Application of a dense genetic map for assessment of genomic responses to selection and inbreeding in *Heliothis virescens.*

Megan L. Fritz¹,²,§, Sandra Paa¹, Jennifer Baltzegar², and Fred Gould¹,²

¹Department of Entomology, North Carolina State University, Raleigh, NC, USA
²Department of Biological Sciences, Program in Genetics, North Carolina State University, Raleigh, NC, USA
§Corresponding Author
Abstract

Adaptation of pest species to laboratory conditions and selection for resistance to toxins in the laboratory are expected to cause inbreeding and genetic bottlenecks that reduce genetic variation. *Heliothis virescens*, a major cotton pest, has been colonized in the laboratory many times, and a few laboratory colonies have been selected for Bt resistance. We developed 350 bp Double-Digest Restriction-site Associated DNA-sequencing (ddRAD-seq) molecular markers to examine and compare changes in genetic variation associated with laboratory adaptation, artificial selection, and inbreeding in this non-model insect species. We found that allelic and nucleotide diversity declined dramatically in laboratory-reared *H. virescens* as compared with field-collected populations. The declines were primarily due to the loss of low frequency alleles present in field-collected *H. virescens*. A further, albeit modest decline in genetic diversity was observed in a Bt-selected population. The greatest decline was seen in *H. virescens* that were sib-mated for 10 generations, where more than 80% of loci were fixed for a single allele. To determine which regions of the genome were resistant to fixation in our sib-mated and Bt-selected lines, we generated a dense intraspecific linkage map containing 3 PCR-based, and 659 ddRAD-seq markers. Markers that retained polymorphism were observed in small clusters spread over multiple linkage groups. These markers are likely associated with genomic regions under balancing selection, thus preventing fixation of deleterious alleles.

Keywords: *Heliothis virescens*, genetic variation, colonization, inbreeding, ddRAD-seq, linkage map
Introduction

Laboratory-reared insect colonies are important resources for many types of entomological experiments. They are used to quantify physiological or behavioral differences between insect populations or species (Dekker et al., 2006; Dobzhansky & Spassky, 1954; Fritz et al., 2015; Groot et al., 2005; Shaw et al., 2000; Sokolowski, 1980; Tomaru et al., 2000), identify the genetic architecture of insect traits (Gahan et al., 2010, Mackay et al., 2012, Oppenheim et al., 2012), develop insect populations that express desirable traits (Collins, 1984; Goldman et al., 1986; Gould et al., 1995; Hoy, 1989; Pradeep et al., 2005), and generate genetically modified species as a means of pest control (de Valdez et al., 2011). A major concern for researchers maintaining insect colonies is the degree to which adaptation to the laboratory environment affects insect genotypic, and thereby phenotypic diversity (Boller, 1972; Huettel, 1976).

The phenotypic consequences of adaptation to the laboratory depend upon the trait of interest, and range from undetectable to severe (Baeshen et al., 2014; Fox et al., 2007; Gerloff et al., 2003; Raulston, 1975; Roush, 1986). Observed phenotypic changes can be attributed to inadvertent selection for traits that are favorable in the laboratory environment (Roush, 1986), inbreeding depression (i.e. reduction in fitness caused by matings between related individuals; reviewed in Charlesworth & Willis, 2009; Mackauer, 1976), or the interaction of the two. Indeed, the selection that occurs during colony establishment creates conditions conducive to inbreeding (Roush, 1986). Families with higher fitness under laboratory conditions contribute disproportionately to the reproductive pool, thereby increasing the probability of matings between related individuals. Where selection is very strong, as in the production of an insecticide resistant colony, measures must often be taken to minimize the effects of inbreeding and thereby inbreeding depression (Gould et al., 1995). Overall, the expectation is that the selection and inbreeding that takes place during insect colonization results in an overall loss of genetic
diversity (Munstermann, 1994), and concomitant genome-wide increase in homozygosity (reviewed in
Etzel & Legner, 1999).

Previous studies that have examined genetic differences between field-collected, laboratory-
adapted, and inbred populations of non-model insects have primarily focused on Dipteran species and
were limited to small numbers of molecular markers (Mukhopadhyay et al., 1997; Munstermann, 1994;
Norris et al., 2001). Such small numbers of markers allow for estimation of the genome-wide average
change in genetic variability across populations, but cannot be used to examine fine-scale patterns of
genomic change. Examination of these patterns allows for identification of where and how genetic
variation, the raw material necessary for environmental adaptation, is maintained (Dobzhansky &
Spassky, 1954). The relatively recent development of high-throughput sequencing combined with
reduced-representation DNA library preparation techniques allows for the discovery of hundreds to
thousands of new molecular markers, even in species for which genomic data are absent (Davey et al.,
2011). Here we used Double-Digest Restriction-Site Associated DNA Sequencing (ddRAD-seq;
Peterson et al., 2012), one type of reduced representation library, for de novo construction of molecular
markers in the non-model species, Heliothis virescens.

The tobacco budworm, H. virescens, is an historically important pest of cotton throughout much
of the Southeastern United States (Blanco, 2012). This non-model Lepidopteran species has been
colonized a number of times for investigations of mating and host-selection behaviors (Sheck & Gould,
1995; Sheck & Gould, 1996; Sheck et al., 2006), as well as detecting the underlying genetic basis for
insecticide resistance (Gahan et al., 2001; Gahan et al., 2010; Taylor et al., 1993). We used our newly
developed ddRAD-seq markers to examine and compare the effects of colonization, selection, and sib-
mating on H. virescens genome-wide measures of genetic diversity. To examine fine-scale patterns of
change in genetic diversity, we also used our ddRAD-seq markers to generate a dense intraspecific
genetic map for *H. virescens*. This map consists of 659 high quality 350-bp markers which will serve as an important genomic resource to the entomological community.

Overall, our research aims to:

1. Quantify overall patterns of change in genomic diversity across field-collected, laboratory-reared (non-selected), Bt-selected, and sib-mated *H. virescens*.

2. Determine whether the observed degree of inbreeding in our sib-mated *H. virescens* calculated from our ddRAD-seq genotypic data matched theoretical expectations (Falconer & Mackay 1996).

3. Examine fine-scale patterns of genomic change by identifying which genomic regions were resistant to fixation in Bt-selected and sib-mated lines using ddRAD-enabled linkage mapping.

