Integrative Zoology 2024; 0: 1–20 doi: 10.1111/1749-4877.12917

ORIGINAL ARTICLE

INTEGRATIVE ZOOLOGY

Sympatric diversity pattern driven by the secondary contact of two deeply divergent lineages of the soybean pod borer *Leguminivora glycinivorella*

Mingsheng YANG,[1,2] Ying WANG,[1] Peng DAI,[3] Dandan FENG,[1] Alice C. HUGHES,[5] Houhun LI[4] and Aibing ZHANG[1]

[1]College of Life Sciences, Capital Normal University, Beijing, P. R. China, [2]College of Life Science and Agronomy, Zhoukou Normal University, Zhoukou, Henan, P. R. China, [3]Institute of Biological Control, Jilin Agricultural University, Changchun, P. R. China, [4]College of Life Sciences, Nankai University, Tianjin, P. R. China and [5]Landscape Ecology Group, Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan, P. R. China

Abstract

The soybean pod borer, *Leguminivora glycinivorella* (Matsumura), is an important tortricid pest species widely distributed in most parts of China and its adjacent regions. Here, we analyzed the genetic diversity and population differentiation of *L. glycinivorella* using diverse genetic information including the standard *cox1* barcode sequences, mitochondrial genomes (mitogenomes), and single-nucleotide polymorphisms (SNPs) from genotyping-by-sequencing. Based on a comprehensive sampling (including adults or larvae of *L. glycinivorella* newly collected at 22 of the total 30 localities examined) that covers most of the known distribution range of this pest, analyses of 543 *cox1* barcode sequences and 60 mitogenomes revealed that the traditionally recognized and widely distributed *L. glycinivorella* contains two sympatric and widely distributed genetic lineages (A and B) that were estimated to have diverged ~1.14 million years ago during the middle Pleistocene. Moreover, low but statistically significant correlations were recognized between genetic differentiation and geographic or environmental distances, indicating the existence of local adaptation to some extent. Based on SNPs, phylogenetic inference, principal component analysis, fixation index, and admixture analysis all confirm the two divergent sympatric lineages. Compared with the stable demographic history of Lineage B, the expansion of Lineage A had possibly made the secondary contact of the two lineages probable, and this process may be driven by the climate fluctuation during the late Pleistocene as revealed by ecological niche modeling.

Key words: DNA barcode, genetic diversity, genotyping-by-sequencing, mitogenome, Tortricidae

INTRODUCTION

Phylogeography reveals genetic diversity and differentiation of species often at the population level and clarifies the mechanisms underlying the current population genetic pattern of a given species in combination with historical

Correspondence: Aibing Zhang, College of Life Sciences, Capital Normal University, Beijing 100048, P. R. China.
Email: zhangab2008@cnu.edu.cn
Mingsheng Yang and Ying Wang contributed equally.

geological events and climatic fluctuations (Avise *et al.* 1987; Beheregaray 2008; Liu *et al.* 2023a). For example, Pleistocene glacial cycles are regarded as having profound effects on the distribution dynamics of species, often leading to complicated genetic variation through allopatric divergence followed by secondary contact (Hewitt 2004; Dong *et al.* 2020). For pest species, population genetics could help understand population differentiation, genetic structure, dispersal dynamic, and local adaption, and thereby provide a reference for decision-making against pests (Beheregaray 2008; Milankov *et al.* 2013). For instance, the key questions revealed by population genetics, such as the number of genetically distinct populations, whether populations are expanding or contracting, and the rate at which individuals are moving between populations, could effectively promote pest management (Rollins *et al.* 2006). The insect family Tortricidae includes numerous important pest species, such as the worldwide invasive pests *Cydia pomonella* (Linnaeus, 1758) and *Grapholita molesta* (Busck, 1916), causing major losses in crop and forest production (van der Geest & Evinhuis 1991; Dombroskie & Sperling 2013). In the past two decades, several investigations of population genetics on tortricid pests especially the *C. pomonella* and *G. molesta* have been conducted utilizing molecular markers such as mitochondrial gene fragments, microsatellite loci, and single-nucleotide polymorphisms (SNPs) generated from genotyping-by-sequencing (GBS) or genome resequencing (Fuentes-Contreras *et al.* 2008; Kirk *et al.* 2013; Li *et al.* 2015; Silva-Brandão *et al.* 2015; Wei *et al.* 2015; Song *et al.* 2018; Basoalto *et al.* 2020; Cao *et al.* 2021, 2022), which greatly contributed to our understanding of population differentiation, dispersal routes, and local adaption of these pests.

The soybean pod borer *Leguminivora glycinivorella* (Matsumura, 1898), belonging to Grapholitini of Tortricidae, is widely distributed in most of China and its adjacent regions (Li 2012). *L. glycinivorella* is a notorious soybean pest (Yang *et al.* 2020), as its larva can enter the soybean pod and feed on young beans, causing considerable loss of soybean yield and a decline in seed quality. In northeastern China, the damage varies between 10% and 20% seed damage on average, and it can exceed 40% for some cultivars (Zhang & Fu 1983; Shi *et al.* 2018). In the United States, *L. glycinivorella* is also listed as a quarantine pest (Areces-Berazain 2022). Given that *L. glycinivorella* has a wide distribution and is of economic importance, several investigations on morphological or genetic differentiation of population samples have been conducted in recent years. Song (2014) reported that the variations in the head length and width of the overwintering larvae and some biological characteristics (such as cooling and freezing points) have happened among populations in China. Based on molecular data, Wang *et al.* (2014) (*cox1* fragment, 657 bp) and Wang *et al.* (2015) (*cytb* and *cox2* fragments, 415 and 683 bp, respectively) investigated the genetic diversity of 10 *L. glycinivorella* populations in Northeast China, both demonstrating low genetic diversity and frequent gene exchange among them. Then, using samples from 19 populations across China (12 of which are located in Northeast China), Zhu *et al.* (2017) (*cox2* fragment, 723 bp) and Shi *et al.* (2018) (*cox1* fragment, 408 bp) recognized high genetic diversity, and populations of Guiyang and Du'An in Southwest China showed significant divergence with other populations. We reanalyzed the *cox1* data used in Shi *et al.* (2018) and interestingly found that the average genetic distance (Kimura 2-parameter model [K2P]) among Du'An and other populations ranged from 0.051 to 0.057, which was much higher than the 0.002–0.011 among the remaining 18 populations including the Guiyang population that is closer to Du'An at a geographic scale. Considering that only a 408-bp *cox1* fragment and sampling of relatively low geographic coverage were used in Shi *et al.* (2018), further investigation of the genetic differentiation of this important pest is necessary based on samples with a larger spatial coverage as well as more genetic information.

In this study, we collected *L. glycinivorella* individuals across most of its distribution range, and especially, the Huang-Huai Valley in central China, rarely sampled in prior studies, was intensely covered. Moreover, extensive genetic information, including the standard *cox1* barcodes, mitogenomes, and SNPs from GBS, was used to analyze the genetic diversity and population differentiation pattern of *L. glycinivorella* and to evaluate its driving factors. In particular, this study also focused on reevaluating the genetic status and distribution pattern of samples from the Du'An population relative to other populations.

MATERIALS AND METHODS

Sampling and DNA extraction

Adults or larvae of *L. glycinivorella* were newly collected at 22 localities in China from 2018 to 2020, and the sampling localities covered most of the known distribution range of this pest (Fig. 1 and Table 1). Specimens were preserved in 100% ethanol under –80°C environment until used for DNA extraction. Specimen identification was conducted based on morphology (Li 2012) or/and the mitochondrial *cox1* barcoding method (Hebert *et al.* 2003). Genomic DNA was extracted from

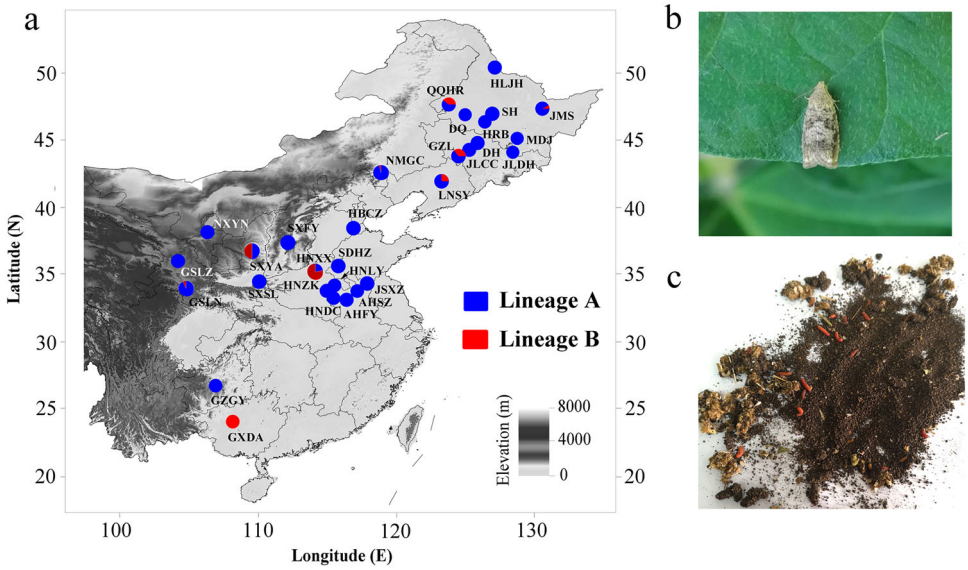

Figure 1 Sampling localities and morphology of *Leguminivora glycinivorella*. (a) Geographical distribution of the 30 sampling localities analyzed in this study; the two colors represent two lineages referring to Fig. 2; the proportion of different lineages in one locality is marked in different lineage colors based on the number of individual belongings to each lineage; the locality codes refer to Table 1 in this study and table 1 in Wang *et al.* (2014). (b) Adult morphology of *L. glycinivorella*. (c). The red larva of *L. glycinivorella*.

the legs or head and thorax (depending on the data types to obtain) tissue from a single individual using a DNeasy tissue kit (Qiagen, Germany), following the manufacturer's instructions. Voucher specimens were deposited in the Biology Laboratory of Zhoukou Normal University, China, under the research project on population genetics of soybean pests.

