REPRODUCTIVE BARRIERS: THEIR ORIGINS AND CONSEQUENCES IN EUROPEAN CORN BORER MOTHS

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

Henry David Kunerth

August 2022

REPRODUCTIVE BARRIERS: THEIR ORIGINS AND CONSEQUENCES IN EUROPEAN CORN BORER MOTHS

Henry David Kunerth Ph. D.

Cornell University 2022

DISSERTATION ABSTRACT

Reproductive barriers drive divergence by restricting gene flow between populations. However, barriers are often semipermeable, allowing some gene flow between groups. Further, populations may vary in the number and coincidence of reproductive barriers. This dissertation focuses on patterns of gene exchange among populations of European corn borer moths (*Ostrinia nubilalis*) that vary in the number of acting barriers. In Chapter One, I performed an empirical test of a 'coupling hypothesis' of reproductive barriers, which examines how their joint effects enhance reproductive isolation and genomic divergence. I found divergence mainly restricted to barrier loci when populations differ by a single barrier, whereas the coincidence of temporal and behavioral barriers is associated with enhanced divergence. Specifically, differentiation at temporal barrier loci increases in the presence of behavioral divergence and differentiation at behavioral barrier loci increases in the presence of temporal divergence. In Chapter Two, I examine how barrier loci can guide life history variation when a species expands its range. To identify whether the genome reflects signatures of natural selection at such phenological loci, I studied the genomes of moths from populations along a latitudinal gradient in the Midwestern US. I found genome-wide signatures that were reflective of a large expansion in both range and population size and demonstrated that two key life history loci showed signatures consistent with both selection and

demographic processes. Finally, in Chapter Three, I considered how introgression and recombination complicate patterns of ancestry of key barrier loci, using genealogical trees for individuals and species of the *Ostrinia* moth group. I developed tree sequence topologies from both genome wide markers and a set of four barrier trait loci and calculated genealogical nearest neighbor estimates, a statistic summarizing relative ancestry proportions of each individual to describe how phenotypic variation has shaped the species and population history of *O. nubilalis*. I showed that much of the genome reflects genealogical sharing across many populations, but genealogies at barrier loci distinguish groups. Finally, there was evidence of introgression of these barrier traits between multiple species within the *Ostrinia* group, suggesting a complex model for barrier trait evolution in this group.

BIOGRAPHICAL SKETCH

Henry David Kunerth was born on May 7th, 1991 in New Ulm, Minnesota. He graduated from Cathedral High School in May 2009. Henry then spent one year living as an exchange student in Mandal, Norway, studying Norwegian culture and language. He attended the University of Minnesota, completing a double major of a Bachelor's Degree of Science in Ecology, Evolution, and Behavior and a Bachelor's Degree of Arts in Scandinavian Languages and Finnish. While attending the University of Minnesota, he worked as an Undergraduate Research Assistant in Dr. Marlene Zuk's lab, studying the phylogeography of parasitic wasps and their cricket hosts. He then moved to Ithaca, New York to study Evolutionary Biology and Genetics at Cornell University in the Ecology and Evolutionary Biology Department. Henry began graduate school advised by Dr. Rick Harrison and completed lab rotations with Dr. Andy Clark and Dr. Philipp Messer in his first year. He began his dissertation focusing on reproductive barriers and their consequences in European corn borer moths. Following Rick's untimely passing in early 2016, Henry joined Dr. Jeremy Searle's lab and continued his work. He collaborated closely with Dr. Erik Dopman at Tufts University and Dr. Genevieve Kozak at the University of Massachusetts – Dartmouth and was guided by committee members Dr. Robert Reed, Dr. Matthew Hare, Dr. Erik Dopman, and Dr. Philipp Messer. Henry relocated to Minneapolis, Minnesota in September 2021 to finish writing his dissertation. He gave his defense seminar and defended his dissertation in Ithaca on May 16th, 2022.

Dedicated to my partner, family, friends, mentors,

teachers, colleagues, nature, and my pets.

ACKNOWLEDGMENTS

I am fortunate to have had an enormous amount of support for this dissertation, without which it could never have been written. It all began with Rick Harrison inviting me to join his lab. His mentorship, though tragically brief, was foundational to any success I've had as a scholar. I owe immense gratitude to Jeremy Searle for continuing to guide me in Rick's absence. Countless colleagues, postdocs, and faculty contributed immensely in shaping my thinking about the natural world, especially Steve Bogdanowicz, Genny Kozak, Erik Dopman, Abby Drake, and my committee members. Students of all levels brought me joy and challenged me in the best ways. My friends in Ithaca and especially Brooktondale made me a better person. My family have supported me at every stage of my life and have inspired me with their strength and generosity. Amanda, my partner, has blessed my life with her presence, and graciously shares her kindness, resilience, patience, and wonder with me every day. Our pets, Lucy and Willow, were ever-present sources of support.

I am also grateful to the National Science Foundation, Andrew W. Mellon Foundation and Cornell President's Life Sciences Fellowship for helping to fund my work.

TABLE OF CONTENTS

CHAPTER ONE

CONSEQUENCES OF COUPLED BARRIERS TO GENE FLOW FOR THE BUILD-UP OF GENOMIC DIFFERENTIATION

ABSTRACT

Theory predicts that when different barriers to gene flow become coincident, their joint effects enhance reproductive isolation and genomic divergence beyond their individual effects, but empirical tests of this 'coupling' hypothesis are rare. Here, we analyze patterns of gene exchange among populations of European corn borer moths that vary in the number of acting barriers, allowing for comparisons of genomic variation when barrier traits or loci are in coincident or independent states. We find that divergence is mainly restricted to barrier loci when populations differ by a single barrier, whereas the coincidence of temporal and behavioral barriers is associated with divergence of two chromosomes harboring barrier loci. Furthermore, differentiation at temporal barrier loci increases in the presence of behavioral divergence and differentiation at behavioral barrier loci increases in the presence of temporal divergence. Our results demonstrate how the joint action of coincident barrier effects leads to levels of genomic differentiation that far exceed those of single barriers acting alone, consistent with theory arguing that coupling allows indirect selection to combine with direct selection and thereby lead to a stronger overall barrier to gene flow. Thus, the state of barriers – independent or coupled – strongly influences the accumulation of genomic differentiation.

INTRODUCTION

Understanding the origin of new species is crucial for explaining patterns of biodiversity. New species are thought to arise as populations diverge and accumulate trait differences that limit interbreeding, genetic exchange, and sharing of evolutionary processes (Mayr 1942, Coyne and

Orr 2004, Harrison and Larson 2014). Consequently, there is long-standing interest in understanding how barriers to reproduction evolve and promote independent evolution of populations in nature, as characterized by the emergence of distinct genetic groups.

Most sister species are separated by multiple barriers to gene flow such as differences in mating time, differences in mate choice, and decreased hybrid survival (Coyne and Orr 2004; Dopman et al. 2010), arguing that speciation depends on the accumulation of barriers and strong reproductive isolation. Different barriers need not, however, initially evolve between the same populations. Instead, independent barriers could restrict gene flow between different populations, potentially as a gradual build-up of barriers with geographic distance (Barton 2013). Without the coincident action of multiple barriers between the same populations, theory indicates that differentiation will be centered at loci underlying barrier traits, reflecting weak overall reproductive isolation and potentially migration-selection balance (Barton 1983; Barton and Bengtsson 1986; Barton and de Cara 2009; Butlin and Smadja 2018; Flaxman et al. 2014; Schilling et al. 2018). At other genomic regions, gene flow and recombination can allow for shared evolutionary processes within what is essentially a single population (Barton 2013).

When multiple barrier traits become coincident in a process known as 'coupling,' their joint effects are believed to enhance reproductive isolation and genomic divergence beyond their individual effects, potentially leading to strong reproductive isolation sufficient for the origin of new species (Barton 1983; Barton and Bengtsson 1986; reviewed in Butlin and Smadja 2018). Recent simulations of divergence with gene flow suggests the presence of tipping-point dynamics, wherein a certain threshold of coupling, achieved by the accumulation of new barrier loci or

increased linkage disequilibrium (LD) among existing loci, initiates a feedback loop of rapid differentiation across the genome (Flaxman et al. 2014, Schilling et al. 2018, reviewed in Nosil et al. 2017). This coupling threshold may correspond to a critical point at which the strength of selection across accumulated barrier loci, arising from both direct fitness effects of each locus and indirect fitness effects of their non-random associations, outweighs the homogenizing impact of recombination (measured as φ, the 'coupling coefficient', Barton 1983). As coupling builds towards this tipping point, recombination is predicted to cause a lag in divergence at neutral loci resulting in heterogeneous genomic differentiation (Nosil et al. 2017; Schilling et al. 2018). Genomic architectures that reduce recombination, including pleiotropy, tight physical linkage, or chromosomal rearrangements, are therefore expected to promote coupling (Servedio et al. 2011; Dagilis and Kirkpatrick 2016, Ortiz-Barrientos et al. 2016; Kirkpatrick & Barton 2006). Although coupling does not require physical linkage, in a widespread pattern known as the 'large X-effect,' sex chromosomes often harbor more barrier loci than autosomes (Coyne and Orr, 1989; Presgraves 2008; 2018). Consequently, the X (or Z) chromosome might be prone to coupling dynamics and rapid chromosomal differentiation (Lasne et al. 2017; Presgraves 2018). Consistent with this, 95% of 129 studies reviewed in Presgraves (2018) find evidence between taxa of elevated differentiation on sex chromosomes relative to autosomes, although there is no shortage of alternative hypotheses for biased sex chromosome differentiation (e.g. Caballero 1995; Pool & Nielsen 2007; Charlesworth 2012; reviewed in Presgraves 2018).

Critical to theoretical predictions of a build-up of genomic divergence that ultimately characterizes new species is the coupling of multiple barriers to gene flow. Understanding how coupling counteracts gene flow to allow genomic divergence requires us to reconstruct the sequence of

events, from barrier independence to barrier coincidence, as speciation progresses. Much empirical work has indirectly studied the evolution of coupling during speciation by quantifying genomic differentiation across lineages at differing stages of divergence (Gagnaire et al. 2013, Seehausen et al. 2014; Shaw & Mullen 2014; Riesch et al. 2017; Xu & Shaw 2019, 2021). Equally valuable are systems allowing direct comparisons of genomic variation when barrier traits or loci are found in either a coincident or independent state. Such direct studies of coupling dynamics may be especially challenging in many systems because barriers to gene flow, trait loci, and recombination landscapes are either unknown or barriers are found exclusively in a coincident state. Consequently, strong empirical tests of coupling dynamics are rare.

The European corn borer moth (ECB) *Ostrinia nubilalis* (Hübner) (Lepidoptera: Crambidae) is useful for studying coupling dynamics because hybridizing population pairs can differ by independent or coincident barriers. In addition to non-coincident barrier effects (Butlin and Smadja 2017), ECB populations differ in combinations of sex pheromone, host-plant association, and phenology (Malausa et al., 2007; Dopman et al. 2010, Coates et al. 2018). Throughout much of its introduced range in North America and its ancestral range in Europe and Asia, two pheromone strains co-exist (Klun and Cooperators, 1975; Anglade and Stockel 1984). In the E strain, females produce and males respond to a sex pheromone comprised mainly of the E isomer of 11-tetradecenyl acetate (E pheromone), whereas Z moths use a blend of mainly Z isomer (Z pheromone) (Klun 1973, Kochansky 1975). The Z strain is more widespread, including areas west of the Appalachian Mountains where it occurs alone (Klun and Cooperators, 1975; Sorenson et al. 1992). Although dietary generalists (Hodgson, 1928; Lewis, 1975), larvae show some host-plant specialization on cultivated maize (Calcagno, et al. 2007; Fisher et al 2017), their predominant

host (O'Rourke et al. 2010; Coates et al. 2019). However, in France, the E strain shows a near complete affiliation with non-maize hosts (Malausa et al., 2007), resulting in its identification as the sister species *O. scapulalis* (Calcagno et al. 2007; Frolov et al. 2007). Finally, three locally adaptive ecotypes have been described in North America that differ in seasonal phenology and voltinism. A northern ecotype ($>42^{\circ}$N) produces one mid-season generation per year (univoltine), a central ecotype (~36–45˚N) makes one early-season and one late-season generation (bivoltine), and a southern ecotype ($<40^{\circ}$N) makes three or more generations per year (multivoltine) (Showers et al. 1976; 1981). The northern ecotype uses the Z pheromone, whereas both Z and E pheromones are used by central and southern ecotypes (Sorenson et al. 1992; Showers 1993). Northern and central ecotypes become sympatric across an area stretching from Wyoming to Maine (Showers et al. 1975; 1981), resulting in population pairs differing by voltinism alone, or both voltinism and pheromone.

The three main axes of phenotypic divergence (phenology, pheromone signaling, host-plant use) result in various barrier effects. Populations of North American ECB moths can experience a temporal barrier from differences in seasonal phenology (bivoltine versus univoltine), a behavioral barrier from differences in pheromone communication (Z versus E), or a coincidence of the two barrier types (bivoltine E versus univoltine Z) (Figure 1.1). Ten other potential barriers to gene flow have been studied that are comparatively weaker (Dopman et al. 2010). North American populations are derived from Hungary and Italy (Smith 1920, Caffrey and Worthly 1927) rather than France, where barrier effects between the ECB moth *O. nubilalis* and its sister species *O. scapulalis* consist of a coincidence of pheromone and host, but not voltinism (Thomas et al. 2003).

Four other aspects of the system are noteworthy because they have been shown by simulations to efficiently reduce gene flow or promote coupling (Flaxman et al. 2014, Nosil et al. 2017, Schilling et al. 2018). First, evidence suggests pheromone signal and preference, controlled by *pgFAR* (pheromone production) and *bab* (preference) genes (Unbehend et al. 2021, Lassance et al. 2010), are 'multiple-effect traits' (Smadja and Butlin 2011) that contribute to both pre- and post-zygotic isolation (mate choice and behavioral sterility) (Glover et al. 1991; Dopman et al. 2010; Unbehend et al. 2021). Second, in an example of 'magic trait' evolution (Servedio et al. 2011), temporal isolation arising from non-random meeting of univoltine and bivoltine adults is an automatic, pleiotropic outcome of divergent selection on diapause phenology, controlled by two interacting clock genes *per* and *Pdfr* (Dopman et al. 2005; Levy et al. 2015; Kozak et al. 2019). Third, three of four identified barrier loci (*per*, *Pdfr*, *bab*) map to the Z chromosome, creating a 'large Z-effect' (Dopman et al. 2004; Kozak et al. 2019; Unbehend et al. 2021). Finally, a chromosomal rearrangement (putatively an inversion(s)) suppressing recombination across ~40% of the Z chromosome has captured at least one barrier locus (*per*) and may be abundant within some E pheromone populations (Wadsworth et al. 2015; Kozak et al. 2017).

Prior population genetic results suggest identified barriers influence the build-up of genomic divergence. Geographic distance is weakly associated with genetic variation throughout the range (Bourguet et al. 2000; Kim et al. 2011; Coates et al. 2019), consistent with numerous examples of long-distance dispersal (30–60 km) (reviewed in Sappington 2018). Instead, genetic structure appears related to barrier effects. In France, differences in pheromone and host use between sympatric *O. nubilalis* and *O. scapulalis* are associated with significant genetic differentiation (Bethenod et al., 2005; Malausa et al., 2007). In an analysis restricted to 65 mostly autosomal SNPs

among 12 localities in North America, pheromone signaling but not host or geography significantly contribute to genetic differentiation, but the influence of phenology was unmeasured (Coates et al. 2019). Kozak et al. (2017) used genome-wide data to study the Z chromosome rearrangement in North America and did find elevated differentiation when populations differed in both pheromone and phenology compared to pheromone alone, but contrasts were between single population pairs. Overall, these prior studies do imply that barrier phenotypes heavily shape patterns of gene flow. However, incomplete genomic data or barrier effect combinations have made it difficult the evaluate the importance of coupling dynamics in this system.

Here, we help resolve some earlier limitations by analyzing pooled genome data from 14 population samples and compare genetic divergence between four types of population pairs characterized by 1) no barrier effects, 2) only phenology differences (voltinism), 3) only pheromone differences, or 4) differences in both phenology and pheromone (joint divergence). In addition, we analyze targeted amplicon sequencing of the 'large-effect' Z chromosome and the autosomal pheromone production locus *pgFAR* from individuals collected from six locations varying in barrier number. Critically, these data allow for comparisons of sympatric populations differing by independent or coincident barriers, allowing us to ask whether we can detect empirical evidence that coupling promotes speciation as suggested by theory. Accordingly, compared to barrier traits when acting alone, upon coincidence we predict evidence of stronger selection or more efficient reduction of gene flow. Two expected signatures are increased divergence around individual barrier loci as well as enhanced coupling (linkage disequilibrium, LD) among barrier loci. If overall selection after coupling outweighs recombination, theory predicts tipping point dynamics and a transition towards differentiation of neutral loci. Differentiation can be genome-

wide or more localized, but the combination of a 'large Z-effect' and a Z-linked rearrangement might suffice in allowing for the build-up of sex chromosome divergence. An alternative possibility is that coincident barriers do not lead to a detectable strengthening of the overall barrier. This could mean that individual barriers are so weak that they evolve independently even when coincident, or possibly that coupling effects are difficult to detect in nature, and theory could explore further quantitative predictions for empirical systems.

MATERIALS AND METHODS

We sampled 13 localities in Eastern United States between 2000 and 2014 (Figure 1.1). Pheromone usage was based on male capture in traps baited with either synthetic E ("New York") or Z ("Iowa") pheromone lures (*Heliothis* traps, Scentry Biologicals, Billings, MO). Traps were separated by ≥ 12 m next to sweetcorn fields. Trap data were verified by gas chromatography of female pheromone gland extracts for the presence of E and Z isomers. For 7 sites, we sampled diapausing larvae from corn stalks and raised individuals under diapause breaking conditions (26˚C, 16h light) until eclosion, allowing classification of voltinism based on post-diapause development (PDD) time (bivoltine: PDD time < 18 days; univoltine: PDD time > 38 days). For one population (Dover), we collected direct developing larvae in July. These individuals were mated as adults, larvae were exposed to diapausing inducing conditions and then phenotyped for PDD. For 5 sites, we determined voltinism by the seasonal timing of male capture in pheromone traps (bivoltine = early and late flight; univoltine = single mid-summer flight) (Glover 1991, Calvin 1994, Zaman 2008, Dopman 2010).

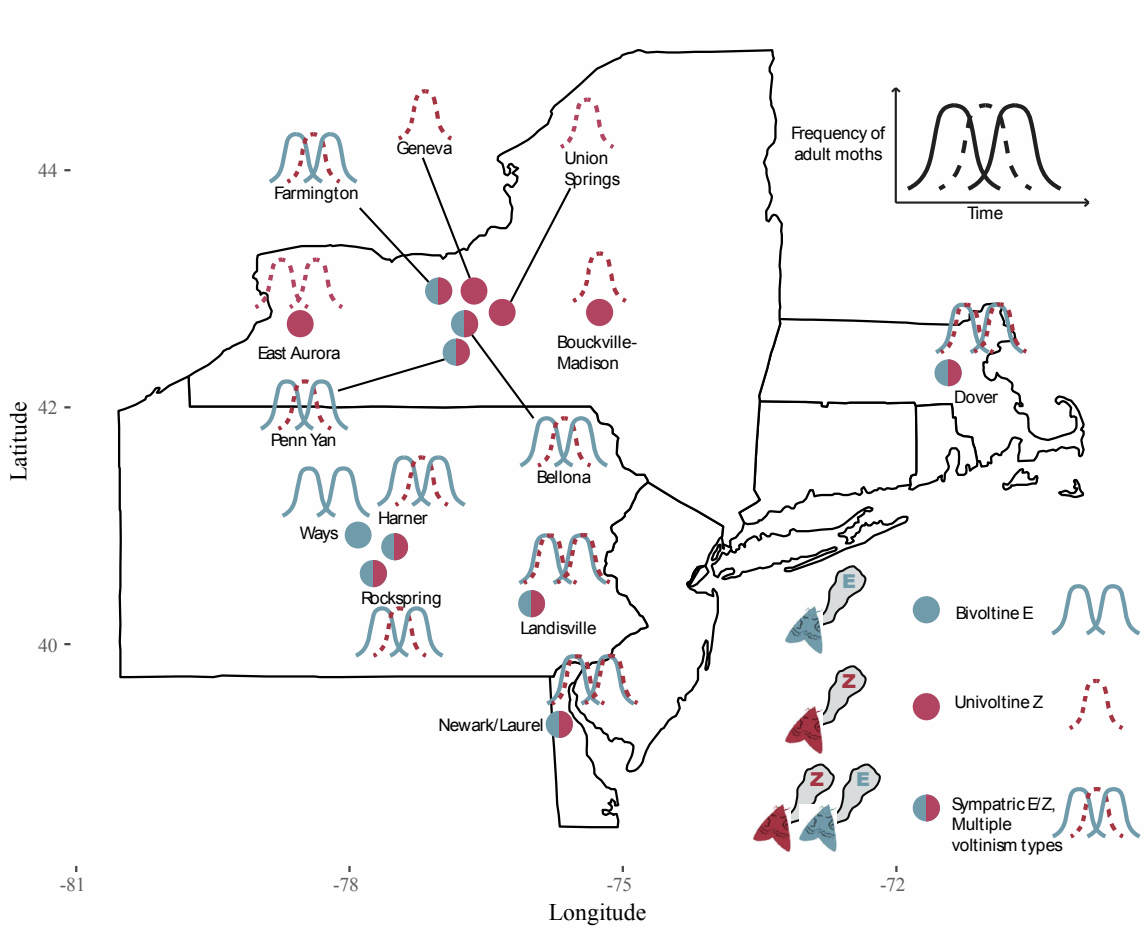

Figure 1.1. Map of collection sites and phenotypic information. Graphs indicate the frequency of adult moths across the growing season and number of distinct generations (voltinism). Graph and circle color indicates whether there were E (blue), Z (red) or both pheromone strains present at each collection site.

Pooled genome sequencing

To evaluate how the transition between independent and joint axes of phenotypic divergence might affect genomic differentiation, we performed pooled genome sequencing on 14 population samples (425 individuals) separated by pheromone and phenology (5 bivoltine E (BE) populations, 3 bivoltine Z (BZ) populations, and 6 univoltine Z (UZ) populations; see Table 1.1). Pheromone strain assignment used *pgFAR* genotype (E or Z homozygotes) from individuals as determined by a diagnostic restriction enzyme digest of *pgFAR* PCR products (Coates et al. 2013). DNA was isolated using Qiagen DNeasy tissue kits (Qiagen, Germantown, MD) without vortexing to preserve high molecular weights, and samples were treated with RNase A (Qiagen). DNA from

samples from Bellona, NY were extracted using Qiagen genomic tips (20 G). After quantifying DNA concentration using a Qubit (Thermo-Fisher), pools combined equal amounts of DNA per individual (Table 1.2). Pooled libraries (2 libraries per lane; 2-5 μg/uL per pool) were sequenced on an Illumina HiSeq3000 (2x150 bp) (except for Rockspring and Ways, which had 1x100 bp data) at Iowa State University DNA Facility, Ames, IA.

Table 1.1. Characteristics and locations for pooled data.

Site	Pheromone	Voltinism	Number of males	Number of females	Collection method	Coverage after alignment
Bouckville-Madison, NY	Z	Univoltine	14	6	DI	37.0
Bellona, NY	E	Bivoltine	31	0	PT	53.18
Bellona, NY	Z	Univoltine	25	0	PT	41.52
Dover, MA	Z	Bivoltine	17	5	DD	15.83
East Aurora, NY	Z	Bivoltine	16	18	DI	32.35
Geneva, NY	Z	Univoltine	13	12	DI	5.94
Harner, PA	E	Bivoltine	31	0	PT	7.25
Harner, PA	Z	Univoltine	27	0	PT	4.59
Landisville, PA	E	Bivoltine	41	0	PT	19.29
Landisville, PA	Z	Bivoltine	39	0	PT	5.04
Penn Yan, NY	Z	Univoltine	11	15	DI	23.54
Rockspring, PA	E	Bivoltine	34	0	PT	10.36
Rockspring, PA	Z	Univoltine	33	0	PT	10.26
Ways, PA	E	Bivoltine	37	0	PT	13.33

PT = pheromone trap, DD = direct-developing larvae, DI = Diapausing larvae

Trimmomatic v.35 removed Illumina adapters (TruSeq2 single-end or TruSeq3 paired-end), and reads with quality score (q) < 15 over a sliding window of 4 and reads < 36 bp long were removed (Bolger et al. 2014). Repetitive regions of the ECB reference genome (GenBank BioProject: PRJNA534504; Accession SWFO00000000; Kozak et al. 2019) were masked by RepeatMasker (using *Drosophila melanogaster* TE library in repbase; http://www.repeatmasker.org/; accessed March 2017), to which trimmed genomic data were aligned using bowtie2 v.2.2.3 (Langmead and Salzberg 2012). Aligned reads were sorted and filtered using Picard (v.2.8.0; http://broadinstitute.github.io/picard/) and SAMtools (v.0.1.18) to remove duplicates and reads with a mapping quality score (Q) < 20 (Li et al. 2009). We calculated coverage from the aligned, filtered, sorted bam files using bedtools v.2.26.0 and the genomeCoverageBed function (Table 1.2)

(Quinlan 2014). As described in Kozak et al. (2019), we determined chromosomal positions of scaffolds by aligning genotyping-by-sequencing data from pedigree families (with and without the recombination suppressor segregating) to the genome scaffolds (Coates and Siegfried 2015, Kozak et al. 2017) and constructing linkage groups from identified SNPs in R v.3.5.1 (R Core Team, 2018) using the R/qtl package (Broman et al. 2003). We considered SNPs from Z chromosome-linked scaffolds that showed no recombination in E x Z strain pedigrees but recombined in Z x Z strain pedigrees to be inside the region of suppression (Wadsworth et al. 2015; Kozak et al. 2017). Twenty-nine scaffolds comprising 8.96 Mb of sequence were determined to be inside the region of recombination suppression while 38 scaffolds (11.94 Mb) were outside the region. We placed a small number (3%) of the Z chromosome-linked scaffolds using positions relative to other markers if they were identified as best hits from BLASTN searches using BLAST v.2.8.1 against the genome scaffolds of markers or BACs from previously published Z-linkage maps and pedigree families in *O. nubilalis* (Kroemer et al. 2009; Levy et al. 2015; Koutroumpa et al. 2016; Yasukochi et al. 2016) and also in BLASTN searches of Z chromosome-linked genome scaffolds in the sister species *O. scapulalis* against the ECB genome (Brousseau et al. 2018). Order and orientation of scaffolds was determined using linkage information from previous maps and gene order from *Bombyx mori* and approximate locations across the Z chromosome was calculated in Mb (described in Kozak et al. 2019). Scaffolds with no evidence for Z chromosome linkage in any pedigree were designated as autosomal. Chromosome assignment of scaffolds to various autosomes (chr. 2–31) used linkage groups in the pedigree families, and synteny of scaffolds to *Bombyx mori* and *Melita cinxia* chromosomes (described in Kozak et al. 2019). Chromosome 12 contains the *pgFAR* gene (Dopman et al. 2004).

Table 1.2. Characteristics and locations for amplicon sequencing data.

				Frequency of pgFAR genotypes			
Site	**Population Type**	**Number of females**	**Number of males**	**EE**	**EZ**	**ZZ**	**Difference in average eclosion time between *pgFAR* ZZ and EE insects (days)** †
Dover, MA	Bivoltine E / Bivoltine Z	22	27	50.0%	29.5%	20.5%	‡
East Aurora, NY	Bivoltine Z	18	15	0.0%	3.5%	96.5%	-
Farmington, NY	Bivoltine E / Univoltine Z	26	19	27.8%	8.3%	63.9%	20.10
Laurel & Newark, DE	Bivoltine E / Bivoltine Z	51	0	12.8%	34.0%	53.2%	1.74
Penn Yan, NY	Bivoltine E / Univoltine Z	20	9	24.0%	16.0%	60.0%	34.63
Union Springs, NY	Univoltine Z	24	29	0.0%	0.0%	100.0%	-

† Time to adult eclosion for diapausing insects (26˚C, 16h light) acts as a proxy for post-diapause development time

‡ Time to eclosion data was not available for these samples but was calculated for the subsequent F_1 generation in the lab and confirmed to be bivoltine (see Material and Methods for more detail).

We used SAMtools to identify single nucleotide polymorphisms (SNPs) in populations (Li et al. 2009). Scripts from Popoolation2 v.1.201 were used to filter (removing SNPs near small indels, minor alleles that did not appear twice in each population), calculate allele frequency, and F_{ST} (Kofler et al. 2011). Pairwise differences in F_{ST} were calculated on 100 kb non-overlapping sliding windows (including only windows with a minimum coverage > 14 reads and a maximum coverage < 300 and at least 20 polymorphisms within the window, as recommended by Fabian et al. 2012). We conducted pairwise F_{ST} comparisons among four categories of populations: no divergence in pheromone or phenology (BZ vs. BZ, UZ vs. UZ), differing only in voltinism (BZ vs. UZ), differing only in pheromone (BE vs. BZ), and differing in both pheromone and voltinism (BE vs. UZ).

We separately compared F_{ST} for autosomal and Z chromosomal loci, and Z loci located inside the region of suppressed recombination ("rearranged") or Z loci located outside this region

("collinear"). We removed and separately analyzed genomic scaffolds containing loci known to contribute to voltinism [*per* (voltinism; scaffold 532; Z chromosome inside "rearranged" region; Kozak et al. 2019), and *Pdfr* (voltinism; scaffold87, Z chromosome "collinear" region; Kozak et al. 2019)] and pheromone trait differences [*pgFAR* (female pheromone production: scaffold178, chr 12; Lassance et al. 2009); *bab* (male flight response; scaffold662, scaffold50; Z chromosome "collinear" region; Unbehend et al. 2021)]. We calculated and plotted F_{ST} separately for 100 kb windows overlapping seasonal timing genes (*per, Pdfr,* N = 4 windows total, 2 in each scaffold), windows overlapping the *bab* locus (N = 4 windows), and windows overlapping the *pgFAR* locus (N = 1 window). We calculated the average F_{ST} for all population pairs of a given type within each 100 kb window and then took the mean of all windows for each chromosome or genomic region. We performed comparison among population types using a Kruskal-Wallis test and pairwise posthoc Dunn's tests in the PMCMRplus package (v1.4.4) with a false discovery rate correction in R (Pohlert 2014).

As some population pairs were collected from the same site ("sympatric") and some from geographically distant sites ("allopatric"), we tested for an effect of geographic distance on the matrix of all pairwise F_{ST} differences using Mantel tests in the R package vegan v.2.5.6 (Oksanen et al. 2019). We used a matrix of geographic distance between sites (circle distance) as the explanatory matrix. We also performed a conservative analysis that was restricted to those population pairs within ~100 km (1-3 generations of dispersal, Sappington 2018) (6 population samples in NY, 7 population samples in PA).

Targeted amplicon sequencing

To specifically test how the transition from independent to coincident barrier effects in sympatry influences the build-up of genomic differentiation, we performed amplicon sequencing of the 'large-effect' Z chromosome and *pgFAR* on population samples at six locations (260 individuals, 6 sites) experiencing either no phenotypic divergence or barrier effects (1 allopatric BZ population, 1 allopatric UZ population), pheromone divergence and behavioral isolation (2 sympatric BE/BZ population pairs), or joint phenotypic divergence and cumulative temporal and behavioral isolation (2 sympatric BE/UZ population pairs) (Table 1.2). Females were classified by gas chromatography of pheromone gland extracts, and both sexes by diagnostic SNPs at the *pgFAR* gene. Phenology classification is described above, using adult eclosion time as a proxy for PDD (except at Dover as described above). At Delaware, only female samples were analyzed, but all other locations had both male and female samples in approximately equal proportion.

We used 67 primers for multiplex PCR of 450–470 bp product sizes (Table S1.1). Primers were distributed across the Z chromosome, including near one of the two temporal barrier genes (*Pdfr*), and one near *pgFAR* on chromosome 12 (Figure S1.1). Only loci successfully amplified from both pheromone strains were carried forward to the final multiplex. We carried out two multiplexed reactions using a Qiagen multiplex PCR kit in two 384-well plates. Following amplification, we pooled products and transferred to one 384-well plate for barcoding. We created a single library by pooling these products after dual barcoding using Nextera N5/N7 primers and OneTaq Polymerase (New England Biolabs). We then cleaned the library using Ampure XP (Beckman

Coulter) for 350–600 bp product sizes. We quantified libraries using a Qubit and sequenced them on a single Illumina MiSeq (2x250 bp) lane at the Cornell Institute of Biotechnology Genomics Facility, Ithaca, NY.

We demultiplexed and trimmed raw reads using a quality threshold of 20 with a custom Perl script. For each locus, we aligned reads using ClustalW (Larkin et al. 2007), generated a haplotype table for each sample at all loci, and called genotypes from the two most common haplotypes (minimum 4 reads total). If the ratio of the second most common genotype to the most common genotype was below 0.2, a homozygous genotype for the most common allele was called. We removed from analysis all individuals with ≥ 50% missing data, then loci with > 35% missing data. Following these cutoffs, we re-filtered all individuals and removed those with ≥ 35% missing data. We tested lower missingness cutoffs but found that a 35% missingness filter produced STRUCTURE and PCA results with no discernable differences to lower filters. Applying quality filters resulted in 217 moths (85 males, 132 females).

Using Z chromosome data, we inferred the presence of distinct populations, assigned individuals to populations, and identified admixed individuals using STRUCTURE (v2.3.4) (Pritchard et al. 2000). Since hemizygous females only carry one Z chromosome, we treated all samples as haploid with males contributing two haploid sets. To assure that this design did not bias results, we conducted separate STRUCTURE runs using only male samples (treated as diploid) or only female samples (treated as haploid). In all cases, we demonstrated qualitatively similar results and could not detect any bias. The presented results include all samples and all markers, unless otherwise specified. We tested $K = 1–7$ with a burn-in phase of 10,000 repetitions and a run phase of 100,000

repetitions and used default model assumptions with no weight given to location as a prior. Fisher's exact tests were used to compare differences in estimated admixture across groups. A principal components analysis (PCA) complemented the analysis of genetic structure using the ade4 package v.1.7.13 in R (Dray & Dufour 2007).

Finally, we measured linkage disequilibrium to quantify the strength of coupling at barrier loci (*pgFAR*, *Pdfr*) and the degree to which selection and assortative mating are counteracting neutral gene flow and recombination at each of the six sites. After filtering for a 0.15 minor allele frequency cutoff, we calculated r^2 values for every pair of biallelic SNPs in the female moth dataset using custom scripts in R. Following calculations, we estimated mean r^2 value per locus and then combined results by barrier effect type (i.e., no barrier, one barrier, two barriers). We thinned this dataset to include only *pgFAR* and a single randomly selected locus per Z chromosome-linked scaffold. This resulted in a final dataset including 27 loci: 11 Z chromosome-linked markers outside the rearrangement (including *Pdfr*), 15 Z chromosome-linked markers within the rearrangement, and *pgFAR*. Mann-Whitney U tests were used to compare differences in LD across groups.

RESULTS

Pooled genome sequencing

There were no significant effects of geography on genetic differentiation, as determined by Mantel tests (for any category: "rearranged" Z, $p = 0.47$; "collinear" Z, $p = 0.57$; autosomal, $p > 0.14$; 1000 permutations), consistent with previous studies (Bourguet et al. 2000; Kim et al. 2011; Coates et al. 2019). We demonstrated qualitatively similar results with our conservative (within ~100km) analysis and could not detect any bias (Figure S1.2). Thus, we included all pairwise comparisons regardless of site of collection to increase the number of comparisons that could be analyzed.

All predicted chromosomes not genetically linked to trait loci showed little genetic differentiation regardless of the number of phenotypic differences between population samples (mean $F_{ST} \leq 0.05$, $p > 0.1$, Figure 1.2). In contrast, differentiation on the Z chromosome and chromosome 12 increased as the number of barriers accumulated. Compared to population pairs lacking phenotypic divergence (mean $F_{ST} = 0.04$), differences in pheromone led to significant elevation (25%) of differentiation on chromosome 12 (mean $F_{ST} = 0.05$, $p = 0.03$), whereas differences in phenology (voltinism) did not (mean $F_{ST} = 0.04$, $p = 0.86$), as expected since phenology barrier loci do not occur on chromosome 12. The largest increase (50%) of chromosome 12 differentiation occurred upon joint divergence of pheromone and phenology (mean $F_{ST} = 0.06$, $p = 6.7 \times 10^{-5}$), although the increase was not significantly greater than pheromone divergence alone ($p = 0.17$).

