

Phylogeographical studies of *Glycine soja*: implicating the refugium during the Quaternary glacial period and large-scale expansion after the Last Glacial Maximum

Yunsheng WANG^{1*}, Muhammad Qasim SHAHID², Faheem Shehzad BALOCH³

¹School of Environmental and Life Science, Kaili University, Kaili, Guizhou, P.R. China

²College of Agriculture, South China Agricultural University, Guangzhou, Guangdong Province, P.R. China

³Department of Field Crops, Faculty of Agricultural and Natural Sciences, Abant İzzet Baysal University, Bolu, Turkey

Received: 10.07.2016 • Accepted/Published Online: 30.09.2016 • Final Version: 14.12.2016

Abstract: Wild soybean (*Glycine soja* Sieb. & Zucc.), the progenitor of cultivated soybean (*Glycine max* (L.) Merr.), is widely distributed in the Sino-Japanese Floristic Region (SJFR). Therefore, it is an excellent model organism to study the effect of Quaternary climatic changes on the distribution and migration of plant taxa in the SJFR. To investigate the phylogeography of *G. soja*, a total of 632 individuals representing 52 populations throughout the species distribution range were genotyped by using the sequencing data of a nuclear locus, AF105221, and a chloroplast locus, *trnQ-rps16*. We detected a total of 23 haplotypes from two loci, including five common and 18 rare haplotypes. Among them, 17 rare haplotypes were limited to Northeast Asia (Northeast China (NEC) and Korea) and the Yangzi River basin (middle and downstream of the Yangzi River (MDYR) and the eastern end of Southwest China (SWC)). Moreover, two common haplotypes were detected throughout the species distribution wide range. All the haplotypes from Japan and the Korean peninsula were also found in China. The haplotype distribution patterns of the two loci suggested that Northeast Asia (NEC and Korea) and the Yangzi River basin (MDYR and the eastern end of SWC) might have been the refugia for *G. soja* during the Quaternary glacial period, and *G. soja* expanded rapidly after the end of the Last Glacial Maximum. There was a land bridge between East China, the Korean peninsula, and Japan that permitted gene flow among populations of these regions in the ice age during the Quaternary.

Key words: Conservation, phylogeography, Pleistocene glaciation, population genetic structure, wild soybean

1. Introduction

The climatic oscillations during the Quaternary profoundly affected the genotype distribution patterns of plant taxa growing in temperate zones of the northern hemisphere (Hewitt, 2000). Fossil data and phylogeographic studies suggest that the distribution ranges of temperate plant species in Europe and North America, and especially those of trees and shrubs, changed with the latitudinal range shifts during glacial and inter-/postglacial periods. They tended to migrate towards lower latitudes in the ice age, and were then reestablished in northern regions after the end of the ice age (Petit et al., 2005). However, unlike Europe and North America, most parts of East Asia had not been covered by extensive ice sheets, and glaciers appeared only in high mountains and the northern part during the Quaternary (Zheng et al., 1998). The severe periodic climatic oscillations also profoundly affected the ecological environment and land topography of East Asia and thus the distribution and genotypic geographical patterns of plant species (Liu, 1988; Harrison et al., 2001).

For example, the large volume of accumulated ice reduced sea levels by about 120 m (Chung, 2007). The lowering of the sea level during ice ages produced land bridges in the east sea basin that linked the Japanese islands and some other islands with East China and the Korean peninsula. Those islands were then isolated from East China and the Korean peninsula again by the sea when sea levels rose during the postglacial period. Hence, gene flow between plant taxa could have happened between China, Japan, and the Korean peninsula (the CJK region) by a so-called land bridge in an ice age, and then was interrupted during the postglacial period (Dobson and Kawamura, 1998). Extreme climate events such as cold/warm spells and dry/wet periods are common in the mainland of East Asia (An, 2000). It is well known that both temperature and humidity have remarkable effects on the distribution of plant and animal taxa.

Phylogeographical studies suggested that the distribution ranges of some plant species in East Asia experienced no latitudinal expansion and multiple cryptic

* Correspondence: wys3269@126.com

refuges may exist in different latitudes (Qiu et al., 2011). Meanwhile, plant taxa were severely affected by habitat fragmentation and population isolation in the CJK region (Li et al., 2008; Qiu et al., 2009a, 2009b; Baloch et al., 2010), North China (Tian et al., 2009; Bai et al., 2010), Southwest China (Wang et al., 2008; Liu et al., 2009), and subtropical China (Gao et al., 2007; Gong et al., 2008; Qiu et al., 2009c; Guan et al., 2010) of the Sino-Japanese Floristic Region (SJFR) during the Quaternary.

Overall, the effects of climatic oscillations on species distribution patterns are not clear in the SJFR during the Quaternary. We inferred the following four possible reasons for this: 1) the taxa for plant phylogeographic studies were not enough; 2) plant taxa were more rich in this region; 3) the heterogeneity of ecological and climatic factors, such as temperature and humidity, showed more pronounced effects on species distribution in the SJFR than Europe or North America; 4) the studied taxa in the SJFR were mainly trees or shrubs with relatively limited geographic distributions (Aizawa et al., 2007; Gao et al., 2007; Wang et al., 2009).

Wild soybean (*Glycine soja*) is the progenitor species of soybean (*G. max*) and one of the world's most important crops (Smil, 2000; Shahid et al., 2009). *G. soja* is widely distributed in East Asia, including China, Korea, Japan, and the Russian Far East, areas located between 24°N and 53°N and 97°E and 143°E, largely belonging to the SJFR. *G. soja* is an annual vine and usually grows in agricultural fields and along field margins, roadsides, riverbanks, lakesides, and marshlands at altitudes ranging from 5 to 2650 m above sea level (Lu, 2004). *G. soja* is a self-fertilized species with outcrossing rates ranging from 2.4% to 19% (Kiang et al., 1992; Fujita et al., 1997). Short-distance dispersal by the means of pod dehiscence is a major way for the expansion of *G. soja* populations, and longer distances of dispersal may occur via water or birds (Kiang et al., 1992; Choi et al., 1999; Kuroda et al., 2006).

Sequence-based studies of population structure, historical demography, and phylogeography offer greater precision and repeatability than molecular marker-based analyses, particularly if both organelle and single-copy nuclear loci are employed (Petit et al., 2005; Wang et al., 2015a). Here, we executed the phylogeographic study of 632 *G. soja* individuals representing 52 populations by using chloroplast and nuclear loci sequences. The major aim was to address the current geographical patterns of the *G. soja* genotypes and the demographical history dynamics of *G. soja* in the Quaternary, such as the possible refugium in the ice age, the gene flow route during the glacial stage, and the expansion during the postglacial period. The results would not reveal only the evolutionary history of this species but would also provide the basic information for the conservation and utilization of *G. soja*

for breeding programs. Moreover, they will contribute to our understanding of the effects of climatic oscillations on species distribution patterns in the SJFR during the Quaternary.

2. Materials and methods

2.1. Sampling

More than 1500 wild soybean individual leaf samples were obtained from 52 populations (43, 5, and 4 populations from China, Korea, and Japan, respectively). The sampled populations were clustered into the following eight groups based on geographical distribution and climatic conditions: South China (SC), the middle and downstream of the Yangzi River (MDYR), Southwest China (SWC), Northwest China (NWC), North China (NC), Northeast China (NEC), Japan, and Korea (Table 1). These groups were largely consistent with the floristic provinces of the Eastern Asiatic region (Wu and Wu, 1998). The young leaves were kept in sealed plastic bags and dried by silica-gel. A total of 632 individuals were randomly selected from 52 populations for PCR, with an average of 4–13 individuals (mostly 12) per population, and sequenced for the final analysis.

