Corresponding Author:
Dr. Megan L. Fritz
Department of Entomology
University of Maryland
4291 Field House Dr.
Plant Sciences Bldg. Rm. 4112
College Park, MD 20742
mfritz13@umd.edu

Mutations in a novel cadherin gene associated with Bt resistance in *Helicoverpa zea*

Megan L. Fritz[1§*], Schyler O. Nunziata[1*], Rong Guo[1], Bruce E. Tabashnik[2], and Yves Carrière[2]

[1]Department of Entomology, University of Maryland, College Park, MD 20742 USA

[2]Department of Entomology, University of Arizona, Tucson, AZ 85721 USA

[§]Correspondence: mfritz13@umd.edu

*Authors contributed equally to the project

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 Cry proteins shape the genome of *H. zea*. We scanned the genomes of Cry1Ac-selected and unselected *H. zea* lines, and identified eleven genes on six scaffolds that showed evidence of selection by Cry1Ac, 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 target of Cry1Ac selection in *H. zea*. Future work should investigate phenotypic effects of these nonsynonymous substitutions and their impact on phenotypic resistance in field populations.

Keywords

Cadherin, Cry1Ac, genome scanning, resistance, selection.

Introduction

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-Lopez 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; Gould et al., 2018; USEPA, 2018; Tabashnik and Carrière, *in press*).

Helicoverpa zea is one lepidopteran species targeted by Bt crops with low 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 molecular 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-Lopez 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 six 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 between the selected and unselected *H. zea* lines. Finally, we discuss the possible functional roles of this gene in field-evolved *H. zea* resistance to Cry proteins.

Methods

Insect material

We conducted our screen for candidate genes associated with Bt resistance in two laboratory-reared lines of *H. zea* with differing susceptibility to Bt toxin. The lines were

85 founded with 180 larvae collected in Georgia from Cry1Ab corn in 2008 (Brévault et al., 2013).
86 F_1 progeny from field-collected individuals gave rise to two lines, GA and GA-R, which were
87 reared on wheat-germ diet and were unexposed to Bt toxins or selected for resistance to Cry1Ac,
88 respectively. From 2008 to 2012, each line was reared with *ca*. 900 individuals per generation.
89 Between 2008 and 2010, GA-R was selected with Cry1Ac nine times as described in Brévault et
90 al. (2013), which yielded 10-fold resistance to Cry1Ac in GA-R relative to GA and significantly
91 higher survival on Cry1Ac and Cry1Ac + Cry2Ab cotton in GA-R than GA (Brévault et al.,
92 2013). From 2010 to 2012, GA-R was selected 10 times with Cry1Ac. In 2012, GA-R was
93 crossed to GA to generate a new GA-R line, and the new GA-R and original GA lines were split
94 into two subsets, each reared with ca. 600 individuals per generation (Orpet et al., 2015a). The
95 two subsets of each line were crossed every second or third generation to generate two new
96 subsets (Orpet et al., 2015a). The new GA-R was selected for seven consecutive generations with
97 Cry1Ac, which yielded 14-fold resistance to Cry1Ac in GA-R relative to GA (Orpet et al.,
98 2015a) and significantly higher survival on Cry1Ac + Cry2Ab cotton in GA-R than GA
99 (Carrière, unpubl. data). Between 2012 and 2016, GA-R was selected 25 times with Cry1Ac,
100 using methods described in Carrière et al. (2018a). Male pupae sampled for genomic analysis
101 were from generations F52 for GA-R (n = 5) and F72 for GA (n = 5) reared in October 2016.
102

103 *DNA isolation and whole genome sequencing*

104 We extracted whole genomic DNA from 5 individuals per line using a Qiagen® DNEasy
105 Blood and Tissue Kit (Qiagen, Inc., Valencia, CA, USA), following a modified version of the
106 mouse-tail protocol. Genomic DNA was submitted to the North Carolina State University
107 Genomic Sciences Laboratory, where it was prepared for sequencing using an Illumina Truseq

LT library prep kit (Illumina, Inc. San Diego, CA). We individually barcoded DNA samples from each individual, after which all DNA samples were pooled for sequencing. We sequenced the prepared pooled library on an Illumina NextSeq500 at 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 et al., 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) using 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 observed (H_O) and expected (H_E) heterozygosity, the proportion of polymorphic SNPs (P_N), and average MAF with PLINK v.1.07 (Purcell et al., 2007). Inbreeding coefficients (F_{IS}) were estimated with VCFtools v.0.1.15. To estimate the proportion of shared alleles within lines we calculated D_{ST} in PLINK, with pairwise genetic distance (D) calculated as $D = 1\text{-}D_{ST}$.

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 ($ZHp < -6$). We used F_{ST} estimates to identify statistically significant genetic divergence between GA and GA-R. These F_{ST} values were calculated two different ways. First, we calculated average F_{ST} corresponding to the sliding window averaged ZHp values, using

PLINK, and then Z-transformed the average F_{ST} values. Next, we calculated sliding window averaged F_{ST} values using the pF_{ST} function in vcflib (https://github.com/vcflib). F_{ST} estimates were calculated over a sliding window size of 5-kb with a 1-kb step, and 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. Regions of the genome with both ZHp < -6 and ZF_{ST} estimates > 1, and an FDR-corrected F_{ST} p-value < 0.05 were considered under selection. In the region with strongest evidence of a selective sweep based on the criteria stated above, 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 specifically 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 95ºC for 3 minutes, followed by 30 cycles of 95ºC for 30 seconds, 60ºC for 30 seconds, 72ºC for 30 seconds and a final elongation step at 72ºC for 1 minute on Bio-Rad T100 Thermal Cycler (Bio-Rad Laboratories, Inc. Hercules, CA, USA). PCR products were cleaned using an ExoAP reaction (see Supplemental Methods) 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) were calculated using the pegas package (v. 0.11; Paradis, 2010) for R (v. 3.5.3; R Development Core Team, 2008). A permutation test, which was custom-coded in R, was used to identify statistically significant differences in π for GA and GA-R at each 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 orthologues 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 Supplementary Data File 1. BtR

orthologues 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 et al., 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.

