


# Patterns of genetic variation reflect multiple introductions and pre-admixture sources of common ragweed (*Ambrosia artemisiifolia*) in China

Feifei Li · Mark van Kleunen · Junmin Li · Xiaoyan Liu · Kexiao Gao · Jinfang Zhu · Xiangjian Zhao · Caiyun Zhao  · Junsheng Li

**Abstract** *Ambrosia artemisiifolia* is native to North America but has become a worldwide invasive weed. It was introduced to China more than 80 years ago and has spread into 20 provinces since then. To assess the population structure of *A. artemisiifolia* in China and whether this invasion involved a single event or multiple events, we investigated patterns of genetic variation for three chloroplast DNA intergenic spacer regions, a nrITS region and five microsatellite loci. Our dataset consists of 370 individuals from 19 sites throughout

China. We compared their cpDNA-haplotypes to those published for native North American populations. The distribution of cpDNA-haplotypes indicates that *A. artemisiifolia* was introduced to China multiple times from different source regions. The numbers of alleles in Chinese populations were not significantly lower than in native populations. Both nrITS-haplotypes and microsatellite alleles showed that there was no evidence for a genetic bottleneck. Four populations were genetically well separated from the other 15 populations. However, the absence of isolation by distance, and the low levels of genetic differentiation and gene flow among the other 15 population suggest that most populations in China come from pre-admixed populations. To find the exact source regions of the Chinese populations, more samples from the native region and other invaded regions will be necessary. Nevertheless, our study provides important insights into the genetic background of *A. artemisiifolia* invasion in China.

F. Li · X. Liu · K. Gao · J. Zhu · X. Zhao · C. Zhao (✉) · J. Li (✉)  
State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, People's Republic of China  
e-mail: zhaocy@craes.org.cn

J. Li  
e-mail: lijsh@craes.org.cn

M. van Kleunen · J. Li  
Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou 318000, People's Republic of China

M. van Kleunen  
Ecology, Department of Biology, University of Konstanz, Universitätsstrasse 10, 78464 Constance, Germany

**Keywords** Alien weed · Bottleneck · Gene flow · Genetic structure · Invasion history · Population differentiation

## Introduction

As a consequence of increased human mobility, economic globalization, trade liberalization and

environmental change, thousands of species have become invasive outside their native ranges (Pyšek et al. 2017; Pagad et al. 2018). Biological invasions are now among the most serious global ecological problems. To prevent new invasions, and to manage and control the species that are already invasive, an understanding of the invasion process and mechanisms underlying invasion success is necessary. While large-scale multi-species comparative studies have revealed some of the general drivers of invasion success (van Kleunen et al. 2018), the mechanisms underlying the invasion success of individual species requires specific case studies. For some individual invasive species, insights into the possible invasion mechanisms have been revealed by studies on the genetic background of the populations (Blossey and Notzold 1995; Müller-Schärer et al. 2004; Dlugosch and Parker 2008; Prentis et al. 2008; Rius and Darling 2014). Such genetic studies may reveal whether genetic processes, such as bottlenecks, drift, admixture and adaptive evolution, played key roles in the establishment and spread of invasive alien species (Dlugosch et al. 2015).

Low within-population genetic variation, e.g. due to genetic bottlenecks, is expected to reduce the adaptability of introduced species to their new environment. Therefore, the fact that some introduced species have become highly invasive despite having gone through genetic bottlenecks poses a “genetic dilemma” (Roman and Darling 2007). Recent research, however, indicates that founding events and low genetic diversity might be less disadvantageous than usually assumed (Dlugosch and Parker 2008). Moreover, Bossdorf et al. (2005) revealed that nearly 69% of invasive plants actually harboured the same or even higher genetic diversity in invasive than in native populations (Roman and Darling 2007). This finding indicates that genetic bottlenecks in invasive populations have not been very common or that genetic diversity has increased due to multiple introductions and admixture of the different genetic sources (Estoup et al. 2016).

Indeed, evidence is accumulating that many invasions involved multiple introduction events (Henry et al. 2009; Pairon et al. 2010; Cao et al. 2017; Fischer et al. 2017). When distantly-related populations come into contact, intraspecific pollen exchange and fertilization will result in genetic admixture (Shi et al. 2018). Admixture between multiple source

populations could benefit invasive populations immediately through heterosis (e.g. Li et al. 2018), and in the long term by increasing genetic diversity and thus evolutionary potential, which allow adaptation to new environmental conditions (Ellstrand and Schierenbeck 2006; Schierenbeck and Ellstrand 2009; Rius and Darling 2014). In principle, admixture could have occurred in the native source region or previously invaded region of the invasive species before the species was introduced to a new place. Such a pre-introduction admixed population may have a greater probability of establishment and successful invasion than a non-admixed one (Gaudeul et al. 2011). Genetic data combined with historical records on establishment can provide insight into whether human-mediated admixture occurs before or during the invasion process (Dlugosch and Parker 2008; Liu et al. 2012; Signorile et al. 2016).

*Ambrosia artemisiifolia* is a herbaceous plant native to North America that has become a notorious global weed in habitats that are frequently disturbed, such as crop fields, road verges, riverbeds, building sites and quarry areas, and this species also causes crop losses and pollen allergies in humans (Makra et al. 2015; Gentili et al. 2017). It can produce many small seeds, which can be easily dispersed by animals, and the transport of soil and agricultural crops. This species was introduced to Europe in the eighteenth century and became invasive during the middle of the twentieth century in both France and Eastern Europe (Gladieux et al. 2011; Gaudeul et al. 2011; Makra et al. 2015). It has now also invaded parts of Asia, Africa, Australia and the Atlantic islands (Lawalrée 1955; Priszter 1960; Gaudeul et al. 2011; Makra et al. 2015; Montagnani et al. 2017).

Genetic marker studies in the introduced European and Australian area have revealed relatively high levels of genetic diversity, most likely due to multiple introductions, in most parts of the invaded range of *A. artemisiifolia* (Genton et al. 2005; Chun et al. 2010; Gaudeul et al. 2011; van Boheemen et al. 2017). It has also been shown that Eastern and Western Europe have different genetic sources of *A. artemisiifolia*, implying that the Eastern European populations did not spread there from France (Gladieux et al. 2011). Martin et al. (2014, 2016) recently found evidence that intraspecific hybridization of *A. artemisiifolia* may have previously occurred in North America before the admixed populations were introduced to other places.

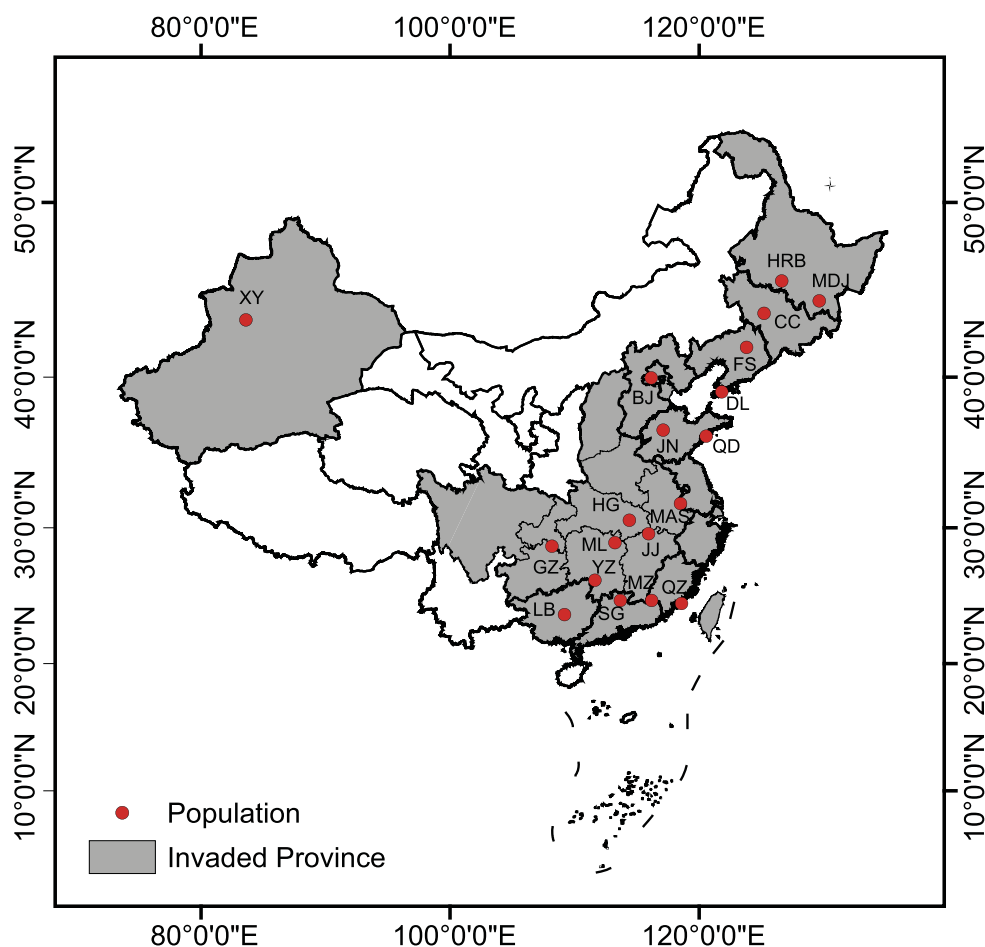
Nevertheless, although a pre-introduction admixed zone has been identified in North America, the European populations were most likely the result of multiple introductions from two non-admixed native populations, and the Australian populations were most likely founded by a single introduction from Eastern Europe (van Boheemen et al. 2017). So, there are different possible introduction scenarios for *A. artemisiifolia* in different parts of the world.

In China, *Ambrosia artemisiifolia* was introduced at the beginning of the twentieth century (Wan et al. 2009). The earliest herbarium records of this weed in China are from Shanwei city, Guangdong Province (He Guanzhou 0096, PE, 1954) and Hangzhou city, Zhejiang Province (Zhang Shaorao 1404, PE, 1957), but it currently occurs in 23 Chinese provinces (Wang 2005; Zhou et al. 2015; Dong et al. 2017; Wan et al. 2017) (Fig. 1, Table S1). A few studies have already assessed genetic components in a limited number of *A. artemisiifolia* populations in China. A microsatellite and amplified fragment length polymorphism (AFLP)

study revealed high genetic diversity and complete outcrossing in five Chinese populations (Li et al. 2012). Furthermore, a study on the molecular basis of flowering time differentiation between Beijing and Wuhan populations suggested that adaptation in this trait occurred during the spread of *A. artemisiifolia* in China (Li et al. 2015). Despite these few published genetic studies on *A. artemisiifolia* in China, large-scale studies on the genetic background of *A. artemisiifolia* in China are still lacking.

