Mutations in a Novel Cadherin Gene Associated with Bt Resistance in *Helicoverpa zea*

Megan L. Fritz,[*,1] Schyler O. Nunziata,[*] Rong Guo,[*] Bruce E. Tabashnik,[†] and Yves Carrière[†]
*Department of Entomology, University of Maryland, College Park, MD 20742 and †Department of Entomology, University of Arizona, Tucson, AZ 85721

ORCID IDs: 0000-0003-0472-4877 (M.L.F.); 0000-0002-2430-2367 (B.E.T.)

ABSTRACT Transgenic corn and cotton produce crystalline (Cry) proteins derived from the soil bacterium *Bacillus thuringiensis* (Bt) that are toxic to lepidopteran larvae. *Helicoverpa zea*, a key pest of corn and cotton in the U.S., has evolved widespread resistance to these proteins produced in Bt corn and cotton. While the genomic targets of Cry selection and the mutations that produce resistant phenotypes are known in other lepidopteran species, little is known about how selection by Cry proteins shape the genome of *H. zea*. We scanned the genomes of Cry1Ac-selected and unselected *H. zea* lines, and identified twelve genes on five scaffolds that differed between lines, including *cadherin-86C* (*cad-86C*), a gene from a family that is involved in Cry1A resistance in other lepidopterans. Although this gene was expressed in the *H. zea* larval midgut, the protein it encodes has only 17 to 22% identity with cadherin proteins from other species previously reported to be involved in Bt resistance. An analysis of midgut-expressed cDNAs showed significant between-line differences in the frequencies of putative nonsynonymous substitutions (both SNPs and indels). Our results indicate that *cad-86C* is a likely target of Cry1Ac selection in *H. zea*. It remains unclear, however, whether genomic changes at this locus directly disrupt midgut binding of Cry1Ac and cause Bt resistance, or indirectly enhance fitness of *H. zea* in the presence of Cry1Ac by some other mechanism. Future work should investigate phenotypic effects of these nonsynonymous substitutions and their impact on fitness of *H. zea* larvae that ingest Cry1Ac.

Transgenic crops are extensively used for the management of both insect and plant pests worldwide, which places extraordinary pressure on pest species to adapt (Tabashnik and Carrière 2017; Gould *et al.* 2018). The first generation of commercially available transgenic crops included corn and cotton bioengineered to produce a single crystalline (Cry) protein from the soil-dwelling bacterium, *Bacillus thuringiensis* (Bt) (Gould 1998). The primary targets of these Cry proteins were difficult-to-manage lepidopteran larvae (Roush 1997). When fed upon Cry-producing plant tissue, larvae often experience

significant reductions in growth and survivorship (Clark *et al.* 2000; Tabashnik *et al.* 2000; Ali *et al.* 2006; Liu *et al.* 2010; Pardo-López *et al.* 2013). Not all targeted lepidopteran species are highly susceptible to Cry proteins, however. Species with low inherent susceptibility are expected (Gould 1998; Tabashnik *et al.* 2004; Carrière *et al.* 2015) and do evolve resistance to Bt crops faster than species with high susceptibility (Tabashnik and Carrière 2017). Accordingly, there is growing concern over the number of species that have evolved significant resistance to Bt crops producing Cry proteins (Tabashnik and Carrière 2017, 2019; Gould *et al.* 2018; USEPA 2018).

Helicoverpa zea is one lepidopteran species targeted by Bt crops with low inherent susceptibility to Cry proteins (Stone and Sims 1993; Ali *et al.* 2006; Sivasupramaniam *et al.* 2008). A major pest of both corn and cotton in the U.S., *H. zea* has evolved resistance to several Cry proteins (Welch *et al.* 2015; Dively *et al.* 2016; Reisig *et al.* 2018; Kaur *et al.* 2019; Yang *et al.* 2019). While some efforts have been made to understand the mechanisms underlying Cry resistance in *H. zea* (Caccia *et al.* 2012; Zhang *et al.* 2019), we still know little about how selection by Cry proteins has shaped genotype frequencies in this important pest.

Work in other Cry-resistant Lepidoptera provides clues as to which genes are potential targets of selection in *H. zea*, however. When a larva ingests Cry proteins, these toxins bind to receptors in the midgut, which lead to pore formation and lethal midgut cell death (Pardo-López *et al.* 2013). Disruption of toxin binding to larval midgut receptors is the most common mechanism of resistance (Peterson *et al.* 2017). Mutations that alter the coding sequence or reduce expression of Cry1A-binding cadherin proteins are associated with resistance to Cry1A toxins in several major lepidopteran pests (Gahan *et al.* 2001; Morin *et al.* 2003; Xu *et al.* 2005; Fabrick *et al.* 2014; Zhang *et al.* 2017; Wang *et al.* 2018).

Here, we used a genome scanning approach to compare previously described Cry1Ac-selected and unselected lines of *H. zea* (Brévault *et al.* 2013, 2015; Orpet *et al.* 2015a, 2015b; Welch *et al.* 2015; Carrière *et al.* 2018a). We identified five regions of the genome showing signatures of selection, one of which includes a novel gene from the cadherin family, which has been shown to comprise genes associated with Cry resistance in other lepidopteran species (Gahan *et al.* 2001; Fabrick *et al.* 2014; Zhang *et al.* 2017). We compared the predicted protein sequence of the novel cadherin with cadherins involved in Bt resistance in other lepidopteran species, and analyzed both the midgut expression patterns and predicted amino acid sequence differences at this gene for selected and unselected *H. zea* lines. Finally, we discuss the possible functional roles of this gene in *H. zea* resistance to Cry proteins.

MATERIALS AND METHODS

Insect material

We conducted our screen for candidate genes associated with Bt resistance in two laboratory-reared lines of *H. zea* that differed in susceptibility to the Bt toxin Cry1Ac. The lines were founded with 180 larvae collected in Georgia from Cry1Ab corn in 2008 (Brévault *et al.* 2013). F_1 progeny from field-collected individuals gave rise to the GA line, which was reared on wheat-germ diet and was unexposed to Bt toxins. After two generations of laboratory rearing, we used a subset of insects from GA to initiate the GA-R line, which was selected for resistance to Cry1Ac. From 2008 to 2012, each line was reared with *ca.* 600 individuals per generation, with approximately 11 generations per year. Between 2008 and 2010, GA-R was selected with Cry1Ac over 9 non-consecutive generations as described in Brévault *et al.* (2013). This yielded 10-fold resistance to Cry1Ac and significantly higher survival on Cry1Ac and Cry1Ac + Cry2Ab cotton in GA-R relative to GA (Brévault *et al.* 2013). From 2010 to 2012, GA-R was selected again with Cry1Ac over 10 non-consecutive generations. In 2012, GA-R was crossed to GA to generate a new GA-R line, and the new GA-R and original GA lines were split into two subsets, each reared with *ca.* 600 individuals per generation (Orpet *et al.* 2015a). The two subsets of each line were crossed every second or third generation to generate two new subsets (Orpet *et al.* 2015a). The new GA-R was selected for seven non-consecutive generations with Cry1Ac, which yielded 14-fold resistance to Cry1Ac in GA-R relative to GA (Orpet *et al.* 2015a) and significantly higher survival on Cry1Ac + Cry2Ab cotton in GA-R than GA (Carrière *et al.* 2019). Between 2012 and 2016, GA-R was selected for an additional 25 non-consecutive generations with Cry1Ac, using the methods described in Carrière *et al.* (2018a). Male pupae sampled for genomic analysis were from generations F52 for GA-R (n = 5) and F72 for GA (n = 5) reared in October 2016.

DNA isolation and whole genome sequencing

We extracted whole genomic DNA from 5 individuals per line using a Qiagen DNEasy Blood and Tissue Kit (Qiagen, Inc., Valencia, CA, USA) following the mouse tail protocol with some modifications. For each individual, DNA was extracted from half of a pupa placed into a 2 mL microcentrifuge tube, containing 180 μL ATL buffer and 20 μL proteinase K. A sterile ceramic bead was added to each tube, and the capped microcentrifuge tubes were placed on a Benchmark Benchmix vortex with a multi-head attachment (Genesee Scientific Corporation, El Cajon, CA, USA). Tube contents were vortexed at the highest speed to grind the tissue, and the ground sample was incubated overnight at 55°. After incubation, samples were centrifuged at 12,000 × *g* for 5 min to precipitate the chitin. RNase A (3 μL at 4 μg/μL) was added to the supernatant and incubated at 37° for 15 min. The supernatant was adsorbed onto a Qiagen column, and the remainder of the preparation followed manufacturer's instructions. The final volume of sample produced per individual was 200 μL, and DNA was stored at −20° until use. Genomic DNA was submitted to the North Carolina State University Genomic Sciences Laboratory, where it was prepared for sequencing using an Illumina TruSeq LT library preparation kit (Illumina, Inc. San Diego, CA). We barcoded DNA samples from each individual, after which all DNA samples were pooled for sequencing. We sequenced the prepared library on an Illumina NextSeq500 at the North Carolina State University Genomic Sciences Laboratory using 150 base-pair (bp) paired-end reads.

Read processing and mapping

Sequences were quality filtered to remove all reads with more than 30% of bases having a quality score below Q20 using NGS QC Toolkit v. 2.3.3 (Patel and Jain 2012). Low-quality ends (<Q20) were trimmed from the 3′ end of remaining reads to improve overall alignment quality (Del Fabbro *et al.* 2013). Remaining filter-trimmed reads were mapped to the *H. zea* reference genome (Pearce *et al.* 2017) with Bowtie v. 2.3.2 (Langmead *et al.* 2009) in end-to-end mode using the highest sensitivity preset parameters (–very-sensitive). Alignment files were cleaned to keep only reads in proper pairs with robust mapping quality (MAPQ ≥ 10) using SAMtools v. 1.5 (Li 2011), and PCR and optical duplicates were identified and removed using Picard v. 2.10.5 (http://broadinstitute.github.io/picard). The cleaned alignment files were used to call SNPs with SAMtools v. 1.5 using the mpileup function, and SNP and indel genotypes in Variant Call Formatted (VCF) files were generated using BCFtools. The VCF files were filtered prior to population genomic analysis to only include loci that were: 1) genotyped in at least 50% of individuals, 2) were sequenced to a minimum depth of coverage of 5 and maximum of 2.5 times the mean genome-wide coverage, 3) had a minor allele frequency (MAF) of > 0.1, and 4) were biallelic, using VCFtools v. 0.1.15 (Danecek *et al.* 2011).

