Novel genetic basis of resistance to Bt toxin Cry1Ac in *Helicoverpa zea*

Kyle M. Benowitz[1,2,*], Carson W. Allan[1], Benjamin A. Degain[1], Xianchun Li[1], Jeffrey A. Fabrick[3], Bruce E. Tabashnik[1], Yves Carrière[1], Luciano M. Matzkin[1,4,5]

[1]Department of Entomology, University of Arizona, Tucson, AZ, USA

[2]Department of Biology, Austin Peay State University, Clarksville, TN, USA

[3]U.S. Department of Agriculture, Agricultural Research Service, U.S. Arid Land Agricultural Research Center, Maricopa, AZ, USA

[4]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

[5]Bio5 Institute, University of Arizona, Tucson, AZ, USA

Corresponding author: Kyle M. Benowitz. Department of Biology, Austin Peay State University, Sundquist Science Center, PO Box 4718, Clarksville, TN 37044. Email: benowitzk@apsu.edu

Running Title: *A novel Bt resistance gene*

Abstract

Crops genetically engineered to produce insecticidal proteins from the bacterium *Bacillus thuringiensis* (Bt) have advanced pest management, but their benefits are diminished when pests evolve resistance. Elucidating the genetic basis of pest resistance to Bt toxins can improve resistance monitoring, resistance management, and design of new insecticides. Here, we investigated the genetic basis of resistance to Bt toxin Cry1Ac in the lepidopteran *Helicoverpa zea*, one of the most damaging crop pests in the United States. To facilitate this research, we built the first chromosome-level genome assembly for this species. Using a genome-wide association study, fine-scale mapping, and RNA-seq, we identified a 250kb QTL on chromosome 13 that was strongly associated with resistance in a strain of *H. zea* that had been selected for resistance in the field and lab. This QTL contains no genes with a previously reported role in resistance or susceptibility to Bt toxins. However, within this QTL, we discovered a premature stop codon in a kinesin gene. We hypothesize that this mutation contributes to resistance. The results indicate the mutation on chromosome 13 was necessary but not sufficient for resistance, and therefore conclude that mutations in more than one gene contributed to resistance. Moreover, we found no changes in gene sequence or expression consistently associated with resistance for 11 genes previously implicated in lepidopteran resistance to Cry1Ac. Thus, the results reveal a novel and polygenic basis of resistance and extend the list of genes contributing to pest resistance to Bt toxins.

Keywords: *Bacillus thuringiensis*; genome wide association study; genome assembly; insecticide resistance; kinesin; Lepidoptera; transgenic crops

41 **Introduction**

42 Crops genetically engineered to produce insecticidal proteins from *Bacillus thuringiensis* (Bt)
43 have provided control of some key pests during the past 25 years while reducing insecticide
44 sprays and conserving arthropod natural enemies (Bravo *et al.* 2011; NASEM 2016; Dively *et al.*
45 2018, Romeis *et al.* 2018, Carrière *et al.* 2020a; Tabashnik *et al.* 2021). However, planting of a
46 cumulative total of more than one billion hectares of Bt crops worldwide (ISAAA 2019) has
47 selected for resistance that has reduced the efficacy of Bt crops against populations of at least
48 nine major pest species (Calles-Torrez *et al.* 2019, Smith *et al.* 2019, Tabashnik and Carrière
49 2019). Knowledge of the genetic basis of pest resistance to Bt toxins can be useful for improving
50 monitoring and management of resistance, as well as for designing new insecticides (Soberón *et*
51 *al.* 2007, Jin *et al.* 2018).

52 Resistance to crystalline (Cry) Bt toxins typically entails mutations that disrupt binding of the
53 toxins to larval midgut receptors and thus block an essential step in the mode of action (Heckel *et*
54 *al.* 2007; Peterson *et al.* 2017; Jurat-Fuentes *et al.* 2021). In particular, research has repeatedly
55 implicated mutations disrupting known or putative Bt toxin receptors from four protein families:
56 ATP-binding cassette (ABC), cadherin, aminopeptidase N (APN), and alkaline phosphatase
57 (ALP). Resistance mediated by mutations that disrupt binding of toxins to receptors is frequently
58 associated with high levels of resistance to one or a few closely related Bt toxins, weak or no
59 cross-resistance to unrelated Bt toxins, and recessive inheritance (Mode 1 resistance; Tabashnik
60 *et al.* 1998). Nonetheless, lepidopteran resistance to Bt toxins also includes examples where
61 proteins from these families are not involved, toxin binding is not reduced, and inheritance of
62 resistance is not recessive (Peterson *et al.* 2017, Jin *et al.* 2018).

63 Here we analyzed the genetic basis of resistance to Bt toxin Cry1Ac in the lepidopteran
64 *Helicoverpa zea* (corn earworm or bollworm), which is one of the most economically important
65 crop pests in the United States (Cook and Threet 2019; Musser *et al.* 2019). This polyphagous
66 pest is the first insect reported to have evolved resistance to a Bt crop, specifically to cotton
67 producing Cry1Ac (Luttrell *et al.* 1999; Ali *et al.* 2006; Tabashnik *et al.* 2008; Reisig *et al.*
68 2018). In contrast with Mode 1 resistance, inheritance of resistance to Cry1Ac in *H. zea* is not
69 completely recessive (Brévault *et al.* 2013, 2015; Carrière *et al.* 2020b and references therein).

70 Caccia *et al.* (2012) concluded that resistance in their lab-selected AR1 strain of *H. zea* was
71 complex, possibly polygenic, and not caused primarily by reduced binding of Cry1Ac to larval
72 midgut membranes. Perera *et al.* (2021) found that knocking out the gene encoding the putative
73 Cry1Ac receptor ABCC2 increased the concentration of Cry1Ac killing 50% of larvae (LC_{50}) by
74 7- to 40-fold. Because >100-fold resistance to Cry1Ac is common in lab- and field-selected *H.*
75 *zea* (Caccia *et al.* 2012, Brévault *et al.* 2013, Reisig *et al.* 2018; Kaur *et al.* 2019), they inferred
76 that mutations disrupting ABCC2 are not the sole or primary mechanism of resistance to Cry1Ac
77 in *H. zea.* Although mechanistic studies of Bt resistance in *H. zea* have emphasized the gene
78 families listed above (Caccia *et al.* 2012; Zhang *et al.* 2019; Perera *et al.* 2021), promising
79 additional candidates have been identified using population genomics and RNA-seq (Fritz *et al.*
80 2020, 2021; Lawrie *et al.* 2020).

81 Our work focuses on the GA-R strain of *H. zea*, which had been selected for resistance to
82 Cry1Ac in the field and lab (Brévault *et al.* 2013, Welch *et al.* 2015). GA-R was derived from
83 the moderately resistant GA strain, which had been selected for resistance to Bt toxins only in
84 the field (Brévault *et al.* 2013). Relative to an unrelated susceptible lab strain (LAB-S) of *H. zea*,
85 the LC_{50} of Cry1Ac was >500 times higher for GA-R and 55 times higher for GA (Brévault *et al.*
86 2013). Previous work identified reduced activation of Cry1Ac by midgut proteases as a potential
87 field-selected mechanism of resistance that could explain part but not all of the resistance in GA-
88 R relative to LAB-S (Zhang *et al.* 2019). Overall, the previous results with GA-R and other
89 strains of *H. zea* summarized above led us to hypothesize that resistance to Cry1Ac in this
90 species is polygenic and entails novel genetic mechanisms. Accordingly, genome-wide mapping
91 approaches are warranted, but have been hindered because the only *H. zea* genome assembly
92 available is highly fragmented (Pearce *et al.* 2017).

93 Here, we generated a chromosome-level genome for *H. zea*, then used a genome wide
94 association study (GWAS), fine-scale mapping, and RNA-seq to elucidate the genetic basis of
95 resistance to Cry1Ac in the GA-R strain of *H. zea.* We identified a quantitative trait locus (QTL)
96 of 250kb on chromosome 13 that was strongly associated with resistance to Cry1Ac. This QTL
97 contains no genes in families previously implicated in resistance or susceptibility to Bt toxins.
98 However, we did identify a premature stop codon in a gene encoding a kinesin-related protein
99 within the QTL. We hypothesize that this mutation contributes to resistance. We additionally

100 found no consistent association between resistance to Cry1Ac and any of 11 genes previously
101 implicated in lepidopteran resistance to Bt toxins. Our results suggest that the mutation on
102 chromosome 13 is necessary, but not sufficient, for resistance in GA-R. Thus, we conclude the
103 genetic basis of resistance to Cry1Ac in the GA-R strain of *H. zea* is both novel and polygenic.

104

105 Materials and methods

106 Insect strains

107 We used four strains of *H. zea*: the highly resistant strain GA-R, its moderately resistant parent
108 strain GA, the unrelated susceptible strain LAB-S, and the heterogeneous strain GA-RS that we
109 created by crossing GA-R with LAB-S as described below. LAB-S was obtained from Benzon
110 Research Inc. (Carlisle, PA, USA) and has not been exposed to Bt toxins or other insecticides.
111 The resistant strain GA-R was derived from the GA strain, which was started with 180 larvae
112 collected on Cry1Ab corn from Tifton, Georgia in 2008 (Brévault *et al.*, 2013). GA-R was
113 initially selected with Cry1Ac for nine generations and with MVPII thereafter (Brévault *et al.*
114 2013, Fritz *et al.* 2020, Carrière *et al.* 2020b). MVPII is a liquid formulation of a hybrid protoxin
115 produced by transgenic *Pseudomonas fluorescens.* The amino acid sequence of the active portion
116 of the protoxin is identical to the hybrid protoxin and Cry1Ac (Welch *et al.* 2015). For
117 simplicity, hereafter we refer to MVPII as Cry1Ac.

118 We conducted all rearing in walk-in growth chambers at 27 ± 1°C with 14h light: 10h dark
119 photoperiod. We reared larvae on a casein-based wheat germ diet (Orpet *et al.* 2015a,b) and
120 conducted larval bioassays on Southland diet (Southland Products, Inc., Lake Village, AR,
121 USA). We use these two different diets to optimize larval development and surface uniformity
122 for toxin overlay in bioassays, respectively (Carrière *et al.* 2020b). Moths were kept in walk-in
123 growth chambers under the same temperature and photoperiod mentioned above but under higher
124 relative humidity than for larvae (60% Rh for moths and 20% Rh for larvae). Moths had access
125 to a 10% sugar water solution for feeding and cheese cloth for oviposition (Welch *et al.* 2015).

126 Relative to GA, GA-R had significantly higher survival on Bt cotton (producing Cry1Ac,

127 Cry1Ac + Cry2Ab, or Cry1Ac + Cry1F) and Bt corn producing Cry1A.105 + Cry2Ab (Brévault
128 *et al.* 2013, 2015; Carrière *et al.* 2018, 2019, 2020b, 2021). At the time we crossed GA-R with
129 LAB-S in May 2018, we had selected GA-R with Cry1Ac for 58 generations. The GA-RS strain
130 was created using reciprocal crosses between GA-R and LAB-S adults. One hundred twenty
131 mating pairs were used for each reciprocal cross (i.e., GA-R females × LAB-S males and LAB-S
132 females × GA-R males). GA-RS was founded with 450 neonates from each reciprocal cross. In
133 the founding and subsequent generations, 900 larvae were reared and 600 adults (sex ratio 1:1)
134 produced the neonates used for propagating the next generation.

