

## ORIGINAL ARTICLE



WILEY

# Population expansion and genomic adaptation to agricultural environments of the soybean looper, *Chrysodeixis includens*

Cleane S. Silva<sup>1</sup> | Erick M.G. Cordeiro<sup>1</sup> | Julia B. de Paiva<sup>1</sup> | Patrick M. Dourado<sup>2</sup> | Renato A. Carvalho<sup>2</sup> | Graham Head<sup>3</sup> | Samuel Martinelli<sup>3</sup> | Alberto S. Correa<sup>1</sup> 

<sup>1</sup>Department of Entomology and Acarology, Luiz de Queiroz College of Agriculture, University of Sao Paulo, Piracicaba, Brazil

<sup>2</sup>Bayer Crop Science, Sao Paulo, Brazil

<sup>3</sup>Bayer Crop Science, Chesterfield, MO, USA

## Correspondence

Alberto S. Correa, Department of Entomology and Acarology, University of Sao Paulo, Luiz de Queiroz College of Agriculture, Piracicaba, 13418900 Sao Paulo, Brazil.

Email: ascorrea@usp.br

## Funding information

Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant/Award Number: 140844/2016-3 and 311465/2018-8; Fundação de Amparo à Pesquisa do Estado de São Paulo, Grant/Award Number: 2014/11495-3, 2016/04159-2 and 2017/02393-0; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Grant/Award Number: Finance Code 001

## Abstract

Evolutionary studies of insect pests improve our ability to anticipate problems in agricultural ecosystems, such as pest outbreaks, control failures, or expansions of the host range. Here, we investigated the mechanisms underlying the evolutionary processes behind the recent census size expansion and local adaptation of *Chrysodeixis includens*. First, we sequenced mitochondrial markers to conduct a phylogeographic investigation of *C. includens* historical processes. Then, we combined a de novo genotyping-by-sequencing approach with a study of agricultural landscapes to uncover recent processes of adaptation. Primarily, we found low genetic diversity across all markers and clear indications of a recent demographic expansion. We also found a lack of significant isolation by distance (IBD), and weak or absent genetic structure considering geographic locations. However, we did find initial signs of population differentiation that were associated with host plant types (i.e., soybean and cotton). Agricultural landscape attributes, including soybean crops, were significantly associated with putative markers under positive selection. Moreover, positive selection associated with host differentiation was putatively linked to digestive enzymes. This study showed how landscape composition and host plants can affect the evolutionary process of agricultural pest insects such as *C. includens*.

## KEYWORDS

demography, insect adaptation, landscape genetics, Noctuidae, selection

## 1 | INTRODUCTION

Detection of early adaptive changes in insect pest populations is in the best interest of farmers and agroindustry and could allow us to anticipate—rather than merely remediate—agricultural problems such as host range expansions or pesticide resistance. However, there is no clear connection between population dynamics (i.e.,

minor pests becoming major pests) and the underlying adaptive process, mainly because of differences in the time scales on which ecological and evolutionary dynamics operate (Cavender-Bares, Kozak, Fine, & Kembel, 2009; Lu et al., 2010). For instance, changes in habitat composition or configuration caused by agricultural practices trigger adaptive responses to the new environments that are not easily noticed in the short term. However, confounding factors do

Silva and Cordeiro contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

**TABLE 1** Information about sampled locations of 12 study areas in Brazil for SNP marker sequencing. Two locations (BACO and MTCV) were sampled during both the soybean- and cotton-growing seasons.  $N_{\text{GEN}}$  refers to the number of insects successfully sequencing using SNP markers

Ecoregion	Location	Code	Latitude	Longitude	Crop	$N_{\text{GEN}}$
Atlantic Forest	Coxilha, RS	RSCO	28°10'54"S	52°44'46"W	Soybean	16
Atlantic Forest	Pitanga, PR	PRPI	24°17'23"S	52°34'16"W	Soybean	14
Atlantic Forest	Casa Branca, SP	SPCB	21°43'34"S	47°09'09"W	Soybean	13
Cerrado	Araguari, MG	MGAR	18°59'23"S	47°33'52"W	Soybean	15
Cerrado	Mineiros, GO	GOMI	17°37'43"S	52°36'52"W	Soybean	13
Cerrado	Campo Grande, MS	MSCG	20°42'31"S	54°30'03"W	Soybean	14
Cerrado	Campo Verde, MT	MTCVS	15°25'34"S	54°48'05"W	Soybean	10
Cerrado	Campo Verde, MT	MTCVC	15°36'40"S	55°14'05"W	Cotton	10
Caatinga	Correntina, BA	BACOS	11°49'21"S	46°10'52"W	Soybean	10
Caatinga	Correntina, BA	BACOC	13°40'39"S	45°41'21"W	Cotton	10
Caatinga	Russas, CE	CERU	04°55'28"S	38°00'14"W	Bean	13
Caatinga	Teresina, PI	PITE	05°02'21"S	42°47'22"W	Bean	12

not always stem from landscape changes, and a population's history and dynamics (i.e., recent speciation events, gene flow, genetic drift, and population demography) can also make it difficult to detect early signs of local adaptation.

Agroecosystem configuration and composition may be responsible for sudden changes in ecological and evolutionary dynamics, imposing strong selective pressure and leading to new forms of pests (Altieri, 1999; Alvarado-Serrano, Van Etten, Chang, & Baucom, 2019). This complex scenario can be accentuated in countries such as Brazil, where agriculture has rapidly intensified and expanded geographically to the Cerrado region (Brazilian savannas) in the past 50 years, compared to places where agriculture is long established (Arvor, Meirelles, Dubreuil, Bégué, & Shimabukuro, 2012; DeFries, Herold, Verchot, Macedo, & Shimabukuro, 2013). However, the cost of this profound economic transformation is the loss and fragmentation of large portions of the native vegetation (Landis, 2017; Ribeiro, Metzger, Martensen, Ponzoni, & Hirota, 2009). The result is extensive uniform cultivation areas where only one or a few crop species dominate the landscape. Beginning in the 1970s, the development of new farming technologies, including new crop varieties and better soil management, allowed farming in new areas (Arvor et al., 2012; Müller, Rufin, Griffiths, Barros Siqueira, & Hostert, 2015). In addition to leading to large areas with only one or a few crops, reducing landscape heterogeneity, the use of crop varieties with early and extra-early cycles made possible two or three crop seasons annually, increasing the challenges for insect management. The succession of suitable hosts cultivated on such large scales can lead to the rapid evolution of insect pest traits such as expansion of the number of host species, insecticide/Bt resistance, and specialization (Andrade et al., 2016; Santos, Specht, Carneiro, Paula-Moraes, & Casagrande, 2017). Furthermore, a continuous agriculture landscape may allow increased gene flow among meta-populations.

In recent decades, the soybean looper, *Chrysodeixis includens* (Walker, [1858]) (Lepidoptera: Noctuidae: Plusiinae), a native to the Americas, has emerged as a serious pest of soybean and cotton (Panizzi, 2013; Santos et al., 2017). This species is polyphagous and shows high fecundity and rapid development (egg-adult period of 30 days in tropical regions). Little is known about its dispersal abilities. Until recently, *C. includens* was described as a minor pest that rarely reached levels requiring control (Kogan, Turnipseed, Shepard, Oliveira, & Borgo, 1977; Panizzi, 2013). Because of its higher population levels during the last decade, *C. includens* has attracted more attention and become a major concern for growers in many parts of Brazil.

Recent studies using inter-simple sequence repeat (ISSR) markers in Brazilian populations of *C. includens* have found low genetic diversity and no genetic structure across the area investigated (Palma, Maebe, Guedes, & Smagghe, 2015). However, not much is known about the adaptive potential of the looper in response to the recent expansion of farming to central and northern Brazil. Here, we used sequence fragments of the genes cytochrome c oxidase subunit I (COI) and cytochrome c oxidase subunit II (COII), in addition to genotyping by sequencing (GBS), to investigate first, if populations are locally adapted, and second, the mechanisms underlying the evolutionary processes behind local adaptation in croplands in Brazil. Our explicit goals were (a) to use mitochondrial markers to reveal details of the *C. includens* phylogeographic history in the Americas, (b) to measure the levels of population genetic structure in important agricultural areas in Brazil, and (c) to test whether positive selection is taking place in different geographic regions and cultivated crops; and if confirmed, then to investigate which environmental features are associated with the ongoing adaptive processes. In a broader context, our goal was to shed more light on the adaptive processes that allow previously unimportant insects to become significant pests.

## 2 | MATERIALS AND METHODS

### 2.1 | Insect sampling and DNA extraction

Specimens of *C. includens* were collected from 15 of the principal common bean, soybean, and cotton-producing regions in Brazil between January and May 2017, by actively collecting larvae or with pheromone traps. Approximately 800 larvae of *C. includens* were collected per sample site in the field. After they matured, 30 insects were transferred to individual 1.5-ml Eppendorf tubes containing 99% EtOH and stored at  $-80^{\circ}\text{C}$  until further sample processing. Considering the geographic range of soybean and cotton cultivation, we directed our sampling efforts to cover the most ecologically or climatically diverse combination of natural areas (the Atlantic Forest, Brazilian Cerrado, and Caatinga domains) and crop types that could provide a realistic view of the Brazilian agriculture mosaic (Table 1 and Appendix S1).

DNA was extracted from the insects by using an adapted CTAB (2%) extraction protocol and resuspended in 70  $\mu\text{l}$  of Milli-Q water (Doyle & Doyle, 1987). DNA extraction quality control was first performed by visual inspection in agarose gel (0.8% w/v), followed by spectrophotometric and fluorometric methods. Samples were tested in a NanoDrop spectrophotometer and agarose gel to check for contaminants and DNA integrity, respectively, followed by a Qubit assay to estimate the DNA concentration in each sample. Only samples with a concentration above 15 ng/ $\mu\text{l}$  were included in the subsequent steps.

