

SCIENTIFIC REPORTS



OPEN

Genetic variation and phylogeographic structure of the cotton aphid, *Aphis gossypii*, based on mitochondrial DNA and microsatellite markers

Xing-Ya Wang^{1,2}, Xian-Ming Yang¹, Bin Lu³, Li-Hong Zhou⁴ & Kong-Ming Wu¹

Aphis gossypii, one of the most important agricultural pests in the world, can cause serious economic losses in the main crop-producing areas. To clarify issues such as the genetic differentiation, genetic structure, and demographic history of *A. gossypii* populations, we used 10 nuclear microsatellite loci (SSR) and two mitochondrial gene sequences (*COI* and *Cytb*) to investigate genetic diversity and population structure of *A. gossypii* populations that were collected from 33 sampling sites in China from different climatic zones. SSR and mtDNA data suggested low to moderate levels of genetic diversity. A star-shaped network of mtDNA haplotypes indicated that the maternal ancestor of China cotton aphids likely originated in Xinjiang. The POPTREE, STRUCTURE and principal coordinate analysis (PCoA) revealed two genetic clusters: an eastern and a western region group. Isolation by distance (IBD) results showed a positive correlation between geographic distance and genetic distance in the vast eastern region but not in the western region. Neutrality testing and mismatch distribution analysis provided strong evidence for a recent rapid expansion in most populations. Genetic bottleneck was not detected in *A. gossypii* populations of China. The present work can help us to develop strategies for managing this pest.

The cotton aphid, *Aphis gossypii* Glover (Hemiptera: Aphididae), is an important cosmopolitan pest that attacks numerous cultivated plants, such as Cucurbitaceae, Rutaceae and Malvaceae, causing serious economic losses in major crop areas around the world. Widely distributed in tropical, subtropical and temperate regions, it is the most common aphid and the primary target pest of cotton in the world¹. Generally, adults and nymphs of this species feed on the underside of leaves or the growing tips of shoots, sucking juices from the plant, causing seedling death, leaf curl and withering, and serious yield loss. Aside from direct damage, this pest is an important vector of various viruses such as mosaic, crinkle, and rosette, and among others^{2,3}. *A. gossypii* is a highly fecund, extremely polyphagous species, and host-adapted races have been identified²⁻⁶. In general, this species can be holocyclic with *Hibiscus*, *Catalpa* or *Rhamnus* trees as primary hosts in China, the United States and Japan. Its life cycle includes both parthenogenetic and sexually reproductive phases². In spring, each egg gives rise to a wingless viviparous, parthenogenetic female, followed by several more parthenogenetic generations during the spring and summer⁷. At the end of autumn, parthenogenetic females give rise to a single sexual generation of males and females, and mated females lay fertilized eggs, which represent the overwintering stage^{2,8}. To our knowledge, *A. gossypii* has developed resistance to all of the original major groups of synthetic insecticides in the last few decades and in virtually all cotton-growing areas worldwide⁹⁻¹². Therefore, this pest has quickly become

¹State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, 100193, P. R. China. ²Plant Protection College, Shenyang Agricultural University, Shenyang, Liaoning, 110866, P.R. China. ³Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, Sichuan, 610041, P.R. China. ⁴Institute of Flower Research, Liaoning Academy of Agricultural Sciences, Shenyang, Liaoning, 110161, P.R. China. Xing-Ya Wang, Xian-Ming Yang and Bin Lu contributed equally to this work. Correspondence and requests for materials should be addressed to K.-M.W. (email: kmwu@ippcaas.cn)

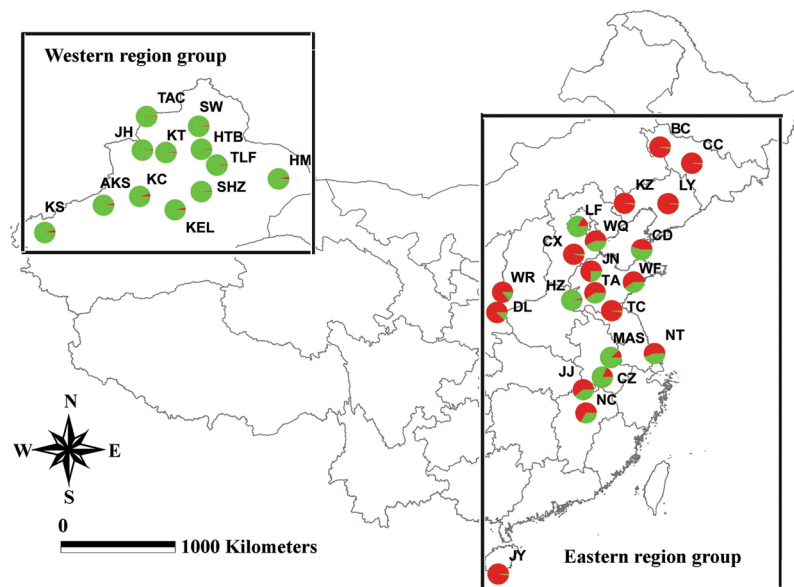


Figure 1. Sampling locations of *Aphis gossypii* and repartition of microsatellite lineages in greater China. Lineage 1 is shown in red and Lineage 2 is shown in green. Groups revealed by STRUCTURE analysis of microsatellite data are shown. For population abbreviations, see Table 2. ArcView GIS (version 3.2, <http://www.resources.esri.com>) was used to produce a distribution map based on the geographical coordinates of the localities, which were obtained with a Global Positioning System receive. The base map (Greater China 1: 4,000,000) for the analysis was obtained the URL: <http://www.diva-gis.org/gdata>.

a major problem when chemical controls fail because of resistance. Thus, an effective control method is essential and urgent.

A better understanding of population genetic structure could help to manage aphid populations by providing more reliable estimates of population dynamics and the risk of resistance genes arising. The population genetic structure of an organism is determined by various factors, such as geographical barriers, ecological difference, and historical processes, as well as the dispersal ability of species¹³. In recent years, many molecular markers have been used to infer the phylogeny and biogeography of species to understand their patterns of evolution¹⁴. Owing to their locus-specific codominant, high abundance and high rates of transferability across species, microsatellite markers have been widely used in population genetics of aphid species^{15–17}. Of course, cytochrome oxidase subunit I (*COI*) and cytochrome *b* (*Cytb*) genes also have a clear evolutionary pattern and evolve at a relatively moderate rate, which makes it suitable for reconstructing species and population-level phylogenies^{18, 19}. Little information on the genetic variation and population genetic structure of *A. gossypii* has been available before the development of methods for identifying polymorphisms of microsatellites, which are now used for analyzing the population genetic structure of many organisms²⁰. Previous studies have shown that genetic diversity of *A. gossypii* is structured by host plants, and its genetic diversity and host specialization are affected by natural selection pressure^{21–23}. Distinct genetic differentiation was found in *A. gossypii* populations from its winter hosts and summer hosts^{4, 6, 24, 25}. In North China, this species has two host-specialized biotypes (cotton and cucumber)²⁶. By contrast, many studies have documented that genetic differentiation between localities was not significant, whereas genetic differentiation was high between host plant families. Low genetic diversity and a lack of significant differentiation have also been found in France^{27, 28}. However, the studies had smaller sample sizes or a limited number of populations, which were insufficient for clearly assessing the genetic diversity and population structure of this species in different climatic zones^{26–28}.

Here, we examined the population genetics, including genetic variation, population structure and demographic history of 33 populations of *A. gossypii* in China (Fig. 1) using ten polymorphic nuclear microsatellite loci and two mitochondrial DNA sequence (*COI* and *Cytb*). We then discuss management strategies for this species based in the results, which also provide a theoretical framework for developing appropriate regional management strategies against this insect pest.

Results

Genetic diversity. In the present study, a total of 782 individuals from 33 populations were genotyped using ten microsatellite loci. Fisher's exact tests showed that 111 of the 330 locus–population combinations deviated significantly from Hardy–Weinberg equilibrium (HWE). Significant linkage disequilibrium was present in 50 of 1485 tests ($\alpha = 0.05$), and the studied loci were not linked with each other and were used in the subsequent analysis. The null allele frequency per locus was mostly less than 0.1 (Supplementary Table S1). The average F_{ST} not using the excluding null alleles (ENA) correction method was 0.224 and 0.213 using ENA per locus (Supplementary Table S2), and these F_{ST} estimations did not differ significantly ($P = 0.4813$). Therefore, the presence of null alleles had little influence on the estimation of F_{ST} . The basic summary statistics of genetic variation

Molecular marker	Source of variation	Sum of squares	Variance components	Percentage variation (%)	F-statistics
Microsatellite	Global analysis				
	Among populations	370.35	0.22839 Va	23.26	$F_{IS} = 0.086^{***}$
	Among individuals within populations	611.27	0.06492 Vb	6.61	$F_{ST} = 0.233^{***}$
	Within individuals	537.00	0.68846 Vc	70.12	$F_{IT} = 0.299^{***}$
	Total	1518.612	0.98177		
	Hierarchical AMOVA (K = 2)				
	Among groups	125.237	0.16467 Va	16.14	$F_{CT} = 0.308^{***}$
	Among populations within groups	239.828	0.15011 Vb	14.71	$F_{SC} = 0.175^{***}$
	Within populations	1077.689	0.70576 Vc	69.16	$F_{IS} = 0.161$
Combined COI and Cytb	Global analysis				
	Among populations	119.253	0.15742 Va	41.62	$F_{ST} = 0.416^{***}$
	Within populations	155.666	0.22080 Vb	58.38	
	Total	274.919	0.37822		
	Hierarchical AMOVA (K = 2)				
	Among groups	3.915	-0.00128 Va	-0.34	$F_{SC} = 0.417^{***}$
	Among populations within groups	115.338	0.15802 Vb	41.86	$F_{ST} = 0.415^{***}$
	Within populations	155.666	0.22080 Vc	58.48	$F_{CT} = -0.003^{***}$