135

136 **Genome sequencing of resistant strain GA-R**

137 We generated a *de novo* genome assembly for GA-R using an approach combining a hybrid
138 short- and long-read assembly with a long-read only assembly (Jaworski *et al.* 2020). For the
139 short-read assembly, we collected and sequenced 30 GA-R larvae as described below. We
140 trimmed reads and generated the assembly using Platanus (Kajitani *et al.* 2014). For the long-
141 read assembly, we extracted DNA from the gut of a single GA-R fifth instar using a chloroform-
142 based extraction method (Jaworski *et al.* 2020). PacBio libraries were constructed at the Arizona
143 Genomics Institute (Tucson, AZ, USA). We then sequenced the library on a single lane of
144 PacBio Sequel II, also at the Arizona Genomics Institute. We formatted raw PacBio reads using
145 bam2fastq (https://github.com/jts/bam2fastq) and SeqKit (Shen *et al.* 2016) before filtering out
146 all reads under 30kb in length using Filtlong (https://github.com/rrwick/Filtlong). We mapped
147 contigs from the short-read assembly to the long reads using DBG2OLC (Ye *et al.* 2016) before
148 running three iterations of Sparc (Ye and Ma 2016) to correct the resulting contigs. We realigned
149 the raw PacBio reads to the resulting assembly with pbmm2
150 (https://github.com/PacificBiosciences/pbmm2) and polished contigs using Arrow (Chin *et al.*
151 2013; https://github.com/PacificBiosciences/GenomicConsensus). Lastly, we mapped raw short
152 reads to the assembly with Bowtie2 (Langmead and Salzberg 2012) to perform a final polishing
153 step using Pilon (Walker *et al.* 2014).

154 We generated the PacBio only assembled with Canu (Koren *et al.* 2017), using the reads longer

155 than 30kb after filtering described above. We then polished the assembly using Arrow and Pilon
156 as described above. Finally, we used Purge Haplotigs (Roach *et al.* 2018) to remove spurious
157 backbones generated due to high heterozygosity.

158 We aligned the two assemblies using nucmer within MUMmer4 (Marçais *et al.* 2018), keeping
159 only alignments greater than 10kb. We then generated the merged assembly using Quickmerge
160 (Chakraborty *et al.* 2016). We performed manual curation of each chromosome after comparison
161 to the *Helicoverpa armigera* genome (Pearce *et al.* 2017; Valencia-Montoya *et al.* 2020) and
162 polished the final assembly using Arrow, Pilon, and Purge Haplotigs as above. Lastly, we
163 ordered and named each chromosome to align with those of *H. armigera*. We generated a
164 synteny plot comparing our genome to the *H. armigera* genome using Dot
165 (https://github.com/marianattestad/dot) after filtering out alignments under 4000 bp in NUCmer.

166 We analyzed the completeness of the GA-R genome using BUSCO v.4 (Seppey *et al.* 2019),
167 comparing genomic content against the lepidodpteraodb10 set of conserved single copy genes.
168 We calculated genome summary statistics with bbmap stats
169 (https://sourceforge.net/projects/bbmap/). We calculated repeat content with RepeatModeler2
170 (Flynn *et al.* 2020) and RepeatMasker (Smit *et al.* 2015). We provide a preliminary annotation
171 produced following the funannotate pipeline (https://github.com/nextgenusfs/funannotate) with
172 transcripts from *H. armigera* and proteins from *B. mori* used as evidence supporting putative
173 annotations.

174

175 **Phenotyping of Cry1Ac-susceptible and -resistant larvae for genetic mapping**

176 We used our standard diet overlay bioassay (Welch *et al.* 2015) to characterize GA-RS larvae as
177 susceptible or resistant to Cry1Ac. We added 40 µl of a dilution containing Triton X-100 and the
178 desired concentration of Cry1Ac to the surface of solidified Southland diet in each well of 128-
179 well bioassay trays (C-D International, Pitman, NJ, USA). One neonate (< 8 h old) was
180 transferred to each well and trays were covered with ventilated plastic lids (C-D International)
181 and held for 7 days under the abiotic conditions mentioned above.

182 We conducted five sets of bioassays using GA-RS neonates from the F2 (July 2018), F12 (July
183 2019), F22 and F23 (July 2020), and F26 (October 2020) generations (Table S1). Neonates were
184 exposed to diet with 0 (control), 1, or 10 μg Cry1Ac per cm^2 diet. After 7 days, live first instar
185 larvae on diet with 1 μg Cry1Ac per cm^2 were considered susceptible because this low toxin
186 concentration inhibited their growth, whereas live larvae on diet with 10 μg Cry1Ac per cm^2 that
187 were third or subsequent instars were considered resistant because they grew well despite this
188 high toxin concentration. For control larvae reared on diet without Cry1Ac, mean survival to
189 third instar was 96% (range: 91-100%, mean n = 99 larvae per bioassay in five bioassays).
190 Susceptible and resistant larvae were transferred individually to plastic cups containing non-Bt
191 diet, reared to fifth instar, transferred individually to 1.5 ml plastic tubes, and frozen at -80°C for
192 subsequent genomic comparison.

193

194 **Genomic comparison of GA-R and LAB-S**

195 We sequenced pools of larvae from the parental strains GA-R and LAB-S, which allowed us to
196 evaluate genetic variation between and within these parental strains. This also allowed us to
197 check if SNPs associated with resistance in the GWAS of GA-RS were more common in GA-R,
198 and if those associated with susceptibility were more common in LAB-S. In April 2019, we
199 collected 30 third instars from each strain, extracted DNA using Qiagen DNeasy Blood and
200 Tissue Kits (Qiagen, Hilden, Germany), and constructed Illumina libraries using KAPA LTP
201 Library Preparation Kits (Roche, Basel, Switzerland). We sequenced both libraries on an
202 Illumina HiSeq4000 at Novogene (Beijing, China). We called variants with Platypus after read
203 trimming and alignment to the genome using Trimmomatic and bwa-mem as described above.
204 To detect potential selective sweeps in each strain, we used SAMtools mpileup (Li 2011) and
205 PoPoolation (Kofler *et al.* 2011) to calculate Tajima's D in overlapping 50kb windows across the
206 genome.

207

208 **Genome-wide association study for Cry1Ac resistance**

209 From the heterogeneous strain GA-RS F12 larvae phenotyped in July 2019, we extracted DNA
210 from 144 resistant and 144 susceptible larvae using ZYMO *Quick-DNA* 96 Plus Kits and
211 quantified the DNA concentration of each individual using a Nanodrop (Thermo Fisher
212 Scientific, Waltham, MA, USA). We combined equal amounts of DNA from each of the 144
213 resistant larvae to make a resistant pool and from each of the 144 susceptible larvae to make a
214 susceptible pool, then generated libraries using KAPA LTP Library Preparation Kits for each
215 pool. We sequenced both libraries on an Illumina HiSeq4000 at Novogene (Beijing, China).

216 We demultiplexed reads and trimmed for adapter contamination and low-quality sequence using
217 Trimmomatic (Bolger *et al.* 2014). We mapped reads to the *de novo H. zea* genome using bwa-
218 mem (Li and Durbin 2009). We used Platypus (Rimmer *et al.* 2014) to call and quantify SNP
219 variants and short INDELs. We extracted biallelic SNPs from the Platypus output for statistical
220 analysis, for which we used two approaches (Benowitz *et al.* 2019). First, we calculated a Z-
221 statistic (Huang *et al.* 2012), using the formula $Z = \frac{\rho_1 - \rho_2}{\sqrt{(\rho_0(1-\rho_0)(\frac{1}{n}+\frac{1}{c_1}+\frac{1}{c_2})}}$, where ρ_1 and ρ_2 are the
222 reference alleles frequencies of each bulk, ρ_0 is the mean allele frequency across bulks, n is the
223 sample size of each bulk, and c_1 and c_2 are the read depth of each bulk. We evaluated statistical
224 significance of Z against the standard normal distribution. Following convention (Barsh *et al.*
225 2012, Welter *et al.* 2014), we used $P < 5 \times 10^{-8}$ as a threshold for significance. To estimate the
226 density of significant sites, we used the R package WindowScanR
227 (https://github.com/tavareshugo/WindowScanR) to calculate the percentage of SNPs with a more
228 liberal threshold of $P < 10^{-5}$ in overlapping 10kb windows. Density of significant sites may be a
229 particularly useful parameter because the magnitude of the *P*-values from a bulk segregant
230 analysis are highly sensitive to noise. Close linkage to the causal allele, however, should result in
231 a higher density of significant sites even if the *P*-value itself varies. Second, we performed a
232 500kb sliding-window analysis using the R package QTLseqr (Mansfeld and Grumet 2018),
233 which implements the G′ method of Magwene *et al.* (2011). This method provides a statistical
234 determination of whether an entire QTL, rather than an individual SNP, is significant, and also
235 defines borders to QTLs deemed significant.

236

237 **Fine-scale mapping within chromosome 13**

238 We conducted fine-scale mapping within the *r1* locus using larvae from the F22 and F23
239 generation (July 2020). Using the methods described above, larvae were scored as resistant or
240 susceptible. After phenotyping, we reared susceptible larvae to fifth instars on untreated diet
241 before storing all larvae (62 resistant and 51 susceptible) at -80°C. We extracted DNA from each
242 sample using Qiagen DNeasy Blood and Tissue Kits and genotyped each larva individually via
243 high-resolution melt curve (HRM) analysis at 12 SNP marker sites within the chromosome 13
244 QTL (Table S2). For each site, we performed PCR in 10uL reactions using Apex Taq DNA
245 polymerase and EvaGreen Dye (Biotium, Fremont, CA, USA) as the intercalating dye. We ran
246 rt-PCR in a QuantStudio 3 Real-Time PCR machine (Thermo Fisher) using continuous capture
247 with a 0.025°C/s ramp. We used QuantStudio Design and Analysis Software (Thermo Fisher) to
248 manually score melt curves for the identity of the focal SNP. We compared test melt curves
249 against control curves generated from the parental LAB-S and GA-R strains. We used Fisher's
250 Exact Test to assess significant differences between allele frequencies of resistant and
251 susceptible individuals.

252 The lack of amplification from some individuals for some markers yielded variation in sample
253 size among the 12 markers. These ranged from 40 to 60 for resistant larvae (mean = 56) and 20
254 to 50 (mean = 45) for susceptible larvae. We also used HRM to obtain genotypes at marker #4
255 for 23 of 25 resistant larvae and 89 of 95 control larvae tested from the F26 generation. The
256 control larvae were not screened with a bioassay and thus contained a mixture of resistant and
257 susceptible individuals.

258 To confirm the accuracy of HRM genotyping, we Sanger sequenced a single site (marker 4;
259 Table S2) for all 62 resistant and a subset (33) of susceptible individuals. We designed new
260 primers to produce a longer amplicon, and confirmed the quality of the resultant amplicons with
261 gel electrophoresis. We cleaned the PCR product with Exonuclease I and Antarctic Phosphatase
262 (New England BioLabs, Ipswich, MA, USA) before sending to Eurofins Genomics (Louisville,
263 KY, USA) for Sanger sequencing.

264 In addition to the fine mapping data from the HRM, we used the data on significant SNP density

265 from the original GWAS experiment as well as the Tajima's D results from comparison of the
266 GA-R and LAB-S strains to provide additional support for narrowing the region associated with
267 resistance within chromosome 13.

268

269 **Inheritance and trajectory of resistance in GA-RS**

270 We performed several analyses of genotype frequencies at marker 4 to understand how *r1* affects
271 resistance. We used the genotypes from the RNA-seq study (see below) as a control group to
272 examine the frequency of marker 4 in the GA-RS strain at the F25 generation. We also used a χ^2
273 test to examine departure from Hardy-Weinberg equilibrium.

274 We used Fisher's exact test to determine if resistant individuals from GA-RS were more likely to
275 be homozygous for alleles from GA-R (GG) than heterozygous with one allele from GA-R and
276 the other from LAB-S (GL) by comparing the frequencies of each genotype in the resistant
277 samples from generations F22, F23, and F26 to the control samples from generation F26. We
278 calculated the dominance parameter *h*, which varies from 0 for recessively inherited resistance to
279 1 for dominantly inherited resistance (Liu and Tabashnik 1997), using the genotype frequencies
280 at marker 4 in the F22, F23, and F26 for resistant, control, and susceptible larvae (Table S3).

281 To evaluate the relationship between genotype and resistance, we compared their trajectories
282 across generations in GA-RS. We used linear regressions in R 4.1.0 to test the null hypothesis of
283 no change in the $\log_{10}$ of percentage survival to third instar. Bioassays used 1 or 10 micrograms
284 of Cry1Ac per cm^2 in generations F2, F22 and F23, and only the higher concentration in the F26
285 test.

