Hybridization and gene flow in the mega-pest lineage of moth, *Helicoverpa*

Craig J. Anderson[a,b,1], John G. Oakeshott[a], Wee Tek Tay[a], Karl H. J. Gordon[a], Andreas Zwick[a], and Tom K. Walsh[a]

[a]Black Mountain Laboratories, Commonwealth Scientific and Industrial Research Organization, Canberra, ACT 2601, Australia; and [b]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, United Kingdom

Within the mega-pest lineage of heliothine moths are a number of polyphagous, highly mobile species for which the exchange of adaptive traits through hybridization would affect their properties as pests. The recent invasion of South America by one of the most significant agricultural pests, *Helicoverpa armigera*, raises concerns for the formation of novel combinations of adaptive genes following hybridization with the closely related *Helicoverpa zea*. To investigate the propensity for hybridization within the genus *Helicoverpa*, we carried out whole-genome resequencing of samples from six species, focusing in particular upon *H. armigera* population structure and its relationship with *H. zea*. We show that both *H. armigera* subspecies have greater genetic diversity and effective population sizes than do the other species. We find no signals for gene flow among the six species, other than between *H. armigera* and *H. zea*, with nine Brazilian individuals proving to be hybrids of those two species. Eight had largely *H. armigera* genomes with some introgressed DNA from *H. zea* scattered throughout. The ninth resembled an F1 hybrid but with stretches of homozygosity for each parental species that reflect previous hybridization. Regions homozygous for *H. armigera*-derived DNA in this individual included one containing a gustatory receptor and esterase genes previously associated with host range, while another encoded a cytochrome P450 that confers insecticide resistance. Our data point toward the emergence of novel hybrid ecotypes and highlight the importance of monitoring *H. armigera* genotypes as they spread through the Americas.

hybridization | gene flow | population genomics | pest | selective sweep

Longstanding assumptions regarding the integrity of species are increasingly coming under challenge as we begin to comprehend the extent of gene flow between species (1, 2). Several cases have now been reported where large portions of the genome have been exchanged between closely related species (1, 3) and, on occasion, with more distantly related taxa (4). Lepidoptera have been useful in exploring the dynamics of hybridization (5), with invasive species in particular offering a rare opportunity to follow the formation of novel ecotypes following secondary contact. Some lepidopteran species are major agricultural pests among which lateral gene transfer could have significant implications for adaptive traits such as pesticide resistance and host range, and therefore, for agricultural production. The recent spread of the cotton bollworm, *Helicoverpa armigera*, into South America presents such a threat (6, 7).

H. armigera is a significant pest of diverse agricultural and horticultural crops throughout temperate and tropical regions of Asia, Europe, and Africa (subspecies *H. armigera armigera*), as well as Australasia (subspecies *H. armigera conferta*). Chemical pesticides and genetically modified crops expressing insecticidal proteins have been widely used to control outbreaks, but the moth has shown a remarkable propensity to develop resistance to many pesticide classes (8); it is claimed to be responsible for more reports of resistance than any other agricultural pest (9). Adult *H. armigera* are also adept at long-range, facultative migration (10–12), and so far the species shows little evidence for population structure below the subspecies level (13), enabling the rapid spread of resistance. In 2013, it was discovered that *H. armigera* had become established in Brazil (6, 7, 14), and while

the origin of the incursion is still debated, previous analyses have implicated several source populations that were likely facilitated by international trade (13, 15). Since then, *H. armigera* is estimated to have caused annual crop losses averaging $1 billion in Brazil alone (16, 17), with losses set to increase as the species spreads through South, Central, and potentially North America (18).

It is estimated that *H. armigera* previously entered the Americas ~1.5 Mya, leading to the formation of the closely related sister species *Helicoverpa zea* (19, 20). While the two species remain similar in many respects, *H. zea* has a narrower host range and lower propensity to develop pesticide resistance. Recent work by the *Helicoverpa* genome consortium (20) has suggested that these differences result, in part, from *H. zea* possessing fewer gustatory receptor and detoxification genes. Thus, one of the major concerns following the recent invasion of *H. armigera* is that it will hybridize with *H. zea*, introducing genes associated with pesticide resistance and host range expansion. Evidence in the field has thus far fallen short of unequivocal identification of hybridization (13, 21), although high levels of synteny between the two species' genomes suggest that genomic compatibility is likely (20). Hybridization in the laboratory has been demonstrated previously (22, 23); Laster and Sheng (24) specifically demonstrated that there were no instances of sterility in reciprocal crosses inbred for two generations or in lines backcrossed for four.

Other *Helicoverpa* species in the heliothine mega-pest lineage that have been classified as pests include the early diverging *Helicoverpa punctigera* in Australia (12), *Helicoverpa assulta* in much of the Old World (12), and *Helicoverpa gelotopoeon* in South
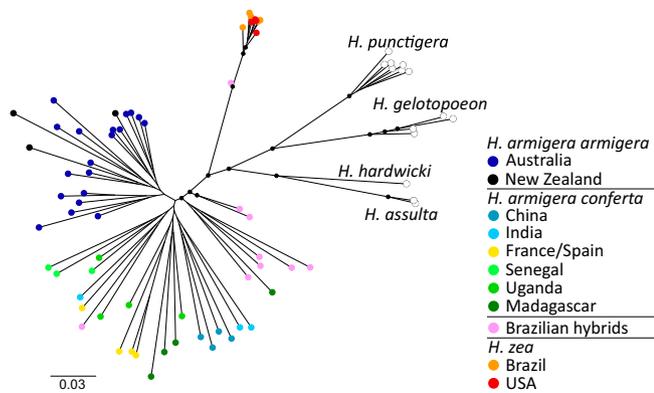
POPULATION BIOLOGY

Fig. 1. Maximum likelihood phylogenies for all six *Helicoverpa* species, calculated using 4D SNPs across autosomes. Interior nodes marked with a black circle represent bootstrap support greater than 50%.

