Genome Resequencing Reveals Rapid, Repeated Evolution in the Colorado Potato Beetle

Benjamin Pélissié,¹,¹ Yolanda H. Chen,² Zachary P. Cohen,¹ Michael S. Crossley,⁴,¹ David J. Hawthorne,³ Victor Izzo,² and Sean D. Schoville*,¹

¹Department of Entomology, University of Wisconsin-Madison, Madison, WI, USA
²Department of Plant and Soil Science, University of Vermont, Burlington, VT, USA
³Department of Entomology, University of Maryland, College Park, MD, USA
¹Present address: Department of Biology, University of Nebraska-Kearney, Kearney, NE, USA
⁴Present address: Department of Entomology & Wildlife Ecology, University of Delaware, Newark, DE, USA

*Corresponding author: E-mail: sean.schoville@wisc.edu.
Associate editor: Victoria Sork

Abstract

Insecticide resistance and rapid pest evolution threatens food security and the development of sustainable agricultural practices, yet the evolutionary mechanisms that allow pests to rapidly adapt to control tactics remains unclear. Here, we examine how a global super-pest, the Colorado potato beetle (CPB), Leptinotarsa decemlineata, rapidly evolves resistance to insecticides. Using whole-genome resequencing and transcriptomic data focused on its ancestral and pest range in North America, we assess evidence for three, nonmutually exclusive models of rapid evolution: pervasive selection on novel mutations, rapid regulatory evolution, and repeated selection on standing genetic variation. Population genomic analysis demonstrates that CPB is geographically structured, even among recently established pest populations. Pest populations exhibit similar levels of nucleotide diversity, relative to nonpest populations, and show evidence of recent expansion. Genome scans provide clear signatures of repeated adaptation across CPB populations, with especially strong evidence of selection on insecticide resistance genes in different populations. Analyses of gene expression show that constitutive upregulation of candidate insecticide resistance genes drives distinctive population patterns. CPB evolves insecticide resistance repeatedly across agricultural regions, leveraging similar genetic pathways but different genes, demonstrating a polygenic trait architecture for insecticide resistance that can evolve from standing genetic variation. Despite expectations, we do not find support for strong selection on novel mutations, or rapid evolution from selection on regulatory genes. These results suggest that integrated pest management practices must mitigate the evolution of polygenic resistance phenotypes among local pest populations, in order to maintain the efficacy and sustainability of novel control techniques.

Key words: population genomics, insecticide resistance, rapid evolution, genetic adaptation, regulatory evolution, insect pest, polygenic trait.

Introduction

Herbivorous pests cause an estimated 18–20% damage to crops and cost nearly $470 billion annually on a global scale (Sharma et al. 2017). The ability of insect pests to evolve resistance to insecticides threatens food security and the development of sustainable agricultural practices, especially when their rate of evolution outstrips the development of novel control strategies (Lewis et al. 1997; Chen and Schoville 2018; Gould et al. 2018). This is the case with insect “super-pests,” which repeatedly evolve insecticide resistance even as they are faced with completely novel insecticides, thus perpetuating the arms race that defines the pesticide treadmill (McCaffery and Nauen 2006). Curiously, particular super-pest species or even select populations are more likely to adapt to new compounds, suggesting that there is a genetic basis in the propensity to evolve resistance (Brevik et al. 2018). Yet, despite more than 60 years of research on the evolution of resistance (Crow 1957), the relative importance of alternative mechanisms that underlie the evolutionary potential for pesticide resistance evolution are still unclear (ffrench-Constant and Bass 2017; Pélissié et al. 2018). Although a considerable effort has been placed on understanding the proximal molecular control of resistance (ffrench-Constant 2013), broader questions about the genetic complexity of resistance, mode of selection, and geographical extent of adaptation have rarely been studied (Daborn et al. 2002; Labbé et al. 2007; Kamdem et al. 2017). Although population genetic models of resistance...
management have been highly effective in certain management scenarios (Tabashnik et al. 2013), observed patterns of insecticide resistance evolution defy many of the assumptions of our evolutionary models (Hoy 1998). Recent genomic resequencing data sets suggest that resistance evolution is sometimes geographically and genetically complex (Cheng et al. 2017; Anderson et al. 2018; Lucas et al. 2019; Calla et al. 2020).

A key goal should be to understand the evolutionary processes that allow species to become pests, particularly the mechanisms underlying phenotypic shifts that result in economically damaging pest outbreaks (Fritz et al. 2018; Péllissié et al. 2018). A prevalent view is that pests, including invasive species, often retain substantial genetic diversity that facilitates evolution in agroecosystems (Lee 2002; Guillemaud et al. 2018). A prevalent view is that pests, including invasive species, often retain substantial genetic diversity that facilitates evolution in agroecosystems (Lee 2002; Guillemaud et al. 2018; Kirk et al. 2013). However, there is increasing recognition that evolution can be rapid irrespective of levels of standing genetic diversity (Hawkins et al. 2019). Although not all pests exhibit rapid rates of adaptation to insecticides (Brevik et al. 2018), insect super-pests often demonstrate repeated rapid evolution of similar resistance phenotypes (Georghiou and Taylor 1986; May and Dobson 1986).

Rapid evolution is defined as a shift in phenotype from underlying variation in exceptionally few generations (Hairston et al. 2005), and can occur as a result of several mechanisms (Whitehead et al. 2017). First, selection can act on novel mutations, which may arise frequently if pests have intrinsically large population sizes (much greater than \( \geq 10^6 \)) and are not mutation-limited (Barton 2010; Cutter et al. 2013). This could lead to repeated evolution of resistance among different populations, often with independent mutations at loci underlying resistance phenotypes (Rinkevich et al. 2012). Second, as a special case of the first mechanism, key mutational changes could affect a master regulatory gene (López-Maury et al. 2008; Roelofs et al. 2010; Reid et al. 2016), where mutations drive expression of the same downstream molecular pathways in different populations. Rapid gene regulatory evolution has been raised as a possible mechanism underlying repeated evolution of pesticide resistance in the spider mite *Tetranychus urticae*, where it has been linked to a transcriptional cascade in xenobiotic detoxification (Dermauw et al. 2013). Third, an alternative pathway of rapid evolution would draw on standing genetic variation (Crow 1957). Standing genetic variation is increasingly viewed as a common source of rapid adaptive variation (Messer and Petrov 2013), where the initial frequency of mutations in a population determine the rate at which populations respond to selection pressures (Barton and Keightley 2002; Pujol et al. 2018). Although population size must typically be large to retain large reservoirs of standing variation, admixture among divergent populations can increase standing variation (De Carvalho et al. 2010; Corrêa et al. 2019). In an extreme case, ancestral standing genetic variation can lead to parallel selection, where the same genes are selected in independent populations resulting in convergent phenotypes (Tennessen and Akey 2011). However, standing genetic variation can also be present in the form of redundancy in molecular pathways that are critical to pesticide resistance phenotypes (e.g., in generalist herbivorous insects that specialize on toxic plants: Hardy et al. 2018; Rane et al. 2019; Láruson et al. 2020). It should be emphasized that these mechanisms of adaptation need not be exclusive, yet it remains unclear how each contributes to the evolutionary success of the top arthropod super-pests. Emerging genomic data sets provide the opportunity to detect and quantify the importance of different mechanisms underlying rapid evolutionary change by screening for genomic signatures of selection (Péllissié et al. 2018).

The Colorado potato beetle (CPB), *Leptinotarsa decemlineata*, is a global super-pest and an especially tractable exemplar of rapid evolution to insecticides. CPB has evolved resistance to over 50 different insecticides in all the major classes, in some cases within the first year of use (Alyokhin et al. 2008). CPB has demonstrated an ability to rapidly evolve in response to a wide range of environmental pressures, including host-plant defenses and climatic variability (Petek et al. 2012; Izzo et al. 2014). This super-pest originated in the Great Plains region of the United States (Izzo et al. 2018), following a host shift to potato (an introduced crop) in the mid-19th century (around 1859) that allowed for rapid spatial expansion from Nebraska to the Eastern US in a 20-year period and colonization of Eurasia by the early 1900s (Walsh 1866; Lu and Lazell 1996; Grapputo et al. 2005). Despite rapid spatial expansion, populations are genetically differentiated (Grapputo et al. 2005; Izzo et al. 2018) and insecticide resistance is geographically heterogeneous (Crossley et al. 2018; Szendrei et al. 2012; Dively et al. 2020), even over local landscape scales (Crossley et al. 2017). In particular, beetles from Long Island, New York are known to have the highest levels of baseline resistance and are typically the first populations to develop resistance to all compounds (Hitchner et al. 2012; Dively et al. 2020), whereas populations in the Pacific Northwest remain susceptible to insecticides despite an equivalent duration of usage and comparable treatment practices (Rondon 2012; Dively et al. 2020). Nonpest populations are native to the Great Plains and Mexico, where they use ancestral host plants (primarily *Solanum rostratum*) (Hsiao 1978). Closely related congeners in the genus *Leptinotarsa* are sympatric in the southern part of CPB’s geographical range (Hsiao 1988). It is unknown whether multiple populations, or species, contribute to CPB pest variation through past episodes of admixture (Izzo et al. 2018). By integrating across this diversity, CPB can serve as a model for understanding evolutionary mechanisms that facilitate and constrain rapid evolution.

Here, we leverage the recent publication of the CPB genome (Schoville et al. 2018) to investigate whether repeatable patterns of evolution occur in highly resistant pest populations. We compare CPB genomic and transcriptomic variation across populations in the United States, Mexico, and Europe, as well as nine closely related *Leptinotarsa* species, to assess three competing models of rapid evolution: pervasive selection on de novo mutation (independent hard
selective sweeps in geographically separate populations), rapid regulatory evolution, or repeated selection on standing genetic variation (see Table 1 for predictions). We also assess whether parallel selection on the same ancestral genomic variation has led to convergent phenotypes underlying neonicotinoid resistance in six geographic regions.

Our approach provides an important insight on the genetic basis for traits that are adaptive in agriculture such as insecticide resistance. Traditional models of insecticide resistance evolution assume that resistance is caused by a simple mutation at a single gene (Roush 1989; Hawkins et al. 2019), whereas repeated selection by insecticides on standing genetic variation likely results in polygenic trait architecture (multiple genes contributing to resistance phenotypes), which we assess based on the number of selected genes and pathways, as well as the size of selective sweeps (Pritchard et al. 2010). In practice, however, it is difficult to determine the cause of polygenic trait architecture. Although polygenic adaptation to a single insecticide (or mode of action) alone can generate polygenic architectures in resistance traits (Crow 1957), its effects are difficult to distinguish from the genomic consequences of independent bouts of insecticide adaptation without appropriate temporal sampling. Since many CPB pest populations have shown some level of resistance to multiple insecticides and/or multiple modes of action (Crossley et al. 2017), we avoid speculating about the potential role of polygenic insecticide adaptation. Instead, we focus on evidence for polygenic architecture of resistance as a product of selection on standing genetic diversity in CPB. To improve our interpretations of insecticide resistance patterns, we provide a detailed description of genomic diversity patterns, evolutionary relationships, and the population history of CPB pest lineages, in order to understand how expansion history or admixture have influenced geographical variation in insecticide resistance. Over the long-term, by improving our understanding of the evolutionary processes and genomic mechanisms underlying the ability to repeatedly evolve insecticide resistance in super-pests, integrated pest management strategies can be developed to provide more sustainable agricultural practices (Kirk et al. 2013; Chen and Schoville 2018).

