1
2
3   Corresponding Author:
4   Dr. Megan L. Fritz
5   Department of Entomology
6   University of Maryland
7   4291 Field House Dr.
8   Plant Sciences Bldg. Rm. 4112
9   College Park, MD 20742
10 mfritz13@umd.edu
11
12

13 **Contemporary evolution of a Lepidopteran species, *Heliothis virescens*, in response to modern**
14 **agricultural practices**

15

16 Megan L Fritz[1,2,§], Alexandra M DeYonke[2], Alexie Papanicolaou[3], Stephen Micinski[4], John Westbrook[5],
17 and Fred Gould[2]
18 [1]Department of Entomology, University of Maryland, College Park, MD 20742 USA
19 [2]Department of Entomology, North Carolina State University, Raleigh, NC 27607 USA
20 [3]Hawkesbury Institute for the Environment, Sydney Australia
21 [4]Louisiana State University AgCenter, Red River Research Station, Bossier City, LA 71112 USA
22 [5]USDA Agricultural Research Service, College Station, TX 77845 USA
23
24
25 [§]Corresponding Author

1

Abstract

Adaptation to human-induced environmental change has the potential to profoundly influence the genomic architecture of affected species. This is particularly true in agricultural ecosystems, where anthropogenic selection pressure is strong. *Heliothis virescens* feeds on cotton in its larval stages and US populations have been declining since the widespread planting of transgenic cotton, which endogenously express proteins derived *Bacillus thuringiensis* (Bt). No physiological adaptation to Bt toxin has been found in the field, so adaptation to this altered environment could involve: 1) shifts in host plant selection mechanisms to avoid cotton, 2) changes in detoxification mechanisms required for cotton-feeding versus feeding on other host plants, or 3) loss of resistance to previously used management practices including insecticides. Here we begin to address the question of whether such changes occurred in *H. virescens* populations between the years 1997 and 2012. As part of our study, we produced an *H. virescens* genome assembly and used this in concert with a ddRAD-seq enabled genome scan to identify loci with significant allele frequency changes over the 15 year period after Bt crops became widespread in the agricultural landscape. Confirmation that a known genetic change at a previously described *H. virescens* target of selection was detectable in our genome scan increased our confidence in this methodology. We quantified the strength of selection required to elicit the observed allele frequency changes at loci under selection. Potential contributions of genes near the loci under selection and their to adaptive phenotypes in the *H. virescens* cotton system are discussed.

Key Words

Heliothis virescens, tobacco budworm, *Bacillus thuringiensis*, cotton, selective sweep

2

47 **Introduction**

48   Human-induced change in the natural landscape places strong selective pressure on populations

49   to adapt over relatively short evolutionary timescales (Palumbi 2001). These changes shape the

50   genomes of local species, providing insight into contemporary evolution and it's implications for

51   affected species. Cultivation of the natural landscape for agricultural purposes is one of the most

52   ubiquitous examples of human-induced environmental change. Modern agricultural practices often

53   involve sweeping changes to the composition of plant species across broad geographic regions, re-

54   sculpting of the physical terrain and chemical inputs into the environment (Tilman 2001). The strong

55   selective pressure placed on species that inhabit agricultural ecosystems make them ideal for examining

56   genetic responses to anthropogenic forces (Taylor *et al.* 1995).

57   One such major change in recent agricultural history is the commercialization of transgenic

58   crops that themselves produce proteins for the management of key insect species. The tobacco

59   budworm, *Heliothis virescens*, feeds on cotton in its larval stages and populations in the Southern

60   United States have been declining since the widespread planting of transgenic cotton (Supplementary

61   Figure 1). These cotton cultivars endogenously express insecticidal proteins derived from the bacterium

62   *Bacillus thuringiensis* (Bt), which are lethal to *H. virescens*. In the Southern United States, Bt-

63   expressing cotton was rapidly adopted after it became commercially available for management of *H.*

64   *virescens* in 1996 (James 2015; Supplementary Figure 2). Prior to the widespread use of Bt-expressing

65   cotton, populations of *H. virescens* had evolved resistance to every insecticide used for their

66   management (Blanco 2012), including pyrethroid insectides (Luttrell *et al.* 1987, Campanhola and

67   Plapp 1989). Concerns over the possibility that *H. virescens* and other insect targets of Bt crops would

68   evolve resistance to the endogenously expressed proteins spawned an entire field of research related to

69   Bt resistance and associated genetic mechanisms (Reviewed in Heckel *et al.* 2007, Tabashnik *et al.*

3

70 2013). Of primary concern was the loss of efficacy of toxic Bt proteins (USEPA 1998, 2001, 2006). In

71 the case of *H. virescens*, no physiological adaptation to the Bt toxin in cotton fields has been detected

72 (Tabashnik *et al.* 2013). Yet widespread adoption of Bt-expressing crops likely placed selective

73 pressure on *H. virescens* in other ways.

74   As one example, widespread planting of Bt cotton cultivars led to an overall decline in

75 insecticide use on cotton in the United States (NASEM 2016, Benbrook 2012), including the use of

76 pyrethroids (personal comm. D. Reisig). In 1999, North Carolina extension entomologists stopped

77 recommending pyrethroids for the damaging generations of *H. virescens* in cotton (Bacheler 1999) and

78 still do not recommend pyrethroids. Similarly, in Louisiana, pyrethroids are no longer recommended

79 for *H. virescens* (LSU Ag Center 2016). In the state of California, one of the only states that makes

80 records of pesticide sales and applications publicly available, the pounds per acre of cotton for three

81 commonly used pyrethroids (deltamethrin, cypermethrin and cyfluthrin) declined from 0.05 to 0.03

82 between the years 2000 and 2012 (http://calpip.cdpr.ca.gov), and it is not clear if any of these

83 pyrethroids sprays are currently used to target *H. virescens*. Prior to Bt cotton adoption, when

84 pyrethroids were heavily used to suppress *H. virescens* populations, the voltage-gated sodium channel

85 gene (*Vgsc*) was identified as one gene target of selection wherein pyrethroid resistance alleles rose to

86 high frequency (Park and Taylor 1997, Park *et al.* 1997). Yet in *H. virescens* and other insect species,

87 *Vgsc* mutations often result in an overall loss of fitness for individuals carrying them (Zhao *et al.* 2000,

88 Foster *et al.* 2005, Kliot and Ghanim 2012, Brito *et al.* 2013). Under these conditions, stability in the

89 frequency of insecticide resistance alleles depends upon whether or not populations are continually

90 exposed to insecticidal pressure. Therefore, one possible effect of Bt adoption in *H. virescens* is a

91 reversion to susceptibility at their pyrethroid resistance locus.

92   Additional inadvertent targets of selection by Bt-expressing cotton could include loci involved

4

93   in feeding and oviposition behaviors. *H. virescens* are well known for the damage they cause to

94   cultivated cotton (reviewed in Blanco 2012), but their host plant range includes tobacco, soybean,

95   garbanzo bean (Fitt 1989), and a number of wild hosts (Sudbrink and Grant 1995).   Heritable, intra-

96   specific variation in host choice has been observed for *H. virescens* (Sheck and Gould 1993, Sheck and

97   Gould 1995, Karpinski *et al.* 2014), as well as other closely-related Lepidopteran species (Jallow and

98   Zalucki 1996, Jallow *et al.* 2004, Oppenheim *et al.* 2012). It is possible that the introduction of Bt-

99   expressing cotton has selected against *H. virescens* individuals that preferentially oviposit and feed

100   upon cotton hosts plants. Allele frequency changes in genes associated with chemosensation, central

101   nervous system function, and metabolism may have occurred as *H. virescens* was driven off of its

102   primary cotton host plant (Blanco 2012).

103   In recent years, identifying genomic change in response to selective forces has been enabled by

104   the development of next-generation sequencing (NGS) technologies. A variety of NGS-enabled marker

105   development techniques are used to generate novel, high density marker sets for model and non-model

106   organisms, including Restriction-site Associated DNA sequencing (RAD-seq; Baird *et al.* 2008),

107   Genotyping-by-Sequencing (GBS; Elshire *et al.* 2011), double-digest RAD-seq (ddRAD-seq; Peterson

108   *et al.* 2012) and others (reviewed in Andrews *et al.* 2016). These marker sets enable scientists to scan

109   the genomes of field-collected organisms in search of the gene targets of selection, particularly where

110   selection is strong and a reference genome assembly for read alignment is available (Lowry *et al.* 2017,

111   Catchen *et al.* 2017). Strong selection for advantageous alleles at target genes can influence allelic

112   composition at physically linked neutral sequences, including nearby marker sites, resulting in a

113   genomic footprint of selection that is much broader than the target gene alone (Nielsen 2005). The

114   breadth of this genomic footprint is influenced by several factors, including the strength of selection,

115   the initial frequency of the advantageous allele, effective pest population size and recombination rate

5

116 (Charlesworth and Charlesworth 2010).

117   Here we scanned the genomes of two *H. virescens* field populations collected in the Southern

118 United States between the years 1997 and 2012 to detect loci that have changed over time. Given that

119 pesticides and transgenic crops impose very strong selection on their target pest species (Onstad 2014),

120 we initially focused on a genomic region known to be associated with insecticide resistance as a

121 confirmation that ddRAD-seq could be used identify genes responsible for adaptive phenotypes under

122 strong selection. To achieve this goal, we produced an annotated draft assembly of the *H. virescens*

123 genome and used it for alignment of ddRAD-seq reads from barcoded individuals collected across

124 space and time. We then tested the hypothesis that changes in a candidate pyrethroid resistance gene,

125 *Vgsc*, could be detected through our ddRAD-enabled genome scanning techniques. Furthermore, we

126 identified additional ddRAD-seq loci with strongly diverging marker allele frequencies, and quantified

127 the strength of selection required to produced the observed changes at these sites. Some of the ddRAD-

128 seq loci identified as under seleciton were linked to genes involved in toxin metabolism and

129 chemosensation. We concluded with a discussion of the adaptive phenotypes that these newly

130 identified gene targets of selection might represent in a field environment.

131

132 **Methods**

133 *Insect Material*

134   For all population genomic analyses, adult male moths were collected by pheromone-baited trap

135 from Bossier Parish, LA, and Burleson County, TX. Collections took place in LA from May through

136 September, and in TX from May through October, in the years 1997, 2002, 2007, and 2012. GPS

137 coordinates from trapping locations can be found in Supplementary Table 1. Moths from each

138 collection date and location were immediately placed together in bottles of 95% ethanol for long-term

6

139 storage. Bottles from 2002, 2007 and 2012 were always held at -20°C until specimens were used,

140 while those from 1997 were initially held at room temperature and then transferred to -20°C. To

141 develop our *H. virescens* genome assembly, individuals from a long-standing colony strain (Gould *et*

142 *al.* 1995) were sib-mated for 10 generations to produce inbred material for sequencing (Fritz *et al.*

143 2016). Siblings from a single inbred family were used for sequencing and analysis. Five sibling pupae

144 were stored at -80°C prior to DNA isolation and library preparation. For all insect samples, DNA was

145 isolated with a Qiagen Blood and Tissue Kit (Qiagen, Inc., Valencia, CA, U.S.A.) using the mouse tail

146 protocol.

