Genetic diversity and population structure of invasive and native populations of *Erigeron canadensis* L.

Saurav Bhattacharya[1], Fernando Hernández[2], Mariana Ferreira Alves[3], Raquel Moura Machado[3], Yan-Yan Sun[1], Meng-Ru Wang[1], Chao-Bin Zhang[4], Jian-Hua Hao[1*]

[1]School of Biology and Food Engineering, Changshu Institute of Technology, Nansanhuanlu 99, Changshu- 215500, Suzhou, Jiangsu Province, China. [2]Departamento de Agronomía, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina. [3]Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, São Paulo, Brazil. [4] Shandong Academy of Agricultural Sciences, Jinan, Shandong Province, China.

These authors contributed equally to the manuscript: Saurav Bhattacharya, Fernando Hernández, Mariana Ferreira Alves, Raquel Moura Machado, Yan-Yan Su，Meng-Ru Wang

*Correspondence author: Jian-Hua Hao

Addres: Changshu Institute of Technology, Nansanhuanlu 99, Changshu- 215500, Suzhou, Jiangsu Province, China.

Tel: 15895609563

Fax: 052-52251562

E-mail: jhhao@cslg.edu.cn, jhhaocs@163.com

This article has been accepted For publication And undergone full peer review but has Not been through the copyediting, typesetting, pagination And proofreading process, which may lead to differences between this

Version And the Version of Record. Please cite this article as doi: 10.1093/jpe/rtac016.

A**bstract**

***Aims*:** Invasive alien plants threaten biodiversity across the world. *Erigeron canadensis* (horseweed) is considered one of the most problematic agricultural weeds and represents a classic example of inter-continental invasion. Here, we studied the genetic diversity and population structure of invasive alien populations from the Jiangsu and Zhejiang Provinces in China and native populations from Alabama, in the USA.

***Methods*:** We used ten polymorphic SSR loci to genotype 312 individuals from five native and five invasive populations to estimate the genetic diversity and structure.

***Important Findings*:** Invasive populations from Jiangsu and Zhejiang Provinces showed, on average, similar genetic diversity to native populations from Alabama, indicating no severe genetic bottlenecks during the invasion. STRUCTURE revealed low population differentiation and only two genetic groupings were detected in both native and invaded ranges. The high diversity observed in the invasive populations suggested multiple introductions and/or the introduction of genetically diverse propagules during initial colonization. Our study provides new insights towards understanding the invasion dynamics of this globally noxious weed in Eastern China. Preventing gene flow via seed dispersal between invasive and native populations should be examined to prevent the introduction and dispersal of herbicide-resistant individuals and inform management practices.

Keywords: Gene flow, Genetic diversity, Invasive alien plant, Introduced population, Native population

Introduction

Genetic and evolutionary processes are important driving forces that influence the successful colonization and spread of invasive alien species (Barrett 2015; Ellstrand and Schierenbeck 2000; van Kleunen *et al*. 2018). Invasive alien plants often pose a major threat due to their growth and high competitiveness that affects the ecology and agricultural productivity in introduced areas (Mack *et al.* 2000; Pyšek *et al.* 2020). The invasion success of introduced populations is generally characterized by adaptation to novel environments and the rapid selection of individuals with greater reproductive fitness and increased phenotypic plasticity driven by underlying genetic variation (Geng *et al.* 2007; Barrett *et al.* 2008; Colautti and Lau 2015).

During introductions and range expansions, it is expected that invasive populations experience demographic bottlenecks, which can negatively impact their genetic variation and, therefore its evolutionary potential (Dlugosch and Parker 2008; Dlugosch *et al.* 2015). There are reports of successful alien plants with lower genetic diversity compared to native populations for species such as *Eichhornia crassipes* (Zhang *et al.* 2010), *Ceratocapnos claviculata* (Voss *et al.* 2012), *Impatiens glandulifera* (Hagenblad *et al.* 2015), *Solanum rostratum* (Zhao *et al.* 2013), and *Senecio vulgaris* (Zhu *et al.* 2017). However, similar or increased genetic diversity of invasive populations has been revealed for species such as *Bromus tectorum* (Novak and Mack 1993), *Ambrosia artemisiifolia* (Genton *et al.* 2005), *Phalaris arundinacea* (Lavergne and Molofsky 2007), *Solidago canadensis* (Zhao *et al.* 2014), *Bunias orientalis* (Patamsytė *et al.* 2018), and *Helianthus annuus* L. (Hernández *et al*. 2019b). Therefore, comparative studies of genetic diversity and population structure of invasive and native populations are desirable to evaluate the contribution of stochastic and deterministic

pressures to the strength and patterns of genetic diversity in invasive species (Zhu *et al*. 2017). Moreover, studies addressing invasions at the intraspecific level have recently been recognized as providing novel insight into the mechanisms underlying plant invasiveness (Pysek *et al*. 2020).

Previous studies have indicated that population genetic structure in non-native plants changes dramatically over the course of invasion and is strongly influenced by genetic admixture as well as mating system, dispersal ability and introduction history (Keller *et al*. 2012; Zhu *et al.* 2017). In seed plants, genetic structure is closely related to two components of gene flow: pollen dispersal dynamics, and biotic or abiotic seed dispersal. Predominant self-fertilization or clonal propagation can limit or entirely prevent opportunities for the admixture of genetically differentiated founder stocks (Barrett *et al.* 2008). Uniparental reproduction reduces gene dispersal via pollen, which, in outcrossing populations, can be more extensive than gene flow via seed (Barrett *et al.* 2008).

The genus *Erigeron* sect. *Conyza* is one of the foremost representatives of inter-continental plant invasion from the New World to the Old World (Thébaud and Abbott 1995; Galkina and Vinogradova 2020). Within the genus, *E. canadensis* (*C. canadensis*), *E. sumatrensis* (*C. sumatrensis*) and *E. bonariensis* (*C. bonariensis*) are the most common taxa and are now widespread throughout the world. *E. canadensis* L. (common name horseweed) is a diploid (2n= 2x= 18) winter or summer annual and among the top ten most invasive species in Europe (Galkina and Vinogradova 2020). It was the first agricultural weed with a reference genome and the first eudicot weed that evolved glyphosate resistance globally (Weaver 2001; Davis *et al.* 2010; Okada *et al.* 2013, 2015; Page *et al.* 2018). In China, *E. canadensis* was first recorded in the 1860s in the Yantai, Shandong Province, and is now

listed as one of the 10 most widespread species growing in disturbed ruderal sites, and is distributed across 28 provinces (Hao *et al.* 2011). It has spread extensively across Eastern and Northern China, where the climate and latitude are similar or the same as the natural habitat of *E. canadensis* in its native range (Wang *et al.* 2017; Rosche *et al.* 2019). The species is highly self-compatible (Weaver 2001; Hao *et al.* 2011; Rosche *et al.* 2019), characterized by large numbers of small, wind-dispersed seeds (ranging to over 200 000 seeds per plant), long-distance dispersal (Weaver 2001; Shields *et al.* 2006; Dauer *et al.* 2007; Liu *et al.* 2018), and widespread distribution along climatic gradients (Wang *et al.* 2017; Rosche *et al.* 2019).