The *cox1*, mitogenome, and SNP data processing

The standard *cox1* barcode sequences for *L. glycinivorella* individuals were amplified using the primer pair LCO1490 (F) and HCO2198 (R) (Folmer *et al.* 1994) through polymerase chain reaction (PCR). Then, the PCR products were sequenced (two directions) with the Sanger method using the same primer pair.

To obtain mitogenomes, the whole genome shotgun method (Covaris, Woburn, MA, USA) was used to fragment the genome to an average size of 400 bases. The libraries were constructed (one library for each *L. glycinivorella*) using TruSeq DNA PCR-Free Sample Preparation Kit (Illumina, USA) and further sequenced to acquire approximately 3 Gb data on an Illumina HiSeq 2500 platform with a strategy of 150 paired-ends. Raw sequences were filtered using Adapter Removal version 2 (Schubert *et al.* 2016) and SOAPdenovo version 2.01 (Luo *et al.* 2012) to generate clean data. The Geneious R11 (Kearse *et al.* 2012) was used to assemble mitogenome sequences, and the "map to reference" strategy was selected to map all cleaned reads to a *cox1* "anchor" amplified before using the primer pair LCO1490 (F) and HCO2198 (R). Then, the contig sequences were circularized with MEGA X (Kumar *et al.* 2018) and further annotated using the MITOS web server with invertebrate genetic code (Bernt *et al.* 2013).

For SNP sequences, 46 *L. glycinivorella* individuals representing the two lineages of *L. glycinivorella* inferred by the *cox1* and mitogenome sequences were selected. The GBS library was prepared using Mse I and EcorI_HaeII restriction enzymes. Sequencing was performed on an Illumina HiSeq 4000 with single reads 150 bp in length. The BWA-MEM v 0.7.17 (Li 2013) was used to map sequencing data to the reference *L. glycinivorella* genome (GCA_023078 275.1) with default parameters, and the duplicate reads generated from PCR were removed with the picard tool (http://broadinstitute.github.io/picard/). The flagstat and coverage functions implemented in SAMtools v 1.13 (Li *et al.* 2009) were used to statistically check the alignment. SNP detection for each sample was conducted with GATK v 4.2.0.0 (McKenna *et al.* 2010), and the SNP data were preliminarily filtered with the parameters "QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 3.0, MQRankSum < -12.5, and ReadPosRankSum < -8.0." To obtain

Table 1 Information on Leguminivora *glycinivorella* samples sequenced in this study

Sample locality	Locality code	Sample size (*cox1*/mitogenome/GBS)	Sampling time	Latitude	Longitude
Du'An, Guangxi, China	GXDA	29/3/7	September 2018	23.9352	108.1011
Heihe, Heilongjiang, China	HLJH	20/3/-	August 2018	50.2533	127.4605
Shenyang, Liaoning, China	LNSY	5/3/-	August 2018	41.8197	123.5483
Changchun, Jilin, China	JLCC	24/3/-	August 2018	43.8061	125.4123
Guiyang, Guizhou, China	GZGY	27/3/-	September 2018	26.5030	106.6565
Suzhou, Anhui, China	AHSZ	8/3/-	September 2018	33.6279	116.9791
Lanzhou, Gansu, China	GSLZ	18/3/-	September 2018	35.5623	104.5952
Longnan, Gansu, China	GSLN	23/3/4	October 2018	33.3931	104.9265
Chifeng, Neimenggu, China	NMGC	24/3/-	October 2018	42.2950	118.8681
Cangzhou, Hebei, China	HBCZ	16/3/-	August 2018	38.2816	116.8175
Zhoukou, Henan, China	HNZK	15/3/7	August 2019	33.6452	114.6781
Yanan, Shaanxi, China	SXYA	5/2/-	August 2019	36.5712	109.4690
Dunhua, Jilin, China	JLDH	3/3/-	August 2019	43.3599	128.2323
Shangluo, Shaanxi, China	SXSL	5/2/-	August 2019	33.8733	109.9258
Fenyang, Shanxi, China	SXFY	7/3/-	August 2019	37.2458	111.7831
Yongning, Ningxia, China	NXYN	11/3/-	August 2019	38.2155	106.2286
Luyi, Henan, China	HNLY	3/3/2	August 2020	33.8756	115.2170
Dancheng, Henan, China	HNDC	2/2/1	August 2020	33.6367	115.2199
Xinxiang, Henan, China	HNXX	53/2/11	August 2019	35.0061	113.6985
Fuyang, Anhui, China	AHFY	1/1/-	August 2019	32.9305	115.7972
Heze, Shangdong, China	SDHZ	13/3/3	August 2019	35.4132	115.2513
Xuzhou, Jiangsu, China	JSXZ	18/3/11	August 2019	34.2795	117.2928
In total	22	390/60/46			

high-quality SNP data, Plink v 1.90 (Purcell *et al.* 2007) was used to retain the biallelic variation sites and filter the SNPs with a missing rate higher than 20% and a minor allele frequency lower than 3% for all samples. To generate loci independent of the linkage disequilibrium, every 50 consecutive SNPs were divided into a window, sliding 10 SNPs each time, and SNPs with an LD coefficient (R^2) greater than 0.5 were rejected.

Molecular data analyses

cox1 *data*

Sequence alignments were conducted using a codon-based mode within the TranslatorX online platform (Abascal *et al.* 2010). Haplotype diversity and nucleotide diversity were calculated using DNASP version 6 (Rozas *et al.* 2017). Pairwise genetic distances under K2P were calculated using MEGA X (Kumar *et al.* 2018). Pairwise differentiation (fixation index [F_{ST}]) values among populations were calculated using Arlequin version 3.5 (Excoffier & Lischer 2010) with 1000 permutations.

Maximum likelihood (ML) and Bayesian inference (BI) analyses were performed based on haplotype sequences using GTR + G model selected by jModelTest version 2.1 (Posada 2008) with Akaike information criterion (AIC) (Akaike 1974). ML tree was constructed using IQ-TREE 2.0.4 (Nguyen *et al.* 2015), and branch supports were calculated using 1000 ultrafast bootstrap replicates (Hoang *et al.* 2018). The BI tree was constructed with MrBayes version 3.2.6 (Ronquist *et al.* 2012), and convergence was considered to be reached when the estimated sample size value was above 200 established by Tracer version 1.7 (Rambaut *et al.* 2018). The median-joining network showing the relationships of

cox1 or mitogenome haplotypes was constructed with the PopART (http://popart.otago.ac.nz) with default settings.

Neutrality tests of Tajima's D (Tajima 1989) and Fu's Fs (Fu 1997) were calculated with DNASP version 6 (Rozas *et al.* 2017), and the sum of square deviations (SSD) and Harpending's raggedness index (*r*) were calculated with Arlequin version 3.5 (Excoffier & Lischer 2010). Analyses of mismatch distribution were performed in DNASP version 6 (Rozas *et al.* 2017) with the constant population size as the model for expected values. Bayesian skyline plots were obtained for *cox1* datasets using the Coalescent Bayesian Skyline model for the prior tree in BEAST version 2.0 (Bouckaert *et al.* 2014). In this analysis, a strict clock model was selected using the proposed conventional mutation rate of the insect *cox1* gene of 2.3% per million years (Brower 1994), and the Tracer software Tracer version 1.7 (Rambaut *et al.* 2018) was used to generate the plot. Based on *cox1* haplotypes, the BEAST version 2.0 (Bouckaert *et al.* 2014) was employed to test the divergence time between the two lineages revealed by phylogenetic trees and networks. Given their close sequence similarity among up to 103 haplotypes defined in Lineage A, six haplotypes were selected from this lineage in the divergence time estimation. Also, a strict clock model was applied to the Yule model (Yule 1925) for the prior tree. Two independent MCMC runs of 1000 million generations were performed, with trees sampled every 1000 generations. The TreeAnnotator program in BEAST version 2.0 (Bouckaert *et al.* 2014) was used to calculate the consensus tree and annotate the divergence times with "height_95%_HPD" after discarding the initial 25 % of trees as burn-in.

To analyze the correlations between F_{ST} and geographic distance (in km) and between F_{ST} and environmental distance, Mantel tests were performed to test the isolation-by-distance (IBD) and isolation-by-environment (IBE) models using the vegan package (https://github.com/vegandevs/vegan) of R software version 4.1.3 (R Core Team 2021) with 999 replicates. Pairwise geographic distances among populations were calculated with the sp package (https://github.com/edzer/sp/) with the population coordinates as input data. In IBE analysis, 19 bioclimatic and elevation environmental layers for historical (near current) conditions downloaded from the WorldClim website were used to extract the environmental data based on the population coordinates (Ran *et al.* 2024).

Mitogenome data

Sequence alignments for each protein-coding gene were conducted with the same method with that of *cox1* data. The tRNA and rRNA genes were aligned with the Q-INS-i algorithm implemented in the MAFFT online platform (Katoh *et al.* 2017). The aligned tRNA and rRNA sequences were filtered using ClipKIT (Steenwyk *et al.* 2020) to delete ambiguously aligned sites with the kpic-gappy algorithm. Pairwise genetic distances under K2P were calculated using MEGA X (Kumar *et al.* 2018). Pairwise differentiation (F_{ST}) values among populations were calculated using Arlequin version 3.5 (Excoffier & Lischer 2010) with 1000 permutations.

Five data partitions were defined for the mitogenome dataset used in phylogenetic analysis, that is, three codon partitions for PCGs, two for rRNAs, and one for tRNAs. The jModelTest was used to select the nucleotide substitution model for each partition that is shown in Table S1, Supporting Information. The mitogenome-based ML and BI trees were constructed with the same softs and parameters as that of *cox1* data.

SNP data

The F_{ST} values were calculated with the vcftools (Danecek *et al.* 2011). The high-quality SNP data were sorted and formatted with Tassel version 5.2.73 (Bradbury *et al.* 2007), and a maximum likelihood phylogeny was constructed using the raxml-ng version 1.0.3 (Kozlov *et al.* 2019) package with the GTR + G model, and the bootstrap analysis was run 1000 times. Based on SNP data, the principal component analysis (PCA) was performed using Plink version 1.90 (Purcell *et al.* 2007). Population structure was analyzed with the Admixture version v 1.3.0 (Alexander *et al.* 2009) under the default parameters. The number of ancestral populations was set from 2 to 7, and the number of optimal ancestral populations was detected based on a cross-validation error (*K*) (Alexander *et al.* 2009).

