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

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

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

[§]Corresponding Author

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 populations in the Southern United States have been declining since the widespread planting of transgenic cotton in the late 1990s. These cotton cultivars endogenously express proteins derived from the bacterium *Bacillus thuringiensis* (Bt), which are lethal to *H. virescens*. No physiological adaptation to Bt toxin has been found, so adaptation to this altered environment could involve: 1) shifts in host plant selection mechanisms to avoid cotton, or 2) changes in detoxification mechanisms required for cotton-feeding versus feeding on other host plants. A decline in pyrethroid use in Bt cotton landscapes likely also led to reversion to susceptible alleles at loci involved in expression of pyrethroid resistance. Here we begin to address the question of whether such changes occurred in *H. virescens* populations between the years 1997 and 2012. In a proof of concept experiment, we confirmed that allele frequency changes at the voltage-gated sodium channel gene, a pyrethroid resistance locus, could be detected through a genomic scanning technique that depends on linkage disequilibrium between molecular markers and gene targets of selection. A direct PCR approach first confirmed a decline in frequency of the pyrethroid resistance allele in *H. virescens* populations over time. We then tested the hypothesis that this known genetic change could be detected via a ddRAD-seq enabled genome scan in concert with our new *H. virescens* genome assembly. One ddRAD-seq marker was physically linked to the sodium channel gene, and the rate of allele frequency change at that marker was similar to that of the sodium channel resistance allele. We then identified additional ddRAD-seq loci with significant allele frequency changes over the 15 year study period. Genes near these ddRAD-seq loci were identified and their potential contributions to adaptive phenotypes are discussed.

Key Words

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

Introduction

Human-induced change in the natural landscape places strong selective pressure on populations to adapt over relatively short evolutionary timescales (Palumbi 2001). Identifying the ways in which human-induced environmental change shapes the genomes of local species can provide insight into contemporary evolution and it's implications for affected species. Cultivation of the natural landscape for agricultural purposes is one of the most ubiquitous examples of human-induced environmental change. Modern agricultural practices often involve sweeping changes to the composition of plant species across broad geographic regions, re-sculpting of the physical terrain and chemical inputs into the environment (Tilman 2001). The strong selective pressure placed on species that inhabit agricultural ecosystems make them ideal for examining genetic responses to anthropogenic forces (Taylor et al. 1995).

One such major change in recent agricultural history is the commercialization of transgenic crops that themselves produce proteins for the management of key insect species. The tobacco budworm, *Heliothis virescens*, feeds on cotton in its larval stages and populations in the Southern United States have been declining since the widespread planting of transgenic cotton (Supplementary Figure 1). These cotton cultivars endogenously express insecticidal proteins derived from the bacterium *Bacillus thuringiensis* (Bt), which are lethal to *H. virescens*. In the Southern United States, Bt-expressing cotton was rapidly adopted after it became commercially available for management of *H. virescens* in 1996 (James 2015; Supplementary Figure 2). Prior to the widespread use of Bt-expressing cotton, populations of *H. virescens* had evolved resistance to every insecticide used for their management (Blanco 2012), including pyrethroid insectides (Luttrell et al. 1987, Campanhola and Plapp 1989). Concerns over the possibility that *H. virescens* and other insect targets of Bt crops would evolve resistance to the endogenously expressed proteins spawned an entire field of research related to

Bt resistance and associated genetic mechanisms (Reviewed in Heckel et al. 2007, Tabashnik et al. 2013). Of primary concern was the loss of efficacy of toxic Bt proteins (USEPA 1998, 2001, 2006). In the case of *H. virescens*, no physiological adaptation to the Bt toxin in the cotton has been detected (Tabashnik et al. 2013). Yet widespread adoption of Bt-expressing crops likely placed selective pressure on *H. virescens* in other ways.

As one example, widespread planting of Bt cotton cultivars led to an overall decline in insecticide use on cotton in the United States (NASEM 2016), including the use of pyrethroids. Prior to Bt cotton adoption, the *H. virescens* voltage-gated sodium channel gene was described as one gene target of selection wherein pyrethroid resistance alleles rose to high frequency (Park and Taylor 1997, Park et al. 1997). Yet in *H. virescens* and other insect species, voltage-gated sodium channel gene mutations often result in an overall loss of fitness for individuals carrying them (Zhao et al. 2000, Foster et al. 2005, Kliot and Ghanim 2012, Brito et al. 2013). Under these conditions, the stability in the frequency of insecticide resistance alleles depends upon whether or not populations are continually exposed to insecticidal pressure. Therefore, one possible effect of Bt adoption in *H. virescens* is a reversion to susceptibility at their pyrethroid resistance locus. Additional inadvertent targets of selection by Bt-expressing cotton could include loci involved in feeding and oviposition behaviors as *H. virescens* was driven off of its primary host plant (Blanco 2012).

In recent years, identifying genomic change in response to selective forces has been enabled by the development of next-generation sequencing (NGS) technologies. A variety of NGS-enabled marker development techniques are used to generate novel, high density marker sets for model and non-model organisms, including Restriction-site Associated DNA sequencing (RAD-seq; Baird *et al.* 2008), Genotyping-by-Sequencing (GBS; Elshire *et al.* 2011), double-digest RAD-seq (ddRAD-seq; Peterson *et al.* 2012) and others (reviewed in Andrews et al. 2016). These marker sets enable scientists to scan the genomes of field-collected organisms in search of the gene targets of selection. Strong selection for advantageous alleles at target genes also frequently influences allelic composition at physically linked neutral sequences, including marker sites (Nielsen 2005). This results in a genomic footprint of selection that is much broader than the target gene alone. The breadth of this genomic footprint is influenced by several factors, including the strength of selection, the initial frequency of the advantageous allele, effective pest population size and recombination rate (Charlesworth and Charlesworth 2010).

Here we scanned the genomes of two *H. virescens* field populations collected in the Southern United States between the years 1997 and 2012 to detect loci that have changed over time. Given that pesticides impose very strong selection on their target pest species (Onstad 2014), we initially focused on genomic regions associated pyrethroid resistance to demonstrate the power of ddRAD-seq to identify genes responsible for adaptive phenotypes. To achieve this goal, we produced an annotated draft assembly of the *H. virescens* genome and used it for alignment of ddRAD-seq reads from barcoded individuals collected across space and time. As proof of concept, we tested the hypothesis that changes in a candidate pyrethroid resistance gene, the voltage-gated sodium channel gene, can be detected through our ddRAD-enabled genome scanning techniques. We then identified several additional genomic regions with strongly diverging marker allele frequencies, some of which are linked to other potentially important insecticide resistance or host plant detoxification genes. Finally, we discuss the adaptive phenotypes that these newly identified gene targets of Bt selection might represent in a field environment.

Methods

Insect Material

For all population genomic analyses, adult male moths were collected by pheromone-baited trap from Bossier Parish, LA, and Burleson County, TX. Collections took place in LA from May through September, and in TX from May through October, in the years 1997, 2002, 2007, and 2012. Moths from each collection date were immediately placed in bottles of 95% ethanol for long-term storage. Bottles from 2002, 2007 and 2012 were always held at -20°C until specimens were used, while those from 1997 were initially held at room temperature and then transferred to -20°C. To develop our *H. virescens* genome assembly, individuals from a long-standing colony strain (Gould et al. 1995) were sib-mated for 10 generations to produce inbred material for sequencing (Fritz et al. 2016). Siblings from a single inbred family were used for sequencing and analysis. Five sibling pupae were stored at -80°C prior to DNA isolation and library preparation. For all insect samples, DNA was isolated with a Qiagen Blood and Tissue Kit (Qiagen, Inc., Valencia, CA, U.S.A.) using the mouse tail protocol.

H. virescens *Candidate Gene Approach*

A polymerase chain reaction (PCR) based upon the methods of Park and Taylor (1997) was used to amplify a 432 bp region in the alpha subunit of the voltage-gated sodium channel gene. The primer pair Nhp3304+ (5' ATGTG GGACT GIATG TTGGT) and Nhp3448- (5' CTGTT GAAGG CCTCT GCTAT) flanked a mutation known as L1029H. In this targeted region of the voltage-gated sodium channel gene, a single nucleotide polymorphism (SNP) caused a Leucine to Histidine amino acid substitution and thereby pyrethroid resistance. Amplicons were digested by restriction enzyme Nla-III, which cut in the presence of the resistance allele (Figure 1). Genotypes were scored by visualizing the digested PCR products on a 3.5% agarose gel (90 to 120 min at 120 V). We examined the genotypes at this pyrethroid resistance locus for *H. virescens* individuals collected from 1997 (n = 194), 2002 (n = 204), 2007 (n = 268), and 2012 (n = 194) in LA, and 1997 (n = 142), 2007 (n = 120), and 2012 (n = 196) in TX. We tested for changes in pyrethroid resistance allele frequencies over time and space using a series of nested generalized linear regression models with binomial error structures in R version 3.1.2 (R Core Team 2014; used here and throughout). The following full model was used to examine the frequency of individual pyrethroid resistance alleles (*i*):

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

for $i = 1, \ldots., n$

where Year represents collection year (e.g. 1997, 2002, 2007, or 2012), Season represents whether the collections were made early (May or June) or late (August through October) in the cotton growing season, and Location represents the collection location of the samples. We identified a model term as statistically significant (α = 0.05) when a comparison of nested models by analysis of deviance indicated that removal of that term significantly influenced model deviance.

