Entomologia
Experimentalis et Applicata

DOI: 10.1111/eea.12215

Population structure of *Spodoptera frugiperda* maize and rice host forms in South America: are they host strains?

M.L. Juárez[1,2]*, G. Schöfl[3], M.T. Vera[1,2], J.C. Vilardi[2,4], M.G. Murúa[2,5], E. Willink[5], S. Hänniger[6], D.G. Heckel[6] & A.T. Groot[6,7]

[1]*Cátedra de Terapéutica Vegetal, Facultad de Agronomía y Zootecnia, Universidad Nacional de Tucumán, Florentino Ameghino s/n, 4000 Tucumán, Argentina,* [2]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CCT-Tucumán, 4000 Tucumán, Argentina,* [3]*Department of Cell- and Molecular Biology, Leibniz Institute for Natural Product Research and Infection Biology, Beutenbergstrasse, 11[a], 07745 Jena, Germany,* [4]*Lab. Genética de Poblaciones Aplicada, Depto. Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires – IEGEBA (CONICET), Intendente Güiraldes 2160, C1428EGA, Buenos Aires, Argentina,* [5]*Sección Zoología Agrícola, Estación Experimental Agroindustrial Obispo Colombres, Las Talitas, T4104AUD, Tucumán, Argentina,* [6]*Department of Entomology, Max Planck Institute for Chemical Ecology, Hans-Knöll Strasse 8, 07745 Jena, Germany, and* [7]*Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands*

Accepted: 10 June 2014

Key words: fall armyworm, host fidelity, AFLP, mtCOI, triose phosphate isomerase, Lepidoptera, Noctuidae, host-associated differentiation

Abstract

Determining which factors contribute to the formation and maintenance of genetic divergence to evaluate their relative importance as a cause of biological differentiation is among the major challenges in evolutionary biology. In *Spodoptera frugiperda* (Smith) (Lepidoptera: Noctuidae) two host strains have been recognized in the 1980s: the corn-strain prefers maize, sorghum, and cotton, whereas the rice-strain prefers rice and wild grasses. However, it is not clear to what extent these so-called 'strains', which have also been called 'host races' or even 'sibling species', are really associated with host plants. Due to the indeterminate evolutionary status, we will use the term 'host forms' (sensu Funk). Here, we characterized populations collected from maize, rice, and wild grasses from three countries in South America. Using two mitochondrial cytochrome oxidase I (mtCOI) markers and 10 polymorphisms in the triose phosphate isomerase (*Tpi*) gene, we found various patterns of host association. Two hundred twenty-seven nuclear amplified fragment length polymorphisms (AFLPs) markers revealed significant genetic differentiation among populations, which was generally correlated to the host from which the larvae were collected. Using a multivariate discriminant analysis and a Bayesian clustering approach, we found that individuals could be grouped into 2–5 genetically distinct clusters, depending on the method. Together, our results indicate that although host-associated differentiation is present in this species, it does not account for all observable genetic variation and other factors must be maintaining genetic differentiation between these forms. Therefore, the term 'host strains' should be abandoned and 'host forms' should be used instead for *S. frugiperda*.

Introduction

Phytophagous insect species often show a population-specific preference for only a few host plant species. This choice and adaptation to a reduced number of host plants may cause reproductive isolation in which the final outcome can be the formation of new species (Walsh, 1864; Bush, 1969; Schluter, 2001; Funk et al., 2006; Feder et al., 2012). One intermediate stage between polymorphic populations and full species along the speciation process is host races. According to Drès & Mallet (2002), host races can be defined as genetically differentiated sympatric populations that are incompletely reproductively isolated with an appreciable rate of gene flow; they exhibit host fidelity by the use of different host taxa in the wild, display a

*Correspondence: M.L. Juárez, Cátedra de Terapéutica Vegetal, Facultad de Agronomía y Zootecnia, Universidad Nacional de Tucumán, Florentino Ameghino s/n, 4000 Tucumán, Argentina. E-mail: lau_zoo@yahoo.com.ar

correlation between host choice and mate choice, and have higher fitness on natal than alternative hosts (Drès & Mallet, 2002). After this definition, several papers have been published reporting the existence of host races in many insect species with few clear examples in which their existence has been recognized; e.g., the apple maggot fly *Rhagoletis pomonella* (Walsh) (Walsh, 1867; Bush, 1969; Feder et al., 1994, 2012), the larch budmoth *Zeiraphera diniana* (Guenée) (Emelianov et al., 1995; Drès & Mallet, 2002), and the leaf beetle *Neochlamisus bebbianae* (Brown) (Funk, 1998, 2012). Recently, Funk (2012) stated that this number is relatively low because the evidence in many biological systems is still inconclusive given the extensive amount of work required to determine whether a certain organism meets the defined criteria. In this sense, Funk emphasizes that it is necessary to introduce terms aimed at describing different kinds of biological variation in entities in which the existence of host-associated differentiation has been proven but its evolutionary status has not yet been determined. One such term is 'host form' which consists of 'a group of individuals or populations exhibiting host-associated biological variation in which the kind of variation has not yet been diagnosed' (Funk, 2012). Determining which factors contribute to the formation and maintenance of genetic divergence to evaluate their relative importance as a cause of biological differentiation is among the major challenges in evolutionary biology (Feder et al., 1988; Berlocher & Feder, 2002; Egan et al., 2008).

The noctuid moth *Spodoptera frugiperda* (Smith) (Lepidoptera: Noctuidae) exemplifies this problem. This species seems to be under a process of ecological divergence in sympatry due to host-associated differentiation, as two so-called 'host strains' have been recognized in the 1980s, which are morphologically indistinguishable but show some genetic differentiation in association with different host plants (Pashley et al., 1985; Pashley, 1986). Larvae of the so-called corn-strain (C) infest maize (*Zea mays* L.), sorghum [*Sorghum bicolor* (L.) Moench subsp. *bicolor*], and cotton (*Gossypium hirsutum* L.) and have been associated to large grasses, whereas larvae of the so-called rice-strain (R) are found mostly on small grasses as rice (*Oryza sativa* L.) and wild grasses, such as Johnson grass [*Sorghum halepense* (L.) Pers.] and Bermuda grass [*Cynodon dactylon* (L.)] (Pashley, 1986, 1988a). Although the term 'strain' or even 'race' and 'sibling species' have been widely used in the literature on this species (Pashley & Martin, 1987; Pashley, 1986, 1988a; Whitford et al., 1988; Pashley et al., 1995; Drès & Mallet, 2002; Prowell et al., 2004; Meagher et al., 2011; Schöfl et al., 2009, 2011), here we will follow Funk (2012) due to the yet indeterminate evolutionary status and use the term 'host form' instead of 'host strain'. These two host forms exhibit some degree of reproductive isolation, including (1) ecological isolation caused by differential use of host plants (Pashley, 1986, 1988a; Prowell et al., 2004), larval performance differences (Pencoe & Martin, 1981; Pashley, 1988b; Whitford et al., 1988; Pashley et al., 1995), and oviposition preference (Whitford et al., 1988; Meagher et al., 2011); (2) temporal isolation caused by temporal partitioning of nocturnal mating activities (Pashley et al., 1992; Schöfl et al., 2009); (3) female-mediated differential mating preferences (Schöfl et al., 2011); and (4) potential sexual isolation caused by differences in the composition of female sex pheromones (Groot et al., 2008; Lima & McNeil, 2009) and directionally biased incompatibility and low viability in hybrids (Pashley & Martin, 1987; Whitford et al., 1988; Groot et al., 2010).

The two host forms of *S. frugiperda* can be identified by a number of genetic markers. These markers include differences in mtDNA sequences identified in the cytochrome oxidase I (COI), and NADH dehydrogenase genes (Pashley, 1989; Lu & Adang, 1996; Levy et al., 2002; Nagoshi et al., 2006a), as well as nuclear DNA differences, including restriction length fragment polymorphisms (RFLPs) (Lu et al., 1992), amplified fragment length polymorphisms (AFLPs) (McMichael & Prowell, 1999; Busato et al., 2004; Clark et al., 2007; Martinelli et al., 2007; Belay et al., 2012), polymorphisms in tandem-repeat sequences (FRs) (Lu et al., 1994; Nagoshi & Meagher, 2003a,b), and 10 polymorphisms in the sex-linked triose phosphate isomerase gene (*Tpi*) (Nagoshi, 2010). Although restriction site polymorphisms in COI have been widely accepted to be the most suitable to characterize populations and assess host association with their respective plants, recently 10 SNPs in the *Tpi* gene have been proposed to be more consistent than COI for these purposes (Nagoshi, 2012).

Irrespective of the markers, the studies described above on this species show that host association is not always absolute, ca. 80% of individuals collected from maize habitats belong to the corn-form, whereas ca. 85–90% of larvae collected from rice habitats belong to the rice-form. In South American populations there seem to be some differences, as we recently found no consistent pattern of host association between the two forms and their respective host plants when using two restriction site polymorphisms in COI (Juárez et al., 2012). The combined use of mitochondrial and nuclear markers which have different inheritance mechanisms allows inferring the rates and directionality of hybridization. In using this combination, about 16% of field-collected samples from Louisiana, Florida (both USA), Puerto Rico, Guadeloupe, and French Guiana (Prowell et al., 2004) were found to be potential hybrids due to discordance for at least one marker

(mtDNA, esterase, and AFLP), with both types of hybrids (RC and CR; first letter always referring to the female) equally frequent, mostly in maize habitats. Similar findings were found with Colombian populations using COI gene and FR-sequence (Saldamando & Vélez-Arango, 2010). Others found mainly RC-hybrids (Nagoshi & Meagher, 2003b; Nagoshi et al., 2006b; Nagoshi, 2012). Most of the work published using molecular markers has been used to identify both forms and assess their host specificity, but provide little information about the genetic diversity and population structure of *S. frugiperda*, with a few exceptions (McMichael & Prowell, 1999; Busato et al., 2004; Clark et al., 2007; Belay et al., 2012).