Ecological niche modeling

The optimized maximum entropy (MaxEnt) method (Phillips *et al.* 2006) was used to predict the past and present potential distribution of *L. glycinivorella*. The 29 occurrence records for Lineage A of *L. glycinivorella* obtained from our field investigation were used as distribution data. Initially, 19 bioclimate variables were downloaded from WorldClim version 1.4 (Hijmans *et al.* 2005). After a Pearson's correlation analysis using the R package "corrplot" (Wei & Simko 2021), one of the two variables with a correlation coefficient $|r| > 0.8$ was randomly removed. Finally, seven variables (BIO02–BIO04, BIO08, BIO14–BIO15, BIO18), representing the present climate conditions during 1960–1990 or the past condition of

the Last Inter-glacial period (LIG; ∼120 000–140 000 years before) (Otto-Bliesner *et al.* 2006), were used in the modeling. To avoid overfitting and improve transferability, the best combination of feature classes (FCs) and regularization multipliers (RMs) was optimized using the R packages "ENMeval version 2" (Muscarella *et al.* 2014; Kass *et al.* 2021) and "dismo version 1.3.5" (https://CRAN.Rproject.org/package=dismo). In this process, six FCs (L, LQ, H, LQH, LQHP, and LQPHT) and eight RMs (from 0.5 to 4.0 in an interval of 0.5) were set to calculate the standardized AIC coefficient (AICc). The area under the curve (AUC) within the receiver operating characteristic curve was used to evaluate the model performance.

RESULTS

Data generation

A total of 543 standard mitochondrial *cox1* barcode sequences were gathered, 330 of which were newly sequenced in this study; 60 were extracted from the 60 mitogenomes sequenced in this study, and the remaining 153 sequences were downloaded from GenBank submitted by Wang *et al.* (2014) with samples collected from eight populations in Northwest China. Sixty mitogenomes were assembled, annotated, and circularized in the present study except for two individuals that we failed to assemble *trnM*, *trnI*, and part of A + T-rich region for AHSZ2M, and part of L-rRNA and A + T-rich region for JSXZ1M. The lengths of fully sequenced mitogenomes ranged from 15 506 to 15 619 bp, and the length differences of A + T-rich regions mainly account for the variation of mitogenome in length (Table S2, Supporting Information). In each fully assembled mitogenome, 37 typical mitochondrial genes were identified, and gene order was identical and typical of Lepidoptera (Fig. S1, Supporting Information). For SNP data, 44 199 SNPs were detected. After Hardy–Weinberg equilibrium tests, 24259 SNPs were maintained in our matrix for 46 individuals and were used in subsequent analyses. The sampling size of molecular data for each population is shown in Table 1.

Genetic diversity based on mitochondrial sequences

The *cox1* barcodes and mitogenome sequences yielded alignment lengths of 657 and 14 948 bp, respectively. Among 543 *cox1* sequences, 109 haplotypes were defined from samples of 30 localities, and one haplotype (Hap1) was shared by 215 individuals from 27 populations (Table S3, Supporting Information). The lowest haplotype diversity was 0.123 for the GXDA population and up to 1 for populations SXFY, HNLY, and HNDC. The nucleotide diversity ranged from 0.00047 (DQ) to 0.02212 (SXYA). Overall, of the two lineages defined by phylogenetic analyses, Lineage A showed comparative haplotype diversity with Linage B, whereas the nucleotide diversity was a little lower than Lineage B (Table 2). However, for each population of the two lineages, the haplotype diversity for Lineage A ranged from 0 to 1, with the populations in central China generally showing higher values, whereas that for Lineage B ranged from 0 to 0.667, generally without a geographic pattern (Fig. S2, Supporting Information). The average pairwise genetic distance between the two lineages was 0.038.

For mitogenome sequences, 58 haplotypes were recognized from 60 sequences. Haplotype diversity was approaching 1 for both lineages, and nucleotide diversity of Lineage A was lower than that of Lineage B, a pattern of genetic diversity similar to that based on *cox1* barcodes (Table 3). To further evaluate the level of genetic divergence between the two lineages, pairwise genetic distances among 60 samples based on each PCG, the combined 13 PCGs, and the combined 37 genes were calculated. As shown in Fig. S3, Supporting Information, except for *atp8*, the average genetic distances between Lineage A and Lineage B were always significantly higher than that within each lineage. The values even reached above 0.04 for *cob*, *nad1*, *nad3*, and *nad6*. In contrast, pairwise distances were generally below 0.01 within Lineage A or Lineage B.

Phylogeny and haplotype network based on mitochondrial sequences

Based on 109 *cox1* haplotypes, ML (Fig. 2a) and BI (Fig. S4, Supporting Information) trees showed two clearly defined lineages with strong supports. Lineage A included 103 of the 109 haplotypes, but the haplotypes from the same or neighbor population generally did not form a cluster. Lineage B contained the remaining six haplotypes, which were distributed in nine populations. Members of the two lineages had a sympatric distribution in eight populations that generally covered the whole distribution range of *L. glycinivorella* in China (Fig. 1a). In the median-joining network (Fig. 2b), two haplotype groups linked by 18 mutation steps were clearly defined, which corresponded to the two lineages of phylogenetic trees. In Lineage A, 103 haplotypes were linked by one

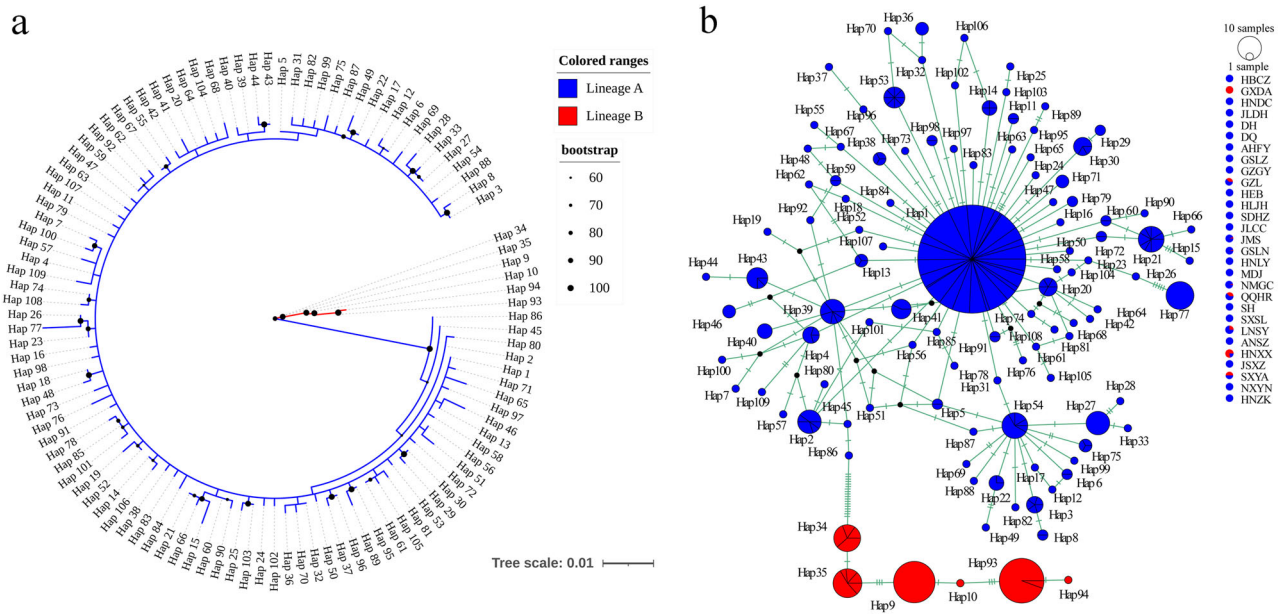

Figure 2 Phylogeny and median-joining network based on 109 haplotypes defined from 543 *cox1* sequences. (a) Maximum likelihood (ML) tree constructed with GTR + G model determined by jModelTest. (b) Median-joining network; circles represent different haplotypes and short lines between haplotypes represent mutation steps; the area of the circle is proportional to the number of individuals belonging to the haplotype; the codes on the right represent sampling localities; the codes marked in blue indicate all individuals for this locality belong to Lineage A; the code marked in red indicate all individuals for this locality belong to Lineage B; the codes marked in both red and blue indicate this locality contain members of both Lineage A and Lineage B.

to five mutation steps, and Hap_1 was shared by up to 27 collecting localities, rendering the star-like haplotype relationships. The remaining six haplotypes constituted Lineage B, which were linked by one to three steps.

Based on 60 mitogenomes, ML (Fig. 3a) and BI (Fig. S5, Supporting Information) analyses yielded identical topologies with two lineages as that of *cox1* data. Lineage A included 55 samples from 20 localities, and the relationships among the 55 samples were generally not related to the geographic distances of sampling localities as well. Lineage B contained five samples from two localities, two from HNXX and three from GXDA, and the samples from the same locality converged together. Among the 58 haplotypes defined from 60 mitogenomes, two haplotype groups (Fig. 3b) linked by 438 mutation steps were clearly defined, corresponding to the two lineages of phylogenetic trees.

Population differentiation based on mitochondrial sequences

Based on *cox1* data, pairwise F_{ST} (Fig. 4a; Table S4, Supporting Information) ranged from −0.069 (between QQHE and GZL and HNZK and JLDH) to 0.998 (between GXDA and MDJ, $P < 0.05$). In Lineage A, pairwise F_{ST} (Table S5, Supporting Information) ranged from −0.069 (between HNZK and JLDH) to 0.763 (between MDJ and SXYA, $P < 0.05$). The F_{ST} between the two lineages was 0.93 ($P < 0.05$). Based on mitogenome data, the pairwise F_{ST} (Fig. 4b; Table S6, Supporting Information) ranged from −0.432 (between SXYA and SXSL) to 0.982 (between GXDA and SDHZ), and the F_{ST} between two lineages was 0.871 ($P < 0.05$).