Illumina WGS Library Preparation and Sequencing

Genomic DNA (gDNA) from one pupa was submitted to the North Carolina State Genomic Sciences Laboratory (Raleigh, NC, USA) for Illumina paired-end (PE) library construction and sequencing. Prior to library preparation, the DNA template was quantified by a Qubit 2.0 Fluorometer (Invitrogen, USA). The PE library with an 800bp insert size was constructed using an Illumina TruSeq Nano Library Kit (Illumina, Inc. San Diego, CA) according to standard protocol. Following enrichment by PCR, the library was checked for quality and final concentration using an Agilent 2100

Bioanalyzer (Agilent Technologies, USA) with a High Sensitivity DNA chip before sequencing on an Illumina HiSeq 2500 (100x2 paired end, rapid run).

Genomic DNA from a second pupa was used for mate-pair (MP) sequencing. Prior to library preparation, whole gDNA was run out on a 0.5% agarose gel at 130v for 2 hours. Fragments 8kb or larger, as compared with Hyperladder I (Bioline USA Inc. Tauton, MA, U.S.A), were excised from the gel and purified using a Zymoclean large fragment recovery kit (Zymo Research Corp. Irvine, CA, U.S.A.). The DNA sample was submitted to the Michigan State University Research and Technology Support Facility (East Lansing, MI, USA) for 8kb MP library preparation and sequencing. The DNA library was prepared using an Illumina Nextera Mate Pair Sample Preparation Kit according to standard protocol. The library was validated using a Qubit dsDNA assay, Caliper LabChipGX (Perkin Elmer, Waltham, MA, U.S.A.) and Kapa Library Quantification qPCR for Illumina Libraries. The library was loaded on one lane of an Illumina HiSeq 2500 High Output flow cell and sequenced in a 2x125bp paired-end format using HiSeq SBS version 4 reagents. For both PE and MP libraries, base calling was done by Illumina Real Time Analysis (RTA) v1.18.64 and the output of RTA was converted to FastQ format with Illumina Bcl2fastq v1.8.4.

PacBio Library Preparation and Sequencing

Genomic DNA from 4 pupae, one of which was also used for Illumina PE sequencing, were prepared into two libraries for PacBio sequencing. For each library, the SMRTbell Template Preparation Kit version 1.0 (Pacific Biosciences, Menlo Park, CA, U.S.A.) was used for gDNA preparation, but shearing and size-selection steps differed. For the first library, shearing was minimal and no size selection was performed. For the second library, shearing prior to DNA concentration was avoided to maximize gDNA fragment length, and a BluePippin (Sage Science Inc., Beverly, MA, U.S.A.) was used to select fragments that were at least 7kb long. This produced sufficient prepared library material for 17 and 5 SMRTcells, respectively. Prior to sequencing, the library concentration and fragment length profiles were checked on a Qubit 2.0 and an Agilent Tapestation 2200 (Agilent Technologies, USA) with a high molecular weight tape. Both libraries were sequenced at the University of North Carolina Chapel Hill Sequencing facility on a PacBio RS II.

H. virescens *Genome Assembly*

Read quality was checked for all Illumina data using FastQC (Babraham Bioinformatics, Cambridge, UK). Low quality ends were trimmed from both PE and MP reads using trimmomatic (v. 0.32; Bolger et al. 2014) and cutadapt (v. 1.9.1; Martin 2011), respectively. Any remaining Illumina adapter sequences and Nextera transposon sequences were also removed. Reads were filtered for potential microbial contaminants and *H. virescens* mitochondrial DNA (Supplementary Data File 1) using BBmap (version 35.10; Bushnell B. - sourceforge.net/projects/bbmap/). For the full list of the screened contaminants, see Supplementary Table 1. SOAPdenovo2 (v. 2.04) was used for assembly, scaffolding and gap closure (Luo et al. 2012). We attempted the assembly with multiple k-mer lengths, where k was set equal to either 47, 55, or 63. For each of the three assemblies, all contigs and scaffolds under 2kb were not used for further analysis. A CEGMA analysis was used to examine completeness of a conserved eukaryotic gene set for all three assemblies (version 2.4.010312; Parra et al. 2007). Further refinement was directed at the K63 assembly because it had better contiguity and similar completeness relative to the other two assemblies (Supplementary Table 2).

RepeatScout (version 1.0.5; Price et al. 2005) was used to find *de novo*, species-specific repeats in the K63 assembly, while RepeatMasker (version open-4.0; Smit et al. 2015) was used to identify other common insect repeats available from Repbase (version 20150807; Jurka et al. 2005). We then soft-masked both repeat classes using BEDTools (version 2.25.0; Quinlan and Hall 2010) and collapsed redundant haplotypes using the default settings in Haplomerger2 (version 3.1; Huang et al. 2012). To fill intra-scaffold gaps, we applied PacBio reads that were over 5kb in length to our Illumina assembly using PBsuite (version 14.9.9; English et al. 2012). Finally, we used BlastStation (TM Software, Inc., Arcadia, CA, U.S.A.) to align 654 mapped ddRAD-seq marker sequences from the F1 parent used to produce an *H. virescens* linkage map (Fritz et al. 2016; Dryad digital repository http://dx.doi.org/10.5061/dryad.567v8) to our scaffolds. All top hits were exported and markers with alignment hit lengths greater than 150bp (of 350 bp total), identities greater than 80%, and e-values below 0.001 were further examined. This enabled us to check for potential misassemblies, and provide additional information about which short scaffolds likely belong together on individual chromosomes (Supplementary Table 3). BlastStation was also used to identify the scaffold to which the alpha-subunit of the voltage-gated sodium channel (GenBank Accession: AH006308.2) aligned.

Structural annotation

The Just_Annotate_My_Genome (JAMg; https://github.com/genomecuration/JAMg) platform was used to generate putative gene models. First, the genome was masked using RepeatMasker (Smit et al. 2013) and RepeatModeler (Smit et al. 2013). Subsequently, RNA-Seq data was obtained from NCBI for *H. subflexa* and *H. virescens* (SRA accessions: ERR738599, ERR738600, ERR738601, ERR738602, ERR738603, ERR738604, ERR738605, SRR1021613), preprocessed using “justpreprocessmyreads” (http://justpreprocessmyreads.sourceforge.net), and assembled with Trinity RNA-Seq 2.1.1 (Haas et al. 2013) using both the ‘*de-novo*’ and ‘genome-guided’ options as implemented in JAMg. The platform made use of multiple lines of evidence to support each gene model: the two Trinity RNA-Seq assemblies integrated with 63,504 publicly available Sanger-sequenced Expressed Sequence Tags using our new version of PASA (Haas et al. 2003); protein domain annotation of putative exons via HHblits (Remmert et al. 2012); the *de-novo* gene predictors GeneMark.HMM-ET (Lomsadze et al. 2014) and Augustus (Stanke et al. 2006) using the assembled and raw RNA-seq and protein domain data as external evidence. These evidence tracks were condensed to an Official Gene Set (OGS) using Evidence Modeler (Haas et al. 2008).

H. virescens *ddRAD-seq library preparation*

DdRAD-seq libraries were prepared according to Fritz et al. (2016) with minor modifications. Briefly, 200 ng of genomic DNA from the thorax of each field-collected specimen was digested with EcoRI and MspI. Overhang sites from each specimen were ligated to Truseq Universal adapters (Illumina, Inc. San Diego, CA) modified to contain a unique barcode (Elshire *et al.* 2011, Fritz et al. 2016). Adapter-ligated DNA fragments from each individual were combined into pools of no more than 24 individuals. A Pippin Prep (Sage Science, Inc., Beverly, MA) was used to select adapter-ligated DNA fragments ranging from 450-650 bp from each pool, and size-selected DNA pools were amplified in a Peltier PTC200 thermalcycler under the following reaction conditions: 72 °C for 5min, 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 each pool, 1 of 4 Illumina indices (1,2,6, or 12) was added via PCR to the MspI adapter. Amplified pools were combined, cleaned with a Qiaquick PCR Purification Kit (Qiagen, Inc., Valencia, CA, U.S.A.), and diluted to 4nM prior to sequencing. Prepared genomic DNA libraries constructed from a total of 177 *H. virescens* individuals were spread across four 2x300 paired-end Illumina MiSeq runs. Individuals from each year and collection location were spread evenly across each MiSeq run to minimize sequencing run bias in our downstream analysis.

