Accurate identification of *Helicoverpa armigera–Helicoverpa zea* hybrids using genome admixture analysis: implications for genomic surveillance

Trujillo, D.[1], Mastrangelo, T.[2], Estevez de Jensen, C.[1], Verle Rodrigues, J.C.[3] Lawrie, R.D.[4], Massey, S.E.[5] *

[1] Department of Agro-Environmental Sciences, University of Puerto Rico - Mayaguez, Mayaguez, PR, USA

[2] Universidade de São Paulo, Centro de Energia Nuclear na Agricultura, 13416-000, Piracicaba, SP, Brazil

[3] USDA-APHIS-PPQ-S&T, 22675 Moorefield Rd, Edinburg, TX 78541, USA

[4] Center for Excellence in Quarantine and Invasive Species (CEQUIS), Estacion Experimental Agricola, Jardin Botanico Sur, San Juan, PR

[5] Department of Biology, University of Puerto Rico - Rio Piedras, San Juan, PR, USA

* Corresponding author: steven.massey@upr.edu

Keywords; *Helicoverpa zea*, *Helicoverpa armigera*, hybrid, F1, admixture, genome, crop pest, genome

Abstract

Helicoverpa armigera, the cotton bollworm moth, is one of the world's most important crop pests, and is spreading throughout the New World from its original range in the Old World. In Brazil, invasive *H.armigera* has been reported to hybridize with local populations of *Helicoverpa zea*. The correct identification of *H.armigera-H.zea* hybrids is important in understanding the origin, spread and future outlook for New World regions that are affected by outbreaks, given that hybridization can potentially facilitate *H.zea* pesticide resistance and host plant range via introgression of *H.armigera* genes. Here, we present a genome admixture analysis of high quality genome sequences generated from two *H.armigera-H.zea* F1 hybrids generated in two different labs. Our admixture pipeline predicts 48.8 % *H.armigera* for both F1 hybrids, confirming its accuracy. Genome sequences from five *H.zea* and one *H.armigera* that were generated as part of the study show no evidence of hybridization. Interestingly, we show that four *H.zea* genomes generated from a previous study are predicted to possess a proportion of *H.armigera* genetic material. Using unsupervised clustering to identify non-hybridized *H.armigera* and *H.zea* genomes, 8511 ancestry informative markers (AIMs) were identified. Their relative frequencies are consistent with a minor *H.armigera* component in the four genomes, however its origin remains to be established. We show that the size and quality of genomic reference datasets are critical for accurate hybridization prediction. Consequently, we discuss potential pitfalls in genome admixture analysis of *H.armigera-H.zea* hybrids, and suggest measures that will improve such analyses.

Introduction

Helicoverpa armigera, the Old World cotton bollworm, is an Old World species of moth, and one of the world's most important plant pests, whose larvae consume plants belonging to at least 68 plant families (Cunningham and Zalucki 2014). In the New World, *H. armigera* was initially observed in Brazil in 2013 (Czepak et al. 2013), and has subsequently spread throughout much of Latin America (Murúa et al. 2014) (Tembrock et al. 2019), appearing to have undergone multiple introduction events into South America from the Old World (Gonçalves et al. 2019) (Arnemann et al. 2019). There has not yet been a formal identification of *H.armigera* in North America, although it has been intercepted at several ports (Kriticos et al. 2015). The potential economic damage that *H.armigera* could cause in North America is large: $78 billion worth of crops in the United States were estimated to be susceptible to the pest in 2015 (Kriticos et al. 2015).

A closely related species, *Helicoverpa zea*, is native to the New World, and does not have such a wide host range, feeding off over 110 host plants (Kogan et al. 1989). *H.zea* does not possess such a high degree of resistance to common pesticides as that observed in *H.armigera* (da Silva et al. 2020) (although resistance to Bt-proteins has been widely documented in *H.zea* (Burd et al. 2003)), implying it does not pose such an economic threat as *H.armigera*.

H.armigera and *H.zea* diverged approximately 1.5 million years ago (Behere et al. 2007), and are able to produce viable hybrids (Laster and Sheng 1995) . *H.armigera-H.zea* hybrids have been reported from Brazil (Anderson et al. 2018)

(Valencia-Montoya et al. 2020) (Cordeiro et al. 2020), but have yet to be identified from elsewhere. Adult *H.armigera* are difficult to distinguish from *H.zea* on the basis of morphology, requiring dissection of genitalia (Pogue 2004). Identifying hybrids using such methods is impossible, while larvae of the two species are likewise indistinguishable using morphology (Tay and Gordon 2019). In addition, such methods are inappropriate for screening large numbers of animals. While pure *H.armigera* and *H.zea* can be differentiated using species-specific PCR of the ITS1 region, this method does not work for hybrids (Perera et al. 2015). Hence, genomic methods have great potential utility for accurate species and hybrid identification.

The occurrence of *H.armigera-H.zea* hybrids in Brazil (Anderson et al. 2018) (Valencia-Montoya et al. 2020) (Cordeiro et al. 2020) has implications for pest management programs. Adaptive introgression of genes from invasive pest species into related local species poses a significant threat to global agriculture (Tay and Gordon 2019). A primary reason for studying *H.armigera-H.zea* hybrids in the field is to monitor the adaptive introgression of pesticide resistance genes to *H.zea* from *H.armigera*, which has been subject to intense selective pressure from synthetic pesticides (Walsh et al. 2022). For example, the *CYP337B3* gene, which confers resistance to pyrethroids, has already introgressed into *H.zea* populations in Brazil (Valencia-Montoya et al. 2020). The frequency of pesticide resistance genes in both *H.zea* and *H.armigera* populations has implications for the choice, duration and intensity of pesticide regimens dedicated to their control.

Genes in addition to those responsible for pesticide resistance may also have a propensity to introgress into local *H.zea* populations. For example, *H.zea* lacks genes for gustatory receptors and detoxification compared to *H.armigera*, which may help to explain its more limited range of host plant species (Pearce et al. 2017). These genes may have the potential to introgress from *H.armigera* into *H.zea*, potentially increasing *H.zea's* agricultural impact by increasing its range of host plants.

H.armigera has not yet been formally identified from North America, partly due to difficulties in distinguishing the species from *H.zea*. *H.armigera* was reported in Puerto Rico in 2014 and 2018, however since that time has not been reported again (Flores-Rivera et al. 2022) The Caribbean represents a major transit route for pests and pathogens between North and South America (Waugh 2009), forming a 'Caribbean corridor' , so Puerto Rico is a critical location for monitoring the potential spread of *H.armigera* from the South American continent into North America.

In this study, we implemented a bioinformatic pipeline to predict hybridization proportions by using whole genome sequences. We used the genomes of two lab generated *H.armigera-H.zea* F1 hybrids to confirm the accuracy of our admixture analysis procedure. We demonstrate that genomes from Puerto Rican and North American *H.zea* genomes generated as part of the study do not show evidence of hybridization with *H.armigera*. However, four attributed North American *H.zea* genomes from a previous study displayed potential evidence of hybridization, representing the potential early presence of *H.armigera* in North America. We show that high quality

genome sequence data, reference genomic datasets and careful SNP filtration approaches are important for the accurate determination of hybridization proportions.