America (25). These three species vary considerably in their host range and resistance to pesticides (26), while the closely related *Helicoverpa hardwicki* persists on a relatively specialized diet and is restricted to northern Australia (22). The natural capacity for gene flow among these species and with *H. armigera* or *H. zea* is not well understood, but hybridization has also been suggested between *H. armigera* and *H. assulta* under laboratory conditions (23).

Here we use whole-genome resequencing to screen for hybridization and introgression among the six *Helicoverpa* species named above. We find little population structure within *H. armigera* or *H. zea*, other than a separation between the two subspecies of *H. armigera*, and no evidence of hybridization involving the other four species. Importantly, we find unequivocal evidence of hybridization between *H. armigera* and *H. zea* in South America. This involves all 31 chromosomes, although the chromosomal regions involved vary widely among individuals. Most cases involve introgression of large genomic regions from *H. zea* into a substantially *H. armigera* genetic background. However, one individual that more closely resembles *H. zea* is nevertheless homozygous for significant genomic regions from *H. armigera*, including ones previously associated with host use and insecticide resistance (20). The data highlight the potential for novel *Helicoverpa* ecotypes to form and spread in South America, presenting new challenges to agriculture in the region.

Results and Discussion

Relationships Among Taxa and Evidence for Hybrids in Brazil. A phylogeny constructed from fourfold degenerate (4D) autosomal sites in 76 individuals from across the six *Helicoverpa* species named above recapitulates the relationships previously demonstrated by Cho et al. (Fig. 1) (26). Additionally, this phylogeny and a more targeted analysis of only *H. armigera* and *H. zea* (Fig. S1) show that nine individuals from Brazil are scattered between the major *H. armigera* and *H. zea* lineages, with eight closer to the former and the ninth closer to *H. zea*. The first eight of these individuals had previously been classified as *H. armigera* and the ninth as *H. zea* on the basis of mitochondrial sequence data (15). Although the two subspecies of *H. armigera* are generally discrete, any further distinction among the diverse geographic samples of *H. armigera* is poorly supported and likely reflects a complex history of migration and gene flow within subspecies. The data suggest that these nine Brazilian individuals may be hybrids, albeit with varying degrees of genetic admixture, and that *H. armigera armigera* was the source of the Brazilian incursion (13, 15). This interpretation is also borne out in the corresponding phylogenies for Z chromosome sites (Fig. S2).

Application of the F_3-test to estimate signals of contemporary gene flow among the six species found evidence for genetic exchange only in a single comparison (Dataset S1). This instance, where the nine Brazilian individuals in question were the target population and *H. armigera armigera* and *H. zea* from the United

States (which *H. armigera* had not penetrated at the time) were sources, yielded a negative f_3 (-0.009, $Z = -7.775$), indicative of gene flow. Henceforth, we refer to these individuals as the "Brazilian hybrids."

We then examined the relationships between *H. zea* and the *H. armigera* subspecies using the variance in allele frequencies among subpopulations relative to the global variance (F_{ST}) (Fig. S3 and Table S1) as a prelude to principal component analysis (PCA) (Fig. S4) and admixture analysis (Fig. S5 and ref. 27) to further investigate the affinities of the Brazilian hybrids. The F_{ST} analysis, which excluded the Brazilian hybrids, showed significantly lower F_{ST} values between *H. armigera armigera* and *H. armigera conferta* than between either subspecies and *H. zea* ($P < 0.001$). Interestingly, the values in all three of those comparisons are slightly, but significantly, higher ($P < 0.001$) for the Z chromosome than for autosomes, indicating increased evolutionary rates in the former (28). The PCA separated *H. zea* from *H. armigera* on principal component 1 (PC1) and clearly distinguished clusters representing both *H. armigera* subspecies on principal component 2 (PC2). Consistent with the phylogenies above, most of the Brazilian hybrids fall between *H. armigera armigera* and *H. zea* but lie closer to the former, while a single individual has closer affinity with *H. zea*. As is consistent with both those analyses, admixture analyses show strong differentiation between the two species, which is supported by the lowest cross-validation (CV) error of any assessed, when the number of populations (K) is two. When K = 3, it is clear that the Brazilian hybrids are derived from *H. armigera armigera* and *H. zea*, while Australian *H. armigera conferta* forms a discrete cluster.

We then employed an outgroup F_3-test to examine which of the sequences within *H. armigera armigera* yielded the highest scores, indicative of the most shared genetic history with Brazilian hybrids. We found that the Brazilian hybrids most resemble the four European individuals, with several African samples presenting broadly comparable but somewhat smaller values for f_3 and followed by Asian individuals that yielded lower values still (Fig. 2).

Genetic Diversity and Effective Population Size. Several recent studies have found unusually high levels of genetic variation both within native populations of *H. armigera* and within the *H. armigera* that have been introduced into South America (13, 15, 29). Our own calculations of average genome-wide nucleotide diversity (Table S2) concur that *H. armigera armigera* and *H. armigera conferta* are unusually polymorphic in autosomal regions (0.008 and 0.010, respectively). Their values are in excess of those of the other species (0.004–0.005) even when randomly subsampled for equivalency ($n = 2$) (Table S2). As is consistent with the findings of Tay et al. (15), who documented several mitochondrial haplotypes from *H. armigera armigera* among Brazilian individuals, our autosomal value for the Brazilian hybrids is also high (0.009). This indicates that the incursion of *H. armigera* into Brazil had the potential to introduce considerable genetic variation via the generation of highly polymorphic hybrids.