Empirical studies have previously explored gene flow and genetic divergence between or among *Erigeron* (*Conyza)* species based on isozyme polymorphism (Thébaud and Abbott 1995; Soares *et al.* 2015) or SSR molecular markers (Okada et al. 2013, 2015; Marochio *et al.* 2017; Rosche *et al.* 2019). High genetic diversity in *E. canadensis* was reported by Okada et al. (2013; 2015) with populations sampled across California. A study on three *Erigeron* species from Southern Brazil showed that *E. canadensis* harbors lower genetic diversity compared to its congeners, *E. sumetrensis* and *E. bonariensis* in terms of number of alleles and expected heterozygosity (Marochio *et al*. 2017). A recent study of genetic diversity, genetic differentiation, and performance between native and nonnative populations across a large biogeographical and climatic gradient in *E. canadensis* showed overall low genetic diversity and high selfing rates for the species, with no differences between native and invaded ranges (Rosche *et al.* 2019). However, there has been no report on genetic variation of populations from lower latitudes (i.e. lower than latitude 35.906) in the native range of *E. canadensis,* which covers much of North America,

including the entire USA (Okada et al. 2013, 2015; Marochio *et al.* 2017). Moreover, climate matching (i.e. long-distance dispersal between climatically similar regions) is recognized as one of the major invasion drivers in large-scale range expansion during biological invasion (Barrett et al. 2008; van Kleunen et al. 2018; Liu et al. 2020) yet no report has specifically focused on a comparison of genetic structure of native populations with that of latitude/climate-matched invasive populations in *E. canadensis*. In this study, we used microsatellite markers to examine intraspecific genetic variation in *E. canadensis* by comparing invasive populations from Eastern China (Jiangsu and Zhejiang Provinces) with native populations from Southeast USA (Alabama). These areas have similar latitude, climate, and *E. canadensis* populations are very common and widespread. In this study we specifically aim to: (1) study genetic diversity and structure between populations from native and invaded areas with similar latitude and climate; and (2) investigate whether genetic patterns and structure of invasive populations in Jiangsu and Zhejiang Provinces have been shaped by demographic factors, such as genetic bottleneck or admixture. The results from our study provide insight towards understanding the invasion dynamics of this globally noxious weed in Jiangsu and Zhejiang Provinces in China, and to better understand dispersal patterns of this species within the invaded area to inform management practices.

Materials and Methods

Study locations and sample collection – In this study we examined 10 populations of *E. canadensis*: five native populations from Alabama (AL, USA) and five invasive populations from Jiangsu and Zhejiang Provinces (China). Among five populations for each region, two

populations were in the north (HV and BH for native, and XY and JH for invasive), one in the center (COS for native and JY for invasive), and two in the south (TR and EP for native, CS and HY for invasive), respectively. The two regions have similar geographic latitudes and humid subtropical climate (Fig. 1, Supplementary Table S1). All studied populations were collected from disturbed habitats and comprised more than 300 individuals. Populations sampled within each region were separated by a minimum of 100 km. The phenotypic performance (plant height, local abundance, and flowering period) is similar in both ranges except for the EP population in the native region (personal observation). The individuals of the EP population are relatively shorter and its distribution is sparser when compared to the other populations.

For each population we randomly selected across an area that was approximately 20×300 m, 20×400 m, or 40×200 m in size and sampled leaf tissue from 35 to 40 individuals from the end of July to early September of 2016. About 10-20 young healthy leaves were collected from each individual and transferred to mesh bags containing dry silica gel. The dry leaf samples were then stored at −18°C until used for DNA extraction. *E. canadensis* is an annual species and reproduces only by seeds, so individuals were sampled at least >10 m apart to reduce the possibility of sampling related (and therefore more genetically similar) samples within a kinship deme. However, some samples failed in DNA extraction and SSR amplification and 24-36 individuals per population were used in the final analysis.

DNA extraction and SSR amplification - For the genetic diversity study, genomic DNA from individual samples was extracted from dry leaf samples using the TIANGEN plant genomic DNA extraction kit (Tiangen Biotech, Beijing, China) following the manufacturer's protocol. Amplification reactions for each sample were carried out in a total volume of 20 µl in *Vapo*

Protect Mastercycler Pro (*Eppendorf*) containing 20 ng template DNA, 1X PCR buffer (*TaKaRa, Takara Bio, Dalian, China)*, 10 mM dNTP mix (*TaKaRa*), and 1 U Taq DNA polymerase (*TaKaRa)*. Each reaction also contained 0.8 µM of reverse and M13 primer labelled with one fluorescent dye (6-FAM/ROX/JOE/TAMRA) (Sangon Biotech, Shanghai, China) along with 0.2 µM of forward primer containing M13 tail at its 5' end (Supplementary Table 2) (Schuelke 2000). Conditions of the PCR amplification were as follows: 94°C (5 min), then 30 cycles at 94°C (50 s) / 55°C (45 s)/72°C (1 min), followed by 15 cycles at 94°C (50 s)/ 52°C (45 s)/ 72°C (1 min), and a final extension at 72°C for 10 mins.

We conducted a preliminary screening with 12 pairs of microsatellite primers previously reported for *E. canadensis* (Wang *et al.* 2009) with 24 samples randomly chosen from the 10 populations. Primer pairs that generated clear, well defined amplified products were sent to a company (Genesky Bio-Tech Co. Ltd., Shanghai, China) and analysed in an ABI3730xl fragment analyser (Applied Biosystem, Foster City, California, USA) using LIZ-500 as the internal standard. For each individual, the amplified fragment sizes or alleles were coded either as homozygous (having two equally sized alleles), or heterozygous (having two different sized alleles). All SSR allele sizing and scoring was performed using Gene-Mapper ver. 4.0 (Applied Biosystems). To analyze the polymorphic loci, polymorphic information content (*PIC*) was estimated using Cervus ver. 3.07 (Kalinowski *et al.* 2007). Finally, ten microsatellite primers with clear, consistent peaks and with a high PIC value (> 0.5) were selected for the genetic diversity study (see, below, Supplementary Table S2).