### Results

#### Extensive Genomic Diversity within CPB

To provide a broad-scale assessment of genomic diversity patterns, we examined short-read whole-genome sequences for a geographically dispersed set of 85 CPB samples, as well as nine additional Leptinotarsa species (fig. 1 and supplementary file 1: table S1 and fig. S2, Supplementary Material online). CPB samples were selected to maximize information about genomic differentiation across the range of CPB, but also include six geographically proximate pairs of neonicotinoid (imidacloprid) susceptible and resistant samples to test for parallel selection to this compound: Maine (R) and Vermont (S), New York (R) and New Jersey (S), Maryland (R and S), Michigan (R and S), Wisconsin (R and S), and Oregon (R and S). However, all pest populations may have evolved resistance to legacies of previous insecticides. Employing best practices in genotype ascertainment, we sequenced each sample (most resulted in coverage $>4×$, supplementary file 1: fig. S3, Supplementary Material online), and found that CPB shows considerable genomic diversity, with 76,647,868 single-nucleotide polymorphisms (SNPs) recovered from the nuclear...
Fig. 1. (A) Unrooted phylogenetic tree of *Leptinotarsa* species obtained with SNPhylo, based on 35,838 SNPs (after LD-based reduction). Node labels represent bootstrap values. The blue arrow highlights a monophyletic clade comprising CPB samples collected in the United States and Europe. (S), imidacloprid susceptible population; (R), imidacloprid resistant population. (B) Geographical sampling of *Leptinotarsa decemlineata* and estimated admixture coefficients. Admixture proportions were estimated with SNMF on the intergenic “CPB” data set for \( k = 6 \) clusters, and are shown as both pie charts and an individual bar-plot. Each pie chart represents a sampled location (small charts for single samples; large ones for populations of five individuals), referenced as a number. Colored boxes around large pie charts differentiate susceptible (green) versus resistant samples (red). (C) Population demographic histories (median \( Ne \) only) estimated from SMC++. Colors correspond to geographical regions.
genome (supplementary file 1: table S2, Supplementary Material online; the CPB genome is estimated to be ~670 Mb in size). Following variant recalibration, there were a total of 47,969,460 SNPs, of which 30,973,249 were in intergenic regions.

We estimated genome-wide nucleotide diversity (π) using a 10-kb sliding window. Within-population genetic diversity of nonpest CPB samples from the inferred ancestral source population, the US Plains region, was high with an average π = 0.005. Field samples of the US pest populations had less nucleotide diversity (average π = 0.0028 and 0.0030 for colored pairs of pesticide resistant and susceptible populations, respectively; supplementary file 1: table S3 and fig. S4, Supplementary Material online), but pooling colored pairs of field samples increased nucleotide diversity in a given agricultural region by 40% (reaching an average π = 0.008). There is thus no difference in π among pest and nonpest regions, showing that high levels of variation have been retained in the beetle’s expansion into agricultural regions. Individual heterozygosity appeared to be reduced relative to expectations under random mating for all populations, as measured by the inbreeding coefficient (FIS). Inbreeding values were slightly larger in pest populations (on average FIS = 0.603 and 0.557 for susceptible and resistant populations, respectively; supplementary file 1: table S4 and fig. S5, Supplementary Material online), relative to Plains individuals (on average FIS = 0.531), but mean values are not significantly different (P = 0.453). The New Jersey lab population, which was maintained as a breeding colony for pesticide assays (since the mid-1980s, but after a history of insecticide exposure and reported cases of resistance), showed the highest level of inbreeding (FIS = 0.723) and was significantly different from the nearby New York population (P = 0.008). Susceptible and resistant pest samples had a comparable number of private alleles (93,407 vs. 89,599, respectively; supplementary file 1: figs. S6 and S7, Supplementary Material online), but Plains samples had three times as many private alleles (93,407 vs. 89,599, respectively; supplementary file 1: tables S5 and S6, Supplementary Material online) and empirical FST quantiles broadly overlapped, but pairwise FST exceeded >0.1 when New Jersey or the Michigan resistant population were compared. Principal component analysis shows clear geographical patterns of population genetic structure (supplementary file 1: figs. S8–S10, Supplementary Material online). Due to their significant genetic divergence, all Mexican CPB, the Arizona sample, and other Leptinotarsa species were removed from analyses of population structure and demographic change, as their inclusion would violate statistical assumptions in those approaches. Admixture-based clustering (fig. 1B and supplementary file 1: figs. S11 and S12, Supplementary Material online) converges with the principal component analysis in supporting six populations, which are identified hierarchically as the New-Jersey population (at K = 2), a western population (Oregon plus Idaho, at K = 3), a distinctive Michigan resistant population (at K = 4), a Plains population (at K = 5), and a Northwestern and Eastern US population (which includes the introduced European samples, at K = 6). Admixture tests using the D statistic were examined among CPB pest and nonpest populations, as well as with other Leptinotarsa species. These tests provided limited evidence of admixture contributing to genetic diversity of pest lineages (fig. 2), ruling out the hypothesis of standing genetic variation increasing in the pest lineage as a result of hybridization. The highest Dmin values suggest historical admixture from Mexico to the Plains (0.036) and Mexico to the Western populations (0.044), and limited ongoing gene flow between New York and the Michigan susceptible population (D = 0.015). Similarly, an assessment of admixture with other Leptinotarsa species (supplementary file 1: fig. S13, Supplementary Material online) does not suggest recent gene flow into pest populations.

**Evolutionary Diversification and Demographic History**

In order to examine the evolutionary history of CPB, we investigated phylogenetic and population genetic divergence patterns. Phylogenetic reconstruction (fig. 1A) clearly separated the Mexican samples from all the US samples, and pairwise FST suggests genomic divergence of Mexican samples relative to the US Plains region (Northern Mexico vs. Plains = 0.187, and Southern Mexico vs. Plains = 0.619). All CPB populations from potato growing regions in the United States and Eurasia formed a well-supported monophyletic group, whereas separate clades contained samples in the Pacific Northwest, Southeast US and the US Plains region. The Arizona CPB sample was found outside the CPB clade, next to the L. lineolata sample, suggesting high divergence. This placement appears to be an effect of long-branch attraction driven by a large number of private alleles (supplementary file 1: fig. S7, Supplementary Material online), as the Arizona sample is genetically related to other Plains samples in a principal component analysis (see below). Interestingly, CPB from Mexico and other Leptinotarsa species were mixed together, suggesting that Mexican CPB are also strongly divergent and may represent unidentified cryptic Leptinotarsa species.

Genetic divergence among the US CPB pest and nonpest populations was generally limited (average pairwise FST = 0.09; supplementary file 1: tables S5 and S6, Supplementary Material online) and empirical FST quantiles broadly overlapped, but pairwise FST exceeded >0.1 when New Jersey or the Michigan resistant population were compared. Principal component analysis shows clear geographical patterns of population genetic structure (supplementary file 1: figs. S8–S10, Supplementary Material online). Due to their significant genetic divergence, all Mexican CPB, the Arizona sample, and other Leptinotarsa species were removed from analyses of population structure and demographic change, as their inclusion would violate statistical assumptions in those approaches. Admixture-based clustering (fig. 1B and supplementary file 1: figs. S11 and S12, Supplementary Material online) converges with the principal component analysis in supporting six populations, which are identified hierarchically as the New-Jersey population (at K = 2), a western population (Oregon plus Idaho, at K = 3), a distinctive Michigan resistant population (at K = 4), a Plains population (at K = 5), and a Northwestern and Eastern US population (which includes the introduced European samples, at K = 6). Admixture tests using the D statistic were examined among CPB pest and nonpest populations, as well as with other Leptinotarsa species. These tests provided limited evidence of admixture contributing to genetic diversity of pest lineages (fig. 2), ruling out the hypothesis of standing genetic variation increasing in the pest lineage as a result of hybridization. The highest Dmin values suggest historical admixture from Mexico to the Plains (0.036) and Mexico to the Western populations (0.044), and limited ongoing gene flow between New York and the Michigan susceptible population (D = 0.015). Similarly, an assessment of admixture with other Leptinotarsa species (supplementary file 1: fig. S13, Supplementary Material online) does not suggest recent gene flow into pest populations.

Demographic reconstruction of CPB populations using SMC++ and Stairway plot analysis (fig. 1C and supplementary file 1: figs. S14 and S15, Supplementary Material online) showed consistent population size fluctuations through time, with similar trajectories of effective size (Ne) for all pest populations. Nearly all agricultural populations exhibited recent population size increases in the SMC++ analysis, most notably in the Eastern US. The split time analysis suggested an early split of the western populations from the Plains region, and subsequent near-simultaneous split times of Midwest and Eastern populations (supplementary file 1: fig. S16, Supplementary...
Applying a mutation rate from other insect taxa \((2.1 \times 10^{-9} \text{ substitutions per site per generation})\) suggests that populations contracted between 300k and 100k years ago, expanded between 200k and 70k years, and declined again until 10k to 5k years ago. The splits of most pest populations from the sampled populations in the Plains occurred between 21k and 11k years, during the transition from the late Pleistocene to early Holocene.

**Figure 2.** Heatmap of \(D\)-statistics, showing the introgression patterns among CPB populations. The color of the heatmap cell indicates the most significant \(D_{\text{min}}\) found with every population pairs: red colors indicate higher \(D\)-statistics, and generally more saturated colors indicate higher \(P\) values. The complete biallelic data set was analyzed.

**Genome-Wide Patterns of Natural Selection**

Evidence of natural selection acting on genomic diversity was first assessed across the entire geographical range of the United States to test hypotheses about de novo mutation and standing genetic variation contributing to insecticide resistance in CPB. Outlier-based tests correcting for population differentiation identified 0.37% of all SNPs as outliers (i.e., 65,815 out of 17,599,906, with a false discovery rate, or FDR, of 0.01%). A total of approximately 32% of the outlier SNPs could be assigned to 8,760 known genes (supplementary file 1: tables S7 and S8, Supplementary Material online; gene list provided in supplementary file 2, Supplementary Material online). Of these genes, 336 were linked to candidate insecticide resistance genes, including 205 genes involved in detoxification pathways, 91 target sites, and 40 genes involved in cuticle development. The well-known voltage-sensitive sodium channel gene (LDEC011942) that provides knockdown resistance to pyrethroids was
included among the target-site genes. We removed the New Jersey samples to see if this population biased our results: 5,044 of the 8,760 genes remained outliers, whereas 224 of the 336 candidate genes remained outliers (see supplementary file 1: Supplementary Material online, for a list of these genes).