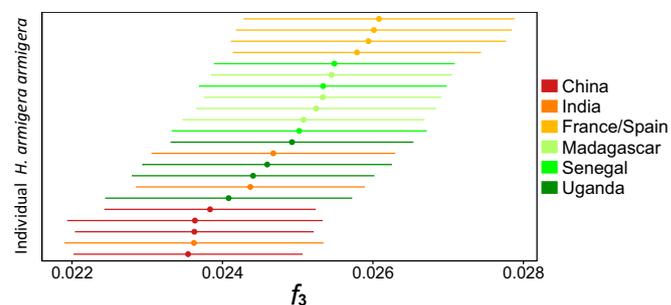**Fig. 2.** Outgroup f_3 statistic for the model f_3 (*H. armigera conferta*; Brazilian hybrids, x), demonstrating shared genetic drift between Brazilian hybrids and individual *H. armigera armigera* (x). Error bars represent SEs.
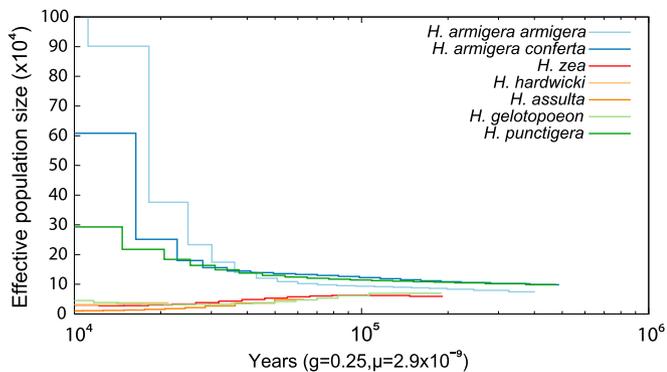
Anderson et al.

Fig. 3. Averages of 100 PSMC bootstrap estimates of effective population size for all species. The hybrids were not included in this analysis.

We then inferred the historical population dynamics for each of our taxa by applying PSMC software (30) to consensus genotype calls representing all samples for each of the *H. armigera* subspecies (excluding the Brazilian hybrids) and the five other species. We calculated that effective population size (Ne) was around 7,000–15,000 for both *H. armigera* subspecies and all four of the other species 100 kya (Fig. 3). Ne then appears to have grown substantially for both *H. armigera* subspecies and *H. punctigera*, with *H. armigera armigera* peaking at ~2,800,000 (Fig. S6). By contrast, Ne for all other species, including *H. zea*, declined over the same period, with all estimates for more recent populations having fallen below 5,000. It is possible that these latter estimates are biased downward because of their smaller sample sizes (31); therefore we subsampled the two *H. armigera* subspecies. While this did indeed reduce the Ne estimates for the most recent populations of the subspecies [~74,000 and 43,000 for *H. armigera armigera* (*n* = 6) and *H. armigera conferta* (*n* = 5), respectively], they remained considerably higher than those of the other species.

We conclude that *H. armigera* has had a large population capable of harboring considerable genetic diversity for many generations and that the size of the hybrid population subsequently formed has been sufficient to retain much of that variation.

Evidence of Selection Within *H. armigera*. Given the species' spectacular success in exploiting diverse agricultural systems, genes subject to recent selective sweeps in *H. armigera* would also be good candidates for variants of adaptive value in hybrids (32). We therefore used SweeD (33) to calculate the composite likelihood ratio (CLR) indicative of regions undergoing selective sweeps for all *H. armigera* samples (excluding the hybrids) (Fig. 4 and Table S3). Only a single sweep was observed in *H. armigera armigera* that was perhaps due to the geographically diverse nature of our samples for that subspecies, meaning only selection common to all the various geographies represented would likely be detected. Importantly, the region found to be swept, a 90-kb stretch containing seven genes half-way along chromosome 16, was also identified as having experienced a selective sweep in *H. armigera conferta*, suggesting it was of very general adaptive value for the species. Other regions found to be subject to selective sweeps in *H. armigera conferta* (Table S3) (which has a significantly smaller geographic range) include one containing a cytochrome P450 gene on chromosome 15 (*CYP337B3*) that is associated with resistance to the insecticide fenvalerate (9). This is not the first evidence for selection at the *CYP337B3* locus in *H. armigera*, with previous analyses demonstrating independently derived alleles that have risen to high frequencies in Asian and African populations of *H. armigera armigera* following historical exposure to fenvalerate (13). European *H. armigera armigera* contain both the African and Asian *CYP337B3* variants, and, notably in the current context, the Brazilian hybrids all possess the Asian variant of that gene (Table S4).

The Genetic Composition of the Hybrids. We used a version of the D statistic, f_D (34), to identify genomic regions that appear most *H. zea*-like in the Brazilian hybrids. Values of f_D calculated for 10-kb windows were plotted for each hybrid and unadmixed representatives of *H. armigera armigera* and *H. zea*. The data derived from each individual are plotted in Fig. 5, and data for some additional unadmixed individuals are shown in Fig. S7). Unadmixed individuals have f_D values around zero for *H. armigera armigera* (e.g., individual A in Fig. 5) or tending toward 0.8 for *H. zea* (e.g., individuals B and C, from Brazil and the United States, respectively). By contrast, the hybrids (individuals D–L) present highly variable patterns.