Methods

Collection and maintenance of parental species

Individual *H.zea* animals were collected by USDA APHIS collaborators (Todd Gilligan) and shipped to our lab in San Juan in ethanol from Colorado in 2015 (HzCol), Illinois in 2016 (HzIll), Maine in 2016 (HzMaine) and North Carolina in 2016 (HzNC). Species identifications were performed using species specific PCR of the ITS1 region, following the methods of Perera et al., 2015.

All live *Helicoverpa* colonies were maintained under the following conditions: 25 ± 2 °C, 57±9 % relative humidity, photoperiod of 15 hours of light and 9 hours of dark (15: 9 LD). Female pupae were placed in incubators at 22.7 ± 1.6 °C, 82 ± 4 % relative humidity, photoperiod of 15:9 LD, females were placed at a lower temperature to synchronize the emergence of adults with males (Armes et al, 1992; Colvin & Cooter, 1994). The larvae were fed with Gypsy Moth Diet (Frontier Agricultural Sciences, Product # F9630B, Newark DE): 140.2 g of dry mix, 20 g of fats and sugars, 1.6 g of vitamin mix, 0.8 g of aureomycin, 1000 ml of distilled water, with the addition of 12 ml of formaldehyde 1%, and 2.5 g of FABCO mold inhibitor (Frontier Agricultural Sciences, Product # F0018, Newark DE); the agar was dissolved, when the temperature was ~50°C the rest of the reagents were added. Each larva was maintained in transparent plastic cups of 30 ml containing diet. The pupae were maintained in the same cups.

Emerged adults and pupae near to emergence were placed in white plastic buckets of 18.9 l, the upper part of the buckets was covered with cheesecloth (DeRoyal, BIDF2012380-BX, Tennessee) for oviposition. Inside each bucket a Petri dish with autoclaved sand a potted tomato plant was placed to increase relative humidity. The adults received the following diet recipe modified from Grzywacz et al. (2002): 500 ml of distilled water, 50 ml of honey, 10 ml of solution 28 % of Vanderzant vitamin mixture (Sigma, V1007, USA), 1 g of methyl-4-hydroxybenzoate (Sigma, H3647, USA), and 1 ml of ethanol 95 %; methyl-4-hydroxybenzoate was dissolved in the 95% ethanol, then all the ingredients were mixed in the water and fed to adult moths using cotton wicks. The cheesecloth with the oviposited eggs was placed in Ziploc bags of 3.8 liters with fine strips of larval diet. Once larvae emerged, they were transferred to cups with diet. Prior to molecular work, all samples were stored in 90% ethanol in a –20 ºC freezer until DNA extractions were performed.

In CEQUIS, separate colonies of *H.armigera* and *H.zea* were maintained. The colony of *H.armigera* was obtained from five larvae and 30 pupae from Brazil courtesy of Dr Thiago Mastrangelo, University of Sao Paolo. The insects were collected from Bahia (12º13’53’’S , 45º44’44’’W) in 2016 and were introduced to quarantine facilities of the Center of Excellence in Quarantine & Invasive Species on February 4, 2017, under Puerto Rico Department of Agriculture Permit number OV-1617-03 and USDA-APHIS Permit number P526P-15-04600 to Dr. José Carlos Verle Rodrigues. The initial colony of *H.zea* was obtained from larvae collected in Isabela, Puerto Rico, from pigeon peas

on November 11, 2015. During the F9 generation, a reintroduction of insects was done, from larvae collected in Isabela in corn on November 22, 2016.

Breeding of the hybrids

The first hybrid included in the study (PRh) was generated in our lab from a male *H.armigera* from Brazil, and a female *H.zea* from Puerto Rico. The resulting hybrid animal was a female. Using the same rearing methods described above, 15 *H.zea* female pupa and 15 *H.armigera* male pupa were placed into a white plastic bucket with cheesecloth lid and allowed to emerge, mate, and oviposit. All surviving F1 hybrids resulting from this cross were labeled and stored in a –20°C freezer.

Genome sequences were generated from parental animals. A sequence from a male *H. armigera* from Brazil (HaM) was generated. This animal was an adult male *H.armigera* from the *H.armigera* colony initiated in CEQUIS, and was one of the parents for the F1 hybrids generated in the lab. A sequence from a female *H.zea* from Puerto Rico (HzF) was also generated. HzF was reared following the conditions described above, and was a parent for the *H.armigera-H.zea* F1 hybrids (PRh in this study) generated in the lab.

The second hybrid included in the study (MAh) was generated from a female *H.armigera* from Spain and a male *H.zea* from the mainland USA by the USDA APHIS Otis Lab in Buzzards Bay, Massachusetts in 2017 by Dr. Hannah Nadel. The Spanish *H.armigera* mother was from a colony maintained in Spain, however the original collection was from Portugal. The *H.zea* father used in the MAh cross were supplied by

Benzon Research Inc. (Carlisle, PA, USA). This hybrid was reared under the same rearing conditions described previously. Both MaH and PRh hybrids were females.

DNA extraction and sequencing

DNA samples were obtained from the animals using QIAGEN blood and tissue DNA extraction kits (QIAGEN INC., Cat No,/ID 69506) following the manufacturer's protocol, with the exception of the Colorado, Illinois and North Carolina samples, which were extracted using the CTAB method (Calderón-Cortés et al. 2010). DNA quality was assessed using a NanoDrop 2000 (ThermoFisher Scientific, Waltham, MA) to assess DNA concentration (ng/uL) and absorbance (A260/280) and gel electrophoresis (1.5 % agarose) to assess integrity and molecular weight. After checking DNA concentration and quality, the eight samples were shipped overnight on ice to the Rapid Genomics sequencing laboratory in Florida (www.rapid-genomics.com).

Paired end sequencing was conducted by RAPID Genomics on the Illumina HiSeq-X platform (sequencing statistics are displayed in Supplementary Table 1). The sequence data has been deposited in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) under the Accession numbers SAMN35038651 (PRh), SAMN35038652 (MAh), SAMN35038653 (HzF), SAMN35038654 (HaM), SAMN35038647 (HzCol), SAMN35038648 (HzIll), SAMN35038649 (HzMaine), SAMN35038650 (HzNC). Additional genomic data was used in the analysis, consisting of 29 *H.armigera,* 9 *H.zea* and 9 *H.armigera-H.zea* hybrids, from (Anderson et al. 2018) (Table 1). Raw sequence data for these animals were obtained from the Commonwealth

Scientific and Industrial Research Organisation (CSIRO; https://data.csiro.au/collection/csiro:29053). Genome sequence analysis was performed on an Amazon Web Services c6g.4xlarge instance (comprising the AWS Graviton2 processor, 16 vCPUs, 32 Gb memory and Amazon Linux platform).

Mapping and SNP calling procedure

Using fastp (Chen et al. 2018) , sequences were removed if they did not fulfill the criteria of 95% nucleotides > Q20, 3' trimming was conducted by quality, and polynucleotide runs (6 or more consecutive). Filtered and trimmed sequences were repaired using the repair.sh script of BBMap (v37.99) (sourceforge.net/projects/bbmap). They were then mapped to the *H.armigera* reference genome (Pearce et al. 2017) (all-chr-r.fasta, obtained from CSIRO at https://data.csiro.au/collection/csiro:29053v1), using BBMap in paired-end mode. The resulting sam files were converted to bam files, and sorted using SAMtools (Danecek et al. 2021).