286

287 **RNA-seq and candidate gene analysis**

288 To generate samples for RNA sequencing, we reared LAB-S, GA-R, and GA-RS (generation 26)
289 individuals on untreated diet as described above in October 2020. When the larvae of the
290 parental strains LAB-S and GA-R reached the third instar, we dissected midguts from 15

291 individuals and froze them in groups of five, generating three biological replicates per strain. For
292 the GA-RS heterogeneous strain, we dissected 95 third instar midguts and froze them
293 individually, while simultaneously freezing the remainder of the carcass separately in wells of a
294 PCR plate. We extracted DNA from each sample using a squish extraction (Gloor *et al.* 1993)
295 using 50uL of buffer. We screened samples HRM as above at marker 4, which was one of two
296 sites we found to be the most strongly associated with resistance. We selected 15 individuals that
297 were homozygous for the LAB-S genotype at this site (henceforth "LL") and 15 individuals
298 homozygous for the GA-R genotype at this site (henceforth "GG") and pooled the midguts
299 corresponding to these samples into groups of five, again generating three biological replicates
300 for each genotype. Selecting genotypes in this way allowed us to isolate the effects that the
301 chromosome 13 QTL has on gene expression, giving us the potential to detect *trans*-regulatory
302 effects. We extracted RNA from all 12 midgut samples using ZYMO Direct-zol RNA Miniprep
303 Kit kits and built paired-end libraries with KAPA stranded mRNA seq kits. Libraries were
304 sequenced in part on an Illumina NextSeq at the University of Arizona Genetics Core (UAGC;
305 Tucson, AZ, USA) and part on an Illumina NovaSeq at Novogene.

306 We trimmed reads using Trimmomatic (Bolger *et al.* 2014) and aligned them to our *H. zea*
307 assembly with Hisat2 (Kim *et al.* 2019). We built genome-guided transcriptome assemblies for
308 each sample using StringTie (Pertea *et al.* 2016) and used StringTie merge to create a single
309 transcriptome. We quantified read abundance for each sample using Salmon (Patro *et al.* 2017)
310 and combined its transcript-level counts into gene-level counts with the R package tximport
311 (Soneson *et al.* 2016). We analyzed differential expression using FDR-corrected *P*-values from
312 negative binomial models at the gene level with edgeR (Robinson *et al.* 2010), after filtering and
313 normalization for library-size differences. We performed statistical comparisons of LAB-S to
314 GA-R and LL to GG. We performed a one-tailed (to account for directionality of gene
315 expression differences) χ^2 test to examine whether the overlap of differentially expressed (DE)
316 genes was more than expected by chance.

317 In addition to analyzing global differential expression, we used the RNA-seq data to better
318 understand gene structure and expression within the candidate QTL. For each sample, we used
319 Trinity (Haas *et al.* 2013) to construct *de novo* transcriptome assemblies for each sample. Using
320 tblastn, we identified the transcripts in each Trinity assembly corresponding to all StringTie

321 transcripts from the region from bp 4,370,000 – 4,620,000. We then took the longest transcript
322 from each Trinity assembly and used orfipy (Singh and Wurtele 2021) to extract the longest open
323 reading frame (ORF) for each gene. Next we compared the ORFs from each of the 12 samples,
324 looking for differences in predicted protein structure between the samples with resistant
325 chromosome 13 genotypes (GA-R and GG) and susceptible ones (LAB-S and LL). We
326 quantified midgut abundance of all genes expressed in this region with average log count per
327 million reads across all samples produced by edgeR.

328 We analyzed the putative *kinesin-12* gene computationally using blastp to find homologous
329 proteins, InterProScan 5 (Jones *et al.* 2014) to search for additional protein domains, SignalP 5.0
330 (Armenteros *et al.* 2019) to examine if the protein contained signal peptides, and DeepGOPlus
331 (Kulmanov and Hoehndorf 2020) to predict gene ontology (GO; Ashburner *et al.* 2000)
332 categorization based on the sequence.

333

334 **Genotyping of GA strain and field samples for kinesin-12 mutation**

335 We collected *H. zea* larvae from the field in Maricopa, Arizona in October 2020 and Tifton,
336 Georgia in July 2021. Both populations had high resistance to Cry1Ac (Yu *et al.* 2021; Y.
337 Carrière, unpubl. data). We reared larvae to adults in the lab and collected tissue from 25 of the
338 Maricopa samples and 39 of the Tifton samples. We extracted DNA using a squish extraction
339 (Gloor *et al.* 1993) in 150 µl of buffer. We analyzed five F72 GA individuals sequenced in
340 October 2016 (Fritz *et al.* 2020). We downloaded raw reads from NCBI (PRJNA599999),
341 trimmed them using Trimmomatic, aligned them to the GA-R genome with bwa-mem, and
342 identified the frequency of the C546T mutation with samtools and VarScan. We additionally
343 collected tissue from 20 F87 adults from the GA strain in May 2018 and extracted DNA with
344 Qiagen DNeasy Blood and Tissue Kits. We designed primers (Table S2) to amplify the region
345 surrounding the *kinesin-12* mutation causing a stop codon. We performed PCR as above,
346 although with the addition of final concentration 0.1 µg/µl bovine serum albumin (Sigma-
347 Aldrich, St. Louis, MO, USA) due to PCR inhibitors. Sanger sequencing was done at Eurofins
348 Genomics as described above. We screened sequences manually in Geneious Prime (Biomatters,

349 Auckland, NZ) for presence of the target mutation.

350

351 **Analysis of 11 genes previously implicated in lepidopteran resistance to Cry1Ac**

352 We used our results from GWAS and RNA-seq to test the hypothesis that one or more of 11
353 genes previously implicated in lepidopteran resistance to Cry1Ac contributed to resistance in our
354 strains. For each gene, we present the lowest *P*-value from the original GWAS of SNPs between
355 the start and end of the gene in the G´ analysis. We also report the FDR corrected *P*-values from
356 differential expression analyses in edgeR comparing GA-R versus LAB-S and within GA-RS
357 comparing individuals with both alleles from GA-R (GG) versus those with both alleles from
358 LAB-S (LL).

359

360 **Results**

361 **Chromosome-level assembly of the genome of resistant strain GA-R**

362 We generated a *de novo* chromosome-level assembly of the GA-R genome, with 31
363 chromosomes largely syntenic with those of *H. armigera* (Pearce *et al.* 2017; Valencia-Montoya
364 *et al.* 2020), although with different inversion karyotypes for 19 chromosomes (Fig. S1). The
365 genome has 375.2 Mb with 33.0% repeat content and 36.9% GC content. Of 5,286 conserved
366 single-copy lepidopteran genes, our genome has 98.7% present, 98.4% complete, and 97.9%
367 complete and single-copy. Thus, our GA-R genome is likely complete and highly accurate.
368 Relative to the previous *H. zea* genome assembly (Pearce *et al.* 2017), the new assembly has a
369 64-fold greater N50, 10% more bp, 15% more complete BUSCO genes, and double the
370 percentage of repeat content.

371

372 **Genomic comparison between GA-R and the susceptible strain LAB-S**

373 Sequencing of 30 larvae from GA-R and 30 larvae from the unrelated susceptible strain LAB-S
374 revealed 165,416 fixed differences between the strains, as well 941,146 variable sites in GA-R
375 and 911,946 in LAB-S. Analysis of Tajima's D showed many regions with low genetic variation
376 throughout the genome in both strains (Fig. S2; S3). For both strains, these regions could reflect
377 genetic drift or selective sweeps caused by adaptation to lab conditions. For GA-R, regions of
378 low variation could also reflect selection for resistance in the lab.

379

380 **Genome-Wide Association Study (GWAS) of resistance to Cry1Ac in GA-RS**

381 We created the GA-RS strain by crossing GA-R and LAB-S. Both the Z-score and G′ sliding
382 window analyses of 1,578,733 SNPs from the GWAS using pools of resistant and susceptible
383 larvae from the F12 generation of GA-RS identified a region associated with resistance on
384 chromosome 13 (Figs. 1 and S4). G′ analysis via QTLseqr defined this QTL as a region from 4.0
385 to 6.5 Mb. This QTL contains 117 SNPs associated with resistance at $P < 5 \times 10^{-8}$. All of these
386 117 SNPs showed the expected relationship with the parental strain. 108 alleles found at higher
387 frequency in the resistant larvae were more common in GA-R than LAB-S and the remaining 9
388 observed at higher frequency in the susceptible larvae were more common in LAB-S than GA-R.

389 Outside of this QTL, only 18 SNPs were associated with resistance at $P < 5 \times 10^{-8}$.
390 Chromosomes 10 and 15 had two of these SNPs. No other chromosome had more than one. In
391 the G′ analysis, only the QTL on chromosome 13 was significantly associated with resistance
392 (Fig. S4). We henceforth refer to this QTL as the *r1* locus (resistance locus 1).

393 Analysis of Z-scores from the GWAS shows that SNP sites that differed significantly between
394 resistant and susceptible larvae were not evenly distributed across the chromosome 13 QTL (Fig.
395 2A). Eighteen of the 25 windows of 10 Kb with the top 5% significant SNP density were
396 clustered between 4.42 and 4.60 Mb. Consistent with the GWAS results, Tajima's D provides
397 evidence for a selective sweep in GA-R between 4.3 and 5.2 Mb on chromosome 13 (Fig. 2B).

398

399 **Fine-scale mapping within the *r1* locus**

400 We used HRM to genotype individual resistant and susceptible larvae from the F22 and F23 of
401 GA-RS for SNPs at 12 marker sites within *r1*. This revealed eight markers (2 to 9) from 4.3 to
402 5.0 Mb significantly associated with resistance (Table 1). For marker 4 at 4.5 Mb (n = 57) and
403 marker 5 at 4.6 Mb (n = 60), all resistant larvae genotyped from GA-RS were either homozygous
404 for the allele from the resistant GA-R strain (GG) or heterozygous, with one allele from GA-R
405 and the other from the susceptible LAB-S strain (GL) (Table 1). A similar test using only
406 resistant larvae from the F26 confirmed this result: all 23 resistant larvae genotyped were either
407 GG (16) or GL (7). By contrast, the three genotypes were in Hardy-Weinberg equilibrium in 89
408 larvae genotyped from a control sample from the F26 that was not screened for resistance and
409 thus contained a mixture of resistant and susceptible individuals (24 GG: 44 GL: 21 LL, χ^2 =
410 0.0091; P = 0.52). Genotype frequency in the F26 larvae differed significantly between the
411 resistant larvae and the control larvae (χ^2 = 22.60, P = 1.2 × 10^{-5}).

412 The results from the GWAS, Tajima's D, the G′ analysis, analysis of SNP density, and fine-scale
413 mapping (Figs. 1 and 2 and Table 2), identify the region between 4.3 to 4.6 Mb as most likely to
414 contain the mutation(s) causing the effect of chromosome 13 on resistance to Cry1Ac.

415

416 **Gene expression in the midgut and a stop codon in *r1***

417 Based on the results above and annotations from funannotate and StringTie, we focused on the
418 10 genes between 4.37 and 4.62 Mb on chromosome 13. Seven of these 10 genes were expressed
419 in the midgut of third instar larvae (Fig. 3A; Table 2). The most highly expressed gene encodes a
420 wild-type protein of 308 amino acids that has sequence identity of 97.4% with *kinesin-related*
421 *protein 12* in *H. armigera* (XP_021193241) and 85.9% with its homolog in *Chloridea virescens*
422 (formerly *Heliothis virescens*; PCG76683.1; Fig. S5). Hereafter, we refer to this gene in *H. zea*
423 as *kinesin-12*.