Genetic diversity analysis - We estimated the genetic diversity for each population-locus combination in native and invaded regions using GenAlEx 6.4 (Peakall and Smouse 2012) with the following indices: number of alleles (N_A), number of effective alleles (N_E), private

alleles (P_A), expected heterozygosity (H_E), and observed heterozygosity (H_O). We used FSTAT 2.9 (Goudet 1995) to calculate allelic richness (A_R) and inbreeding coefficient (F_{IS}) for each population averaged across all 10 loci and the significance was tested with 10000 randomization using Bonferroni corrections. Selfing rate (s) was estimated as 2 F_{IS} /(1+F_{IS}), following Okada *et al*. (2013). In addition, to test the overall genetic difference between native and invaded regions, we compared the diversity parameters A_R, H_O, H_E, and F_{IS} values with an independent *t*-test using SPSS 16.0 (IBM SPSS Statistics).

Genetic structure and gene flow analysis - We performed an analysis of molecular variance (AMOVA) to find out how much genetic diversity was partitioned within and among populations and between regions. Genetic differentiation (F_{ST}) and gene flow from pair-wise migrants per generation (Nm) in the invasive and native populations were also calculated. To examine whether there was any isolation-by-distance pattern, we used the Mantel test to examine the correlation between genetic (pairwise F_{ST}) and geographic distances. Patterns of isolation by distance in the native and invaded regions were separately tested based on 999 permutations. Principal coordinate analysis (PCoA) was performed from genetic distances calculated for individuals and populations. All the above analyses were performed using GenAlEx 6.5. A dendrogram based on an unweighted pair-group method with arithmetic averages (UPGMA) clustering method was constructed using Nei's genetic similarity matrix for all 10 populations. The cluster analysis was performed using XLSTAT 2018 (https://www.xlstat.com/). To further analyze the genetic structure, we used a Bayesian clustering based software, STRUCTURE v.2.3.4 (Pritchard *et al.* 2000). Twenty independent runs for each K value ranging from 1 to 10 were performed using the allele data matrix of all 10 native and invasive populations with a burn-in length of 200,000

followed by 500,000 MCMC iterations. The admixture model was applied with correlated allele frequency (Falush *et al.* 2003). We also performed the STRUCTURE analysis separately for the native and invasive populations, where K values from 1 to 10 were tested with a burn-in length of 200,000 followed by 500,000 MCMC iterations and replicated 20 times. The value of K was estimated using STRUCTURE HARVESTER and CLUMPP in a web-based program, CLUMPAK server (Kopelman et al. 2015). We inspected plots for each value of K following the approach by Janes et al. (2017). The final graphs were viewed using a web-based program, CLUMPAK server (Kopelman et al. 2015).

Results

Genetic diversity in invasive alien and native populations

Ten microsatellite primers that generated high PIC values (0.517 - 0.881) were used successfully to genotype 312 individuals from the 10 native and invasive populations. From the ten microsatellite loci we detected a total of 655 alleles among the *E. canadensis* populations (Table 1). The number of alleles per population ranged from 52 to 84 (Table 1). The average number of alleles detected per locus ranged from 5.20 to 7.50 in the invaded region and from 5.30 to 8.40 in the native region (Table 1). The number of effective alleles ranged from 2.14 to 3.68 in the invaded region and 2.42 to 4.65 in the native regions, and the number of private alleles from 5 to 14 in the invaded region and 5 to 21 in the native region (Table 1). The invasive population with the highest number of private alleles was XY (P_A= 14), while the native populations HV, BH, and COS, showed the highest number of private alleles as well as greater allelic richness (Table 1). As expected for self-fertilizing

species, observed heterozygosity *(Ho)* showed overall low values (*Ho* < 0.1 in eight out of 10 populations), but two native populations (EP and HV) showed unusually higher values (Table 1). The inbreeding coefficient (F_{IS}) was high and ranged from 0.87 to 0.94 in the introduced area and 0.73 to 0.90 in the native region (Table 1). Selfing rate (*s*) ranged from 0.88 to 0.96 in the invaded region and from 0.81 to 0.95 in the native region (Table 1) and there was no significant difference between ranges (t= 0.848, df= 8, p= 0.421). The t-test showed no significant difference between native and invaded regions for expected heterozygosity H_E (t= -1.417, df= 8, p= 0.194), allelic richness A_R (t=-1.153, df=8, p= 0.282), and inbreeding rates F_{IS} (t= 1.650, df= 5.616, p= 0.153) (Supplementary Table S3).

Genetic structure and gene flow in invasive alien and native populations

AMOVA indicated significant genetic structure in invaded (F_{ST}= 0.169; P < 0.01) and native areas (F_{ST} = 0.146; P < 0.01) (Table 2). Overall, 76% of genetic diversity was distributed within populations while 14% was distributed among populations within regions (China and USA) (Table 2). Only a small portion (10%) of genetic diversity was explained by differences between the two regions, indicating narrow genetic differentiation between the two continental regions (Table 2). Pairwise F_{ST} between populations ranged from 0.012 (CS-JY) to 0.48 (CS-TR) (Table 3). The mean F_{ST} was slightly higher in the invaded than in the native region (0.204 vs. 0.173) but this difference was not statistically significant (Supplementary Table S4). Mantel test revealed no significant correlations between spatial and genetic distances either in the invaded (R_{XY} = 0.53; *P* = 0.117) or the native region (R_{XY} = 0.42; *P* = 0.109) (Fig. S1). .

Two independent PCoAs were performed, the first to analyze the genetic relationship among individuals (using all samples from both invaded and native regions)

and the second to analyze the mean genetic differences among populations. In the first PCoA, we observed overall low population structure (Fig. 2A), especially within regions. PCoA 1 explained 11.8% of the total variation and separated all the native populations from the Southern and Central invasive populations (JY, CS, and HY; Fig. 2A, 2B). PCoA 2 explained 7.1% of the total variation and separated native populations from the northern invasive populations (XY and JH; Fig. 2A, 2B). A similar pattern was observed when individuals were assigned to populations. Native and invasive populations were well separated between each other (Fig. 2C). For invasive populations, two genetic clusters were found (Fig. 2C), one with three Central and Southern populations (JY, CS, HY; Fig. 2C) and the other with the two Northern populations (XY and JH; Fig. 2C). Similar results were obtained in the UPGMA dendrogram where the cluster analysis divided the ten populations into two major clusters corresponding to their native or invaded regions (Fig. 3). STRUCTURE identified K= 2 when all individuals from invaded and native regions were analyzed together (Fig. 4A, Supplementary Fig. S2). The invasive and native populations were each represented by a single major genetic component, which agrees with the PCoA and UPGMA analyses. When we analyzed invasive and native individuals separately, STRUCTURE identified K= 2 in both invasive (Fig. 4B, Supplementary Fig. S3) and native populations (Fig. 4C, Supplementary Fig. S4). At K= 2, invasive individuals were clearly distinguished into two groups with a clear geographic pattern: Northern populations (XY and JH) were separated from the Central and Southern populations (JY, CS, and HY). The HY populations had individuals from both genetic clusters, and all populations contained admixed individuals (Fig. 4B). In the native populations, individuals were also assigned to two clusters (Fig. 4C) with some level of admixture between individuals in the BH, COS, and TR population, while in the EP population there were low levels of admixture.