BCFtools mpileup (Danecek et al. 2021) was used for variant calling. Bam files for all *Helicoverpa* genomes in the study were processed together, to improve the accuracy of calls of SNPs shared across genomes. After SNP calling, the resulting vcf files were filtered using vcftools (Danecek et al. 2011), removing those SNPs that possessed mean read depth (min-meanDP < 5), Q value (Q < 20) and minor allele frequency (MAF < 0.05).

Admixture analysis

Admixture v1.3.0 (Danecek et al. 2021) was used for genome admixture analysis. Sex chromosomes (chromosome 1) were excluded from the analysis. Plink (Purcell et al. 2007) was used to convert the combined vcf file into bed format, which was used as input for the Admixture analysis, which was run using K=2. Admixture output was visualized using the R ggplot2 package.

Identification of Ancestry Informative Markers

34 *H.armigera* and 7 *H.zea* genomes were identified using the unsupervised clustering approach of Admixture, described above. The genotype data from these genomes was then used to identify SNPs that possessed a minor allele count (MAC) of 7 for the *H.zea* genomes and 1 for *H.armigera* genomes, using vcftools. These were then pruned by removing all SNP positions, where a SNP was completely absent (GT = 0/0) from one or more *H.zea* genome.

Results and Discussion

Genome sequencing results are shown in Table 2, and show that the quality of the raw sequences was high for all eight genomes. For consistency, SNP calling was jointly conducted on the *Helicoverpa* raw sequence reads generated by (Anderson et al. 2018), and on the sequences generated as part of this study. Filtering resulted in the removal of a large proportion of SNPs (83 %); this might be reduced in future by increasing sequence coverage in the overall dataset.

Predicted hybridization proportion of the two lab-reared F1 hybrids

In the hybrid animals, approximately equal proportions of the genome originate from both *Helicoverpa* species (51.2 % *H.zea* : 48.8 % *H.armigera* in both PRh and MAh). These data are displayed on the Admixture plot (Figure 2). In both cases, the Admixture prediction is not exactly 50% *H.zea* : 50% *H.armigera* for either hybrid, even though in the case of PRh, genomes derived from the parental populations were 100% *H.zea* (HzF) and 100% *H.armigera* (HaM). This may be due to two reasons. Firstly, the (male) parental insect population may have possessed a degree of hybridization because they were collected originally from Brazil near where early hybrids have since been detected (Valencia-Montoya et al., 2020). However, it is notable that the MAh F1 hybrid also has the same *H.armigera* : *H.zea* ratio (48.8 % : 51.2 %). This would mean that the *H.zea* from the USA, used to generate the hybrid would also have to have had a low level of *H.armigera* admixture; this seems more unlikely than for an *H.zea* insect from Brazil, where the presence of hybrids has been validated.

Secondly, the Admixture analysis may lack exact precision. This may be the result of a limited number of pure *H.zea* in the dataset (seven), which means that the genetic diversity of the species is not adequately represented.This is supported by the observation that the *H.armigera* : *H.zea* ratio is the same for both F1 hybrids: this indicates a systemic bias in admixture prediction.

Encouragingly, even though *H.armigera* and *H.zea* are closely related species, the Admixture analysis is capable of accurately identifying the relative proportions present in an F1 hybrid genome. In future, accuracy may be improved by refinements in SNP

calling, increasing the sequencing depth in the overall dataset and adding additional genomes, particularly from *H.zea*. Admixture analysis may be affected by a small sample size of one or more of the reference populations (Lawson, van Dorp, and Falush 2018). In the analysis, even after the addition of five *H.zea* genomes generated in this study, only seven non-hybridized *H.zea* genomes were apparent.

Predicted hybridization proportions of other Helicoverpa spp. genomes

From the new genome data generated by the study, the analysis indicated that the animals identified as *H.armigera* (HaM), and *H.zea* (HzF, HzCol, HzIll, HzMaine, HzNC), were non-hybridized animals. All Old World *H.armigera* datasets from (Anderson et al. 2018) were identified as non-hybridized, as expected.

The Admixture analysis reveals some discrepancies with those previously published for 47 previously sequenced *Helicoverpa* genomes (Anderson et al. 2018), which were used as a reference dataset here and in other studies. Most of the animals previously identified as 100 % *H.zea* (Anderson et al. 2018; Valencia-Montoya et al. 2020) are predicted in our analysis to have a *H.armigera* component (132, 133, 134, HZRL10, HZRL12, HZRL17, HZRL20), while several specimens previously identified as hybrids (Anderson et al. 2018; Valencia-Montoya et al. 2020) were identified here as 100% *H.armigera* (131, 144, BRA2, TPG2) (Table 1). TMG4, previously described as a *H.zea* hybrid (Anderson et al. 2018; Valencia-Montoya et al. 2020), is also predicted by our analysis as a hybrid and appears to be F1, given its predicted proportion of 48.8%

H.armigera. Given that this animal was collected in August 2013, this implies that hybridization occurred one generation previous to the collection date.

A key difference between our study and (Anderson et al. 2018; Valencia-Montoya et al. 2020) is that our inclusion of two lab-reared F1 hybrids allows us to verify the accuracy of our analysis. Potential explanations for differences in predicted hybridization proportions reported in (Anderson et al. 2018) may include lack of filtration after SNP calling, and the lower number of *H.zea* in the dataset (leading to a limited reference population for this species). In addition, in (Anderson et al. 2018) SNPs were called on a dataset which included *Helicoverpa punctigera*, *Helicoverpa gelotopoeon*, *Helicoverpa hardwicki* and *Helicoverpa assulta*. In our method, our simultaneous SNP calling procedure only included *H.armigera* and *H.zea* datasets. In addition, in our analysis we chose not to include the Z sex chromosome (chromosome 1), focussing only on autosomes.

The reason for differences between our study and the predicted species proportions described in (Valencia-Montoya et al. 2020) (Table S4, Valencia-Montoya et al; Table 1 this study) is less clear, given that the authors used filtration criteria similar to our own, and only called SNPs against *H.armigera* and *H.zea* genomes, rather than including additional *Helicoverpa* spp. in their analyses. However, the ancestry proportions that they report in Table S4 are derived from ~1 million SNPs identified as segregating between the two species, whereas we base our ancestry proportions on Admixture

analysis, consequently methodological differences may provide the source of the discrepancy.

Potential H.armigera hybridization detected in North American H.zea from 2005

The identification in the reference dataset of potential *H.armigera-H.zea* hybrids from North America (HZRL10, HZRL12, HZRL17, HZRL20), with predicted *H.armigera* proportions of 12.7 %, 18.7 %, 12.5 % and 15.4 %, respectively (Table 1) is interesting, given that *H.armigera* has not been formally identified in the mainland US, and that *H.armigera* was first detected in the Americas in 2013 in Brazil (Czepak et al. 2013). This may therefore represent an early presence of *H.armigera* in the Americas.

The samples were originally described in a 2007 study that constructed a phylogeny of *Helicoverpa* spp. using mitochondrial DNA (Behere et al. 2007), and their genome sequences, used in the study described here, were described in (Anderson et al. 2018). The samples are recorded as having been collected from 'Riverland, NY' (Anderson et al. 2018), however this location is unclear. Dr Daniel Gilrein supplied the *H.zea* samples (Behere et al. 2007), and is based at the Long Island Horticultural Research and Extension Center (LIHREC), Riverhead, NY. The origin of the samples is confirmed as Riverhead, NY (personal communication, Dr Dan Gilrein).