424 In GA-R we found a point mutation (C546T) in *kinesin-12* that introduces a premature stop
425 codon expected to truncate the protein at 183 amino acids (Fig. 3B). This mutation occurred in

426 100% of reads from 30 GA-R larvae and in 0% of reads from 30 LAB-S larvae that were
427 sequenced in the genomic comparison between these strains. In the GWAS with GA-RS, this
428 mutation was more common in resistant larvae (71%) than susceptible larvae (32%; $P = 7.48 \times$
429 10^{-6}). It occurs at bp 4,547,246, between the two markers (4 and 5) most tightly associated with
430 resistance to Cry1Ac (Table 1). Furthermore, RNA-seq near marker 4 detected the C546T
431 mutation in 100% of reads of 15 GG larvae and 0% of reads of 15 LL larvae, confirming
432 complete linkage between this mutation and marker 4. All of this evidence identifies the C546T
433 mutation in *kinesin-12* as a strong candidate for underlying the *r1* allele and contributing to
434 resistance to Cry1Ac.

435 *Kinesin-12* is a taxonomically-restricted gene with no homology outside of Lepidoptera and poor
436 conservation (< 65% sequence identity) outside of the *Helicoverpa*/*Chloridea* clade. The
437 identification as a kinesin derives from the annotation in *H. armigera* of a conserved *EnvC*
438 domain
439 (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE=XP_02119
440 3241.1) with roles in cell cycle control, cell division, and chromosome partitioning. This domain
441 extends from amino acids 116-210 and would be disrupted in the mutant GA-R form of the
442 protein, indicating a potentially major effect. Likewise, InterProScan found a coil domain (aa
443 124-194) that would be disrupted. Analysis with SignalP found no evidence for a signal peptide,
444 indicating that this gene is not likely to be integrated into or secreted through the cell membrane.
445 The most specific of the GO terms associated with this protein (Table S4) is GO:0043232
446 (intracellular non-membrane-bounded organelle), which is most closely associated with kinesin-
447 related proteins in *Drosophila melanogaster*
448 (http://amigo.geneontology.org/amigo/term/GO:0043232).

449

450 ***Kinesin-12* mutation in the GA strain and in field samples**

451 To test the hypothesis that the C546T mutation in *kinesin-12* originated in the field, we
452 determined its frequency in the GA strain of *H. zea*, which was selected for resistance in the
453 field, but not in the lab (Brévault *et al.* 2013, Welch *et al.* 2015). In GA, the frequency of the

454 C546T mutation was 80% in five larvae from the F72 generation (three with homozygous TT
455 genotypes and two with heterozygous CT genotypes) and 97.5% in 20 larvae from the F87
456 generation (19 homozygous TT and one heterozygous CT). These results imply the C546T
457 mutation originated in the field population from which GA was established. The high frequency
458 of C546T in GA after rearing for many generations in the lab without exposure to Bt toxins
459 implies this mutation caused little or no fitness cost when reared in the lab in the absence of Bt
460 toxins. However, we did not find this mutation in 39 individuals collected from the field in
461 Georgia in July 2021 or in 25 individuals derived from the field in Arizona in 2020, despite the
462 resistance to Cry1Ac in both of these field populations (Yu *et al.* 2021; Y. Carrière, unpubl.
463 data). Whereas all individuals from the Georgia 2021 sample had the same sequence as LAB-S at
464 the codon starting with bp 546, three individuals from Arizona had a single base pair substitution
465 (C546A) changing the encoded amino acid from glutamine to lysine. In the field samples from
466 Arizona in 2020 and Georgia in 2021, we detected no insertions, deletions, or other mutations
467 introducing a stop codon in the 200 bp upstream or downstream from the C546T mutation.

468

469 **Inheritance and trajectory of resistance in GA-RS**

470 The genotype frequencies at marker 4 in resistant and susceptible larvae indicate at least one
471 GA-R allele at this locus was necessary for survival at 10 mcg Cry1Ac per cm^2 diet (Table 1).
472 However, a substantial proportion of susceptible larvae were homozygous for the GA-R allele at
473 marker 4 (Table 1). Together these results suggest that the *r1* allele was necessary, but not
474 sufficient for resistance to Cry1Ac in our screening bioassay.

475 Compared to 89 control samples not screened on diet tested with Cry1Ac, 69 resistant larvae
476 from the F22, F23, and F26 generations had a significantly higher ratio of the GG genotype to
477 the GL genotype (Fisher's Exact Test; $P < 0.0001$). At marker 4, the *r1* allele had a value for
478 dominance (h) of 0.23 (Table S3), which is intermediate between completely recessive
479 inheritance ($h = 0$) and additive inheritance ($h = 0.5$).

480 To test the hypothesis that alleles at one or more other loci contributed to resistance, we
481 compared the trajectory of resistance based on bioassay data with the trajectory of the GA-R

482 allele at marker 4. Resistance to Cry1Ac decreased substantially over time (Fig 4; Table S1),
483 indicating that in the absence of Cry1Ac, a pleiotropic fitness cost was associated with one or
484 more alleles contributing to resistance. However, the frequency of the GA-R allele at marker 4
485 was 0.52 in 89 control larvae from the F26, which is not different than the expected 0.50 in the
486 F1. This suggests no fitness cost was associated with the *r1* allele, which is tightly linked with
487 GA-R allele at marker 4. In the F26, marker 4 was in Hardy-Weinberg equilibrium as noted
488 above, confirming the absence of selection at this locus in GA-RS. The significant decrease in
489 resistance over time despite no decline in the frequency of the GA-R allele at marker 4 implies
490 the decrease in resistance was caused by reduced frequency of one or more resistance alleles that
491 carry a fitness cost in the absence of Cry1Ac and are not linked with the *r1* allele.

492

493 **Analysis of gene expression using RNA-seq**

494 To test the hypothesis that differential gene expression contributes to resistance, we used RNA-
495 seq to compare transcript levels between GA-R and LAB-S and between the GG and LL
496 genotypes within GA-RS. After filtering, we analyzed expression of 12,965 genes (Table S5).
497 We found 2,173 differentially expressed (DE) genes between the unrelated strains LAB-S and
498 GA-R (Table S6) and 23 DE genes between the GG and LL genotypes within GA-RS (Table
499 S7). None of the genes associated with *r1* in chromosome 13 differed significantly in expression
500 between GA-R and LAB-S or between GG and LL.

501 Twelve of the 23 DE genes between GG and LL were also among the set of 2,173 DE genes in
502 the parental strain comparison, of which nine were DE in the same direction in both comparisons
503 (higher expression in GA-R than LAB-S and in GG than LL and vice versa; Tables S6 and S7).
504 The overlap in DE genes between these two datasets is significantly greater than expected by
505 chance ($\chi^2 = 3.86$; $P = 0.025$), implying the within-strain comparison between GG and LL
506 reflects meaningful differences between the parental strains. However, none of the 23 DE genes
507 between GG and LL are among the 11 genes previously implicated in resistance to Cry1Ac in
508 lepidopterans (Table S7). Thus, we cannot rule out transcription as a mechanism mediating the
509 effect of *r1* on resistance, but the RNA-seq data provide no compelling evidence for this.

510

511 **Analysis of 11 genes previously implicated in lepidopteran resistance to Cry1Ac**

512 We used our results from GWAS and RNA-seq to evaluate potential contributions to resistance
513 of 11 genes previously implicated in lepidopteran resistance to Cry1Ac (Table 3). None of these
514 candidate genes are in the resistance-associated QTL on chromosome 13 or the putative minor
515 effect QTL on chromosome 10 (Table 3). In addition, none of them had any SNPs that were
516 significantly associated with resistance in the GWAS (Table 3). Although *tetraspanin-1* is on
517 chromosome 10 in *H. zea*, it is outside the regions of this chromosome that were moderately
518 associated with resistance in the GWAS. As noted above, none of the 23 DE genes between GG
519 and LL are among the 11 candidate genes (Table S7). Only one of the 11 pairwise comparisons
520 between strains based on RNA-seq showed a significant difference. Expression of *ABCC1* was
521 significantly lower in GA-R than LAB-S ($P = 0.0014$, Table 3; Table S6). However, within GA-
522 RS, expression of *ABCC1* did not differ significantly between GG and LL ($P = 0.29$; Table 3;
523 Table S7), which indicates reduced expression of *ABCC1* was not genetically linked with
524 resistance.

525

526 **Discussion**

527 We report three key results demonstrating a novel genetic basis of Cry1Ac resistance in the GA-
528 R strain of *H. zea*, which resulted from field selection followed by lab selection (Brévault *et al.*
529 2013, Welch *et al.* 2015). First, resistance was associated with a 250kb region of chromosome 13
530 that contains no genes with a previously identified role in Bt resistance or toxicity. Second,
531 within this region, resistance to Cry1Ac was associated with a point mutation that introduces a
532 premature stop codon in a novel candidate gene, *kinesin-12*. Third, we report evidence that one
533 or more other loci also contributed to resistance to Cry1Ac. To facilitate these advances, we built
534 the first chromosome-level genome assembly for *H. zea*, adding to a growing set of highly
535 contiguous genomes for lepidopteran pests (Chen *et al.* 2019; Ward *et al.* 2021; Yan *et al.* 2021).
536 This genome was essential for the genetic mapping reported here and will serve as a resource for
537 other genomic investigations into the biology of *H. zea*.

538 Our finding of a novel gene associated with resistance to Cry1Ac in *H. zea* adds to an
539 increasingly diverse list of genes associated with Bt resistance in lepidopterans (Jin *et al.* 2018;
540 Guo *et al.* 2021; Jurat-Fuentes *et al.* 2021). However, the novel genetic basis of resistance does
541 not necessarily imply a novel biochemical mechanism of resistance. The effects of *r1* could be
542 mediated by either decreased toxin activation or reduced binding of Cry1Ac to larval midgut
543 membranes, which are well known mechanisms of Bt resistance (Peterson *et al.* 2017; Jurat-
544 Fuentes *et al.* 2021). Indeed, previous studies of *H. zea* have found decreased protoxin activation
545 in GA-R (Zhang *et al.* 2019) and reduced binding of Cry1Ac to larval midgut preparations in
546 strains with resistance caused by knocking out the putative receptor ABCC2 (Perera *et al.* 2021).

547 The location of *r1* on chromosome 13 is noteworthy because it corresponds closely to the region
548 under the strongest selection in *H. zea* populations from Louisiana that were exposed to Bt crops
549 over the past 19 years (Fritz *et al.* 2021). Although Fritz *et al.* (2021) identified a narrow region
550 near but not containing *r1* as the most likely site of selection (~4.0 Mb in our assembly), the
551 broader region associated with the selective sweep in their data includes *r1* (~3.8 to 5.8 Mb) and
552 aligns with both our original and refined QTL for resistance. Thus, we hypothesize that *r1* is
553 associated with Cry1Ac resistance in the field populations of *H. zea* from Louisiana studied by
554 Fritz *et al.* (2021), as well as in GA-R and its parent strain GA (Fritz *et al.* 2020), which
555 originated from a field-selected population in Georgia in 2008 (Brévault *et al.* 2013).

556 We hypothesize that the effect of *r1* on resistance is mediated by the premature stop codon in the
557 *kinesin-12* gene, which shortens the predicted protein by roughly 40%. Of the 10 genes in the
558 region tightly associated with resistance, midgut expression was highest for *kinesin-12*,
559 indicating a likely midgut function for the protein it encodes. Protein functional prediction
560 algorithms including DeepGoPlus provide moderate support for the original annotation as a
561 kinesin with a function in intracellular transport or structure. Nonetheless, we do not know the
562 normal function of the kinesin-12 protein and cannot infer that the C546T mutation causes
563 resistance. In future work, we aim to determine if knocking out this gene in a susceptible strain
564 causes resistance to Cry1Ac and reduces binding of Cry1Ac to larval midgut membranes, and
565 conversely if knocking in the wild type sequence into kinesin-12 from GA-R restores
566 susceptibility to and binding of Cry1Ac.