Table 1. Results of analysis of molecular variance (AMOVA) test on microsatellite and mtDNA markers in different populations of *Aphis gossypii* in China. $***P < 0.001$. Two groups including eastern region group and western region group based on STRUCTURE analysis (See Fig. 4).

among different geographic populations of *A. gossypii* in China based on the 10 microsatellite loci are presented in Supplementary Table S3. Overall, a low level of genetic diversity for the total populations in these study regions was obtained. The observed number of alleles (N_a) across microsatellite loci ranged from 3.900 in Tacheng (TAC) to 8.400 in Baicheng (BC) and Changchun (CC), and had a mean of 6.103. The effective number of alleles (N_e) values across microsatellite loci ranged from 2.013 in Heze (HZ) to 4.129 in Taian (TA) and had a mean of 3.223. The mean observed heterozygosity ($H_o = 0.512$) was similar to the mean expected heterozygosity ($H_e = 0.574$) in the total populations. The observed heterozygosity ranged from 0.291 in Liaoyang (LY) to 0.697 in Jinan (JN), whereas the expected heterozygosity ranged from 0.388 in HZ to 0.681 in TC. The unbiased expected heterozygosity (uH_E) ranged from 0.395 in HZ to 0.698 in TC. Shannon's information index (I) values across loci ranged from 0.748 in HZ to 1.537 in BC. Allelic richness (A_R) across loci was 3.735, ranging from 2.455 (HZ) to 4.615 (BC).

A region of 579 bases in the mtDNA *COI* genes from 782 individuals (deposited in GenBank under accession nos. KY069907–KY069921) and 629 bases of *Cytb* gene from 738 individuals (accession nos. KY069922–KY069936) were obtained. For subsequent analyses, we combined *COI* and *Cytb* and treated them as a single locus because of the lack of recombination in the mtDNA. Of the 1208 total characters, 1172 were conserved, and 36, that included 22 singleton polymorphic sites and 14 parsimonious informative sites, were variable. The distribution of haplotypes and genetic diversity among different populations of *Aphis gossypii* based on the combined *COI* and *Cytb* sequences was shown in Fig. 2a and Supplementary Table S4. Twenty-nine haplotypes were observed in 738 individuals. The average number of haplotypes in each population was 2.85 ± 1.39 , ranging from 1 to 5. Weifang (WF), Changdao (CD), Jiujiang (JJ) and Nanchang (NC) had the highest number of haplotypes (5). In addition, H10 was the most common haplotype and shared by 551 samples and present in 32 of the 33 populations (except LY population). Thus, H10 is most likely to be the ancestral haplotype. The second most frequent haplotype (10.84%) was H3, which was shared by 80 individuals and present in 22 populations. Overall, the average Hd was 0.420, and the average Pi was 0.00062. The highest genetic diversity (Hd = 0.667, Pi = 0.00066) was detected in Taian (TA). However, in 5 localities among 3 provinces, including Baicheng (BC), Liaoyang (LY), Kuitun (KT), Tacheng (TAC) and Hutubi (HTB) genetic diversity was zero because no variable sites were found from these locations.

Origin and distribution of mtDNA haplotypes. Phylogenetic analyses were conducted to determine the relationships among 29 *A. gossypii* haplotypes of the combined *COI* and *Cytb* and to detect any groups that can be explained by their geographic distribution. ModelTest indicated that the best substitution model was HKY + G. High congruence was observed between the phylogenies derived from maximum parsimony (MP) and maximum likelihood (ML) analysis, and these analyses generated only one inclusive group, but simultaneously provided many unresolved branches (Supplementary Fig. S1). These 29 mtDNA haplotypes were from 33 locations, which covered more than 4,000 km. These results reflected the random distribution of mtDNA haplotypes in *A. gossypii*, which may indicate a high level of gene flow in this species. Furthermore, the median-joining network of haplotypes showed that there was no apparent geographical clustering of mtDNA haplotypes (Fig. 2b). The haplotype network of combined genes obviously displayed a star-like pattern with the most common haplotype (H10) in the

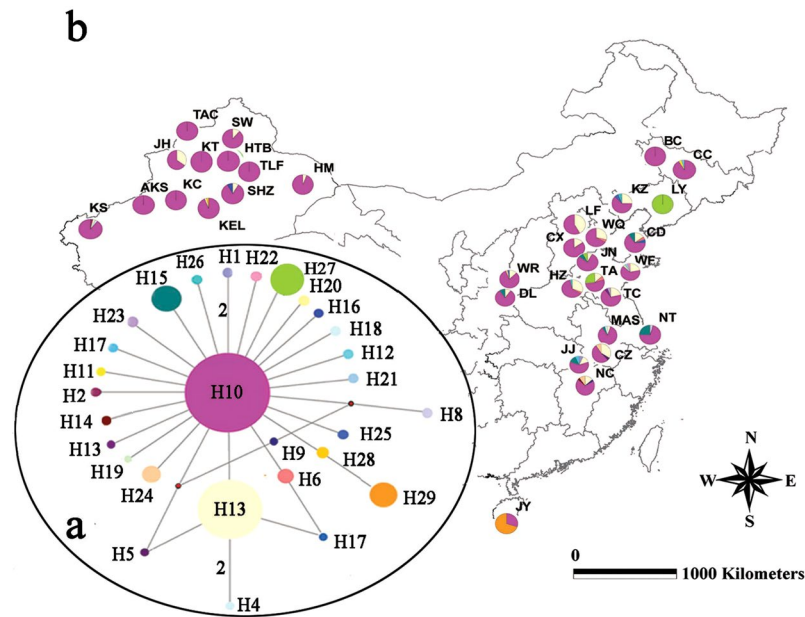


Figure 2. MtDNA haplotype network and haplotype frequencies in the examined populations of *Aphis gossypii*. (a) Haplotype network based on combined *COI* and *Cytb* sequences. Each haplotype was represented by a circle and identified by the codes H1–H29. The sizes of circles are proportional to the haplotype frequencies. Small red circles represent internal (unsampled) nodes. The median-joining network of the haplotypes of the combined *COI* and *Cytb* were generated by Network 2.0. (b) Haplotype frequencies and distribution in 33 populations. The color of each haplotype was the same as the network. For population abbreviations, see Table 2. ArcView GIS (version 3.2, <http://www.resources.esri.com>) was used to produce a distribution map based on the geographical coordinates of the localities. The base map (Greater China 1: 4,000,000) for the analysis was obtained the URL: <http://www.diva-gis.org/gdata>.

star's center. H10 can be observed in all 6 cotton region of China (in 32 populations) and mainly distributed in Northwest China (Xinjiang Province), which raises the possibility that the maternal ancestor of China population was from Xinjiang. For most cases, only a 1-step mutation was found between the most common haplotype and the other haplotypes, which was often associated with demographic expansion.

Population genetic differentiation. There was a moderate level of genetic differentiation among populations. Based on the microsatellite genotype data, pairwise F_{ST} values between the studied populations ranged from -0.04 to 0.58 , and exact tests showed that 391 of the 561 pairwise populations differed significantly (Supplementary Table S5). For combined genes of *COI* and *Cytb*, pairwise F_{ST} values (ranging from -0.09 to 1.00) showed statistically significant genetic differentiation for 496 pairwise comparisons ($P < 0.05$) (Supplementary Table S6).

Population genetic structure. Five methods were used to investigate the population structure and infer the relationships among *A. gossypii* populations.

POPTREE analysis based on SSR. The unrooted tree defined two major clades among 33 populations (Fig. 3). Populations from Xinjiang Province, Northwest China, including KT, TAC, TLF, JH, HTB, SW, KS, SHZ, KC, AKS, HM and KEL grouped together (bootstraps = 83). The other 21 populations from eastern China can be treated as one group because of the absence of a robust phylogeographic structure, which suggested frequent gene flow in the vast eastern region.

Bayesian clustering. We used an admixture model implemented in STRUCTURE 2.3.3 software to explore different numbers of populations K to the population structure based on microsatellite data. STRUCTURE analyses produced stable solutions for K between 1 and 10, and the ΔK method determined that the best K was 2 according to the plateau criterion (Fig. 4), which was consistent with the hypothesis that these populations could be divided into two groups: eastern region group and western region group (Figs 1 and 4). Further structuring indicated the divergence of CC, BC, LY, KZ, LF, TC, and CD from the other eastern populations at $K = 3$ (Supplementary Fig. S2). A more fine-grained population structure was present at $K = 4$, and the subdivision still closely followed the geography. The Bayesian clustering method detected significant genetic clusters among the Northeast populations, with locations CC, BC, LY, and KZ in one cluster; LF, TC, CD, and DL in a second cluster, and the remaining populations, except Xinjiang, in a third cluster; the western populations (all from Xinjiang) grouped in a fourth cluster at $K = 4$ (Supplementary Fig. S2).