The estimated gene flow across population pairs indicated moderate to high gene flow existing between invasive (overall N_m= 3.435) and native (overall N_m= 1.537) populations. No significant difference (t= 1.31, df=18, p= 0.204) was detected for migrants (N_M) among pairwise populations in the native and introduced regions (Supplementary Table S4). A high level of genetic exchange was detected between invasive populations: XY-JH (Nm= 2.699), HY-JY (Nm= 2.488), HY-CS (Nm= 2.791) and CS-JY (Nm= 20.795). Similarly high genetic exchange was found across native populations: EP-TR (Nm= 2.337) and BH-EP (Nm= 2.157) (Table 4).

Discussion

Comparative studies utilizing genetic markers across native and invaded regions provide the necessary information for understanding the amount of genetic variation introduced to the invaded range, and the major demographic factors underlying the invasion process (Zhu *et al.* 2017). In this study we used molecular markers to compare the genetic diversity and population structure of native and invasive populations of *E. canadensis*.

In our study, we found overall high genetic diversity in both native and introduced ranges. High genetic diversity in *E. canadensis* has been previously reported by Okada *et al.* (2013; 2015) with populations sampled across native populations of California. In contrast, Rosche *et al*. (2019) reported overall low genetic diversity for this species, but also found that genetic diversity largely varied geographically, with some populations from either the native or the invaded ranges showing high genetic diversity. Therefore, it is not surprising that both native and invasive populations from our study harbor high genetic diversity.

The breeding system is a main determinant of patterns of genetic variation in plants, and self-pollinating species typically have reduced genome-wide diversity, with a tendency to be more homogeneous (Nybom 2004; Glémin *et al.* 2006; Barrett 2010). *E. canadensis* is self-pollinating, with 1 to 4% outcrossing occurrence (Davis *et al*. 2010). However, selfing rates estimated in our study suggest that outcrossing may be more frequent (4% to 19%). In line with this, we observed admixed individuals in every population, suggesting that admixture is common, both in the native and invaded regions. Admixture may catalyze invasiveness by increasing the genetic diversity of the invaders, creating novel genetic variants on which selection can act, and generating heterozygosity-fitness correlations (Ellstrand and Schierenbeck 2000; Szücs *et al*. 2017; van Kleunen *et al*. 2018).

Demographic bottlenecks are often associated with unintentional introductions, but many other processes may help to counteract genetic bottlenecks, such as multiple introductions, admixture, and high propagule pressure (Dlugosch and Parker 2008; Keller *et al*. 2012; Dlugosch *et al*. 2015; Hernández *et al*. 2019b). Here, we found no evidence of genetic bottlenecks during introduction, as native and invasive populations showed similar (and high) genetic diversity. Similar results have been observed in other studies comparing native and invasive populations in a diverse range of taxa, e.g., *Ambrosia artemisiifolia* (Genton *et al.* 2005) and *Helianthus annuus* (Hernández *et al*. 2019b). In *E. canadensis,* Rosche *et al*. (2019), found no differences in genetic diversity between native and invasive populations, but large differences within both native and invaded ranges.

Population genetic tools are useful to identify the major processes underlying genetic diversity and structure of invasive species. Overall, we observed low population differentiation in both native and introduced ranges. In the native range, F_{ST} ranged from

0.097 to 0.198 (Table 3), which were much lower than those previously reported (Okada *et al*. 2013; 2015; Rosche *et al*. 2019). In the invaded range, pairwise F_{ST} values ranged from 0.012-0.262, and were very low within both the Southern and Northern regions, but higher between them (Table 3). This pattern of F_{ST} variation is often associated with multiple introductions, with genetic differences between clusters resulting from historical population structure in the native range rather than to events associated with invasions (Hernández *et al*. 2019b; Keller *et al*. 2012). Multiple introductions may not only rescue invasive populations from genetic bottlenecks by increasing the founding size (demographic effect), but also can generate novel variability if introduced genetic clusters hybridize in the invaded range (Keller *et al*. 2012; Dlugosch *et al*. 2015; Hernández *et al*. 2019b). We observed admixed individuals in every invasive population, but whether admixture occurred in the invaded range or admixed individuals were introduced cannot be answered from our study. A previous study with populations from the native range found that outcrossing occurs more frequently than previously realized in *E. canadensis* and suggested that admixture may play a role in creating novel variability and enhancing herbicide resistance (Okada *et al.* 2013).

Along with multiple introductions, other processes such as propagule pressure (the simultaneous introduction of many individuals), adaptive divergence, and genetic drift may affect the genetic diversity and structure of invasive populations (Dlugosch and Parker 2008; Keller *et al*. 2012; Dlugosch *et al*. 2015). Propagule pressure (including number of colonists and multiple introductions) is one of the best predictors of invasion success (Lockwood *et al*. 2005). High propagule pressure can alleviate genetic bottlenecks and drive a successful invasion (Szücs *et al*. 2017). The high diversity observed in the invasive populations from our

study (especially in XY and JH) suggests multiple introductions and/or the introduction of genetically diverse propagules during initial colonization. In addition, natural selection can drive adaptive divergence post-introduction (Xu *et al*. 2010). Page *et al*. (2018) observed that glyphosate-resistance largely explains genetic clustering of *E. canadensis* populations from Ontario, demonstrating that local selection can rapidly drive population divergence. Further studies are needed to test whether herbicide-resistance, or other factors such as patterns of disturbance and cropping systems, are playing a role in population divergence in Eastern China.

Another process that should be considered is gene flow via seed dispersal. Seed dispersal is the major mechanism by which *E. canadensis* spreads (Wang *et al*. 2017). Previous studies showed that seeds of *E. canadensis* may travel as far as 550 km (Shields *et al.* 2006; Liu *et al.* 2018). In fact, at the end of the seed-shedding season, seeds were found to spread further than 186 km under field conditions in Illinois (Liu *et al.* 2018). In addition, human travel and transport may disperse seeds across larger areas than any natural dispersal vector as found in many invasive species (Pandolfo *et al*. 2018; Hernández *et al*. 2019b). Admixture via seeds may play a role in creating novel variability and spreading herbicide resistance alleles (Okada *et al.* 2013). Therefore, gene flow via seed dispersal between invasive populations should be examined to prevent the introduction and dispersal of herbicide-resistant individuals and to inform management practices. Further studies investigating the presence, frequency, and geographic distribution of herbicide resistance in the eastern China are needed.