Genetic diversity

Following genotype quality filtering, levels of genetic diversity within GA and GA-R were estimated using various metrics. We estimated the proportion of polymorphic sites (P_N), and average MAF with PLINK v.1.07 (Purcell *et al.* 2007). We thinned our dataset to 1 SNP per 5kb to reduce the number of SNPs in linkage disequilibrium in our dataset, and then calculated observed (H_O) and expected (H_E) heterozygosity and inbreeding coefficients (F_{IS}) for each line using the hierfstat package (v. 0.04-22, Goudet 2005) in R (v. 3.5.3; R Development Core Team 2008). To estimate overall within population genetic diversity, we calculated Nei's pairwise genetic distance (d_X and d_Y) using the StAMPP package (v. 1.5.1; Pembleton *et al.* 2013).

Identifying selective sweeps

To identify putative selective sweeps, we searched for regions of the genome with high divergence between GA and GA-R and exceptionally

low genetic variation in GA-R, as measured by heterozygosity. We calculated the heterozygosity of within-population pools of individuals (Hp; Rubin *et al.* 2010) using 40-kb windows with 20-kb of overlap. Hp was calculated as follows:

$$H_p = \frac{2 \sum n_{MAJ} \sum n_{MIN}}{\left(\sum n_{MAJ} \sum n_{MIN}\right)^2}$$

where n_{MAJ} is the number of major alleles and n_{MIN} the number of minor alleles in the window. To prevent spurious signals from few SNPs, we excluded windows with fewer than 10 SNPs. According to Rubin *et al.* (2010), we standardized estimates of Hp using a Z transformation (ZHp), and putative regions under selection were identified as being over 6 standard deviations from the mean (ZHp < -6). Using the same sliding window and step size, we also calculated F_{ST} to quantify genetic divergence between GA and GA-R in PLINK. Sliding-window averaged F_{ST} values were also Z-transformed, although no regions with very low heterozygosity in GA-R (ZHp < -6) also showed F_{ST} values greater than 6 standard deviations from the mean F_{ST} value calculated for the GA and GA-R comparison. However, some regions of low heterozygosity in GA-R did display F_{ST} values greater than 1 standard deviation from the mean ($ZF_{ST} > 1.1$), and for these regions, the F_{ST} values ranged from 0.44-0.74.

We used msms simulations (Ewing and Hermisson 2010) to calculate the probability that the observed genetic divergence between strains ($F_{ST} > 0.44$) could have occurred along a 40kb window without selection, given a population demographic scenario similar to what our *H. zea* lines experienced. As a test of sensitivity, additional simulated genotypic datasets were also developed to test population demographic scenarios with smaller population sizes per generation, as well as initial collection sizes. The distributions of F_{ST} values were compared among scenarios. The full description of these simulations can be found in File S1.

Finally, we calculated sliding window averaged F_{ST} values for the pair of lines using the pF_{ST} function in vcflib (https://github.com/vcflib) over a sliding window size of 5-kb with a 1-kb step size. These values were compared with an empirically-derived estimate of genome-wide divergence between GA and GA-R. A Benjamini and Hochberg (1995) false discovery rate (FDR) correction was applied to the p-values for each of these statistical tests. All of these tests taken together indicated that there was one region of the genome where multiple 40kb genomic windows had ZHp values < -6, statistically significant FDR-corrected F_{ST} values and ZF_{ST} estimates > 1.1, the latter of which appeared unlikely in the absence of selection, according to msms simulations. In this region with the strongest evidence of a selective sweep based on the aforementioned criteria, one of the genes we found was called *cadherin-86C* (*cad-86C*). Because of strong evidence of selection in and around this gene in GA-R, and evidence of cadherin involvement in Bt resistance in other lepidopteran species (Gahan *et al.* 2001; Morin *et al.* 2003; Wang *et al.* 2005; Zhang *et al.* 2017; Wang *et al.* 2018), we examined this region further.

Verification of heterozygosity and genetic divergence at cad-86C by Sanger sequencing

Further evidence for selection at the *cad-86C* gene in GA-R was gathered using a Sanger sequencing approach. Because our whole genome resequencing data were from 5 individuals per line, we amplified and sequenced two ~600 base-pair (bp) regions of the targeted gene across 24 new individuals in each line to confirm the evidence for selection. These 600 bp target sequences were 1,222 bp

apart in a non-coding region of the gene. Primers were designed with PrimerQuest (www.idtdna.com/Primerquest) and are listed in Table S1. Twenty microliter polymerase chain reactions (PCRs) were conducted using 300 ng of genomic DNA per individual, 0.7uM of each forward and reverse primer, 0.2mM dNTPs, and 4uL 5× GoTaq buffer and 1.25 U GoTaq DNA polymerase (Promega Corporation, Madison, WI, USA) according to the recommended protocol. Cycling conditions included denaturation at 95° for 3 min, followed by 30 cycles of 95° for 30 sec, 60° for 30 sec, 72° for 30 sec and a final elongation step at 72° for 1 min on Bio-Rad T100 Thermal Cycler (Bio-Rad Laboratories, Inc. Hercules, CA, USA). PCR products were cleaned using an ExoAP reaction (see File S2) prior to submission for sequencing. Sanger sequencing was performed using BigDye Terminator v3.1 chemistry (Applied Biosystems) and SNPs were called using PolyPhred (Nickerson *et al.* 1997). Sequences were manually edited to remove low quality nucleotide calls at their ends, as well as to incorporate variant SNPs for heterozygotes using PolyPhred output. Trimmed and edited sequences were aligned using Clustal Omega available through the European Bioinformatics Institute (EMBL-EBI), and estimates of within population genetic diversity (π, θ_W) and Tajima's D were calculated using the pegas package (v. 0.11; Paradis 2010) for R. A p-value was also calculated for each Tajima's test, which indicated whether the D statistic was significantly different from 0. Permutation tests, which were custom-coded in R, were used to identify statistically significant differences in π for GA and GA-R at each *cad-86C* locus.

Comparison of CAD-86C to other cadherin proteins by sequence alignment

We compared our *H. zea* CAD-86C protein sequence (ID = 537580 from the *H. zea* assembly annotation) with CAD-86C orthologs from other species, as well as other cadherins (BtR and CAD2) known to be involved in Bt resistance. All available cadherin protein sequences from 6 lepidopteran species (*H. zea*, *Pectinophora gossypiella*, *Heliothis virescens*, *Helicoverpa armigera*, *Chilo suppressalis*, *and Bombyx mori*), as well as CAD-86C from *Drosophila melanogaster* were acquired from NCBI. Protein sequences used for analysis, along with their GenBank accession numbers, can be found in File S3. BtR orthologs and the *C. suppressalis* CAD2 were first aligned to our *H. zea* CAD-86C sequence, and a percentage identity matrix was calculated using T-Coffee available through EMBL-EBI (Madeira *et al.* 2019). All available BtR, CAD2, and CAD-86C sequences were then aligned to one another using MUSCLE (Edgar 2004), and a phylogenetic analysis was conducted using the phangorn package (v.2.5.3, Schliep 2011) in R. We used a maximum likelihood approach to distance matrix calculation, assuming a Whelan and Goldman model of molecular protein evolution (Whelan and Goldman 2001). An unrooted phylogeny was produced using a neighbor-joining approach, and bootstrap support values for the tree nodes were calculated using 1000 resampling events.

Expression of cad-86C in the H. zea midgut

If *cad-86C* were involved in Bt resistance, we reasoned that it should be expressed in the larval midgut. We performed reverse transcriptase quantitative PCR (RT-qPCR) on dissected midgut samples from GA (F81 generation) and GA-R (F62 generation) reared in October 2017. GA-R had been selected for resistance to Cry1Ac in 6 additional non-consecutive generations between the F52 generation used for genomic analysis (see above) and the F62 generation. Resistance to Cry1Ac was verified in dissected GA-R larvae by selecting F62 neonates with a diet

overlay bioassay using 40 µg Cry1Ac per cm² of diet surface (Carrière *et al.* 2018a). Susceptibility to Cry1Ac was also verified in dissected GA larvae by selecting F81 neonates with a diet overlay bioassay using a concentration of 30 µg Cry1Ac per cm² of diet surface. Seven days after these bioassays were initiated, GA-R larvae that reached third instar or larger were considered resistant, while GA larvae that remained at first instar were considered susceptible to Cry1Ac. Resistant larvae from GA-R and susceptible larvae from GA were transferred to non-Bt diet and their midguts were dissected upon reaching 4th or 5th instar. Dissections were done in ice-cold RNAlater+PBS mixture. Whole midguts were stored individually in RNAlater for total RNA extraction.

Total RNA from dissected midguts was isolated using a Zymo Direct-zol RNA miniprep according to the protocol recommended by the manufacturer. We verified RNA quality and determined RNA concentration on an Experion automated electrophoresis station. We synthesized first strand cDNA using RevertAid H minus reverse transcriptase and 1µg of total RNA in a 20µL reaction. Gene specific primers were used to amplify *cad-86C* and an endogenous control gene, α-Tubulin (see Table S2 for primer sequences). We performed 20µL qPCR reactions using 10ng of cDNA, 0.5µM primers, and 10uL PowerUp SYBR Green PCR Master Mix (Applied Biosystems). Cycling conditions included initial incubation at 50° for 2 min, followed by denaturation at 95° for 2 min, and then 40 cycles at 95° for 15 sec, 60° for 1 min on a 7300 Applied Biosystems Real Time PCR System. All reactions were run in triplicate for the 13 GA-R and 14 GA individuals. We observed a single peak in the dissociation curves for all reactions, and PCR efficiencies > 90% for each primer pair using LinRegPCR v.2017.1 (Ruijter *et al.* 2009). Gene transcript levels were normalized to α-Tubulin and relative expression was standardized using the gene transcript levels detected in GA individuals. Expression of *cad-86C* in GA-R relative to GA was calculated using $2^{-\Delta\Delta Ct}$.