Eight of the nine hybrids (D–K) appear to be the products of initial hybridization between *H. armigera* and *H. zea* followed by varying degrees of backcrossing into predominantly *H. armigera* backgrounds. This results in f_D profiles across the genome that generally resemble that of *H. armigera* but are interspersed with more *H. zea*-like spikes. Although we find evidence of hybridization on all chromosomes, those bearing f_D spikes vary widely among the different hybrids, suggesting different histories of recombination and introgression following the initial hybridization event(s). All individuals except E have several relatively broad regions derived from *H. zea* (between three and nine, each peaking around f_D values of ~0.4), with some extending across hundreds of kilobases and even entire chromosomes (e.g., chromosome 14, 9.9 Mb in individual D of Fig. 5). The f_D profile across the genome of individual E differs, in which we observe more numerous but generally smaller spikes that are likely representative of a longer hybrid ancestry, providing more opportunity for recombination to break up blocks of *H. zea* DNA.

As each of these eight individuals appears to have undergone hybridization followed by several generations of backcrossing with *H. armigera*, we sought to determine the proportion of each of the eight genomes that was ancestrally derived from *H. zea*. We used 127,880 SNPs that strongly segregated between the two parent taxa to calculate the proportion of alleles derived from *H. zea*, which we found ranged from 2.0% in individual K to 8.9% in individual D. Multistratum permutation analysis of average chromosomal f_D demonstrated no significant difference (*P* > 0.05) among chromosomes in the proportions of *H. zea*-derived DNA; although the small sample size limits the sensitivity of this analysis, our results suggest there are no major differences among the chromosomes in the persistence of introgressed material since hybridization.

The genome of the ninth hybrid, individual L (Fig. 5), has significantly more *H. zea* DNA than would be expected in an F1 hybrid of *H. armigera* and *H. zea*, with 51.4% (*P* < 0.001) of its genome
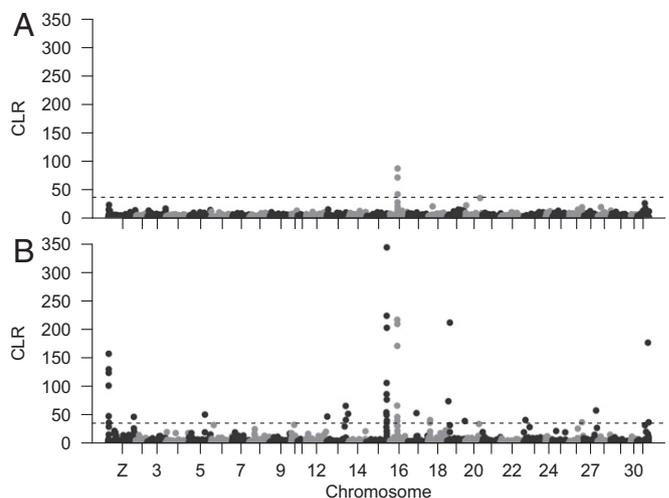

Fig. 4. CLR plots screening for selective sweeps in (*A*) *H. armigera armigera* and (*B*) *H. armigera conferta*, where the significance thresholds are 36.6 and 34.9, respectively, as shown by dashed lines.
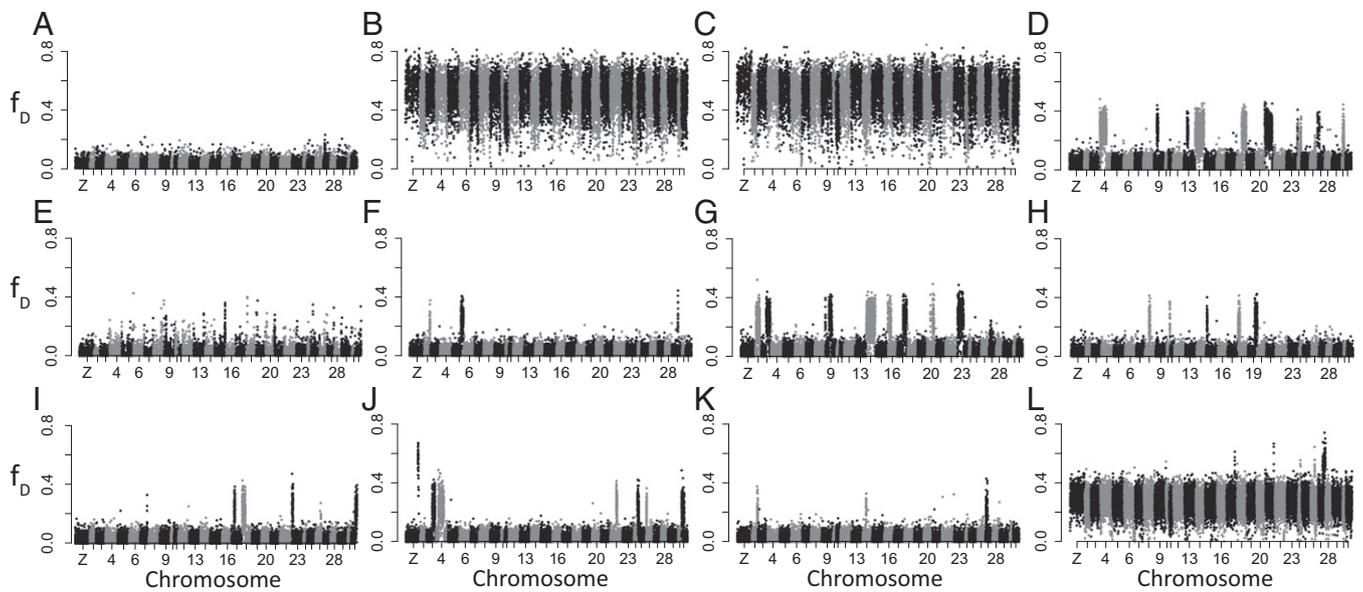
Anderson et al.