Cad-86C *expression in the* H. zea *midgut*

We reasoned that *cad-86C* should be expressed in the larval midgut to be involved in Bt resistance. To verify midgut expression, we performed reverse transcriptase quantitative PCR (RT-qPCR) on samples from GA and GA-R. We dissected midguts from the F62 generation of GA-R and F81 generation of GA reared in October 2017. GA-R had been selected for resistance to Cry1Ac six times 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 a concentration of 40 µg/cm2 of diet (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/cm^2 of diet. Seven days after these bioassays were initiated, third instar (or larger) GA-R larvae were considered resistant while first instar GA larvae were considered susceptible to Cry1Ac. For GA and GA-R, first and third instar (or larger) larvae were respectively 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^{TM} 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 50ºC for 2 minutes, followed by denaturation at 95ºC for 2 minutes, and then 40 cycles at 95ºC for 15 seconds, 60ºC for 1 minute 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 mastermix 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 95ºC for 3 minutes, followed by 30 cycles of 95ºC for 30 seconds, 60ºC for 30 seconds, 72ºC for 3:40 min and a final elongation step at 72ºC for 1 minute 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 et al., 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.

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 variant markers 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).

290

291 *Genetic Diversity and Selective Sweeps*

292 Genome-wide average values of genetic diversity, observed heterozygosity (H_O), and
293 pairwise genetic distance (D) were similar for GA and GA-R (Table 1). However, GA-R had a
294 lower inbreeding coefficient (F_{IS}). Overall, the two lines were moderately diverged from each
295 other with a genome-wide F_{ST} value of 0.23. Based on heterozygosity estimates among GA-R
296 individuals, a total of 38 genomic windows across 24 scaffolds were identified as putative
297 regions under selection (Figure 1). These candidate regions corresponded to 41 genomic
298 annotations located either upstream or downstream of the identified window (Table S5). The
299 right skewed distribution of ZHp values indicated a high degree of heterozygosity across most
300 40-kb windows (Figure 1d) throughout the genome. We then examined the Z-transformed F_{ST}
301 values across these 40-kb genomic windows, but no 40-kb window had $ZF_{ST} > 6$, which we
302 expected given the moderate genome-wide divergence between lines (Table 1). The maximum
303 ZF_{ST} was 3.79 and the ZF_{ST} distribution was right skewed (Figure 1c), indicating that there were
304 a number of genomic regions with moderate to high genetic divergence between lines (Figure
305 1a).

306 When we parsed genomic regions displaying both low Hp (ZHp < -6), high broad-
307 window F_{ST} ($ZF_{ST} > 1$), and statistically significant narrow-window F_{ST} ($p < 0.05$), we identified
308 six gene-containing regions with putative selective sweeps (Table S5). These regions were
309 found on six different scaffolds and contained eleven predicted genes: *sol-1* (suppressor of
310 lurcher-1*)*, *cad-86C* (cadherin-86C), *map3k15* (mitogen-activated protein kinase kinase kinase
311 15), *msp20 (*muscle-specific protein 20*)*, *not1* (CCR4-Not complex subunit 1), *allfused 10*, *cpr47*
312 (cuticular RR1 motif 47), *cgrrf1* (cell growth regulator with RING finger domain 1), *itf46*

(intraflagellar transport 46)**,** *cfap100* (cilia and flagellar associated protein 100), and an uncharacterized protein. The F_{ST} values for the 40kb regions containing these genes ranged from 0.44 to 0.72, whereas the genome-wide average level of divergence was 0.23. 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 2), a member of a gene 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.

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 homologues rather than by species (Figure 3). High bootstrap support for the gene clusters demonstrated that CAD-86C was not homologous to other cadherins involved in Bt resistance.

Cad-86C *expression in* 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 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) than GA (1.0%, N = 512) (Fisher's exact test, $p < 0.0001$) 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 (Figure 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. By contrast, only one individual from GA was homozygous for the reference. 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 references 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 population (GA) and reselected for Cry1Ac resistance, while crossing the two subsets of GA and GA-R every second or third generation to counteract the effects of drift. This led to similar genome-wide observed heterozygosity within each line (0.31 and 0.35 for GA and GA-R, respectively), 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 Perez-Figueroa (F_{ST} = 0.2) for a genome scan to identify adaptive loci (Peréz-Figueroa et al., 2010). Therefore, we used both reduced heterozygosity and statistically significant genetic divergence (F_{ST}) calculated across both broad and narrow sliding genomic windows, to identify genomic regions under selection.

We detected 6 genomic regions with statistically significant evidence of a selective sweep. These genomic regions contained 11 genes, several of which drive various aspects of animal behavior (Table S5). For example, *sol-1* encodes 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 locomotory 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 form GA consumed more of the suboptimal non-Bt diet (Orpet et al., 2015b). This suggests 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 the 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.

We focused on a cadherin gene, *cad-86C,* in part 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 the second highest level of genetic divergence between lines and second lowest heterozygosity estimates in GA-R in the entire genome (Table S5; Fig. 1). The *H. zea* CAD-86C protein sequence showed only 17-21% identity with other cadherin proteins involved in Bt resistance (this according to 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).