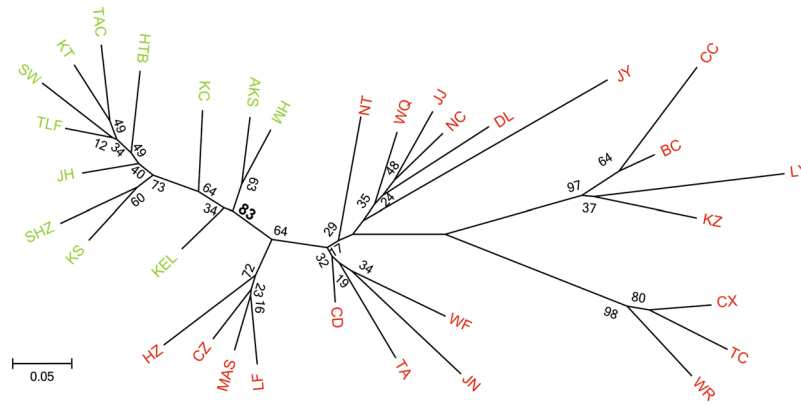


Figure 3. Unrooted NJ tree based on SSR data from 33 *Aphis gossypii* populations collected in China. Numbers beside the nodes are bootstrap values. Populations from Xinjiang Province, Northwest China with relatively stronger support, are marked as green, and the other populations, from eastern China, are marked as red.

BayesAss analysis based on SSR data indicated that very limited and asymmetric gene flow existed between the eastern and the western region groups. Gene flow from west to east (0.0055) was greater than in the opposite direction (0.0016). Limited gene flow may be associated with moderate migration capability of the cotton aphid, habitat differences and lack of commercial trade between the east and the more isolated west, although potential dispersion routes (gene flow) from west to east are presented.

Principal coordinate analysis (PCoA). The PCoA result (Fig. 5) showed a western–eastern pattern of genetic structure, which was similar with the POPTREE (Fig. 3) and STRUCTURE results (Fig. 4). The first and second axes explained 29.83% and 15.09% of the overall variance in microsatellite data, respectively (Fig. 5). Bayesian clustering based on data collected 782 individuals showed two distinct groups and supported the efficacy of the PCoA approach. We noticed that most of the geographically close populations resembled each other.

Analysis of molecular variance (AMOVA). Results of the AMOVA test on microsatellite and mtDNA markers in different populations and regional groups of *A. gossypii* are shown in Table 1. The global AMOVA of the data for the two molecular markers revealed that 23.16% (microsatellite) and 16.14% (mtDNA) genetic variation could be explained by the variation among populations, whereas the remaining came from variation within populations. Significant variability was found among microsatellites between the eastern and western groups (16.14%, $P < 0.001$) as inferred from above analyses. In contrast to SSR, mtDNA did not support the western–eastern pattern (−0.34%). Considering the limited variability (information) of the mtDNA here, which may due to the slower evolutionary rate of mtDNA and the especially rapid expansion of China cotton aphid populations (see the demographic results that follow), the AMOVA result from the SSR data showing moderate variability are more reasonable and can represent more recent events.

Mantel test for isolation by distance (IBD). A Mantel test for IBD was performed to assess the correlation between the genetic distance matrix and the corresponding geographic distance matrix of *A. gossypii*. If the dispersal of *A. gossypii* is limited by distance, genetic and geographical distances should be positively correlated, producing a pattern of isolation by distance. Applying the Mantel test, significant positive relationships between genetic and geographical distances were found over the 33 populations of China cotton aphids based on both kinds of markers (microsatellite genotypes: $Z = 349.537$, $r = 0.433$, $P = 0.000$, Fig. 6a; combined *COI* and *Cytb*: $Z = 239.218$, $r = 0.150$, $P = 0.013$, Fig. 6b). Given that the isolated western and the eastern regions are far from each other and lack of gene flow, to minimize the distance influence, we performed separate IBD analyses for the western and the eastern group. Interesting, an IBD pattern was detected in the eastern region group of 21 populations (microsatellite genotypes: $Z = 113.402$, $r = 0.315$, $P = 0.006$, Fig. 6c; the combined genes of *COI* and *Cytb*: $Z = 114.490$, $r = 0.357$, $P = 0.033$, Fig. 6d) rather than in the western region group of 12 populations (microsatellite genotypes: $Z = 10.609$, $r = -0.200$, $P = 0.799$, Fig. 6e; combined *COI* and *Cytb*: $Z = 5.443$, $r = -0.060$, $P = 0.560$, Fig. 6f). The results were consistent with the moderate migration capability of cotton aphids. For the western region, the geographical scope is not large, and the simple habitat also makes migration feasible. Compared with the mtDNA data, the SSR data was more in line with the IBD pattern, perhaps as a result of the limited variability (information) of the mtDNA here, which may again be due to the slower evolution rate of mtDNA and especially rapid expansion of China cotton aphid populations.

Demographic history analysis. The cotton aphid populations in China appear to have expanded demographically. Widespread single haplotype (Hap10) and classic “Star” shape of mtDNA haplotype network provide preliminary but strong evidence of expansion (Fig. 2). In addition, Tajima’s D^{29} and Fu’s F_s^{30} were calculated from the total number of segregating sites and used to assess the evidence for population expansion. For most populations in China, Tajima’s D value was significantly negative (Supplementary Table S4). When the 33 populations

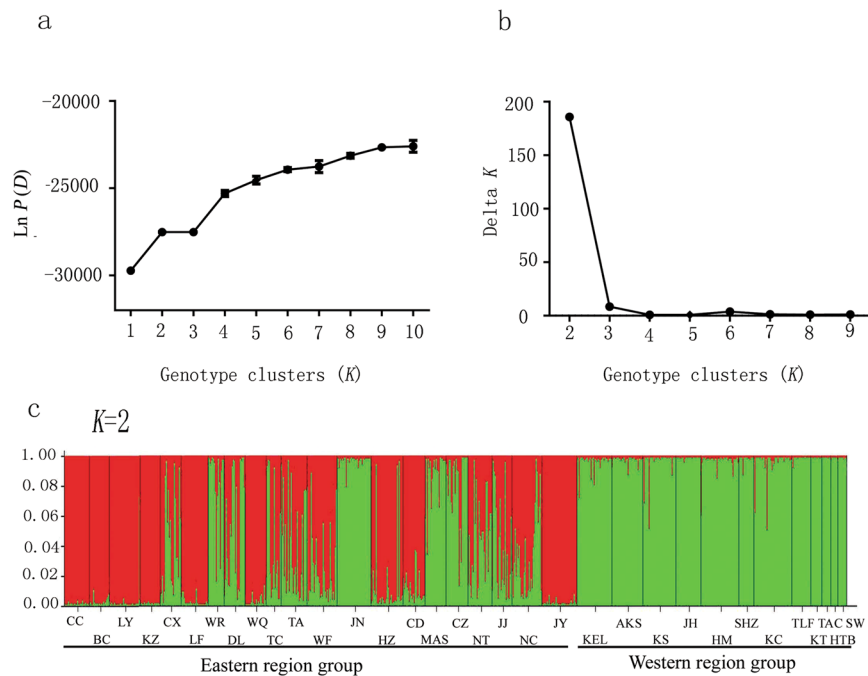


Figure 4. Bayesian inference to determine suitable cluster (K) and estimated cluster proportion using STRUCTURE for *Aphis gossypii*. The likelihood of the data given $\ln P(D)$ (a) and ΔK (b) are plotted against the number of genetic clusters (K). Error bars represent standard deviations over 10 runs. For the assignment proportion in all populations, each individual is represented by a thin vertical line, which was partitioned into K segments that represent its estimated population group membership fractions (c). Population codes are given in Table 2.

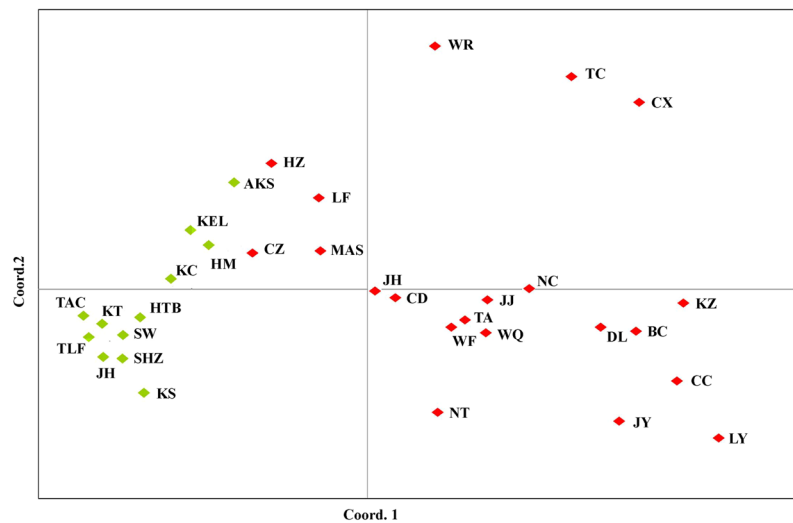


Figure 5. Principal coordinate analysis (PCoA) based on the genetic distance matrix of F_{ST} values for microsatellite data. Population codes are given in Table 2. Colors within the diamond: red, eastern region group; green, western region group.