Demultiplexing and Genome Alignment of DdRAD-seq Markers

Illumina-generated read 1 and 2 files were merged using FLASH version 1.2.7 (Magoc and Salzburg 2011), then demultiplexed and filtered for quality using the process_radtags script from Stacks version 1.09 (Catchen et al. 2011, 2013). Quality filtering entailed removal of reads when: 1) they did not have an intact EcoRI cut site, 2) had a quality score < 30, or 3) were smaller than 350 bp. We disabled the rescue reads feature in the process_radtag script, and therefore no read containing errors in the barcode sequence was used for downstream analysis. All remaining merged reads were truncated to a maximum length of 350 bp. Filtered demultiplexed reads were aligned to our *H. virescens* genome assembly using Bowtie 2 (version 2.2.4; Langmead and Salzberg 2012). All reads were aligned in end-to-end mode using the preset parameters with the highest sensitivity (--very-sensitive).

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

We first identified whether any raw ddRAD sequencing reads aligned to the scaffold containing the voltage-gated sodium channel gene using Integrative Genomic Viewer (IGV; Robinson et al. 2011). Following identification of potential ddRAD-seq markers near the voltage-gated sodium channel, we inspected stacks of ddRAD-seq reads for individuals with genotypic data at the pyrethroid resistance locus. Particular attention was paid to individuals that were homozygous for the pyrethroid resistance allele. Through an IGV visual inspection of ddRAD-seq raw reads, we identified one 350bp ddRAD-seq locus (hereafter Hv_11322), for which a single 350bp sequence (hereafter Hv_11322_hap1) was commonly associated with the L1029H mutation at the voltage-gated sodium channel. Filtered, genome-aligned reads from all specimens were then fed into the Stacks v. 1.09 (Catchen et al. 2011, 2013) pipeline for read clustering. Custom R and python scripts were used to call 350bp ddRAD-seq genotypes at Hv_11322 for all field-collected individuals, which were then manually inspected and edited to include any insertions and deletions that were omitted by the Stacks software. For purposes of genotype calling at Hv_11322, individuals with a read count of 6 or higher for a single 350bp sequence were considered homozygotes, with two copies of that allele. Where individuals carried fewer than 6 reads for a single 350bp sequence, their genotypes were scored as a single copy of that observed allele plus one null allele.

We postulated that if the breadth of the "selective sweep" surrounding the voltage-gated sodium channel resistance allele included Hv_11322, such that Hv_11322_hap1 was associated with the L1029H mutation, the rates of their decline in frequency should be similar, if indeed there was a decline. We therefore examined whether the frequencies of the L1029H mutation and Hv_11322_hap1 differed in their rate of decline over time. Specifically, we used a series of nested generalized linear models with binomial error structures to examine whether locus and collection year interacted to influence individual allele (i). In the case of the Hv_11322 response, Hv_11322_hap1 was scored as a 1 and all other alleles were scored as a zero. Our full statistical model was as follows:

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

for $i = 1,\ldots., n$

where Year represented the years during which the moths were collected and Locus indicated either the voltage-gated sodium channel or ddRAD-seq marker Hv_11322. As before, we identified a model term as statistically significant when a comparison of nested models by analysis of deviance indicated that removal of that term significantly influenced model deviance. No significant difference between a model with and without the interaction term might indicate that the slope of the decline in the L1029H mutation was similar to that of Hv_11322_hap1.

We also analyzed the distribution of Hv_11322_hap1 for groups of individuals that were homozygous for either the resistant or susceptible alleles at the voltage-gated sodium channel locus. In total, 32 individuals were homozygous in our target region of the voltage-gated sodium channel gene and contained sufficient ddRAD-seq data at nearby locus Hv_11322 to call at least one allele. Of these 32 individuals, two Hv_11322 alleles could be called for 26 individuals, whereas only a single allele could be called for 6 of the individuals given our previously mentioned criteria. In total, 58 haplotypes (from 32 individuals), which contained genotypic information for both the voltage-gated sodium channel locus and the nearby ddRAD-seq marker were examined. A Fisher's exact test of independence was used to determine whether there was an association between the frequencies of Hv_11322_hap1 and the L1029H mutation.

H. virescens *ddRAD-seq Enabled Genome Scan*

Samtools (version 0.1.18; Li et al. 2009, Li 2011) view was used to convert SAM files output by Bowtie 2 to BAM files, and SNPs were called using mpileup. BCFtools was used to generate SNP and indel genotypes, as well as genotype likelihoods in a Variant Call Formatted (VCF) file. This VCF file was first filtered and then analyzed by VCFtools (version 0.1.15, Auton and Marcketta 2009, https://vcftools.github.io) for allele frequency changes over time. The filtered dataset included loci that: 1) were sequenced to a depth of 3 or more reads, 2) had a minor allele frequency of 0.1 or greater, 3) were represented in at least 50% of individuals, and 4) included only SNP variant sites (indels were excluded). VCFtools was then used to calculate Weir and Cockerham's FST for between-population comparisons on a per variant-site basis, as well as produce a weighted Weir and Cockerham's FST across all loci to examine overall genomic divergence between populations. P-values for the Weir and Cockerham's FST estimates were calculated by likelihood ratio test for allele frequency differences using GPAT++ (available at https://github.com/zeeev/vcflib/wiki/Association-testing-with-GPAT). P-values were adjusted for multiple comparisons using fdrtool (Strimmer 2008) in R.

Scaffolds containing SNP markers that strongly diverged over time (FST > 0.4) were identified for further analysis (Table 1). Predicted structural genes on each scaffold identified in Table 1 were examined for evidence of previous involvement in insecticide resistance. Protein sequences corresponding to each annotation along a scaffold where divergent markers were present were aligned to the NCBI Insecta database (taxid: 50557) via Blastp. Top alignments for each annotation are provided in Supplementary Table 4.

Results

H. virescens *Candidate Gene Analysis*

In 1997, the frequency of the L1029H mutation was 0.66 in LA and 0.63 in TX. By the year 2012, the frequency of this resistance allele declined to 0.44 in LA and 0.36 in TX (Figure 2). This decline in the resistance allele frequency over our 15 year sampling period was statistically significant ($p < 0.001$). Neither the interaction between year and season ($p = 0.36$), season itself ($p = 0.21$), nor sampling location ($p = 0.25$) significantly influenced the frequency of the resistance allele.

Genome Sequencing and Assembly

In total, the Illumina sequencing runs produced 122,433,923 and 232,607,659 reads for PE and MP libraries, respectively. After read trimming and filtering, 115,374,414 and 227,857,423 reads from the PE and MP libraries were used for assembly. An additional 482,464 PacBio reads with an average length of 7560 (s.d. = 2663) bp were applied to our Illumina assembly using PBsuite software for gap filling. The final *H. virescens* genome assembly was comprised of 8826 scaffolds with a total length of 403,154,421 bp, similar to the previously estimated *H. virescens* genome size of 401 Mbp (Gregory and

Herbert 2003). The scaffold N50 was 102,214 bp (mean size = 45,678 bp; range = 659 – 628,964 bp). A CEGMA analysis of our final assembly demonstrated that 186 of the 248 core conserved eukaryotic genes were complete. When we examined our previously mapped *H. virescens* ddRAD-seq markers (Fritz et al. 2016), a total of 562 out of 654 met the aforementioned alignment criteria relative to the reference genome and were used to examine and group scaffolds into chromosomes. Of these 562 markers, 557 aligned uniquely to a single scaffold, while 5 markers (4851, 5891, 13906, 22644, 29612) aligned well to multiple scaffolds. This suggested that either those scaffolds were allelic, or that the marker sequences contain repetitive DNA. Four-hundred eighty three of the 8826 scaffolds present in our assembly were aligned to at least 1 mapped marker. In most cases (n = 421 scaffolds), a single scaffold was associated with a single mapped marker. However, 62 scaffolds could be aligned to multiple mapped markers, which enabled us to check the quality of our assembly against our linkage map. Of these 62 scaffolds, only 5% (3 scaffolds) aligned to markers that originally mapped to different linkage groups. A summary of the scaffold names and groupings by linkage group can be found in Supplementary Table 3. One scaffold, numbered 4600, contained the entire voltage-gated sodium channel gene sequence available from GenBank accession AH006308.2.

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

We located one ddRAD-seq marker, called Hv_11322, spanning bp 11,397 through 11,747 of Scaffold 4600 which is *ca.* 37 kb upstream from the pyrethroid resistance locus. In total, 55 unique Hv_11322 marker sequences could be identified from 138 individuals using Stacks. Fifty of these 55 alleles were found fewer than three times in our field-collected populations. The remaining five most common alleles (Figure 3) were found 5, 5, 6, 12, and 161 times, respectively. According to an NCBI blast, all of these sequences aligned well with an *Helicoverpa armigera* sequence (GenBank accession

DQ458470.1) that also contained the voltage-gated sodium channel gene.

No statistically significant difference existed between the slope of the decline in the L1029H mutation and Hv_11322_hap1 (deviance = -0.355, df = 1, p = 0.551). These allele frequency declines are plotted in Figure 2. When we examined the full haplotypes (e.g. containing both the Hv_11322 locus and the pyrethroid resisance locus), specifically in homozygotes at the pyrethroid resistance locus, we identified 20 unique ddRAD-seq alleles in the 32 total individuals (58 total haplotypes). When haplotypes containing the L1029H mutation were examined, 85% (29 of 34 haplotypes) also carried the ddRAD-seq allele Hv_11322_hap1 (Figure 4). Only 5% (1 of 24) of the haplotypes bearing the wild-type voltage-gated sodium channel allele also carried Hv_11322_hap1. A Fisher's exact test indicated that there was a statistically significant association between the presence of Hv_11322_hap1 and the L1029H mutation (p < 0.001).