We demonstrated that *cad-86C* is transcribed in the midgut of GA and GA-R larvae, which is consistent with a potential role in Bt resistance. 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 amino acid sequence DNTAT, rose from 0.5 in GA to 0.95 in GA-R (Table 3; Figure S4). These frequencies include GA-R individuals with amino acid variants 2 and 3, which also include the 15bp insertion relative to GA variant 4. The same trend was observed in our whole genome sequencing data. The presence of this insertion in the GA line indicated that selection for this mutation in GA-R took place from standing genetic variation in the original field-collected population from Georgia. Interestingly, the 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). A second SNP causing an amino acid substitution appeared to be in strong LD with the 15 bp indel, and the threonine-producing variant rose from 0.55 in GA to 0.95 in GA-R. Putative protein-coding changes, in concert with evidence of midgut gene expression, and the breadth of the putative selective sweep identified by our whole genome sequencing data analysis, provide compelling evidence that this gene is under selection by Cry1Ac in GA-R.

There are a few caveats to interpretation of our findings, however. The first is that identification of a selective sweep at any one of these genes, including *cad-86C*, cannot demonstrate that they are directly involved in Cry1Ac resistance. Instead, it remains possible that some or all of these genomic regions may be under indirect selection. For example, genes in our putative sweep regions may allow for recovery of fitness in individuals bearing mutations with otherwise deleterious effects. Many insecticide resistance mutations are thought to negatively impact fitness in the absence of insecticide pressure (Kliot and Ghanim, 2012), and Bt resistance mutations are no exception (Gassmann et al., 2009; Carrière et al., 2018b). 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.

Secondly, our approach cannot quantify the impact of individual mutations on the trait itself. Rather, by comparing the genomes of multiple populations, it allows us to detect signals that show selection has taken place. It is interesting that the 15bp insertion, which produces a DNTAT amino acid sequence in GA-R, is present in the reference genome. The strain of *H. zea*

474 used for genome sequencing was described as highly susceptible to Cry1A toxins by Pearce et al.
475 (2017), which is true relative to levels of resistance recently described in *H. zea* (Brévault et al.,
476 2013; Dively et al., 2016; Reisig et al., 2018; Yang et al., 2019). Yet in the early 1990s, before
477 Bt crops were commercialized, this long-term laboratory-reared strain was less susceptible to
478 Cry1Ac than most strains derived recently from the field, including a >400-fold difference in one
479 case (Luttrell et al., 1999). Thus, we cannot exclude the possibility that the *cad-86C* allele found
480 here in GA-R and in the strain sequenced by Pearce et al. (2017) decreases susceptibility to
481 Cry1Ac. Additional work remains to characterize the functional role of this and other alleles.
482 Finally, there are inherent tradeoffs to using laboratory-selected versus field-selected
483 populations to identify genes under selection (Ffrench-Constant, 2013). On the one hand,
484 laboratory-selected populations can provide insights into the genetic mechanisms underlying
485 resistance (Gahan et al., 2001), and Bt resistance mutations identified in laboratory-selected
486 populations have been found in field-selected populations (Zhang et al., 2012; Jin et al., 2018;
487 Mathew et al., 2018; Wang et al., *in press*). On the other hand, some mutations that produce
488 resistant phenotypes in the lab cannot be found in field-collected individuals (Zhang et al., 2012;
489 Fabrick et al., 2014; Mathew et al., 2018). Here, we used a pair of laboratory-reared *H. zea* lines
490 for identification of genes under selection by Cry1Ac. Although we determined that a novel
491 midgut-expressed cadherin gene shows signatures of selection in a Cry1Ac-selected line, it is
492 unclear whether this gene has a significant role in field-evolved Cry1Ac resistance. Future work
493 should examine the importance of *cad-86C* to Bt resistance in field-selected *H. zea*.
494

Availability of data and other materials

Whole genome sequencing data and PacBio data are available on the NCBI Sequence Read Archive under accession numbers XXXXXX and XXXXXX (to be updated following acceptance of the publication). Raw Sanger and qPCR data can be found in Dryad Digital Repository XXXXX (to be updated following acceptance of the publication). All scripts used for data analysis can be found at https://github.com/mcadamme/GA_and_GAR.

Competing interests

The authors declare that they have no competing interests.

Funding

Funding for this research was provided by the USDA NIFA Biotechnology and Risk Assessment Grant 2016-33522-25640. The funding source did not participate in design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

YC and BT generated the *H. zea* lines. MLF, SON, YC, and BT designed the experiments. MLF, SON, RG, and YC collected the data. MLF, SON, and RG analyzed the data. MLF, SON, and RG wrote the manuscript. All authors edited and reviewed the final manuscript.

513 **Acknowledgments**

514 Alex DeYonke and the North Carolina Genomic Sciences Lab prepared samples and sequencing
515 libraries. Anahí Espindola provided valuable feedback on the cadherin phylogeny. Fred Gould
516 read and provided valuable comments on the manuscript.

References

Ali, M.I., Luttrell, R.G., Young III, S.Y., 2006. Susceptibilities of *Helicoverpa zea* and *Heliothis virescens* (Lepidoptera: Noctuidae) populations to Cry1Ac insecticidal protein. J. Econ. Entomol. 99(1), 164-75.

Ayme-Southgate, A., Lasko, P., French, C., Pardue, M.L., 1989. Characterization of the gene for mp20: a *Drosophila* muscle protein that is not found in asynchronous oscillatory flight muscle. J. Cell Biol. 108(2), 521-31.

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

Blankenberg, D., Kuster, G.V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., Taylor J., 2010. Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. 89(1), 19-0.

Benedict, J.H., Altman, D.W., Umbeck, P.F., Ring, D.R., 1992. Behavior, growth, survival, and plant injury by *Heliothis virescens* (F.)(Lepidoptera: Noctuidae) on transgenic Bt cottons. J. Econ. Entomol. 85(2), 589-93.

Brévault, T., Heuberger, S., Zhang, M., Ellers-Kirk, C., Ni, X., Masson, L., Li, X., Tabashnik, B.E., Carrière, Y., 2013. Potential shortfall of pyramided transgenic cotton for insect resistance management. Proc. Nat. Acad. Sci. U. S. A. 110(15), 5806-11.