**Results**

We sequenced 204 *H. virescens* individuals from a total of 6 populations that were used in a population-level analysis of genomic change associated with laboratory colonization, artificial selection and inbreeding. These populations were comprised of 2 field-collected, 2 laboratory-reared, and 1 Bt-selected population, as well as specimens from a single inbred family following 10 generations of full-sibling mating (Table 1). Three of these populations, YDK, YHD2, and the inbred line were founded from a collection in Yadkin County, NC, in 1988 (Gould *et al.*, 1995), but were thereafter subjected to different rearing conditions, allowing us to make comparisons of population-genomic change within the same genetic background. In addition, 99 individuals (2 parents and 97 progeny) were sequenced for linkage analysis. This produced a total of 105,487,499 Illumina MiSeq reads (38,221,995 and 67,265,504 for linkage- and population-level analyses, respectively) that passed quality filters (data available upon request).
Genomic diversity among *H. virescens* populations

On average, 338,892 sequencing reads (s.d. = 113397) were produced per individual, and the variation in average read counts was spread uniformly across populations (Supplementary Figure 1). Ninety-four percent of individuals had read counts between 90,000 and 688,000, and were fed into the Stacks pipeline (Catchen *et al.*, 2011; 2013) for *de novo* locus construction (Supplementary Figure 1). Loci constructed by Stacks were sequenced to a 6× average depth of coverage per individual. In total, 4,281 polymorphic 350-bp ddRAD-seq markers (hereafter loci) were detected in at least one individual per population across all populations. Two well-documented challenges commonly encountered when working with moderate coverage reduced representation library data like ours are: 1) uneven distribution of missing data across sets of loci (Davey *et al.*, 2012; Xu *et al.*, 2014), and 2) under-sampling of heterozygotes (Li *et al.*, 2009; Nielsen *et al.*, 2011). Both reduce confidence in final genotypes called by genotyping-by-sequencing SNP calling algorithms, including the algorithm used in Stacks. To overcome these challenges, we examined several subsets of these 4,281 polymorphic loci for our downstream population genomic analyses. These subsets contained between 125-1231 loci, and were chosen based upon the overall proportion missing genotype calls in the subset. The smallest subset consisted of loci for which over 75% of individuals per population had genotypic data present and were therefore likely sequenced to greatest depth of coverage. Each larger subset allowed additional loci at the expense of coverage (*i.e.* more missing genotypic data were allowed; Table 2). By using multiple datasets, we were able to examine whether the presence of missing genotype calls influenced overall estimates of genomic diversity across populations.

We first examined all subsets of loci and determined the mean and maximum numbers of alleles per locus (Table 2). Across all sequenced populations (n = 192 total *H. virescens*), the mean numbers of unique alleles detected per locus ranged from 29 to 34 depending upon the number of loci included.
in the analysis. As more loci were included, the average number of unique alleles detected per locus decreased. However, the maximum number of unique alleles detected increased from 86 in the smallest subset of loci to 94 in the 3 larger subsets. We also examined the proportion of loci that were fixed (i.e. only a single allele present) across populations. Across subsets, few loci were fixed for a single allele in laboratory-reared (5.6-10.9%), Bt-selected (5.3-7.3%) and field-collected (0-2.4%) populations (Table 2). Yet over 80% of loci were fixed in the inbred line following 10 generations of sib-mating (Table 2). Of the 125 loci with the fewest missing genotype calls, 86% were fixed in the inbred line. Expanding the number of loci to include those with more missing genotypes (n = 378, 573, 1231) reduced the percentage of fixed loci in the inbred line by up to 5% (Table 2).

We then determined the mean number of unique alleles present per locus for each subset of loci (Supplementary Figure 2). In general, we found no within population differences in the mean numbers of unique alleles detected among subsets of loci, and therefore we used a single, conservatively chosen subset of loci (n = 378) where at least 10 individuals were genotyped per population per locus for further analysis. The mean numbers of unique alleles per locus were 2.1 for the inbred line, 5.3 for the Bt-selected population, 5.4 and 4.4 for the non-selected, laboratory-reared populations (YDK and BENZ, respectively), and 18.4 and 17 for the field-collected populations (LA and TX, respectively). However, our sample sizes (i.e. numbers of individuals sequenced; see Table 1) differed for each population, and it was unclear whether differences between the aforementioned means were caused by sample size or population-level differences. To determine whether the mean number of unique alleles per locus differed according to population, where sample sizes were equal across populations, we randomly sampled 6, 12, 18, and 24 alleles per population per locus without replacement. We found that increasing the total number of alleles sampled led to an increase in the mean numbers of unique alleles per locus for all but the inbred line (Figure 1). Regardless of the number of alleles sampled,
field populations always exhibited the greatest allelic diversity, followed by selected and non-selected colony populations. The lowest allelic diversity was observed in the inbred line.

When 18 alleles were randomly sampled per population per locus, we detected an average of just over 1 unique allele per locus in the inbred line, indicating that most loci were fixed for a single allele. For the inbred line, 52 loci did not reach fixation. Of these, forty-seven had 2 unique alleles, four had 3 unique alleles, and one had 4 unique alleles out of the 18 that were randomly sampled. On average, Bt-selected and non-selected colony strains each had ca. 3 unique alleles per locus, and field-collected populations had ca. 9 unique alleles per locus (Figure 1). The majority of unique alleles present in the field-collected populations (70.3% and 68.7% for LA and TX populations, respectively) were observed only once (of 18 alleles; Figure 2). Low frequency alleles were less common in the Bt-selected and non-selected laboratory populations (35.1% and 21.5%, respectively) and rare (0.7%) in the sib-mated inbred line (Figure 2).

To further quantify and compare genetic diversity by population, we used the same random sampling regime (n = 18 alleles without replacement), paired with the same conservative subset of loci (n = 378) to calculate sample-size corrected $S_K$ (Charlesworth & Charlesworth, 2010), and $\pi$ (Nei 1978; Nei & Li, 1979). These two measures are complementary: $\pi$ is calculated as the proportion of nucleotides that differ per two randomly chosen DNA sequences, averaged across all pairwise comparisons per marker per population, and $S_K$ is calculated as the number of unique single nucleotide variants in a population at a single locus. When averaged across all markers (n = 378), the number of variant sites ($S_K$) per 350 bp marker was 0.15 for the inbred line, and the maximum $S_K$ was 2.95. Bt-selected and non-selected laboratory populations had, on average, just over 1 nucleotide variant per 350 bp locus, with a maximum of ca. 6. Field-collected populations had the greatest number of variant sites per 350 bp locus, where the genome-wide average was just over 5 nucleotide variants per locus, with a
maximum of ca. 15. Similar trends were observed for genome-wide and maximum nucleotide diversity (\(\pi\)) values. Relative to the laboratory-reared populations, genome-wide estimates of \(\pi\) were nearly an order of magnitude lower for the inbred line. The genome-wide \(\pi\) estimate for laboratory-reared populations ranged from \(4.0 \times 10^{-3}\) (Bt-selected population) to \(6.7 \times 10^{-3}\) (non-selected, YDK population), and \(6.2 \times 10^{-4}\) for the inbred line. Field-collected populations exhibited genome-wide \(\pi\) estimates of \(9.3 \times 10^{-3}\) and \(9.2 \times 10^{-3}\) for the LA and TX populations, respectively. All genome-wide and maximum \(\pi\) and \(S_K\) estimates, along with their corresponding 95% non-parametric bootstrapped confidence intervals (\(N = 5000\)) are reported in Table 3.

