Insect science

Insect Science (2024) 0, 1–14, DOI 10.1111/1744-7917.13320

ORIGINAL ARTICLE

Assessing genome-wide adaptations associated with range expansion in the pink rice borer, *Sesamia inferens*

Hongran Li[1,#], Yan Peng[1,#], Chao Wu[1,#], Zhimin Li[1], Luming Zou[1], Kaikai Mao[1], Junfen Ping[1], Ryan Buck[2,3], Scott Monahan[2], Arun Sethuraman[2] and Yutao Xiao[1]

[1]*Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Gene Editing Technologies (Hainan), Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong Province, China;* [2]*Department of Biology, San Diego State University, CA, USA and* [3]*Department of Ecology & Evolutionary Biology, University of California, Los Angeles, CA, USA*

Abstract Understanding the genetic basis of adaptive evolution following habitat expansion can have important implications for pest management. The pink rice borer (PRB), *Sesamia inferens* (Walker), is a destructive pest of rice that was historically restricted to regions south of 34° N latitude in China. However, with changes in global climate and farming practices, the distribution of this moth has progressively expanded, encompassing most regions in North China. Here, 3 highly differentiated subpopulations were discovered using high-quality single-nucleotide polymorphism and structural variant datasets across China, corresponding to northern, southern China regions, and the Yunnan-Guizhou Plateau, with significant patterns of isolation by geographic and environmental distances. Our estimates of evolutionary history indicate asymmetric migration with varying population sizes across the 3 subpopulations. Selective sweep analyses estimated strong selection at insect cuticle glycine-rich cuticular protein genes which are associated with enhanced desiccation adaptability in the northern group, and at the histone-lysine-N-methyltransferase gene associated with range expansion and local adaptation in the Shandong population. Our findings have significant implications for the development of effective strategies to control this pest.

Key words desiccation adaptability; evolutionary history; genetic adaptation; phylogeography; *Sesamia inferens*

Introduction

Exploring the relative contributions of demography and local adaptation on spatial patterns of genetic variation has been a recurrent and crucial theme in evolutionary biology (Sork, 2016; Gao *et al.*, 2017; Wang *et al.*, 2020). Evolution by natural selection may be best evidenced by the existence of local adaptation, that is, when individuals have a higher fitness in their local environment than individuals from elsewhere (Bessette *et al.*, 2022; Mérot, 2022). Specifically, the selection for ecological characters related to environmental niches can drive the rapid accumulation of adaptive genetic differentiation in pest populations (Szűcs *et al.*, 2017). However, an in-depth understanding of the role of natural selection in shaping the patterns of adaptive evolution is still poorly understood in many agricultural pests, yet important for

Correspondence: Yutao Xiao, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Gene Editing Technologies (Hainan), Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518116, Guangdong Province, China. Email: xiaoyutao@caas.cn

[#]These authors contributed equally to this work.

sustainable pest management. With advances in genomic sequencing technologies, it is now becoming increasingly feasible to generate population genomic data that can serve as a complementary strategy to examine local adaptation throughout the geographical range of a targeted species (Ding *et al.*, 2018; Cheng *et al.*, 2021). Although there may be millions of variants across the genome within any specific species, only some of these variants are expected to be related to local climactic adaptations (Savolainen *et al.*, 2013). The process of identifying climate-associated genetic variation is not only critical for resolving fine-scale patterns of local adaptation but also facilitates a more mechanistic understanding of how species respond to climate change. This is particularly important while studying agricultural pests, in which the effective management strategies are designed and adopted based on studies of genetic divergence, local adaptation, and spread (Bay *et al.*, 2018; Capblancq *et al.*, 2020; Chen *et al.*, 2021).

The pink rice borer (PRB), *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae), is a polyphagous pest which damages a wide range of food crops including rice, maize sorghum, and barley (Mahesh *et al.*, 2014; Baladhiya *et al.*, 2018). The PRB is widely distributed in Asian countries such as China, Japan, India, Laos, and Pakistan (Karim & Riazuddin, 1999; Reddy *et al.*, 2003; Li *et al.*, 2011; Xu *et al.*, 2011; Chai & Du, 2012; Sun *et al.*, 2014), and was recently discovered in Hawaii and Guam (United States) according to the Center for Agriculture and Bioscience International (CABI) (https://www.cabi.org/isc/datasheet/49751#REFDDB-202162). In China, the natural range of this species was traditionally confined to areas south of 34° N latitude (Zhang, 1965). During winter, the PRB predominantly overwinters as mature larvae on the roots of rice, water bamboo, and various weed species (Tang *et al.*, 2022). To our knowledge, the PRB exhibits limited ability for migratory flight, as 75.5% of the moths had a cumulative flight distance of $\leq$5.0 km (Han *et al.*, 2012). However, due to global climate change and increased agriculture across China, the distribution of the moth has progressively exceeded its traditional limit of 34° N and expanded to include a majority of regions in North China (Chen & Lu, 2015; Tang *et al.*, 2022). Due to the overwintering ability being one of the main factors restricting the distribution of the moth, the climactic differences between the high humidity Pearl River basin and the low humidity northern regions have significant research value in understanding the impact on the overwintering process of the moth. Therefore, it is particularly important to study the desiccation adaptability of overwintering larvae of the moth in low-humidity environments in North China.

While the severe damage to rice and other crops caused by PRB has gained significant attention, the current pest management efforts primarily focus on field dynamics and forecast, environmental adaptation, biological control and pesticides resistance assessment (Jin *et al.*, 2014; Sun *et al.*, 2014; Yang *et al.*, 2014; Li *et al.*, 2015; Wu *et al.*, 2018; Mansoor *et al.*, 2019; Huang *et al.*, 2020). The knowledge of the population structure and genetic bases of the rapid adaptation of PRB in its expanded habitat are limited. Only a few studies have been conducted to explore the genetic diversity of PRB, using a smaller number of molecular markers (Yao *et al.*, 2008; Zhang *et al.*, 2013). Furthermore, 2 prominent phylogeographic clusters were identified, representing the northern and southern regions of China, respectively (Tang *et al.*, 2014, 2022). These findings were based on mitochondrial DNA or microsatellite markers, which either reflected the maternal history or were limited by a lack of sufficient genetic information.

Here, we conducted whole genome resequencing on 143 individuals of PRB across 14 natural populations in various habitats within China. We analyzed genome-wide variation encompassing single-nucleotide polymorphisms (SNPs) and large structural variants (SVs). Our objectives using these genomic datasets were to: (1) infer the spatial patterns of genetic diversity, population structure, and phylogeography; (2) reveal the demographic dynamics and movement patterns of *S. inferens* populations; and (3) identify putative genes under selection that are likely involved in their evolutionary adaptation to local environments. These data should help us understand genetic mechanisms behind the PRB's evolutionary adaptations across a variety of habitats and agro-ecosystems spanning China, with implications for its future impact and control.

Materials and methods

Sample preparation

Pink rice borer populations were mainly collected from rice fields, covering almost all their habitats across China. A total of 143 PRB individuals were collected in 14 geographic locations (Table 1). All PRB individuals were identified by morphological features and the molecular marker *mtCOI* (Tang *et al.*, 2022). The larva can be readily identified within the rice stem by its pinkish hue, smooth texture, and full-grown length of approximately 30 mm (Das & Padmaja, 2016). The pink rice borer samples were placed in 95% ethanol and stored at −80 °C for later resequencing.

Table 1 Details of sampling collections of the pink rice borer, *Sesamia inferens* in different geographical sites across China.