Thus, even though the two host forms of the fall armyworm have been considered as host races or even as sibling species, it is not clear whether these forms are associated with specific host plants along the entire range of their distribution or whether there is a constant level of genetic differentiation between populations from different host plants. Therefore, in this study we characterized populations of *S. frugiperda* in the southern limit of its distribution obtained from different hosts, to determine whether these host forms can be considered host races. To do so, we analyzed the combination of two genetic markers that are generally used to distinguish the two strains (mtCOI and *Tpi* genes) and we studied the genetic structure among the various populations using 227 AFLP markers.

Materials and methods

Insect collection

Fall armyworm larvae were collected from three hosts at six localities from Argentina, one from Brazil, and two from Paraguay (Figure 1). The sampling design aimed at sampling two regions: the eastern region comprising Northeast Argentina, Paraguay, and Southern Brazil, and the western region comprising Northwest Argentina. Within these two regions, one of the characteristic hosts of each form was chosen. In the eastern region, larvae were collected from maize and rice. These two crops are widely cultivated next to each other in extensive areas. In the western region, rice is not cultivated and for this reason the alternative hosts sampled for the rice-form were Bermuda grass and Guinea grass [*Panicum maximum* (Jacq)], that grow spontaneously in the surroundings of maize plantations. Collections took place during November (spring) to February (summer) from 2007 to 2010. Each population was assigned a code denoting the host plant of each form (i.e., C for maize and R for rice or wild grasses), the year of collection, and the region, as detailed in Table 1.

In a given field, ca. 30 sites with 10 plants each were sampled randomly. To avoid any homogenization effect, at least 250 larvae were collected (one per plant) and placed individually in glass tubes (12 cm high, 1.5 cm diameter) with leaves of the host plant. Larvae were taken to the laboratory and reared in chambers at 27 ± 2 °C, 70–75% r.h., and L14:D10 photoperiod until adult emergence. Late instars and adults were examined to confirm that all individuals were fall armyworm based on diagnostic taxonomic characters. Populations from each sampled host in each locality were maintained separately and 200 adults were used from each population to establish laboratory colonies. In separate mating cages (30 cm high, 10-cm-diameter cylindrical polyethylene-terephthalate cages with nylon mesh cloth) 4–5 females of <24 h old and 4–5 males were introduced. We had in total about 20 mating cages per population. The cages contained pieces of paper that allowed females to rest and to lay eggs. Food was provided via a cotton plug saturated with a 1:1 (vol: vol) mixture of honey and water, which was renewed every day. Cages were checked daily for oviposition and adult mortality. To minimize loss of genetic variability, once females started to lay egg masses, ca. 15 egg masses from each cage were collected and deposited in glass tubes (12 cm high, 1.5 cm diameter). Once emerged, 15 neonate larvae from each of the egg masses were placed individually

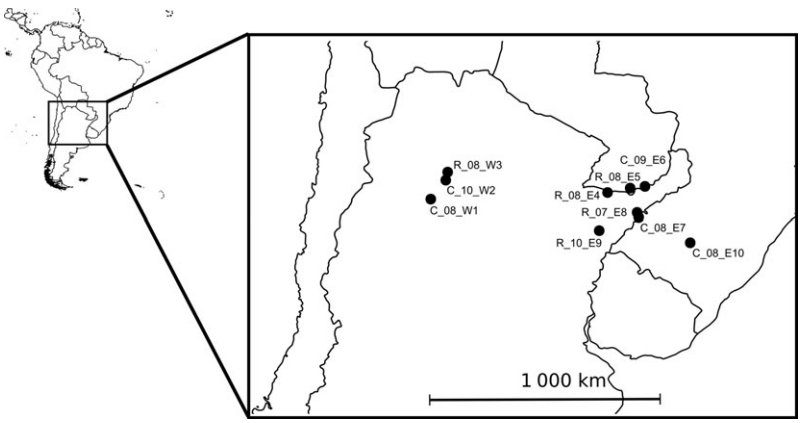

Figure 1 *Spodoptera frugiperda* sampling sites in Argentina, Brazil, and Paraguay.

Table 1 Collection sites, years, and host plants of *Spodoptera frugiperda*

Region	Country	Site	Longitude	Latitude	Host plant	Year	Population code
Western	Argentina	La Cocha	65°34′47.04′W	27°46′25.38′S	Maize	2008	C_08_W1
		Los Pereyra	64°53′36.9′W	26°55′09.0′S	Maize	2010	C_10_W2
		Benjamín Aráoz	64°48′26.79′W	26°33′28.64′S	Grass	2008	R_08_W3
Eastern	Argentina	Berón de Astrada	57°29′53.90′W	27°28′35.01′S	Rice	2008	R_08_E4
	Paraguay	San Cosme y Damián	56°27′46.1′W	27°16′43.8′S	Rice	2008	R_08_E5
		Capitán Miranda	55°47′12.4′W	27°12′02.5′S	Maize	2009	C_09_E6
	Argentina	Santo Tomé	56°04′26.8′W	28°34′58.3′S	Maize	2008	C_08_E7
		Santo Tomé	56°08′47.1′W	28°22′12.9′S	Rice	2007	R_07_E8
		Mercedes	57°52′42.4′W	29°10′23.1′S	Rice	2010	R_10_E9
	Brazil	Santa María	53°43′05.2′W	29°43′10.4′S	Maize	2008	C_08_E10

Population code: first letter denotes the host plant the larvae were collected from [C for maize (corn), R for rice and wild grasses]; number refers to year of collection; second letter denotes the geographic location of population (eastern or western); the number behind it is the population number. Bermuda grass and Guinea grass are referred to as grass.

in glass tubes with artificial diet (Osores et al., 1982), which was renewed every 2–3 days. As larvae pupated, they were placed in cylindrical cages until adult emergence. On average, 200 adults were used again to initiate a new generation. After establishing a colony from each population and host, larvae from the second generation were stored at −20 °C until DNA extraction.

DNA extraction and identification by COI markers

Total DNA was extracted using a modification of Black & DuTeau (1997) CTAB (hexadecyltri-methylammonium bromide) method. Buffer and running conditions were performed according to Sambrook et al. (1989). All samples were characterized using two mtCOI markers by amplifying a 600-bp fragment and digesting separately with *Msp*I (producing 510- and 90-bp fragments in the corn-form) and *Sac*I (producing 450- and 150-bp fragments in the rice-form) (Juárez et al., 2012). Populations were considered to belong to one of the two host forms if the frequency of the corresponding haplotype was above 80%. If the frequency was between 0.8 and 0.2, we characterized the population as a mixture of haplotypes (Juárez et al., 2012).

Characterization of fall armyworm host form by *Tpi* polymorphisms

Identification of two haplotypes of the fall armyworm was performed using the 10 polymorphic nucleotide sites (SNPs) located in the Z-linked *Tpi* gene as described by Nagoshi (2010). Primers *Tpi*-282F (5-GGTGAAATCTC CCCTGCTATG-3) and *Tpi*-850gR (5-AATTTTA TTACCTGCTGTGG-3) (Nagoshi, 2010) were synthesized by Metabion (Martinsried, Germany). PCR amplicons from genomic DNA, generated using these primers, were sequenced from both ends using the same primers in separate reactions. Sequencing was performed at the Entomology Department of the Max Planck Institute for Chemical Ecology (Jena, Germany). The DNA sequences were aligned and compared using the program Geneious Pro 5.4.3 (Biomatters, Auckland, New Zealand) (Drummond et al., 2011).

As previously described, each of the 10 sites has a specific nucleotide associated with each host form, making it possible to obtain a consensus sequence for the corn-form (*Tpi*-C) and the rice-form (*Tpi*-R). We followed the criterion proposed by Nagoshi (2010), in which at least seven of the 10 sites must match to the consensus *Tpi*-C or *Tpi*-R sequence in order to identify an individual as corn- or rice-form, respectively. Nagoshi (2010) defined all other configurations of the 10 sites as intermediate, *Tpi*-int. Due to sex-linkage, all females carry only one *Tpi* allele and can be classified in this way. However, males carry two *Tpi* alleles, and therefore can be homozygous or heterozygous. Nagoshi (2010) classified homozygous males in the same manner as females, but only classified heterozygous males if both of their *Tpi* alleles were of the same strain category, e.g., *Tpi*-C or *Tpi*-R with three or fewer double peaks indicating heterozygous SNPs. *Tpi*-C/*Tpi*-R heterozygotes were not distinguished from *Tpi*-C/*Tpi*-int, *Tpi*-R/*Tpi*-int, or *Tpi*-int/*Tpi*-int heterozygotes, and none of these heterozygotes were included in the analysis of Nagoshi (2010). Here, we scored all 10 of the polymorphic SNPs, because we sequenced all amplicons from both ends, and so we distinguished among genotypic classes in the following way. C, R, and IHo refer to individuals hemizygous (females) or homozygous (males) for a *Tpi*-C, *Tpi*-R, or *Tpi*-int sequence, respectively; i.e., with no double peaks in the sequencing chromatogram. CHe and RHe refer to individuals heterozygous for two different *Tpi*-C or *Tpi*-R sequences, respectively; i.e., showing double peaks at one, two, or three sites, but matching the consensus *Tpi*-C or

Tpi-R sequences at the other sites. IHe (heterozygous intermediates) include all other heterozygous classes. Our analysis included all these classes.

To consider the possibility of hybridization, we denote the COI and *Tpi* types of individuals by a configuration code in which the first letter represents the COI haplotype (C or R) and the rest represents the *Tpi* type as defined above. For example, the configuration C/RHo has a COI haplotype of C, and is hemizygous or homozygous for *Tpi*-R.