Based on a gene set enrichment analysis of the full outlier data set, overrepresented gene ontology terms were linked to insecticide resistance and/or stress (supplementary file 1: fig. S17, Supplementary Material online). For biological processes, GO terms included oxidation–reduction process and response to oxidative stress (and multiple nested terms), among others. For cellular components, terms linked to insecticide resistance included voltage-gated sodium channel and acetylcholine-gated channel complexes, presynaptic active zone and synapse, and integral component of the membrane. Among the molecular functions, terms such as heme and zinc ion binding (including iron ion binding), extracellular ligand-gated ion channel activity (including voltage-gated sodium channel and acetylcholine receptor activity), glutathione transferase activity, ATP-binding cassette (ABC) transporter activity via the term ATPase activity coupled to transmembrane movement, and peroxidase activity (including cytochrome p450 [CYP] monooxygenase activity) were associated with insecticide resistance.

To leverage another genome scan approach to test for patterns of natural selection, we also employed gene environment association association analysis to examine five different, ecologically relevant environmental predictors of natural selection on the genome: latitude, elevation, precipitation, minimum temperature in the coldest month, and potato land cover. Only 0.02% of the analyzed SNPs (4,098 out of 17,599,906 SNPs, with an FDR of 0.01%) were significantly associated with at least one environmental variable (supplementary file 1: table S9 and fig. S18, Supplementary Material online). SNPs were grouped into Supplementary Material online). On an average (across the genome), only 4.87% of these selection events were singularly associated (supplementary file 1: fig. S22, Supplementary Material online). These singular regions were shared in both tests, and 29 of these are insecticide resistance genes (table 2).

Among the biological processes, terms included chemical synaptic transmission, oxidation–reduction process, proteolysis, DNA repair, and chloride transport, transmembrane transport, and ion transport. Terms associated with cellular components included integral component of membrane, synapse, and presynaptic active zone. Terms associated with molecular functions included pathways such as heme and iron ion binding, ATPase activity coupled to transmembrane movement, GABA and G-protein coupled receptor activity, and monooxygenase and oxidoreductase activity.

Local Adaptation to Insecticides

In order to examine how frequently selection events are population-specific and if they arise from de novo mutations or standing genetic variation, we employed a haplotype-based approach that examines shifts in haplotype frequency along branches of a population tree. Due to the fragmented nature of the reference genome, we examined haplotype frequencies on the longest 95 genomic scaffolds (encompassing ~21% of the genome, all >1 Mb). Our analyses show that 0.01% (7,826 SNPs), 0.01% (7,826 SNPs), and 0.16e−3% (1,106 SNPs) of the markers were significant at \( \alpha = 0.01, 0.001, \) and 0.0001, respectively (see list in supplementary file 1: Supplementary Material online). SNPs were grouped into regions, where each region was separated by at least 1 kb up- and downstream. This resulted in 1,169 selection regions at \( \alpha = 0.01, 140 \) regions at \( \alpha = 0.001, \) and 24 regions at \( \alpha = 0.0001 \) (supplementary file 1: table S11, Supplementary Material online).

Excluding one extremely long region (35 kb in length), the average length of the most significant haplotypes (\( \alpha = 0.0001 \)) was 3.1 kb (supplementary file 1: fig. S21, Supplementary Material online). On an average (across \( \alpha \) levels and branch association thresholds), these regions in the genome were repeatedly selected in multiple populations (on an average, in 6.06 branches of the population tree, and only 4.87% of these selection events were singularly associated with one population; supplementary file 1: table S12, Supplementary Material online). These singular regions tended to be short (1.1 kb on an average). Out of all selected regions (\( \alpha = 0.01 \)), 319 were found in 224 genes and only 3.8% were singular (supplementary file 1: fig. S22, Supplementary Material online).

Among the 224 genes, 16 were candidate insecticide resistance-associated genes, including six ABC transporters, comparison of the results shows that a total of 557 genes (see supplementary file 2, Supplementary Material online) were shared in both tests, and 29 of these are candidate insecticide resistance genes (table 2). The largest group represents xenobiotic detoxification genes, with nine ABC transporters, seven CYP genes, two esterase genes, one MFS gene, and one glutathione S-transferase (GST) gene represented. Target-site genes included four voltage-dependent channel genes and two nicotinic acetylcholine receptors, and three cuticle genes overlap in both tests. Gene set enrichment analysis of significant genes identified as overlapping in the two genome scan tests (supplementary file 1: fig. S20, Supplementary Material online) showed enrichment of gene ontology terms associated with insecticide resistance and stress. Among biological processes, terms included chemical synaptic transmission, oxidation–reduction process, proteolysis, DNA repair, and chloride transport, transmembrane transport, and ion transport. Terms associated with cellular components included integral component of membrane, synapse, and presynaptic active zone. Terms associated with molecular functions included pathways such as heme and iron ion binding, ATPase activity coupled to transmembrane movement, GABA and G-protein coupled receptor activity, and monooxygenase and oxidoreductase activity.
**Table 2.** Candidate Resistance Genes Identified in Both PCAdapt and LFMM.

| Mechanisms/Metabolic detoxification | Categories | Gene ID | Annotated Gene Name | Principal Component |
|-------------------------------------|------------|---------|---------------------|---------------------|
| Metabolic detoxification            | ABC transporters | LDEC015007 | ABC subfamily a member 5-like isoform | 1 |
| Metabolic detoxification            | ABC transporters | LDEC002775 | ABC subfamily g member 1-like | 2 |
| Metabolic detoxification            | ABC transporters | LDEC005530 | ABC subfamily g member 4-like | 1 |
| Metabolic detoxification            | ABC transporters | LDEC019090 | MULTIDRUG resistance-associated 1 | 1 |
| Metabolic detoxification            | ABC transporters | LDEC020530 | Multidrug resistance-associated 4-like | 1 |
| Metabolic detoxification            | ABC transporters | LDEC022533 | Multidrug resistance-associated protein 4 | 3 |
| Metabolic detoxification            | ABC transporters | LDEC005086 | Multidrug resistance-associated protein 4 | 2 |
| Metabolic detoxification            | ABC transporters | LDEC002518 | Probable multidrug resistance-associated protein lethal 03659 | 1 |
| Metabolic detoxification            | ABC transporters | LDEC012031 | Probable multidrug resistance-associated protein lethal 03659 | 5 |
| Metabolic detoxification            | CYPs        | LDEC018533 | Cytochrome P450 | 1 |
| Metabolic detoxification            | CYPs        | LDEC019188 | Cytochrome P450 | 1 |
| Metabolic detoxification            | CYPs        | LDEC019765 | Cytochrome P450 | 1 |
| Metabolic detoxification            | CYPs        | LDEC019766 | Cytochrome P450 | 1 |
| Metabolic detoxification            | CYPs        | LDEC015048 | Cytochrome P450 | 4 |
| Metabolic detoxification            | CYPs        | LDEC018119 | Cytochrome P450 | 5 |
| Metabolic detoxification            | CYPs        | LDEC005460 | Cytochrome P450 6BQ10 | 1 |
| Target sites                        | Voltage-dependent channels | LDEC009862 | Glutamate receptor 2-like isoform | 1 |
| Target sites                        | Voltage-dependent channels | LDEC000112 | Voltage-dependent calcium channel subunit alpha-2 delta-3 | 1 |
| Target sites                        | Voltage-dependent channels | LDEC021584 | Voltage-dependent calcium channel subunit alpha-1 | 1 |
| Target sites                        | Voltage-dependent channels | LDEC015955 | Voltage-dependent calcium channel type D subunit alpha-1-like protein | 1 |
| Target sites                        | Voltage-dependent channels | LDEC016101 | Nicotinic acetylcholine receptor a9 subunit | 2 |
| Target sites                        | Voltage-dependent channels | LDEC007707 | Nicotinic acetylcholine receptor subunit alpha4 | 1 |
| Growth factors                      | Cuticle proteins | LDEC003392 | Cuticle protein 19 | 2 |
| Growth factors                      | Cuticle proteins | LDEC014693 | Cuticular protein analogous to peritrophins 1-j precursor | 1 |
| Growth factors                      | Cuticle proteins | LDEC010803 | Larval cuticle protein 8-like | 3 |

**Note.—** The loading of each gene on a principal component is indicated (see supplementary file 1: fig. S9, Supplementary Material online).

*Among these genes, three (the CYP gene LDEC015048, the cuticle protein LDEC010803, and the voltage-dependent calcium channel gene LDEC000112) were found as candidate genes among field populations within Wisconsin (Crossley et al. 2017). The nicotinic acetylcholine receptor subunit alpha4 (LDEC007707) was also found as a candidate gene in a comparative genomic analysis of *Leptinotarsa* by Cohen et al. (2021).

Over 150 of the 224 genes (68.5%) identified in the haplotype-based test were also identified in the outlier test (supplementary file 1: fig. S26, Supplementary Material online), including 11 of the 16 candidate insecticide resistance genes. Seven of these were ABC transporters, and notably three were significant in all selection tests (the ABC subfamily C/multidrug associated gene LDEC002518, and two ABC subfamily G genes, LDEC002775 and LDEC005530). The other ABC genes included: the subfamily F gene LDEC004565, and three additional multidrug associated genes LDEC005089, LDEC003183, and LDEC002116. The remaining genes included one target site gene, the acetylcholine receptor β subunit (LDEC002850), one cuticle protein (LDEC003397), one transient receptor potential (TRP) gene (LDEC003216), and one odorant-binding gene (LDEC003898).

To determine whether parallel selection (positive selection on ancestral genomic variation in the same genes) drives repeated insecticide adaptation across the geographical range...
of CPB, we examined six pairs of neonicotinoid (imidacloprid) resistant and susceptible pest populations. $F_{ST}$ outliers for each pair, identified in 1-kb windows across the genome using a conservative threshold of the 99th percentile, show limited overlap (parallelism) among geographical regions (0.7 to 2.7% overlap; supplementary file 1: table S14, Supplementary Material online). A small proportion (4–6%) of these outlier regions intersect with candidate insecticide resistance genes (supplementary file 1: table S15, Supplementary Material online), with few genes (one to three) showing evidence of parallel selection in more than one region. These include a voltage-sensitive sodium channel gene (LDEC011942), the nicotinic acetylcholine receptor subunit alpha 6 (LDEC000437), five esterase genes (LDEC009237, LDEC011823, LDEC019607, LDEC024201, and LDEC014120), and three ABC transporters (LDEC008906, LDEC005089, and LDEC008907). Although few genes are commonly selected among geographical regions, gene set enrichment shows that insecticide resistance pathways are commonly targeted (see supplementary results in supplementary file 1, Supplementary Material online), as five of the six regions show enrichment of terms for metal ion binding, zinc ion binding, integral component of membrane, and intracellular anatomical structure.