2.1.1. DNA extraction and selection of loci/genes

Total genomic DNA was extracted from silica gel-dried leaves using the CTAB method (Doyle and Doyle, 1987), dissolved in 0.1X TE buffer (10 mmol/L Tris-HCl, pH 8.0, 1 mmol/L EDTA) to a final concentration of 50–100 ng/μL for PCR. In this study, a nuclear gene locus (AF105221) and a chloroplast gene locus (*trnQ-rps16*) were selected for investigations. Previous study had shown that AF105221 is a single-copy gene in *G. soja* populations (Wang et al., 2015b) and that cpDNA locus *trnQ-rps16* evolved at a fast rate (Shaw et al., 2007).

2.2. PCR amplification and sequencing

The PCR amplification was done by using a gene-specific pair of primers for AF105221 and *trnQ-rps16* as described by Van et al. (2004) and Shaw et al. (2007), and *trnQ-rps16* is a universal primer, i.e. one pair of the primer is likely to amplify nearly all angiosperm taxa (Shaw et al., 2007). PCR amplification was carried out in a volume of 50 μL containing 50–100 ng of DNA template, 0.1–0.2 μmol of each primer, 0.25 mM of each dNTP, 1.5 mM MgCl₂, and 0.5 U of Taq DNA polymerase. The PCR protocol was used as follow: 94 °C for 5 min; 35 cycles of 94 °C for 30 s, 56 °C for 45 s, and 72 °C for 90 s; and a final extension at 72 °C for 10 min. The PCR products were purified with PCR Purify Kit UNIQ-10 (Sangon Biotech, China) following the manufacturer's protocol. The purified PCR products were sequenced with the 3730/ABI sequencer (Thermo Fisher Scientific, USA). PCR amplifications and sequencing were repeated twice for each product to confirm the sequence variation.

Table 1. The location and natural habitats of *G. soja* populations.

| Geographical region | Population name | Location of sampling | Longitude (N) | Latitude (E) | Altitude (m) | Habitat |
|---------------------|-----------------|--|---------------|--------------|--------------|-------------------|
| SC | Population AF | Anfu County, Jiangxi Province | 27°23'17" | 114°36'06" | 85 | Beside road |
| | Population JO | Jianou County, Fujian Province | 27°01'28" | 118°17'34" | 126 | Beside river |
| | Population HY | Hengyang County, Hunan Province | 26°57'43" | 112°09'12" | 123 | Beside river |
| | Population RY | Ruyuan County, Guangdong Province | 24°55'12" | 113°08'12" | 510 | Beside road |
| | Population QZ | Quanzhou County, Guangxi Province | 25°52'21" | 110°51'43" | 722 | Beside road |
| MDYR | Population WC | Wuchang District, Hubei Province | 30°31'16" | 114°23'44" | 19 | Beside road |
| | Population XU | Xuanwu District, Jiangsu Province | 32°02'44" | 118°51'39" | 28 | Wasteland |
| | Population WQZ | Duqing County, Zhejiang Province | 30°33'00" | 119°58'20" | 15 | Beside canal |
| | Population SC | Shucha County, Anhui Province | 31°18'51" | 117°07'40" | 45 | Beside road |
| SWC | Population TB | Tongbai County, Henan Province | 32°22'12" | 113°24'01" | 33 | Beside road |
| | Population CK | Chengkou County, Chongqing | 31°58'98" | 108°40'02" | 805 | Valley |
| | Population YJ | Yinjiang County, Guizhou Province | 28°00'03" | 108°24'36" | 458 | Valley |
| | Population GH | Guangan City, Sichuan Province | 30°59'49" | 104°20'55" | 458 | Beside river |
| | Population CY | Chayu County, Xizang Province | 27°03'53" | 100°50'03" | 1685 | Valley |
| SWC | Population NL1 | Ninglang County, Yunnan Province | 28°30'08" | 97°00'38" | 2600 | Beside field |
| | Population NL2 | Ninglang County, Yunnan Province | 28°06'54" | 99°51'02" | 2550 | Beside field |
| NWC | Population BX | Bingxian County, Shaanxi Province | 35°02'26" | 108°04'39" | 835 | Valley |
| | Population HX | Huixian County, Gansu Province | 33°53'36" | 105°49'33" | 1126 | Canal |
| | Population YY | Youyu County, Shanxi Province | 40°08'38" | 112°20'22" | 1279 | Beside road river |
| | Population LW | Lingwu County, Ningxia Province | 38°08'44" | 106°19'33" | 1103 | Canal |
| | Population WS | Wenshui County, Shanxi Province | 37°25'02" | 112°01'02" | 759 | Beside canal |
| | Population YA | Ya'an County, Shaanxi Province | 36°37'29" | 109°27'18" | 962 | Along river |
| | Population YL | Yulin city, Shanxi Province | 38°16'50" | 109°44'18" | 1051 | Along river |
| | Population LXS | Linxian County, Shanxi Province | 37°56'51" | 110°59'15" | 948 | Along river |
| | Population YC | Yichuan County, Shanxi Province | 36°03'31" | 110°11'20" | 829 | Along river |
| | Population BJ | Baoji city, Shanxi Province | 34°22'04" | 107°0'05" | 586 | Along river |
| NC | Population WN | Weinan city, Shaanxi Province | 34°31'01" | 109°30'00" | 355 | Along river |
| | Population JZ | Jizhou County, Hebei Province | 37°44'31" | 115°41'12" | 23 | Beside road |
| | Population DY | Dongying City, Shandong Province | 37°34'25" | 118°31'27" | 6 | Beside ditches |
| | Population WQT | Wuqing District, Tianjing | 39°26'57" | 117°14'57" | -6 | Beside ditches |
| | Population XH | Xuanhua County, Hebei Province | 40°35'34" | 115°01'17" | 601 | Beside river |
| | Population QHD | Qinghuangdao city, Hebei Province | 39°48'29" | 119°25'54" | 18 | Beside river |
| NEC | Population YT | Yantai city, Shandong Province | 37°28'65" | 121°27'14" | 10 | Wasteland |
| | Population LXH | Lanxi County, Heilongjiang Province | 46°13'06" | 126°20'16" | 139 | Beside pond |
| | Population JH | Jiaohe County, Jinlin Province | 43°48'28" | 127°14'15" | 126 | Beside river |
| | Population XM | Xinming County, Liaoning Province | 41°57'19" | 122°51'29" | 28 | Beside pond |
| | Population MDJ | Mudanjiang city, Heilongjiang Province | 44°34'07" | 129°36'26" | 253 | Beside river |
| | Population BS | Baishan city, Jilin Province | 42°27'00" | 128°07'44" | 701 | Roadside |
| | Population KS | Keshan County, Heilongjiang Province | 48°17'00" | 125°29'54" | 325 | Beside field |
| | Population QQHE | Qiqihaer city, Heilongjiang Province | 47°20'27" | 123°56'24" | 304 | Beside river |
| | Population HG | Hegan city, Heilongjiang Province | 47°12'20" | 130°13'46" | 213 | Beside road |
| Japan | Population DH | Dunhua city, Jilin Province | 43°22'28" | 128°13'36" | 495 | Beside road |
| | Population HL | Hulin city, Heilongjiang Province | 42°02'24" | 128°03'16" | 73 | Beside field |
| | Population J1 | Kanagawa, Japan | 34°57'33" | 137°08'20" | 12 | Wetland |
| | Population J2 | Tokyo, Japan | 34°49'41" | 135°46'13" | 35 | Wetland |
| Korea | Population J3 | Hirakata, Osaka, Japan | 34°48'52" | 135°29'02" | 11 | Wetland |
| | Population J4 | Okazaki, Japan | 34°57'33" | 137°08'20" | 37 | Wetland |
| Korea | Population K1 | Gangwon-do, South Korea | 37°35'15" | 128°24'33" | 520 | Wetland |
| | Population K2 | Gangwon-do, South Korea | 37°54'47" | 128°29'56" | 340 | Wetland |
| | Population K3 | Incheon, South Korea | 37°27'30" | 126°21'53" | 11 | Wetland |
| | Population K4 | Yeongcheon-si city, South Korea | 37°54'47" | 128°29'56" | 102 | Along road |
| | Population K5 | Moonkyeong-si city, South Korea | 37°27'30" | 126°21'53" | 77 | Along road |

SC = South China; MDYR = Middle and downstream of the Yangzi River; SWC = Southwest China; NWC = Northwest China; NC = North China; NEC = Northeast China.