The four samples were collected in 2005, in September / October (personal communication, Dr Dan Gilrein). Significantly, this date predates the first reports of *H.armigera* in the New World in 2013 in Brazil (Czepak et al. 2013). In order to confirm

this result, 8511 AIMs were identified, as described in Methods. Unsupervised clustering allowed the *a priori* identification of 34 *H.armigera* and 7 *H.zea* non-hybridized genomes (Table 1). These were used to identify SNPs that preferentially segregate in one species or the other (AIMS). The 8511 AIMS thus identified indicate a *H.armigera* component ranging from 25.8 to 31.1 % in the four genomes (Table 3). The predicted presence of a *H.armigera* component is consistent with the results from the Admixture analysis.

Regarding the accuracy of this approach, using a reduced set of SNPs is not expected to give the same accuracy as the whole genome considerations utilized by Admixture, however the unsupervised clustering approach represents an independent manner of assessing a potential *H.armigera* contribution to *H.zea* genomic datasets. The *H.armigera* component is higher than predicted by the Admixture approach, which gives 12.5 to 18.7 % *H.armigera*. Notably, the predicted *H.armigera* proportion for the two F1 hybrids is 34.8% (PRh) and 38.9% (MAh) (Table 3), which underestimates the true proportion of 50%. One potential source of error is uneven distribution of AIMs along the chromosomes. Another is that the *H.zea* dataset was limited in size, and so this reduces the accuracy in identifying species-specific AIMs. The low level (0.6 %) of *H.zea* AIMs detected in most of the *H.armigera* genomes reflects the AIM selection approach: the *H.zea* AIMs were present in all 7 *H.zea* genomes, and were also found to be present in at most one *H.armigera* genome in the reference dataset.

Finally, it is possible that low sequencing depth may affect the predicted hybridization proportions. Given that the SNPs are called against a *H.armigera* reference genome,

then if a SNP position has low or no read depth in a particular genome, the SNP calling software will call the *H.armigera* genotype at that position. This means a bias toward calling *H.armigera* AIMs when sequencing depth is low. For example, HZRL10 has 333 AIM positions where there is no sequence coverage, reflecting its low average sequencing depth of 21.8 for the AIM positions. In total there are 986 AIM positions where DP < 5, and so cannot be called with confidence; these constitute 11 % of the total number of AIMs. The AIM positions where there is no sequence coverage are by default identified as *H.armigera* (reflecting the reference genome sequence at those positions). This therefore can account for a proportion of *H.armigera* AIMs in the HZRL10 genome sequence, but not all. This observation may also account for a proportion of the *H.armigera* ancestry in HZRL10 detected by the Admixture analysis.

Further work will be required to validate or discount these observations. In particular, the approaches described are not able to distinguish sample contamination from hybridization. Larger, high quality datasets will be necessary in order to distinguish these two alternative scenarios. Development of such fine-grained methods will have value in screening of historic samples and detection of contamination in hybridization studies, which is currently difficult to detect (a method developed by SEM for detecting contamination of NGS datasets, mitoscan https://github.com/semassey/Scanning-NGS-datasets-for-mitochondrial-and-coronavirus-contaminants/blob/main/mitoscan.sh, maps reads against all NCBI mitochondrial genomes, however it is not able to distinguish contamination by closely related species, due to cross-mapping between closely related mitochondria).

The use of genome admixture analysis for the identification and control of Helicoverpa infestations

We have shown the efficacy of genome admixture analysis for verifying the identity of *Helicoverpa* hybrids, which are morphologically cryptic, and so recalcitrant to traditional identification methods, as is the identification of the two *Helicoverpa* species themselves. We found that increasing the number of *H.zea* genomes in the analysis improved the accuracy of admixture prediction, for the *H.zea* and *H.armigera* genomes, and the two F1 hybrid genomes generated in the study. Likewise, filtering based on sequencing depth also had a similar effect, although we were restricted in increasing filtering stringency, given limitations in sequencing depth in the dataset. Future improvements in accuracy will arise from greater average sequencing depth in the reference genomes used in admixture analyses. Finally, for accurate hybrid identification, whole genome approaches are most likely to yield the precision necessary for understanding the dynamics of *H.armigera* invasivity in the field.

In addition to the indirect detection of *H.armigera* in a region via identification of *H.armigera-H.zea* hybrids, determining the presence of the hybrids will have utility for monitoring the occurrence and spread of pesticide resistance. This is desirable because *H.armigera* populations in the Old World have typically been subjected to significant pesticide exposure, thus leading to the evolution of resistance (Valencia-Montoya et al. 2020). Hybridization with local *H.zea* populations is expected to lead to the introgression of pesticide resistance genes from the *H.armigera* genomic component

(Valencia-Montoya et al. 2020). The phenomenon of rapid introgression of pesticide resistance genes between sister species has been observed in *Anopheles* spp. exposed to selection pressure from pesticide exposure (Norris et al. 2015). The evolutionary dynamics would be expected to be rather similar in crop pests such as *Helicoverpa* spp.

Host plant preference is another agriculturally relevant phenotype that may be influenced by hybridization and gene introgression is that of host plant preference. *H.armigera* has a considerably more extensive plant host range than *H.zea*, apparently partly due to its larger number of gustatory receptor and detoxification genes compared to *H.zea* (Pearce et al. 2017). Adaptive introgression of these genes from *H.armigera* into local populations of *H.zea* may cause changes in the host plant preferences of *H.zea*, a process consistent with the 'hybrid bridge' hypothesis of host shifting of herbivorous insect pests (Floate and Whitham 1993). Furthermore, increasing ease of *H.armigera-H.zea* hybrid detection will allow for the collection of empirical evidence for whether hybridization will influence changes in pesticide susceptibility or feeding behavior. Currently, because hybrids are extremely difficult to identify, empirical data for these phenotypic changes are near impossible to collect.

Puerto Rico is a stepping stone between North and South America, given its geographic location and possession of a major port in San Juan, through which agricultural produce enters and exits the United States. This transit route for agricultural pests and pathogens comprises part of a 'Caribbean corridor'. So far, there are no reports in the

literature on sustained *H.armigera* populations in North America or Puerto Rico. One potential route for the spread of *H.armigera* into North America from South America may be through Puerto Rico.

The detection of *H.armigera-zea* hybrids can reveal aspects of the population dynamics of both species and help inform control strategies. The accurate determination of hybrid proportions can also indicate whether species boundaries are maintained, given that hybridization is often maladaptive.

Accurate admixture prediction methods for *Helicoverpa* species are essential for the design of accurate high throughput hybrid identification tools, and so the datasets generated as part of this study will be useful in the development of tools for the rapid, economical and accurate identification of pure species or hybrids. Future detection of hybrids from Puerto Rico and potentially North America will help inform control regimens, facilitated by the development of rapid molecular tests to accurately determine hybrids. In particular, if there is detection of *H.armigera* in North America, screening of local *H.zea* populations for hybridization could be used to assess whether breeding has occurred.

Acknowledgements

This work was supported by USDA/APHIS Agreement AP20PPQS&T00C161. It may not necessarily express APHIS' views. We would like to thank Dr Todd Gilligan, USDA-APHIS, for providing USA specimens, Dr Hannah Nadel, Otis Laboratory MA, for

providing the MAh hybrid, Dr Fernando Rodrigues da Silva, University of Florida, for support during rearing insect colonies, and Patricia Caligari, UPR-CEQUIS, for assistance during DNA preparations for sequencing. We would also like to thank Dr Tom Walsh and Dr Tek Tay (CSIRO), and Dr Dan Gilrein (LIHREC, Cornell University, Riverhead, NY) for valuable discussion regarding the specimens collected from NY state. Lastly, we thank an anonymous researcher for helpful comments prior to submission.