Genetic group	Code	Collection site	Longitude (°E)	Latitude (°N)	No. of individuals	Collection year
NCN	SDLY	Linyi, Shandong	118.21	34.35	13	2021
	SDDY	Dongying, Shandong	118.52	37.59	12	2021
	JSJY	Jiangyan, Jiangsu	120.04	32.41	8	2022
	HNXY	Xinyang, Henan	113.89	32.24	12	2022
	ZJJX	Jiaxing, Zhejiang	119.92	30.08	9	2022
	SXHZ	Hanzhong, Shannxi	107.11	33.17	4	2022
	HBJZ	Jingzhou, Hubei	112.03	30.34	17	2022
	HNCS	Changsha, Hunan	113.05	28.18	6	2022
GZCN	GZGY	Guiyang, Guizhou	106.58	26.00	12	2022
	GZAS	Anshun, Guizhou	106.08	26.15	11	2022
SCN	GXNN	Nanning, Guangxi	108.33	23.03	11	2022
	GDSZ	Shenzhen, Guangdong	114.48	22.59	16	2022
	HNSY	Sanya, Hainan	109.12	18.37	9	2022
	YNKM	Kunming, Yunnan	103.25	25.56	3	2022
	Total				143	

GZCN, Guizhou Province, China; NCN, northern China; SCN, southern China.

DNA extraction

Genomic DNA was extracted using a Qiagen DNA purification kit (Qiagen, Valencia, CA, USA), following the manufacturer's instructions. DNA quantity and quality controls were validated by Qubit2.0, Agilent 2100, and quantitative polymerase chain reaction (PCR) method. The extracted genomic DNA was stored at −80 °C until population genome resequencing. Short insertion fragment libraries of 300–500 bp were constructed at BGI (Beijing Genomics Institute)-Shenzhen, then sequenced on the BGISEQ-500 platform utilizing 150-bp paired-end reads, following standard protocols.

SNP and SV calling

All clean reads were filtered using SOAPnuke (v1.5.6) software with parameters (-nRate 0.1 - qualRate 0.5 - -lowQua 12 - qualSys 2) and then mapped to the reference genome of *S. inferens* using BWA v0.7.17 with the following parameters: "bwa mem - Mark secondary alignments (-M) - Read Group information (-R)." (Li, 2013; Chen *et al.*, 2018). These reads were sorted using SAMtools (v1.7) (Li *et al.*, 2009) software. Duplicate reads were removed using Picard (http://broadinstitute.github.io/picard/). The Genome Analysis Toolkit (GATK) v4.2.3.0 (McKenna *et al.*, 2010) was used to sort the alignment results and remove PCR duplicate reads. The HaplotypeCaller was used to identify SNPs for each individual and generate single GVCF (genomic variant call format) files. The GVCF files were merged into a VCF file by the CombineGVCFs command. SNPs were filtered using a custom script and then hard filtered using the VariantFiltration command of GATK. The filtration standard was "QD < 2.0 || MQ < 40.0 || FS > 60.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0". To further obtain high-quality SNPs, we used VCFTools v0.1.16 (Danecek *et al.*, 2011) to further filter with the following parameter: bi-allelic SNPs with missing data rate less than 80%, minor allele frequency (MAF) greater than 0.01. We employed SnpEff v4.3t (Cingolani *et al.*, 2012) for SNP annotation to classify SNPs into exons, introns, intergenic regions, and upstream or downstream regions.

SV calling was performed with Delly v1.1.6 (Rausch *et al.*, 2012) using mapping results in BAM (binary alignment map) format from resequencing data. First, the SV calling was run on each individual from scratch, and then the results were merged into 1 VCF file. Second, the SV calling was run again guided by the combined VCF file. After combining all samples of SVs using BCFTools v1.13 (Danecek & McCarthy, 2017), we retained SVs with PASS tag and length greater than 50 bp for further analysis. Translocations were excluded due to their potential uncertainty called from short reads of Illumina sequencing technology (Yang *et al.*, 2022). We further filtered with a missing rate ≥ 20% to make sure the accu-

racy of SVs using VCFtools (v0.1.16) (Danecek *et al.*, 2011).

Phylogenetic and population structure analysis

Both SV and SNP dataset of 143 samples were generated for phylogenetic and population structure analyses. A maximum likelihood (ML) tree was constructed using FastTree (v2.1.10) with GTR (generalized time-reversible) model and visualized with Interactive Tree of Life (iTOL) v6 (Letunic & Bork, 2021), with 1 000 rapid bootstrap replicates conducted to determine branch confidence values. Population structure was investigated using ADMIXTRUE (v1.3.0) and the cluster number K value was set from 1 to 10 (Alexander *et al.*, 2009). Following the description in Yang *et al.* (2022), the most suitable number of subpopulations was identified by selecting the K value associated with the smallest CV error. Additionally, individuals exhibiting a primary genetic component of less than 60% were excluded from subsequent analyses on genetic diversity, differentiation, and selection.

Genetic diversity and differentiation

We randomly selected 20 individuals from each subpopulation using python library random (https://docs.python.org/3/library/random.html) to calculate nucleotide diversity (π), population genetic differentiation (F_{ST}), with 10-kb non-overlapping windows using VCFtools (v0.1.16) (Danecek *et al.*, 2011). To further infer population-level splits and mixtures for *S. inferens* populations from China, we used Treemix 1.13 to investigate the admixture and gene flow events between populations, with migration edges from 1 to 20 (Pickrell & Pritchard, 2012).

Isolation by distance and environment

For each sampling location, we obtained environmental data from publicly available databases. These variables included 19 climate variables downloaded from WorldClim (Table S1) (Fick & Hijmans, 2017). For all locations with at least 5 individuals remaining after filtering for missing data (2 locations were removed), we calculated pairwise F_{ST} across all quality filtered SNPs using the R package hierfstat (Goudet, 2005). We calculated pairwise geographic distances from longitude and latitude using the R package geosphere. A measure of environmental distance between each pair of sites was calculated by first scaling and centering each environmental variable to account for differences in magnitude, then calculating pairwise Euclidean differences between sites. To reduce the impact of group size differences, we selected populations with more than 4 individuals to calculate genetic distance ($F_{ST}/1–F_{ST}$). To analyze the relationship between genetic distance ($F_{ST}/1–F_{ST}$), geographic distance (km) and environmental distance in populations, we performed a Mantel test using the package vegan (https://cran.r-project.org/web/packages/vegan/index.html) in R version 4.0.2 (R Core Team, 2020). The significance of the correlation was estimated based on 9 999 permutations.

Estimation of evolutionary demographic history

First, SMC++ was designed to analyze a large number of samples simultaneously while requiring only unphased genomes (Terhorst *et al.*, 2017). We utilized SMC++ to infer effective population size histories and divergence time based on the unphased high-quality SNP datasets. All samples from each population of *S. inferens* were included to run the SMC++, respectively. The -m parameter was used to mask the uncalled regions in the VCF file, which was filtered without the MAF parameter. A generation time of 0.1 year and a mutation rate of 0.84×10^{-8} mutations per nucleotide per year were adopted to convert coalescence generations into time in years (Haag-Liautard *et al.*, 2007).

