# Analysis of average standardized SSR allele size supports domestication of soybean along the Yellow River

Genetic Resources and Crop Evolution

Li, Y.H.; Zhang, C.; Smulders, M.J.M.; Li, W.; Ma, Y.S. et al

https://doi.org/10.1007/s10722-012-9873-z

Genet Resour Crop Evol
DOI 10.1007/s10722-012-9873-z

RESEARCH ARTICLE

# Analysis of average standardized SSR allele size supports domestication of soybean along the Yellow River

**Ying-hui Li · Chen Zhang · Marinus J. M. Smulders · Wei Li · Yan-song Ma · Qu Xu · Ru-zhen Chang · Li-juan Qiu**

Received: 16 December 2011 / Accepted: 21 May 2012


**Abstract** Soybean (*Glycine max*) was domesticated in China from its wild progenitor *G. soja*. The geographic region of domestication is, however, not exactly known. Here we employed the directional evolution of SSR (microsatellite) repeats (which mutate preferentially into longer alleles) to analyze the domestication process and to infer the most ancestral soybean landraces. In this study, the average standardized SSR allele sizes across 42 SSR loci in 62 accessions of *G. soja* were determined, and compared with those in 1504 landraces of *G. max*, collected from all over China and representing the diversity in the gene bank. The standardized SSR allele size in the landraces (0.009) was significantly ($P = 8.63 \times 10^{-58}$) larger than those in *G. soja* (−0.406). Pairwise comparisons between inferred clusters and sub-clusters of Chinese landraces indicated that the average standardized SSR allele size also increased with the further differentiation of landraces populations. Spring-sowed types had the shortest size, followed by summer-sown types, while the sub-cluster of autumn-sown type had the largest length. The spring-sowed landraces located near the middle region along the Yellow River had the smallest allele sizes, indicating that this is the most ancestral population of cultivated soybean. We concluded that soybean was most likely domesticated in the middle region of the Yellow River in central China, initially as a spring-sown type.

**Keywords** Directional evolution · Domestication · *Glycine max* · *Glycine soja* · Landrace · SSR

**Abbreviations**

| | |
|---|---|
| NESpM | NorthEast region, Spring-sowing type, Model-based |
| NSpM | North region, Spring-sowing type, Model-based |
| HSuM | Huanghuai (Yellow river) region, Summer-sowing type, Model-based |
| SSpM | South region, Spring-sowing type, Model-based |
| SSpSM | South region, Spring-sowing type, Southwest, Model-based |

Y. Li · C. Zhang · W. Li · R. Chang · L. Qiu (✉)
The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI), Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, People's Republic of China
e-mail: qiu_lijuan@263.net

Y. Li
e-mail: dadousoybean@yahoo.com.cn

M. J. M. Smulders
Plant Research International, Wageningen UR Plant Breeding, P.O. Box 16, 6700 AA Wageningen, The Netherlands

Y. Ma
Soybean Research Institute, Heilongjiang Academy of Agricultural Sciences (HAAS), Harbin 150086, People's Republic of China

Q. Xu
High School Attached to Capital Normal University, Beijing 100048, People's Republic of China

Published online: 23 June 2012

| | |
|---|---|
| SSuM | South region, Summer-sowing type, Model-based |
| SSuSM | South region, Summer-sowing type, Southwest, Model-based |

## Introduction

Soybean (*Glycine max* (L.) Merr.) is one of most important crops for food, feed, industrial materials and bio-energy world-wide. Most workers would agree that soybean was domesticated from its annual wild progenitor *Glycine soja* Sieb. et Zucc., although an undomesticated population of *G. max* was predicted as the direct ancestor of the domesticated one based on a comparison of the whole genome sequence of *G. max* and *G. soja* (Kim et al. 2010). The distribution of *G. soja* is limited to China, Japan, Korea, and the Far East Area of Russia. It is generally accepted that cultivated *G. max* was domesticated in China (Fukuda 1933; Vavilov 1951; Hymowitz and Newell 1981) approximately 5,000 years ago (Carter et al. 2004), based on linguistic, geographical, and historical evidence and on morphological or molecular analysis. Most studies suggested that cultivated soybean was domesticated only once, except Xu et al. (2002) who proposed domestication on multiple sites. However, the exact geographic region of origin has not been defined for cultivated soybean, although Northern-East China (Fukuda 1933; Li 1994), the Yellow River valley (Central China) (Zhao and Gai 2004; Dong et al. 2004; Hymowitz and Newell 1981; Vavilov 1951; Chang 1989; Li et al. 2008, 2010a), Southern China (Gai et al. 2000; Guo et al. 2010) have been proposed as candidate locations.