Fig. 5. Profiles of f_D values calculated for 10-kb nonoverlapping windows across the genome for individuals A–L. Individual A is a European *H. armigera*. Individuals B and C are *H. zea* from the United States and Brazil, respectively. Individuals D–L are Brazilian hybrids, and their profiles provide evidence for hybridization between *H. armigera* and *H. zea*. Individual-specific metadata, including mitochondrial genotypes, are listed in Table S4.

derived from the latter. Values of f_D generally fell approximately midway between the reference values for the two parent species; however, a few regions approached values seen in reference *H. zea*. A sliding-window plot of a hybrid index (HI) (35) confirmed that most of the genome comprised alleles from both parent taxa, with stretches clearly homozygous for *H. zea* DNA as well as some small but significant signals of *H. armigera* homozygosity (Fig. 6). A parsimonious interpretation of the pedigree of this individual would involve parents which were both hybrids, one with a majority of *H. zea* DNA and the other with a smaller majority of *H. armigera* DNA.

The regions of homozygosity in individual L (listed in Table S5) include tracts of *H. zea* DNA with significantly high HI values totaling about 7.75 Mb over several chromosomes, with the largest stretch occurring over 2.75 Mb of chromosome 27. A few other windows demonstrate significantly low HI, with the lowest occurring at the beginning of chromosome 20 (HI = 0.42), highlighting a short (~11 kb) stretch of homozygosity derived from *H. armigera*. This region is notable for containing genes encoding a clade 16 detoxifying esterase and two gustatory receptors that have been implicated in host use (20). HI was also low (0.45) where we observed the selective sweep on chromosome 16 in *H. armigera armigera*, although this was not considered statistically significant. The HI analysis also failed to identify significant levels of *H. armigera* homozygosity in the region encompassing the *CYP337B3* fenvalerate-resistance gene (HI = 0.48) which had been subject to a selective sweep *in H. armigera conferta*, but targeted analysis using Sanger sequencing of PCR products demonstrated that this individual was homozygous for this gene (Table S4).

As noted, eight of the nine hybrids were previously classified as *H. armigera* and the ninth as *H. zea* by Tay et al. (15), who sequenced mitochondrial DNA. Five other individuals also collected from Brazil in the same year, which Tay et al. (15) classified as *H. zea* from their mitochondrial sequences, all proved to have pure *H. zea* nuclear genomes in our analysis. By contrast, no Brazilian individuals that Tay et al. (15) had classified as *H. armigera* were found to have pure *H. armigera* nuclear genomes in our analysis. Overall, these data indicate that females of both parent species have given rise to fertile hybrid lineages, but they leave open the question of whether the mitochondria from the two parent species differ in their effects on the fitness of the hybrids.

Conclusions

The repeated introduction of one of the most significant agricultural pests in the Old World into South America is clearly cause for concern. *H. armigera* appears to be unlike the other *Helicoverpa* species in that it possesses relatively high genetic diversity, likely contributing to its very high adaptive and invasive capacities. A population with European origins is clearly part of an ongoing, multigenerational hybridization with *H. zea* in a region of the New World that contributes a significant proportion of the world's crops. It is too early to reliably detect signals of selection in the resulting hybrids, but clear evidence for the transfer of a wide range of *H. zea* genes into *H. armigera* backgrounds and *H. armigera* genes related to host use and pesticide resistance into an *H. zea* background is already apparent. Ongoing evolution in hybrid populations could generate new ecotypes in which features of the *H. zea* genome that may be important to adaptation in the New World are retained alongside genes from *H. armigera* that are capable of augmenting its properties as a pest.
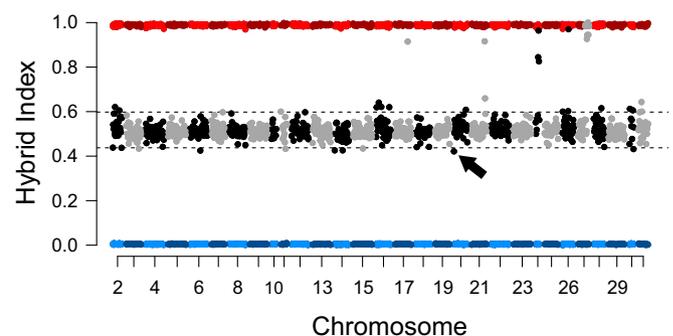
Fig. 6. HI across 250-kb windows for individual L, highlighting homozygosity indicating previous hybridization. SNPs segregating strongly between *H. armigera armigera* and *H. zea* were used to infer ancestry of individual L (black and gray), while calculations using all *H. armigera armigera* and *H. zea* (red and blue, respectively) are plotted for reference. Windows are considered significantly different from 0.5 where HI is above 0.598 or below 0.4375 ($P < 5 \times 10^{-7}$). The arrow highlights the lowest value of HI for individual L.

Anderson et al.

Future analyses of South American samples could provide insight into the progression of hybridization and potential for displacement of *H. zea* as a discrete species. Given the variation in introgressed regions observed, monitoring populations using genes that are likely under selection, such as *CYP337B3*, offer the highest throughput and most relevant way to follow the spread of *H. armigera* genotypes across the Americas. Given increased trade-associated movements of *Helicoverpa* around the world, we should also be prepared to see genetic variation originating from *H. zea* appearing in Old World *H. armigera*.