Our study aimed to assess large-scale genetic patterns among and within Chinese populations and to gain insights into the genetic background of *A. artemisiifolia* invasion in China. We combined multilocus genotypic data obtained with three different molecular markers to analyse the genetic diversity and population structure of Chinese *A. artemisiifolia* populations. To compare our cpDNA sequences with 56 haplotype sequences from North American populations published by Martin et al. (2014), we used the same cpDNA markers that were used in their study.

**Fig. 1** The Chinese provinces invaded by *Ambrosia artemisiifolia* and the distributions of the populations sampled in this study



The nuclear ribosomal ITS (nrITS) was also used due to its faster mutation rate and biparental inheritance. However, one nuclear region cannot provide sufficient information to analyze the genetic background of species, and most other nuclear markers were not single-copy markers. Therefore, simple sequence repeat (SSR) loci were used in this study as the third kind of molecular marker. This kind of codominant marker is very suitable and widely used to analyze the genetic structure of populations of diploid species population, such as *A. artemisiifolia* (Ciappetta et al. 2016; Meyer et al. 2017). Based on the results from three different molecular markers, we aimed to answer three main questions: (1) Did *A. artemisiifolia* populations in China undergo genetic bottlenecks? (2) Have multiple introductions happened during the invasion process in China and shaped the current genetic patterns? (3) Do *A. artemisiifolia* populations in China show genetic structure?

## Materials and methods

### Sample collection and genotyping

In 2016, we collected leaf samples of 370 individuals of *A. artemisiifolia* from 19 populations (5–30 samples per population) representing most of the invaded range in China (Table 1, Fig. 1). The habitats of those populations are ruderal sites. Within each population, the distance between each sampled individual was more than 3 m. Genomic DNA was extracted from the leaves using the EasyPure Plant Genomic DNA Kit (Beijing TransGen Biotech Co., Ltd., Beijing, China) according to the manufacturer's instructions. We amplified the nuclear ribosomal Internal Transcriber Spacer (nrITS) region and three chloroplast DNA (cpDNA) intergenic spacer regions (*psbA-trnH* spacer, *atpH-atpF* spacer, *psbK-psbI* spacer). The primer sequences are shown in Table S2. The 40  $\mu$ l PCR mixtures consisted of 20  $\mu$ l 2  $\times$  Taq PCR MasterMix (TIANGEN), 10–20 ng of template DNA and 1  $\mu$ l of each primer. The PCR-cycle protocol for amplification of the cpDNA regions was the same as the one in the study of Martin et al. (2014) on native North American population of *A. artemisiifolia*. The nrITS PCR protocol was as follows: 5 min at 94  $^{\circ}$ C for initial denaturation, 30 cycles of 30 s at 94  $^{\circ}$ C, 45 s at 57  $^{\circ}$ C and 1 min at 72  $^{\circ}$ C, followed by a final

extension for 10 min at 72  $^{\circ}$ C. The PCR products were purified and sequenced directly with an ABI 3730xl DNA Analyzer by Sangon Biotech (Shanghai, China).

In addition, we used five microsatellite primer pairs, including Ambart04, Ambart06, Ambart09, Ambart17 and Ambart27 (GenBank accession nos. FJ595149-FJ595151, FJ595152 and FJ595156) (Table S2) (Molecular Ecology Resources Primer Development Consortium 2009). The M13 tail of the forward primer was removed and labelled with fluorescent HEX or FAM. The PCR mixture contained 20  $\mu$ l 2  $\times$  Taq PCR MasterMix (TIANGEN), 10 ng of template DNA and 0.8  $\mu$ l of each primer. The PCR cycle protocol was as follows: 4 min at 95  $^{\circ}$ C for initial denaturation, 35 cycles of 40 s at 93  $^{\circ}$ C, 40 s at 54  $^{\circ}$ C and 30 s at 72  $^{\circ}$ C, followed by extension for 4 min at 72  $^{\circ}$ C. The PCR products were run with an ABI 3730xl DNA Analyzer (Sangon Biotech). Alleles were scored with GeneMapper 3.5 and GeneMarker V1.91 Demo. This was done at least twice to reduce genotyping errors.

### Genetic diversity analysis

The cpDNA and nrITS sequences were assembled using MEGA version 7 (Kumar et al. 2016). The heterozygous sites of nrITS sequences were identified by analysing the double peaks in chromatograms, and the allelic phases of nrITS were inferred using PHASE (Stephens et al. 2001; Stephens and Donnelly 2003; Harrigan et al. 2008;) and SeqPHASE (Flot 2010). For cpDNA-haplotype analysis, we combined our cpDNA sequences with 56 cpDNA-haplotype sequences (three concatenated chloroplast spacers) that represent all cpDNA-haplotypes from the 45 North American extant populations and historical *A. artemisiifolia* herbarium specimen in the study of Martin et al. (2014). The number of polymorphic sites ( $S$ ), the number of haplotypes ( $h$ ), haplotype diversity ( $Hd$ ), nucleotide diversity ( $\pi$ ), the average number of nucleotide differences ( $K$ ) and the population pairwise  $F_{ST}$  based on cpDNA and nrITS sequences were analyzed in DnaSP version 5.10 (Librado and Rozas 2009) without gap consideration. The cpDNA-haplotype and nrITS-haplotype networks were constructed by Network 5 using the minimum-evolution parsimony method (Polzin and Daneshmand 2003). We used pairwise  $F_{ST}$  values of cpDNA and nrITS to build

**Table 1** Location information, number of sampled individuals, cpDNA-haplotype and nrITS-haplotype diversity of *Ambrosia artemisiifolia* in each sampled population

Pop.	Collection location	Longitude (°E)	Latitude (°N)	No.	cpDNA					nrITS				
					<i>S</i>	<i>h</i>	<i>Hd</i>	$\pi$	<i>K</i>	<i>S</i>	<i>h</i>	<i>Hd</i>	$\pi$	<i>K</i>
BJ	Mentougou, Beijing	115.983	39.971	24	2	3	0.518	0.001	0.572	6	8	0.774	0.002	1.419
CC	Changchun, Jilin	125.218	43.873	18	0	1	0	0	0	11	8	0.686	0.003	1.546
DL	Dalian, Liaoning	121.821	39.102	14	9	2	0.538	0.004	4.846	8	8	0.802	0.003	1.561
FS	Fushun, Liaoning	123.822	41.847	12	0	1	0	0	0	6	6	0.641	0.002	0.989
ML	Miluo, Hunan	113.253	28.967	20	11	3	0.353	0.003	3.400	8	9	0.823	0.003	1.963
GZ	Tongren, Guizhou	108.065	28.625	15	0	1	0	0	0	5	5	0.526	0.002	1.469
HG	Huanggang, Hubei	114.541	30.996	20	9	2	0.100	0.001	0.900	4	6	0.524	0.001	0.835
HRB	Haerbin, Heilongjiang	126.639	45.734	19	9	2	0.199	0.002	1.789	5	7	0.688	0.002	1.131
JJ	Jiujiang, Jiangxi	115.943	29.591	22	11	4	0.658	0.002	1.797	7	7	0.754	0.002	1.390
JN	Jinan, Shandong	117.128	36.647	22	0	1	0	0	0	7	6	0.636	0.002	0.957
LB	Laibin, Guangxi	109.202	23.711	30	0	1	0	0	0	5	6	0.659	0.002	0.923
MAS	Maanshan, Anhui	108.515	31.710	25	1	2	0.220	0	0.220	4	5	0.716	0.002	1.359
MDJ	Mudanjiang, Heilongjiang	129.659	44.597	17	10	3	0.676	0.005	5.147	6	6	0.729	0.002	1.344
MZ	Meizhou, Guangdong	116.197	24.771	26	4	4	0.483	0.001	1.332	5	7	0.522	0.002	1.026
QD	Qindao, Shandong	120.569	36.252	5	1	2	0.600	0.001	0.600	4	4	0.644	0.002	1.378
QZ	Quanzhou, Fujian	118.590	24.545	12	1	2	0.167	0	0.167	3	3	0.663	0.002	1.152
SG	Shaoguan, Guangdong	113.689	24.769	22	2	3	0.177	0	0.182	5	5	0.698	0.002	1.027
XY	Xinyuan, Xinjiang	83.722	43.436	29	12	3	0.310	0.002	1.714	8	12	0.581	0.002	1.091
YZ	Yongzhou, Hunan	111.633	26.254	18	0	1	0	0	0	5	7	0.684	0.002	1.090
USA	–	–	–	56	40	35	0.969	0.005	5.117	–	–	–	–	–
Mean (China)	–	–	–	–	–	–	0.263	0.001	1.193	–	–	0.671	0.002	1.245
Total	–	–	–	426	41	36	0.669	0.003	3.255	32	35	0.759	0.002	1.442

No.: the number of sampled individuals, *S*: the number of polymorphic sites, *h*: the number of haplotypes, *Hd*: haplotype diversity,  $\pi$ : nucleotide diversity, *K*: average number of nucleotide differences

unrooted neighbor-joining (NJ) trees in MEGA version 7, reflecting genetic relationships among populations. The geographical distribution of cpDNA-haplotypes, nrITS-haplotypes and two NJ trees were visualized using the GenGIS software version 2.5.1 with a partial national map of China (Parks et al. 2013).

Based on the microsatellite loci, Hardy–Weinberg equilibrium (*HWE*) across loci and sampling localities, inbreeding coefficient (*F<sub>IS</sub>*) and heterozygosity excess and deficit were assessed using GENALEX 6.503 (Peakall and Smouse 2006, 2012). This software

was also used to estimate the number of different alleles (*Na*), the number of effective alleles (*Ne*), the Shannon's information index (*I*), the number of alleles unique to a single population (*Pa*), the expected heterozygosity (*He*) and the observed heterozygosity (*Ho*). We used Arlequin version 3.1 (Excoffier et al. 2006) to calculate the population specific fixation indices (*F<sub>ST</sub>*) with 10,000 permutations, the population pairwise *F<sub>ST</sub>* (with *P*-values). We did analysis of molecular variance (AMOVA) using three hierarchical levels: regions (as identified by GENELAND, see below), populations, and individuals. We used

analysis of variance (ANOVA) in the SPSS software to compare the total number of alleles and the number of private alleles between Chinese and native populations, using only the four microsatellite loci that had been used both by our study and the one of (Martin et al. 2014): Ambart04, Ambart06, Ambart17 and Ambart27.