were considered as one or two groups, significantly negative values for combined *COI* and *Cytb* were obtained (total population: Tajima's $D = -2.218$, $P < 0.05$, Fu's $F_s = -2.805$; eastern region group: Tajima's $D = -2.014$, $P < 0.05$, Fu's $F_s = -18,678$; western region group: Tajima's $D = -1.764$, $P < 0.05$, Fu's $F_s = -5.964$). Furthermore, the unimodal mismatch distribution of all populations and the eastern region group and western region group indicated a sudden demographic expansion (Supplementary Fig. S3). The bottleneck analysis based on the microsatellite data, no significant heterozygote excess in 31 of the 33 populations (exceptions: HTB and SW) under the infinite allele model (IAM), two-phase model (TPM), and the strict stepwise mutation model (SMM). Meanwhile,

Geographic region	Site code	Site/Climatic zones	Latitude/Longitude	Collection date	Altitude (m asl)	Group
Northeast China	CC	1 Changchun, Jilin/MTZ	43.81 °N/125.41 °E	12 June 2012	219	E
	BC	2 Baicheng, Jilin/WTZ	44.82 °N/123.08 °E	12 July 2012	148	E
	LY	3 Liaoyang, Liaoning/WTZ	41.27 °N/123.23 °E	9 June 2012	27	E
	KZ	4 Kazuo, Liaoning/WTZ	41.29 °N/119.92 °E	10 July 2012	249	E
North China	CX	5 Cangxian, Hebei/WTZ	38.32 °N/116.60 °E	2 August 2012	35	E
	LF	6 Langfang, Hebei/WTZ	39.51 °N/116.58 °E	23 August 2012	39	E
	WR	7 Wanrong, Shanxi/WTZ	35.42 °N/110.56 °E	19 June 2012	491	E
	DL	8 Dali, Shanxi/WTZ	34.81 °N/110.11 °E	25 July 2013	334	E
	WQ	9 Wuqing, Tianjin/WTZ	39.08 °N/117.19 °E	31 August 2012	10	E
East China	TC	10 Tancheng, Shandong/WTZ	34.61 °N/118.37 °E	8 August 2012	148	E
	TA	11 Taian, Shandong/WTZ	36.18 °N/117.08 °E	10 August 2013	167	E
	WF	12 Weifang, Shandong/WTZ	36.38 °N/119.75 °E	10 August 2012	143	E
	JN	13 Jinan, Shandong/WTZ	36.68 °N/117.02 °E	11 August 2012	148	E
	HZ	14 Heze, Shandong/WTZ	35.26 °N/115.46 °E	12 August 2012	54	E
	CD	15 Changdao, Shandong/WTZ	38.40 °N/120.92 °E	27 August 2014	84	E
	MAS	16 Maanshan, Anhui/WTZ	31.70 °N/118.11 °E	21 August 2013	19	E
	CZ	17 Cizhou, Anhui/WTZ	30.66 °N/117.49 °E	22 August 2012	15	E
	NT	18 Nantong, Jiangsu/SZ	31.95 °N/121.31 °E	2 August 2012	2	E
Central China	JJ	19 Jiujiang, Jiangxi/SZ	29.71 °N/115.99 °E	10 July 2012	41	E
	NC	20 Nancang, Jiangxi/SZ	28.68 °N/115.89 °E	13 July 2012	16	E
South China	JY	21 Jiyang, Hainan/TZ	18.28 °N/109.58 °E	11 December 2015	15	E
Northwest China	KEL	22 Korla, Xinjiang/WTZ	41.73 °N/86.17 °E	July 2015	934	W
	AKS	23 Aksu, Xinjiang/WTZ	41.17 °N/80.26 °E	July 2015	1107	W
	KS	24 Kashi, Xinjiang/WTZ	39.47 °N/75.99 °E	July 2015	1298	W
	JH	25 Jinghe, Xinjiang/WTZ	44.60 °N/82.89 °E	July 2015	320	W
	HM	26 Hami, Xinjiang/WTZ	42.82 °N/93.51 °E	July 2015	759	W
	SHZ	27 Shihezi, Xinjiang/WTZ	44.31 °N/86.08 °E	July 2015	482	W
	KC	28 Kuche, Xinjiang/WTZ	41.72 °N/82.96 °E	July 2015	1071	W
	TLF	29 Tulufan, Xinjiang/WTZ	43.83 °N/87.61 °E	July 2015	900	W
	KT	30 Kuitun, Xinjiang/MTZ	44.42 °N/84.90 °E	July 2015	481	W
	TAC	31 Tacheng, Xinjiang/WTZ	46.75 °N/82.98 °E	July 2015	548	W
	HTB	32 Hutubi, Xinjiang/WTZ	44.01 °N/87.31 °E	July 2015	520	W
	SW	33 Shawan, Xinjiang/WTZ	44.87 °N/87.20 °E	July 2015	528	W

Table 2. Habitat location, sample size and regional group information of *Aphis gossypii* in different geographic populations. Sites are labeled according to abbreviated location names for *Aphis gossypii*. MTZ, mid-temperature zone; WTZ, warm temperate zone; TZ, tropical zone; and SZ, subtropical zone; E, eastern region group; W, western region group.

these results were consistent with a normal L-shaped distribution of allelic frequency, indicating that *A. gossypii* had expanded demographically without a severe bottleneck in most regions of China (Supplementary Table S7).

It is more likely that the mitochondrial loci reflect the patterns of resistance of other genes and are thus a by-product of selection. When we scanned for signals of selection, the Z-tests of selection rejected the null hypothesis of strict-neutrality ($dN = dS$) in favour of purifying selection ($dN < dS$; $P = 0.002$) rather than for positive selection ($dN > dS$; $P = 1.000$).

Discussion

Genetic diversity. Deviations from HWE were observed at multiple loci in multiple populations. Fisher's exact tests showed that 111 of the 330 locus/population combinations deviated significantly from HWE. The specific reproductive mode of parthenogenesis might also result in deviation from HWE among the studied populations. In general, aphid population genetic studies have more often documented populations not in HWE because of heterozygote deficits (i.e., homozygote excess)^{15,31}, presumably because of clonal selection. Permanent parthenogenetic clones have high heterozygosity as a consequence of accumulated mutations^{31–34}. The F_{IS} values of the 10 microsatellite loci evaluated in this study were a little less than zero, which indicated a deficit of heterozygotes (Supplementary Table S3), which in turn reflected a balance in the asexual ratio of the population. Overall, the average expected heterozygosity (H_e) was similar to the observed (H_o). Note that a low H_o was obtained in some populations such as Liaoyang ($H_o = 0.291$), which suggested a high genetic identity within the population. These results could be attributed to the specificity of the geography and climate in Liaoning in northeastern China because aphids usually reproduce sexually at relatively high latitudes and low temperatures.

Genetic variability is fundamental for adaptive processes in a species. The genetic diversity found in *A. gossypii* through the use of molecular markers has shown that the clonal diversity is structured by its host plant^{34,35}.

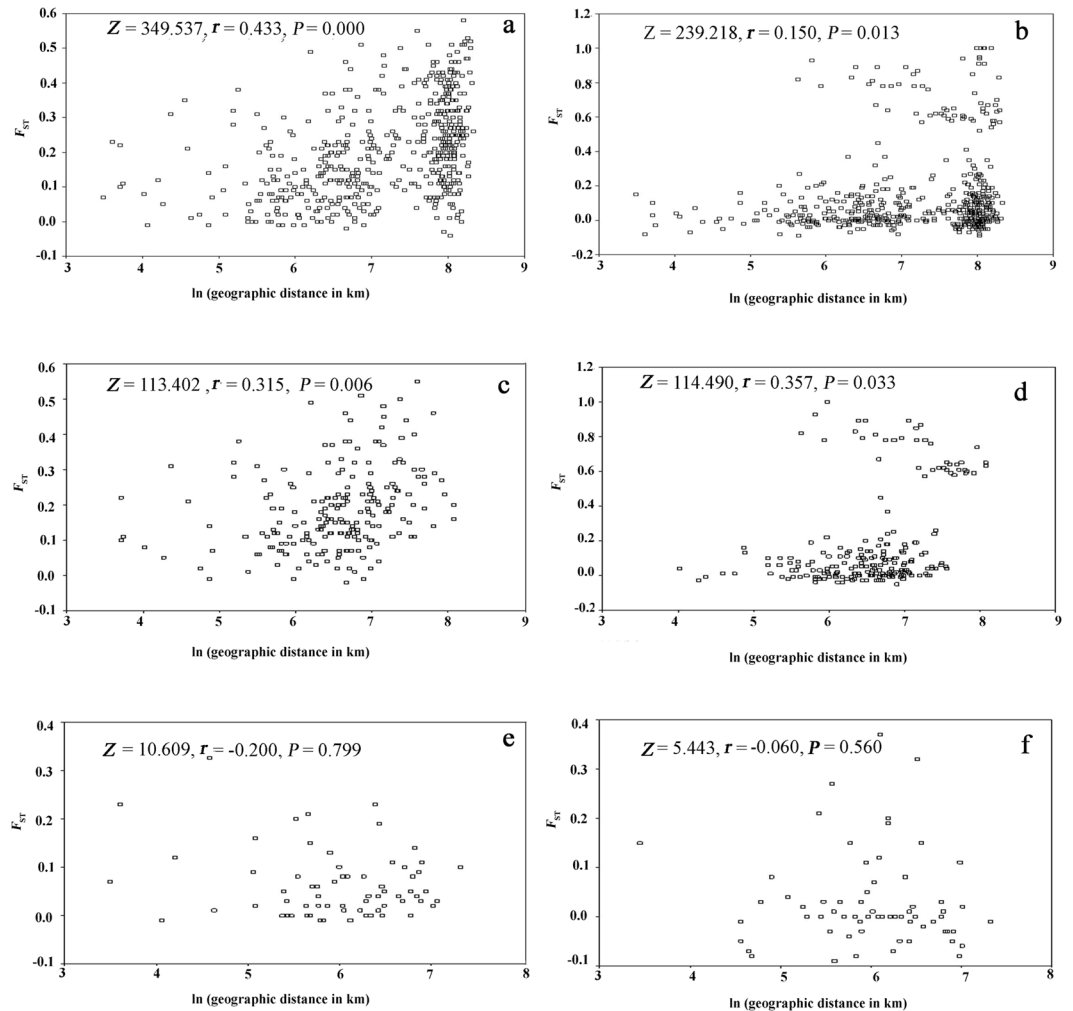


Figure 6. Correlation analysis between pairwise linearized F_{ST} values and the logarithm of geographic distance in Chinese populations of *Aphis gossypii* based on microsatellites ((a) total population; (b) eastern region group; (c) western region group) and combined *COI* and *Cytb* ((d) total population; (e) eastern region group; (f) western region group).