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

537 Bushnell, B., 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National
538 Lab (LBNL), Berkeley, CA (United States).

539 Caccia, S., Moar, W.J., Chandrashekhar, J., Oppert, C., Anilkumar, K.J., Jurat-Fuentes, J.L.,
540 Ferré, J., 2012. Association of Cry1Ac toxin resistance in *Helicoverpa zea* (Boddie) with
541 increased alkaline phosphatase levels in the midgut lumen. Appl. Environ. Microbiol.
542 78(16), 5690-8.

543 Carrière, Y., Crickmore, N., Tabashnik, B.E., 2015. Optimizing pyramided transgenic Bt crops
544 for sustainable pest management. Nat. Biotechnol. 33, 161-168.

545 Carrière, Y., Degain, B., Unnithan, G.C., Harpold, V., Heuberger, S., Li, X., Tabashnik, B.E.,
546 2018a. Effects of seasonal changes in cotton plants on evolution of resistance to pyramided
547 cotton producing the toxins Cry1Ac and Cry1F by *Helicoverpa zea.* Pest Manag. Sci. 74,
548 627–637.

549 Carrière, Y., Williams, J.L., Crowder, D.W., Tabashnik, B.E., 2018b. Genotype-specific fitness
550 cost of resistance to Bt toxin Cry1Ac in pink bollworm. Pest Manag. Sci. 74(11), 2496-
551 2503.

552 Chandler, C.H., Chari, S., Dworkin, I., 2013. Does your gene need a background check? How
553 genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet.
554 29(6), 358-66.

555 Clark, T.L., Foster, J.E., Kamble, S.T., Heinrichs, E.A., 2000. Comparison of Bt (*Bacillus*
556 *thuringiensis* Berliner) maize and conventional measures for control of the European corn
557 borer (Lepidoptera: Crambidae). J. Entomol. Sci. 35(2),118-28.

558 Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E.,
559 Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., 2011. The variant call format and
560 VCFtools. Bioinformatics. 27(15), 2156-8.

561 Del Fabbro, C., Scalabrin, S., Morgante, M., Giorgi, F.M., 2013. An extensive evaluation of read
562 trimming effects on Illumina NGS data analysis. PLoS One. 8(12), e85024.

563 Ding, D., Liu, G., Hou, L., Gui, W., Chen, B., Kang, L., 2018. Genetic variation in PTPN1
564 contributes to metabolic adaptation to high-altitude hypoxia in Tibetan migratory locusts.
565 Nat. Commun. 9(1), 4991.

566 Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high
567 throughput. Nucleic Acids Res. 32(5), 1792–1797.

568 Fabrick, J.A., Ponnuraj, J., Singh, A., Tanwar, R.K., Unnithan, G.C., Yelich, A.J., Li, X.,
569 Carrière, Y., Tabashnik, B.E., 2014. Alternative splicing and highly variable cadherin
570 transcripts associated with field-evolved resistance of pink bollworm to Bt cotton in India.
571 PLoS One. 9(5), e97900.

572 Ffrench-Constant, R.H., 2013. The molecular genetics of insecticide resistance. Genetics. 194,
573 807–815.

574 Gahan, L.J., Gould, F., Heckel, D.G., 2001. Identification of a gene associated with Bt resistance
575 in *Heliothis virescens*. Sci. 293(5531), 857-60.

576 Gassmann, A.J., Carrière, Y., Tabashnik, B.E., 2009. Fitness costs of insect resistance to *Bacillus*
577 *thuringiensis*. Ann. Rev. Entomol. 54, 147-163.

578 Goldstein, J.A., Mason, C.E., Pesek, J., 2010. Dispersal and movement behavior of neonate
579 European corn borer (Lepidoptera: Crambidae) on non-Bt and transgenic Bt corn. J. Econ.
580 Entomol. 103(2), 331-9.

581 Gould, F., 1998. Sustainability of transgenic insecticidal cultivars: integrating pest genetics and
582 ecology. Ann. Rev. Entomol. 43(1), 701-26.

583 Gould, F., Brown, Z.S., Kuzma, J., 2018. Wicked evolution: Can we address the sociobiological
584 dilemma of pesticide resistance?. Sci. 360(6390), 728-32.

585 Hahn, M.W., 2019. Molecular Population Genetics, first ed. Oxford University Press.

586 Hua, G., Jurat-Fuentes, J.L., and Adang, M.J., 2004. Bt-R1a Extracellular Cadherin Repeat 12
587 Mediates *Bacillus thuringiensis* Cry1Ab Binding and Cytotoxicity. J. Biol. Chem. 279,
588 28051-28056.

589 Izutsu, M., Zhou, J., Sugiyama, Y., Nishimura, O., Aizu, T., Toyoda, A., Fujiyama, A., Agata,
590 K., Fuse, N., 2012. Genome features of “Dark-fly”, a Drosophila line reared long-term in a
591 dark environment. PLoS One. 7(3), e33288.

592 Jana, S.C., Bettencourt-Dias, M., Durand, B., Megraw, T.L., 2016. *Drosophila melanogaster* as a
593 model for basal body research. Cilia. 5(1), 22.

594 Jha, A.R., Miles, C.M., Lippert, N.R., Brown, C.D., White, K.P., Kreitman, M., 2015. Whole-
595 genome resequencing of experimental populations reveals polygenic basis of egg-size
596 variation in *Drosophila melanogaster*. Mol. Biol. Evol. 32(10), 2616-32.

597 Jin, L., Zhang, H., Lu,Y., Yang, Y., Wu, K., Tabashnik, B.E., Wu, Y., 2015. Large-scale test of
598 the natural refuge strategy for delaying insect resistance to transgenic Bt crops. Nat.
599 Biotechnol. 33(2), 169.