#### Population genetic structure based on microsatellite loci

The microsatellite data were also used to analyse population genetic structure of *A. artemisiifolia* in China. The genetic structure of the Chinese populations was assessed using Bayesian model-based clustering analysis implemented in STRUCTURE 2.3 (Pritchard et al. 2010). The initial range of potential genetic clusters ( $K$ ) was set to range from 1 to 10 with 10 independent runs under the admixture model with 200,000 MCMC iterations and a 100,000 burn-in period. The most probable number of clusters ( $K$ ) was selected by calculating an *ad hoc* statistic  $\Delta K$  based on comparing the log probability of the data ( $\text{LnP}(D)$ ) for each value of  $K$  as described by Evanno et al. (2005), and was implemented in the online program STRUCTURE HARVESTER (Earl 2012). The highest  $\Delta K$  value was selected to determine the number of clusters. The software CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) was used to calculate the average membership coefficient ( $Q$ ) for each individual, permute population-membership-coefficient matrices ( $Q$ -matrices) for 10 replicate cluster analyses, and to produce a mean of the permuted matrices. Each individual's probability of assignment to each cluster was visualized using the Distruct software version 1.1 (Rosenberg 2004). Only the individuals with  $Q \geq 0.6$  membership coefficients to a cluster were considered to be independent, as suggested by previous studies (Evanno et al. 2005; Naegele et al. 2015; Karn and Jasieniuk 2017). We also used a spatially explicit Bayesian clustering algorithm implemented in the R package GENELAND 3.1.4 (Guillot et al. 2008). We did 19 independent runs under the spatial model and the assumption of uncorrelated allele frequencies between samples for a total of 1000,000 MCMC iterations, and thinned to one in every 1000th. Post-processing was used to extract information from MCMC simulations with a horizontal and a vertical discretization of the spatial domain in 100 pixels and a

**Fig. 2 a** The cpDNA-haplotype network derived from three cpDNA intergenic spacer regions (*psbA-trnH* spacer, *atpH-atpF* spacer, *psbK-psbI* spacer) of 370 sample individuals from 19 *Ambrosia artemisiifolia* populations in China and 56 published post-concatenation cpDNA sequences from native North American populations (Martin et al. 2014). The circle size is proportional to the cpDNA-haplotype frequency. The cpDNA-haplotypes that are in yellow have only been recorded in North America. The cpDNA-haplotypes that have been recorded in China have colors different from yellow, but also have yellow in them if some of their sequences correspond to those of North American cpDNA-haplotypes. The proportion of yellow corresponds to the proportion of North American sequences. **b** Sampling location and cpDNA-haplotypes (pie charts) distribution of *A. artemisiifolia* in China with the cpDNA-NJ tree. The size of each pie chart is proportional to sample size. The color proportions of pie charts represent the frequency of each cpDNA-haplotype in a population, and the color of each cpDNA-haplotype corresponds to the non-yellow color of the cpDNA-haplotype in the cpDNA-haplotype network (a)

burn-in of 500 iterations. Isolation by distance (IBD) was tested using a Mantel test on the matrices of the pairwise  $F_{ST}$  values and the pairwise geographical distances between populations in the software GenA-IEx 6.503 (Peakall and Smouse 2006).

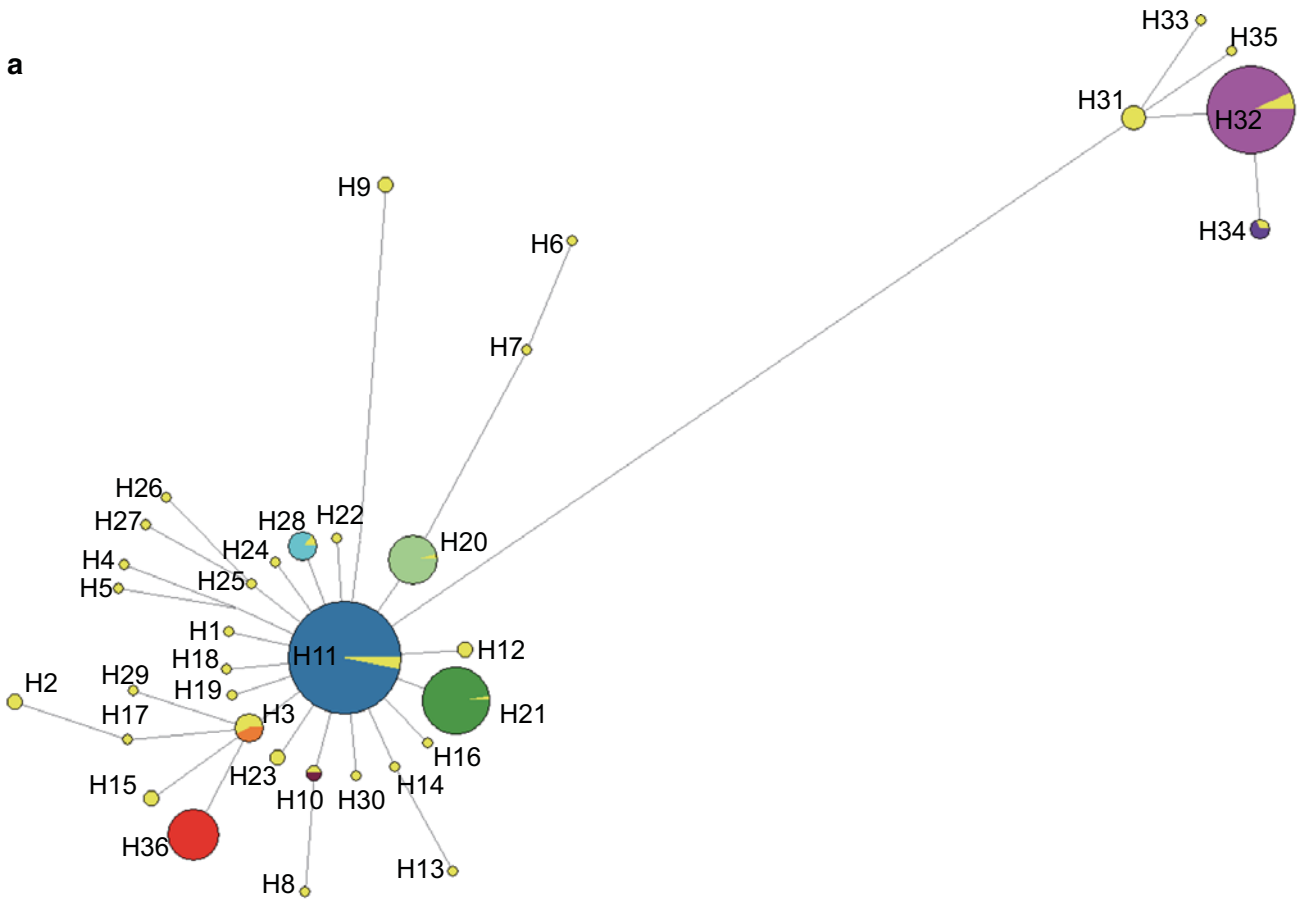
#### Gene flow

We used the software package BAYESASS v 3.0 (Wilson and Rannala 2003) to estimate pairwise gene flow between populations based on the microsatellite loci. This software uses Bayesian statistics to calculate a posterior probability distribution of the proportion of migrants from one population to another. We ran the MCMC for 10 million iterations, the first 1 million iterations were discarded as burn-in, and the remaining 9 million were thinned to one in every 1000th. Migration rates (gene flow), allele frequencies, and inbreeding coefficients were adjusted by changing the values of the respective mixing parameters as 0.50, 0.95 and 0.50 to achieve recommended acceptance rates (Wilson and Rannala 2003).

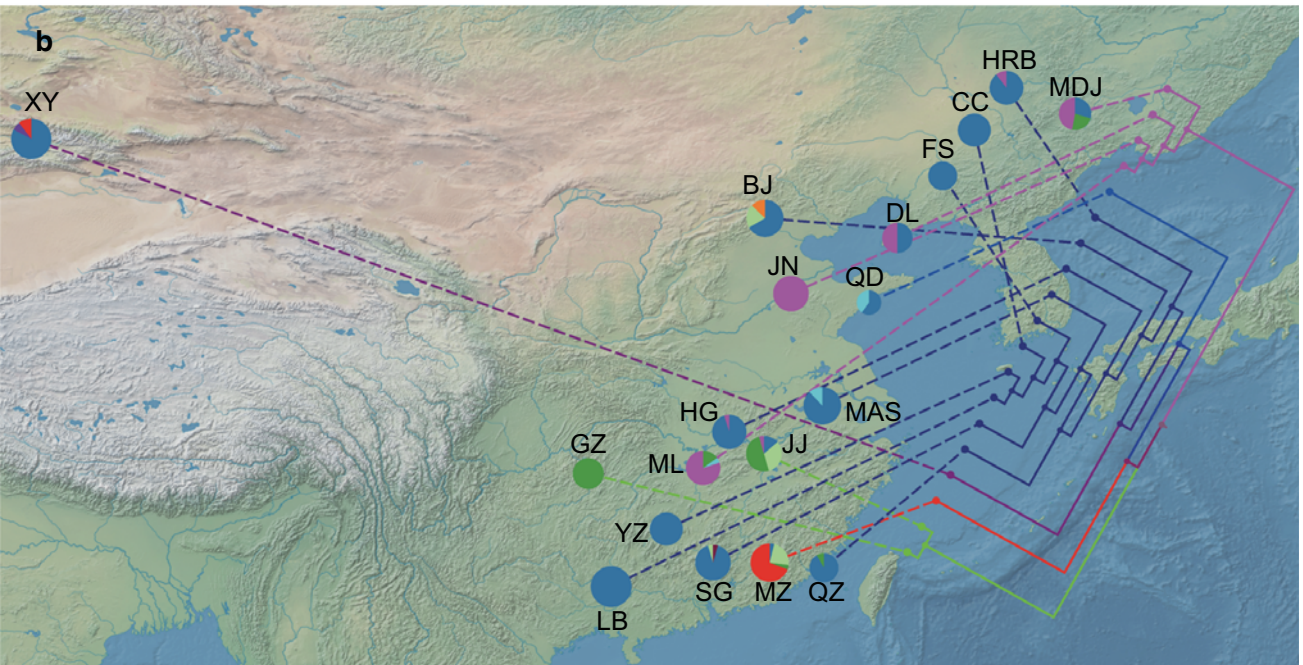
#### Genetic bottlenecks analysis

The BOTTLENECK v.1.2.02 software (Cornuet and Luikart 1996) was used to detect potential bottlenecks for microsatellite data. Because few microsatellite loci follow the strict Stepwise Mutation Model (SMM) and the infinite alleles model (IAM), we chose to use the

a



b



Two-phased Mutation Model (TPM, with 70% SMM), which is considered to be the most suitable one for analyzing microsatellite data (Di Rienzo et al. 1994). Significance was tested with the Wilcoxon sign-rank test for each population.