Methods

Sample Collection and DNA Extraction. Heliothine moths were collected between 2004 and 2014 from 16 countries and included 52 that had previously been classified as *H. armigera*, 10 previously classified as *H. zea*, seven *H. punctigera*, four *H. gelotopoeon*, two *H. assulta*, and one *H. hardwicki*. Those classified as *H. armigera* included 22 from Australia and New Zealand [conventionally designated as *H. armigera conferta* (22)], 22 from seven countries in Asia, Europe, and Africa [designated as *H. armigera armigera* (22)], plus eight collected from Brazil in 2013, which had been identified previously using mitochondrial markers (15) but which we show herein are actually hybrids with *H. zea*. The 10 *H. zea* moths included six collected from Brazil in 2013 identified as such from mitochondrial markers (15), one of which we show here is a hybrid with *H. armigera armigera*, plus four from the United States in 2004 that are considered to be geographically and temporally isolated from the current *H. armigera* invasion (15). There was no heteroplasmy in mitochondrial DNA among Brazilian hybrids, thereby eliminating the possibility that contamination contributed to any potential signals of hybridization (15). Sample collection data are listed in Table S4; all samples were preserved in ethanol/RNAlater or stored at −20 °C following collection. DNA was extracted using DNeasy blood and tissue kits (Qiagen) before quantification with a Qubit 2.0 fluorometer (Thermo Fisher Scientific).

Genome Resequencing and SNP Genotyping. Nextera (Illumina) libraries were produced following the manufacturer's instructions, and 100-bp paired-end reads were generated (Illumina HiSeq. 2000, Biological Resources Facility, Australian National University, Australia). Sample and sequencing data are included in Table S4. Raw sequence reads were aligned to the *H. armigera* genome using BBMap v. 33.43 (sourceforge.net/projects/bbmap/). Reads were trimmed when quality in at least two bases fell below Q10, and only uniquely aligning reads were included in the analysis. Resulting BAM files were sorted before duplicate reads were removed, and files were annotated using Picard v. 1.138 (broadinstitute.github.io/picard) before indexing with SAMtools v. 1.1.0 (36). UnifiedGenotyper in gatk v. 3.3-0 (37) was used to estimate genotypes across all individuals simultaneously, implementing a heterozygosity value of 0.01. VCFtools (38) was used to calculate mean coverage statistics for each sample.

Phylogenetic Analyses. A reduced SNP file consisting of 4D sites, which are essentially neutral (39), was used to derive phylogenetic and population structure end points. We annotated SNPs using SnpEff v. 4.0 (40) with default parameters to create a database before transferring annotations from the genome (20). Bases were then selected as in Martin et al. (41), where the first and second codon positions were invariable, providing us with 1,015,455 SNPs.

RAxML v. 8.2.10 (42) was used to plot phylogenetic trees for both the Z-chromosome and autosomes of all individuals, implementing a generalized time-reversible model of sequence evolution. The best topology was selected from 20 maximum likelihood searches, and its statistical support was assessed with 100 bootstrap pseudoreplicates. When only *H. armigera* and *H. zea* were used to construct a phylogeny, RAxML (42) was invoked using the -N autoMRE -f a -x options for autosomes and the Z chromosome, resulting in 600 and 804 rapid bootstrap searches, respectively, before a thorough maximum likelihood search.

Population Structure. We used markers from all chromosomes simultaneously for analysis of population structure in *H. armigera* and *H. zea*. EIGENSOFT v. 6.0.1 (43) was used to perform PCA. A Tracy–Widom distribution was used to infer statistical significance for principal components, with a threshold of 1×10^{-12}. The admixture software (27) was implemented using default parameters over K-values ranging from 1 to 10, with cross-validation enabled. Results were plotted using Distruct2.pl (www.crypticlineage.net/pages/distruct.html).

A more complete dataset was used for subsequent analyses, whereby phasing and imputation of missing bases was performed on all SNPs from all samples using default parameters in Beagle (44). Linkage disequilibrium (LD)-based pruning conducted using PLINK v. 1.07 (45) was used to filter one of a pair of SNPs using a pairwise LD threshold ($r^2 = 0.5$) within windows of 50 SNPs, moving forward five SNPs per iteration, resulting in 22,790,059 SNPs.

Selective sweeps in each *H. armigera* subspecies were detected by calculating CLRs from allele frequency data using SweeD (33). SweeD was run for each chromosome with a grid size of 1,000 blocks, and analyses were continued only where CLRs over 30 were observed. A significance threshold was calculated for both subspecies of *H. armigera* (13) using the inferred effective population size histories as detailed by the pairwise sequentially Markovian coalescent (PSMC) analyses as the basis of MSMS (v. 1.3) simulations (46), opting for 100 simulated datasets over 1 Mb. In addition, nonnegative, weighted F_{ST} was calculated across 10-kb windows (≥50 SNPs) in autosomes and the Z chromosome for *H. armigera* and *H. zea* using VCFtools v. 0.1.14 (38). *H. armigera* subspecies were randomly downsampled to $n = 9$. Average nucleotide diversity was also calculated for each species using VCFtools v. 0.1.14 (38), using both the full data and following random downsampling, where $n = 2$. The extent of variation between species was examined with the *t* test function in R v. 3.4.3 (47).

Demographic Inference. We used the PSMC software (30) to gain insights into historical differences in effective population size. The diploid consensus required by PSMC was generated from the mpileup command (-Q 20 -q 25) of SAMtools (36), followed by the bcftools call command for all samples from each species as well as subspecies of *H. armigera* (excluding the Brazilian hybrids). Vcfutils.pl (d 5 -D 34 -Q 25) was employed to prepare the data for PSMC. We used fq2psmcfa from the PSMC package on whole-genome consensus diploid sequences to create the input file before running PSMC (-N25 -t15 -r5) with the partition 6 + 2 × 4 + 3 + 13 × 2 + 3 + 2 × 4 + 6. Raw PSMC outputs were scaled to time and population sizes assuming a generation time of 0.25 y and a mutation rate of $2.9e^{-09}$ (48), plotting the average over 100 bootstraps.