We then explored models of evolutionary demographic history of the most supported clustered subpopulations determined by the phylogenetic and ADMIXTURE analyses, which suggested clustering of PRB individuals into 3 geographical subpopulations – NCN (Northern China), SCN (southern China), and GZCN (Guizhou Province, China). We designed 4 topologies using these geographical clusters, shown here as Newick trees – (1) ((NCN, SCN), GZCN), (2) ((GZCN, SCN), NCN), (3) ((NCN, GZCN), SCN), and (4) (NCN, GZCN, SCN), such that GZCN was derived from both NCN and SCN. Each topology was also tested using an Isolation model (no gene flow between populations), and an Isolation with migration model (asymmetric gene flow between populations), to a total of 8 demographic scenarios (Fig. S1) in *fastsimcoal* 2.7 and used easySFS (https://github.com/isaacovercast/easySFS) to generate folded 2-dimensional site frequency spectra (SFS) (Excoffier *et al.*, 2013; Excoffier *et al.*, 2021). Each demographic model was replicated 100 times using 40 expectation–conditional maximization (ECM) cycles, a log precision of 18, and 200 000 simulations to calculate the likelihood. A generation time range of 5 generations/year (Keightley

et al., 2015) was used to convert output parameter estimates (e.g., 100 generations = 20 years ago) and a mutation rate of 2.9×10^{-9} per site per generation was implemented as an average of the reported nuclear mutation rate for *Heliconius* (Keightley *et al.*, 2015). Effective population sizes were bounded at 2×10^6 (Keightley *et al.*, 2015) and divergence times were bounded at 515 000 years (Tang *et al.*, 2022). We used Akaike's information criterion (AIC) to find the most probable model given the observed data. Confidence intervals on parameter estimates for the best-supported model were then generated using parametric bootstrapping. Briefly, we simulated 100 bootstrap replicate datasets from the best fitting model, thereon these 100 SFS were then run in *fsc*2.7 using the same parameters and conditions as the original runs. The estimates from the best runs (highest likelihood) out of those replicated were then used to compute confidence intervals.

Genome scan for selective signals

To identify the genomic regions potentially undergoing strong selective sweeps during adaptation to local environments, 2 statistical methods were employed to identify selective signals based on high-quality biallelic SNPs. First, the population branching statistic (PBS) approach was used to identify incomplete selective sweeps over short divergence times by comparing the pairwise F_{ST} (Weir & Cockerham, 1984) with 10-kb windows sizes using VCFtools (Danecek *et al.*, 2011) version 0.1.14100. The PBS was calculated as follows:

$(T^{TP-CP}+TP^{TP-CO}-T^{CP-CO})/2$, where T represents the population divergence time in units scaled by the population size, which is negative log transformed $(1-F_{ST})$ between 2 populations. TP means the targeted population; CP means the control population; and CO means the outgroup. The windows were selected as the candidate region then the PBS value of the comparative windows was significant ($P < 0.01$, *Z*-test). The formula utilized for this computation thus was: $(T^{NCN-GZCN}+TP^{NCN-SCN}-T^{GZCN-SCN})/2$.

Second, we used RAiSD (Alachiotis *et al.*, 2018) with a window size of 10-kb to detect candidate genomic regions under positive selection for 3 populations (NCN, GZCN, SCN). RAiSD identifies outlier genomic regions while accounting for genetic diversity, the SFS, and the level of linkage disequilibrium. Outlier loci were then identified at the top 1% quantile.

Given that the moth species can typically survive the winter in Shandong Province and has undergone significant adaptive evolution in recent decades, we further employed SweeD (v3.3.1) analysis to examine signatures of positive selection in the SDDY and SDLY populations. These 2 populations occur north of 34° N latitude in China. To compute likelihood in 10-kb windows on across chromosomes, proper grid values were selected for each chromosome. We considered the highest 1% likelihood genomic regions as being under positive selection for each subpopulation (Pavlidis *et al.*, 2013).

Results

Genomic polymorphism

To estimate genomic variation in *S. inferens* across China, a total of 143 individuals were sequenced across China, which generated 683 Gb of clean PE150 paired-end data. The average sequencing depth of these samples was 27.5×. Initially, we acquired over a hundred million SNPs from all samples. After filtering, a total of 1 048 575 high-quality SNPs were identified, of which 43.30% SNPs were located in intergenic regions, 33.82% were in intronic regions, and 10.04% and 9.03% were located in upstream and downstream regions. We obtained a set of 18 743 SVs that were larger than 50 bp, including deletion (DEL), duplication (DUP) and insertion (INS) and inversions (INV), of which the DEL accounted for the majority of SVs (94.3%) (Fig. S2A, B; Table S2).

Population genetic structure

To clarify the phylogenetic relationship and population genetic structure of different geographical populations of *S. inferens*, ADMIXTURE was applied to both SNP and SV datasets, and the results were highly consistent (Fig. 1B, C). All populations were divided into 3 groups, which corresponded to their geographical distribution, namely, NCN, GZCN, and SCN (Figs. S3–S5). Phylogenetic trees constructed with SNP and SV datasets showed similar phylogenetic relationships for the main branches, and good correspondence to the major genetic groups of ADMIXTURE results (Fig. 1A, D). In addition, all collected populations formed 3 separate clades, which were also supported by principal-component analyses (PCA) based on the SNP and SV datasets (Fig. 1E, F).

Genetic differentiation and gene flow

To investigate genetic differentiation among the identified groups, we calculated genetic distances (F_{ST}) between 3 groups. The F_{ST} value between NCN and