### 2.2 | Landscape and landscape genetic analyses

We used only satellite images with a maximum of two months' difference between the sampling time and the satellite image capture. This criterion avoided problems with cloudy or unavailable images, thus improving the landscape classifications. We used images from two satellites, CBERS 4 and LANDSAT 8, available from The National Institute of Spatial Research (INPE) and the United States Geological Survey EarthExplorer (USGS EarthExplorer) repositories, respectively. A buffer of a 25-km radius from the sampled point was used to delimit the size of the classified landscapes used. Landscapes were classified in the software ArcGIS 10.2.2, using the "create signature" option that uses different spectral responses that can be associated with the presence of a class (i.e., agriculture, native vegetation, urban areas, and exposed soil). Supervised classification was performed using a maximum-likelihood classification and was manually inspected for errors and misclassifications. At two geographic locations (BACO and MTCV), two sample collections were performed (soybean versus predominantly cotton areas) in overlapping landscapes. We only performed the analyses on the soybean landscape at those two locations, to ensure a minimum difference in the time frame between sample collections, avoiding the effect of migration between the landscapes. Soybean and cotton landscapes were contrasted later in separate analyses.

We used the LecoS plug-in, which is based on metrics taken from FRAGSTATS v4, to estimate the landscape metrics based on the composition of each attribute in the landscape (Jung, 2016; McGarigal, Cushman, Neel, & Ene, 2012). For "patch-based metrics" (e.g., bare soil, native forest, soybean), land cover, landscape proportion, edge length, edge density [ $\text{ED} = (\text{Total length (m) of the edge in landscape} / \text{Total landscape areas (m}^2) * 10,000]$ , number of patches, and mean patch area were collected to describe landscape structure and composition. For the "landscape-level metrics," the Shannon diversity index (SHDI) and Shannon evenness index (SHEI) were estimated (Shannon & Weaver, 1949).

The SHDI is a measure of diversity that is commonly applied in community-ecology studies but can also be used to measure diversity in landscapes. When the SHDI is zero, the landscape contains only one patch, and therefore, no diversity is observed at that particular scale. As the value of the SHDI increases, a larger number of different patch types are observed, and consequently, a higher patch richness. A second index used here was the SHEI, which expresses the degree of dominance of a patch type in relation to other patch types present in the landscape. Low values can be interpreted as meaning that the landscape is dominated by one or a few patch types, whereas high values mean that patch types are evenly distributed in the landscape.

Principal component analysis (PCA) was used to reduce the dimensionality and to simplify pattern recognition of the landscape and climate variables. We selected land cover as our patch-based metric to study the response to the landscape composition at the genomic level. The 14 land-cover attributes were transformed with PCA to reduce the dimensionality, and the first PCA coordinate was used as a predictor. We selected land cover as our primary landscape variable based on our a priori knowledge that *C. includens* strongly responds to host availability in the field (Andrade et al., 2016; Santos et al., 2017). The same approach was also used for the climate variables; 19 WorldClim variables ([www.worldclim.org/bioclim](http://www.worldclim.org/bioclim)) had their dimensional space reduced by PCA transformation. This approach allowed us to assess the position of each putative population in environmental space (climate + hosts). We used the coordinates of the projections on the first PC axis as our independent variables. The importance of each variable in the PCA was presented as the variable contribution to the first PC axis, which was given by squaring the product of the loadings by the component standard deviation and dividing it by the square of the total PCA loadings multiplied by the component standard deviation (Kassambara, 2017).

### 2.3 | Mitochondrial COI and COII amplification and sequencing

Two mitochondrial fragments from the COI and COII genes were amplified in PCR for 61 individuals from 15 sample sites (Appendix S1) (Monteiro et al., 1999; Simon et al., 1994). Raw mitochondrial COI and COII sequence chromatograms were inspected and edited to 522 bp (NCBI accession: MK256487 – MK256546) and 697 bp

(NCBI accession: MK256547 – MK256606), respectively, using Sequencher 4.0.1 (Gene Codes Corp.). Next, the COI and COII sequences were concatenated and aligned using MUSCLE with the default options in MEGA 7.0 (Kumar, Stecher, & Tamura, 2016).

## 2.4 | GBS library preparation and sequencing

Two libraries containing 75 samples each (150 insects in total), from 12 sample sites (Table 1), were processed using the GBS protocol (Elshire et al., 2011). Samples from three locations did not meet the quality standards and were therefore not included in the GBS library preparation. Only one enzyme (*Pst*II, CTGCAG) was used during the genome digestion step to reduce genome complexity. After library preparation, DNA samples were quantified by qPCR, and the size distribution checked using an Agilent BioAnalyzer. Libraries were sequenced on an Illumina HiSeq 2500 (100-bp single-end reads) at the Animal Genome Centre at USP/ESALQ (see Appendix S3 for details).

## 2.5 | mtDNA sequence analysis

For COI-COII concatenated sequences, we calculated the haplotype diversity and nucleotide diversity parameters using DnaSP ver. 5 (Librado & Rozas, 2009). The analysis of molecular variance (AMOVA), using the samples collected as a putative population, was also carried out using Arlequin with parametric bootstrapping (1,000 replicates) and significance at the 5% level (Excoffier, Smouse, & Quattro, 1992). A network of median-joining haplotypes was generated using the PopART software program to reconstruct the genealogical relationships among the COI-COII haplotypes. The hypothesis of demographic expansion was tested using Tajima's D and Fu's  $F_s$  tests of selective neutrality in Arlequin 3.1 (Excoffier, Laval, & Schneider, 2006). Both tests used 1,000 random samples, using coalescent simulations and the "Infer from distance matrix" option for "Haplotype definition." We also tested the goodness of fit of the observed mismatch distribution to that expected under the spatial expansion model. The sum of squared deviations (SSD) and Harpending's Raggedness index statistics and their associated p-values were calculated using Arlequin 3.1 (Excoffier et al., 2006) with 1,000 replicates. Finally, we used additional COI sequences downloaded from BOLD Systems to estimate *C. includens* COI haplotype relationship and genetic diversity throughout the Americas. Details of the method are shown in Appendix S2.

## 2.6 | GBS data analysis

Raw sequences were demultiplexed, trimmed to 90 bp, and filtered according to the *Phred* scores using *process-radtags* in STACKS 2.2, using default settings (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). We proceeded with the analysis using the five major modules available in STACKS in a de novo approach. During the first

step, the reads were used to build RAD loci (restriction site-associated DNA marker loci) using *ustacks* to form putative loci present in each sample file. At this stage, two critical parameters are the minimum number of raw reads to form a putative allele ( $m = 3$ ) and the number of mismatches allowed between stacks to merge them into a single allele ( $M = 4$ ). Setting  $m$  to 3 has shown good performance for a broad range of data and represents a good trade-off between coverage and polymorphs (Paris, Stevens, & Catchen, 2017; Rochette & Catchen, 2017). The criterion to merge stacks into a single locus ( $M$ ) is highly dependent on the data and required a balance on the homolog and paralogue loci detection. In general, the presence of a higher degree of polymorphs requires higher values for  $M$ , but here we tested a range of values ( $M1$ – $M8$ ) to identify where the parameter plateaus (Paris et al., 2017), which proved to be  $\sim 4$  (Figure S3). Next, STACKS creates a catalog containing consensus loci using *cstacks*, using 4 as the number of mismatches allowed between sample loci ( $n = 4$ ). We used the general criteria of  $M = n$ . In the following step, STACKS matches loci found in each individual against the catalog, using default settings. Using this approach, STACKS genotyped a total of 340,089 loci with mean effective per-sample coverage of  $97.7\times$  (stdev =  $27.9\times$ ), minimum coverage  $10.1\times$ , and maximum coverage  $164.8\times$ .

Last, we used the *population* module to filter missing data, retaining only loci present in 80% or more individuals within a sampling location ( $r = .8$ ) and  $\sim 80\%$  or more across sampling locations ( $p = 8$ ). Minor allele frequency cutoff was set at 0.05, and maximum observed heterozygosity limited to 50%. The resulting dataset was saved as VCF, genepop, and structure files and converted to other formats using PGDSPIDER 2.0 when necessary (Lischer & Excoffier, 2012). The percentage of missing data allowed and the number of markers implemented to estimate parameters can significantly affect estimates of genetic diversity; therefore, we calculated genetic diversity parameters varying both the overall percentage of missing genotypes and the number of SNPs to find the most stable values. In our dataset, the most stable estimates (i.e., that stop varying as we increase the number of SNPs used) were acquired with more than 5,000 SNPs, while no more than 80% of overall missing genotypes were allowed (see Appendix S3).

## 2.7 | Demographic history

We reconstructed the demographic history of *C. includens* using the multi-locus method Extended Bayesian Skyline Plot (EBSP) implemented in BEAST version 1.8.2 (Bouckaert et al., 2014; Drummond, Suchard, Xie, & Rambaut, 2012). Because of the lack of genetic structure found in the  $F_{ST}$  and structure analysis, we randomly selected 50 neutral sequences (not flagged in our analysis as putative candidates under selection) from a simple pooled population to perform the demographic reconstruction analysis (Calderón et al., 2016). Site model and clock model partitions were linked, whereas the tree partitions remained unlinked. The HKY substitution model was implemented because it allowed us to work with fewer parameters. A strict clock

model was implemented using a fixed mutation rate across the genome of  $2.9 \times 10^{-9}$  mutations/site/generation. The mutation rate was selected based on similar organisms, as no value is available for this species (El-Shehawi & Elseehy, 2016; Keightley, Ness, Halligan, & Haddrill, 2014; Oppold & Pfenninger, 2017). We assumed that the species has 10 generations per year. The chain length of 100 million generations was sampled every 4,000 states, and a 20% burn-in was implemented in three independent runs. The results were inspected in Tracer, examining the posterior ESS (effective sample size) and the convergence among the runs.

## 2.8 | $F_{ST}$ , population structure, and gene flow

We assessed genetic partitioning by means of the software STRUCTURE, using a single SNP per locus (i.e., --write\_single\_snp was used in population module in STACKS) (Pritchard, Stephens, & Donnelly, 2000). An initial burn-in of  $1.5 \times 10^5$  was used, followed by  $5 \times 10^5$  Markov chain Monte Carlo (MCMC) interactions, using 10 replications for each K value without prior population information. The analyses were run from  $K = 1$  to  $K = 10$ , allowing allele correlation and an admixture model. The most likely number of clusters was determined using the Evanno method implemented in STRUCTURE HARVESTER 0.6.93 (Earl & vonHoldt, 2012). The STRUCTURE results were averaged using CLUMPP (Jakobsson & Rosenberg, 2007) and visualized using DISTRUCT (Rosenberg, 2004). Additionally, we used principal component analysis (PCA) implemented in *ade4* and *ade4genet* to investigate the underlying structure.