Inbreeding among laboratory-reared H. virescens populations

To determine whether the observed degree of homozygosity in our inbred line was consistent with that which would be expected following 10 generations of sib-mating, we compared the inbreeding coefficient \(F\), as calculated according to pedigree- (Falconer & Mackay, 1996) and DNA marker-based information (Keller & Waller, 2002; Kim et al., 2007). The expected inbreeding coefficient \((F_t)\), following 10 generations of sib-mating was 0.89. This expected value fell within the bootstrapped 95% confidence intervals for marker-based inbreeding coefficients \((F_{IT})\) calculated from all subsets of markers. This indicated that there was no significant difference between the expected inbreeding coefficient and the observed inbreeding coefficient calculated using ddRAD-seq marker data. The genome-wide \(F_{IT}\) values (95% CIs) were 0.92 (0.88, 0.96), 0.89 (0.86, 0.92), 0.89 (0.87, 0.92), and 0.88 (0.86, 0.89), for the inbred line as calculated from 125, 378, 573, and 1231 ddRAD-seq loci, respectively.
Linkage mapping

Few genomic resources are available for *H. virescens*. Therefore, we determined the genomic location of loci which were resistant to fixation by generating a dense ddRAD-seq linkage map. The map was produced using a male informative cross (reviewed in Baxter *et al.*, 2009), and was comprised of 659 informative polymorphic loci generated by ddRAD-seq, plus 3 partial gene sequences of *ABCC2*, *HevCaLP*, and *DesatI*. Adding these partial gene sequences to our linkage map, all with known locations in the *B. mori* genome allowed us to validate marker groupings for our linkage map.

All informative ddRAD-seq loci were grouped into 33 linkage groups, two more than the expected 30 *H. virescens* autosomes, and one segregating Z chromosome from the hybrid male parent used in our cross. Linkage groups ranged in size from 7cM to 110cM (Figure 3), and yielded a total map length of 1919.5 cM. On average, there were 20 ddRAD-seq loci per linkage group, and the average spacing was one locus per 3.5 cM. The smallest and largest linkage groups contained 3 and 53 loci, respectively. The *HevCaLP*, *DesatI*, and *ABCC2* genes were grouped with linkage groups 15, 16, and 22, respectively. These linkage groups corresponded to *B. mori* chromosomes 6, 23, and 15 (Table 4), where these candidate genes are known to reside (Gahan *et al.*, 2001; Gahan *et al.*, 2010; Mita *et al.*, 2004).

In total, 99 of the 659 mapped ddRAD-seq loci could be aligned uniquely to a single locus in the *B. mori* genome. Twenty-two linkage groups contained ddRAD-seq loci that could be aligned to a single *B. mori* chromosome (Table 4), while 8 linkage groups did not contain any that could be aligned. Linkage groups 19, 25, and 33 contained ddRAD-seq loci that aligned uniquely to more than one *B. mori* chromosome. This was unlikely caused by spurious associations between ddRAD-seq loci; increasing the LOD score to 8 failed to break up associations for those three linkage groups.
Identification of genomic regions resistant to fixation

To determine where heterozygosity was being maintained in the genome, we examined nucleotide diversity at the 659 mapped ddRAD-seq loci for Bt-selected (YHD2), non-selected (YDK), and inbred line. Of these 659 mapped loci, 302 (46%) were previously included in our population-level analyses, and 357 were newly discovered. Nucleotide diversity ($\pi$) was only calculated for a population if at least 3 individuals were genotyped at a locus. Therefore only 441 loci were examined for the inbred line, and 658 loci were examined for both Bt-selected and non-selected populations. In total, 60 (13%) of the mapped loci retained polymorphism in the inbred line, whereas 645 (98%) and 583 (86%) of mapped loci retained polymorphism in the Bt-selected and non-selected (YDK) populations, respectively.

DdRAD-seq loci that retained polymorphism in the inbred line were spread over 16 linkage groups (Figure 4), and often clustered together within a linkage group (Supplementary Figure 3). We used Tajima's D, a statistical test which identifies departures from the neutral model of molecular evolution, to determine whether any of these polymorphic loci were under balancing selection in the inbred line. Of these 60 polymorphic sites, 20 had statistically significant Tajima's D values ($q > 0.05$) indicative of balancing selection (Table 5). Linkage groups 3, 11, 14, and 15 each contained multiple polymorphic sites that appeared to be under balancing selection. Whereas 5 polymorphic loci on linkage group 11 were spread across the length of the linkage group, linkage groups 3, 14, and 15 contained multiple polymorphic loci clustered together within 5cM of one another.

We examined which, if any, of these polymorphic loci with strongly positive Tajima's D values were shared between the inbred line, the Bt-selected population, and their ancestral laboratory-reared H. virescens population (YDK). Following a bayesian adjustment for multiple comparisons (Efron 2001), no loci showed significantly positive Tajima's D values for the Bt-selected population, but 35
loci had positive Tajima's D values in the non-selected YDK population. While only 3 loci under balancing selection were shared between the inbred line and the non-selected (YDK) populations (n = 3), many of the same genomic regions were shared between populations. For example, both populations shared signatures of balancing selection on chromosome 15, for locus Hv20150 at 61 cM. A complete list of the shared genomic regions showing signatures of balancing selection can be found in Table 6.

Discussion

Here, we examined the degree to which colonization, artificial selection, and intense inbreeding influence genome-wide and fine-scale patterns of diversity. In the absence of a publicly available *H. virescens* reference genome, we used ddRAD-seq de novo locus construction to identify multiple subsets of polymorphic loci ranging in size from 125 (where at least 75% of individuals per population contained genotypic data), to 1231 (where at least 25% of individuals had genotypes present per population). Genome-wide measures of allelic diversity, $F_{IT}$ values, and the degree of homozygosity were either unaffected (Supplementary Figure 2), or minimally affected (Table 2) when markers with high levels of missing genotypic data were present. Therefore, any biased genotype calls made by the Stacks SNP calling algorithm due to our moderate depth of sequencing coverage had little impact on our overall genome-wide estimates of diversity. Our results demonstrate that moderate coverage ddRAD-seq data can be used with confidence when conducting population genomic comparisons of genome-wide means.

We observed a precipitous decline in nucleotide and allelic diversity following long-term laboratory colonization, selection, and inbreeding for *H. virescens*. Despite the decline in genomic diversity for non-selected and Bt-selected laboratory-reared populations, fewer than 10% of loci were
fixed. While our Bt-selected population did not retain the level of genetic diversity that their ancestral
(YDK) laboratory-reared population did, they consistently had higher measures of genomic diversity
than did the non-selected (BENZ) population. Retention of higher levels of polymorphism in our Bt-
selected line was likely due to the measures taken during its generation to ensure genomic diversity was
maintained in the face of strong selection (Gould, 1995). Alternatively, strong genetic bottlenecks in
the non-selected (BENZ) population prior to their use in our study could explain why our Bt-selected
line was more genetically diverse than the non-selected line. Overall, differences among Bt-selected
and non-selected laboratory-reared populations were modest; when 18 total alleles were sampled,
laboratory-reared populations retained ca. 3 alleles per 350-bp locus relative to 9 alleles per 350-bp
locus present in field-collected *H. virescens*. However, few low frequency alleles remained in the
laboratory-reared populations relative to the field-collected populations (Figure 2), which has been
observed elsewhere (Munstermann, 1994). For this reason, laboratory-reared populations are generally
considered inbred (Roush, 1986). In the case of *H. virescens*, our results clearly show that a great deal
of genomic diversity is retained, even following decades in colony. Few genome assembly algorithms
accommodate polymorphism well (Kajitani et al., 2014), and it is clear that the reductions in
heterozygosity in our inbred line will be useful for production of a high quality *H. virescens* reference
genome assembly.