Simple sequence repeats (SSRs) or microsatellites are used to infer evolutionary process of crops using diversity and phylogenetic analysis. Recently the directional evolution of SSRs has been used to analyze species origins (Anthony et al. 2002; Maughan et al. 1995; Matsuoka et al. 2002; Grassi et al. 2003; Vigouroux et al. 2003; Garris et al. 2005). The phenomenon of directional evolution of SSRs was first observed between humans and nonhuman primates (Rubinsztein et al. 1995). The bias of SSR repeats to mutate towards a larger repeat size rather than to become shorter may lead to larger average allele sizes in "derived" groups (Petit et al. 2005). Vigouroux et al. (2003) showed that there was an increase in the average allele size of SSRs in the geographically derived North and South American maize groups relative to the ancestral Mexican group. Garris et al. (2005) demonstrated that the average standardized allele size was greater in temperate than in tropical japonica (*Oryza sativa* L. ssp. japonica Katoi), supporting the hypothesis that the temperate japonica group was derived from the tropical japonica group. Recently, this change in SSRs was used as a molecular clock to identify phylogenetic relationships during the speciation process in the genus *Secale* (Ren et al. 2011).

Soybean in China has the most extensive distribution and the richest genetic diversity in the world. In order to accelerate evaluation and utilization of soybean germplasm, a primary core collection was constructed with 70–80 % of genetic variation present in the original entire collection (Qiu et al. 2003, 2009). In previous studies, the hierarchical patterns of population structure of landraces in the soybean primary core collection in China have been clarified (Li et al. 2008, 2010b). Seven clusters and twenty sub-clusters were inferred based on SSR analysis. The subdivisions, not only on the cluster level but also on the sub-cluster level, were generally in accordance with differences in geographic origin and sowing seasons. The distribution pattern of genetic diversity among clusters or among sub-clusters was uneven. The HSuM (Huanghuai region, Summer-sowing type, Modeled) sub-cluster from central China showed a significantly higher level of genetic diversity. Based on the genetic structure and diversity analysis, the central and downstream parts of the Yellow River (in central China) were inferred as the candidate domestication site of cultivated soybean. However we also discovered some "noise", for example, some South China clusters or sub-clusters also showed significantly highly level of genetic diversity.

In this study, we employed the directional evolution of SSRs length on a large sample with 1566 soybean accessions to clarify the ancestral domestication of cultivated soybean in China. The genetic diversity in *G. soja* was determined and compared with those of soybean landrace clusters and sub-clusters. Thus, we provided molecular evidence for the domestication site of soybean.

## Materials and methods

### Plant material

A total of 1566 accessions were analyzed in this study, including 1504 *G. max* landraces and 62 *G. soja* accessions (Table 1) collected evenly across China. Based on 59 SSR markers the 1504 *G. max* landraces have been assigned to seven clusters using the program Structure (Li et al. 2008), named NESpM (NorthEast region, Spring-sowing type, Model-based), NSpM (North region, Spring-sowing type, Model-based), HSuM (Huanghuai region, Summer-sowing type, Model-based), SSpM (South region, Spring-sowing type, Model-based), SSpM (South region, Spring-sowing type, Southwest, Model-based), SSuM (South region, Summer-sowing type, Model-based) and SSuSM (South region, Summer-sowing type, Southwest, Model-based). The clusters coincided partly with a geographic structure (north, central and south China) and partly with type of soybean (sown in spring or summer). In the follow-up study, Li et al. (2010b) distinguished twenty sub-clusters within these seven clusters, namely NESpM1-4, NSpM1-3, HSuM1 and 2, SSpM1 and 2, SSpSM1 and 2, SSuM1-4, and SSuSM1-3 (Fig. 1).

The set of 62 *G. soja* accessions (Table 1) were analyzed in the present study as representatives of the ancestral state, which were obtained from the Chinese National Soybean GeneBank (CNSGB). In order to be able to integrate the *G. soja* SSR data with the dataset of *G. max*, which already existed, two landraces (Huipizhiheidou and Xiaolimoshidou) from the 1504 were included during SSR genotyping of *G. soja* with the same 59 SSR markers. As 17 SSR loci had more than 10 % missing data and were subsequently deleted from the dataset, all analysis were based on data from 42 SSR loci (Table 2). Nearly all of these consist of (ATT)n trinucleotide repeats, and their alleles fit in a perfect stepwise series.

### SSR genotyping

For each *G. soja* accession, the second or third young trifoliate leave was harvested from greenhouse-grown seedling. DNA extraction and PCR amplification followed the procedure described by (Wang et al. 2006). SSR marker detection was performed using a HAD-GT12 capillary electrophoresis system with the GCK-5000 gel cartridge kit (eGene Inc.) as described (Liu and Amirkhanian 2003). Allele binning was performed by rounding off the BioCalculator software assigned values to the nearest whole base-pair integer to give a base pair estimate for the allele. To ensure the veracity of the dataset, randomly sampled SSR amplification products (10 % of the total) were also resolved by electrophoresis in 7 % denaturing polyacrylamide gels containing 8.5 M urea and visualized using silver staining. Here the bins were assigned manually. The data from eGene were compared to and where necessary corrected with the result from polyacrylamide gels.

### Genetic diversity

Since sample sizes of two species were largely different, the number of distinct alleles and private alleles (that is, not found in other populations) of *G. soja* and *G. max* was estimated by ADZE (Szpiech et al. 2008), which employed a rarefaction approach to obtain sample-size corrected estimates. Effective number, observed heterozygosity and expected heterozygosity were estimated using POPGENE 1.31 (available at http://www.ualberta.ca/~fyeh/index.htm).