Finally, to test for patterns of rapid gene regulatory evolution, we examined gene expression profiles for a subset of CPB populations sampled among geographical regions (supplementary file 1: table S16, Supplementary Material online). As most of the samples represent previously published results, we first assessed whether differences in experimental design influenced the expression of candidate insecticide resistance genes (see detailed results in supplementary file 1, Supplementary Material online). Based on these comparisons, we determined that regional population differences could be compared for adults from field populations irrespective of generation sampled, but lab reared larvae needed to be compared separately. For the larval comparison, we removed samples representing an insecticide induction treatment, focusing our analysis on constitutive differences in gene expression. These comparisons showed strong geographical differences in overall gene expression profiles despite local populations representing both neonicotinoid resistant and susceptible phenotypes (supplementary file 1: figs. S27 and S28, Supplementary Material online). Focusing on significantly differentially expressed candidate insecticide resistance genes, geographical regions showed divergent patterns of constitutive upregulation among populations (fig. 4 and supplementary file 1: fig. S29, Supplementary Material online; see gene list in supplementary file 2, Supplementary Material online). Seven differentially expressed candidate insecticide resistance genes were found among the adults and 84 among larvae (supplementary file 1: fig. S29 and table S17, Supplementary Material online), with only one esterase (LDEC019310) and one ABC transporter (LDEC004154) common to both data sets. A gene set enrichment analysis of the differentially expressed gene list in among-population comparisons of adults (supplementary file 1: fig. S30, Supplementary Material online) showed enrichment of terms associated with insecticide detoxification and/or stress, including heme and iron binding, oxidoreductase activity and dioxygenase activity, proteolysis, transport, defense response, and integral component of the membrane. Among larvae (supplementary file 1: fig. S31, Supplementary Material online), there was enrichment of terms associated with regulatory changes to gene networks underlying insecticide detoxification or stress, such as glutathione transferase activity.
hexachlorocyclohexane metabolism, oxidoreductase and monooxygenase activity, gap junction channel activity, proteolysis, substrate-specific transmembrane transporter activity, heme and iron ion binding, innate immune response, and integral component of the membrane. Although 26 transcription factors were significantly differentially expressed among the four larval populations (supplementary file 1: fig. S32 and table S18, Supplementary Material online), they are not known to be associated with detoxification pathways.

Although these results support gene regulatory evolution, they do not support rapid evolution of master regulatory genes, as different protein-coding genes, and in some cases different molecular pathways, are favored in each regional population. Gene set enrichment analysis of the overlapping set of significant genes in larvae and adult differential expression analysis gene sets (supplementary file 1: fig. S33, Supplementary Material online) showed shared regulatory changes in gene networks linked to insecticide detoxification and stress, such as oxidoreductase and dioxygenase activity, heme and iron ion binding, proteolysis and integral component of membrane. The shared enrichment of lipid metabolism might also be related to insecticide detoxification (through an interaction with oxidation-reduction or membrane-transport processes), rather than metabolism per se.

Finally, we examined a global intersection of gene ontology terms that were enriched in the range-wide genome scans and differential expression data sets (supplementary file 1: fig. S34, Supplementary Material online). Six gene ontology terms were identified, each of which has been related to insecticide resistance in prior studies: biological process terms include
oxidation–reduction process and proteolysis, whereas cellular components include integral component of the membrane, and molecular functions include oxidoreductase activity acting on paired donors with incorporation or reduction of molecular oxygen, heme binding, and iron ion binding.

**Discussion**

Population genomics is increasingly providing insight into the evolutionary mechanisms that give rise to super-pests and holds promise for improving pest management practices (Péllissié et al. 2018). Our study provides the first comprehensive genomic and geographical assessment of genome-wide patterns of genetic variation for a super-pest in the center of its origin. Whole-genome variation demonstrates that the CPB is geographically structured, including among pest populations, and corroborates evidence from microsatellite and mitochondrial markers that CPB pest populations are most closely related to populations in the Great Plains instead of Mexico (Edgerton 1861; Casagrande 1987; Izzo et al. 2018). In evaluating the competing (but not mutually exclusive) mechanisms of rapid evolution, our data suggest that selection occurs independently across growing regions, and even among different populations within regions, with statistical tests providing a consistent pattern of repeated evolution at many candidate insecticide resistance genes. The observed pattern does not involve parallel selection on ancestral variation in the case of recent adaptation to imidacloprid, but instead the recruitment of different genes underlying the similar molecular resistance mechanisms across geographical regions. Therefore, our results suggest that insecticide resistance in CPB readily evolves from an underlying polygenic trait architecture. Below we discuss in turn the evidence that supports insecticide resistance arising from standing genetic variation, in contrast to hypotheses of selection on de novo mutation and rapid regulatory evolution. We close by discussing the pest management consequences of these modes of rapid evolution.

**Evidence for Repeated Selection on Standing Genetic Variation**

Population clustering analyses demonstrate that regional CPB pest populations (western US, eastern US, multiple lineages in the Midwestern US, and Europe) are genetically distinct, with $D$-statistics suggesting limited ongoing gene flow. This alone supports previous observations that insecticide resistance evolves locally among CPB populations (Alyokhin et al. 2008; Szendrei et al. 2012; Crossley et al. 2017; Dively et al. 2020). However, genome scans, using both outlier-based and environmental association-based methods, as well as haplotype-based tests, provide clear signatures of adaptation occurring repeatedly across different populations of CPB. Focusing on candidate genes of insecticide resistance, the considerable overlap among pathways, but not specific genes, suggests that different genes are selected in different populations (see PC loadings in table 2 and parallel selection test in supplementary file 1: table S13, Supplementary Material online). Similarly, candidate insecticide resistance genes are constitutively upregulated, but in distinct patterns in different populations (fig. 4). Selected genes also encompass multiple resistance mechanisms, including metabolic detoxification, target site resistance, and cuticle proteins, which is broadly supported by gene set enrichment analyses across multiple tests (supplementary file 1: fig. S34, Supplementary Material online).

An important caveat to these results is that tests of selection are prone to false positives (Mallick et al. 2009). For example, hapFLK tends to show a high false-positive rate under population models with continuous migration (Fariello et al. 2013; Vatsiou et al. 2016), whereas it has performed well in species (e.g., domestic livestock and dogs) that experience strong bottlenecks (Boitard et al. 2016; Schlamp et al. 2016). We did not integrate simulations to assess the false-positive rate of our genome scans due to the complex population structure and demography underlying our study design, although this should be a major goal in future work. However, we did employ conservative testing thresholds to mitigate false positives, as well as multiple independent tests. Although chance false positives might contribute to our observed patterns, it is highly unlikely that so many would occur in candidate insecticide resistance genes. Furthermore, these results are consistent with previous genetic studies that have documented many of these same genes (see table 2 and supplementary file 2, Supplementary Material online) or mechanisms in CPB resistance (Hawthorne 2003; Alyokhin et al. 2008; Rinkevich et al. 2012; Clements, Schoville, Peterson, Lan, et al. 2016; Gaddelapati et al. 2018). We also identify compelling new candidates, such as the multiple ABC transporters, as well as acetylcholine receptor β subunit (LDEC002850), which are likely linked to neonicotinoid resistance phenotypes (Tan et al. 2008; Qu et al. 2016).

Although our analysis focuses on pesticide resistance candidates, we note that other interesting genes emerged from our study. In particular, an octopamine receptor (LDEC006841), which was identified as a gene associated with pest behavior in a comparative genomics analysis of CPB (Cohen et al. 2021), emerged as a significant target in contrasts of nonpest and pest populations in both LFMM and PCAdapt. The observed patterns of repeated local adaptation are most consistent with evolution from standing genetic variation. Using a more sophisticated haplotype-based method (focusing on ~21% of the genome), we found 24 highly significant ($P < 0.0001$) haplotype blocks suggestive of selective sweeps. Only one selected region exceeded 4 kb in length (35 kb), whereas the remaining regions averaged 3.1 kb in length. These shorter sweep lengths are similar to those found in at least one other insect pest, Spodoptera frugiperda (on an average, sweeps are 4.1 kb: Nam et al. 2020), but are much smaller than the strong selective sweeps found in other prominent cases of insecticide resistance (Schlenke and Begun 2004; Calla et al. 2020; Weedall et al. 2020). Sweep lengths typically scale inversely with the rate of recombination as a product of increasing effective population size, so we might have expected sweep lengths similar to those found in insects with large population size (typically >10 kb in Drosophila melanogaster) (Garud et al. 2021). Instead, nearly
levels of standing variation at a regional level (Crossley et al. 2017) and the polygenic trait architecture of insecticide resistance. Traditionally, the host-shift to potato, coupled with agroecosystem invasion over a broad geographical scale, has been presumed to reduce genetic variation in CPB (Zehnder et al. 1992; Lu and Lazell 1996). Instead, we find that CPB pest populations, despite having recently invaded agroecosystems, exhibit no reductions in nucleotide diversity and only modest declines in private alleles that most likely reflects their spatial expansion history. Dense population sampling of CPB at a landscape scale in the Midwestern US also supports high levels of standing variation at a regional level (Crossley et al. 2017).

Early population genetic research on insecticide resistance by J.F. Crow suggested that resistance evolution arises from polygenic standing variation (Crow 1957). The quantitative nature of most phenotypic traits has been documented in many cases of adaptation (Dermauw et al. 2013; Faucon et al. 2015; Cheng et al. 2017; Barghi et al. 2020), and since large-effect mutations often require compensatory fitness changes that are unlikely to evolve quickly enough to prevent such mutations from being removed by antagonistic selection, the most likely outcome of evolution in theoretical models is polygenic adaptation (Wellenreuther and Hansson 2016). Quantitative variation in insecticide tolerance, among populations and individuals within populations, is frequently observed, even under controlled laboratory conditions (Tabashnik and Cushing 1989; Kobiela and Snell-Rood 2020). The action of multiple genes and their regulatory elements can be difficult to detect using population genomics methods, as only modest changes in allele frequencies are expected under soft or partial selective sweeps (Teshima et al. 2006; Messer et al. 2016; Stephan 2016). Yet, our results suggest that insecticide resistance in CPB has a polygenic trait architecture, with many different genes (but shared molecular pathways) rapidly evolving across the geographical range. This is in broad support with the emerging view that polygenic architecture is now common in contemporary cases of insecticide resistance (Schmidt et al. 2017; Kreiner et al. 2018; Hawkins et al. 2019). Unfortunately, in our results, it is difficult to determine whether polygenic evolution occurs as resistance evolves to a single insecticide, or if instead the pattern of polygenic architecture merely reflects a legacy of evolved resistance to many different insecticides. In other words, it is possible that individual episodes of selection to more than 50 compounds contribute to a signature of polygenic adaptation in present-day genetic patterns (Crossley et al. 2017). The main line of evidence supporting polygenic adaptation at least some insecticides is that we find many genes under selection ($\gg$50, much more than expected for monogenic evolution to $\sim$50 compounds) and haplotype signatures of selected alleles are consistent with a pattern of soft sweeps that are expected under polygenic adaptation (Barghi et al. 2020). Future efforts to test these different explanations could leverage temporal sampling and quantitative modeling to investigate the age of alleles and modes of selection underlying the complex genetic architecture of insecticide resistance.

Selection on De Novo Mutation: Is CPB Mutation Limited?