147

148 H. virescens *Candidate Gene Approach*

149   A polymerase chain reaction (PCR) based upon the methods of Park and Taylor (1997) was

150 used to amplify a 432 bp region in the alpha subunit of the *Vgsc*. The primer pair Nhp3304+ (5'

151 ATGTG GGACT GIATG TTGGT) and Nhp3448- (5' CTGTT GAAGG CCTCT GCTAT) flanked a

152 mutation known as L1029H. In this targeted region of the *Vgsc*, a single nucleotide polymorphism

153 (SNP) caused a Leucine to Histidine amino acid substitution and thereby pyrethroid resistance.

154 Additional mutations associated with pyrethroid resistance have been detected in *H. virescens*,

155 including D1561V + E1565G and V421M (Rinkevich *et al.* 2013). We specifically targeted L1029H

156 for our research because D1561V+ E1565G has not yet been functionally confirmed using ectopic

157 expression assays (Rinkevich *et al.* 2013), and V421M was not found in our study populations, even in

158 1997 when phenotypic pyrethroid resistance was at its peak.

159   Amplicons from PCRs targeting the L1029H mutation were digested by restriction enzyme Nla-

160 III, which cut in the presence of the resistance allele (Supplementary Figure 3). Genotypes were scored

161 by visualizing the digested PCR products on a 3.5% agarose gel (90 to 120 min at 120 V). We

7

162 examined the genotypes at this pyrethroid resistance locus for *H. virescens* individuals collected from

163 1997 (n = 194), 2002 (n = 204), 2007 (n = 268), and 2012 (n = 194) in LA, and 1997 (n = 142), 2007 (n

164 = 120), and 2012 (n = 196) in TX. Changes in pyrethroid resistance allele frequencies over time and

165 space were examined using a series of nested generalized linear regression models with binomial error

166 structures in R version 3.1.2 (R Core Team 2014; used here and throughout). The following full model

167 was used to examine the frequency of individual pyrethroid resistance alleles (*i*):

168 $$\Pr(y_i = 1) = \text{logit}^{-1}(\beta_{0i} + \beta_{1\ Year_i} + \beta_{2\ Season_i} + \beta_{3\ Location_i} + \beta_{4\ Year\ x\ Season_i}),$$

169 for *i* = 1,...., n

170 where Year represented collection year (e.g. 1997, 2002, 2007, or 2012), Season represented whether

171 the collections were made early (May or June) or late (August through October) in the cotton growing

172 season, and Location represented the collection location of the samples. We identified a model term as

173 statistically significant ($\alpha = 0.05$) when a comparison of nested models by analysis of deviance

174 indicated that removal of that term significantly influenced model deviance.

175

176 *Strength of Selection against the* Vgsc *Resistance Allele*

177 Following the discovery of a decline in frequency of the pyrethroid resistance allele, we quantified the

178 strength of selection associated with the decline in pyrethroid pressure. We used the following

179 equation to calculate the selection coefficient against the recessive resistance allele for our field

180 populations from TX and LA over 15 years after the introduction of Bt cotton into the landscape:

181
182 $$\triangle p = \frac{s\ p\ q^2}{1 - s\ q^2}$$

183 where p was the frequency of the susceptible allele, q was the frequency of the resistance allele, and *s*

184 was the selection coefficient. To calculate the change in p over 1 generation, we took the difference in

8

185 allele frequency over the 15 year period, and divided it by the total number of generations. For this, we

186 assumed 4 generations (Barber 1937) per year for each of the 15 years examined.

187

188 *Illumina WGS Library Preparation and Sequencing*

189   Genomic DNA from one pupa was submitted to the North Carolina State Genomic Sciences

190 Laboratory (Raleigh, NC, USA) for Illumina paired-end (PE) library construction and sequencing.

191 Prior to library preparation, the DNA template was quantified by a Qubit 2.0 Fluorometer (Invitrogen,

192 USA). The PE library with an 800bp insert size was constructed using an Illumina TruSeq Nano

193 Library Kit (Illumina, Inc. San Diego, CA) according to standard protocol. Following enrichment by

194 PCR, the library was checked for quality and final concentration using an Agilent 2100 Bioanalyzer

195 (Agilent Technologies, USA) with a High Sensitivity DNA chip before sequencing on an Illumina

196 HiSeq 2500 (100x2 paired end, rapid run).

197   Genomic DNA from a second pupa was used for mate-pair (MP) sequencing. Prior to library

198 preparation, whole genomic DNA was run out on a 0.5% agarose gel at 130v for 2 hours. Fragments

199 8kb or larger, as compared with Hyperladder I (Bioline USA Inc. Tauton, MA, U.S.A), were excised

200 from the gel and purified using a Zymoclean large fragment recovery kit (Zymo Research Corp. Irvine,

201 CA, U.S.A.). The DNA sample was submitted to the Michigan State University Research and

202 Technology Support Facility (East Lansing, MI, USA) for 8kb MP library preparation and sequencing.

203 The DNA library was prepared using an Illumina Nextera Mate Pair Sample Preparation Kit according

204 to standard protocol. The library was validated using a Qubit dsDNA assay, Caliper LabChipGX

205 (Perkin Elmer, Waltham, MA, U.S.A.) and Kapa Library Quantification qPCR for Illumina Libraries.

206 The library was loaded on one lane of an Illumina HiSeq 2500 High Output flow cell and sequenced in

207 a 2x125bp paired-end format using HiSeq SBS version 4 reagents. For both PE and MP libraries, base

9

208   calling was done by Illumina Real Time Analysis v1.18.64, the output of which was converted to FastQ

209   format with Illumina Bcl2fastq v1.8.4.

210

211   *PacBio Library Preparation and Sequencing*

212   Genomic DNA from 4 pupae, one of which was also used for Illumina PE sequencing, were

213   prepared into two libraries for PacBio sequencing. For each library, the SMRTbell Template

214   Preparation Kit version 1.0 (Pacific Biosciences, Menlo Park, CA, U.S.A.) was used for gDNA

215   preparation, but shearing and size-selection steps differed. For the first library, shearing was minimal

216   and no size selection was performed. For the second library, shearing prior to DNA concentration was

217   avoided to maximize fragment length, and a BluePippin (Sage Science Inc., Beverly, MA, U.S.A.) was

218   used to select fragments that were at least 7kb long. This produced sufficient prepared library material

219   for 17 and 5 SMRT cells, respectively. Prior to sequencing, the library concentration and fragment

220   length profiles were checked on a Qubit 2.0 and an Agilent Tapestation 2200 (Agilent Technologies,

221   USA) with a high molecular weight tape. Both libraries were sequenced at the University of North

222   Carolina Chapel Hill Sequencing facility on a PacBio RS II.

223

224   H. virescens *Genome Assembly*

225   Read quality was checked for all Illumina data using FastQC (Babraham Bioinformatics,

226   Cambridge, UK). Low quality ends were trimmed from both PE and MP reads using trimmomatic (v.

227   0.32; Bolger *et al.* 2014) and cutadapt (v. 1.9.1; Martin 2011), respectively. Any remaining Illumina

228   adapter sequences and Nextera transposon sequences were also removed. Reads were filtered for

229   potential microbial contaminants and *H. virescens* mitochondrial DNA (Supplementary Data File 1)

230   using BBmap (version 35.10; Bushnell B. - sourceforge.net/projects/bbmap/). For the full list of the

10

231 screened contaminants, see Supplementary Table 2. SOAPdenovo2 (v. 2.04) was used for assembly,

232 scaffolding and gap closure (Luo *et al.* 2012) with a k-mer length set to 63. Contigs and scaffolds over

233 2kb were used for further analysis.

234 RepeatScout (version 1.0.5; Price *et al.* 2005) was used to find *de novo*, species-specific repeats

235 in the K63 assembly, while RepeatMasker (version open-4.0; Smit *et al.* 2013) was used to identify

236 other common insect repeats available from Repbase (version 20150807; Jurka *et al.* 2005). We soft-

237 masked both repeat classes using BEDTools (version 2.25.0; Quinlan and Hall 2010) and collapsed

238 redundant haplotypes using the default settings in Haplomerger2 (version 3.1; Huang *et al.* 2012). To

239 fill intra-scaffold gaps, we applied PacBio reads over 5kb in length to our Illumina assembly using

240 PBsuite (version 14.9.9; English *et al.* 2012).

241 As an assembly quality check, BlastStation (TM Software, Inc., Arcadia, CA, U.S.A.) was used

242 to align 654 mapped ddRAD-seq marker sequences from the F1 parent used to produce an *H. virescens*

243 linkage map (Fritz *et al.* 2016; Dryad digital repository http://dx.doi.org/10.5061/dryad.567v8) to our

244 scaffolds. All top hits were exported and markers with alignment hit lengths greater than 150bp (of 350

245 bp total), identities greater than 80%, and e-values below 0.001 were further examined. This enabled

246 us to check for potential misassemblies, and provide additional information about which short scaffolds

247 likely belong together on individual chromosomes (Supplementary Table 3). BlastStation was also

248 used to identify the scaffold to which the alpha-subunit of the *Vgsc* (GenBank Accession: AH006308.2)

249 aligned.

250

251 *Structural annotation*

252 The Just_Annotate_My_Genome (JAMg; https://github.com/genomecuration/JAMg) platform

253 was used to generate putative gene models. First, the genome was masked using RepeatMasker (Smit

11

254 *et al.* 2013) and RepeatModeler (Smit *et al.* 2013). Subsequently, RNA-Seq data was obtained from

255 NCBI for *H. subflexa* and *H. virescens* (SRA accessions: ERR738599, ERR738600, ERR738601,

256 ERR738602, ERR738603, ERR738604, ERR738605, SRR1021613), preprocessed using

257 "justpreprocessmyreads" (http://justpreprocessmyreads.sourceforge.net), and assembled with Trinity

258 RNA-Seq 2.1.1 (Haas *et al.* 2013) using both the '*de-novo*' and 'genome-guided' options as

259 implemented in JAMg. The platform made use of multiple lines of evidence to support each gene

260 model: the two Trinity RNA-Seq assemblies integrated with 63,504 publicly available Sanger-

261 sequenced Expressed Sequence Tags using our new version of PASA (Haas *et al.* 2003); protein

262 domain annotation of putative exons via HHblits (Remmert *et al.* 2012); the *de-novo* gene predictors

263 GeneMark.HMM-ET (Lomsadze *et al.* 2014) and Augustus (Stanke *et al.* 2006) using the assembled

264 and raw RNA-seq and protein domain data as external evidence. These evidence tracks were condensed

265 to an Official Gene Set (OGS) using Evidence Modeler (Haas *et al.* 2008). A quantitative assessment

266 of our assembly and annotation completeness was conducted using BUSCO software using the

267 metazoan lineage setting (version 2.0.1; Simao *et al.* 2015).

268

269 H. virescens *ddRAD-seq library preparation*

270   DdRAD-seq libraries were prepared according to Fritz *et al.* (2016) with minor modifications.

271 Briefly, 200 ng of genomic DNA from the thorax of each field-collected specimen was digested with

272 EcoRI and MspI. Overhang sites from each specimen were ligated to Truseq Universal adapters

273 (Illumina, Inc. San Diego, CA) modified to contain a unique barcode (Elshire *et al.* 2011, Fritz *et al.*

274 2016). Adapter-ligated DNA fragments from each individual were combined into pools of no more

275 than 24 individuals. A Blue Pippin (Sage Science, Inc., Beverly, MA) was used to select adapter-

276 ligated DNA fragments ranging from 450-650 bp from each pool, and size-selected DNA pools were

12

277 amplified in a Peltier PTC200 thermalcycler under the following reaction conditions: 72 °C for 5min,

278 18 cycles of 98 °C for 30 sec, 65 °C for 20 sec, 72 °C for 30 sec, followed by 72 °C for 5 min. For

279 each pool, 1 of 4 Illumina indices (1,2,6, or 12) was added via PCR to the MspI adapter. Amplified

280 pools were combined, cleaned with a Qiaquick PCR Purification Kit (Qiagen, Inc., Valencia, CA,

281 U.S.A.), and diluted to 4nM prior to sequencing. Prepared genomic DNA libraries constructed from a

282 total of 177 *H. virescens* individuals were spread across four 2x300 paired-end Illumina MiSeq runs.