Phylogeny, population differentiation, and genetic structure based on SNPs

ML tree (Fig. 5a) based on SNPs recovered two clusters corresponding to the two lineages yielded by mitochondrial genes. The F_{ST} results (Fig. 5b) showed the F_{ST} values of the GXDA and HNXX with other populations ranged from 0.164 (between HNXX and HNLY) to 0.255 (between GXDA and HNXX, GXDA and SDHZ), whereas the F_{ST} values among the remaining populations were all near zero. PCA analyses showed that GXDA and HNXX samples were assigned into two separate groups, respectively, and other populations were assembled into one cluster (Fig. 5c) and were not even separated from each other when independently analyzed

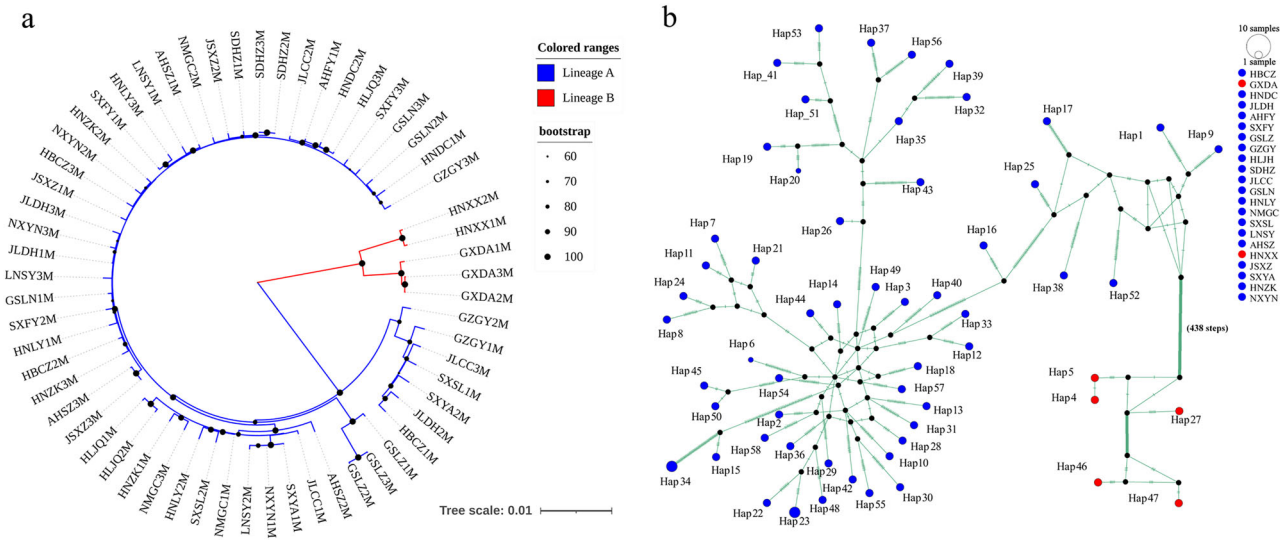

Figure 3 Maximum likelihood (ML) tree and median-joining network based on the 37-gene dataset and 58 haplotypes defined from these sequences. (a) ML tree with partitioned models as shown in Table S1, Supporting Information. (b) Median-joining network; circles represent different haplotypes and short lines between haplotypes represent mutation steps; the area of the circle is proportional to the number of haplotypes; the codes on the right represent the sampling localities referring to Table 1; the codes marked in blue indicate all individuals for this locality belong to Lineage A; the codes marked in red indicate all individuals for this locality belong to Lineage B.

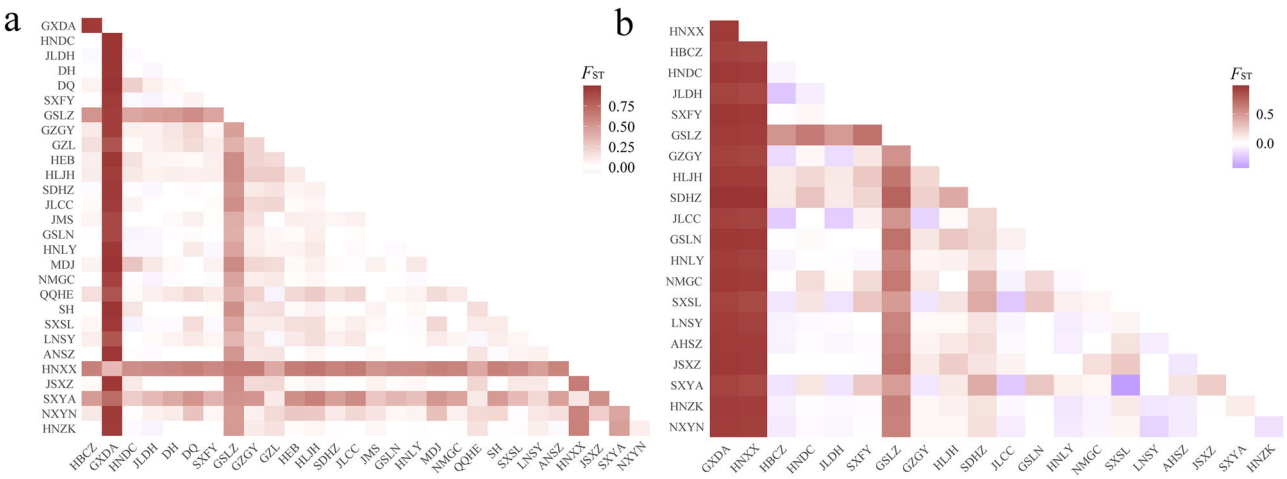

Figure 4 Pairwise differentiation among populations inferred from (a) *cox1* data and (b) mitogenome data.

(Fig. 5d). When *K* was increased from 2 to 7, admixture analysis showed the optimal *K* was 2 for the 46 individuals. In this situation, one cluster included all 7 samples from GXDA and 10 samples from HNXX, and the other 29 samples from the remaining populations constituted the second cluster (Fig. 5e).

Population demography and divergence time

Neutrality tests were conducted for all samples or lineages using both *cox1* and mitogenome data. When all samples were treated as a whole, the Fu's Fs values were significantly negative for both data types. For Lineage A, Tajima's D and Fu's Fs were always significantly

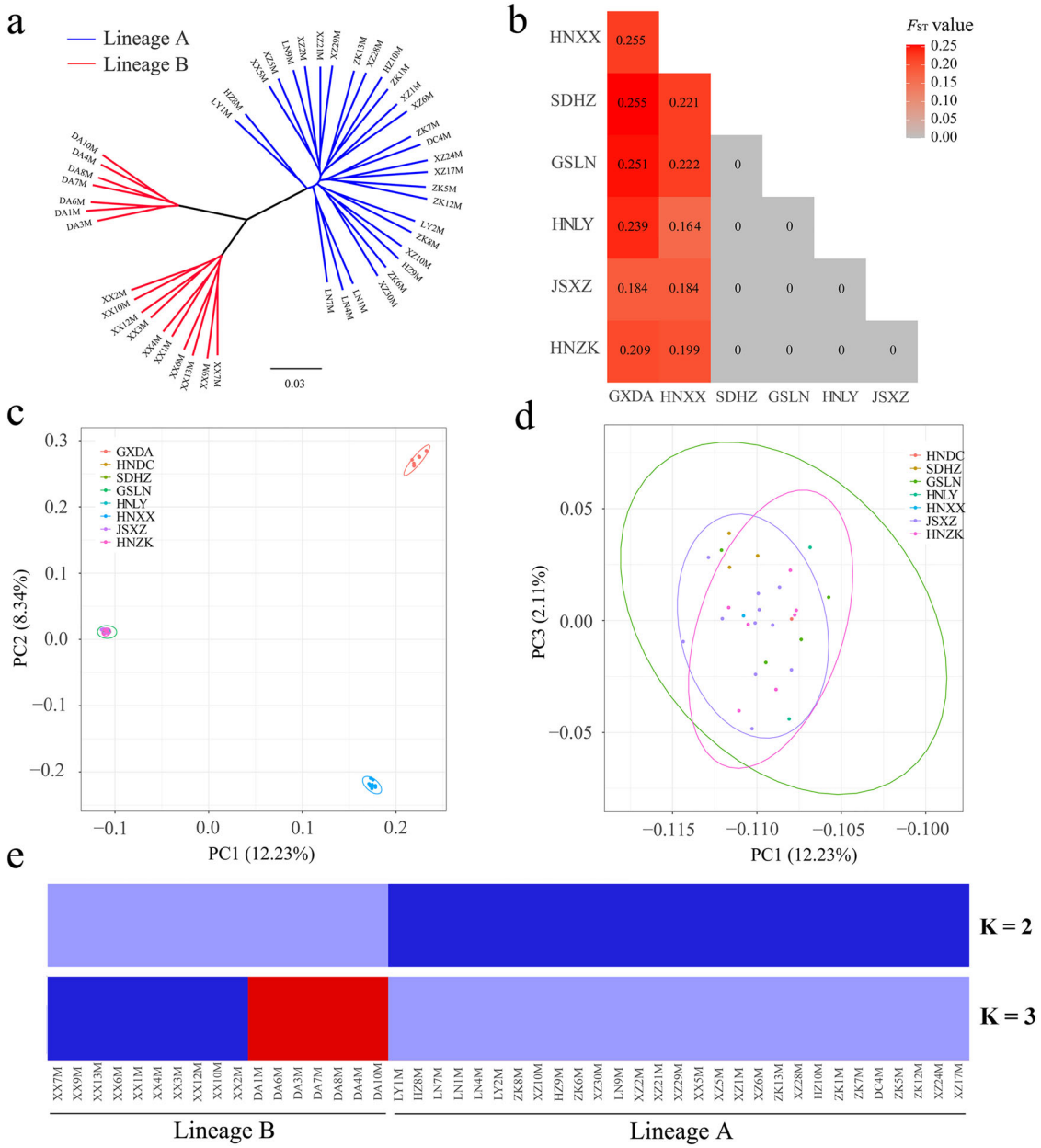

Figure 5 Population genetic analyses of *Leguminivora glycinivorella* inferred from genotyping-by-sequencing (GBS) data. (a) Maximum likelihood (ML) tree constructed with the GTR + G model. (b) Pairwise F_{ST} matrix among populations. (c) Principal component analysis (PCA) analysis based on 46 individuals from eight populations. (d) PCA analysis based on 29 individuals after excluding the samples from Lineage B. (e) The ancestry composition of each individual represented by a vertical bar.

negative. In contrast, most tests showed positive values for Lineage B. These results indicated Lineage A had possibly undergone population expansion (Tables 2,3). Also, the small *r* index and the lack of significant SSD exhibited that populations of Lineage A probably experienced historical expansion events. Analyses of mismatch distribution showed a unimodal structure for Lineage A, also suggesting that this lineage underwent demographic expansion (Fig. 6a–c; Fig. S6, Supporting Information). Moreover, the effective population size of Lineage A increased rapidly at ~ 0.1 million years ago (Ma) corresponding to the periods of climate fluctuation during the