Following 10 generations of inbreeding, > 80% of markers went to fixation in our sib-mated *H.
virescens* population. Indeed, our inbreeding coefficient *F*<sub>IT</sub>, as observed from our ddRAD-seq data,
mets theoretical expectations for all subsets of loci. Our *H. virescens* laboratory population was more
amenable to inbreeding than other insect species (Munstermann, 1994; Rumball et al., 1994; Turissini
et al., 2014; You et al., 2013), despite their relatively high levels of genomic diversity (Figure 1) and
genetic load (Supplementary Figure 4). Higher than expected allelic diversity has been observed in
several other insect species following experimental inbreeding attempts (Munstermann, 1994; Rumball et al., 1994; Turissini et al., 2014; You et al., 2013). As one example, only 57% of the An. gambiae genome went to fixation, as observed according to SNP markers, following 10 generations of inbreeding (Turissini et al., 2014). Observed differences between our H. virescens population and other insects could be species specific, but is more likely related to the proportion of the genome containing balanced lethal systems (Falconer & Mackay, 1996).

To determine where heterozygosity was being maintained in our inbred line, we developed a high density genetic linkage map for H. virescens. Our map contained 659 newly developed 350 bp ddRAD-seq markers that are long enough for future primer design and direct sequencing. This map represents a new tool for an historically important pest species that lacks genomic resources. No ddRAD-seq markers remained unlinked following mapping, which indirectly speaks to the quality of our linkage map. However, the number of groups in our linkage map was 2 more than expected (n = 31). This is likely due to the relatively small mapping population size used in this work (Silva et al., 2007). Other mapping studies that analyzed segregating populations of a similar size have also reported genetic maps with excess numbers of linkage groups (Pootakham et al., 2015; Singh et al., 2009). Another possible explanation for the disparity between our observed and expected number of linkage groups is that polymorphic markers may not be evenly distributed over the chromosomes (Paterson, 1996). Perhaps the markers in several of our smaller linkage groups are found on opposite ends of H. virescens chromosomes, and thus are unlikely to show strong associations.

When we applied our linkage map to examine fine-scale patterns of genomic change following 10 generations of sib-mating, we found several linkage groups contained clusters of loci that did not go to fixation (Figure 4). Twenty out of the 60 ddRAD-seq loci not fixed for a single allele showed signatures of balancing selection (Table 5). These loci were likely genetically linked to balanced lethal
systems, or perhaps chromosomal inversions. Twelve loci spanning the entire length of linkage group 11 (syntenic to *B. mori* chromosome 9) did not go to fixation in the inbred line, and 5 appeared to be under balancing selection. Furthermore, several loci on this linkage group were under balancing selection in our non-selected (YDK) population, suggesting that a large chromosomal inversion may be present on *H. virescens* linkage group 11.

Several smaller genomic regions with signatures of balancing selection were also shared among our non-selected population (YDK) and inbred line (Table 6). These regions are likely associated with deleterious alleles that were sampled from the original field population in Yadkin County, NC from which our YDK population derived. It is possible that those deleterious alleles were never purged from YDK, but instead increased in frequency creating several balanced lethal systems. YDK is ancestral to the inbred line, so these balanced lethals were likely transmitted the inbred line as it was generated.

**Conclusions**

This work serves as one of the most thorough attempts to quantify the effects of genomic responses to selection and inbreeding in a non-model insect species. We demonstrated that laboratory-reared *H. virescens* have reduced allelic and nucleotide diversity relative to field-collected populations, and that inbreeding further diminishes genetic diversity. Although we identified several loci that did not go to fixation in *H. virescens* following 10 generations of inbreeding, our ddRAD-seq marker-based $F_{IT}$ values met theoretical expectations. The genomic regions that retained polymorphism were likely due to deleterious alleles that were present in the founding individuals, and would unlikely go to fixation. This work demonstrates the difficulty involved in producing fully homozygous insect strains, which are currently critical to producing high-quality, complete reference genomes.
Methods

Field-collected H. virescens

Adult male moths were collected from Bossier Parish, Louisiana, and College Station, Texas using pheromone-baited live traps. Collections took place in LA from May through September, 2012, and in TX from May through October, 2012. Moths from each collection date were immediately placed in bottles of 95% ethanol for long-term storage. All bottles were held at -20 °C until DNA isolations took place.

H. virescens colonies

H. virescens were collected from Yadkin County, NC in 1988 (Gould et al., 1995). This original population founded two of the colonies used in this study. YHD2 was selected for high levels of Bt resistance for 4 years on MVP- treated (0.864 mg/mL diet; Mycogen, San Diego, CA) corn-soy diet (Gould et al., 1995), whereas a non-selected population (YDK) was reared on corn-soy diet alone. A third population (BENZ) originating from Stoneville, MS, was acquired from Benzon Research Incorporated (Carlisle, PA). BENZ H. virescens were acquired in their pupal stage, and newly eclosed adults were used for population-level comparisons. To produce an inbred population, single pair matings (SPMs) were set up between YDK siblings for 10 generations. An initial 37 SPMs were used to establish 29 lineages in filial generation one (8 single pair matings did not produce progeny). When SPMs failed to produce offspring, likely due to inbreeding depression, surviving lineages were expanded (Supplementary Figure 4). This was done to extend inbreeding for as many generations as possible, thus promoting complete reduction in heterozygosity as possible. Adult males from each laboratory-reared population were killed by freezing (-20 °C), and stored at -80 °C until DNA isolation.
A non-selected female from the BENZ population was crossed to a Bt-selected (YHD2) male in a single pair mating. One hybrid male offspring was then back-crossed to a Bt-selected (YHD2) female, and their progeny were reared to adulthood on untreated corn-soy diet according to Joyner and Gould (1985). Of the 120 progeny, 97 reached adulthood. Parents and their 97 adult progeny were killed by freezing and stored until DNA isolation as described above.

Genomic DNA library preparation

All DNA was isolated from the adult thorax using a Qiagen Dneasy Blood and Tissue Kit (Qiagen, Inc., Valencia, CA, U.S.A.). Genomic DNA samples were prepared for Illumina sequencing according to the Poland et al., (2012) protocol with minor modifications. Two-hundred ng of DNA per individual were digested with EcoRI and MspI. For each individual, the overhang sites were ligated to standard Truseq Universal adapters (Illumina, Inc. San Diego, CA). These adapters, which were ligated to EcoRI overhang sites, contained one of 48 unique barcodes (Elshire et al., 2011; Supplementary Table 1). DNA fragments from each individual were assigned a unique barcode, and individuals were combined into pools of no more than 48 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. Size-selected DNA fragments were amplified in a Peltier PTC200 thermalcycler (here and throughout) using Illumina primers (Supplementary Table 2) under the following 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 was added via PCR to the MspI adapter. Therefore, sequences from each individual could be identified by the unique combination of barcode and index. A complete
list of barcodes and indices used in this study can be found in the Supplementary Tables 1 and 2, respectively. Amplified libraries were pooled, 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 303 *H. virescens* individuals were spread across 9 full and partial Illumina MiSeq runs.