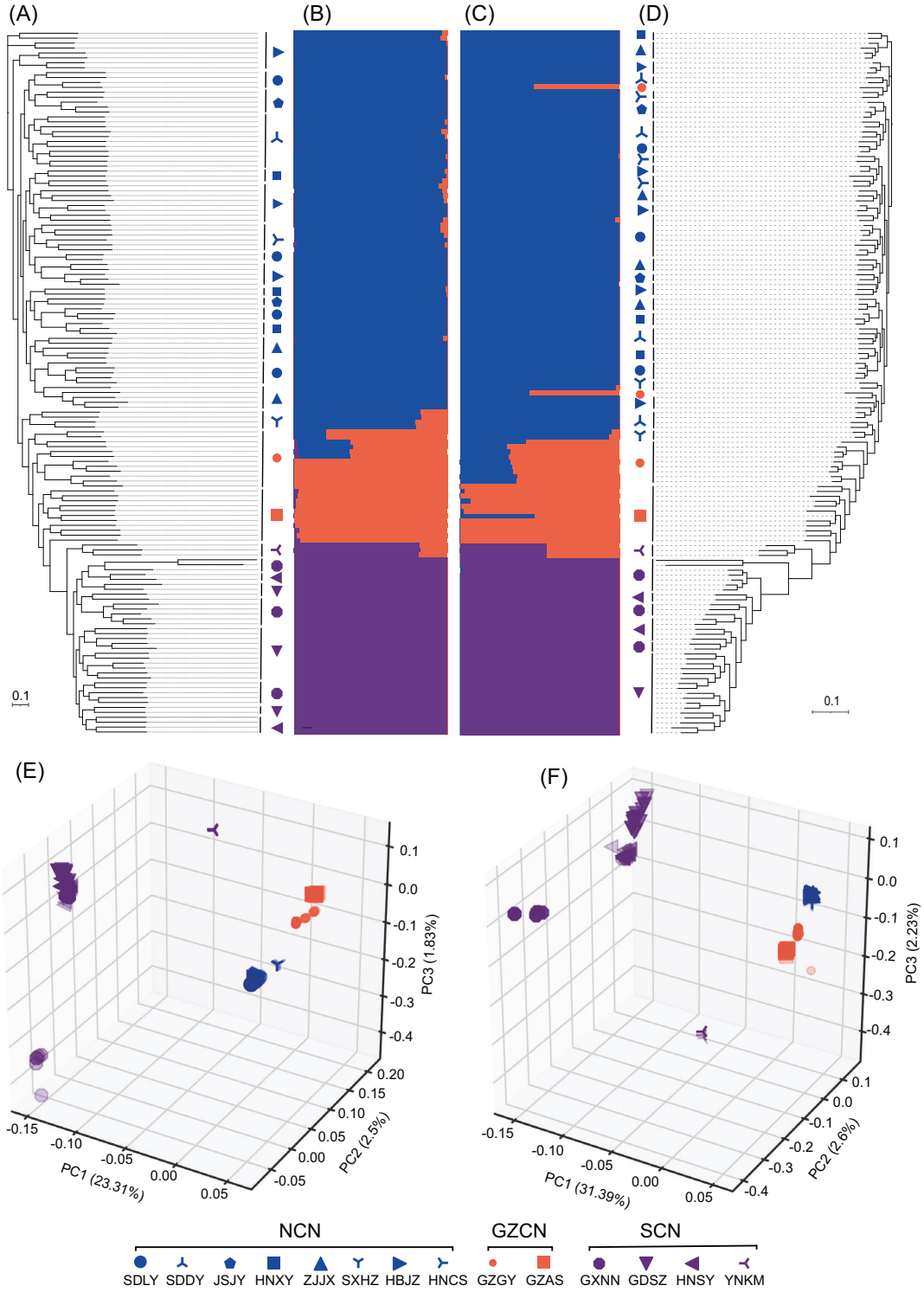

Fig. 1 Population genomic analyses of 14 representative collections in the pink rice borer, *Sesamia inferens* based on SNPs and SVs. (A) SNP-based phylogenetic tree. (B) SNP-based ADMIXTURE analysis at K = 3. (C) SV-based ADMIXTURE analysis at K = 3. (D) SV-based phylogenetic tree. (E) SNP-based PC analysis. (F) SV-based PC analysis. Colors and shapes indicate the genetic groups in *S. inferens*. PC, principal-component; SNP, single-nuclseotide polymorphism; SV; structural variant.

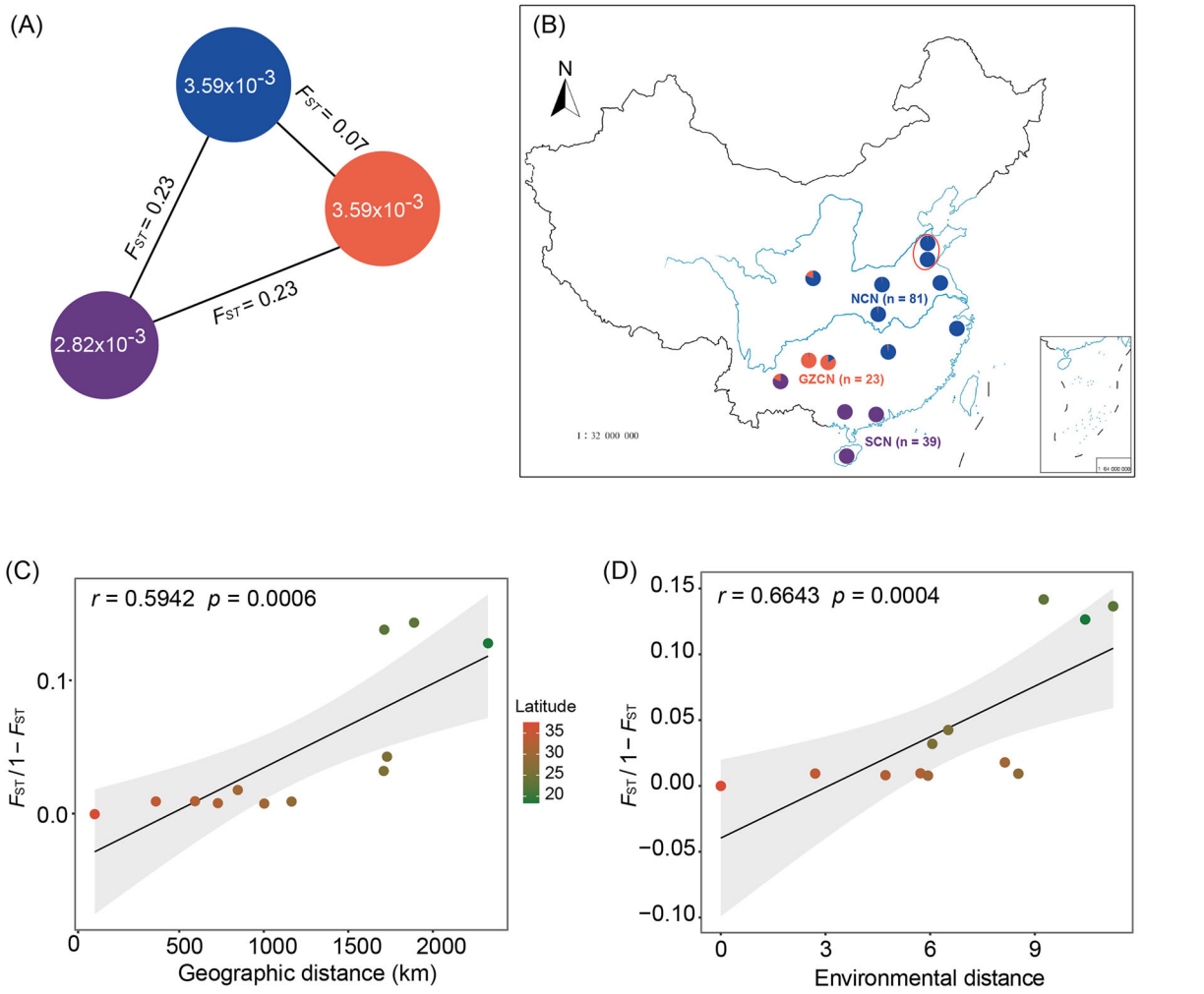

Fig. 2 Geographic and genetic distribution of the pink rice borer, *Sesamia inferens* in China. (A) Genetic diversity (π) and genetic differentiation (F_{ST}) across groups, where color indicates phylogenetic group; radius of pie indicates genetic diversity; and dashed line length indicates F_{ST} value between the 2 groups. (B) Geographic distribution of 14 populations where colors represent subpopulation structure inferred by ADMIXTURE (according to the substructure at K = 3). The red shadow circle represents the expanded habitat distribution in North China in recent years. (C) Pairwise genetic distance ($F_{ST}/1–F_{ST}$) is associated with geographical distance using isolation-by-distance (IBD) analysis (Mantel test, 2-sided), the shadow of linear regression denotes the 95% confidence interval. (D) Pairwise genetic distance ($F_{ST}/1–F_{ST}$) is associated with environmental distance using isolation-by-environment (IBE) analysis (Mantel test, 2-sided); the shadow of linear regression denotes the 95% confidence interval. GS (2023)2453.

GZCN ($F_{ST} = 0.09$) was smaller than that between SCN with NCN ($F_{ST} = 0.27$) or GZCN ($F_{ST} = 0.24$), suggesting that the GZCN group has the closest genetic distance with the NCN group (Fig. 2A). This result is consistent with the phylogenetic analysis (Fig. 1). We further investigated genetic diversity by calculating the nucleotide diversity (π) within each group: NCN group ($\pi = 3.37 \times 10^{-3}$), GZCN group ($\pi = 3.60 \times 10^{-3}$), and SCN group ($\pi = 2.86 \times 10^{-3}$) (Fig. 2A). TreeMix analysis was conducted to reveal possible gene flow events among geographic populations. Interestingly, 1 gene flow event was observed from the Yunnan populations (YNKM) to Guizhou population (GZAS and GZGY) with a high migration weight and another gene flow event from the Guizhou population to Shannxi population (SXHZ) when assuming 3 gene flow events (Fig. S6), further supporting the genetic admixture and phylogenetic relationships of *S. inferens* populations.