Long-read sequencing of cad-86C cDNAs

To analyze the expressed gene product and putative protein sequence, we designed barcoded PCR primers in the 5′ and 3′ UTR regions of *cad-86C*, amplified the complete cDNA, and sequenced this cDNA by single molecule sequencing. The midgut cDNAs from 13 GA-R and 14 GA individuals were subjected to high fidelity PCRs using the barcoded primer sequences found in Table S3. Full length cDNAs were amplified using Q5 High Fidelity DNA polymerase in master-mix format (New England Biolabs Inc., Ipswich, MA, USA). The 50µL reactions contained 25ng cDNA, 0.5µM primers, and 25µL of 2× Q5 mastermix, and cDNAs were amplified under the following cycling conditions: initial denaturation at 95° for 3 min, followed by 30 cycles of 95° for 30 sec, 60° for 30 sec, 72° for 3:40 min and a final elongation step at 72° for 1 min on Bio-Rad T100 Thermal Cycler (Bio-Rad Laboratories, Inc. Hercules, CA, USA). Amplicons were run on a 1% agarose gel and the highest molecular weight fragments (> 5kb) were excised, purified using a Zymoclean Gel DNA recovery kit (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol, and the quantity of purified cDNA was measured using an Agilent D1000 Screentape System (Agilent Technologies, Inc. Santa Clara, CA, USA). Amplicons were then pooled in equimolar amounts, and sequenced on a single Pacific Biosciences (PacBio) SMRT cell at the North Carolina State University Genomic Sciences Laboratory.

Identification of CAD-86C amino acid substitutions

A PacBio-generated bam file was converted to fastq file by bedtools (Quinlan and Hall 2010), and sequences from individuals were demultiplexed using the bbduk.sh script from bbmap (Bushnell

2014). Parameters k and restrictleft were set to 18, which was equal to the primers' lengths, so that the software only looked for primers matching the leftmost 18 bp. Filtering was done by the FASTA manipulation tool of Galaxy (https://usegalaxy.org; Blankenberg *et al.* 2010), and the minimal and maximal length parameters were set to 5000 and 0, respectively, to return all sequences longer than 5000 bp. The correction phase of Canu (Koren *et al.* 2017) was used to improve the accuracy of base calls in PacBio long reads. All error-corrected reads were aligned to the cDNA sequence of *H. zea*'s *cad-86C* gene using the mapPacBio.sh script of bbmap. The results were viewed in Integrative Genomics Viewer (IGV) (Robinson *et al.* 2011).

Consensus cDNA sequences were identified for each GA and GA-R individual, and ambiguous regions with two alleles, indicating an individual was heterozygous at a locus, were manually edited based upon visual inspection of sequences in IGV. Nucleotide sequences were translated to protein sequences using the translate tool in ExPASy (https://web.expasy.org/translate/), and amino acid sequences were aligned to each other with T-Coffee (Notredame *et al.* 2000) to identify potential amino acid substitutions. A two-sided Fisher's exact test was used to compare the distribution of CAD-86C amino acid variants between lines.

Data availability

Whole genome sequencing data are available on NCBI under the BioProject ID PRJNA599999. Sanger, qPCR, and PacBio data can be found in Dryad Digital Repository https://doi.org/10.5061/dryad.1c59zw3s1. Scripts used for analysis of PacBio data can be found at https://github.com/rguo1990/Hzea_PacBio_sequencing. Scripts used for all other analyses can be found at https://github.com/mcadamme/GA_and_GAR. Supplemental material available at figshare: https://doi.org/10.25387/g3.11549403.

Description of figshare contents

Figures S1 and S2 contain the permutation test resampling distributions used to identify statistically significant differences between nucleotide diversity estimates for amplicons of *cad-86C* 1b and 2b primer pairs. Figure S3 is a gel image showing the 5,539 bp *cad-86C* cDNA amplicons from the *H. zea* larval midguts. Table S1 contains the *cad-86C* 1b and 2b primer sequences. Table S2 contains the qRT-PCR primers for *cad-86C* and the α-Tubulin control. Table S3 contains the barcodes used for identification of *cad-86C* sequences from each individual. Table S4 contains a summary of the number of Illumina reads produced per individual, the number that remained after filtering, and coverage depth. Table S5 contains a summary of the genomic regions with low heterozygosity in GA-R and high genetic divergence between GA and GA-R. Table S6 contains a percent identity matrix for the *H. zea* CAD-86C with other cadherin proteins known to be involved in Bt resistance. Table S7 contains a summary of the PacBio sequencing reads produced for each *H. zea* individual, and the number of reads remaining after filtering. Table S8 contains haplotype frequency information for *cad-86C* cDNA sequences. File S1 describes the methods used to run msms. File S2 contains our ExoAP clean protocol. File S3 contains the cadherin sequences used for phylogenetic analysis. File S4 shows the cadherin protein sequences aligned by T-Coffee.

RESULTS

WGS data and variant call

A total of 427,136,781 raw PE reads were generated, with 356,202,508 remaining after quality filtering, and 179,040,743 remaining after mapping to the reference genome (Table S4). Uniquely placed reads

covered 74.3% of the 335.5 megabases (Mb) genome, with a mean genome wide depth of coverage of 16.1× (range 7.96-23.2×) across all individuals. The initial number of variants before filtering was 5,106,839. After filtering, the total number of SNPs used for population genomic analysis was 1,986,042, and the total number of short indels was 422,149 (2,408,191 total markers).

Genetic diversity and selective sweeps

Genome-wide average values of pairwise genetic distance (d_X, d_Y) were similar for GA and GA-R (Table 1) as determined by their overlapping 95% confidence intervals. However, GA-R had significantly higher observed heterozygosity (H_O), and a significantly lower inbreeding coefficient (F_{IS}). Overall, the two lines were moderately diverged from each other with a genome-wide F_{ST} value of 0.23. Based on heterozygosity estimates among GA-R individuals, a total of 38 genomic windows across 24 scaffolds were identified as putative regions under selection (Figure 1). These candidate regions corresponded to 38 genomic annotations located either upstream or downstream of the identified window (Table S5). The distribution of ZHp values in GA-R indicated a high degree of heterozygosity across most 40-kb windows throughout the genome (Figure 1B). We then examined the Z-transformed F_{ST} values across these 40-kb genomic windows, but no 40-kb window had $ZF_{ST} > 6$ (Figure 1A). The maximum ZF_{ST} was 3.79 and the ZF_{ST} distribution was right skewed (Figure 2A), indicating that there were a number of genomic regions with moderate to high genetic divergence between lines (Figure 1A).

For those genomic regions displaying low Hp (ZHp < -6) in GA-R, several showed regions with $ZF_{ST} > 1$ between lines. The untransformed F_{ST} values across these 40Kb windows ranged from 0.44 - 0.74. Simulations with msms indicated that 40Kb window-averaged F_{ST} values of ≥ 0.44 were very unlikely under neutral expectations, given the population demographic conditions under which GA and GA-R were cultured. Under these conditions, none of the 10,000 simulations without selection produced $F_{ST} \geq 0.44$. Reducing the initial collection size in the simulations from the actual value of 180 to 90 or 45 individuals did not produce $F_{ST} > 0.44$ in 10,000 runs. When the simulated population sizes per generation were reduced to 50 or 25% of those used in our experiment, F_{ST} values exceeded 0.44 in two and 130 simulations of 10,000 ($P = 0.0002$ and 0.013), respectively. Thus, even with much smaller population sizes, the probability of producing $F_{ST} \geq 0.44$ was still low.

As further confirmation of genomic divergence in these regions, we observed statistically significant F_{ST} values ($P < 0.05$) in five of the already identified gene-containing regions using pF_{ST}, and we identified these as putative selective sweeps (Table S5). These regions were found on five different scaffolds and contained twelve predicted genes: *sol-1* (suppressor of lurcher-1), *cad-86C* (cadherin-86C), *map3k15* (mitogen-activated protein kinase kinase kinase 15), *msp20* (muscle-specific protein 20), *not1* (CCR4-Not complex subunit 1), *cpr47* (cuticular RR1 motif 47), *cgrrf1* (cell growth regulator with RING finger domain 1), *itf46* (intraflagellar transport 46), *cfap100* (cilia and flagellar associated protein 100), *mocs2* (molybdopterin synthase catalytic subunit), *zcchc24* (zinc finger CCHC domain containing 24), and an uncharacterized protein. The longest of these putative sweeps extended from 560,000bp on scaffold 20, just upstream of *cad-86C* (ID537580), through the entire length of *cad-86C*, as well as two other genes (an uncharacterized protein and *map3k15*), and ended at bp 740,000. Although the region was broad, the greatest reduction in heterozygosity in GA-R and greatest divergence (F_{ST}) between lines overlapped with *cad-86C* (Figure 3), a member of a gene

■ **Table 1** P_N is proportion of polymorphic sites, MAF is the mean minor allele frequency, H_O and H_E are observed and expected heterozygosity, F_{IS} is the inbreeding coefficient, and d_X and $_Y$ are Nei's average pairwise genetic distance between individuals in a population. Bias-corrected and accelerated bootstrapped 95% confidence intervals (N = 1000) are presented in parentheses. Bolded lines are statistically significant at the $P < 0.05$ level. H_O, H_E, F_{IS}, and $d_{X,Y}$ were calculated with a thinned SNP dataset of 51,011 SNPs to reduce bias in these estimates associated with linkage disequilibrium

	Susceptible	Resistant
Sample Size	5	5
P_N	0.738	0.823
MAF	0.225	0.241
H_O	**0.312 (0.310, 0.315)**	**0.347 (0.344, 0.349)**
H_E	**0.329 (0.327, 0.330)**	**0.345 (0.343, 0.347)**
F_{IS}	**0.008 (0.004, 0.012)**	**-0.027 (-0.030, -0.024)**
$d_{X,Y}$	0.180 (0.140, 0.209)	0.182 (0.137, 0.211)

family implicated in Bt resistance in other lepidopteran species (Gahan *et al.* 2001; Morin *et al.* 2003; Wang *et al.* 2005; Zhang *et al.* 2017; Wang *et al.* 2018). This led us to further examine whether *cad-86C* had the potential to serve as target of Cry1Ac selection in *H. zea*.