Pairwise differentiation between sampling locations was estimated based on  $F_{ST}$  estimation calculated in STACKS (Weir, 1996). A heatmap was produced to illustrate the  $F_{ST}$  relationships, using the R program for both the complete set of SNPs and outliers. To test the correlation between genetic distance and geographic distance, a Mantel test, including an all-pairwise-comparison matrix, was performed with 10,000 permutations, using the R libraries *ecodist* (Goslee & Urban, 2007) and *ade4* (Dray & Dufour, 2007).

## 2.9 | Outlier detection and associating genomic variation with environmental characteristics

We used Lositan to detect putative loci under natural selection. The method implemented in Lositan simulates the joint distribution between  $F_{ST}$  and heterozygosity ( $H_e$ ) under neutrality (Antao, Lopes, Lopes, Beja-Pereira, & Luikart, 2008), assuming the island model of migration with neutral markers. The neutral mean  $F_{ST}$  was calculated using 100,000 simulations, with a confidence interval of 0.99 and a false discovery rate (FDR) of 0.05. To assure more stringent detection parameters, we selected the “force mean  $F_{ST}$ ” and “neutral mean  $F_{ST}$ ” options. We also used a Bayesian approach that estimated the posterior probability of a given locus to be under selection, contrasting this assumption with the neutral model. A second analysis was implemented in BayeScan v2.1 (Foll & Gaggiotti, 2008). A total of

20 pilot runs with 5,000 interactions each were generated before computing Markov chains. Burn-in was set to 50,000 steps, and the chains had 5,000 interactions with a thinning interval of 10. We considered prior odds of 10 and a threshold of a posterior P of >0.95 to select an SNP as “under positive selection.” In both analyses, we contrasted sampling locations, except the cotton areas (BACOC and MTCVC) that partially overlapped in time and space with soybean areas (BACOS and MTCVS). Contrasts between soybean locations and cotton locations were performed separately.

We used a series of partial Mantel tests with 10,000 randomizations to test the correlation between the variation in  $F_{ST}$  outliers (putative loci under directional selection) detected by both Lositan and BayeScan, and landscape and climate variables. The partial Mantel test accounts for the nonindependence of points by creating a random distribution to use as a null distribution and test for significance (Mantel, 1967). Additionally, a partial correlation for each variable is calculated by controlling for any variables already in the model. Four hypotheses may explain the driver(s) of outlier RAD locus divergence: (a) isolation by habitat hosts and habitat composition (landscape distance), (b) isolation by the environment (IBE) due to climate differences, and (c) isolation by geographic distance (IBD).

Additionally, the presence of loci that correlate with environmental variables (climate and landscape composition) was also tested using a latent factor mixed-effect model (LFMM) implemented in R package LEA (Frichot & François, 2015). The advantage of the LFMM algorithm is to correlate markers and environmental values while taking into account possible levels of population structure (i.e., latent factors). We defined latent factor 2 ( $K = 2$ ) based on “*snmf*.” Again, for this analysis, individuals collected from cotton landscapes were removed, as were individuals with more than 10% missing data. The two variables were used to test the associations with climate and landscape. To test the association, we used the coordinates of the first PC of WorldClim variables and the landscape composition (percentage of each component of the landscape) described in the previous section (“Landscape and landscape genetic analyses”). The Gibbs sampler algorithm was run for 10,000 cycles, following a burn-in of 5,000 cycles. To control for false-positive detection, we used the Benjamin-Hochberg algorithm, adjusting the false discovery rate (FDR)  $q = 10\%$ . Putative loci under selection detected by more than one method were used as queries in a BLASTx search against the NCBI database, using Blast2GO (Götz et al., 2008). Ontologies with  $e$ -value  $< 1 \times 10^{-3}$  were used for annotation.

## 3 | RESULTS

### 3.1 | Mitochondrial COI and COII variation

The mitochondrial network (COI-COII) of Brazilian *C. includens* revealed a star-like topology of the haplotype relationships, with a central haplotype (haplotype H1) that is widely distributed and in a high frequency, and additional haplotypes at low frequencies (11 total), connected to the central haplotype by a single mutation step



(Figure 1b). The ubiquitous haplotype H1 was recovered from 73.8% of the individuals and was present at all sample sites. Eight haplotypes were singletons (i.e., they occurred only once in our dataset), and three were present in low frequency (H2, H6, and H12) and geographically restricted to certain sample sites.

The genetic diversity of *C. includens* in Brazil seems to be evenly distributed through the species' range and is moderate to low. The haplotype and nucleotide diversities across the locations sampled were  $H_d = 0.455$  and  $\pi = 0.00048$ , respectively. Haplotype and nucleotide diversities were highest in the Cerrado ( $H_d = 0.622$ ;  $\pi = 0.00067$ ) and Atlantic Forest biomes ( $H_d = 0.505$  and  $\pi = 0.00059$ ) and lowest in the Caatinga biome, where each sample site had only a single haplotype (Appendix S1). The AMOVA confirmed the lack of genetic structure at the regional spatial scale, that is, the differences within populations accounted for almost all the genetic variance for *C. includens* in Brazil (98.1%;  $\Phi_{ST} = 0.019$ ) (Appendix S4).

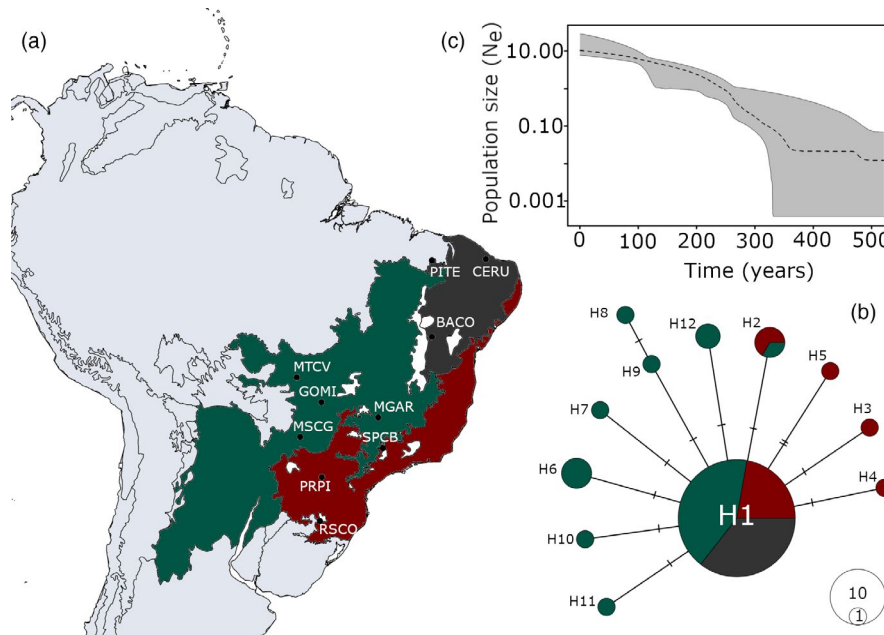
Similar trends were observed at a broader scale when 173 COI sequences from 8 different countries were analyzed (Appendix S2). The analysis using curated sequences from the BOLD database also revealed the widespread occurrence of the single dominant haplotype A1, found throughout the Americas. This analysis also showed haplotypes with a lower frequency derived from A1 by a single mutational step (i.e., star-like topology), similar to the findings in Brazil.

### 3.2 | Demographic history

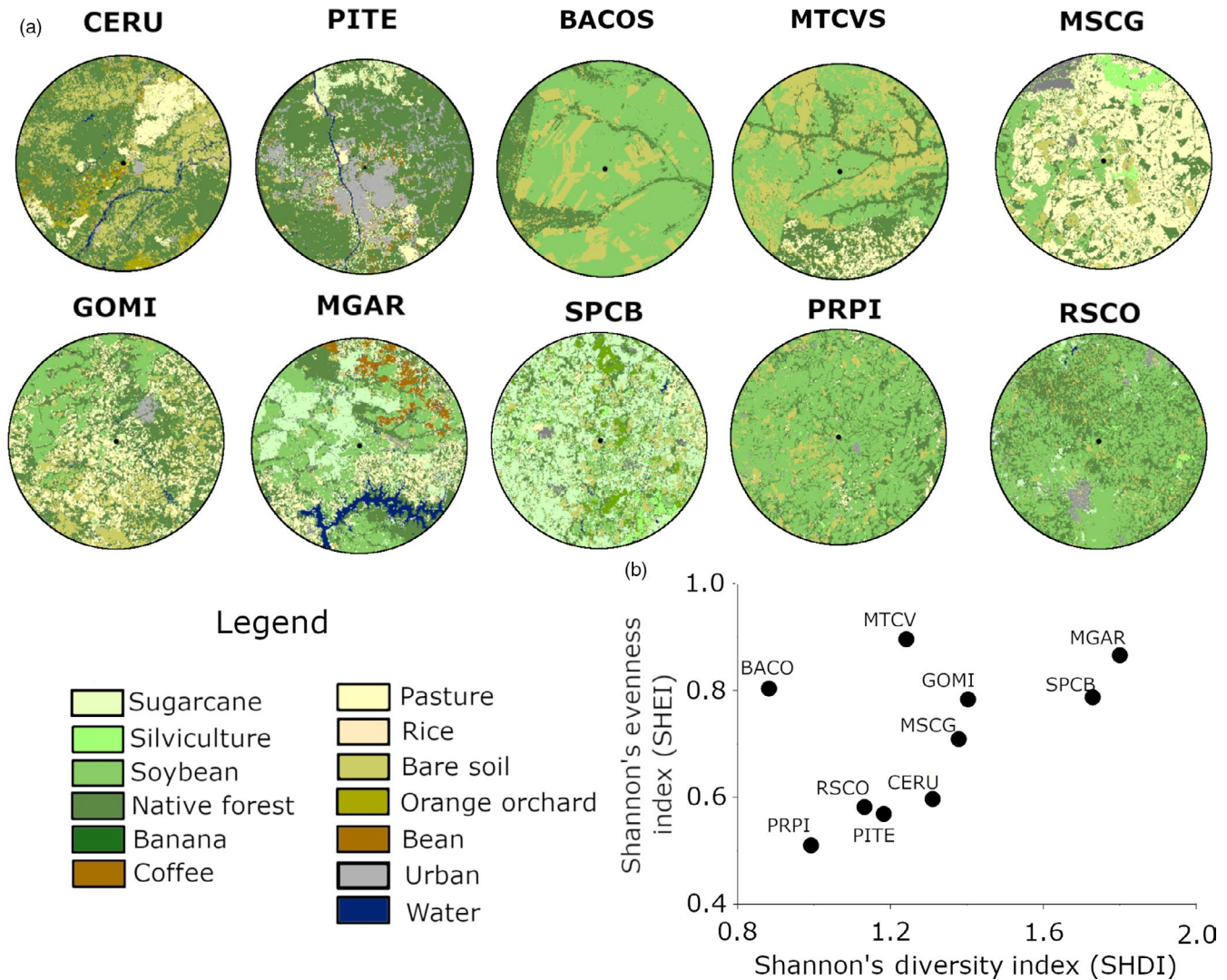
The analyses of both mitochondrial sequences and SNP markers indicated a demographic expansion of *C. includens* in Brazil. Neutrality tests based on COI-COII indicated a population expansion, with Tajima's  $D$  ( $D = -2.21$ ;  $p < .001$ ) and Fu's  $F_s$  ( $F_s = -12.16$ ;  $p < .001$ ). In addition, the model of a sudden expansion did not reject the hypothesis of a spatial expansion for *C. includens* ( $SSD = 0.0002$ ;  $p = .60$ ; Raggedness = 0.1090,  $p = .60$ ). Similarly, the EBS analysis based on 50 nuclear sequences showed a signature of recent demographic expansion in the *C. includens* populations. Based on the assumed molecular clock calibration ( $2.9 \times 10^{-9}$  mutations/site/generation), the expansion started ~300 years before the present (Figure 1c).