Genome-wide random nuclear markers

AFLP markers were developed following Vos et al. (1995) with some modifications. Genomic DNA (200 ng) was digested with restriction enzymes, *Eco*RI (5 U) and *Mse*I (3 U) in a 12.5-μl reaction mix. *Eco*RI adapter (5 pmol μl^{-1}) and *Mse*I adapter (50 pmol μl^{-1}) were ligated to generate template DNA for the amplification of DNA fragments by PCR. The adapters had the following sequences: *Eco*RI adapter: 5′-CTCGTAGACTGCG TACC, 5′-AATTGGTACGCAGTCTAC, and *Mse*I adapter: 5′-GACGATGAGTCCTGAG, 5′-TACTCAGGACT CAT) (Metabion). After the pre-amplification step, the selective amplifications were conducted using 11 primer combinations (Table 2). Two 96-well gels were used for each primer combination, where the samples of one population were equally divided among the two gels, as well as within the gels. Twelve individuals were represented on both plates. In this way, we included a total of 177 individuals in the analysis. For visualization in the polyacrylamide gel, all *Eco*RI primers were labeled with an infrared dye (IRD) of 700 or 800 nm. AFLP fragments were separated based on size with a Li-Cor 4300 DNA analyzer that simultaneously detects infrared DNA fragments of 700 and 800 nm. The samples were run on a 6.5% polyacrylamide gel and loaded into 96 wells with a Hamilton syringe (Hamilton, Reno, NV, USA). A labeled standard (Li-Cor STR marker, 50–700 bp) was loaded in the first and last well of each gel (1–100). We scored the gels using image analysis software AFLP-Quantar Pro 1.0 (KeyGene Products, Wageningen, The Netherlands). AFLP markers were identified by scoring the presence (1, indicating the dominant homozygote or the heterozygote) or absence (0, indicating the recessive homozygote) of the bands for every selective primer combination in each gel. The repeated 12 individuals were used to indicate the same markers on both gels. Only those markers that were scored consistently on both gels were used for subsequent analysis.

Table 2 AFLP primer combinations used to evaluate genetic diversity in *Spodoptera frugiperda*

Primer combinations	
Mse	*Eco*R
AAG	AAG700/ACG800
AAG	ACC700/ACT800
ACA	AAG700/ACG800
ACA	ACC700/ACT800
ACG	ACC700/ACT800
CGA	AAG700/ACG800
CGA	ACC700/ACT800
ACA	TAC700/GTA800
CAT	AAG700/ACG800
CAT	ACC700/ACT800
CTT	ACC700/ACT800

Genetic diversity and genetic structure

To assess whether the available loci allow for an acceptable precision for genetic analyses, the software BOOTSIE (https://code.google.com/p/bootsie/) was used to calculate the coefficient of variation for genetic distances across 100 bootstrap samples for a decreasing number of loci. Population genetic parameters were estimated based on the AFLP markers using the program AFLPsurv 1.0 (Vekemans, 2002). We used two criteria to define populations and assign individuals to each population: (1) 10 populations were defined depending on their origin (i.e., based on sampling site, year of collection, and host plant from which the larvae were collected), and (2) 21 populations were defined depending on their origin (as above) and the combination of mitochondrial (COI) and nuclear (*Tpi* genes) genotypes obtained with both markers. To estimate allele frequencies, a Bayesian method with non-uniform prior distribution (Zhivotovsky, 1999) was used. The parameters of genetic diversity and population genetic structure estimated were: total gene diversity (Ht), average gene diversity within populations (Hw), average gene diversity among populations (Hb), and Wright's F_{ST}. Parameters were estimated using the approach of Lynch & Milligan (1994) and assuming Hardy-Weinberg equilibrium. Pairwise Wright's fixation indices (F_{ST}) and pairwise Nei's distances were used to estimate the genetic differentiation and distance between populations. To test the significance level of genetic differentiation among populations, a permutation test using 2 000 replications was performed. Based on pairwise F_{ST} values, a phenogram representing genetic differentiation between populations was reconstructed using the *bionj* neighbor-joining algorithm (Gascuel, 1997) and visualized using the R package APE (Paradis et al., 2004). To infer bootstrap confidence on tree nodes, Neighbor and Consense procedures from the PHYLIP software ver. 3.6 were used (Felsenstein, 2005). Consensus was obtained

using the 'Majority rule' option, from 1 000 matrices of pairwise F_{ST} generated by AFLPsurv. Nodes are considered well supported if they occur in at least 500 (50%) bootstrap tree reconstructions.

The significance of the correlation between geographic distance and genetic distance matrices was estimated using the Mantel test implemented in the R package ADE4 (Chessel et al., 2004). The P-value for the Mantel coefficient r was obtained after performing 2 000 permutations.

To determine the presence of outlier-F_{ST} loci, we performed two sets of analyses using the software MCHEZA (Antao & Beaumont, 2011) with 50 000 simulations. The first set of analyses was made to detect outliers that contribute to geographic and host differentiation among all populations. For that, first we considered the 10 populations defined by their origin and then we considered nine populations excluding the population from grasses. The second set of analyses was made to detect outliers that contribute to differentiation between hosts only (i.e., excluding geographic differentiation). For that, first we pooled all 10 populations from the same host (maize or rice/grass) and then we pooled the nine populations without the grass population (i.e., maize or rice). Once all outlier loci were identified, we estimated the population genetic parameters from only neutral loci using the program AFLP-SURV 1.0 and performed the same two hierarchical analysis of variance (ANOVA) components (see above).

Hierarchical analysis of population structure

We performed two hierarchical ANOVA components (Wright, 1978) using the HIERFSTAT package (Goudet, 2006) from the statistical software R (R Development Core Team, 2012). In the first analysis, the estimation of hierarchical variance components considered four levels: populations (defined by their origin as shown in Table 1), regions (western and eastern), populations within each region, and individuals within each population. In the second analysis, the hierarchical components considered populations (defined by their origin as shown in Table 1), host plant species (maize, rice, and wild grasses), populations within each host plant species, and individuals within each population. Thus, in the first case the highest hierarchical level tested was the geographic distance, whereas in the second case this level was represented by the host plant from which individuals were collected.

Probabilistic analysis of population structure

To identify the level of clustering of genetically related individuals we applied two approaches: an exploratory multivariate method and a model-based Bayesian method.

Discriminant analysis of principal components. Discriminant analysis of principal components (DAPC) (Jombart et al., 2010) combines the multivariate principal component analysis with a discriminant analysis and makes no assumptions about Hardy-Weinberg equilibrium or linkage disequilibrium. The analysis was performed using the R package ADEGENET (Jombart, 2008). The clustering of individuals was determined without prior information on population groupings using the function 'find.clusters', which runs successive K-means clustering with increasing number of clusters (k) to achieve the optimal number of groups (Jombart et al., 2010). The optimal number was based on the minimum value of the Bayesian information criterion (BIC). The association of individuals in clusters and the correspondence between clusters and their original populations was shown by means of a scatter plot of individuals on the first two components of the DAPC, where the grouping factor was defined by the clusters recognized by 'find.clusters'. Each individual was identified by a color key of the sampling population. This scatter plot was obtained with the function 's.class' of the package ADE4 of R (Dray & Dufour, 2007). The reliability of the results was corroborated by comparing the a priori assignment with the a posteriori assignment of each individual.

Bayesian clustering analysis. Bayesian approaches to genotypic clustering of individuals typically use explicit population genetic models to sort individuals into clusters such that deviations from equilibrium within clusters are minimized. We estimated the number of clusters and the assignment of individuals into clusters without prior information on population groupings using the methods implemented in the programs STRUCTURE 2.3.1 (Pritchard et al., 2000; Falush et al., 2003) and STRUCTURAMA (Huelsenbeck & Andolfatto, 2007).

Using STRUCTURE, the most likely number of clusters is estimated by determining the change in the marginal likelihood of the data Pr (X|K) when the number of clusters (K) is fixed to different values (K = 1, 2,. . ., 10). We used an ancestry model that allowed for admixture and correlated allele frequencies between populations. Under this model, individuals are fractionally assigned to clusters using a membership coefficient. We ran eight replicate Markov chains with a burn-in period of 200 000 iterations followed by a sampling period of 800 000 iterations for each K. We also used the ΔK method of Evanno et al. (2005) to detect the amount of structuring beyond which a further subdivision does not substantially improve the fit of the admixture model. An individual was assigned to the cluster for which it had the highest average membership coefficient across runs, after 'label switching' heterogeneity

had been accounted for using the software CLUMPP (v. 1.1.1) (Jakobsson & Rosenberg, 2007).

Using STRUCTURAMA, the number of clusters and the assignment of individuals to clusters were estimated simultaneously by applying a Dirichlet process prior, which treats both the assignment of individuals to populations and the number of populations as random variables. STRUCTURAMA implements the basic no-admixture model of STRUCTURE and additionally allows setting the concentration parameter α of the Dirichlet process prior (which shapes the prior probability of the number of clusters) by specifying the prior mean of the number of clusters. We performed eight analyses, varying the prior mean of the number of clusters from two to nine. Each analysis consisted of a single Markov chain run for 2 000 000 cycles. Samples were drawn from the chain every 100th cycle. The first 10 000 of the resulting 20 000 samples were removed as burn-in prior to analysis. The posterior probabilities of the number of populations given the data Pr(K|X) were averaged across runs. At each step in the MCMC chain, each individual was assigned to a cluster. To summarize the results of this partitioning of individuals, the partition that minimizes the squared distance to all sampled partitions during an MCMC-run was calculated and reported (mean partition). The distance measure is the number of individuals that must be deleted between two partitions to make them the same (Huelsenbeck & Andolfatto, 2007).

Results

Coefficient of variation of AFLP markers

The relationship between the coefficient of variation and the number of AFLP markers allows determining the robustness of genetic variability estimates. The coefficient of variation calculated for all 227 markers was 4.8%, indicating that this number of markers was sufficient to perform unbiased analyses of genetic structure and diversity (Figure S1).

Identification of the corn and rice genotypes with the Z-linked nuclear marker *Tpi*

The association between *Tpi* genotypes and the host plant from which larvae were collected was more pronounced for populations from maize and grasses than for those collected from rice. The analysis identified homozygous and heterozygous individuals for *Tpi*-C (C and CHe), *Tpi*-R (R and RHe), and intermediate (IHo and IHe) types (Tables 3 and S1–S11; Figures 2 and S2). In three of the five populations collected from maize fields, more than 80% of the individuals could be classified as *Tpi*-C; of the two remaining populations, one also showed a high percentage of *Tpi*-C individuals (75%) whereas the other population showed a broader distribution of *Tpi* types with 52, 4, and 44% of the individuals belonging to *Tpi*-C, *Tpi*-R, and intermediate, respectively. In the population collected from grasses, 86% of the individuals were *Tpi*-R, consistent with the preference of this type for this habitat. In contrast, none of the four populations collected from rice were characterized as *Tpi*-R. In two of these populations, the *Tpi*-C type was predominant whereas in the other two, intermediate types prevailed.