H. virescens *ddRAD-seq Enabled Genome Scan*

Of the 1,682,114 SNPs in our ddRAD-seq dataset, the total number of filtered SNPs included in the analyzed dataset was 41,744. Based upon this filtered dataset, overall population genomic divergence between years was low. Weir and Cockerham's weighted FST values were 0.005, 0.001, and 0.004 for the between year comparisons 1997-2007, 1997-2012, and 2007-2012, respectively. We first examined SNPs along Scaffold 4600, where the voltage-gated sodium channel gene was located, for evidence of genomic divergence between years. Between the years 1997 and 2012, two SNPs at positions 11655 and 11706 on Scaffold 4600 had Weir and Cockerham FST values of 0.2042 and 0.2101, respectively. The probabilities that genomic divergence at SNPs 11655 and 11706 were due to random chance were low ($p = 6.42 \times 10^{-4}$ and 1.97×10^{-5}, respectively). When a false discovery rate correction was applied, the SNP at 11706 remained significantly diverged in *H. virescens* populations between the years 1997 and 2012 with a corrected p-value of 0.01. These same SNPs were not significantly diverged between the years 1997 and 2007, or between the years 2007 and 2012.

Among the three pairwise comparisons, 541 scaffolds (6%) contained SNPs for which allele frequencies had significantly diverged over time. When broken down by pairwise comparison, the years 2007 to 2012 had the greatest number of scaffolds with diverging SNPs (n = 369 unique scaffolds), followed by 1997 to 2007 (n = 220) and 1997 to 2012 (n = 52). Of these scaffolds, only 5 contained multiple SNPs with high FST values (FST > 0.4) in at least one by-year comparison. These scaffolds were 167, 688, 3242, 3424, and 8088 (Table 1). Examination of the nearby annotated sequences revealed that two of these scaffolds contained genes involved in pyrethroid resistance in other species (Na et al. 2007, Yang et al. 2008, Hou et al. 2014). Predicted gene sequences evm.model.Contig167.9, evm.model.Contig167.10, and evm.model.Contig167.11 were found between bp 155,728 and 179,640 of Scaffold 167. The protein sequences of all three aligned well to protein sequences in the Trypsin-like serine protease superfamily (e.g. *Danaus plexippus, Operophtera brumata,* and *Papillio xuthus* sequences with GenBank Accession numbers EHJ76740.1, XP_013200191.1, and KPI91088.1, respectively) with > 60% identity, query covers of > 85%, and e-values lower than 6.43×10^{-67}. Another predicted gene sequence, evm.model.Contig3424.3, aligned with a cytochrome p450 protein sequence (*CYP6AE12*) from *Helicoverpa armigera* (GenBank Accession AID54888.1) with a 100% query cover and 83% identity.

Discussion

Double-digest RAD-seq and other NGS marker-development methods have been used to detect signatures of local adaptation in a number of non-model plant and animal species (e.g. Hohenlohe et al. 2010, Nadeau et al. 2013, Pujolar et al. 2014, Ruegg et al. 2014, Pais et al. 2016). Here we demonstrated the power of ddRAD-seq to identify genomic regions that have diverged over short evolutionary time scales in a landscape characterized by human-induced environmental change. We postulated that widespread adoption and cultivation of Bt cotton in the Southern United States would likely impose strong selection on Lepidopteran herbivore and cotton pest, *H. virescens*, through shifts in host plant composition and insecticide use. In a proof of concept experiment, we first identified allele frequency changes at a likely gene target of selection, the pyrethroid resistance locus, in field-collected populations of *H. virescens*. We then demonstrated that this change could be detected using a nearby ddRAD-seq marker. Allele frequencies at several other regions of the *H. virescens* genome also diverged over time, likely in response to selection pressures imposed by widespread adoption of Bt cotton. Furthermore, we sequenced and assembled the first *H. virescens* draft genome to help us identify potential structural genes involved in adaptation to agricultural inputs, and made it publicly available at (LINK TO GENOME COMING SOON).

Our initial examination of the voltage-gated sodium channel gene, a candidate gene likely to be impacted by the decline in pyrethroid use that followed Bt cotton adoption, demonstrated that the resistance-conferring L1029H mutation declined in frequency over time. This seemed reasonable given the fitness cost associated with carrying this resistance allele (Zhao et al. 2000). However, the frequency of the resistance allele plateaued in the year 2007 and remained at *ca.* 0.4 through the year 2012. There are several possible explanations for this. One explanation is that pyrethroid pressure has declined but remains sufficiently high such that maintenance of the resistance allele in field populations is advantageous, in spite of the fitness cost to individuals that carry it (Zhao et al. 2000). Alternatively, as the resistance allele frequency declines there are relatively fewer homozyogous resistant genotypes. If the fitness cost is only associated with homozygotes then the decline in resistance allele frequency could level off, even in the absence of pyrethroid selection.

Using a ddRAD-seq dataset, we identified one marker that aligned to our reference genome 37 kb upstream of the voltage-gated sodium channel gene. One allele of this 350bp marker, called Hv_11322_hap1, was strongly associated with the L1029H mutation that confers pyrethroid resistance. This suggests that Hv_11322_hap1 is in linkage disequilibrium with the L1029H mutation. Furthermore, the breadth of the selective sweep in this genomic region extends at least 37 kb on one side of the voltage-gated sodium channel gene. Upon further examination of Scaffold 4600, which contains this region under selection, we identified three cytochrome p450s that are found between Hv_11322 and the voltage-gated sodium channel gene. This confirmed previous reports of tight linkage between the voltage-gated sodium channel gene and *CYP6B10* in *H. virescens* (Park and Brown 2002). It is possible that these cytochrome p450s could also be targets of selection by pyrethroid insecticides, and future work could be directed at whether or not they play any roles in the expression of pyrethroid resistance phenotypes. Work in other closely-related Lepidopteran species suggests that cytochrome p450s linked to the voltage-gated sodium channel are not involved in pyrethroid resistance, however (Grubor et al. 2007).

Single nucleotide polymorphism data from our ddRAD-seq marker Hv_11322 enabled us to rediscover changes at the voltage-gated sodium channel gene associated with the L1029H mutation over time. To our knowledge, ours is the first demonstration of the utility of ddRAD-sequencing to detect genomic changes associated with insecticide use over short time scales in an insect species. While the SNP outlier in the Hv_11322 marker demonstrated significant allelic divergence relative to the genome-wide average FST value, SNP outliers from ddRAD-seq markers on other *H. virescens* scaffolds showed much greater allelic divergence over the 15 year period. It is likely that markers on each of these 5 scaffolds are in linkage disequilibrium with gene targets of selection as host plant composition, or more specifically management of cotton ecosystems has led to the replacement of conventional cotton cultivars with Bt-expressing varieties. Interestingly, two of these scaffolds contain genes that are previously known to be associated with insectide resistance in other species.

Based upon our structural annotation, Scaffold 167 contains genes that are homologous to the trypsin-like serine protease superfamily. Genes in this family have been previously involved in Bt resistance in other Lepidopteran species (Reviewed in Oppert 1999, Rodriguez-Cabrera et al. 2010). Perhaps these genes contribute to low-level Bt resistance in *H. virescens* as well. Allelic divergence at SNPs on this scaffold was detected in *H. virescens* populations collected during and after the year 2007, when dual Bt cotton cultivars became commercially available. While this was somewhat surprising due to the lack of observed Bt resistance in field-collected *H. virescens* (Tabashnik et al. 2013), it is possible that changes in the composition of Bt cultivars in the landscape led to changes in selection pressure at these target genes.

Scaffold 3424 also contained SNPs that diverged significantly over time in our field-collected populations of *H. virescens*. Divergence was strongest in by-year comparisons from 1997-2007, and 1997-2012. This suggests that most genomic change occurred between the years 1997 and 2007, and that allele frequencies remained stable between the years 2007 and 2012. Blast results for predicted gene sequences found on this scaffold revealed homology with the cytochrome p450 superfamily. The predicted sequence aligned well with an *H. armigera CYP6A* gene, which is a cytochrome p450 family known to be involved in detoxification (Zhou et al. 2010). The best alignment was to an H. armigera *CYP6AE12*. Expression levels of this gene in *H. armigera* are modified in response to pyrethroid insecticides (Li-Na et al. 2007, Zhou et al. 2010). It is possible that allelic changes on this scaffold are a response to reduced pyrethroid use in the Southern United States as a result of Bt cotton deployment in the agricultural landscape.

Alternative explanations exist for both of the candidate genes on these two scaffolds, however.