### 3.3 | Landscape/environmental composition

There was a tendency for agricultural landscapes with higher Shannon diversity indexes to also show more evenly distributed proportions of vegetation cover types across the different classes, although this relationship was not statistically significant ( $r = .47$ ,  $p = .17$ ) (Figure 2b). Landscapes with a higher diversity of patch-fragment classes and more evenly distributed patches were found in MGAR and SPCB. The least diverse and least evenly distributed compositions of landscape attributes were found in BACOS (i.e., least diverse) and PRPI (i.e., least



**FIGURE 1** *Chrysodeixis includens* population demography. (a) Geographic distribution of the 12 collection sites of *C. includens* (b) Median-joining network of *C. includens* haplotypes in Brazil, based on two mitochondrial gene fragments (COI-COII). The size of each haplotype circle reflects the haplotype frequency, and hatch marks reflect mutation steps. Colors indicate the three Brazilian biomes: Caatinga (Ca), Cerrado (Ce), and Atlantic Forest (AF). (c) Bayesian skyline plot (BSP) showing population size dynamics for *C. includens* in Brazil based on 50 loci randomly sampled throughout the genome. The y-axis indicates an effective population size ( $N_e$ ) scaled by mutation rate ( $\mu$ ) as a function of time (years). Black dotted line shows the median BSP estimate, and gray area shows the upper and lower 95% highest posterior density limits. The soybean looper can produce up to 12 generations per year in a tropical climate.



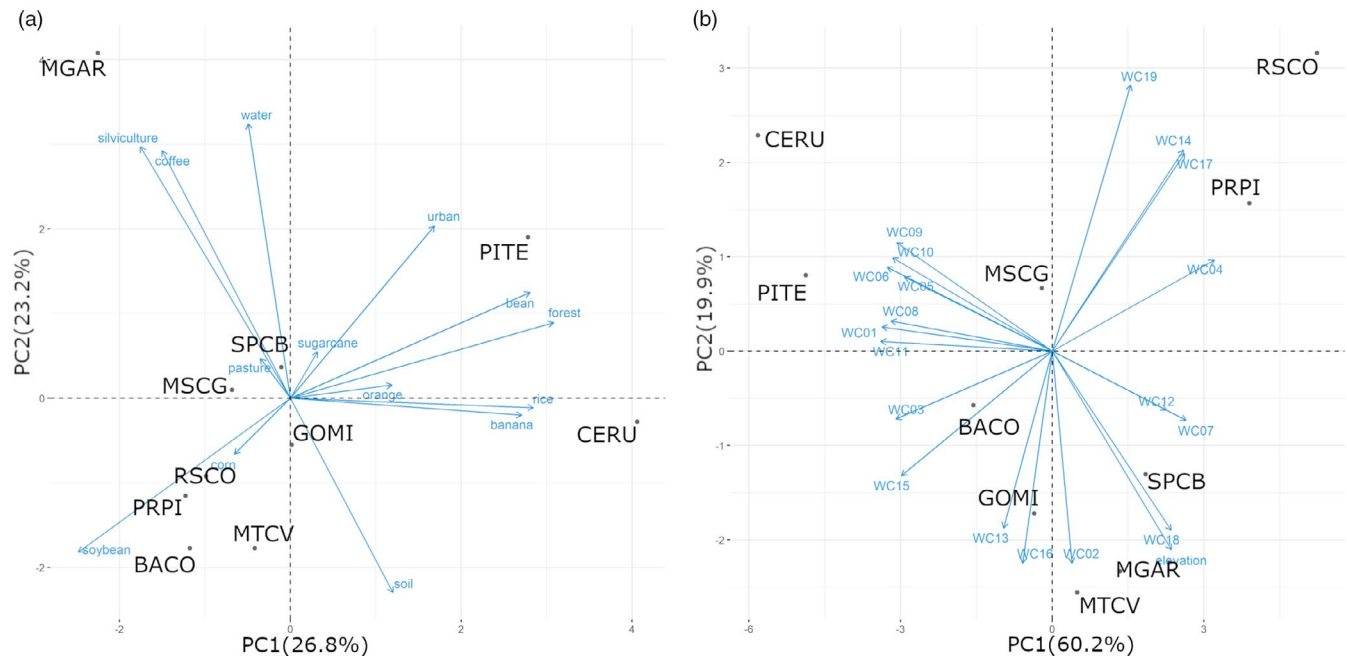
**FIGURE 2** Agriculture landscapes and crop diversity. (a) Landscape classification (25-km radius buffer) of 10 sampling sites for *Chrysodeixis includens* (in soybean crops) in Brazil. Images were obtained from two satellites, CBERS 4 and LANDSAT 8, available at the National Institute of Space Research (INPE) and the United States Geological Survey EarthExplorer (USGS EarthExplorer) repositories, respectively. For details, see Appendix S12. (b) Plot of Shannon diversity index (SHDI) against Shannon evenness index (SHEI) for 10 agricultural landscapes where *Chrysodeixis includens* is commonly found in soybean crops in Brazil

even), respectively (Figure 2). Soybean farms were the predominant agriculture class in 80% of the landscapes, ranging from 7% (SPCB) to 52% (PRPI), with a mean proportion of 28.80% across all landscapes (Appendix S5). The second most common class was native forest, ranging from 12% (BACOS) to 51% (PITE), with a mean proportion of 25.75%. Bare soil and native forest showed the most patches, indicating high levels of fragmentation of the native vegetation (Appendix S5). Based solely on the lack of heterogeneity, we can predict that insects from BACO in the state of Bahia and PRPI in the state of Paraná might be under greater host selection pressure due to monoculture, if we consider only land cover and patch size.

The PCA of the agricultural landscapes, based on the land-cover attributes, revealed similarities and dissimilarities between the landscapes and was highly useful for pattern recognition (Figure 3a). The first two components were statistically significant and together explained 50.0% of the overall variance (Appendix S6). The most

important factors in the first component (PC1) were native forest (18.65% contribution), rice (15.88%), common bean (15.45%), bananas (14.41%), and soybeans (12.12%). The most important factors in the second PCA component (PC2) were water (23.90%), silviculture (20.03%), and coffee (19.44%). Soybeans were more associated with corn (maize), while coffee was more associated with silviculture and pasture. PRPI and RSCO had similar landscapes, where soybean was the predominant component. MGAR had a unique landscape, where silviculture and coffee were the predominant components.

The PCA of the WorldClim data defined a climate space of reduced dimensionality where the agricultural landscapes are located. The first two components were significant and together explained 80.1% of the overall variance (Figure 3b, Appendix S6). The first component was more associated with temperature (WC01 = "Annual Mean Temperature" and WC11 = "Mean Temperature of Coldest Quarter"), while the second component was more associated with



**FIGURE 3** Environmental distance according to landscape composition and climate. Principal component analysis (PCA) loading plots for 15 landscape attributes and 19 WorldClim environmental variables. Solid lines (blue) show the loading of landscape attributes (a) or climate variables (b), and points show the scores for individual landscapes from each site

**TABLE 2** Summary of genetic statistics considering only variant positions in populations of *Chrysodeixis includens* assessed using SNP markers: observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), Wright's inbreeding coefficient ( $F_{IS}$ ), and nucleotide diversity ( $\pi$ )

Sample location	$H_o$	$H_e$	$F_{IS}$	$\pi$
RSCO	0.174	0.240	0.247	0.248
PRCM	0.167	0.236	0.243	0.246
SPCB	0.171	0.238	0.239	0.249
MGAR	0.172	0.244	0.251	0.253
GOMI	0.181	0.239	0.222	0.249
MSCG	0.182	0.245	0.240	0.254
MTCVC	0.177	0.218	0.162	0.234
MTCVS	0.170	0.228	0.211	0.241
BACOC	0.171	0.219	0.182	0.236
BACOS	0.175	0.228	0.194	0.244
PIEB	0.178	0.237	0.218	0.247
CERU	0.185	0.245	0.217	0.256

precipitation (WC19 = "Precipitation of Coldest Quarter"). In the climate space, the pairs PRPI and RSCO, MGAR and SPCB, and CERU and PITE occupied very similar environments.