The Yangtze River is a natural geographical barrier between the JY and CS populations in the invaded region. Previous studies have shown that it plays a major role in

the distribution of genetic diversity of plant species endemic to China, such as *Parrotia subaequalis* (Geng *et al*. 2015) and *Vitex negundo* (Zhang *et al*. 2007). However, to our knowledge, the Yangtze River has not been previously shown to be a major biogeographical barrier for invasive species. In our study, we found that invasive populations belonged to two major genetic clusters (the Northern: XY and JH, and the Southern and Central populations: HY, CS, and JY). High levels of genetic exchange were detected between Southern and Central populations, for example between JY-HY (Nm= 2.488) and JY-CS (Nm= 20.795). Migration across the Yangtze riversides has been demonstrated for endemic plant species (Zhang *et al*. 2007; Geng *et al*. 2015), indicating that the Yangtze River limits but does not prevent gene flow. Considering the highly developed transport network among these areas, human-mediated transport of seeds is likely to be the reason for genetic migration across the Yangtze River for *E. canadensis*.

We found admixed individuals in all of the *E. canadensis* populations sampled (Fig. 4A and 4B). Once distinct genetic clusters come into contact, admixture can produce genetically diverse weeds with high evolutionary potential (Keller *et al*. 2014; van Boheemen *et al*. 2017). The presence of admixed (*i.e.* heterozygous) individuals within populations could be the result of long-distance pollen flow, or seed dispersal. The species is highly self-compatible (Weaver 2001; Hao *et al.* 2011; Rosche *et al.* 2019), and uniparental reproduction reduces gene dispersal via pollen, which for outcrossing populations can be more extensive than gene flow via seed (Barrett *et al.* 2008). Given the major mechanism of spreading for *E. canadensis* is via seeds, this is likely the most plausible reason for the presence of admixed individuals within the population, resulting from cross-pollination between incoming individuals (germinating from seeds from other regions) and individuals

already present in the population. Unfortunately, the microsatellites we used in present study can only identify admixed individuals, and cannot distinguish whether admixed individuals are the result of seed or pollen-mediated gene flow. Further studies using genome-wide markers (e.g., from genotyping by sequencing) and non-nuclear markers (mitochondrial and chloroplast) are needed to evaluate the relative contribution of seed and pollen-mediated gene flow in *E. canadensis* and to better understand the invasion history of this weed in the eastern China.

Conclusions

In our study, we found no evidence of genetic bottlenecks associated with invasive *E. canadensis* in Jiangsu and Zhejiang Provinces in China. We provide evidence supporting multiple introductions of *E. canadensis* into eastern China. Our findings suggest that subsequent gene flow and admixture between differentiated clusters has generated genetic diversity similar to that in native range populations.. Seed and pollen-mediated gene flow between invasive populations of *E. canadensis* should be examined in detail to aid in controlling this noxious weed, and to prevent the introduction and dispersal of other herbicide-resistant weeds.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant No. 31370548), the Natural Science Foundation of Jiangsu Province, China (Grant No. BK20131192), the Talented Young Scientist Program, China (TYSP, India-17-001), and the Science and Technology Research Program of Animal and Plant Quarantine Agency of the People's Republic of China (Grant No. JSCIQ_2014IK013). We thank Dr. Prof. Bao-Rong LU and Dr Zhe WANG of Fudan University for designing experiments, guiding the method of DNA extraction and SSR amplification, and data analysis. No conflicts of interest have been declared.

Reference

Barrett SCH, Colautti RI, Eckert CG (2008) Plant reproductive systems and evolution during biological invasion. Mol Ecol **17**:373-383.

Barrett SCH (2010) Darwin's legacy: the forms, function and sexual diversity of flowers. Phil Trans R Soc B **365**: 351-368.

Barrett SCH (2015). Foundations of invasion genetics: the Baker and Stebbins legacy. Mol Ecol **24**: 1927–1941.

Blair AC, Wolfe LM (2004). The evolution of an invasive plant: an experimental study with *Silene latifolia*. Ecology 85: 3035-3042

Colautti RI, Lau JA (2015) Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. Mol Ecol **24**:1999-2017.

Dar TUH, Bhat BA, Khuroo AA, et al. (2020) Genetic diversity and population structure of an invasive plant species differ in two non-native regions with differing climate and invasion success. Nord. J. Bot. 38:1-9.

Dauer TJ, Mortensen AD, Vangessel JM (2007) Temporal and spatial dynamics of long-distance *Conyza canadensis* seed dispersal. J Appl Ecol **44**:105-114.

Davis VM, Kruger GR, Hallett SG, *et al.* (2010). Heritability of glyphosate resistance in Indiana horseweed (*Conyza canadensis*) populations. Weed Sci. 58: 30-38.

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, *et al.* (2015) The devil is in the details: genetic variation in introduced populations and its contributions to invasion. Mol Ecol **24**: 2095–2111.

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci **97**:7043-7050.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics **164**:1567-1587.

Galkina MA, Vinogradova YK (2020). Invasive species of *Erigeron* sect. *Conyza* in the Mediterranean and their hybridogenic activity. Biol Bull Russ Acad Sci **47**: 40–48.

Geng Q, Yao Z, Yang J, *et al*. (2015). Effect of Yangtze River on population genetic structure of the relict plant *Parrotia subaequalis* in eastern China. Ecology and Evolution **5:** 4617–4627.

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.

Glémin S, Bazin E, Charlesworth D (2006) Impact of mating systems on patterns of sequence polymorphism in flowering plants. P Roy Soc Biol Sci **273**:3011-3019.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-Statisitics. J Hered **86**: 485-486.

Hagenblad J, Hülskötter J, Acharya KP, *et al.* (2015) Low genetic diversity despite multiple introductions of the invasive plant species *Impatiens glandulifera* in Europe. BMC Genetics **16:**1-6.

Hao J, Qiang S, Chrobock T, *et al.* (2011) A test of Baker's Law: breeding systems of invasive species of Asteraceae in China. Biol Invasions **13**:571-580.

Hernández F, Poverene M, Garayalde A, *et al.* (2019a) Re-establishment of latitudinal clines and local adaptation within the invaded area suggest rapid evolution of seed traits in Argentinean sunflower (*Helianthus annuus* L.). Biol Invasions **8**: 2599-2612.

Hernández F, Presotto A, Poverene M, *et al.* (2019b) Genetic diversity and population structure of wild sunflower (*Helianthus annuus* L.) in Argentina: reconstructing its invasion history. J Hered **6**: 746-759.

Janes JK, Miller JM, Dupuis JR, *et al.* (2017) The K = 2 conundrum. Mol Ecol. **26**:3594-3602.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol **16**:1099-1106.