Genomic variation explained by geography and environment

The 3 genetic groups displayed clearly different geographic distributions. The NCN group, distributed in the majority of North China, mainly contained collections from Shandong, Jiangsu, Henan, Shannxi, Hubei, Hunan, and Zhejiang Provinces. GZCN group is distributed in Guizhou Province including 2 populations (GZGY and GZAS). SCN group distributed in South China, mainly included populations collected from Yunnan, Guangxi, Guangdong and Hainan Provinces (Fig. 2B). We also further examined the genetic differentiation of 143 PRB individuals from 12 locations ($n \geq 5$) using isolation-by-distance (IBD) and isolation-by-environment (IBE) analysis. We found pairwise genetic distance between locations ($F_{ST}/1–F_{ST}$) was highly correlated with geographic distance, suggesting a strong signal of isolation by distance (Fig. 2C, Mantel's $r = 0.5942$, $P = 0.0006$). Genetic distance was also significantly associated with environmental distance, based on climactic variables downloaded from public environmental databases (Fig. 2D, Mantel's $r = 0.6643$, $P = 0.0004$). These results suggested that geographical and environmental factors are important contributors to population genetic differences between geographical populations.

Estimation of evolutionary demographic history

To infer the more recent demographic history (e.g., 10 kya) of *S. inferens*, SMC++ software was first applied to estimate the changes of the effective population size (*Ne*). The real time was scaled with a generation time of 0.1 year and a mutation rate of 0.84×10^{-8}. We found that the demographic history of 3 subpopulations of *S. inferens* followed a similar trajectory (Fig. S7). The arrival of the last glacial maximum (LGM; 19–26.5 kya) have had significant effects on the *Ne* of *S. inferens*, with a strong population decline. *fastsimcoal* 2.7 analyses best supported a model of recent split of the northern and Guizhou populations of *S. inferens*, around 384 000 ybp, and a much more ancient split of the southern China population around 1 million ybp, with significant bidirectional asymmetric gene flow (Fig. 3; Table S3, S4). Interestingly, high rates of gene flow were detected unidirectionally from the Guizhou population into southern China (mean = 0.358, 97.5% CI of 0.12−0.60, Table S3). Contemporary effective population sizes of all 3 populations (NCN, GZCN, and SCN) were estimated to be several folds larger than their ancestral sources, with the SCN population estimated to be the largest (*Ne* ~270 000), compared to the relatively smaller GZCN (*Ne* ~101 000) and NCN (*Ne* ~190 000) populations.

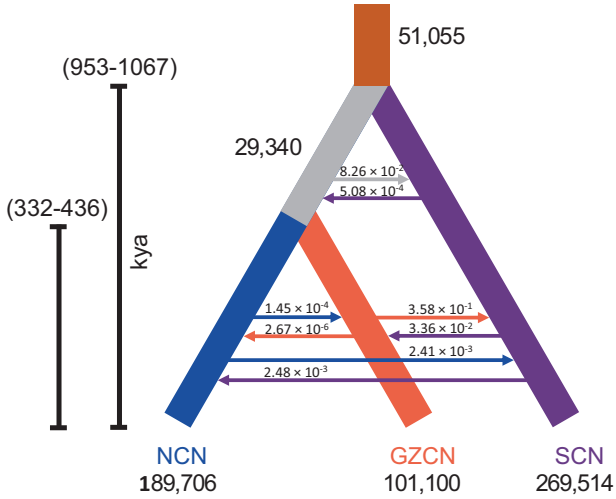

Fig. 3 Population demographic history and genome-wide patterns of differentiation of major groups of *Sesamia inferens*. Schematic of the best demographic scenario modeled using FSC27 analyses. The arrows indicate the direction of gene flow. Point estimates of parameters and 95% confidence intervals are reported in Table S3.

Local selective sweeps of northern populations

To identify genomic regions in PRB showing signatures of selection during the adaptations, the PBS and *u*-test methods were conducted between NCN and SCN geographic populations, which had maximum genetic differentiation ($F_{ST} = 0.27$) (Fig. 4). A total of 127 genes and 1 302 genes showed significant signatures of positive selection in the NCN population based on PBS and *u*-test methods, respectively. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis showed that these genes are mainly associated with Nucleotide excision repair (ko03420; $P < 0.0001$), Axon guidance (ko0360; 0004930; $P = 0.00002$), Antigen processing and presentation (ko04612; $P = 0.00009$), Mannose type O-glycan biosynthesis (ko00515; $P = 0.0015$), Hippo signaling pathway – fly (ko04391; $P = 0.00265$), Apoptosis – multiple species (ko04215; $P = 0.00671$), Fatty acid elongation (ko00062; $P = 0.0121$) and Biosynthesis of unsaturated fatty acids (ko01040; $P = 0.02529$) (Fig. S8). We focus on the glycine-rich cuticular proteins (*GRP*) gene because of its strong selection signal identified in PBS analysis. This finding was also supported by its detection in the *u*-test analysis (Fig. 4A). In the local region,