283 Individuals from each year and collection location were spread evenly across each MiSeq run to

284 minimize sequencing run bias in our downstream analysis.

285

286 *Demultiplexing and Genome Alignment of DdRAD-seq Markers*

287   Illumina-generated read 1 and 2 files were merged using FLASH version 1.2.7 (Magoc and

288 Salzburg 2011), then demultiplexed and filtered for quality using the process_radtags script from

289 Stacks version 1.09 (Catchen *et al.* 2011, 2013). Quality filtering entailed removal of reads when: 1)

290 they did not have an intact EcoRI cut site, 2) had a quality score < 30, or 3) were smaller than 350 bp.

291 We disabled the rescue reads feature in the process_radtag script, and therefore no read containing

292 errors in the barcode sequence was used for downstream analysis. All remaining merged reads were

293 truncated to a maximum length of 350 bp. Filtered demultiplexed reads were aligned to our *H.*

294 *virescens* genome assembly using Bowtie 2 (version 2.2.4; Langmead and Salzberg 2012). All reads

295 were aligned in end-to-end mode using the preset parameters with the highest sensitivity (--very-

296 sensitive).

297

298 *Association of DdRAD-seq Marker Genotypes with the Pyrethroid Resistance Allele*

299   We first identified whether any raw ddRAD sequencing reads aligned to the scaffold containing

13

300   the *Vgsc* using Integrative Genomic Viewer (IGV; Robinson *et al.* 2011). Following identification of

301   potential ddRAD-seq markers near the *Vgsc*, we inspected stacks of ddRAD-seq reads for individuals

302   with genotypic data at the pyrethroid resistance locus. Particular attention was paid to individuals that

303   were homozygous for the pyrethroid resistance allele. Through an IGV visual inspection of ddRAD-

304   seq raw reads, we identified one 350bp ddRAD-seq locus (hereafter Hv_11322), for which a single

305   350bp sequence (hereafter Hv_11322_hap1) was commonly associated with the L1029H mutation at

306   the *Vgsc*. Filtered, genome-aligned reads from all specimens were then fed into the Stacks v. 1.09

307   (Catchen *et al.* 2011, 2013) pipeline for read clustering. Custom R and python scripts were used to call

308   350bp ddRAD-seq genotypes at Hv_11322 for all field-collected individuals, which were then

309   manually inspected and edited to include any insertions and deletions that were omitted by the Stacks

310   software. For purposes of genotype calling at Hv_11322, individuals with a read count of 6 or higher

311   for a single 350bp sequence were considered homozygotes, with two copies of that allele. Where

312   individuals carried fewer than 6 reads for a single 350bp sequence, their genotypes were scored as a

313   single copy of that observed allele plus one null allele. This threshold was chosen because individuals

314   with 6 or more reads can be called homozygotes with greater than 95% certainty (Buerkle and Gompert

315   2013).

316   We postulated that if the breadth of the "selective sweep" surrounding the *Vgsc* resistance allele

317   included Hv_11322, such that Hv_11322_hap1 was associated with the L1029H mutation, the rates of

318   their decline in frequency should be similar, if indeed there was a decline. We therefore examined

319   whether the frequencies of the L1029H mutation and Hv_11322_hap1 differed in their rate of decline

320   over time. Specifically, we used a series of nested generalized linear models with binomial error

321   structures to examine whether locus and collection year interacted to influence individual allele (i). In

322   the case of the Hv_11322 response, Hv_11322_hap1 was scored as a 1 and all other alleles were scored

14

323   as a zero. Our full statistical model was as follows:

324 $$\Pr(y_i = 1) = \text{logit}^{-1}(\beta_{0i} + \beta_{1\,Year_i} + \beta_{2\,Locus_i} + \beta_{3\,Year\,x\,Locus_i}),$$

325   for $i = 1,...., n$

326   where Year represented the years during which the moths were collected and Locus indicated either the

327   *Vgsc* or ddRAD-seq marker Hv_11322. As before, we identified a model term as statistically

328   significant when a comparison of nested models by analysis of deviance indicated that removal of that

329   term significantly influenced model deviance. No significant difference between a model with and

330   without the interaction term indicated that the slope of the decline in the L1029H mutation was similar

331   to that of Hv_11322_hap1.

332   We also analyzed the distribution of Hv_11322_hap1 for groups of individuals that were

333   homozygous for either the resistant or susceptible alleles at the *Vgsc*. In total, 32 individuals were

334   homozygous in our target region of the *Vgsc* and contained sufficient ddRAD-seq data at nearby locus

335   Hv_11322 to call at least one allele. Of these 32 individuals, two Hv_11322 alleles could be called for

336   26 individuals, whereas only a single allele could be confidently called for 6 of the individuals due to

337   their lower than 6X depth of coverage. In total, 58 haplotypes (from 32 individuals), which contained

338   genotypic information for both the *Vgsc* and the nearby ddRAD-seq marker were examined. A Fisher's

339   exact test of independence was used to determine whether there was an association between the

340   frequencies of Hv_11322_hap1 and the L1029H mutation.

341

342   H. virescens *ddRAD-seq Enabled Genome Scan*

343   Previous work by Groot *et al.* showed that genetic differentiation among North American

344   populations of *H. virescens* was low (Groot *et al.* 2008). Given these already documented high levels

345   of gene flow between collection sites, and our goal to detect genomic change over time, we specifically

346   focused on analyzing allele frequency changes between years.   Samtools (version 0.1.18; Li *et al.*

347   2009, Li 2011) view was used to convert SAM files output by Bowtie 2 to BAM files, and SNPs were

348   called using mpileup.   BCFtools was used to generate SNP and indel genotypes, as well as genotype

349   likelihoods in a Variant Call Formatted (VCF) file. This VCF file was filtered by VCFtools (version

350   0.1.15, Danacek *et al*. 2011, https://vcftools.github.io) prior to downstream population genomic

351   analysis. The filtered dataset included loci that: 1) were sequenced to a depth of 3 or more reads, 2)

352   had a minor allele frequency of 0.1 or greater, 3) were represented in at least 50% of individuals, and 4)

353   included only SNP variant sites (indels were excluded). The number of SNPs was thinned such that no

354   more than 2 were examined per ddRAD-seq locus. This thinned SNP dataset was transformed from

355   VCF format to genepop format using PGDSpider (version 2.1.1.0, Lischer and Excoffier 2012).

356   Pairwise SNP outlier analyses of these thinned SNP datasets were made using Lositan (Antao *et al.*

357   2008) with the following parameter settings: 1) "neutral" and "forced" mean FST settings were

358   engaged, 2) the Infinite Alleles model was assumed, 3) the false discovery rate was set to 0.1, and 4)

359   the type I error was set to 0.01.

360   Scaffolds containing SNP markers with statistically significant pairwise-genetic divergence

361   were identified for further analysis (Supplementary Data File 2). Using the physical distance between

362   the ddRAD-seq locus Hv_11322 and the *Vgsc* as our guide, the predicted structural genes on each

363   scaffold identified in our Supplementary Data File 2 and found within 36 kb of each SNP outlier were

364   examined. Protein sequences corresponding to each annotation along a scaffold where divergent

365   markers were present were aligned to the NCBI Arthropod database (taxid: 6656) via blastp using

366   Blast2GO software. Gene ontology (GO) assignments for each predicted gene sequence within 36 kb

367   of an outlier locus are also provided in Supplementary Data File 3.

368   Given the large number of outlier loci (Figure 3), and that the potential traits under selection in

16

369   this agricultural system (e.g. metabolic detoxification of insecticides and/or host plant defensive

370   compounds, host volatile detection) are often quantitative, we reasoned that certain gene families may

371   be over-represented near our outliers with respect to their overall distribution throughout the *H.*

372   *virescens* genome. Therefore, we examined the distribution of GO categorizations (54 level two

373   categorizations in each of three GO domains: "Biological Process", "Molecular Function", and

374   "Cellular Component") for the subset of predicted genes found near outlying SNPs for each by-year

375   comparison. We compared those distributions to the numbers of genes found in each of these same

376   categories in the overall *H. virescens* genome using a series of Fisher's Exact tests. Due to the large

377   number of comparisons (n = 54), we used a Bonferroni-corrected alpha value of 0.0009 to establish

378   statistically significant over-representation in any one GO category. For each by-year comparison, we

379   examined the subsets of genes within 2 different intervals from SNP outliers: 10 kb ("moderate"

380   linkage according to Lowry *et al.* 2017) and 36 kb (extended linkage, based upon the distance between

381   Hv_11322 and the *Vgsc* target of selection).

382

383   *Strength of Selection on Outlier Loci*

384   Finally, we examined what the strength of selection must have been to produce the observed

385   allele frequency changes at all identified outlier loci for each by-year comparison. We calculated the

386   coefficient of selection (*s*) against q, the declining allele frequency, assuming either dominance of p or

387   co-dominance of p and q according to Falconer and Mackay (1996). Custom scripts written in R were

388   used to calculate *s* for all outliers with the exception of those with initial values of p that were very low

389   (p > 0.05). These were excluded because they are most susceptible to under-sampling due to our small

390   sample sizes, and small biases in these values have the potential to significantly influence the selection

391   coefficient.

17

392

393 **Results**

394 H. virescens *Candidate Gene Analysis*

395 In 1997, the frequency of the L1029H mutation was 0.66 in LA and 0.63 in TX. By the year

396 2012, the frequency of this resistance allele declined to 0.44 in LA and 0.36 in TX (Figure 1). This

397 decline in the resistance allele frequency over our 15 year sampling period was statistically significant

398 ($p < 0.001$). Neither the interaction between year and season ($p = 0.36$), season itself ($p = 0.21$), nor

399 sampling location ($p = 0.25$) significantly influenced the frequency of the resistance allele.

400

401 *Strength of Selection against the* Vgsc *Resistance Allele*

402 In LA, the frequency of the resistance allele declined by 0.22 over the 15 years between 1997

403 and 2012. In TX, the decline in the resistance allele frequency was 0.27. Therefore, the increase in p

404 per generation was 0.004 and 0.005 for LA and TX, respectively. With this information, we were able

405 to calculate a selection coefficient of 0.03 for each of the LA and TX populations.