Tests for Admixture. To evaluate the possibility of admixture across the *Helicoverpa* phylogeny, we calculated the f_3 statistic for all possible three-species combinations using AdmixTools v. 3.0 (49). Only trios where f_3 is negative are indicative of admixture between two source populations against a resulting target population. To populate a tree-like scenario for the estimation of gene flow between sister taxa, populations with more than two individuals were split randomly into two groups, and f_3 was calculated for all sister taxa and also for species that shared geographic distributions. All trios used in the analysis are listed in Dataset S1.

As our analyses suggest that the Brazilian hybrids are ancestrally derived from *H. armigera armigera*, an outgroup F3-test was conducted to determine which individual *H. armigera armigera* most resembled the hybrids. Under a specific tree-like scenario (outgroup; A, B), the expected value of f_3 will represent the genetic drift shared between populations A and B following divergence from the outgroup (50). In this instance, we used *H. armigera conferta* from Australia as the outgroup in tests for a shared history between the hybrids (A) and various *H. armigera armigera* individuals (B). The *H. armigera armigera* individuals generating the highest positive values of f_3 will therefore be those most related to the hybrids.

To identify introgressed genomic regions resulting from hybridization between *H. armigera armigera* and *H. zea* in the hybrid individuals, we used a version of Patterson's *D*. The *D*-statistic is a measure of admixture between populations that is robust to biases associated with SNP ascertainment and demographic history (49). Recent work by Martin et al. (34) improved the utility of the statistic for calculation across small, kilobase-level regions, terming this variant statistic "f_D." Our calculation of f_D implements a model of [*H. punctigera*, North American *H. zea*; *x*, *H. armigera armigera*] where *x* are hybrid individuals or *H. armigera conferta* and *H. zea* not otherwise included in the analysis. The model assumes there is no gene flow between the outgroup and *H. armigera armigera*, but gene flow is permitted between *x* and either *H. zea* or *H. armigera armigera*, which results in positive or negative f_D, respectively.

We used f_D to measure gene flow between individual moths and *H. zea* from United States. This statistic was calculated for 10-kb nonoverlapping windows using functions of the genomics.py script as used by Martin et al. (34) and plotted using a modified version of qqman (51). Only windows where the average ABBA and BABA counts were equal to or greater than two were considered. Significant differences between chromosomal averages for f_D were investigated using multistratum permutation testing in the R package lmPerm (52), invoking default settings.

We determined the proportion of the genome in each hybrid individual that was ancestrally derived from *H. zea* using SNPs that strongly segregated between that species and *H. armigera armigera*. These were determined via an association test in PLINK v1.90b (45), taking only SNPs where $P < 3.83 \times 10^{-14}$, thereby limiting the maximum number of alleles shared between the

two populations to the equivalent of one full homozygote or two heterozygotes. This resulted in a SNP set of 127,880 SNPs. We calculated the proportion of alleles from *H. zea* for each hybrid using a custom python script, considering the sum of *H. zea* alleles divided by all alleles in the analysis. For individual L, we then tested whether the estimated proportion of its genome derived from *H. zea* was significantly higher than that expected in an F1 hybrid using the variance of a binomial distribution, where $n = 127,880$.

We used a version of the HI (35) calculated using a custom python script to test for regions of homozygosity indicative of additional historical admixture events across the genome of individual L. HI calculations incorporated the panel of 127,880 segregating SNPs above, with average values for each SNP being inferred for *H. armigera armigera* and *H. zea*, which therefore achieved HI values of ~0 and 1, respectively. For windows consisting entirely of heterozygous SNPs, synonymous with an F1 hybrid, HI = 0.5. HI was calculated across 250-kb nonoverlapping sliding windows, considering only those windows where there were more than 20 SNPs. Sliding-window averages were plotted using a modified version of qqman (51). To infer regions

that varied significantly beyond neutral expectations, simulations were conducted using 112 SNPs (the average number contributing to a 250-kb window) randomly sampled from individual L. After 100,000 iterations, scores ranged from 0.598 to 0.438, and significance was calculated by dividing alpha (0.05) by the number of simulations with scores above or below these thresholds, respectively.

Hybrid individuals and *H. armigera* controls were screened for the presence of the *CYP337B3* gene as in Joussen et al. (9). Individual L was screened in triplicate, and heterozygote/homozygote status was determined through relevant band detection on 1.5–2% agarose gel followed by confirmation with Sanger sequencing.