In the present study, *A. gossypii* had low to moderate genetic diversity based on microsatellite (Supplementary Table S3) and mitochondrial DNA data (Supplementary Table S4), and H_e varied between 0.388 (in HZ) and 0.681 (in TC). Similar cases of moderate genetic diversity have been documented in this species; H_e ranged from 0.409 to 0.873 on the basis of eight microsatellite loci¹⁷. Within a glasshouse population of *A. gossypii*, clonal diversity declined significantly as the spring/summer season progressed²². Low levels of genetic diversity were also found previously for other aphid species such as *Eriosoma lanigerum*³⁶ and *Myzus persicae*³⁷, which resulted from the adaptation of the aphids to two heavy selection pressures, the distribution of host plants and the use of insecticides^{22, 27}. In general, a critical feature of aphid biology is their viviparous reproduction by cyclical parthenogenesis, which can result in numerous overlapping generations in a short period and an extremely rapid increase in population size, which could explain the low level of genetic variability detected in aphids as compared with other insects⁷. Additionally, aphids are also very sensitive to selection. *A. gossypii* is a particularly polyphagous species; it has been described on almost 300 host plants from various botanical families³. Because the eastern populations had more genetic diversity than in the western populations (Supplementary Table S4; eastern: $H_d = 0.549$; $P_i = 0.00086$; western: $H_d = 0.141$; $P_i = 0.00014$), fragmented habitats, migration between populations and genetic drift probably contributed to the loss of diversity in the western populations.

Single origin and genetic structure. Compared with conventional phylogenetic trees, haplotype networks based on the combined *COI* and *Cytb* provide an enhanced view of the relationships among haplotypes and are preferable for intraspecific analyses because ancestral haplotypes may still be extant³⁸. H10 was the most common haplotype of mtDNA and shared by 551 samples and present in 32 of the 33 populations (Fig. 2). Combined with the classic “star” shape of the network, the results indicated that these China cotton aphid populations arose from one or only a few maternal ancestors in Xinjiang and then quickly spread geographically. The mtDNA tree and network did not reveal any distinct geographic distribution pattern for this pest (Fig. 2), which may be due to

the relatively slow evolutionary rate of the aphid and the “maternal inheritance” pattern of this marker. In contrast, microsatellite DNA shows more variation information than mitochondrial gene, which may reflect more recent events, including population divergence, gene flow and bottleneck. The China cotton aphids separated into two genetic groups as supported by the PCoA, POPTREE, STRUCTURE analysis (Figs 3, 4 and 5) and AMOVA result (Table 1). Beside the fact that the ancestor haplotype (H10) of mtDNA from Xinjiang was widely distributed in eastern populations, we also detected limited and asymmetric nuclear gene flow from the western to eastern region (0.0055), indicated that the eastern populations migrated from the western region over time. Geographic barriers and climatic factors might contribute to the isolation of the studied population and play an important role in preventing gene flow among groups^{39, 40}, for example, mountains shaped the population structure of the terrestrial species *Locusta migratoria*⁴¹ and *Paeonia rockii*⁴². Therefore, we speculate that the differing climates and large mountain systems (Qilianshan, Qinling, and Taihang Mountains) in the western and eastern region might have prevented or slowed nuclear gene flow between the eastern and western populations, resulting in lineage differentiation. Furthermore, for the JY population in south China (Jiyang, Hainan province), there is one mtDNA haplotype (H29) which was more divergent from H10 (Fig. 2). H29 were found in 23 of the 33 individuals in JY population. Taking into account the larger geographical distance between JY and other populations, distance may also play a role in population differentiation of China cotton aphid.

Moderate flight and dispersal ability of the aphids might allow short-distance movement. Usually, IBD effects are pronounced in moderately mobile species but weak in both low- and high-mobility species, since extensive gene flow homogenizes populations across large geographic areas⁴³. Therefore, we have reason to assume that there is a positive correlation between genetic and geographic distances among the populations. However, an IBD pattern was detected in the eastern region, not in the western. For the western region, geographical scope is not large, and the simple habitat and climate also make migration and gene flow feasible. For the vast eastern region, air-flow direction and monsoons can affect long-distance migration behavior though the difference in climate from north to south is great. Human-mediated dispersal of this pest might be an important mode of spread along the eastern coastline of China, perhaps through the transport of seedlings. Pairwise nuclear F_{ST} between populations of JY (Hainan Province) and Northeast China (CC, BC, LY, KZ) was not large though the two sites are separated by more than 3000 km (Fig. 1). Considering the insect's moderate migration ability, commercial transport may be responsible. For this reason, lack of commercial trades as well as geographic barriers between the east and the more isolated west likely decreased gene exchange between the two regions. That is also an important reason for our defining all the populations in the eastern region as one genetic group. Generally speaking, IBD still played a major role in genetic differentiation in the eastern region though the factors already mentioned also have an effect.

Pest management implications and future work. Increasing our understanding of the population dynamics of *A. gossypii* may improve predictions of its outbreaks and enhance management efforts. The present work provided robust evidence of population expansion. The widespread occurrence of single mtDNA haplotype (H10), which was detected in all 33 sampled populations, is the strongest evidence of expansion. The large negative values of Tajima's D and Fu's F_s statistic and the unimodal mismatch distribution are also evidence of population expansion in most populations of cotton aphids in China. SSR data indicated that these cotton aphids experienced a recent demographic expansion without a severe bottleneck in most regions in China. As noted already, the results will provide essential information for understanding possible local adaptation and dispersal patterns and for further clarifying the relationship between genetic variation and outbreaks of *A. gossypii*. In the future, we will focus on monitoring the population dynamics of this species and understanding its seasonal occurrence in the main crop-producing areas of China to provide more useful data for forecasting outbreaks and managing this species.

Over the past several decades, pest management programs, such as cultural controls, biological controls, and the use of insecticides have been used in China and have partially prevented the damage and spread of *A. gossypii*. Unfortunately, *A. gossypii* has developed resistance to insecticides such as carbamate and organophosphates^{9, 44–46}, and the exponential growth associated with parthenogenesis in this aphid species favours rapid selection for insecticide resistance mechanisms as a consequence of intense insecticide use^{9, 11, 12}. It is more likely that the mitochondrial loci reflect the patterns of resistance of other genes and are thus a byproduct of natural selection. However, we did not detect any signal for positive selection on *COI* or *Cytb*. Positive selection may act on detoxification enzymes and metabolic genes like P450 family. High gene flow among the cotton aphid populations may explain the rapid spread of insecticide-resistance genes, and many resistance characteristics also accumulate through long-distance migration. Therefore, to survey for populations of *A. gossypii* with insecticide resistance over large temporal and spatial scales and investigate the distribution of resistance genes relative to the genetic structure and level of gene flow among populations of *A. gossypii* in China, quantifying the level of resistance to insecticides will be important for devising application strategies for chemical pesticides. In addition to geography and climate, host plant specialization and insecticide treatment can also affect the distribution of genetic variation in aphid populations²⁷.

The study of the temporal dynamics of genotypic diversity of *A. gossypii* populations in the cotton-growing area in Xinjiang over several consecutive years would help to reveal the relationship between the low genetic diversity and the adaptation of the aphids to two heavy selection pressures, the distribution of host plants and the use of insecticides. Additionally, we also found evidence of differentiation between the Xiuyan population, which was collected from *Rhamnus koraiensis*, and other populations collected from cotton (Wang *et al.* unpublished). Obviously, a broad-scale sampling strategy on different host plant families is needed. Therefore, future work should concentrate on the genetic differentiation between winter hosts and summer hosts through extensive sampling in China. Simultaneously, we will also focus on the relationship of the incidence of aphid populations on cotton, natural selection pressure (e.g., host rotation, host resources and pesticides), and genetic structure. Our

genetic data suggest significant genetic differentiation among different strains of *Wolbachia* symbiosis, which is present between the eastern region and western region (Wang *et al.* unpublished). Therefore, future experiments should shed light on the influence of the bacterium on the diversity and genetic structure of cotton aphid populations. Developing additional microsatellite markers and more extensive sampling, including samples at a finer distance scale, should improve the resolution of the understood population genetic structure and our understanding of intra- and inter-race diversity of this species.