In the case of Scaffold 167, serine proteases are also known to be generally involved in host plant detoxification and digestion (Vogel et al. 2014). In the case of Scaffold 3424, *CYP6AE12* expression is also modified in response to the plant compound xanthotoxin (Zhou et al. 2010). If the trypsin-like serine proteases and the cytochrome p450 are indeed the targets of selection, it is possible that the divergence in allele frequencies between the years 1997 and 2012 could be the result of selection for an *H. virescens* population that feeds on alternative, wild host plants. Perhaps widespread planting of Bt-expressing cotton drove *H. virescens* off of their previously abundant cotton host and back to alternative wild host plants, some of which may naturally produce antifeedant chemicals. Allelic changes in these genomic regions could be caused by *H. virescens* adaptation to alternative wild host plants, rather than adaptation to changes in pyrethroid use or Bt toxins. To determine whether phenotypes resulting from these molecular shifts are directly associated with changing *H. virescens* management practices (e.g. pyrethoid or Bt toxin use), further work could involve measuring associations between Bt or pyrethoid response phenotypes and genotypes at target genes on these scaffolds.

The *H. virescens* draft assembly that we produced was instrumental in identifying these potential gene targets of selection, in spite of it's imperfections. For example, two scaffolds contained markers from our *H. virescens* linkage map that originally were found on different linkage groups (LGs). Scaffold 121 aligned to markers found in both LGs 22 and 25, and Scaffold 209 aligned to markers found in both LGs 13 and 30. Future work to improve the contiguity, completeness and correctness of our *H. virescens* assembly will help us to identify additional gene targets of selection of the 3 other contigs, where gene targets could not currently be identified.

In conclusion, we conducted a proof-of-concept experiment to demonstrate that ddRAD-seq enabled genomic scanning can be used to identify organismal responses to anthropogenic changes in agricultural ecosystems. We used a ddRAD-seq dataset to rediscover *H. virescens* genes known to respond to cotton management practices in the Southern United States, and identified additional genomic regions in this Lepidopteran species that are likely changing in response to shifts from conventional cotton planting to widespread Bt-cotton adoption. Our results suggest that ddRAD-seq genome scans may be useful for monitoring pest populations for real-time changes in allele frequencies at loci responding to management practices. This could be useful for identifying resistance alleles, and then acting to mitigate widespread phenotypic resistance to management practices across plant and insect species in agricultural ecosystems.

Data Availability

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

Scripts used for population genomic analysis can be found at: COMING SOON

Raw sequence data have been deposited in the NCBI SRA as: COMING SOON

Our *H. virescens* draft 1 assembly can be found at: COMING SOON

Acknowledgements

Thanks to Dr. J. Schaff and Dr. D. Baltzegar of the North Carolina State University Genomic Sciences Lab for their insightful suggestions on ways to improve our methods. We also thank Dr. Keith Hopper for early discussions of our Gabrielle Beaudry, Emma Thompson, Kelsey Mckinney, Wilfred Wong, Sharyar Samir, and Xuechun Wang isolated gDNA from the moths used in this project. Nico Olegario assisted with the L1029H genotyping. Mr. S. Micinski, Dr. J. Lopez, and Dr. J. Westbrook collected the moths used in this project. R. Waples wrote one of the custom scripts used in our data pipeline. This project was supported by the Biotechnology Risk Assessment Program competitive grant number 2012-33522-19793 from the USDA - National Institute of Food and Agriculture.

References

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81-92.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS One*, **3**, e3376.

BBMap - Bushnell B. - sourceforge.net/projects/bbmap/

Benbrook CM (2012) Impacts of genetically engineered crops on pesticide use in the U.S. – the first sixteen years. *Environmental Sciences Europe*, **24**, 24.

Blanco CA (2012) *Heliothis virescens* and Bt cotton in the United States. *GM Crops & Food: Biotechnology in Agriculture and the Food Chain*, **3**, 201-212.

Bolger, A. M., Lohse, M., and Usadel, B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Brito LP, Linss JGB, Lima-Camara TN, Belinato TA, Peixoto AA, Lima JBP, Valle D, Martins AJ (2013) Assessing the effects of Aedes aegypit kdr mutations on pyrethroid resistance and its fitness cost. *PloS ONE*, **8**, e60878.

Campanhola C and Plapp FW (1989) Pyrethroid resistance in the tobacco budworm (Lepidoptera: Noctuidae): insecticide bioassays and field monitoring. *J Econ Entomol*, **82**, 22-28.

Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J, De Koning, D (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes Genomes Genetics*, **1**, 171-182.

Catchen J, Hohenlohe P, Bassham S, Amores A, and Cresko W (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124-3140.

Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics*. Roberts and Company Publishers, Greenwood Village, Colorado, USA.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K., Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS ONE*, **6**, e19379.

English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PloS One*, **7**, e47768.

Foster SP, Denholm I, Thompson R, Poppy GM, Powell W (2005) Reduced response of insecticide-resistant aphids and attraction of parasitoids to aphid alarm pheromone; a potential fitness trade-off. *Bulletin of Entomological Research*, **95**, 37–46.

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

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

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

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

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

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

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

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

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

Hou MZ, Shen GM, Wei D, Li YL, Dou W, Wang JJ (2014) Characterization of *Bactrocera dorsalis* Serine Proteases and Evidence for Their Indirect Role in Insecticide Tolerance. *International Journal of Molecular* Sciences, **15,** 3272–3286.

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

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

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

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.

Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, **42**, e119.

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 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, Yang H, Wang J, Lam T, Wang, J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, **1**, 18.

Luttrell RG, Roush RT, Ali A, Mink JS, Reid MR (1987) Pyrethroid resistance in field populations of Heliothis virescens (Lepidoptera: Noctuidae) in Mississippi in 1986. *J Econ Entomol*, **80**, 985-989.

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.

Magoc T, Salzberg S (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**, 2957-2963.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.

Nadeau NJ, Ruiz M, Salazar P, Counterman B, Medina JA, Ortiz-Zuazaga H, Morrison A, McMillan WO, Jiggins CD, Papa R (2014) Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res*, **24**, 1316-1333.

NASEM, National Academies of Sciences, Engineering, and Medicine (2016) *Genetically Engineered Crops: Experiences and Prospects.* Washington, DC: The National Academies Press.

Nielsen R (2005) Molecular signatures of natural selection. *Annu. Rev. Gen.*, **39**, 97-218.

Onstad, DW (2014) Insect Resistance Management (Second Edition) Biology, Economics, and Prediction. Elsevier Ltd. ISBN: 978-0-12-396955-2

Oppert B (1999) Protease interactions with *Bacillus thuringiensis* insecticidal toxins. *Archives of Insect Biochemistry and Physiology*, **42**, 1-12.

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

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

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

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

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

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, **23**, 1061-1067.

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

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

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

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

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

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

Rodríguez–Cabrera L, Trujillo–Bacallao D, Borrás–Hidalgo O, Wright DJ, Ayra–Pardo, C (2010) RNAi-mediated knockdown of a Spodoptera frugiperda trypsin-like serine-protease gene reduces susceptibility to a Bacillus thuringiensis Cry1Ca1 protoxin. *Environmental microbiology*, **12**, 2894-2903.

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Mol Ecol*, **23**, 4757-4769.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative Genomics Viewer. *Nature Biotechnology* **29**, 24–26.

Smit AFA, Hubley R, Green P (2013) *RepeatMasker Open-4.0*.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B (2006) AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Research*, **34**.

Strimmer K (2008) fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics Applications*, **24**, 1461-1462.

Tabashnik BE, Brevault T, Carriere Y (2013) Insect resistance to Bt crops: lessons from the first billion acres. Nature Biotechnology, **31**, 510-521.

Taylor M, Shen Y, Kreitman M (1995) A population genetic test of selection at the molecular level. *Science*, **270**, 1497-1499.

The Variant Call Format and VCFtools, Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert Handsaker, Gerton Lunter, Gabor Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin and 1000 Genomes Project Analysis Group, Bioinformatics, 2011

Tilman D, Fargione J, Wolff B, D'Antonio C, Dobson A, Howarth R, Schindler D, Schelesinger WH, Simberloff D, Swackhamer D (2001). Forecasting agriculturally driven global environmental change.

Science, **292**, 281-284.

[USEPA] U.S. Environmental Protection Agency (1998) The environmental protection agency's white paper on Bt plant-pesticide resistance management. Washington: http://www.epa.gov/scipoly/sap/meetings/1998/february/finalfeb.pdf

[USEPA] U.S. Environmental Protection Agency (2001) Biopesticides registration action document: *Bacillus thuringiensis* plant-incorporated protectants. http://www.epa.gov/pesticides/biopesticides/pips/bt_brad.htm

[USEPA] U.S. Environmental Protection Agency (2006) Analysis of a Natural Refuge of Non-Cotton Hosts for Monsanto's Bollgard II Cotton. FIFRA Scientific Advisory Panel Meeting June 13-15, 2006. Arlington, Virginia. 101pgs.

Vogel H, Musser RO, Celorio-Mancera M (2014). Transcriptome responses in herbivorous insects towards host plant and toxin feeding. *Annual Plant Reviews*, **47**, 197-234.

Yang Q, Zhou D, Sun L, Zhang D, Qian J, Xiong C, Sun Y, Ma L, Zhu C (2008) Expression and characterization of two pesticide resistance-associated serine protease genes (NYD-tr and NYD-ch) from *Culex pipiens* pallens for metabolism of deltamethrin. *Parasitol Res*, **103**, 507-516.