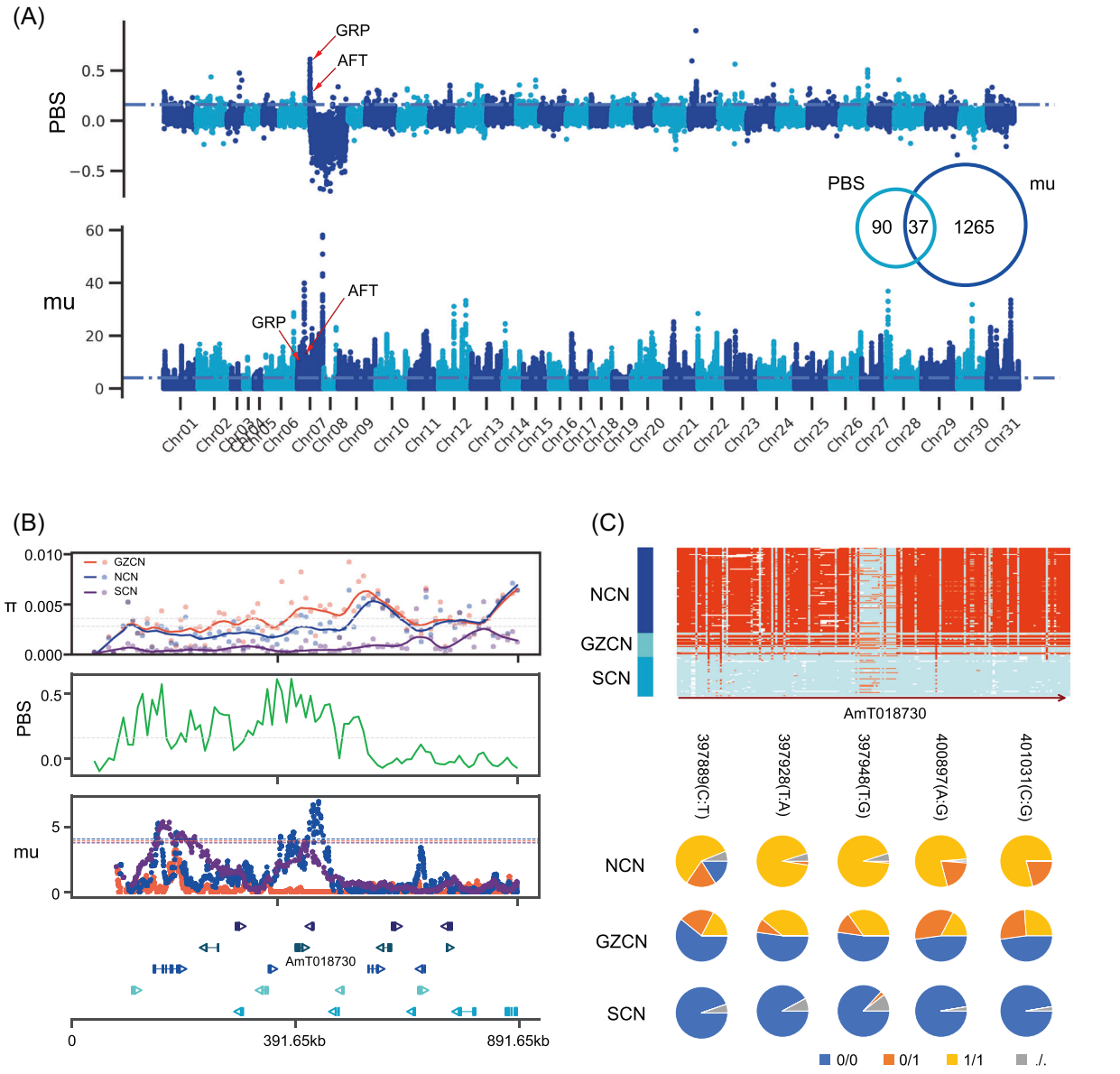

Fig. 4 Genome-wide scan signals of selective sweeps in *Sesamia inferens* populations. (A) Manhattan plot showing the distribution of PBS and mu across 31 linkage groups of *S. inferens* in NCN versus SCN group. Genomic regions that showed strong signal of positive selection were determined as the outliers. The blue horizontal line indicates the cutoff of top 95% quantile in the genome-wide scale. The arrowhead indicates the peak signal containing the candidate genes. (B) Detailed plots of PBS, mu, and nucleotide diversity (π) for *GRP* gene (AmT018730) and gene locations on the chromosomes for each population. Statistics were calculated using 10-kb windows. (C) Haplotypes found at the *GRP* gene in NCN, GZCN, and SCN group. The pie chart shows the frequency of non-synonymous mutations at this locus, and the colors represent the types of alleles. GZCN, Guizhou Province, China; NCN, northern China; SCN, southern China; PBS, population branching statistic. "./." represents missing genotype.

the genetic diversity was lower in the southern populations as identified by the π method (Fig. 4B). To further identify haplotypes of the *GRP* gene, we rearranged the order based on the similarity of variations in this gene. We clearly distinguished 2 haplotypes (Fig. 4C). Additionally, we found that 5 non-synonymous mutations, including R192C (397889 C > T), S205T (397928 T > A), F211L (397948 T > G), D228G (400897 A > G), and

L273V (401031 C > G) in the *GRP* gene, exhibited high significant allele frequency differences between the NCN and SCN populations.

Given that Shandong Province is located north of the northern distribution line of PRB in China, rapid environmental adaptation plays a crucial role in facilitating the moth's life cycle and ensuring its normal development. To elucidate a potential genetic candidate for these rapid adaptations, we used the composite likelihood ratio (CLR) method to scan the genomes for signatures of selection and a total of 547 candidate genes were identified (Fig. S9A; Table S5). KEGG enrichment analysis showed that these genes are mainly associated with the Peroxisome proliferator -activated receptor signaling pathway (ko03320; $P = 9.17E{-}05$), Plant-pathogen interaction (ko04626, 0.00036), Porphyrin and chlorophyll metabolism (ko00860; $P = 0.00096$), Pentose and glucuronate interconversions (ko00040; $P = 0.00588$), ABC transporters (ko02010; $P = 0.00624$) and Fatty acid biosynthesis (ko00061; $P = 0.043144$) (Fig. S9B). The CLR analysis showed that the histone-lysine-N-methyltransferase (*SETMAR*; AmT018558) locus has putatively undergone strong selection in Shandong populations (Fig. S9A), with significant allele frequency differences of *SETMAR* gene between the NCN and SCN populations (Fig. S9C).

Discussion

The application of population-based whole genome resequencing has proven successful in unraveling the genetic structure and local adaptation of various insect species (Chen *et al.*, 2021; Peng *et al.*, 2023). In this study, we employed whole genome resequencing data from geographic populations spanning the native and invasive range of PRB in China. Our aim was to uncover the genetic composition of these populations and identify genes that displayed significant signatures of selection. Our results indicate a clear division of all populations into 3 distinct genetic groups (NCN, GZCN, and SCN), as evidenced by PCA and phylogenetic relationship analysis, which correspond to their geographical distribution. However, previous study using both mitochondrial and microsatellite datasets indicated that 2 large phylogeographic clusters were identified, corresponding to northern and southern China regions, respectively (Tang *et al.*, 2022). We speculate that the GZCN genetic group identified here may be related to small cryptic refugia or microrefugia in southern or southwestern China, particularly in mountainous mosaic areas (Zeng *et al.*, 2015; Tang *et al.*, 2018). The Sichuan Basin and Yunnan-Guizhou Plateau are located in the eastern part of the Qinghai-Tibet Plateau, the highest plateau in the world. Despite the Sichuan Basin having a higher latitude compared to southern China, its lower altitude has resulted in warmer temperatures and a stable climate throughout the Pleistocene period (Ju *et al.*, 2007). This favorable environment could have facilitated the persistence of populations in small refugia within the region. Additionally, the presence of PRB subgroups in the Guizhou Province may have arisen due to geographical separation facilitated by microrefugia (Tang *et al.*, 2022).

Our analysis of evolutionary history supports a model in which the northern and Guizhou populations of *S. inferens* experienced a recent split approximately 384 000 years ago, while the southern China population underwent a much earlier split around 1 000 000 years ago. These splits were accompanied by significant bidirectional asymmetric gene flow. These findings are consistent with previous analyses of PRB using Tajima's D and Fu's Fs statistics, which examined concatenated mitochondrial DNA genes (Tang *et al.*, 2022). In that study, the divergence time between the 2 major clades was estimated to be 385.0 kya (95% highest posterior density: 266.9−514.7 kya), coinciding with the first stage (Marine Isotope Stage 10, MIS10) of the Penultimate Glacial Maximum (PGM) during the Pleistocene epoch (Yi *et al.*, 2005). This temporal correlation with Pleistocene climatic oscillations suggests that significant climate changes during glaciation likely played a crucial role in the separation and divergence of PRB populations in China.

Evidence of local adaptation can be found using selection scans (Peng *et al.*, 2023). Selective sweep analysis between populations in South China (SCN) and North China (NCN) showed that the strongest genetic differentiation occurred in the *GRP* gene, suggesting that GRP plays a major role in the environmental adaptation of PRB. The cuticle is responsible for slow movement of water from the insect body to the environment (Pelletier, 1995; William *et al.*, 2005). Insect adaptation to desiccation conditions through minimizing dehydration may be achieved by changing the chemical composition of the cuticle. The change in cuticle permeability has been linked to adaptation to arid environments, not only among insect species but also to the adaptive difference of individuals from the same species living in different niches (Cloudsley-Thompson, 1975). A previous study had confirmed that insect cuticular *GRP* genes were highly induced by desiccation stress (Zhang *et al.*, 2008). Given that, we speculate that *GRP* genes under natural selection in northern populations are associated with local

adaptation in response to desiccation environmental stress. It is essential to emphasize that the northward movement of PRB is also closely linked to increasing global temperatures. In our upcoming research, we intend to extensively investigate the definitive contribution of the *GRP* gene through genome-editing techniques. This could establish a crucial foundation for curtailing the onward expansion of PRB into northern or northeastern regions, thus playing a pivotal role in the management of this pest. Meanwhile, suppression of PRB individuals in northern and southern China is essential as well, since PRB has resided in these regions for an extensive period of time and constitutes a continuing threat to rice production.

Also, we found *SETMAR* locus has putatively undergone strong selection in Shandong populations (Fig. S9A), with significant allele frequency differences between the NCN and SCN populations. This locus, which has been implicated in double-stranded DNA break repair and control of replication, is involved in reproductive regulation, epigenetic maternal imprinting, and polymorphism regulation in other insects, such as the rice brown planthopper *Nilaparvata lugens* (Liu *et al.*, 2021) and the leafcutter ant *Acromyrmex echiniator* (Howe *et al.*, 2020). Interestingly, this locus is found to be associated with range expansion and local adaptation across a variety of novel environments in the damselfly, *Ischnura elegans* (Dudaniec *et al.*, 2018). Therefore, we hypothesize that the *SETMAR* gene may play a crucial role in the local adaptation of PRB in Shandong Province, given that the moth has recently been observed in this region.