567 Kinesins and kinesin-related proteins are motor proteins important in microtubule function,
568 chromosomal movement, and organelle transport (Ali and Wang 2020) that have not been
569 associated previously with Bt resistance. Several kinesins are involved in MAPK signaling
570 cascades (Liang and Yang 2019) and MAPK signaling is implicated in Bt resistance (Guo *et al.*
571 2015, 2020, 2021; Qin *et al.* 2021). Furthermore, a case of xenobiotic resistance in mice involved
572 a mutant kinesin acting downstream of a map kinase (Watters *et al.* 2001). Thus, one hypothesis
573 is that *kinesin-12* acts downstream of MAPK as part of the signaling cascade initiated when
574 Cry1Ac binds to a midgut receptor. However, MAPK influences Bt resistance via
575 downregulation of Bt receptors (Guo *et al.* 2015, 2020, 2021; Qin *et al.* 2021). Here we see no
576 evidence for reduced transcription of putative receptors, making this an unlikely explanation for
577 the link between *kinesin-12* and resistance.

578 Kinesins also play a role in the localization of transmembrane proteins to the cell membrane
579 (Jana *et al.* 2021) and in intracellular cadherin trafficking (Phang *et al.* 2014). The transport
580 functions of kinesins and kinesin-related proteins entail motor complexes of three or more
581 proteins (Phang *et al.* 2014), suggesting interactions between different proteins could be
582 interrupted to interfere with receptor transport to the membrane. We hypothesize that in *H. zea,* a
583 truncated *kinesin-12* in combination with mutations affecting one or more of its interacting
584 partners blocks proper localization of a Bt receptor on the membrane.

585 Although necessary for resistance, the *r1* allele alone was not sufficient to cause resistance,
586 implying contributions from one or more additional loci. Polygenic resistance to Cry1Ac also has
587 been reported in strains of *H. zea* unrelated to GA-R (Caccia *et al.* 2012; Fritz *et al.* 2021; Perera
588 *et al.* 2021) and is consistent with the decline in resistance over time in GA-RS. While resistance
589 to Cry1Ac declined significantly across generations in GA-RS, the frequency of C546T and
590 other *r1* linked alleles did not. After 22, 23 and 26 generations without exposure to Bt toxins, it
591 was not lower than its expected initial frequency of 0.50. Thus, the decline in resistance reflects a
592 decreased resistance allele frequency at one or more other loci. Unlike the C546T mutation,
593 which did not have a substantial fitness cost in the lab, the observed decline in resistance
594 suggests that a fitness cost was associated with at least one mutation at another locus that
595 contributed to resistance in GA-RS.

596 The results summarized above have implications for understanding the trajectory of the C546T
597 mutation in the field. The high frequency of the C546T mutation in the field-selected GA strain
598 implies this mutation was common in 2008 in the moderately resistant field population in
599 Georgia from which GA was derived (Brévault *et al.* 2013). However, we did not detect this
600 mutation in Cry1Ac-resistant populations from the same site in Georgia in 2021 or in Arizona in
601 2020. Thus, this mutation is not associated with resistance to Cry1Ac in all field populations of
602 *H. zea*. Also, its frequency apparently decreased in the field in Georgia from 2008 to 2021. One
603 hypothesis is that the frequency of this mutation decreased in Georgia because it has a substantial
604 fitness cost under field conditions, such as reduced overwintering survival (Carrière *et al.* 2001),
605 that would not be evident in the lab. The C546T mutation could have been replaced by one or
606 more mutations that have lower fitness costs (Guillemaud *et al.* 1998), confer higher resistance to
607 Cry1Ac, and/or confer resistance without contributions from mutations at other loci. More
608 research is needed to determine the function of kinesin-12 and its role in resistance, as well as the
609 genetic basis of Cry1Ac resistance in current field populations of *H. zea.*

610

611 **Data availability**

612 All sequence data will be made available at NCBI (Bioproject: PRJNA767434). Phenotypic data
613 will be made available at OSF.

614 **Author contributions**

615 J.F., Y.C., B.E.T., and L.M.M. conceived and designed project. K.M.B., C.W.A., B.A.D., and
616 X.L. performed experiments and collected the data. K.M.B., C.W.A., Y.C., and B.E.T. analyzed
617 data. K.M.B. and B.E.T. wrote the manuscript with input from all authors.

618

619 Funding

620 This work was supported by grants from the USDA National Institute of Food and Agriculture
621 (Agriculture and Food Research Initiative 2020-67013-31924 and Biotechnology Risk

Assessment Research Grants Program 2020-33522-32268), Corteva Agriscience, and the Cotton Insect Resistance Management (IRM) Technical Subcommittee of the Agricultural Biotechnology Stewardship Technical Committee (ABSTC).

Conflicts of interest

As noted above, support for this study was provided in part by Corteva Agriscience and the Cotton IRM Technical Subcommittee of the ABSTC.

Acknowledgments

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer. We thank Xinzhi Ni for sending *H. zea* from Georgia, Alex Yelich and Chandran Unnithan for help with insect rearing and dissections, Yidong Wu and David Heckel for valuable discussion on preliminary versions of the manuscript, Jon Galina-Mehlman, Jayson Talag, and Dave Kudrna for their assistance and advice regarding PacBio and Illumina sequencing.

Literature cited

Ali MI, Luttrell RG, Young III SY. 2006. Susceptibilities of *Helicoverpa zea* and *Heliothis virescens* (Lepidoptera: Noctuidae) populations to Cry1Ac insecticidal protein. J Econ Entomol. 99:164-175.

Ali I, Yang W-C. 2020. The functions of kinesin and kinesin-related proteins in eukaryotes. Cell Adh Migr. 14: 139-152.

Armenteros JJA, Tsirigos KD, Sønderby CK, Peterson TN, Winther O, *et al.* 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotech. 37:420-

647 423.
648 Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, *et al.* 2000. GeneOntology: tool for the
649 unification of biology. Nat Genet. 25:25-29.
650 Barsh GS, Copenhaver GP, Gibson G, Williams SM. 2012. Guidelines for genome-wide
651 association studies. PLoS Genet. 8:e1002812.
652 Baxter SW, Badenes-Pérez FR, Morrison A, Vogel H, Crickmore N, *et al.* Parallel evolution of
653 *Bacillus thuringiensis* toxin resistance in Lepidoptera. Genetics. 189:675-679.
654 Benowitz KM, Coleman JM, Matzkin LM. 2019. Assessing the architecture of *Drosophila*
655 *mojavensis* locomotor evolution with bulk segregant analysis. G3. 9:1767-1775.
656 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
657 data. Bioinformatics. 30:2114-2120.
658 Bravo A, Likitvivatanavong S, Gill SS, Soberón M. 2011. *Bacillus thuringiensis*: a story of a
659 successful bioinsecticide. Ins Biochem Mol Biol. 41:423-431.
660 Brévault T, Heuberger S, Zhang M, Ellers-Kirk C, Ni X, *et al.* 2013. Potential shortfall of
661 pyramided transgenic cotton for insect resistance management. Proc Natl Acad Sci USA.
662 110:5806-5811.
663 Brévault T, Tabashnik BE, Carrière Y. 2015. A seed mixture increases dominance of resistance
664 to Bt cotton in *Helicoverpa zea*. Sci Rep. 5:9807.
665 Caccia S, Moar WJ, Chandrashekhar J, Oppert C, Anilkumar KJ, *et al.* 2012. Association of
666 Cry1Ac toxin resistance in *Helicoverpa zea* (Boddie) with increased alkaline phosphatase
667 levels in the midgut lumen. Appl Env Microbiol. 78:5690-5698.
668 Calles-Torrez V, Knodel JJ, Boetel MA, French BW, Fuller BW, *et al.* 2019. Field-evolved
669 resistance of northern and western corn rootworm (Coleoptera: Chrysomelidae)
670 populations to corn hybrids expressing single and pyramided Cry3Bb1 and Cry34/35Ab1
671 Bt proteins in North Dakota. J Econ Entomol. 112:1875-1886.
672 Carrière Y, Ellers-Kirk C, Patin AL, Sims M, Meyer S, *et al.* 2001. Overwintering cost
673 associated with resistance to transgenic cotton in the pink bollworm. J. Econ. Entomol.
674 94:935-941.
675 Carrière Y, Degain BA, Unnithan GC, Harpold VS, Heuberger S, Li X, *et al.* 2018. Effects of
676 seasonal changes in cotton plants on the evolution of resistance to pyramided cotton
677 producing the Bt toxins Cry1Ac and Cry1F in *Helicoverpa zea*. Pest Man Sci. 74:627-

678 637.
679 Carrière Y, Degain B, Unnithan GC, Harpold VS, Li X, *et al.* 2019. Seasonal declines in Cry1Ac
680 and Cry2Ab concentration in maturing cotton favor faster evolution of resistance to
681 pyramided Bt cotton in Helicoverpa zea. J Econ Entomol. 112:2907–2914.
682 Carrière, Y, Brown Z, Aglasan S, Dutilleul P, Carroll M, *et al.* 2020a. Crop rotation mitigates
683 impacts of corn rootworm resistance to transgenic Bt corn. Proc Natl Acad Sci USA.
684 117:18385-18392.
685 Carrière Y, Degain BA, Harpold VS, Unnithan GC, Tabashnik BE. 2020b. Gene flow between
686 Bt and non-Bt plants in a seed mixture increases dominance of resistance to pyramided Bt
687 corn in *Helicoverpa zea* (Lepidoptera: Noctuidae). J Econ Entomol. 113:2041–2051.
688 Carrière Y, Degain BA, Tabashnik BE. 2021. Effects of gene flow between Bt and non-Bt plants
689 in a seed mixture of Cry1A.105 + Cry2Ab corn on performance of corn earworm in
690 Arizona. Pest Manag Sci. 77: 2106–2113
691 Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016. Contiguous and accurate de
692 novo assembly of metazoan genomes with modest long read coverage. Nuc Ac Res. 44:
693 e147.
694 Chen L, Wei J, Liu C, Niu L, Zhang C, *et al.* 2019. Effect of midgut specific binding protein
695 ABCC1 on Cry1Ac toxicity against *Helicoverpa armigera*. Sci Ag Sin. 52:3337-3345.
696 Chen W, Yang X, Tetreau G, Song X, Coutu C, *et al.* 2019. A high-quality chromosome-level
697 genome assembly of a generalist herbivore, *Trichoplusia ni*. Mole Ecol Res. 19:485-496.
698 Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, *et al.* 2013. Nonhybrid, finished
699 microbial genome assemblies from long-read SMRT sequencing data. Nat Methods.
700 10:563-569.
701 Cook DR, Threet M. 2019. Cotton insect losses – 2019.
702 https://www.entomology.msstate.edu/resources/2019loss.php
703 Dively GP, Venugopal PD, Bean D, Whalen J, Holmstrom K, *et al.* 2018. Regional pest
704 suppression associated with widespread Bt maize adoption benefits vegetable growers.
705 Proc Natl Acad Sci USA. 115:3320–3325.
706 Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, *et al.* 2020. RepeatModeler2 for automated
707 genomic discovery of transposable element families. Proc Natl Acad Sci USA. 117:9451-
708 9457.