Yue LN, Yang YH, Wu SW, WU YD (2007) Cloning and mRNA expression levels of cytochrome P450 genes *CYP6AE12*and *CYP9A18* in the cotton bollworm, *Helicoverpa armigera*. *ACTA Enotomologica Sinica*, **50**, 234-240.

Zhao Y, Park Y, and Adam ME (2000). Functional and evolutionary consequences of pyrethroid resistance mutations in S6 transmembrane segments of a voltage-gated sodium channel. Biochemical and Biophysical Research Communications 278: 516-521.

Zhou X, Sheng C, Li M, Wan H, Liu D, Qui X (2010). Expression responses of nine cytochrome p450 genes to xenobiotics in the cotton bollworm *Helicoverpa armigera*. Pesticide Biochemistry and Physiology 97:209-213.

Tables

Table 1 – List of *H. virescens* contigs containing multiple SNPs with high, statistically significant pairwise FST values between years. Not all by-year comparisons contained SNPs with high FST values, as indicated in the final table column. The number of annotations present represents the number of nucleotide sequences identified during the automated annotation process on the contig, including genes, mRNA, peptide, and repetitive sequences. Two contigs contained annotated peptide sequences were orthologous to previously implicated as insecticide resistance genes.

Scaffold Number	Scaffold Length (bp)	Number Annotations Present	Predicted Insecticide Resistance Genes Present	Significant Pairwise Comparisons
167	273,034	13	Trypsin-like serine protease superfamily	2007-2012
688	215,790	7	-	1997-2012 1997-2007
3242	41,624	3	-	1997-2012 1997-2007
3424	78,310	3	Cytochrome p450	1997-2012 2007-2012
8088	47,238	0	-	1997-2012 2007-2012

Figures

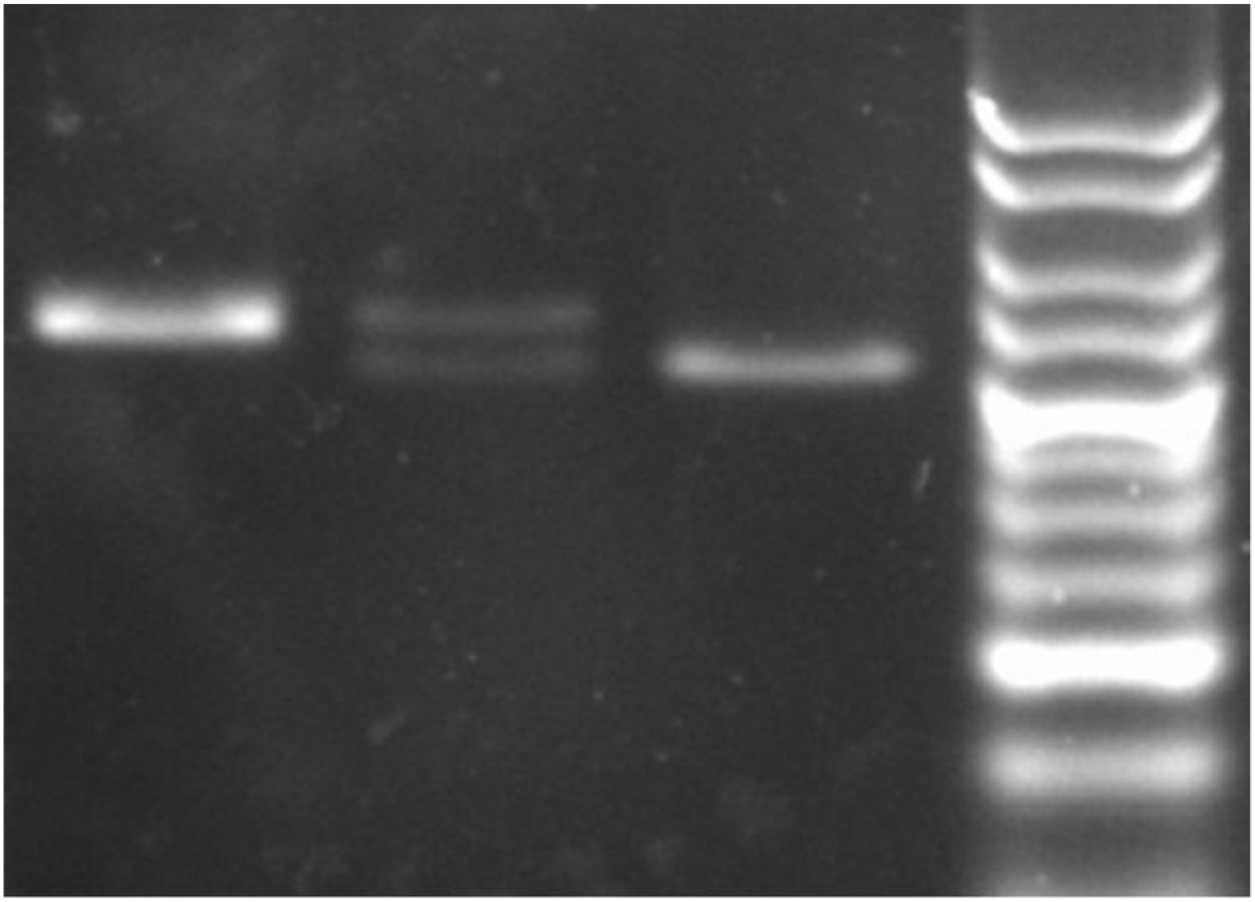


Figure 1 - Three *H. virescens* individuals from Louisiana in 2012 were genotyped using Sanger sequencing. Amplicons from these same samples were digested with NlaIII and electrophoresed on a 3.5% agarose gel alongside Hyperladder V. From left to right, the genotypes as viewed on the agarose gel confirmed the Sanger genotypes, TT, AT, and AA.