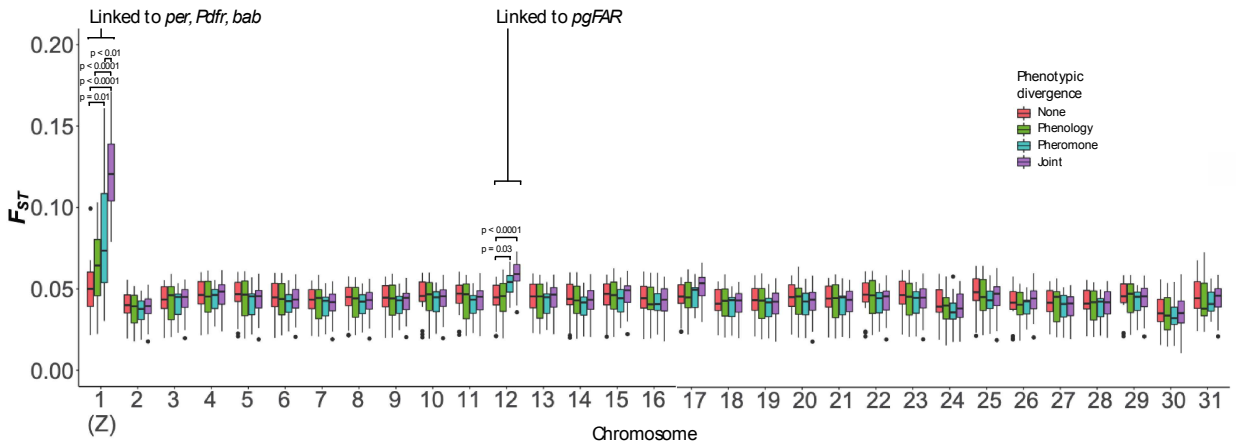

Figure 1.2. Genetic differentiation (mean F_{ST}) estimated in 100 kb windows plotted by predicted chromosome between population pairs lacking differences in barrier traits (Z/Z; red) or differing in phenology (BZ/UZ; green), pheromone signaling (BE/BZ; blue), or both phenology and pheromone (BE/UZ; purple). Scaffolds linked to barrier trait loci *pgFAR*, *per*, *Pdfr*, and *bab* were excluded from F_{ST} calculations on the Z chromosome and chromosome 12. Statistically significant differences assessed via Kruskal-Wallis tests with Dunn's test for false discovery rates are labeled with brackets.

Compared to population pairs lacking phenotypic divergence on the Z chromosome (mean F_{ST} = 0.05, Figure 1.2), phenology divergence led to a weak (20%) non-significant increase of differentiation (mean F_{ST} = 0.06, p = 0.28), whereas a significant (80%) increase was seen in the presence of pheromone divergence (mean F_{ST} = 0.09, p = 0.01). Joint divergence led to the greatest (140%) increase of Z chromosome differentiation (mean F_{ST} = 0.12, p = 5.3 x 10^{-9}) and was statistically significant compared to phenology alone (p = 2.7 x 10^{-6}) and pheromone alone (p = 8.2 x 10^{-3}). This effect was primarily driven by elevated F_{ST} inside the region of reduced recombination among populations that differed in both pheromone and phenology (Figure 1.3D, 1.3E). F_{ST} of the rearranged region was elevated in joint divergence comparisons relative to the remaining three population categories (p < 0.007, Figure 3D). F_{ST} of collinear regions was significantly elevated between populations with joint divergence compared to populations lacking trait differences and populations differing only in phenology (p < 0.007, Figure 3E). However, F_{ST} was not quite significantly elevated compared to pairs that differed only in pheromone (p = 0.11). In populations that differed in just pheromone, F_{ST} inside the region of reduced recombination was

elevated compared to populations lacking barrier effects ($p = 0.007$), but not populations differing in phenology ($p = 0.23$, Figure 1.3D).

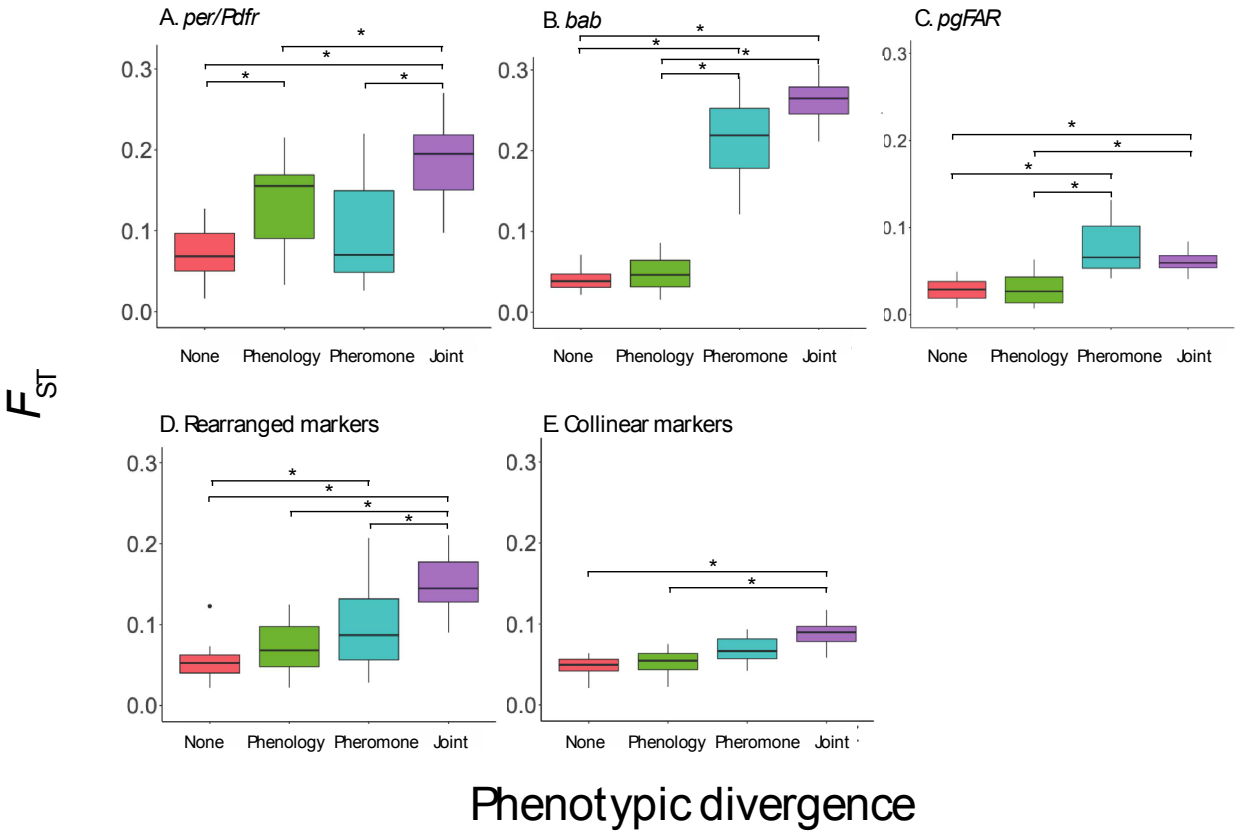

Figure 1.3. A. Genetic differentiation (F_{ST}) at phenology barrier loci *per* and *Pdfr* between population pairs lacking differences in barrier traits (Z/Z; red) or differing in phenology (BZ/UZ; green), pheromone signaling (BE/BZ; blue), or both phenology and pheromone (BE/UZ; purple). B. Genetic differentiation at pheromone response barrier locus *bab* across population pairs. C. Genetic differentiation at pheromone production barrier locus *pgFAR* across population pairs. D. Genetic differentiation at neutral Z chromosome loci inside a ~10 Mb "rearranged" region of suppressed recombination across population pairs. E. Genetic differentiation at neutral Z chromosome loci outside a ~10 Mb region of recombination suppression ("collinear" region) across population pairs. Asterisks indicate significantly different comparisons following Kruskal-Wallis tests and Dunn's tests accounting for false discovery rates (corrected $p < 0.05$).

With respect to barrier loci, the phenology loci (*per/Pdfr*) had significantly elevated F_{ST} in jointly diverged population pairs compared to population pairs without phenology differences ($p < 0.0016$), as well as those differing only in phenology ($p = 0.017$, Figure 1.3A). F_{ST} in the *bab* region significantly increased in all populations that differed in pheromone compared to populations that did not ($p < 0.0003$, Figure 3B), but F_{ST} was not quite significantly increased upon

joint divergence compared to pheromone alone (p = 0.11). F_{ST} in the *pgFAR* region was broadly similar to that seen at *bab* across population pairs. The 100 kb window with the highest mean F_{ST} on the Z chromosome in populations differing only in phenology overlapped the voltinism gene *per* (scaffold532: F_{ST} = 0.23; Figure S1.4), in populations differing only in pheromone signaling it was within the *bab* region (scaffold50: F_{ST} = 0.39), and the highest window in jointly diverged populations was inside the region of recombination suppression (scaffold804: F_{ST} = 0.61).

Targeted amplicon sequencing

The number of distinct genotypic clusters and degree of admixed ancestry varied across localities by the number of acting barriers. The most likely number of clusters across populations was two for all STRUCTURE runs (K = 2, Evanno et al. 2005). The two locations lacking any known barrier effect (Table 1.1) showed strong assignment to a single cluster. For all 73 individuals, more than 98% of their Z chromosome loci belonged to the same cluster (Qs > 0.98; Figure 1.4A). The two locations differing in pheromone but not phenology (BE/BZ) showed evidence of two genotypic clusters and greater admixture. At these sites, 22% (19 of 83) of individuals showed admixture (defined as 0.1 < Q < 0.9). The two locations where pheromone and phenology could act coincidentally (BE/UZ) also showed two genotypic clusters, but unlike at sites where insects differed only in pheromone, most individuals showed limited admixture of the Z chromosome and the proportion of admixed individuals was reduced. 96% of individuals at BE/UZ sites had a high probability of belonging to either cluster 1 (N = 21, Q1 ≥ 0.9) or cluster 2 (N = 37, Q2 ≥ 0.9).

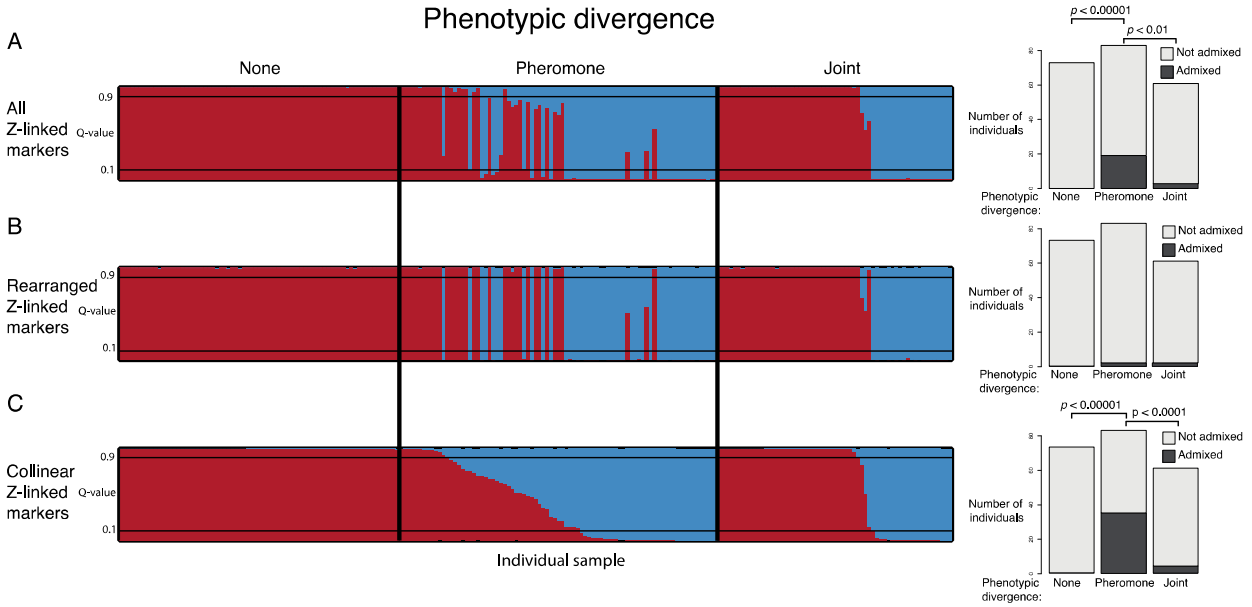

Figure 1.4. STRUCTURE plots partitioned by population category. Plotted on the right are the number of individuals that are considered admixed (Q-value greater than 0.1 and less than 0.9) and not-admixed (Q value greater than 0.9 or less than 0.1). A. All Z chromosome markers; B. rearranged Z markers; C. collinear Z markers. Red indicates assignment to cluster 1 (Z clade), blue indicates assignment to cluster 2 (E clade). Comparisons are between populations lacking differences in barrier traits (None; Z/Z), or pheromone signaling (Pheromone; BE/BZ), or both phenology and pheromone (Joint; BE/UZ).

Separate analyses of Z chromosome loci inside and outside the rearranged region showed that admixture depended on marker location (Figure 1.4), in addition to barriers to gene flow. Population types significantly differed in the degree of admixed collinear regions ($p = 7.63 \times 10^{-14}$), with an excess of admixture at sites with insects differing only in pheromone compared to sites without phenotypic divergence or sites with joint differences in phenology and pheromone. Admixture of rearranged regions was similarly low across all three population types ($p = 0.386$). At sites where insects differed only in pheromone, nearly half (42%) of 83 individuals were estimated to have admixed collinear loci ($0.1 < Q < 0.9$), whereas only 2 showed appreciable admixture at loci in the rearranged region ($p = 1.64 \times 10^{-10}$). In contrast, at sites differing in both pheromone signaling and phenology, both collinear and rearranged regions were relatively free of admixture and the estimated proportion of admixed Z chromosomes did not significantly differ

(3.3% vs. 6.6%, $p = 0.68$). Consistent with a lower per generation rate of admixture across the rearrangement, a large percentage of individuals of mixed ancestry inferred from the rearranged region resembled F_1 hybrids (Q = 0.4 – 0.6; 3 of 4 admixed, 75%) compared to regions with normal recombination rates (11 of 39 admixed, 28.2%) ($p = 0.094$). The former may represent F_1 heterokaryotype male offspring.

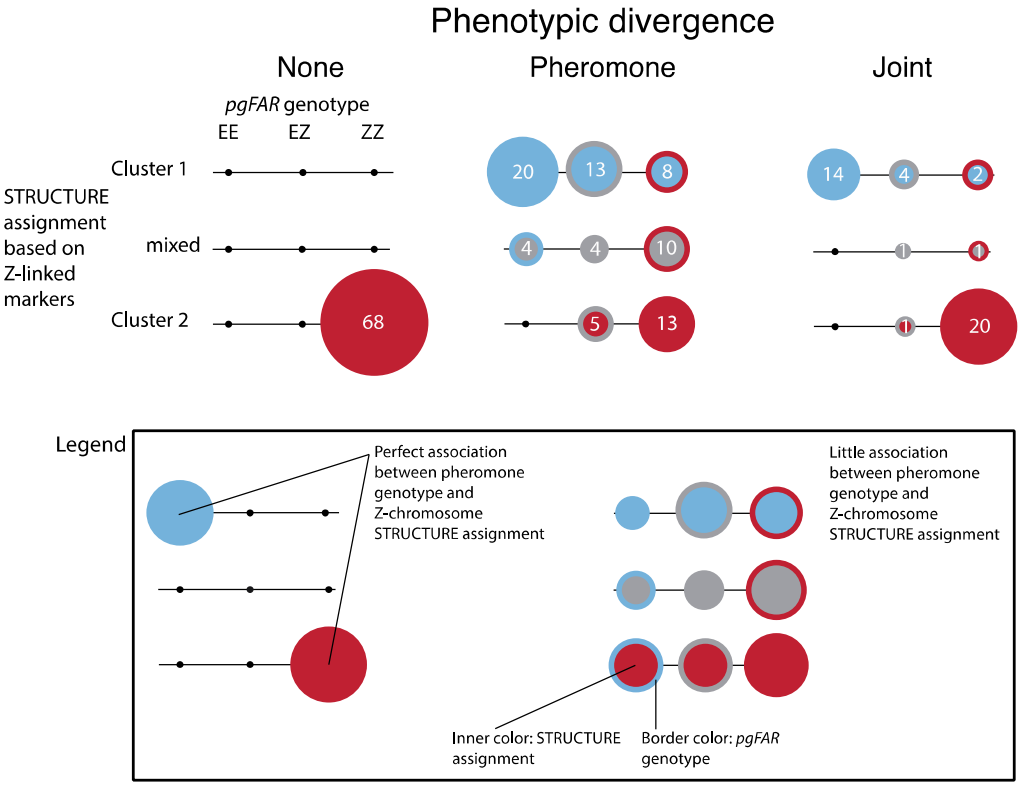

Figure 1.5. The accumulation of barrier traits leads to increase concordance between *pgFAR* and STRUCTURE clade assignment. Bead-and-string plots depict association between sample genotypes at *pgFAR*, the autosomal locus controlling pheromone production (border color) and Z chromosome marker-based STRUCTURE assignment (inner color). Samples were considered E-clade if they were assigned to an E-strain associated STRUCTURE background at a 0.9 threshold or above, mixed if between 0.1 and 0.9, and Z-clade if below 0.1 (see Figure 1.4). Number of individuals listed inside beads. Comparisons are between populations lacking differences in barrier traits (None; Z/Z), or pheromone signaling (Pheromone; BE/BZ), or both phenology and pheromone (Joint; BE/UZ).

We evaluated Z chromosome admixture relative to allelic associations at the autosomal gene *pgFAR* and found that sex-linked genotypic clusters 1 and 2 mainly corresponded to E and Z

pheromone alleles, respectively. Moths assigned with high probability to either of the two clusters (Q1 or Q2 ≥ 0.9, N = 168) were usually homogenous for E or Z alleles at *pgFAR* (145; 86%), whereas 25 percent of moths showing evidence of admixture (0.1 < Q < 0.9) had heterozygote *pgFAR* SNP genotypes. We used a simple network to visualize the E genotypic cluster (Q1 > 0.9), the Z genotypic cluster (Q2 > 0.9), and the admixed cluster (0.1 < Q1 < 0.9) separated by *pgFAR* genotype (Figure 1.5). Most moths demonstrating a mismatch between their autosomal *pgFAR* genotype and their genotypic cluster assignment (N = 48) were collected at sites differing only in pheromone (40 of 77) (Figure 1.5). In contrast, sites differing in both pheromone and phenology had fewer discordant individuals (8 of 43). Results from the PCA of the Z chromosome (Figure 1.6, Figure S1.3) were consistent with these results. The first coordinate axis was broadly consistent with *pgFAR* assignment (Figure S1.3), and whereas intermediates in coordinate space of an analysis of collinear markers were primarily admixed individuals at BE/BZ sites (Figure 1.6A), few intermediates occurred when analyzing solely rearranged loci (Figure 1.6B).

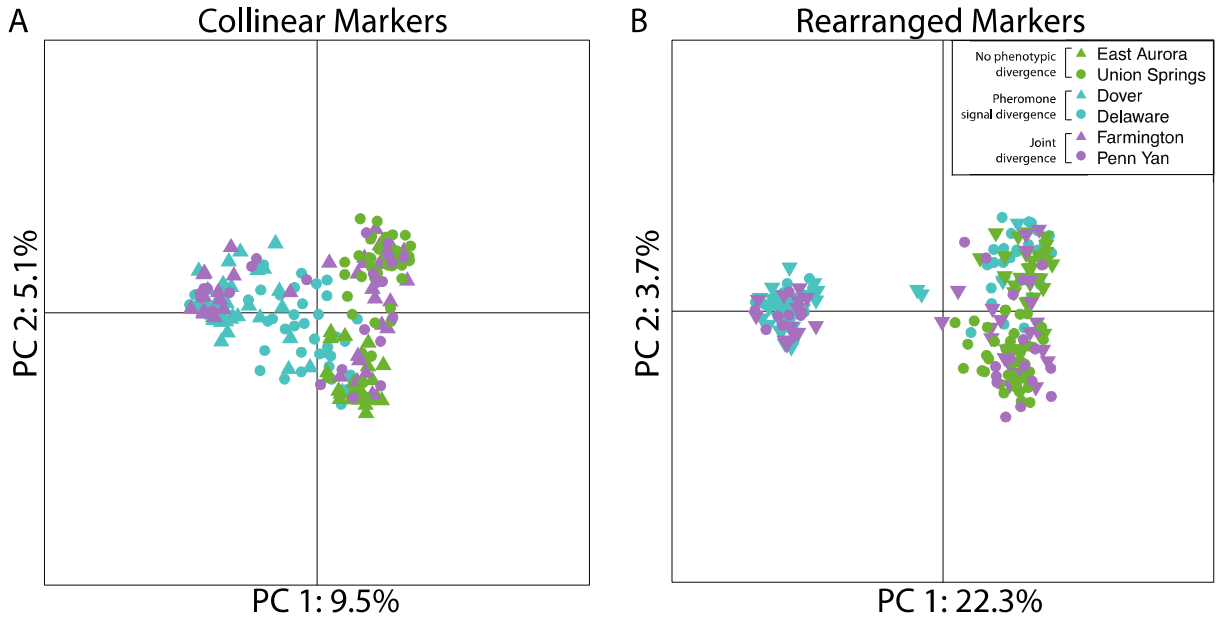

Figure 1.6. Principal components analyses of (A) collinear and (B) rearranged Z chromosome markers. Individual labels are based on geographic location and population type. Percentages indicate the amount of variance explained by that axis. Comparisons are between populations lacking differences in barrier traits (Z/Z; green), or pheromone signaling (BE/BZ; blue), or both phenology and pheromone (BE/UZ; purple).

LD across the entire Z chromosome increased as the number of barrier traits increased. r^2 values were low at locations without barriers (mean r^2 =0.027), but progressively grew with the presence of a single barrier (~120%) (BE/BZ, mean r^2 = 0.06, p = 1.00 x 10^{-20}) and coincident barriers (~310%) (BE/UZ, mean 0.112, p = 5.64 x 10^{-54}) (Figure 1.7). Finally, LD between *pgFAR* and *Pdfr* increased primarily upon joint divergence, as did LD between *pgFAR* and the rest of the Z chromosome (Figure 1.8). Splitting loci by collinear and rearranged location showed that the influence of barrier effects on LD were more pronounced in rearranged intervals (Figure S5).

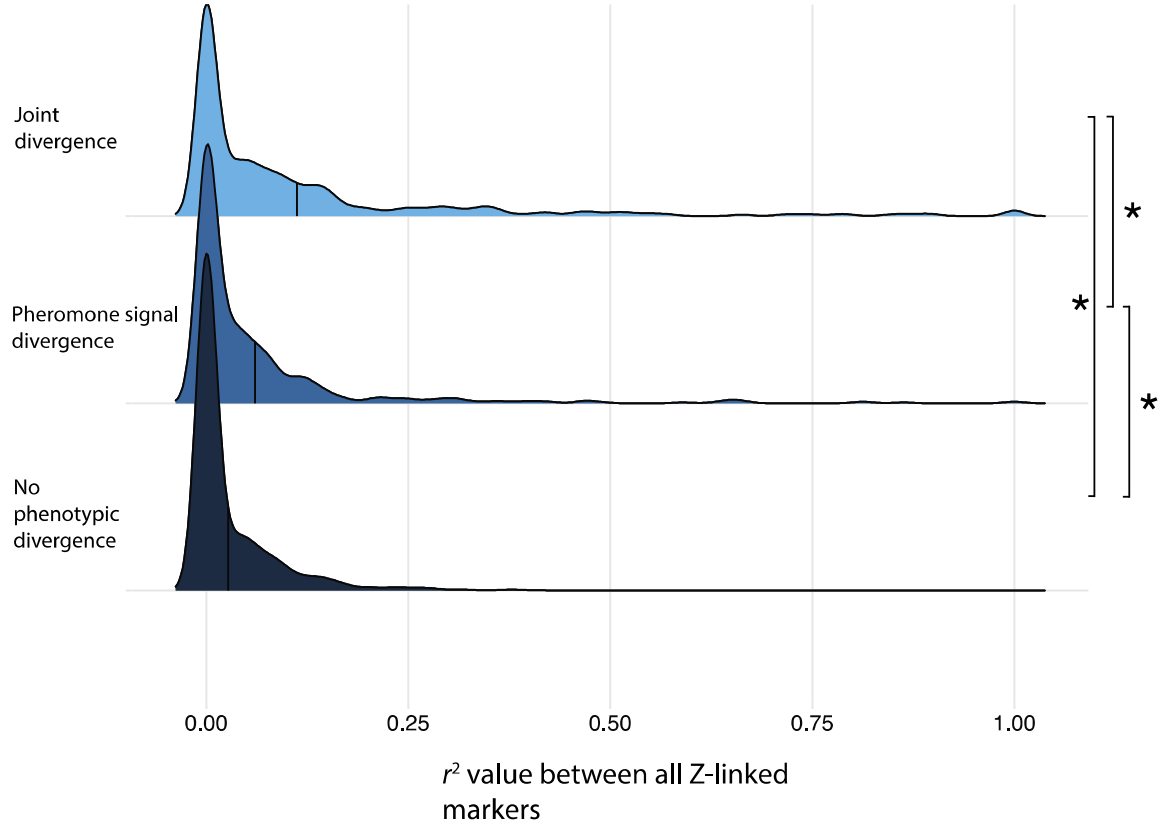

Figure 1.7. Linkage disequilibrium (r^2) between all Z chromosome SNPs separated by population type. The mean value is indicated by a vertical black line. Significantly different distributions measured using Mann-Whitney U tests are marked with asterisks (*). Comparisons are between populations lacking differences in barrier traits (Z/Z), or pheromone signaling (BE/BZ), or both phenology and pheromone (BE/UZ).

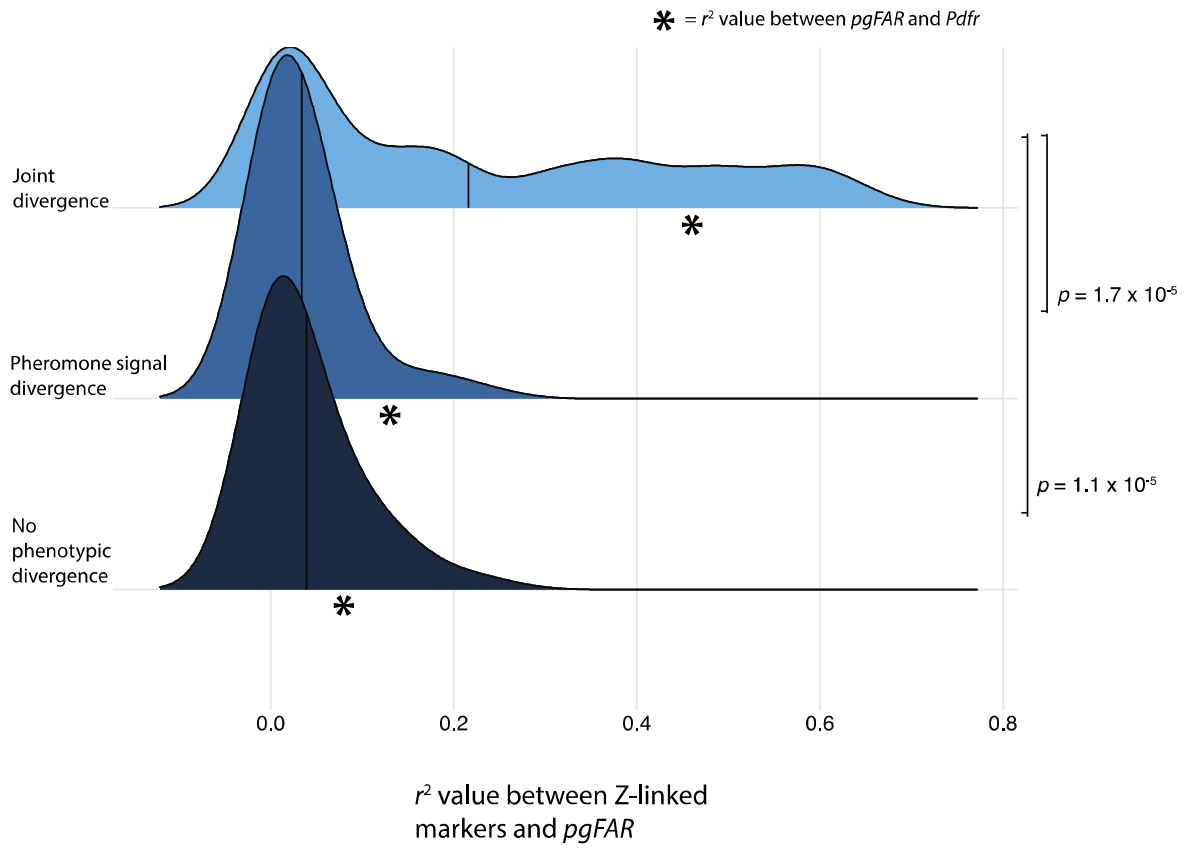

Figure 1.8. Linkage disequilibrium (r^2) calculated between the autosomal *pgFAR* gene and SNPs on the Z-chromosome, separated by population type. The mean value is indicated by a vertical black line. The star depicts mean LD between SNPs at *pgFAR* and the Z-linked *Pdfr* gene. Significance from Mann-Whitney U tests comparing population types are shown. Comparisons are between populations lacking differences in barrier traits (Z/Z), or pheromone signaling (BE/BZ), or both phenology and pheromone (BE/UZ).

DISCUSSION

The origin of distinct genotypic clusters as speciation progresses is thought to depend on 1) the number and genomic architecture of segregating barrier loci, 2) strength of selection or assortative mating on each locus, and 3) features that reduce the rate of recombination among barrier loci (Butlin et al. 2021, Felsenstein 1981, Nosil et al. 2017, Seehausen et al. 2014). The ECB moth system is interesting because many of these factors are at least partially understood. Genomic locations of barrier loci and a Z-chromosomal rearrangement are known, the strength of relevant

barriers have been estimated, and barrier effects among populations can be independent or coincident, allowing insight into how the transition from a weaker barrier(s) to a stronger overall barrier relates to the emergence of exclusive genetic groups.

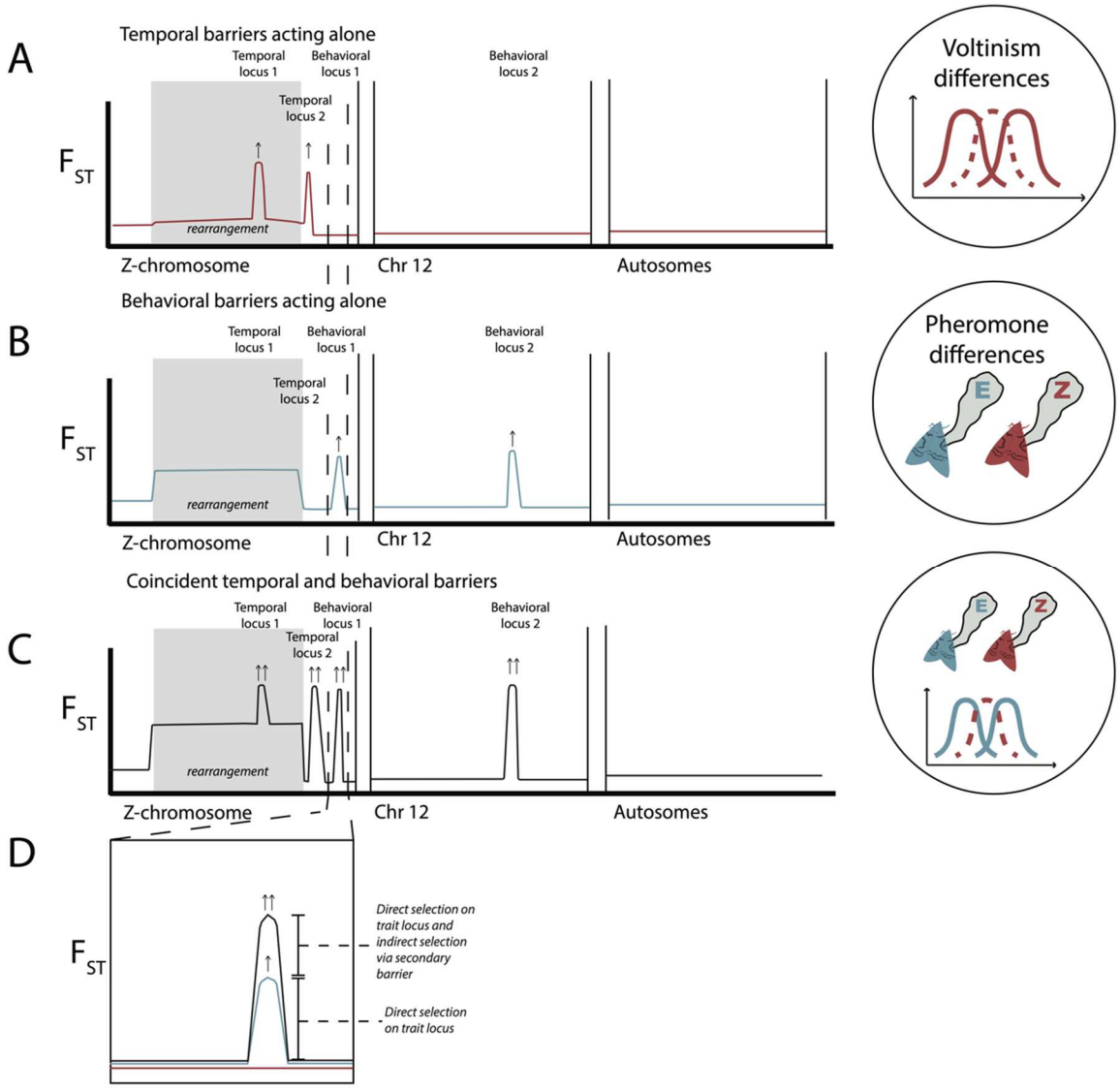

Figure 1.9. A graphical overview of genomic differentiation as barriers to gene flow accumulate. Patterns reflect key findings but are not real data. A. Differentiation at locations differing only in phenology (temporal barrier), B. only in pheromone signaling (behavioral barrier), or C. joint divergence along both phenotypic axes (temporal and behavioral barriers). D. An illustration of how coincidence shapes differentiation of barrier loci by allowing both direct and indirect selection.

Phenology divergence alone

Our findings indicate that a >20 day shift in phenology between univoltine Z (UZ) and bivoltine Z (BZ) populations results largely in regions of differentiation overlapping Z-linked *per* and *Pdfr* loci controlling the timing of diapause termination (Figure 1.2, 1.3A, 1.3C, 1.3D). Lack of widespread genetic differentiation was somewhat unexpected because the estimated strength of assortative mating using Coyne and Orr's (1989) method indicates intermediate reproductive isolation (mean $RI = 0.66$, Dopman et al. 2010). Moreover, as a 'magic trait,' reproductive isolation would be enhanced if hybrid individuals suffer low fitness because their intermediate phenology causes a temporal mismatch with the seasonal environment (Kozak et al. 2017; 2019; Servedio et al. 2011). Among interacting populations, the chromosomal rearrangement spanning ~10 Mb of the Z-chromosome (from ~6 Mb to ~16 Mb) (Wadsworth et al. 2015; Kozak et al. 2017) might conceivably help reduce the number of such intermediate genotypes, as it includes *per* located at ~13.5 Mb (but not *Pdfr* at ~18 Mb) (Kozak et al. 2019; Unbehend et al. 2021). Yet we found limited evidence of enhanced differentiation in this part of the Z chromosome (Figure 1.3C), which would have been expected if a rearrangement differed in frequency and was helping maintain high fitness genotypes by preventing recombination (Kirkpatrick & Barton 2006).

Shallow phenology-associated genetic structure is also seen in other insect systems, for example summer and winter forms of the pine processionary moth (Santos et al. 2007, 2011), as well as marine midge populations living in different phenological niches (Fuhrmann et al. 2021). One possible explanation for weak, localized genetic isolation in these systems is if temporal barriers are commonly conditionally dependent, and therefore unstable. Like other insect species, year-to-

year differences in weather alters breeding cycle phenology in the ECB moth (Beck 1983), which can lead to occasional increased overlap of allochronic mating flights or possibly ease fitness consequences of intermediate hybrid phenology, thereby allowing genomic regions not tightly linked to selected barrier loci to be exchanged. In possible agreement with conditional dependence, Dopman et al. (2010) found the barrier effect to vary by as much as two-fold across years. Despite the potential for divergent selection on phenology to efficiently reduce gene flow, allochrony does not therefore appear to be strong enough to promote divergence by itself, a question recently considered across a variety of taxa (Taylor and Friesen 2017). Instead it may more commonly function when combined with other barriers, such as habitat or resource partitions (Doellman et al. 2018), or when reinforcing pre-existing divergence (Lowry et al. 2008; Møller et al. 2011).