Authors contributions

Conceptualization: JR, TM, SM; Obtain and maintain biological material: TM, JR, DT; DNA preparation: JR, RL; Funding: JR, CE, SM; Data analysis: SM; Drafting MS: SM; Writing and revision: All.

References

Anderson, Craig J., John G. Oakeshott, Wee Tek Tay, Karl H. J. Gordon, Andreas Zwick, and Tom K. Walsh. 2018. “Hybridization and Gene Flow in the Mega-Pest Lineage of Moth, Helicoverpa.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (19): 5034–39.

Arnemann, Jonas Andre, Stephen Roxburgh, Tom Walsh, Jerson Guedes, Karl Gordon, Guy Smagghe, and Wee Tek Tay. 2019. “Multiple Incursion Pathways for Helicoverpa Armigera in Brazil Show Its Genetic Diversity Spreading in a Connected World.” *Scientific Reports* 9 (1): 19380.

Behere, Gajanan T., Wee Tek Tay, Derek A. Russell, David G. Heckel, Belinda R. Appleton, Keshav R. Kranthi, and Philip Batterham. 2007. “Mitochondrial DNA Analysis of Field Populations of Helicoverpa Armigera (Lepidoptera: Noctuidae) and of Its Relationship to H. Zea.” *BMC Evolutionary Biology* 7 (1): 117.

Burd, Anthony D., Fred Gould, J. R. Bradley, John W. Van Duyn, and William J. Moar. 2003. “Estimated Frequency of Nonrecessive Bt Resistance Genes in Bollworm, Helicoverpa Zea (Boddie) (Lepidoptera: Noctuidae) in Eastern North Carolina.” *Journal of Economic Entomology* 96 (1): 137–42.

Calderón-Cortés, Nancy, Mauricio Quesada, Horacio Cano-Camacho, and Guadalupe Zavala-Páramo. 2010. “A Simple and Rapid Method for DNA Isolation from Xylophagous Insects.” *International Journal of Molecular Sciences* 11 (12): 5056–64.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics* 34 (17): i884–90.

Cordeiro, Erick M. G., Laura M. Pantoja-Gomez, Julia B. de Paiva, Antônio R. B. Nascimento, Celso Omoto, Andrew P. Michel, and Alberto S. Correa. 2020. “Hybridization and Introgression between Helicoverpa Armigera and H. Zea: An Adaptational Bridge.” *BMC Evolutionary Biology* 20 (1): 61.

Cunningham, John Paul, and Myron P. Zalucki. 2014. “Understanding Heliothine (Lepidoptera: Heliothinae) Pests: What Is a Host Plant?” *Journal of Economic Entomology* 107 (3): 881–96.

Czepak, C., K. C. Albernaz, L. M. Vivan, H. O. Guimaraes, and T. Carvalhais. 2013. “First Reported Occurrence of Helicoverpa Armigera (Hubner) (Lepidoptera: Noctuidae) in Brazil.” *Pesquisa Agropecuaria Tropical* 43: 110–13.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156–58.

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2). https://doi.org/10.1093/gigascience/giab008.

Floate, K. D., and T. G. Whitham. 1993. “The ‘Hybrid Bridge’ Hypothesis: Host Shifting via Plant Hybrid Swarms.” *The American Naturalist* 141 (4): 651–62.

Flores-Rivera, X. L., S. V. Paula-Moraes, J. J. Johnson, C. J. Jack, and O. P. Perera. 2022. “Helicoverpa Genus on the Edge of the Continental US: Flight Phenology, Analysis of Hybrid Presence, and Insecticide Performance in High Input Field Crops ….” *Frontiers in Insect Science* 2.

Gonçalves, Rogério Martins, Thiago Mastrangelo, José Carlos Verle Rodrigues, Daniel Fernando Paulo, Celso Omoto, Alberto Soares Corrêa, and Ana Maria Lima de Azeredo-Espin. 2019. “Invasion Origin, Rapid Population Expansion, and the Lack of Genetic Structure of Cotton Bollworm (Helicoverpa Armigera) in the Americas.” *Ecology and Evolution* 9 (13): 7378–7401.

Kogan, M., C. G. Helm, J. Kogan, and E. Brewer. 1989. “Distribution and Economic Importance of Heliothis Virescens and Heliothis Zea in North, Central, and South America and of Their Natural Enemies and Host Plants.” In *Proceedings of the Workshop on Biological Control of Heliothis: Increasing the Effectiveness of Natural Enemies, New Delhi, India 11 15 November 1985*, edited by E. G. King and R. D. Jackson, 241–97. Far Eastern Regional Research Office, Office of International Cooperation & Development, USDA.

Kriticos, Darren J., Noboru Ota, William D. Hutchison, Jason Beddow, Tom Walsh, Wee Tek Tay, Daniel M. Borchert, Silvana V. Paula-Moraes, Cecília Czepak, and Myron P. Zalucki. 2015. “The Potential Distribution of Invading Helicoverpa Armigera in North America: Is It Just a Matter of Time?” *PloS One* 10 (3): e0119618.

Laster, M. L., and C. F. Sheng. 1995. “Search for Hybrid Sterility for Helicoverpa Zea in Crosses between the North-American Heliothis Zea and Helicoverpa Armigera (Lepidoptera, Noctuidae) from China.” *Journal of Economic Entomology* 88: 1288–91.

Lawson, Daniel J., Lucy van Dorp, and Daniel Falush. 2018. “A Tutorial on How Not to over-Interpret STRUCTURE and ADMIXTURE Bar Plots.” *Nature Communications* 9 (1): 3258.

Murúa, M. G., F. S. Scalora, F. R. Navarro, and L. E. Cazado. 2014. “First Record of Helicoverpa Armigera (Lepidoptera: Noctuidae) in Argentina.” *Florida*. https://bioone.org/journals/florida-entomologist/volume-97/issue-2/024.097.0279/First-Record-of-Helicoverpa-armigera-Lepidoptera-Noctuidae-in-Argentina/10.1653/024.097.0279.full.

Norris, Laura C., Bradley J. Main, Yoosook Lee, Travis C. Collier, Abdrahamane Fofana, Anthony J. Cornel, and Gregory C. Lanzaro. 2015. “Adaptive Introgression in an African Malaria Mosquito Coincident with the Increased Usage of Insecticide-Treated Bed Nets.” *Proceedings of the National Academy of Sciences of the United States of America* 112 (3): 815–20.

Pearce, S. L., D. F. Clarke, P. D. East, S. Elfekih, K. H. J. Gordon, L. S. Jermiin, A. McGaughran, et al. 2017. “Genomic Innovations, Transcriptional Plasticity and Gene Loss Underlying the Evolution and Divergence of Two Highly Polyphagous and Invasive Helicoverpa Pest Species.” *BMC Biology* 15 (1). https://doi.org/10.1186/s12915-017-0402-6.

Perera, Omaththage P., Kerry C. Allen, Devendra Jain, Matthew Purcell, Nathan S. Little, and Randall G. Luttrell. 2015. “Rapid Identification of Helicoverpa Armigera and Helicoverpa Zea (Lepidoptera: Noctuidae) Using Ribosomal RNA Internal Transcribed Spacer 1.” *Journal of Insect Science* 15 (1). https://doi.org/10.1093/jisesa/iev137.