## Results

### Genetic diversity of *Ambrosia artemisiifolia* in China

We observed 41 polymorphic sites and 36 cpDNA-haplotypes among the 370 post-concatenation cpDNA sequences from China and the 56 post-concatenation cpDNA sequences from North America (Table 1, Fig. 2). Invasive populations in China contained nine cpDNA-haplotypes, and eight of them were also found in North America. The cpDNA-haplotype H36 was found only in China (populations MZ and XY) (Fig. 2). Another 27 cpDNA-haplotypes from the native range (North America) were not found in any of the invasive populations that we sampled.

The mean cpDNA-haplotype diversity of our 370 sampled individuals, representing 19 populations in China, was 0.263, and the mean nucleotide diversity was very low ( $\pi$ -mean = 0.001). The average number of nucleotide differences ( $K$ ) was 1.193. The populations MZ and JJ had the largest number of cpDNA-haplotypes (four cpDNA-haplotypes), whereas the populations CC, FS, GZ, JN, LB and YZ each had only one cpDNA-haplotype. Population MDJ, with three cpDNA-haplotypes at generally equal proportions, had the highest cpDNA-haplotype diversity ( $Hd = 0.676$ ), and nucleotide diversity ( $\pi = 0.005$ ) and average number of nucleotide differences ( $K = 5.147$ ).

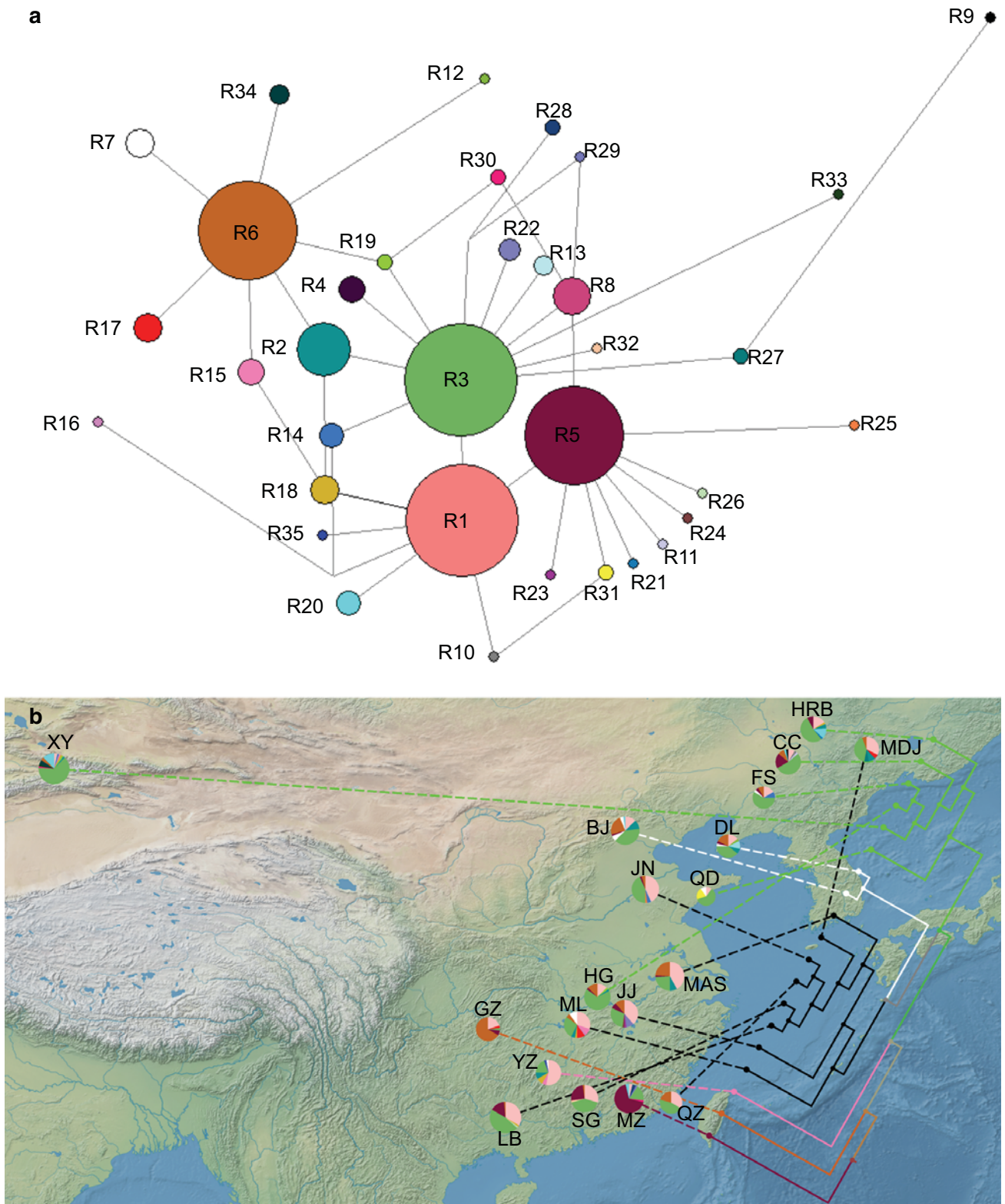
The cpDNA-haplotype network displayed a nearly starlike topology, similar to the networks for North America (Martin et al. 2014). Among all the cpDNA-haplotypes, H11 was the central cpDNA-haplotype in the network (Fig. 2a) and was connected to most other cpDNA-haplotypes by a single mutation. This cpDNA-haplotype was also widespread in China (61.1% of all individuals) and North America (Martin et al. 2014), and it was the dominant cpDNA-haplotype in 12 of the 19 populations. H32 was the second most common cpDNA-haplotype (15.4% of all individuals), and was the dominant cpDNA-haplotype

**Fig. 3** a nrITS-haplotype network derived from the nrITS region of 370 sampled individuals from 19 *Ambrosia artemisiifolia* populations in China. The circle size is proportional to the nrITS-haplotype frequency. Each nrITS-haplotype has a single color. b Sampling location and nrITS-haplotypes (pie charts) distribution of *A. artemisiifolia* in China with the nrITS-NJ tree. The size of the pie charts is proportional to the sample size. The color proportions of each pie chart represent the frequency of each nrITS-haplotype in a population, and the color of each nrITS-haplotype corresponds to the color of the nrITS-haplotype in the nrITS-haplotype network (a)

in populations JN, ML and MDJ (Fig. 2b). The cpDNA-haplotype H21 was mainly distributed in the south of China, except for four individuals that were found in MDJ. It was the dominant cpDNA-haplotype in GZ (100% of individuals) and JJ (50% of individuals). The cpDNA-haplotype H36 was dominant in MZ (69.23% of individuals in MZ) and was not found in any other populations except for XY (three individuals). The cpDNA-haplotype H3 was found only in BJ (three individuals), H34 was found only in XY (two individuals), and H10 was found only in SG (one individual).

Based on the nrITS data, 32 polymorphic (segregating) sites and 35 nrITS types were identified among all sampled individuals in China (Table 1, Fig. 3). Populations in China had a high nrITS-haplotype diversity (mean  $Hd = 0.671$ ), a low nucleotide diversity (mean  $\pi = 0.002$ ), and 1.245 nucleotide differences ( $K$ ) on average. The number of nrITS-haplotypes per population ranged from 3 (population QZ) to 12 (population XY). Population ML had the highest genetic diversity, as indicated by all three indices ( $Hd = 0.823$ ,  $\pi = 0.003$ , and  $K = 1.963$ ).

The nrITS-haplotypes network revealed complex genetic lineages of *A. artemisiifolia* in China (Fig. 3a). nrITS-haplotypes R3 (51.9% of the individuals), R1 (34.1%), R5 (14.9%) and R6 (14.3%) were the most frequent and widespread nrITS-haplotypes, and three of them (R3, R1 and R5) were successively connected by a single mutation (Fig. 3a). The nrITS-haplotypes R3 and R1 occurred across the entire range of *A. artemisiifolia* in China, and both of them were common in most populations, with the exception of populations XY, MZ and GZ (Fig. 3b). Thirteen of the 19 populations had 21 private nrITS-haplotypes, and 16 of these nrITS-haplotypes belonged to one individual respectively. The highest proportion of individuals with private nrITS-haplotypes was observed in population HRB (26.3%).



We detected a total of 114 alleles at the five nuclear microsatellite loci in the Chinese individuals. The average number of alleles at each locus in China was

$7.92 \pm 0.36$  (Table 2). For the four loci shared between our study and the study of Martin et al. (2014), the total number of alleles and the number of

**Table 2** Genetic diversity and bottlenecks of *Ambrosia artemisiifolia* populations in China based on microsatellite data, bottlenecks with significant results ( $P < 0.05$ ) are in bold and underlined

The population codes are the same as in Table 1

$N_a$ : the number of alleles at each locus,  $N_e$ : number of effective alleles,  $I$ : Shannon's information index,  $P_A$ : number of private alleles,  $H_o$ : observed heterozygosity,  $H_e$ : expected heterozygosity,  $uH_e$ : unbiased expected heterozygosity,  $F_{IS}$ : inbreeding coefficient,  $TPM_{W2}$ : the Two-phased Mutation Model under Wilcoxon sign-rank test