Behavioral divergence alone

There is general agreement that many of the ~160,000 moth species remain separate because of changes in pheromone preference and pheromone blend. However, as with ongoing discussions on phenology, current debate centers on whether this phenotypic axis suffices to initiate the origin of new species (reviewed in Allison and Cardé 2016). The buildup of genome-wide differentiation that ultimately characterizes species is thought to be favored by 'multiple-effect traits' like pheromone signaling in ECB moths, because they are highly efficient at reducing gene flow (Smadja and Butlin 2011). Indeed, the two routes of reproductive isolation that evolved from pheromone divergence (mate choice and behavioral sterility) are predicted to eliminate almost 90% of gene flow between Z and E strains, at least under laboratory conditions (mean $RI = 0.89$; Glover et al. 1991; Dopman et al. 2010). In the field, Unbehend et al. (2021) recently showed that

male moths carry either E type or Z type alleles at *bab* (pheromone preference) and at *pgFAR* (pheromone blend) genes ~80% of the time, indicating coupled pre- and post-zygotic isolation is strong enough in nature to counteract randomization of these physically unlinked genes.

Our findings corroborate prior inferences of selection and assortative mating on *bab* and *pgFAR* barrier loci in field populations (Figure 1.3B, 1.3C), but they go on to suggest that divergence of pheromone signaling alone may not necessarily lead to the emergence of genome-wide differentiation (Figure 1.2). Nevertheless, bivoltine E (BE) and bivoltine Z (BZ) populations do exhibit broad differentiation of chromosomes harboring *bab* and *pgFAR* genes (Z chromosome and autosome 12) (Figure 1.2, 1.3D, 1.3E), as well as long-range LD between sympatric populations (Figure 1.7). Differentiation was unexpectedly high within rearranged regions of the Z chromosome (Figure 1.3D, 1.3E), despite *bab* (~18.7 Mb) being collinearly located ~3 Mb away from the nearest estimated rearrangement breakpoint (Figure S1.1). This might indicate that divergent selection acts on another component of the pheromone system controlled by loci within the rearrangement. If true, two prime suspects are the tandemly repeated odorant receptor loci *OR4* and *OR6* (~12.14 Mb) required for physiological perception of E and Z pheromone components within the male antenna (Koutroumpa et al. 2014). Overall, the apparent stability of chromosome but not genome-wide differentiation implies the sex pheromone system of moths might reduce gene flow enough to allow a partial build-up of differentiation, but this phenotypic axis alone may not lead to substantial progress towards the emergence of well-defined species.

Joint phenotypic divergence

Our analyses of bivoltine E (BE) and univoltine Z (UZ) moths indicate that joint divergence along multiple phenotypic axes (phenology and pheromone signaling) is associated with intensified genetic differentiation. The expected progression towards enhanced coupling as behavioral and temporal barriers coincide in sympatry was detected as elevated LD between unlinked behavioral barrier and temporal barrier genes (*pgFAR* and *Pdfr*) (Figure 1.8). Differentiation at temporal barrier loci (*per*, *Pdfr*) increased in the presence of behavioral divergence (Figure 1.3A, phenology vs. joint divergence). Likewise for the behavioral barrier locus *bab* in the presence of phenology divergence (Figure 1.3B, pheromone vs. joint divergence). Unlike *bab*, differentiation was not elevated at *pgFAR* (Figure 1.3C, pheromone vs. joint divergence), a result likely explained by exclusion of *pgFAR* heterozygotes when creating phenotypically divergent population pools for sequencing. As shown by the amplicon data (Figure 1.5), *pgFAR* heterozygotes are more than twice as frequent at sites differing only in pheromone signaling (BE/BZ), compared to sites that also differ in phenology (BE/UZ). Therefore their removal may have artificially inflated estimated *pgFAR* genetic differentiation in BE/BZ comparisons (e.g., Landisville, Dover). In contrast, levels of differentiation at this locus in comparisons of jointly diverged sites may be more reflective of reality, since heterozygotes are relatively rare (e.g., Rockspring, Harner).

The wider genomic landscape appears to be similarly influenced by joint divergence. On the Z chromosome, both long-range LD (Figure 1.7) and inter-chromosomal LD with *pgFAR* (Figure 1.5, 1.8, S1.4) increased upon the coincidence of behavioral and temporal barriers in sympatry. Furthermore, whereas evidence of reduced admixture (or greater clustering) on the Z was mainly

limited to the rearrangement when sympatric populations differed in pheromone signaling, this signature expanded into collinear regions when sympatric populations also differed in phenology (Figure 1.4, Figure 1.6). Finally, overall differentiation of the two chromosomes containing the four barrier loci (Z chromosome and autosome 12) increased by more than two-fold (Figure 1.2), although differentiation is disproportionate on the Z chromosome where coupling leads to a 'large Z-effect' and divergence of three barrier loci (*per–Pdfr–bab* complex), compared to just one on chromosome 12 (*pgFAR*).

All these features are congruent with prior estimates of cumulative reproductive isolation. When differences in sex pheromone communication as well as phenology act as barriers, their cumulative effects are estimated to result in almost complete reproductive isolation (mean $RI = 0.96$, Dopman et al. 2010). The process underlying heightened genetic differentiation is unlikely to simply involve a sudden increase in fitness effect on each barrier locus (direct selection). Instead, as barrier loci increase their non-random associations during coupling, a higher mean fitness differential can be achieved between alternative multi-locus genotypes (indirect selection). Since each locus can have its own fitness effect and also be indirectly influenced by the fitness effect of other loci (Barton 1983), overall selection is more effective at each locus, as is the overall barrier to gene flow at each locus, making it more difficult for neutral alleles to cross population boundaries (Barton and Bengtsson 1986).

Conclusions

We argue that our study provides support for the coupling hypothesis of speciation (Felsenstein 1981; Smadja and Butlin 2011; Seehausen et al. 2014) and helps close a gap between theoretical predictions of this hypothesis and empirical systems. Our population comparisons show that whether barriers are independent or coupled make important contributions to a hallmark of the speciation process, the build-up of genomic differentiation. Genetic distances fit expectations from estimates of reproductive isolation, with locus-specific differentiation associated with the weakest barrier (temporal), differentiation across much of a chromosome with the barrier of moderate strength (behavioral), and chromosome-wide differentiation when the two barriers operate cumulatively. These results demonstrate that the joint action of multiple, coincident barrier effects lead to levels of genomic differentiation that far exceed those of single barriers acting alone, consistent with theory in which coupling allows indirect selection to combine with direct selection and thereby result in a stronger overall barrier.

Finding chromosome-wide but not genome-wide differentiation when populations diverge along multiple phenotypic axes suggests that the overall barrier upon coupling may only partly outweigh gene flow and recombination. Thus, a critical tipping point or threshold of coupling that could allow for sudden differentiation of most of the genome was not reached (Nosil et al. 2017). When barriers are coincident, much of the genome appears well mixed and moderate genetic differentiation (F_{ST} = 0.05 – 0.15) is observed for only two chromosomes containing barrier loci, making further divergence challenging without evolution of either more barrier loci or stronger selection/assortative mating. Simulations demonstrate that, in certain cases, many divergently

selected loci may be required for genome-wide differentiation (Feder et al. 2014, Flaxman et al. 2014). Some empirical studies seem to support these predictions, including Midas cichlids, where a transition from locus-specific to genome-wide divergence coincides with a transition from simple to polygenic barriers (Kautt et al. 2020). Strengthening of the overall barrier in the ECB moth could in time be facilitated by the evolution of highly polygenic traits, perhaps in the form of intrinsic post-zygotic barriers. However, it is also possible that many intermediate stages of speciation seen in nature, including the ECB moth, are either stably maintained or even temporary, and that barrier traits may not remain coincident should ecological or geographic shifts take place (Seehausen 2006). If so, then coupling and current population differentiation will disappear or rebound according to external environmental factors.

Understanding how and why barrier traits evolve and become coincident or disappear remains as a critical problem for speciation biology, and few empirical systems have been able to directly assess coincidence of known barrier traits in different stages of coupling. Future work on this study system will reconstruct the evolutionary and demographic history of this group and provide a framework for inferring the timing and selective history of genes underlying barrier phenotypes. Ultimately, we must reconstruct not only the sequence of evolution of individual barriers, but also the timing and causal factors responsible for their coupling. Although byproduct explanations have been proposed that do not require coupling itself to be selectively favored (Butlin and Smadja 2017), adaptive coupling seems likely in the ECB moth and can be experimentally studied by measuring fitness of full allelic combinations in nature. Indirect approaches are equally promising for identifying factors associated with changes of coupling in nature. Since it is now possible in this system to precisely estimate population differentiation and the extent of coupling of all major

barrier loci (*bab, pgFAR, per, Pdfr*) across various spatial, ecological, and historical contexts, the causal links that establish barrier coupling and facilitate neutral divergence during speciation appear to be within reach.

CHAPTER TWO

PHENOLOGY OF EUROPEAN CORN BORER MOTHS: RANGE EXPANSION ALONG A LATITUDINAL GRADIENT AND ITS EFFECT ON THE GENOME

ABSTRACT

Seasonal variation in climate creates selection pressure for organisms on phenological traits that control timing of life history events. In their introduced range in North America, European corn borer moths *Ostrinia nubilalis* are active during summer months and overwinter after entering diapause as larvae. They demonstrate variation in post-diapause developmental rate, which leads to variation in the number of generations that occur each year. Depending on geographic locality, different selective forces are expected to operate on loci that underlie post-diapause developmental timing. To identify whether the genome reflects signatures of natural selection at such phenological loci, we studied the genomes of moths from populations along a latitudinal gradient in the Midwestern US. We analyzed ten field collected populations and one lab colony using pooled whole genome sequencing. We found genome-wide signatures that were reflective of a large expansion in both range and population size. We demonstrated that key life history loci *period* and *Pdfr* showed signatures that were consistent with selection and shaped by neutral demographic processes. We then identified other candidate loci that were associated with environmental variables or showed signatures of balancing selection: either through an excess of nearby intermediate frequency alleles or through oscillations in SNP frequency along the transect. Our results provide a basis for future investigation of potential candidates for adaptation to novel climatic gradients.

INTRODUCTION

Theory and empirical studies show that the expansion of a species into a novel range shapes the organism's genome (Crow & Kimura 1970, Barton & Hewitt 1985, Excoffier 2009, Peter & Slatkin 2013). As ranges expand, these genomic changes occur as a result of demographic shifts such as population bottlenecks and rapid population growth. As the populations expand, they are expected to show excesses in rare and low frequency alleles, reflected by genome-wide reductions in summary statistics such as Tajima's D. Meanwhile, selection on traits allowing local adaptation to environmental conditions could leave distinct genomic signatures such as local reductions in nucleotide diversity and increases in localized genomic differentiation between populations. Research has demonstrated genetic responses to range expansion in many taxa, including invasive insects (Lindroth et al. 2012), birds (Pruett et al. 2011), and even humans (Tishkoff et al. 2009; Sousa et al. 2014). These studies and others have shown that the interplay between neutral and selective processes can be complex and merits study in more taxa.

In relation to the responses to environmental conditions revealed in the genome, selection acting on the genes that control adaptation to climatic variation will generate allele frequency changes over the species range reflecting local conditions. In the case of continuous environmental variables, allele frequencies may vary continuously across the range, forming the basis of a genetic cline (Slatkin 1970, Barton & Hewitt 1985). Genetic clines have been demonstrated in classic studies such as those investigating the alcohol dehydrogenase gene (Berry and Kreitman 1993)

and the *couch potato* gene (Schmidt et al. 2008) in *Drosophila*.

Organisms that exhibit distinct life history stages through the year and that occur in locations where there is seasonal variation in climate are shaped by selection on the precise timing and characteristics of these stages. This is dependent on local temperature and other climatic conditions (Varpe 2017). In temperate areas, for holometabolous insects, there are distinct life stages that overwinter, and a certain number of generations will utilize the active period through the spring, summer and fall. Over a latitudinal gradient, with the concomitant smooth climatic change, there is likely to be change in the number of generations during the active period (Roff 1980). For a species that expands its range into such regions, change in the number of generations is expected to introduce selective forces depending on the interaction between generation number and local environmental conditions such as climate and season length. Here we investigate genomic and local signatures of range expansion and selection in the European Corn Borer (ECB) moth, *Ostrinia nubilalis* Hübner, a holometabolous insect.

ECB moths were introduced into eastern North America from Europe in the early 20th century, and have subsequently expanded westward to the Rocky Mountains (Smith 1920; Caffrey and Worthly 1927; Klun and Cooperators, 1975). Across this North American range, ECB moths display distinct numbers of generations during each growing season, occurring as early as April and as late as October, with variation across the range (Showers et al. 1975). In the summer, adult moths emerge from pupation to mate and lay eggs. If the season length is long enough, these eggs may hatch, develop, pupate, and another generation may occur. However, if temperature and photoperiod cues are met, the larval ECB will enter diapause, a period of reduced metabolic rate,

to overwinter in their final larval instar. Following diapause-breaking cues in the spring, larvae will resume development and pupate, then complete their life cycle as adult moths (Dopman et al. 2005; Levy et al. 2015; Kozak et al. 2019).

Control of developmental rate following a vernal break in diapause is governed by two major sex linked factors that contribute variation in the phenotype PDD (post-diapause development) (Glover et al. 1991; Calvin et al. 1994; Zaman et al. 2008; Dopman et al. 2010). This variation leads to up to a 30-day difference in developmental timing, and thus a corresponding offset in the timing of emergence of adult moths (Wadsworth et al. 2013; Kozak et al. 2019). Individuals in populations with short post-diapause development phenotypes take ~14 days to pupate after exposure to diapause-breaking cues allowing two generations per year (bivoltine) in northern parts of the ECB moth range. Individuals in long PDD populations take ~45 days to pupate and only have one generation (univoltine) in northern areas. Since traits associated with diapause are responsible for patterns of voltinism, the number of generations per year, across the species range, variation in these traits corresponds to variation in the number of discrete generations at a given locality.

Selection is expected to govern the frequency of the fast and slow developing phenotypes, allowing for ECB populations to fit in the optimal number of generations per season at any locality. Theory suggests that when there is selection on the basis of optimizing discrete traits in relation to a continuously varying environmental factor (here, growing season length), the result can be non-linear change along the environmental gradient (Roff 1980). Data on ECB moths using a small number of markers suggest that some genetic loci may exhibit such non-linear genetic clines (Levy

et al. 2015). There is reason to expect PDD phenotypes to vary in this manner. Variation in PDD may be driven by selection on moths to match generations to the length of the growing season. As the length of the growing season increases with a decrease in latitude, more generations are possible, and short developmental time may be favored so that an extra generation may be completed before winter (Van Dyck et al. 2015). As latitude decreases slightly more, longer growing seasons may not allow for another generation, but rather for a shift to the long developmental timing phenotype while maintaining the same number of generations. This may indicate that short and long PDD phenotypes and genotypes may alternate across latitudes. Recent work has highlighted the action of two interacting clock genes, *period* and *Pdfr*, both located on the Z (sex) chromosome, in controlling PDD (Dopman et al. 2005, Levy et al. 2015, Kozak et al. 2019). These genes are among a catalog of circadian clock genes such as *clock*, *period*, *timeless*, and *pigment dispersing factor* that are involved in daily timing and have been hypothesized to be involved in seasonal timing of life-history, since they often vary among populations that differ in phenology (Meuti et al. 2015; Gotthard & Wheat 2019; Abrieux et al. 2020; Lindestad et al. 2021; Pruisscher et al. 2021).

This previous work on *period* and *Pdfr* alleles was conducted on samples drawn from northeastern populations (New York and Pennsylvania, Kozak et al. 2019), which maintain multiple alleles that correspond to different PDD phenotypes, but contemporary ECB populations are widespread throughout eastern North America. Because ECB moths rapidly expanded their range following their introduction to North America (Smith 1920; Caffrey and Worthley 1927; Klun and Cooperators 1975), establishing populations across a broad latitudinal span in the Midwest, signatures of this expansion are expected to be left in their genomes (Excoffier et al. 2009). Further,

adaptation to environmental conditions at different latitudes may also shift allele frequencies at underlying trait loci such as *period* and *Pdfr*.

To follow up on these aspects of range expansion and environmental selection in ECB months, the present study examines allele frequencies of SNPs across the entire genome to search for demographic and selective signatures following range expansion along a climatic gradient. We sampled the Midwestern US, where populations have occupied their current localities for fewer than 90 years. We focus on the allele frequencies of *period* and *Pdfr,* key phenological timing loci, along a latitudinal transect of these Midwestern populations (Figure 2.1). The data we use are pooled genomic sequences of 10 populations from the transect as well as one lab colony that breeds true for a univoltine (long PDD) phenotype. This univoltine colony was leveraged for comparative purposes, since we know the causal alleles underlying PDD shifts and we know the phenotype of this colony. If PDD is shaped by selection during range expansion, we also expect that PDD-associated phenotypes should respond as well, showing changes in allele frequencies at underlying genotypes. Changes are also expected at loci that respond to latitudinal factors such as degree days (a statistic incorporating time-adjusted climatic conditions: see methods). We also wish to test a potential role for a "step cline" in which two or more populations differ from the rest. Early work on this suggested the presence of a distinct ecotype found only in the northernmost localities (Sparks et al. 1966a, Showers et al. 1975). In these populations, photoperiod cues are hypothesized to have a stronger effect than temperature cues in inducing the diapause break. In more southern populations, moths rely on a more balanced combination of temperature and photoperiod cues to break diapause.

Our study aims to inform us of the genetics underlying range expansions and the selective control of life history phenotypes in ECB moths. In particular, we expect to observe genome wide patterns consistent with population growth as the species spread as well as shifts in allele frequencies at loci controlling PDD timing along the latitudinal cline. These shifts are expected from the action of selection optimizing the number of generations of each population to a local optimum dictated by the growing season length. We pose the following questions: (a) How have demographic processes such as range and population size expansion shaped the genomes of ECB moths in the Midwest populations? (b) Can we demonstrate the selective maintenance of alleles at *period* and *Pdfr* along the latitudinal gradient? (c) What other loci may be targets of selection, due to a direct response to environmental factors varying along the latitudinal gradient, or due to covariation with known changes in life history traits such as voltinism?

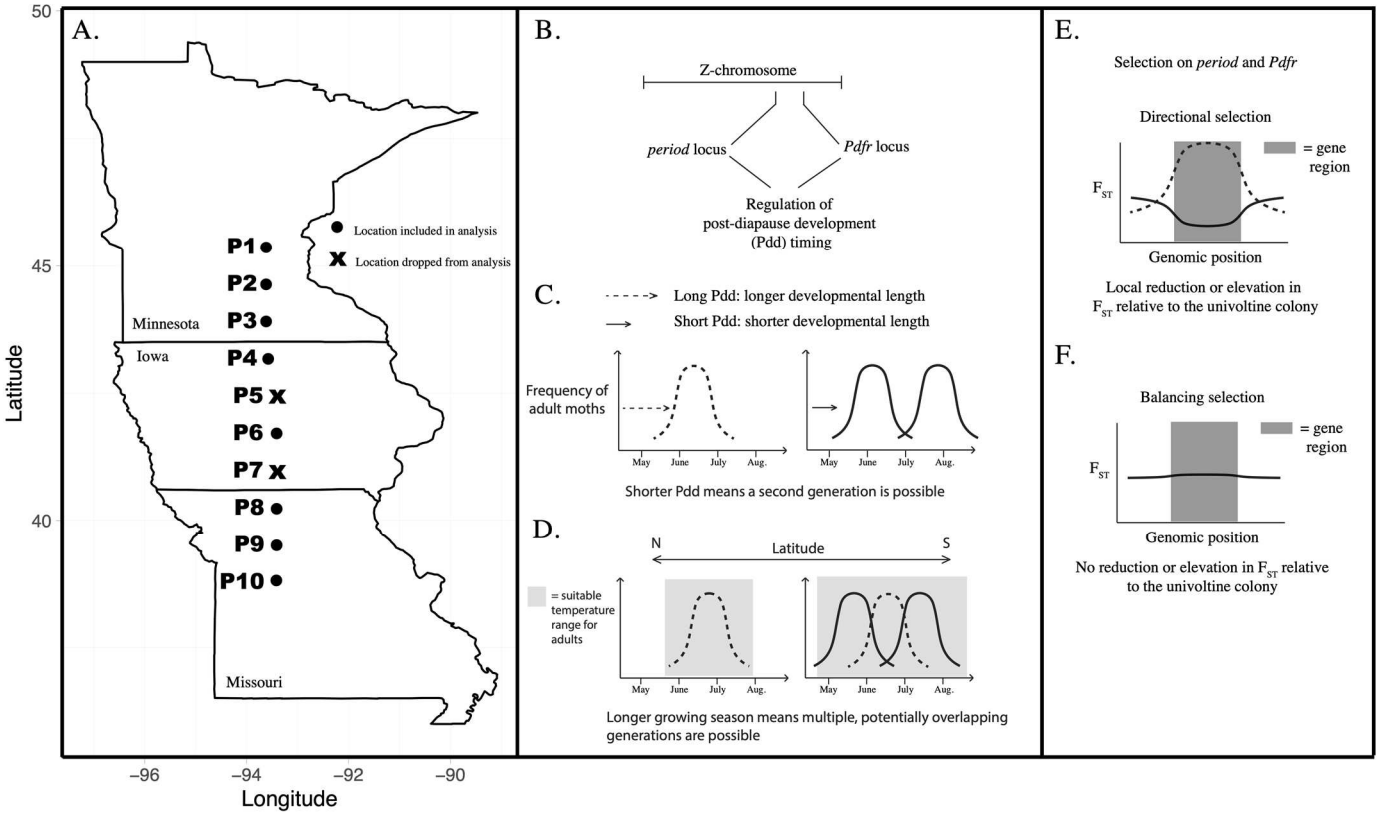

Figure 2.1. A. Map of collection localities of European corn borer moths along a latitudinal gradient in the Midwestern US. Each dot represents a sampled population, from which a pool of ~60 individual moths was sequenced to ~30x coverage. Populations from locations marked with an X did not clear post sequencing quality filters and were removed from subsequent analysis. B. Locations of key life history loci *period* and *Pdfr* on the Z chromosome. Each contribute to post diapause development (PDD) timing. C. Variation in PDD governs the number of accessible generations each growing season. Shorter PDD lengths allow more generations (Dopman et al. 2005, Levy et al. 2015, Kozak et al. 2019). D. Growing season lengths increase as latitude decreases, allowing for more generations and potential maintenance of multiple life history phenotypes. E, F. Expectations for potential signatures of selection on life history loci. If one allele is favored at a locality, F_{ST} between that population and the univoltine colony may be decreased or elevated (shown by a solid and dashed line, respectively), depending on the direction of selection. If both alleles are maintained via balancing selection, F_{ST} may be intermediate relative to the univoltine colony, and other statistics such as Tajima's D may be more indicative of selection.

METHODS

Sample collection

Adult ECB moths were collected in 2005 along a latitudinal gradient in the Midwestern US, from central Minnesota (45.3° N) to northern Missouri (38.8° N) (Figure 2.1). Collections were made using pheromone traps placed at 50-mile (80.5-km) intervals over this range, resulting in 10 population samples across roughly 500 miles (804.7 km) and 4.5 degrees latitude. Specimens were stored at -20° C prior to sequencing. We also utilized DNA from lab colony specimens that had been bred for long PDD phenotypes (see Kozak et al. 2019 for details), treating this as an eleventh population.

Pooled sequencing

At Cornell University we sequenced complete genomes from 11 pooled samples each representing ~60 ECB moths from one population. For the transect population samples, DNA from each moth specimen was sent from the Dopman lab at Tufts University, having previously been extracted by Kim et al. (2009). The DNA concentration of each moth sample was measured using a Qubit fluorometric quantification and these samples were diluted prior to pooling to assure DNA from each sample was equally concentrated within each pool. For the univoltine colony, tissues of ~50 adult male moths were pooled and ground to a powder in liquid nitrogen. DNA was then isolated from this pooled sample. For all pooled DNA samples, approximately one microgram of genomic

DNA was sheared to 700 bp with a Covaris S2 sonicator. Sheared DNA was treated with 1x Ampure XP and eluted in 30 µl Qiagen 0.5 x AE buffer. Samples were end polished with T4 polynucleotide kinase and T4 DNA polymerase, adenylated with Taq polymerase, and ligated to an Illumina Truseq +A adapter. Individual samples were indexed by PCR, pooled, and sequenced on an Illumina NextSeq 500 (2 x 150 bp reads) across two lanes at the Cornell Institute of Biotechnology Genomics Facility. To avoid a bias induced by either lane, we split the 11 pooled samples so each was sequenced to lower coverage across both lanes. Sequence data were analyzed for quality by FastQC (Andrews 2010) and multiQC softwares.

We processed reads using the GATK pipeline (Van der Auwera et al. 2013) using default and recommended parameters, with a slight modification to recombine the pooled sample split between the two lanes once adapters were removed and reads were trimmed. Sequence data obtained for each pool were trimmed, removing adapters, low quality regions, and unpaired reads. The data were aligned to the *O. nubilalis* draft reference genome (GenBank: PRJNA534504; BioSample SAMN11491597; accession SWFO00000000). The pooled samples from two localities (P5 and P7) did not meet the sequencing thresholds and read quality scores of the other pooled samples and were dropped from subsequent analysis. A total of 13,031,157 SNP markers were included in the primary dataset following quality filters. We used Popoolation 2.0 software (Kofler et al. 2011) to generate allele frequencies and calculate π, Tajima's D, and pairwise F_{ST} between each of the 9 populations studied. Tajima's D is a metric used to infer and distinguish the action of balancing selection, indicated by an overrepresentation of intermediate frequency polymorphisms, and directional selection, indicated by an overrepresentation of low frequency polymorphisms. We utilized comparisons between each transect population and the univoltine Z (UZ) colony to identify

genomic regions with similar allele frequencies and nucleotide diversity. We focused on scaffolds containing a known life history locus, *period* or *Pdfr*, and attempted to distinguish localized genomic changes in F_{ST}, Tajima's D, or π from genome-wide estimates of these summary statistics. We utilized 1000bp sliding window averages across each scaffold throughout the entire genome, following best practices protocols established in Kofler et al. (2011). Following these estimates, we filtered windows to those that contained at least 20 SNPs and a minimum fraction of the window covered by reads that met quality filters greater than 0.2. Scaffolds on the Z chromosome were identified and ordered as described in Kozak et al. (2019) and Kunerth et al. (2022) (Table S2.1). To compare summary statistics between genomic regions, we applied Mann Whitney U tests.

To test for geographic effects on genetic distance, we first calculated genetic distance using the method of $F_{ST}(1 - F_{ST})$ established in Rousset 1997. We then used the R package vegan v2.6-2 (Oksanen et al. 2013) to conduct separate partial Mantel tests for Z-linked and autosomal datasets.

To search for signals of balancing selection, we used the program Betascan (Siewert & Voight 2017, 2020) on each transect population and the UZ colony. Betascan utilizes counts of intermediate frequency SNPs within a genomic region surrounding a focal SNP to detect departures from neutrality consistent with the maintenance of balanced polymorphisms. We used the folded site frequency spectrum and default parameters. Since Betascan must be run independently on each chromosome or, in our case, genomic scaffold, we ran separate runs for each scaffold and concatenated the results afterwards. We separated Z-linked and autosomal markers and searched each dataset for outliers using a strict cutoff of the top 0.01% of Beta scores.

We then used an annotated ECB genome (see Kozak et al. 2019) to identify matches between these candidates and genes of known or inferred function.

We used the program pool-HMM v. 1.4.4 (Boitard et al. 2013), which utilizes each nucleotide site in a pooled sequencing dataset to infer the locations of putative selective sweeps in the scaffolds containing *period* and *Pdfr*. Pool-HMM uses a hidden Markov model with each nucleotide site modeled as a hidden state of "neutral", "intermediate", or "selection". As suggested by the authors, we utilized a range of k-values, the transition probabilities between hidden states between 0.001 and 0.1 in independent runs for both the *period* and *Pdfr* scaffolds. We applied the model to each transect population and the UZ colony.

Using the software BayPass 2.2 (Gautier 2015), we scanned the pooled genome-wide allele frequencies for alleles that a) show signatures consistent with departures from the background population covariance structure, suggestive of selection, and b) for alleles that covary with two environmental variables and two phenotypic variables. We tested for significant associations with four population-specific covariables after accounting for the background population structure (N = 317,856 tests for Z-linked markers and N = 9,484,284 tests for autosomal markers). These covariables were tested independently from one another, as they may themselves covary. The population covariables were:

- **Step cline**, a break between the northernmost two locations and the others, corresponding to a previously documented change in ecotype (Showers et al. 1975).

- **Number of generations**, an estimate of the average number of generations at each locality in the range, based on trapping data on the host plant (sweet corn).
- **Degree days**, a proxy statistic estimating the range of suitable days for ECB moths at a location over a growing season. A degree day for European corn borer is stated as degrees above 50°F over a 24-hour period. For example, if the average temperature for a 24-hour period was 70°F, then 20 degree days would have accumulated (70 - 50 = 20) on that day.
- **Latitude,** which serves as a proxy for many environmental conditions such as temperature, photoperiod, host plant variation, insecticide use, and more.

The degree days and latitude factors were coded as continuous variables for the eight transect populations, while the step cline and number of generations were coded as categorical. Both types of covariables were scaled and adjusted to a mean of zero before running.

The final dataset for BayPass included eight transect populations across 450 miles (720 km) and 4.5 degrees of latitude with four covariates. We ran scaffolds linked to the Z-chromosome separately from autosomal-linked scaffolds. To identify SNPs with significant departures from the background covariance matrix, we searched for markers with significant XtX values. To assess statistical significance, we generated 120 sets of 100,000 SNPs based on omega values from the empirical dataset and ran BayPass again on these simulated datasets. We considered SNPs in the empirical datasets with XtX values higher the 95th percentile in the simulated set (recommended in Gautier 2015).

We searched for data that showed significant associations with covariables by filtering SNPs with high BayesFactors (as measured by BF.dB) relative to the four covariables. We used a cutoff of BF.db > 15 for both the Z-linked and autosomal datasets in keeping with recent similar studies (Chaturvedi et al. 2018; Miller et al. 2019; Sokolkova et al. 2021). BayPass was applied separately on datasets containing 79,464 Z-linked SNPs and 2,371,071 autosomal SNPs. Each population along the transect corresponded to a particular covariable value which impacted the BayesFactor score if alleles at that locality were found to covary with the environmental/phenotypic covariable. We then mapped statistically significant hits to the annotated ECB genome to identify genes underlying putatively selected phenotypes.

A previous study (Levy et al. 2015) had demonstrated SNPs that varied in allele frequency in a sinusoidal pattern across the range, consistent with varying selective strength on alternative voltinism phenotypes. We followed this approach but for genome wide markers, using R package JTKcycle 3.0 (Hughes et al. 2010) to scan our dataset for sinusoidal patterns of allele frequencies. We allowed a search space of period length corresponding to a complete cycle between every third population to a maximum of a complete cycle over six transect populations. We filtered the resulting dataset by removing any oscillation with an amplitude of less than 0.1, as these are likely the result of sampling error and noise, rather than a biological signal. We also applied a p-value cutoff of 0.5 after applying a Bonferroni correction for multiple tests.

For all final datasets (Betascan, JTKcycle, BayPass), we utilized BEDTOOLS (version 2.30.0) to compare to an annotation of the ECB genome using the 'intersect' command in order to find genes that overlapped with significant hits. We also considered values at genes of known phenological

importance to ECB moths such as *period* and *Pdfr*.

RESULTS

Genomic level statistics

We used pooled whole genome sequencing to analyze 8 populations of ECB moths along a latitudinal cline in the midwestern United States (Figure 2.1) as well as one lab colony. The 8 populations along the latitudinal transect did not display substantial substructure. We used the perl script fst-sliding.pl from the Popoolation2 package (Kofler et al. 2011) to calculate F_{ST} in 1000 bp sliding windows across all scaffolds. The mean pairwise F_{ST} between populations along the latitudinal cline was 0.039 for autosomal markers and 0.032 for Z-linked markers, lower than recent pairwise estimates of ECB populations in Pennsylvania and New York (Kozak et al. 2019, $F_{ST} > 0.05$; Kunerth et al. 2022, $F_{ST} > 0.04$). We did not find evidence of highly differentiated outlier loci between any pairwise comparison of transect populations (Figure S1). Mean pairwise F_{ST} between the inbred UZ colony and populations along the cline was 0.199 for autosomal markers and 0.162 for Z-linked markers (Tables S2.2 and S2.3). We should note that Pool-seq has been estimated to show three to five times higher F_{ST} estimates than other next-generation sequencing methods (Dorant et al. 2019), so we encourage caution in comparing the absolute estimates to other studies.

To test for the effects of geographic distance on genetic structure, we used a partial Mantel test using the R package vegan (Oksanen et al. 2013). We compared pairwise genetic distance

estimates for each transect population. These tests were applied separately to autosomal and Z-linked markers. In each case, no significant differences were found ($p = 0.453$ for autosomal markers and $p = 0.414$ for Z-linked markers).

We calculated π for each transect population and the UZ colony across the entire genome using the Variance-sliding.pl script in the Popoolation package (Kofler et al. 2011). The mean π value for populations along the cline was 0.018 for autosomal markers and 0.008 for Z-linked markers, suggestive of low overall levels of nucleotide diversity. Consistent with expectations, the UZ colony, which had been selectively maintained for the univoltine phenotype, had even lower levels of diversity: 0.006 for autosomes and 0.001 for Z-linked SNPs.

In Figure 2.2 we show Tajima's D for 1000 bp windows across the entire genome for each transect population and the lab colony. Distributions of the D statistic were separated by autosomal markers and Z-linked markers. Mean Tajima's D was slightly negative for both autosomal markers (-0.169) and Z-linked markers (-0.239) for populations along the cline. It was marginally negative in the UZ lab colony for autosomal markers (-0.022) and strongly negative for Z-linked markers (-0.414), the location of the genes that were the targets of selection during maintenance of this colony. This Z-linked mean value was driven by highly negative values of Tajima's D, since the median value was 0.0 (Figure 2.2).

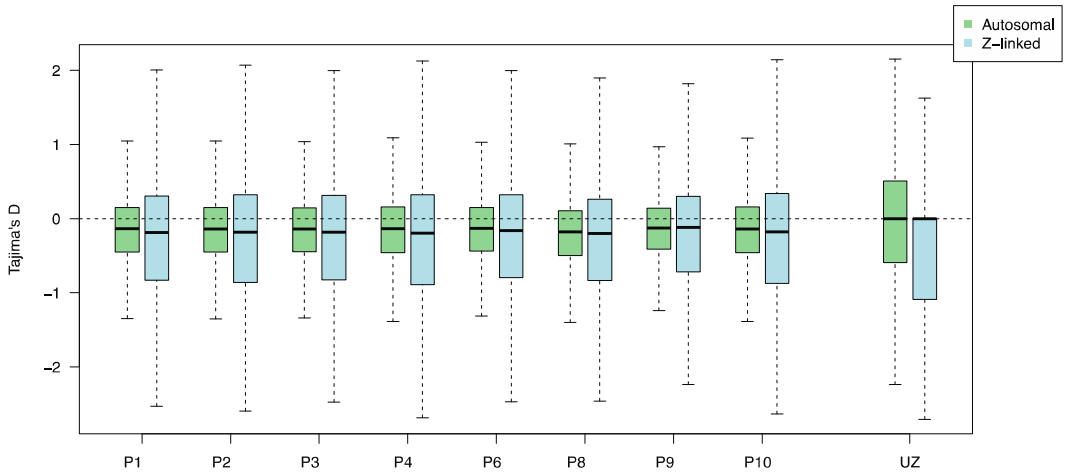

Figure 2.2. Box-plots of Tajima's D based on all autosomal and Z-linked loci. Solid line shows median, box shows interquartile range and error bars show absolute range of values.

Locus specific tests

The key barrier loci *period* and *Pdfr* showed similarities in direction but differences in magnitude when assessing estimates of allele frequency and diversity patterns in the populations along the latitudinal gradient. The comparison with the UZ lab colony was important particularly to estimate pairwise F_{ST} as a proxy for the frequency of the PDD haplotypes of these genes relating to developmental rate. The expectation is that transect populations that are low in F_{ST} compared to the UZ colony are high in frequency of long PDD alleles.