600 Jin, L., Wang, J., Guan, F., Zhang, J., Yu, S., Liu, S., Xue, Y., Li, L., Wu, S., Wang, X., Yang,
601 Y., Abdelgaffar, H., Jurat-Fuentes, J.L., Tabashnik, B.E., Wu, Y., 2018. Dominant point
602 mutation in a tetraspanin gene associated with field-evolved resistance of cotton bollworm
603 to transgenic Bt cotton. Proc. Nat. Acad. Sci. U.S.A. 115(46), 11760-11765.

604 Kaur, G., Guo, J., Brown, S., Head, G.P., Price, P.A., Paula-Moreaes, S., Ni, X., Dimase, M.,
605 Huang, F., 2019. Field-evolved resistance of *Helicoverpa zea* (Boddie) to transgenic maize
606 expressing pyramided Cry1A.105/Cry2Ab2 proteins in northeast Louisiana. J. Inv. Pathol.
607 163, 11-20.

608 Kliot, A., Ghanim, M., 2012. Fitness costs associated with insecticide resistance. Pest Manag.
609 Sci. 68(11), 1431-1437.

610 Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu:
611 scalable and accurate long-read assembly via adaptive k-mer weighting and repeat
612 separation. Genome Res. 27(5), 722-36.

613 Langmead, B., Trapnell, C., Pop, M., Salzberg, S.L., 2009. Ultrafast and memory-efficient
614 alignment of short DNA sequences to the human genome. Genome Biol. 10(3), R25.

615 Li, H., 2011. A statistical framework for SNP calling, mutation discovery, association mapping
616 and population genetical parameter estimation from sequencing data. Bioinformatics.
617 27(21), 2987-93.

618 Liu, C., Li, Y., Gao, Y., Ning, C., Wu, K., 2010. Cotton bollworm resistance to Bt transgenic
619 cotton: a case analysis. Sci. China Life Sci. 53(8), 934-41.

620 Lovegrove, B., Somoes, S., Rivas, M.L., Sotillos, S., Johnson, K., Knust, E., Jacinto, A.,
621 Hombria, J.C., 2006. Coordinated control of cell adhesion, polarity, and cytoskeleton
622 underlies Hox-induced organogenesis in *Drosophila*. Curr. Biol. 16(22), 2206-16.

623 Luttrell, R.G., Wan, L., Knighten, K., 1999. Variation in susceptibility of noctuid (Lepidoptera)
624 larvae attacking cotton and soybean to purified endotoxin proteins and commercial
625 formulations of *Bacillus thuringiensis*. J. Econ. Entomol. 92(1), 21-32.

626 Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey,
627 A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence
628 analysis tools APIs in 2019. Nucleic Acids Res.

629 Mathew, L.G., Ponnuraj, J., Mallappa, B., Chowdary, L.R., Zhang, J., Tay, W.T., Walsh, T.K.,
630 Gordon, K.H.J., Heckel, D.G., Downes, S., Carrière, Y., Li, X., Tabashnik, B.E., Fabrick,
631 J.A., 2018. Disruption of an ABC transporter associated with resistance to Bt toxin Cry2Ab
632 in lab- and field-selected pink bollworm from U.S. and India. Sci. Rep. 8, 13531.
633 DOI:10.1038/s41598-018-31840-5.

634 Men, X., Ge, F., Yardim, E.N., Parajulee, M.N., 2005. Behavioral response of *Helicoverpa*
635 *armigera* (Lepidoptera: Noctuidae) to cotton with and without expression of the CrylAc δ-
636 endotoxin protein of *Bacillus thuringiensis* Berliner. J. Insect Behav. 18(1), 33-50.

637 Morin, S., Biggs, R.W., Sisterson, M.S., Shriver, L., Ellers-Kirk, C., Higginson, D., Holley, D.,
638 Gahan, L.J., Heckel, D.G., Carrière, Y., Dennehy, T.J., 2003. Three cadherin alleles

associated with resistance to *Bacillus thuringiensis* in pink bollworm. Proc. Nat. Acad. Sci. U. S. A. 100(9), 5004-9.

Nickerson, D.A., Tobe, V.O., Taylor, S.L., 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res. 25(14), 2745-51.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302(1), 205-17.

Orpet, R.J., Degain, B.A., Unnithan, G.C., Welch, K.L., Tabashnik, B.E., Carrière, Y., 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.

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

Paradis, E., 2010. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 26(3), 419-20.

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

Patel, R.K., Jain, M., 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 7(2), e30619.

659 Pearce, S.L., Clarke, D.F., East, P.D., Elfekih, S., Gordon, K.H., Jermiin, L.S., McGaughran, A.,
660 Oakeshott, J.G., Papanikolaou, A., Perera, O.P., and Rane, R.V., 2017. Genomic
661 innovations, transcriptional plasticity and gene loss underlying the evolution and divergence
662 of two highly polyphagous and invasive *Helicoverpa* pest species. BMC Biol. 15(1), 63.

663 Pérez-Figueroa, A., García-Pereira, M.J., Saura, M., Rolán-Alvarez, E., and Caballero, A., 2010.
664 Comparing three different methods to detect selective loci using dominant markers. J. Evol.
665 Biol. 23, 2267-2276.

666 Prasifka, J.R., Hellmich, R.L., Sumerford, D.V., Siegfried, B.D., 2009. *Bacillus thuringiensis*
667 resistance influences European corn borer (Lepidoptera: Crambidae) larval behavior after
668 exposure to Cry1Ab. J. Econ. Entomol. 102(2), 781-7.

669 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar,
670 P., De Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome
671 association and population-based linkage analyses. Am. J. Hum. Genet. 81(3), 559-75.

672 Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic
673 features. Bioinformatics. 26(6), 841-2.