406

407 *Genome Sequencing and Assembly*

408 In total, the Illumina sequencing runs produced 122,433,923 and 232,607,659 reads for PE and

409 MP libraries, respectively. After read trimming and filtering, 115,374,414 and 227,857,423 reads from

410 the PE and MP libraries were used for assembly. An additional 482,464 PacBio reads with an average

411 length of 7560 (s.d. = 2663) bp were applied to our Illumina assembly using PBsuite software for gap

412 filling. The final *H. virescens* genome assembly was comprised of 8,826 scaffolds with a total length

413 of 403,154,421 bp, similar to the previously estimated *H. virescens* genome size of 401 Mbp (Gregory

414 and Herbert 2003). The scaffold N50 was 102,214 bp (mean size = 45,678 bp; range = 659 – 628,964

18

415   bp). A BUSCO analysis of our final assembly indicated that 865 (88%) of the 978 core conserved

416   eukaryotic genes were complete. Further specifications for our genome assembly can be found in

417   Supplementary Table 4. When we examined our previously mapped *H. virescens* ddRAD-seq markers

418   (Fritz *et al.* 2016), a total of 562 out of 654 met the aforementioned alignment criteria relative to the

419   reference genome and were used to examine and group scaffolds into chromosomes. Of these 562

420   markers, 557 aligned uniquely to a single scaffold, while 5 markers (4851, 5891, 13906, 22644, 29612)

421   aligned well to multiple scaffolds. This suggested that either those scaffolds were allelic, or that the

422   marker sequences contain repetitive DNA. Four-hundred eighty three of the 8826 scaffolds present in

423   our assembly were aligned to at least 1 mapped marker. In most cases (n = 421 scaffolds), a single

424   scaffold was associated with a single mapped marker. However, 62 scaffolds could be aligned to

425   multiple mapped markers, which enabled us to check the quality of our assembly against our linkage

426   map. Of these 62 scaffolds, only 5% (3 scaffolds) aligned to markers that originally mapped to

427   different linkage groups. A summary of the scaffold names and groupings by linkage group can be

428   found in Supplementary Table 3. One scaffold, numbered 4600, contained the entire *Vgsc* sequence

429   available from GenBank accession AH006308.2.

430

431   *Association of DdRAD-seq Marker Genotypes with the Pyrethroid Resistance Allele*

432   We located one ddRAD-seq marker, called Hv_11322, spanning bp 11,397 through 11,747 of

433   Scaffold 4600 which is *ca.* 37 kb upstream from the pyrethroid resistance locus. In total, 55 unique

434   Hv_11322 sequences (alleles) could be identified from 138 individuals using Stacks. Fifty of these 55

435   alleles were found fewer than three times in our field-collected populations. The remaining five most

436   common alleles (Supplementary Figure 4) were found 5, 5, 6, 12, and 161 times, respectively.

437   According to an NCBI blast, all of these sequences aligned well with an *Helicoverpa armigera*

19

438   sequence (GenBank accession DQ458470.1) that also contained the *Vgsc*.

439   No statistically significant difference existed between the slope of the decline in the L1029H

440   mutation and Hv_11322_hap1 (deviance = -0.355, df = 1, p = 0.551). These allele frequency declines

441   are plotted in Figure 1. When we examined the full haplotypes (e.g. containing both the Hv_11322

442   locus and the pyrethroid resisance locus), specifically in homozygotes at the pyrethroid resistance

443   locus, we identified 20 unique Hv_11322 alleles in the 32 total individuals (or N = 58 chromosomes).

444   When chromosomes containing the L1029H mutation (pyrethroid resistance-conferring *Vgsc* allele)

445   were examined, 85% (29 of 34) also carried the ddRAD-seq allele Hv_11322_hap1 (Figure 2). The

446   remaining 5 chromosomes bearing the L1029H mutation contained 4 unique Hv_11322 alleles, none of

447   which were the 5 most common Hv_11322 alleles. Only 5% (1 of 24) of the chromosomes bearing the

448   wild-type *Vgsc* allele also carried Hv_11322_hap1. Fifteen unique Hv_11322 alleles of the 20 alleles

449   found in homozygotes were associated with the wild-type *Vgsc* allele. A Fisher's Exact test indicated

450   that there was a statistically significant association between the presence of Hv_11322_hap1 and the

451   L1029H mutation (p < 0.001).

452

453   H. virescens *ddRAD-seq Enabled Genome Scan*

454   Of the 1,682,114 SNPs in our ddRAD-seq marker dataset, the total number of filtered SNPs

455   included in the analyzed dataset was 8,963. Based upon this filtered dataset, overall population

456   genomic divergence between years was low. Pairwise Weir and Cockerham's overall FST values were

457   0.005, 0.004, and 0.013 for the comparisons 1997-2007, 1997-2012, and 2007-2012, respectively.

458   We first examined SNPs along Scaffold 4600, where the *Vgsc* was located, for evidence of

459   genomic divergence between years. Between the years 1997 and 2012, one SNP at position 11706 on

460   Scaffold 4600, which was part of Hv_11322, showed signs of statistically significant genetic

20

461 divergence with a Weir and Cockerham's FST value of 0.205 (p < 0.001). This SNP, a cytosine to

462 thymine transition, went from a cytosine allele frequency of 0.875 (n = 44 diploid individuals) in 1997,

463 fell to a frequency of 0.724 in 2007 (n = 49) and further declined to a frequency of 0.548 in 2012 (n =

464 52; Fig. 1).

465   In addition to this FST outlier on Scaffold 4600, our genome scan revealed a number of

466 additional diverging allele frequencies for each by-year comparison. In total, we detected a total of 351

467 SNPs on 314 scaffolds (3.6% of the 8,826 total scaffolds) as outliers in at least one by-year

468 comparison. Table 1 shows the number of genomic outliers for each comparison, as well as the number

469 of unique scaffolds on which these unique outliers were found. Between the years 1997 and 2007, 201

470 SNPs (2.2% of the 8,963 total SNPs examined) showed signs of significant allele frequency

471 divergence, whereas only 35 SNPs (0.4% of the 8,963 total SNPs) significantly diverged between the

472 years 2007 and 2012. When a comparison was made over the total time period, between the years 1997

473 and 2012, 184 SNPs showed signs of statistically significant allelic divergence.

474   In some cases, the same SNPs showed signs of divergence in two different by-year

475 comparisons, but no SNP outliers were shared among all 3 comparisons (Supplementary Figure 5).

476 Fifty-three SNPs were considered outliers between the years 1997 and 2007, and again between 1997

477 and 2012. This indicated that significant allele frequency changes occurred between the years 1997

478 and 2007, followed by stability or small, non-significant allele frequency changes through the year

479 2012. Likewise, 5 SNPs showed significant allele frequency divergence between the years 1997 and

480 2012, and again between the years 2007 and 2012, indicating that allele frequencies were stable, or

481 underwent modest changes between the years 1997 and 2007 followed by a significant change in the

482 year 2012.

483   When we examined the 36 kb of DNA flanking the SNP outliers on these 314 scaffolds (72 kb

21

484   total), the mean number of putative genes identified in these broader genomic regions was fewer than 3,

485   and the maximum was 12 (see Table 1 for a complete breakdown by comparison). A complete list of

486   outlier SNPs and their genome positions for each by-year comparison, as well as the putative genes and

487   their GO assignments within the 36 kb window surrounding these SNPs can be found in Supplementary

488   Data files 2 and 3, respectively.

489   Some outliers were found near putative genes with functions related to either insecticide

490   resistance or changes in host use. For example, one predicted gene sequence within 10kb of two

491   outliers, JAMg_model_7841, aligned to a cytochrome p450 protein sequence (*Cyp6AE12*) from

492   *Helicoverpa armigera* (GenBank Accession AID54888.1) with a 100% query cover and 83% identity.

493   It was identified on Scaffold 3424 and the associated SNPs were detected as outliers between 1997 and

494   2012, and also between 2007 and 2012. As a second example, two predicted gene sequences,

495   JAMg_model_4652 and JAMg_model_4653 aligned with at least 70% query cover and 78% identity to

496   *Athetis lepigone* olfactory receptor (OR60; GenBank Accession KT588155.1) and *Helicoverpa assulta*

497   olfactory receptor (OR33; GenBank Accession KJ542684.1) sequences, respectively. These gene

498   sequences were found within 10kb of a SNP outlier from 1997-2007 and 1997-2012 comparisons and

499   was located on Scaffold 2173. We examined the ontology of genes near SNP outliers, with an eye

500   toward those predicted to be important for detoxification and behavior, to identify whether any GO

501   category was over-represented near outlier SNPs relative to the overall genome.

502   GO assignments by Blast2GO were only achieved for a subset of predicted genes. The

503   proportions of putative genes within 36 kb of outliers for which a function could be predicted were

504   0.37, 0.38, and 0.32 for the by-year comparisons 1997-2007, 1997-2012, and 2007-2012, respectively.

505   While low, these were similar to the proportion of gene predictions assigned a function by Blast2GO in

506   the overall genome assembly (0.40). When Fisher's Exact tests were applied to identify whether any

22

507 level 2 GO term was over-represented in either 10 or 36kb windows near outlier SNPs from each by-

508 year comparison, no significant difference was found relative to their distribution in the overall

509 genome.

510

511 *Strength of Selection on Outlier Loci*

512   Selection coefficients (s) were calculated for 164, 150, and 31 outliers for the following by-year

513 comparisons, respectively: 1997 and 2007, 1997 and 2012, and 2007 and 2012. For selection against q,

514 the declining allele, s ranged from 0.009 – 0.301 across all by-year comparisons when dominance of p

515 was assumed. When co-dominance of p and q was assumed, selection coefficients ranged from 0.012

516 to 0.623. The mean selection coefficients (± standard deviation) calculated across outliers were

517 0.047(0.020), 0.077 (0.029) and 0.165 (0.062) assuming dominance of p for the 1997 - 2012, 1997 -

518 2007, and 2007 - 2012 by-year comparisons, respectively. Mean selection coefficients assuming

519 codominance of p and q were 0.058 (0.037), 0.088 (0.058), and 0.200 (0.150) for these same time

520 periods. In general, SNPs with higher selection coefficients were associated with greater rates of

521 change in the frequency of q, as well as higher initial starting values of q (Figure 4).

522

523 **Discussion**

524   Double-digest RAD-seq and other NGS marker-development methods have been used to detect

525 signatures of local adaptation in a number of non-model plant and animal species (e.g. Hohenlohe *et al*.

526 2010, Nadeau *et al*. 2014, Pujolar *et al.* 2014, Ruegg *et al.* 2014, Pais *et al.* 2016). Here we have

527 demonstrated the power of ddRAD-seq to identify genomic regions that have diverged over short

528 evolutionary time scales (fewer than 2 decades) in a landscape characterized by human-induced

529 environmental change. We postulated that widespread adoption and cultivation of Bt cotton in the

23

530 Southern United States would likely impose strong selection on Lepidopteran herbivore and cotton

531 pest, *H. virescens*, through shifts in host plant composition and insecticide use. We first identified

532 allele frequency changes at a likely gene target of selection, the *Vgsc*, in field-collected populations of

533 *H. virescens*. We then demonstrated that this change could be detected using a nearby ddRAD-seq

534 marker. Allele frequencies at many other regions of the *H. virescens* genome also diverged over time,

535 likely in response to selection pressures imposed by widespread adoption of Bt cotton. We calculated

536 selection coefficients for SNPs that were detected as having changed significantly over time, to

537 demonstrate the strength of selection encountered by organisms found in agricultural ecosystems.

538 Furthermore, we sequenced and assembled an *H. virescens* genome to help us identify potential

539 structural genes involved in adaptation to agricultural inputs, and made it publicly available at NCBI.

540 Our initial examination of the *Vgsc*, a candidate gene likely to be impacted by the decline in

541 pyrethroid use that followed Bt cotton adoption, demonstrated that the resistance-conferring L1029H

542 mutation declined in frequency over time. Indeed, examination of allele frequency changes at the *Vgsc*

543 locus yielded a selection coefficient of 0.03. This indicated that the resistance allele was moderately to

544 strongly deleterious in field populations where pyrethroid pressure was low (Eyre-Walker and

545 Keightley 2007, Kim *et al.* 2017). A selection coefficient of this magnitude seemed reasonable given

546 the previously reported fitness cost associated with carrying this resistance allele (Zhao *et al.* 2000).