We found that the Z-linked *Pdfr* locus was extremely similar in F_{ST} for all transect populations relative to the UZ colony (Figure 2.3, column 1). The mean F_{ST} within the genic region was 0.065, while the mean F_{ST} on the scaffold outside the genic region was significantly different at 0.17 ($p = 2.2 \times 10^{-16}$). F_{ST} on the scaffold but outside the genic region was similar to mean pairwise F_{ST} between the UZ colony and transect populations for all Z-linked markers (0.156). At *Pdfr,* reductions in F_{ST} were accompanied by a dramatic reduction in nucleotide diversity (π)

within the genic region (Figure 2.3, column 3). This pattern held for all populations along the transect. The UZ colony also shows a lower nucleotide diversity in the genic region (combined mean π within genic region = 0.0016, mean π on scaffold outside genic = 0.0087, $p = 8.29 \times 10^{-11}$).

Similarly, at *period,* pairwise F_{ST} between transect populations and the UZ colony was lower within the genic region relative to the scaffold regions outside of the gene (within = 0.11, outside = 0.20, $p = 4.0 \times 10^{-09}$) (Figure 2.4, column 1). Relative to those observed at *Pdfr,* less dramatic reductions in nucleotide diversity were observed at *period* (Figure 2.4, column 3; mean π within = 0.0029, mean π outside = 0.0046, p = 0.00014). Despite these within-gene estimates at *period* not showing as low values of π as those at *Pdfr*, they were substantially lower than overall Z-linked estimates of nucleotide diversity among transect populations (mean Z-linked π = 0.0083).

Estimates of Tajima's D showed contrasting patterns for *period* and *Pdfr*. At *Pdfr*, Tajima's D was significantly negative within the genic region relative to the genomic scaffold (mean D within = -1.14, mean D outside = -0.29, $p = 2.3 \times 10^{-05}$) (Figure 2.3). In contrast, at *period,* there was no significant difference between the genic region and the linked scaffold (mean D within = -0.504, mean D outside = -0.359, p = 0.13) (Figure 2.4). For *Pdfr,* the pairwise comparisons between the genic region and the surrounding scaffold for every population along the transect was significantly different. At *period*, no comparison between the genic and surrounding region was significantly different, assessed via Mann Whitney U tests. For the UZ colony, no significant differences between genic and surrounding differences were found for either *period* or *Pdfr.*

The signature of Tajima's D at the *Pdfr* locus was distinct compared with the rest of its scaffold (Figure 2.3, column 2). To visualize this chromosome-wide, we plotted every 1000bp windowed estimate for all Z-chromosomal markers on scaffolds, ordered based on linkage (Table S2.1), for every population (see Figure 2.5 for an example at population P1). A moving average for 100 point estimate windows (fifty 1000 bp estimates on either side of a focal marker: Figure 2.5, red line). The region containing *Pdfr* was clearly distinctive in the context of the broader chromosome patterns of Tajima's D, reaching its lowest moving average along the entire chromosome in the genic region plus the 120 kb region downstream.

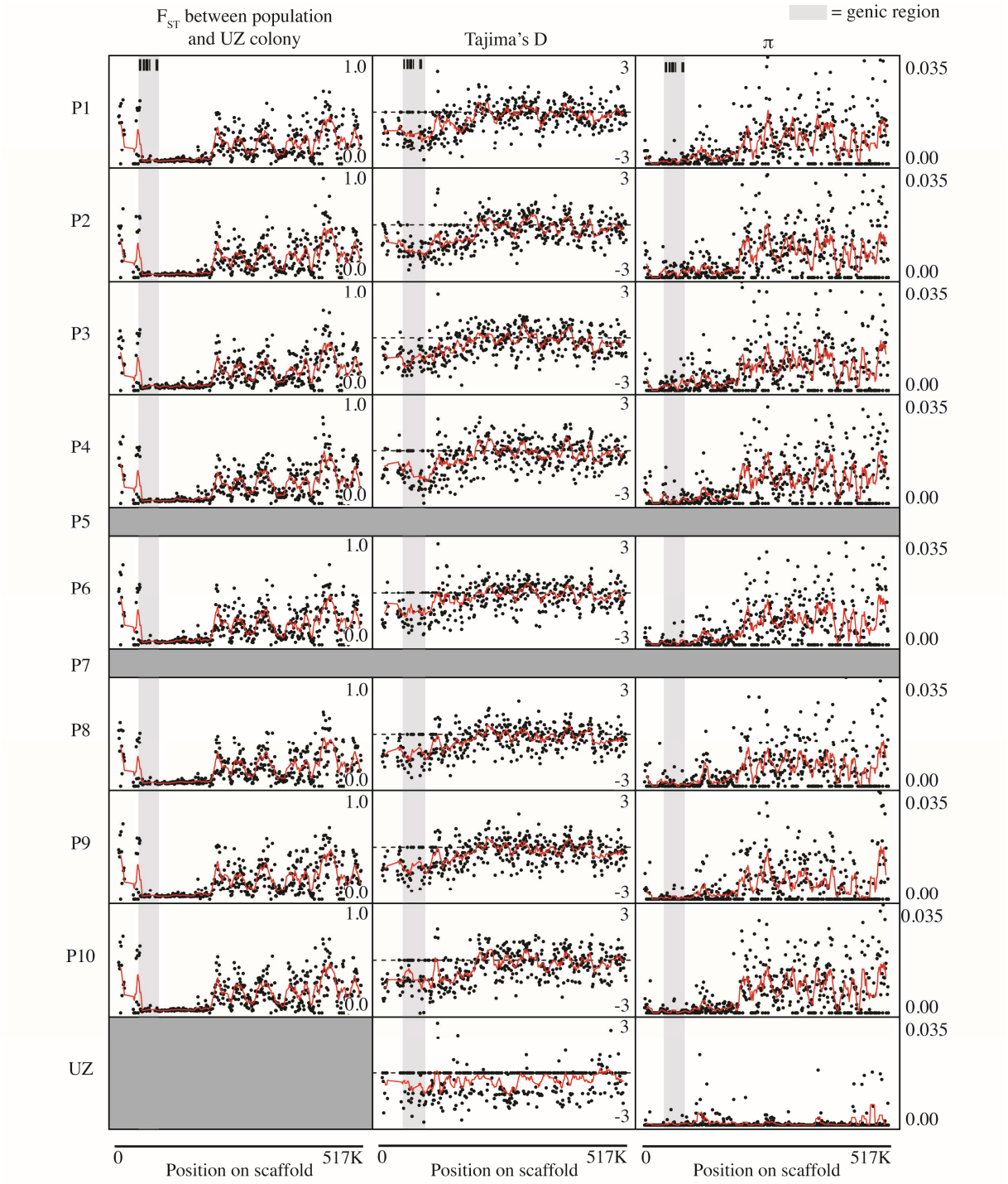

Figure 2.3. Genomic summary statistics across scaffold 87, containing *Pdfr* (Table S2.1). The first column depicts pairwise F_{ST} (each point is one 1000 bp sliding window average) between each population along the latitudinal gradient (Figure 2.1) and the univoltine Z (UZ) colony. The second column shows Tajima's D (each point shows an average of a 1000 bp sliding window), and the third column shows nucleotide diversity (π) (also averaged across 1000 bp sliding windows). Red lines indicate 10-window moving averages for all columns. The vertical bar and black tick marks (top) indicate the location of the exons for each gene. Horizontal grey bars indicate populations that were sequenced but removed after failing to pass data quality filters.

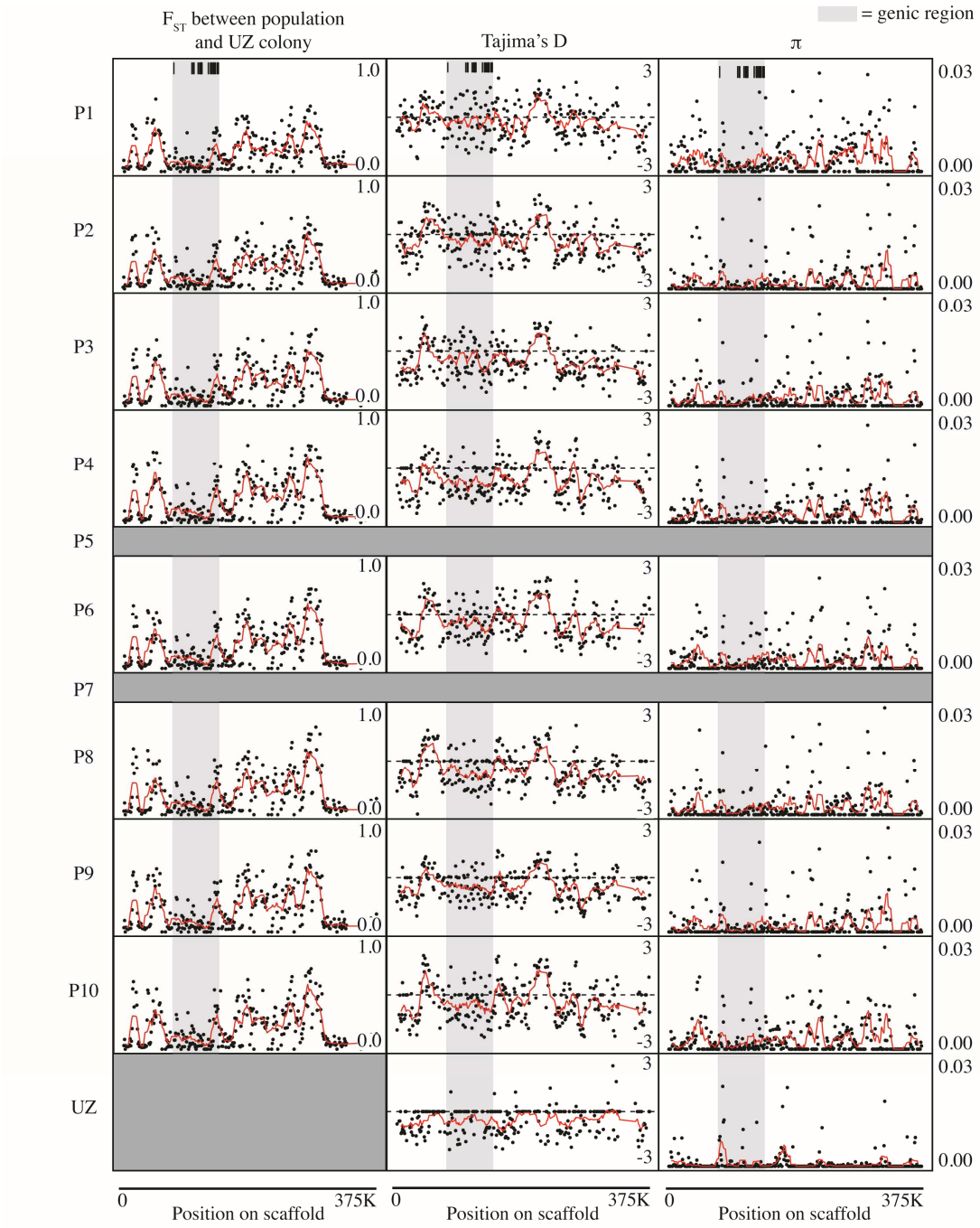

Figure 2.4. Genomic summary statistics across scaffold 532, containing *period* (Table S2.1). The first column depicts pairwise F_{ST} (averaged over 1000 bp sliding windows) between each population along the latitudinal gradient (Figure 2.1) and the univoltine Z (UZ) colony. The second column shows Tajima's D averaged over 1000 bp sliding windows, and the third column shows nucleotide diversity (π) averaged across 1000 bp sliding windows. Red lines indicate 10-window moving averages for all columns. The vertical bar and black tick marks (top) indicate the location of the exons for each gene. Horizontal grey bars indicate population that were sequenced but removed after failing to pass data quality filters.

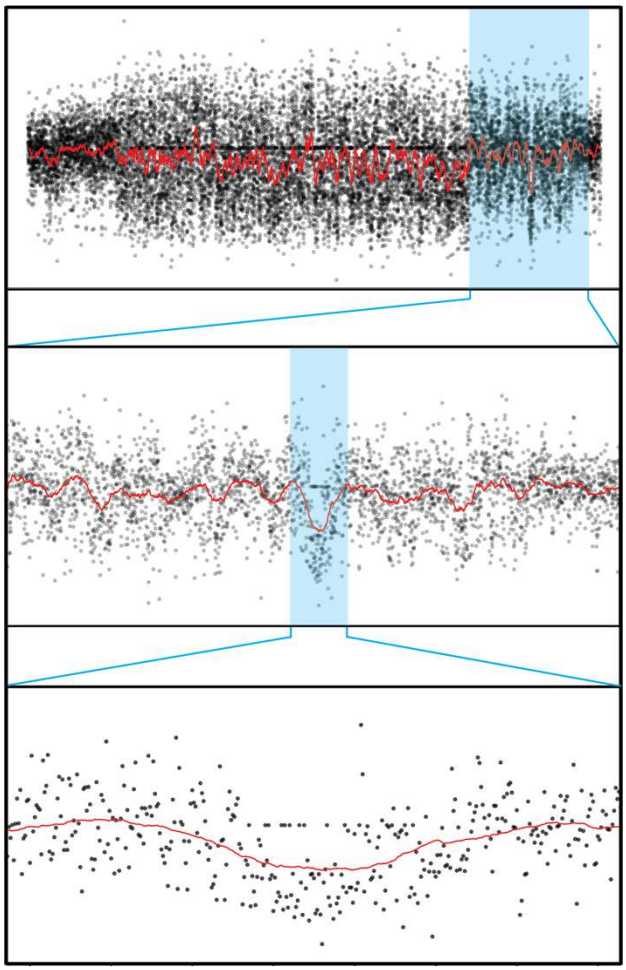

Figure 2.5. Estimates of Tajima's D across the entire Z-chromosome (top) for population P1. Each point shows Tajima's D calculated from a 1000 bp window within genomic scaffolds. Scaffolds are arranged in physical order along the entire chromosome (see Table S2.1). The red line shows a moving average from 100 estimate windows. The middle and lower plots are the same data, but zoomed in on the 'valley' underneath the *Pdfr* gene.

Tests of selective sweeps at *period* and *Pdfr* using the hidden Markov model implemented in pool-HMM (Boitard et al. 2013) did not reveal evidence of these for either *period* or *Pdfr* for any transect population, whether considering the scaffold within the genic region or outside. We attempted multiple model runs using different k-values (transition probability between hidden states) ranging from 0.001 to 0.1 but did not detect evidence of a sweep in any run.

Genome wide scans for selection

Following our assessment of broad, genome-wide summary statistics and more targeted analysis of known life history loci *period* and *Pdfr*, we applied methods to search for putative targets of selection across the entire genome: BayPass, Betascan, and JTKcycle. We used BayPass (Gautier et al. 2015) in an attempt to find outlier SNPs that, after controlling for background population covariance, have significant associations with population specific covariables. We used Betascan to look for regions that show signatures consistent with balancing selection: genomic windows with an excess of intermediate frequency alleles found in surrounding SNP sites. Finally, we used JTKcycle to identify SNPs that showed sinusoidal patterns in allele frequency along the transect.

BayPass

We ran BayPass models for the eight populations collected along the geographic cline, excluding the lab colony. For the smaller Z-linked dataset, we identified 39 SNPs with BayesFactor scores (BF.dB) greater than 15, a commonly used cutoff in the literature to assess significance (Chaturvedi et al. 2018; Miller et al. 2019; Sokolkova et al. 2021). Nineteen of the 39 SNPs were found to coincide with annotated genes (Table S2.4). Among these, four were associated with the "step cline" covariate, nine with degree days, four with latitude, and two with the number of generations.

Among Z-linked markers, we found two genes of known function among our outlier loci that covaried with a "step cline" shape (Table S2.4). The first is *bric-a-brac*, a gene of known function in controlling male pheromone response behavior (Dopman et al. 2004, Koutroumpa et al. 2016, Unbehend et al. 2021). While ECB moths in the Midwest have not been shown to display the variation in pheromone classes as seen in populations in the Northeastern United States or Europe, it may be possible that earlier selection on this locus has shaped patterns that were then carried into Midwestern populations. However, it is also possible that the *Pdfr* locus, known to have functional significance in ECB phenology and located less than 0.5 Mb from *bric-a-brac* is shaping patterns at surrounding loci. We also found an association between a zinc-finger domain and the step cline. While zinc-finger regions can be involved in many functions given their role in binding DNA, RNA, proteins, and other cellular molecules, the specific type of protein identified by our scan has been demonstrated as a target of selection in *Bombyx mori*, the silk moth, and controls sex-determination and alternative splicing (Gopinath et al. 2016).

The degree days covariable is a broad proxy for the summary environmental conditions experienced by ECB populations, as it combines both cumulative days and average temperature in its calculation. We identified seven Z-linked loci of known function that showed association with this covariable. While functions varied, a shared theme among these loci was a role in either cellular structure, transport, or motor function (Gee et al. 1997, Hu et al. 2021). One associated locus encodes a proton-coupled amino acid transport protein that has been demonstrated to vary in gene expression level in response to insecticide application (Gao et al. 2020). Outliers with the latitude covariable were a subset of the same loci that covaried with the degree days covariable. This is expected, given latitude and degree days covary with each other.

Finally, the number of generations covariable was associated with two Z-linked genes. The first, *clock*, is a circadian rhythm controlling gene. While circadian rhythm itself could have functional visibility to selection, it should also be noted that there is increasing evidence in the cooption of circadian genes in controlling diapause behaviors (Kozak et al. 2019; Gotthard and Wheat 2019). Because developmental rate following the diapause break controls phenology in ECB moths (Dopman et al. 2005; Levy et al. 2015; Kozak et al. 2019), selection on circadian genes that have expanded their function into diapause pathways is plausible. We also found a correlation with the LIM homeobox gene. It is possible that LIM plays a role in a developmental process under selection, but, similarly to *bric-a-brac, lim* is estimated to fall within 0.5 Mb of *Pdfr*, so it is also possible that selection at *Pdfr* has shaped this genomic region.

Among autosomal markers, we found 629 SNPs with BayesFactor scores (BF.dB) greater than 15. Among these, 246 mapped to locations of annotated genes. Table S2.5 shows the top 25 sorted by BF.dB. Among these hits, eight were associated with the step cline variable, five with the degree days covariable, six with latitude, and six with the number of generations.

Outlier loci that were associated with the step cline variable were often related to motor control and developmental gene pathways. *Patched* is associated with development, cell growth, and differentiation (Ingham et al. 1991) and the homeobox protein extradenticle is a transcription factor associated with *Hox* gene pathways (Mann et al. 1998). The *Caskin 2* protein is involved in *wnt* singaling and, in *Drosophila*, plays a role in guiding motor axons. While other hits had less deducible roles, these may indicate potential for selection to have acted on development and

motor pathways, especially if they have a role during diapause or after it is broken and larval development resumes.

Latitude and degree day associations had functions such as stress response (Zhao et al. 2020), oxygen binding (Kawaoka et al. 2009), innate immunity in moths (Lin et al. 2017), and metabolic regulation of ECB moths (Uzelac et al. 2020). While validating these loci is not feasible with our dataset, putative roles in adaptation to thermal and seasonal stress may be suggested.

Several autosomal outliers associated with the number of generations could potentially indicate roles in immune function and developmental pathways. The protein lap4 plays a role in cell development as has been shown to be differentially expressed in viral-resistant silkmoths (Guo et al. 2015), *Chordin* controls developmental patterning (Holley et al. 1995), and an E3 ubiquitin protein ligase MYLIP-A degrades lipoprotein receptors (Calkin et al. 2011).

Finally, the largest outlier in our BayPass dataset was found under the malate dehydrogenase gene. This locus varied across the transect and covaried with latitude and degree days. Both *Ostrinia nubilalis* and the copepod *Calanus glacialis* have demonstrated downregulation of this gene during diapause, suggestive of a role in metabolic regulation (Koštál et al. 2006; Uzelac et al. 2020).

Identifying potential loci under balancing selection using Betascan

We utilized Betascan software (Siewert and Voight 2017, 2020) to scan the genome for evidence of balancing selection at each locality along the transect. Table S2.6 shows the top 0.01% of the dataset after application of Betascan. Particularly notable are 10 outliers, several of which were replicated across multiple localities, that show patterns consistent with the action of balancing selection. We are also aware of the potential for mapping failures associated with gene duplications and paralogous sequences generating false positives in the data. This appears to be the case with some outliers identified: notably the large number of transposable elements and retrotransposons in the top 0.01% of the dataset (Table S2.6).

The functional significance of the non-TE/RTE Betascan outliers ranged from transcription regulation (Dickson et al. 1999; Laity et al. 2001; Boube et al. 2002) to transmembrane transport (François et al. 2012; Hu et al. 2021), cellular architecture, and signaling (Sharabi et al. 2013).

One candidate locus was an isoform of tropomodulin and was amongst the top 0.01% of Betascores in six localities (P2, P3, P4, P6, P8, and P10). Tropomodulin is involved in actin filament architecture in invertebrate muscles (Hooper et al. 2008; Yamashiro et al. 2012). Tropomodulin has been found to be a target of plant retroviruses, which exploit it to regulate cell tubule motility in insect vectors (Chen et al. 2017). If ECB is a potential vector for host-plant-associated viruses, this gene may have functional significance in deterring viral attack.

Tests for latitudinally oscillating allele frequencies

We analyzed allele frequencies for every SNP in the post-filter dataset (N = 8,489,446 SNPs) to

determine whether they showed significant sinusoidal patterns along the latitudinal gradient using the R package JTKcycle (v3.0). Following filters for amplitude (> 0.1), we narrowed our results to just 2,166 SNPs. Of these, 118 were found to have a significant ($p < 0.05$) association with a sinusoidal curve: we show the top 25 in Table S2.7 along with annotation information. The joint third most significant hit in our list fell within the *period* locus (multiple test adjusted $p = 0.0017$) and showed a pattern of variation with an amplitude of 0.113 across allele frequencies between 0.57 and 0.88 (Figure 2.6).

Other significant hits largely fell among developmental regulatory genes such as *wnt,* neuropeptide receptor A3*,* dual specificity protein kinase pom1*,* and interferon regulatory factor 2-binding proteins 1 & 2 (Nusse & Varmus 1992; Martin et al. 2009; Stirnimann et al. 2010; Soppa et al. 2015; Ding et al. 2019; Li et al. 2020). Other hits included a protein controlling salinity and osmotic stress response through sodium/potassium ATPase channels (Sebastian et al. 2018) and a sulfotransferase gene putatively involved in pesticide response mechanisms (Coughtrie 2002; Jia et al. 2020).

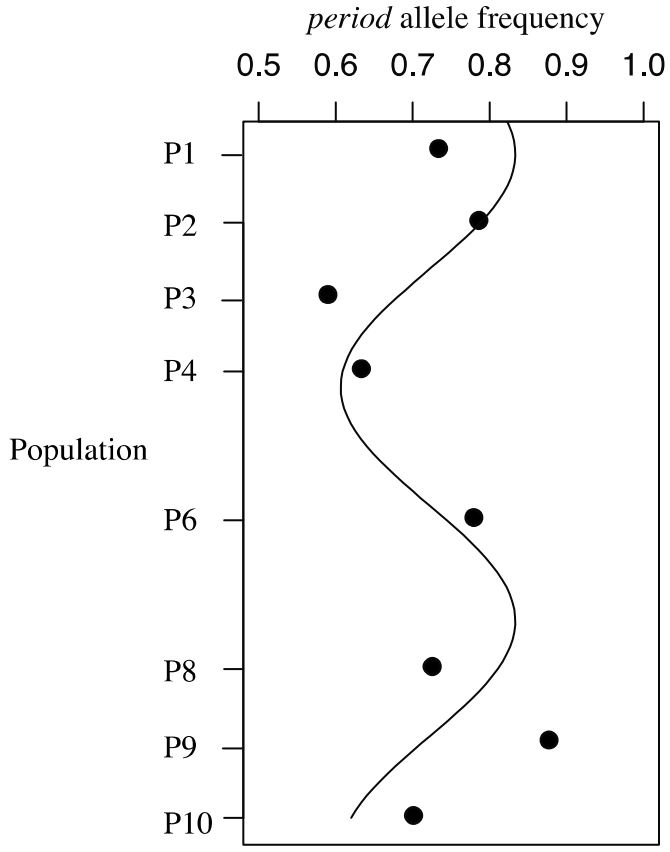

Figure 2.6. Latitudinal sequence of European corn borer populations from north to south in the Midwestern US (see Figure 2.1) showing frequencies of an allele within the *period* gene identified using JTKcycle. The sinusoidal curve is the inferred fit for the sequence generated by JTKcycle.

DISCUSSION

Identifying signatures of range expansion and selection along a latitudinal gradient

Phenotypes that control timing, rhythm and seasonality form the basis of an organism's phenology. Here we attempted to understand whether a recent, rapid, and widespread range expansion by the European corn borer (ECB) moth into the Midwestern United States could leave signatures of natural selection throughout the genome, and if these signatures could help us to understand the

forces driving the phenological variation in this species. To do this we studied populations along a latitudinal gradient of about 720 km and 4.5 degrees latitude (Figure 2.1).

The demographic factors associated with the recent and rapid range expansion of ECB moths, occurring in the last 100 years since their introduction to North America left a clear mark on the genome. ECB populations have been documented in the Midwestern transect for at least 90 years (Sparks et al. 1966), so their range rapidly expanded following arrival into North America in the early 1920s (Smith 1920, Caffrey and Worthley 1927). This process would have been shaped by potential bottlenecking, leading to strong drift and the loss of variation, followed by rapid demographic growth. This expansion will lead to an excess of rare alleles, as initial variation will be low and the majority of new mutations that arise will not have spread throughout the expanding population. Traditional estimates of selection that rely on nucleotide site frequencies to infer departures from neutrality, such as Tajima's D reflected this pattern. Tajima's D is expected to be negative for expanding populations (Tajima 1989, Excoffier et al. 2009), and our data showed this pattern throughout the autosomal and sex-linked averages (Figure 2.2C). However, despite our sampling drawing from a large geographic range and with at least 80 km separating each locality (greater than the typical ECB dispersal range of 30-60km: Sappington 2018), little population structure was apparent, and populations showed no evidence of genomic regions containing fixed differences (Figure S2.1) or isolation-by-distance effects.

We investigated these genome-wide patterns and quantified their effects. In addition to consistently negative Tajima's D estimates for both autosomal and Z-linked markers, we found that allelic diversity was relatively low, especially on the sex chromosome. We also noted that F_{ST}

between transect populations was low and we failed to identify outlier windows. An exception was F_{ST} between the transect populations and the UZ lab colony, which showed higher values in every pairwise comparison. This is likely due to the lack of diversity from the selected lab colony, since depletions in local diversity can elevate pairwise F_{ST} (Cruickshank and Hahn 2014) and the UZ colony showed extremely low levels of nucleotide diversity.

While these demographic effects can lead to a greater challenge in inferring the effects of selection, we developed predictions about the two known loci that form the basis of distinct seasonal timing classes in ECB. Prior studies (Dopman et al. 2004, Levy et al. 2015, Kozak et al. 2019) have demonstrated that post-diapause development rate leads to variation in the number of generations occurring at a particular locality. This phenotype is regulated by two interacting genes in the circadian clock pathway, *period* and *Pdfr.* Variation in these alleles leads to variation in generation number in the Northeastern United States, and our expectation is that some of this variation (1) would have been carried with the species during the range expansion into the Midwest and (2) would allow ECB moths to adapt to novel environments by selecting for phenology phenotypes that were advantageous at a particular locality. This might take the form of balancing selection, should both phenotypes be maintained through factors such as seasonal variation, tradeoffs associated with developmental rates, or a combination thereof. Alternatively, individual localities might have a local optimum that would drive directional selection for different phenotypes, or a single phenotype could be favored throughout the new range.

Because we established that F_{ST} was low between transect populations and no outlier loci were found, we focused instead on how *period* and *Pdfr* varied across the range. Pairwise comparisons

between transect populations did not show evidence of differentiation, but comparisons between the UZ colony and transect populations showed a dramatic reduction in F_{ST} at *Pdfr* and a modest reduction at *period* (Figures 2.3 & 2.4). The similarity of the lab colony to the field populations is accompanied by a dramatic reduction in nucleotide diversity in the *Pdfr* locus relative to the surrounding genomic region. This is consistent with the action of strong and recent selection on *Pdfr* across the entire transect, potentially a selective sweep (Smith and Haigh 1974). An explicit test for signatures of a selective sweep (Pool-hmm) did not return any hits, but methods to identify signatures of selective sweeps in pooled sequencing datasets are limited because long haplotypes cannot be called.

The observed changes in allele frequencies were also reflected in patterns at Tajima's D, with *Pdfr* showing significant differences from the rest of the genome in a negative direction, which is consistent with strong directional selection (Tajima 1989). The region containing *Pdfr* shows a broad, localized reduction in Tajima's D larger in magnitude than any other on the Z-chromosome (Figure 2.5). At *period*, while there are overall similarities in the allele frequencies at locations along the cline to allele frequencies in the UZ colony, we did not observe a reduction in nucleotide diversity within and surrounding the genomic region, suggesting that variation has not been lost. While the modest reduction in F_{ST} suggested that there may be some degree of similarity between *period* alleles in transect populations and in the UZ colony, this effect appears weaker than the effect observed at *Pdfr*. This weaker effect is the result of variation in allele frequencies at *period* within transect populations that was not observed in the UZ colony.

Some alleles of *period* showed significant sinusoidal oscillations in frequency along the transect

(Figure 2.6). This is consistent with the idea that each population may have a local optimal frequency of an allele or alleles associated with the number of generations per year. Favored alleles may not reach fixation in a particular locality due to factors such as annual variation, plasticity, dispersal and strength of selection.

Given the significant drop in diversity and low F_{ST} between the UZ colony and all transect populations, is possible that the same *Pdfr* allele that was selected in the colony is also favored across the transect populations. This does not preclude phenotypic variation in PDD timing along the transect. We know from prior studies that diversity at the *period* locus alone can give rise to multiple phenotypes even if *Pdfr* is invariant (see Kozak et al. 2019, Figure S2.1, panel E). It is therefore possible that as the ECB range expanded, *Pdfr* experienced strong, directional selection while *period* retained variation, potentially through balancing selection or spatially/temporally varying directional selection. This finding would be consistent with the observation of intermediate to low frequency allele classes at *period* and the oscillation of allele frequency in these classes along the transect.

Identification of other loci under selection

We also leveraged the entire dataset to identify outlier loci beyond *period* and *Pdfr*, using BayPass to search for associations with population specific environmental covariables, Betascan to look for signatures of balancing selection, and JTKcycle to identify loci that show oscillating patterns in allele frequency. Outlier loci were found in circadian clock pathways, or among genes associated with thermal tolerance, or in positions that could potentially indicate divergence hitchhiking.

Circadian clock genes

Circadian clock genes have long been hypothesized to also control annual rhythms (Bünning 1936), and support appears to be growing for this idea. Our study identified two genes in addition to *period* that are known to have a role in circadian rhythms: *clock* and *slowpoke.* Our BayPass analysis identified *clock* as significantly associated with the average number of generations at a locality as a covariate. *Clock* is a gene with known functions as a transcriptional operator of the circadian clock (Sandrelli et al. 2008; Kozak et al. 2019). Prior work has indicated that the gene may also be involved in changing voltinism patterns, having been identified as an outlier locus corresponding to voltinism changes in two different butterfly species, *Pieris napi* and *Pararge aegeria* (Pruisscher et al. 2018, 2021, Lindestat et al. 2021). Our Betascan analysis identified *slowpoke,* a gene that has been demonstrated to have a significant role in regulating rest cycles in *Drosophila melanogaster* (Jaramillo et al. 2004, Ruiz et al. 2021). In this species, expression was demonstrated to be reduced in the neurons that lack pigment dispersing factor (PDF), a protein that binds to *Pdfr* (pigment dispersing factor receptor). While to our knowledge, it has not been demonstrated to have an explicit role in timing annual rhythms, the connection with one known trait locus merits note.

Thermal tolerance

Among the highest scores in our Betascan analysis for balancing selection was the gene Tropomodulin. This regulates actin filament architecture in invertebrate muscles (Hooper et al.

2008; Yamashiro et al. 2012, Chen et al. 2017). Tropomodulin may also have a role in adaptation to thermal stress, as previously suggested for another invertebrate, the invasive green crab (Tepolt and Palumbi 2020). Thermal stress is likely to vary along the Midwestern latitudinal transect and is particularly likely to vary with differing life history and voltinism phenotypes, since the climate experienced by adults during early June or late August flights could be quite different from that experienced following mid-July emergence. Given ECB in a new distribution area may experience a broad range of temperatures, according to the timing and number of generations within a locality, variation for thermal tolerance via muscle architecture could potentially be regulated by balancing selection. Two other outlier candidates are involved in actin filament binding and recruitment: unconventional myosin and FERM (Bird et al. 2014, Gaspar et al. 2015). These loci were both identified as significant associates with the degree days covariable. Degree days are our closest proxy for the strength of temperatures experienced by insects at a given locality, and thermal stress could have acted as a selective force driving a change in muscle architecture.

Genetic hitchhiking

In our BayPass dataset, we found *bric-a-brac* to be an outlier associated with the step cline and potentially discriminating between the two northernmost populations and the six southern populations. *Bric-a-brac* is a gene with a known function in male pheromone response (Glover et al. 1991; Dopman et al. 2010; Unbehend et al. 2021). While it is to our knowledge unlikely that multiple pheromone strains are present in the Midwestern United States, *bric-a-brac* could potentially have effects beyond control of pheromone response. *Bric-a-brac* is located in close genomic proximity to *Pdfr* (< 0.5 Mb), and our BayPass result could reflect hitchhiking relative

to that locus (Smadja et al. 2008). Likewise, the LIM homeobox gene could play a role in a developmental process under selection, but, similarly to *bric-a-brac, lim* is estimated to fall within 0.5 Mb of *Pdfr*. It seems plausible that selection at *Pdfr* has shaped this entire genomic region.

Conclusions

Over the course of the past 110 years, ECB moths were introduced and spread across North America, westwards to the Rocky Mountains (Smith 1920; Klun and Cooperators 1975; Sorenson et al. 1992). During this expansion, they encountered novel and widely variable climatic conditions. Despite the challenges these conditions might present, they quickly established new populations. To do so, they utilized existing variation in voltinism, a critical trait for timing their life cycles with the annual growing season length (Showers et al. 1975). Our findings here are reflective of the historical demography of this species as it expanded its range. The range expansion left indelible marks on the genome that are reflected in genome wide summary statistics such as reduced π, and negative Tajima's D. While these same genome wide forces may present challenges in distinguishing selective forces from neutrality, we were able to find signatures that are consistent with selection on known loci corresponding to adaptive phenotypes: phenology loci *period* and *Pdfr*. We also found novel candidate genes that may be under selection along the climatic gradient. We maintain that alleles underlying phenotypes involved in species phenology are expected to be subject to strong selective forces as ranges expand (Śniegula et al. 2014). The recency and speed of the ECB moth spread through North America has resulted in signatures in our data that are subtle, yet consistent with departures from neutrality. We encourage further study to further

disentangle the effects of range expansions on neutral and selective processes. These processes have implications for our understanding how organisms utilize shifts in phenology to expand their range (Macgregor et al. 2019) and how this may enable the persistence of species in changing climates (Bradshaw and Holzapfel 2001, 2010).

CHAPTER THREE

INTROGRESSION OF SPECIATION TRAIT LOCI DRIVES DIFFERENTIATION IN THE EUROPEAN CORN BORER, *OSTRINIA NUBILALIS*

ABSTRACT

Reproductive barriers drive divergence by restricting gene flow between populations. However, barriers are often semipermeable, allowing some gene flow between groups. In cases such as these, introgression and recombination will complicate emerging genealogical patterns throughout the genome. By inferring genealogical trees for individuals and species of the *Ostrinia* moth group, we explore how four trait loci involved in reproductive isolation drive patterns of ancestry. We utilize whole genome sequences to infer changes in population size over time. We then develop tree sequence topologies from both genome wide markers and the set of four trait loci. We calculate genealogical nearest neighbor estimates, a statistic that summarizes the relative ancestry proportions of each individual in order to describe how phenotypic variation has shaped the species and population history of European Corn Borer (*O. nubilalis*) moths. We show that much of the genome reflects genealogical sharing across many populations, but genealogies at barrier loci distinguish groups. Finally, we calculate D statistics and demonstrate evidence of introgression of these barrier traits between multiple species within the *Ostrinia* group, suggesting a complex model for barrier trait evolution in this group. We argue that the outsized impact of few trait loci and their movement between lineages promotes speciation in this clade.