### De novo marker formation

Overlapping paired-end reads were merged with FLASH (Magoc & Salzberg, 2011), and Stacks v. 1.09 (Catchen *et al.*, 2011; 2013) was used for demultiplexing and *de novo* formation of loci.

Merged paired-end reads were filtered for quality using the process_radtags script. Further 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 did not allow process_radtags to rescue reads where barcode sequences contained an error. All remaining merged reads were truncated at a length of 350 bp, and fed into the Stacks pipeline.

### Stacks parameter settings

Reads from all individuals were run through ustacks with the following parameter settings: -m 3, -M 14 (allowing for 5% nucleotide mismatch rate between alleles per individual), -max_locus_stacks 2, --alpha 0.05. A consensus catalog of loci was first formed using the parents of the mapping cross with cstacks, where the -n 14 parameter allowed for a 5% between individual nucleotide mismatch rate. For the mapping family, genotype calls were made using sstacks prior to field- and colony-strain alleles being added to the catalog. Progeny genotypes were automatically corrected using the Stacks genotypes script. Twenty-four individuals of each colony and field-collected strain collected in 2012...
were later added to the catalog, and all field-collected and laboratory-reared populations were also genotyped using sstacks.

Data analyses

All population genomic and linkage analysis were conducted in R version 3.1.2 (R core team, 2014).

Genomic diversity among H. virescens populations

In total, we sequenced the 13 surviving males from an inbred line subjected to 10 generations of sib-mating, 42-46 males per colony strain, and 30 males per field-collected population (Table 1). Prior to running sequence data through the Stacks pipeline, we checked individual read counts across populations to ensure uniformity (Supplementary Figure 1). Twelve of the 204 individuals sequenced had too few (< 90,000) or too many reads (> 688,000) and were removed from the dataset prior to analysis, following Bi et al. (2013). From our Stacks output, we constructed 4 different sets of consensus loci present across populations. These subsets, containing a core overlapping set of 125 loci, and increasing in size from 125 to 1231, consisted of marker sets with varying percentages of missing genotype calls (range = 11.2-29.5%) (Table 2). We used these 4 different subsets to examine and compare changes in genomic diversity across populations.

We estimated the mean number of unique alleles present per locus, and corresponding 95% non-parametric bootstrapped confidence intervals (N = 5000) across populations using a custom-written R script. Within a population, each locus was randomly sampled without replacement to a depth of 6, 12, 18, and 24 haplotypes, and then unique alleles per locus were counted. Due to their small sample size resulting from intensive inbreeding, we only sampled 6, 12, and 18 haplotypes per locus for the inbred
We focused our analysis on the subset of loci containing 378 consensus loci because genotype calls were present for at least 10 individuals per population. Increasing the number of loci analyzed did not significantly change the mean number of unique alleles per population (Supplementary Figure 2).

We also calculated two measures of nucleotide diversity per 350 bp locus using the R package, pegas (v. 0.6; Paradis, 2010): $\pi$ (Nei, 1987) and $S_K$ corrected for sample size (Charlesworth & Charlesworth, 2010; Watterson, 1975). We then generated population-level genome-wide means and 95% non-parametric bootstrapped ($N = 5000$) confidence intervals for each metric (Table 3).

### Estimating the inbreeding coefficient

To estimate our marker-based inbreeding coefficient, we examined multiple sets of loci (Table 2) and found that trends across all datasets were similar (data not shown). However, we reported $F_{IT}$ values from a set of 378 loci because the dataset contains few missing genotypes per population, while still making inferences from several hundred loci. We calculated $F_{IT}$ for the inbred line relative to the non-selected (YDK) population after Keller and Waller (2002), where $(1-F_{IS})(1-F_{ST}) = 1-F_{IT}$. $F_{IS}$ was the level of inbreeding within the inbred line, calculated as $F_{IS} = 1-(H_O/H_E)$, where $H_O$ and $H_E$ were calculated for each locus using the R package adegenet (v. 1.4-2; Jombart, 2008). $F_{ST}$ was the accumulated effect of inbreeding over time, calculated as $1-(H_E$(inbred line)/$H_E$(YDK)).

### Amplification and genotyping of PCR-based markers for the mapping family

Progeny from the mapping family were genotyped at three additional loci, and these loci were mapped alongside our ddRAD-seq markers to validate our linkage groupings. We targeted the previously described *H. virescens* genes DesatI, ABCC2, and HevCaLP via PCR followed by gel electrophoresis, or direct sequencing. Amplification and genotyping protocols were as follows.
A 468 bp fragment from *DesatI* was amplified in a 30 µl reaction with forward and reverse primers [5'-TGAGGGACCATCGTCTCCAT-3'] and [5'-CACTGCTACATTTTGGGCAG-3'], respectively (Ward, 2009). Each reaction contained 6 µL of 5× GoTaq buffer (Promega), 29 µM per dNTP, 92 ng per primer, 0.75 U GoTaq polymerase, and ca. 1 µg genomic DNA. Sample DNA was amplified alongside a negative control (here and throughout), where pcr-grade H2O was substituted for genomic DNA. Reactions were incubated at 95°C for 1 min followed by 35 cycles of 95°C for 1 min, 52°C for 1 min, and 72°C for 2 min. PCR products were purified using a standard ethanol precipitation, and directly sequenced on an ABI3730xl (Applied Biosystems, San Francisco, CA). A single nucleotide polymorphism (cytosine to tyrosine substitution) at bp 36 was found in the YHD2 parent of the mapping cross. Offspring were genotyped at this locus using PolyPhred (Nickerson et al., 1997), and genotype calls were visually confirmed using consed (Gordon et al., 1998).

An intronic region of the *ABCC2* gene previously described by Gahan et al. (2010) was amplified using primers Hs-ABC2dU02-F1 [5’ – TGGTTACAAGAAATAGAAAATGCAAC-3’] and Hs-ABC2eU03-R2 [5’ – CTTTCAAACGTGAACCGCATC – ’3]. Each 30 µl reaction volume consisted of 6 µL of 5× GoTaq buffer, 29 µM per dNTP, 73 ng per primer, 0.75 U GoTaq polymerase, and 1 µg genomic DNA. Reactions were held at 95°C for 2 min followed by 30 cycles of 95°C for 30 sec, 58°C for 30 sec, and 72°C for 40 sec, and the resulting products were cleaned via ethanol precipitation. Following sequencing on an ABI3730xl, chromatogram files were visualized using FinchTV (version 1.3.1, PerkinElmer, Inc., Seattle, WA). As described by Gahan et al. (2010), the YHD2 parent was homozygous for a 22 bp deletion, whereas the F1 parent was heterozygous for this deletion. Therefore, we examined the segregation of this deletion in the mapping family offspring, which was detectable by the presence of a TAT sequence near amplicon bp 40.