Correlation between *Tpi* and COI markers and their association with host plants

The combined analysis using both molecular markers showed the following patterns (Figures 2 and S2; Tables S1–S11).

Concordance between Tpi *and COI markers with association to the host plant.* In three of the 10 populations, the *Tpi* types were largely consistent with both the previously characterized mitochondrial haplotypes and the host plants from which larvae were obtained. Two of these

Table 3 Number and percentage (in parenthesis) of *Tpi* genotypes in populations of *Spodoptera frugiperda* collected from Argentina, Brazil, and Paraguay

Population	n	*Tpi*-C	*Tpi*-CHe	*Tpi*-R	*Tpi*-RHe	*Tpi*-IHe	*Tpi*-IHo
C_08_W1	25	7 (28)	6 (24)	0	1 (4)	10 (40)	1 (4)
C_10_W2	15	9 (60)	6 (40)	0	0	0	0
R_08_W3	14	0	0	7 (50)	5 (36)	2 (14)	0
R_08_E4	21	0	1 (5)	4 (20)	2 (9)	12 (57)	2 (9)
R_08_E5	19	2 (10)	0	3 (16)	1 (5)	13 (70)	0
C_09_E6	19	4 (21)	15 (79)	0	0	0	0
C_08_E7	20	6 (30)	13 (65)	0	0	1 (5)	0
R_07_E8	15	13 (86)	1 (7)	1 (7)	0	0	0
R_10_E9	5	1 (20)	3 (60)	0	0	1 (20)	0
C_08_E10	16	2 (12)	10 (63)	0	0	4 (25)	0

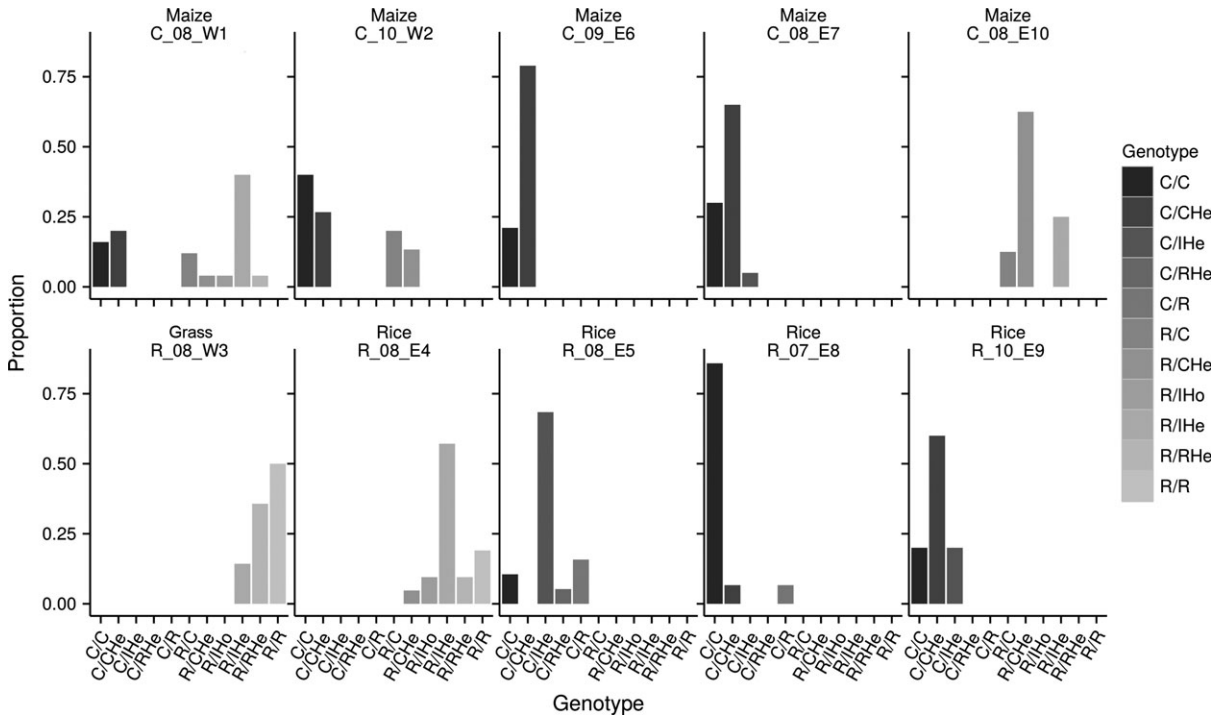

Figure 2 Proportion of different configurations of COI and *Tpi* types in the populations of *Spodoptera frugiperda* collected from Argentina, Brazil, and Paraguay.

populations were collected from maize (C_09_E6 and C_08_E7) with 100 and 95% individuals characterized as corn-form (C/C and C/CHe) by both markers, and one population was collected from grasses (R_08_W3) where 86% of the individuals were characterized as rice-form (R/R and R/RHe) by both markers.

Concordance between Tpi *and COI markers without association to the host plant.* In two of the 10 populations, the *Tpi* types were largely consistent with the mitochondrial haplotypes, but not with the host plants. These populations were collected from rice (R_07_E8, R_10_E9), with 93 and 80% of individuals identified as corn-form (C/C and C/CHe), respectively.

Discordance between Tpi *and COI markers.* In the remaining five populations, the *Tpi* types were not consistent with the mitochondrial haplotypes and indicated different configurations. In C_08_W1, there were 36% COI-C and 64% COI-R types, and 44 and 16% of individuals had an R/I (R/IHo and R/IHe) and R/C (R/C and R/CHe) configuration, respectively (Figure 2). Similar results were found for C_10_W2 from maize, with 67% COI-C and 33% COI-R types, and 33% of individuals having the RC (R/C and R/CHe) configuration. In C_08_E10 from maize, 100% of larvae analyzed carried COI-R, whereas 75% of them had the *Tpi*-C (C and CHe) and 25% had the intermediate (IHe) type. The remaining two populations were collected from rice. In R_08_E4, 100% of larvae analyzed were COI-R, with 28% having *Tpi*-R (R/R and R/RHe), 67% as intermediate (R/IHe and R/IHo), and 5% with *Tpi*-C (R/CHe). Population R_08_E5 showed the opposite pattern, with 100% of larvae as COI-C and only 11% *Tpi*-C (C/C), 68% intermediate (C/IHe), and 21% *Tpi*-R (C/R and C/RHe) (Figure 2).

Taken together, of the 169 individuals analyzed from the 10 populations, 46% were determined as corn-form with both markers (C/C and C/CHe), 11% as rice-form with both markers (R/R), whereas the remaining 43% showed discordance between the mitochondrial and nuclear markers, C/R (3%), C/I (9%), R/C (13%), and R/I (18%).

Genetic diversity and population structure

We scored a total of 227 genomic AFLP markers, 215 (94%) of which were polymorphic and were thus used for the population genetic analysis.

Populations defined according to their origin. The overall gene diversity (Ht) was 0.28, with the highest component represented by the within-population diversity

(Hw = 0.20) and a relatively low among-populations diversity (Hb = 0.084). The overall genetic differentiation among populations was highly significant (F_{ST} = 0.31, P<0.0001; Table 5). The most differentiated population compared to all others was the population obtained from wild grasses (R_08_W3). When this population was excluded from the analysis, F_{ST} and Hb decreased to 0.18 and 0.045, respectively. Thus, roughly half of the total genetic differentiation among populations was due to the differentiation between the samples from wild grasses and all other populations. The differentiation between populations still remained significant after the samples from wild grasses were removed (P<0.0001; Table 5).

The lowest genetic distance between populations as estimated by Nei's distance and F_{ST} coefficients (Table S12) was observed between C_08_E7 and C_09_E6 (both from maize). The largest genetic distance was observed between R_08_W3 and R_07_E8 (from wild grasses and rice, respectively). Correlation analysis using a Mantel test showed an absence of significant isolation by distance between *S. frugiperda* populations (r = 0.148, P = 0.31) (Figure S3).

A neighbor-joining tree derived from F_{ST}-values showed no support for clustering of populations by geographic region (Figure 3A). It did, however, indicate a separation of populations by host plant. Samples collected from rice or maize plants were clearly separated from the wild grass population, which had the highest bootstrap support.

Populations defined according to their origin and the combination of mitochondrial and nuclear types. The overall gene diversity (Ht = 0.29), was mainly represented by the within-population diversity (Hw = 0.20), with relatively low between-population diversity (Hb = 0.09). The overall genetic differentiation among populations was highly significant (F_{ST} = 0.31, P<0.0001). After excluding the diverged samples collected from wild grass, F_{ST} and Hb decreased to 0.21 and 0.06, respectively, but F_{ST} remained significant. The genetic distances between populations as estimated by Nei's distance and F_{ST} coefficients are shown in Table S13. The lowest genetic distance was observed between C_08_W1 (C/C) and C_08_W1 (C/CHe), R_08_W3 (R/R) and R_08_W3 (R/RHe), C_10_W2 (C/CHe) and C_10_W2 (C/C), C_08_E7 (C/C) and C_08_E7 (C/CHe), C_09_E6 (C/CHe) and C_09_E6 (C/C), C_08_W1 (R/C) and C_08_W1 (R/IHe), all these from the same collection site. The largest genetic distance was observed between R_08_W3 (R/RHe) and R_07_E8 (C/C) (Nei's distance), and R_08_W3 (R/R) and C_08_W1 (C/C) (F_{ST}), respectively.