547 We were surprised to see the frequency of the resistance allele plateau in the year 2007 and remain at

548 *ca.* 0.4 through the year 2012, however. There are several possible explanations for this. One

549 explanation is that pyrethroid pressure has declined but remains sufficiently high such that maintenance

550 of the resistance allele in field populations is advantageous, in spite of the fitness cost to individuals

551 that carry it (Zhao *et al.* 2000). Alternatively, as the resistance allele frequency declines there are

552 relatively fewer homozyogous resistant genotypes. If the fitness cost is only associated with

24

553 homozygotes then the decline in resistance allele frequency could level off, even in the absence of

554 pyrethroid selection.

555   Using a ddRAD-seq dataset, we identified one marker that aligned to our reference genome 37

556 kb upstream of the *Vgsc*. One allele of this 350bp marker, called Hv_11322_hap1, was associated with

557 the L1029H mutation that confers pyrethroid resistance. This suggested that Hv_11322_hap1 was in

558 linkage disequilibrium with the L1029H mutation. Furthermore, the breadth of the selective sweep in

559 this genomic region extended at least 36 kb on one side of the *Vgsc*. Upon further examination of

560 Scaffold 4600, which contains this region under selection, we identified three cytochrome p450s that

561 are found between Hv_11322 and the *Vgsc*. This confirmed previous reports of tight physical linkage

562 between the *Vgsc* and *Cyp6B10* in *H. virescens* (Park and Brown 2002). It is possible that these

563 cytochrome p450s could also be targets of selection by pyrethroid insecticides, and future work could

564 be directed at whether or not they play any roles in the expression of pyrethroid resistance phenotypes.

565 Work in another closely-related Lepidopteran species, *H. armigera*, suggests that this particular

566 cytochrome p450 is not involved in pyrethroid resistance, however (Grubor *et al.* 2007).

567   SNP data from our ddRAD-seq marker Hv_11322 enabled us to rediscover changes at the *Vgsc*

568 associated with the L1029H mutation over time. While the SNP outlier in the Hv_11322 marker

569 demonstrated significant allelic divergence relative to the genome-wide average FST value, additional

570 SNP outliers on other *H. virescens* scaffolds were also detected. In spite of the fact that we applied a

571 correction to reduce detection of false positives, it is possible that up to 35 of our total 351 SNP outliers

572 are false positives given our false discovery rate of 0.1. This correction threshold was selected to

573 minimize false positive detection, while retaining true positives (Verhoeven *et al.* 2005). To safeguard

574 against pursuing potential false positives, further research could initially focus on genes near outliers

575 with significant allele frequency changes in multiple by-year comparisons. Fifty-three SNPs showed

25

576   significant allelic divergence across multiple by-year comparisons, where major allele frequency shifts

577   took place between the years 1997 and 2007. Between 2007 and 2012, 5 SNPs showed significant

578   shifts in allele frequency and these changes were detected in multiple by-year comparisons as well. It

579   is likely that these SNP markers are in linkage disequilibrium with gene targets of selection as

580   management of cotton ecosystems has led to the replacement of conventional cotton cultivars with Bt-

581   expressing varieties. Interestingly, two of these SNPs were found on scaffolds containing genes with

582   plausible roles in detoxification (of insecticides and plant defensive compounds) or host plant

583   detection.

584   Scaffold 3424 contained SNPs that diverged significantly over time in our field-collected

585   populations of *H. virescens*. Divergence was strongest in by-year comparisons from 1997-2007, and

586   1997-2012. This suggests that most genomic change occurred between the years 1997 and 2007, and

587   that allele frequencies remained stable between the years 2007 and 2012. Blast results for predicted

588   gene sequences found on this scaffold revealed homology with the cytochrome p450 superfamily. The

589   predicted sequence aligned well with an *H. armigera Cyp6AE* gene, which is a cytochrome p450

590   family known to be involved in detoxification (Zhou *et al.* 2010). The best alignment was to an *H.*

591   *armigera Cyp6AE12*, and expression levels of this gene in *H. armigera* are modified in response to

592   pyrethroid insecticides (Yue *et al.* 2007, Zhou *et al.* 2010). It is possible that allelic changes on this

593   scaffold are a response to reduced pyrethroid use in the Southern United States as a result of Bt cotton

594   deployment in the agricultural landscape.

595   An alternative explanation exists for changes in allele frequencies at this candidate gene,

596   however. *Cyp6AE12* expression is also modified in response to the plant compound xanthotoxin (Zhou

597   *et al.* 2010). Another *Cyp6AE* monoxygenase, *Cyp6AE14*, has 68% protein sequence similarity to *H.*

598   *armigera Cyp6AE12*, is highly expressed in the midgut of *H. armigera,* and is likely involved in

26

599 detoxification of the cotton defensive compound gossypol (Mao *et al.* 2007). Indeed, RNAi silencing

600 of *Cyp6AE14* in *H. armigera* led to a decline in larval growth when gossypol was present in their diet

601 (Mao *et al.* 2007). If this *H. virescens* cytochrome p450 is indeed the target of selection, it is possible

602 that the divergence in allele frequencies between the years 2007 and 2012 was the result of selection for

603 improved larval performance in *H. virescens* populations that feed on alternative hosts. When

604 widespread planting of Bt-expressing cotton drove *H. virescens* off of their previously abundant cotton

605 host to alternative host plants, *H. virescens* would have been exposed to new plant defensive

606 compounds. Genes associated with larval performance in one host plant genera may or may not be

607 associated with improved larval performance on other host plant genera (Sheck and Gould 1993).

608 Therefore, allelic changes near our *H. virescens* SNP outlier may be caused by relaxation of selection

609 for gossypol detoxification as they moved out of cotton, or adaptation to new host plants and their

610 antifeedant chemicals. To determine whether phenotypes resulting from these molecular shifts are

611 directly associated with changing *H. virescens* management practices (e.g. pyrethoid or Bt toxin use) or

612 shifts in host plant use, further work could involve measuring associations between Bt or pyrethoid

613 response phenotypes and genotypes at target genes on these scaffolds.

614   Equally as important to host shifts is the ability of adult females to find a suitable host for

615 oviposition. Chemosensation is important for host plant identification in phytophagous insects, and

616 genes involved in olfaction and gustation are often targets of selection when host shifts occur (Dworkin

617 and Jones 2009, Smajda *et al.* 2012). An outlier SNP was detected on *H. virescens* Scaffold 2173 that

618 underwent allele frequency changes between the years 1997 and 2007, then remained stable between

619 the years 2007-2012. Within 10 kb of this outlier was a pair of olfactory receptor genes. It is plausible

620 that allele frequency changes at these olfactory receptor genes may reflect changes in the olfactory

621 percept, enabling females to identify non-cotton host plants. Future work directed at elucidating the

27

622 phenotypic effects of changes at these genes will be critical to determining their role, if any, in host

623 plant detection.

624   Both toxin metabolism and chemosensation can be complex traits with multiple loci involved.

625 Arrays of duplicated cytochrome p450s (Li *et al.* 2002, David *et al.* 2013) and carboxylesterases (Field

626 *et al.* 1988, Guillemaud *et al.* 1997) have been implicated in metabolic detoxification of chemicals in

627 other insect species. Likewise, clusters of olfactory and gustatory receptors are thought to contribute to

628 host-plant utilization in some insect species (Smajda *et al.* 2012). Therefore, we examined GO level 2

629 categorizations, particularly those related to detoxification and chemosensation, for the putative genes

630 surrounding our SNP outliers to determine if any GO categories were over-represented near our outliers

631 relative to the rest of the genome. Our results did not suggest there was any statistically significant

632 over-representation of any GO category near our outliers for any of our by-year comparisons. There

633 are at least two possible explanations for this: 1.) The targets of selection linked to our SNP outliers do

634 not necessarily involve duplicated, amplified, or arrayed genes of similar function. Instead these

635 targets may involve single copy genes or gene regulatory regions associated with adaptive phenotypes

636 in *H. virescens*. 2.) Only 30-40% of the putative genes in our genome were assigned a function in our

637 annotation pipeline. While our *H. virescens* draft assembly was instrumental to identifying selection at

638 the *Vgsc*, as well as identifying novel genomic regions under selection in field-collected populations,

639 there is room for improvement. Increasing the contiguity of our assembly will enable chromosome-

640 level identification of runs of homozygosity, a hallmark of selection. Likewise, manual curation efforts

641 will improve official gene set and the likelihood of identifying the gene targets near our SNP outliers.

642   While the gene targets of selection during these time periods require further validation, the

643 strength of selection near these targets, as measured by *s,* ranged from 0.009 and 0.623, depending

644 upon by-year comparison and degree of dominance assumed. Following theoretical expectations,

28

645 selection against q required higher selection coefficients when initial frequencies of q were high

646 (Figure 4). Furthermore, degree of dominance impacted *s*, where values ranged from 14-23% higher

647 when co-dominance was assumed, rather than dominance of p. Based upon theoretical expectations,

648 the coefficient of selection against q should be greater under assumptions of a recessive p and a

649 dominant q. Indeed, examples of selection against a dominant susceptible allele (q), resulting in the

650 increase in frequency of a recessive resistance allele (p) have been described in the insecticide

651 resistance literature (for example, see Ferré and Van Rie 2002). Across all outlier SNPs discovered

652 within each by year comparison, 1997-2012, 1997-2007, and 2007-2002, the average selection

653 coefficients were 0.047 and 0.058, 0.077 and 0.088, and 0.165 and 0.200, respectively, depending on

654 degree of dominance, and these are considered to be moderate to high values of *s* according to Eyre-

655 Walker and Keightley (2007), and Kim *et al.* (2017). Ultimately, these values of *s* demonstrate the

656 strong selection imposed on species found in agricultural ecosystems.

657   In conclusion, we demonstrated that ddRAD-seq enabled genomic scanning can be used to

658 identify organismal responses to anthropogenic changes in agricultural ecosystems, even on short, 15

659 year time scales. We then identified additional genomic regions in this Lepidopteran species that are

660 likely changing in response to shifts from conventional cotton planting to widespread Bt-cotton

661 adoption. From an applied perspective, our results suggest that ddRAD-seq genome scans may be

662 useful for monitoring pest populations for real-time changes in allele frequencies at loci responding to

663 the very strong selection imposed by insect management practices. We conclude that this technology

664 could be useful for identifying strong selection for resistance alleles, and then acting to mitigate

665 widespread phenotypic resistance to management practices across plant and insect species in

666 agricultural ecosystems.

667   **Data Availability**

668   Scripts and configuration files used for genome assembly can be found at:

669   https://github.com/mcadamme/Hv_Genome_Assembly_Draft1

670

671   Scripts used for population genomic analysis can be found at:

672   https://github.com/mcadamme/FieldHv_Pop_Genomics

673

674   Raw sequence ddRAD-seq data have been deposited in the NCBI SRA as: (updated following

675   acceptance of the manuscript)

676

677   Our *H. virescens* draft 1 assembly and associated read sequences have been deposited in the NCBI

678   database under BioProject number PRJNA379496.