In conclusion, genome resequencing analysis highlights 3 highly differentiated genetic groups of PRB in China, indicating asymmetric migration with varying population sizes across these 3 groups. Furthermore, we discovered numerous genes linked to local environmental adaptation and stress resistance through selective sweep analysis. Our findings present a valuable asset for broader genomic investigations on the pink rice borer. This will streamline forthcoming functional studies and contribute to devising efficient strategies for managing *S. inferens* pests in the future.

Acknowledgments

This work was supported by STI 2030–Major Projects (2022ZD04021), Shenzhen Science and Technology Program (Grant No. KQTD20180411143628272) and the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences. The funders had no role in study design, data collection and analysis, decision to publish or manuscript preparation.

Data Availability Statement

The raw data used for genome resequencing analysis are available at the Dryad Digital Repository: https://doi.org/10.5061/dryad.3n5tb2rpq.

Disclosure

The authors declare they have no competing interests.

References

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.

Alachiotis, N. and Pavlidis, P. (2018) RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communication Biology*, 1, 79.

Baladhiya, H., Sisodiya, D. and Pathan, N. (2018) A review on pink stem borer, *Sesamia inferens* Walker: a threat to cereals. *Journal of Entomology and Zoology Studies*, 6, 1235–1239.

Bay, R.A., Harrigan, R.J., Buermann, W., Underwood, V.L., Gibbs, H.L., Smith, T.B. *et al.* (2018) Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, 359, 83–86.

Bessette, M., Ste-Croix, D.T., Brodeur, J., Mimee, B. and Gagnon, A.È. (2022) Population genetic structure of the carrot weevil (*Listronotus oregonensis*) in North America. *Evolutionary Applications*, 15, 300–315.

Capblancq, T., Fitzpatrick, M.C., Bay, R.A., Exposito-Alonso, M. and Keller, S.R. (2020) Genomic prediction of (mal) adaptation across current and future climatic landscapes. *Annual Review of Ecology Evolution and Systematics*, 51, 245–269.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L. *et al.* (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; iso-2; iso-3. *Flyer*, 6, 80–92.

Chai, H.N. and Du, Y.Z. (2012) The complete mitochondrial genome of the pink stem borer, *Sesamia inferens*, in comparison with four other noctuid moths. *International Journal of Molecular Sciences*, 13, 10236–10256.

Chen, X.J. and Lu, D.H. (2015) Study advances in occurrence and control of the polyphagous pink stem borer *Sesamia inferens*. *China Agriculture Science Bulletin*, 31, 171–175.

Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S. *et al.* (2018) SOAPnuke: a MapReduce acceleration-supported

software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience*, 7, 120.

Chen, Y., Liu, Z., Régnière, J., Vasseur, L., Lin, J., Huang, S. *et al.* (2021) Large-scale genome-wide study reveals climate adaptive variability in a cosmopolitan pest. *Nature Communications*, 12, 7206.

Cheng, Y.L., Miller, M.J., Zhang, D., Xiong, Y., Hao, Y., Jia, C. *et al.* (2021) Parallel genomic responses to historical climate change and high elevation in East Asian songbirds. *Proceedings of the National Academy of Sciences USA*, 118, e2023918118.

Cloudsley-Thompson, J.L. (1975) Adaptations of arthropoda to arid environments. *Annual Review of Entomology*, 20, 261–283.

Danecek, P. and McCarthy, S.A. (2017) BCFtools/csq: haplotype-aware variant consequences. *Bioinformatic*, 33, 2037–2039.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.

Das, I.K. and Padmaja, P.G. (2016) *Biotic stress resistance in millets*. Academic Press, USA.

Ding, D., Liu, G.J., Hou, L., Gui, W.Y., Chen, B. and Kang, L. (2018) Genetic variation in *PTPN1* contributes to metabolic adaptation to high-altitude hypoxia in Tibetan migratory locusts. *Nature Communications*, 9, 4991.

Dudaniec, R.Y., Yong, C.J., Lancaster, L.T., Svensson, E.I. and Hansson, B. (2018) Signatures of local adaptation along environmental gradients in a range-expanding damselfy (*Ischnura elegans*). *Molecular Ecology*, 27, 2576–2579.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V.C. and Foll, M. (2013) Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9, e1003905.

Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A. and Sousa, V.C. (2021) Fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics*, 37, 4882–4885.

Fick, S.E. and Hijmans, R.J. (2017) WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302–4315.

Gao, F., Zou, W., Xie, L. and Zhan, J. (2017) Adaptive evolution and demographic history contribute to the divergent population genetic structure of potato virus Y between China and Japan. *Evolutionary Applications*, 10, 379–390.

Goudet, J. (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology*, 5, 184–186.

Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D.L., Houle, D. *et al.* (2007) Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature*, 445, 82–85.

Han, L.Z., Peng, Y.F. and Wu, K.M. (2012) Studies on larval dispersal ability in the field and flight capacity of the pink stem borer, *Sesamia inferens*. *Plant Protection*, 38, 9–13. (In Chinese)

Howe, J., Schiøtt, M., Li, Q., Wang, Z., Zhang, G. and Boomsma, J.J. (2020) A novel method for using RNA-seq data to identify imprinted genes in social Hymenoptera with multiply mated queens. *Journal of Evolutionary Biology*, 33, 1770–1782.

Huang, J., Li, G., Lei, H., Fan, C., Tian, C., Chen, Q. *et al.* (2020) Low-temperature derived temporal change in the vertical distribution of *Sesamia inferens* larvae in winter, with links to its latitudinal distribution. *PLoS ONE*, 15, e0236174.

Jin, J.Y., Li, Z.Q., Zhang, Y.N., Liu, N.Y. and Dong, S.L. (2014) Different roles suggested by sex-biased expression and pheromone binding affinity among three pheromone binding proteins in the pink rice borer, *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae). *Journal of Insect Physiology*, 66, 71–79.

Ju, L., Wang, H. and Jiang, D. (2007) Simulation of the last glacial maximum climate over East Asia with a regional climate model nested in a general circulation model. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 248, 376–390.

Karim, S. and Riazuddin, S. (1999) Rice insect pests of Pakistan and their control: a lesson from past for sustainable future integrated pest management. *Pakistan Journal of Biological Sciences*, 2, 261–276.

Keightley, P.D., Pinharanda, A., Ness, R.W., Simpson, F., Dasmahapatra, K.K., Mallet, J. *et al.* (2015) Estimation of the spontaneous mutation rate in *Heliconius melpomene*. *Molecular Biology Evolution*, 32, 239–243.

Letunic, I. and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49, W293–W296.

Li, B., Xu, Y., Han, C., Han, L., Hou, M. and Peng, Y. (2015) *Chilo suppressalis* and *Sesamia inferens* display different susceptibility responses to Cry1A insecticidal proteins. *Pest Management Science*, 71, 1433–1440.

Li, C.X., Cheng, X. and Dai, S.M. (2011) Distribution and insecticide resistance of pink stem borer, *Sesamia inferens* (Lepidoptera: Noctuidae), in Taiwan. *Formosan Entomologist*, 31, 39–50.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics,n. pag. https://arxiv.org/abs/1303.399

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Liu, X., Chen, G., He, J., Wan, G., Shen, D., Xia, A. *et al.* (2021) Transcriptomic analysis reveals the inhibition of reproduction in rice brown planthopper, *Nilaparvata lugens*, after silencing the gene of MagR (IscA1). *Insect Molecular Biology*, 30, 253–263.