## Verification of heterozygosity and genetic divergence at cad-86C by Sanger sequencing

Sanger sequencing of 421 and 520 bp within non-coding regions of *cad-86C* revealed 20 and 9 SNPs in GA and 21 and 14 SNPs in GA-R, respectively (n = 22-24 additional individuals per line). The variation in the numbers of SNPs is reflected in the Watterson's theta values (θ_W), which were 4.5 and 2.0 for GA, and 4.8 and 3.2 for GA-R, respectively (Table 2). Within population values of π, however, demonstrated that the allele frequency distributions were different between lines. π was always significantly lower for GA-R than GA according to a permutation test ($P < 0.001$; Table 2; Figures S1 and S2), indicating a reduction in intermediate frequency alleles in GA-R. For the 421 and 520 bp sequences, nucleotide diversity (π) values were 0.016 and 0.006 for GA, and 0.009 and 0.004 for GA-R, respectively. Tajima's D test statistics were negative for GA-R, which could indicate a reduction in genetic diversity by purifying selection or a population contraction. D was positive for GA, which could signal greater genetic diversity than expected due to balancing selection or a recent population expansion. Only one Tajima's D statistic, calculated with the 520 bp GA sequence data were significantly different from 0 at the $\alpha < 0.05$ level, however (Table 2).

Comparison of CAD-86C to other cadherin proteins by sequence alignment

Amino acid sequence identity was 56–84% among five cadherin proteins involved in Bt resistance in other Lepidoptera, but only 17–22% between CAD-86C from *H. zea* and each of these five proteins (Table S6). When we reconstructed the phylogenetic relationships among 14 cadherin proteins, clustering occurred by putative homologs rather than by species (Figure 4). High bootstrap support for the gene clusters demonstrated that CAD-86C was not homologous to other cadherins involved in Bt resistance.

Expression of cad-86C in the H. zea midgut

Quantitative PCR revealed that *cad-86C* was transcribed in the midgut of GA and GA-R larvae, with 1.6 fold higher expression in
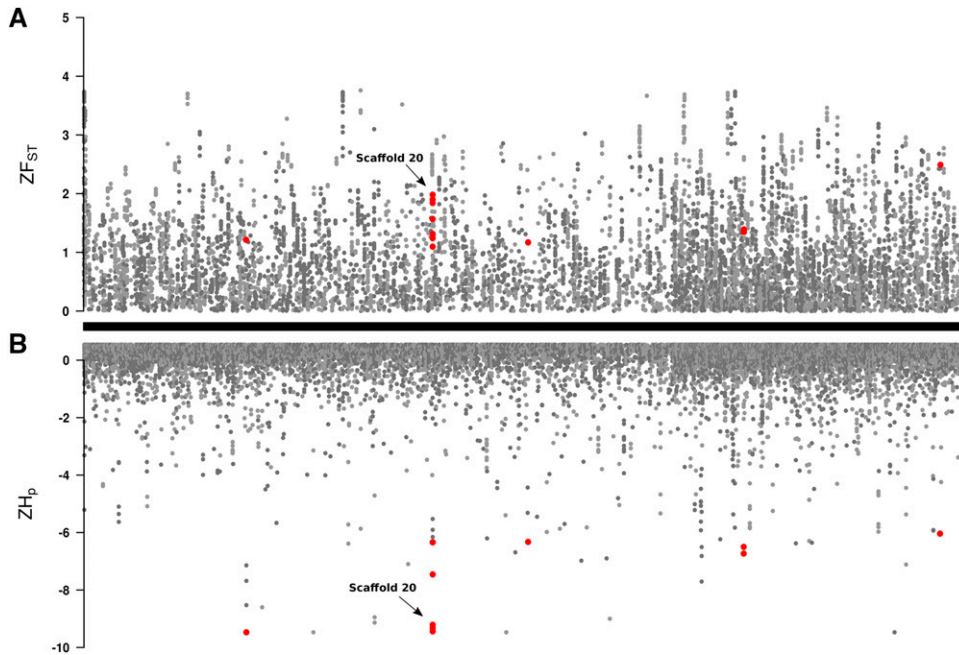
Figure 1 (A) Z-transformed F_{ST} values were used to estimate genetic divergence between GA and GA-R in 40-kb sliding windows with a 20-kb step size along the *H. zea* genome. (B) Z-transformed pooled heterozygosity in the GA-R line for the same 40-kb sliding windows with a 20-kb step size. Each of the 2,975 scaffolds that comprise the *H. zea* reference genome are indicated by the alternating light and dark gray points. Genomic windows identified as potentially under selection are in red.

GA-R (Mann-Whitney *U*-test, *P* = 0.037). Survival to at least 3rd instar was significantly higher for GA-R (61.3%, N = 3840) as compared to GA (1.0%, N = 512) (Fisher's exact test, *P* < 0.0001) reared on Cry1Ac treated diet at the time of midgut dissection.

### Identification of CAD-86C amino acid substitutions

A 5,539 base pair PCR amplicon, which corresponded to the *cad-86C* coding sequence, was produced for most GA and GA-R individuals (Figure S3). Following PacBio sequencing of these amplicons, a total of 1,081,073 PacBio long reads were generated from 30 *H. zea* individuals. After demultiplexing and quality and length filtering, 14,862 reads remained (minimum length = 5000 bp, Table S7), and 22 individuals (11 per line) were used for analysis.

Using 11 individuals from each line (total n = 22), we identified five predicted CAD-86C protein variants (File S4). The distribution of these variant sequences differed significantly between GA and GA-R (Fisher's exact test, *P* = 0.002; Table 3). Variant 1, the full-length reference protein, was the most common variant in both lines, accounting for 86% of the sequences for GA-R (nine homozygotes and one heterozygote), and 50% for GA (five homozygotes and one heterozygote (Tables 4 and S8). Variants 2 and 3, which occurred in only one individual from GA-R, were identical to the reference sequence except for an insertion at 600,850-600,883 bp of scaffold 20. This insertion introduced a premature stop codon, and was expected to yield a truncated protein. Variant 4, which accounted for 41% of the sequences for GA and 5% for GA-R, had a 15 bp deletion (5'- GAT AAT ACT GCA ACA - 3') at position 599,832-599,846 bp, and a SNP at 603,408 bp. The deletion was expected to cause the loss of a 5 amino acid sequence (DNTAT) from cadherin repeat domain five, part of the extracellular protein domain which is proximal to the membrane-spanning region. The SNP was expected to produce a threonine to lysine substitution as compared to the reference (T1706K). Variant 5, which appeared in only one individual from GA, had the same 15 bp deletion present in variant 4, and an extra exon in position 600,522-600,574 bp that would terminate the amino acid sequence, producing a truncated protein.

Upon identification of these variants by midgut cDNA sequencing, we revisited our WGS data. All five GA-R individuals were homozygous for the reference sequence at positions 599,832-599,846 and 603,408, and were expected to produce a DNTAT+T haplotype. By contrast, only one individual from GA was homozygous for this
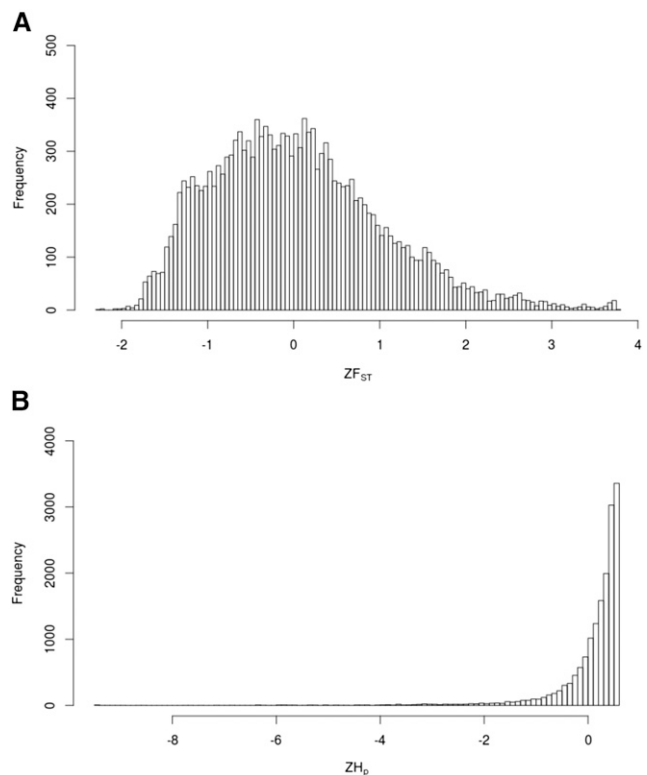
Figure 2 Frequency distributions of (A) Z-transformed F_{ST} values for the comparison of genetic divergence between GA and GA-R, and (B) Z-transformed pooled heterozygosity in the GA-R line.
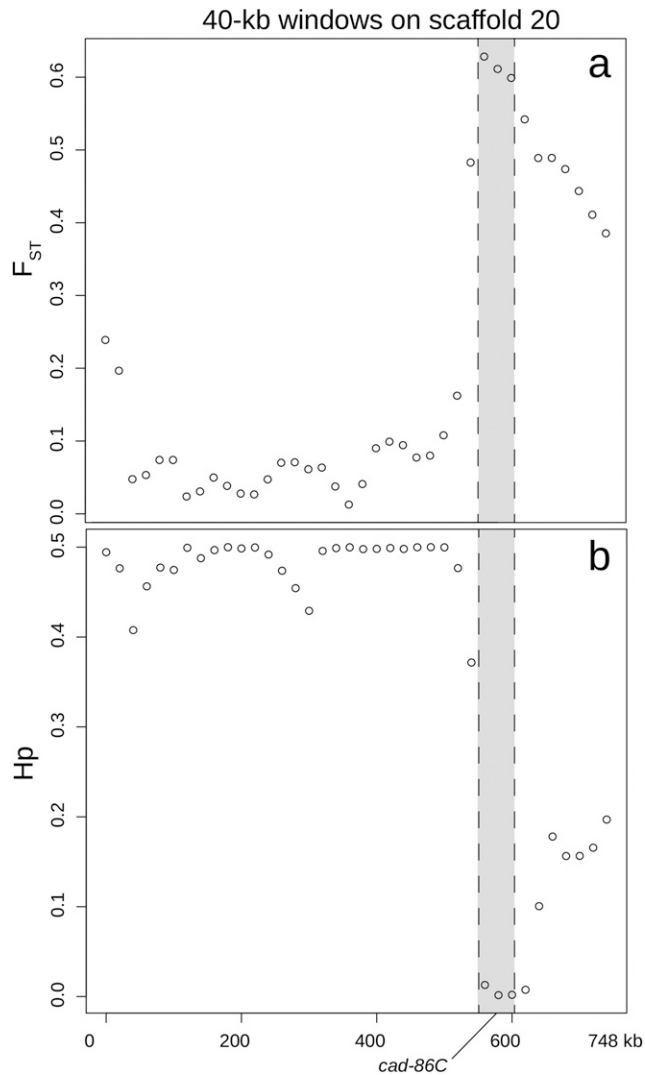
Figure 3 (a) F_{ST} between lines and (b) pooled heterozygosity in the GA-R line along 40-kb sliding windows surrounding the putative selective sweep on scaffold 20. *cad-86C* is in gray.