709 Fritz ML, Nunziata SO, Guo R, Tabashnik BE, Carrière Y. 2020. Mutations in a novel cadherin
710 gene associated with Bt resistance in *Helicoverpa zea*. G3. 10:1563-1574.
711 Fritz ML, Hamby KA, Taylor K, DeYonke AM, Gould F. 2021. Genome evolution in an
712 agricultural pest following adoption of transgenic crops. bioRxiv.
713 DOI:10.1101/2020.10.05.326801.
714 Gahan LJ, Gould F, Heckel DG. 2001. Identification of a gene associated with Bt resistance in
715 *Heliothis virescens*. Science. 293:857-860.
716 Gloor GB, Preston CR, Johnson-Schlitz DM, Nassif NA, Phillis RW, *et al.* 1993. Type I
717 repressors of *P* element mobility. Genetics. 135:81-95.
718 Guillemaud T, Lenormand T, Bourguet D, Chevillon C, Pasteur N, *et al.* 1998. Evolution of
719 resistance in *Culex pipiens*: allele replacement and changing environment. Evolution.
720 52:443-453.
721 Guo Z, Kang S, Chen D, Wu Q, Wang S, *et al.* 2015. MAPK signaling pathway alters expression
722 of midgut ALP and ABCC genes and causes resistance to *Bacillus thuringiensis* Cry1Ac
723 toxin in diamondback moth. PLoS Genetics. 11:e1005124.
724 Guo Z, Kang S, Sun D, Gong L, Zhou J, *et al.* 2020. MAPK-dependent hormonal signaling
725 plasticity contributes to overcoming *Bacillus thuringiensis* toxin action in an insect host.
726 Nat Comms. 11:3003.
727 Guo Z, Kang S, Wu Q, Wang S, Crickmore N, *et al.* 2021. The regulation landscape of MAPK
728 signaling cascade for thwarting *Bacillus thuringiensis* infection in an insect host. PLoS
729 Path. 17:e1009917.
730 Haas BJ, Papanicolau A, Yassour M, Grabherr M, Blood PD, *et al.* 2013. *De novo* transcript
731 sequence reconstruction from RNA-seq using the Trinity platform for reference
732 generation and analysis. Nat Prot. 8:1494-1512.
733 Heckel DG, Gahan LJ, Baxter SW, Zhao J-Z, Shelton AM, *et al.* 2007. The diversity of Bt
734 resistance genes in species of Lepidoptera. J Inv Biol. 95:192-197.
735 ISAAA. 2019. Global status of commercialized biotech/GM crops: 2019. ISAAA Brief No. 55.
736 ISAA: Ithaca, NY.
737 Jaworski CC, Allan CW, Matzkin LM. 2020. Chromosome-level hybrid de novo genome
738 assemblies as an attainable option for nonmodel insects. Mol Ecol Res. 20:1277-1293.
739 Jin L, Wei Y, Zhang L, Yang Y, Tabashnik BE, *et al.* 2013. Dominant resistance to Bt cotton

740 and minor cross-resistance to Bt toxin in Cry2Ab in cotton bollworm from China. Evol
741 Appl. 6:1222-1235.
742 Jones P, Binns D, Chang HY, Fraser M, Li W, *et al.* 2014. InterProScan 5: genome-scale protein
743 function classification. Bioinformatics. 30:1236-1240.
744 Jurat-Fuentes JL, Karumbaiah L, Jakka SRK, Ning C, Liu C, *et al.* 2011. Reduced levels of
745 membrane-bound alkaline phosphatase are common to lepidopteran strains resistant to
746 Cry toxins from *Bacillus thuringiensis*. PLoS ONE. 6:e17606.
747 Jurat-Fuentes JL, Heckel DG, Ferré J. 2021. Mechanisms of resistance to insecticideal proteins
748 from *Bacillus thuringiensis*. Annu Rev Entomol. 66:121-140.
749 Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, *et al.* 2014. Efficient *de novo*
750 assembly of highly heterozygous genomes from whole-genome shotgun short reads.
751 Genom Res. 24:1384-1395.
752 Kaur G, Guo J, Brown S, Head GP, Price PA, *et al.* 2019. Field-evolved resistance of
753 *Helicoverpa zea* (Boddie) to transgenic maze expressing pyramided Cry1A.105.Cry2Ab2
754 proteins in northeast Louisiana, the United States. J Inv Pathol. 163:11-20.
755 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and
756 genotyping with HISAT2 and HISAT-genotype. Nat Biotech. 37:907-915.
757 Kofler R, Orozco-terWengel P, De Maio N, Pandey RV, Nolte V, *et al.* 2011. PoPoolation: a
758 toolbox for population genetic analysis of next generation sequencing data from pooled
759 individuals. PLoS ONE. 6:e15925.
760 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, *et al.* 2017. Canu: scalable and accurate
761 long-read assembly via adaptive k-mer weighting and repeat separation. Genom Res.
762 27:722-736.
763 Kulmanov M, Hoehndorf R. 2020. DeepGOPlus: improved protein function prediction from
764 sequence. Bioinformatics. 36:422-429.
765 Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods. 9:357-
766 359.
767 Lawrie RD, Mitchell III RD, Deguenon JM, Ponnusamy L, Reisig D, *et al.* 2020. Multiple
768 known mechanisms and a possible role of an enhanced immune system in Bt-resistance
769 in a field population of the bollworm, *Helicoverpa zea*: differences in gene expression
770 with RNAseq. Int J Mol Sci. 21:6528.

771 Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping
772 and population genetical parameter estimation from sequencing data. Bioinformatics.
773 27:2987-2993.
774 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
775 Bioinformatics. 25:1754-1760.
776 Liang Y-J, Yang W-X. 2019. Kinesins in MAPK cascade: how kinesin motors are involved in
777 the MAPK pathway? Gene. 684:1-9.
778 Liu Y, Tabashnik BE. 1997. Inheritance of resistance to the *Bacillus thuringiensis* toxin Cry1C
779 in the diamondback moth. Appl Env Microbiol. 63:2218-2223.
780 Liu C, Xiao Y, Li X, Oppert B, Tabashnik BE, *et al.* 2014. *Cis*-mediated down-regulation of a
781 trypsin gene associated with Bt resistance in cotton bollworm. Sci Rep. 4:7219.
782 Luttrell RG, Wan L, Knighten K. 1999. Variation in susceptibility of Noctuid (Lepidoptera)
783 larvae attacking cotton and soybean to purified endotoxin proteins and commercial
784 formulations of *Bacillus thuringiensis*. J Econ Entomol. 92:21-32.
785 Magwene PM, Willis JH, Kelly JK. 2011. The statistics of bulk segregant analysis using next
786 generation sequencing. PLoS Comp Biol. 7:e1002255.
787 Mansfeld BN, Grumet R. 2018. QTLseqr: an R package for bulk segregant analysis with next-
788 generation sequencing. Plant Genom. 11:180006.
789 Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, *et al.* 2018. MUMmer4: a fast
790 and versatile genome alignment system. PLoS Comp Biol. 14:e1005944.
791 Musser FR, Catchot AL, Conley SP, Davis JA, DiFonzo C, *et al.* 2019. 2018 Soybean insect
792 losses in the United States. *Midsouth Entomol* 12: 1-24.
793 National Academies of Sciences, Engineering, and Medicine. 2016. *Genetically Engineered*
794 *Crops: Experiences and Prospects*. Washington DC, National Academies Press.
795 Orpet RJ, Degain BA, Unnithan GC, Welch KL, Tabashnik BE, *et al.* 2015a. Effects of dietary
796 protein to carbohydrate ratio on Bt toxicity and fitness costs of resistance in *Helicoverpa*
797 *zea*. Entomol Exp Appl. 156:28-36.
798 Orpet RJ, Degain BA, Tabashnik BE, Carrière Y. 2015b. Balancing Bt toxin avoidance and
799 nutrient intake by *Helicoverpa zea* (Lepidoptera: Noctuidae) larvae. J Econ Entomol.
800 108:2581-2588.
801 Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-

802 aware quantification of transcript expression. Nat Meth. 14:417-419.
803 Pearce SL, Clarke DF, East PD, Elfekih S, Gordon KHJ, *et al.* 2017. Genomic innovations,
804 transcriptional plasticity and gene loss underlying the evolution and divergence of two
805 highly polyphagous and invasive *Helicoverpa* pest species. BMC Biol. 15:63.
806 Perera OP, Little NS, Abdelgaffar H, Jurat-Fuentes JL, Reddy GVP. 2021. Genetic knockouts
807 indicate that the ABCC2 protein in the bollworm *Helicoverpa zea* is not a major receptor
808 for the Cry1Ac insecticidal protein. Genes. 12:1522.
809 Pertea M, Pertea GM, Antonescu CM, Chang TC. Mendell JT, *et al.* 2015. StringTie enables
810 improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotech. 33:290-
811 295.
812 Peterson B, Bezuidenhout CC, Van den Berg J. 2017. An overview of mechanisms of Cry toxin
813 resistance in lepidopteran insects. J Econ Entomol. 110:362-377.
814 Qin J, Guo L, Ye F, Kang S, Sun D, *et al.* 2021. MAPK-activated transcription factor PxJun
815 suppresses *PcABCB1* expression and confers resistance to *Bacillus thuringiensis* Cry1Ac
816 toxin in *Plutella xylostella* (L.). Apple Env Microbiol. 87:13.
817 Rajagopal R, Arora N, Sivakumar S, Rao NGV, Nimbalkar SA, *et al.* 2009. Resistance of
818 *Helicoverpa armigera* to Cry1Ac toxin from *Bacillus thuringiensis* is due to improper
819 processing of the protoxin. Biochem J. 419:309-316.
820 Reisig DD, Huseth AS, Bacheler JS, Aghaee M-A, Braswell L, *et al.* 2018. Long-term and
821 observational evidence of practical *Helicoverpa zea* resistance to cotton with pyramided
822 Bt toxins. J Econ Entomol. 111:1824-1833.
823 Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, *et al.* 2014. Integrating mapping-,
824 assembly-, and haplotype-based approaches for calling variants in clinical sequencing
825 applications. Nat Genet. 46:912-918.
826 Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for
827 third-gen diploid genome assemblies. BMC Bioinf. 19:460.
828 Robinson MD, McCarthy DJ, Smythe GK. 2010. edgeR: a Bioconductor package for differential
829 expression analysis of digital gene expression. Bioinformatics. 26:139-140.
830 Romeis J, Naranjo SE, Meissle M, Shelton AM. 2018. Genetically engineered crops help support
831 conservation biological control. Biol Contr. 130:136–154.
832 Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation

833 completeness. In: Kollmar M (ed.) Gene Prediction. Methods in Molecular Biology, vol
834 1962. Humana, New York, NY. pp. 227-245.
835 Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q
836 manipulation. PLoS ONE. 11:e0163962.
837 Singh U, Wurtele ES. 2021. orfipy: a fasta and flexible tool for extracting ORFs. Bioinformatics.
838 37:3019-3020.
839 Smith JL, Farhan Y, Schaafsma AW. 2019. Practical resistance of *Ostrinia nubilalis*
840 (Lepidoptera: Crambidae) to Cry1F *Bacillus thuringiensis* maize discovered in Nova
841 Scotia, Canada. Sci Rep. 9:18247.
842 Soberón M, Pardo-López L, López I, Gómez I, Tabashnik BE, *et al.* 2007. Engineering modified
843 Bt toxins to counter insect resistance. Science. 318:1640-1642.
844 Soneson C, Love MI, Robinson MD. 2016. Differential analyses for RNA-seq: transcript-level
845 estimates improve gene-level inferences. F100Research. 4:1521.
846 Tabashnik BE and Carrière Y. 2019. Global patterns of resistance to Bt crops highlighting Pink
847 Bollworm in the United States, China, and India. J Econ Entomol. 112:2513-2523.
848 Tabashnik BE, Liu Y-B, Malvar T, Heckel DG, Masson L, *et al.* 1998. Insect resistance to
849 *Bacillus thuringiensis*: uniform or diverse? Phil Trans R Soc B. 353:1751-1756.
850 Tabashnik BE, Gassmann AJ, Crowder DW, Carrière Y. 2008. Insect resistance to Bt crops:
851 evidence versus theory. Nat Biotech. 26:199-202.
852 Tabashnik BE, Liesner LR, Ellsworth PC, Unnithan GC, Fabrick JA, *et al.* 2021. Genetically
853 engineered cotton synergizes eradication of the pink bollworm a century after its invasion
854 of the United States. Proc Natl Acad Sci USA. 118:e2019115118.
855 Valencia-Montoya WA, Elfekih S, North HL, Meier JI, Warren IA, *et al.* 2020. Adaptive
856 introgression across semipermeable species boundaries between local *Helicoverpa zea*
857 and invasive *Helicoverpa armigera* moths. Mol Biol Evol. 37:2568-2583.
858 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, *et al.* 2014. Pilon: an integrated tool for
859 comprehensive microbial variant detection and genome assembly improvement. PLoS
860 ONE. 9:e112963.
861 Wang B, Wei J, Wang Y, Chen L, Liang G. 2020. Polycalin is involved in the toxicity and
862 resistance to Cry1Ac toxin in *Helicoverpa zea*. Arch Ins Biochem Physiol. 104:e21661.
863 Ward CM, Perry KD, Baker G, Powis K, Heckel DG, *et al.* 2021. A haploid diamondback moth

(*Plutella xylostella* L.) genome assembly resolves 31 chromosomes and identifies a diamide resistance mutation. Ins Biochem Mol Biol. 138:103622.