2.3. Data analysis

DNA sequences were aligned using ClustalX version 2.0 (Thompson et al., 1997) and then adjusted manually by the program BioEdit 4.10 (Hall, 1999). All sequences were deposited in GenBank (accession numbers: GU594719–GU594728 and GU594681–GU594692, KX060992). For further analyses, insertion and deletion mutations were excluded from aligned sequences.

The haplotype occurrences and their frequencies were identified with the program DAMBE (Xia and Xie, 2001). Median-joining haplotype networks were constructed for the two loci using Network version 4.2.0.1 (Bandelt et al., 1999) (<http://www.fluxus-engineering.com>). The geographical distribution of the different haplotypes identified in *G. soja* was mapped using ArcMap 9.1 (<http://www.esri.com/software/arcgis/>).

The parameters of genetic diversity (H_s and H_T) and of genetic differentiation (G_{ST} and N_{ST}) were estimated by using the program PERMUT (available at <http://www6.bordeaux-aquitaine.inra.fr/biogeco/Production-scientifique/Logiciels/Contrib-Permut/Permut>). G_{ST} (coefficient of gene differentiation) and N_{ST} (frequencies and similarities among individuals) were compared by using a permutation test with 1000 permutations (Burban et al., 1999).

3. Results

3.1. Haplotype statistics and evolutionary network

The sequences of the AF105221 and *trnQ-rps16* loci were obtained with aligned lengths of 541 bp and 476 bp after eliminating the insertion/deletion sites from 616 and 617 individuals, respectively. We identified 11 and 9 polymorphic sites and 13 (A1–A13) and 10 (B1–B10) different haplotypes at loci AF105221 and *trnQ-rps16*, respectively (Tables 2 and 3). Among A1 to A13, A1 exhibited the highest frequency and accounted for 71.75% of total sequences of locus AF105221, followed by A3 and A2, which accounted for 9.42% and 6.17% of individuals, respectively. Five haplotypes, A5, A9, A10, A11, and A13, appeared only one time, and all other haplotypes were found to be rare haplotypes (<5%) (Table 4). A1 appeared in the central position of the evolutionary network, and eight haplotypes, including the second and third most frequent haplotypes (A3 and A2), linked directly with the A1 haplotype (Figure 1a). Among B1 to B10, two main haplotypes, B1 and B2, accounted for 34.36% and 51.38% of total sequences of locus *trnQ-rps16*, respectively. The third highest was B3, which accounted for 9.40%, and the other seven were rare haplotypes (<5%) (Table 5). B2 was found in the central position of the evolutionary network, and six haplotypes, including the second and third most frequent haplotypes, B1 and B3, exhibited a direct link with the B2 haplotype (Figure 1b).

3.2. Geographical distribution patterns of haplotypes

For AF105221, the dominant haplotype A1 was distributed widely in the species range, appearing in 43 populations of China and two Japanese and two Korean populations, excluding two populations of SWC, one population of NC, and four populations of NEC. B1 was the dominant haplotype (Freq. > 50%) in all populations, and it was the only haplotype found in 19 populations. The A2 haplotype was distributed in five populations across four geoecological regions (i.e. one population of SWC, one population of MDYR, two populations of SC, and one population of Korea). A3, the second highest frequent haplotype, was mainly distributed (37.93%) in five populations of NEC and was also found in four populations of MDYR, one population of NC, two populations of NWC, one population of the Korean region, and two populations of the Japanese region. Among the rare haplotypes of locus AF105221, A4 was distributed mainly in NEC, but was also found in MDYR, Korea, and Japan in one or two populations with low frequency. A7 was distributed across four geoecological regions but just appeared in one or two populations with very low frequency. Other rare haplotypes, such as A5, A6, A8, A9, A10, A11, and A13, just appeared in one population and belonged to different geographical regions. A11 was the dominant haplotype but was limited to the Korean region (Figure 2a; Table 4).

For *trnQ-rps16*, B1 was distributed mainly in SC, SWC, MDYR, and the Korean region, and it was also found in other geoecological regions. B1 occupied the whole SC area alone and dominated the SWC and MDYR regions. B2 was distributed mainly in NEC, NC, NWC, and SWC and it was the dominant haplotype in NEC, NC, and NWC. B3 was detected in the MDYR and Japanese regions. Interestingly, the shore of the Bohai Sea, Jiaodong peninsula, was also occupied by B3. Other rare haplotypes were limited to one geoecological region or a population. NEC had the highest haplotype diversity, and six haplotypes, including two common haplotypes, B2 and B3, and four rare haplotypes, B7–B10, were limited to three populations of this region. The geoecological groups with the second highest haplotype diversity were MDYR and SWC, and four haplotypes were detected in each region. We found three common haplotypes (B1, B2, and B3) and one rare haplotype (B4) in MDYR, which was limited to one population. There were two common haplotypes (B1 and B2) and two rare haplotypes (B5 and B6) in SWC, and each rare haplotype was limited to one population and located in the eastern end of this region, near the MDYR region. NWC, NC, Korea, and Japan shared two common haplotypes (Figure 2b; Table 5).

3.3. Genetic diversity of geographical structure

We detected approximately 2.15 and 2.7 times higher total genetic diversity, H_T , than average diversity within

Table 2. Nucleotide substitutions detected at AF105221 locus in *G. soja* and their accession numbers.

| Haplotype | 22* | 34 | 44 | 63 | 166 | 208 | 221 | 224 | 243 | 413 | 500 | Accession number |
|-----------|-----|----|----|----|-----|-----|-----|-----|-----|-----|-----|------------------|
| A1 | A | A | C | T | C | A | G | G | T | A | A | GU594681 |
| A2 | A | A | C | T | C | A | A | G | T | A | A | GU594682 |
| A3 | A | A | C | T | C | A | G | G | C | A | A | GU594683 |
| A4 | A | A | C | C | C | A | G | G | T | G | A | GU594684 |
| A5 | A | A | C | T | C | G | G | G | T | A | A | GU594685 |
| A6 | A | A | A | T | C | A | G | G | T | A | A | GU594686 |
| A7 | A | G | C | C | C | A | G | G | T | G | A | GU594687 |
| A8 | T | A | C | T | C | A | G | G | T | A | A | GU594688 |
| A9 | A | A | C | T | C | A | G | G | T | G | A | GU594689 |
| A10 | A | A | C | T | C | A | G | T | T | A | A | GU594690 |
| A11 | A | G | C | T | C | A | A | G | T | A | A | GU594691 |
| A12 | A | A | C | T | T | T | A | A | G | A | A | GU594692 |
| A13 | A | A | C | T | T | T | A | A | G | A | T | KX060992 |

*Numbers indicate the order of polymorphic sites in gene segment.