haplotype. Two GA individuals were homozygous for the 15 bp deletion leading to the loss of the DNTAT amino acid sequence from the expressed protein, and bore the SNP that caused a threonine to lysine substitution at amino acid position 1706. The final two individuals were heterozygous bearing one copy of the reference allele and the alternate allele that produced the DNTAT deletion. This gave a DNTAT + T haplotype frequency of 1 in GA-R, and 0.4 in GA, which was consistent with our PacBio sequencing data.

DISCUSSION

Here we used whole genome sequencing to identify putative genomic targets of Cry1Ac selection in laboratory-reared *H. zea*, an important agricultural insect pest. This approach has been used previously to identify signatures of selection in other laboratory-selected insects (Izutsu *et al.* 2012; Jha *et al.* 2015; Ding *et al.* 2018). Reduced heterozygosity in an otherwise heterozygous selected population, combined with increased genetic divergence between selected and unselected lines, serves as a signal that a region of the genome was under selection. Because of small population size, genetic drift often occurs among laboratory-reared lines (Schoen *et al.* 1998; Fritz *et al.* 2016).

When drift occurs, loss of heterozygosity and random fixation of alleles can lead to heightened genome-wide divergence between lines, which can impede separating effects of drift and selection (Hahn 2019). As recommended by Chandler *et al.* (2013), we backcrossed GA-R to its ancestral line (GA) and reselected for Cry1Ac resistance. To counteract the effects of drift, we crossed the two subsets within each line every second or third generation. We found a slight, but statistically significant, increase in the genome-wide heterozygosity for GA-R (H_O = 0.35) relative to GA (H_O = 0.31), which was not observed in a pair of laboratory-selected Bt resistant (YHD2) and susceptible (YDK) *H. virescens* lines maintained without backcrossing (Fritz *et al.* 2016). Furthermore, genome-wide divergence between lines was lower for GA and GA-R (F_{ST} = 0.23, n = 5 per line) than YHD2 and YDK (F_{ST} = 0.28, n = 43-46 per line). Even so, the genetic divergence between GA and GA-R was slightly higher than that recommended by Pérez-Figueroa (F_{ST} = 0.2) for a genome scan to identify adaptive loci (Pérez-Figueroa *et al.* 2010). Therefore, we quantified the extent of genetic divergence (F_{ST}) between lines for regions of lowest heterozygosity in GA-R.

We detected 5 genomic regions with combined ZHp < -6 and ZF_{ST} values >1.1. These regions contained 12 genes, several of which drive various aspects of animal behavior (Table S5). For example, *sol-1* encodes a product that is required for proper function of an ionotropic glutamate receptor, which regulates synaptic transmission and ultimately locomotory behavior in *Caenorhabditis elegans* (Zheng *et al.* 2004). *Muscle-specific 20*, another potential locomotory target of selection, is expressed in the muscle tissue of insect larvae and is likely involved in actin binding (Ayme-Southgate *et al.* 1989). *Map3k15* is a gene from a family involved in the regulation of other genes and ultimately cellular responses to external environmental stimuli (Treisman 1996). Finally, *cfap* genes are thought to play a role in the motility of cilia, such as those on type-1 sensory neurons in *Drosophila*. Ciliary defects are thought to impact a number of sensory processes (*e.g.*, touch, coordination, taste, olfaction and hearing; Jana *et al.* 2016).

Previous studies indicate that some species of Lepidoptera show distinct behaviors when exposed to plant tissue or diet treated with Cry proteins as compared to those exposed to untreated tissue or diet. For example, increased locomotor activity and ballooning behaviors are thought to be the result of larval toxin detection and avoidance (Benedict *et al.* 1992; Men *et al.* 2005; Prasifka *et al.* 2009; Goldstein *et al.* 2010; Ramalho *et al.* 2014). Furthermore, in binary choice tests, larvae from GA-R preferred to feed on a nutritionally optimal Cry1Ac diet relative to a non-Bt suboptimal diet, while larvae from GA consumed more of the suboptimal non-Bt diet (Orpet *et al.* 2015b). This suggested that detection and consumption of a nutritionally optimal diet supersedes toxin avoidance behaviors in GA-R. Genes involved in sensory and locomotory function may have served as targets of Cry1Ac selection in GA-R, conferring these behavioral changes. Future work on the function of genes linked to animal behavior identified here could provide clues as to how behavioral changes occurred in GA-R as a result of Cry1Ac selection.

Much of our analysis focused on a cadherin gene, *cad-86C*, because disruption of the coding sequence of cadherin proteins or reduced cadherin gene expression confers resistance to Cry1 toxins in other Lepidoptera (Gahan *et al.* 2001; Morin *et al.* 2003; Xu *et al.* 2005; Fabrick *et al.* 2014; Zhang *et al.* 2017; Wang *et al.* 2018). Moreover, this gene had high levels of genetic divergence between lines (F_{ST} = 0.44-0.63) and the second lowest heterozygosity estimates in GA-R in the entire genome (Table S5; Figure 1). Simulations with msms using a realistic population demographic scenario under neutral expectations

Primer Pair	Sequence Length	Line	N	θ_W	π	Tajima's D
Cad_1b	421	GA	48	4.5	0.016	1.5
		GA-R	44	4.8	0.009	−0.74
Cad_2b	520	GA	48	2.0	0.006	2.3*
		GA-R	48	3.2	0.004	−0.86

$P < 0.05$ *

indicated that F_{ST} values > 0.44 were not likely to occur due to drift alone ($P < 0.0001$). Furthermore, our Sanger data indicated that GA-R *cad-86C* sequences always had significantly lower π values than did GA, opposite the overall genome-wide trends in other measures of genetic diversity for these two populations. Tajima's D values also trended in opposite directions, with positive values associated with GA and negative values associated with GA-R. At one 520 bp *cad-86C* sequence, the GA Tajima's D statistic was significantly higher than zero, indicating that there was higher than expected genetic diversity in GA at *cad-86C*, which was also opposite genome-wide trends for these two populations. Taken altogether, these results indicated that distinct evolutionary processes were likely occurring at this gene for the two populations.

The *H. zea* CAD-86C protein sequence showed only 17–21% identity with other cadherin proteins involved in Bt resistance (Table S6). As far as we know, the function of CAD-86C has been described only for embryonic organogenesis in *D. melanogaster* (Lovegrove *et al.* 2006; Fung *et al.* 2008; Schlichting *et al.* 2008). To demonstrate that *cad-86C* has a potential role in Bt resistance, we quantified and compared transcription of this gene in the midguts of GA and GA-R larvae. This was of particular interest because of the role of midgut-expressed cadherins other Lepidoptera (Gahan *et al.* 2001; Morin *et al.* 2003; Wang *et al.* 2005; Zhang *et al.* 2017; Wang *et al.* 2018). The 1.6-fold higher expression of *cad-86C* in GA-R relative to GA was statistically significant, but it seemed unlikely this was the primary cause of the >10-fold difference in resistance between these lines (Orpet *et al.* 2015a). Therefore, further experiments quantified the difference between lines in the frequency of predicted protein sequence variants.

We found that the presence of a 15bp indel, whose *in silico* translation produced an amino acid sequence DNTAT, was at a frequency of 0.5 in GA and 0.95 in GA-R according to PacBio sequencing of 22 individuals (n = 11 per line; Table 3; Figure S4). These frequencies included GA-R individuals with amino acid variants 2 and 3, which also contained the 15bp insertion relative to GA variant 4. The same trend was present in individuals from each line sequenced by Illumina short reads. Under one possible evolutionary scenario, the DNTAT haplotype was present as a standing genetic variant in the original field-collected population from Georgia, and increased in frequency under the Cry1Ac selection regime experienced by GA-R, but remained at lower frequency in the unselected GA.