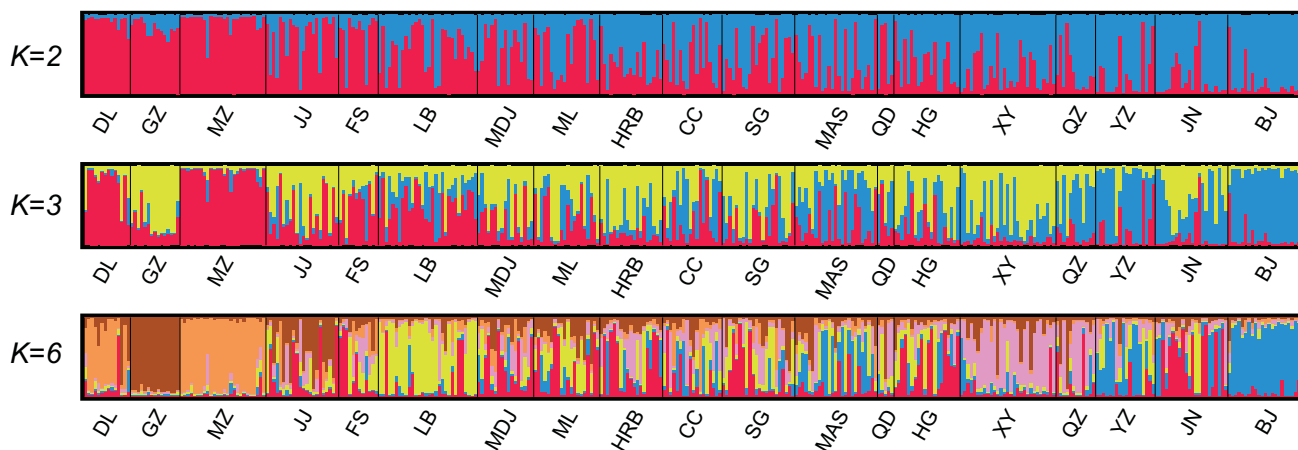
Pop.	$N_a$	$N_e$	$I$	$P_A$	$H_o$	$H_e$	$uH_e$	$F_{IS}$	$TPM_{W2}$
BJ	9.200	5.906	1.883	0.200	0.466	0.795	0.812	0.441	0.063
CC	9.400	6.229	1.880	0.400	0.500	0.782	0.804	0.380	0.625
DL	5.000	3.572	1.297	0.200	0.543	0.661	0.685	0.180	<b>0.031</b>
FS	5.800	3.353	1.381	0.400	0.433	0.656	0.685	0.352	0.156
ML	8.400	4.937	1.688	0.000	0.530	0.731	0.750	0.283	1.000
GZ	3.800	2.595	1.045	0.200	0.587	0.568	0.587	-0.063	<b>0.031</b>
HG	8.400	5.233	1.732	0.600	0.620	0.749	0.768	0.189	0.625
HRB	9.400	6.393	1.953	0.600	0.658	0.821	0.843	0.203	<b>0.031</b>
JJ	9.400	4.967	1.776	0.400	0.494	0.754	0.771	0.345	0.156
JN	8.600	4.626	1.605	0.000	0.373	0.691	0.707	0.466	0.625
LB	8.200	4.984	1.692	0.000	0.573	0.756	0.769	0.252	0.094
MAS	11.800	6.372	2.048	0.200	0.432	0.811	0.828	0.478	0.219
MDJ	8.800	5.801	1.838	0.000	0.659	0.798	0.822	0.167	0.156
MZ	8.600	4.163	1.593	1.000	0.491	0.724	0.738	0.343	0.219
QD	3.800	3.286	1.141	0.000	0.510	0.632	0.707	0.277	0.063
QZ	6.400	3.910	1.466	0.000	0.595	0.683	0.713	0.123	0.625
SG	9.200	5.145	1.765	0.600	0.403	0.743	0.760	0.448	0.813
XY	9.400	5.761	1.787	0.600	0.586	0.752	0.765	0.231	0.625
YZ	6.800	4.051	1.543	0.000	0.522	0.725	0.745	0.303	0.813
Mean	7.916	4.804	1.638		0.525	0.728	0.751	0.284	
SE	0.359	0.245	0.053		0.023	0.015	0.015	0.029	
Total	22.800	8.920	2.333	27.000	0.523	0.833	0.834	0.379	

private alleles did not differ between the Chinese and North American populations (for the total number of alleles:  $F = 2.769$ ,  $P = 0.147$ ; for the number of private alleles:  $F = 0.728$ ,  $P = 0.426$ , Table S3). In China, the largest number of different alleles ( $N_a$ ), number of effective alleles ( $N_e$ ) and Shannon's information index ( $I$ ) were found in the populations MAS ( $N_a = 11.8$ ,  $N_e = 6.4$ , and  $I = 2.05$ ) and HRB ( $N_a = 9.4$ ,  $N_e = 6.4$ , and  $I = 1.95$ ). The lowest values of these three indices were found in the population GZ ( $N_a = 3.8$ ,  $N_e = 2.6$ , and  $I = 1.05$ ). The mean number of private alleles ( $P_A$ ) at each locus was the highest in population MZ ( $P_A = 1$ ), whereas no private alleles were detected in the populations ML, JN, LB, MDJ, QD and QZ. The highest values of the observed and expected heterozygosity were also found in population HRB. We detected 52 heterozygote deficiencies in 95 population-locus combinations (19 populations  $\times$  5 loci). Significant deviation from Hardy-Weinberg equilibrium was observed in each of the populations with positive inbreeding coefficient ( $F_{IS}$ ) values, with the exception of population GZ ( $F_{IS} = -0.063$ ).

## Genetic structure

The genetic relationships among the 19 populations were visualized by neighbor joining (NJ) trees (Fig. 2). The topologies of the two trees were partly incongruent. For cpDNA, the populations GZ and JJ were grouped together and appeared to be divergent from the other populations, whereas for the nrITS, GZ (but not JJ) was separated from the others. Moreover, for the nrITS, YZ was differentiated and formed a single clade. The populations JN, ML, DL and MDJ, which had higher proportions of cpDNA-haplotype H32 than the other populations, formed one clade in the cpDNA-NJ tree but not in the nrITS-NJ tree. However, the divergence of population MZ was consistent in the NJ trees of both the nuclear and chloroplast markers.

The non-spatial algorithm implemented in STRUCTURE based on microsatellite loci identified the highest  $\ln Pr(X|K)$  value at  $K = 2$  (i.e., two clusters represented by red and blue), followed by  $K = 3$  (i.e., three clusters represented by red, blue and yellow) and  $K = 6$  (i.e., six clusters represented by six colours)



**Fig. 4** Assignment of the 370 individuals of *Ambrosia artemisiifolia* to genetic clusters using  $K = 2, 3$  and  $6$  based on STRUCTURE simulations using the microsatellite data

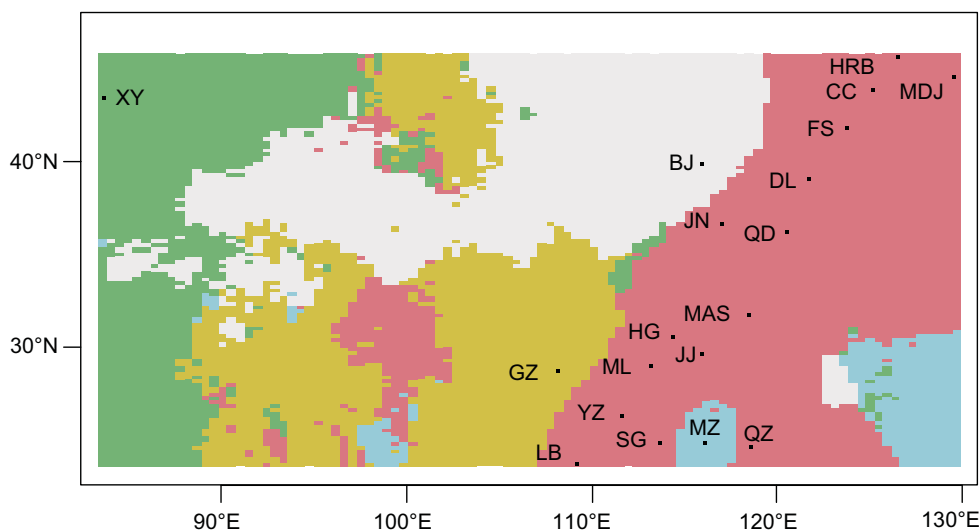
(Fig. 4, Fig. S1). At  $K = 2$ , almost all of the individuals belonging to the populations DL, GZ and MZ were grouped together in the red cluster. Population BJ mainly consisted of individuals of the blue cluster, and the other populations were mixed. Thirty-nine of the 370 individuals (in 13 of the 19 populations) could not be assigned to a single cluster (i.e., had a  $Q < 0.6$ ), indicating admixture between the two clusters.

When using  $K = 3$  or  $K = 6$ , MZ and BJ were clearly distinct from the other populations (population  $Q > 0.8$ ). At  $K = 6$ , population GZ no longer grouped with populations DL and MZ (brown cluster, population  $Q > 0.8$ ). LB had a lower population  $Q$  value (0.64) than the other populations, and it was assigned to the yellow cluster without strong confidence. The other populations had each at least three gene pools.

The spatially explicit model (with uncorrelated allele frequencies) implemented in GENELAND identified five genetic clusters (Fig. 5, Fig S2, S3). Similar to the results from STRUCTURE at  $K = 3$  and  $K = 6$ , the populations GZ, MZ and BJ formed three distinct clusters. However, unlike the results from STRUCTURE, population XY was identified as an additional distinct cluster, whereas LB was not a distinct cluster. The other 14 populations formed a single cluster with LB.

Neither STRUCTURE nor GENELAND revealed clear geographic patterns in the genetic structure of *A. artemisiifolia* in China. Pairwise  $F_{ST}$  values ranged from 0.022 (ML vs. HG populations) to 0.224 (GZ vs. YZ populations) with an average of 0.099, and most of them were statistically significant. The population

**Fig. 5** GENELAND results for  $K = 5$  using the spatial model with correlated allele frequencies. Map of estimated posterior probability of *Ambrosia artemisiifolia* population membership (by posterior mode)



fixation index ( $F_{ST}$ ) averaged over all loci was 0.095 (Table S4). A Mantel test of pairwise  $F_{ST}$  values and pairwise geographical distances revealed no evidence for isolation by distance ( $r = -0.003$ ,  $P = 0.550$ ). Analysis of molecular variance (AMOVA) revealed that most of the variance occurred within individuals (61.0%) rather than among individuals within populations (27.8%), among populations (6.4%) and among the five regions identified by GENELAND (4.81%) (Table S5).