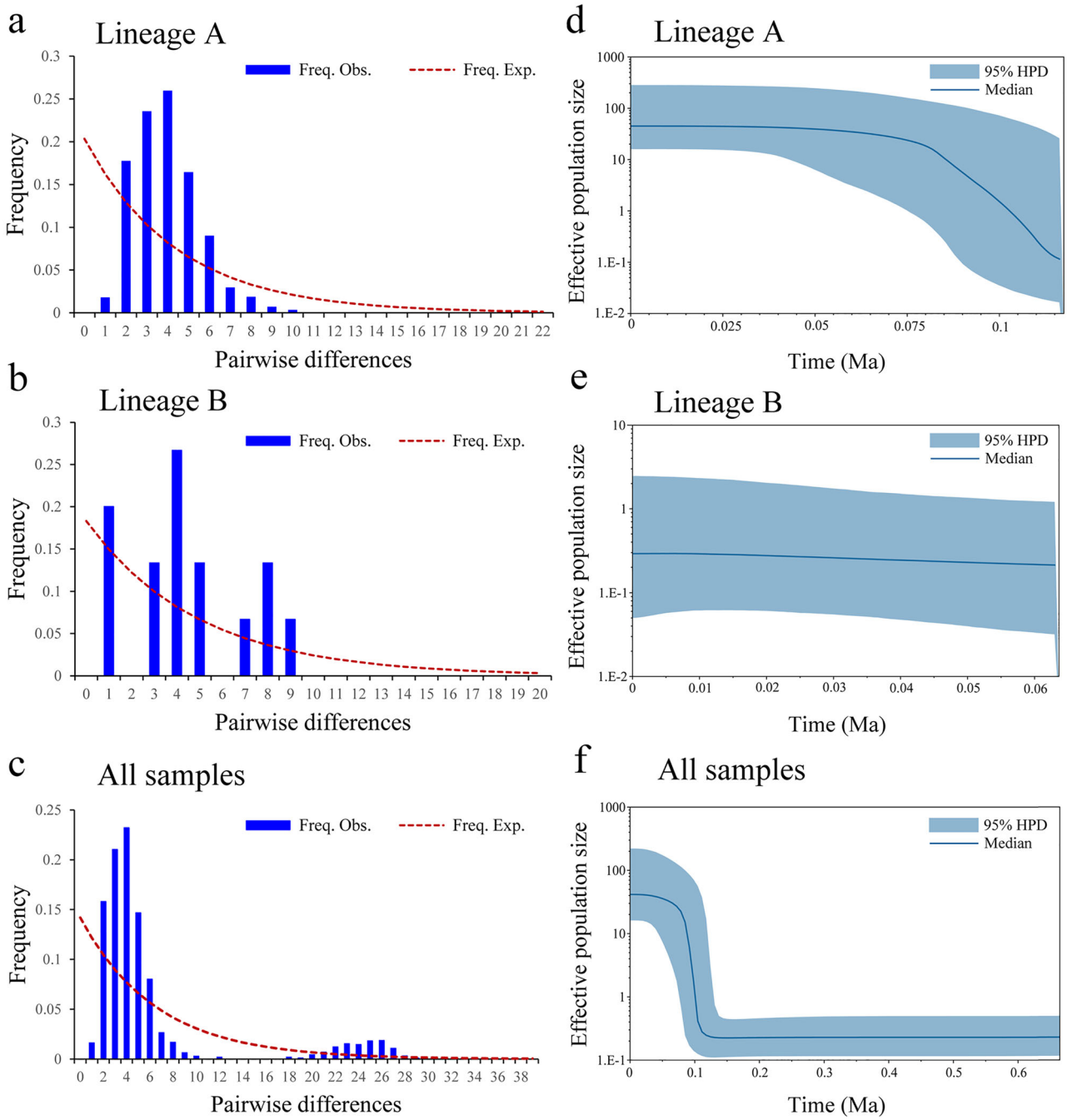

Figure 6 Mismatch distribution analyses and Bayesian skyline plots for different lineages or samples based on *cox1* data. Ma, million years ago.

late Pleistocene, whereas the effective population size of Lineage B generally remained stable (Fig. 6d–f).

A BEAST analysis suggested the estimated time of divergence between the two lineages defined herein was ~1.14 Ma (95% HPD: 0.74–1.59 Ma) (Fig. 7).

Mantel tests

Since the individuals of Lineage B recognized from *L. glycinivorella* samples in this study were relatively limited, the Mantel tests of IBD and IBE were conducted

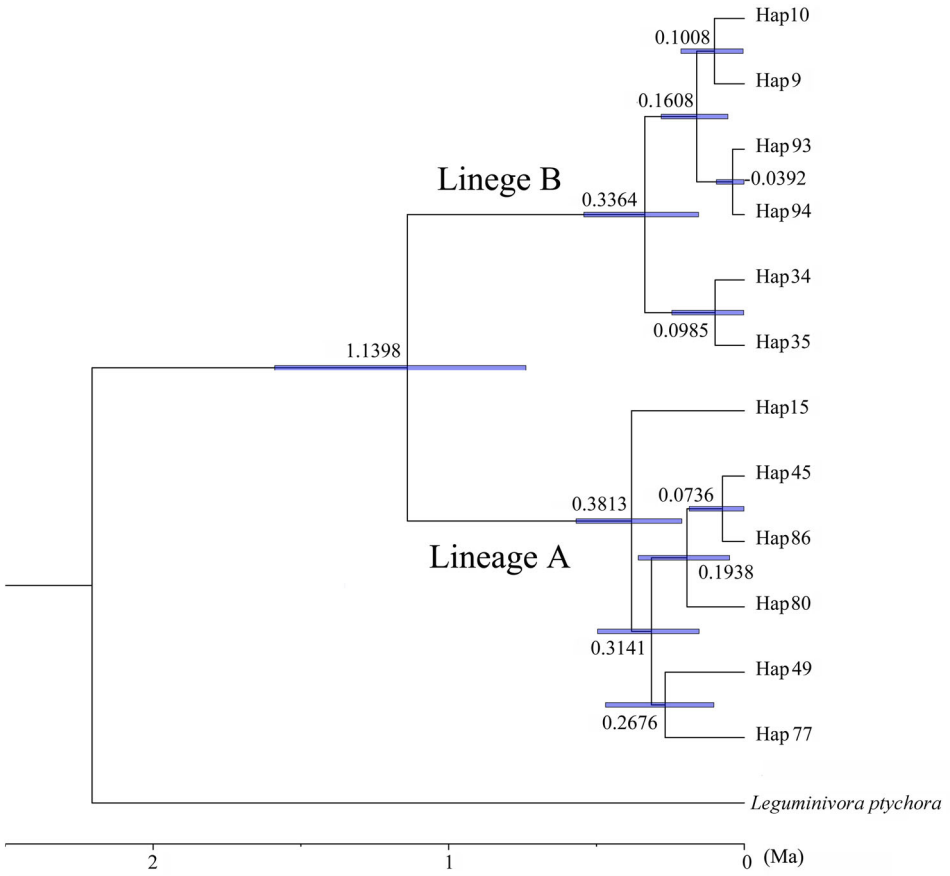

Figure 7 Divergence time of the two lineages defined in *Leguminivora glycinivorella* based on *cox1* data. The purple bars on nodes indicate the 95% highest posterior density intervals and the values above the bars indicate mean divergence time. Ma, million years ago.

for both all samples as a whole and Lineage A. For all samples, significant positive relationships were observed between genetic distance and geographic distance ($R = 0.264$, $P = 0.001$) (Fig. 8a) as well as environmental distance ($R = 0.289$, $P = 0.013$; Fig. 8c). The same situation was found in Lineage A, and both IBD and IBE models were manifested. These results indicated that both geography and environment are associated with population differentiation of *L. glycinivorella* to some extent.

The present and past potential distribution

The average test AUC value was 0.844, indicating that the performance of ecological niche modeling (ENM) analysis was good, and the projection results were reliable. The distribution of predicted suitable habitat under present conditions was generally identical to the geographical range defined by previous studies, especially in central China (Fig. 9a). During LIG, *L. glycinivorella* had a broader distribution range with the major distribution region in South China (Fig. 9b).

DISCUSSION

Incorporating extensive genetic data including mitochondrial *cox1*, mitogenome, and nuclear SNP data, as well as samplings from larger coverage, we recognized the soybean pest *L. glycinivorella* contains two clearly divergent lineages that are widely distributed and sympatric. The *L. glycinivorella* individuals from the Du'An population (coded GXDA in this study), that showed clear genetic differentiation from other sampling localities as revealed by Zhu *et al.* (2017) and Shi *et al.* (2018), are genetically closely related to partial individuals from other populations such as HNXX in central China and SXYA and GSLN in Northwest China and other four localities