Interestingly, the DNTAT mutation was found in the cadherin repeat domain that was most proximal to the membrane-spanning region of the protein. Previous investigations of BtR indicated that the repeat domain most proximal to the membrane-spanning protein domain was critical for toxicity, and mutations in this region promoted Cry1Ab resistance (Hua *et al.* 2004). If *cad-86C* were involved in Cry1A toxin binding, a mutation of this nature may reduce toxin binding and activation, ultimately leading to resistance in GA-R. The putative protein-coding changes, in concert with evidence of midgut

gene expression, and the breadth of the selective sweep identified by our whole genome sequencing data analysis, provide compelling evidence for this evolutionary scenario, and suggest that this gene is under selection by Cry1Ac in GA-R.

There are a few caveats to interpretation of our findings, however. Given the nature of our experiments, it is difficult to infer the directionality of the change in allele frequency experienced by GA and GA-R after the lines were established in the laboratory. A second evolutionary scenario, distinct from the one described above, is that the DNTAT mutation was already at high frequency in the moderately resistant founder population when they were collected from Cry1Ab-expressing corn in 2008 (Brévault *et al.* 2013). Due to experimentally imposed Cry1Ac selection, the DNTAT mutation may have been maintained at high frequency in GA-R, but declined in the GA line reared on untreated diet. Many insecticide resistance mutations reduce fitness in the absence of insecticides (Kliot and Ghanim 2012), and Bt resistance mutations are no exception (Gassmann *et al.* 2009; Carrière *et al.* 2018b). The changes in allele frequency described under this scenario could have occurred if the Cry1Ac resistance-conferring mutation was costly to maintain in the absence of selection. Previous work has indicated that the fitness cost of resistance mutations in GA-R is not high enough to cause rapid loss of a resistance phenotype when selection is removed for 1-2 generations (Carrière *et al.* 2018a). GA and GA-R survival and development times were similar on untreated artificial diets with various casein:sucrose ratios, and survivorship, pupal weight and development time did not differ on young non-Bt cotton plants (Carrière *et al.* 2018a 2019). Nonetheless, fitness components were lower in GA-R in the absence of Bt toxins under certain conditions.
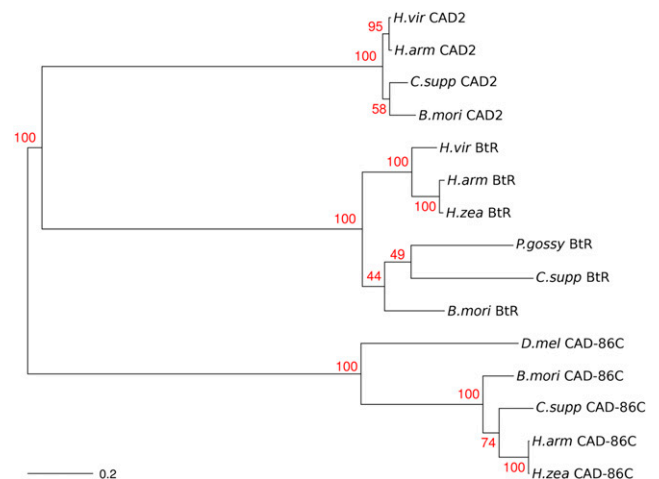
Figure 4 Unrooted neighbor-joining tree indicating the phylogenetic relationships between CAD-86C, CAD2, and BtR. Numbers in red are bootstrap support values (N = 1000) for the tree nodes. A scale bar for genetic distance is in the lower left corner.

■ **Table 3 CAD-86C amino acid sequence variant counts for the 22 *H. zea* individuals from the GA and GA-R lines. Counts represent chromosomes sampled (n = 2 per individual)**

Variants	GA	GA-R
1	11	19
2	0	1
3	0	1
4	9	1
5	2	0

Pupal weight, for example, was generally lower for GA-R than GA on untreated artificial diet (Orpet *et al.* 2015a). Furthermore, survival was lower for GA-R larvae fed upon old non-Bt cotton plants as compared to GA larvae (Carrière *et al.* 2018a). Such fitness costs, if present in the ancestral population of GA and GA-R, could underlie the decline in resistance allele frequency described for GA in this second evolutionary scenario.

Identification of a selective sweep which includes *cad-86C* or any one of the other 11 genes we identified in our study, cannot demonstrate that they are directly involved in Cry1Ac resistance. Instead, some or all of these genomic regions may be under linked or even indirect selection. It is interesting that the DNTAT amino acid sequence is present in *H. armigera* CAD-86C reference sequences used in our analyses, as well as in the reference genome of *H. zea*. This suggests that the mutation is not novel, but was present in the common ancestor of these two species. Furthermore, both *H. armigera* (Tabashnik & Carrière 2017), and the strain of *H. zea* used for genome sequencing (Pearce *et al.* 2017) are generally considered to be susceptible to Cry1A toxins. Therefore, a third evolutionary scenario may be that the DNTAT mutation is not itself the target of selection, but instead is physically linked to an unidentified but nearby target of Cry1A selection. A fourth evolutionary scenario may be that genes in our identified putative sweep regions, including *cad-86C*, allow for recovery of fitness in individuals bearing mutations with otherwise deleterious effects. Perhaps selection for a mutation that conferred resistance in GA-R was followed by selection for mutations that improve the fitness of individuals bearing such a mutation. In any case, our approach cannot quantify the phenotypic impact of individual mutations on a trait itself. Rather, by comparing the genome-wide changes in allele frequency between the two lines, we can detect signals that indicate selection has taken place.

Finally, there are inherent tradeoffs to using laboratory-selected *vs.* field-selected populations to identify genes under selection (Ffrench-Constant 2013). On the one hand, laboratory-selected populations can provide powerful insights into the genetic mechanisms underlying resistance (Gahan *et al.* 2001), and some Bt resistance mutations identified in laboratory-selected populations have been found in field-selected populations (Zhang *et al.* 2012; Jin *et al.* 2018; Mathew *et al.* 2018; Wang *et al.* 2019). On the other hand, some mutations or molecular mechanisms that produce resistant phenotypes in the lab cannot be found in field-collected individuals (Zhang *et al.* 2012; Fabrick *et al.* 2014; Mathew *et al.* 2018), or even other lab-selected strains. As one example, Caccia *et al.* (2012) determined that increased alkaline phosphatase levels in the midguts of laboratory-selected *H. zea* strain AR1 were associated with Cry1Ac resistance. Yet this mechanism did not explain the resistance in GA and GA-R relative to a susceptible strain (Zhang *et al.* 2019). Instead, changes in the expression of midgut proteases were responsible for resistance to Cry1Ac in both GA and GA-R relative to the susceptible strain, but not for the 14-fold increase in Cry1Ac resistance for GA-R relative to GA.

Here, we further examined GA and GA-R to identify the genes under selection by Cry1Ac, and to determine the molecular mechanisms underlying this 14-fold higher Bt resistance in GA-R. We identified several regions of the GA-R genome with signatures of selection, including one that overlapped with a novel midgut-expressed cadherin. Future work should quantify the phenotypic impacts of *cad-86C* variants in this pair of lines, as well as evaluate the importance of these variants to Cry1A resistance in field-selected *H. zea*.

LITERATURE CITED

Ali, M. I., R. G. Luttrell, and S. Y. Young, III, 2006   Susceptibilities of *Helicoverpa zea* and *Heliothis virescens* (Lepidoptera: Noctuidae) populations to Cry1Ac insecticidal protein. J. Econ. Entomol. 99: 164–175. https://doi.org/10.1603/0022-0493(2006)099[0164:SOHZAH]2.0.CO;2

Ayme-Southgate, A., P. Lasko, C. French, and M. L. Pardue, 1989   Characterization of the gene for mp20: a *Drosophila* muscle protein that is not found in asynchronous oscillatory flight muscle. J. Cell Biol. 108: 521–531. https://doi.org/10.1083/jcb.108.2.521

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

Blankenberg, D., G. V. Kuster, N. Coraor, G. Ananda, R. Lazarus *et al.*, 2010   Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. Chapter 19: Unit 19.10.1-21. https://doi.org/10.1002/0471142727.mb1910s89

Benedict, J. H., D. W. Altman, P. F. Umbeck, and D. R. Ring, 1992   Behavior, growth, survival, and plant injury by *Heliothis virescens* (F.)(Lepidoptera: Noctuidae) on transgenic Bt cottons. J. Econ. Entomol. 85: 589–593. https://doi.org/10.1093/jee/85.2.589

Brévault, T., S. Heuberger, M. Zhang, C. Ellers-Kirk, X. Ni *et al.*, 2013   Potential shortfall of pyramided transgenic cotton for insect resistance management. Proc. Natl. Acad. Sci. USA 110: 5806–5811. https://doi.org/10.1073/pnas.1216719110

Brévault, T., B. E. Tabashnik, and Y. Carrière, 2015   A seed mixture increases dominance of resistance to Bt cotton in *Helicoverpa zea*. Sci. Rep. 5: 9807. https://doi.org/10.1038/srep09807

Bushnell, B., 2014   *BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab*, LBNL, Berkeley, CA.