Selection on novel mutations could lead to repeated evolution of resistance in pests with intrinsically large population sizes (much greater than $10^7$) that are not considered mutation-limited (Barton 2010; Cutter et al. 2013). Analyses of nucleotide diversity in protein-coding genes is known to range widely in animal species and is most strongly correlated with reproductive strategy and effective population size, with highly fecund ($r$-selected) species like CPB having the greatest diversity (Romiguier et al. 2014). Schoville et al. (2018) found evidence for a high rate of polymorphism in protein-coding regions of CPB (nucleotide diversity, $\pi$, was $\sim0.01$), suggesting a high level of standing genetic diversity and large effective size. Furthermore, CPB shows a larger effective size, higher rate of positive selection, and greater levels of standing variation compared with other species in the genus Leptinotarsa (Cohen et al. 2021). In this study, irrespective of pest or nonpest status across different geographical regions, we identify an average genome-wide nucleotide diversity of 0.005 within CPB. Although this rate of nucleotide diversity remains high relative to most vertebrates (median 0.0025), it is not exceptional among arthropods (median 0.0125 all sites, 0.00204 synonymous sites; Leffler et al. 2012). Despite its super-pest status, CPB nucleotide diversity falls within the range of empirical estimates for insect species (0.0023–0.0288). Among Coleoptera, nucleotide diversity found in the bark beetles Dendroctonus ponderosae and Dendroctonus brevicornis (0.0023 and 0.008, respectively; Keeling et al. 2013; Bracewell et al. 2018) and the horned scarab beetle Onthophagus taurus (0.0056, Choi et al. 2010) is similar to CPB. Lepidopteran pest genomes are more polymorphic than CPB; five species of Helicoverpa range in nucleotide diversity from 0.004 to 0.010 in autosomal regions, with the super-pest H. armigera as the most polymorphic (Anderson et al. 2018), whereas two sympatric strains of the pest species Spodoptera frugiperda range from 0.043 to 0.044 (Kiwoong et al. 2020) and the closely related pest Spodoptera litura has nucleotide diversity as high as 0.016 (Cheng et al. 2017). These results suggest that CPB is not exceptional in terms of the potential for rapid evolution. However, the amount of nucleotide diversity is not a perfect proxy to measure mutation limitation. There is an upper limit on nucleotide diversity in species with large effective population sizes (e.g., Drosophila melanogaster), as nucleotide variation at neutral sites is removed as a result of selection on nearby linked sites (Corbett-Detig et al. 2015). In fact, patterns of reduced variation in species with large effective population sizes might reflect selection from recurrent adaptive mutation (Karasov et al. 2004).
resistant populations of the beetle population of CPB (Gaddelapati et al. 2018) and deltamethrin (Wilding 2018), including the highly resistant Long Island to regulate multiple detoxification loci in many insects (Ingham et al. 2017; Peng et al. 2017; Smith et al. 2018; Hu et al. 2019). Known transacting transcription factors involved in the xenobiotic detoxification pathway include CncC, Maf-S, AhR, ARNT, and Met. CncC forms a heterodimer with Maf-S to regulate multiple detoxification loci in many insects (Wilding 2018), including the highly resistant Long Island population of CPB (Gaddelapati et al. 2018) and deltamethrin resistant populations of the beetle Tribolium castaneum (Kalsi and Palli 2015). However, in comparing gene expression profiles of CPB populations throughout several growing regions, we found that expression levels of insecticide detoxification genes vary across populations, suggesting that varied transcriptional patterns are most-likely achieved through cis-regulatory evolution (Wittkopp and Kalay 2011; Verta and Jones 2019). In a similar analysis, comparison of transcriptional profiles of Anopheles gambiae across Africa revealed the recruitment of many population-specific candidate insecticide resistance genes (Ingham et al. 2018), suggesting cis-regulatory evolution of these pathways may be common among insect pests.

Our data do not support the role of a single master-regulatory switch driving insecticide resistance, as we see no evidence for differential expression of key transacting transcription factors despite comparing insecticide resistant and susceptible populations. Furthermore, the role of different genes in metabolic resistance in CPB has been demonstrated by RNAi experiments where knockdown of different upregulated CYP genes restores susceptibility in different pesticide resistant CPB populations (Clements, Schoville, Peterson, Huseth, et al. 2016; Kalsi and Palli 2017). Altogether, our results suggest that a simple upstream shift in Cap-n-collar expression is not sufficient to explain all cases of metabolic resistance in CPB and that, instead, additional cis-regulatory changes are required to account for the heterogeneity and diversity of resistance pattern among populations (Wilding 2018). This is consistent with widespread evidence that cis-regulatory evolution is more common in adaptive evolution, whereas transacting gene regulation is typically constrained by strong stabilizing selection (Wray 2007; Signor and Nuzhdin 2018).

Implications for Pest Management and Novel Control Tacticts

Current resistance management strategies assume that the evolution of resistance is a rare event, caused by simple (single gene) mutations (Roush 1989), thereby ignoring the importance of alternative mechanisms involving multiple loci (Denholm and Rowland 1992). Resistance management models also assume that resistance involves fitness tradeoffs (Tabashnik 1990), a rationale underlying high dose/refuge strategies where gene flow from susceptible “refuge” populations into insecticide-treated fields delays resistance evolution (Sudo et al. 2018). However, it is increasingly difficult to understand the rate of pesticide resistance evolution using conventional models of single, large-effect mutations in conferring a resistance phenotype (Haridas and Tenhumberg 2018). Rates of insecticide resistance evolution in CPB are among the highest observed in agricultural pests (Brevik et al. 2018) and often lead to failure unless implemented in an integrated pest management framework (Alyokhin et al. 2015). From our results, evolution of multiple resistance loci from standing variation appears to best explain this pattern, although we note that some alternative mechanisms were not investigated and could contribute to rapid evolutionary change. Notably, recent work in CPB has shown that changes in DNA methylation patterns might drive trans-generational epigenetic mechanisms of regulatory evolution that lead to pesticide resistance (Brevik et al. 2021).

As society faces the challenge of global food security, there is a prevailing view that insect pests have won the arms race involving conventional chemical pesticide control (Gould et al. 2018). Novel chemical modes of action are needed to avoid target-site resistance, yet there is a high cost to such efforts, both in terms of development costs and environmental impacts. In addition, the widespread emergence of cross-resistance, especially through metabolic detoxification, suggests novel modes of action may have limited efficacy and durability (Vontas et al. 2020). Although gene drives have been raised as promising novel control tactics (Scott et al. 2018; Gui et al. 2020), most recent work in CPB has focused on gene-targeted insecticides via the RNAi pathway (Palli et al. 2014). Gene knock-down via RNAi could allow for highly effective, species-specific management if multiple genes are targeted simultaneously, and such products are currently under development (Zhu and Palli 2019).

How do genome-wide population genetic data shed light on RNAi implementation? Drawing on the mechanisms of evolution described in this paper, where standing variation is substantial, the likelihood of resistance evolution to RNAi might be high unless population-specific approaches are developed. One target site mutagenesis experiment in CPB has shown that a mismatch rate of 3% or less still allows for effective gene target suppression (He et al. 2020), but clearly some CPB target genes would be problematic due to high
levels of diversity. Additionally, alternative pathways of RNAi resistance might emerge. For example, experiments with cell lines of CPB have shown that resistance evolving from mutations altering the uptake and transport of dsRNA (Yoon et al. 2018). CPB is known to utilize both sid-1 transmembrane channel-mediated uptake and clathrin-mediated endocytosis in processing dsRNA (Cappelle et al. 2016), suggesting that there are multiple targets for resistance evolution in the RNAi pathway. Comparison of dsRNA efficacy among European CPB populations also suggested variation in the RNAi pathway itself (involving the multiple homologs of dicer, argonaut, and staufen) was more likely to evolve than at target site loci (Mehlhorn et al. 2020). However, knockdown experiments in other Coleopteran pests have shown that loss of function of genes in the RNAi pathway impair development and reduce reproductive fitness (Davis-Vogel et al. 2018), thus the inherent trade-off in resistance may be too great. One interesting RNAi resistance pathway involves selection on gut nuclease activity that alters the sensitivity of CPB to RNAi (Spit et al. 2017). In other insects, such as lepidopteran pests (Guan et al. 2018), a single nuclease is responsible for dsRNA tolerance. Though clearly RNAi provides a novel mode of action for controlling CPB pests, the propensity to draw on reservoirs of standing genetic variation to rapidly evolve suggests that multiple mechanisms of resistance are likely to occur.

**Conclusion**

Understanding the molecular mechanisms underlying pesticide adaptation has become increasingly important because of the widespread occurrence of the “pesticide treadmill” phenomenon in agricultural pests (Luck et al. 1977), wherein the repeated and escalating use of pesticides, and the search for new chemistries (Alyokhin et al. 2015), is required to keep pace with pest evolution. We provide strong evidence that repeated evolution from standing variation could explain how insects rapidly overcome multiple classes of pesticides, and that resistance phenotypes can be underpinned by polygenic trait architectures. Although the importance of polygenic adaptation remains mostly theoretical in the insecticide resistance literature and has proven challenging to identify in empirical case studies (Péllissié et al. 2018), polygenic adaptation has also been broadly implicated in the evolution of herbicide resistance in agricultural weeds (Kreiner et al. 2018) and antibiotic resistance in bacterial pathogens (Nyerges et al. 2018). From an applied perspective, we provide evidence that CPB evolves insecticide resistance repeatedly across agricultural regions, and often through the same genetic pathways. An important future goal will be to understand how rapidly resistance phenotypes spread among local pest populations, in order to refine integrated pest management practices to maintain the efficacy and sustainability of novel control techniques.

**Materials and Methods**

**Study Design and Aim**

We collected a geographically dispersed set of 88 samples, selected to maximize information about genomic differentiation across the range of CPB (supplementary file 1: fig. S1 and table S1, Supplementary Material online). An additional ten samples comprising nine species of Leptinotarsa were also collected for relevant information on outgroup variation and possible sources of hybridization. Within the 88 CPB samples, we sampled six geographically proximate pairs of neonicotinoid resistant (R) and susceptible (S) populations, five beetles per R/S site: Maine (R and S), Vermont (S), New York (R and New Jersey (S), Maryland (R and S), Michigan (R and S), Wisconsin (R and S), Oregon (R and S). The resistance status of beetles was ascertained by topical exposure to a neonicotinoid insecticide (imidacloprid) or from published records at those sites (see supplementary file 1, Supplementary Material online, for detailed methods). Notably, published data (Dively et al. 2020) suggest populations in Maine, New York, Maryland, Michigan, and Wisconsin are resistant to multiple classes of insecticides (neonicotinoid, ryanoid, semicarbazone, and spinosyn modes of action). However, all pest populations have potentially evolved resistance to other insecticides (i.e., organochlorides, organophosphates, and pyrethroids), as insecticide use was widespread starting in the late 1940s and cases of resistance have been reported globally among CPB populations (Alyokhin et al. 2008). As each diploid individual represents $N = 2$ genomes, our sample size exceeds the requirements for most population genomic tests and allows for accurate estimation of frequencies for all but the most rare (and therefore, presumably, less important) alleles in key potato growing regions of the United States (Nevado et al. 2014).