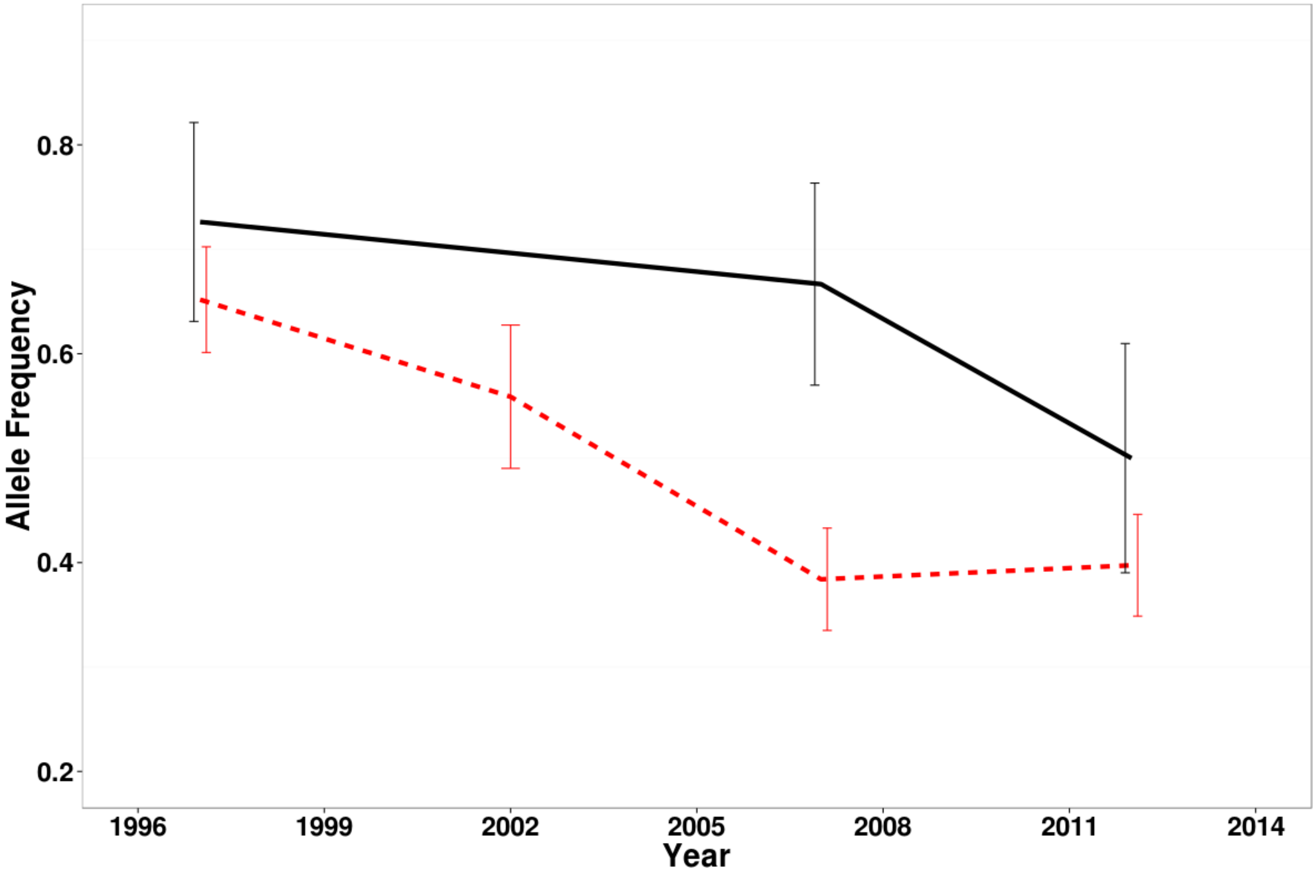
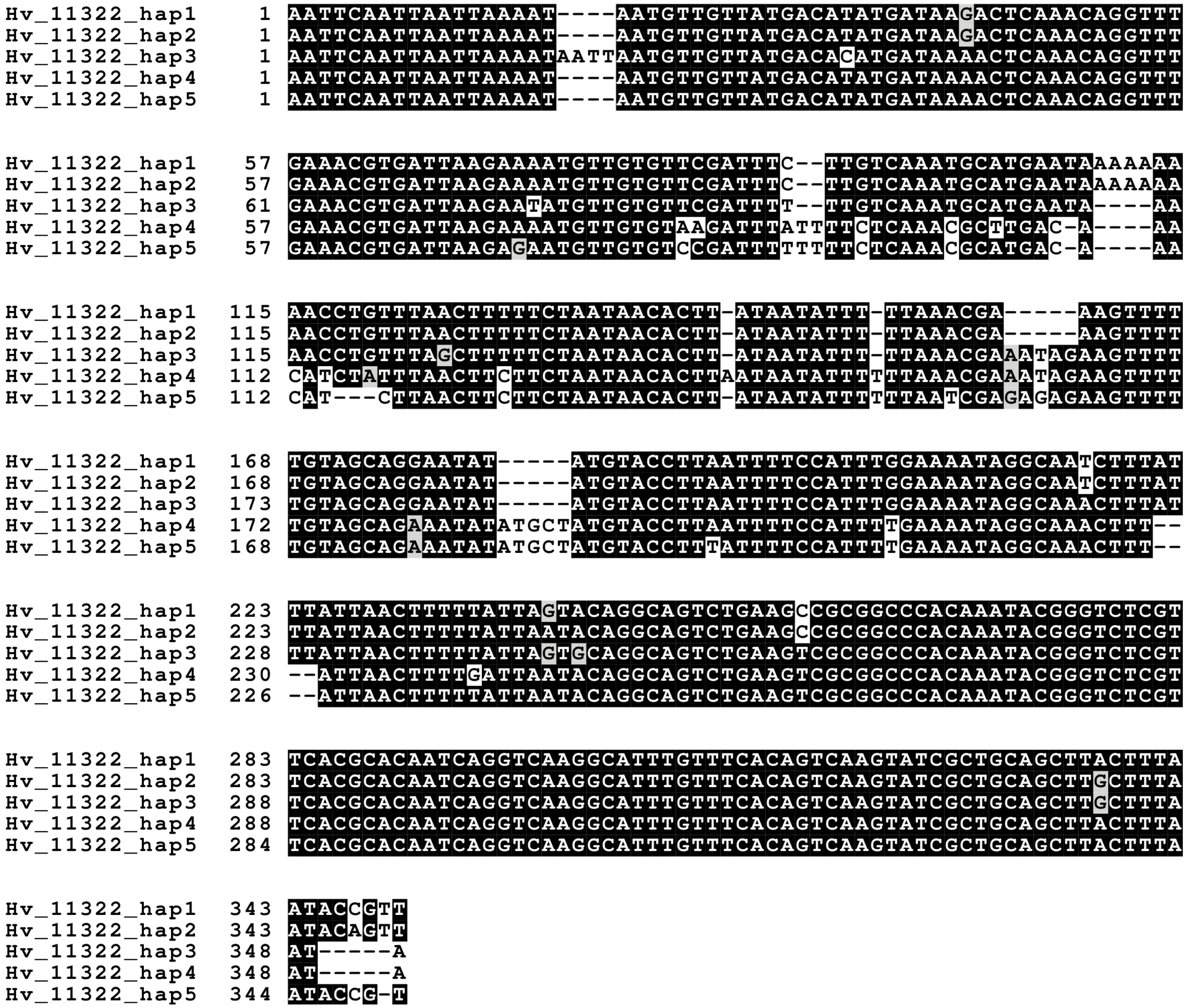


Figure 2 – The decline in the frequency of the pyrethroid resistance allele in *H. virescens* (pooled LA and TX samples), represented by the solid black line, was statistically significant over the course of our 15 year sampling period (n = 659). A unique ddRAD-seq haplotype, represented by the dashed red line, was found *ca.* 37Kb upstream from the alpha subunit of the voltage-gated sodium channel gene and also declined in frequency in the subset of individuals (n = 141) sequenced for our genome scan. Error bars represent bootstrapped 95% confidence intervals (N = 5000) around the mean of each year.



```
Hv_11322_hap1     1 AATTCAATTAATTAAAAT----AATGTTGTTATGACATATGATAAGACTCAAACAGGTTT
Hv_11322_hap2     1 AATTCAATTAATTAAAAT----AATGTTGTTATGACATATGATAAGACTCAAACAGGTTT
Hv_11322_hap3     1 AATTCAATTAATTAAAATAATTAATGTTGTTATGACACATGATAAAACTCAAACAGGTTT
Hv_11322_hap4     1 AATTCAATTAATTAAAAT----AATGTTGTTATGACATATGATAAAACTCAAACAGGTTT
Hv_11322_hap5     1 AATTCAATTAATTAAAAT----AATGTTGTTATGACATATGATAAAACTCAAACAGGTTT

Hv_11322_hap1    57 GAAACGTGATTAAGAAAATGTTGTGTTCGATTTC--TTGTCAAATGCATGAATAAAAAAA
Hv_11322_hap2    57 GAAACGTGATTAAGAAAATGTTGTGTTCGATTTC--TTGTCAAATGCATGAATAAAAAAA
Hv_11322_hap3    61 GAAACGTGATTAAGAATATGTTGTGTTCGATTTT--TTGTCAAATGCATGAATA----AA
Hv_11322_hap4    57 GAAACGTGATTAAGAAAATGTTGTGTAAGATTTATTTTCTCAAACGCTTGAC-A----AA
Hv_11322_hap5    57 GAAACGTGATTAAGAGAATGTTGTGTCCGATTTTTTTTCTCAAACGCATGAC-A----AA

Hv_11322_hap1   115 AACCTGTTTAACTTTTTCTAATAACACTT-ATAATATTT-TTAAACGA-----AAGTTTT
Hv_11322_hap2   115 AACCTGTTTAACTTTTTCTAATAACACTT-ATAATATTT-TTAAACGA-----AAGTTTT
Hv_11322_hap3   115 AACCTGTTTAGCTTTTTCTAATAACACTT-ATAATATTT-TTAAACGAAATAGAAGTTTT
Hv_11322_hap4   112 CATCTATTTAACTTCTTCTAATAACACTTAATAATATTTTTTAAACGAAATAGAAGTTTT
Hv_11322_hap5   112 CAT---CTTAACTTCTTCTAATAACACTT-ATAATATTTTTTAATCGAGAGAGAAGTTTT

Hv_11322_hap1   168 TGTAGCAGGAATAT-----ATGTACCTTAATTTTCCATTTGGAAAATAGGCAATCTTTAT
Hv_11322_hap2   168 TGTAGCAGGAATAT-----ATGTACCTTAATTTTCCATTTGGAAAATAGGCAATCTTTAT
Hv_11322_hap3   173 TGTAGCAGGAATAT-----ATGTACCTTAATTTTCCATTTGGAAAATAGGCAAACTTTAT
Hv_11322_hap4   172 TGTAGCAGAAATATATGCTATGTACCTTAATTTTCCATTTTGAAAATAGGCAAACTTT--
Hv_11322_hap5   168 TGTAGCAGAAATATATGCTATGTACCTTTATTTTCCATTTTGAAAATAGGCAAACTTT--

Hv_11322_hap1   223 TTATTAACTTTTTATTAGTACAGGCAGTCTGAAGCCGCGGCCCACAAATACGGGTCTCGT
Hv_11322_hap2   223 TTATTAACTTTTTATTAATACAGGCAGTCTGAAGCCGCGGCCCACAAATACGGGTCTCGT
Hv_11322_hap3   228 TTATTAACTTTTTATTAGTGCAGGCAGTCTGAAGTCGCGGCCCACAAATACGGGTCTCGT
Hv_11322_hap4   230 --ATTAACTTTTGATTAATACAGGCAGTCTGAAGTCGCGGCCCACAAATACGGGTCTCGT
Hv_11322_hap5   226 --ATTAACTTTTTATTAATACAGGCAGTCTGAAGTCGCGGCCCACAAATACGGGTCTCGT

Hv_11322_hap1   283 TCACGCACAATCAGGTCAAGGCATTTGTTTCACAGTCAAGTATCGCTGCAGCTTACTTTA
Hv_11322_hap2   283 TCACGCACAATCAGGTCAAGGCATTTGTTTCACAGTCAAGTATCGCTGCAGCTTGCTTTA
Hv_11322_hap3   288 TCACGCACAATCAGGTCAAGGCATTTGTTTCACAGTCAAGTATCGCTGCAGCTTGCTTTA
Hv_11322_hap4   288 TCACGCACAATCAGGTCAAGGCATTTGTTTCACAGTCAAGTATCGCTGCAGCTTACTTTA
Hv_11322_hap5   284 TCACGCACAATCAGGTCAAGGCATTTGTTTCACAGTCAAGTATCGCTGCAGCTTACTTTA

Hv_11322_hap1   343 ATACCGTT
Hv_11322_hap2   343 ATACAGTT
Hv_11322_hap3   348 AT------A
Hv_11322_hap4   348 AT------A
Hv_11322_hap5   344 ATACCG-T
```

Figure 3 – Multiple sequence alignment for the top 5 ddRAD-seq haplotypes found upstream from the *H. virescens* pyrethroid resistance locus. Hv_11322_hap1 appeared to be in strong linkage disequilibrium with voltage-gated sodium channel mutation L1029H.