INTRODUCTION

The natural world consists of organisms that are genetically connected with other organisms, forming populations. Speciation, the formation of groups of organisms that are novel and distinct from others, occurs when genetic ties between populations become severed (Shaw and Mullen 2011). An understanding of how and why such species boundaries are formed is a long-standing question (Mayr 1963; Wu 2001; Harrison & Larson 2014). In particular, genes that cause, allow for, or maintain the severance of genetic connections and promote reproductive isolation can be seen to be critical for the generation of diversity (Butlin & Ritchie 2001; Orr et al. 2004). Despite significant progress in the past several decades, we still do not have a complete picture of the origin of variation in the traits that underlie phenotypes that contribute to speciation and the genes that control this variation.

Part of the challenge comes from identification of the original reproductive barriers and the genetic loci involved. While demonstrating genetic divergence at loci between related taxa is useful and can begin to narrow down candidate loci that may contribute to speciation, divergence alone is not evidence of a functional role for these loci in the speciation process (Harrison 1998; Shaw and Mullen 2011). Loci that are genetically distinct may not have become so until after species barriers emerged through the action at other loci or may have diverged even before the rest of the genome became distinct, but played no functional role in separating species.

Key allies in identifying how barriers arise and generate new species are identifiable forms in the early stages of speciation that have not yet become completely distinct and, therefore, have not yet formed "good" species (Barton & Hewitt 1989; Harrison 1990; Jiggins & Mallet 2000; Coyne & Orr 2004). These forms may still exchange genes throughout all or portions of their genome yet have evolved differences that act directly or indirectly in generating and/or maintaining distinct genetic groups. A major benefit of careful study of these forms is that the action of individual "speciation phenotypes" (Shaw & Mullen 2011, 2014) can often be tested directly by quantifying the levels of gene flow between groups with and without the phenotypes (Barton 1983; Dopman et al. 2010; Butlin & Smadja 2018).

Barriers that act prior to copulation ("pre-mating barriers") may be especially important at the very beginnings of the speciation process, as has been demonstrated in a number of taxa, such as *Rhagoletis* flies (Feder et al. 1994), *Mimulus* monkeyflowers (Sobel & Streisfeld 2015), *Sporophila* seedeaters (Turbek et al. 2021) and many others. These barriers may begin as intraspecific variation within a species, but if multiple barriers arise independently (or arise at once, for example in a "magic trait" model: (Gavrilets 2004; Maan & Seehausen 2011; Servedio et al. 2011; Smadja & Butlin 2011), their effect may amplify and lead to the acceleration of speciation (Flaxman et al. 2014; Nosil et al. 2017). Pre-mating barriers often act on an organism's ability to locate suitable mates, for example by shifts in the timing of mate availability, or on the signals that organisms use to locate a mate in geographic space (Mayr 1947; Coyne & Orr 2004).

Seasonal timing (phenology) and reproductive signaling (behavior) constitute a combination of

traits that are tractable in the study of species formation. The European Corn Borer moth, *Ostrinia nubilais* Hübner, exemplifies variation in both of these traits and the combinations of the two in its introduced range of North America. The complexity and variation of these mating traits leads to our ability to explicitly test their effect. Behavioral barriers are controlled by variation in pheromone signal and response (Klun and Cooperators 1975, Anglade et al. 1984). *Ostrinia* females produce a blend of E and Z isomers of tetradecenyl acetate. Within a species, variation in the amount of E and Z isomer in the blend gives rise to different pheromone classes (Klun et al. 1973, Kochansky et al. 1975). Different species also demonstrate variation in the placement of the double bond that distinguishes the cis and trans isomers (Lassance et al. 2013). This blend is controlled by variation in the pheromone gland fatty acid reductase gene (*pgFAR*), an autosomal locus of major effect (Lassance et al 2010). Male moths show correspondent preferences to the E and Z pheromone blends, a flight response trait controlled by the Z-linked *bric-a-brac* (=*bab*) locus (Unbehend et al. 2021). Together these loci can act to greatly reduce gene flow between sympatric *Ostrinia* pheromone strains (Dopman et al. 2010). Temporal barriers are produced through two epistatically interacting circadian clock genes that control post-diapause developmental (PDD) timing (Dopman et al. 2005; Levy et al. 2015; Kozak et al. 2019). These genes, *period* and pigment dispersal factor receptor (*Pdfr)*, map to the Z chromosome and, within *Ostrinia nubilalis*, regulate the number of annual generations (voltinism) of a population at a particular locality. These two traits differ across the North America range of *Ostrinia* and the particular combinations of E and Z pheromone types and univoltine and bivoltine temporal classes can result in variable levels of genetic differentiation across ECB chromosomes (Kunerth et al. 2022).

The recent characterization of alleles regulating these four genes (Lassance et al. 2010, Kozak et al. 2019, Unbehend et al. 2021), has invited their study through evolutionary inference and population genetic modeling in the *Ostrinia* group. In doing so we can address how variation arises in reproductive barriers. Our specific aim in this study is to describe the sequence evolution of four genes that encode two traits critical to reproductive isolation in the *O. nubilalis* species group as it diversified. To further our understanding of these genes we analyzed them within the complete genomes of 128 individual moths, thereby allowing study of neutral variation across the genome to infer demographic history. We then attempted to reconstruct the genealogical histories of individuals at the four trait loci. In this the context, we explored how the broader *Ostrinia* moth groups and the intraspecific variation within *O. nubilalis* have been shaped by gene flow and introgression at trait loci. All these approaches contribute to the reconstruction of the evolutionary pathways that have promoted incipient speciation in the *O. nubilalis* group.

RESULTS & DISCUSSION

After filters, we generated complete whole genome sequences for 124 individuals, totaling 26,491,541 SNPs within 454.7 Mb of scaffolds. These individuals represented a variety of species (abbreviated to ALB, ACB, ABB and ECB) and populations (Table 3.1). In total, the *O. nubilalis* draft genome contains 8,843 scaffolds (N50 = 392.5 kb).

Table 3.1. Details of populations and species used for genome sequencing

Population / Species	**Population tag**	**Number of individuals**	**Country / State**	**Number E-strain**	**Number Z-strain**
American Lotus Borer	ALB	3	USA	NA	NA
Asian Corn Borer	ACB	4	China	NA	NA
Adzuki Bean Borer	ABB	1	Japan	NA	NA
European Corn Borer					
Rockspring	ECB	30	USA - PA	15	15
Landisville	ECB	32	USA - NY	16	16
East Aurora	ECB	14	USA - NY	0	14
Farmington	ECB	4	USA - NY	4	0
Geneva	ECB	22	USA - NY	10	12
Perry City	ECB	3	USA - NY	0	3
Madison	ECB	4	USA - NY	0	4
Penn Yan	ECB	4	USA - NY	0	4
Dover	ECB	2	USA - MA	2	0
E colony	ECB	1	NA	1	0
Total		124		48	68

We visualized population structure using principal coordinates analysis (PCA) divided between six genomic regions targeted for analysis: autosomes, the sex chromosome, and, at a finer scale, the scaffolds for four barrier traits (Figure 3.1). In every PCA, the outgroup ALB separates from the majority of the other taxa. A PCA of the autosomal scaffolds distinguished two of the outgroup species. ALB clustered in the bottom right of coordinate space, ECB samples all in the bottom right, and ACB in the top right. The ABB individual clustered near the ECB samples, but was distinguishable from their cluster. PCAs of trait scaffolds showed some differentiation between ECB samples driven primarily by E and Z pheromone types (Figure 3.2). *O. nubilalis* showed little differentiation between samples when using autosomal markers to generate the PCA. Pheromonal trait loci *pgFAR* and *bab* were able to distinguish groups of E and Z ECB.

Intermediates between these groups appeared to be heterozygotes or backcross individuals. At phenological trait loci *period* and *Pdfr*, separation between groups was weaker. Samples with known phenotypic information for both phenological timing (long vs short PDD) and pheromone class (E vs Z) did appear to be distinguishable within PCAs based on *period* and *Pdfr* genic locations (Figure S3.1). Geographic structure did not appear to drive differences between groups (Figure S3.2).

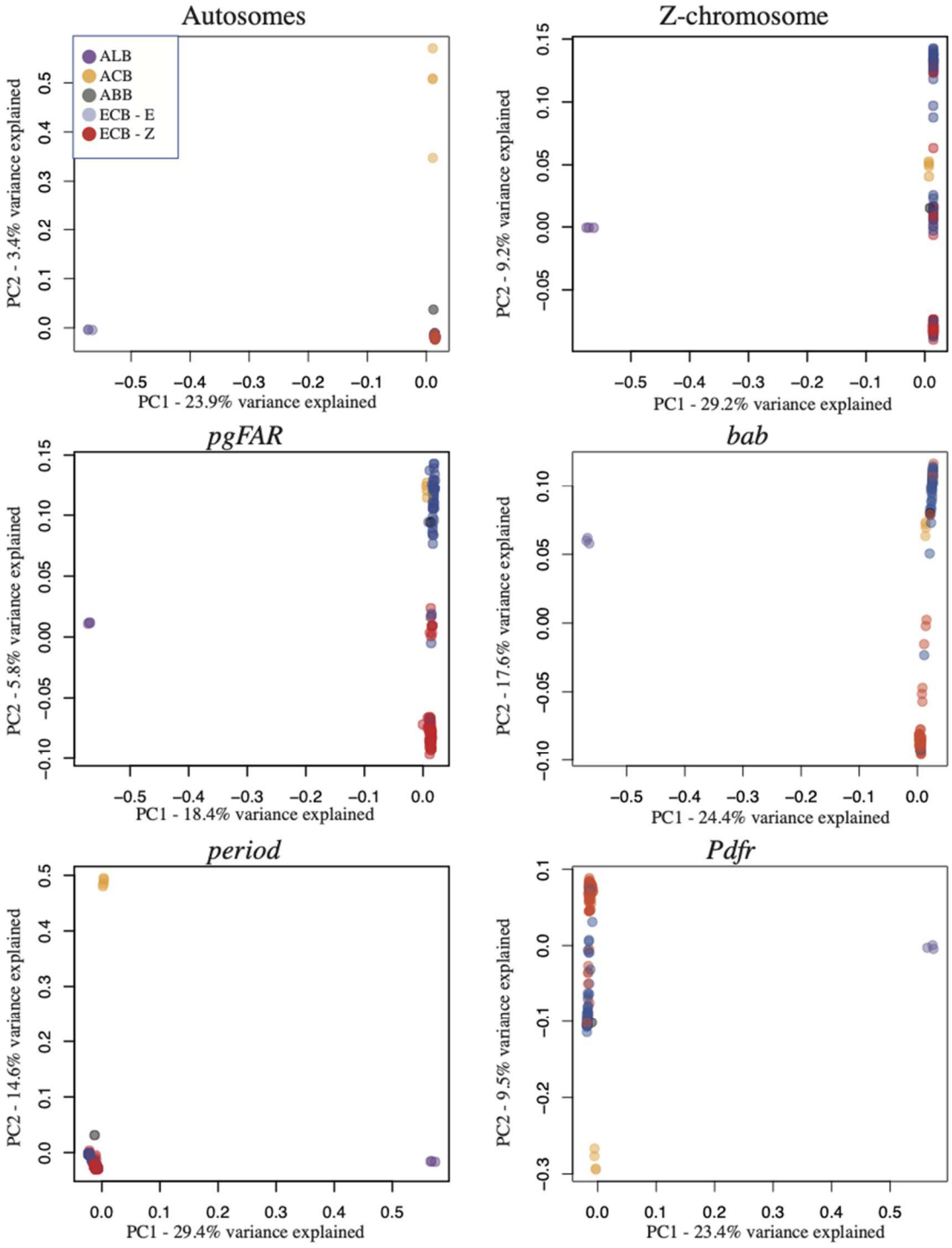

Figure 3.1. PCAs of all *Ostrinia* sequenced for the six genomic regions colored by sample taxa: ALB, ACB, ABB, ECB-E pheromone type, and ECB-Z. All samples are depicted on the plots, but in some cases the ECB individuals cluster so tightly that they are hard to distinguish. See Figure 3.2 for PCAs with outgroups removed for a more detailed examination of intraspecific variation in ECB.

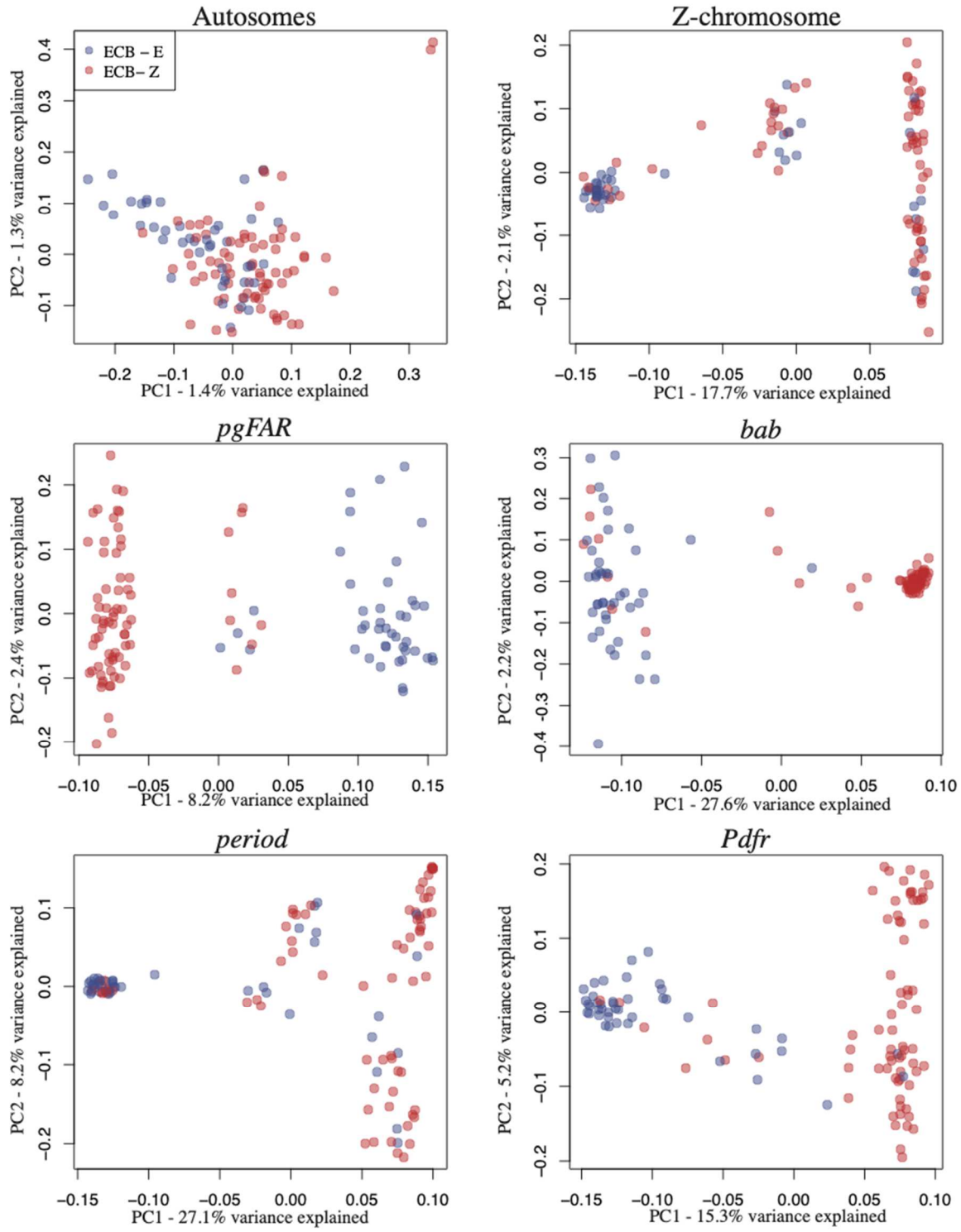

Figure 3.2. Principal components analysis (PCA) of six genomic regions with outgroup species removed. Individual samples colored by estimated pheromone category.

We inferred genealogical tree sequences using *tsinfer* for each genomic region. We then estimated genealogical nearest neighbors (GNN) for each individual in the dataset. GNNs are a statistical summary of the topology of a tree sequence centered around a focal individual or node in the tree. For this focal individual or node, GNNs are computed with respect to a user-defined list of reference sets along the tree sequence, beginning with the focal sample, until another node that has one or more descendants in the reference set or set is located. This process is repeated in order to find the proportion of each reference set among the descendants from that node.

Here we consider GNNs to be an approximation of the ancestry proportions of a given individual in our dataset that is shared with other individuals in the dataset, subdivided by the reference partitions of our choosing. Since recombination and gene flow should change the genealogical patterns and thus the inference of the tree topology in different portions of the genome, we can estimate the tree sequence topology for genomic regions of interest, calculate GNNs with respect to sets of known or estimated phenotypic classes, and in doing so demonstrate which portions of the genome are associated with those phenotypic groups.

We tested multiple reference sets to infer whether the ancestry proportion of a given individual at a chosen genomic location was defined on the basis of known phenotypes of the other individuals. For example, we first estimated whether the geographic population to which an individual belonged influenced the ancestry patterns by calculating GNNs using the geographic population as a reference set. With the exceptions of the outgroup taxa, geography was not a strong driver of ancestry patterns (Figure 3.3). Each ECB individual drew GNN proportions from a broadly consistent mix of geographic groups. While the relative proportions of the groups changed from

population to population, genealogical tree sequences were clearly not a strong reflection of geography as a whole.

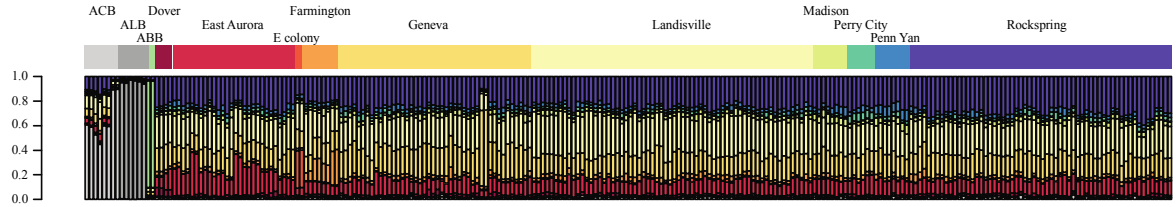

Figure 3.3. GNN proportions of each individual using autosomal markers. Each individual bar shows a haplotype, so every two adjacent bars is an individual. Individuals are ordered by geographic population (top bar). Colors correspond to population reference set, and the proportion of color in a vertical bar shows the proportion of GNNs drawn from each population reference set.

We then considered whether pheromone type drove ancestry patterns, by generating GNNs utilizing four reference sets: ACB, ALB, ECB-E and ECB-Z. For autosomal markers, GNNs based on pheromone category did not affect local ancestry, as every ECB sample showed GNN proportions drawn from both E and Z sample sets in roughly similar proportions (Figure 3.4, Panel A).

However, for trait loci, especially pheromone trait loci *pgFAR* and *bab*, GNN proportions of ECB individual samples were drawn heavily from either ECB-E or ECB-Z, indicative of the role these loci play in contributing to the genealogical tree topology that distinguishes groups. Since E and Z pheromone phenotypes depend on variation at these markers, it is unsurprising that tree sequences based on these genome regions reflect these patterns. Phenological trait loci *period* and *Pdfr* also showed concordance with pheromone category, suggesting that the tree topologies formed from markers near these phenological loci were shaped by the effects of the pheromonal loci. This is consistent with prior work that has demonstrated the effects of the coupling of these

trait loci with the promotion of genomic ‘congealing’ (Kunerth et al. 2022, process reviewed in Butlin & Smadja 2018).

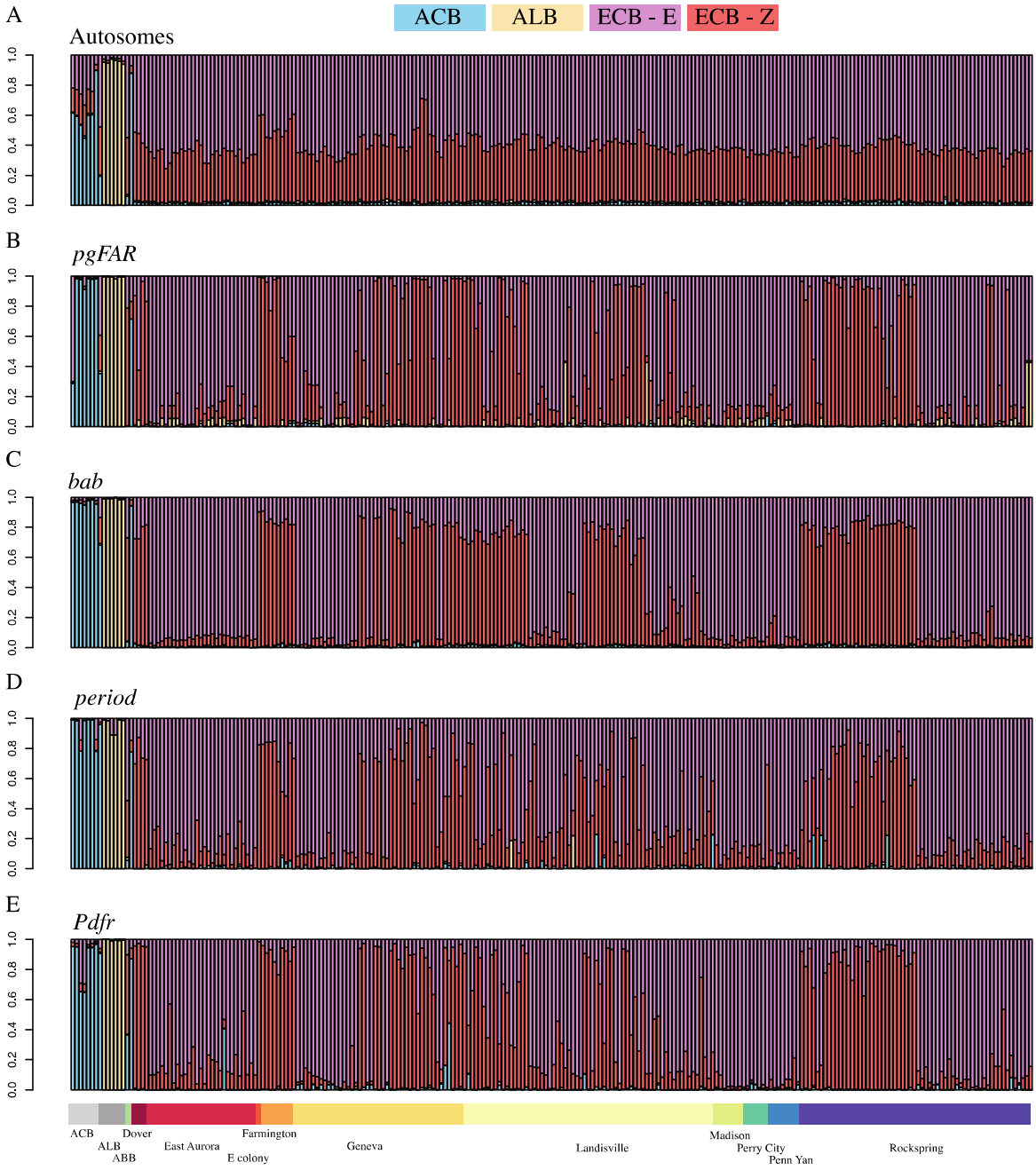

Figure 3.4. GNN proportions for each haplotype in the dataset, divided by tree topologies based on markers located on autosomes (Panel A) and markers located within genic regions of the four trait loci (Panels B-E). The horizontal colored bars at the top correspond to the sets tested and their colors within each GNN plot: ACB, ALB, ECB-E and ECB-Z. Colored bars at the bottom correspond to geographic populations (see Figure 3.3).

D statistics and detection of introgression

To test for introgression between groups, we used Dsuite (Malinsky et al. 2021) to calculate D for both autosomal and genic regions using entire scaffolds broken up into 20 jacknife blocks. We tested autosomal scaffolds of comparable sizes to the scaffolds containing genic regions. We also calculated the *df* statistic in sliding windows for autosomal and trait scaffolds and plotted the results from one representative scaffold here (Figure 3.5A). We estimated D using a tree with ALB as an outgroup, ACB branching next, and ABB and ECB as sister taxa, following the relationship inferred from past studies of *Ostrinia* phylogenies (Kim et al. 1999; Lassance et al. 2013; Yang et al. 2021). Using this topology, we were able to test for evidence of introgression between ACB and ABB. Positive values of D indicate an excess in ABBA tree topologies relative to BABA topologies. Under a model of incomplete lineage sorting and no introgression, we would expect these topologies in equal proportion, so an excess in ABBA topologies suggests introgression between taxa P2 and P3. Using this test, we demonstrated evidence of introgression between ACB and ABB at both autosomal markers and trait loci (Table 2). While all values were significant, the relative strength of the introgression was substantially higher at some barrier trait loci, suggesting that alleles at these loci may have adaptive significance in the recipient species.

Table 3.2. D statistics to detect introgression between ABB and ACB.

Region	P1	P2	P3	Dstatistic	Z-score	p-value
Autosomes	ECB	ABB	ACB	0.0371256	4.60807	**4.06e-06**
pgFAR	ECB	ABB	ACB	0.422881	6.64779	**< 10e-10**
bab	ECB	ABB	ACB	0.610141	13.1012	**< 10e-10**
period	ECB	ABB	ACB	0.219795	2.73911	**0.00616055**
Pdfr	ECB	ABB	ACB	0.451177	5.16666	**2.38e-07**

We then subdivided the ECB samples into groups associated with each pheromone category,

having previously demonstrated the importance of this trait in distinguishing ECB population structure at these genomic regions. We tested both ABB and ACB as potential sources of introgression in separate tests. We were not able to detect evidence of introgression between ACB or ABB into either E or Z types using markers from autosomal scaffolds (Table 3.3). However, applying the test to scaffolds containing a trait locus demonstrated excess sharing of alleles, consistent with positive D statistics and a signature of introgression between E clade ECB and both ACB and ABB (Table 3.3). The only exception was found at the *Pdfr* trait locus, where D statistics showed elevation in tests of introgression between ABB and the E type ECB, but not between ACB and either pheromone type.

Table 3.3. D statistic for trios of ECB-E, ECB-Z, ACB, and ABB

	P1	**P2**	**P3**	**D statistic**	**Z-score**	**p-value**
Autosome	Z	E	ABB	-0.0109257	1.81226	0.0699462
	Z	E	ACB	-0.0075653	1.25101	0.210931
pgFAR	Z	E	ABB	0.577758	6.52563	**< 10e-10**
	Z	E	ACB	0.327589	5.03494	**4.78E-07**
bab	Z	E	ABB	0.816847	25.2699	**< 10e-10**
	Z	E	ACB	0.718913	19.0884	**< 10e-10**
period	Z	E	ABB	0.40921	5.03583	**4.76E-07**
	Z	E	ACB	0.39644	6.64815	**< 10e-10**
Pdfr	Z	E	ABB	0.29647	4.60414	**4.14E-06**
	Z	E	ACB	0.0330259	0.779027	0.435964

To determine whether these scaffold wide signals were driven by alleles at the trait loci themselves, we applied a windowed test across each scaffold and a similarly sized autosomal scaffold for comparative purposes. We plotted the *df* statistic (Pfeifer and Kapan 2019) since the null expectation of no introgression should be symmetric around 0 and the statistic is bounded by -1 and 1 in all scenarios, with values near 1 suggestive of introgression between taxa P2 and P3 and

values near -1 suggestive of excess shared alleles between taxa P1 and P3. We considered whether the similarity between E and ACB/ABB was centered on genic windows by plotting statistics with Z, E, and ACB or ABB as taxa P1, P2, and P3 (Figure 3.5). At our comparative autosomal scaffold, we did not recover any signature consistent with introgression between groups (Figure 3.5A). At *pgFAR*, a clear peak was centered over the trait locus, suggesting that pheromone production alleles were shared between ECB-E and both ACB and ABB (Figure 3.5B). The *bab* scaffold was characterized by a broad elevation in *df* (Figure 3.5C) across the length of the scaffold. *bab* exons are dispersed across this scaffold, separated by long (>100kb) introns. These regions contain SNPs demonstrated in previous studies to be significantly differentiated between pheromone response phenotypes. They also show elevated cross-chromosomal LD with *pgFAR,* suggesting of strong selective pressure to maintain this association (Unbehend et al. 2021). The highest density of these SNPs occurs around the 300kb mark on the scaffold, which appears to be the *df* "peak" in our data. This could indicate a mechanism for adaptive introgression of *bab* alleles between *Ostrinia* taxa.

While mean values of *df* were positive across the *period* and *Pdfr* scaffolds (Table 3.3), indicative of excess sharing of alleles between ECB-E and ACB or ABB, there were no clear signals of a particular portion of the scaffold that contained a "peak" (Figure 3.5D,E). However, it is notable that all three Z-linked trait loci linked showed broad, scaffold wide elevations in *df*, while the autosomal *pgFAR* showed a peak localized on the genic region. The sex chromosome in ECB has been shown to play an outsized role in driving differentiation ECB subpopulations (Dopman et al. 2010, Kozak et al. 2019, Kunerth et al. 2022), as it harbors multiple barrier trait loci and, in many populations, appears to contain a large chromosomal rearrangement. These features lead to elevated LD across long stretches of the chromosome. This suggests that sharing of alleles between

Ostrinia taxa may be subject to 'divergence hitchhiking' (Via & West 2008, Via 2012), wherein introgression of one trait locus carries with it long stretches of shared variation in a broad genomic region surrounding it. In contrast, the freely recombining autosomes are quicker to break up this linkage, limiting the signal of shared excess variation to the source itself, in this case the *pgFAR* gene.

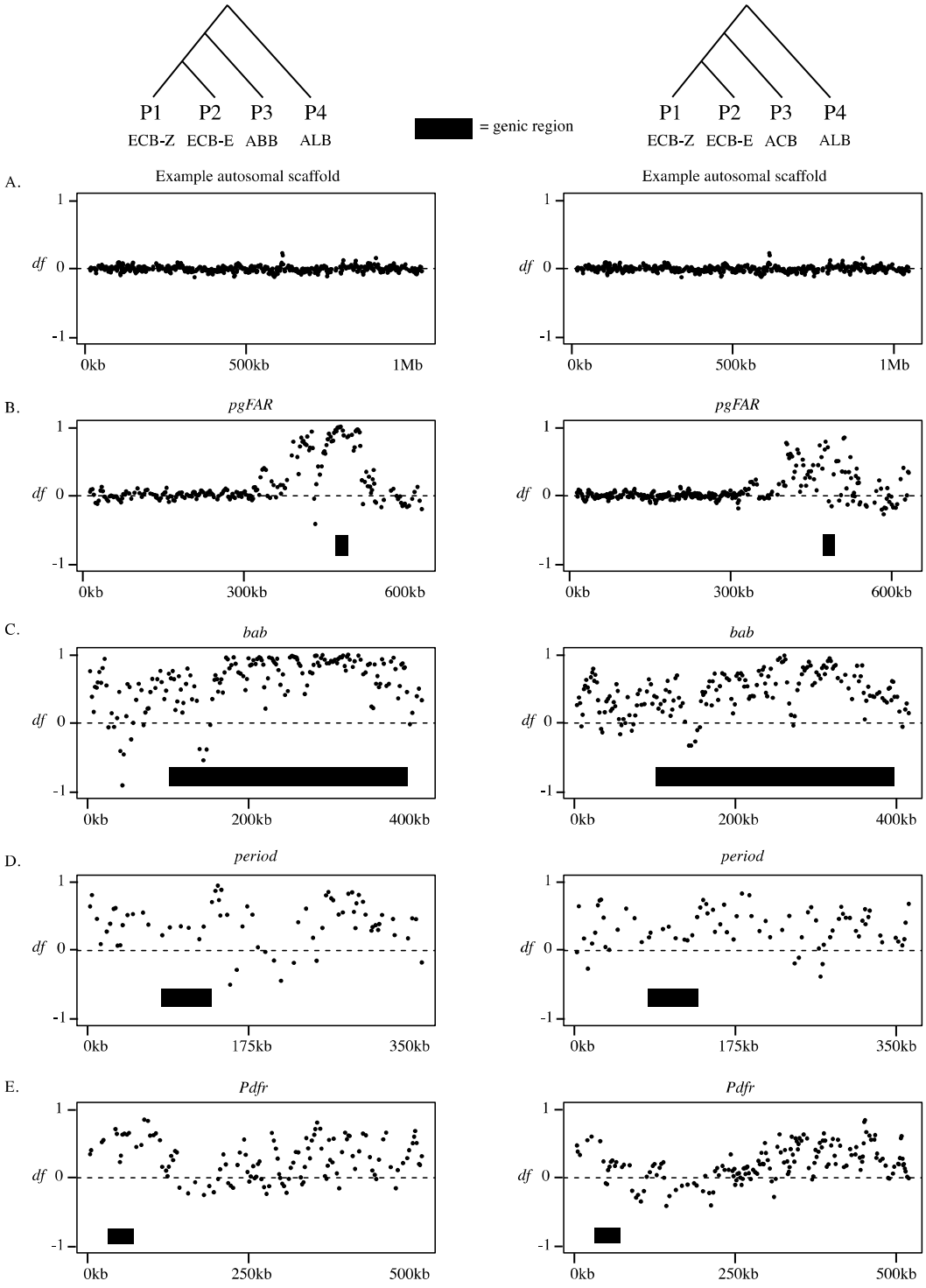

Figure 3.5. *df* statistics for an example autosomal scaffold and the four trait scaffolds in comparisons between ECB-Z, ECB-E, and either ABB or ACB, with ALB as the outgroup. Positive *df* scores indicate excess allele sharing between taxa P2 and P3.

Conclusions

In the absence of gene flow, isolated populations will accrue genetic differences. As these differences are passed on to subsequent generations, they will eventually result in unique shared ancestry patterns throughout the genome. The rate at which a population's ancestry separates it from other isolated populations is dependent on the demography, mutation rate, mating systems, and selection within the populations. This process can be further complicated by introducing gene flow between populations, even those separated for thousands or even millions of generations. Should introgression occur, ancestry patterns will reflect this in shared genealogical sequences across populations or even species as the genomic regions move between these groups.

Here we attempt to infer this process in a system in which some barrier loci are well characterized. We explore genome wide tracts of shared ancestry and compare them to the ancestry patterns underlying the known barrier traits. We demonstrate how the signatures underlying these reproductive isolation phenotype loci explain the population structure within *O. nubilalis*. We then demonstrate that, within the *Ostrinia* group, there is evidence of introgression between species. This result is consistent with prior studies that have also demonstrated introgression between sympatric species of *Ostrinia* moths (Malausa et al. 2007; Wang et al. 2017). Here we show strong patterns of introgression between Asian Corn Borer (ACB) and Adzuki Bean Borer (ABB) across all genomic regions (Table 3.2). These species are both a part of a shallowly diverged clade, and a recent phylogenetic analysis by Yang et al. (2021) found discordance between ACB and ABB in their relationship to ECB using different tree building

methods (Maximum likelihood vs. ASTRAL). They argue that similarities between ACB and ECB may be the result of convergence. Here we show that the difficulty in resolving species tree relationships may be driven by introgression as well.

We also demonstrated evidence of the introgression of key trait loci between both ECB-E pheromone moths and both ACB and ABB, based on excess sharing of derived alleles relative to sharing between Z types and ACB or ABB. ECB and ABB both utilize the same E and Z isomers in blends of 11-14 monounsaturated tetradecyl acetates in their pheromone blends, while ACB utilizes a 12-14 molecule (the numbers indicate the placement of the double bond) (Lassance et al. 2013). This is consistent with a history in which alleles that control pheromone phenotypes arose in one species before introgressing into one or more sister taxa. It is notable that this pattern holds for both the autosomal *pgFAR* locus and the *bab* locus. If variation for both pheromone production (*pgFAR*) and response (*bab*) evolved in one taxon, it would appear reasonable for selection to favor the introgression of both associated variants into the recipient species, since the evolution or introgression of just one novel sex-communication phenotype should generally be disfavored, as there would be a lack of either signal senders or receivers within the population (Groot et al. 2016; De Pasqual et al. 2021). In a scenario in which reproductive barriers are incomplete and either signal or response is adaptive in the recipient, introgression from a source species allows this challenge to be overcome, since both genes can be transmitted together.