### 3.4 | Population diversity and structure

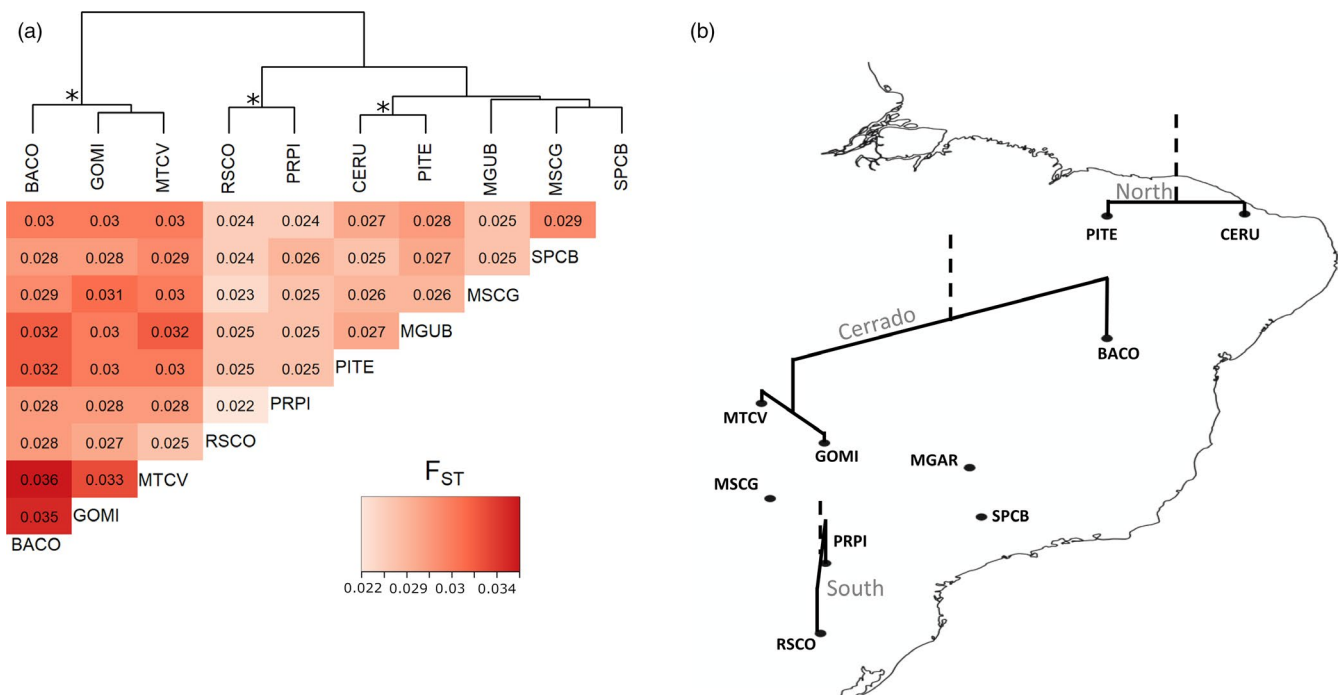
SNP-estimated diversity indices were largely concordant with the low diversity found in the mitochondrial data, and a narrow range

of values for nucleotide diversity ( $\pi = 0.006\text{--}0.007$ ), observed heterozygosity ( $H_o = 0.005 \pm 0.000$ ), and inbreeding coefficient ( $F_{IS} = 0.004\text{--}0.007$ ) were found across the 12 sites where samples collected in the same regions were more similar (Table 2). Analysis based on pairwise  $F_{ST}$  revealed weak structuring (i.e., range of 0.022–0.036 and mean of 0.028), but a consistent separation of collecting sites according to broad regions, where samples collected in the same regions were more similar according to  $F_{ST}$  values (Figure 4a). However, nonsignificant isolation by distance at the 5% level was detected ( $r = .30$ ,  $p = .06$ ). The STRUCTURE analysis confirmed a low degree of differentiation among the sampling locations. However, a higher value of  $K$  ( $K = 4$ ) was detected by the Evanno method (Earl & vonHoldt, 2012), even though populations were fully admixed (Appendix S7).

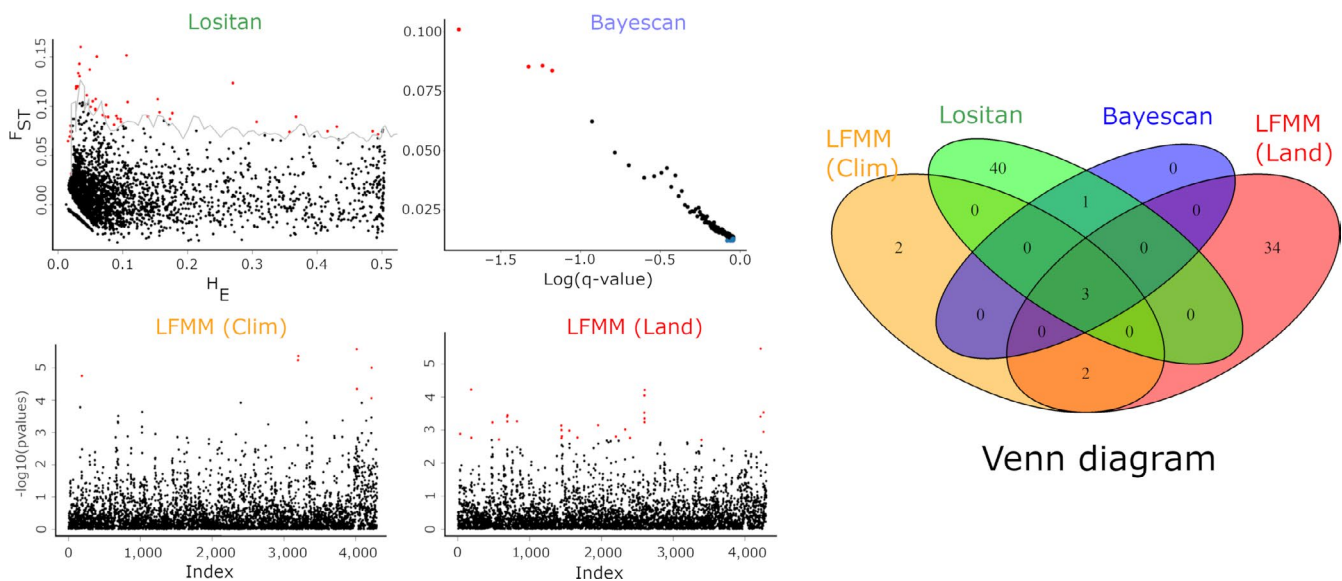
### 3.5 | Landscape/Environmental correlation and association study

Lositan yielded 44 loci that are putatively under positive selection, out of the 3,503 SNP evaluated; while BayeScan detected 4 candidates, all present in the Lositan list of candidates (Figure 5). The four shared loci were not identified in a Blastx search. The Mantel tests demonstrated that landscape distance alone correlated significantly with genetic distance, based on  $F_{ST}$  outliers (putative loci under positive selection) among population pairs, independently of the method used (Table 3). Conversely, environmental distances based on climate variables showed no correlation with the genetic distance, while a significant correlation with geographic distance was present only in Lositan markers. When partial Mantel tests





**FIGURE 4** Genetic differentiation according to the fixation index ( $F_{ST}$ ) of *Chrysodeixis includens* populations. (a) Heatmap of pairwise  $F_{ST}$  values based on 1,520 neutral makers. (b) Map showing grouping pattern of samples collected in the Cerrado, North, and South regions according to  $F_{ST}$  values. \*Indicates on the map, clusters according to broad geographic regions (Cerrado, North, and South)



**FIGURE 5** Detection of outlier SNP under positive selection of *Chrysodeixis includens* populations, using three different approaches. Plots showing outlier loci putatively under positive selection (red) detected by (a) Lositan ( $N = 43$ ) and (b) BayeScan (7). Manhattan plot of candidate loci that showed a significant association (red) with (c) landscape variables ( $N = 39$ ) and (d) climate variables ( $N = 7$ ) by the LMFF method. On the right-hand side, the Venn diagram shows the SNP candidate overlap between the different methods

were used to control for secondary factors, a slight increase in the correlation between the landscape distance and genetic distance was observed (Table 3).

Using a latent factor of 2 and controlling FDR at 10%, we obtained 39 candidate SNPs associated with the landscape variables, and 7 associated with climate variables, using the LFMM approach (Figure 5). The  $p$ -values of the putative candidates ranged from .002

to  $3.4 \times 10^{-6}$ , showing overlap with the Lositan and BayeScan candidates (Figure 5, Venn diagram). From the list of candidates associated with landscapes, 13 were successfully identified in a Blast search, 10 mapped, and 7 annotated. Among potential candidates, a locus associated with glucuronosyltransferase activity was recognized (UDP-glycosyltransferase). Associated with climate variables, no loci containing outlier SNP were identified.

**TABLE 3** Partial Mantel tests for partitioning of genetic variation of 43 markers under directional selection selected by Lositan and 4 markers selected by BayeScan in *Chrysodeixis includens* based on landscape variables, environmental variables, and geographic distance

Models	<i>r</i>	<i>p</i> -values
Lositan		
Landscape distance	0.31	.05 <sup>a</sup>
Environmental distance	0.19	.12
Geographic distance	0.33	.03 <sup>a</sup>
Landscape distance, controlling for environmental distance	0.35	.03 <sup>a</sup>
Landscape distance, controlling for geographic distance	0.36	.02 <sup>a</sup>
Environmental distance, controlling for landscape distance	−0.06	.64
Environmental distance, controlling for geographic distance	0.04	.32
BayeScan		
Landscape distance	0.47	.01 <sup>b</sup>
Environmental distance	0.17	.13
Geographic distance	0.20	.10
Landscape distance, controlling for environmental distance	0.47	.01 <sup>b</sup>
Landscape distance, controlling for geographic distance	0.49	.00 <sup>b</sup>
Environmental distance, controlling for landscape distance	−0.13	.85
Environmental distance, controlling for geographic distance	0.03	.32

<sup>a</sup>Significance at 5% probability.

<sup>b</sup>Significance at 1% probability.