Finally, the *HevCaLP* locus described by Gahan et al. (2007) was amplified in a multiplexed
reaction using three primers: the universal reverse primer [5’-
ATACGAGCTGACGACACGCTGGGAGA-3’], one forward primer that targets a retrotransposon
insertion conferring resistance to *Bacillus thuringiensis* [5’ –
CGCAACGCGATCTACTTTGTCACC – 3’], and another forward primer that targets wild-type
sequence [5’ – AAGTGTCCCAGTCGATGCTGAA – 3’]. An initial 20-µl reaction contained 4 µl 5×
GoTaq buffer, 29 μM per dNTP, 56 ng per primer, 0.5 U GoTaq polymerase, and 1 μg genomic DNA.
Reactions were incubated at 95°C for 2 min, followed by 30 cycles of 95°C for 30 sec, 58°C for 20 sec,
and 72°C for 40 sec. A reconditioning reaction, aimed at reducing heteroduplex formation, was set up
as above, but incubated for 3 rather than 30 cycles. These reactions were capable of producing two
amplicons, which differed in length by 76 bp. The YHD2 parent was homozygous for the long
amplicon (*ca.* 800 bp) containing the insertion that confers resistance to *Bacillus thuringiensis*, whereas
the F₁ parent was heterozygous for a long and short amplicon. PCR products from mapping family
offspring were run on a 2% agarose gel alongside Hyperladder I (Bioline, Taunton, MA) for
visualization and genotype scoring.

**Linkage mapping**

Double-digest RAD-seq markers present in fewer than 75% of the mapping family offspring
were filtered out, and the remainder were checked using a chi-square test for mendelian segregation (*α*
= 0.01). PCR-based markers, as well as those ddRAD-seq markers that segregated in a mendelian
fashion were assigned to linkage groups (LOD = 5, maximum recombination fraction = 0.3) using the
onemap package (Margarido *et al.*, 2007) in R. We validated groupings by aligning all markers to the
*Bombyx mori* genome using Blastn in Kaikobase version 3.2.2 (http://sgp.dna.affrc.go.jp/KAIKObase/).
Furthermore, we confirmed that the locations of the *DesatI*, *ABCC2*, and *HevCaLP* pcr-based markers,
as well as ddRAD-seq markers found in their respective linkage groups aligned to the same *B. mori*
chromosomes (Table 4). Markers on each linkage group were ordered using the recombination
counting and ordering algorithm (RECORD; Van Os *et al*., 2005). RECORD was chosen based upon
previous studies demonstrating the reliability of it’s performance (Collard *et al*., 2009, Mollinari *et al*.,
2009).

Assessment of fine-scale differences in nucleotide diversity across laboratory-reared populations

Markers that retained polymorphism in the inbred line were examined for signatures of
balancing selection in the following populations: non-selected (YDK), Bt-selected (YHD2), and the
inbred line. We used a Tajima's D test as calculated by the R package, pegas (v. 0.6; Paradis, 2010). A
bayesian false discovery rate methodology (Efron *et al*., 2001) implemented with the R package fdrtool
(v.1.2.15; Strimmer 2008) was applied to Tajima's D p-values to account for multiple hypothesis tests.

Acknowledgements

Thanks to Dr. R. Whetten of North Carolina State University, and Dr. J. Schaff of the NCSU
Genomic Sciences Lab for their insightful suggestions on ways to improve our methods. Dr. S.
Micinski, Dr. J. Lopez, and Dr. J. Westbrook collected the moths used in this project. R. Waples
provided one of the custom scripts used in our data pipeline. This project is supported by the
Biotechnology Risk Assessment Program competitive grant number 2012-33522-19793 from the
USDA - National Institute of Food and Agriculture.
References


subflexa (Lepidoptera). Evolution, 50(2), pp. 831-841.


Figures

Figure 1 – Mean numbers of unique alleles detected among 378 loci depend upon the number of haplotypes sampled per population. For all but the inbred line (in green), as sample size increases, so does the number of unique alleles detected per locus. Due to low survivorship in the inbred line, no mean was computed for a haplotype sample size per locus of 24.
Figure 2 – Total numbers of unique alleles detected in sib-mated (INT), Bt-selected (SEL), non-selected (NS), and field-collected (FIELD) populations based upon random sampling of 18 haplotypes per population per locus. Alleles were binned and color-coded according to the frequencies at which they were present out of 18 total haplotypes. Numbers above each bar represent the total number of unique alleles found per population out of 378 loci.
Figure 3 – Heliothis virescens linkage map with a total length of 1919.5 cM. Centimorgan lengths are below each linkage group. Each tick mark represents an individual marker that mapped to a particular position in the linkage group. Red circles next to linkage groups 15 and 22 represent the positions of the HevCaLP, and the ABCC2, respectively. The blue square represents the position of the delta-11-desaturase on linkage group 16.
Figure 4 – Nucleotide diversity per mapped marker site. Grey and blue circles represent nucleotide diversity in non-selected (YDK), and Bt-selected (YHD2) populations, respectively. Red circles represent nucleotide diversity in the inbred line following 10 generations of sib-mating.
### Tables

**Table 1** - Population history, sample sizes (N) and ddRAD-seq read summary data. Filtered N refers to the population sample size following removal of individuals with low read counts.

<table>
<thead>
<tr>
<th>Population</th>
<th>Population History</th>
<th>Total N</th>
<th>Filtered N</th>
<th>Mean Number Reads per Individual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbred Line</td>
<td>10 generations of sib-mating</td>
<td>13</td>
<td>13</td>
<td>334689</td>
</tr>
<tr>
<td>YHD2</td>
<td>Bt-selected</td>
<td>43</td>
<td>41</td>
<td>346264</td>
</tr>
<tr>
<td>BENZ</td>
<td>Unselected</td>
<td>42</td>
<td>40</td>
<td>318636</td>
</tr>
<tr>
<td>YDK</td>
<td>Unselected</td>
<td>46</td>
<td>44</td>
<td>335728</td>
</tr>
<tr>
<td>LA2012</td>
<td>Field-collected</td>
<td>30</td>
<td>27</td>
<td>398336</td>
</tr>
<tr>
<td>TX2012</td>
<td>Field-collected</td>
<td>30</td>
<td>28</td>
<td>306164</td>
</tr>
</tbody>
</table>
Table 2 – Multiple sets of consensus loci used to calculate population genetics parameters. Consensus sets of loci containing 125, 378, and 583 loci are subsets of the largest set containing 1231 loci. The mean and maximum numbers of alleles per marker reported represent summary statistics for the entire multi-population dataset. The abbreviations Bt-sel and NS stand for Bt-selected and non-selected, respectively.

<table>
<thead>
<tr>
<th>Number of Loci</th>
<th>Missing Genotypes (%)</th>
<th>Mean Number of Alleles per Marker</th>
<th>Max Number of Alleles per Marker</th>
<th>Fixed Loci (% of Total Examined)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Inbred line</td>
<td>Bt-Sel (YHD2)</td>
<td>NS (YDK)</td>
</tr>
<tr>
<td>125</td>
<td>11.2</td>
<td>86</td>
<td>86.4</td>
<td>7.2</td>
</tr>
<tr>
<td>378</td>
<td>18.6</td>
<td>94</td>
<td>85.7</td>
<td>5.3</td>
</tr>
<tr>
<td>583</td>
<td>19.3</td>
<td>94</td>
<td>84.2</td>
<td>6.9</td>
</tr>
<tr>
<td>1231</td>
<td>29.5</td>
<td>94</td>
<td>81.3</td>
<td>7.3</td>
</tr>
</tbody>
</table>
Table 3 – Genome-wide nucleotide diversity values per 350 bp locus across populations. Values were generated by randomly sampling (without replacement) 18 haplotypes per population, then calculating $\pi$ and $S_K$ at each locus. Genome-wide values represent population-level $\pi$ and $S_K$ averaged across all loci. The abbreviations Bt-sel and NS stand for Bt-selected and non-selected, respectively.