679

690 **References**
691 Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of
692 RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics,* **17**, 81-92.
693
694 Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect
695 molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 28(9): 328.
696
697 Bacheler, J (1999) Pyrethroids no longer recommended for second generation tobacco budworms.
698 *Carolina Cotton Notes,* http://www.cotton.ncsu.edu/ccn/1999/ccn-99-5f.htm
699
700 Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson
701 EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One,* **3**,
702 e3376.
703
704 Barber GW (1937) Seasonal availability of food plants of two species of *Heliothis* in eastern Georgia.
705 Journal of Economic Entomology, **30**, 150-158.
706
707 BBMap - Bushnell B. - sourceforge.net/projects/bbmap/
708
709 Benbrook CM (2012) Impacts of genetically engineered crops on pesticide use in the U.S. – the first
710 sixteen years. *Environmental Sciences Europe,* **24**, 24.
711
712 Blanco CA (2012) *Heliothis virescens* and Bt cotton in the United States. *GM Crops & Food:*
713 *Biotechnology in Agriculture and the Food Chain,* **3**, 201-212.
714
715 Bolger, A. M., Lohse, M., and Usadel, B (2014) Trimmomatic: A flexible trimmer for Illumina
716 Sequence Data. *Bioinformatics,* btu170.
717
718 Brito LP, Linss JGB, Lima-Camara TN, Belinato TA, Peixoto AA, Lima JBP, Valle D, Martins AJ
719 (2013) Assessing the effects of *Aedes aegypti* kdr mutations on pyrethroid resistance and its fitness
720 cost. *PloS ONE,* **8**, e60878.
721
722 Buerkle CA and Gompert Z (2013) Population genomics based on low coverage sequencing: how low
723 should we go? *Molecular Ecology* **22**, 3028-3035.
724
725 Campanhola C and Plapp FW (1989) Pyrethroid resistance in the tobacco budworm (Lepidoptera:
726 Noctuidae): insecticide bioassays and field monitoring. *J Econ Entomol,* **82**, 22-28.
727
728 Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J, De Koning, D (2011) Stacks: Building
729 and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes Genomes Genetics,* **1**, 171-
730 182.
731

31

732 Catchen J, Hohenlohe P, Bassham S, Amores A, and Cresko W (2013) Stacks: an analysis tool set for
733 population genomics. *Molecular Ecology*, **22**, 3124-3140.
734
735 Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW (2017) Unbroken:
736 RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations.
737 *Mol Ecol Res* doi:10.1111/1755-0998.12669
738
739 Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics*. Roberts and Company
740 Publishers, Greenwood Village, Colorado, USA.
741
742 David J, Ismail H, Chandor-Proust A, Paine MJI (2013) Role of cytochrome P450s in insecticide
743 resistance: impact on the control of mosquito-borne diseases and use of insecticides on Earth. *Philos*
744 *Trans R Soc London B Biol Sci*, **368(1612)**, 20120429.
745
746 Dworkin I, and Jones CD (2009) Genetic changes accompanying the evolution of host specialization in
747 *Drosophila sechellia*. *Genetics* **181(2)**, 721-736.
748
749 Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K., Buckler ES, Mitchell SE (2011) A robust,
750 simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS ONE*, **6**, e19379.
751
752 English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs
753 RA (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing
754 technology. *PloS One*, **7**, e47768.
755
756 Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature*
757 *Reviews Genetics*, **8**, 610-618.
758
759 Falconer DS and Mackay TFC (1996) Introduction to Quantitative Genetics, 4th ed. Addison Wesley
760 Longman Limited. Essex, England.
761
762 Ferré F and Van Rie J (2002) Biochemistry and genetics of insect resistance to *Bacillus thuringiensis*.
763 Ann Rev Entomol, **47**, 501-533.
764
765 Field LM, Devonshire AL, Forde BG (1988) Molecular evidence that insecticide resistance in peach-
766 potato aphids (*Myzus persicae* Sulz.) results from amplification of an esterase gene. *Biochem J* **251**,
767 309-321.
768
769 Fitt GP (1989) The ecology of *Heliothis* species in relation to agroecosystems. *Ann. Rev. Entomol.* **34**,
770 17-52.
771
772 Foster SP, Denholm I, Thompson R, Poppy GM, Powell W (2005) Reduced response of insecticide-
773 resistant aphids and attraction of parasitoids to aphid alarm pheromone; a potential fitness trade-off.

774 *Bulletin of Entomological Research*, **95**, 37–46.

776 Fritz ML, Paa S, Baltzegar J, Gould F (2016) Application of a dense genetic map for assessment of
777 genomic responses to selection and inbreeding in *Heliothis virescens*. *Insect Molecular Biology,* **25(4)**,
778 385-400.

780 Gould F, Anderson A, Reynolds A, Bumgarner L, Moar W (1995) Selection and genetic
781 analysis of a *Heliothis virescens* (Lepidoptera: Noctuidae) strain with high levels of resistance to
782 *Bacillus thuringiensis* toxins. *Journal of Economic Entomology*, **88**, 1545-1559.

784 Gregory TR, Hebert PD (2003) Genome size variation in lepidopteran insects. *Canadian Journal of*
785 *Zoology*, **81**, 1399-1405.

787 Groot AT, Classen A, Inglis O, Blanco CA, Lopez J, Teran Vargas A, Schal C, Heckel DG, Schofl G
788 (2011) Genetic differentiation across North America in the generalist moth *Heliothis virescens* and the
789 specialist *H. subflexa*. *Molecular Ecology* **20 (13)**, 2676-2692.

791 Grubor VD, Heckel DG (2007) Evaluation of the role of CYP6B cytochrome p450s in pyrethroid
792 resistance in Australian *Helicoverpa armigera*. *Insect Mol Biol*, **16**, 15-23.

794 Guillemaud T, Makate N, Raymond M, Hirst B, Callaghan A (1997) Esterase gene amplification in
795 *Culex pipiens*. *Insect Molecular Biology*, **6**, 319-327.

797 Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood P D, Bowden J, Regev A (2013) De novo
798 transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation
799 and analysis. *Nature Protocols*, **8**, 1494–1512.

801 Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, Wortman JR (2008) Automated eukaryotic
802 gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments.
803 *Genome Biology*, **9**, R7.

805 Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, White O (2003) Improving the
806 Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids*
807 *Research*, **31**, 5654–5666.

809 Heckel DG, Gahan LJ, Baxter SW, Zhao J, Shelton AM, Gould F, Tabashnik BE (2007) The diversity
810 of Bt resistance genes in species of Lepidoptera. *Journal of Invertebrate Pathology,* **95**,192-197.

812 Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics
813 of parallel adaptation in threespine stickleback using sequenced RAD tags. *PloS Genet*, **6**, e1000862.

815 Huang S, Chen Z, Huang G, Yu T, Yang P, Li J, Fu Y, Yuan S, Chen S, Xu A (2012) HaploMerger:

33

reconstructing allelic relationships for polymorphic diploid genome assemblies. *Genome Res*, **22**, 1581-1588.

Jallow MFA, Zalucki MP (1996) Within- and between-population variation in host-plant preference and specificity in Australian *Helicoverpa armigera* (Hubner)(Lepidoptera: Noctuidae). *Aust. J. Zool.*, **44**, 503-519.

Jallow MFA, Cunningham JP, Zalucki MP (2004) Intra-specific variation for host plant use in *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae): implications for management. *Crop Protection*, **23**, 955-964.

James, C (2015) 20th Anniversary (1996 to 2015) of the Global Commercialization of Biotech Crops and Biotech Crop Highlights in 2015. *ISAAA Brief* No. 51. ISAAA: Ithaca, NY.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenetic and Genome Research 110:462-467

Karpinski A, Haenniger S, Schofl G, Heckel DG, Groot AT (2014) Host plant specialization in the generalist moth *Heliothis virescens* and the role of egg imprinting. *Evol Ecol.*, **28**, 1075-1093.

Kim BY, Huber CD, Lohmueller KE (2017) Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics*, **206 (1)**, 345-361.

Kliot A, Ghanim M (2012) Fitness costs associated with insecticide resistance. *Pest. Manag. Sci.*, **68**, 1431–1437.

Langmead B, Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357-359.

Li X, Berenbaum MR, Schuler MA (2002) Plant allelochemicals differentially regulate *Helicoverpa zea* cytochrome P450 genes. *Insect Molecular Biology*, **11(4)**: 343-351.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27, 2987-2993.

Lischer HEL and Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28: 298-299.

858 Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-Seq reads into automatic
859 training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, **42**, e119.
860
861 Louisiana State University Agricultural Center (2016) Louisiana Insect Pest Management Guide.
862 http://www.laca1.org/Pub%201838%20Insect%20Pest%20Management%20Guide%202016_complete
863 %20book.pdf
864
865 Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A (2017) Breaking RAD:
866 an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of
867 adaptation. *Mol Ecol Res* 17: 142-152.
868
869 Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi
870 Y, Liu Y, Yu C, Wang B, Yao L, Han C, Cheung DW, Yiu S, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y,
871 Yang H, Wang J, Lam T, Wang, J (2012) SOAPdenovo2: an empirically improved memory-efficient
872 short-read de novo assembler. *GigaScience*, **1**, 18.
873
874 Luttrell RG, Roush RT, Ali A, Mink JS, Reid MR (1987) Pyrethroid resistance in field populations of
875 *Heliothis virescens* (Lepidoptera: Noctuidae) in Mississippi in 1986. *J Econ Entomol,* **80**, 985-989.
876
877 Magoc T, Salzberg S (2011) FLASH: Fast length adjustment of short reads to improve genome
878 assemblies. *Bioinformatics*, **27**, 2957-2963.
879
880 Mao Y, Cai W, Wang J, Hong G, Tao X, Want L, Huang Y, Chen X (2007) Silencing a cotton bollworm
881 P450 monooxygenase gene y plant-mediated RNAi impairs larval tolerance of gossypol. *Nature
882 Biotechnology* **25**, 1307-1313.
883
884 Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads.
885 *EMBnet.journal*, **17**, 10.
886
887 Nadeau NJ, Ruiz M, Salazar P, Counterman B, Medina JA, Ortiz-Zuazaga H, Morrison A, McMillan
888 WO, Jiggins CD, Papa R (2014) Population genomics of parallel hybrid zones in the mimetic
889 butterflies, *H. melpomene* and *H. erato*. *Genome Res,* **24**, 1316-1333.
890
891 NASEM, National Academies of Sciences, Engineering, and Medicine (2016) *Genetically Engineered
892 Crops: Experiences and Prospects.* Washington, DC: The National Academies Press.
893
894 Nielsen R (2005) Molecular signatures of natural selection. *Annu. Rev. Gen.,* **39**, 97-218.
895
896 Onstad, DW (2014) Insect Resistance Management (Second Edition) Biology, Economics, and
897 Prediction. Elsevier Ltd. ISBN: 978-0-12-396955-2
898
899 Oppenheim SJ, Gould F, Hopper KR (2012) The genetic architecture of a complex ecological trait in

35

900   the specialist moth *Heliothis subflexa*. *Evolution*, **66(11)**, 3336-3351.

901

902   Pais AL, Whetten RW, Xiang Q (2016) Ecological genomics of local adaptation in *Cornus florida* L. by
903   genotyping by sequencing. Ecology and Evolution, **00**, 1–25. doi:10.1002/ece3.2623

904

905   Palumbi SR (2001) Humans as the world's greatest evolutionary force. *Science*, **293**, 1786-1790.

906

907   Park S, Brown TM (2002) Linkage of genes for sodium channel and cytochrome P450 (CYP6B10) in
908   *Heliothis virescens*. *Pest management science*, **58**, 209-212.