Table 3. Nucleotide substitutions detected at *trnQ-rps16* locus in *G. soja* and their accession numbers.

| Haplotype | 4* | 124 | 203 | 236 | 237 | 287 | 313 | 375 | 383 | Accession number |
|-----------|----|-----|-----|-----|-----|-----|-----|-----|-----|------------------|
| B1 | A | T | A | A | T | T | T | C | G | GU594719 |
| B2 | A | T | A | A | T | G | T | C | G | GU594720 |
| B3 | A | T | A | A | T | G | T | A | G | GU594721 |
| B4 | C | T | A | A | T | G | T | A | G | GU594722 |
| B5 | A | T | A | A | T | T | T | C | A | GU594723 |
| B6 | A | G | A | A | T | G | T | C | G | GU594724 |
| B7 | A | T | A | T | T | T | T | C | A | GU594725 |
| B8 | A | T | A | A | G | G | T | C | G | GU594726 |
| B9 | A | T | C | A | T | G | T | C | G | GU594727 |
| B10 | A | T | A | A | T | G | G | C | G | GU594728 |

*Numbers indicate the order of polymorphic sites in gene segment.

populations (H_s) across populations for the AF105221 and *trnQ-rps16* loci. However, there was a nonsignificant difference between G_{ST} and N_{ST} for both loci, although N_{ST} was higher than G_{ST} (Table 6), which suggested the presence of a relatively weak phylogeographic structure within the species.

4. Discussion

4.1. The possible refugia of *G. soja* during the ice age

Phylogeographic studies suggested that plant taxa did not experience long-distance dispersal southward to northward during the interglacial cycle of the Quaternary

glacial period and the Last Glacial Maximum (LGM) in the East Asia region, and these plant taxa were believed to survive in multiple cryptic refugia near their natural habitats during the interglacial cycle (Qiu et al., 2011). Similar behavior was noticed in *G. soja*. Here, we demonstrated that there were at least two refugia for *G. soja*: one refugium was in Northeast Asia, including the NEC region and the Korean peninsula, and another was in the Yangtze River basin, including the MDYR region and the eastern region of SWC. The reasons for this conclusion are as follows: two characters, such as high haplotype diversity and appearance of unique haplotypes, were regarded as

Table 4. Number of haplotypes per population identified at AF105221 locus.

| Population | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | A10 | A11 | A12 | A13 | Total |
|------------|-----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-------|
| SC group | 55 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 59 |
| G1_AF | 11 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G1_JO | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G1_HY | 8 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G1_RY | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G1_QZ | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| MDYR group | 45 | 3 | 9 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 60 |
| G2_WC | 8 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_XU | 7 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_WQZ | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_SC | 8 | 0 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_TB | 10 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| SWC group | 50 | 10 | 0 | 0 | 1 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 71 |
| G3_CK | 2 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G3_YJ | 11 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G3_GH | 1 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G3_CY | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G3_NL1 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G3_NL2 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| NWC group | 127 | 0 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 131 |
| G4_BX | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_HX | 11 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YY | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_LW | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_WS | 10 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YA | 11 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YL | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G4_LXS | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YC | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_BJ | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_WN | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| NC group | 66 | 0 | 12 | 0 | 0 | 0 | 1 | 4 | 0 | 0 | 0 | 0 | 0 | 83 |
| G5_JZ | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5_DY | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5-WQT | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5_XH | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G5_QHD | 11 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5_YT | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5_JN | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| NEC group | 73 | 0 | 22 | 19 | 0 | 0 | 3 | 0 | 1 | 1 | 0 | 0 | 0 | 119 |
| G6_LXH | 7 | 0 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_JH | 9 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |

Table 4. (Continued).

| | | | | | | | | | | | | | | |
|---------------|-------|------|------|------|------|------|------|------|------|------|------|------|------|-----|
| G6_XM | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_MDJ | 5 | 0 | 6 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_BS | 4 | 0 | 3 | 1 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G6_KS | 9 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 12 |
| G6_QQHE | 10 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_HG | 3 | 0 | 5 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_DH | 3 | 0 | 3 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_HL | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 12 |
| Japan group | 11 | 18 | 10 | 3 | 0 | 0 | 5 | 0 | 0 | 0 | 1 | 0 | 0 | 48 |
| J1 | 2 | 0 | 8 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| J2 | 9 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| J3 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| J4 | 0 | 6 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 | 0 | 0 | 12 |
| Korea group | 15 | 3 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 1 | 45 |
| K1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 10 |
| K2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 12 |
| K3 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| K4 | 0 | 0 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| K5 | 4 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 8 |
| Total | 442 | 38 | 58 | 27 | 1 | 10 | 10 | 4 | 1 | 1 | 1 | 22 | 1 | 616 |
| Frequency (%) | 71.75 | 6.17 | 9.42 | 4.38 | 0.16 | 1.62 | 1.62 | 0.65 | 0.16 | 0.16 | 0.16 | 3.57 | 0.16 | |

SC = South China; MDYR = Middle and downstream of the Yangzi River; SWC = Southwest China; NWC = Northwest China; NC = North China; NEC = Northeast China.

the indication of refugia of species in the ice age (Avisé, 1994; Hewitt, 1996; Schmitt, 2007; Schimpf et al., 2014). The highest haplotype diversity for two loci was found in these two regions. Moreover, most unique haplotypes were also limited to these two regions. In Northeast Asia, a total of 23 different haplotypes were detected in the Yangtze River basin. Of these 23 haplotypes, 10 were distinguished with locus *trnQ-rps16*, except three common haplotypes, and the NEC region and Yangtze River basin possessed four and three unique haplotypes, respectively. Among 13 haplotypes detected by locus AF105221, four and two rare haplotypes were limited to Northeast Asia and the Yangtze River basin, respectively. Interestingly, two unique haplotypes were found in the Japanese region, which indicates that Japan may have been one of the refugia of *G. soja* during LGM.

Why was one of the refugia located in Central China and another one in NEC, when these regions are far away from each other? We speculate that the situation was not only caused by the climate change in East Asia in the Quaternary, but also by the biological characteristics of wild soybean. *G. soja* is widely distributed in East Asia and its distributional range is between 24°N and

53°N, characterized by three climatic zones, including subtropical, warm temperate, and cold temperate zones. There is at least a 20 °C difference in the average annual temperature between southern (Heyuan, Guangdong Province: 19.9 °C) and northern (Mohe, Heilongjiang Province: 0 °C) regions of China (China Meteorological Data Network, <http://data.cma.gov.cn/>). These data show that *G. soja* has a very wide temperature range and photoperiodic adaptability, and strong tolerance to low temperatures. However, the ecological niches of *G. soja* populations are always located near places where water is in abundance, such as rivers, ponds, ditch edges, and watered fields, which ensures a sufficient quantity of moisture required for seed germination, because the seed coat of wild soybean is dense and hard, and it needs high moisture contents for germination. Thus, *G. soja* is resistant to cold but not to drought, at least during the seed germination stage. In spite of NEC and the Korean peninsula being situated at high altitudes, this area could still have been the refugium of *G. soja* during the glacial period in the Quaternary because this region is located on the shore of the West Pacific, and we deduce that the moisture was enough for growth of *G. soja*. In the Quaternary, the most