A neighbor-joining tree derived from F_{ST}-values (Figure 3B) showed that most populations grouped together on the basis of their geographic origin and not on the basis of their genotype composition. The only exception was the individuals from C_08_W1, which formed two clearly separated groups based on their COI haplotype. Individuals carrying the COI-C haplotype clustered with the rest of the samples derived from maize fields, whereas individuals carrying the COI-R haplotype formed a cluster relatively basal in the tree indicating a very distant position from all other groups. The sample C_08_W1 thus seems to be comprised of individuals from two genetically diverged populations. This may explain why C_08_W1 is not clustered with the remaining populations collected from maize

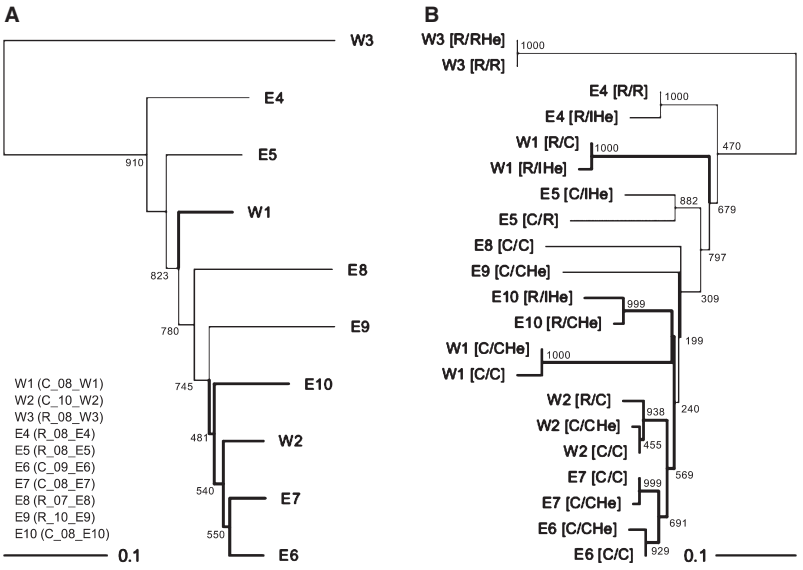

Figure 3 Neighbor-joining trees, (A) based on F_{ST} values calculated from AFLP markers, from *Spodoptera frugiperda* populations defined according to their origin: sampling site, year of collection, and host plant, and (B) with populations defined by their origin and combination of mitochondrial and nuclear haplotypes. Populations collected from maize are indicated by black branches, populations collected from rice and grasses by gray branches.

in Figure 3A. In addition, high bootstrap support was found between the populations R_08_W3 (R/RHe) and R_08_W3 (R/R), R_08_E4 (R/R) and R_08_E4 (R/IHe), C_08_W1 (R/C) and C_08_W1 (R/IHe), and between C_08_W1 (C/Che) and C_08_W1 (C/C) populations.

Analysis to detect the presence of outlier-F_{ST} loci that contribute to geographic and host differentiation among populations (first set of analyses) showed that, considering the 10 populations together, no outliers were found but when we considered the nine populations, six of the 227 loci were identified as outliers and thus candidates to be under positive selection (Figure S4AB). In this case, the overall genetic differentiation among populations was 0.19 ($P<0.0001$). The hierarchical analysis revealed significant differences among regions and hosts, whereas the differentiation among populations within each region and host was highly significant (Tables S14–S16). No outlier-F_{ST} loci were found when all populations from the same host were considered, including or not the wild grass population (second analysis) (Figure S5AB).

Hierarchical analysis of population structure

A hierarchical ANOVA component considering populations, regions, populations within regions, and individuals within populations revealed no significant differences between regions; in contrast, the differentiation among populations and populations within each region was highly significant (Table 4). The hierarchical analysis considering populations, host plant species, populations within each host plant species, and individuals within populations revealed borderline significant differences among host species and highly significant differences among populations and among populations within host species (Table 5).

Table 4 Hierarchical analysis of *Spodoptera frugiperda* populations among populations, among regions, among populations within regions, and among individuals within each population

	Variance	F	95% confidence interval	P
Populations – total	33.707	0.416	0.387–0.445	<0.0005
Regions – total	1.040	0.013	−0.005 to 0.031	0.46
Populations – regions	32.667	0.408	0.381–0.437	<0.0005
Individuals – populations	47.384			

'Regions' denotes eastern (incl. R_08_E4, R_08_E5, C_09_E6, C_08_E7, R_07_E8, R_10_E9, C_08_E10) or western (incl. C_08_W1, C_10_W2, R_08_W3) geographic location of populations (see also Table 1).

Table 5 Hierarchical analysis of *Spodoptera frugiperda* populations among populations, among host plants, among populations within host plants, and among individuals within each population

	Variance	F	95% confidence interval	P
Populations – total	39.663	0.456	0.426–0.487	<0.0005
Hosts – total	1808.641	0.208	0.172–0.244	0.041
Populations – hosts	21.577	0.313	0.287–0.344	0.001
Individuals – populations	47.384			

'Hosts' denotes the plant species from which populations were sampled: maize (incl. C_08_W1, C_10_W2, C_09_E6, C_08_E7, C_08_E10), rice (incl. R_08_E4, R_08_E5, R_07_E8, R_10_E9), and grasses (incl. R_08_W3) (see also Table 1).

Probabilistic analysis of population structure

Discriminant analysis of principal components by 'find.clusters' function. Applying the BIC, the total sample was best divided into 10 groups (Figure 4). Mostly, these groups were composed of individuals from the same locality and sampling year. Major exceptions were groups A and K, which were composed of individuals from C_08_W1; group A was composed of individuals carrying the C/C and C/CHe configurations, whereas group K was composed of individuals with R/RHe, R/C, R/CHe, R/Iho, and R/IHe configurations. Group I was composed of all individuals from C_08_E7 and R_10_E9, which carried the C/C, C/CHe, and C/IHe configuration, and by some individuals from C_09_E6 and R_08_E5 with C/C, C/CHe, C/R, and CIHe configurations. In addition, DAPC indicated five clusters of groups (gray ellipses in Figure 4). The largest cluster was composed of five groups: G (R_07_E8), D (C_10_W2), H (most individuals from R_08_E5), I, and F (most individuals from C_09_E6). All these groups were represented by individuals collected from maize and rice and carrying the C/C, C/CHe, and C/IHe configurations. The cluster represented by group J (R_08_E4) was very close to the former cluster and included individuals collected from rice and carrying mostly the R/IHe, R/R, and R/RHe configurations. A third cluster was formed by group A (C_08_W1), composed of individuals carrying the C/C and C/CHe configurations, and group B (C_08_E10), composed of individuals carrying the R/C, R/CHe, and R/IHe configurations; both were collected from maize and exhibited a slight overlap.

The more isolated clusters are group L, consisting of R_08_W3, including individuals from wild grasses carrying R/R, R/RHe, and R/IHe configurations, and group K (C_08_W1), that includes individuals from maize carrying the R/C, R/CHe, R/RHe, R/IHo, and R/IHe configurations. The distribution of the groups represented here seems to constitute a hierarchical islands model (Jombart et al., 2010). The agreement comparing the prior and posterior assignments was 84%.

Bayesian analysis. The Bayesian inference of structural patterns among the individuals gave no consistent results. The minimal number of genetic clusters necessary to explain the data as suggested by Evanno's ΔK was 2 (Figure S6). The posterior probability of the number of clusters derived from STRUCTURAMA was highest at K = 4 (Figure S6). Average log-likelihoods across 10 replicate STRUCTURE runs showed no marked plateau before K = 8 [Figure S6; empirical evidence suggests that a biologically meaningful number of K may be indicated by a declining rate of increase in Pr(X|K) as K increases, rather than by the absolute maximum likelihood; Pritchard et al., 2000; Evanno et al., 2005]. Figure 5 presents the assignments of individuals to different clusters by both programs for K≤4. With K≥4, assignments became increasingly inconsistent across replicate runs and hence difficult to interpret and summarize. At K = 2, R_08_W3 (from wild grasses) consistently formed one cluster, whereas all other samples were joined in a second cluster (Figure 5). At K = 3, a third cluster was split off, comprising R_08_E4, R_08_E5 (both from rice) and most individuals from C_08_W1 (from maize). At K = 4, four alternative clusters were observed across 10 runs, all of which introduced a split among populations R_08_E4, R_08_E5, R_07_E8, and C_08_W1 (Figure 5). These are most populations collected from rice and the genotypically mixed population C_08_W1. When K was treated as a random variable (using STRUCTURAMA), the overall patterns were largely similar. However, as noted previously (Groot et al., 2011), STRUCTURAMA tends to introduce additional populations comprised of only very few individuals, which often lack a biologically meaningful interpretation. Thus, despite STRUCTURAMA detecting between four and six clusters as the most likely number of K across multiple runs (compare Figure 5), the vast majority of individuals were assigned to either two or three clusters (Figure 5). As with STRUCTURE, if two major clusters were inferred, one cluster contained all individuals from R_08_W3, whereas most other individuals were placed in the second cluster. If three major clusters were inferred, the third comprised R_08_E4, R_08_E5, R_07_E8, and C_08_W1 in varying combinations (Figure 5).

Discussion

The present study aimed to determine whether *S. frugiperda* is under a process of speciation through host associated differentiation, to support or reject the existence of host races. We characterized different South-American

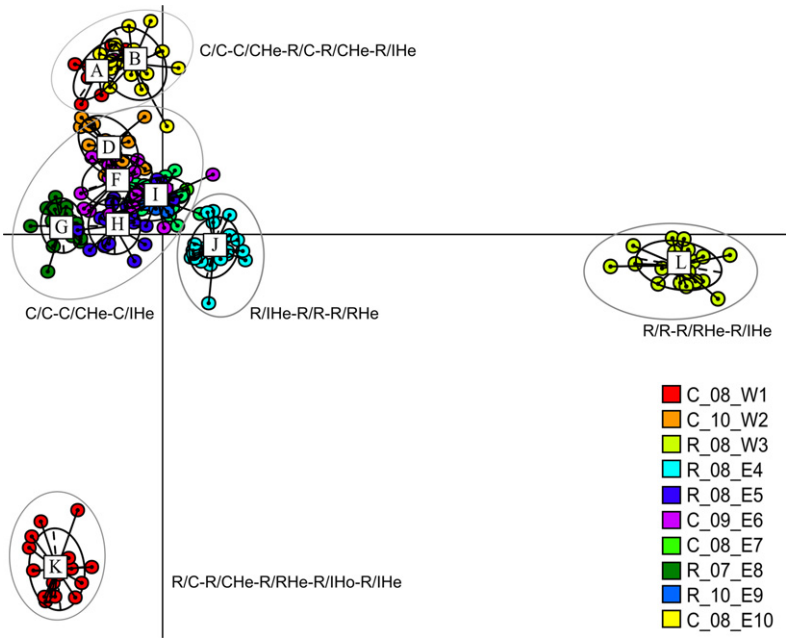

Figure 4 Population analysis by discriminant analysis of principal components (DAPC). Groups of individuals were identified by the '*find.clusters*' function without prior information on population groupings: A (C_08_W1), J (R_08_E4), L (R_08_W3), G (R_07_E8), B (C_08_E10), K (C_08_W1), D (C_10_W2), H (R_08_E5), I (including all individuals from C_08_E7, all individuals from R_10_E9, and some individuals from R_08_E5 and C_09_E6), and F (C_09_E6).