909

910   Park Y, Taylor MFJ (1997) A novel mutation L1029H in sodium channel gene hscp associated with
911   pyrethroid resistance for *Heliothis virescens* (Lepidoptera Noctuidae). *Insect Biochem. Mol. Biol.*, **27**,
912   9-13.

913

914   Park Y, Taylor MFJ, Feyereisen R (1997) A Valine421 to Methionine mutation in IS6 of the hscp
915   voltage-gated sodium channel associated with pyrethroid resistance in *Heliothis virescens* F. *Biochem.*
916   *Biophy. Res. Com.*, **239**, 688-691.

917

918   Peterson BK, Weber JN, Kay EH, Fisher H., Hoekestra HE (2012) Double digest RADseq: An
919   inexpensive method of de novo SNP discovery and genotyping in model and non-model species. *PloS*
920   *ONE*, **7**, e37135.

921

922   Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. To
923   appear in Proceedings of the 13th Annual International conference on Intelligent Systems for Molecular
924   Biology (ISMB-05). Detroit, Michigan.

925

926   Pujolar JM, Jacobson MW, Als TD, Frydenberg J, Munch K, Jonsson B, Jian JB, Cheng L, Maes GE,
927   Bernatchez L, Hansen MM (2014) Genome-wide single-generation signatures of local selection in the
928   panmictic European eel. *Molecular Ecology*, **23**, 2514-2528.

929

930   Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features.
931   *Bioinformatics*. **26**, 841–842.

932

933   R Core Team (2014). R: A language and environment for statistical computing. R Foundation for
934   Statistical Computing, Vienna, Austria.

935

936   Remmert M, Biegert A, Hauser A, Soding J (2012) HHblits: lightning-fast iterative protein sequence
937   searching by HMM-HMM alignment. *Nat Meth*, **9**, 173–175.

938

939   Rinkevich FD, Du Y, and Dong K (2013) Diversity and convergence of sodium channel mutations
940   involved in resistance to pyrethroids. *Pesticide Biochemistry and Physiology*, **106**(3), 93-100.

941

942 Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and
943 genomic islands in divergence of a songbird. *Mol Ecol*, **23**, 4757-4769.
944
945 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011)
946 Integrative Genomics Viewer. *Nature Biotechnology* **29**, 24–26.
947
948 Sheck AL, Gould F (1993) The genetic basis of host range in *Heliothis virescens*: larval surival and
949 growth. *Entomol. Exp. Appl.*, **69**, 157-172.
950
951 Sheck AL, Gould F (1995) Genetic analysis of differences in oviposition preferences of *Heliothis*
952 *virescens* and *H. subflexa* (Lepidoptera: Noctuidae). *Environ. Entomol.*, **24(2)**, 311-347.
953
954 Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zbodnov EM (2015) BUSCO: assessing
955 genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* doi:
956 10.1093/bioinformatics/btv351
957
958 Smajda CM, Canback B, Vitalis R, Gautier M, Ferrari J, Zhou J, Butlin RK (2012) Large-scale
959 candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation
960 in the pea aphid. *Evolution* **66(9)**, 2723-2738.
961
962 Smit AFA, Hubley R, Green P (2013) *RepeatMasker Open-4.0.*
963
964 Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: Ab initio
965 prediction of alternative transcripts. *Nucleic Acids Research*, **34**.
966
967 Sudbrink DL, Grant JF (1995) Wild host plants of *Helicoverpa zea* and *Heliothis virescens*
968 (Lepidoptera: Noctuidae) in Eastern Tennessee. *Environ. Entomol.*, **24(5)**, 1080-1085.
969
970 Tabashnik BE, Brevault T, Carriere Y (2013) Insect resistance to Bt crops: lessons from the first billion
971 acres. Nature Biotechnology, **31**, 510-521.
972
973 Taylor M, Shen Y, Kreitman M (1995) A population genetic test of selection at the molecular level.
974 *Science*, **270**, 1497-1499.
975
976 Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo,
977 Robert Handsaker, Gerton Lunter, Gabor Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin
978 and 1000 Genomes Project Analysis Group (2011) The Variant Call Format and VCFtools.
979 *Bioinformatics*, **27 (15),** 2156-2158.
980
981 Tilman D, Fargione J, Wolff B, D'Antonio C, Dobson A, Howarth R, Schindler D, Schelesinger WH,
982 Simberloff D, Swackhamer D (2001). Forecasting agriculturally driven global environmental change.
983 *Science,* **292**, 281-284.
984

985   [USEPA] U.S. Environmental Protection Agency (1998) The environmental protection agency's white
986   paper on Bt plant-pesticide resistance management. Washington: http://www.epa.gov/scipoly/sap/
987   meetings/1998/february/finalfeb.pdf
988
989   [USEPA] U.S. Environmental Protection Agency (2001) Biopesticides registration action document:
990   *Bacillus thuringiensis* plant-incorporated protectants. http://www.epa.gov/pesticides/biopesticides/
991   pips/bt_brad.htm
992
993   [USEPA] U.S. Environmental Protection Agency (2006) Analysis of a Natural Refuge of Non-Cotton
994   Hosts for Monsanto's Bollgard II Cotton. FIFRA Scientific Advisory Panel Meeting June 13-15, 2006.
995   Arlington, Virginia. 101pgs.
996
997   Verhoeven KJF, Simonsen KL, McIntyre LM (2005) Implementing false discovery rate control:
998   increasing your power. Oikos, 108(3), 643-647.
999
1000 Yue LN, Yang YH, Wu SW, WU YD (2007) Cloning and mRNA expression levels of cytochrome P450
1001 genes *CYP6AE12* and *CYP9A18* in the cotton bollworm, *Helicoverpa armigera*. *ACTA Enotomologica*
1002 *Sinica*, **50**, 234-240.
1003
1004 Zhao Y, Park Y, and Adam ME (2000). Functional and evolutionary consequences of pyrethroid
1005 resistance mutations in S6 transmembrane segments of a voltage-gated sodium channel. Biochemical
1006 and Biophysical Research Communications 278: 516-521.
1007
1008 Zhou X, Sheng C, Li M, Wan H, Liu D, Qui X (2010). Expression responses of nine cytochrome p450
1009 genes to xenobiotics in the cotton bollworm *Helicoverpa armigera*. Pesticide Biochemistry and
1010 Physiology 97:209-213.
1011

1012

1013 **Tables**

1014 **Table 1 –** Number of SNPs with pairwise FST values deemed as significantly divergent according to

1015 Lositan analysis for each by-year comparison. The number of unique scaffolds (of 8826) containing at

1016 least 1 significantly diverged SNP, as well as the average and maximum numbers of putative genes

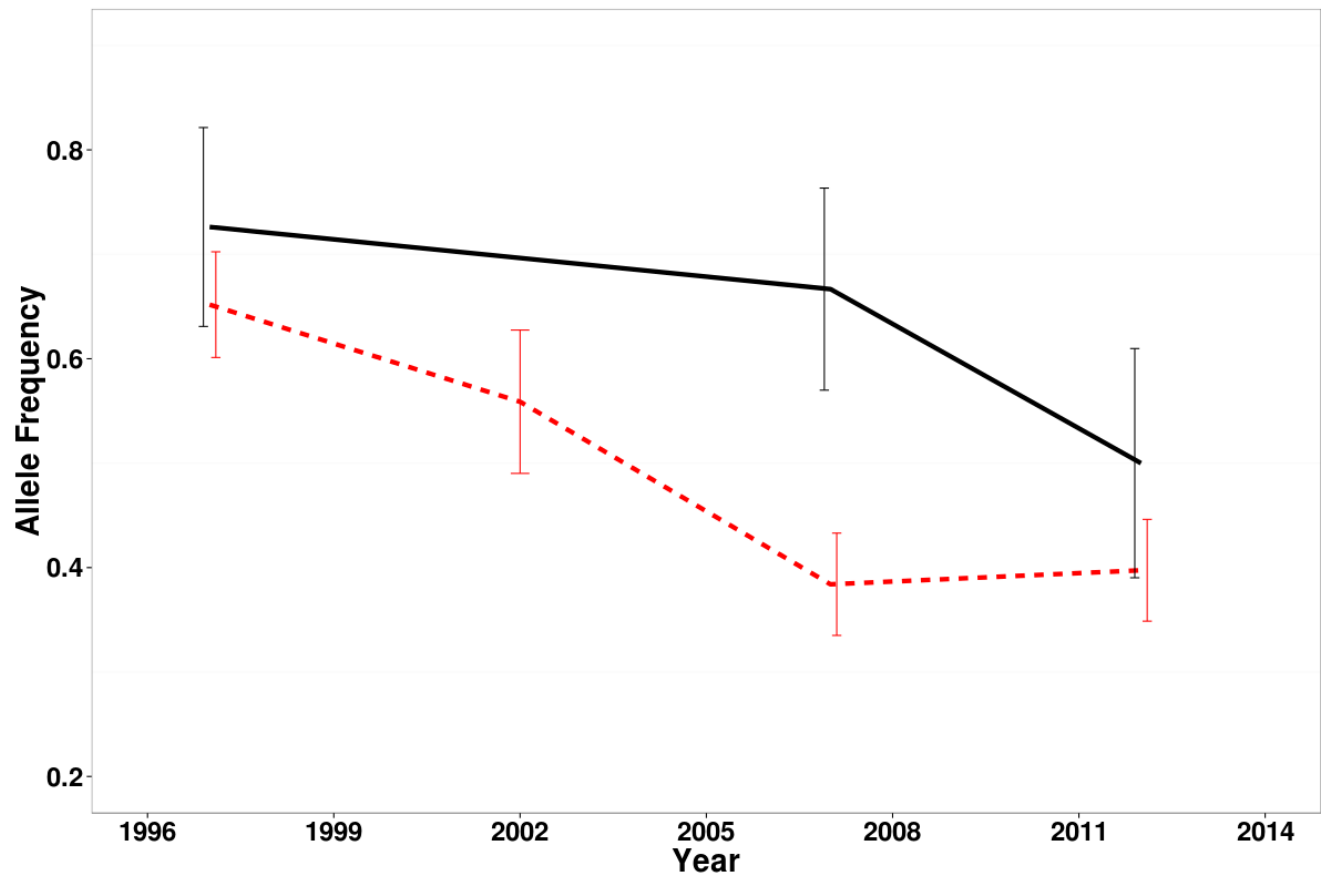1017 within 36 kb of the outlier SNP is also included.

| Pairwise Comparisons	Total SNPs	Number Outlier SNPs	Number Scaffolds Containing Outliers	Mean Number Putative Genes within 36 kb of Outlier SNPs	Max Number Putative Genes within 36 kb of Outlier SNPs
1997-2007	8963	201	190	2.6	12
2007-2012	8963	35	33	2.6	10
1997-2012	8963	184	170	2.7	10

1018

1019

1020 **Figures**

1021

1023 **Figure 1 –** The decline in the frequency of the pyrethroid resistance allele in *H. virescens* (pooled LA

1024 and TX samples), represented by the dashed red line, was statistically significant over the course of our

1025 15 year sampling period (n = 659). A unique ddRAD-seq haplotype, represented by the solid black

1026 line, was found *ca.* 36Kb upstream from the alpha subunit of the *Vgsc* and also declined in frequency in

1027 the subset of individuals (n = 141) sequenced for our genome scan. Error bars represent bootstrapped

1028 95% confidence intervals (N = 5000) around the mean of each year.