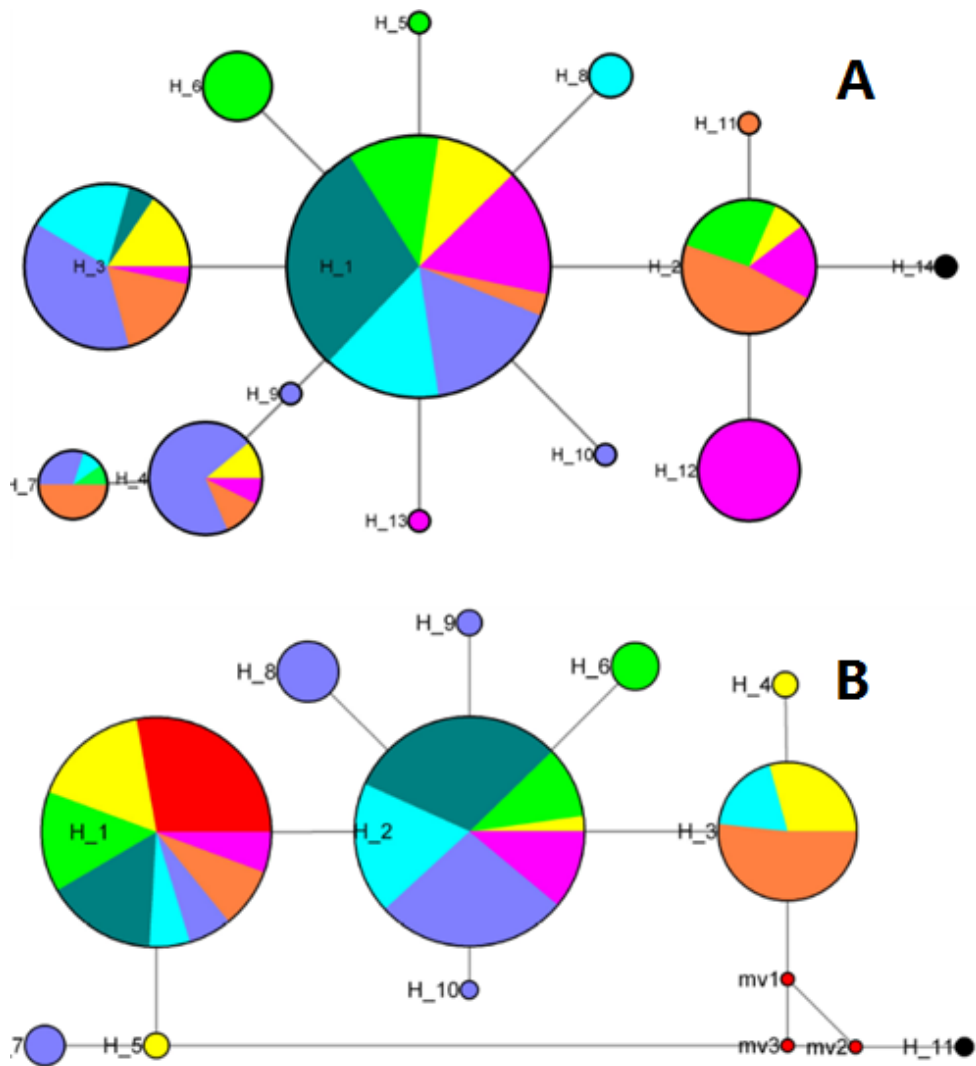


Figure 1. Median-joining haplotype networks of *G. soja* haplotypes: a) AF105221, b) *trnQ-rps16*.

notable climate feature was periodic warm/wet and cold/dry spells (Lu et al., 2013). In the cold stage, called an ice age, annual average temperature dropped about 10 °C, but not evenly in different East Asian regions (Shi, 2002). Meanwhile, annual precipitation was also greatly reduced in the ice age (Shi, 1989). A previous study showed that the NC and NWC regions were very arid and became desert and grasslands (Yu et al., 2000). Therefore, these two regions should not have been suitable habitats for *G. soja* in the ice age; in fact, in the LGM, because of sea regression, the Bohai Sea and the East China Sea changed to land, and all of NC was situated away from the sea, not the shore of the sea like today. The haplotypes detected by two loci in these two regions were rather common haplotypes, which widely appeared in the *G. soja* populations and represent a recent migration event happening after the last glacial period. Although we sampled 16 populations from regions

of NC and NWC, only one unique haplotype was detected from NC with AF105221. The extreme dry temperate climate might have eliminated wild soybean from most of the habitats of these two regions. These two regions shared haplotypes with the regions of NEC and MDYR, suggesting that the migration of wild soybean may have happened from NEC and MDYR after the LGM. As for the MDYR region and its contiguous zone, the east end of SWC, they still had adequate moisture to supply the needs of seed germination even in the condition of a dry climate, because the Yangzi River had plenty of water during the ice age, and this makes the MDYR region and its contiguous zone be one of the refugia for *G. soja*.

4.2. Gene flow of *G. soja* by land bridge among the refugia during ice age

The CJK region was separated by the East China Sea with higher sea level during inter-/postglacial times.

Table 5. Number of haplotypes per population identified at *trnQ-rps16* locus.

| Population | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B9 | B10 | Total |
|------------|----|----|----|----|----|----|----|----|----|-----|-------|
| SC group | 59 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 59 |
| G1_AF | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G1_JO | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G1_HY | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G1_RY | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G1_QZ | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| MDYR group | 35 | 6 | 17 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 60 |
| G2_WC | 5 | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_XU | 10 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_WQZ | 9 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_SC | 5 | 0 | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G2_TB | 6 | 3 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| SWC group | 30 | 32 | 0 | 0 | 2 | 7 | 0 | 0 | 0 | 0 | 71 |
| G3_CK | 0 | 10 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 12 |
| G3_YJ | 0 | 5 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 12 |
| G3_GH | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G3_CY | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G3_NL1 | 11 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G3_NL2 | 7 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| NWC group | 33 | 98 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 131 |
| G4_BX | 4 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_HX | 5 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YY | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_LW | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_WS | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YA | 3 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YL | 1 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G4_LXS | 3 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_YC | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_BJ | 5 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G4_WN | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| NC group | 12 | 59 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 82 |
| G5_JZ | 6 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5_DY | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5-WQT | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5_XH | 6 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G5_QHD | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G5_YT | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| G5_JN | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| NEC group | 13 | 87 | 0 | 0 | 0 | 0 | 5 | 11 | 2 | 1 | 119 |
| G6_LXH | 6 | 4 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 12 |
| G6_JH | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |

Table 5. (Continued).

| | | | | | | | | | | | |
|-------------|-------|-------|------|------|------|------|------|------|------|------|-----|
| G6_XM | 3 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_MDJ | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 12 |
| G6_BS | 4 | 5 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 12 |
| G6_KS | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 12 |
| G6_QQHE | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 12 |
| G6_HG | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| G6_DH | 0 | 3 | 0 | 0 | 0 | 0 | 3 | 4 | 0 | 1 | 11 |
| G6_HL | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| Japan group | 18 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 48 |
| J1 | 4 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| J2 | 2 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| J3 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| J4 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| Korea group | 12 | 35 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 47 |
| K1 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| K2 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| K3 | 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| K4 | 1 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| K5 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| Total | 212 | 317 | 58 | 2 | 2 | 7 | 5 | 11 | 2 | 1 | 617 |
| Frequency | 34.36 | 51.38 | 9.40 | 0.32 | 0.32 | 1.13 | 0.81 | 1.78 | 0.32 | 0.16 | |

SC = South China; MDYR = Middle and downstream of the Yangzi River; SWC = Southwest China; NWC = Northwest China; NC = North China; NEC = Northeast China.