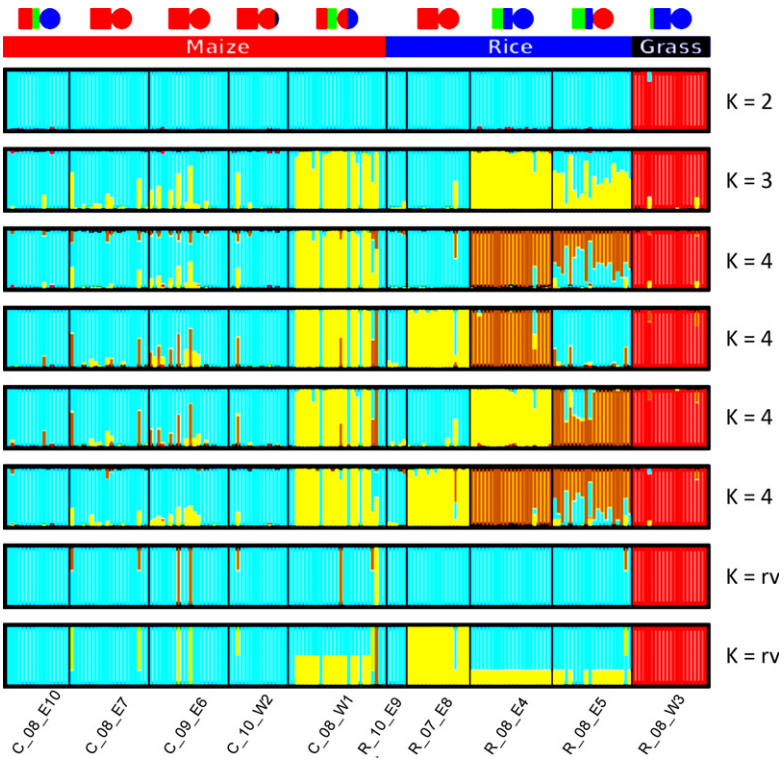

Figure 5 Estimated population structure of *Spodoptera frugiperda* populations. The analysis with STRUCTURE was performed under an admixture model with the number of clusters (K) fixed to different values (K = 2, 3, and 4). The analysis with STRUCTURAMA was performed under a no-admixture model where the assignment of individuals to populations and the number of populations as random variables (rv).

populations with the Z-linked nuclear marker *Tpi* and 227 nuclear AFLP markers to complement a study by Juárez et al. (2012) that used two restriction site polymorphisms in the mtDNA COI. In this study, we found a different pattern of association between the host species from which the populations were sampled and the two molecular markers, COI and *Tpi*, and highly significant genetic variability, with strong genetic differentiation of some populations.

Population characterization and the association with host plants

Populations collected from maize and wild grasses mostly showed the expected *Tpi* genotypes. This result is in agreement with Nagoshi (2012), who found that this marker is an accurate indicator and therefore should be considered as the most appropriate marker developed so far to assign host form identity. However, the combined analysis of *Tpi* and COI revealed that only two out of the five populations collected from maize were characterized as belonging to the corn-form (i.e., *Tpi* and COI correlated). The other three showed various combinations of the corn-, rice-, or intermediate forms of the markers, indicating some sort of hybridization. Because the individuals we sequenced come from the second generation of laboratory-maintained colonies founded by field-collected individuals, we cannot determine how much of this hybridization occurred in the field, and how much is due to crossing in the laboratory of different, non-hybrid forms collected from the same field. However, the large difference in COI haplotype frequencies and *Tpi* allele frequencies in three of the populations indicates that some hybridization must be occurring in the field, because these frequency differences cannot have arisen within two generations. In any event, this means that *Tpi* marker alone is not sufficient to characterize the populations as it is unable by itself to provide information on possible mixed genomes. Populations collected from rice revealed an unexpected situation. None of the populations were characterized by both markers as rice-form, and two populations (R_07_E8 and R_10_E9) were genetically identified as corn-form (both by COI and *Tpi*).

The high frequency of corn-form individuals in rice plants is unexpected; this result is unaffected by the fact that the analyzed individuals come from the second generation of a field collection. Previous studies found that corn-form individuals seemed restricted to their host plant (Lu & Adang, 1996; Nagoshi & Meagher, 2003b, 2004; Vélez-Arango et al., 2008) and this could be attributed to the low levels of toxic cyanogenic compounds found in this plant compared to wild grasses (Hay-Roe et al., 2011). It could be argued that rice is also less toxic than wild grasses and hence this host can be exploited by corn-form individuals which lack the capacity to cope with high levels

of these toxic compounds (Hay-Roe et al., 2011). The remaining samples derived from two populations collected from rice were composed mostly of individuals bearing intermediate *Tpi* types. In one of these populations the haplotype COI-R was predominant, whereas in the other the haplotype COI-C prevailed. This revealed a more complex pattern and the presence of individuals with mixed genomes. In the populations we sampled overall, rice occurs as a host in which the pure rice-form of *S. frugiperda* was almost absent and in which pure corn-form individuals and individuals with mixed genomes can develop.

Genetic variability and population structure

We found a high genetic diversity between the sampled populations, which could be clustered into 2–5 genetically distinct groupings. This poses the question of what maintains this genetic differentiation: geographical distance, host fidelity, or strain identity, and how they are related. The highest amount of diversity was found among individuals and the correlation analysis confirmed no genetic isolation by geographic distance between *S. frugiperda* populations, which is in agreement with Martinelli et al. (2007), Clark et al. (2007), and Belay et al. (2012). The ANOVA components revealed no differences between eastern and western regions and showed a marginally significant differentiation between host plants. In the neighbor-joining trees based on F_{ST}-values, populations grouped together mainly based on their geographic origin and not on their haplotype composition or region (Figure 3B). To a lesser extent, there was also an association based on the host. There were, however, two main exceptions. One was R_08_W3, collected from wild grass, which never grouped with any population collected from rice and appeared in a separate branch showing that the least amount of gene flow is between this population and the rest of populations. The other was C_08_W1, where individuals formed two clearly separated groups; one included those individuals bearing the COI-R haplotype and the other those bearing the COI-C haplotype irrespective of their *Tpi*. The same pattern was observed with the DAPC and Bayesian clustering methods. It can be assumed that COI-R individuals from the C_08_W1 population were recent immigrants. In this region (Western Argentina) rice is not cultivated and hence, the migrant individuals possibly have derived from the surrounding grasses (C_08_W1 and R_08_W3 are only 160 km apart). However, the neighbor-joining analysis and the DAPC failed to merge these individuals, revealing the need of more sampling to determine whether wild grasses act as a reservoir of pure rice individuals or even of other genetically isolated populations. DAPC grouped the individuals mostly based on the site and year of sampling. Additionally, the groups with C-mitochondrial haplotype tended to cluster, whereas most of the groups with the R-mitochondrial haplotype were more isolated. STRUCTURE and STRUCTURAMA also showed that the populations collected from maize clustered and were homogeneous. Similar to the DAPC analysis, populations collected from rice formed three distinct groups, indicating a much higher level of heterogeneity. The outlier analysis showed that they did not contribute significantly to the geographical and host differences and their removal did not alter the conclusions compared to the analysis that included the total number of loci.

A possible cause of bias in our study was the introduction of the populations into the laboratory, which might have caused changes in the distribution of allele frequencies. To minimize this problem, we maximized the number of adults from which we started the colonies and sampled an equal number of larvae from egg masses and mating cages to reduce any skew in reproductive success. In addition, if we consider that the mtCOI is maternally transmitted and does not recombine, a significant change in allele frequencies is not expected within two generations of laboratory rearing. The *Tpi* marker is sex-linked and may recombine only in males, but, due to the close linkage between SNPs, a significant number of intermediate patterns by recombination in only one or two generations is highly unlikely. Therefore, we assume that the *Tpi* frequencies have also not changed significantly during two generations in the laboratory. The other aspect of sampling second-generation laboratory populations is a loss of information about the naturally occurring frequency of hybrids. From samples with large discrepancies in frequencies of mitochondrial and nuclear markers, we can infer that some hybridization is occurring in the field, but we cannot estimate its frequency. Our study and the work done by Nagoshi (2012) infer hybridization values greater than 40% with individuals from the laboratory, whereas the hybridization rates found by Prowell et al. (2004) are near 16%. As the study of Prowell et al. (2004) is based on field-collected samples, hybridization rates of 16% are likely to be an accurate estimation.

Evolutionary and ecological implications

Our results provide additional information for understanding the population structure and the host-associated differentiation in the two host forms of *S. frugiperda*. In this study, the utility of the host as indicator of population identity was variable as shown by the different molecular markers. The genotypes identified by *Tpi* revealed high frequency of populations matching their respective hosts, and the AFLP analysis showed that populations collected

from the same host tended to be more associated, but this was not confirmed by the mtCOI markers. In addition, the majority of individuals from populations collected from rice were more heterogeneous than individuals collected from maize.