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.

Keller SR, Gilbert KJ, Fields PD, *et al.* (2012) Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. Mol Ecol **21:**4721–4734.

Kopelman NM, Mayzel J, Jakobsson M, *et al.* (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour **15**:1179-91.

Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. Proc Natl Acad Sci USA **104:** 3883–3888

Liu J, Qi M, Wang J. (2018) Long-distance and dynamic seed dispersal from horseweed (*Conyza canadensis*). Écoscience 10.1080/11956860.2018.1455371

Liu W, Zhang Y, Chen X, *et al.* (2020). Contrasting plant adaptation strategies to latitude in the native and invasive range of *Spartina alterniflora*. New Phytol **226:** 623–634.

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. Trends Ecol Evol. **20:** 223–228.

Marochio CA, Bevilaqua MRR, Takano HK, *et al.* (2017) Genetic admixture in species of *Conyza* (Asteraceae) as revealed by microsatellite markers. Acta Sci Agron **39**:437-445.

Novak SJ, Mack RN (1993) Genetic variation in *Bromus tectorum* (Poaceae): comparison between native and introduced populations. Heredity **71:** 167–176.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol **13**:1143-55.

Okada M, Hanson BD, Hembree KJ, *et al.* (2013) Evolution and spread of glyphosate resistance in *Conyza canadensis* in California. Evol Appl. **6**:761-777.

Okada M, Hanson BD, Hembree KJ, *et al.* (2015) Evolution and spread of glyphosate resistance in *Conyza bonariensis* in California and a comparison with closely related *Conyza canadensis*. Weed Res **55:** 173-184.

Page ER, Grainger CM, Laforest M, *et al* (2018). Target and non–target site mechanisms confer resistance to glyphosate in Canadian accessions of *Conyza canadensis*. Weed Sci. **66:** 234-245.

Pandolfo CE, Presotto A, Carbonell FT, *et al* (2018) Transgene escape and persistence in an agroecosystem: the case of glyphosate-resistant *Brassica rapa* L. in central Argentina. Environ Sci Pollut Res Int **25:** 6251–6264.

Patamsytė J, Naugžemys D, Čėsnienė T, *et al* (2018). Evaluation and comparison of the genetic structure of *Bunias orientalis* populations in their native range and two non-native ranges. Plant Ecol **219**:101–114.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics **28**: 2537-2539.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics **155**: 945-959.

Pyšek P, Čuda J, Šmilauer P, *et al.* (2020). Competition among native and invasive *Phragmites australis* populations: an experimental test of the effects of invasion status, genome size, and ploidy level. Ecol Evol **10**: 1106-1118.

Rosche C, Hensen I, Schaar A, *et al.* (2019). Climate outweighs native vs. nonnative range effects for genetics and common garden performance of a cosmopolitan weed. Ecol Monog **89**:e01386.10.1002/ecm.1386

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol **18**: 233-234.

Shields EJ, Dauer JT, VanGessel MJ, *et al.* (2006) Horseweed (*Conyza canadensis*) seed collected in the planetary boundary layer. Weed Sci **54**:1063-1067.

Soares FAA, Fregonezi A, Bassi D, *et al.* (2015) Evidence of high gene flow between samples of horseweed (*Conyza canadensis*) and hairy fleabane (*Conyza bonariensis*) as revealed by isozyme polymorphisms. Weed Sci **63**:604-612.

Szűcs M, Melbourne BA, Tuff T, *et al.* (2017). Genetic and demographic founder effects have long-term fitness consequences for colonising populations. Eco Lett **20:** 436–444.

Thébaud C, Abbott RJ (1995) Characterization of invasive *Conyza* species (Asteraceae) in Europe: quantitative trait and isozyme analysis. Am J Bot **82**: 360-368

van Boheemen LA, Lombaert E, Nurkowski KA, Gauffre B, Rieseberg LH, Hodgins KA. 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.

Voss N, Eckstein RL, Durka W (2012) Range expansion of a selfing polyploid plant despite widespread genetic uniformity. Ann Bot **110**:585-593.

Wang X, Abercrombie LG, Wadl AP, *et al.* (2009) Microsatellites from *Conyza canadensis* (Horseweed). Mol Ecol Resour **9**:1375-1379.

Wang C, Zhou J, Liu J, *et al.* (2017). Functional traits and reproductive allocation strategy of *Conyza canadensis* as they vary by invasion degree along a latitude gradient. Polish J Envir Stu. **26:** 1289–1297.

Wang J, Qi M, Huang H, *et al.* (2017) Atmospheric pollen dispersion from herbicide-resistant horseweed (*Conyza canadensis* L). J Aerobiologia **33**:393-406.

Weaver SE (2001) The biology of Canadian weeds. 115. *Conyza canadensis*. Can J Plant Sci **81**: 867-875.

Xu CY, Julien MH, Fatemi M, *et al.* (2010) Phenotypic divergence during the invasion of *Phyla canescens* in Australia and France: evidence for selection-driven evolution. Ecol Lett. **13:** 32–44.

Zhang YY, Zhang DY, Barrett SCH (2010) Genetic uniformity characterizes the invasive spread of water hyacinth (*Eichhornia crassipes*), a clonal aquatic plant. Mol. Ecol. **19**: 1774-1786.

Zhang ZY, Zheng XM, Ge S. 2007. Population genetic structure of *Vitex negundo* (Verbenaceae) in Three-Gorge Area of the Yangtze River: The riverine barrier to seed dispersal in plants. Biochemical Systematics and Ecology **35:** 506–516.

Zhao J, Solís-Montero L, Lou A, *et al.* (2013) Population structure and genetic diversity of native and invasive populations of *Solanum rostratu*m (Solanaceae). Plos one **8**:e79807.

Zhao SY, Sun SG, Dai C, *et al.* (2014) Genetic variation and structure in native and invasive *Solidago canadensis* populations. Weed Res. **55**:163-172.

Zhu BR, Barrett SCH, Zhang DY, *et al.* (2017) Invasion genetics of *Senecio vulgaris*: loss of genetic diversity characterizes the invasion of a selfing annual, despite multiple introductions. Biol Invasions **19**: 255-267.

Figure legends

Figure 1: Diagram showing the *Erigeron canadensis* populations sampled in native Alabama of the USA (A) and invasive Jiangsu and Zhejiang of China (B).

Figure 2. Principal coordinate analysis (PCoA) of 312 individuals from 10 populations of *Erigeron canadensis.* Genetic distances were calculated for individuals (A and B) and for populations (C). A: individuals labelled by population, squares and circles for invasive and native populations, respectively. B: individuals labelled by range, red and blue circles for invasive and native populations, respectively. C: PCoA with genetic by spatial distance between sampled populations, red and blue represent invasive and native populations, respectively.