However, these regions had been connected by the East China Sea basin when sea level fell by c. 85–130/140 m during the glacial period (Millien-Parra and Jaeger, 1999; Qian and Ricklefs, 2001), especially during the LGM (Shi, 2002). The land bridge of the East China Sea served as a ‘dispersal corridor’ (Simpson, 1940; Lomolino et al., 2006), allowing intermittent migration of most warm-temperate deciduous forest-restricted plant species from the Asian mainland into Japan (or vice versa), and/or periodic secondary contact and gene flow among formerly isolated populations, possibly up to the last shelf submergence (c. 16,000–7,000 years BP) (Chung, 2007). This land bridge has been suggested to be covered by temperate deciduous forest during glacial intervals (Harrison et al., 2001). Here, we inferred that refugia of *G. soja*, such as Northeast Asia (NEC and the Korean peninsula), Japan, and the Yangzi River basin, were connected by the land bridge of the East Hai Sea, and gene flows among these regions were also detected. In the present study, eight haplotypes appeared in four Japanese populations, and two common haplotypes, B1 and A1, were found in the populations of all other ecogeological regions, while only one unique haplotype (A11) was detected with both loci. This revealed recent

migration during the LGM, and two common haplotypes, B3 and A3, with a mid-level of frequency, showed similar distribution patterns in populations of Japan, the east of NC, and the MDYR region, which suggested that an older migration event happened by the land bridge between Japan and the East Asian mainland. Wild soybean is widely distributed across Japan and more populations being sampled would depict a clear picture of gene flow in wild soybean between Japan and the East Asian mainland.

Phylogeographic studies on the genus *Croonia* (Li et al., 2008), *Kirengeshoma palmata* (Qiu et al., 2009b), and *Platycrater arguta* (Qiu et al., 2009c) suggested that deep allopatric-vicariant differentiation of isolated lineages appeared in the CJK region (Qiu et al., 2011). In spite of one unique haplotype from Korea, most of the Korean wild soybean haplotypes were similar to those of East China, and this revealed that the gene flow among the three regions occurred frequently and not so long ago.

4.3. Population expansion of *G. soja* in the Holocene after the Last Glacial Maximum

Among 23 haplotypes detected by AF105221 and *trnQ-rps16* loci in *G. soja* populations, three haplotypes, A1, B1, and B2, appeared in more than half of the total samples

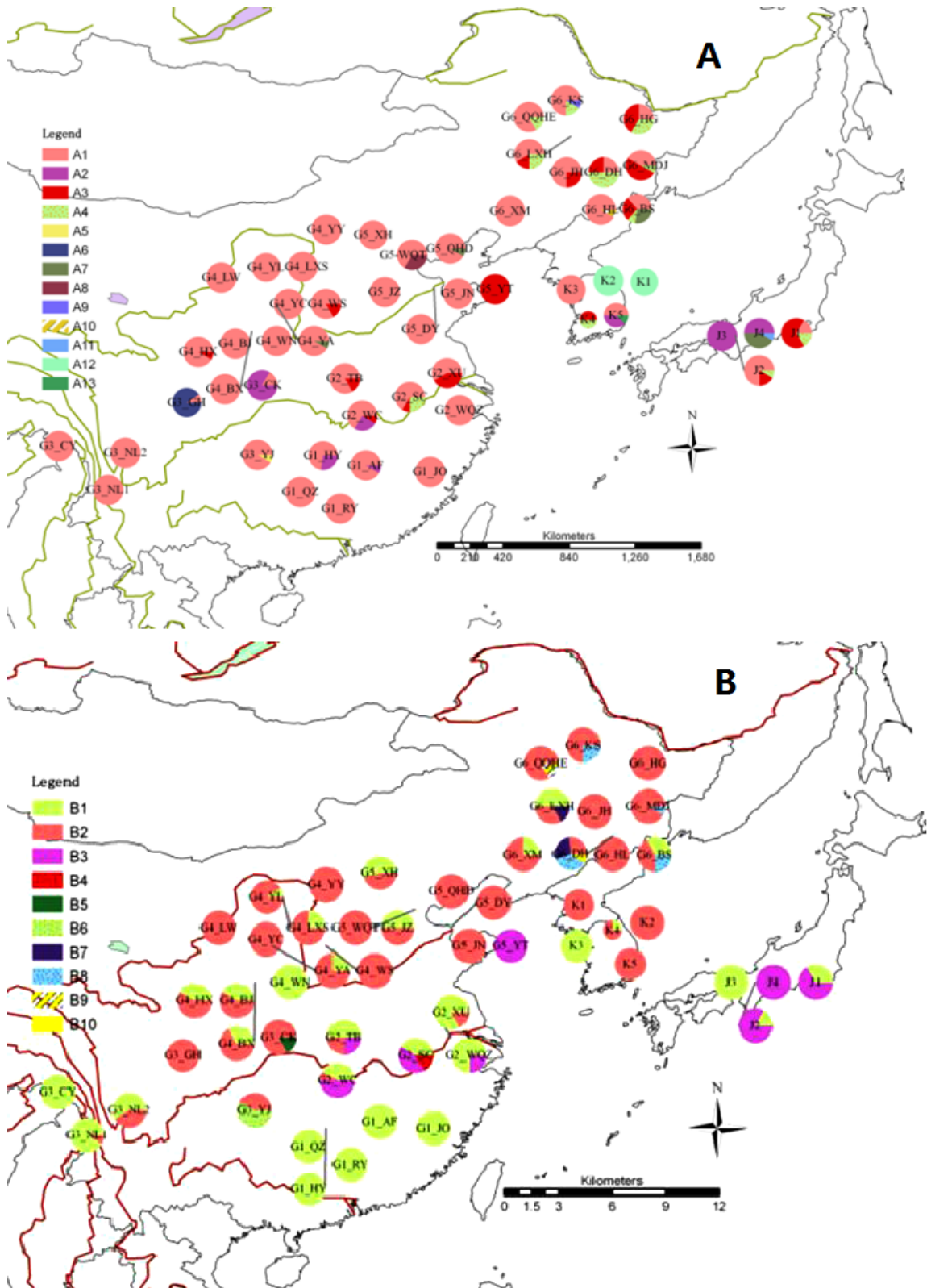


Figure 2. Frequencies and geographical distribution of *G. soja* haplotypes: a) AF105221, b) *trnQ-rps16*. Numbers at the end of Chinese population names indicate the region from which they derived: G1–G6 represent SC, MDYR, SWC, NWC, NC, and NEC regions, respectively.

Table 6. Genetic differentiation in the *G. soja* populations detected by using nuclear and chloroplast loci.

| Locus | H_s | H_T | G_{ST} | N_{ST} |
|-------------------|----------------|----------------|----------------|----------------|
| <i>trnQ-rps16</i> | 0.238 (0.0362) | 0.614 (0.0341) | 0.613 (0.0562) | 0.564 (0.0638) |
| AF105221 | 0.229 (0.0359) | 0.491 (0.0641) | 0.534 (0.0680) | 0.589 (0.1093) |

H_s , Gene diversity within populations; H_T , Total gene diversity; G_{ST} , Coefficient of gene differentiation; N_{ST} , Frequencies and similarities among haplotypes. The values in parentheses represent standard errors.

and in almost all populations. For example, haplotype A1 was distributed in almost the whole species range. B1 and B2 were also distributed widely; B1 was just found to be absent in NC, and B2 in SC and Japan. These results showed the presence of a large-scale migration and expansion in *G. soja* populations in the recent warm Holocene after the LGM. In general, an expanding population would show a star-like haplotype network (Harpending, 1994). In this study, haplotype network analysis of both AF10522A and *trnQ-rps16* loci revealed a star-like haplotype network (Figure 1), and this result suggested that *G. soja* is an expanding population.