Mahesh, P., Srikanth, J. and Chandran, K. (2014) Pattern of pink stem borer *Sesamia inferens* (Walker) incidence in different crop seasons and *Saccharum* spp. *Journal Sugarcane Research*, 4, 91–95.

Mansoor, M.M., Raza, A.B.M. and Afzal, M.B.S. (2019) Fipronil resistance in pink stem borer, *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae) from Pakistan: cross-resistance, genetics and realized heritability. *Crop Protection*, 120, 103–108.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A. *et al.* (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, 1297–1303.

Mérot, C. (2022) Evolution: how important is the dimensionality of natural selection in local adaptation? *Current Biology*, 32, 274–276.

Pavlidis, P., Živković, D., Stamatakis, A. and Alachiotis, N. (2013) SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, 30, 2224–2234.

Pelletier, Y. (1995) Effects of temperature and relative humidity on water loss by the Colorado potato beetle, *Leptinotarsa decemlineata* (Say). *Journal of Insect Physiology*, 41, 235–239.

Peng, Y., Jin, M., Li, Z., Li, H., Zhang, L., Yu, S. *et al.* (2023) Population genomics provide insights into the evolution and adaptation of the Asia corn borer. *Molecular Biology and Evolution*, 40, 112.

Pickrell, J.K. and Pritchard, J.K. (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8, e1002967.

R, Core Team. (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. https://www.r-project.org

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V. and Korbel, J.O. (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*, 28, i333–i339.

Reddy, M.L., Babu, T.R. and Venkatesh, S. (2003) A new rating scale for *Sesamia inferens* (Walker) (Lepidoptera: Noctuidae) damage to maize. *International Journal of Tropical Insect Science*, 23, 293–299.

Savolainen, O., Lascoux, M. and Merilä, J. (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14, 807–820.

Sork, V.L. (2016) Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. *Evolutionary Applications*, 9, 291–310.

Sun, M., Tang, X.T., Lu, M.X., Yan, W.F. and Du, Y.Z. (2014) Cold tolerance characteristics and overwintering strategy of *Sesamia inferens* (Lepidoptera: Noctuidae). *Florida Entomologist*, 97, 1544–1553.

Szűcs, M., Vahsen, M.L., Melbourne, B.A., Hoover, C., Weiss-Lehman, C. and Hufbauer, R.A. (2017) Rapid adaptive evolution in novel environments acts as an architect of population range expansion. *Proceedings of the National Academy of Sciences USA*, 114, 13501–13506.

Tang, C.Q., Matsui, T., Ohashi, H., Dong, Y.F., Momohara, A., Herrando-Moraira, S. *et al.* (2018) Identifying long-term stable refugia for relict plant species in East Asia. *Nature Communications*, 9, 4488.

Tang, X.T., Lu, M.X. and Du, Y.Z. (2022) Molecular phylogeography and evolutionary history of the pink rice borer (Lepidoptera: Noctuidae): implications for refugia identification and pest management. *Systematic Entomology*, 47, 371–383.

Tang, X.T., Xu, J., Sun, M., Xie, F.F. and Du, Y.Z. (2014) First microsatellites from *Sesamia inferens* (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America*, 107, 866–871.

Terhorst, J., Kamm, J.A. and Song, Y.S. (2017) Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49, 303–309.

Wang, X., Bernhardsson, C. and Ingvarsson, P.K. (2020) Demography and natural selection have shaped genetic variation in the widely distributed conifer Norway spruce (*Picea abies*). *Genome Biology and Evolution*, 12, 3803–3817.

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution; International Journal of Organic Evolution*, 38, 1358–1370.

William, J.B., Ruehl, N.C. and Lee, R.E. Jr (2005) Partial link between the seasonal acquisition of cold-tolerance and desiccation resistance in the goldenrod gall fly *Eurosta solidaginis* (Diptera: Tephritidae). *Journal of Experimental Biology*, 207, 4407–4414.

Wu, S.F., Zhao, D.D., Huang, J.M., Zhao, S.Q., Zhou, L.Q. and Gao, C.F. (2018) Molecular characterization and expression profiling of ryanodine receptor gene in the pink stem borer, *Sesamia inferens* (Walker). *Pesticide Biochemistry and Physiology*, 146, 1–6.

Xu, L., Li, C., Hu, B., Zhou, Z. and Li, X. (2011) Review of history, present situation and prospect of pink stem borer in China. *Chinese Agricultural Science Bulletin*, 27, 244–248.

Yang, G.Q., Du, S.G., Li, L., Jiang, L.B. and Wu, J.C. (2014) Potential positive effects of pesticides application on *Sesamia inferens* (Walker) (Lepidoptera: Insecta). *International Journal of Insect Science*, 6, S16485.

Yang, T., Liu, R., Luo, Y., Hu, S., Wang, D., Wang, C. *et al.* (2022) Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nature Genetics*, 54, 1553–1563.

Yao, Y.H., Du, Y.Z., Zheng, F.S. and Wang, L.P. (2008) The variation of mtDNA COII sequences in 9 geo-populations of rice stem borer, *Sesamia inferens*. *Journal of Environmental Entomology*, 30, 39–43.

Yi, C., Cui, Z. and Xiong, H. (2005) Numerical periods of quaternary glaciations in China. *Quaternary Sciences*, 25, 609–619.

Zeng, Y.F., Wang, W.T., Liao, W.J., Wang, H.F. and Zhang, D.Y. (2015) Multiple glacial refugia for cool-temperate deciduous trees in northern East Asia: the Mongolian oak as a case study. *Molecular Ecology*, 24, 5676–5691.

Zhang, J., Goyer, C. and Pelletier, Y. (2008) Environmental stresses induce the expression of putative glycine-rich insect cuticular protein genes in adult *Leptinotarsa decemlineata* (Say). *Insect Molecular Biology*, 17, 209–216.

Zhang, S.M. (1965) Discussion the boundary line between the ancient North and the Eastern regions in the east of the Qinling Mountains in China from the distribution of some agricultural insects. *Acta Entomologica Sinica*, 14, 411–419.

Zhang, Z.C., Luo, G.H., Zhang, G.F., Han, G.J., Liu, B.S. and Fang, J.C. (2013) Genetic diversity of different geographical populations of *Sesamia inferens* as determined by AFLP. *Chinese Journal of Applied Entomology*, 50, 693–699.

Manuscript received October 10, 2023
Final version received November 20, 2023
Accepted December 18, 2023

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1 All evolutionary models and competing topologies of Sesamia inferens tested in FSC27 under the three subpopulations model (NCN, GZCN, SCN) as estimated by ADMIXTURE.

Fig. S2 The density of different sizes for each SV type.

Fig. S3 Population structure analysis (K = 2–10) based on SNPs.

Fig. S4 Population structure analysis (K = 2–10) based on SVs.

Fig. S5 The estimated cross-validation error of possible clusters (K) from 1 to 10.

Fig. S6 Gene flow analysis across different geographical populations in China.

Fig. S7 Effective population size of S. inferens populations inferred using SMC++ on the basis of whole genomes.

Fig. S8 The scatterplot shows the KEGG pathway enrichment of identified genes in significant pathways (FDR < 0.05) of natural selected genes in PBS analysis.

Fig. S9 Whole-genome scan on the basis of CLR analyses.

Table S1 Description of the environmental variables that were considered in the IBE analysis in this study.

Table S2 Length distribution of SV categories in *S. inferens*.

Table S3 Bootstrap confidence intervals of evolutionary demographic history parameters of *S. inferens* computed using the best fitting model (asymmetric gene flow) and topology of ((NCN, GZCN), SCN) using fastsimcoal27.

Table S4 Maximum likelihood estimates across 8 demographic models tested in fastsimcoal27, with the best fitting model indicated in red.

Table S5 A total of 547 candidate genes were identified using CLR analysis. See attached Excel file.