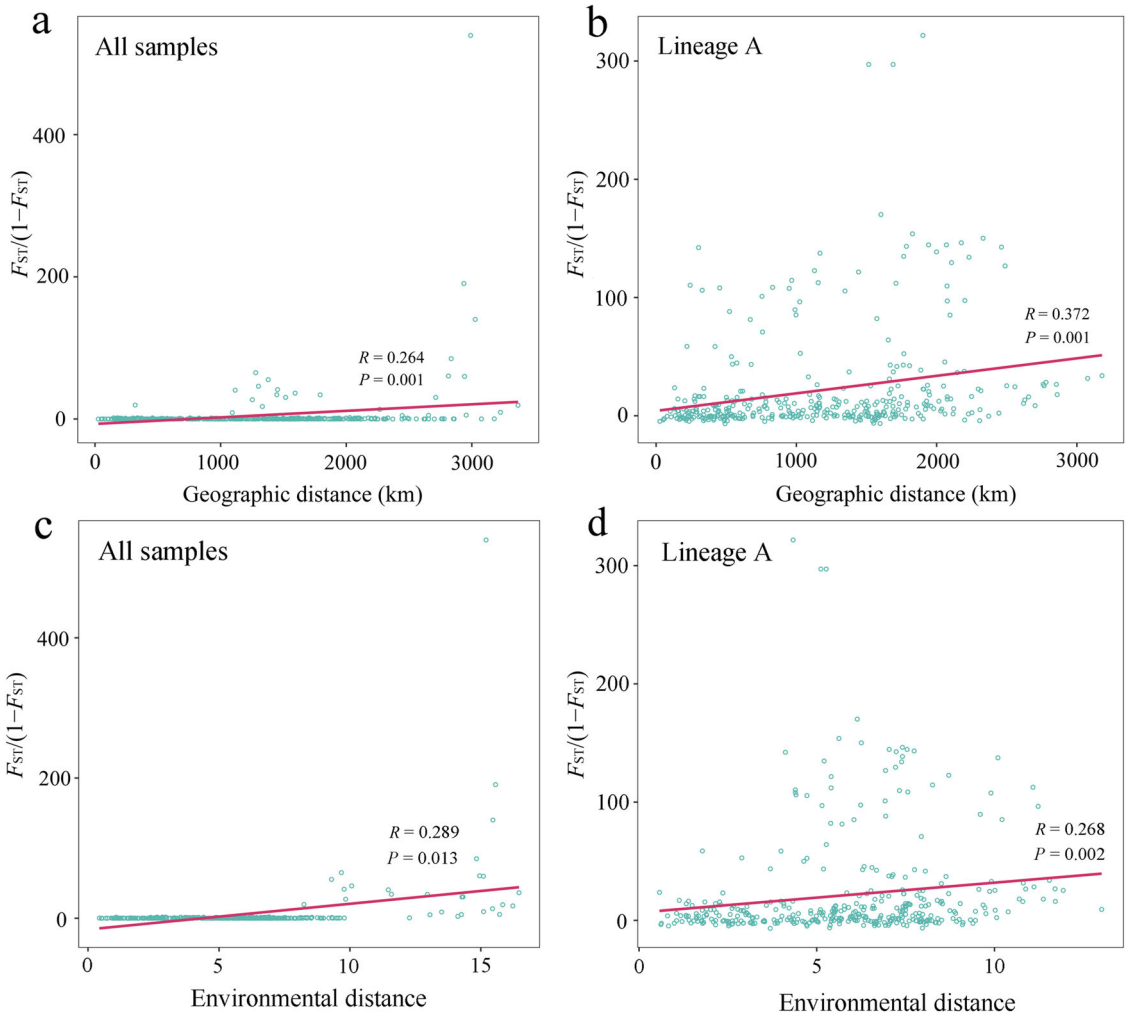

Figure 8 Mantel tests of isolation-by-distance model and isolation-by-environment model among *Leguminivora glycinivorella* populations based on *cox1* data.

in Northeast China (all these individuals are designated as Lineage B in this study). Our results highlight that the members of Lineage B are widely distributed across China and sympatric with Lineage A, rather than only in GXDA in Southwest China as recognized by Zhu *et al.* (2017) and Shi *et al.* (2018).

Geological events and climatic fluctuations during the Quaternary period play a crucial role in shaping geographic distribution and population differentiation (Hewitt 1996, 2004) often with complex genetic variations of species such as the existence of divergent genetic lineages in sympatry. One hypothesis providing explanations for the genetic discontinuities in sympatry is historically secondary contact among divergent populations (Avise *et al.* 1987) and has been used to account for the current genetic patterns for many insects. For instance, the lack of phylogeographic structure among three genetic lineages in a hemipteran species *Geisha distinctissima* across the Zhoushan Archipelago in China may be a consequence of geographic isolation driven by the rise of sea level during the Holocene and subsequent human-mediated secondary contact among divergent populations (Li *et al.* 2020). Similarly, the wide coexistence of L3 and L5 haplotype clusters of the flower thrips *Frankliniella intonsa* is presumably caused by human activities after their divergence occurred during the late Pleistocene (Liu *et al.* 2023b). For *L. glycinivorella*, the estimated time of divergence between the two lineages was ~1.14 Ma (95% HPD: 0.74–1.59 Ma), a history period generally corresponding to uplift of the Yungui Plateau and Kunlun-Huanghe

Table 2 Genetic diversity and neutrality tests of 543 *cox1* sequences and two lineages

	Number of samples	Number of haplotypes	Haplotype diversity	Nucleotide diversity	Tajima's D (*P*-value)	Fu's Fs (*P*-value)	SSD (*P*-value)	*r* (*P*-value)
All samples	543	109	0.837	0.01322	−1.031 (>0.1)	−67.714 (0.000)	0.027 (0.330)	0.025 (0.640)
Lineage A	446	103	0.771	0.00337	−2.387 (<0.01)	−175.487 (0.000)	0.001 (0.960)	0.01 (0.990)
Lineage B	97	6	0.726	0.00519	0.248 (0.632)	5.097 (0.013)	0.105 (0.000)	0.238 (0.010)

SSD, sum of square deviations; *r*, Harpending's raggedness index.

Table 3 Genetic diversity and neutrality tests of 60 mitochondrial genomes and two lineages

	Number of samples	Number of haplotypes	Haplotype diversity	Nucleotide diversity	Tajima's D (*P*-value)	Fu's Fs (*P*-value)	SSD (*P*-value)	*r* (*P*-value)
All samples	60	58	0.999	0.00827	−1.530 (>0.1)	−8.043 (0.000)	0.004 (0.900)	0.001 (0.660)
Lineage A	55	53	0.999	0.00385	−2.110 (<0.05)	−14.807 (0.000)	0.005 (0.660)	0.001 (0.810)
Lineage B	5	5	1	0.00665	1.472 (>0.1)	2.252 (0.905)	0.135 (0.070)	0.100 (1.000)

SSD, sum of square deviations; *r*, Harpending's raggedness index.

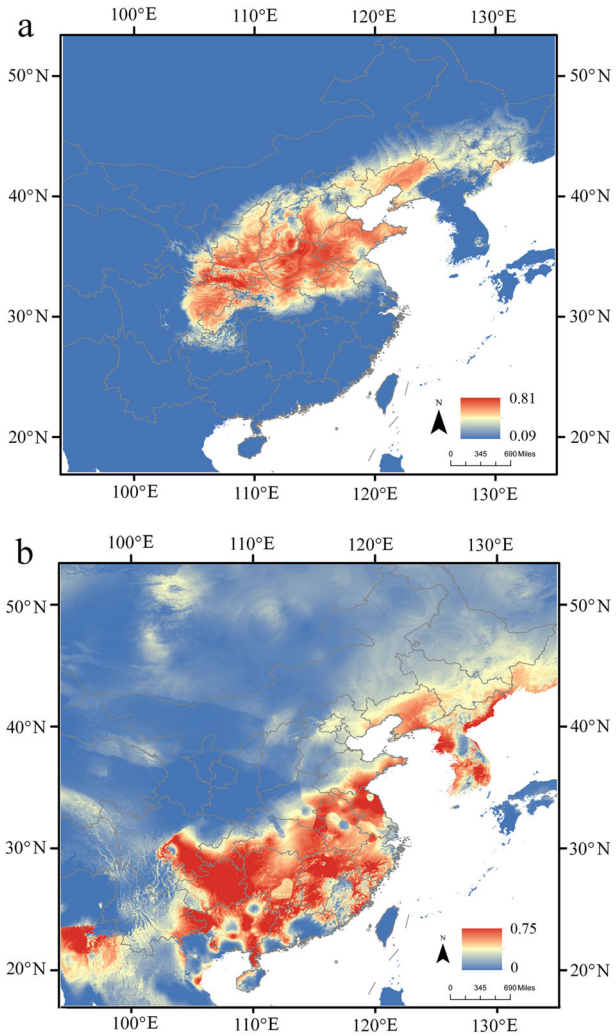

Figure 9 Potential distribution of the suitable habitat of *Leguminivora glycinivorella* predicted under (a) the nearly present condition (during 1960–1990) and (b) the Last Inter-glacial period (~120 000–140 000 years before).

tectonic movement in China (Cui *et al.* 1998; Zachos *et al.* 2001), indicating the geological event may play an important role in the division of *L. glycinivorella* into two lineages. Then, members of two lineages began to evolve independently and finally resulted in genetic divergence. It was ~0.1 Ma, which generally corresponds to the last interglacial period (Otto-Bliesner *et al.* 2006). Climate warming afforded the expansion of the distribution range of Lineage A as revealed by demographic analyses from South China or Southwest China toward central China as revealed by ENM analysis, which made the secondary contact of the two lineages probable. Further, from the perspective of genetic diversity, the members of Lineage A from Northeast China may be the newly colonized individuals because of their low haplotype diversity (Pang *et al.* 2009). In contrast to Lineage A, Lineage B has a relatively stable population history. Unfortunately, we failed to further reveal the demographic history using the ENM methods due to the limited occurrence points we recognized in the present study.

Although gene exchange might occur between two divergent groups upon contact, resulting in the genetic homogenization of populations, there is still a relatively high level of genetic differentiation ($F_{ST} = 0.93$) between the two lineages in *L. glycinivorella* as we revealed. This result may be because the time to the present (~0.1 Ma) was insufficient for gene exchange between them, especially in the context of the long life history (one generation per year) of *L. glycinivorella*, which limited the rate of genetic recombination. Alternatively, the two lineages before secondary contact might represent incipient speciation and the capacity of reproductive communication may have already been limited, which raises the issue of whether the two lineages represent different species.

Species represent the fundamental unit for various biological studies such as diversity analysis, ecology, and evolution (Isaac *et al.* 2004; de Queiroz 2007). For pest species, it is particularly imperative to delimit species because accurate species identification may facilitate pest management. Regarding the divergent status of the two lineages, analyses of 543 barcode sequences showed the average pairwise genetic distance between the two lineages was 0.038, and the deep genetic divergence is also revealed by other mitochondrial genes of *L. glycinivorella*. For the *cox1* barcode, a threshold of 3% interspecies divergence was proposed by Hebert *et al.* (2003) and clarified by later investigations (Hebert *et al.* 2004; Behere *et al.* 2007) in Lepidoptera. The 3.8% between the two lineages found herein indicates the clear divergence of *L. glycinivorella*. However, selecting a threshold of divergence to delimit species is both challenging and often contentious (Wiemers & Fiedler 2007; van Nieukerken *et al.* 2012). *Leguminivora* Obraztsov, 1960 included seven described extant species, and *L. glycinivorella* is the only *Leguminivora* species recorded in China (Li 2012; Gilligan *et al.* 2018). Since the other six *Leguminivora* species (Gilligan *et al.* 2018) have not been barcoded on GenBank or BOLD systems, 733 *cox1* barcode sequences of 63 *Cydia* spp., which are close to *Leguminivora*, were downloaded and analyzed from the BOLD systems. The average pairwise distances among these species ranged from 0.015 to 0.18, and there were 11 species pairs with distances $\leq$ 0.038, indicating the two lineages detected

herein within *L. glycinivorella* hold the possibility of representing two closely related species. However, we tentatively refrain from concluding that the two lineages represent distinct species. Further investigations based on morphology and biology are necessary to clarify this issue.