Methods

Sample collection and DNA extraction. All individuals of *A. gossypii* were collected from 12 provinces at 33 geographic locations in China, covering four climatic regions—mid-temperature zone (MTZ), warm temperate zone (WTZ), tropical zone (TZ), and subtropical zone (SZ)—from June 2012 to December 2015 (Table 2; Fig. 1). Samples in Changdao were collected with a suction trap (Keyun ST-1B), which was developed by the Institute of Zoology, Chinese Academy of Sciences. For preventing over-representation of siblings from each locality, each aphid was collected at least 1 m from the next sample. All the samples were preserved and stored at -20°C in 95% ethanol and kept at the Plant Protection Institute, Chinese Academy of Agriculture Sciences (Beijing). Total DNA was extracted separately from each *A. gossypii* using a DNeasy extraction kit (Qiagen, Germantown, MD, USA) according to the manufacturer's protocols *Aphis glycines* (GenBank accession nos. JQ860254 and GU205350) and *Aphis craccivora* (GenBank accession nos. AB506714 and AM085376) were used as outgroup species in the phylogenetic analysis.

Mitochondrial DNA amplification and sequencing. We sequenced two partial regions of the mitochondrial genes *COI* and *Cytb*. To amplify *COI*, we used published primers CIS (5'-ACCAGTTTGTAGCAGGTGCTATTAC-3') and CIA (5'-GTATATCGACGAGGTATACCATT-3')⁴⁷ and for *Cytb*, CP1 (5'-GATGATGAAATTTGGATC-3') and CP2 (5'-CTAATGCAATAACTCCTCC-3')⁴⁸. Each PCR mixture contained 0.25 μL EasyTaq DNA polymerase (5 U/ μL), 2.5 μL 10 \times Easy Taq buffer, 0.5 μL dNTP mixture (2.5 mM of each), 0.5 μL of each primer (10 pM), 1 μL of DNA template and 19.75 μL of distilled water, making a final volume of 25 μL . The reactions were performed in a GeneAmp PCR System 9700 (Perkin Elmer Applied Biosystems, Foster City, CA, USA) under the following conditions: 94 $^{\circ}\text{C}$ for 5 min; 35 cycles at 94 $^{\circ}\text{C}$ for 1 min, 52 $^{\circ}\text{C}$ for *COI* or 48 $^{\circ}\text{C}$ for *Cytb* for 1 min, 72 $^{\circ}\text{C}$ for 1 min; and a final extension at 72 $^{\circ}\text{C}$ for 10 min. All specimens were sequenced on an Applied Biosystems ABI 3730 DNA sequencer (Applied Biosystems). Chromatograms, including sense and antisense, were edited and assembled using DNASTAR 5.0 (DNASTAR, Madison, WI, USA) to obtain a single consensus sequences.

Microsatellite amplification and genotyping. Ten microsatellite loci were used in this study using the methods of Michel *et al.*⁴⁹ for loci *ago59* and *ago89*; Kim *et al.*⁵⁰ for loci AgI2-6, AgI1-10, AgI1-11 and AgI1-2; and Gauffre & Coeur⁵¹ for loci AF-93, AF-45, AF-F and AF-beta, which were assigned unique fluorophores for fluorescent tagging of DNA in a PCR reaction. For these isolated microsatellite, each 20 μL reaction volume for the genotyping PCR contained 0.2 μL EasyTaq DNA polymerase (5 U/ μL), 2.0 μL 10 \times Easy Taq buffer, 0.2 μL dNTP mixture (2.5 mM of each), 0.4 μL of reverse primers (10 pM), 0.4 μL of forward primer, which were labeled with fluorochromes (HEX or FAM) (10 pM), 0.5 μL of DNA template and 16.3 μL of distilled water. For the thermal profile, an initial denaturation step at 94 $^{\circ}\text{C}$ for 4 min was followed by 30 cycles of denaturing at 94 $^{\circ}\text{C}$ for 30 s, annealing at 58 $^{\circ}\text{C}$ for 30 s, and extension at 72 $^{\circ}\text{C}$ for 30 s; and a final 5 min extension at 72 $^{\circ}\text{C}$. Following amplification, the products were visualized at Sangon Biotech Co., Ltd. (Shanghai, China) using an ABI 3730XL automated sequencer (Applied Biosystems). Microsatellite alleles were analyzed using GeneMapper 4.0 (Applied Biosystems).

Data analysis

Mitochondrial data. *Genetic variation and diversity.* The sequences were first aligned with CLUSTAL X 1.81⁵² using the multiple alignment default parameters and corrected by hand. Nucleotide composition, conserved sites, variable sites, parsimony informative sites were analyzed with MEGA 6⁵³. Molecular diversity indices such as haplotype diversity (Hd) and nucleotide diversity (Pi) were estimated using DnaSP 4.0⁵⁴.

Phylogenetic and network relationships among haplotypes. Maximum parsimony (MP), maximum-likelihood (ML) and Bayesian inference (BI) phylogenetic analyses were used to identify major clades and to evaluate the relationships between the haplotypes separately. MP analyses were performed in PAUP* 4.10b⁵⁵ using a heuristic search with 1000 random sequence repetitions and tree-bisection-reconnection (TBR) branch swapping. Consensus trees (50% majority rule) were obtained if more than one equally parsimonious tree was found. ModelTest 3.7⁵⁶ and the Akaike information criterion (AIC)⁵⁷ were used to identify the appropriate nucleotide substitution models, and the selected models of sequence evolution were used for the ML phylogeny reconstruction. The ML analysis was performed in PAUP* 4.10b using a heuristic search strategy with 10 random additions of sequences and TBR branch swapping. Bootstrap analysis was performed under the same model with 100 pseudoreplicates, 10 random addition sequences per replicate and TBR branch swapping. Bayesian analysis was also carried out using the model selected by Modeltest 3.7. The analysis used a random starting tree and proceeded for one million Markov chain Monte Carlo generations, and trees were sampled every 100 generations. The first 2,500 generations (25% of the total) were later discarded as burn-in; 50% majority-rule consensus trees were generated from the remaining trees and posterior probabilities computed. Compared with conventional phylogenetic trees, haplotype networks can enhance the relationships among haplotypes and are preferable for intraspecific

analyses. The median-joining network of the haplotypes of *COI* and *Cytb* were generated in Network 2.0 by a median-joining method⁵⁸.

Genetic differentiation and population genetic structure. Analysis of molecular variance (AMOVA) and calculation of population genetic differentiation (F_{ST}) statistics between the populations were conducted in Arlequin 3.0⁵⁹. Mantel tests were performed to determine the significance between genetic and geographic distances.

Neutrality test and demographic history. We performed codon-based Z-tests of selection to estimate the probabilities of neutral ($dN = dS$), positive ($dN > dS$) and purifying selection ($dN < dS$) using MEGA 6⁵³. Demographic history changes were analyzed for *A. gossypii* using two neutrality tests, Tajima's D ²⁹ and Fu's F_S ³⁰, which were calculated from the total number of segregating sites and used to assess evidence for population expansion. Mismatch distributions were calculated between the observed and expected mismatch distributions used as a test statistic. According to coalescent theory, a population usually exhibits a unimodal mismatch distribution following a recent population demographic or range expansion⁶⁰. To determine whether movement patterns were influenced by spatial scale, we conducted a spatial analysis of genetic structure by performing the isolation by distance (IBD) analysis for each site type using a Mantel test implemented in the web-based IBDWS 3.32⁶¹. An IBD analysis was used to determine the relationship between calculated pairwise F_{ST} and pairwise geographic distances (km) between collection sites.

Microsatellite data. *Gene variation and genetic diversity.* Genotypic linkage disequilibrium (LD) was tested among all pairs of loci across all populations using GenePop 3.4⁶² and exact probability tests. An exact test for HWE was conducted per locus and over all loci in each population using the same program. Corrections for multiple tests were performed by Bonferroni corrections. Null allele frequencies for each microsatellite locus were estimated following the expectation maximum algorithm of Dempster *et al.*⁶³ using FreeNA⁴¹. The excluding null alleles (ENA) correction method was used to provide accurate estimation of F_{ST} in presence of null alleles. The population genetic diversity indices such as mean number of alleles (N_a), effective number of alleles (N_e), Shannon's information index (I), observed heterozygosity (H_o), expected heterozygosity (H_e), and unbiased expected heterozygosity (uH_E) were assessed using GenAlEx 6.41⁶⁴. Allelic richness (A_R), fixation index (F_{ST}), inbreeding coefficient (F_{IS}) among these regions were calculated in FSTAT 2.9.3⁶⁵. The levels of genetic differentiation between pairs of populations were also estimated in Arlequin 3.0⁵⁹.

Population genetic structure. Population structure was estimated using principal coordinate analysis (PCoA) implemented in GenAlEx 6.41⁶⁴ and by using the Bayesian model-based clustering method implemented in STRUCTURE 2.2.3⁶⁶. We also used POPTREE 2⁶⁷ to construct a unrooted tree using the neighbour-joining method. PCoA was based on the covariance of the genetic distance matrix. STRUCTURE sorts individuals into K optimal region groups, according to their genetic similarity. The allele frequencies of different populations were assumed to be correlated due to historical migration and common ancestry⁶⁸. In addition, the admixture model of individual ancestry was used to assign hybrid individuals into population clusters; no prior population information was used⁶⁹. Ten independent runs for K values ranging from 1 to 10 were performed with a burn-in length of 30,000 followed by 1,000,000 Markov chain Monte Carlo replicates. To identify the optimal number of groups (K), measures of change in the likelihood function (ΔK) and F_{ST} (ΔF_{ST}) were calculated using the package CORRSIEVE^{70,71}, and the plateau criterion was applied⁵⁷. The software CLUMPP 1.1⁷² was used for model averaging of individual ancestry coefficients across the 10 independent runs. Then clusters were visualized using DISTRUCT 1.1⁷³. The consistency of the clusters identified through the STRUCTURE approach was tested by a dissimilarity analysis, which was calculated from allelic data. The significance of the hierarchical partitioning of genetic structure among the geographic groups was examined using an analysis of molecular variance (AMOVA) in Arlequin 3.0⁵⁹. Three hierarchical levels were defined: (1) between regional groups (western and eastern region), (2) between populations within regional groups, (3) between individuals within populations. BayesAss 3⁷⁴ was used to estimate recent migration between regional groups. We ran the MCMC for 20,000,000 iterations, discarding the first 2,000,000 iterations and sampling per 200 iterations. We then adjusted the mixing parameters to ensure an acceptance rate of about 30% and carried out three independent runs with different random number seeds. Additionally, we conducted an IBD analysis between microsatellite and geographic distance data using a Mantel test implemented in IBDWS 3.23⁶¹.