Pogue, Michael G. 2004. “A New Synonym of Helicoverpa Zea (Boddie) and Differentiation of Adult Males of H. Zea and H. Armigera (Hübner) (Lepidoptera: Noctuidae: Heliothinae).” *Annals of the Entomological Society of America* 97 (6):

1222–26.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *American Journal of Human Genetics* 81 (3): 559–75.

Silva, Fernando R. da, Dario Trujillo, Oderlei Bernardi, Jose Carlos Verle Rodrigues, Woodward D. Bailey, Todd M. Gilligan, and Daniel Carrillo. 2020. “Comparative Toxicity of Helicoverpa Armigera and Helicoverpa Zea (Lepidoptera: Noctuidae) to Selected Insecticides.” *Insects* 11 (7). https://doi.org/10.3390/insects11070431.

Tay, Wee Tek, and Karl Heinrich Julius Gordon. 2019. “Going Global - Genomic Insights into Insect Invasions.” *Current Opinion in Insect Science* 31 (February): 123–30.

Tembrock, Luke R., Alicia E. Timm, Frida A. Zink, and Todd M. Gilligan. 2019. “Phylogeography of the Recent Expansion of Helicoverpa Armigera (Lepidoptera: Noctuidae) in South America and the Caribbean Basin.” *Annals of the Entomological Society of America* 112 (4): 388–401.

Valencia-Montoya, Wendy A., Samia Elfekih, Henry L. North, Joana I. Meier, Ian A. Warren, Wee Tek Tay, Karl H. J. Gordon, et al. 2020. “Adaptive Introgression across Semipermeable Species Boundaries between Local Helicoverpa Zea and Invasive Helicoverpa Armigera Moths.” *Molecular Biology and Evolution* 37 (9): 2568–83.

Walsh, T. K., D. G. Heckel, Yidong Wu, S. Downes, K. H. J. Gordon, and J. G. Oakeshott. 2022. “Determinants of Insecticide Resistance Evolution: Comparative Analysis Among Heliothines.” *Annual Review of Entomology* 67 (January): 387–406.

Waugh, J. D. 2009. “Trade and Invasive Species in the Caribbean: A Universe of Risk.” The International Union for Conservation of Nature, Gland, Switzerland.

Figure 1 Admixture analysis of *Helicoverpa armigera* and *Helicoverpa zea* genomes

The bar plot shows the relative proportions of *H.armigera* and *H.zea* present in *Helicoverpa* genomes generated in this study (PRh, MAh, HaM, HzF, HzCol, HzIll, HzMaine, HzNC), and from (Anderson et al. 2018). The Admixture analysis used K=2, and excluded sex chromosomes.

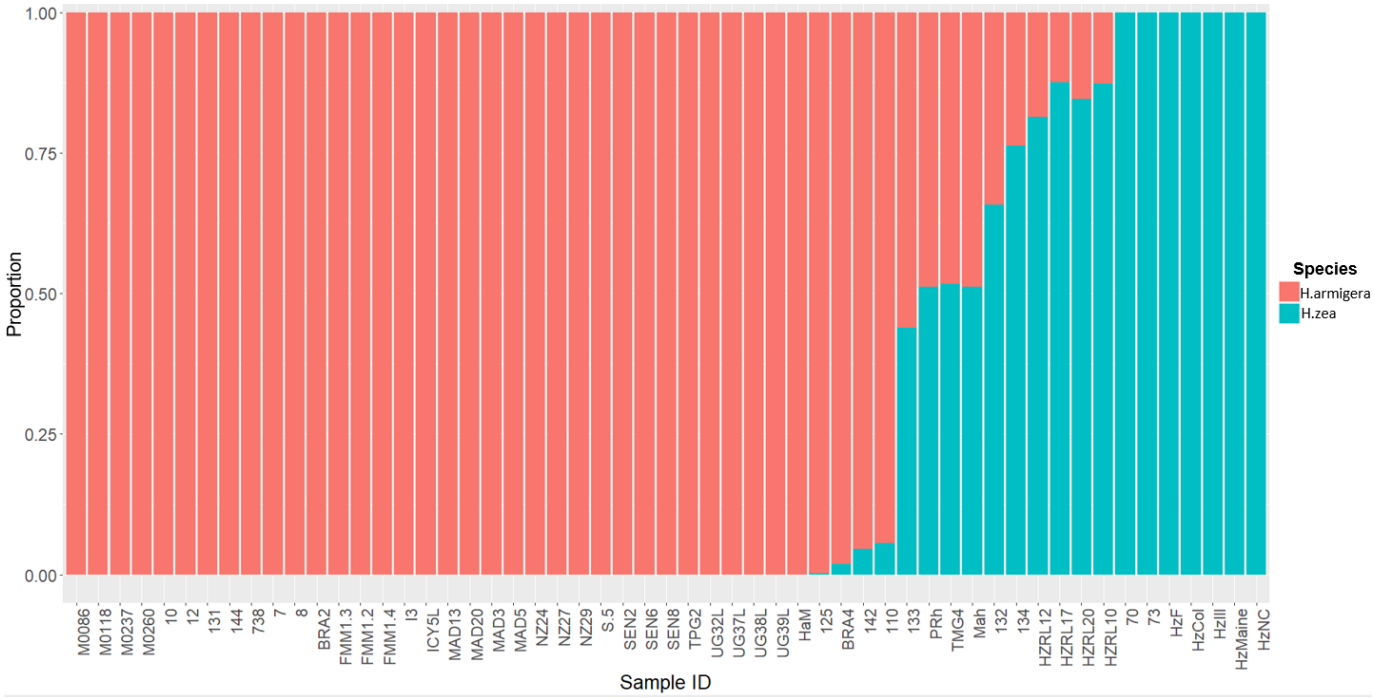

Table 1 Details of 47 additional *Helicoverpa* genomes used in the Admixture analysis. Genome data was obtained from (Anderson et al. 2018). 'NA' means 'not available'.

Assigned species	Sample name	Sequence ID	Species assignment, Anderson et al 2018 (in brackets from Valencia-Montoya et al. 2020) % *H.zea*	Species assignment from Admixture analysis (this work) % *H.zea*	Sample origin
Helicoverpa zea	70	Index_70_703_503_1	NA (98.0)	100	Brazil
Helicoverpa zea	73	Index_73_702_503_1	NA (97.3)	100	Brazil
Helicoverpa zea	132	Index_132_705_502_1	NA (99.9)	65.7	Brazil
Helicoverpa zea	133	133_N702_S502_CGTACTAG-CTCTCTAT_L002_R1_001	NA	43.9	Brazil
Helicoverpa zea	134	Index_134_705_503_1	NA (99.8)	76.3	Brazil
Helicoverpa zea / hybrid	TMG4	TMG4_N701_S501_TAAGGCGA-TAGATCGC_L002_R1_001	51.4 (47.4)	51.6	Brazil
Helicoverpa zea	HZRL10	Index_HZRL10_703_502_1	NA (100)	87.3	USA
Helicoverpa zea	HZRL12	Index_HZRL12_701_503_1	NA (100)	81.3	USA
Helicoverpa zea	HZRL17	Index_HZRL17_705_504_1	NA (100)	87.5	USA
Helicoverpa zea	HZRL20	Index_HZRL20_704_502_1	NA (100)	84.6	USA
Helicoverpa armigera	HM0002				Australia
Helicoverpa armigera	HM0003				Australia
Helicoverpa armigera	HM0004				Australia
Helicoverpa armigera	M0001				Australia
Helicoverpa armigera	M0086	HaM0086_R1	NA	0	Australia
Helicoverpa armigera	M0118	HaM0118_R1	NA	0	Australia
Helicoverpa armigera	M0163				Australia
Helicoverpa armigera	M0237	HaM0237_R1	NA	0	Australia
Helicoverpa armigera	M0243				Australia
Helicoverpa armigera	M0250				Australia
Helicoverpa armigera	M0251				Australia
Helicoverpa armigera	M0254				Australia
Helicoverpa armigera	M0260	HaM0260_R1	NA	0	Australia
Helicoverpa armigera	M0261				Australia
Helicoverpa armigera	M0270				Australia
Helicoverpa armigera	M0272				Australia
Helicoverpa armigera	M0273				Australia
Helicoverpa armigera	M0276				Australia
Helicoverpa armigera	M0299				Australia