Watters JW, Dewar K, Lehoczky J, Boyartchuk V, Dietrich WF. 2001. Kif1C, a kinesin-like motor protein, mediates mouse macrophage resistance to anthrax lethal factor. Curr Biol. 11: 1503-1511.

Welch KL, Unnithan GC, Degain BA, Wei J, Zhang J, *et al.* 2015. Cross-resistance to toxins used in pyramided Bt crops and resistance to Bt sprays in *Helicoverpa zea*. J Invert Pathol. 132:149-156.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, *et al.* 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nuc Ac Res. 42:D1001-D1006.

Xu H, Zhao X, Yang Y, Chen X, Mei Y, *et al.* 2021. Chromosome-level genome assembly of an agricultural pest, the rice leaffolder *Cnaphalocrocis exigua* (Crambidae, Lepidoptera). Mol Ecol Res. 21:561-572.

Yan B, Ou H, Wei L, Wang X, Yu X, *et al.* 2021. A chromosome-level genome assembly of *Ephestia elutella* (Hübner, 1796) (Lepidoptera: Pyralidae). Genom Biol Evol. 13:evab114.

Ye C, Ma ZS. 2016. Sparc: a sparsity-based consensus algorithm for long erroneous sequencing reads. PeerJ. 4:e2016.

Ye C, Hill CM, Wu S, Ruan J, Ma ZS. 2016. DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci Rep. 6:31900.

Yu W, Lin S, Dimase M, Niu Y, Brown S, *et al.* 2021. Extended investigation of field-evolved resistance of the corn earworm *Helicoverpa zea* (Lepidoptera: Noctuidae) to *Bacillus thuringiensis* Cry1A.105 and Cry2Ab2 proteins in the southeastern United States. J Inv Pathol. 183: 107560

Zhang S, Cheng H, Gao Y, Wang G, Liang G, *et al.* 2009. Mutation of an aminopeptidase N gene is associated with *Helicoverpa armigera* resistance to *Bacillus thuringiensis* Cry1Ac toxin. Ins Biochem Mol Biol. 39:421-429.

Zhang M, Wei J, Ni X, Zhang J, Jurat-Fuentes JL, *et al.* 2019. Decreased Cry1Ac activation by midgut proteases associated with Cry1Ac resistance in *Helicoverpa zea*. Pest Man Sci. 75:1099-1106.

895 Zhang S, Shen S, Peng J, Zhou X, Kong X, *et al.* 2020. Chromosome-level genome assembly of
896 an important pine defoliator, *Dendrolimus punctatus* (Lepidoptera; Lasiocampidae). Mol
897 Ecol Res. 20:1023-1037.
898

899 **Table 1** Genotype and allele frequencies at 12 markers on the chromosome 13 QTL for resistant
900 and susceptible GA-RS larvae from generations F22 and F23.
901

Marker	Bp (Chr13)	Resistant larvae		Susceptible larvae		
		Genotypes (GG / GL / LL)[a]	G allele frequency	Genotypes (GG / GL / LL)[a]	G allele frequency	p^b
1	4,110,560	41 / 15 / 4	0.81	23 / 23 / 3	0.70	1.0
2	4,281,434	30 / 23 / 7	0.69	9 / 24 / 16	0.43	9.6 e-03
3	4,379,538	41 / 18 / 1	0.83	13 / 24 / 12	0.51	2.4 e-04
4	4,475,706	40 / 17 / 0	0.85	13 / 24 / 12	0.51	4.0 e-05
5	4,596,970	43 / 17 / 0	0.86	13 / 24 / 12	0.51	3.0 e-05
6	4,720,567	39 / 17 / 2	0.82	12 / 13 / 12	0.50	1.3 e-04
7	4,829,519	40 / 18 / 1	0.83	13 / 23 / 13	0.50	1.0 e-04
8	4,902,478	38 / 21 / 1	0.81	13 / 23 / 13	0.50	1.0 e-04
9	4,998,799	38 / 19 / 1	0.82	12 / 23 / 14	0.48	5.0 e-05
10	5,226,319	27 / 19 / 3	0.75	14 / 24 / 5	0.61	0.47
11	5,727,295	11 / 22 / 7	0.55	5 / 14 / 1	0.60	0.29
12	6,232,397	15 / 19 / 8	0.58	3 / 33 / 8	0.44	1.0

[a] G indicates the allele was more common in the resistant GA-R strain,
L indicates the allele was more common in the susceptible LAB-S strain.
[b] From Fisher's exact test of the null hypothesis that allele frequency did not differ between the resistant and susceptible larvae.

902

903 **Table 2** Larval midgut expression of genes in the region of chromosome 13 QTL associated with
904 resistance to Cry1Ac.
905

Gene ID	Start - Stop (orientation)	Name	Larval midgut expression (mean log CPM)
MSTRG.8092	4,371,598 – 4,491,148 (+)	*Cyclic AMP-response element-binding protein A*	2.44
MSTRG.8095	4,491,241 – 4,520,601 (-)	*Heparan-alpha-glucosaminide-N-acetyltransferase*	-0.02
MSTRG.8097	4,504,492 – 4,507,681 (+)	*Juvenile hormone esterase-like Carboxyl/cholinesterase CCE006D*	4.84
hz_G0000107	4,536,278 – 4,541,251 (+)	*Lipase member H-B-like*	N/A
MSTRG.8100	4,541,250 – 4,593,328 (-)	*Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha*	3.25
MSTRG.8101	4,546,116 – 4,547,856 (+)	Kinesin-related protein 12-like	7.08
hz_G0000111	4,593,482 – 4,595,264 (-)	Uncharacterized protein	N/A
hz_G0000112	4,600,520 – 4,606,939 (+)	Uncharacterized protein	N/A
MSTRG.8102	4,607,861 – 4,617,975 (-)	*Ubiquitin-protein ligase E3A*	3.88
MSTRG.8103	4,618,419 – 4,620,800 (+)	*Retinal rod rhodopsin-sensitive cGMP 3`, 5`-cyclic phosphodiesterase subunit delta*	2.98

906

907 Gene IDs refer to the StringTie annotation when expressed for correspondence with the RNA-
908 seq data. Funannotate IDs refer to non-expressed genes. N/A indicates the gene was not
909 expressed.
910

911 **Table 3** GWAS and RNA-seq results for 11 genes previously implicated in lepidopteran
912 resistance to Cry1Ac.
913

Gene	Key reference	*H. zea* genome location[a]	Gene ID[b]	GWAS: *P*-value[c]	RNA-seq: GA-R vs. LAB-S *P*-value[d]	RNA-seq: GG vs. LL *P*-value[d]
ABCC1	Chen *et al.* 2019	12: 9.24-9.29	7445	0.16	**0.0014**	0.29
ABCC2	Baxter *et al*. 2011	15: 7.07-7.09	9800	0.35	0.95	0.56
APN1	Zhang *et al.* 2009	9: 11.33-11.37	5943	0.53	0.32	0.33
Cadherin	Gahan *et al.* 2001	6: 1.89-1.97	4086	0.46	1.0	0.13
Cadherin-86C	Fritz *et al.* 2020	12: 4.60-4.62	7702	0.49	0.69	0.75
MAPK4	Guo *et al.* 2015	12: 6.80-6.82	9824	0.21	0.12	0.43
mALP	Jurat-Fuentes *et al.* 2011	8: 10.41-10.43	4895	0.37	0.11	0.62
Polycalin	Wang *et al.* 2020	25: 2.72-2.73	16287	0.48	0.73	0.14
SP2	Rajagopal *et al.* 2009	7: 1.34-1.35	4260	0.53	0.85	0.88
Tetraspanin1	Jin *et al.* 2018	10: 11.58-11.59	6160	0.087	0.061	0.20
TryR	Liu *et al.* 2014	27: 2.04-2.09	17235	0.60	0.97	1.0

914 [a] Chromosome: Mb
915 [b] Full Gene ID begins with MSTRG.
916 [c] Lowest *P*-value for any SNP in the gene based on the G' analysis
917 [d] *P*-values are FDR-corrected, bold indicates significant at < 0.05.
918
919

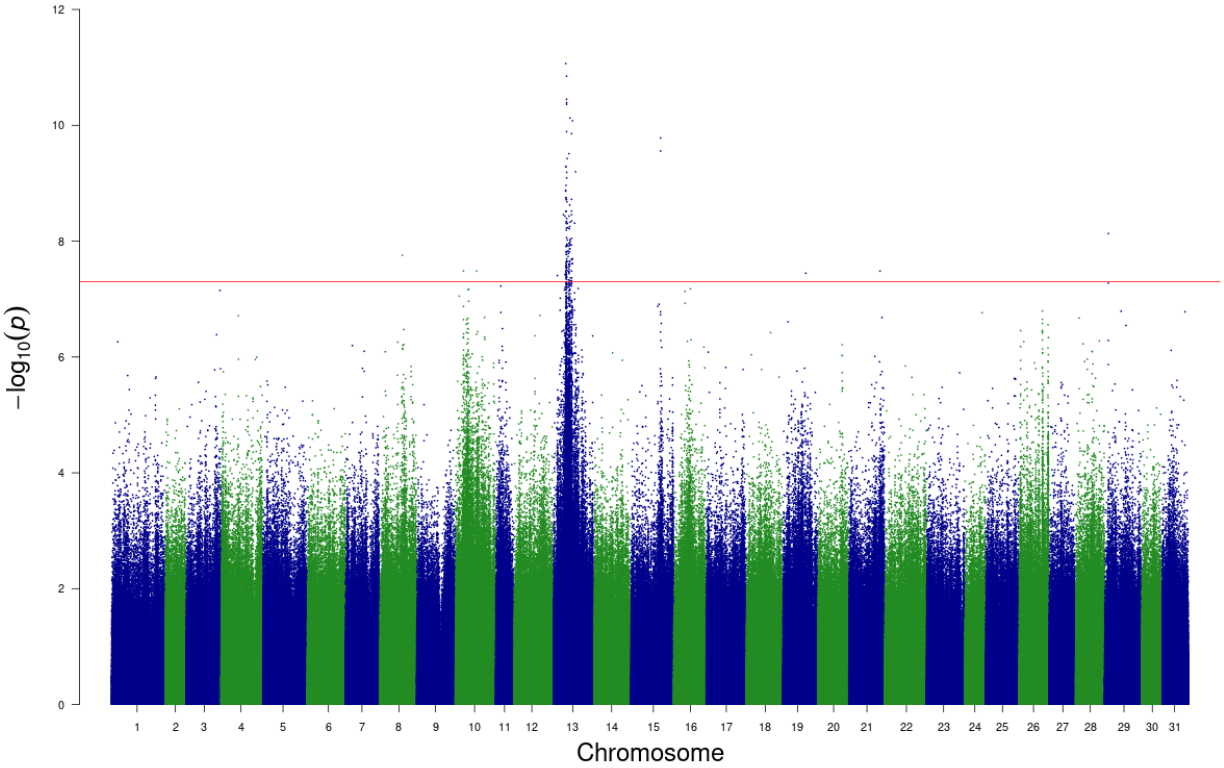


Figure 1 Manhattan plot of GWAS results showing $-\log_{10}$ of the *P*-values for Z-scores comparing allele frequencies between resistant and susceptible larvae. The red line indicates the threshold for significant association ($P = 5e^{-8}$).