Caccia, S., W. J. Moar, J. Chandrashekhar, C. Oppert, K. J. Anilkumar *et al.*, 2012   Association of Cry1Ac toxin resistance in *Helicoverpa zea* (Boddie) with increased alkaline phosphatase levels in the midgut lumen. Appl. Environ. Microbiol. 78: 5690–5698. https://doi.org/10.1128/AEM.00523-12

Carrière, Y., N. Crickmore, and B. E. Tabashnik, 2015   Optimizing pyramided transgenic Bt crops for sustainable pest management. Nat. Biotechnol. 33: 161–168. https://doi.org/10.1038/nbt.3099

Carrière, Y., B. Degain, G. C. Unnithan, V. Harpold, S. Heuberger *et al.*, 2018a   Effects of seasonal changes in cotton plants on the evolution of resistance to pyramided cotton producing the Bt toxins Cry1Ac and Cry1F by *Helicoverpa zea*. Pest Manag. Sci. 74: 627–637. https://doi.org/10.1002/ps.4746

Carrière, Y., J. L. Williams, D. W. Crowder, and B. E. Tabashnik, 2018b Genotype-specific fitness cost of resistance to Bt toxin Cry1Ac in pink bollworm. Pest Manag. Sci. 74: 2496–2503. https://doi.org/10.1002/ps.4928

Carrière, Y., B. DeGain, G. C. Unnithan, V. S. Harpold, X. Li *et al.*, 2019 Seasonal declines in Cry1Ac and Cry2Ab concentration in maturing cotton favor faster evolution of resistance to pyramided Bt Cotton in *Helicoverpa zea* (Lepidoptera: Noctuidae). J. Econ. Entomol. 112: 2907–2914. https://doi.org/10.1093/jee/toz236

Chandler, C. H., S. Chari, and I. Dworkin, 2013 Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet. 29: 358–366. https://doi.org/10.1016/j.tig.2013.01.009

Clark, T. L., J. E. Foster, S. T. Kamble, and E. A. Heinrichs, 2000 Comparison of Bt (*Bacillus thuringiensis* Berliner) maize and conventional measures for control of the European corn borer (Lepidoptera: Crambidae). J. Entomol. Sci. 35: 118–128. https://doi.org/10.18474/0749-8004-35.2.118

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Del Fabbro, C., S. Scalabrin, M. Morgante, and F. M. Giorgi, 2013 An extensive evaluation of read trimming effects on Illumina NGS data analysis. PLoS One 8: e85024. https://doi.org/10.1371/journal.pone.0085024

Ding, D., G. Liu, L. Hou, W. Gui, B. Chen *et al.*, 2018 Genetic variation in PTPN1 contributes to metabolic adaptation to high-altitude hypoxia in Tibetan migratory locusts. Nat. Commun. 9: 4991. https://doi.org/10.1038/s41467-018-07529-8

Dively, G. P., P. D. Venugopal, and C. Finkenbinder, 2016 Field-evolved resistance in corn earworm to Cry proteins expressed by transgenic sweet corn. PLoS One 11: e0169115. https://doi.org/10.1371/journal.pone.0169115.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797. https://doi.org/10.1093/nar/gkh340

Ewing, G., and J. Hermisson, 2010 MSMS: a coalescent simulation program including recombination, demographic structure, and selection at a single locus. Bioinformatics 26: 2064–2065. https://doi.org/10.1093/bioinformatics/btq322

Fabrick, J. A., J. Ponnuraj, A. Singh, R. K. Tanwar, G. C. Unnithan *et al.*, 2014 Alternative splicing and highly variable cadherin transcripts associated with field-evolved resistance of pink bollworm to Bt cotton in India. PLoS One 9: e97900. https://doi.org/10.1371/journal.pone.0097900

Ffrench-Constant, R. H., 2013 The molecular genetics of insecticide resistance. Genetics 194: 807–815. https://doi.org/10.1534/genetics.112.141895

Fritz, M. L., S. Paa, J. Baltzegar, and F. Gould, 2016 Application of a dense genetic map for assessment f genomic responses to selection and inbreeding in *Heliothis virescens*. Insect Mol. Biol. 25: 385–400. https://doi.org/10.1111/imb.12234

Fung, S., F. Wang, M. Chase, D. Godt, and V. Hartenstein, 2008 Expression profile of the cadherin family in the developing *Drosophila* brain. J. Comp. Neurol. 506: 469–488.

Gahan, L. J., F. Gould, and D. G. Heckel, 2001 Identification of a gene associated with Bt resistance in *Heliothis virescens*. Sci. 293: 857–860. https://doi.org/10.1126/science.1060949

Gassmann, A. J., Y. Carrière, and B. E. Tabashnik, 2009 Fitness costs of insect resistance to *Bacillus thuringiensis*. Annu. Rev. Entomol. 54: 147–163. https://doi.org/10.1146/annurev.ento.54.110807.090518

Goldstein, J. A., C. E. Mason, and J. Pesek, 2010 Dispersal and movement behavior of neonate European corn borer (Lepidoptera: Crambidae) on non-Bt and transgenic Bt corn. J. Econ. Entomol. 103: 331–339. https://doi.org/10.1603/EC09304

Goudet, J., 2005 Hierfstat, a package for R to compute and test hierarchical F-statistics. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Gould, F., 1998 Sustainability of transgenic insecticidal cultivars: integrating pest genetics and ecology. Annu. Rev. Entomol. 43: 701–726. https://doi.org/10.1146/annurev.ento.43.1.701

Gould, F., Z. S. Brown, and J. Kuzma, 2018 Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance? Sci. 360: 728–732. https://doi.org/10.1126/science.aar3780

Hahn, M. W., 2019 *Molecular Population Genetics*, Ed. 1st. Oxford University Press, Oxford.

Hua, G., J. L. Jurat-Fuentes, and M. J. Adang, 2004 Bt-R1a Extracellular Cadherin Repeat 12 Mediates *Bacillus thuringiensis* Cry1Ab Binding and Cytotoxicity. J. Biol. Chem. 279: 28051–28056. https://doi.org/10.1074/jbc.M400237200

Izutsu, M., J. Zhou, Y. Sugiyama, O. Nishimura, T. Aizu *et al.*, 2012 Genome features of "Dark-fly", a Drosophila line reared long-term in a dark environment. PLoS One 7: e33288. https://doi.org/10.1371/journal.pone.0033288

Jana, S. C., M. Bettencourt-Dias, B. Durand, and T. L. Megraw, 2016 *Drosophila melanogaster* as a model for basal body research. Cilia 5: 22. https://doi.org/10.1186/s13630-016-0041-5

Jha, A. R., C. M. Miles, N. R. Lippert, C. D. Brown, K. P. White *et al.*, 2015 Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in *Drosophila melanogaster*. Mol. Biol. Evol. 32: 2616–2632. https://doi.org/10.1093/molbev/msv136

Jin, L., J. Wang, F. Guan, J. Zhang, S. Yu *et al.*, 2018 Dominant point mutation in a tetraspanin gene associated with field-evolved resistance of cotton bollworm to transgenic Bt cotton. Proc. Natl. Acad. Sci. USA 115: 11760–11765. https://doi.org/10.1073/pnas.1812138115

Kaur, G., J. Guo, S. Brown, G. P. Head, P. A. Price *et al.*, 2019 Field-evolved resistance of *Helicoverpa zea* (Boddie) to transgenic maize expressing pyramided Cry1A.105/Cry2Ab2 proteins in northeast Louisiana. J. Inv. Pathol. 163: 11–20. https://doi.org/10.1016/j.jip.2019.02.007

Kliot, A., and M. Ghanim, 2012 Fitness costs associated with insecticide resistance. Pest Manag. Sci. 68: 1431–1437. https://doi.org/10.1002/ps.3395

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman *et al.*, 2017 Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27: 722–736. https://doi.org/10.1101/gr.215087.116

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25. https://doi.org/10.1186/gb-2009-10-3-r25

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Liu, C., Y. Li, Y. Gao, C. Ning, and K. Wu, 2010 Cotton bollworm resistance to Bt transgenic cotton: a case analysis. Sci. China Life Sci. 53: 934–941. https://doi.org/10.1007/s11427-010-4045-x

Lovegrove, B., S. Somoes, M. L. Rivas, S. Sotillos, K. Johnson *et al.*, 2006 Coordinated control of cell adhesion, polarity, and cytoskeleton underlies Hox-induced organogenesis in *Drosophila*. Curr. Biol. 16: 2206–2216. https://doi.org/10.1016/j.cub.2006.09.029

Madeira, F., Y.M. Park, J. Lee, N. Buso, T. Gur, *et al.*, 2019 The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47: W636–W641. https://doi.org/10.1093/nar/gkz268

Mathew, L. G., J. Ponnuraj, B. Mallappa, L. R. Chowdary, J. Zhang *et al.*, 2018 ABC transporter mis-splicing associated with resistance to Bt toxin Cry2Ab in laboratory- and field-selected pink bollworm. Sci. Rep. 8: 13531. https://doi.org/10.1038/s41598-018-31840-5

Men, X., F. Ge, E. N. Yardim, and M. N. Parajulee, 2005 Behavioral response of *Helicoverpa armigera* (Lepidoptera: Noctuidae) to cotton with and without expression of the CrylAc δ-endotoxin protein of *Bacillus thuringiensis* Berliner. J. Insect Behav. 18: 33–50. https://doi.org/10.1007/s10905-005-9345-9

Morin, S., R. W. Biggs, M. S. Sisterson, L. Shriver, C. Ellers-Kirk *et al.*, 2003 Three cadherin alleles associated with resistance to *Bacillus thuringiensis* in pink bollworm. Proc. Natl. Acad. Sci. USA 100: 5004–5009. https://doi.org/10.1073/pnas.0831036100

Nickerson, D. A., V. O. Tobe, and S. L. Taylor, 1997 PolyPhred: automating the detection and genotyping of single nucleotide substitutions using

fluorescence-based resequencing. Nucleic Acids Res. 25: 2745–2751. https://doi.org/10.1093/nar/25.14.2745

Notredame, C., D. G. Higgins, and J. Heringa, 2000 T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302: 205–217. https://doi.org/10.1006/jmbi.2000.4042

Orpet, R. J., B. A. Degain, G. C. Unnithan, K. L. Welch, B. E. Tabashnik *et al.*, 2015a Effects of dietary protein to carbohydrate ratio on Bt toxicity and fitness costs of resistance in *Helicoverpa zea*. Entomol. Exp. Appl. 156: 28–36. https://doi.org/10.1111/eea.12308

Orpet, R. J., B. A. DeGain, B. E. Tabashnik, and Y. Carrière, 2015b Balancing Bt toxin avoidance and nutrient intake by *Helicoverpa zea* (Lepidoptera: Noctuidae) larvae. J. Econ. Entomol. 108: 2581–2588. https://doi.org/10.1093/jee/tov226

Paradis, E., 2010 pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics 26: 419–420. https://doi.org/10.1093/bioinformatics/btp696

Pardo-López, L., M. Soberón, and A. Bravo, 2013 *Bacillus thuringiensis* insecticidal three-domain Cry toxins: mode of action, insect resistance and consequences for crop protection. FEMS Microbiol. Rev. 37: 3–22. https://doi.org/10.1111/j.1574-6976.2012.00341.x

Patel, R. K., and M. Jain, 2012 NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7: e30619. https://doi.org/10.1371/journal.pone.0030619

Pearce, S. L., D. F. Clarke, P. D. East, S. Elfekih, K. H. Gordon *et al.*, 2017 Genomic innovations, transcriptional plasticity and gene loss underlying the evolution and divergence of two highly polyphagous and invasive *Helicoverpa* pest species. BMC Biol. 15: 63. https://doi.org/10.1186/s12915-017-0402-6

Pembleton, L. W., N. O. I. Cogan, and J. W. Forster, 2013 StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol. Ecol. Resour. 13: 946–952. https://doi.org/10.1111/1755-0998.12129

Pérez-Figueroa, A., M. J. García-Pereira, M. Saura, E. Rolán-Alvarez, and A. Caballero, 2010 Comparing three different methods to detect selective loci using dominant markers. J. Evol. Biol. 23: 2267–2276. https://doi.org/10.1111/j.1420-9101.2010.02093.x

Peterson, B., C. C. Bezuidenhout, and J. Van den Berg, 2017 An overview of mechanisms of Cry toxin resistance in lepidopteran insects. J. Econ. Entomol. 110: 362–377.