### Genetic bottlenecks analysis and gene flow

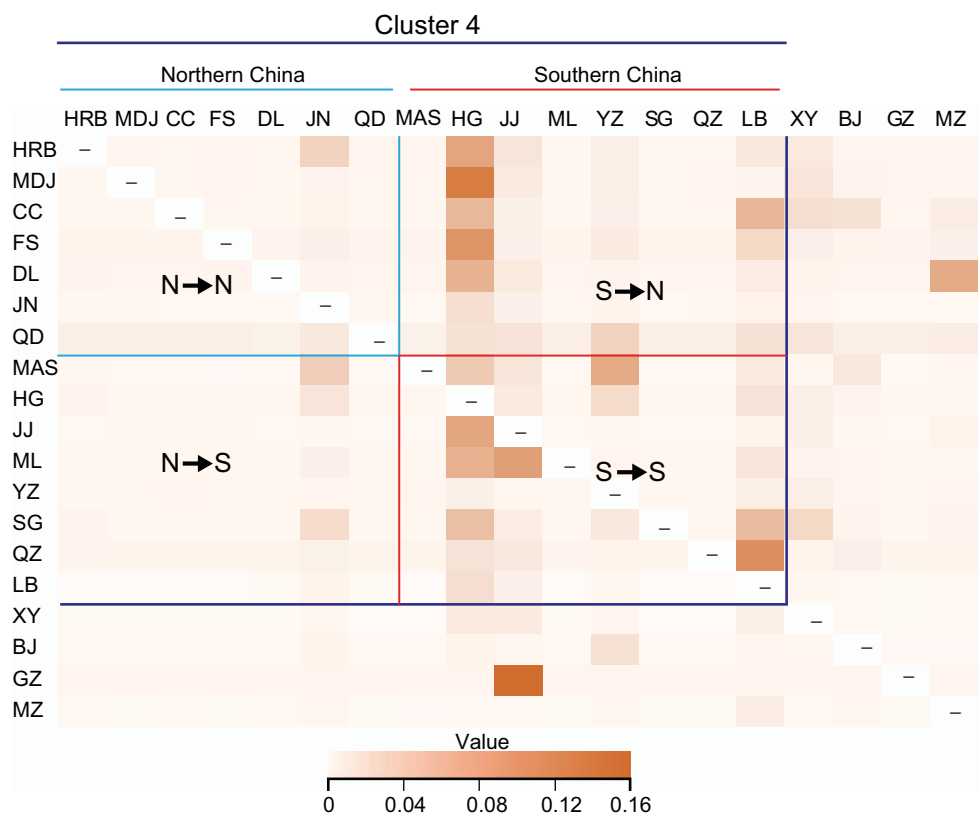
Based on microsatellite alleles, bottlenecks were only detected in DL, GZ and HRB using the Two-phased Mutation (TPM) model (Table 2). The inferred (posterior mean) migration rates of the 19 populations based on microsatellite loci are provided as a heat map (Fig. 6, Table S6). Only low levels of gene flow were found between populations; estimates ranged from 0.006 to 0.156 with an average of 0.015. Self-recruitment dominated with proportions of over 0.8 for most populations. The gene flow within Cluster 4 (as identified by GENELAND; 0.016 on average) was slightly higher than the overall average). However,

within cluster 4, there was limited evidence for gene flow from northern China to southern China (0.010 on average) and within northern China (0.011 on average), whereas more gene flow was found from southern China to northern China (0.022 on average) and within southern China (0.021 on average). Population JJ might be a major source of gene flow to GZ with a relatively high proportion of migrants per generation ( $0.156 \pm 0.029$ ), while HG might be a major source of gene flow to MDJ ( $0.135 \pm 0.035$ ) (Table S6).

### Discussion

In this study, we used three different kinds of molecular markers to gain insight into the occurrence of genetic variation, bottlenecks, multiple introductions, admixture and gene flow in *A. artemisiifolia* in its non-native range in China. The numbers of alleles in Chinese populations were not significantly lower than those in native populations, and in two of the 19 sampled Chinese populations (MZ and XY), a new cpDNA-haplotype not found in the North American native range was identified. Additionally, significant

**Fig. 6** Contemporary gene flow among populations of *Ambrosia artemisiifolia* in China, the color shades represented levels of gene flow



genetic bottlenecks were not detected in most of the sampled populations of *A. artemisiifolia* in China. Four populations (XY, BJ, MZ and GZ) formed a distinct genetic cluster separated from the other 15 populations, which showed low levels of gene flow. Overall, these findings suggest that *A. artemisiifolia* has not undergone major bottlenecks and has been introduced multiple times into China.

#### Genetic diversity and structure of *Ambrosia artemisiifolia* populations in China

The number of cpDNA-haplotypes in the 19 populations that we sampled across China (nine cpDNA-haplotypes) was lower than the number reported for the native range (35 cpDNA-haplotypes; Martin et al. 2014). Nevertheless, based on nrITS-haplotypes and microsatellite alleles, we did not detect any declines in genetic diversity that could be indicative of bottlenecks in the Chinese populations. Our analyses, revealed a significant heterozygote deficiency in most Chinese populations. This finding is in line with the high levels of heterozygote deficiency observed in *A. artemisiifolia* populations in its native range in North America and its non-native ranges in Europe and Australia (Chun et al. 2011; Gaudeul et al. 2011; Gladioux et al. 2011; Martin et al. 2014, 2016; van Boheemen et al. 2017). A possible explanation for this heterozygote deficiency could be the Wahlund effect,—i.e., a reduction in heterozygosity as a consequence of population substructuring. However, as wind-pollinated *A. artemisiifolia* shows extensive pollen dispersal, population substructuring is unlikely; therefore some studies have argued that the heterozygote deficiency must be due to selfing (Genton et al. 2005; Martin et al. 2009, 2014; van Boheemen et al. 2017). This explanation is at odds with the observation that native *A. artemisiifolia* populations in Ontario (Friedman and Barrett 2008) and invasive populations in China (Li et al. 2012) are completely outcrossing. Therefore, the causes of widespread heterozygote deficiency in *A. artemisiifolia* should be further studied.

The 19 invasive populations of *A. artemisiifolia* varied in the amount of genetic diversity, as indicated by several genetic diversity parameters (Table 2). The populations HRB, MAS, MDJ and XY, which occurred in different parts of China (Fig. 1), all had relatively high levels of genetic diversity. In contrast,

population GZ in Guizhou Province had low genetic variation, and this low variation was consistent across the three molecular marker types we used. The bottleneck analysis indicated that the low genetic diversity of this population was due to by a recent bottleneck. Interestingly, this population was also the only one that did not deviate from Hardy–Weinberg equilibrium at any locus.

Consistent with prior studies on populations of *A. artemisiifolia* in its native range (Martin et al. 2014) and in non-native ranges (Gaudeul et al. 2011), we found that in China most genetic variation was found within populations instead of among populations (and among the five spatial genetic clusters identified with GENELAND). High within-population genetic diversity has been found in numerous highly invasive species, particularly in monoecious, wind-pollinated species with a predominantly outcrossing breeding system, such as *A. artemisiifolia* (Groves and Burdon 1986; Bossdorf et al. 2005; Li et al. 2015). The proportion of genetic variation among populations in China (8.69%) (Table S5) was similar to the value reported for the non-native range in Eastern Europe (8.79%), but higher than the values reported for the native range in North America (6.39%) and the non-native range in France (4.81%) (Gladioux et al. 2011). This more pronounced among-population genetic structure in China might imply that *A. artemisiifolia* in China originates from multiple native populations and that gene flow among Chinese populations has been limited. The latter hypothesis is supported by the results of our gene-flow analysis.

The cpDNA-haplotypes, nrITS-haplotypes and microsatellite alleles partly identified different genetic lineages. Discrepancies were particularly apparent when comparing the genetic lineages identified with cpDNA and the genetic lineages identified with nuclear loci. Similar discrepancies between the results for cpDNA and the results for nuclear loci have been reported for many other taxonomic groups (Tsutsui et al. 2009; Wang et al. 2011). This inconsistency most likely reflects the fact that cpDNA is non-recombinant and maternally inherited, and therefore transmitted through seeds, whereas nuclear DNA is biparentally inherited, and therefore transmitted through seeds and pollen. Moreover, the mutation rate of cpDNA is slower than the mutation rate of nrITS (Wolfe et al. 1987; Jin et al. 2016). While the patterns based on nuclear DNA were more difficult to interpret, the

cpDNA-haplotypes that were shared among distant populations imply that there has been at least long-distance seed dispersal of *A. artemisiifolia* in China.

Population MZ in Guangdong Province was identified by all three molecular markers as a separate genetic cluster distinct from the other 18 populations. This population had the largest numbers of private alleles and cpDNA-haplotypes. It also had a high proportion of cpDNA-haplotype H36, which was otherwise found only in population XY in Xinjiang Province, and it had the lowest proportion of nrITS-haplotype R1, which was common in many of the other populations. Population MZ was collected in Meizhou city. More than seven million Chinese people who live overseas are originally from Meizhou (the People's Government of Meizhou, 2016). The associated international traffic may have increased the chance of a unique haplotype being introduced in the city. Furthermore, Meizhou city is close to Shanwei city (c. 200 km), which is among the regions in China first invaded by *A. artemisiifolia*. If the MZ population descends from the early introduction in Shanwei, our results may imply that the genetic background of early founders was very different from that of later introductions. However, we did not find a clear link between the age of the herbarium specimens in a certain region and genetic variation.

Although population GZ in Guizhou Province was identified as a separate genetic cluster at the nuclear loci (SSR and nrITS), both cpDNA-haplotypes and microsatellite alleles showed that it had a close genetic relationship with population JJ (Jiangxi Province), which is located c. 800 km to the east. The single cpDNA-haplotype (H21) in GZ was also the most dominant cpDNA-haplotype in JJ, and three of the four nrITS-haplotypes in GZ also occurred in JJ. Consistently, our gene-flow analysis also revealed that the highest proportion of migrants per generation to GZ came from JJ (Fig. 6). This finding suggests that population JJ was a major source of genetic material for population GZ.

Most studies on *A. artemisiifolia* in China have focused on populations in eastern China. However, *A. artemisiifolia* has recently been reported in Xinjiang Province in Northwest of China (first recorded in 2010), and in 2016 the invaded area was over 1000 km<sup>2</sup>, which was 10,150 times larger than that in 2010 (Dong et al. 2017). Our study is the first to analyze the genetic diversity of a population (XY) in

Xinjiang. The XY population had the largest number of nrITS-haplotype and was identified as a separate cluster by GENELAND. Interestingly, the cpDNA-haplotype H36, which distinguished population MZ in Guangdong Province from most other populations, was also found in XY. This finding may suggest that the MZ population was one of the source populations for the XY population, but it could also be that both populations came from the same source.