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

677 Ramalho, F.S., Pachú, J.K., Lira, A.C., Malaquias, J.B., Zanuncio, J.C., Fernandes, F.S., 2014.
678 Feeding and dispersal behavior of the cotton leafworm, *Alabama argillacea*

679 (Hübner)(Lepidoptera: Noctuidae), on Bt and non-Bt cotton: implications for evolution and
680 resistance management. PLoS One. 9(11), e111588.

681 Reisig, D.D., Huseth, A.S., Bacheler, J.S., Aghaee, M.A., Braswell, L., Burrack, H.J., Flanders,
682 K., Greene, J.K., Herbert, D.A., Jacobson, A., Paula-Moraes, S.V., 2018. Long-term
683 empirical and observational evidence of practical *Helicoverpa zea* resistance to cotton with
684 pyramided Bt toxins. J. Econ. Entomol. 111(4), 1824-33.

685 Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M,, Lander, E.S., Getz, G.,
686 Mesirov, J.P., 2011. Integrative genomics viewer. Nat. Biotechnol. 29(1), 24.

687 Roush, R.T., 1997. Bt-transgenic crops: just another pretty insecticide or a chance for a new start
688 in resistance management?. Pestic. Sci. 51(3), 328-34.

689 Rubin, C.J., Zody, M.C., Eriksson, J., Meadows, J.R., Sherwood, E., Webster, M.T., Jiang, L.,
690 Ingman, M., Sharpe, T., Ka, S., Hallböök, F., 2010. Whole-genome resequencing reveals
691 loci under selection during chicken domestication. Nature. 464(7288), 587.

692 Ruijter, J.M., Ramakers, C., Hoogaars, W.M., Karlen, Y., Bakker, O., Van den Hoff, M.J.,
693 Moorman, A.F., 2009. Amplification efficiency: linking baseline and bias in the analysis of
694 quantitative PCR data. Nucleic Acids Res. 37(6), e45-e45.

695 Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27(4), 592-593.

696 Schoen, D.J., David, J.L., Bataillon, T.M., 1998. Deleterious mutation accumulation and the
697 regeneration of genetic resources. Proc. Nat. Acad. Sci. U. S. A. 95(1), 394-9.

698 Sivasupramaniam, S., Moar, W.J., Ruschke, L.G., Osborn, J.A., Jiang, C., Sebaugh, J.L., Brown,
699 G.R., Shappley, Z.W., Oppenhuizen, M.E., Mullins, J.W., Greenplate, J.T., 2014. Toxicity
700 and characterization of cotton expressing *Bacillus thuringiensis* Cry1Ac and Cry2Ab2
701 proteins for control of lepidopteran pests. J. Econ. Entomol. 101(2), 546-54.

702 Stone, T.B., Sims, S.R., 1993. Geographic susceptibility of *Heliothis virescens* and *Helicoverpa*
703 *zea* (Lepidoptera: Noctuidae) to *Bacillus thuringiensis*. J. Econ. Entomol. 86(4), 989-94.

704 Tabashnik, B.E., Patin, A.L., Dennehy, T.J., Liu ,Y-B., Carrière, Y., Antilla, L., 2000. Frequency
705 of resistance to *Bacillus thuringiensis* in field populations of pink bollworm. Proc, Natl.
706 Acad. Sci. U.S.A. 21, 12980-12984.

707 Tabashnik, B.E., Gould, F., Carrière, Y., 2004. Delaying evolution of resistance to transgenic
708 crops by decreasing dominance and heritability. J. Evol. Biol. 17, 904-912.

709 Tabashnik, B.E., Carrière, Y., 2017. Surge in insect resistance to transgenic crops and prospects
710 for sustainability. Nat. Biotechnol. 35(10), 926.

711 Tabashnik, B. E. Carrière, Y., *in press*. Global patterns of resistance to Bt crops highlighting
712 pink bollworm in the United States, China and India. J. Econ. Entomol.

713 Treisman, R., 1996. Regulation of transcription by MAP kinase cascades. Curr. Opin. Cell Biol.
714 8(2), 205-15.

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

719 Wang, G., Wu, K., Liang, G., Guo, Y., 2005. Gene cloning and expression of cadherin in midgut
720 of *Helicoverpa armigera* and its Cry1A binding region. Sci. China C Life Sci. 48(4), 346-
721 56.

722 Wang, L., Ma, Y., Wan, P., Liu, K., Xiao, Y., Wang, J., Cong, S., Xu, D., Wu, K., Fabrick, J.A.,
723 Li, X., Tabashnik, B.E., 2018. Resistance to *Bacillus thuringiensis* linked with a cadherin
724 transmembrane mutation affecting cellular trafficking in pink bollworm from China. Insect
725 Biochem. Mol. Biol. 94, 28-35.

726 Wang, J., D. Xu, L. Wang, S. Cong, P. Wan, C. Lei, J. A. Fabrick, X. Li, B. E Tabashnik, and K.
727 Wu. *in press*. Bt resistance alleles in field populations of pink bollworm from China:
728 Similarities with the United States and decreased frequency from 2012 to 2015. Pest
729 Manag. Sci..

730 Welch, K.L., Unnithan, G.C., Degain, B.A., Wei, J., Zhang, J., Li, X., Tabashnik, B.E., Carrière,
731 Y., 2015. Cross-resistance between toxins in pyramided Bt crops and resistance to Bt
732 sprays in *Helicoverpa zea*. J. Invert. Pathol. 132, 149-156.

733 Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from
734 multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18(5),
735 691-699.

736 Xu, X., Yu, L., Wu, Y., 2005. Disruption of a cadherin gene associated with resistance to
737 Cry1Ac-endotoxin of *Bacillus thuringiensis* in *Helicoverpa armigera*. App. Environ.
738 Microbiol. 71, 948-954.