The near absence of pure rice-form populations on rice is interesting, given that previous reports indicate that this form is predominant in ca. 85–90% of the collections from rice or wild grasses and only in ca. 20% of the collections from maize (Lu & Adang, 1996; McMichael & Prowell, 1999; Nagoshi & Meagher, 2003b, 2004; Machado et al., 2008; Vélez-Arango et al., 2008). Most studies on the rice-form have focused on its association with wild grasses. These findings, together with the physiological and biological evidence from laboratory studies showing the capacity of both strains to develop equally well on maize and rice plants (Pashley et al., 1995; Meagher et al., 2004; Groot et al., 2010), suggest that host plants do not exert the same selective pressure toward differential host use. It also suggests that other factors are relevant in the process of host shift. For example, Pashley et al. (1995) reported that over 2 years of sampling, *S. frugiperda* larval mortality caused by parasites and pathogens was higher in pasture than in maize fields, suggesting that the maize habitat may constitute a more protected environment compared to rice habitats. The unexpected high frequency of corn individuals in rice raises doubt whether host-driven selection can still create population divergence even when host fidelity is weak.

The high genetic variability that we found within and among populations may have arisen from new genotypic combinations, providing a high capacity to adapt to changes in agricultural environments with an evolutionary potential (Domingues, 2011). Another possible explanation is that divergent genotypes are not of recent origin and they existed and still exist on native grasses (grass-form) and some simply expanded their host ranges into maize and rice with the domestication and introduction of these crops in America. In some areas of the distribution of *S. frugiperda*, this has resulted in *Tpi* and COI haplotype patterns that correlate with host plant, but in other places these markers do not show a consistent pattern. The greater homogeneity found in the populations collected from maize, and the greater heterogeneity found in the populations collected from rice, suggests that the rice-form is the ancestral type.

Conclusions

Reiterating the definition of host races (Drès & Mallet, 2002), our study sheds light on whether the two forms of *S. frugiperda* can be considered host races. The two forms use different host taxa in the wild, at least to some extent; the two forms consist of a number of individuals that exhibit host fidelity; the two forms coexist in sympatry in at least part of their distribution range; and they are genetically differentiated at more than one locus. However, we found that the two forms are not always more genetically differentiated from populations on another host in sympatry than from some geographically distant populations on the same host. Our study did not examine whether the two forms display a correlation between host choice and mate choice, but we did find that they undergo gene flow at an appreciable rate. Previous studies have shown the two forms do not have higher fitness on natal than on alternative hosts, or more accurately, inconsistent results were found across studies (Pashley, 1988b; Whitford et al., 1988; Pashley et al., 1995; Meagher et al., 2004; Stuhl et al., 2008; Groot et al., 2010). However, hybrid incompatibility has been shown for RC hybrid females, which mate at much lower rates and few matings result in fertile egg clutches (Pashley & Martin, 1987; Whitford et al., 1988; Groot et al., 2010). Overall, we have insufficient information to conclude that the two forms are true host races, and thus 'host form' (Funk, 2012) is the appropriate terminology at this stage of our knowledge. Our results indicate that although host-associated differentiation is confirmed as one of the diverging mechanisms, this species is composed of genetically distinct entities that are most likely diverging due to (additional) factors other than host specialization.

Acknowledgements

We are grateful to Sofía Torres, Carmen Urueña, and Alejandro Vera at EEAOC for their collaborations in sampling and technical support. We also thank Antje Schmaltz, Susanne Donnerhacke, and Melanie Unbehend in the Department of Entomology, MPICE, for their technical support and assistance in the lab. We are grateful to Fernando Navarro (Facultad de Ciencias Naturales e Instituto Miguel Lillo, Universidad Nacional de Tucumán) for the taxonomic identification of *S. frugiperda* larvae. This research was partly funded by the National Science Foundation (award IOS-1052238) and by the Max-Planck-Gesellschaft.

References

Antao T & Beaumont MA (2011) Mcheza: a workbench to detect selection using dominant markers. Bioinformatics 27: 1717–1718.

Belay DK, Clark P, Skoda SR, Isenhour DJ, Molina-Ochoa J et al. (2012) Spatial genetic variation among *Spodoptera frugiperda*

(Lepidoptera: Noctuidae) sampled from the United States, Puerto Rico, Panama, and Argentina. Annals of the Entomological Society of America 105: 359–367.

Berlocher SH & Feder JL (2002) Sympatric speciation in phytophagous insect: moving beyond controversy? Annual Review of Entomology 47: 773–815.

Black WC & DuTeau NM (1997) RAPD-PCR and SSCP analysis for insect population genetic studies. The Molecular Biology of Insect Disease Vectors: A Methods Manual (ed. by JM Crampton, CB Beard & C Louis), pp. 361–373. Chapman & Hall, London, UK.

Busato GR, Grützmacher AD, De Oliveira AC, Vieira EA, Zimmer PD et al. (2004) Análise da estrutura e diversidade molecular de populações de *Spodoptera frugiperda* (J.E. Smith) (Lepidoptera: Noctuidae) associadas às culturas de milho e arroz no Rio Grande do Sul. Neotropical Entomology 31: 525–529.

Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). Evolution 23: 237–251.

Chessel D, Dufour AB & Thioulouse J (2004) The ade4 package-I-one-table methods. R News 4: 5–10.

Clark PL, Molina-Ochoa J, Martinelli S, Skoda SR, Isenhour DJ et al. (2007) Population variation of the fall armyworm, *Spodoptera frugiperda,* in the western hemisphere. Journal of Insect Science 7: 1–10.

Domingues FA (2011) Variabilidade Genética em Populações de *Heliothis virescens* (Lepidoptera: Noctuidae) no Brasil Inferida por Marcadores Microssatélites. MSc Thesis, University of São Paulo, Piracicaba, SP, Brazil.

Dray S & Dufour A (2007) The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22: 1–20.

Drès M & Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. Philosophical Transactions of the Royal Society of London B 357: 471–492.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A et al. (2011) Geneious v5.4. Available at: http://www.geneious.com/ (accessed 4 July 2014).

Egan SP, Nosil P & Funk DJ (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host-association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. Evolution 62: 1162–1181.

Emelianov I, Mallet J & Baltensweiler W (1995) Genetic differentiation in the larch budmoth *Zeiraphera diniana* (Lepidoptera: Tortricidae): polymorphism, host races or sibling species? Heredity 75: 416–424.

Evanno G, Regnaut S & Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611–2620.

Falush D, Stephens M & Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

Feder JL, Chilcote CA & Bush GL (1988) Genetic differentiation between sympatric host races of *Rhagoletis pomonella*. Nature 336: 61–64.

Feder JL, Opp S, Wazlo B, Reynolds K, Go W & Spizak S (1994) Host fidelity as an effective premating barrier between sympatric races of the apple maggot fly. Proceedings of the National Academy of Sciences of the USA 91: 7990–7994.

Feder JL, Egan SP & Forbes AA (2012) Ecological adaptation and speciation: the evolutionary significance of habitat avoidance as a postzygotic reproductive barrier to gene flow. International Journal of Ecology 2012: e456375.

Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) Version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Funk DJ (1998) Isolating a role for natural selection in speciation: host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. Evolution 52: 1744–1759.

Funk DJ (2012) Of 'host forms' and host races: terminological issues in ecological speciation. International Journal of Ecology 2012: e506957.

Funk DJ, Nosil P & Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proceedings of the National Academy of Sciences of the USA 103: 3209–3213.

Gascuel O (1997) BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Molecular Biology and Evolution 14: 685–695.

Goudet J (2006) hierfstat: Estimation and tests of hierarchical F-statistics. R package version 0.04-4. Available at: http://www.r-project.org; http://www.unil.ch/popgen/softwares/hierfstat.htm (accessed 4 July 2014).

Groot AT, Marr M, Schöfl G, Lorenz S, Svatoš A & Heckel DG (2008) Host strain specific sex pheromone variation in *Spodoptera frugiperda*. Frontiers in Zoology 5: 20.

Groot AT, Marr M, Heckel DG & Schöfl G (2010) The roles and interactions of reproductive isolation mechanisms in fall armyworm (Lep.: Noctuidae) host strains. Ecological Entomology 35: 105–118.

Groot AT, Classen A, Inglis O, Blanco C, Lopez JJ et al. (2011) Genetic differentiation across North America in the generalist moth *Heliothis virescens* and the specialist *H. subflexa*. Molecular Ecology 20: 2676–2692.

Hay-Roe MM, Meagher RL & Nagoshi RN (2011) Effects of cyanogenic plants on fitness in two host strains of the fall armyworm (*Spodoptera frugiperda*). Journal of Chemical Ecology 37: 1314–1322.

Huelsenbeck JP & Andolfatto P (2007) Inference of population structure under a Dirichlet process model. Genetics 175: 1787–1802.

Jakobsson M & Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403–1405.

Jombart T, Devillard S & Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11: 94.

Juárez ML, Murúa MG, García MG, Ontivero M, Vera MT et al. (2012) Host association of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) corn and rice strains in Argentina, Brazil, and Paraguay. Journal of Economic Entomology 105: 573–582.

Levy HC, García Maruniak A & Maruniak JE (2002) Strain identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) insects and cell line: PCRRFLP of cytochrome oxidase C subunit I gene. Florida Entomologist 85: 186–190.

Lima ER & McNeil JN (2009) Female sex pheromones in the host races and hybrids of the fall armyworm, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). Chemoecology 19: 29–36.

Lu Y & Adang MJ (1996) Distinguishing fall armyworm (Lepidoptera: Noctuidae) strains using a diagnostic mitochondrial DNA marker. Florida Entomologist 79: 48–55.

Lu Y, Adang MJ, Isenhour DJ & Kochert GD (1992) RFLP analysis of genetic variation in North American population of the fall armyworm moth *Spodoptera frugiperda* (Lepidoptera: Noctuidae). Molecular Ecology 1: 199–208.

Lu Y, Kochert GD, Isenhour DJ & Adang MJ (1994) Molecular characterization of a strain-specific repeated DNA sequence in the fall armyworm *Spodoptera frugiperda* (Lepidoptera: Noctuidae). Insect Molecular Biology 3: 123–130.

Lynch M & Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Molecular Ecology 3: 91–99.

Machado V, Wunder M, Baldissera VD, Oliveira JV, Fiúza LM & Nagoshi RN (2008) Molecular characterization of host strains of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) in Southern Brazil. Annals of the Entomological Society of America 101: 619–626.