Prasifka, J. R., R. L. Hellmich, D. V. Sumerford, and B. D. Siegfried, 2009 *Bacillus thuringiensis* resistance influences European corn borer (Lepidoptera: Crambidae) larval behavior after exposure to Cry1Ab. J. Econ. Entomol. 102: 781–787. https://doi.org/10.1603/029.102.0240

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira *et al.*, 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81: 559–575. https://doi.org/10.1086/519795

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Development Core Team, 2008 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, URL http://www.R-project.org.

Ramalho, F. S., J. K. Pachú, A. C. Lira, J. B. Malaquias, J. C. Zanuncio *et al.*, 2014 Feeding and dispersal behavior of the cotton leafworm, *Alabama argillacea* (Hübner)(Lepidoptera: Noctuidae), on Bt and non-Bt cotton: implications for evolution and resistance management. PLoS One 9: e111588. https://doi.org/10.1371/journal.pone.0111588

Reisig, D. D., A. S. Huseth, J. S. Bacheler, M. A. Aghaee, L. Braswell *et al.*, 2018 Long-term empirical and observational evidence of practical *Helicoverpa zea* resistance to cotton with pyramided Bt toxins. J. Econ. Entomol. 111: 1824–1833. https://doi.org/10.1093/jee/toy106

Robinson, J. T., H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander *et al.*, 2011 Integrative genomics viewer. Nat. Biotechnol. 29: 24–26. https://doi.org/10.1038/nbt.1754

Roush, R. T., 1997 Bt-transgenic crops: just another pretty insecticide or a chance for a new start in resistance management? Pestic. Sci. 51: 328–334.

https://doi.org/10.1002/(SICI)1096-9063(199711)51:3<328::AID-PS650>3.0.CO;2-B

Rubin, C. J., M. C. Zody, J. Eriksson, J. R. Meadows, E. Sherwood *et al.*, 2010 Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 464: 587–591. https://doi.org/10.1038/nature08832

Ruijter, J. M., C. Ramakers, W. M. Hoogaars, Y. Karlen, O. Bakker *et al.*, 2009 Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. 37: e45. https://doi.org/10.1093/nar/gkp045

Schlichting, K., and C. Dahmann, 2008 Hedgehog and Dpp signaling induce cadherin Cad86C expression in the morphogenetic furrow during *Drosophila* eye development. Mech. Dev. 125(8): 712–728.

Schliep, K. P., 2011 phangorn: phylogenetic analysis in R. Bioinformatics 27: 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schlichting, K., and C. Dahmann, 2008 Hedgehog and Dpp signaling induce cadherin Cad86C expression in the morphogenetic furrow during *Drosophila* eye development. Mech. Dev. 125: 712–728.

Schoen, D. J., J. L. David, and T. M. Bataillon, 1998 Deleterious mutation accumulation and the regeneration of genetic resources. Proc. Natl. Acad. Sci. USA 95: 394–399. https://doi.org/10.1073/pnas.95.1.394

Sivasupramaniam, S., W. J. Moar, L. G. Ruschke, J. A. Osborn, C. Jiang *et al.*, 2008 Toxicity and characterization of cotton expressing *Bacillus thuringiensis* Cry1Ac and Cry2Ab2 proteins for control of lepidopteran pests. J. Econ. Entomol. 101: 546–554. https://www.ncbi.nlm.nih.gov/pubmed/18459423

Stone, T. B., and S. R. Sims, 1993 Geographic susceptibility of *Heliothis virescens* and *Helicoverpa zea* (Lepidoptera: Noctuidae) to *Bacillus thuringiensis*. J. Econ. Entomol. 86: 989–994. https://doi.org/10.1093/jee/86.4.989

Tabashnik, B. E., A. L. Patin, T. J. Dennehy, Y.-B. Liu, Y. Carrière *et al.*, 2000 Frequency of resistance to *Bacillus thuringiensis* in field populations of pink bollworm. Proc. Natl. Acad. Sci. USA 97: 12980–12984. https://doi.org/10.1073/pnas.97.24.12980

Tabashnik, B. E., F. Gould, and Y. Carrière, 2004 Delaying evolution of insect resistance to transgenic crops by decreasing dominance and heritability. J. Evol. Biol. 17: 904–912. https://doi.org/10.1111/j.1420-9101.2004.00695.x

Tabashnik, B. E., and Y. Carrière, 2017 Surge in insect resistance to transgenic crops and prospects for sustainability. Nat. Biotechnol. 35: 926–935. https://doi.org/10.1038/nbt.3974

Tabashnik, B. E., and Y. Carrière, 2019 Global patterns of resistance to Bt crops highlighting pink bollworm in the United States, China and India. J. Econ. Entomol. 112: 2513–2523. https://doi.org/10.1093/jee/toz173

Treisman, R., 1996 Regulation of transcription by MAP kinase cascades. Curr. Opin. Cell Biol. 8: 205–215. https://doi.org/10.1016/S0955-0674(96)80067-6

U.S. Environmental Protection Agency, 2018 Resistance of Lepidopteran Pests to *Bacillus thuringiensis* (Bt) Plant Incorporated Protectants (PIPs) in The United States. SAP Minutes No. 2018–06. FIFRA Scientific Advisory Panel Meeting, 17–July 19, 2018 Arlington, VA.

Wang, G., K. Wu, G. Liang, and Y. Guo, 2005 Gene cloning and expression of cadherin in midgut of *Helicoverpa armigera* and its Cry1A binding region. Sci. China C Life Sci. 48: 346–356. https://doi.org/10.1360/03yc0273

Wang, L., Y. Ma, P. Wan, K. Liu, Y. Xiao *et al.*, 2018 Resistance to *Bacillus thuringiensis* linked with a cadherin transmembrane mutation affecting cellular trafficking in pink bollworm from China. Insect Biochem. Mol. Biol. 94: 28–35. https://doi.org/10.1016/j.ibmb.2018.01.004

Wang, J., D. Xu, L. Wang, S. Cong, P. Wan *et al.*, 2019 Bt resistance alleles in field populations of pink bollworm from China: Similarities with the United States and decreased frequency from 2012 to 2015. Pest Manag. Sci. https://doi.org/10.1002/ps.5541

Welch, K. L., G. C. Unnithan, B. A. Degain, J. Wei, J. Zhang *et al.*, 2015 Cross-resistance between to toxins used in pyramided Bt crops and resistance to Bt sprays in *Helicoverpa zea*. J. Invertebr. Pathol. 132: 149–156. https://doi.org/10.1016/j.jip.2015.10.003

Whelan, S., and N. Goldman, 2001 A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18: 691–699. https://doi.org/10.1093/oxfordjournals.molbev.a003851

Xu, X., L. Yu, and Y. Wu, 2005 Disruption of a cadherin gene associated with resistance to Cry1Ac-endotoxin of *Bacillus thuringiensis* in *Helicoverpa armigera*. Appl. Environ. Microbiol. 71: 948–954. https://doi.org/10.1128/AEM.71.2.948-954.2005

Yang, F., J. C. González, J. Williams, D. C. Cook, and R. T. Gilreath, 2019 Occurrence and ear damage of *Helicoverpa zea* on transgenic *Bacillus thuringiensis* maize in the field in Texas, US and its susceptibility to Vip3A protein. Toxins (Basel) 11: 102. https://doi.org/10.3390/toxins11020102

Zhang, H., W. Tian, J. Zhao, L. Jin, J. Yang *et al.*, 2012 Diverse genetic basis of field-evolved resistance to Bt cotton in cotton bollworm from China.

Proc. Natl. Acad. Sci. USA 109: 10275–10280. https://doi.org/10.1073/pnas.1200156109

Zhang, Z., X. Teng, W. Ma, and F. Li, 2017 Knockdown of two cadherin genes confers resistance to Cry2A and Cry1C in *Chilo suppressalis*. Sci. Rep. 7: 5992. https://doi.org/10.1038/s41598-017-05110-9

Zhang, M., J. Wei, X. Ni, J. Zhang, J. L. Jurat-Fuentes *et al.*, 2019 Decreased Cry1Ac activation by midgut proteases associated with Cry1Ac resistance in *Helicoverpa zea*. Pest Manag. Sci. 75: 1099–1106. https://doi.org/10.1002/ps.5224

Zheng, Y., J. E. Mellem, P. J. Brockie, D. M. Madsen, and A. V. Maricq, 2004 SOL-1 is a CUB-domain protein required for GLR-1 glutamate receptor function in *C. elegans*. Nature 427: 451–457. https://doi.org/10.1038/nature02244

Communicating editor: A. Wong