Wild soybean of the modern age has two isolated distribution areas: one in Ninglang County in the northeastern Yunnan Province and another one in Chayu County of the southeastern Tibet Province; the former is also the highest distribution habitat with an altitude of 2600 m. These two regions are separated by over 450 km of distance from the wild soybean's main distribution. Two populations sampled from Ninglang County and one population from Chayu County contain one to two haplotypes, but without a unique haplotype, which suggests the recent colonization of wild soybean in these regions. The possible reason for this is that humans transferred these wild soybeans through the ancient "Tea Horse Road", a trade road that lasted for thousands of years and connected Yunnan, Sichuan, and Tibet provinces of China with Burma, Nepal, and India. Horses were used as the major transportation source, and wild soybean is a good source of forage for horses. At the same time, sites of wild soybean populations were present along the ancient Tea Horse Road in Ninglang and Chayu counties. To offer the best forage to their horses and to complete the hard journey, perhaps ancient caravans brought the seeds of *G. soja* and planted them along the roadside of important posts. If our hypothesis is true, this is the first report of plant transfer along the ancient Tea Horse Road by human beings.

4.4. Implications for conservation

Preservation of genetic diversity and evolutionary potential, and prevention of inbreeding depressions, are key goals in conservation (Frankel and Soulé, 1981; Hufford and Mazer, 2003; Rauch and Bar-Yam, 2005;

Wang et al., 2015a; Baloch et al., 2016). Wild soybean has undergone considerable habitat fragmentation and population extinction due to human activity in the recent past (Li et al., 2005), and our field investigations failed to find populations at many sites where wild soybean had been reported previously. For example, wild soybean was widely distributed along the Fenhe River in Shanxi Province and different lines have distinguished characteristics: one variety had a 100-grain weight as high as 14.7 g; another had gray hairs, white flowers, and obvious stems; and another had pale green, yellow, and brown colored seeds. These previously recorded *G. soja* populations disappeared after the expansion of cities and conversion of land for urban and farming purposes (Li et al., 2005). Our field investigations also revealed that wild soybean is usually found in small populations in these regions. The regions containing the MDYR and NEC groups deserve special conservation considerations because these two regions have the most abundant genetic diversity and the largest number of private alleles. The MDYR region is located in the most economically developed area of China, and the frequent industrial and agricultural activities and the expansion of cities are threatening the natural habitats of *G. soja*. The NEC region, the most important agricultural area in China, has seen heavy fragmentation and loss of *G. soja* habitat, resulting in reductions and/or extinctions of distinctive wild populations from this region (Li et al., 2005). Group SWC is another region worthy of concern, where we detected a high number of private alleles, although only a limited number of samples were included in this study from SWC. The populations of NL in Yunnan Province and CY in Xizang Province are isolated from the main *G. soja* distribution area by more than 400 km in linear distance, and wild soybean from population CY is an important breeding resource because of its high protein contents.

4.5. Conclusion

We used two loci to investigate the distribution patterns and possible refugia of wild soybean and the results revealed that the climatic oscillations during the Quaternary profoundly affected the geographical distribution patterns of *G. soja*. The haplotype distribution

patterns suggested that Northeast Asia (NEC and Korea) and the Yangzi River basin (MDYR and east end of SWC) might have been the refugia for *G. soja* during the Quaternary glacial period, and *G. soja* expanded rapidly after the LGM. There was a land bridge among East China, the Korean peninsula, and Japan that permitted gene flow among populations of these regions in the ice age during the Quaternary. We inferred that NEC and MDYR are the key regions for the protection of *G. soja* germplasm.

References

- Aizawa M, Yoshimaru H, Saito H, Katsuki T, Kawahara T, Kitamura K, Shi F, Kaji M (2007). Phylogeography of a northeast Asian spruce, *Picea jezoensis*, inferred from genetic variation observed in organelle DNA markers. *Mol Ecol* 16: 3393-3405.
- An Z (2000). The history and variability of the East Asian paleomonsoon climate. *Quaternary Sci Rev* 19: 171-187.
- Avise JC (1994). *Molecular Markers, Natural History and Evolution*. New York, NY, USA: Chapman and Hall.
- Bai WN, Liao WJ, Zhang DY (2010). Nuclear and chloroplast DNA phylogeography reveal two refuge areas with asymmetrical gene flow in a temperate walnut tree from East Asia. *New Phytol* 188: 892-901.
- Baloch FS, Alsaleh A, Andeden EE, Hatipoğlu R, Nachit M, Özkan H (2016). High levels of segregation distortion in the molecular linkage map of bread wheat representing the West Asia and North Africa region. *Turk J Agric For* 40: 352-364.
- Baloch FS, Kurt C, Arioğlu H, Özkan H (2010). Assaying of diversity among soybean (*Glycine max* (L.) Merr.) and peanut (*Arachis hypogaea* L.) genotypes at DNA level. *Turk J Agric For* 34: 285-301.
- Bandelt HJ, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37-48.
- Burban C, Petit RJ, Carcreff E, Jactel H (1999). Rangewide variation of the maritime pine bast scale *Matsucoccus feytaudi* Duc. (Homoptera: Matsucoccidae) in relation to the genetic structure of its host. *Mol Ecol* 8: 1593-1602.
- Choi IY, Kang JH, Song HS, Kim NS (1999). Genetic diversity measured by simple sequence repeat variations among the wild soybean, *Glycine soja*, collected along the riverside of five major rivers in Korea. *Genes Genet Syst* 74: 169-177.
- Chung CH (2007). Vegetation response to climate change on Jeju Island, South Korea, during the last deglaciation based on pollen record. *Geosci J* 11:147-155.
- Dobson M, Kawamura Y (1998). Origin of the Japanese land mammal fauna: Allocation of extant species to historically-based categories. *Quaternary Res* 37: 385-395.
- Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11-15.
- Frankel OH, Soulé ME (1981). *Conservation and Evolution*. Cambridge, UK: Cambridge University Press.
- Fujita R, Ohara M, Okazaki K, Shimamoto Y (1997). The extent of natural cross-pollination in wild soybean (*Glycine soja*). *J Hered* 88: 124-128.
- Gao LM, Möller M, Zhang XM, Hollingsworth ML, Liu J, Mill RR, Gibby M, Li DZ (2007). High variation and strong phylogeographic pattern among cpDNA haplotypes in *Taxus wallichiana* (Taxaceae) in China and North Vietnam. *Mol Ecol* 16: 4684-4698.
- Gong W, Chen C, Dobe C, Fu CX, Koch MA (2008). Phylogeography of a living fossil: Pleistocene glaciations forced *Ginkgo biloba* L. (Ginkgoaceae) into two refuge areas in China with limited subsequent postglacial expansion. *Mol Phylogenet Evol* 48: 1094-1105.
- Guan BC, Fu CX, Qiu YX, Zhou SL, Comes HP (2010). Genetic structure and breeding system of a rare understory herb, *Dysosma versipellis* (Berberidaceae), from temperate deciduous forests in China. *Am J Bot* 97: 111-122.
- Hall TA (1999). BIOEDIT: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41: 95-98.
- Harpending H (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum Biol* 66: 591e600.
- Harrison SP, Yu G, Takahara H, Prentice IC (2001). Diversity of temperate plants in East Asia. *Nature* 413: 129-130.
- Hewitt GM (1996). Some genetic consequences of ice ages and their role in divergence and speciation. *Biol J Linn Soc* 58: 247-276.
- Hewitt GM (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405: 907-913.
- Hufford KM, Mazer SJ (2003). Plant ecotypes: genetic differentiation in the age of ecological restoration. *Trends Ecol Evol* 18: 147-155.
- Kiang YT, Chiang YC, Kaizuma N (1992). Genetic diversity in natural populations of wild soybean in Iwate prefecture. *Japan J Hered* 83: 325-329.
- Kuroda Y, Kaga A, Tomooka N, Vaughan DA (2006). Population genetic structure of Japanese wild soybean (*Glycine soja*) based on microsatellite variation. *Mol Ecol* 15: 959-974.
- Li EY, Yi S, Qiu YX, Guo JT, Comes HP, Fu CX (2008). Phylogeography of two East Asian species in *Crotonia* (Stemonaceae) inferred from chloroplast DNA and ISSR fingerprinting variation. *Mol Phylogenet Evol* 49: 702-714.