In such a scenario we can also explain the finding of introgression of the *period* locus between E strain ECB and both ACB and ABB as well as introgression of the *Pdfr* locus between E strain

ECB and ABB. These loci are both Z-linked, and found within (in the case of *period*) and nearby (in the case of *Pdfr*), a major recombination suppressor (Kozak et al. 2017). Further, the *bab* locus is also located within 1 Mbp of the inferred breakpoints of this putative inversion. In a scenario in which any one of these alleles was adaptive in a recipient species, it may have led to a cascade of introgression of the other loci, even as the rest of the genome remained distinct. Such models of "speciation supergenes" have found recent support in other taxa (Thompson & Jiggins 2014; Cohen & Privman 2020, reviewed in Taylor & Campagna 2016).

While our findings are limited by the number of samples in non-ECB *Ostrinia* species, especially those in known Z-type pheromone categories, we argue that our results are consistent with sharing of alleles between taxa within this group at key barrier loci. We suggest that future studies could explore a broader dataset of *Ostrinia* moths and potentially estimate the relative ages of the variation in barrier alleles. Determining the order of evolution and selective strength of the barrier phenotypes may contribute toward our understanding of a largely unexplored path for the process of speciation.

METHODS

Sample collection

We re-sequenced complete genomes of 128 *Ostrinia*: 119 *O. nubilalis* (ECB) were collected from populations in the Northeastern United States, either from pheromone traps as adults or from corn

stubble as larva. Eight samples came from three outgroup taxa within the *Ostrinia* group: one Adzuki bean borer (ABB), *O. scapulalis,* four Asian Corn Borer (ACB), *O. furnacalis*, and three American lotus borer (ALB), *O. penitalis.* We also sequenced one E-strain *O. nubilalis* female pupa from an inbred lab culture maintained at Tufts University in Boston, Massachusetts.

Library preparation and sequencing

Approximately one microgram of genomic DNA was sheared to 700 bp with a Covaris S2 sonicator. Sheared DNA was treated with 1x Ampure XP and eluted in 30 µl Qiagen 0.5 x AE buffer. Samples were end polished with T4 polynucleotide kinase and T4 DNA polymerase, adenylated with Taq polymerase, and ligated to an Illumina Truseq +A adapter. Individual samples were indexed by PCR, pooled, and sequenced on an Illumina NextSeq 500 (2 x 150 bp reads). Sequence data were analyzed for quality by FastQC (Andrews 2010) and multiQC software.

Reads were processed following standard protocols using the GATK pipeline
(Van der Auwera et al. 2013) using recommendations for best practice
(https://www.broadinstitute.org/gatk/guide/best-practices). Briefly, demultiplexed reads that passed Illumina sequencing quality filters were trimmed of any remaining adapter sequences using Trimmomatic v.35. Reads with a phred quality score (q) less than 15 and reads less than 36 base pairs long were removed. Remaining reads were mapped to the ECB draft genome assembly (GenBank: PRJNA534504; BioSample SAMN11491597; accession SWFO00000000) using bwa (Li 2014). Duplicated sequences were removed (Li et al. 2009) and SNPs and indels were called. We filtered the resulting genotypes for quality (>30) and coverage depth (>10) based on the best

practices recommendations of the Broad Institute.

Using VCFtoolsv0.1.16 (Danecek et al. 2011), we separated the genome into different regions for subsequent analysis: (1) fully diploid autosomal scaffolds, (2) the Z-chromosome, which is diploid in males and haploid in females, (3) autosomal scaffold 178 (627,257 bp), containing the pheromone production locus *pgFAR*, (4) Z-linked scaffold bab (x bp), containing the pheromone response locus *bric-a-brac*, (5) Z-linked scaffold 87 (517,236 bp), containing the voltinism locus *pdfr*, and (6) the Z-linked scaffold 532 (371,787 bp), containing the voltinism locus *period*. We compared each of these regions separately for heterozygosity, site missingness and individual missingness for quality control assurance.

Phasing and ancestral allele estimation

Scaffolds were indexed using BCFtoolsv1.15.1 (Danecek et al. 2021) and phased using SHAPEITv4.2.2 (Delaneau et al. 2019). To generate recombination maps for each scaffold, we assumed a uniform recombination rate of 3.3 cM/Mb for autosomes and 1.6 cM/Mb for the sex chromosomes, based on prior estimates in *O. nubilalis* (Dopman et al. 2004). We estimated ancestral allele frequencies using est-sfs v 2.01, which uses the approach developed in Keightley et al. (2016) and Keightley & Jackson (2018). We utilized the ALB and ABB samples as outgroups. Since est-sfs can only utilize one allele from each outgroup, we selected the most common allele within each outgroup sample and split ties randomly.

Identification of population structure

To visualize baseline genetic structure, we performed principal components analysis (PCA) using the smartpca package of eigenstrat (Price et al. 2006). We tested each genomic region separately, utilizing all SNPs across the scaffold that contained the trait locus. Separate plots were created so individuals could be colored by geographic location or pheromone/phenology category (Figure 3.2, Figure S3.1, S3.2). Individuals were assigned to pheromone, geographic, and PDD categories based on prior phenotype estimation whenever possible.

Inference of tree sequences and haplotype ages

We utilized *tsinfer* v 0.2.3 (Kelleher et al. 2019) to generate a unified genealogical tree sequence for each scaffold. We utilized each tree sequence to find each individual's genealogical nearest neighbors (GNN), a method utilized in the reconstruction of human evolutionary history (Kelleher et al. 2019). For our purposes, we examined, for every target region of the genome, the GNNs of every individual moth in the dataset, defining the reference sets by taxon, pheromone category, and/or phenology category. We then plotted the resulting GNN proportions, which will total 1 (similar to a STRUCTURE plot, coloring each proportion by the reference taxa/phenotype examined).

D-statistics and introgression

We calculated D-statistics utilizing the software package Dsuite v 0.4 (Malisky et al. 2021). D-

statistics are used in the inference of historical introgression between trios of focal taxa or populations and an outgroup. Most commonly implemented as an ABBA-BABA test (Green et al. 2010, Durand et al. 2011), D-statistics are implemented on SNPs throughout the genome and utilize construction of tree topologies between the four taxa at every site. If no introgression has occurred, the only differences from the species tree should be the result of incomplete lineage sorting. However, introgression between taxa since the split from the outgroup is theorized to result in an excess of tree topologies uniting the taxa between which the introgression occurred. Here we estimated D-statistics using ALB as the outgroup taxon and performed two sets of tests between the remaining trios. First, we tested for introgression between trios of ALB, ABB, and ECB, using all ECB individuals together. Next, we tested trios after separating ECB into an E and a Z category. We calculated both the D statistic and the *df* statistic (Pfeifer and Kaplan 2019), which is better suited for exploring sliding windows along a scaffold. *df* utilizes correlated allele frequencies and is bounded by -1 and 1, which allows for more accurate comparisons of the amount of introgression than related measures.

LITERATURE CITED

Abrieux, A, Xue, Y, Cai, Y, Lewald, KM, Nguyen, HN, Zhang, Y, Chiu, JC. 2020. EYES ABSENT and TIMELESS integrate photoperiodic and temperature cues to regulate seasonal physiology in Drosophila. *Proceedings of the National Academy of Sciences USA*. 117, 15293-15304.

Allison, JD, Cardé, RT. 2016. Pheromones: reproductive isolation and evolution in moths. *Pheromone Communication in Moths.* 11-24. Berkeley: University of California Press.

Andrews, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anglade, P, Stockel, J, I.W.G.O. Cooperators. 1984. Intraspecific sex-pheromone variability in the European corn borer, Ostrinia nubilalis Hübner (Lepidoptera, Pyralidae). *Agronomie*. 4(2), 183-187.

Barton, N, Bengtsson, BO. 1986. The barrier to genetic exchange between hybridising populations. *Heredity*. 57(3), 357-376.

Barton, NH, De Cara, MAR. 2009. The evolution of strong reproductive isolation. *Evolution*. 63(5), 1171-1190.

Barton, NH, Hewitt, GM. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics*. 16, 113-148.

Barton, NH, Hewitt, GM. 1989. Adaptation, speciation and hybrid zones. *Nature*. 341(6242), 497-503.

Barton, NH. 1983. Multilocus clines. *Evolution*. 37(3), 454-471.

Barton, NH. 2013. Does hybridization influence speciation? *Journal of Evolutionary Biology*. 26(2), 267-269.

Becherel, OJ, Yeo, AJ, Stellati, A, Heng, EY, Luff, J, Suraweera, AM, Woods, R, Fleming, J, Carrie, D, McKinney, K, Xu, X. 2013. Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. *PLoS Genetics*. 11, e1003435.

Berry, A, Kreitman, M. 1993. Molecular analysis of an allozyme cline: alcohol dehydrogenase in Drosophila melanogaster on the east coast of North America. *Genetics*. 134, 869-893.

Bethenod, MT, Thomas, Y, Rousset, F, Frérot, B, Pélozuelo, L, Genestier, G, Bourguet, D. 2005. Genetic isolation between two sympatric host plant races of the European corn borer, Ostrinia nubilalis Hübner. II: assortative mating and host-plant preferences for oviposition. *Heredity*. 94(2), 264-270.

Bird, JE, Takagi, Y, Billington, N, Strub, MP, Sellers, JR, Friedman, TB. 2014. Chaperone-enhanced purification of unconventional myosin 15, a molecular motor specialized for stereocilia protein trafficking. *Proceedings of the National Academy of Sciences USA*. 111, 12390-12395.

Bolger, AM, Lohse, M, Usadel, B. 2014. Trimmomatic, A flexible trimmer for Illumina Sequence Data. *Bioinformatics*. 30(15), 2114-2120.

Boube, M, Joulia, L, Cribbs, DL, Bourbon, HM. 2002. Evidence for a mediator of RNA polymerase II transcriptional regulation conserved from yeast to man. *Cell*. 110, 143-151.

Bourguet, D, Genissel, A, Raymond, M. 2000. Insecticide resistance and dominance levels. *Journal of Economic Entomology*. 93(6), 1588-1595.

Bradshaw, WE, Holzapfel, CM. 2001. Genetic shift in photoperiodic response correlated with global warming. *Proceedings of the National Academy of Sciences USA*. 98, 14509-14511.

Bradshaw, WE, Holzapfel, CM. 2010. What season is it anyway? Circadian tracking vs. photoperiodic anticipation in insects. *Journal of Biological Rhythms*. 25, 155-165.

Brodsky, B, Shah, NK. 1995. The triple□helix motif in proteins. *The FASEB Journal*. 9, 1537-1546.

Broman, KW, Wu, H, Sen, Ś, Churchill, GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* 19, 889-890.

Brousseau, L, Nidelet, S, Streiff, R. 2018. New WGS data and annotation of the heterosomal vs. autosomal localization of Ostrinia scapulalis (Lepidoptera, Crambidae) nuclear genomic scaffolds. *Data in Brief*. 20, 644-648.

Butlin, R, Ritchie, MG. 2001. Searching for speciation genes. *Nature.* 412(6842), 31-33.

Butlin, RK, Servedio, MR, Smadja, CM, Bank, C, Barton, NH, Flaxman, SM, ... Qvarnström, A. 2021. Homage to Felsenstein 1981, or why are there so few/many species? *Evolution*. 75(5), 978-988.

Butlin, RK, Smadja, CM. 2018. Coupling, reinforcement, and speciation. *The American Naturalist*. 191(2), 155-172.

Caballero, A. 1995. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics*. 139(2), 1007-1011.

Caffrey, DJ, Worthley, LH. 1927. A progress report on the investigations of the European corn borer (No. 1476). Washington, D.C.: US Government Printing Office.

Calcagno, V, Thomas, Y, Bourguet, D. 2007. Sympatric host races of the European corn borer: adaptation to host plants and hybrid performance. *Journal of Evolutionary Biology*. 20(5), 1720-1729.

Calkin, AC, Goult, BT, Zhang, L, Fairall, L, Hong, C, Schwabe, JW, Tontonoz, P. 2011. FERM-dependent E3 ligase recognition is a conserved mechanism for targeted degradation of lipoprotein receptors. *Proceedings of the National Academy of Sciences USA*. 108, 20107-20112.

Calvin, DD, Song, PZ. 1994. Variability in postdiapause development periods of geographically separate Ostrinia nubilalis (Lepidoptera: Pyralidae) populations in Pennsylvania. *Environmental Entomology*. 23, 431-436.

Charlesworth, B. 2012. The effects of deleterious mutations on evolution at linked sites. *Genetics*. 190(1), 5-22.

Chen, Q, Zhang, L, Zhang, Y, Mao, Q, Wei, T. 2017. Tubules of plant reoviruses exploit tropomodulin to regulate actin-based tubule motility in an insect vector. *Scientific Reports*. 7, e38563.

Coates, BS, Johnson, H, Kim, KS, Hellmich, RL, Abel, CA, Mason, C, Sappington, TW. 2013. Frequency of hybridization between Ostrinia nubilalis E□ and Z□ pheromone races in regions of sympatry within the United States. *Ecology and Evolution*. 3(8), 2459-2470.

Coates, BS, Kozak, GM, Kim, K, Sun, J, Wang, Y, Fleischer, SJ, Dopman, EB, Sappington, TW. 2019. Influence of host plant, geography and pheromone strain on genomic differentiation in sympatric populations of Ostrinia nubilalis. *Molecular Ecology*. 28(19), 4439-4452.

Coates, BS, Siegfried, BD. 2015. Linkage of an ABCC transporter to a single QTL that controls Ostrinia nubilalis larval resistance to the Bacillus thuringiensis Cry1Fa toxin. *Insect Biochemical Molecular Biology*. 63, 86-96.

Cohen, P, Privman, E. 2020. The social supergene dates back to the speciation time of two Solenopsis fire ant species. *Scientific Reports*. 10, 11538.

Coughtrie MW. 2002. Sulfation through the looking glass—recent advances in sulfotransferase research for the curious. *The Pharmacogenomics Journal*. 2, 297-308.

Coyne, JA, Orr, HA. 1989 Two rules of speciation. *Speciation and Its Consequences.* 180-207. Sunderland, MA: Sinauer.

Coyne, JA, Orr, HA. 2004. *Speciation*. Sunderland, MA: Sinauer Associates.

Crow, JF, Kimura, M. 1970. *An Introduction to Population Genetics Theory*. Scientific Publishers.

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*. 23, 3133-3157.

Dagilis, AJ, Kirkpatrick, M. 2016. Prezygotic isolation, mating preferences, and the evolution of chromosomal inversions. *Evolution*. 70(7), 1465-1472.

Danecek, P, Auton, A, Abecasis, G, Albers, CA, Banks, E, DePristo, MA, Handsaker, RE, Lunter, G, Marth, GT, Sherry, ST, McVean, G. 2011. The variant call format and VCFtools. *Bioinformatics*. 27(15), 2156-2158.

Danecek, P, Bonfield, JK, Liddle, J, Marshall, J, Ohan, V, Pollard, MO, Whitwham, A, Keane, T, McCarthy, SA, Davies, RM, Li, H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience*. 10(2), giab008.

De Pasqual, C, Groot, AT, Mappes, J, Burdfield-Steel, E. 2021. Evolutionary importance of intraspecific variation in sex pheromones. *Trends in Ecology and Evolution*. 36(9), 848-859.

Delaneau, O, Zagury, JF, Robinson, MR, Marchini, JL, Dermitzakis, ET. 2019. Accurate, scalable and integrative haplotype estimation. *Nature Communications*. 10, 5436.

Dickson, KS, Bilger, A, Ballantyne, S, Wickens, MP. 1999. The cleavage and polyadenylation specificity factor in Xenopus laevis oocytes is a cytoplasmic factor involved in regulated polyadenylation. *Molecular and Cellular Biology*. 19, 5707-5717.

Ding X, Liu J, Zheng L, Song J, Li N, Hu H, Tong X, Dai F. 2019. Genome-wide identification and expression profiling of Wnt family genes in the silkworm, Bombyx mori. *International Journal of Molecular Sciences.* 20, 1221.

Doellman, MM, Ragland, GJ, Hood, GR, Meyers, PJ, Egan, SP, Powell, TH, Lazorchak, P, Glover, MM, Tait, C, Schuler, H, Hahn, DA. 2018. Genomic differentiation during speciation-with-gene-flow: Comparing geographic and host-related variation in divergent life history adaptation in Rhagoletis pomonella. *Genes*. 9(5), 262.

Dopman, EB, Bogdanowicz, SM, Harrison, RG. 2004. Genetic mapping of sexual isolation between E and Z pheromone strains of the European corn borer (Ostrinia nubilalis). *Genetics*. 167, 301-309.

Dopman, EB, Pérez, L, Bogdanowicz, SM, Harrison, RG. 2005. Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proceedings of the National Academy of Sciences USA*. 102, 14706-14711.

Dopman, EB, Robbins, PS, Seaman, A. 2010. Components of reproductive isolation between North American pheromone strains of the European corn borer. *Evolution.* 64(4), 881-902.

Dopman, EB. 2011. Genetic hitchhiking associated with life history divergence and colonization of North America in the European corn borer moth. *Genetica*. 139(5), 565-573.

Dorant, Y, Benestan, L, Rougemont, Q, Normandeau, E, Boyle, B, Rochette, R, Bernatchez, L. 2019. Comparing Pool□seq, Rapture, and GBS genotyping for inferring weak population structure: The American lobster (Homarus americanus) as a case study. *Ecology and Evolution*. 9, 6606-6623.

Dray, S, Dufour, AB. 2007. The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*. 22(4), 1-20.

Durand, EY, Patterson, N, Reich, D, Slatkin, M. 2011. Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*. 28(8), 2239-2252.

Excoffier L, Foll M, Petit RJ. 2009. Genetic consequences of range expansions. *Annual Review*

of Ecology, Evolution, and Systematics. 40, 481-501.

Farkaš, R, Knopp, J. 1997. Ecdysone□modulated response of Drosophila cytosolic malate dehydrogenase to juvenile hormone. *Archives of Insect Biochemistry and Physiology*. 35, 71-83.

Feder, JL, Nosil, P, Wacholder, AC, Egan, SP, Berlocher, SH, Flaxman, SM. 2014. Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *Journal of Heredity*. 105(S1), 810-820.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*. 17(6), 368-376.

Feng X, Lu T, Li J, Yang R, Hu L, Ye Y, Mao F, He L, Xu J, Wang Z, Liu Y. 2020. The tumor suppressor interferon regulatory factor 2 binding protein 2 regulates hippo pathway in liver cancer by a feedback loop in mice. *Hepatology*. 71, 1988-2004.

Fisher, KE, Mason, CE, Flexner, JL, Hough□Goldstein, J, McDonald, JH. 2017. Survivorship of Z pheromone strain European corn borer (Lepidoptera: Crambidae) on a range of host plants varying in defensive chemistry. *Journal of Economic Entomology*. 110(3), 978-985.

Flaxman, SM, Wacholder, AC, Feder, JL, Nosil, P. 2014. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*. 23(16), 4074-4088.

François, A, Grauso, M, Demondion, E, Bozzolan, F, Debernard, S, Lucas, P. 2012. Bestrophin-encoded Ca2+-activated Cl− channels underlie a current with properties similar to the native current in the moth Spodoptera littoralis olfactory receptor neurons. *PLoS One*. 7, 52691.

Friesen, PD, Nissen, MS. 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. *Molecular and Cellular Biology*. 10, 3067-3077.

Frolov, AN, Bourguet, D, Ponsard, S. 2007. Reconsidering the taxomony of several Ostrinia species in the light of reproductive isolation: a tale for Ernst Mayr. *Biological Journal of the Linnean Society*. 91(1), 49-72.

Fuhrmann, N, Prakash, C, Kaiser, TS. 2021. Polygenic adaptation from standing genetic variation allows rapid ecotype formation. bioRxiv.

Gagnaire, PA, Pavey, SA, Normandeau, E, Bernatchez, L. 2013. The genetic architecture of reproductive isolation during speciation□with□gene□flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*. 67(9), 2483-2497.

Gaspar, P, Holder, MV, Aerne, BL, Janody, F, Tapon, N. 2015. Zyxin antagonizes the FERM protein expanded to couple F-actin and Yorkie-dependent organ growth. *Current Biology*. 25, 679-689.

Gautier, M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*. 201, 1555-1579.

Gee, MA, Heuser, JE, Vallee, RB. 1997. An extended microtubule-binding structure within the dynein motor domain. *Nature*. 390, 636-639.

Giannakouros, T, Nikolakaki, E, Mylonis, I, Georgatsou, E. 2011. Serine□arginine protein kinases: a small protein kinase family with a large cellular presence. *The FEBS Journal*. 278, 570-586.

Glover, TJ, Campbell, MG, Linn, CE, Roelofs, WL. 1991. Unique sex chromosome mediated behavioral response specificity of hybrid male European corn borer moths. *Experientia*. 47(9), 980-984.

Glover, TJ, Robbins, PS, Eckenrode, CJ, Roelofs, WL. 1992. Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. *Archives of Insect Biochemistry and Physiology*. 20, 107-117.

Gopinath, G, Arunkumar, KP, Mita, K, Nagaraju, J. 2016. Role of Bmznf-2, a Bombyx mori CCCH zinc finger gene, in masculinisation and differential splicing of Bmtra-2. *Insect Biochemistry and Molecular Biology*. 75, 32-44.

Gotthard, K, Wheat, CW. 2019. Diapause: Circadian clock genes are at it again. *Current Biology*. 29, R1245-R1246.

Green, RE, Krause, J, Briggs, AW, Maricic, T, Stenzel, U, Kircher, M, Patterson, N, Li, H, Zhai, W, Fritz, MHY, Hansen, NF. 2010. A draft sequence of the Neandertal genome. *Science*. 328(5979), 710-722.

Groot, AT, Dekker, T, Heckel, DG. 2016. The genetic basis of pheromone evolution in moths. *Annual Review of Entomology*. 61, 99-117.

Guo, R, Wang, S, Xue, R, Cao, G, Hu, X, Huang, M, Zhang, Y, Lu, Y, Zhu, L, Chen, F, Liang, Z. 2015. The gene expression profile of resistant and susceptible Bombyx mori strains reveals cypovirus-associated variations in host gene transcript levels. *Applied Microbiology and Biotechnology*. 99, 5175-5187.

Guo, Z, Qin, J, Zhou, X, Zhang, Y. 2018. Insect transcription factors: a landscape of their structures and biological functions in Drosophila and beyond. *International Journal of Molecular Sciences*. 19, 3691.

Hajek, KL, Friesen, PD. 1998. Proteolytic processing and assembly of gag and gag-pol proteins of TED, a baculovirus-associated retrotransposon of the gypsy family. *Journal of Virology*. 72, 8718-8724.

Harrison, RG, Larson, EL. 2014. Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*. 105(S1), 795-809.

Harrison, RG. 1990. Hybrid zones: windows on evolutionary process. *Oxford Surveys in Evolutionary Biology*. 7, 69-128.

Harrison, RG. 1998. Linking evolutionary pattern and process. *Endless Forms*. 19-31. Oxford University Press.

Hodgson, BE. 1928. *The Host Plants of the European Corn Borer in New England* (No. 77). US Department of Agriculture.

Holley, SA, Jackson, PD, Sasai, Y, Lu, B, Robertis, EMD, Hoffmann, FM, Ferguson, EL. 1995. A conserved system for dorsal-ventral patterning in insects and vertebrates involving sog and chordin. *Nature*. 376, 249-253.

Hooper, SL, Hobbs, KH, Thuma, JB. 2008. Invertebrate muscles: thin and thick filament structure; molecular basis of contraction and its regulation, catch and asynchronous muscle. *Progress in Neurobiology*. 86, 72-127.

Hu, J, Du, Y, Meng, M, Dong, Y, Peng, J, 2021. Development of two continuous hemocyte cell sublines in the Asian corn borer Ostrinia furnacalis and the identification of molecular markers for hemocytes. *Insect Science*. 28, 1382-1398.

Hughes ME, Hogenesch JB, Kornacker K. 2010. JTK_CYCLE: an efficient nonparametric algorithm for detecting rhythmic components in genome-scale data sets. *Journal of Biological*

Rhythms. 25, 372-80.

Ikegaya, M, Miyazaki, T, Park, EY. 2021. Biochemical characterization of Bombyx mori α□N□ acetylgalactosaminidase belonging to the glycoside hydrolase family 31. *Insect Molecular Biology*. 30, 367-378.

Ingham, PW, Taylor, AM, Nakano, Y. 1991. Role of the Drosophila patched gene in positional signalling. *Nature*. 353, 184-187.

Jaramillo, AM, Zheng, X, Zhou, Y, Amado, DA, Sheldon, A, Sehgal, A, Levitan, IB. 2004. Pattern of distribution and cycling of SLOB, Slowpoke channel binding protein, in Drosophila. *BMC Neuroscience*. 5, 3.

Jia ZQ, Liu D, Peng YC, Han ZJ, Zhao CQ, Tang T. 2020. Identification of transcriptome and fluralaner responsive genes in the common cutworm Spodoptera litura Fabricius, based on RNA-seq. *BMC Genomics*. 21, 120.

Jiggins, CD, Mallet, J. 2000. Bimodal hybrid zones and speciation. *Trends in Ecology and Evolution*. 15(6), 250-255.

Jongsma, MA, Bolter, C. 1997. The adaptation of insects to plant protease inhibitors. *Journal of Insect Physiology*. 43, 885-895.

Jurenka, RA. 2021. Lepidoptera: Female sex pheromone biosynthesis and its hormonal regulation. *Insect Pheromone Biochemistry and Molecular Biology.* Academic Press.13-88.

Kautt, AF, Kratochwil, CF, Nater, A, Machado-Schiaffino, G, Olave, M, Henning, F, ... Meyer, A. 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature*. 588(7836), 106-111.

Kawaoka, S, Katsuma, S, Meng, Y, Hayashi, N, Mita, K, Shimada, T. 2009. Identification and characterization of globin genes from two lepidopteran insects, Bombyx mori and Samia cynthia ricini. *Gene*. 431, 33-38.

Keightley, PD, Campos, JL, Booker, TR, Charlesworth, B. 2016. Inferring the frequency spectrum of derived variants to quantify adaptive molecular evolution in protein-coding genes of Drosophila melanogaster. *Genetics*. 203(2), 975-984.

Keightley, PD, Jackson, BC. 2018. Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics*. 209(3), 897-906.

Kelleher, J, Wong, Y, Wohns, AW, Fadil, C, Albers, PK, McVean, G. 2019. Inferring whole-genome histories in large population datasets. *Nature Genetics*. 51(9), 1330-1338.

Kim, CG, Hoshizaki, S, Huang, YP, Tatsuki, S, Ishikawa, Y. 1999. Usefulness of mitochondrial COII gene sequences in examining phylogenetic relationships in the Asian corn borer, Ostrinia furnacalis, and allied species (Lepidoptera: Pyralidae). *Applied Entomology and Zoology*. 34(4), 405-412.

Kim, Y, Jung, S, Kim, Y, Lee, Y. 2011. Real-time monitoring of oriental fruit moth, Grapholita molesta, populations using a remote sensing pheromone trap in apple orchards. *Journal of Asia-Pacific Entomology*. 14(3), 259-262.

Kirkpatrick, M, Barrett, B. 2015. Chromosome inversions, adaptive cassettes and the evolution of species' ranges. *Molecular Ecology*. 24(9), 2046-2055.

Kirkpatrick, M, Barton, N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics.* 173(1), 419-434.

Klun, JA, Chapman, OL, Mattes, KC, Wojtkowski, PW, Beroza, M,, Sonnet, PE. 1973. Insect sex pheromones: minor amount of opposite geometrical isomer critical to attraction. *Science.* 181, 661–663.

Klun, JA. 1975. Insect sex pheromones: intraspecific pheromonal variability of *Ostrinia nubilalis* in North America and Europe. *Environmental Entomology*. 4, 891–894.

Kochansky, J, Cardé, RT, Liebherr, J, Roelofs, WL. 1975. Sex pheromone of the European corn borer, Ostrinia nubilalis (Lepidoptera: Pyralidae), in New York. *Journal of Chemical Ecology*. 1(2), 225-231.

Kofler, R, Pandey, RV, Schlötterer, C. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*. 27, 3435-3436.

Koštál, V. 2006. Eco-physiological phases of insect diapause. *Journal of Insect Physiology*. 52, 113-127.

Kozak, GM, Wadsworth, CB, Kahne, SC, Bogdanowicz, SM, Harrison, RG, Coates, BS, Dopman, EB. 2017. A combination of sexual and ecological divergence contributes to rearrangement spread during initial stages of speciation. *Molecular Ecology*. 26(8), 2331-2347.

Kozak, GM, Wadsworth, CB, Kahne, SC, Bogdanowicz, SM, Harrison, RG, Coates, BS, Dopman, EB. 2019. Genomic basis of circannual rhythm in the European corn borer moth. *Current Biology*. 29(20), 3501-3509.

Kozak, GM, Wadsworth, CB, Kahne, SC, Bogdanowicz, SM, Harrison, RG, Coates, BS, Dopman, EB. 2017. A combination of sexual and ecological divergence contributes to rearrangement spread during initial stages of speciation. *Molecular Ecology*. 26(8), 2331-2347.

Kroemer, JA, Coates, BS, Nusawardani, T, Rider, SD, Fraser, LM, Hellmich, RL. 2011. A rearrangement of the Z chromosome topology influences the sex-linked gene display in the European corn borer, Ostrinia nubilalis. *Molecular Genetics & Genomics*. 286, 37-56

Kunerth, HD, Bogdanowicz, SM, Searle, JB, Harrison, RG, Coates, BS, Kozak, GM, Dopman, EB. 2022. Consequences of coupled barriers to gene flow for the build□up of genomic differentiation. *Evolution*. 76(5), 985-1002

Laity, JH, Lee, BM, Wright, PE. 2001. Zinc finger proteins: new insights into structural and functional diversity. *Current Opinion in Structural Biology*. 11, 39-46.

Langmead, B, Salzberg, SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9(4), 357.

Larkin, MA, Blackshields, G, Brown, NP, Chenna, R, McGettigan, PA, McWilliam, H, Valentin, F, Wallace, IM, Wilm, A, Lopez, R, Thompson, JD, 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*. 23(21), 2947-2948.

Lasne, C, Sgrò, CM, Connallon, T. 2017. The relative contributions of the X chromosome and autosomes to local adaptation. *Genetics*. 205(3), 1285-1304.

Lassance, JM, Groot, AT, Liénard, MA, Antony, B, Borgwardt, C, Andersson, F, Hedenström, E, Heckel, DG, Löfstedt, C. 2010. Allelic variation in a fatty-acyl reductase gene causes divergence in moth sex pheromones. *Nature*. 466(7305), 486-489.

Lassance, JM, Liénard, MA, Antony, B, Qian, S, Fujii, T, Tabata, J, Ishikawa, Y, Löfstedt, C. 2013. Functional consequences of sequence variation in the pheromone biosynthetic gene

pgFAR for Ostrinia moths. *Proceedings of the National Academy of Sciences USA*. 110(10), 3967-3972.

Lee, MN, Kweon, HY, Oh, GT. 2018. N-α-acetyltransferase 10 (NAA10) in development: the role of NAA10. *Experimental & Molecular Medicine*. 50, 1-11.

Levy, RC, Kozak, GM, Wadsworth, CB, Coates, BS, Dopman, EB. 2015. Explaining the sawtooth: latitudinal periodicity in a circadian gene correlates with shifts in generation number. *Journal of Evolutionary Biology*. 28, 40-53.

Lewis, LC. 1975. Natural regulation of crop pests in their indigenous ecosystems and in Iowa agroecosystems: Bioregulation of economic insect pests. *Iowa State Journal of Research*. 49, 435–445.

Li, F, Zhao, X, Zhu, S, Wang, T, Li, T, Woolfley, T, Tang, G. 2020. Identification and expression profiling of neuropeptides and neuropeptide receptor genes in Atrijuglans hetaohei. *Gene*. 743, 144605.

Li, H, Handsaker, B, Wysoker, A, Fennell, T, Ruan, J, Homer, N, Marth, G. Abecasis, G, Durbin, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25(16), 2078-2079.

Li, H. 2014. Toward better understanding of artifacts in variant calling from high-coverage samples. *Bioinformatics*. 30(20), 2843-2851.

Li, R, Li, Y, Fang, X, Yang, H, Wang, J, Kristiansen, K, Wang, J. 2009. SNP detection for massively parallel whole-genome resequencing. *Genome Research*. 19(6), 1124-1132.

Lin, H, Lin, X, Zhu, J, Yu, XQ, Xia, X, Yao, F, Yang, G, You, M. 2017. Characterization and expression profiling of serine protease inhibitors in the diamondback moth, Plutella xylostella (Lepidoptera: Plutellidae). *BMC Genomics*. 18, 162.

Lindestad, O, Nylin, S, Wheat, CW, Gotthard, K, 2022. Local adaptation of life cycles in a butterfly is associated with variation in several circadian clock genes. *Molecular Ecology*. 31, 1461-1475

Lindroth, E, Hunt, TE, Skoda, SR, Culy, MD, Lee, D, Foster, JE. 2012. Population genetics of the western bean cutworm (Lepidoptera: Noctuidae) across the United States. *Annals of the Entomological Society of America*. 105, 685-692.

Lowry, DB, Rockwood, RC, Willis, JH. 2008. Ecological reproductive isolation of coast and inland races of Mimulus guttatus. *Evolution*. 62(9), 2196-2214.

Macgregor, CJ, Thomas, CD, Roy, DB, Beaumont, MA, Bell, JR, Brereton, T, Bridle, JR, Dytham, C, Fox, R, Gotthard, K, Hoffmann, AA. 2019. Climate-induced phenology shifts linked to range expansions in species with multiple reproductive cycles per year. *Nature Communications*. 10, 4455.

Malausa, T, Dalecky, A, Ponsard, S, Audiot, P, Streiff, R, Chaval, Y, Bourguet, D. 2007. Genetic structure and gene flow in French populations of two Ostrinia taxa: host races or sibling species? *Molecular Ecology*. 16(20), 4210-4222.

Malinsky, M, Matschiner, M, Svardal, H. 2021. Dsuite□fast D□statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*. 21(2), 584-595.

Mann, RS, Affolter, M, 1998. Hox proteins meet more partners. *Current Opinion in Genetics & Development*. 8, 423-429.

Martin SG, Berthelot-Grosjean M. 2009. Polar gradients of the DYRK-family kinase Pom1 couple cell length with the cell cycle. *Nature*. 459, 852-856.

Martin, SH, Davey, JW, Jiggins, CD. 2015. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*. *32*(1), 244-257.

Matsumoto, K, Okamoto, K, Okabe, S, Fujii, R, Ueda, K, Ohashi, K, Seimiya, H. 2021. G□quadruplex□forming nucleic acids interact with splicing factor 3B subunit 2 and suppress innate immune gene expression. *Genes to Cells*. 26, 65-82.

Matsuo, T, Iida, T, Ishiura, M. 2012. N-terminal acetyltransferase 3 gene is essential for robust circadian rhythm of bioluminescence reporter in Chlamydomonas reinhardtii. *Biochemical and Biophysical Research Communications*. 418, 342-346.

Mayr, E. 1942. *Systematics and the Origin of Species*. New York: Columbia University Press.

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

Meisel, RP, Connallon, T. 2013. The faster-X effect: integrating theory and data. *Trends in Genetics*. 29(9), 537-544.

Meuti, ME, Stone, M, Ikeno, T, Denlinger, DL. 2015. Functional circadian clock genes are essential for the overwintering diapause of the Northern house mosquito, Culex pipiens. *Journal of Experimental Biology*. 218, 412-422.

Møller, AP, Antonov, A, Stokke, BG, Fossøy, F, Moksnes, A, Røskaft, E, Takasu, F. 2011. Isolation by time and habitat and coexistence of distinct host races of the common cuckoo. *Journal of Evolutionary Biology*. 24(3), 676-684.

Nefedova, LN, Kim, AI. 2009. Molecular phylogeny and systematics of Drosophila retrotransposons and retroviruses. *Molecular Biology*. 43, 747-756.

Nosil, P, Feder, JL, Flaxman, SM, Gompert, Z. 2017. Tipping points in the dynamics of speciation. *Nature Ecology & Evolution*. 1, 0001.

Nusse R, Varmus HE. Wnt genes. 1992. *Cell.* 26, 1073-1087.

O'Rourke, M.E, Sappington, TW, Fleischer, SJ. 2010. Managing resistance to Bt crops in a genetically variable insect herbivore, Ostrinia nubilalis. *Ecological Applications*. 20(5), 1228-1236.

Oksanen, J, Blanchet, FG, Kindt, R, Legendre, P, Minchin, PR, O'Hara, RB, Simpson, GL, Solymos, P, Stevens, MHH, Wagner, H, Oksanen, MJ. 2013. Package 'vegan'. *Community Ecology Package*. 2, 1-295.