739 Yang, F., González, J.C., Williams, J., Cook, D.C., Gilreath, R.T., 2019. Occurrence and ear
740 damage of *Helicoverpa zea* on transgenic *Bacillus thuringiensis* maize in the field in Texas,
741 US and its susceptibility to Vip3A protein. Toxins. 11(2), 102.

742 Zhang, H., Tian, W., Zhao, J., Jin, L., Yang, J., Liu, C., Yang, Y., Wu, S., Wu, K., Cui, J.,
743 Tabashnik, B.E., Wu, Y., 2012. Diverse genetic basis of field-evolved resistance to Bt
744 cotton in cotton bollworm from China. Proc. Natl. Acad. Sci. U.S.A. 109(26), 10275–10280.
745 doi:10.1073/pnas.1200156109

746 Zhang, Z., Teng, X., Ma, W., Li, F., 2017. Knockdown of two cadherin genes confers resistance
747 to Cry2A and Cry1C in *Chilo suppressalis*. Sci. Rep. 7(1), 5992.

748 Zhang, M., Wei, J., Ni, X., Zhang, J., Jurat-Fuentes, J.L., Fabrick, J.A., Carrière, Y., Tabashnik,
749 B.E., and Li, X., 2019. Decreased Cry1Ac activation by midgut proteases associated with
750 Cry1Ac resistance in *Helicoverpa zea*. Pest Manag. Sci. 75, 1099-1106.

751 Zheng, Y., Mellem, J.E., Brockie, P.J., Madsen, D.M., Maricq, A.V., 2004. SOL-1 is a CUB-
752 domain protein required for GLR-1 glutamate receptor function in *C. elegans*. Nature
753 427(6973), 451.

754

Figure Captions

Figure 1. (a) F_{ST} between lines and (b) pooled heterozygosity in the GA-R line along 40-kb sliding windows in the *H. zea* genome. The location of the putative *cad-86c* selective sweep is indicated with an arrow in panel b. Panels c and d show the distribution of F_{ST} values between lines throughout the genome, and Hp values within the GA-R line throughout the genome, respectively.

Figure 2. (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 grey.

Figure 3. 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.