Figure 3. UPGMA clustering showing the structure of the invasive and native populations of *Erigeron canadensis,* constructed from Nei genetic similarity based on ten SSR loci.

Figure 4. STRUCTURE clustering results for 312 individuals in 10 populations of *Erigeron canadensis* based on ten nuclear microsatellite loci. Colours represent distinct clusters (a) Genetic structure (K= 2) for joint analysis of invasive and native populations. (b) Genetic structure (K= 2) of 162 individuals in five invasive populations. (c) Genetic structure (K =2) of 150 individuals in five native populations. All STRUCTURE plots for populations analyzed are shown in Supplementary Figures S2-S4.

Table 1. Characterization of genetic variability in invasive and native populations of *Erigeron canadensis* estimated from ten nuclear microsatellite loci from 312 individuals (mean ± SE). N= number of individuals; Nt= number of alleles; N_A= mean number of alleles per locus; N_E= mean effective number of alleles per locus; A_R= allelic richness; $P_{A=}$ mean number of private alleles; H_E= expected heterozygosity; H_O=observed heterozygosity; F_{IS}= inbreeding coefficient; *s*= selfing rate estimated as 2 F_{IS}/(F_{IS}+1). * Significant differences (p < 0.001).

Country	Population	N	N_t	N_A	N_E	A_R	P_A	*He*	*Ho*	*FIS*	*s*
China	XY	31	75	7.50±1.30	3.40±0.51	6.23	14	0.62±0.07	0.08±0.03	0.87±0.05*	0.88
China	JH	31	67	6.70±0.94	3.68±0.61	5.97	9	0.64±0.07	0.08±0.03	0.87±0.02*	0.91
China	JY	32	57	5.70±0.70	2.22±0.17	4.93	9	0.53±0.02	0.05±0.02	0.91±0.03*	0.95
China	CS	36	52	5.20±0.55	2.14±0.15	4.38	8	0.51±0.03	0.03±0.01	0.94±0.04*	0.96
China	HY	32	59	5.90±0.70	2.68±0.33	5.20	5	0.57±0.05	0.03±0.02	0.94±0.02*	0.96
China	Mean	32.4	62	6.20±0.83	2.82±0.35	5.34	9.0±0.4	0.57±0.05	0.05±0.02	0.90±0.03*	0.93±0.02
USA	HV	32	78	7.80±1.08	3.55±0.48	6.68	14	0.66±0.04	0.14±0.06	0.80±0.03*	0.89
USA	BH	32	84	8.40±1.21	3.50±0.42	6.92	14	0.65±0.07	0.07±0.02	0.90±0.02*	0.95
USA	COS	24	76	7.60±0.79	4.65±0.61	7.13	21	0.74±0.03	0.08±0.02	0.90±0.08*	0.94
USA	TR	30	53	5.30±0.71	2.42±0.19	4.78	5	0.55±0.05	0.06±0.03	0.89±0.03*	0.94
USA	EP	32	54	5.40±0.80	2.67±0.36	4.80	12	0.57±0.05	0.16±0.02	0.73±0.05*	0.81
USA	Mean	30	69	6.90±0.91	3.36±0.41	6.06	13.2±0.5	0.63±0.05	0.10±0.03	0.84±0.04	0.90±0.03
	Total		655								
p value of t-test between two regions				0.386	0.317	0.282	0.191	0.194	0.080	0.153	0.421
USA+China	Mean	31.2		6.55±0.87	3.09±0.38	5.70	11.1±0.4	0.60±0.05	0.07±0.02	0.87±0.03	0.92±0.02

Table 2. Analysis of molecular variance (AMOVA) in invasive and native populations of *Erigeron canadensis*.

Source of Variation	*d.f*	SS	Estimated variance	% Variation	F_{ST}	*p-value**
Invaded region					0.169	<0.001
Among populations	4	169.930	0.610	17		
Within populations	319	957.132	3.000	83		
Native region					0.146	<0.001
Among populations	4	150.915	0.570	15		
Within populations	295	991.072	3.360	85		
Invaded + Native region						
Among regions	1	171.479	0.42	10	0.242	<0.001
Among populations	8	320.844	0.59	14		
Within populations	614	1.948.203	3.17	76		

Table 3. Pairwise F_{ST} matrix across 10 populations of *Erigeron canadensis* genotyped using ten microsatellite loci.

Native region						Invaded region				
HV	**BH**	**COS**	**TR**	**EP**		**XY**	**JH**	**JY**	**CS**	**HY**
0.000					**HV**					
0.123	0.000				**BH**					
0.169	0.165	0.000			**COS**					
0.198	0.157	0.157	0.000		**TR**					
0.153	0.104	0.142	0.097	0.000	**EP**					
0.155	0.181	0.221	0.295	0.235	**XY**	0.000				
0.164	0.164	0.203	0.260	0.218	**JH**	0.085	0.000			
0.276	0.276	0.237	0.315	0.258	**JY**	0.252	0.262	0.000		
0.289	0.287	0.256	0.324	0.272	**CS**	0.250	0.247	0.012	0.000	
0.226	0.202	0.201	0.272	0.216	**HY**	0.181	0.144	0.091	0.082	0.000

Table 4. Pairwise N_m matrix across invasive and native populations of *Erigeron canadensis* genotyped using ten microsatellite loci.

Native region						Invaded region				
HV	**BH**	**COS**	**TR**	**EP**		**XY**	**JH**	**JY**	**CS**	**HY**
0.000					**HV**					
1.786	0.000				**BH**					
1.227	1.263	0.000			**COS**					

1.014	1.344	1.347	0.000		**TR**					
1.383	2.157	1.512	2.337	0.000	**EP**					
1.366	1.132	0.879	0.597	0.813	**XY**	0.000				
1.272	1.277	0.981	0.713	0.897	**JH**	2.699	0.000			
0.656	0.655	0.803	0.544	0.720	**JY**	0.740	0.704	0.000		
0.615	0.621	0.726	0.521	0.668	**CS**	0.748	0.762	20.795	0.000	
0.859	0.985	0.992	0.668	0.908	**HY**	1.133	1.490	2.488	2.791	0.000