Helicoverpa armigera	7	Index_7_703_504_1	NA	0	China
Helicoverpa armigera	8	Index_8_704_504_1	NA	0	China
Helicoverpa armigera	10	Index_10_705_501_1	NA	0	China
Helicoverpa armigera	12	Index_12_705_502_1	NA	0	China
Helicoverpa armigera	FMM1.3	Index_FFM1.3_706_502_1	NA	0	France
Helicoverpa armigera	FMM1.2	Index_FMM1.2_706_501_1	NA	0	France
Helicoverpa armigera	FMM1.4	Index_FMM1.4_706_503_1	NA	0	France
Helicoverpa armigera	738	Index_738_705_504_1	NA	0	India
Helicoverpa armigera	I3	Index_I3_701_502_1	NA	0	India
Helicoverpa armigera	ICY5L	Index_ICY5L_702_502_1	NA	0	India
Helicoverpa armigera	MAD13	Index_MAD13_702_501_1	NA	0	Madagascar
Helicoverpa armigera	MAD20	Index_MAD20_703_501_1	NA	0	Madagascar
Helicoverpa armigera	MAD3	Index_MAD3_701_503_1	NA	0	Madagascar
Helicoverpa armigera	MAD5	Index_MAD5_701_501_1	NA	0	Madagascar
Helicoverpa armigera	NZ24	Index_NZ24_701_501_1	NA	0	New Zealand
Helicoverpa armigera	NZ27	Index_NZ27_706_504_1	NA	0	New Zealand
Helicoverpa armigera	NZ29	Index_NZ29_702_501_1	NA	0	New Zealand
Helicoverpa armigera	SEN2	Index_SEN2_704_501_1	NA	0	Senegal
Helicoverpa armigera	SEN6	Index_SEN6_701_502_1	NA	0	Senegal
Helicoverpa armigera	SEN8	Index_SEN8_702_502_1	NA	0	Senegal
Helicoverpa armigera	S.5	Index_S.5_703_503_1	NA	0	Spain
Helicoverpa armigera	UG32L	Index_UG32L_705_503_1	NA	0	Uganda
Helicoverpa armigera	UG37L	Index_UG37L_701_503_1	NA	0	Uganda
Helicoverpa armigera	UG38L	Index_UG38L_702_503_1	NA	0	Uganda
Helicoverpa armigera	UG39L	Index_UG39L_703_503_1	NA	0	Uganda
Helicoverpa armigera / hybrid	110	110_N704_S504_TCCTGAGC-AGAGTAGA_L002_R1_001	8.9 (4.5)	5.6	Brazil
Helicoverpa armigera / hybrid	125	Index_125_702_501_1	3.2 (0)	0.2	Brazil
Helicoverpa armigera / hybrid	131	Index_131_704_501_1	2.4 (0)	0	Brazil
Helicoverpa armigera / hybrid	142	142_N703_S503_AGGCAGAA-TATCCTCT_L002_R1_001	7.9 (0.8)	4.6	Brazil
Helicoverpa armigera / hybrid	144	Index_144_706_501_1	2.8	0	Brazil
Helicoverpa armigera / hybrid	BRA2	Index_BRA2_704_503_1	3.2 (0.5)	0	Brazil
Helicoverpa armigera / hybrid	BRA4	Index_BRA4_701_504_1	4.6 (0.2)	1.9	Brazil
Helicoverpa armigera / hybrid	TPG2	Index_TPG2_701_501_1	2.1 (0)	0	Brazil
Helicoverpa assulta	343	Index_343_704_504_1	NA	NA	Australia

Helicoverpa assulta	YC3	Index_YC3_706_503_1	NA	NA	China
Helicoverpa gelotopoeon	ARG1	Index_ARG1_704_503_1	NA	NA	Argentina
Helicoverpa gelotopoeon	ARG2	Index_ARG2_701_504_1	NA	NA	Argentina
Helicoverpa gelotopoeon	ARG3	Index_ARG3_702_504_1	NA	NA	Argentina
Helicoverpa gelotopoeon	ARG4	Index_ARG4_703_501_1	NA	NA	Argentina
Helicoverpa hardwicki	Hh	Hh_NoIndex_L007_R1_001	NA	NA	Australia
Helicoverpa punctigera	M0087	M0087_R1	NA	NA	Australia
Helicoverpa punctigera	M0236	M0236_R1	NA	NA	Australia
Helicoverpa punctigera	M0239	M0239_R1	NA	NA	Australia
Helicoverpa punctigera	M0244	M0244_R1	NA	NA	Australia
Helicoverpa punctigera	M0245	M0245_R1	NA	NA	Australia
Helicoverpa punctigera	M0263	M0263_R1	NA	NA	Australia
Helicoverpa punctigera	M0264	M0264_R1	NA	NA	Australia
NA	M0238	M0238_R1	NA	NA	NA

Table 2 Mapping and SNP statistics, and hybridization proportions, of eight new genome sequences generated in the study

Sample name	Average genome sequencing depth	Number of SNPs after filtering	Species assignment from Admixture analysis (this work) % *H.zea*	Geographic origin
PRh	31.4	9230976	51.2	F1 hybrid of male *H.armigera* from Brazil and a female *H.zea* from Puerto Rico
MAh	34.7	8990547	51.2	F1 hybrid of a female *H.armigera* from Spain and a male *H.zea* from USA
HaM	43.7	7189237	0	Brazil
HzF	37.4	5139536	100	Puerto Rico
HzCol	25.0	5185015	100	Colorado, USA
HzIll	25.0	5190809	100	Illinois, USA
HzMaine	25.3	5188817	100	Maine, USA
HzNC	26.0	5793330	100	North Carolina, USA

Table 3 Proportion of *H.zea* AIMs present in different genomic datasets

8511 AIMs were identified as described in Methods. The proportion of *H.armigera* - specific AIMs identified in the different genomes is listed. The proportions were determined by comparison with the filtered SNPs produced from the SNP calling procedure described in Methods.