<table>
<thead>
<tr>
<th>Population</th>
<th>Genome-wide $\pi$ (2.5, 97.5% CIs)</th>
<th>Max $\pi$</th>
<th>Genome-wide $S_K$ (2.5, 97.5% CIs)</th>
<th>Max $S_K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbred line</td>
<td>0.0006 (0.0004, 0.0008)</td>
<td>0.013</td>
<td>0.15 (0.10, 0.19)</td>
<td>2.95</td>
</tr>
<tr>
<td>Bt-Sel (YHD2)</td>
<td>0.0040 (0.0035, 0.0043)</td>
<td>0.020</td>
<td>1.67 (1.54, 1.81)</td>
<td>6.53</td>
</tr>
<tr>
<td>NS (YDK)</td>
<td>0.0067 (0.0061, 0.0073)</td>
<td>0.028</td>
<td>1.66 (1.53, 1.80)</td>
<td>6.20</td>
</tr>
<tr>
<td>NS (BENZ)</td>
<td>0.0051 (0.0046, 0.0056)</td>
<td>0.026</td>
<td>1.35 (1.24, 1.47)</td>
<td>5.46</td>
</tr>
<tr>
<td>Field (LA)</td>
<td>0.0093 (0.0086, 0.0100)</td>
<td>0.034</td>
<td>5.23 (4.92, 5.55)</td>
<td>15.40</td>
</tr>
<tr>
<td>Field (TX)</td>
<td>0.0092 (0.0085, 0.0099)</td>
<td>0.033</td>
<td>5.02 (5.15, 5.33)</td>
<td>15.63</td>
</tr>
</tbody>
</table>
Table 4 – Linkage group (LG) correspondence with *B. mori* chromosome (Chr). Linkage groups with an asterisk contained one or more markers that aligned uniquely to an unmapped *B. mori* sequence. Where linkage groups contained markers that aligned to more than one *B. mori* chromosome, italicized marker names correspond to the italicized *B. mori* chromosome.

<table>
<thead>
<tr>
<th>LG</th>
<th><em>B. mori</em> Chr</th>
<th>Number Markers Aligned to <em>B. mori</em> Chr</th>
<th>Total markers in LG</th>
<th>LG Length (cM)</th>
<th>Average Marker Spacing (cM)</th>
<th>Names of Markers Aligned to <em>B. mori</em> Chr</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>3</td>
<td>25</td>
<td>81.6</td>
<td>3.3</td>
<td>19754, 29329, 20282</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>6</td>
<td>23</td>
<td>52.3</td>
<td>2.3</td>
<td>22394, 27499, 21475, 17667, 17852, 25500</td>
</tr>
<tr>
<td>3*</td>
<td>17</td>
<td>3</td>
<td>24</td>
<td>74.5</td>
<td>3.1</td>
<td>376, 21931, 25112</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>7</td>
<td>35.6</td>
<td>107.3</td>
<td>3.0</td>
<td>2556, 29595, 22095, 8695, 15160</td>
</tr>
<tr>
<td>5*</td>
<td>4</td>
<td>5</td>
<td>36</td>
<td>95.5</td>
<td>1.9</td>
<td>13710, 2328, 1443, 3654, 13880, 22897, 3123, 2430, 19858</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>6</td>
<td>40</td>
<td>79.5</td>
<td>2.0</td>
<td>18575, 4268, 21041, 23679, 66, 5723</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>9</td>
<td>49</td>
<td>95.5</td>
<td>1.9</td>
<td>13710, 2328, 1443, 3654, 13880, 22897, 3123, 2430, 19858</td>
</tr>
<tr>
<td>8*</td>
<td>11</td>
<td>9</td>
<td>25</td>
<td>60.2</td>
<td>1.7</td>
<td>22584, 939, 3060, 18820, 23156, 17932, 202, 22290, 21173</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>2</td>
<td>7</td>
<td>42.4</td>
<td>6.1</td>
<td>4275, 1286</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>5</td>
<td>26</td>
<td>69.6</td>
<td>2.7</td>
<td>5015, 5329, 4265, 4545, 23262</td>
</tr>
<tr>
<td>11</td>
<td>9</td>
<td>8</td>
<td>37</td>
<td>97.8</td>
<td>2.6</td>
<td>19938, 6113, 16969, 19392, 1572, 89, 22588, 19343</td>
</tr>
<tr>
<td>12</td>
<td>-</td>
<td>18</td>
<td>18</td>
<td>89.5</td>
<td>5.0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>6</td>
<td>31</td>
<td>86.8</td>
<td>2.8</td>
<td>2074, 15962, 26388, 1193, 16684, 18825</td>
</tr>
<tr>
<td>14</td>
<td>19</td>
<td>1</td>
<td>21</td>
<td>47.0</td>
<td>2.2</td>
<td>14242</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>7</td>
<td>35</td>
<td>78.9</td>
<td>2.3</td>
<td>25943, 4579, 25083, 1752, 3385, 2784, 18178</td>
</tr>
<tr>
<td>16</td>
<td>23</td>
<td>4</td>
<td>20</td>
<td>55.9</td>
<td>2.8</td>
<td>2485, 16350, 806, 20599</td>
</tr>
<tr>
<td>17</td>
<td>7</td>
<td>2</td>
<td>14</td>
<td>42.5</td>
<td>3.0</td>
<td>29343, 343</td>
</tr>
<tr>
<td>18</td>
<td>26</td>
<td>1</td>
<td>14</td>
<td>57.9</td>
<td>4.1</td>
<td>1113</td>
</tr>
<tr>
<td>19</td>
<td>11, 22</td>
<td>2</td>
<td>4</td>
<td>9.5</td>
<td>2.4</td>
<td>3536, 16185</td>
</tr>
<tr>
<td>20*</td>
<td>10</td>
<td>6</td>
<td>25</td>
<td>57.0</td>
<td>2.3</td>
<td>17820, 2978, 18970, 12277, 16316, 12984</td>
</tr>
<tr>
<td>LG</td>
<td>B. mori Chr</td>
<td>Number Markers Aligned to B. mori Chr</td>
<td>Total markers in LG</td>
<td>LG Length (cM)</td>
<td>Average Marker Spacing (cM)</td>
<td>Names of Markers Aligned to B. mori Chr</td>
</tr>
<tr>
<td>----</td>
<td>-------------</td>
<td>---------------------------------------</td>
<td>---------------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>21</td>
<td>21</td>
<td>3</td>
<td>21</td>
<td>68.4</td>
<td>3.3</td>
<td>26113, 20542, 7624</td>
</tr>
<tr>
<td>22</td>
<td>15</td>
<td>2</td>
<td>33</td>
<td>82.3</td>
<td>2.5</td>
<td>21411, 5773</td>
</tr>
<tr>
<td>23</td>
<td>14</td>
<td>1</td>
<td>17</td>
<td>65.2</td>
<td>3.8</td>
<td>4156</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>7</td>
<td>7</td>
<td>7.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>12, 1</td>
<td>4</td>
<td>53</td>
<td>110.8</td>
<td>2.0</td>
<td>20280, 9211, 12382, 23604</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>7</td>
<td>7</td>
<td>51.0</td>
<td>7.3</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>18.6</td>
<td>6.2</td>
<td>5781</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>6</td>
<td>6</td>
<td>25.3</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>16</td>
<td>1</td>
<td>12</td>
<td>65.2</td>
<td>5.4</td>
<td>23481</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>6</td>
<td>6</td>
<td>32.4</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td>7</td>
<td>7</td>
<td>50.3</td>
<td>7.2</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>3</td>
<td>3</td>
<td>6.8</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>14, 28</td>
<td>2</td>
<td>6</td>
<td>18.1</td>
<td>3.0</td>
<td>6601, 20184</td>
</tr>
</tbody>
</table>
Table 5 – Markers that retained polymorphism following 10 generations of inbreeding, and showed statistically significant ($\alpha = 0.05$) signatures of balancing selection according to Tajima’s D. N represents the number of individuals (of 13 total) that were genotyped per locus. The q-value is derived from a bayesian false discovery rate methodology (Strimmer, 2008) that accounts for multiple comparisons.