When gene exchange among populations is limited by geographical and climatic barriers, local adaptation can be induced through natural selection under different ecological environments (Egger *et al.* 2017; Cao *et al.* 2021; Liang *et al.* 2023). Among *L. glycinivorella* populations, we found low, but statistically significant support for both IBD and IBE, indicating the existence of local adaptation to some extent, which provides support to the results of Zhu *et al.* (2017) and Shi *et al.* (2018). This evolutionary process is also revealed by the high genetic diversity of *L. glycinivorella* populations recognized herein, which indicates accumulation of genetic heterozygosity generally representing high adaptability to local habitats (Liang *et al.* 2023). As one of the important environmental factors responsible for genetic variation and further leading to local adaptation, topographical isolation has played a key role in driving the genetic differentiation and even speciation of populations (Xu *et al.* 2021; Liu *et al.* 2023a; Zhao *et al.* 2023). Interestingly, *L. glycinivorella* samples from mountainous regions in western China (such as populations GSLZ, GZGY, and SXYA) show higher differentiation from other samples. In particular, the F_{ST} values of the GSLZ population (about 2000 m above sea level) are even greater than 0.5 with significant levels ($P < 0.05$). In addition to climate heterogeneity and geographic barrier, we speculate that local adaptation of *L. glycinivorella* is closely associated with its low dispersal ability in adults (Wang *et al.* 2014). Moreover, the opportunity for *L. glycinivorella*, like some other pest species such as the oriental fruit moth *G. molesta* that can spread long distance through host transport and trade, to spread over a long distance through human activity is limited because the larvae exit the soybean pods to hibernate in the soil prior to seed harvest and it is unlikely to be transported through the seed trade (Areces-Berazain 2022).

Molecular markers provide crucial support for the investigations of insect phylogeography and population genetics. As one of the earliest and most efficient molecular markers, mitochondrial gene fragments have gained wide use (Avise *et al.* 1987; Hewitt 2004; Liang *et al.* 2023). Based on mitochondrial *cox1*, *cox2*, and *cytb* fragments, several studies (Wang *et al.* 2014, 2015; Zhu *et al.* 2017; Shi *et al.* 2018) attempted to investigate the genetic divergence of this pest, which greatly advanced our understanding of its genetic pattern. In the present study, our results, based on the standard *cox1* sequences but samplings from larger coverage, support the findings of Zhu *et al.* (2017) and Shi *et al.* (2018) in the existence of local adaptation but show differences in the population demography. Likewise, mitogenome has proven effective for resolving intraspecific divergence history (Morin *et al.* 2010; Ma *et al.* 2012; Du *et al.* 2019, 2023). We sequenced and assembled the mitogenome for *L. glycinivorella* for the first time, and comparative analyses of 60 mitogenomes show generally identical genetic patterns with that based on *cox1* data. In addition, we analyzed nuclear SNP data from samples representing the two lineages and typical sampling localities given that mitochondrial genes cannot provide complete information representing the evolutionary history of both parents and may experience introgression in given species and produce mito-nuclear discordance (Linnen & Farrell 2007; Weigand *et al.* 2017; Hinojosa *et al.* 2019; Campbell *et al.* 2020; Liang *et al.* 2023). Results of phylogeny, PCA, and Admixture all confirm the two divergent lineages of *L. glycinivorella* as revealed by mitochondrial genes.

CONCLUSIONS

For pest species, information on genetic diversity and population structure is particularly important since it can guide pest management (Beheregaray 2008; Milankov *et al.* 2013; Liang *et al.* 2023). The two lineages of *L. glycinivorella* recovered herein that are deeply divergent and sympatric highlight that perhaps different control strategies against this pest should be developed. Besides, we presume that the wide sympatry of the two lineages may be due to secondary contact driven by climatic fluctuation during the late Pleistocene. In the field, it is very difficult to distinguish the members of the two lineages. Using molecular evidence, we found only less than 20% of *L. glycinivorella* individuals belonged to Lineage B. Besides, the population haplotype diversity of Lineage B is lower than most of the populations of Lineage A. These results indicate Lineage A may have a higher adaptive ability to the coexisting environment. In the future, genome-wide molecular markers, together with morphology and biology data, should be considered to gain a more comprehensive understanding of the divergent mechanisms of the two lineages and their ecological adaptabilities.

ACKNOWLEDGMENTS

We thank all the institutions and individuals who provided assistance in the sample collection. This work was supported by the Natural Science Foundation of China (31702046 and 32170421), the Beijing Municipal Natural Science Foundation (5232001), the Support Project of High-level Teachers in Beijing Municipal Universities in the Period of 14th Five-year Plan (BPHR20220114), and the Young Backbone Teacher Guiding Foundation in Colleges and Universities in Henan Province (2020GGJS211).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The genetic data of *Leguminivora glycinivorella* used in this study have been deposited in GenBank with accession numbers OQ651860–OQ651968 for *cox1* haplotypes, MH013480–MH013481 and MZ506761–MZ506818 for mitochondrial genomes, and SRR28713114–SRR28713159 for genotyping-by-sequencing data.

REFERENCES

Abascal F, Zardoya R, Telford MJ (2010). TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* **38**, 7–13.

Akaike H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–23.

Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655–64.

Areces-Berazain F (2022). Leguminivora glycinivorella *(soybean pod borer)*. CABI Compendium. Available from URL: https://doi.org/10.1079/cabicompendium.29902

Avise JC, Arnold J, Ball RM *et al.* (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology & Systematics* **18**, 489–522.

Basoalto A, Ramírez CC, Lavandero B *et al.* (2020). Population genetic structure of codling moth, *Cydia pomonella* (L.) (Lepidoptera: Tortricidae), in different localities and host plants in Chile. *Insects* **11**, 285.

Behere GT, Tay WT, Russell DA *et al.* (2007). Mitochondrial DNA analysis of field populations of *Helicoverpa armigera* (Lepidoptera: Noctuidae) and of its relationship to *H. zea*. *BMC Evolutionary Biology* **7**, 117.

Beheregaray LB (2008). Twenty years of phylogeography: The state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology* **17**, 3754–74.

Bernt M, Donath A, Jühling F *et al.* (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**, 313–19.

Bouckaert R, Heled J, Kühnert D *et al.* (2014). BEAST2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**, e1003537.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–35.

Brower AVZ (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *PNAS* **91**, 6491–95.

Campbell EO, Gage EV, Gage RV, Sperling FAH (2020). Single nucleotide polymorphism-based species phylogeny of greater fritillary butterflies (Lepidoptera: Nymphalidae: Speyeria) demonstrates widespread mitonuclear discordance. *Systematic Entomology* **45**, 269–80.

Cao LJ, Li BY, Chen JC, Zhu JY, Hoffmann AA, Wei SJ (2021). Local climate adaptation and gene flow in the native range of two co-occurring fruit moths with contrasting invasiveness. *Molecular Ecology* **30**, 4204–19.

Cao LJ, Song W, Chen JC, Fan X L, Hoffmann AA, Wei SJ (2022). Population genomic signatures of the oriental fruit moth related to the Pleistocene climates. *Communications Biology* **5**, 142.

Cui ZJ, Wu YQ, Liu GN, Ge DK, Pang QQ, Xu QH (1998). On Kunlun-yellow river tectonic movement. *Science in China Series D Earth Sciences* **41**, 592–600.

Danecek P, Auton A, Abecasis GR *et al.* (2011). The variant call format and VCFtools. *Bioinformatics* **27**, 2156–58.

de Queiroz K (2007). Species concepts and species delimitation. *Systematic Biology* **56**, 879–86.

Dombroskie J, Sperling FAH (2013). Phylogeny of the tribe Archipini (Lepidoptera: Tortricidae: Tortricinae) and evolutionary correlates of novel secondary sexual structures. *Zootaxa* **3729**, 1–62.

Dong F, Hung CM, Yang XJ (2020). Secondary contact after allopatric divergence explains avian speciation and high species diversity in the Himalayan-Hengduan Mountains. *Molecular Phylogenetics and Evolution* **143**, 106671.

Du Z, Hasegawa H, Cooley JR *et al.* (2019). Mitochondrial genomics reveals shared phylogeographic patterns and demographic history among three periodical cicada species groups. *Molecular Biology and Evolution* **36**, 1187–200.

Du Z, Zhao Q, Wang X *et al.* (2023). Climatic oscillation promoted diversification of spinous assassin bugs during Pleistocene glaciation. *Evolutionary Applications* **16**, 880–94.

Egger B, Roesti M, Böhne A, Roth O, Salzburger W (2017). Demography and genome divergence of lake and stream populations of an east African cichlid fish. *Molecular Ecology* **26**, 5016–30.

Excoffier L, Lischer HE (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–67.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994). DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294–99.

Fu YX (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–25.

Fuentes-Contreras E, Espinoza JL, Lavandero B, Ramírez CC (2008). Population genetic structure of codling moth (Lepidoptera: Tortricidae) from apple orchards in central Chile. *Journal of Economic Entomology* **101**, 190–98.

Gilligan TM, Baixeras J, Brown JW (2018). *T@RTS: Online World Catalogue of the Tortricidae* (Version 4.0). Available from URL: http://www.tortricidae.com/catalogueGenusList.asp?gcode=448

Hebert PD, Cywinska A, Ball SL, de Waard JR (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270**, 313–21.

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004). Ten species in one: DNA barcode reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *PNAS* **101**, 14812–17.

Hewitt GM (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247–76.

Hewitt GM (2004). Genetic consequences of climatic oscillations in the quaternary. *Philosophical Transactions of the Royal Society of London Series B* **359**, 183–95.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**, 1965–78.

Hinojosa JC, Koubínová D, Szenteczki MA *et al.* (2019). A mirage of cryptic species: Genomics uncover striking mitonuclear discordance in the butterfly *Thymelicus sylvestris*. *Molecular Ecology* **28**, 3857–68.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**, 518–22.