Species assignment	Sample	Proportion of *H.zea* specific AIMs (%)	Average sequence depth of the 8767 AIMS
H.armigera-H.zea F1 hybrid	PRh	65.2	102.3
H.armigera-H.zea F1 hybrid	MAh	61.1	107.2
H.armigera	HaM	0.6	98.8
H.zea	HzF	100	107.0
H.zea	HzCol	100	85.7
H.zea	HzIll	100	83.1
H.zea	HzMaine	100	86.3
H.zee	HzNC	100	25.7
H.zea	70	100	23.7
H.zea	73	100	29.1
H.zea	132	61.0	10.1
H.zea	133	52.2	8.7
H.zea	134	66.9	11.9
H.zea / hybrid	TMG4	66.0	76.1
H.zea	HZRL10	73.7	21.8
H.zea	HZRL12	69.9	14.9
H.zea	HZRL17	74.2	17.2
H.zea	HZRL20	70.9	21.9
H.armigera	M0086	0.6	39.7

H.armigera	M0118	0.6	26.1
H.armigera	M0237	0.6	30.7
H.armigera	M0260	0.6	34.7
H.armigera	7	0.6	14.5
H.armigera	8	0.6	18.1
H.armigera	10	0.6	12.2
H.armigera	12	0.6	16.4
H.armigera	FMM1.3	0.6	11.7
H.armigera	FMM1.2	0.6	9.0
H.armigera	FMM1.4	0.6	13.3
H.armigera	738	0.6	24.9
H.armigera	I3	0.6	16.2
H.armigera	ICY5L	0.6	14.2
H.armigera	MAD13	0.6	13.5
H.armigera	MAD20	0.6	14.1
H.armigera	MAD3	0.6	15.3
H.armigera	MAD5	0.6	9.9
H.armigera	NZ24	0.6	10.7
H.armigera	NZ27	0.6	28.0
H.armigera	NZ29	0.6	16.0
H.armigera	SEN2	0.6	12.7
H.armigera	SEN6	0.6	16.1
H.armigera	SEN8	0.6	22.5
H.armigera	S.5	0.6	15.2
H.armigera	UG32L	0.6	19.7
H.armigera	UG37L	0.6	13.8
H.armigera	UG38L	0.6	18.8
H.armigera	UG39L	0.6	15.3

H.armigera / hybrid	110	11.4	49.6
H.armigera / hybrid	125	3.4	11.8
H.armigera / hybrid	131	0.6	16.4
H.armigera / hybrid	142	11.1	33.5
H.armigera / hybrid	144	0.6	14.2
H.armigera / hybrid	BRA2	0.6	23.0
H.armigera / hybrid	BRA4	4.6	21.8
H.armigera / hybrid	TPG2	0.6	19.8

Supplementary Material

Supplementary Table 1 Sequencing statistics

Fastq files corresponding to this table can be found in the SRA under BioProject ID: PRJNA973566 or SRA submission number SUB13362524. The latter four genomes were sequenced in two separate lanes. After filtration and repair, reads from the two lanes were merged before mapping.

Sample name	Sample code	Number of reads	Number of filtered reads	Average length of filtered reads (bp)	Q20 bases of filtered reads (%)	Q30 bases of filtered reads (%)
PRh	RAPiD-Genomics_F061_UPR_134801_P001_WA01_i5-505_i7-59_S1442_L008_R1_001.fastq	68237942	68215872	149	98.6	96.0
	RAPiD-Genomics_F061_UPR_134801_P001_WA01_i5-505_i7-59_S1442_L008_R2_001.fastq	68237942	68160626	149	96.4	92.0
MAh	RAPiD-Genomics_F061_UPR_134801_P001_WA02_i5-505_i7-27_S1443_L008_R1_001.fastq	69981718	69481299	149	98.6	96.2
	RAPiD-Genomics_F061_UPR_134801_P001_WA02_i5-505_i7-27_S1443_L008_R2_001.fastq	69981718	69462248	148	95.1	89.2
HaM	RAPiD-Genomics_F071_UPR_134802_P001_WA01_i5-508_i7-59_S2885_L004_R1_001.fastq	81978423	81952805	148	97.5	93.4
	(RAPiD-Genomics_F071_UPR_134802_P001_WA01_i5-508_i7-59_S2885_L004_R2_001.fastq)	81978423	81950055	147	95.6	90.4
HzF	RAPiD-Genomics_F071_UPR_134802_P001_WA02_i5-508_i7-27_S2886_L004_R1_001.fastq	74909547	74854750	148	97.6	93.6
	RAPiD-Genomics_F071_UPR_134802_P001_WA02_i5-508_i7-27_S2886_L004_R2_001.fastq	74909547	74856157	147	95.8	90.7
HzCol	RAPiD-Genomics_F131_UPR_134803_P001_WA01_i5-512_i7-97_S184_L002_R1_001.fastq.gz	40950840	40909180	148	98.4	95.1
	RAPiD-Genomics_F131_UPR_134803_P001_WA01_i5-512_i7-97_S184_L002_R2_001.fastq.gz	41688962	41645905	148	98.5	95.1
	RAPiD-Genomics_F134_UPR_134803_P001_WA01_i5-512_i7-97_S5_L001_R1_001.fastq.gz	9003282	9002312	149	96.4	90.4
	RAPiD-Genomics_F134_UPR_134803_P001_WA01_i5-512_i7-97_S5_L001_R2_001.fastq.gz	9003282	8999277	149	95.2	88.1
HzIll	RAPiD-Genomics_F131_UPR_134803_P001_WB01_i5-512_i7-109_S185_L002_R1_001.fastq.gz	41688962	41645904	148	98.5	95.1
	RAPiD-Genomics_F131_UPR_134803_P001_WB01_i5-512_i7-109_S185_L002_R2_001.fastq.gz	41688962	41670503	148	98.0	93.9
	RAPiD-Genomics_F134_UPR_134803_P001_WB01_i5-512_i7-109_S6_L001_R1_001.fastq.gz	8180764	8179779	149	96.4	90.5
	RAPiD-Genomics_F134_UPR_134803_P001_WB01_i5-512_i7-109_S6_L001_R2_001.fastq.gz	8180764	8177487	149	95.4	88.6
HzMaine	RAPiD-Genomics_F131_UPR_134803_P001_WC01_i5-512_i7-121_S186_L002_R1_001.fastq.gz	41936744	41894121	148	98.5	95.1
	RAPiD-Genomics_F131_UPR_134803_P001_WC01_i5-512_i7-121_S186_L002_R2_001.fastq.gz	41936744	41916752	148	98.1	94.0
	RAPiD-Genomics_F134_UPR_134803_P001_WC01_i5-512_i7-121_S7_L001_R1_001.fastq.gz	8823812	8822863	149	96.4	90.5
	RAPiD-Genomics_F134_UPR_134803_P001_WC01_i5-512_i7-121_S7_L001_R2_001.fastq.gz	8823812	8819941	149	95.4	88.6
HzNC	RAPiD-Genomics_F131_UPR_134803_P001_WD01_i5-512_i7-133_S187_L002_R1_001.fastq.gz	43364690	43319834	149	98.4	95.1

	RAPiD-Genomics_F131_UPR_134803_P001_WD01_i5-512_i7-133_S187_L002_R2_001.fastq.gz	43364690	43341100	149	98.0	93.9
	RAPiD-Genomics_F134_UPR_134803_P001_WD01_i5-512_i7-133_S8_L001_R1_001.fastq.gz	8677566	8676286	150	96.4	90.3
	RAPiD-Genomics_F134_UPR_134803_P001_WD01_i5-512_i7-133_S8_L001_R2_001.fastq.gz	8677566	8673245	150	95.3	88.4