Bottleneck analysis. We tested the bottleneck effect at the population level to explore population demography using different models and testing methods implemented in Bottleneck 1.2.02, and the infinite allele model (IAM), two-phase model (TPM), and the strict stepwise mutation model (SMM) were applied with 10,000 replications⁷⁵. A qualitative descriptor of the allele frequency distribution (mode-shift indicator), which discriminates between bottlenecked and stable populations was also used. Under the three models, the standardized differences test was removed from this study because this test is typically only used when at least 20 polymorphic loci are available.

References

- Schirmer, S., Sengonca, C. & Blaeser, P. Influence of abiotic factors on some biological and ecological characteristics of the aphid parasitoid *Aphelinus asychis* (Hymenoptera: Aphelinidae) parasitizing *Aphis gossypii* (Sternorrhyncha: Aphididae). *Eur. J. Entomol.* **105**, 121–129, doi:10.14411/eje.2008.017 (2008).
- Kersting, U., Satar, S. & Uygun, N. Effect of temperature on development rate and fecundity of apterous *Aphis gossypii* Glover (Hom., Aphididae) reared on *Gossypium hirsutum* L. *J. Appl. Entomol.* **123**, 23–27, doi:10.1046/j.1439-0418.1999.00309.x (1999).
- Ebert, T. A. & Cartwright, B. Biology and ecology of *Aphis gossypii* Glover (Homoptera: aphididae). *Southwest. Entomol.* **22**, 116–153 (1997).

4. Inaizumi, M. Life cycle of *Aphis gossypii* (Homoptera, Aphididae) with special reference to biotype differentiation on various host plants. *Jpn. J. Entomol.* **49**, 219–240 (1981).
5. Guldemond, J. A., Tigges, W. T. & De Vrijer, W. F. Host races of *Aphis gossypii* (Homoptera, Aphididae) on cucumber and chrysanthemum. *Environ. Entomol.* **23**, 1235–1240, doi:10.1093/ee/23.5.1235 (1994).
6. Wool, D., Hales, D. & Sunnucks, P. Host plant relationships of *Aphis gossypii* Glover (Homoptera: aphididae) in Australia. *J. Aust. Entomol. Soc.* **34**, 265–271, doi:10.1111/aen.1995.34.issue-3 (1995).
7. Hales, D. F., Tomiuk, J., Wöhrmann, K. & Sunnucks, P. Evolutionary and genetic aspects of aphid biology: A review. *Eur. J. Entomol.* **94**, 1–55 (1997).
8. Blackman, R. L. The simplification of aphid terminology. *Eur. J. Entomol.* **91**, 139–141 (1994).
9. Delorme, R., Augé, D., Béthenod, M. T. & Villatte, F. Communication to the editor insecticide resistance in a strain of *Aphis gossypii* from Southern France. *Pestic. Sci.* **49**, 90–96 (1997).
10. Herron, G. A., Powis, K. & Rophail, J. Insecticide resistance in *Aphis gossypii* Glover (Hemiptera: Aphididae), a serious threat to Australian cotton. *Aust. J. Entomol.* **40**, 85–91, doi:10.1046/j.1440-6055.2001.00200.x (2001).
11. Ahmad, M., Arif, M. I. & Denholm, I. High resistance of field populations of the cotton aphid *Aphis gossypii* Glover (Homoptera: aphididae) to pyrethroid insecticides in Pakistan. *J. Econ. Entomol.* **96**, 875–878, doi:10.1093/jee/96.3.875 (2003).
12. Ai, Y., Qiu, X. H. & He, F. Q. A review of insecticide resistance mechanism of *Aphis gossypii*. *Entomol. Knowl.* **40**, 385–391 (2003).
13. Sun, J. T., Lian, C., Navajas, M. & Hong, X. Y. Microsatellites reveal a strong subdivision of genetic structure in Chinese populations of the mite *Tetranychus urticae* Koch (Acari: Tetranychidae). *BMC Genet.* **13**, 1–13 (2012).
14. Susanta, K. B. Molecular marker systems in insects: current trends and future avenues. *Mol. Ecol.* **15**, 3087–3113, doi:10.1111/mec.2006.15.issue-11 (2006).
15. Simon, J. C. *et al.* Reproductive mode and population genetic structure of the cereal aphid *Sitobion avenae* studied using phenotypic and microsatellite markers. *Mol. Ecol.* **8**, 531–545, doi:10.1046/j.1365-294x.1999.00583.x (1999).
16. Massonnet, B., Leterme, N., Simon, J. C. & Weisser, W. W. Characterization of microsatellite loci in the aphid species *Macrosiphoniella tanacetaria* (Homoptera, Aphididae). *Mol. Ecol. Notes.* **1**, 14–15, doi:10.1046/j.1471-8278.2000.00006.x (2001).
17. Llewellyn, K. S. *et al.* Migration and genetic structure of the grain aphid (*Sitobion avenae*) in Britain related to climate and clonal fluctuation as revealed using microsatellites. *Mol. Ecol.* **12**, 21–34, doi:10.1046/j.1365-294x.2003.01703.x (2003).
18. Astorga, M. & Galleguillos, R. Genetic divergence of jack mackerel of genus *Trachurus* from northwestern and southeastern Pacific. *Rev. Biol. Mar. Oceanogr.* **33**, 155–161 (1998).
19. Barraclough, T. G. & Nee, S. Phylogenetic and speciation. *Trends. Ecol. Evol.* **16**, 391–399, doi:10.1016/S0169-5347(01)02161-9 (2001).
20. Zhang, F. & Liu, X. D. Variation of host-specialized and migratory biotypes of *Aphis gossypii* Glover based on mtDNA COI gene sequences. *J. Nanjing Agr. Univ.* **35**, 65–70 (2012).
21. Vanlerberghe-Masutti, F. & Chavigny, P. Host-based genetic differentiation in the aphid *Aphis gossypii* Glover, evidenced from RAPD fingerprints. *Mol. Ecol.* **7**, 905–914, doi:10.1046/j.1365-294x.1998.00421.x (1998).
22. Fuller, S. J., Chavigny, P., Lapchin, L. & Vanlerberghe-Masutti, F. Variation in clonal diversity in glasshouse infestations of the aphid, *Aphis gossypii* Glover in southern France. *Mol. Ecol.* **8**, 1867–1877, doi:10.1046/j.1365-294x.1999.00782.x (1999).
23. Charabi, K. *et al.* Genotypic diversity of the cotton-melon aphid *Aphis gossypii* (Glover) in Tunisia is structured by host plants. *B. Entomol. Res.* **98**, 333–341, doi:10.1017/S0007485307005585 (2008).
24. Takada, H. Interclonal variation in photoperiodic response for sexual morph production of Japanese *Aphis gossypii* Glover (Hemiptera, Aphididae). *J. Appl. Entomol.* **106**, 188–197, doi:10.1111/jen.1988.106.issue-1-5 (1988).
25. Gong, P., Yang, X. W., Zhang, X. X., Liu, X. D. & Chen, X. F. Microsatellite primer-PCR studies on the population differentiation of *Aphis gossypii* in relation to host plants and seasons. *Acta. Ecol. Sin.* **21**, 765–771 (2001a).
26. Gong, P., Zhang, X. X., Yang, X. W. & Chen, X. F. Microsatellite DNA polymorphism in different forms of the cotton aphid. *Acta. Ecol. Sin.* **44**, 416–421 (2001b).
27. Brévault, T., Carletto, J., Linderme, D. & Vanlerberghe-Masutti, F. Genetic diversity of the cotton aphid *Aphis gossypii* in the unstable environment of a cotton growing area. *Agr. Forest. Entomol.* **10**, 215–223, doi:10.1111/afe.2008.10.issue-3 (2008).
28. Thomas, S. *et al.* Genetic diversity of the cotton-melon aphid, *Aphis gossypii* Glover in different melon growing areas of France. *Redia.* **92**, 47–50 (2008).
29. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA Polymorphism. *Genetics* **123**, 585–595 (1989).
30. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
31. Delmotte, F., Leterme, N., Gauthier, J. P., Rispe, C. & Simon, J. C. Genetic architecture of sexual and asexual populations of the aphid *Rhopalosiphum padi* based on allozyme and microsatellite markers. *Mol. Ecol.* **11**, 711–723, doi:10.1046/j.1365-294x.2002.01478.x (2002).
32. Sunnucks, P., England, P. R., Taylor, A. C. & Hales, D. F. Microsatellite and chromosome evolution of parthenogenetic *Sitobion* aphids in Australia. *Genetics* **144**, 747–756 (1996).
33. Normark, B. B. Evolution in a putatively ancient asexual aphid lineage: recombination and rapid karyotype change. *Evol.* **53**, 1458–1469, doi:10.2307/2640892 (1999).
34. Wilson, A. C., Sunnucks, P. & Hales, D. F. Microevolution, low clonal diversity and genetic affinities of parthenogenetic *Sitobion* aphids in New Zealand. *Mol. Ecol.* **8**, 1655–1666, doi:10.1046/j.1365-294x.1999.00751.x (1999).
35. Vanlerberghe-Masutti, F., Chavigny, P. & Fuller, S. J. Characterization of microsatellite loci in the aphid species *Aphis gossypii* Glover. *Mol. Ecol.* **8**, 865–702, doi:10.1046/j.1365-294x.1999.00876.x (1999).
36. Timm, A. E., Pringle, K. L. & Warnich, L. Genetic diversity of woolly apple aphid *Eriosoma lanigerum* (Hemiptera: Aphididae) populations in the Western Cape, South Africa. *B. Entomol. Res.* **95**, 187–191, doi:10.1079/BER2004348 (2005).
37. Zamoum, T. *et al.* Does insecticide resistance alone account for the low genetic variability of asexually reproducing populations of the peach-potato aphid *Myzus persicae*? *Heredity.* **94**, 630–639, doi:10.1038/sj.hdy.6800673 (2005).
38. Posada, D. & Crandall, K. A. Intraspecific gene genealogies: trees grafting into networks. *Trends. Ecol. Evol.* **16**, 37–45, doi:10.1016/S0169-5347(00)02026-7 (2001).
39. Gascon, C. *et al.* Riverine barriers and the geographic distribution of Amazonian species. *P. Natl. Acad. Sci. USA* **97**, 13672–13677, doi:10.1073/pnas.230136397 (2000).
40. Cozzolino, S., Cafasso, D., Pellegrino, G., Musacchio, A. & Widmer, A. Fine-scale phylogeographical analysis of *Mediterranean Anacamptis palustris* (Orchidaceae) populations based on chloroplast minisatellite and microsatellite variation. *Mol. Ecol.* **12**, 2783–2792, doi:10.1046/j.1365-294x.2003.01958.x (2003).
41. Chapuis, M. P. & Estoup, A. Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* **24**, 621–631, doi:10.1093/molbev/msl191 (2007).
42. Yuan, J. H., Cheng, F. Y. & Zhou, S. L. Genetic structure of the tree peony (*Paeonia rockii*) and the Qinling Mountains as a geographic barrier driving the fragmentation of a large population. *PLoS One* **4**, e34955, doi:10.1371/journal.pone.0034955 (2012).
43. Peterson, M. A. & Denno, R. F. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. *Am. Nat.* **152**, 428–446, doi:10.1086/286180 (1998).
44. Meinke, L. J. & Ware, G. W. Tolerance of three beet armyworm strains in Arizona to methomyl. *J. Econ. Entomol.* **71**, 645–646, doi:10.1093/jee/71.4.645 (1978).