#### Multiple sources were revealed in Chinese *Ambrosia artemisiifolia* populations

Of all 35 North American cpDNA-haplotypes from the study of Martin et al. (2014), only eight were found in our 19 Chinese populations. Unfortunately, we do not have any geographical information on the occurrence of the haplotypes in North America. Because of this limitation and because it is known that admixture happened within North America (Martin et al. 2014), we cannot say with certainty whether the Chinese cpDNA-haplotypes came from multiple genetically distinct populations or a single admixed population in the native range. However, the North American cpDNA-haplotype H3 was detected only in population BJ in Beijing Province, and cpDNA-haplotype H10 was detected only in population SG in Guangdong Province. This result suggests that *A. artemisiifolia* populations in China most likely did not result from a single introduction.

One of the haplotypes (H36) that we found in China (populations MZ and XY) has not been reported in the native range. This apparently novel cpDNA-haplotype may have evolved in the invasive range or may be relatively rare in North America and thus has not been detected there. The cpDNA-haplotype network (Fig. 2a) showed that H36 was most closely related to H3, and thus might have evolved from H3. However, the populations that included cpDNA-haplotype H36 (MZ and XY) did not include H3. The latter haplotype was found only in population BJ, which is geographically very distant from both MZ and XY. Therefore, we conclude that the novel cpDNA-haplotype most likely came from either a North American population that was not included in the study of Martin et al. (2014) or an invaded region outside of China.

There was no link between geographical location (longitude and latitude) and the occurrence of cpDNA-

and nrITS-haplotypes. Most populations had very low levels of genetic differentiation even though they were far apart. For example, the southernmost population (SG) and the northernmost population (HRB) were in the same clade in the cpDNA-NJ tree (Fig. 2). Consistent with this pattern, we did not find any evidence for isolation-by-distance, neither across all 19 study populations nor across the 15 populations that were in the largest cluster identified by GENELAND based on microsatellite alleles. This finding is congruent with the results of previous studies in the invaded range of *A. artemisiifolia* (Genton et al. 2005; van Boheemen et al. 2017). The spatial and non-spatial genetic structure analyses also did not reveal any obvious spatial genetic patterns among the Chinese populations. This result suggests that there is not much natural spread or gene flow between the populations included in this study. However, this possibility does not preclude such spread and gene flow happening at smaller spatial scales (the shortest distance between two populations in our study was 180.53 km), which were not considered in our study. Nevertheless, the result does suggest that most of our study populations are likely to have resulted from multiple long-distance introductions by humans, either from native populations or from other invasive populations. The production of many hard-coated, very small seeds with spines and their ability to form large seed banks in the soil, are likely to enable the species to spread over long distances by transport of, for example, soil, harvested grains and nursery stock.

Multiple introductions have been frequently reported for different invasive species (Dlugosch and Parker 2008; Pointier et al. 2008; Berbegal et al. 2013). Invasive populations of *A. artemisiifolia* in Western and Eastern Europe also came from multiple sources (Genton et al. 2005; Chun et al. 2010, 2011; Gladieux et al. 2011; Ciappetta et al. 2016; Meyer et al. 2017). Australian populations, in contrast, most likely are the result of a single introduction from the invaded range in Europe (van Boheemen et al. 2017). It is likely that multiple introductions from different sources helped *A. artemisiifolia* invade China, as it might overcome negative effects of bottlenecks and help to maintain a high level of genetic variation (Dlugosch and Parker 2008; Lockwood et al. 2007; Fischer et al. 2017). Allelic richness is considered to be a sensitive index for detecting founder effects as low levels could reflect the loss of rare alleles (Nei

et al. 1975; Allendorf 1986; Leberg 1992; Dlugosch and Parker 2008;). Therefore, the similar total numbers of alleles and numbers of private alleles in the Chinese and North American populations for the same microsatellite loci indicate that the genetic diversity in Chinese populations is similar to that in North American populations. Bottleneck analysis also indicated that most populations in China have not undergone genetic bottlenecks. Overall, our findings thus strongly suggest that *A. artemisiifolia* in China has been introduced from multiple sources.

Most *Ambrosia artemisiifolia* in China came from pre-admixed source populations

The results of our non-spatial (STRUCTURE) and spatial (GENELAND) structural analyses revealed genetic admixture in 15 of the 19 Chinese populations, which resulted in low genetic differentiation. The small number of private alleles that we found also shows that many of the alleles are shared among populations. Admixture and high genetic variation within populations (and low genetic differentiation among populations) may be caused by multiple introductions and/or high gene flow. The latter cause has been found for European populations of *A. artemisiifolia* (Chun et al. 2010). Recently, at a more local scale, Kropf et al. (2018) also found highly dynamic gene flow among *A. artemisiifolia* populations in Vienna, where the Danube River acts as a potential dispersal corridor. However, we detected only a small amount of gene flow among our study populations, especially among the northern populations. This suggests that low genetic differentiation among the Chinese populations is not the result of genetic exchange through seeds or pollen among these populations. Lack of genetic structure and low gene flow suggest that the Chinese populations most likely originated from admixed source populations. Previous studies found evidence for an admixture zone of *A. artemisiifolia* in its native range, and it has also been shown that European populations, which are most likely post-introduction admixed from various source populations, have been a bridgehead for subsequent introductions elsewhere (Martin et al. 2014; van Boheemen et al. 2017). Therefore, it is likely that the native North American admixture zone or the European populations were the sources of introduction for most populations of *A. artemisiifolia* in China.

However, the occurrence of additional introductions from other sources, such as non-native populations in Japan and Korea, cannot be excluded.

The potential importance of multiple introductions and genetic admixture for successful invasion has been intensively discussed (Keller and Taylor 2010; Keller et al. 2014; Rius and Darling 2014). Ideally the importance of admixture would be tested by linking measures of genetic diversity in the populations to their expansion rates, but such data are usually not available for natural populations. Admixture might directly increase fitness via heterosis, and might indirectly do so by creating novel gene and allele combinations, helping invaders adapt to their new environments and overcome founder effects. However, admixture can also dilute locally adapted genomes in the admixed offspring and lead to hybrid breakdown (Keller et al. 2000; Verhoeven et al. 2010). Our data do not allow us to assess whether admixture promoted the invasion success of *A. artemisiifolia* in China. However, if the benefits of pre-introduction admixture are similar to those of post-introduction admixture (van Boheemen et al. 2017), it is likely that individuals from admixed native populations of the species have enhanced invasion success in China.

## Conclusions

Overall, we found low levels of genetic differentiation among invasive populations of *A. artemisiifolia* in China. However, there was considerable genetic variation within populations, despite limited gene flow among the populations. Therefore, we conclude that *A. artemisiifolia* was most likely introduced multiple times from pre-admixed populations. As it is likely that this has contributed to the invasion success of the species in China, further introductions and admixture should be prevented.

**Acknowledgements** The authors would like to thank Dr. Michael D. Martin at the Centre for GeoGenetics, Natural History Museum of Denmark for providing haplotypes sequences from the 45 North American extant populations and historical *Ambrosia artemisiifolia* herbarium specimen.

**Funding** This work was financially supported by grants from The National Key Research and Development Program of China 2016YFC1201100, 2017YFC0506200, and The National Natural Science Foundation of China 41701026.

## References

- Allendorf FW (1986) Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biol* 5:181–190
- Berbegal M, Armengol J, Grünwald NJ (2013) Evidence for multiple introductions and clonality in Spanish populations of *Fusarium circinatum*. *Phytopathology* 103:851–861
- Blossey B, Notzold R (1995) Evolution of increased invasive competitive ability nonindigenous a hypothesis plants: a hypothesis. *J Ecol* 83:887–889
- Bossdorf O, Auge H, Lafuma L, Rogers WE, Siemann E, Prati D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* 144:1–11
- Cao LJ, Wang ZH, Gong YJ, Zhu L, Hoffmann AA, Wei SJ (2017) Low genetic diversity but strong population structure reflects multiple introductions of western flower thrips (Thysanoptera: *Thripidae*) into China followed by human-mediated spread. *Evol Appl* 10:391–401
- Chun YJ, Fumanal B, Laitung B, Bretagnolle F (2010) Gene flow and population admixture as the primary post-invasion processes in common ragweed (*Ambrosia artemisiifolia*) populations in France. *New Phytol* 185:1100–1107
- Chun YJ, Le Corre V, Bretagnolle F (2011) Adaptive divergence for a fitness-related trait among invasive *Ambrosia artemisiifolia* populations in France. *Mol Ecol* 20:1378–1388
- Ciappetta S, Ghiani A, Gilardelli F, Bonini M, Citterio S, Gentili R (2016) Invasion of *Ambrosia artemisiifolia* in Italy: assessment via analysis of genetic variability and herbarium data. *Flora* 223:106–113
- Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144:2001–2014
- Di Rienzo A, Peterson AC, Garzat JC, Valdes AM, Slatkint M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 91:3166–3170
- Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Mol Ecol* 17:431–449
- Dlugosch KM, Anderson SR, Braasch J, Cang FA, Gillette HD (2015) The devil is in the details: genetic variation in introduced populations and its contributions to invasion. *Mol Ecol* 24:2095–2111
- Dong H, Zhou M, Liu Z, Hao X, Liu Y, Abdulvai A, Liu T (2017) Diffusion and intrusion features of *Ambrosia*