**Genomic Resequencing, Quality Control, and Variant Calling**

High-quality genomic DNA was isolated from adult beetle thoracic muscle tissue using DNeasy Blood & Tissue kits (Qiagen) and then submitted to the University of Wisconsin-Madison Biotechnology Center. Libraries were sequenced using paired-end, 125-bp sequencing on a HiSeq2500 sequencer (see supplementary file 1, Supplementary Material online, for detailed methods). We predetermined sequencing effort to yield $>6 \times$ average coverage for each of our CPB genomes, a quantity sufficient to identify SNPs with reasonable accuracy (Li et al. 2009).

Each sample was demultiplexed prior to downstream analysis, and we followed GATK’s “Best Practices” guidelines (https://software.broadinstitute.org/gatk/best-practices/). Using the L. decemlineata reference genome v1.0 (GCA_000500325.1; Schoville et al. 2018), we aligned demultiplexed reads using BWA v0.7.101 (Li and Durbin 2009), and converted sam files to bam format using SAMTOOLS v1.3.12 (Li and Durbin 2009). We generated one .ubam file (i.e., unmapped .bam file) per forward–reverse pair of the .fastq raw reads using FastToSam and then marked Illumina adapters with MarkIllluminaAdapters, both functions available from PICARD v2.2.4 (https://github.com/broadinstitute/picard). We then reverted .bam files to .fastq format with PICARD’S SamToFastq, aligned the new .fastq files to the reference genome with the BWA-mem algorithm and merged all alignments into one .bam file per sample with PICARD’s MergeBamAlignment tool. We marked PCR and

doi:10.1093/molbev/msac016
optical duplicates using PICARD's MarkDuplicates tool but, for some of our samples, we marked duplicates first at the lane level (i.e., per replicate), then at the sample level (merging duplicates into a unique bam output). Finally, we realigned reads around insertions and deletions with GATK's RealignerTargetCreator and IndelRealigner tools. In order to assess the quality of our bam files, we used GATK's Flagstat and DepthOfCoverage tools. Among our 88 CPB samples, three samples (two susceptible samples from Oregon: CPBWGS_59 and CPBWGS_63, and one susceptible sample from Vermont: CPBWGS_93) had few successfully mapped reads and were removed (supplementary file 1: fig S2 and table S1, Supplementary Material online).

Genotyping was split into two steps: per-individual variant calling, followed by joint genotyping. Variant calling was conducted with GATK’s HaplotypeCaller tool, which generates a likelihood score for all reference sites (–ERC GVCF option), including nonvariant sites. Two different joint genotyping procedures were performed: one excluding non-CPB samples (“CPB” data set; N = 85), and one including non-CPB samples, but keeping only one susceptible and one resistant sample (chosen at random) for populations from Oregon, Wisconsin, Michigan, Maryland, New-Jersey/New-York, and Vermont/Maine (“Leptinotarsa” data set; N = 50; supplementary file 1: table S2, Supplementary Material online). For joint genotyping, we employed variant quality score recalibration (VQSR) using a training data set. VQSR is based on applying machine learning algorithms and clustering methods to examine the overlap of the raw call set and a training data set. It is composed of two steps: 1) ApplyRecalibration describes the multidimensional annotation profile of variants and calculates (for each variant in both data sets) a new, well-calibrated quality score called VQSLOD (for variant quality score log-odds). 2) ApplyRecalibration uses VQSLOD to apply a new cutoff to retain only high likelihood variants from the call set, based on a proportion of the variants in the training set that are present in the call set (e.g., 99.9% to enhance sensitivity or 90% to enhance specificity). As this approach requires a well-validated, independent data set to be used as a training set, we used 41,454 SNPs generated from a published genotyping-by-sequencing experiment (Crossley et al. 2017). These data represent 188 samples from 24 Midwestern populations (supplementary file 1: fig. S35, Supplementary Material online), which were hard filtered for depth of coverage $\geq 10x$, polymorphism in at least 30% of the individuals of each population, minor allele frequency (MAF) $\geq 5\%$, and less than 20% missing genotypes across all individuals. We calculated VQSLODs based on the following annotations: QD, MQ, MQRankSum, ReadPosRankSum, FS, SOR, DP, and InbreedingCoeff. The recalibrated score provides a continuous estimate for the probability of each variant, which can then be partitioned into quality tranches. Tranchce plots for the “CPB” and “Leptinotarsa” data sets are based on a 90% threshold that maximizes specificity over sensitivity (supplementary file 1: fig. S36, Supplementary Material online). Finally, we used GATK’s VariantsToTable tool to assess the quality of our inferred SNP data set. We plotted the improvement in the distribution of QualByDepth (QD) following the VQSR procedure for the “CPB” data set (supplementary file 1: fig. S37, Supplementary Material online) using the ggplot2 package in R v3.6 (Wickham 2016; R Core Team 2019). After removing sites representing mitochondrial DNA (1,756 total variant sites), our data set contained 47,969,460 SNPs in the “CPB” data set and 69,680,768 in the “Leptinotarsa” data set. Some analyses, like demographic reconstruction, require analyses with neutrally evolving loci. To mitigate the effect of nonneutral loci in these analyses, we created an “intergenic CPB data set” from our “CPB” data set, by considering only SNPs located outside of known genes from the L. decemlineata Official Gene Set (OGS) v1.1 (Schoville et al. 2018).

Genomic Diversity
To estimate the genetic diversity of populations, we used the “CPB” data set (i.e., not including Mexican samples), removed multiallelic SNPs and those SNPs with a MAF < 5% (using GATK’s SelectVariants tool; final SNP number = 17,599,906). For these analyses, we examine paired populations (susceptible/resistant) and grouped individuals from the US Plains region (Colorado, Nebraska, Kansas, Missouri, New Mexico, and Texas). We estimated genome-wide nucleotide diversity ($\pi$) using a 10-kb sliding window with VCFtools v0.1.15 (Danecek et al. 2011). We also estimated heterozygosity by calculating the inbreeding coefficient ($F$) for each individual, using the method of moments estimator in VCFtools. Individual estimates were then averaged per population, keeping susceptible and resistant individuals separate. Finally, we used the “singletons” function in VCFtools to calculate the number of singletons and private doubletons for each sample.

Evolutionary Divergence
In order to reconstruct the evolutionary origins of CPB populations, we conducted a phylogenomic analysis of the “Leptinotarsa” data set. This data set comprised 48 samples, including ten L. decemlineata samples from Mexico, 28 L. decemlineata from the United States and Europe, and ten samples of closely related Leptinotarsa species. Since the data set was quite large, we used only the first 100 scaffolds comprising 16,519,065 SNPs. We used SNPhylo v.20160204 (Lee et al. 2014) to construct a phylogeny, after pruning the SNPs for linkage disequilibrium (LD). In order to ensure the relative independence of the SNPs used in the analysis, we tested several values of the LD threshold parameter and selected 0.5 for downstream analyses, which resulted in 35,838 SNPs. The SNP data were concatenated and a maximum likelihood phylogeny was estimated using DNAML in PHYLIP v3.6 (Felsenstein 2005; Shimada and Nishida 2017). Support values for nodes in the tree were determined by bootstrap resampling 100 times.

To estimate population structure, we examined both the “CPB” data set and the “interenic CPB” data set, but the results were biologically consistent. We used three approaches: classical $F_{ST}$ estimates, principal component
analysis using PCAdapt (Duforet-Frebourg et al. 2016; Luu et al. 2017), and ancestry analysis with sNMF (Frichot et al. 2014). For the $F_{ST}$ analyses, we first assessed pairwise genetic divergence of Northern and Southern Mexican populations to the Plains. We then focused on paired populations (susceptible/resistant) and Plains individuals (Colorado, Nebraska, Kansas, Missouri, New Mexico, and Texas) for additional comparisons. We measured population-specific $F_{ST}$ using the “fs.dose” function in the hierfstat package in R (Goudet 2005), measuring the genetic divergence between each focal population and all other populations. Some paired populations were quite similar, so we then grouped those and measured mean weighted pairwise $F_{ST}$ in 1-kb sliding windows using Weir and Cockerham’s estimator in VCFtools (Daneczek et al. 2011). For the principal component analysis, we assessed the percentage of variance explained by the first 20 principal components in a scree plot. We followed Cattell’s rule (Cattell 1966), where eigenvectors are retained up to the point of inflection in the scree plot, above which additional terms provide diminishing returns in terms of explained variance. For sNMF, we inferred individual patterns of ancestry by estimating ancestral population allele frequencies and admixture coefficients using the R package LEA (Frichot and François 2015). We first converted our vcf files to PLINK’s ped format using VCFtools, then to the geno file format using LEA’s ped2geno tool. We implemented 10 runs per $k$ value and combined the different runs with CLUMPACK (http://clumpaktau.ac.il/). We plotted cross-entropy values to assess the number of $k$ values.

Finally, to test for possible admixture among ancestral populations during the invasion of CPB into agroecosystems, we assessed evidence of gene flow: 1) from Mexican CPB populations and 2) from other Leptinotarsa species, into the pest lineage. We used Dsuite (Malinsky et al. 2021) to calculate the genome-wide D-statistic ($D$) on allele frequency data, estimating the strength of introgression based on the ABBA/BABA test (Green et al. 2010). $D$ ranges from zero (no introgression) to one (complete introgression), and is calculated using a set of three focal populations or taxa (P1, P2, P3) with one additional outgroup. For the first test (1), we only considered populations with sample sizes $\geq 5$, grouping susceptible and resistant samples from Vermont/Maine, Maryland, Wisconsin, and Oregon as $F_{ST}$ between these subpopulations was negligible (supplementary file 1: table S6, Supplementary Material online), grouping samples from the “Plains” region ($N = 7$), and grouping all CPB samples from Mexico ($N = 10$). We tested every possible focal trio of populations, retaining the lowest $D$-statistic for every given trio ($D_{min}$, a conservative estimate of $D$). We assessed whether $D$ was significantly different from zero by calculating a $P$ value based on jackknifing. We analyzed the complete biallelic SNP data set with the two L. juncta samples as outgroups (the results were almost identical when using L. undecemlineata instead; data not shown). For the second test (2), we included the nine other Leptinotarsa species using the same data set from the SNPhylo analysis, comprising the 100 first scaffolds and 16,519,065 SNPs. The data set contained one susceptible and one resistant sample for each of the six paired populations. In order to analyze a balanced data set, we limited the “Plains” population to the samples from Colorado. We also created two Mexican populations, representing the two distinct Mexican clades recovered in our phylogeny: “Mexico_City” (containing two samples) and “Mexico_South” (containing the sample from Oaxaca and the one from Guerrero; supplementary file 1: table S1, Supplementary Material online). We used L. lineolata as the outgroup, as it was recovered as the most basal and distantly related taxon to the US CPB clade.