45. Chaufaux, J. & Ferron, P. Sensibilité différente de deux populations de *Spodoptera exigua* Hubner (Lepidoptera: Noctuidae) aux baculovirus et aux pyrethrinoides de synthèse. *Agronomie* **6**, 99–104, doi:10.1051/agro:19860109 (1986).
46. Moores, G. D., Gao, X. W., Denholm, I. & Devonshire, A. L. Characterization of insensitive acetylcholinesterase in insecticideresistant cotton aphids, *Aphis gossypii* Glover (Homoptera: Aphididae). *Pestic. Biochem. Phys.* **56**, 102–110, doi:10.1006/pest.1996.0064 (1996).
47. Favret, C. & Voegtlin, D. J. Speciation by host-switching in *Cinara* (Insecta: Hemiptera: Aphididae). *Mol. Phylogenet. Evol.* **32**, 139–151, doi:10.1016/j.ympev.2003.12.005 (2004).
48. Harry, M., Solignac, M. & Lachaise, D. Molecular evidence for parallel evolution of adaptive syndromes in fig-breeding *Lissocephala* (Drosophilidae). *Mol. Phylogenet. Evol.* **9**, 542–551, doi:10.1006/mpev.1998.0508 (1998).
49. Michel, A. P., Zhang, W., Jinky, J., Kang, S. T. & Mian, M. A. R. Population genetic structure of *Aphis glycines*. *Environ. Entomol.* **38**, 1301–1311, doi:10.1603/022.038.0442 (2009).
50. Kim, K. S., Hill, C. B., Hartman, G. L., Mar, M. & Diers, B. W. Discovery of soybean aphid biotypes. *Crop. Sci.* **48**, 923–928, doi:10.2135/cropsci2007.08.0447 (2008).
51. Gauffre, B. & D'Acier, A. C. New polymorphic microsatellite loci, cross-species amplification and PCR multiplexing in the black aphid, *Aphis fabae* Scopoli. *Mol. Ecol. Notes* **6**, 440–442, doi:10.1111/men.2006.6.issue-2 (2006).
52. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic. Acids. Res.* **25**, 4876–4882, doi:10.1093/nar/25.24.4876 (1997).
53. Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729, doi:10.1093/molbev/mst197 (2013).
54. Rozas, J. S., Sánchez-DelBarrio, J. C., Messegyer, X. & Rozas, R. DnaSP: DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496–2497, doi:10.1093/bioinformatics/btg359 (2003).
55. Swofford, D. L. PAUP*. *Phylogenetic analysis using parsimony (*and other methods)*, Version 4.0. Sinauer Associates, Sunderland, Massachusetts (2003).
56. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818, doi:10.1093/bioinformatics/14.9.817 (1998).
57. Posada, D. & Buckley, T. R. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Syst. Biol.* **53**, 793–808, doi:10.1080/10635150490522304 (2004).
58. Bandelt, H. J., Forster, A. R. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48, doi:10.1093/oxfordjournals.molbev.a026036 (1999).
59. Excoffier, L., Laval, G. & Schneider, S. Arlequin, ver. 3.0: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).
60. Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
61. Jensen, J. L., Bohonak, A. J. & Kelley, S. T. Isolation by distance, web service. *BMC Genet.* **6**, 13, doi:10.1186/1471-2156-6-13 (2005).
62. Raymond, M. & Rousset, F. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249, doi:10.1093/oxfordjournals.jhered.a111573 (1995).
63. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM-algorithm. *J. R. Stat. Soc. B* **39**, 1–38 (1977).
64. Peakall, R. & Smouse, P. E. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* **6**, 288–295, doi:10.1111/men.2006.6.issue-1 (2006).
65. Goudet, J. FSTAT (version 1.2): a computer program to calculate F-statistics. *J. Hered.* **86**, 485–486, doi:10.1093/oxfordjournals.jhered.a111627 (1995).
66. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotypic data. *Genetics* **155**, 945–959 (2000).
67. Takezaki, N., Nei, M. & Tamura, K. POPTREE 2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows-interface. *Mol. Bio. Evol.* **27**, 747–752, doi:10.1093/molbev/msp312 (2010).
68. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587 (2003).
69. Pritchard, J. & Falush, D. Documentation for structure software: version 2.2. (Chicago, USA, University of Chicago, 2007).
70. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620, doi:10.1111/mec.2005.14.issue-8 (2005).
71. Campana, M. G., Hunt, H. V., Jones, H. & White, J. CorrSieve: software for summarizing and evaluating structure output. *Molecular Ecology Resources* **11**, 349–352, doi:10.1111/men.2011.11.issue-2 (2011).
72. Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinform.* **23**, 1801–1806, doi:10.1093/bioinformatics/btm233 (2007).
73. Rosenberg, N. A. Distruct: a program for the graphical display of population structure. *Mol. Ecol. Notes* **4**, 137–138, doi:10.1046/j.1471-8286.2003.00566.x (2004).
74. Wilson, G. A. & Rannala, B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* **163**, 1177–1191 (2003).
75. Piry, S., Luikart, G. & Cornuet, J. M. Bottleneck: A computer program for detecting recent reductions in the effective population size using allele frequency data. *J. Hered.* **90**, 502–503, doi:10.1093/jhered/90.4.502 (2001).

Acknowledgements

We are grateful to Professor Yan-Hui Lu at Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China, for providing valuable comments and suggestions. We are grateful to Dr. Jing Chen of the Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China for providing valuable comments and suggestions. We are also grateful to many experts for supplying much important information about *A. gossypii* and enthusiastic help during our field surveys. This work was supported by the National Natural Science Foundation of China (Grant No. 31321004, No. 31101626, No. 31471770 and No. 31401961); Cultivation Plan for Youth Agricultural Science and Technology Innovative Talents of Liaoning Province (Grant No. 2014025); West Light Foundation of the Chinese Academy of Sciences (Y5C3021100).

Author Contributions

Conceived and designed the experiments: X.Y.W., X.M.Y. and K.M.W. Performed the experiments: X.Y.W. and L.H.Z. Analyzed the data: X.Y.W., X.M.Y. and B.L. Wrote the paper: X.Y.W. and X.M.Y. Contributed reagents/materials/analysis tools: K.M.W. Acquisition of data: X.Y.W. and X.M.Y. Contributed intellectually to the interpretation and discussion of results: K.M.W.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-02105-4](https://doi.org/10.1038/s41598-017-02105-4)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017