<table>
<thead>
<tr>
<th>Marker</th>
<th>LG</th>
<th>Position (cM)</th>
<th>N</th>
<th>$\pi$</th>
<th>Tajima’s D</th>
<th>Unadjusted p-value</th>
<th>q-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>856</td>
<td>3</td>
<td>20</td>
<td>11</td>
<td>$4.3 \times 10^{-3}$</td>
<td>2.2</td>
<td>0.028</td>
<td>0.043</td>
</tr>
<tr>
<td>19130</td>
<td>3</td>
<td>22</td>
<td>13</td>
<td>$8.7 \times 10^{-3}$</td>
<td>2.7</td>
<td>0.006</td>
<td>0.038</td>
</tr>
<tr>
<td>2599</td>
<td>3</td>
<td>26</td>
<td>11</td>
<td>$7.2 \times 10^{-3}$</td>
<td>2.5</td>
<td>0.011</td>
<td>0.042</td>
</tr>
<tr>
<td>6113</td>
<td>11</td>
<td>11</td>
<td>9</td>
<td>$6.0 \times 10^{-3}$</td>
<td>2.4</td>
<td>0.017</td>
<td>0.043</td>
</tr>
<tr>
<td>19392</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>$7.8 \times 10^{-3}$</td>
<td>2.1</td>
<td>0.006</td>
<td>0.050</td>
</tr>
<tr>
<td>20645</td>
<td>11</td>
<td>45</td>
<td>12</td>
<td>$3.0 \times 10^{-3}$</td>
<td>2.1</td>
<td>0.037</td>
<td>0.050</td>
</tr>
<tr>
<td>13922</td>
<td>11</td>
<td>54</td>
<td>10</td>
<td>$6.0 \times 10^{-3}$</td>
<td>2.5</td>
<td>0.014</td>
<td>0.042</td>
</tr>
<tr>
<td>21320</td>
<td>11</td>
<td>86</td>
<td>10</td>
<td>$6.8 \times 10^{-3}$</td>
<td>2.1</td>
<td>0.032</td>
<td>0.049</td>
</tr>
<tr>
<td>68</td>
<td>14</td>
<td>29</td>
<td>11</td>
<td>$5.8 \times 10^{-3}$</td>
<td>2.4</td>
<td>0.017</td>
<td>0.043</td>
</tr>
<tr>
<td>13550</td>
<td>14</td>
<td>34</td>
<td>11</td>
<td>$1.5 \times 10^{-2}$</td>
<td>3.1</td>
<td>0.002</td>
<td>0.031</td>
</tr>
<tr>
<td>18240</td>
<td>15</td>
<td>48</td>
<td>10</td>
<td>$1.0 \times 10^{-2}$</td>
<td>2.6</td>
<td>0.009</td>
<td>0.041</td>
</tr>
<tr>
<td>17301</td>
<td>15</td>
<td>50</td>
<td>11</td>
<td>$4.3 \times 10^{-3}$</td>
<td>2.2</td>
<td>0.030</td>
<td>0.048</td>
</tr>
<tr>
<td>18676</td>
<td>15</td>
<td>59</td>
<td>12</td>
<td>$4.5 \times 10^{-3}$</td>
<td>2.4</td>
<td>0.017</td>
<td>0.043</td>
</tr>
<tr>
<td>20150</td>
<td>15</td>
<td>61</td>
<td>11</td>
<td>$6.9 \times 10^{-3}$</td>
<td>2.3</td>
<td>0.023</td>
<td>0.045</td>
</tr>
<tr>
<td>18178</td>
<td>15</td>
<td>61</td>
<td>11</td>
<td>$5.9 \times 10^{-3}$</td>
<td>2.5</td>
<td>0.012</td>
<td>0.042</td>
</tr>
<tr>
<td>18240</td>
<td>16</td>
<td>32</td>
<td>10</td>
<td>$4.3 \times 10^{-3}$</td>
<td>2.1</td>
<td>0.034</td>
<td>0.041</td>
</tr>
<tr>
<td>13777</td>
<td>18</td>
<td>27</td>
<td>10</td>
<td>$2.4 \times 10^{-2}$</td>
<td>3.0</td>
<td>0.002</td>
<td>0.031</td>
</tr>
<tr>
<td>20662</td>
<td>20</td>
<td>0</td>
<td>11</td>
<td>$4.5 \times 10^{-3}$</td>
<td>2.3</td>
<td>0.021</td>
<td>0.045</td>
</tr>
<tr>
<td>29612</td>
<td>23</td>
<td>36</td>
<td>8</td>
<td>$1.2 \times 10^{-2}$</td>
<td>2.7</td>
<td>0.007</td>
<td>0.040</td>
</tr>
<tr>
<td>4851</td>
<td>26</td>
<td>1</td>
<td>12</td>
<td>$8.0 \times 10^{-3}$</td>
<td>2.2</td>
<td>0.027</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Table 6 – Genomic regions that share signatures of balancing selection for the inbred and non-selected (YDK) populations of *H. virescens*. Numbers in parenthesis following cM ranges indicate the numbers of markers with significant Tajima's D values at the corrected significance threshold of $\alpha = 0.05$ within the cM ranges.

<table>
<thead>
<tr>
<th>Linkage Group</th>
<th>Inbred Line</th>
<th>Non-selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>20-26 (3)</td>
<td>22-26 (2)</td>
</tr>
<tr>
<td>11</td>
<td>11-86 (5)</td>
<td>14-86 (4)</td>
</tr>
<tr>
<td>14</td>
<td>29-34 (2)</td>
<td>30-31 (2)</td>
</tr>
<tr>
<td>15</td>
<td>48-61 (5)</td>
<td>61 (1)</td>
</tr>
<tr>
<td>23</td>
<td>36 (1)</td>
<td>22 (1)</td>
</tr>
<tr>
<td>26</td>
<td>1 (1)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>