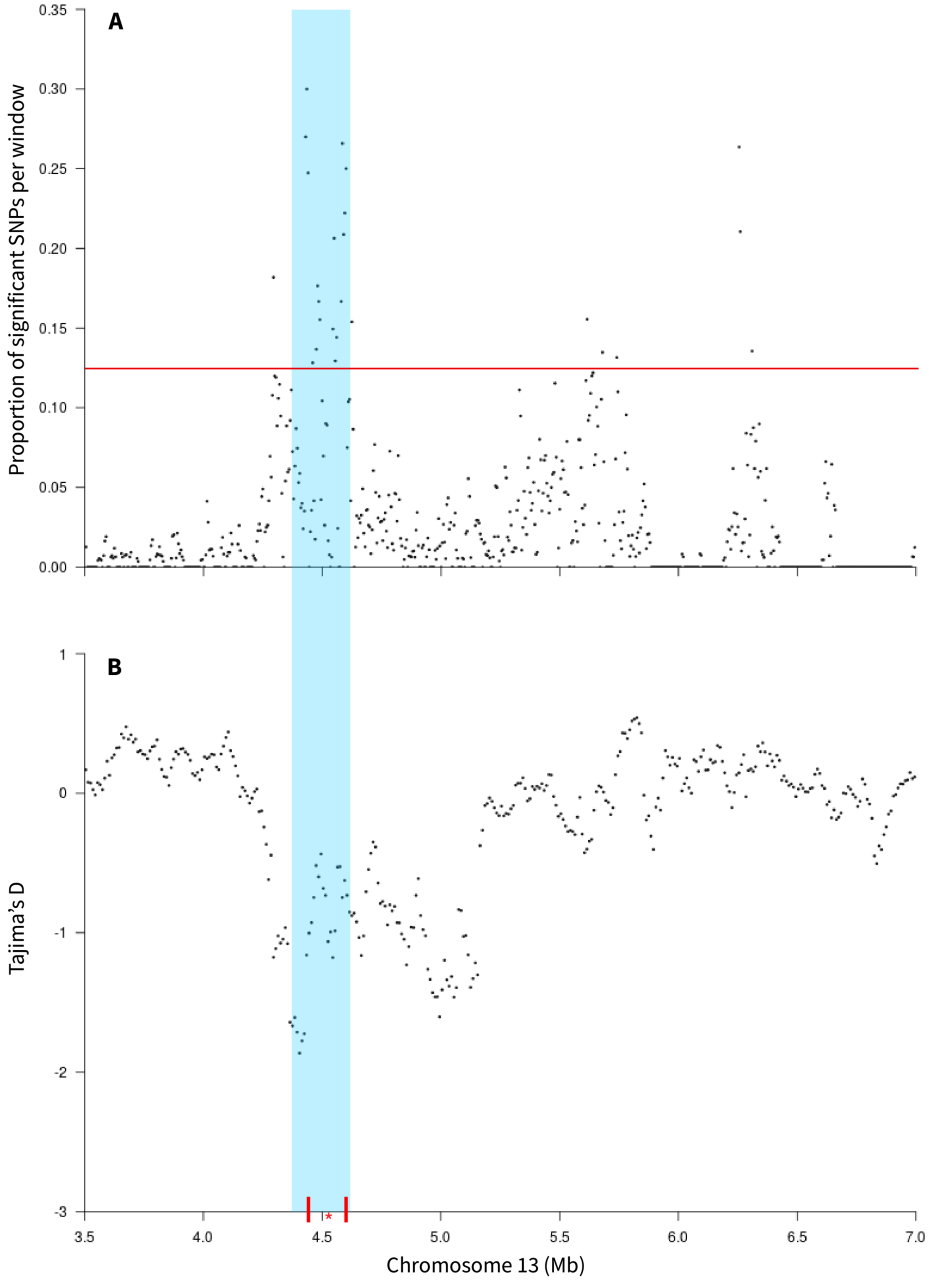
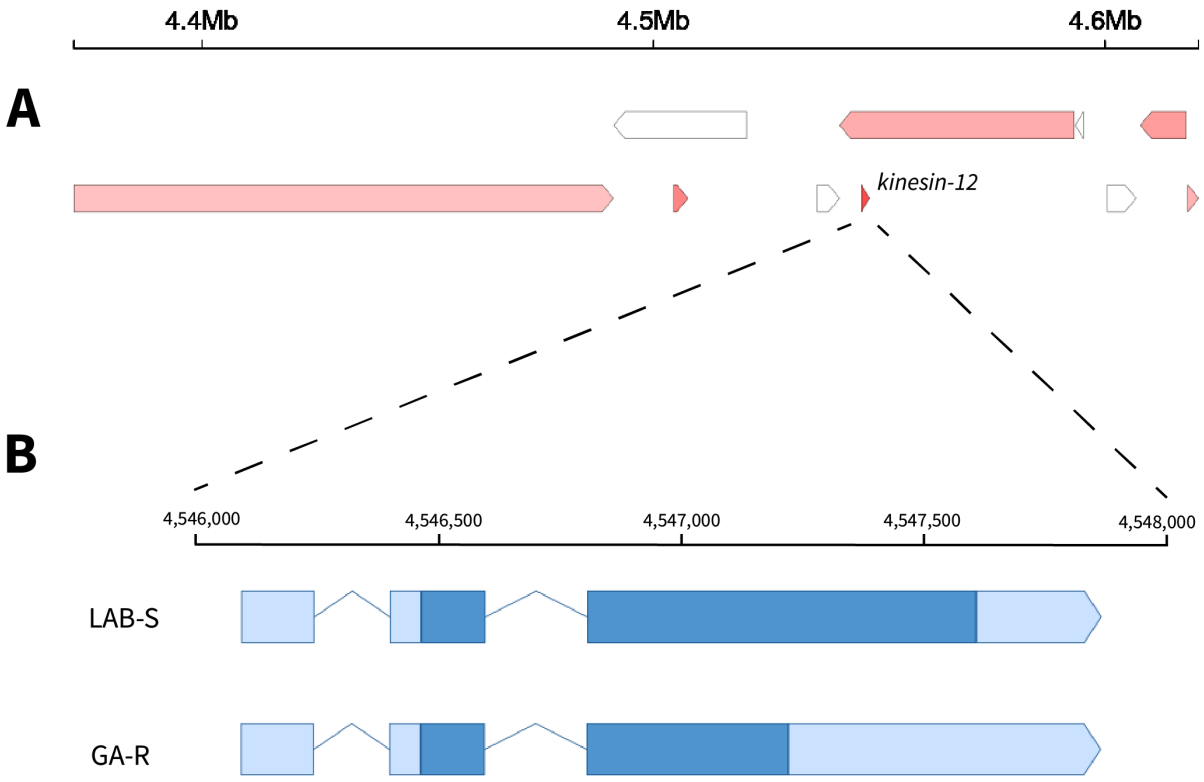


Figure 2 Association between resistance to Cry1Ac and SNPs within the chromosome 13 QTL. **A.** Proportion of significant SNPs ($P < 1e^{-5}$) from the Z-score analysis of the QTL data in 10 kb sliding windows. The horizontal red line indicates the 95th percentile of the distribution. **B.** Evidence of a selective sweep in GA-R from Tajima's D in 50 kb sliding windows. Blue shading covers the QTL from 4.37 to 4.62 Mb. The vertical red bars show the locations of markers 4 and 5 (Table 1). The red asterisk indicates the location of *kinesin-12*.



934
935
936
937 **Figure 3** Ten genes including *kinesin-12* in the resistance-associated QTL on chromosome 13.
938 **A.** The four genes at the top are in the (–) orientation, the other six below are in the (+)
939 orientation (Table 2). The four genes in white were not expressed in the midgut. Darker red
940 indicates higher expression in the midgut (Table 2). **B.** The structure of *kinesin-12* in LAB-S and
941 GA-R. Boxes represent exons, light blue indicates UTRs, and dark blue signifies coding regions.
942
943
944

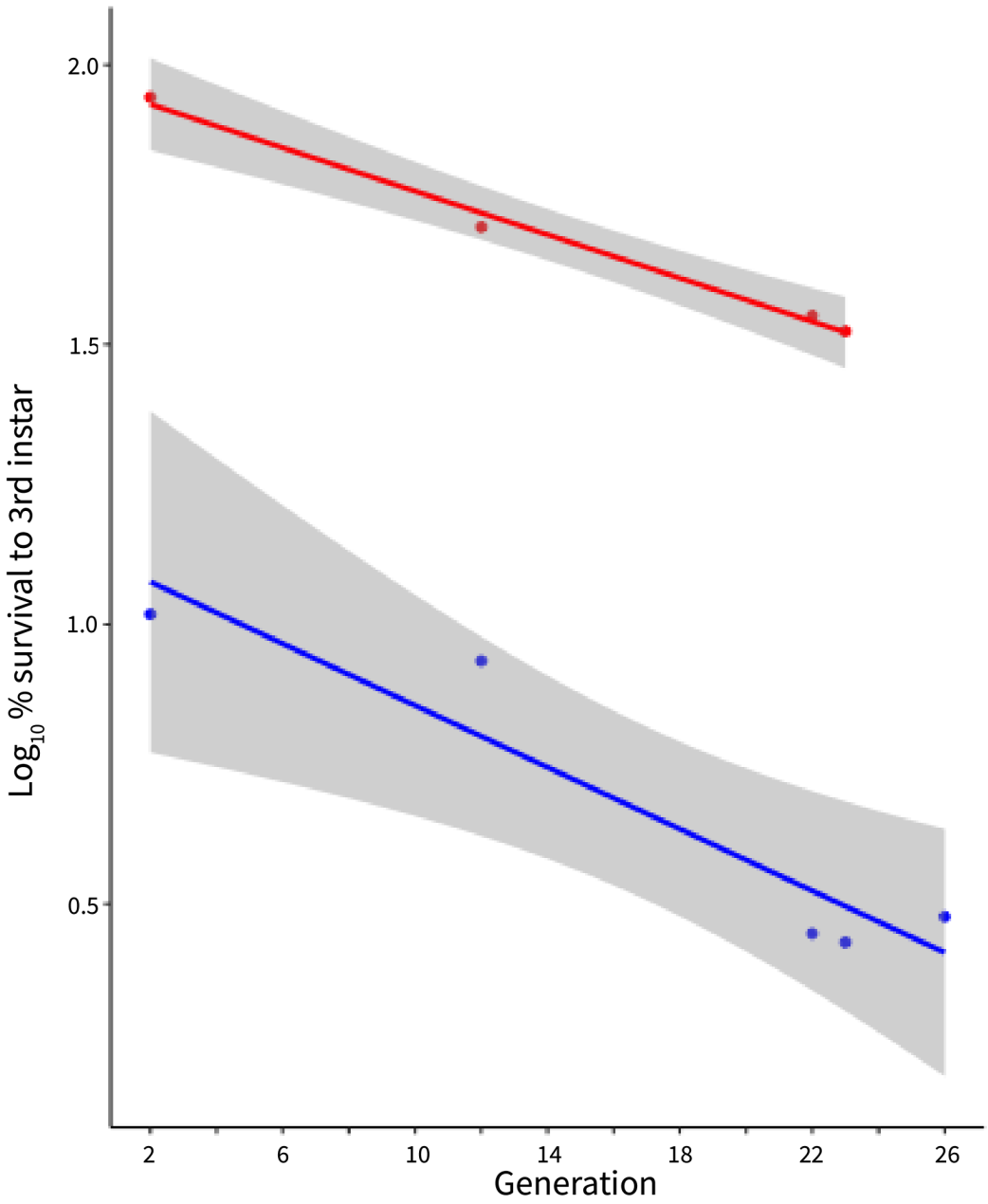


Figure 4 Survival of the heterogeneous GA-RS strain of *H. zea* tested on diet with 1 (blue) or 10 (red) micrograms Cry1Ac per cm^2 diet. Survival at each test concentration decreased significantly. Regressions of $\log_{10}$ (% survival to third instar) on generation: y = -0.019x + 1.97, $R^2 = 0.99$, df = 2, $P = 0.004$ and y = -0.028x + 1.13, $R^2 = 0.90$, df = 3, $P = 0.015$ for 1 and 10 micrograms Cry1Ac per cm^2 diet, respectively. Generation 26 was tested only at the higher concentration. Shaded areas represent 95% confidence intervals.