Figure 1

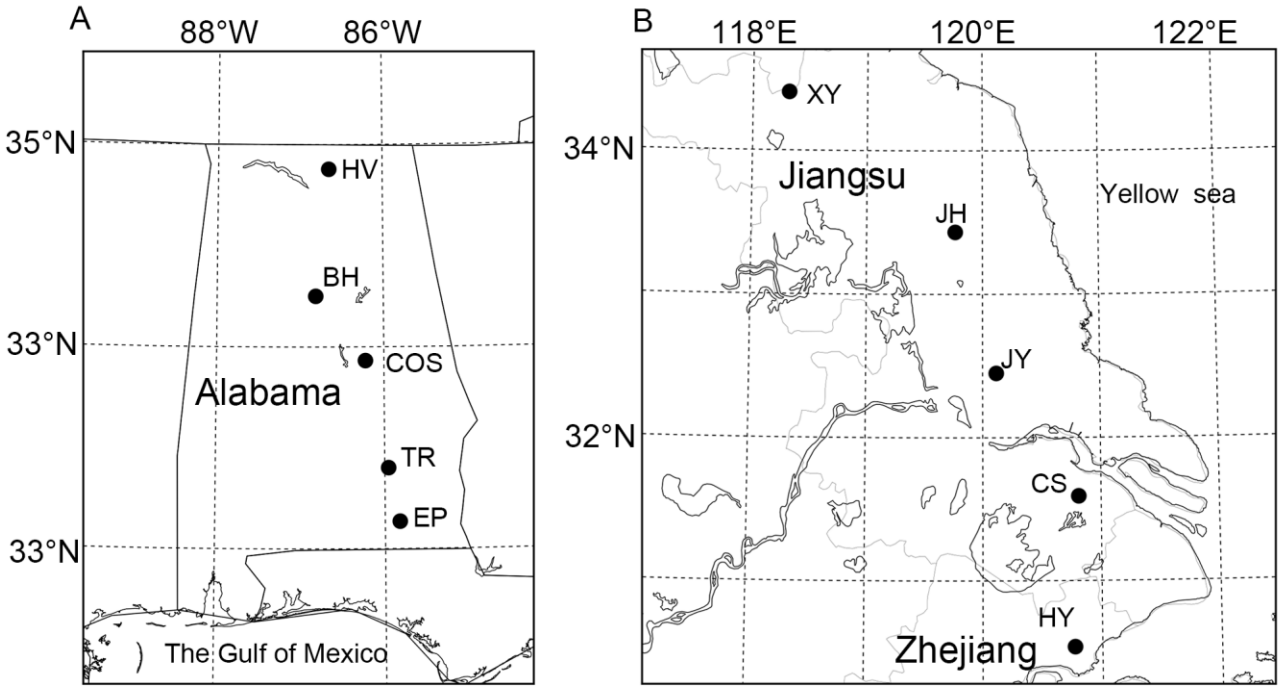

Figure 2

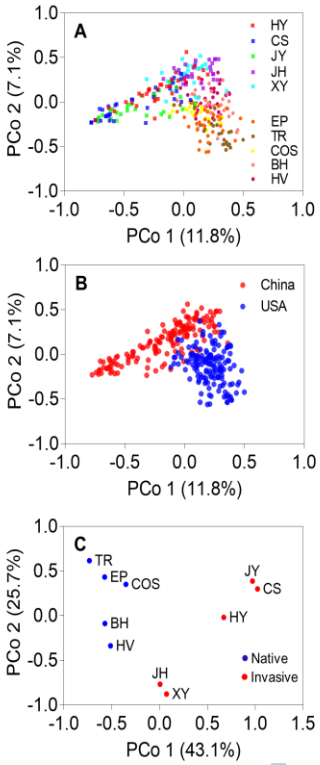

Figure 3

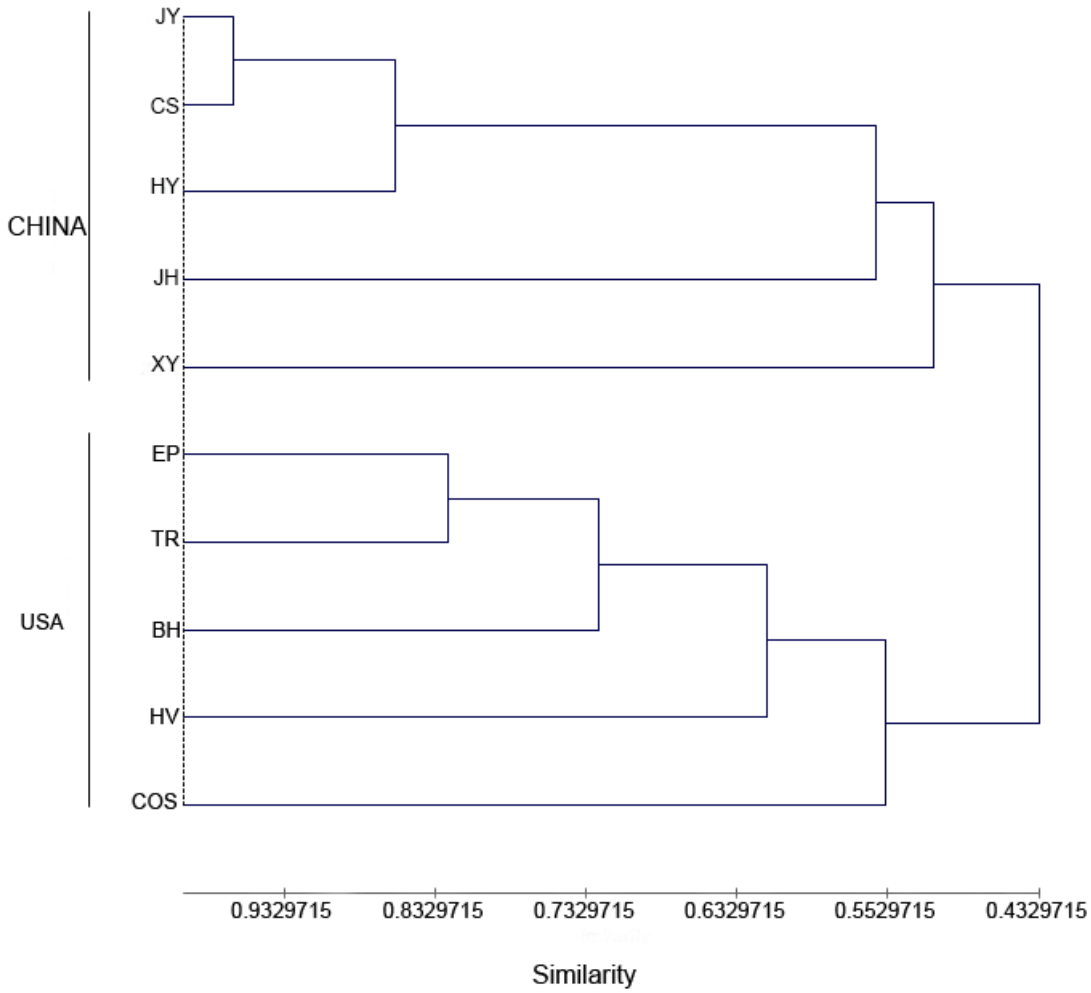

Figure 4