Orr, HA, Masly, JP, Presgraves, DC. 2004. Speciation genes. *Current Opinion in Genetics and Development*. 14(6), 675-679.

Ortiz-Barrientos, D, Engelstädter, J, Rieseberg, LH. 2016. Recombination rate evolution and the origin of species. *Trends in Ecology & Evolution*. 31(3), 226-236.

Peter, BM, Slatkin, M. 2013. Detecting range expansions from genetic data. *Evolution*. 67, 3274-3289.

Pfeifer, B, Kapan, DD. 2019. Estimates of introgression as a function of pairwise distances. *BMC Bioinformatics*. 20, 207.

Pohlert, T. 2014. The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). R package, https://CRAN.R-project.org/package=PMCMR.

Pool, JE, Nielsen, R. 2007. Population size changes reshape genomic patterns of diversity. *Evolution*. 61(12), 3001-3006.

Presgraves, DC. 2008. Sex chromosomes and speciation in Drosophila. *Trends in Genetics*. 24(7), 336-343.

Presgraves, DC. 2018. Evaluating genomic signatures of "the large X□effect" during complex speciation. *Molecular Ecology*. 27(19), 3822-3830

Price, AL, Patterson, NJ, Plenge, RM, Weinblatt, ME, Shadick, NA, Reich, D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*. 38(8), 904-909.

Pritchard, JK, Stephens, M, Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics*. 155(2), 945-959.

Pruett, CL, Tanksley, SM, Small, MF, Taylor, JF, Forstner, MR. 2011. The effects of range expansion on the population genetics of white-winged doves in Texas. *The American Midland Naturalist*. 166, 415-425.

Pruisscher, P, Nylin, S, Gotthard, K, Wheat, CW. 2018. Genetic variation underlying local adaptation of diapause induction along a cline in a butterfly. *Molecular Ecology*. 27, 3613-3626.

Pruisscher, P, Nylin, S, Wheat, CW, Gotthard, K. 2021. A region of the sex chromosome associated with population differences in diapause induction contains highly divergent alleles at clock genes. *Evolution*. 75, 490-500.

Quinlan, AR. 2014. BEDTools: the Swiss□army tool for genome feature analysis. *Current Protocols in Bioinformatics*. 47(1), 11-12.

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Riesch, R, Muschick, M, Lindtke, D, Villoutreix, R, Comeault, AA, Farkas, TE, ... Nosil, P. 2017. Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*. 1, 0082.

Roff, D. 1980. Optimizing development time in a seasonal environment: the 'ups and downs' of

clinal variation. *Oecologia*. 45, 202-208.

Rousset, F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics*. 145, 1219-1228.

Saini, RK, Keum, YS. 2018. Omega-3 and omega-6 polyunsaturated fatty acids: Dietary sources, metabolism, and significance—A review. *Life Sciences*. 203, 255-267.

Sandrelli, F, Costa, R, Kyriacou, CP, Rosato, E, 2008. Comparative analysis of circadian clock genes in insects. *Insect Molecular Biology*. 17, 447-463.

Santos, H, Burban, C, Rousselet, J, Rossi, JP, Branco, M, & Kerdelhue, C. 2011. Incipient allochronic speciation in the pine processionary moth (Thaumetopoea pityocampa, Lepidoptera, Notodontidae). *Journal of Evolutionary Biology*. 24(1), 146-158.

Santos, H, Rousselet, J, Magnoux, E, Paiva, MR, Branco, M, Kerdelhué, C. 2007. Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B.* 274(1612), 935-941.

Schiffels, S, Durbin, R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*. 46(8), 919-925.

Schilling, MP, Mullen, SP, Kronforst, M, Safran, RJ, Nosil, P, Feder, JL, ... Flaxman, SM. 2018. Transitions from single-to multi-locus processes during speciation with gene flow. *Genes*. 9(6), 274.

Schmidt, PS, Zhu, CT, Das, J, Batavia, M, Yang, L, Eanes, WF. 2008. An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in Drosophila melanogaster. *Proceedings of the National Academy of Sciences USA.* 105, 16207-16211.

Sebastian W, Sukumaran S, Zacharia PU, Gopalakrishnan A. 2018. Isolation and characterization of aquaporin 1 (AQP1), sodium/potassium-transporting ATPase subunit alpha-1 (Na/K-ATPase α1), heat shock protein 90 (HSP90), heat shock cognate 71 (HSC71), osmotic stress transcription factor 1 (OSTF1) and transcription factor II B (TFIIB) genes from a euryhaline fish, Etroplus suratensis. *Molecular Biology Reports.* 45, 2783-2789.

Seehausen, O, Butlin, RK, Keller, I, Wagner, CE, Boughman, JW, Hohenlohe, PA, ... Widmer, A. 2014. Genomics and the origin of species. *Nature Reviews Genetics*. 15(3), 176-192.

Servedio, MR, Van Doorn, GS, Kopp, M, Frame, AM, Nosil, P. 2011. Magic traits in speciation: 'magic' but not rare? *Trends in Ecology & Evolution.* 26(8), 389-397.

Shabbir, MZ, Zhang, T, Prabu, S, Wang, Y, Wang, Z, Bravo, A, Soberón, M, He, K. 2020. Identification of Cry1Ah-binding proteins through pull down and gene expression analysis in Cry1Ah-resistant and susceptible strains of Ostrinia furnacalis. *Pesticide Biochemistry and Physiology*. 163, 200-208.

Sharabi, O, Ventura, T, Manor, R, Aflalo, ED, Sagi, A. 2013. Epidermal growth factor receptor in the prawn Macrobrachium rosenbergii: Function and putative signaling cascade. *Endocrinology*. 154, 3188-3196.

Shaw, KL, Mullen, SP. 2011. Genes versus phenotypes in the study of speciation. *Genetica*. 139(5), 649-661.

Shaw, KL, Mullen, SP. 2014. Speciation continuum. *Journal of Heredity*. 105(S1), 741-742.

Showers, W.B. 1981. Geographic variation of the diapause response in the European corn borer. *Insect Life History Patterns.* 97-111. New York: Springer.

Showers, WB, Chiang, HC, Keaster, AJ, Hill, RE, Reed, GL, Sparks, AN, Musick, GJ. 1975. Ecotypes of the European corn borer in North America. *Environmental Entomology*. 4, 753-760.

Showers, WB, Reed, GL, Robinson, JF, Derozari, MB. 1976. Flight and sexual activity of the European corn borer. *Environmental Entomology.* 5(6), 1099-1104.

Showers, WB. 1993. Diversity and variation of European corn borer populations. Evolution of *Insect Pests: Patterns of Variation.* 287-309. Chichester, West Sussex, UK: Wiley.

Slatkin, M. 1970. Selection and polygenic characters. *Proceedings of the National Academy of Sciences USA*. 66, 87-93.

Smadja, C, Galindo, J, Butlin, R. 2008. Hitching a lift on the road to speciation. *Molecular Ecology*. 17, 4177-4180.

Smith, HE. 1920. Broom corn, the probable host in which Pyrausta nubilalis Hubn. reached America. *Journal of Economic Entomology*. 13, 425-430.

Smith, JM, Haigh, J. 1974. The hitch-hiking effect of a favourable gene. *Genetical Research*. 23, 23-35.

Śniegula, S, Drobniak, SM, Gołąb, MJ, Johansson, F. 2014. Photoperiod and variation in life history traits in core and peripheral populations in the damselfly Lestes sponsa. *Ecological Entomology*. 39, 137-148.

Soll JM, Brickner JR, Mudge MC, Mosammaparast N. 2018. RNA ligase-like domain in activating signal cointegrator 1 complex subunit 1 (ASCC1) regulates ASCC complex function during alkylation damage. *Journal of Biological Chemistry*. 1, 13524-13533.

Soppa U, Becker W. DYRK protein kinases. 2015. *Current Biology*. 15, 488-489.

Sorenson, CE, Kennedy, GG, Schal, C, Walgenbach, JF. 2005. Geographical variation in pheromone response of the European corn borer, Ostrinia nubilalis (Lepidoptera: Crambidae), in North Carolina: a 20-year perspective. *Environmental Entomology.* 34, 1057-1062.

Sorenson, CE, Kennedy, GG, Van Duyn, W, Bradley Jr, JR, Walgenbach, JF. 1992. Geographical variation in pheromone response of the European corn borer, Ostrinia nubilalis, in North Carolina. *Entomologia Experimentalis et Applicata*. 64(2), 177-185.

Sousa, V, Peischl, S, Excoffier, L. 2014. Impact of range expansions on current human genomic diversity. *Current Opinion in Genetics & Development*. 29, 22-30.

Stirnimann, CU, Petsalaki, E, Russell, RB, Müller, CW. 2010. WD40 proteins propel cellular networks. *Trends in Biochemical Sciences*. 3, 565-574.

Szczepaniak, K, Bukala, A, da Silva Neto, AM, Ludwiczak, J, Dunin-Horkawicz, S. 2020. A library of coiled-coil domains: from regular bundles to peculiar twists. *Bioinformatics*. 36, 5368-5376.

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

Tay, WT, Behere, GT, Batterham, P, Heckel, DG. 2010. Generation of microsatellite repeat families by RTE retrotransposons in lepidopteran genomes. *BMC Evolutionary Biology*. 10, 144.

Taylor, RS, Friesen, VL. 2017. The role of allochrony in speciation. *Molecular Ecology*. 26(13),

3330-3342.

Taylor, S, Campagna, L. 2016. Avian supergenes. *Science*. 351(6272), 446-447.

Tepolt, CK, Palumbi, SR. 2020. Rapid adaptation to temperature via a potential genomic island of divergence in the invasive green crab, Carcinus maenas. *Frontiers in Ecology and Evolution*. 8, 411.

Thomas, Y, Bethenod, MT, Pelozuelo, L, Frérot, B, & Bourguet, D. 2003. Genetic isolation between two sympatric host□plant races of the European Corn Borer, Ostrinia nubilalis Hubner I. Sex pheromone, moth emergence timing, and parasitism. *Evolution*. 57(2), 261-273.

Thompson, MJ, Jiggins, C. 2014. Supergenes and their role in evolution. *Heredity*. 113(1), 1-8.

Tishkoff, SA, Reed, FA, Friedlaender, FR, Ehret, C, Ranciaro, A, Froment, A, Hirbo, JB, Awomoyi, AA, Bodo, JM, Doumbo, O, Ibrahim, M. 2009. The genetic structure and history of Africans and African Americans. *Science*. 324, 1035-1044.

Unbehend, M, Kozak, GM, Koutroumpa, F, Coates, BS, Dekker, T, Groot, AT, Heckel, DG, Dopman, EB. 2021. bric à brac controls sex pheromone choice by male European corn borer moths. *Nature Communications*. 12, 2818.

Uzelac, I, Avramov, M, Čelić, T, Vukašinović, E, Gošić-Dondo, S, Purać, J, Kojić, D, Blagojević, D, Popović, ŽD. 2020. Effect of cold acclimation on selected metabolic enzymes during diapause in the European corn borer Ostrinia nubilalis (Hbn.). *Scientific Reports*. 10, 9085.

Van der Auwera, GA, Carneiro, MO, Hartl, C, Poplin, R, Del Angel, G, Levy□Moonshine, A, Jordan, T, Shakir, K, Roazen, D, Thibault, J, Banks, E. 2013. From FastQ data to high□ confidence variant calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 43(1110),11.10.1-11.10.33.

Van Dyck, H, Bonte, D, Puls, R, Gotthard, K, Maes, D. 2015. The lost generation hypothesis: could climate change drive ectotherms into a developmental trap? *Oikos*. 124, 54-61.

Varpe, Ø. 2017. Life history adaptations to seasonality. *Integrative and Comparative Biology*. 57, 943-960.

Via, S, West, J. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*. 17(19), 4334-4345.

Via, S. 2012. Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B*. 367(1587), 451-460.

Wadsworth, CB, Li, X, Dopman, EB. 2015. A recombination suppressor contributes to ecological speciation in Ostrinia moths. *Heredity*. 114(6), 593-600.

Wadsworth, CB, Woods Jr, WA, Hahn, DA, Dopman, EB. 2013. One phase of the dormancy developmental pathway is critical for the evolution of insect seasonality. *Journal of Evolutionary Biology*, 26, 2359-2368.

Wang, H, Yang, Y, Huang, F, He, Z, Li, P, Zhang, W, Zhang, W, Tang, B. 2020. In situ fluorescent and photoacoustic imaging of Golgi pH to elucidate the function of transmembrane protein 165. *Analytical Chemistry*. 92, 3103-3110.

Wang, RL, Xia, QQ, Baerson, SR, Ren, Y, Wang, J, Su, YJ, Zheng, SC, Zeng, R.S. 2015. A novel cytochrome P450 CYP6AB14 gene in Spodoptera litura (Lepidoptera: Noctuidae) and its potential role in plant allelochemical detoxification. *Journal of Insect Physiology*. 75, 54-62.

Wang, Y, Kim, KS, Guo, W, Li, Q, Zhang, Y, Wang, Z, Coates, BS. 2017. Introgression between divergent corn borer species in a region of sympatry: implications on the evolution and adaptation of pest arthropods. *Molecular Ecology*. 26(24), 6892-6907.

Weng, YL, Liu, N, DiAntonio, A, Broihier, HT. 2011. The cytoplasmic adaptor protein Caskin mediates Lar signal transduction during Drosophila motor axon guidance. *Journal of Neuroscience*. 31, 4421-4433.

Werck-Reichhart, D, Feyereisen, R. 2000. Cytochromes P450: a success story. *Genome Biology*. 1, reviews3003.1.

Wu, CI. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*. 14(6), 851-865.

Xiong, Y, Burke, WD, Eickbush, TH. 1993. Pao, a highly divergent retrotransposable element from Bombyx mori containing long terminal repeats with tandem copies of the putative R region. *Nucleic Acids Research*. 21, 2117-2123.

Xu, M, Shaw, KL. 2021. Extensive linkage and genetic coupling of song and preference loci underlying rapid speciation in Laupala crickets. *Journal of Heredity*. 112(2),204-213.

Xu, M, Shaw, KL. 2019. Genetic coupling of signal and preference facilitates sexual isolation during rapid speciation. *Proceedings of the Royal Society B*. 286(1913), 20191607.

Yamashiro, S, Gokhin, DS, Kimura, S, Nowak, RB, Fowler, VM. 2012. Tropomodulins: pointed□end capping proteins that regulate actin filament architecture in diverse cell types. *Cytoskeleton*. 69, 337-370.

Yang, Z, Plotkin, D, Landry, JF, Storer, C, Kawahara, AY. 2021. Revisiting the evolution of Ostrinia moths with phylogenomics (Pyraloidea: Crambidae: Pyraustinae). *Systematic Entomology*. 46(4), 827-838.

Yasukochi, Y, Ohno, M, Shibata, F, Jouraku, A, Nakano, R, Ishikawa, Y, & Sahara, K. 2016. A FISH-based chromosome map for the European corn borer yields insights into ancient chromosomal fusions in the silkworm. *Heredity.* 116(1), 75-83.

Zaman, M. 2007. A comparison of univoltine and multivoltine European corn borer (Ostrinia nubilalis Hübner): Life history characters, Bt toxin susceptibility, parasitoid impact, and population pattern. Ph.D. diss. Pennsylvania State University, University Park, PA.

Zhao, L, Sun, X, Qin, S, Kong, Y, Wang, X, Li, M. 2021. The *Vps13d* gene affects the silk yield of Bombyx mori. Available at SSRN online.

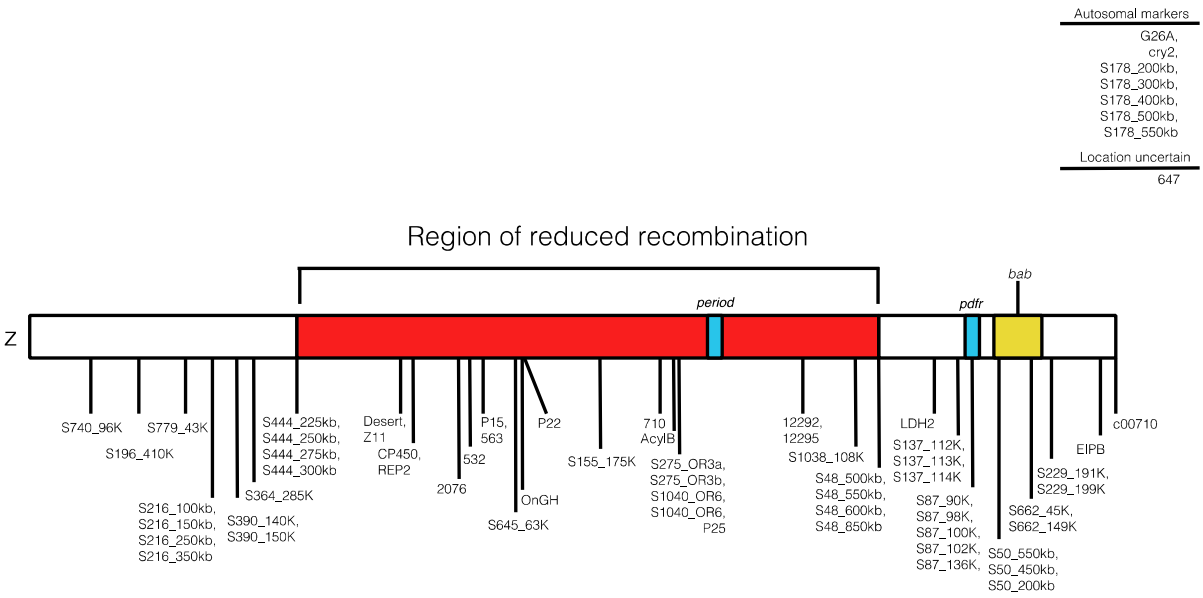

Figure S1.1. Map of marker locations on the Z-chromosome. Indicated are the large region of reduced recombination, *per* and *Pdfr*, loci responsible for seasonal timing and the male pheromone response locus *bab*.

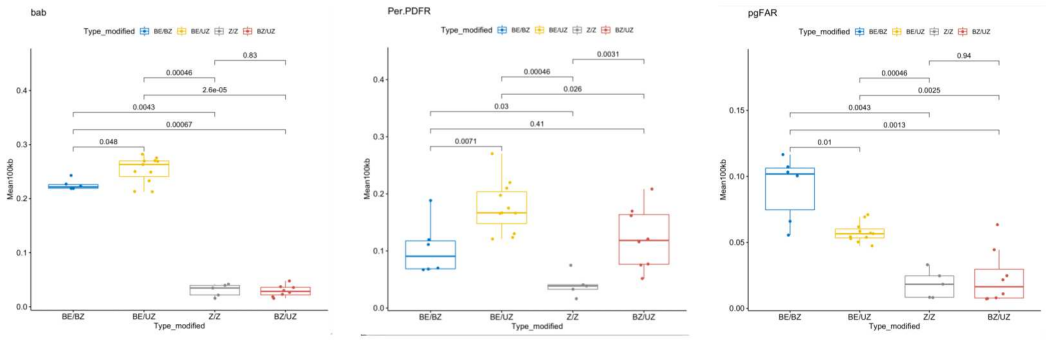

Figure S1.2. Pairwise F_{ST} comparisons of barrier loci restricted to populations within ~100km, or roughly 1-3 generations of dispersal.

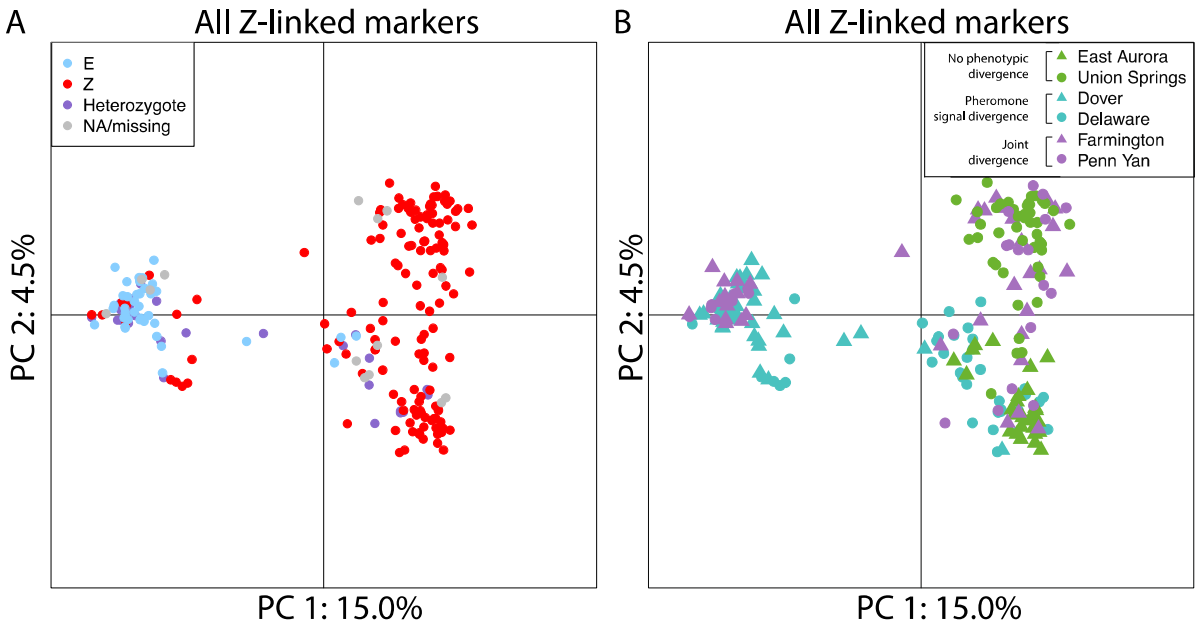

Figure S1.3. Principal components analyses based on Z-linked marker datasets. On the left, individual labels are based on *pgFAR* genotype. On the right, individual labels are based on geographic location and population type. Percentages indicate the amount of variance explained by that axis.

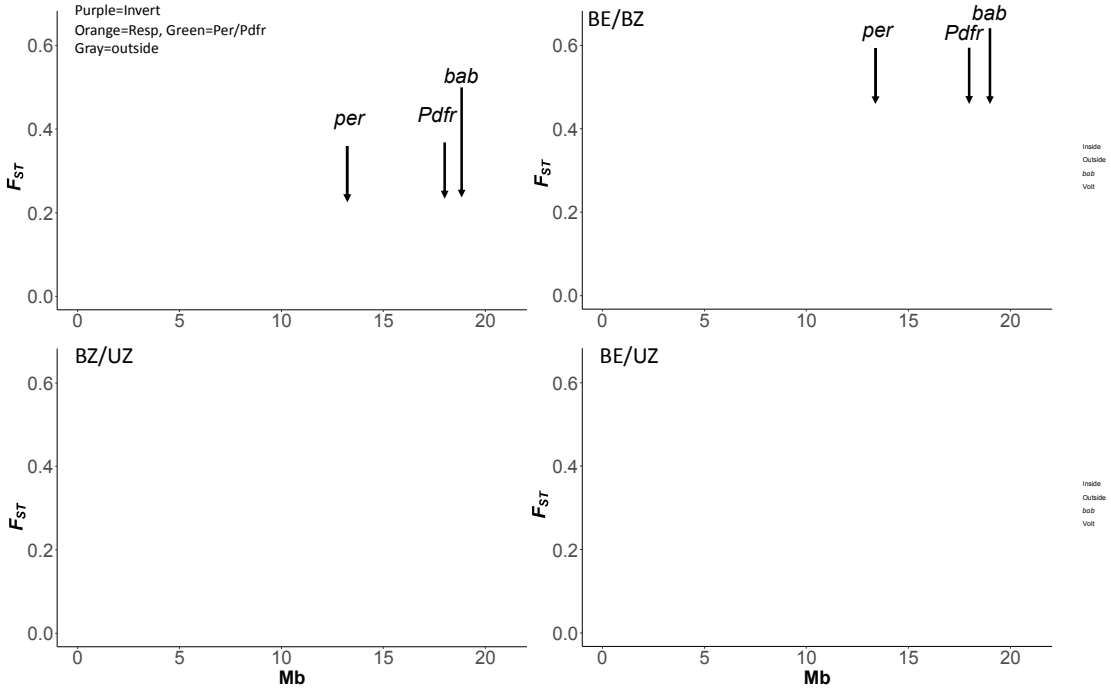

Figure S1.4. Fst 100kb windows across the Z-chromosome using pooled sequencing data for four population type comparisons. Colors distinguish the rearrangement (purple) and key barrier traits (teal and orange).

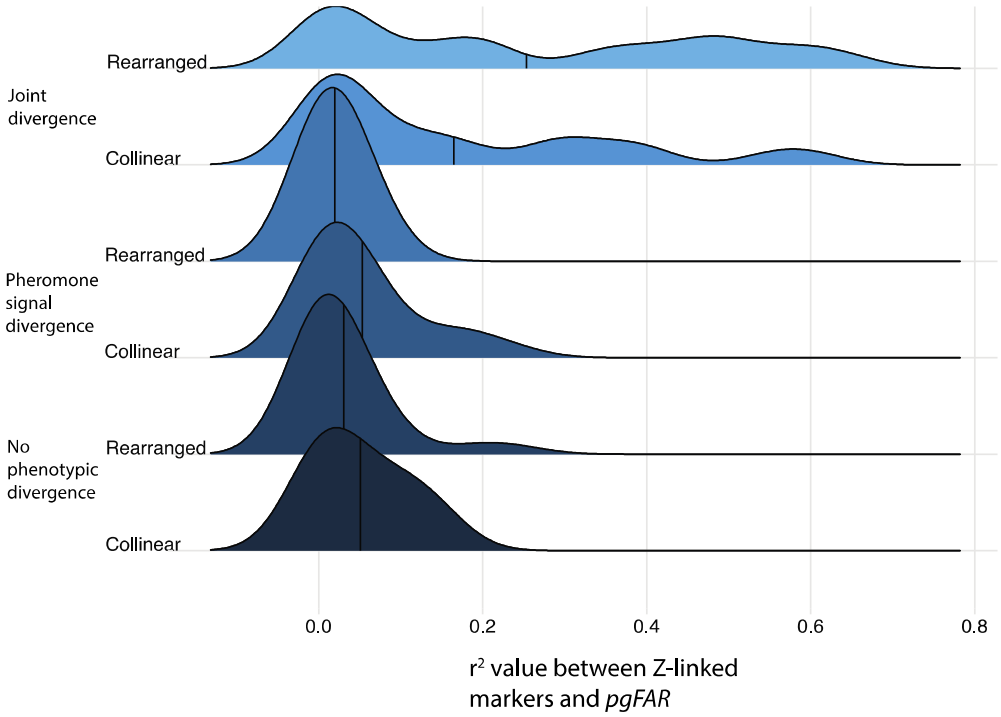

Figure S1.5. Linkage disequilibrium calculated within geographic populations between pgFAR and SNPs on the Z-chromosome, pooled by population type and separated by collinear vs rearranged markers. The mean is indicated by the vertical black line.

Table S1.1. Genome localization and other details of sequences generated by PCR primers tested and used in analysis

Locus	Z-chromosome position Mb	Z-chromosome position cM	Features	Forward sequence	Reverse sequence
S740_96K	1.165	4.45	-	GATGCCTACGTCATAACATCTACCT	TAAGCCTCAGTTTTAAGTCCATTTG
S196_410K	2.061	10.85	-	TTTGTGGTTTTGGTGTTTTATTTT	CTGTCCTGATAAACAATGTGAAATG
S779_43K	2.929	10.85	-	TCCAGCTACGTGTCAAATATCAA	AGCGTCAGTTTACAAGCTGAAAT
sc216_350kb	3.036	10.85	-	CGGTAACTAAAGGCGGAAGA	TGTCAACTTTTTGGCCACTG
sc216_250kb	3.137	10.85	-	CACAAACAAGCTGCATTTCG	AGCAGAAAATCAAACGCACA
sc216_150kb	3.237	10.85	-	TGTTCGGAGCACCATTATGA	CGCCCTACTCGAACAAAAAT
sc216_100kb	3.287	10.85	-	CATAGGGCTGTCATCCATCC	AACCAGTCATTTTCAGGGACA
S390_140K	3.917	10.85	-	CAGTCTACCCTGAAAATGGTTAGTG	TTGTATGGTGTTTGACAGATTTTTG
S390_159K	4.155	10.85	-	ACTCTACCTACTTCCATTGTTGACG	CATTACGACATGTGATTAACTTGGA
S364_285K	4.214	17.59	-	ACGCGAACTTTAAAGATTTTCTTCT	ACGTAATCAACTACACTCGACATCA
sc444_225kb	5.269	25.06	Rearrangement	CAAGACATGGTCTCCATTCG	TTTTCAATCGGGCTAAATGG
sc444_250kb	5.294	25.06	Rearrangement	ATTTGGCCCATCAAAAATTG	AGACAGTGGACGGTGTGATG

sc444_275kb	5.319	25.06	Rearrangement	TTAAGCACTCGCAATGATCG	GAACAGTAGCGAGCCACCAC
sc444_300kb	5.344	25.06	Rearrangement	GGACCTTGTTAGCTGGGAAA	CTGATGAATTCGCCGCTATT
Desert	6.988	25.06	Rearrangement	TTCAGTGCCCTCATCGTTTCTT	CCCGAAGGTGCTGTAACAAA
Z11	6.988	25.06	Rearrangement	ATCTTCCTTATTGCGTACCA	GCATTCAGGAGCAGGAGGA
CP450	7.231	25.06	Rearrangement	AGGCCTAAAGGGATCTTCTACCAGA	ACCCACCATATCAATAGCAGAGAAT
REP2	7.231	25.06	Rearrangement	AAATGGACGATGCTGGCTGA	TTGACCGGCCAAAGTTAGG
2076	8.076	25.06	Rearrangement	GAAGCCCCAGTGTGGTGCCC	AGGGTCGCATCGAGTCCAGCT
532	8.298	25.06	Rearrangement	TGGATGTCCTCCTCTGTTTGGCT	TCACCATCATCGGATTCATCGGACA
563	8.547	25.06	Rearrangement	TGTTGGTAAGCGCATCACTCA	ACATAATTAGGGCCGTTCCACATA
P15	8.547	25.06	Rearrangement	CGCGACCACTCCAGTTGTTT	TGGTTTGGTCTGGAAATCTCG
S645_63K	9.146	26.51	Rearrangement	ACGAGTCTGGGTGAATTTTATTGTA	CCAACTGAGGCTAGTTATGTAAGGA
OnGH	9.285	26.51	Rearrangement	ACGAGGGTGGCAACGTGT	TTTCTAAGTACAGCATGTCCGTATCC
P22	9.315	26.51	Rearrangement	GATCGCGAAGCAGAAGTCAGT	CCAAAGCGGCGGAACCAC

				C	
S155_175K	10.75	32.5	Rearrangement	TGATTGATGACTGATTTGGTGAT	TTGTGAATACCGGCGTAAATATC
710	11.880	39.27	Rearrangement	GCGTGCGACCCAACTCTGGA	CCATGAGCTGATCAAGGCTGATCTC
AcylB	12.130	39.27	Rearrangement	AAACTCCCGTCATCGCCG	TCTGGCCGTTCAGGATCCATTCG
S1040_OR6a	12.240	40.72	Rearrangement	TTACCATTCCATACCACCATCAT	TGTATTCCACAAGGACAAACCTC
S275_OR3a	12.240	40.72	Rearrangement	ATTATGTCCGCCGTGTAAGTAAA	GCCAAGAACGTTGATAGAATCAT
S275_OR3b	12.240	40.72	Rearrangement	AATGCATTTTGCACATAAAGGTC	TAGAACCAAATCTAATCCGCACA
P25	12.240	40.72	Rearrangement	TACGCGGGTAAGTACAGGGA	GTCGAGCGCGGCACTCATAC
period	13.250	40.72	Rearrangement	AAGAACGTCAGCGATGAAGACGGA	CAGGTAGGGCACCGACTCAGGG
12295	14.580	40.72	Rearrangement	CTGCTGATGATGAGTGATTTCCGGT	TGGTGGAGAACCATACACACTGGG
12292	14.580	40.72	Rearrangement	GATGAGCATGCGCAGACGCG	AGACGTGGTCCGACCGATCA
S1038_108K	15.570	58.22	Rearrangement	AGACGTGGTCCGACCGATCA	ATTATGCGCCGATATTTAATAACAA
S48_500kb	16.010	59.87	Rearrangement	GGTTTCCACCAAGACACCAT	CATCTTCGACACAGAACAGCA

S48_550kb	16.010	59.87	Rearrangement	CGGGTGT CAAACAT TGGTTA	TCTTGGC CCAAATT TGTATGT
S48_600kb	16.010	59.87	Rearrangement	TGGGCAG AAAATGA TGATGA	CATATCG TCGTGGT GGTTTG
S48_850kb	16.010	59.87	Rearrangement	CCGCCGG AAGTACT TAATAGC	TCCCAA ACACGT GAGAGT TG
LDH2	17.06	70.87	-	CGATACC TTCTGTCC GAGAAAC	CATGGG TCTCCTT CCAGTTC
S137_112K	17.500	62.22	-	TTTTGTTT AATCCGT TTTTCACC	ACCTCA AGGTCA TGGGGT ATCT
S137_112K	17.500	62.22	-	GGCTTTCC CAACACA GTAATTC	TGTCGCT AAGAGA AAAAGC ACA
S137_112K	17.500	62.22	-	TTGGGTG AATGCTTT TGATGTA	GGCAGA CAAAAT CACAAA CAAG
S87_90K	17.760	62.22	-	TTGAGTTT TGGTTAA GGCCAAT	TGGGTTA CCAGGG TACCATA AA
S87_98K	17.760	62.22	-	TTACGTG GTTCATA CTGGAAC G	ACGCAA ATTAAA AGCGAA TTGT
S87_100K	17.760	62.22	-	TTCAACC GAAACCA TTATCAG A	AACGTTC TTGCTGT GTTGAA AA
S87_102K	17.760	62.22	-	TGTTTTCG TCTTTGTT TGTCGT	TTCCGTC TTACGTA TTCCACA A
S87_163K	17.760	62.22	-	AACAGGG CACTATT AATAACG AAA	GGGATTT CAGAAA GGAAGC ATA
S50_200kb	18.270	70.87	*bab* (Pheromone	ATCATTG GTTCCCA AAGTCG	AGGGGG CGTATCT CATTGTT

			response locus)		
S50_450kb	18.270	70.87	*bab*	TTCCGTGATTGCAAAAACAA	CCGGAGTCACTTTGGCTAGT
S50_550kb	18.270	70.87	*bab*	GACGGAACCCGTTGTCTTAG	TCGTTACGATGCCAGTACGA
S662_45K	18.880	70.87	*bab*	GCATTTGAAATCTCACCTTGCTA	CTGCTCCCTCCTGGAAATCTA
S662_149K	18.880	70.87	*bab*	TTAGACATCCAAACTTTCGCATT	TCCGCTAAATACTATTGG
S229_191K	19.260	74.6	-	AAATACCGAAAGGTCCTAACATACC	AGGCTCGCGTAAGTATCATTT
S229_199K	19.260	74.6	-	GCTTTGGTATCATTATGTCCTTTGT	CCTCTCGAGTTCTTTTTCAC
EIPB	20.190	82.7	-	GTACAGCAGGTCGTGGAGTT	CCATCATCGAAGACGAGGC
c00710	20.470	82.7	-	CTAGCTCGGTTTCACATTTCAAGT	TACCCACCGTCCAACACATAAG
647	Likely autosomal	-	-	TGCAAGAAGTGGTCCAGGCACT	GCAAGATGCCGAAGTCGCTC
pgFAR	Autosomal	-	Pheromone production gene	GGTGGCATGGGGACGGTACAG	TTTGGATTAAATTAATTTTAAAC
sc178_200kb	Autosomal	-	Pheromone production scaffold	ATGAGATGTGGTTGCACGAA	TCATTGTTTCAGCCACTGGA
sc178_300kb	Autosomal	-	Pheromone production scaffold	ACAAGTTATGCCGGACAAGG	CTTACAGATGCTGGCAATCG
sc178_400kb	Autosomal	-	Pheromone production	CGTCTGCGAGTGTC	CCCAATCCCAACG

			scaffold	CATTTA	AAAATAA
sc178_500kb	Autosomal	-	Pheromone production scaffold	TGATGATGAATCGCTTTTGC	CCGTTTCGAACCACTACACC
sc178_550kb	Autosomal	-	Pheromone production scaffold	TAGCTGCTGGAGGATTTTGG	TGGATGAGAAAGCGGAAGAT
cry2	Autosomal	-	-	CGGTGCGGCCGTCGATCCATC	AACCAGTCATTTTCAGG
G26A	Autosomal	-	-	AGATCTCGCACGAATTTTGAATG	ACGCCTATTATTTACGTCAGG

Table S1.2. STRUCTURE output statistics based on the Evanno method (Evanno et al. 2005).