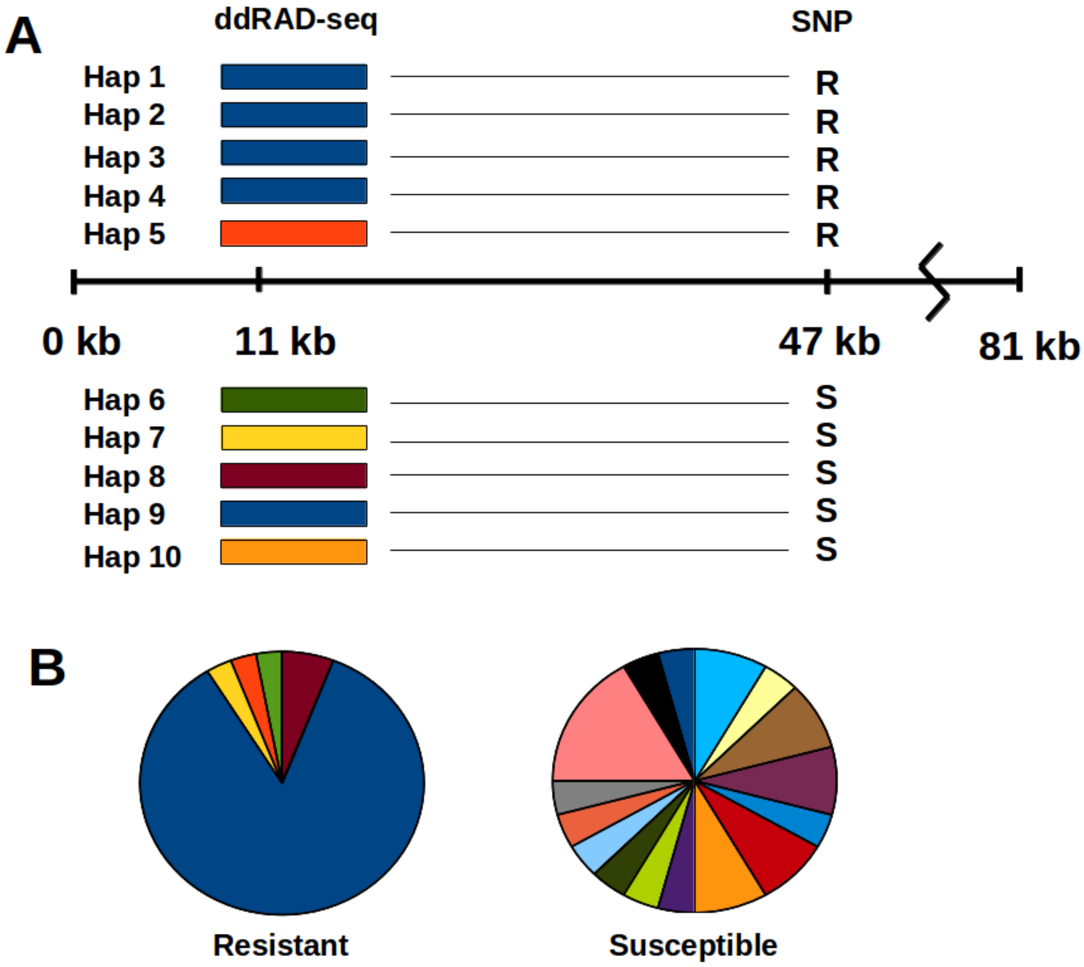


Figure 4 – Significantly greater diversity was observed from ddRAD-seq haplotypes linked to the susceptible voltage-gated sodium channel allele relative to those linked to the resistance allele. (A) depicts the relationship of the ddRAD-seq marker to the L1029H voltage-gated sodium channel SNP along an 81kb genome scaffold. The different colored bars at 11kb represent unique alleles at the ddRAD-seq locus. (B) The unique colors in the pie charts depict the number and proportion of ddRAD-seq alleles linked to the voltage-gated sodium channel resistant (n = 34) and susceptible (n =24) SNP alleles. The dark blue wedges always represent Hv_11322_hap 1.

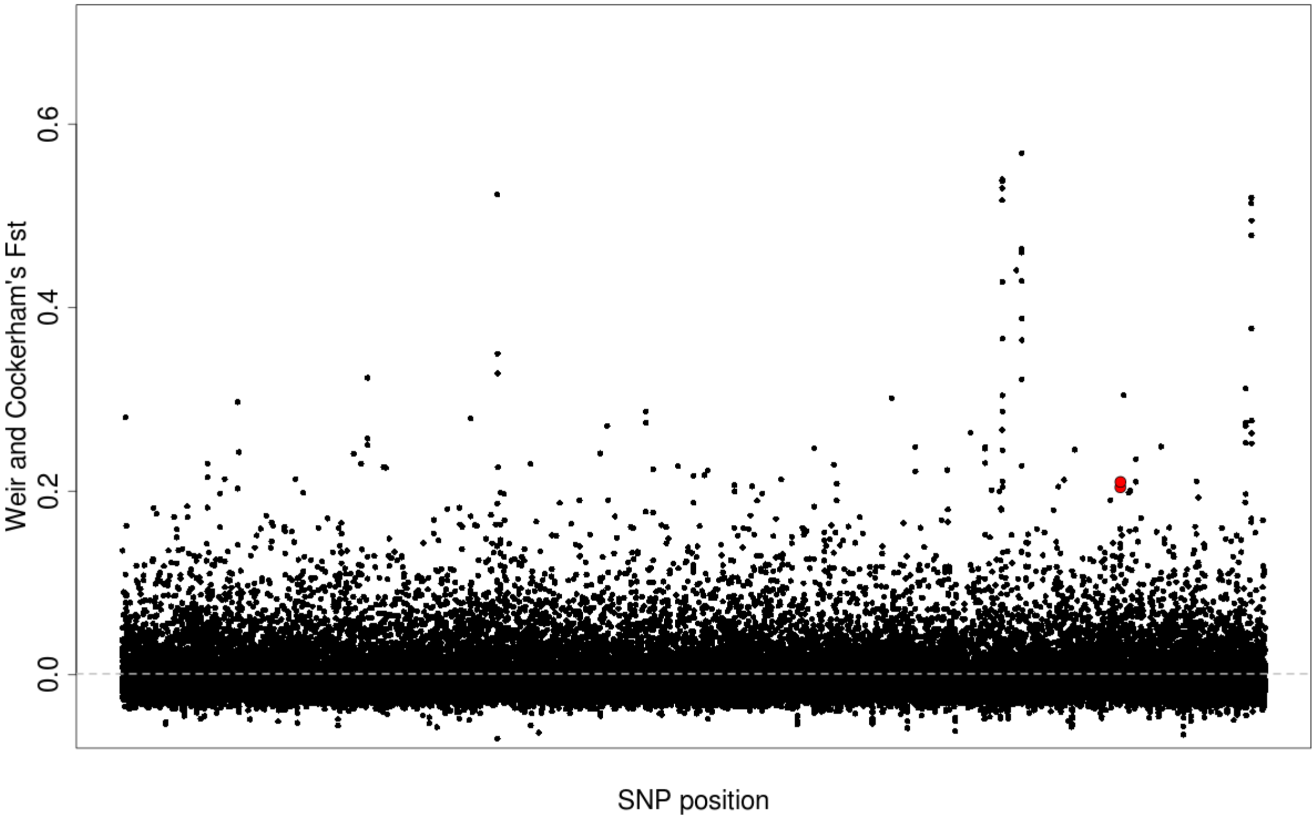


Figure 5 – Genetic divergence according to Weir and Cockerham's FST between populations collected in 1997 and 2012. Each black point represents one SNP of 41742 along the *H. virescens* genome. Two additional SNPs, 11655 and 11706 on Contig 4600, are represented in red and are physically linked to the point mutation at the *H. virescens* voltage-gated sodium channel gene. Genetic divergence at one of these SNPs (11706) remains statistically significant following a correction for false discovery rate (p = 1.97 x 10 -5; corrrected-p = 0.01).