Demographic Analysis

To reconstruct the population history of CPB, we used the CPB intergenic SNP data set and employed two coalescent approaches: the Stairway plot method (Liu and Fu 2015) and SMC++ (Terhorst et al. 2017). Demographic reconstructions rely on an accurate estimate of the mutation rate. Most estimates of nuclear mutation rate in insects fall into the range of $2 \times 10^{-9}$ to $7 \times 10^{-9}$ substitutions per site per generation (Keightley et al. 2015; Allo et al. 2017). As there is no genome-wide mutation rate estimate for CPB or related beetles, we chose to use a mutation rate of $2.1 \times 10^{-9}$, estimated recently in the nonbiting midge (Oppold and Pfenninger 2017). We set the generation time of 0.5 per year (i.e., two generations per year) for all our samples. The Stairway plot approach relies on the calculation of the expected composite likelihood of a given site frequency spectrum (SFS), which reduces the computational burden of inferring population parameters. It is also suitable for estimating recent population histories with low-coverage genomic data. We analyzed the resistant and susceptible paired populations, both separately and pooled together, and also considered pooled sample from a geographically cohesive groupings that had some evidence of genetic distinctiveness: the Plains (Colorado, Nebraska, Kansas, Missouri, New Mexico, and Texas), East (Florida, Tennessee, North Carolina, Virginia, Kentucky, and Ohio), and Europe (Italy, Russia). For each population, we estimated the folded SFS in dadi (Gutenkunst et al. 2009), calculated for a genome length of 678 Mb (Schoville et al. 2018), and used 200 bootstraps to assess confidence intervals. We then conducted the Stairway plot analysis with default parameters and plotted the estimated median (and 95% confidence intervals) effective population size ($Ne$) through time.

We chose SMC++ (Terhorst et al. 2017) as an alternative approach, as it incorporates estimates of recombination and LD in an SFS framework. Even though this method is relatively computationally efficient, long run times prevented us from using the entire CPB data set. Due to the fragmented nature of the genome, we analyzed the longest 95 genomic scaffolds of the CPB data set, including all intergenic and nonintergenic biallelic SNPs. This encompasses approximately 21% of the genome ($\sim 140$ Mb out of $\sim 670$ Mb, mean length of 1.9 Mb, and all $\geq 1$ Mb) and contains 6,669,259 SNPs. This subset of highly contiguous reference genomic data ensures we can accurately infer demography using the sequential Markov coalescent (Nadachowska-Brzyska et al. 2016), although we note that analyses considering all $\geq 10$-kb scaffolds (covering
Selection Analyses
We used four different approaches to study genomic signatures of selection: outlier detection with PCAdapt (Duforet-Frebourg et al. 2016; Luu et al. 2017), genome–environment association with LFMM (Frichot et al. 2013), haplotype-based tests using hapFLK (Fariello et al. 2013), and a $F_{ST}$ outlier approach for paired susceptible and resistant populations. Simulation studies have shown that PCAdapt, LFMM, and hapFLK retain power and have a controlled false-positive rate under scenarios of recent range expansion (Frichot et al. 2015; Luu et al. 2017), which we expect based on the history of rapid geographical expansion of CPB (Hsiao 1985). To detect outlier SNPs using PCAdapt, Mahalanobis distances were transformed into $P$ values and then the FDR was controlled by transforming the $P$ values into $q$ values and considering an FDR of 0.01% ($\alpha = 0.0001$). We initially filtered SNPs using a MAF of 0.05 and a conservative setting of $K = 10$, but the number of SNPs suggested a high rate of false positives (see supplementary file 1, Supplementary Material online). We therefore refined our filtering steps using LD clumping (choosing a window size of 500 SNPs and a squared-correlation coefficient threshold of 0.2). We examined the screeplot of the principle components and, following Cattell’s rule, selected $K = 6$ as the optimal clustering level. Finally, we adjusted the data set by setting a more conservative MAF setting of 0.1. Using these settings, we ran PCAdapt with and without the New Jersey population to test for any bias due to inclusion of this inbred population.

As an alternative genome scan approach, we employed a genome-environmental association method using latent factor mixed models (LFMM; Frichot et al. 2013). This test identifies SNP allele frequencies that are significantly associated with environmental predictor variables, whereas simultaneously modeling the confounding effect of population structure as latent factors. To account for the population structure observed in our data, we modeled $k = 6$ latent factors, as suggested by our PCAdapt and sNMF results. We adjusted $P$ values by an empirically determined genomic inflation factor, while controlling the false discovery rate at 0.01%. We explored five different environmental variables: elevation, precipitation, minimum temperature in the coldest month, and potato land cover. We reasoned that genes containing SNPs associated with climate variables could be related to $L. decemlineata$ adaptation to northern climates during range expansion, whereas associations with potato land cover might reveal genes responding to selective pressures faced in potato agroecosystems, such as novel host plants, natural enemy pressure and insecticide exposure. We obtained historic, county-level potato land cover data (between 1850 and 2012), as detailed in Crossley et al. (2021). For latitude, elevation, precipitation, and minimum temperature in the coldest month, we obtained the data from the PRISM climate group (http://prism.oregonstate.edu). For each environmental variable, we took the average value within a 75 km radius around each sampling site, using functions available in the rgdal and raster packages in R (Bivand et al. 2014; Hijmans et al. 2017). For potato land cover, we summarized the average proportion of area planted with potato within a 75 km radius of each sample site (2018).

In order to integrate information from linked SNPs in tests of selection, we used the haplotype frequency-based method hapFLK (Fariello et al. 2013). This approach models haplotype frequency changes in a population tree, with population branch lengths representing the amount of genetic drift relative to a shared ancestral node. The population tree is estimated using genome-wide haplotype data, and deviation from the background tree topology is tested to identify candidates of selection. The hapFLK method has been shown to be relatively robust to confounding effects of population structure and variable population size (Schlamp et al. 2016; Vatsiou et al. 2016). It also allows selection events to be pinpointed to specific branches of the population tree, so that for each identified signature of selection, a local tree is re-estimated using significant haplotypes while being constrained by the overall topology of the population tree. Statistical significance is computed for the difference between the branch lengths estimated from the focal region and from the global tree. We analyzed the first (longest) 95 genomic scaffolds of the CPB data set (representing ~21% of the genome), including all biallelic SNPs. From a vcf file produced with GATK’s SelectVariants function, we produced a ped file and associated map file with VCFtools’ plink function. In order to use the multipoint LD model, hapFLK requires the number of haplotype clusters ($K$) to be specified and a population tree. We compared hapFLK results for different $K$ values, ultimately selecting $K = 20$ to minimize imputation errors (see supplementary file 1, Supplementary Material online, for detailed methods; supplementary file 1: fig. S38, Supplementary Material online). The population tree was estimated from a kinship matrix (supplementary file 1: fig. S39, Supplementary Material online). We standardized hapFLK values and computed corresponding $P$ values from a standard normal distribution, examining results at three nominal levels ($\alpha = 0.01, 0.001, and 0.0001$; supplementary file 1: fig. S40, Supplementary Material online). We grouped significant SNPs into selected regions, where each region was separated by at least 1 kb up- and downstream.

Finally, in order to determine whether parallel selection (positive selection acting on ancestral variation in the same genes) drives insecticide adaptation across the geographical range of pest populations, we first identified $F_{ST}$ outliers in pairs of imidacloprid resistant and susceptible pest populations: Oregon, Wisconsin, Michigan, Maryland, New York (with New Jersey), and Maine (with Vermont). Pairwise $F_{ST}$ was measured in 1-kb sliding windows across the genome, using a conservative threshold of the 99th percentile (the 0.995 quantile in a one-tailed test) to identify outlier regions. To identify parallel selection, we examined whether these outlier regions overlapped with candidate insecticide genes and exhibited geographical overlap among the six pairs.
Candidate Genes and Gene Network Analysis

For each selection test, we obtained functional information for candidate SNPs using manual annotation of the OGS supplemented by Blast2GO annotations (Conesa et al. 2005), which have been previously published (Crossley et al. 2017; Schoville et al. 2018). To develop a list of candidate insecticide resistance genes, Crossley et al. (2017) identified 664 genes associated with processes or functions potentially linked to known mechanisms of insecticide resistance (supplementary file 1: table S19, Supplementary Material online); metabolic detoxification including CYPs, esterases, GSTs, and ABC transporters (Feyereisen et al. 2015); target-site insensitivity including most of the modes of actions classified by the Insecticide Resistance Action Committee (IRAC: http://www.irac-online.org/modes-of-action/) such as TRPC channels, sodium channels, calcium channels, glutamate receptors, and acetylcholine receptors; and reduced cuticle penetration, including genes involved in chitin production or cuticle development.

One frequent concern in analyses of large data sets is the inclusion of false positives (Type I error) resulting from the large number of statistical tests. Frequently, this is resolved by adjusting P values to more conservative values, for example, by implementing multiple testing recalibrations such as Bonferroni or the Benjamini–Hochberg false discovery rate. However, in functional genomic studies, gene lists provide objective hypotheses that can be easily assessed in follow-up studies, while removing them based on statistical criteria can sometimes be challenging (François et al. 2016). Although we employ stringent statistical criteria in all tests, we additionally leverage our gene lists by comparing results from different tests. As we expect that both regulatory and structural changes might lead to genetic adaptation, we tested for overrepresentation of specific gene networks in selection tests, gene expression analyses, and combinations of both approaches. Given multiple data sources, overlap in gene identity and function provides a measure of support for repeated evolution and polygenic adaptation. We curated gene ontology terms associated with significant genes in lists from genome-wide selection tests and differential expression tests, and used a one-sided hypergeometric Fisher’s exact test (Rivals et al. 2007; De Leeuw et al. 2016) to test for overrepresentation (enrichment) of gene ontology terms, with P value < 0.05 used as the statistical significance threshold. To further refine this analysis, we used REVIGO (Supek et al. 2011), a clustering algorithm that relies on semantic similarity measures, to summarize the list of enriched gene ontology terms. Gene ontology terms associated with biological processes, cellular components, and molecular function were separately clustered using the simRel score for functional similarity, allowing for redundancy in similar terms up to a value of 0.7 before removal, and then compared with the UniProt database to find the percentage of genes annotated with each gene ontology term. The results were visualized using a CirGO plot (Kuznetsova et al. 2019). To provide a context for interpreting these results, we used our list of candidate genes (supplementary file 1: table S19, Supplementary Material online) to generate a CirGO plot of gene ontology terms associated with insecticide resistance (supplementary file 1: fig. S41, Supplementary Material online). Major biological processes include response to insecticide/response to oxidative stress, endocytosis, glycerolipid metabolism, sensory perception, and DNA integration. Major cellular components include the plasma membrane/integral component of the membrane and transcription factor complex. Finally, major molecular functions include metalloendopeptidase activity, monoxygenase activity, acetylcholine binding, lipid binding, DNA polymerase binding, tetracycline transporter activity, chromatin binding, chitin binding, and structural component of cuticle.