- artemisiifolia* and *Ambrosia trifida* in Yili River Valley. *J Arid Land Res Environ* 11:175–180
- Earl DA (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361
- Ellstrand NC, Schierenbeck KA (2006) Hybridization as a stimulus for the evolution of invasiveness in plants? *Euphytica* 148:35–46
- Estoup A, Ravigné V, Hufbauer R, Vitalis R, Gautier M, Facon B (2016) Is there a genetic paradox of biological invasion? *Annu Rev Ecol Evol Syst* 47:51–72
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Excoffier L, Laval G, Schneider S (2006) ARLEQUIN version 3.1: a software for population genetic data analysis. Computational and molecular population genetics laboratory. Institute of Zoology, University of Berne, Bern
- Fischer ML, Salgado I, Beninde J et al (2017) Multiple founder effects are followed by range expansion and admixture during the invasion process of the raccoon (*Procyon lotor*) in Europe. *Divers Distrib* 23:409–420
- Flot JF (2010) SeqPHASE: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Mol Ecol Resour* 10:162–166
- Friedman J, Barrett SCH (2008) High outcrossing in the annual colonizing species *Ambrosia artemisiifolia* (Asteraceae). *Ann Bot* 101:1303–1309
- Gaudeul M, Giraud T, Kiss L, Shykoff JA (2011) Nuclear and chloroplast microsatellites show multiple introductions in the worldwide invasion history of common ragweed, *Ambrosia artemisiifolia*. *PLoS ONE* 6:e17658
- Gentili R, Montagnani C, Gilardelli F, Guarino MF, Citterio S (2017) Let native species take their course: *Ambrosia artemisiifolia* replacement during natural or “artificial” succession. *Acta Oecol* 82:32–40
- Genton BJ, Shykoff JA, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Mol Ecol* 14:4275–4285
- Gladieux P, Giraud T, Kiss L, Genton BJ, Jonot O, Shykoff JA (2011) Distinct invasion sources of common ragweed (*Ambrosia artemisiifolia*) in Eastern and Western Europe. *Biol Invasions* 13:933–944
- Groves RH, Burdon JJ (1986) Ecology of biological invasions: an Australian perspective. Australian Academy of Science, Canberra
- Guillot G, Santos F, Estoup A (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* 24:1406–1407
- Harrigan RJ, Mazza ME, Sorenson MD (2008) Computation vs cloning: evaluation of two methods for haplotype determination. *Mol Ecol Resour* 8:1239–1248
- Henry P, Le Lay G, Goudet J, Guisan A, Jahodová S, Besnard G (2009) Reduced genetic diversity, increased isolation and multiple introductions of invasive giant hogweed in the western Swiss Alps. *Mol Ecol* 18:2819–2831
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806
- Jin DP, Lee JH, Xu B, Choi BH (2016) Phylogeography of East Asian *Lespedeza buergeri* (Fabaceae) based on chloroplast and nuclear ribosomal DNA sequence variations. *J Plant Res* 129:793–805
- Karn E, Jasieniuk M (2017) Genetic diversity and structure of *Lolium perenne* ssp. *multiflorum* in California vineyards and orchards indicate potential for spread of herbicide resistance via gene flow. *Evol Appl* 10:616–629
- Keller SR, Taylor DR (2010) Genomic admixture increases fitness during a biological invasion. *J Evol Biol* 23:1720–1731
- Keller M, Kollmann J, Edwards PJ (2000) Genetic introgression from distant provenances reduces fitness in local weed populations. *J Appl Ecol* 37:647–659
- Keller SR, Fields PD, Berardi AE, Taylor DR (2014) Recent admixture generates heterozygosity-fitness correlations during the range expansion of an invading species. *J Evol Biol* 27:616–627
- Kropf M, Huppenberger AS, Karrer G (2018) Genetic structuring and diversity patterns along rivers—local invasion history of *Ambrosia artemisiifolia* (Asteraceae) along the Danube River in Vienna (Austria) shows non-linear pattern. *Weed Res* 58:131–140
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Ecol Resour* 33:1870–1874
- Lawalrée A (1955) Note complémentaire sur les *Ambrosia* adventices en Europe occidentale. *Bull Soc R Bot Belg* 87:207–208
- Leberg PL (1992) Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* 46:477–494
- Li XM, Liao WJ, Wolfe LM, Zhang DY (2012) No evolutionary shift in the mating system of North American *Ambrosia artemisiifolia* (Asteraceae) following its introduction to china. *PLoS ONE* 7:1–6
- Li XM, Zhang DY, Liao WJ (2015) The rhythmic expression of genes controlling flowering time in southern and northern populations of invasive *Ambrosia artemisiifolia*. *J Plant Ecol* 8:207–212
- Li Y, Stift M, Kleunen M (2018) Admixture increases performance of an invasive plant beyond first generation heterosis. *J Ecol* 106:1595–1606
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452
- Liu J, Chen H, Kowarik I, Zhang Y, Wang R (2012) Plant invasions in China: an emerging hot topic in invasion science. *NeoBiota* 15:27–51
- Lockwood JL, Hoopes MF, Marchetti MP (2007) Invasion ecology. Wiley, Oxford
- Makra L, Matyasovszky I, Hufnagel L, Tusnady G (2015) The history of ragweed in the world. *Appl Ecol Env Res* 13:489–512
- Martin MD, Chamecki M, Brush GS, Meneveau C, Parlange MB (2009) Pollen clumping and wind dispersal in an invasive angiosperm. *Am J Bot* 96:1703–1711
- Martin MD, Zimmer EA, Olsen MT, Foote AD, Gilbert MTP, Brush GS (2014) Herbarium specimens reveal a historical

- shift in phylogeographic structure of common ragweed during native range disturbance. *Mol Ecol* 23:1701–1716
- Martin MD, Olsen MT, Samaniego JA, Zimmer EA, Gilbert MTP (2016) The population genomic basis of geographic differentiation in North American common ragweed (*Ambrosia artemisiifolia* L.). *Ecol Evol* 6:3760–3771
- Meyer L, Causse R, Pernin F et al (2017) New gSSR and EST-SSR markers reveal high genetic diversity in the invasive plant *Ambrosia artemisiifolia* L. and can be transferred to other invasive *Ambrosia* species. *PLoS ONE* 12:e97
- Molecular Ecology Resources Primer Development Consortium (MERPDC) (2009) Permanent genetic resources added to Molecular Ecology Resources database 1 January 2009–30 April 2009. *Mol Ecol Resour* 9:1375–1429
- Montagnani C, Gentili R, Smith M, Guarino MF, Citterio S (2017) The worldwide spread, success, and impact of ragweed (*Ambrosia* spp.). *Crit Rev Plant Sci* 36:139–178
- Müller-Schärer H, Schaffner U, Steinger T (2004) Evolution in invasive plants: implications for biological control. *Trends Ecol Evol* 19:417–422
- Naegel RP, Tomlinson AJ, Hausbeck MK (2015) Evaluation of a diverse, worldwide collection of wild, cultivated, and landrace pepper (*Capsicum annuum*) for resistance to phytophthora fruit rot, genetic diversity, and population structure. *Phytopathology* 105:110–118
- Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10
- Pagad S, Genovesi P, Carnevali L, Schigel D, McGeoch MA (2018) Data descriptor: introducing the global register of introduced and invasive species. *Sci Data* 5:170202
- Pairon M, Petitpierre B, Campbell M et al (2010) Multiple introductions boosted genetic diversity in the invasive range of black cherry (*Prunus serotina*; Rosaceae). *Ann Bot* 105:881–890
- Parks DH, Mankowski T, Zangooei S et al (2013) GenGIS 2: geospatial analysis of traditional and genetic biodiversity, with new gradient algorithms and an extensible plugin framework. *PLoS ONE* 8:e69885
- Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in excel population genetic software for teaching and research. *Mol Ecol Resour* 6:288–295
- Peakall R, Smouse PE (2012) GenA1Ex 65: genetic analysis in excel population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Pointier J, Jarne P, Sarda V, David P, Spe I (2008) Report high genetic variance in life-history strategies within invasive populations by way of multiple introductions. *Curr Biol* 18:363–367
- Polzin T, Daneshmand SV (2003) On Steiner trees and minimum spanning trees in hypergraphs. *Oper Res Lett* 31:12–20
- Prentis PJ, Wilson JR, Dormontt EE, Richardson DM, Lowe AJ (2008) Adaptive evolution in invasive species. *Trends Plant Sci* 13:288–294
- Prisizer S (1960) *Adventív gymnövényeink terjedése*. Mezőgazdasági Kiadó Agricultural Publishing Ltd, Budapest
- Pritchard J, Wen X, Falush D (2010) Documentation for STRUCTURE software, version 2.3. University of Chicago, Chicago
- Pyšek P, Pergl J, Essl F et al (2017) Naturalized alien flora of the world: species diversity, taxonomic and phylogenetic patterns, geographic distribution and global hotspots of plant invasion. *Preslia* 89:203–274
- Rius M, Darling JA (2014) How important is intraspecific genetic admixture to the success of colonising populations? *Trends Ecol Evol* 29:233–242
- Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends Ecol Evol* 22:454–464
- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Resour* 4:137–138
- Schierenbeck KA, Ellstrand NC (2009) Hybridization and the evolution of invasiveness in plants and other organisms. *Biol Invasions* 11:1093–1105
- Shi J, Macel M, Tielbörger K, Verhoeven KJ (2018) Effects of admixture in native and invasive populations of *Lythrum salicaria*. *Biol Invasions* 20:2381–2393
- Signorile AL, Lurz PWW, Wang J, Reuman DC, Carbone C (2016) Mixture or mosaic? Genetic patterns in UK grey squirrels support a human-mediated “long-jump” invasion mechanism. *Divers Distrib* 22:566–577
- Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 73:1162–1169
- Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68:978–989
- The People’s Government of Meizhou (2016) Meizhou overview. [https://www.meizhou.gov.cn/zjzmz/mzgz/mzgzk1/t20160825\\_130.htm](https://www.meizhou.gov.cn/zjzmz/mzgz/mzgzk1/t20160825_130.htm). Accessed 12 Dec 2018
- Tsutsui K, Suwa A, Sawada KI, Kato T, Ohsawa TA, Watano Y (2009) Incongruence among mitochondrial, chloroplast and nuclear gene trees in *Pinus* subgenus *Strobos* (Pinaceae). *J Plant Res* 122:509–521
- van Boheemen LA, Lombaert E, Nurkowski KA et al (2017) Multiple introductions, admixture and bridgehead invasion characterize the introduction history of *Ambrosia artemisiifolia* in Europe and Australia. *Mol Ecol* 26:5421–5434
- van Kleunen M, Bossdorf O, Dawson W (2018) The ecology and evolution of alien plants. *Annu Rev Ecol Evol Syst* 49:25–47
- Verhoeven KJ, Macel M, Wolfe LM, Biere A (2010) Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc R Soc B* 278:2–8
- Wan FH, Guo JY, Zhang F (2009) *Research on biological invasions in China*. Science Press, Beijing
- Wan F, Jiang M, Zhan A (2017) *Biological invasions and its management in china*, vol 2. Springer, Singapore
- Wang MY (2005) Current situation of common ragweed and control strategy. *J Anhui Agric Sci* 33:1771–1786
- Wang J, Wu Y, Ren G, Guo Q, Liu J, Lascoux M (2011) Genetic differentiation and delimitation between ecologically diverged *Populus euphratica* and *P. pruinosa*. *PLoS ONE* 6:e26530
- Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163:1177–1191

- Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci USA* 84:9054–9058
- Zhou ZS, Guo JY, Wan FH (2015) *Ambrosia artemisiifolia* L. In: Wan FH, Hou YM, Jiang MX (eds) *Invasion biology*. Science Press, Beijing, pp 182–185