Acknowledgments

This research was supported by the National Natural Science Fund (31560091) and the Key Project of the Education Department of Guizhou Province [KY(2013)186]. The authors thank Okada Hiroshi and Naoko Ishikawa for providing Japanese samples and Chunghee Lee for providing Korean samples.

- Li X, Wang K, Li F, Yan M (2005). Research progress of wild soybean (*Glycine soja*) and suggestions for improving its effective utilization and protection. *Soybean Sci* 24: 305-309 (in Chinese with English abstract).
- Liu KB (1988). Quaternary history of the temperate forests of China. *Quaternary Sci Rev* 7: 1-20.
- Liu Y, Wang Y, Huang H (2009). Species-level phylogeographical history of *Myricaria* plants in the mountain ranges of western China and the origin of *M. laxiflora* in the Three Gorges mountain region. *Mol Ecol* 18: 2700-2712.
- Lomolino MV, Riddle BR, Brown JH (2006). *Biogeography*. 3rd ed. Sunderland, MA, USA: Sinauer Associates.
- Lu BR (2004). Conserving biodiversity of soybean gene pool in the biotechnology era. *Plant Spec Biol* 19: 115-125.
- Lu HY, Yi SW, Xu ZW, Zhou YL, Zeng L, Zhu FY, Feng H, Dong LN, Zhuo H, Yu KF et al (2013). Chinese deserts and sand fields in Last Glacial Maximum and Holocene Optimum. *Chin Sci Bull* 58: 2775-2783.
- Millien-Parra V, Jaeger JJ (1999). Island biogeography of the Japanese terrestrial mammal assemblages: an example of a relict fauna. *J Biogeography* 26: 959-972.
- Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D, Vendramin GG (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol Ecol* 14: 689-701.
- Qian H, Ricklefs RE (2001). Palaeovegetation - Diversity of temperate plants in East Asia - Reply. *Nature* 413: 130.
- Qiu YX, Fu CX, Comes HP (2011). Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. *Mol Phylogenet Evol* 59: 225-244.
- Qiu YX, Guan BC, Fu CX (2009c). Did glacials and/or interglacials promote allopatric incipient speciation in East Asian temperate plants? Phylogeographic and coalescent analyses on refugial isolation and divergence in *Dysosma versipellis*. *Mol Phylogenet Evol* 51: 281-293.
- Qiu YX, Sun Y, Zhang XP (2009a). Molecular phylogeography of East Asian *Kirengeshoma* (Hydrangeaceae) in relation to Quaternary climate change and land bridge configurations. *New Phytol* 183: 480-495.
- Qiu YX, Qi XS, Jin XF (2009b). Population genetic structure, phylogeography, and demographic history of *Platycrater arguta* (Hydrangeaceae) endemic to East China and South Japan, inferred from chloroplast DNA sequence variation. *Taxon* 58: 1226-1241.
- Rauch EM, Bar-Yam Y (2005). Estimating the total genetic diversity of a spatial field population from a sample and implications of its dependence on habitat area. *P Natl Acad Sci USA* 102: 9826-9829.
- Schmitt T (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Front Zool* 4: 11.
- Schrimpf A, Theissinger K, Dahlem J (2014). Phylogeography of noble crayfish (*Astacus astacus*) reveals multiple refugia. *Freshwater Biol* 59: 761-776.
- Shahid MQ, Saleem MF, Khan HZ, Anjum SA (2009). Performance of soybean (*Glycine max* L.) under different phosphorus levels and inoculation. *Pak J Agri Sci* 46: 1-5.
- Shaw J, Lickey EB, Schilling EE, Small RL (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. *Am J Bot* 94: 275-288.
- Shi Y (1989). Discussion on the reconstruction of paleosnowline during glacial period. In: Shi Y, Cui ZJ, Li JJ, editors. *Problems on Quaternary Glaciation and Environments in Eastern China*. Beijing, China: Science Press, pp. 363-374 (in Chinese with English abstract).
- Shi Y (2002). Characteristics of late Quaternary monsoonal glaciation on the Tibetan Plateau and in East Asia. *Quatern Int* 97-98: 79-91.
- Simpson GG (1940). Mammals and land bridges. *J Washington Acad Sci* 30: 137-163.
- Smil V (2000). Magic beans. *Nature* 407: 467.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24: 4876-4882.
- Tian B, Liu R, Wang L, Qui Q, Chen K, Liu J (2009). Phylogeographic analyses suggest that a deciduous species (*Ostryopsis davidiana* Decne., Betulaceae) survived in northern China during the last Glacial Maximum. *J Biogeogr* 36: 2148-2155.
- Van K, Hwang EY, Kim MY (2004). Discovery of single nucleotide polymorphisms in soybean using primers designed from ESTs. *Euphytica* 139: 147-157.
- Wang J, Gao P, Kang M (2009). Refugia within refugia: the case study of a canopy tree (*Eurycorymbus cavaleriei*) in subtropical China. *J Biogeogr* 36: 2156-2164.
- Wang KJ, Li XH, Li FS (2008). Phenotypic diversity of the big seed type subcollection of wild soybean (*Glycine soja* Sieb. et Zucc.) in China. *Genet Resour Crop Ev* 55: 1335-1346.
- Wang YS, Shahid MQ, Ghouri F, Baloch FS, Wang Y, Huang H (2015a). Evaluation of the geographical pattern of genetic diversity of *Glycine soja* and *Glycine max* based on four single copy nuclear gene loci: for conservation of soybean germplasm. *Biochem Syst Ecol* 62: 229-235.
- Wang YS, Shahid MQ, Huang H, Wang Y (2015b). Nucleotide diversity patterns of three divergent soybean populations: evidences for population-dependent linkage disequilibrium and taxonomic status of *Glycine gracilis*. *Ecol Evol* 5: 3969-3978.
- Wu ZY, Wu SG (1998). A proposal for a new floristic kingdom (realm) the Asiatic kingdom, its delineation and characteristics. In: Zhang AL, Wu SG, editors. *Floristic Characteristics and Diversity of East Asian Plants*. Beijing, China: China Higher Education Press & Springer-Verlag, pp. 3-42.
- Xia X, Xie Z (2001). DAMBE: Software package for data analysis in molecular biology and evolution. *J Hered* 92: 371-373.
- Yu G, Chen X, Ni J, Cheddadi R, Guiot J, Han H, Harrison SP, Huang C, Ke M, Kong Z et al (2000). Palaeovegetation of China: a pollen data-based synthesis for the mid-Holocene and last glacial maximum. *J Biogeogr* 27: 635-664.
- Zheng Z, Yuan B, Petit-Maire N (1998). Palaeoenvironments in China during the Last Glacial Maximum and the Holocene Optimum. *Episodes* 21: 152-158.