Martinelli S, Clark PL, Zucchi MI, Silva MC, Foster JE & Omoto C (2007) Genetic structure and molecular variability of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) collected in maize and cotton fields in Brazil. Bulletin of Entomological Research 97: 225–231.

McMichael M & Prowell DP (1999) Differences in amplified fragment-length polymorphisms in fall armyworm (Lepidoptera: Noctuidae) host strains. Annals of the Entomological Society of America 92: 175–178.

Meagher RL, Nagoshi RN, Stuhl C & Mitchell ER (2004) Larval development of fall armyworm (Lepidoptera: Noctuidae) on different cover crop plants. Florida Entomologist 87: 454–460.

Meagher RL, Nagoshi RN & Stuhl CJ (2011) Oviposition choice of two fall armyworm (Lepidoptera: Noctuidae) host strains. Journal of Insect Behavior 24: 337–347.

Nagoshi RN (2010) The fall armyworm Triose Phosphate Isomerase (*Tpi*) gene as a marker of strain identity and interstrain mating. Annals of the Entomological Society of America 103: 283–292.

Nagoshi RN (2012) Improvements in the identification of strains facilitate population studies of fall armyworm subgroups. Annals of the Entomological Society of America 105: 351–358.

Nagoshi RN & Meagher RL (2003a) FR tandem-repeat sequence in fall army-worm (Lepidoptera: Noctuidae) host strains. Annals of the Entomological Society of America 96: 329–335.

Nagoshi RN & Meagher RL (2003b) Fall armyworm *FR* sequences map to sex chromosomes and their distribution in the wild indicate limitations in interstrain mating. Insect Molecular Biology 12: 453–458.

Nagoshi RN & Meagher RL (2004) Behavior and distribution of the fall armyworm host strains in Florida. Florida Entomologist 87: 440–449.

Nagoshi RN, Meagher RL, Adamczyk JJ Jr, Braman SK, Brandenburg RL & Nuessly G (2006a) New restriction fragment length polymorphisms in the cytochrome oxidase I gene facilitate host strain identification of fall armyworm (Lepidoptera: Noctuidae) populations in the southeastern United States. Journal of Economic Entomology 99: 671–677.

Nagoshi RN, Meagher RL, Nuessly G & Hall D (2006b) Effects of fall armyworm (Lepidoptera: Noctuidae) interstrain mating in wild populations. Environmental Entomology 35: 561–568.

Osores V, Willink E & Costilla M (1982) Cría de *Diatraea saccharalis* F. en laboratorio. Boletín de la Estación Experimental Agroindustrial Obispo Colombres, Tucumán 139: 1–10.

Paradis E, Claude J & Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289–290.

Pashley DP (1986) Host-associated genetic differentiation in fall armyworm (Lepidoptera: Noctuidae). A sibling species complex? Annals of the Entomological Society of America 79: 898–904.

Pashley DP (1988a) Current status of fall armyworm host strains. Florida Entomologist 71: 227–234.

Pashley DP (1988b) Quantitative genetics, development and physiological adaptation in sympatric host strains of fall armyworm. Evolution 42: 93–102.

Pashley DP (1989) Host-associated differentiation in armyworms: an allozymic and mtDNA perspective. Electrophoretic Studies on Agricultural Pests (ed. by H Loxdale & J den Hollander), pp. 103–114. Oxford University Press, Oxford, UK.

Pashley DP & Martin JA (1987) Reproductive incompatibility between host strains of fall armyworm (Lepidoptera: Noctuidae). Annals of the Entomological Society of America 80: 731–733.

Pashley DP, Johnson SJ & Sparks AN (1985) Genetic population structure of migratory moths: the fall armyworm (Lepidoptera: Noctuidae). Annals of the Entomological Society of America 78: 756–762.

Pashley DP, Hammond AB & Hardy TN (1992) Reproductive isolation mechanisms in fall armyworm host strains (Lepidoptera: Noctuidae). Annals of the Entomological Society of America 85: 400–405.

Pashley DP, Hardy TN & Hammond AM (1995) Host effects on developmental and reproductive traits in the fall armyworm strains (Lepidoptera: Noctuidae). Annals of the Entomological Society of America 88: 748–755.

Pencoe NL & Martin PB (1981) Development and reproduction of Fall Armyworm on several wild grasses. Environmental Entomology 10: 999–1002.

Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Prowell DP, McMichael M & Silvain JF (2004) Multilocus genetic analysis of host use, introgression, and speciation in host strains of fall armyworm (Lepidoptera: Noctuidae). Annals of the Entomological Society of America 97: 1034–1044.

R Development Core Team (2012) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Saldamando CI & Vélez-Arango AM (2010) Host plant association and genetic differentiation of corn and rice strains of *Spodoptera frugiperda* Smith (Lepidoptera: Noctuidae) in Colombia. Neotropical Entomology 39: 921–929.

Sambrook J, Fritsch EF & Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, Long Island, NY, USA.

Schluter D (2001) Ecology and the origin of species. Trends in Ecology and Evolution 16: 372–380.

Schöfl G, Heckel DG & Groot AT (2009) Time-shifted reproductive behaviours among fall armyworm (Noctuidae: *Spodoptera frugiperda*) host strains: evidence for differing modes of inheritance. Journal of Evolutionary Biology 22: 1447–1459.

Schöfl G, Dill A, Heckel DG & Groot AT (2011) Allochronic separation versus mate choice: nonrandom patterns of mating between fall armyworm host strains. American Naturalist 177: 470–485.

Stuhl CJ, Meagher RL & Nagoshi RN (2008) Genetic variation in neonate behavior of fall armyworm (Lepidoptera: Noctuidae). Florida Entomologist 91: 151–158.

Vekemans X (2002) AFLP-SURV version 1.0. Distributed by the author. Laboratoire de Génétique et Ecologie Végétale, Université Libre de Bruxelles, Brussels, Belgium.

Vélez-Arango AM, Arango RE, Villanueva D, Aguilera E & Saldamando CI (2008) Identificación de biotipos de *Spodoptera frugiperda* (Lepidoptera: Noctuidae) mediante marcadores mitocondriales y nucleares. Revista Colombiana de Entomología 34: 145–150.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T et al. (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23: 4407–4414.

Walsh BD (1864) On phytophagic varieties and phytophagic species. Proceedings of the Entomological Society of Philadelphia 3: 403–430.

Walsh BD (1867) The apple-worm and the apple maggot. Journal of Horticulture 2: 338–343.

Whitford F, Quisenberry SS, Riley TJ & Lee JW (1988) Oviposition preference, mating compatibility and development of two fall armyworm strains. Florida Entomologist 71: 234–243.

Wright S (1978) Evolution and the Genetics of Populations, Vol 4: Variability Within and Among Natural Populations. University of Chicago Press, Chicago, IL, USA.

Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. Molecular Ecology 8: 907–913.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. The coefficient of variation of *Spodoptera frugiperda* AFLP markers with 100 bootstrap replicates. The horizontal dashed line indicates a coefficient of variation of 5%.

Figure S2. Count of different configurations of COI and *Tpi* types in populations of *Spodoptera frugiperda* collected from maize, rice, and grasses in Argentina, Brazil, and Paraguay.

Figure S3. Isolation by distance for *Spodoptera frugiperda* populations collected from Argentina, Brazil, and Paraguay.

Figure S4. (A) F_{ST}/H_E plot showing neutral loci, loci candidate for balancing selection, and loci candidate for positive selection derived from 10 populations defined by their origin. (B) F_{ST}/H_E plot showing neutral loci, loci candidate for balancing selection, and loci candidate for positive selection derived from nine populations defined by their origin and (excluding the population from grasses).

Figure S5. (A) F_{ST}/H_E plot showing neutral loci, loci candidate for balancing selection, and loci candidate for positive selection derived from all populations from the same host (maize or rice/grass). (B) F_{ST}/H_E plot showing neutral loci, loci candidate for balancing selection, and loci candidate for positive selection derived from nine populations from the same host without the grass population (maize or rice).

Figure S6. Estimates of the most likely number *of Spodoptera frugiperda* populations. Black squares show the marginal log likelihoods of the data Pr (X|K) when the number of clusters (K) is fixed to different values. The gray squares denote ΔK, an ad hoc indicator of the uppermost hierarchical level of structure detected, based on the rate of change in Pr (X|K) between successive K values. The gray bar denotes the posterior probability distributions Pr (K|X) for the number of populations where K is treated as a random variable.

Table S1. Proportion of different configurations of COI haplotypes and *Tpi* genotypes in populations of *Spodoptera frugiperda* collected from Argentina, Brazil, and Paraguay.

Table S2. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population C_08_W1.

Table S3. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population C_10_W2.

Table S4. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population R_08_W3.

Table S5. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population R_08_E4.

Table S6. Polymorphic nucleotide sites (SNPs) located

in the *Tpi* gene region present in population R_08_E5.

Table S7. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population C_09_E6.

Table S8. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population C_08_E7.

Table S9. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population R_07_E8.

Table S10. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population R_10_E9.

Table S11. Polymorphic nucleotide sites (SNPs) located in the *Tpi* gene region present in population C_08_E10.

Table S12. Matrix of genetic distance between populations of *Spodoptera frugiperda* defined according to sampling site, year of collection, and host plant estimated by pairwise Wright's indices F_{ST} (above diagonal) and pairwise Nei's genetic distance (below diagonal).

Table S13. Matrix of genetic distance between populations of *Spodoptera frugiperda* defined by site and year of sampling and combination of nuclear and mitochondrial haplotype (COI/TPI) estimated by pairwise Wright's indices F_{ST} (above diagonal) and pairwise Nei's genetic distance (below diagonal).

Table S14. Analysis of genetic structure of nine *Spodoptera frugiperda* populations excluding the population from grasses and outlier-F_{ST} loci.

Table S15. Hierarchical analysis of nine *Spodoptera frugiperda* populations among regions, among populations within regions, and among individuals within each population, excluding the population from grasses and outlier-F_{ST} loci.

Table S16. Hierarchical analysis of nine *Spodoptera frugiperda* populations among host plants, among populations within host plant, and among individuals within each population, excluding the population from grasses and outlier-F_{ST} loci.