1. Martin SH, et al. (2013) Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res* 23:1817–1828.
2. Sankararaman S, et al. (2014) The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* 507:354–357.
3. Martin SH, Jiggins CD (2017) Interpreting the genomic landscape of introgression. *Curr Opin Genet Dev* 47:69–74.
4. Zhang W, Dasmahapatra KK, Mallet J, Moreira GRP, Kronforst MR (2016) Genome-wide introgression among distantly related Heliconius butterfly species. *Genome Biol* 17:25.
5. Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487:94–98.
6. Czepak C, Albernaz KC, Vivan LM, Guimarães HO, Carvalhais T (2013) First reported occurrence of Helicoverpa armigera (Hubner) (Lepidoptera: Noctuidae) in Brazil. *Pesqui Agropecu Trop* 43:110–113.
7. Tay WT, et al. (2013) A brave new world for an old world pest: Helicoverpa armigera (Lepidoptera: Noctuidae) in Brazil. *PLoS One* 8:e80134.
8. Downes S, et al. (2017) A perspective on management of Helicoverpa armigera: Transgenic Bt cotton, IPM, and landscapes. *Pest Manag Sci* 73:485–492.
9. Joußen N, et al. (2012) Resistance of Australian Helicoverpa armigera to fenvalerate is due to the chimeric P450 enzyme CYP337B3. *Proc Natl Acad Sci USA* 109:15206–15211.
10. Feng H, Wu X, Wu B, Wu K (2009) Seasonal migration of Helicoverpa armigera (Lepidoptera: Noctuidae) over the Bohai Sea. *J Econ Entomol* 102:95–104.
11. Jones CM, et al. (2015) Genomewide transcriptional signatures of migratory flight activity in a globally invasive insect pest. *Mol Ecol* 24:4901–4911.
12. Fitt GP (1989) The ecology of Heliothis species in relation to agroecosystems. *Annu Rev Entomol* 34:17–52.
13. Anderson CJ, Tay WT, McGaughran A, Gordon K, Walsh TK (2016) Population structure and gene flow in the global pest, *Helicoverpa armigera*. *Mol Ecol* 25:5296–5311.
14. Sosa-Gómez DR, et al. (2016) Timeline and geographical distribution of Helicoverpa armigera (Hübner) (Lepidoptera, Noctuidae: Heliothinae) in Brazil. *Rev Bras Entomol* 60:101–104.
15. Tay WT, et al. (2017) Mitochondrial DNA and trade data support multiple origins of Helicoverpa armigera (Lepidoptera, Noctuidae) in Brazil. *Sci Rep* 7:45302.
16. Mastrangelo T, et al. (2014) Detection and genetic diversity of a heliothine invader (Lepidoptera: Noctuidae) from north and northeast of Brazil. *J Econ Entomol* 107:970–980.
17. Lopes-da-Silva M, Sanches MM, Stancioli AR, Alves G, Sugayama R (2014) The role of natural and human-mediated pathways for invasive agricultural pests: A historical analysis of cases from Brazil. *Agric Sci* 5:634–646.
18. Kriticos DJ, et al. (2015) The potential distribution of invading Helicoverpa armigera in North America: Is it just a matter of time? *PLoS One* 10:e0119618.
19. Mallet J, Korman A, Heckel DG, King P (1993) Biochemical genetics of Heliothis and Helicoverpa (Lepidoptera: Noctuidae) and evidence for a founder event in Helicoverpa zea. *Ann Entomol Soc Am* 86:189–197.
20. Pearce S, 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 Biol* 15:63, and erratum (2017) 15:69.
21. Leite NA, et al. (2017) Pan-American similarities in genetic structures of helicoverpa armigera and helicoverpa zea (lepidoptera: Noctuidae) with implications for hybridization. *Environ Entomol* 46:1024–1034.
22. Matthews M (1999) *Heliothine Moths of Australia. A Guide to Pest Bollworms and Related Noctuid Groups* (CSIRO Publishing, Collingwood, Australia).
23. Hardwick DF (1965) The corn earworm complex. *Mem Entomol Soc Can* 97:5–247.
24. Laster M, Sheng C (1995) Search for hybrid sterility for Helicoverpa zea in crosses between the North-American Heliothis zea and Helicoverpa armigera (Lepidoptera, Noctuidae) from China. *J Econ Entomol* 88:1288–1291.
25. Murúa MG, et al. (2016) Species from the Heliothinae complex (Lepidoptera: Noctuidae) in Tucumán, Argentina, an update of geographical distribution of Helicoverpa armigera. *J Insect Sci* 16:61.
26. Cho S, et al. (2008) Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. *Syst Entomol* 33:581–594.
27. Alexander DH, Lange K (2011) Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* 12:246.
28. Kronforst MR, et al. (2013) Hybridization reveals the evolving genomic architecture of speciation. *Cell Rep* 5:666–677.
29. Leite NA, Alves-Pereira A, Corrêa AS, Zucchi MI, Omoto C (2014) Demographics and genetic variability of the new world bollworm (Helicoverpa zea) and the old world bollworm (Helicoverpa armigera) in Brazil. *PLoS One* 9:e113286.
30. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475:493–496.
31. Gattepaille L, Günther T, Jakobsson M (2016) Inferring past effective population size from distributions of coalescent times. *Genetics* 204:1191–1206.
32. Norris LC, et al. (2015) Adaptive introgression in an African malaria mosquito co-incident with the increased usage of insecticide-treated bed nets. *Proc Natl Acad Sci USA* 112:815–820.
33. Pavlidis P, Živkovic D, Stamatakis A, Alachiotis N (2013) SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol* 30:2224–2234.
34. Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol Biol Evol* 32:244–257.
35. Goulet BE, Roda F, Hopkins R (2017) Hybridization in plants: Old ideas, new techniques. *Plant Physiol* 173:65–78.
36. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
37. McKenna A, et al. (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303.
38. Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
39. Künstner A, Nabholz B, Ellegren H (2011) Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. *Genome Biol Evol* 3:1381–1389.
40. Cingolani P, et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92.
41. Martin SH, et al. (2016) Natural selection and genetic diversity in the butterfly Heliconius melpomene. *Genetics* 203:525–541.
42. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
43. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2:e190.
44. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* 81:1084–1097.
45. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.
46. Ewing G, Hermisson J (2010) MSMS: A coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26:2064–2065.
47. R Core Team (2014) R: A Language and Environment for Statistical Computing (R Found Stat Comput, Vienna).
48. Keightley PD, et al. (2015) Estimation of the spontaneous mutation rate in Heliconius melpomene. *Mol Biol Evol* 32:239–243.
49. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192:1065–1093.
50. Sankararaman S, Patterson N, Li H, Pääbo S, Reich D (2012) The date of interbreeding between Neandertals and modern humans. *PLoS Genet* 8:e1002947.
51. Turner S (2017) qqman: Q-Q and Manhattan Plots for GWAS Data, R Package Version 0.1.4. Available at https://cran.r-project.org/web/packages/qqman/. Accessed March 21, 2018.
52. Wheeler B, Torchiano M (2016) lmPerm: Permutation Tests for Linear Models, R Package Version 2.1.0. Available at https://cran.r-project.org/web/packages/lmPerm/index.html. Accessed March 21, 2018.