### 3.6 | Soybean versus cotton

Principal component analysis revealed that larvae collected in cotton fields were slightly different from those collected in soybean fields in BACO (~14% of cotton, distributed in 925 patches), but not in MTCV (~5% of cotton, distributed in 1,259 patches) (Figure 6). When contrasting locations where both cotton and soybean were sampled (i.e., BACO and MTCV), BayeScan found signs of natural selection only in BACO. A total of 43 candidate loci under positive selection were identified by the BayeScan method (Figure 6a, upper-left corner), of which 19 were successfully identified by Blastx, 4 mapped, and 2 annotated. One of the loci successfully annotated is associated with proteolysis (serine-type D-Ala-D-Ala carboxypeptidase activity). Other nonannotated Blastx results included aminotransferase, amidases, and carboxypeptidase.

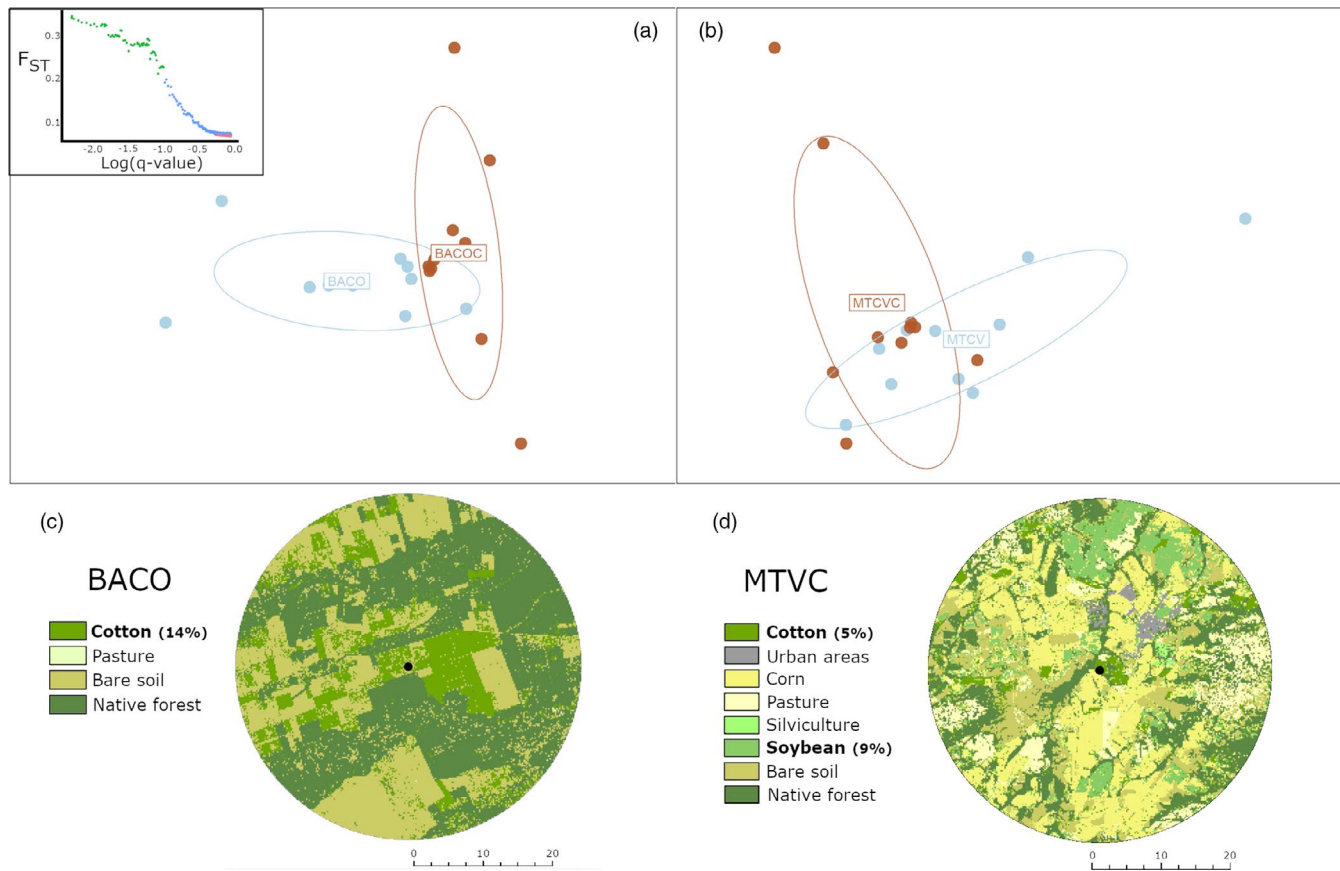
## 4 | DISCUSSION

Our results based on the mtDNA and SNP markers provided strong support for the assumption that the *C. includens* population is

undergoing demographic expansion. The results of the analysis (considering only individuals sampled in Brazil) are consistent with the history of diversification and demographic expansion observed in the Americas (i.e., similar haplotype network patterns). The demographic reconstruction traces the beginning of the *C. includens* population expansion back to the year 1,700, coinciding with the period of colonization of the Americas. However, the demographic reconstruction (i.e., Skyline analysis) must be cautiously interpreted, given that the genome-wide mutation rate for this particular species is not known (El-Shehawi & Elseehey, 2016; Keightley et al., 2014; Oppold & Pfenninger, 2017).

The pattern of low diversity and genetic structure observed with the mtDNA markers is congruent with nuclear markers and supports previous findings with ISSR markers (Palma et al., 2015). The lack of a clear genetic structure, despite the large geographic distances between sample sites, may be evidence of a recent range expansion, crop-adapted lineages, source-sink/meta-population dynamics, or even *Wolbachia*-associated lineages that could attenuate the signals of isolation by distance (IBD) (Forbes et al., 2017; Gonçalves et al., 2019; Hurst & Jiggins, 2005; Peterson & Denno, 1998). Large-scale panmixia is not commonly reported and must not be considered the default pattern, given the limitations on the dispersal of most organisms that would prevent them from interbreeding without restrictions (Ritchie, Cousins, Cregeen, & Piernney, 2013; Teixeira, Serrão, & Arnaud-Haond, 2012). Furthermore, in Brazil, we do not have reports of migratory behavior in Noctuidae moths that could increase the gene flow among geographically distant populations, but no conclusive study has been conducted to investigate this behavior in *C. includens*. However, our preliminary data of gossypol derived biomarker indicate that *C. includens* do not often engage in long-distance dispersal in Brazil (Silva, 2020). Another potentially important variable is the antagonistic interaction between *C. includens* and the entomopathogenic fungus *Nomuraea rileyi* (Sosa-Gómez, Delpin, Moscardi, & Nozaki, 2003). It is not known whether *N. rileyi* can significantly impact the population dynamics of *C. includens* on a large scale, affecting the effective population size and host interactions; however, fungicides might disrupt this interaction and favor *C. includens* in soybean systems as fungicide spraying increases to control soybean fungal diseases (Sosa-Gómez et al., 2003).

Signals of differentiation by distance or geographic regions (i.e.,  $F_{ST}$  grouping pattern) and by host (i.e., cotton and soybean) can be identified and indicate a process of differentiation driven by both genetic drift and positive selection. In our study, soybean fields were the predominant agricultural type in all landscapes, covering up to 50% of the total area in some cases (e.g., PRPI and BACO). The existing extensive soybean fields might be reducing genetic diversity in *C. includens*. If this is the case, then the levels of genetic diversity of *C. includens* may be affected by genetic drift, promoted by continuous processes of (re)colonization of soybean crops and rapid increases in population density within a short period. In a recent study, the number of moths captured by light traps in farmland in the Brazilian Cerrado was positively correlated with the



**FIGURE 6** Population differentiation by host plants (soybean versus cotton). (a) Principal component analysis (PCA) using 7,923 SNP markers of *Chrysodeixis includens* collected from soybean and cotton plants in successive growing seasons at BACO, and (b) using 10,010 SNP markers for insects collected at MTCV. The small panel (upper left) shows putative markers under positive selection when cotton versus soybean insects from BACO were contrasted. No sign of selection was found at MTCV. Ellipse colors according to host plants: cotton (brown) and soybean (blue). Figures (c) and (d) show the landscape composition of BACO and MTCV during the cotton season. Samples labeled according to the host type are inside their 95% inertia ellipses

presence and phenological stage of local soybean crops, where the period of high moth density in the area lasted only 60 days (Santos et al., 2017). This would indicate an abrupt population decrease, followed by a complete absence of insects from the area for a prolonged period of time. Depending on the scale of this phenomenon, the wide fluctuation in the number of insects can contribute to genetic drift and stochastic loss of genetic diversity. Furthermore, the number of individuals of *C. includens* has decreased in recent years at several sites (not to be confused with effective population size), along with the continuing adoption of *Bt* soybeans in Brazil (Santos et al., 2017). This effect was not captured in our study, which showed only signs of population demographic expansion, and continued monitoring of natural populations is necessary to measure the impact of *Bt* soybean crops on the effective population size of *C. includens*. Our data suggested, however, that the composition of the agricultural landscape, soybean and cotton in particular, is shaping the pattern of genetic diversity in *C. includens*. Therefore, a study of the temporal variation or meta-population could offer additional insights into how seasonal gene flow and genetic drift might be affecting the process of natural selection in this species.

#### 4.1 | Natural selection and local adaptation

We have found compelling pieces of evidence that natural selection is shaping populations of *C. includens* in Brazil. These adaptive processes can be linked to the recent change in the pest status of *C. includens*. The insects are widely distributed and therefore experience environmental and habitat-associated factors differently at different locations. Hence, both abiotic factors such as temperature and humidity, and biotic factors such as host plants can be putative drivers of the evolutionary process through natural selection. Based on the selection screenings, landscape composition seems to be a more important driver of the adaptive process than climate conditions.