a

b

c

d

ZF_{ST}

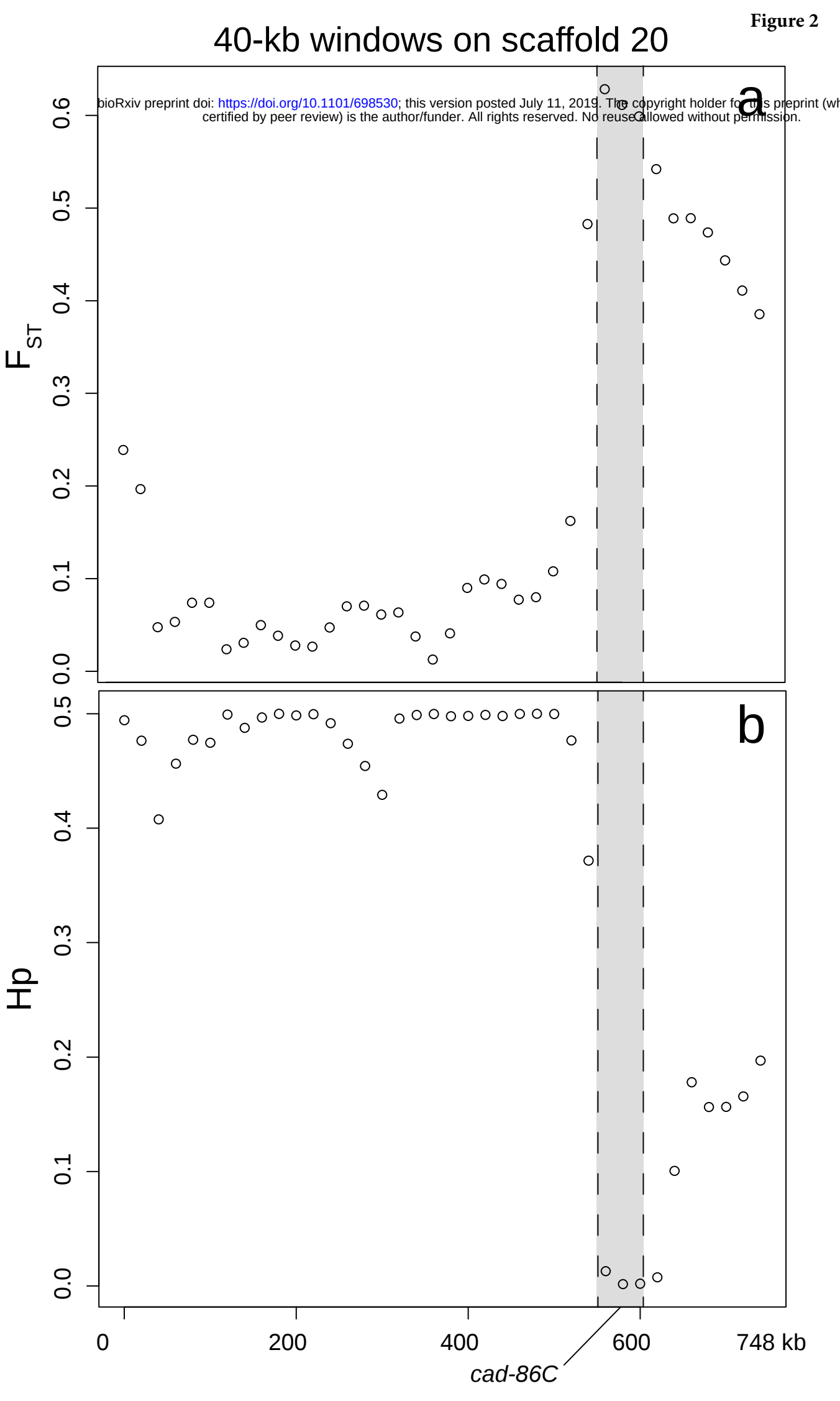
ZHp

cad-86C

40-kb windows along scaffolds

Number of 40-kb windows

Figure 2



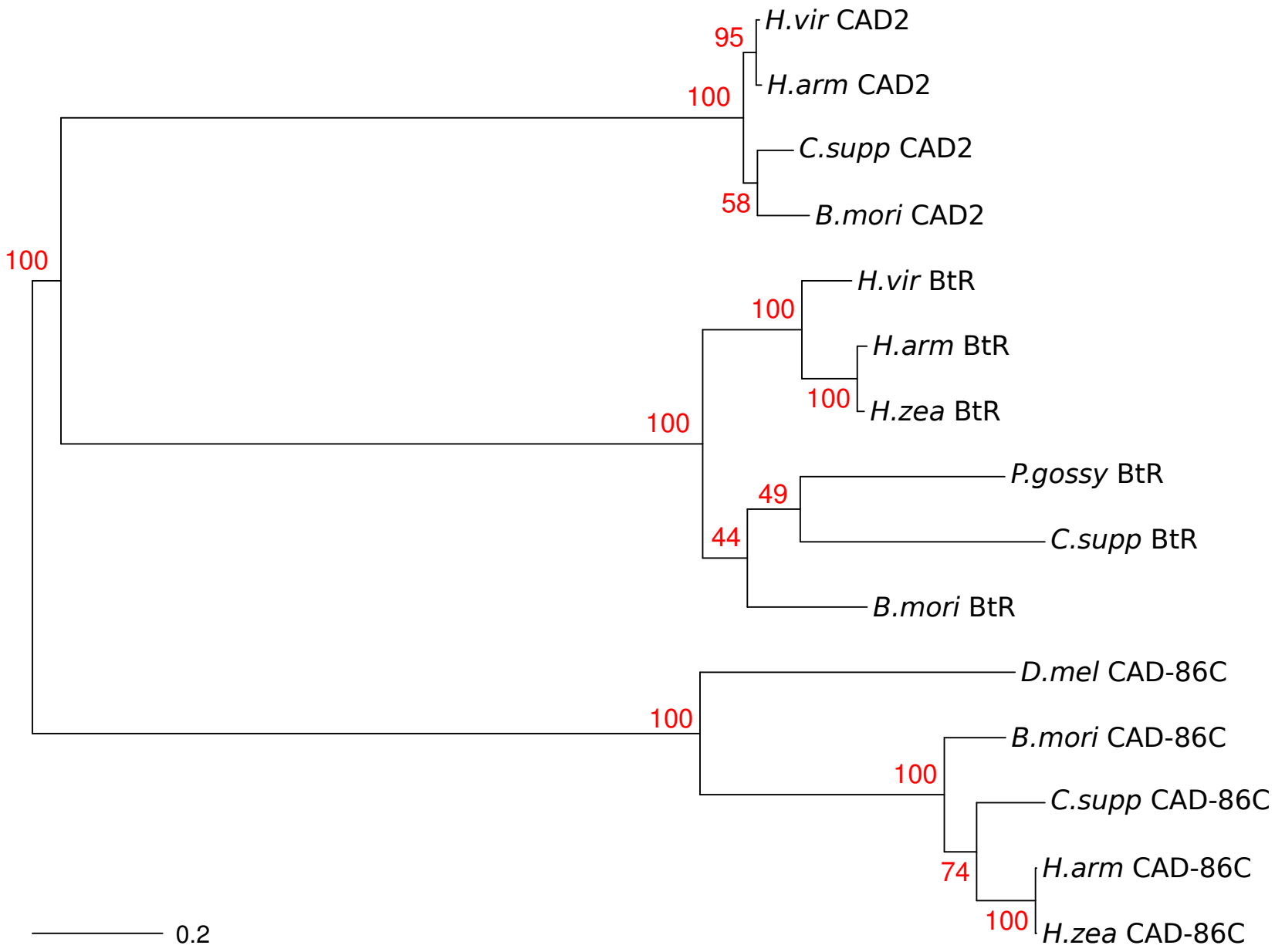

H.vir CAD2
H.arm CAD2
C.supp CAD2
B.mori CAD2
H.vir BtR
H.arm BtR
H.zea BtR
P.gossy BtR
C.supp BtR
B.mori BtR
D.mel CAD-86C
B.mori CAD-86C
C.supp CAD-86C
H.arm CAD-86C
H.zea CAD-86C
95
100
58
100
100
100
100
49
44
100
100
74
100
0.2

Table 1. P_N is proportion of polymorphic SNPs, MAF is the mean minor allele frequency, H_O and H_E are observed and expected heterozygosity, and F_{IS} is the inbreeding coefficient, and D is the average pairwise genetic distance between individuals in a population. Standard deviations are presented in parentheses.

	Susceptible	**Resistant**
Sample Size	5	5
P_N	0.738	0.823
MAF	0.225	0.241
H_O	0.312 (0.273)	0.350 (0.266)
H_E	0.289 (0.194)	0.313 (0.177)
F_{IS}	0.250 (0.060)	0.161 (0.039)
D	0.263 (0.015)	0.275 (0.008)

Table 2. Estimates of genetic diversity within the GA and GA-R lines in two non-coding regions of the *cad-86C* gene. N represents the number of chromosomes sampled.

Primer Pair	**Sequence Length**	**Line**	**N**	θ_W	π
Cad_1b	421	GA	48	4.5	0.016
		GA-R	44	4.8	0.009
Cad_2b	520	GA	48	2.0	0.006
		GA-R	48	3.2	0.004

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) rather than individuals.

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