANTONOVICS. Oxford Univ. Press, Oxford, UK.

HARRISON, R. G., and A. T. VAWTER, 1977 Allozyme differentiation between pheromone strains of the European corn borer, *Ostrinia nubilalis*. Ann. Entomol. Soc. Am. 70: 717-720.

HUDSON, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. 50: 245-250.

HUDSON, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18: 337-338.

HUDSON, R. R., and N. L. KAPLAN, 1994 Gene trees with background selection, pp. 140-153 in Non-Neutral Evolution: Theories and Molecular Data, edited by B. GOLDING. Chapman & Hall, London.

HUDSON, R. R., and N. L. KAPLAN, 1995 The Coalescent Process and Background Selection. Philosophical Transactions Of the Royal Society Of London Series B- Biological Sciences 349: 19-23.

HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153-159.

JIGGINS, C. D., R. E. NAISBIT, R. L. COE and J. MALLET, 2001 Reproductive isolation caused by colour pattern mimicry. Nature 411: 302-305.

KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The hitchhiking effect revisited. Genetics 123: 887-899.

KELLY, J. K., 1997 A test of neutrality based on interlocus associations. Genetics 146: 1197-1206.

KIMURA, M., 1983 The Neutral Theory of Molecular Evolution. Cambridge Univ. Press, Cambridge.

KLUN, J. A., O. L. CHAPMAN, J. C. MATTES, P. W. WOJTOWSKI, M. BEROZA *et al.*, 1973 Insect sex pheromones: Minor amount of opposite geometrical isomer critical to attraction. Science 181: 661-663.

KLUN, J. A., and COOPERATORS, 1975 Insect sex pheromones: Intraspecific pheromonal variability of *Ostrinia nubilalis* in North America and Europe. Environmental Entomology 4: 894-894.

KLUN, J. A., and M. D. HUETTEL, 1988 Genetic regulation of sex pheromone production and response: Interaction of sympatric pheromonal types of European corn borer, *Ostrinia nubilalis* (Lepidoptera: Pyralidae). J. Chem. Ecol. 14: 2047-2061.

KLUN, J. A., and S. MAINI, 1979 Genetic basis of an insect chemical communication system: The European corn borer. Environ. Entomol. 8: 423-426.

LIEBHERR, J., 1974 Studies on two strains of European corn borer, *Ostrinia nubilalis* (Hübner), pp. 1-70 in Masters Thesis. thesis, Cornell University, Ithaca, NY.

LIEBHERR, J., and W. L. ROELOFS, 1975 Laboratory hybridization and mating period studies using two pheromone strains of *Ostrinia nubilalis*. Ann. Entomol. Soc. Am. 68: 305-309.

LINN, C., E. JR., M. S. YOUNG, M. GENDLE, T. J. GLOVER and W. L. ROELOFS, 1997 Sex pheromone blend discrimination in two races and hybrids of the European corn borer moth, *Ostrinia nubilalis*. Physiol. Entomol. 22: 212-223.

MARCON, P., D. B. TAYLOR, C. E. MASON, R. L. HELLMICH and B. D. SIEGFRIED, 1999 Genetic similarity among pheromone and voltinism races of *Ostrinia nubilalis* (Hubner) (Lepidoptera: Crambidae). Insect Mol. Biol. 8: 213-221.

MAYNARD SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

NEI, M., 1987 Molecular evolutionary genetics. Columbia University Press, New York.

NIELSEN, R., 2001 Statistical tests of selective neutrality in the age of genomics. Heredity 86: 641-647.

NOOR, M. A., 1999 Reinforcement and other consequences of sympatry. Heredity 83 (Pt 5): 503-508.

PORNKULWAT, S., S. R. SKODA, G. D. THOMAS and J. E. FOSTER, 1998 Random amplified polymorphic DNA used to identify genetic variation in ecotypes of the European corn borer (Lepidoptera: Pyralidae). Ann. Entomol. Soc. Am 91: 719-725.

PRESGRAVES, D. C., L. BALAGOPALAN, S. M. ABMAYR and H. A. ORR, 2003 Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature 423: 715-719.

PRZEWORSKI, M., 2002 The signature of positive selection at randomly chosen loci. Genetics 160: 1179-1189.

ROELOFS, W., T. GLOVER, X. H. TANG, I. SRENG, P. ROBBINS *et al.*, 1987 Sex-pheromone production and perception in European Corn Borer moths is determined by both autosomal and sex-linked genes. Proc. Natl. Acad. Sci. U.S.A. 84: 7585-7589.

ROELOFS, W. L., J.-W. DU, X.-H. TANG, P. ROBBINS and C. J. ECKENRODE, 1985 Three European corn borer populations in New York based on sex pheromones and voltinism. J. Chem. Ecol. 11: 829-936.

ROZAS, J., J. C. SÁNCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496-2497.

SCHNEIDER, S., D. ROESSLI and L. EXCOFFIER, 2000 Arlequin ver. 2.000: A software for population genetics data analysis.

SMITH, H. E., 1920 Broom corn, the probable host in which *Pyrausta nubilalis* Hubn. reached America. J. Econ. Entomol. 13: 425-430.

SORENSON, C. E., G. C. KENNEDY, W. VAN DUYN, J. R. BRADLEY, JR. and J. F. WALGENBACH, 1992 Geographical variation in pheromone response of the European corn borer, *Ostrinia nubilalis*, in North Carolina. Entomol. exp. appl. 64: 177-185.

SWOFFORD, D. L., 2000 PAUP*. Phylogenetic analysis using parisomy (*and other methods). pp. Sinauer, Sunderland, MA.

TAJIMA, F., 1989a The effect of change in population size on DNA polymorphism. Genetics 123: 597-601.

TAJIMA, F., 1989b Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

TAKAHATA, N., and M. SLATKIN, 1990 Genealogy of neutral genes in two partially isolated populations. Theor Popul Biol 38: 331-350.

TAMURA, K., and M. NEI, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512-526.

THORNTON, K., 2003 Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19: 2325-2327.

WAKELEY, J., 2000 The effects of subdivision on the genetic divergence of populations and species. Evolution Int J Org Evolution 54: 1092-1101.

WALL, J. D., P. ANDOLFATTO and M. PRZEWORSKI, 2002 Testing models of selection and demography in Drosophila simulans. Genetics 162: 203-216.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7: 256-276.

WILLETT, C. S., and R. G. HARRISON, 1999 Insights into genome differentiation: Pheromone-binding protein variation and population history in the European corn borer (*Ostrinia nubilalis*). Genetics 153: 1743-1751.

WITTBRODT, J., D. ADAM, B. MALITSCHEK, W. MAUELER, F. RAULF *et al.*, 1989 Novel Putative Receptor Tyrosine Kinase Encoded By the Melanoma-Inducing Tu Locus In Xiphophorus. Nature 341: 415-421.

WU, C., 2001 The genic view of the process of speciation. J. Evol. Biol 14: 851-865.

WU, C. I., and C. T. TING, 2004 Genes and speciation. Nat Rev Genet 5: 114-122.