Isaac NJ, Mallet J, Mace GM (2004). Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology and Evolution* **19**, 464–69.

Kass JM, Muscarella R, Galante PJ *et al.* (2021). ENMeval 2.0: Redesigned for customizable and reproducible modelling of species' niches and distributions. *Methods in Ecology and Evolution* **12**, 1602–8.

Katoh K, Rozewicki J, Yamada KD (2017). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics* **20**, 1160–66.

Kearse M, Moir R, Wilson A *et al.* (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–49.

Kirk H, Dorn S, Mazzi D (2013). Worldwide population genetic structure of the oriental fruit moth (*Grapholita molesta*), a globally invasive pest. *BMC Ecology* **13**, 12.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–55.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–49.

Li H (2012). *Microlepidoptera of Qinling Moutains (Insecta: Lepidoptera)*. Beijing Science Press, China.

Li H (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, 1303.3997.

Li H, Handsaker B, Wysoker A *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–79.

Li K, Lin CP, Liang AP (2020). Comparative phylogeography of two hemipteran species (*Geisha distinctissima* and *Megacopta cribraria*) in the Zhoushan Archipelago of China reveals contrasting genetic structures despite concordant historical demographies. *Heredity* **124**, 207–22.

Li Y, Duan X, Qiao X *et al.* (2015). Mitochondrial DNA revealed the extent of genetic diversity and invasion origin of populations from two separate invaded areas of a newly invasive pest, *Cydia pomonella* (L.) (Lepidoptera: Tortricidae) in China. *Bulletin of Entomological Research* **105**, 485–96.

Liang Y, Du S, Jin Z *et al.* (2023). Population genetics and ecological niche modelling provide insights into management strategies of the herbivorous pest *Phytomyza horticola* (Diptera: Agromyzidae). *Diversity and Distributions* **29**, 1371–87.

Linnen CR, Farrell BD (2007). Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution* **61**, 1417–38.

Liu X, He J, Du Z, Zhang R, Cai W, Li H (2023b). Weak genetic structure of flower thrips *Frankliniella intonsa* in China revealed by mitochondrial genomes. *International Journal of Biological Macromolecules* **231**, 123301.

Liu Y, Bu Y, Wang J, Wei C (2023a). Geological events and climate change drive diversification and speciation of mute cicadas in eastern continental Asia. *Molecular Phylogenetics and Evolution* **184**, 107809.

Luo R, Liu B, Xie Y *et al.* (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18.

Ma C, Yang P, Jiang F *et al.* (2012). Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. *Molecular Ecology* **21**, 4344–58.

McKenna A, Hanna M, Banks E *et al.* (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297–303.

Milankov V, Ludoški J, Francuski L, Stahls G, Vujic A (2013). Genetic and phenotypic diversity patterns in *Merodon albifrons* Meigen, 1822 (Diptera: Syrphidae): Evidence of intraspecific spatial and temporal structuring. *Biological Journal of the Linnean Society* **110**, 257–80.

Morin PA, Archer FI, Foote AD *et al.* (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research* **20**, 908–16.

Muscarella R, Galante PJ, Soley-Guardia M *et al*. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models. *Methods in Ecology and Evolution* **5**, 1198–205.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**, 268–74.

Otto-Bliesner BL, Marshall SJ, Overpeck JT, Miller GH, Hu A, CAPE Last Interglacial Project members (2006). Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751–53.

Pang JF, Kluetsch C, Zou XJ *et al.* (2009). mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Molecular Biology and Evolution* **26**, 2849–64.

Phillips SJ, Anderson RP, Schapire RE (2006). Maximum entropy modelling of species geographical distributions. *Ecological Modelling* **190**, 231–59.

Posada D (2008). jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253–56.

Purcell S, Neale B, Todd-Brown K *et al.* (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–75.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**, 901–4.

Ran WW, Luo GM, Zhao YQ, Li C, Dietrich CH, Song YH (2024). Climate change may drive the distribution of tribe Zyginelline pests in China and the Indo-China Peninsula to shift towards higher latitude river-mountain systems. *Pest Management Science* **80**, 613–26.

Rollins LA, Woolnough AP, Sherwin WB (2006). Population genetic tools for pest management: A review. *Wildlife Research* **33**, 251–61.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**, 539–42.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC *et al.* (2017). DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution* **34**, 3299–302.

Schubert M, Lindgreen S, Orlando L (2016). AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Research Notes* **9**, 88.

Shi S, Cui J, Zhu S, Xu W, Wang X (2018). Genetic differentiation among geographic populations of *Leguminivora glycinivorella* (Lepidoptera: Olethreutidae) based on mitochondrial COI gene sequences. *Acta Entomologica Sinica* **45**, 214–22.

Silva-Brandão KL, Silva OABN, Brandão MM, Omoto C, Sperling FAH (2015). Genotyping-by-sequencing approach indicates geographic distance as the main factor affecting genetic structure and gene flow in Brazilian populations of *Grapholita molesta* (Lepidoptera, Tortricidae). *Evolutionary Applications* **8**, 476–85.

Song PX (2014). Study on the population differentiation of *Leguminivora glycinivorella* Matsumura and the adaption of larvae to low temperature (Dissertation). Jilin Agricultural University, Changchun.

Song W, Cao LJ, Li BY (2018). Multiple refugia from penultimate glaciations in East Asia demonstrated by phylogeography and ecological modelling of an insect pest. *BMC Evolutionary Biology* **18**, 152.

Steenwyk JL, Buida TJ, Li YN, Shen XX, Rokas A (2020). ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLoS Biology* **18**, e3001007.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–95.

van der Geest LPS, Evinhuis HH (1991). *Tortricid Pests, Their Biology, Natural Enemies and Control*. Elsevier Science Publishers, Amsterdam.

van Nieukerken EJ, Doorenweerd C, Stokvis FR, Groenenberg DSJ (2012). DNA barcoding of the leafmining moth subgenus *Ectoedemia* s. str. (Lepidoptera: Nepticulidae) with COI and EF1-α: Two are better than one in recognising cryptic species. *Contributions to Zoology* **81**, 1–24.

Wang H, Han LL, Xu ZX *et al.* (2015). Absence of geographic population structure in the soybean pod borer *Leguminivora glycinivorella* (Tortricidae). *Journal of the Lepidopterists' Society* **69**, 173–82.

Wang H, Xu ZX, Han LL, Wang KQ, Zhao KJ (2014). Analysis of the genetic diversity in geographic populations of *Leguminivora glycinivorella* (Lepidoptera: Olethreutidae) from northeastern China based on mitochondrial DNA COI gene sequences. *Acta Entomologica Sinica* **57**, 1051–60.

We T, Simko V (2021). *R package 'corrplot': Visualization of a Correlation Matrix* (Version 0.92). Available from URL: https://github.com/taiyun/corrplot

Wei S, Cao L, Gong Y *et al.* (2015). Population genetic structure and approximate Bayesian computation analyses reveal the southern origin and northward dispersal of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) in its native range. *Molecular Ecology* **24**, 4094–111.

Weigand H, Weiss M, Cai HM *et al.* (2017). Deciphering the origin of mito-nuclear discordance in two sibling caddisfly species. *Molecular Ecology* **26**, 5705–15.

Wiemers M, Fiedler K (2007). Does the DNA barcoding gap exist?—A case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology* **4**, 8.

Xu W, Dong WJ, Fu TT *et al.* (2021). Herpetological phylogeographic analyses support a Miocene focal point of Himalayan uplift and biological diversification. *National Science Review* **8**, nwaa263.

Yang M, Wang Z, Wang R *et al.* (2020). Transcriptomic and proteomic analyses of the mechanisms of overwintering diapause in soybean pod borer (*Leguminivora glycinivorella*). *Pest Management Science* **76**, 4248–57.

Yule GU (1925). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London Series B* **213**, 21–87.

Zachos JC, Pagani M, Sloan L, Thomas E, Billups K (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–93.

Zhang ZJ, Fu CQ (1983). Inheritance of resistance to soybean pod borer in soybean. In: Irwin BJ, Sinclair JB, Wang J, eds. Soybean Research in China and the United States: Proceedings of the First China/USA Symposium and Working Group Meeting. University of Illinois at Urbana-Champaign, Urbana, IL, pp. 74–79.

Zhao Y, Su C, He B *et al.* (2023). Dispersal from the Qinghai-Tibet plateau by a high-altitude butterfly is associated with rapid expansion and reorganization of its genome. *Nature Communications* **14**, 8190.

Zhu SY, Xu W, Gao Y, Cui J, Ou SQ, Shi SS (2017). Analysis of the genetic differentiation among geographic populations of *Leguminivora glycinivorella* (Lepidoptera: Olethreutidae) in China based on mitochondrial COII gene sequences. *Acta Entomologica Sinica* **60**, 475–86.

SUPPLEMENTARY MATERIALS

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

Figure S1 Mitochondrial genome map of *L. glycinivorella* (with the individual JMC1M as an example).

Figure S2 Haplotype diversity of *L. glycinivorella* populations. (a) Lineage A. (b) Lineage B.

Figure S3 Pairwise distances (K2P) within each lineage or between lineages for 13 PCGs and the datasets of combined 13 PCGs and 37 genes. The A and B on horizontal axis indicate Lineage A and Lineage B, respectively.

Figure S4 BI tree based on 109 haplotypes defined from 543 *cox1* sequences.

Figure S5 BI tree based on the 37-gene dataset.

Figure S6 Mismatch distribution analyses based on mitochondrial genomes.

Table S1 Nucleotide substitution models for data partitions selected by jModelTest

Table S2 Information on the 60 mitogenomes generated in this study

Table S3 Population genetic diversity of *L. glycinivorella* based on *cox1* sequences

Table S4 Pairwise differentiation (F_{ST}) values among populations based on *cox1* data

Table S5 Pairwise differentiation (F_{ST}) values among Lineage A populations based on *cox1* data

Table S6 Pairwise differentiation (F_{ST}) values among populations based on mitogenome data

Cite this article as:

Yang M, Wang Y, Dai P *et al.* (2024). Sympatric diversity pattern driven by the secondary contact of two deeply divergent lineages of the soybean pod borer *Leguminivora glycinivorella*. *Integrative Zoology* **00**,1–20. https://doi.org/10.1111/1749-4877.12917