K	Replicates	Mean LnP(K)	Standard deviation LnP(K)	Ln'(K)	|Ln"(K)|	Delta K
1	50	-57459.252	6.8696	NA	NA	NA
2	50	-49713.204	4.7239	7746.048	6152.868	1302.50176
3	50	-48120.024	9.548	1593.18	2566.848	268.836187
4	50	-49093.692	1586.3296	-973.668	339.462	0.213992
5	50	-49727.898	2741.6036	-634.206	53.446	0.019494
6	50	-50308.658	1384.6757	-580.76	259.016	0.187059
7	50	-50630.402	1451.1248	-321.744	NA	NA

The grey bar indicates the most likely number of inferred clusters.

Table S2.1. Z-linked scaffolds and physical start positions along the length of the Z-chromosome. Asterisks (*) indicate a scaffold containing *period* (532) or *Pdfr* (87).

Scaffold	**Size**	**Mb**
scaffold4435	4078	0
scaffold751	275148	0.004
scaffold35	722381	0.279
scaffold572	163719	1.002
scaffold740	235087	1.165
scaffold1546	33109	1.4
scaffold247	627634	1.434
scaffold196	868841	2.061
scaffold779	92251	2.93

scaffold2358	14287	3.022
scaffold600	417131	3.037
scaffold216	461870	3.454
scaffold390	298204	3.916
scaffold364	585662	4.214
scaffold1895	21730	4.799
scaffold409	222884	4.821
scaffold444	384115	5.044
scaffold325	886211	5.428
scaffold15	673834	6.314
scaffold169	242750	6.988
scaffold348	845032	7.231
scaffold865	221853	8.076
scaffold518	233254	8.298
scaffold2225	15954	8.531
scaffold546	598492	8.547
scaffold645	139123	9.146
scaffold1618	30317	9.285
scaffold789	91266	9.315
scaffold218	453784	9.406
scaffold1660	63095	9.86
scaffold1044	193178	9.923
scaffold581	122005	10.12
scaffold1237	50895	10.24
scaffold1246	194460	10.29
scaffold1756	66030	10.48
scaffold1161	202793	10.55
scaffold155	350779	10.75
scaffold282	321509	11.1
scaffold804	458571	11.42
scaffold502	248588	11.88
scaffold3599	6641	12.13
scaffold1040	105542	12.14
scaffold275	1004024	12.24
scaffold4243	4645	13.25
scaffold532*	371787	13.25
scaffold459	265770	13.62
scaffold1889	21771	13.89
scaffold221	590889	13.91

scaffold931	77297	14.5
scaffold140	987430	14.58
scaffold1038	208510	15.57
scaffold598	233855	15.78
scaffold48	965936	16.01
scaffold1486	80768	16.98
scaffold193	439059	17.06
scaffold137	260670	17.5
scaffold87*	517236	17.76
scaffold50	611115	18.27
scaffold662	196882	18.88
scaffold728	179971	19.08
scaffold229	374464	19.26
scaffold330	554422	19.64
scaffold673	283851	20.19
scaffold812	88987	20.47
scaffold1046	116232	20.56

Table S2.2. Mean pairwise F_{ST} between populations. Values below the diagonal are from autosomal markers and values above the diagonal are from Z-linked markers.

Locality	**P1**	**P2**	**P3**	**P4**	**P6**	**P8**	**P9**	**P10**	**UZ**
P1	NA	0.0293	0.0302	0.0290	0.0311	0.0335	0.0358	0.0290	0.1586
P2	0.0372	NA	0.0299	0.0278	0.0306	0.0332	0.0354	0.0282	0.1605
P3	0.0379	0.0376	NA	0.0291	0.0314	0.0338	0.0359	0.0294	0.1614
P4	0.0360	0.0353	0.0362	NA	0.0298	0.0326	0.0351	0.0264	0.1644
P6	0.0387	0.0384	0.0391	0.0373	NA	0.0335	0.0356	0.0298	0.1627
P8	0.0398	0.0397	0.0403	0.0389	0.0407	NA	0.0367	0.0325	0.1610
P9	0.0442	0.0441	0.0446	0.0439	0.0446	0.0452	NA	0.0349	0.1652
P10	0.0362	0.0356	0.0367	0.0336	0.0375	0.0388	0.0438	NA	0.1634
UZ	0.1978	0.1991	0.1992	0.2008	0.1997	0.1976	0.2021	0.1997	NA

Table S2.3. Mean pairwise genetic distances ($F_{ST}(1 - F_{ST})$; see methods) between populations. Values below the diagonal are from autosomal markers and values above the diagonal are from Z-linked markers.

Locality	**P1**	**P2**	**P3**	**P4**	**P6**	**P8**	**P9**	**P10**	**UZ**
P1	NA	0.0302	0.0311	0.0299	0.0322	0.0347	0.0371	0.0298	0.1885
P2	0.0386	NA	0.0308	0.0286	0.0316	0.0344	0.0367	0.0291	0.1912
P3	0.0394	0.0391	NA	0.0299	0.0324	0.0349	0.0373	0.0303	0.1924
P4	0.0374	0.0366	0.0375	NA	0.0308	0.0337	0.0364	0.0271	0.1967

P6	0.0403	0.0399	0.0407	0.0387	NA	0.0346	0.0369	0.0307	0.1943
P8	0.0415	0.0413	0.0420	0.0404	0.0425	NA	0.0381	0.0336	0.1919
P9	0.0463	0.0462	0.0467	0.0459	0.0466	0.0473	NA	0.0361	0.1979
P10	0.0376	0.0369	0.0380	0.0348	0.0389	0.0404	0.0458	NA	0.1953
UZ	0.2466	0.2486	0.2487	0.2512	0.2495	0.2462	0.2532	0.2496	NA

Table S2.4. Top Z-linked SNPs (by BF.dB) with regards association with one of four specified covariables (see text).

Gene	Annotation	Scaffold	Location on scaffold	Location of scaffold on Z-chromosome (Mbp)	Covariable	BF.dB	Function	Literature cited
bab	bric-a-brac1	bab_gene_region_CCv02	318417	18.27	Step cline	20.74	Involved in male pheromone response behavior. Located within 0.5 Mb of *pdfr*	Dopman et al. 2004; Koutroumpa et al. 2016; Unbehend et al. 2021
bab	bric-a-brac1	bab_gene_region_CCv02	388040	18.27	Step cline	17.19		
ostnu00g0112725.0.7	Function unknown	scaffold137	20421	17.5	Latitude	26.28	unknown	
ostnu00g0112725.0.7	Function unknown	scaffold137	198112	17.5	Latitude	19.28	unknown	
ostnu00g0112725.0.4	Coiled-coil domain containing protein	scaffold137	20421	17.5	Degree Days	20.97	Widespread in process, from transcription regulation to structural functions	Szczepaniak et al. 2020
ostnu00g0103382.0.2	Triple functional domain protein	scaffold193	191518	17.06	Degree Days	18.49	Structural function	Brodsky et al. 1995
ostnu00g0120008.0.1	High affinity cAMP-specific 3' / *dunce*	scaffold229	292169	19.26	Degree Days	17.73	Physiological and motory signal transduction. *dunce* was shown to be in high linkage disequilibrium with AMPK/SNF4 gamma in Kozak et al. 2019. This LD value was comparatively similar in strength to the *period-Pdfr* association.	Müller 1997; Kozak et al. 2019
ostnu00g0118594.0.1	Function unknown	scaffold247	134601	1.43	Degree Days	17.87	unknown	
ostnu00g0118594.0.1	Function unknown	scaffold247	171322	1.43	Step cline	15.35	unknown	
ostnu00g0115831.0.1	Clock	scaffold275	786038	12.24	Number of Generations	17.79	Circadian rhythms, diapause timing	Sandrelli et al. 2008; Kozak et al. 2019

ostnu00g0123284.0.2	Vacuolar protein sorting-associated protein 13B-like protein	scaffold330	466384	19.64	Degree Days	15.38	Host-virus interaction, Protein transport, transport	Zhao et al. 2021
ostnu00g0123284.0.2	Vacuolar protein sorting-associated protein 13B-like protein	scaffold330	469220	19.64	Latitude	18.62		
ostnu00g0123284.0.1	Vacuolar protein sorting-associated protein 13B-like protein	scaffold330	469220	19.64	Degree Days	18		
ostnu00g0116926.0.1	RING finger and CCCH-type zinc finger domain-containing protein 2-like isoform 1	scaffold390	230817	3.92	Step cline	16.09	Sex determination, differential splicing in Bombyx	Gopinath et al. 2016
ostnu00g0120772.0.1	Proton-coupled amino acid transporter 4-like protein	scaffold48	572753	16.01	Degree Days	16.73	Amino-acid transport, Symport, Transport	Gao et al. 2020; Hu et al. 2021
ostnu00g0125649.0.1	LIM homeobox transcription factor 1	scaffold50	97765	18.27	Number of Generations	24.18	Transcription factor, responsible for development of limbs, membranes, eyes, and neurons in Drosophila	Guo et al. 2018
ostnu00g0111445.0.1	Dynein heavy chain 6	scaffold518	123609	8.3	Latitude	22.05	Motor protein, responsible for intracellular movements	Gee et al. 1997
ostnu00g0111445.0.2	Dynein heavy chain 6	scaffold518	123609	8.3	Degree Days	20.74		
ostnu00g0126236.0.1	Acyl-coa desaturase (stearoyl)	scaffold598	188552	15.78	Degree Days	15.41	Biosynthesis of omega 3- and omega 6-polyunsaturated fatty acids	Saini & Keum 2018

Table S2.5. Top autosomal SNPs (by BF.dB) with regards association with one of four specified covariables (see text).

Gene	Annotation	Scaffold	Location on scaffold	Covariable	BF.dB	Function	Literature cited

ostnu00g01191 81.0.2	Malate dehydrogenase	scaffold 184	481756	Degree Days	39.64	Lipid biosynthesis, hormonally regulated during development in fruit flies. In *Ostrinia*, malate dehydrogenase is downregulated in the winter. Found exclusively in resistant strains of *O. furnicalis*.	Farkaš & Knopp 1997; Koštál et al. 2006; Shabbir et al. 2020; Uzelac et al. 2020
ostnu00g01073 29.0.1	Unconventional myosin-XV-like protein	scaffold 104	1330706	Degree Days	30.52	Actin filament binding, ATP binding, microfilament motor activity	Bird et al. 2014
ostnu00g01135 99.0.1	Neuropeptide receptor A10	scaffold 554	99375	Degree Days	27.54	G-protein coupled receptor, involved in feeding and stress response in *Drosophila*. Putatively involved in pheromone regulation in Lepidoptera.	Li et al. 2020; Jurenka et al. 2021
ostnu00g01286 06.0.1	Function unknown	scaffold 595	355081	Degree Days	26.4	unknown	
ostnu00g01086 00.0.4	FERM	scaffold 478	76829	Degree Days	24.21	Tissue growth, eye development, actin recruitment	Gaspar et al. 2015
ostnu00g01191 81.0.2	Malate dehydrogenase	scaffold 184	481756	Latitude	33.53	Lipid biosynthesis, hormonally regulated during development in fruit flies. In *Ostrinia*, malate dehydrogenase is downregulated in the winter. Found exclusively in resistant strains of *O. furnicalis*.	Farkaš & Knopp 1997; Koštál et al. 2006; Shabbir et al. 2020; Uzelac et al. 2020
ostnu00g01188 22.0.1	N-alpha-acetyltransferase 15	scaffold 185	26876	Latitude	29.48	RNA and ribosome binding	
ostnu00g01286 06.0.1	Function unknown	scaffold 595	355081	Latitude	28.95	unknown	
ostnu00g01275 36.0.1	Globin	scaffold 557	64083	Latitude	25.99	Oxygen binding and transport	Kawaoka et al. 2009
ostnu00g01135 99.0.1	Neuropeptide receptor A10	scaffold 554	99375	Latitude	25.49	G-protein coupled receptor, involved in feeding and stress response in *Drosophila*. Putatively	Li et al. 2020; Jurenka et al. 2021

						involved in pheromone regulation in Lepidoptera.	
ostnu00g01126 72.0.2	Serpin (serine protease inhibitor)	scaffold 34	925582	Latitude	25.26	Potentially involved in innate immunity	Lin et al. 2017
ostnu00g01089 53.0.2	Function unknown	scaffold 47	1656579	Number of Generations	33.01	unknown	
ostnu00g01263 06.0.3	Chordin	scaffold 326	905338	Number of Generations	31.83	Patterning early in embryonic development	Holley et al. 1995
ostnu00g01226 03.0.1	Function unknown	scaffold 310	457929	Number of Generations	30.26	-	
ostnu00g01033 98.0.1	Splicing factor 3B subunit 2	scaffold 2463	8185	Number of Generations	28.16	Gene regulation, immune suppression	Matsumoto et al. 2021
ostnu00g01055 28.0.1	E3 ubiquitin-protein ligase MYLIP-A	scaffold 519	91986	Number of Generations	26.27	Lipoprotein receptor degradation	Calkin et al. 2011
ostnu00g01022 86.0.11	Protein lap4-like isoform 4	scaffold 455	859278	Number of Generations	25.57	Many functions in cell development, regulation, and proliferation	Guo et al. 2015
ostnu00g01253 93.0.1	Serine/arginine repetitive matrix protein 1	scaffold 1446	12309	Step cline	30.92	Pre-mRNA processing, gene regulation, cell signaling	Giannakouros et al. 2011
ostnu00g01128 22.0.1	Alpha-galactosidase/alpha-n-acetylgalactosaminidase	scaffold 669	269681	Step cline	29.91	N-glycan processing	Ikegaya et al. 2021
ostnu00g01088 19.0.4	Homeobox protein extradenticle	scaffold 432	247003	Step cline	29.59	Transcription factor with downstream targets such as wingless, teashirt, and decapentaplegic. Involved in eye development	Mann & Affolter 1998
ostnu00g01300 52.0.3	Transmembrane protein 165	scaffold 134	137276	Step cline	28.56	Transmembrane transporter	Wang et al. 2020
ostnu00g01092 72.0.2	Proteinase inhibitor I1	scaffold 45	21429	Step cline	27.91	Found in plants and involved in defense strategies	Jongsma & Bolter 1997
ostnu00g01104 68.0.4	Zinc finger	scaffold 479	79167	Step cline	24.44	DNA expression regulation	Laity et al. 2001

ostnu00g01074 03.0.3	Caskin-2-like isoform X3	scaffold 104	404436	Step cline	24.44	Protein kinase regulator activity, participates in wnt signaling	Weng et al. 2011
ostnu00g01264 28.0.2	Patched 1	scaffold 1415	53528	Step cline	24.1	Development, cell growth and differentiation	Ingham et al. 1991

Table S2.6. Betascan outliers (top 0.01%) for each locality and putative functional significance. Organized with inferred transposable elements and retrotransposons at the bottom.

Scaffold	Position on scaffold	Localities with hit	Beta score	Annotation	Likely TE / RTE	Function	References
scaffold182	64859	P1, P2, P3, P4, P6, P8, P10	60.405279	near to Putative cleavage and polyadenylation specificity factor 100 kDa subunit	n	Regulation of polyadenalation in the cytoplasm	Dickson et al. 1999
scaffold43	118132	P1, P2, P3, P4, P6, P8, P9, P10	41.871068	near to Cytochrome P450 49a1-like protein	n	Heme-binding, catalytic properties. Cytochromes have been demonstrated to have a role in lepidopteran strategies to overcome host-plant defenses.	Wang et al. 2015; Werck-Reichhart & Feyereisen 2000
scaffold444	75283	P1, P2, P3, P4, P6, P8, P9, P10	39.497657	Putative epidermal growth factor receptor	n	Developmental signaling network receptor. Involved in growth rate, limb, eye, and gonad development, as well as body size.	Sharabi et al. 2013
scaffold572	98878	P1, P2, P3, P4, P6, P8, P9, P10	39.205223	Bestrophin 2	n	Formation of Cl- channels. These anion channels have been shown to regulate lepidopteran olfactory receptor neuron function.	François et al. 2012
scaffold332	135040	P2, P3, P4, P6, P8, P10	37.331009	Tropomodulin isoform 1	n	Regulation of actin filament architecture, a key gene in invertebrate muscle structure. Found to be under selection for thermal tolerance in green crabs. Evidence for selection in plant-retrovirus-insect vector interactions.	Hooper et al. 2008; Yamashiro et al. 2012; Chen et al. 2017; Tepolt & Palumbi 2020
scaffold504	96453	P2, P8	36.975003	Proton-coupled amino acid transporter 4-like protein	n	Transmembrane protein involved in amino acid transport. Found differentially expressed in *O. furnicalis* cell lines and may play a role in cellular immune function.	Hu et al. 2021
scaffold17	1108594	P4, P10	41.615798	Slowpoke-binding protein-like protein	n	Protein kinase activity. Involved in negative regulation of ATPase and ion transport, neuromuscular synaptic transmission, and starvation response. Involved in circadian pathways in *Drosophila*. Expression is reduced in neurons of flies that lack pigment dispersing factor (PDF).	Jaramillo et al. 2004
scaffold185	58920	P4	35.650366	N-alpha-acetyltransferase	n	Cell growth, development, and differentiation. An N-acetyltransferase has been shown to have functional significance in circadian rhythms in algae.	Matsuo et al. 2012; Lee et al. 2018

scaffold38	327014	P10	35.840015	Mediator of RNA polymerase II transcription subunit 13-like protein	n	"Universal" eukaryotic mediator. Provides a large physical interface for regulation of novel, gene specific transcription factors.	Boube et al. 2002
scaffold242	959176	P1, P2	40.36236	Zinc finger	n	DNA binding and regulation	Laity et al. 2001
scaffold556	478120	P1, P2, P3, P4, P6, P8, P9, P10	46.158632	Endonuclease/exonuclease/ phosphatase	y	Intracellular signaling	Tay et al. 2010
scaffold17	248883	P1, P2, P3, P6, P8, P9	44.236502	Gag-pol polyprotein	y	Retrovirus protein subunits	Hajek & Friesen 1998
scaffold261	917	P1, P6, P8, P9	44.085783	Pol polyprotein	y	Retrovirus protein subunits	Friesen & Nissen 1990
scaffold40	186803	P1, P2, P3, P4, P6, P8, P9, P10	43.641313	Similar to gag-pol polyprotein	y	Retrovirus protein subunits	
scaffold390	297081	P1, P2, P3, P6, P8, P9, P10	40.375472	near to Pao retrotransposon peptidase	y	Pao is a common retrotransposible element.	Xiong et al. 1993
scaffold110	514179	P1, P2, P3, P4, P9	40.280123	Similar to putative gag-pol protein	y	Retrovirus protein subunits	
scaffold274	451500	P1, P2, P3, P10	39.104401	Similar to putative gag-pol protein	y	Retrovirus protein subunits	
scaffold109	101161	P1, P8, P10	38.38246	BEL12|AG transposon polyprotein	y	Retrotransposon with single ORF. May have functional importance in DNA regulation.	Nefedova & Kim 2009
scaffold18	305664	P2, P3, P4, P6, P8, P9, P10	38.06523	Similar to RNA-directed DNA polymerase	y	Retroviral domain	Tay et al. 2010
scaffold564	284732	P2, P4, P6, P8, P9	37.579089	Retroelement polyprotein	y	Retrovirus protein subunits	
scaffold305	20617	P2, P3, P6, P8, P9, P10	37.411221	BEL12|AG transposon polyprotein	y	Retrovirus protein subunits	
scaffold363	380187	P2	37.402839	Reverse transcriptase	y	Retroviral domain	
scaffold46	171522	P8, P9	38.742735	Reverse transcriptase	y	Retroviral domain	
scaffold538	453704	P8	38.525999	Retroelement polyprotein	y	Retrovirus protein subunits	
scaffold194	346124	P9	37.867899	Endonuclease-reverse transcriptase	y	Retroviral domain	

Table S2.7. Top 25 candidate loci identified by JTKcycle 3.0 and putative functional significance.

Scaffold	Location	Gene	Annotation	Adjusted p-value	Period length (km)	Amplitude	Function	Literature cited
scaffold204	1001207	ostnu00g0128318	Wnt inhibitory factor 1	0.0000827	402.5	0.102	Signal transduction pathway gene, involved in embryonic development and cell differentiation	Nusse & Varmus 1992; Ding et al. 2019
scaffold498	63488	ostnu00g0113040	Neuropeptide receptor A31	0.0004464	483	0.112	Regulator of physiological functions, involved in development, reproduction, behavior, and feeding	Li et al. 2020
scaffold186	251199	6kb from ostnu00g0126144	Activating signal cointegrator 1 complex subunit 1-like protein	0.0017278	402.5	0.157	DNA damage repair pathway protein	Soll et al. 2018
scaffold20	330474	30 kb from ostnu00g0102080	function unknown	0.0017278	402.5	0.145	-	
scaffold14	466226	9kb from ostnu00g0107007	Dual specificity protein kinase pom1-like isoform X3	0.0017278	402.5	0.137	Cell length and mitotic entry control via stabilization of microtubules	Martin et al. 2009; Soppa et al. 2015
scaffold718	225614	ostnu00g0115867	Helicase senataxin-like protein	0.0017278	322	0.114	DNA damage repair pathway protein	Becherel et al. 2013
scaffold532	79303	PERIOD	period	0.0017278	322	0.113	Control of post-diapause development timing (PDD)	Dopman et al. 2005; Levy et al. 2015; Kozak et al. 2019
scaffold669	340260	1kb from ostnu00g0112822	Alpha galactosidase / alpha n acetylgalactosaminidase	0.0017278	483	0.109	N-glycan processing	Ikegaya et al. 2021
scaffold202	1078152	ostnu00g0128318	Wnt inhibitory factor 1	0.0017278	483	0.106	Signal transduction pathway gene, involved in embryonic development and cell differentiation	Nusse & Varmus 1992; Ding et al. 2019
scaffold884	42374	ostnu00g0115594	Sulfotransferase 1C4-like protein	0.0017278	483	0.102	Formation of sulfate conjugates: downregulated in response to insecticides	Coughtrie 2002; Jia et al. 2020
scaffold24	75928	ostnu00g0106776	WD domain	0.0017278	402.5	0.102	Structural motif of WD proteins, which are involved in signal transduction, transcription regulation, and cell cycle control	Stirnimann et al. 2010
scaffold1372	21918	ostnu00g0111315	Function unknown	0.0035466	322	0.126		

scaffold2794	2385	ostnu00g0120086	Sodium/potassium-transporting ATPase subunit alpha-like protein	0.0035466	483	0.112	Salinity control and osmotic stress response	Sebastian et al. 2018
scaffold335	93869	0.2 kb from ostnu00g0114276	Interferon regulatory factor 2-binding protein 1 & 2	0.0053654	483	0.150	Catalytic activity, organ size regulation, tissue homeostasis, and tumorigenesis	Feng et al. 2020
scaffold871	264677	3kb from ostnu00g0130547	Function unknown	0.0053654	483	0.134		
scaffold429	12542	ostnu00g0104091	Function unknown	0.0053654	483	0.128		
scaffold444	248362	46kb from ostnu00g0104313	Function unknown	0.0053654	483	0.119		
scaffold363	379163	10kb from ostnu00g0103615	Function unknown	0.0053654	483	0.116		
scaffold815	155707	1 kb from ostnu00g0126399	Function unknown	0.0053654	402.5	0.116		
scaffold2794	2338	ostnu00g0120086	Sodium/potassium-transporting ATPase subunit alpha-like protein	0.0053654	483	0.111	Salinity control and osmotic stress response	Sebastian et al. 2018
scaffold1102	64801	51kb from ostnu00g0113372	Function unknown	0.0053654	483	0.111		
scaffold82	992028	ostnu00g0119712	Cell division control protein 45 homolog	0.0053654	483	0.109	Cell cycle control	Gautier et al. 1988
scaffold1009	41881	ostnu00g0120478	Function unknown	0.0053654	483	0.109		
scaffold226	536526	ostnu00g0105567	P94-like protein	0.0053654	483	0.106	Muscle specific protein, modulator protease in cellular processes	Ono et al. 2004
scaffold212	98353	0.1 kb from ostnu00g0109904	Protein hairy-like protein	0.0053654	483	0.102	Transcription regulation and control of cell fate: neurogenesis and myogenesis	Fisher et al. 1996

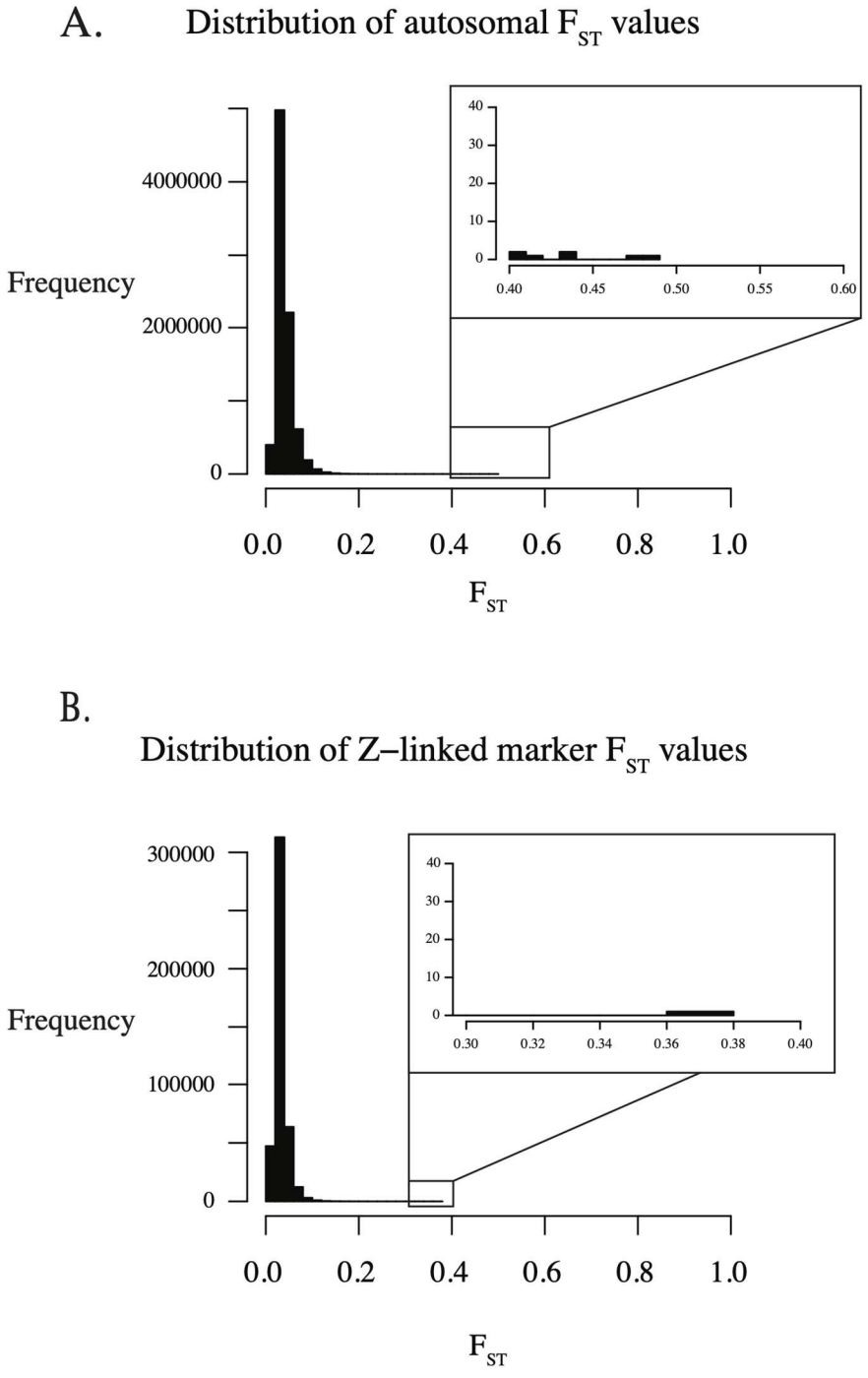

Figure S2.1. Distribution of F_{ST} values for all pairwise comparisons between transect populations. Mean F_{ST} calculated from 1000 bp windows with > 20 SNPs and > 0.2 coverage within the window using the Weir and Cockerham method. Panel A shows data from autosomal markers and Panel B shows data from Z-linked markers. Inset figures show the tail ends of the distribution (F_{ST} > 0.3) for each set of comparisons. Note the change in axes between outer and inlaid figures.

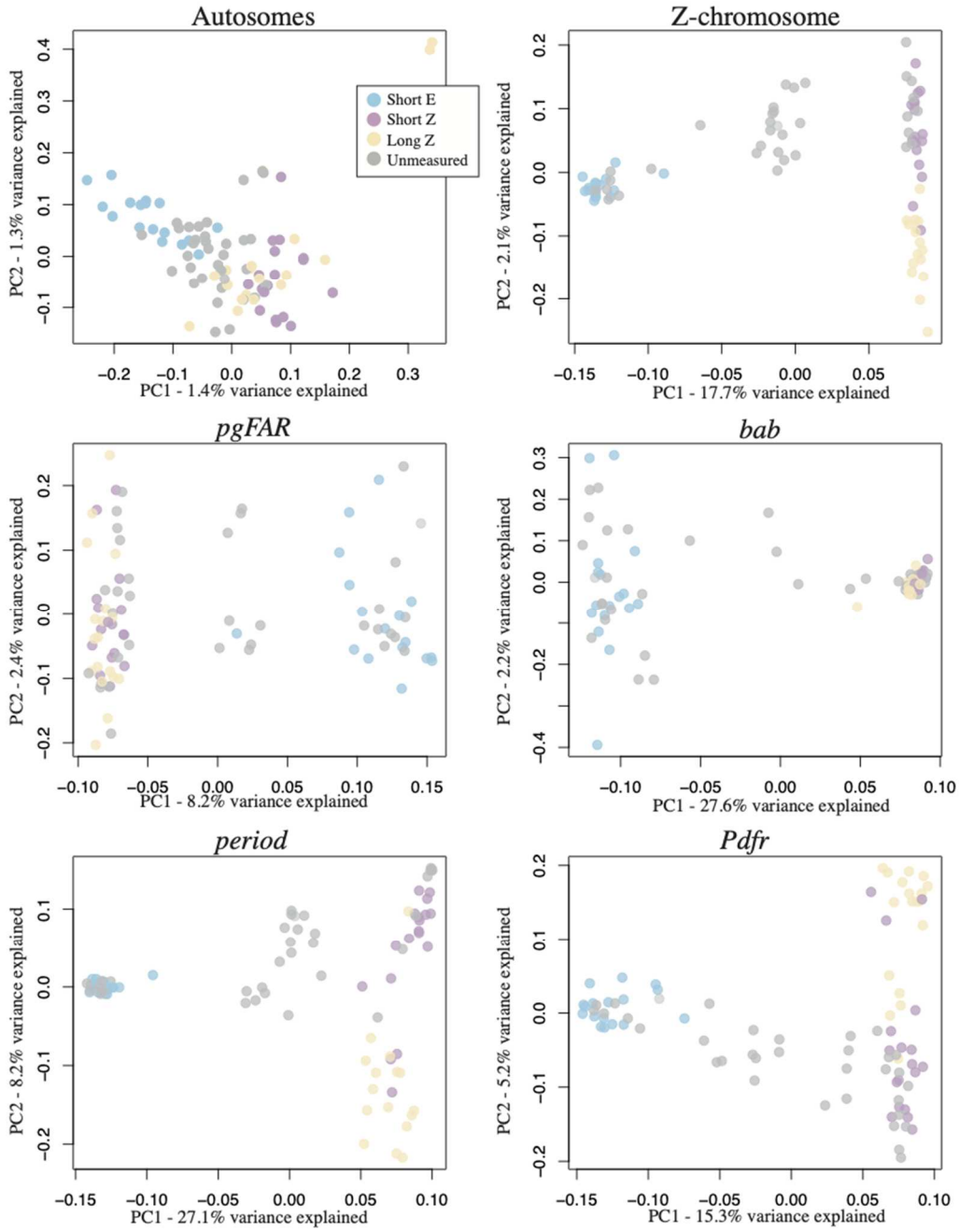

Figure S3.1. PCAs of a subset of samples with outgroups ALB, ABB, and ACB removed. Each sample represents an individual, with the color corresponding to the combined pheromone and phenology phenotype of the individual, if known. Short E corresponds to E pheromone types that have been measured to have short developmental timing. Short Z corresponds to Z pheromone types that have been measured to have short developmental timing. Long Z corresponds to Z pheromone types that have been measured to have long developmental timing. Some samples were not measured for PDD timing, and are indicated in grey.

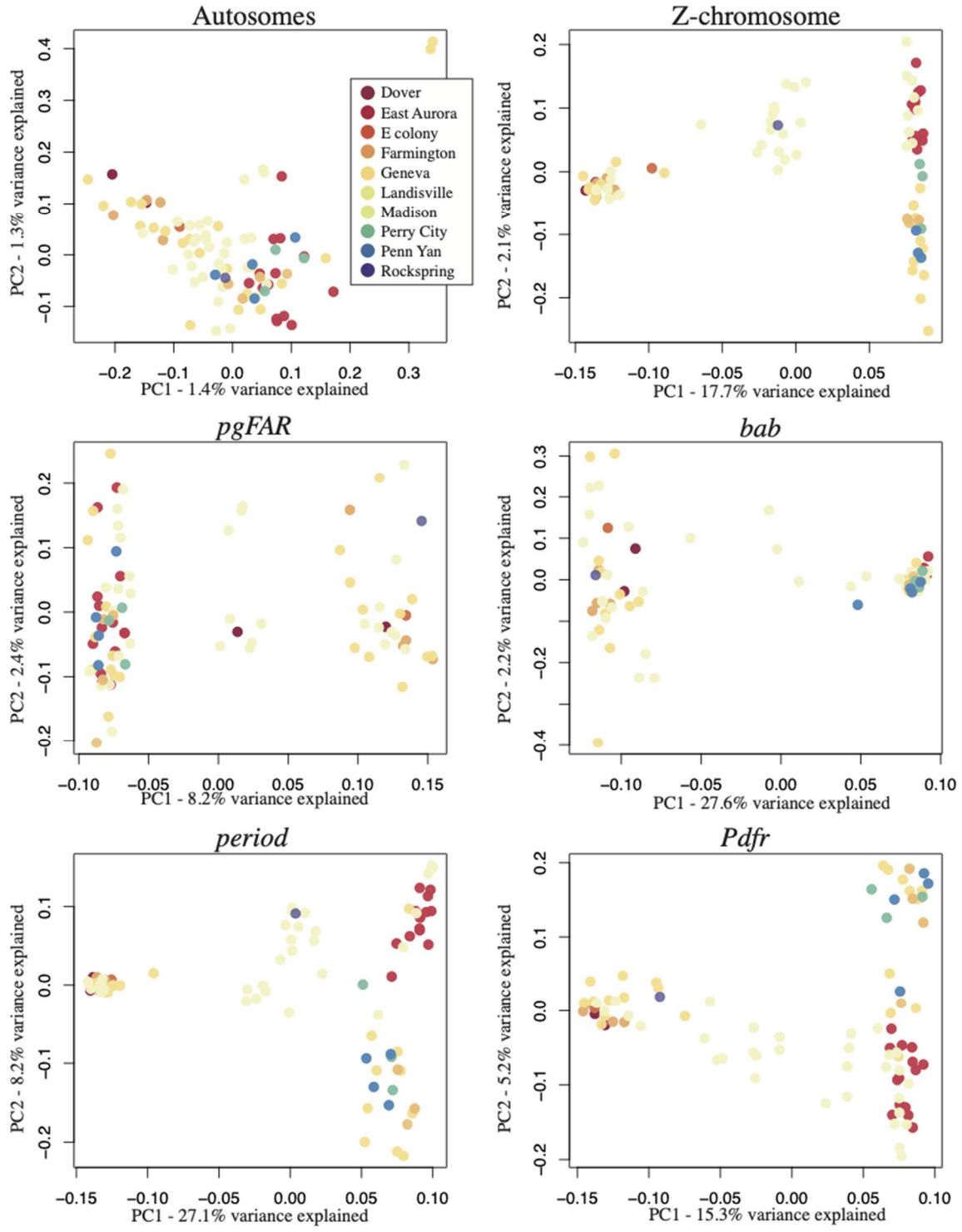

Figure S3.2. PCAs of a subset of samples with outgroups ALB, ABB, and ACB removed. Each sample represents an individual, with the color corresponding to their original collection location.