Gene Expression Analyses

In order to test for regulatory evolution, we compared gene expression data from RNA sequencing (RNAseq) experiments across the geographical range of CPB, including original data from the Plains region (a Colorado population) and previously published pest CPB population samples (see supplementary file 1: table S16, Supplementary Material online). The Colorado population was raised under greenhouse conditions on potato plants (~25 °C, 16:8 light:dark cycle), but represents the first generation derived from wild-caught adults feeding on an ancestral hostplant Solanum rostratum. RNAseq studies are recognized as robust estimators of whole-genome gene expression profiles (Everaert et al. 2017) that are highly responsive to experimental conditions (Conesa et al. 2016). As the original experiments varied in their design, we conducted a set of initial analyses to determine if sampling conditions might bias geographical comparisons. First, we assessed whether sampling of an overwintering (postdiaspore) population versus a summer (nondiaspore) generation at the same field in Wisconsin altered gene expression patterns (first generation vs. second generation). Second, we examined whether direct exposure to imidacloprid was necessary to induce insecticide resistance gene expression responses, by comparing a set of lab-reared individuals from Wisconsin, Oregon, and Long Island populations (control vs. an imidacloprid-induction treatment). Third, we compared samples of larvae and adults from a susceptible population in Wisconsin. Based on these comparisons, we determined that regional population differences could be compared for adults from field collected populations irrespective of generation sampled, but lab reared larvae needed to be compared separately. In both the adult and larval regional analysis, populations were included notwithstanding imidacloprid tolerance. We compared constitutive levels of gene expression in six adult populations: Colorado (CO), Wisconsin (WI), Michigan (MI), New York (NY), New Jersey (NJ), and eastern Canada (CAN). We compared constitutive levels of gene expression in four larval populations: Oregon (OR), Wisconsin (WI), New York (NY), and New Jersey (NJ). As some of the adult samples were sequenced as pair-end and single-end reads, we analyzed only the first read of a set of pair-end samples representing CO and WI.

We aligned short read data from each sample to the L. decemlineata reference genome using HISAT2 (Kim et al. 2015). SAMTOOLs was used to convert sam files to bam files. Read counts per gene per sample were generated using the
function `featureCounts` available in the `Rsubread` package (Liao et al. 2019), with reference to the `L. decemlineata` OGS. Using the resulting counts, we evaluated evidence for differential gene expression for each region using `DESeq2` (Love et al. 2014). `DESeq2` first estimates the dispersion among a set of replicated samples and then the logarithmic fold change of transcript counts among sample groups. It then employs a generalized linear model based on the negative binomial distribution of transcript counts and a binomial Wald statistic to test for differences among experimental contrasts. We first trimmed the read count matrix to remove genes with less than five reads and then conducted the differential expression analysis. We retained differentially expressed genes if read counts were >2-fold and the significance level `α` = 5% was reached. We then adjusted the false discovery rate to 1% level, using a Benjamini–Hochberg correction (Benjamini and Hochberg 1995). Heatmaps of differentially expressed genes were generated in `R` using the `pheatmap` package (Kolde and Kolde 2015). As described previously, we curated gene ontology terms associated with significant genes and used a one-sided hypergeometric Fisher’s exact test to test for overrepresentation (enrichment) of gene ontology terms, with a `P` value <0.05 used as the statistical significance threshold.

Supplementary Material
Supplementary data are available at Molecular Biology and Evolution online.

Acknowledgments
The authors wish to thank the reviewers and editorial staff for their assistance with our manuscript. The University of Wisconsin Biotechnology Center DNA Sequencing Facility provided essential facilities and services. The authors thank Russell Groves, Justin Clements, Sylvia Rondon, Kristian Brevik, and Andrei Alyokhin for engaging discussions on CPB. Finally, they thank Margarethe Brummerman for help with the manuscript. The authors thank Russell Groves, Justin Clements, Sylvia Rondon, Kristian Brevik, and Andrei Alyokhin for engaging discussions on CPB. Finally, they thank Margarethe Brummerman for help with the manuscript.

Data Availability
The data underlying this article are publicly available in NCBI Short Read Archive at https://www.ncbi.nlm.nih.gov/sra/ (Bioproject PRJNAS80490) and as supplementary files 1 and 2, Supplementary Material online.

References
Allio R, Donega S, Galtier N, Nabholz B. 2017. Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. Mol Biol Evol. 34(11):2762–2772.

Alyokhin A, Baker M, Mota-Sanchez D, Dively G, Grafius E. 2008. Colorado potato beetle resistance to insecticides. Am J Pot Res. 85(6):395–413.

Alyokhin A, Mota-Sanchez D, Baker M, Snyder WE, Menasha S, Whalon M, Dively G, Moarsi WF. 2015. The Red Queen in a potato field: integrated pest management versus chemical dependency in Colorado potato beetle control. Pest Manag Sci. 71(3):343–356.

Anderson CJ, Oakeshott JG, Tay WT, Gordon KHL, Zwick A, Walsh TK. 2018. Hybridization and gene flow in the mega-pest lineage of moth, Helicoverpa. Proc Natl Acad Sci U S A. 115(19):5034–5039.

Barghi N, Hermisson J, Schlötterer C. 2020. Polygenic adaptation: a unifying framework to understand positive selection. Nat Rev Genet. 21(12):769–781.

Barton N. 2010. Understanding adaptation in large populations. PLoS Genet. 6(6):e1000987.

Barton NH, Keightley PD. 2002. Understanding quantitative genetic variation. Nat Rev Genet. 3(1):11–21.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Methodol. 57(1):289–300.

Bivand R, Keitt T, Rowlingson B, Pebesma E. 2014. Package ‘rgdal’. Bindings for the Geospatial Data Abstraction Library. Available from: https://cran.r-project.org/web/packages/rgdal/index.html. Accessed December 01, 2019.

Boitard S, Boussaha M, Capitan A, Rocha D, Servin B. 2016. Uncovering adaptation from sequence data: lessons from genome resequencing of four cattle breeds. Genetics 203(1):433–450.

Bracewell R, Vanderpool D, Good J, Six D. 2018. Cascading speciation among mutualists and antagonists in a tree–beetle–fungi interaction. Proc R Soc B. 285(1811):20180694.

Brevik K, Bueno EM, McKay S, Schoville SD, Chen YH. 2021. Insecticide exposure affects intergenerational patterns of DNA methylation in the Colorado potato beetle, Leptinotarsa decemlineata. Eval Appl. 14(3):746–757.

Brevik K, Schoville SD, Mota-Sanchez D, Chen YH. 2018. Pesticide durability and the evolution of resistance: a novel application of survival analysis. Pest Manag Sci. 74(8):1953–1963.

Calia B, Demkovich M, Siegel JP, Viana JGP, Walden KKO, Robertson HM, Berenbaum MR. 2020. Selective sweeps in a nutshell: the genomic footprint of rapid insecticide resistance evolution in the almond agroecosystem. Genome Biol Evol. 13(3):evaa234.

Campos JL, Zhao L, Charlesworth B. 2017. Estimating the parameters of background selection and selective sweeps in Drosophila in the presence of gene conversion. Proc Natl Acad Sci U S A. 114(24):E4762–E4771.

Cappelle K, de Oliveira CFR, Van Eynde B, Christiaens O, Smagghe G. 2021. Insecticide resistance and the use of mitochondrial DNA as a molecular marker. Mol Biol Evol. 34(11):2762–2772.
select populations of *Leptinotarsa decemlineata* in the Central Sands region of Wisconsin. *PloS One* 11(1):e0147844.

Cohen ZP, Brevik K, Chen YH, Hawthorne DJ, Weibiel BD, Schoville SD. 2021. Elevated rates of positive selection drive the evolution of pestiferousness in the Colorado potato beetle (*Leptinotarsa decemlineata*, Say). *Mol Ecol.* 30(1):237–254.

Conesa A, Gótz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21(18):3674–3676.

Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczekan M, Gaffney DJ, Elollo L, Zhang X, et al. 2016. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17:13.

Corbett-Detig RB, Hartl DL, Schoville SD, Rondon SI, Schoville SD. 2017. Landscape genomics and adaptation from standing variation in the European aspen (*Populus tremula* L.), a widespread forest tree. *Mol Ecol.* 26(22):6284–6300.

Crossley MS, Chen YH, Groves RL, Schoville SD. 2017. Landscape genomics of Colorado potato beetle provides evidence of polygenic adaptation to insecticides. *Mol Ecol.* 26(22):6284–6300.

Crossley MS, Rondon SJ, Schoville SD. 2018. A comparison of resistance to imidacloprid in Colorado potato beetle (*Leptinotarsa decemlineata* Say) populations collected in the Northwest and Midwest US. *Am J Potato Res.* 95(5):495–499.

Crow JF. 1957. Genetics of insect resistance to chemicals. *Annu Rev Entomol.* 2(1):227–246.

Cutter AD, Jovelin R, Dey A. 2013. Molecular hyperdiversity and evolution in very large populations. *Mol Ecol.* 22(8):2074–2095.

Daborn PJ, Yen JL, Bogwitz MR, Le Goff G, Feil E, Jeffers S, Tijet N, Perry T, Cutter AD, Jovelin R, Dey A. 2013. Molecular hyperdiversity and evolution in arthropods: theory and practice. *Am J Potato Res.* 8(1):7858.

Davis-Vogel C, Ortiz A, Procyk L, Robeson J, Kassa A, Wang Y, Huang E, Guillemaud T, Ciosi M, Lombaert E, Estoup A. 2011. Biological invasions and implications for pest management. In: National Research Council, Board on Agriculture, Committee on Strategies for the Management of Pesticide Resistant Pest Populations. Pesticide resistance: strategies and tactics for management. Washington (DC): National Academy Press. p. 157–169.

Dedhia SN, Nair AV, Robles M. 2005. A nuclease specific to lepidopteran insects suppresses RNAi. *Curr Opin Insect Sci.* 17:1–16.

Denholm I, Rowland M. 1992. Tactics for managing pesticide resistance. In: National Research Council, Board on Agriculture, Committee on Strategies for the Management of Pesticide Resistant Pest Populations. Pesticide resistance: strategies and tactics for management. Washington (DC): National Academy Press. p. 157–169.

Denu eck P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Denny N, Li H, Zhai W, Fritz MH-Y, et al. 2010. A draft sequence of the *Drosophila* genome and 186. 2006. Factors influencing the evolution of resistance. In: National Research Council, Board on Agriculture, Committee on Strategies for the Management of Pesticide Resistant Pest Populations. Pesticide resistance: strategies and tactics for management. Washington (DC): National Academy Press. p. 157–169.

Gould J. 2005. Hierfstat, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes.* 5(1):184–186.

Faria M, Brown ZS, Kuzma J. 2018. Wicked evolution: can we address the sociobiological dilemma of pesticide resistance? *Science* 360(6390):728–732.

Faye FR, Krause J, Briggs JW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MY-H, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328(5979):710–722.

Guan R-B, Li H-C, Fan Y-J, Hu S-R, Christiaens O, Smagghe G, Miao X-X. 2010. A nuclease specific to lepidopteran insects suppresses RNAi. *J Biol Chem.* 293(16):6011–6021.

Gui S, Taning CNT, Wei D, Smagghe G. 2020. First report on CRISPR/Cas9-targeted mutagenesis in the Colorado potato beetle, *Leptinotarsa decemlineata*. *J Insect Physiol.* 121:161–177.

Guerilloud T, Ciosi M, Lombaert E, Estoup A. 2011. Biological invasions in agricultural settings: insights from evolutionary biology and population genetics. *C R Biol.* 334(3):237–246.