1029

40
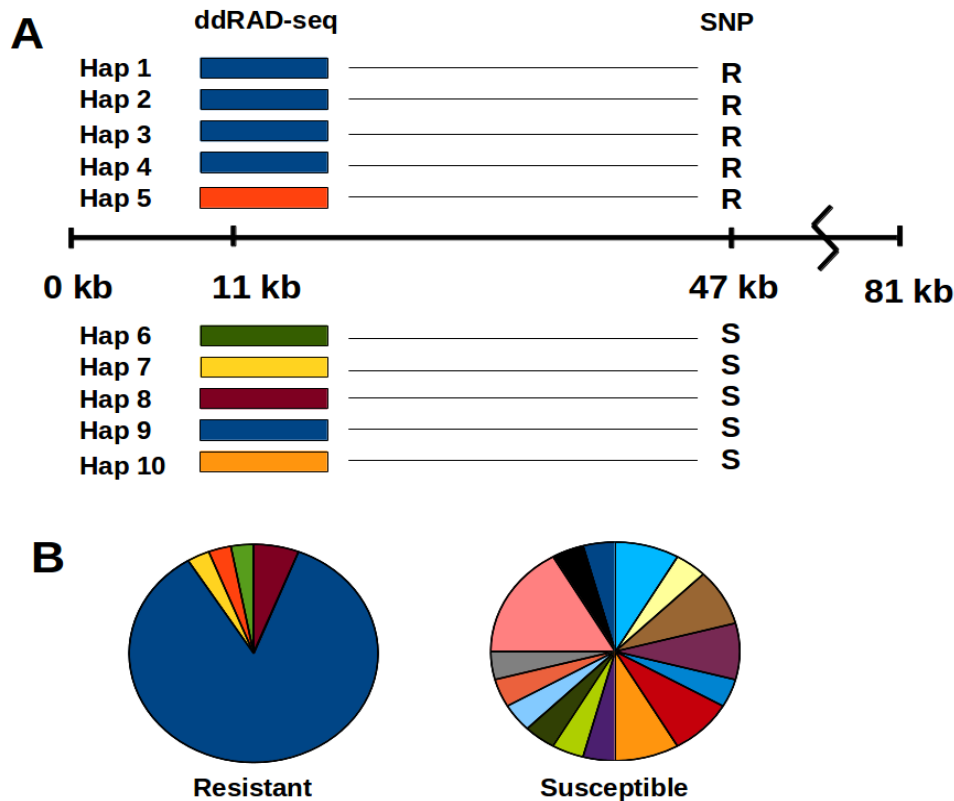
1030 **Figure 2 –** Significantly greater diversity was observed from ddRAD-seq haplotypes linked to the

1031 susceptible *Vgsc* allele relative to those linked to the resistance allele. (A) is a visualization of the

1032 relationship of Hv_11322 to the L1029H *Vgsc* SNP along an 81kb genome scaffold. The different

1033 colored bars at 11kb represent unique alleles at the Hv_11322 locus, showing a greater diversity of

1034 Hv_11322 alleles were found to be associated with the *Vgsc* wild-type allele, relative to the resistance

1035 allele as would be indicative of a selective sweep. Due to the number of unique Hv_11322 alleles

1036 found to be associated with each *Vgsc* allele (n = 5 and n = 15 for resistant and wild-type, respectively),

1037 full representation of this diversity could not be incorporated into (A). However, the unique colors in

1038 (B) depict the number and proportion of Hv_11322 alleles linked to the *Vgsc* resistant (n = 34) and

1039 susceptible (n =24) SNP alleles. Each unique color represents a different Hv_11322 allele. The dark

1040 blue wedges always represent Hv_11322_hap 1.
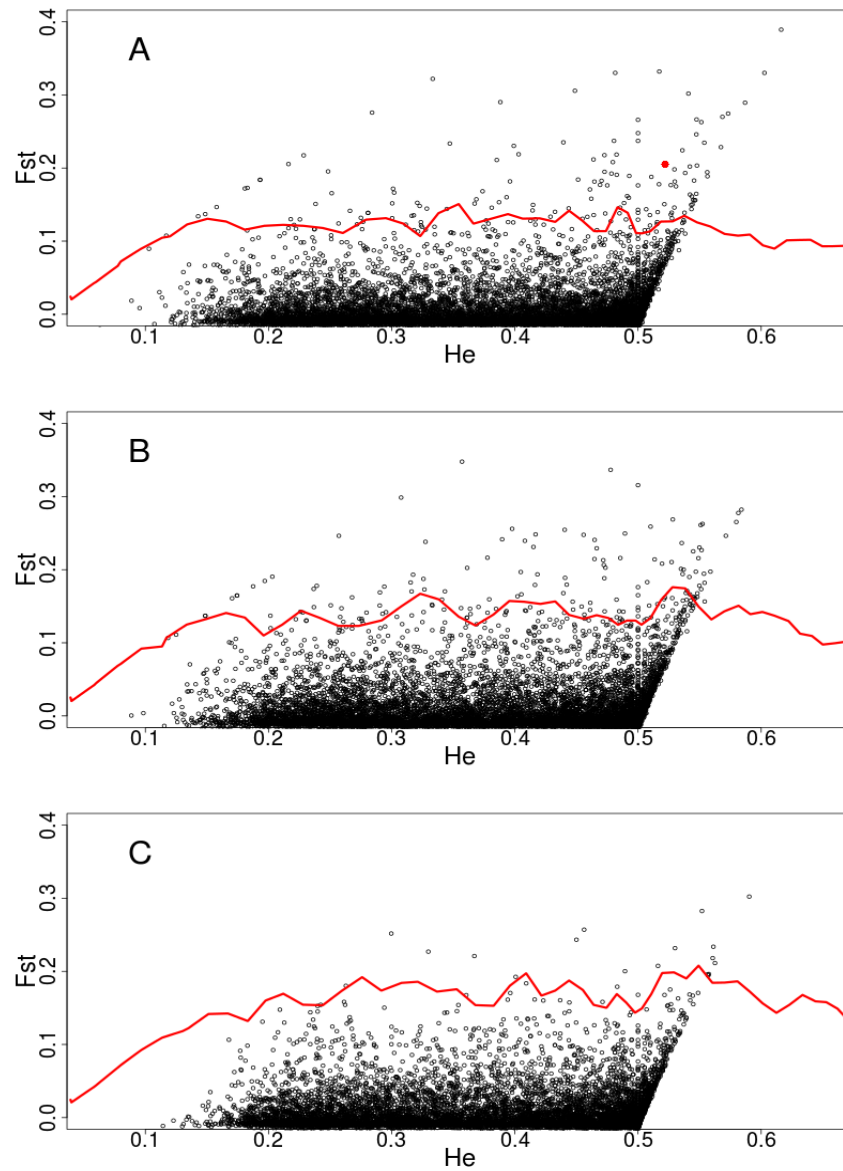
1041 **Figure 3** – Pairwise genetic divergence according to Weir and Cockerham's FST for populations of *H.*

1042 *virescens* collected in the years A) 1997 and 2012, B) 1997 and 2007, C) 2007 and 2012. Each black

1043 point represents one SNP of 8,963 along the *H. virescens* genome. Points above the red line represent

1044 loci with pairwise genetic divergence that is statistically significant at the $\alpha = 0.01$ level following

1045 correction for false discovery. Pairwise genetic divergence at the SNP near the *Vgsc* on Scaffold 4600

1046 is represented by the red point on panel A.
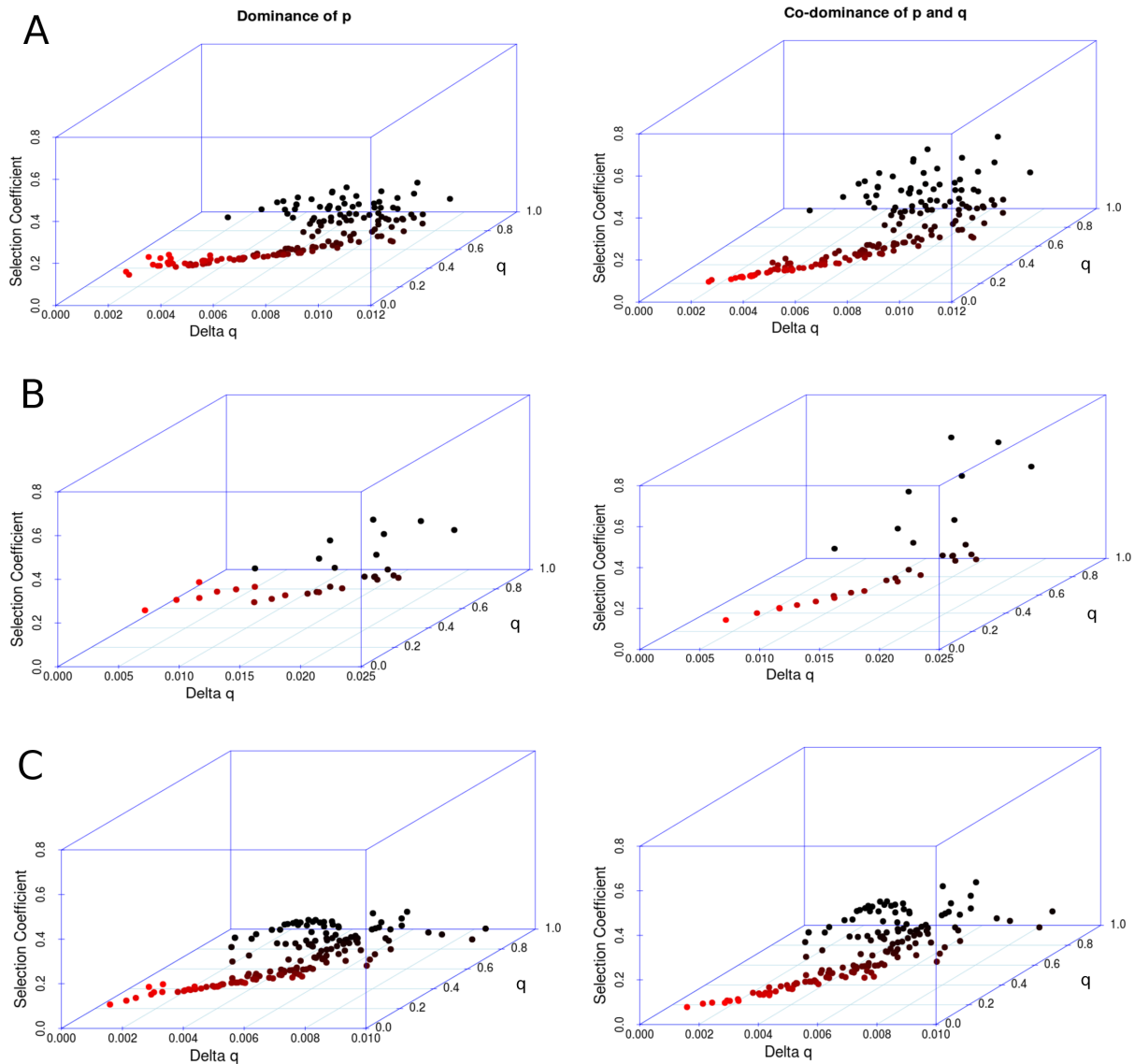
Figure 4 – Selection coefficients associated with SNP outliers where the initial frequency of p was greater than 0.05. Selection against q is plotted against the rate of allele frequency change and the initial starting frequency of q. Colors for each plotted point help to visualize the initial starting frequency of q, where red is low and black is high. Plots in rows A, B, and C represent selection coefficients calculated assuming dominance and codominance of p for each of the by-year comparisons 1997-2007, 2007-2012, and 1997-2012, respectively.