Inevitably, extensive monoculture areas will not only favor the performance of certain herbivores but impose strong selection pressures on the populations of pest insects (Via, 1990; Wetzel, Kharouba, Robinson, Holyoak, & Karban, 2016). In addition to the extensive areas and the continuous succession of suitable hosts, soybean fields in the southwest and southern regions are cultivated in very intensive management systems with high input of fertilizers, reduced plant spacing, short plant cycles, and in recent years, the use of *Bt* crops. Specifically, *C. includens* is still very sensitive

to *Bt* crops and has different levels of resistance to the different pesticides used on soybean and cotton, although no control failures have been reported (Bernardi et al., 2012; Stacke et al., 2019; Yano et al., 2016). However, this scenario can become more problematic when we consider the advance of agricultural frontiers into northern Brazil, increasing the soybean-producing area even further. Considering the rapid habitat transformations in the Cerrado region in recent decades, local adaptation is an expected response for a population under such intensive selective forces in highly homogeneous areas (Mopper & Strauss, 1998). The greatest concern is that, once present, resistance to *Bt* or insecticides would spread quickly through the population. Among the proteins that may be associated with the management of pest-insect resistance and host adaptation, a UDP-glycosyltransferase protein (detected by LFMM and Lositan) has been detected (Kaplanoglu, Chapman, Scott, & Donly, 2017; Kreml et al., 2016). Further functional analysis and continued resistance monitoring may help to assess whether resistance is developing and spreading.

#### 4.2 | Host selection and agricultural landscapes

Understanding how genetic diversity is distributed over the range of a species is essential to study patterns of gene flow and to assess adaptive trends. Knowledge of the genetic structure of populations will enable us to more reliably infer which biotic and abiotic factors may be responsible for gene-pool isolation or the lack of it (Avice, 1992; Slatkin, 1987). Habitat patchiness and host plant selection and distribution can counteract the effect of gene flow, producing genetic structure at fine spatial scales (Mopper, 1996). The evolution of host-specialized races within polyphagous species is often reported (Forbes et al., 2017; Mopper, 1996; Thompson, 1988; Wood, Tilmon, Shantz, Harris, & Pesek, 1999). The formation of sympatric races in pest populations is not well understood, but extensive crop fields impose strong selection pressures on herbivory-related adaptations, causing evolutionary divergence at fine scales (Carroll & Boyd, 1992; Feder et al., 1994; Gouin et al., 2017; Via, 1991). It is difficult to predict whether the differentiation by host reported here will increase or will be eroded by gene flow in the long term, particularly in unstable environments such as the agricultural systems. Moreover, the difference between insects collected from cotton and soybean plants was evident in only one of the two sampling locations (BACOC), for which the Blast2GO results revealed that feeding behavior associated with aminotransferase, amidases, carboxypeptidase, and carboxylase might be changing. BACOC had a much less diverse landscape, where cotton crops cover up to 14% of the land, and soybean was absent during the cotton-growing period, whereas MTCVC had a more diverse landscape, where cotton covered only 5% of the total area, sharing the landscape with soybean fields. The smaller proportion of cotton in MTCVC, the greater degree of fragmentation (smaller patch sizes), and the presence of soybean could explain the different outcomes regarding the natural selection process in *C. includens* when cotton plants were the main driver. Further studies are

needed to reinforce the evidence of host-related race formation and the functional importance of the candidate genes listed here.

## 5 | CONCLUSIONS

Both the mitochondrial and SNP markers revealed a pattern of low genetic diversity and a weak or absent genetic structure in *C. includens* in Brazil. This pattern could be attributed to recurrent evolutionary processes such as gene flow, or to a historical process associated with demographic expansion and host-related adaptations. Analyses of the landscape composition suggested that the increase in the soybean crop area is likely important for the process of local adaptation, in particular in relation to host-specific adaptations. Landscape composition and configuration may explain the still-incipient differentiation between the hosts; and no effect of climate variables was detected. Our results reinforced the general recommendation for polyculture and for more diverse agroecosystems as strategies to prevent the rapid evolution of undesirable traits. Polyculture including one or more nonhost plants often limits the total number of pest insects relative to the numbers in monoculture systems, by reducing food resources and offering more opportunities for antagonistic ecological interactions with natural enemies, for instance (Risch 1981).

The present data increase our understanding of how the host range and future problems of *Bt* and pesticide resistance could affect *C. includens*. Understanding how adaptive processes shape the evolutionary trajectory of pest populations will not prevent the ongoing transformation of natural populations, but can guide us in a search for better resistance monitoring and crop-succession management over time.

## ACKNOWLEDGEMENTS

We thank São Paulo Research Foundation (FAPESP, fellowship #2017/02393-0 to EMGC and fellowship #2017/02393-0 to JBP, and grant #2014/11495-3 to ASC), by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001; and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, fellowship #140844/2016-3 to CSS and #311465/2018-8 to ASC). We thank also to R. Querino, K.D.S. Sombra, F. Hickmann, R.M. Pereira, N.A. Leite, and PROMIP (SISBIO License Number 40380-5) for helping to collect insect samples in different Brazilian regions. Research carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by FAPESP (grant 2013/07375-0).

## CONFLICT OF INTEREST

None declared.

## AUTHORS' CONTRIBUTION

ASC, CSS, EMGC, GH, PMD, and RAC conceived and designed the study. CSS collected the data. ASC, CSS, and EMGC analyzed the data. ASC, PMD, and RAC provided reagents, analytical tools, and



financial support. ASC, CSS, and EMGC wrote the manuscript. All authors read, corrected, and approved the manuscript.

## DATA AVAILABILITY STATEMENT

Detailed information regarding the parameter selection in STACKS, data curation, haplotype information, and landscape descriptions is included in the supplementary material. Raw fasta files of Illumina sequences were included in the SRA-NCBI repository (PRJNA576576).

## ORCID

Alberto S. Correa  <https://orcid.org/0000-0002-3788-7480>

## REFERENCES

- Altieri, M. A. (1999). The ecological role of biodiversity in agroecosystems. *Agriculture, Ecosystems & Environment*, 74, 19–31. [https://doi.org/10.1016/S0167-8809\(99\)00028-6](https://doi.org/10.1016/S0167-8809(99)00028-6)
- Alvarado-Serrano, D. F., Van Etten, M. L., Chang, S., & Baucom, R. S. (2019). The relative contribution of natural landscapes and human-mediated factors on the connectivity of noxious invasive weed. *Heredity*, 122, 29–40. <https://doi.org/10.1038/s41437-018-0106-x>
- Andrade, K., Bueno, A. F., da Silva, D. M., Stecca, C. S., Pasini, A., & de Oliveira, M. C. N. (2016). Bioecological characteristics of *Chrysodeixis includens* (Lepidoptera: Noctuidae) fed on different hosts. *Austral Entomology*, 55, 449–454. <https://doi.org/10.1111/aen.12208>
- Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a  $F_{ST}$ -outlier method. *BMC Bioinformatics*, 9, 323. <https://doi.org/10.1186/1471-2105-9-323>
- Arvor, D., Meirelles, M., Dubreuil, V., Bégué, A., & Shimabukuro, Y. E. (2012). Analyzing the agricultural transition in Mato Grosso, Brazil, using satellite-derived indices. *Applied Geography*, 32, 702–713. <https://doi.org/10.1016/j.apgeog.2011.08.007>
- Avise, J. C. (1992). Molecular population structure and the biogeographic history of a regional fauna: A case history with lessons for conservation biology. *Oikos*, 63, 62–76. <https://doi.org/10.2307/3545516>
- Bernardi, O., Malvestiti, G. S., Dourado, P. M., Oliveira, W. S., Martinelli, S., Berger, G. U., ... Omoto, C. (2012). Assessment of the high-dose concept and level of control provided by MON 87701 × MON 89788 soybean against *Anticarsia gemmatilis* and *Pseudoplusia includens* (Lepidoptera: Noctuidae) in Brazil. *Pest Management Science*, 68, 1083–1091. <https://doi.org/10.1002/ps.3271>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Calderón, L., Campagna, L., Wilke, T., Lormee, H., Eraud, C., Dunn, J. C., ... Quilfeldt, P. (2016). Genomic evidence of demographic fluctuations and lack of genetic structure across flyways in a long distance migrant, the European turtle dove. *BMC Evolutionary Biology*, 16, 237. <https://doi.org/10.1186/s12862-016-0817-7>
- Carroll, S. P., & Boyd, C. (1992). Host race radiation in the soapberry bug: Natural history with the history. *Evolution*, 46, 1052–1069. <https://doi.org/10.1111/j.1558-5646.1992.tb00619.x>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, 12, 693–715. <https://doi.org/10.1111/j.1461-0248.2009.01314.x>
- DeFries, R., Herold, M., Verchot, L., Macedo, M. N., & Shimabukuro, Y. (2013). Export-oriented deforestation in Mato Grosso: Harbinger or exception for other tropical forests? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368, 20120173. <https://doi.org/10.1098/rstb.2012.0173>
- dos Santos, S. R., Specht, A., Carneiro, E., de Paula-Moraes, S. V., & Casagrande, M. M. (2017). Interseasonal variation of *Chrysodeixis includens* (Walker, [1858]) (Lepidoptera: Noctuidae) populations in the Brazilian Savanna. *Revista Brasileira De Entomologia*, 61, 294–299. <https://doi.org/10.1016/j.rbe.2017.06.006>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15. citeulike-article-id:678648
- Dray, S., & Dufour, A.-B.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22, 1–20. <https://doi.org/10.1.1.177.8850>
- Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29, 1969–1973. <https://doi.org/10.1093/molbev/mss075>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A web-site and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- El-Shehawi, A. M., & Elseehy, M. M. (2016). Estimation of genome evolution time by mutation rate. *Journal of Phylogenetics & Evolutionary Biology*, 4, 172. <https://doi.org/10.4172/2329-9002.1000172>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, 6, e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Excoffier, L., Laval, G., & Schneider, S. (2006). ARLEQUIN version 3.1: A software for population genetic data analysis. Berne, Switzerland: Computational and Molecular Population Genetics Laboratory, Institute of Zoology, University of Berne.
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.
- Feder, J. L., Opp, S. B., Wlazlo, B., Reynolds, K., Go, W., & Spisak, S. (1994). Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 7990–7994. <https://doi.org/10.1073/pnas.91.17.7990>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Forbes, A. A., Devine, S. N., Hippee, A. C., Tvedte, E. S., Ward, A. K. G., Widmayer, H. A., & Wilson, C. J. (2017). Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*, 71, 1126–1137. <https://doi.org/10.1111/evo.13164>
- Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, 6, 925–929. <https://doi.org/10.1111/2041-210X.12382>
- Gonçalves, R. M., Mastrangelo, T., Rodrigues, J. C. V., Paulo, D. F., Omoto, C., Corrêa, A. S., & de Azeredo-Espin, A. M. L. (2019). Invasion origin, rapid population expansion, and the lack of genetic structure of cotton bollworm (*Helicoverpa armigera*) in the Americas. *Ecology and Evolution*, 9, 7378–7401. <https://doi.org/10.1002/ece3.5123>
- Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, 22, 1–19. citeulike-article-id:12008924
- Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., ... Conesa, A. (2008). High-throughput functional

- annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, 36, 3420–3435. <https://doi.org/10.1093/nar/gkn176>
- Gouin, A., Bretaudeau, A., Nam, K., Gimenez, S., Aury, J.-M., Duvic, B., ... Fournier, P. (2017). Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. *Scientific Reports*, 7, 11816. <https://doi.org/10.1038/s41598-017-10461-4>
- Hurst, G. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: The effects of inherited symbionts. *Proceedings of the Royal Society B: Biological Sciences*, 272, 1525–1534. <https://doi.org/10.1098/rspb.2005.3056>
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jung, M. (2016). LecoS — A python plugin for automated landscape ecology analysis. *Ecological Informatics*, 31, 18–21. <https://doi.org/10.1016/j.ecoinf.2015.11.006>
- Kaplanoglu, E., Chapman, P., Scott, I. M., & Donly, C. (2017). Overexpression of a cytochrome P450 and a UDP-glycosyltransferase is associated with imidacloprid resistance in the Colorado potato beetle, *Leptinotarsa decemlineata*. *Scientific Reports*, 7, 1762. <https://doi.org/10.1038/s41598-017-01961-4>
- Kassambara, A. (2017). *Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra*. Scotts Valley, CA: STHDA, CreateSpace.
- Keightley, P. D., Ness, R. W., Halligan, D. L., & Haddrill, P. R. (2014). Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics*, 196, 313–320. <https://doi.org/10.1534/genetics.113.158758>
- Kogan, M., Turnipseed, S. G., Shepard, M., De Oliveira, E. B., & Borgo, A. (1977). Pilot insect pest management program for soybean in southern Brazil. *Journal of Economic Entomology*, 70, 659–663. <https://doi.org/10.1093/jee/70.5.659>
- Krempl, C., Sporer, T., Reichelt, M., Ahn, S.-J., Heide-Fischer, H., Vogel, H., ... Joußen, N. (2016). Potential detoxification of gossypol by UDP-glycosyltransferases in the two Heliothine moth species *Helicoverpa armigera* and *Heliothis virescens*. *Insect Biochemistry and Molecular Biology*, 71, 49–57. <https://doi.org/10.1016/j.ibmb.2016.02.005>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Landis, D. A. (2017). Designing agricultural landscapes for biodiversity-based ecosystem services. *Basic and Applied Ecology*, 18, 1–12. <https://doi.org/10.1016/j.baae.2016.07.005>
- Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299. <https://doi.org/10.1093/bioinformatics/btr642>
- Lu, Y., Wu, K., Jiang, Y., Xia, B., Li, P., Feng, H., ... Guo, Y. (2010). Mirid bug outbreaks in multiple crops correlated with wide-scale adoption of Bt cotton in China. *Science*, 328, 1151–1154. <https://doi.org/10.1126/science.1187881>
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220. <https://doi.org/10.1038/214637b0>
- McGarigal, K., Cushman, S. A., Neel, M. C., & Ene, E. (2012). FRAGSTATS v4: Spatial pattern analysis program for categorical and continuous maps. Amherst, MA: University of Massachusetts. Retrieved from <http://www.umass.edu/landeco/research/fragstats/fragstats.html>
- Monteiro, F. A., Pérez, R., Panzera, F., Dujardin, J.-P., Galvão, C., Rocha, D., ... Beard, C. B. (1999). Mitochondrial DNA variation of *Triatoma infestans* populations and its implication on the specific status of *T. melanosoma*. *Memórias do Instituto Oswaldo Cruz*, 94, 229–238. <https://doi.org/10.1590/S0074-02761999000700037>
- Mopper, S. (1996). Adaptive genetic structure in phytophagous insect populations. *Tree*, 11, 235–238. [https://doi.org/10.1016/0169-5347\(96\)10036-7](https://doi.org/10.1016/0169-5347(96)10036-7)
- Mopper, S., & Strauss, S. Y. (1998). *Genetic structure and local adaptation in natural insect populations*. Dordrecht, The Netherlands: Springer.
- Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A. J., & Hostert, P. (2015). Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. *Remote Sensing of Environment*, 156, 490–499. <https://doi.org/10.1016/j.rse.2014.10.014>
- Oppold, A. M., & Pfenninger, M. (2017). Direct estimation of the spontaneous mutation rate by short-term mutation accumulation lines in *Chironomus riparius*. *Evolution Letters*, 1, 86–92. <https://doi.org/10.1002/evl3.8>
- Palma, J., Maebe, K., Guedes, J. V. C., & Smagghe, G. (2015). Molecular variability and genetic structure of *Chrysodeixis includens* (Lepidoptera: Noctuidae), an important soybean defoliator in Brazil. *PLoS ONE*, 10, e0121260. <https://doi.org/10.1371/journal.pone.0121260>
- Panizzi, A. R. (2013). History and contemporary perspectives of the integrated pest management of soybean in Brazil. *Neotropical Entomology*, 42, 119–127. <https://doi.org/10.1007/s13744-013-0111-y>
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for STACKS. *Methods in Ecology and Evolution*, 8, 1360–1373. <https://doi.org/10.1111/2041-210X.12775>
- Peterson, M. A., & Denno, R. F. (1998). The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *The American Naturalist*, 152, 428–446. <https://doi.org/10.1086/286180>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959. <https://doi.org/10.1111/j.1471-8286.2007.01758.x>
- Ribeiro, M. C., Metzger, J. P., Martensen, A. C., Ponzoni, F. J., & Hirota, M. M. (2009). The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, 142, 1141–1153. <https://doi.org/10.1016/j.biocon.2009.02.021>
- Risch, S. J. (1981). Insect herbivore abundance in tropical monocultures and polycultures: An experimental test of two hypotheses. *Ecology*, 62, 1325–1340. <https://doi.org/10.2307/1937296>
- Ritchie, H., Cousins, N. J., Cregeen, S. J., & Piernney, S. B. (2013). Population genetic structure of the abyssal grenadier (*Coryphaenoides armatus*) around the Mid-Atlantic Ridge. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 98, 431–437. <https://doi.org/10.1016/j.dsr.2013.06.014>
- Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, 12(12), 2640. <https://doi.org/10.1038/nprot.2017.123>
- Rosenberg, N. A. (2004). DISTRICT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4, 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.
- Silva, C. S. (2020). *Historical and contemporary population dynamics of Chrysodeixis includens (Lepidoptera: Noctuidae) in Brazil (in Portuguese)*. Ph.D. Dissertation, Universidade De São Paulo, Escola Superior De Agricultura “Luiz De Queiroz” (ESALQ-USP).
- Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, 87, 651–701. <https://doi.org/10.1093/aesa/87.6.651>

- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236, 787–792. <https://doi.org/10.1126/science.3576198>
- Sosa-Gómez, D. R., Delpin, K. E., Moscardi, F., & Nozaki, M. D. H. (2003). The impact of fungicides on *Nomuraea rileyi* (Farlow) Samson epizootics and on populations of *Anticarsia gemmatilis* Hübner (Lepidoptera: Noctuidae), on soybean. *Neotropical Entomology*, 32, 287–291. <https://doi.org/10.1590/S1519-566X2003000200014>
- Stacke, R. F., Giacomelli, T., Bronzatto, E. S., Halberstadt, S. A., Garlet, C. G., Muraro, D. S., ... Bernardi, O. (2019). Susceptibility of Brazilian Populations of *Chrysodeixis includens* (Lepidoptera: Noctuidae) to Selected Insecticides. *Journal of Economic Entomology*, 112, 1378–1387. <https://doi.org/10.1093/jee/toz031>
- Teixeira, S., Serrão, E. A., & Arnaud-Haond, S. (2012). Panmixia in a fragmented and unstable environment: The hydrothermal shrimp *Rimicaris exoculata* disperses extensively along the Mid-Atlantic Ridge. *PLoS ONE*, 7, e38521. <https://doi.org/10.1371/journal.pone.0038521>
- Thompson, J. N. (1988). Variation in preference and specificity in monophagous and oligophagous swallowtail butterflies. *Evolution*, 42, 118–128. <https://doi.org/10.2307/2409006>
- Via, S. (1990). Ecological genetics and host adaptation in herbivorous insects: The experimental study of evolution in natural and agricultural systems. *Annual Review of Entomology*, 35, 421–446. <https://doi.org/10.1146/annurev.en.35.010190.002225>
- Via, S. (1991). Specialized host plant performance of pea aphid clones is not altered by experience. *Ecology*, 72, 1420–1427. <https://doi.org/10.2307/1941114>
- Weir, B. S. (1996). *Genetic data analysis II: Methods for discrete population genetic data*. Sunderland, MA: Sinauer Associates.
- Wetzel, W. C., Kharouba, H. M., Robinson, M., Holyoak, M., & Karban, R. (2016). Variability in plant nutrients reduces insect herbivore performance. *Nature*, 539, 425. <https://doi.org/10.1038/nature20140>
- Wood, T. K., Tilmon, K. J., Shantz, A. B., Harris, C. K., & Pesek, J. (1999). The role of host-plant fidelity in initiating insect race formation. *Evolutionary Ecology Research*, 1, 317–332.
- Yano, S. A., Specht, A., Moscardi, F., Carvalho, R. A., Dourado, P. M., Martinelli, S., ... Sosa-Gómez, D. R. (2016). High susceptibility and low resistance allele frequency of *Chrysodeixis includens* (Lepidoptera: Noctuidae) field populations to Cry1Ac in Brazil. *Pest Management Science*, 72, 1578–1584. <https://doi.org/10.1002/ps.4191>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Silva CS, Cordeiro EMG, de Paiva JB, et al. Population expansion and genomic adaptation to agricultural environments of the soybean looper, *Chrysodeixis includens*. *Evol Appl*. 2020;00:1–15. <https://doi.org/10.1111/eva.12966>