### Phylogenetic analysis

Phylogenetic reconstruction for sub-clusters of soybean landraces rooted with *G. soja* population was based on the neighbor-joining method (with 1000 bootstraps) implemented in PowerMarker (Liu and Muse 2005) and displayed by Mega4 (Tamura et al. 2007).

### Directional evolution for SSR size

Average standardized allele sizes in each population were calculated as in Vigouroux et al. (2003). First, the average allele size and its standard deviation were calculated for each locus. Then the standardized size of the alleles for each landraces at the locus was calculated as [actual allele size minus the mean allele size]/[standard deviation]. Finally, the average individual size of its SSRs as the mean of the standardized size of 42 SSR loci were calculated for a sample of 50 *G. soja* accessions and a sample of 50 landraces which were randomly drawn 30 times from 62 *G. soja*

**Table 1** The geographical distribution pattern of the 62 *G. soja* accessions sampled across China and analyzed in this study

| Accession number in CNSGB genebank | Longitude (°E) | Latitude (°N) | Altitude (m) | Origin (Province) |
|---|---|---|---|---|
| ZYD00755 | 121.2 | 46.0 | 499 | Jilin |
| ZYD00847 | 125.4 | 44.3 | 185 | Jilin |
| ZYD00851 | 125.4 | 44.3 | 185 | Jilin |
| ZYD01147 | 123.3 | 43.3 | 115 | Jilin |
| ZYD01550 | 125.4 | 42.2 | 362 | Jilin |
| ZYD01712 | 124.0 | 42.3 | 98 | Liaoning |
| ZYD01825 | 124.0 | 42.3 | 98 | Liaoning |
| ZYD02391 | 122.3 | 42.2 | 79 | Liaoning |
| ZYD02677 | 125.2 | 41.2 | 240 | Liaoning |
| ZYD02739 | 117.6 | 40.6 | 311 | Hebei |
| ZYD02749 | 119.1 | 39.4 | 13 | Hebei |
| ZYD02750 | 119.4 | 39.6 | 2 | Hebei |
| ZYD02797 | 105.4 | 37.3 | 1,185 | Ningxia |
| ZYD02798 | 106.4 | 38.5 | 1,102 | Ningxia |
| ZYD03100 | 112.5 | 36.5 | – | Shanxi |
| ZYD03120 | 112.3 | 35.3 | – | Shanxi |
| ZYD03149 | 110.4 | 35.4 | – | Shanxi |
| ZYD03217 | 110.5 | 36.3 | – | Shanxi |
| ZYD03223 | 122.2 | 37.1 | 38 | Shandong |
| ZYD03247 | 118.3 | 36.4 | 80 | Shandong |
| ZYD03262 | 118.1 | 37.3 | 11 | Shandong |
| ZYD03294 | 104.4 | 36.3 | 1,300 | Gansu |
| ZYD03299 | 108.2 | 35.3 | 1,250 | Gansu |
| ZYD03313 | 106.4 | 35.3 | 1,340 | Gansu |
| ZYD03386 | 114.3 | 35.4 | 65 | Henan |
| ZYD03390 | 112.3 | 35.3 | 52 | Henan |
| ZYD03489 | 113.0 | 33.5 | 136 | Henan |
| ZYD03658 | 114.0 | 34.4 | 78 | Henan |
| ZYD03670 | 114.0 | 34.4 | 78 | Henan |
| ZYD03809 | 110.3 | 35.3 | – | Shaanxi |
| ZYD03888 | 107.1 | 34.2 | – | Shaanxi |
| ZYD04073 | 109.0 | 32.4 | 220 | Shaanxi |
| ZYD04104 | 118.5 | 34.1 | 8 | Jiangsu |
| ZYD04132 | 119.0 | 33.4 | 16 | Jiangsu |
| ZYD04159 | 119.2 | 32.2 | 7 | Jiangsu |
| ZYD04200 | 115.5 | 33.5 | 36 | Anhui |
| ZYD04217 | 117.5 | 33.1 | 15 | Anhui |
| ZYD04231 | 116.4 | 32.4 | 19 | Anhui |
| ZYD04318 | 106.5 | 32.2 | 1,400 | Sichuan |
| ZYD04320 | 108.1 | 32.1 | 850 | Sichuan |
| ZYD04321 | 108.4 | 31.6 | 1,040 | Sichuan |
| ZYD04348 | 103.6 | 30.3 | 664 | Sichuan |
| ZYD04351 | 115.1 | 29.5 | – | Hubei |
| ZYD04366 | 110.0 | 29.5 | – | Hubei |
| ZYD04398 | 110.5 | 32.5 | – | Hubei |

**Table 1** continued

| Accession number in CNSGB genebank | Longitude (°E) | Latitude (°N) | Altitude (m) | Origin (Province) |
|---|---|---|---|---|
| ZYD04410 | 111.3 | 30.2 | – | Hubei |
| ZYD04432 | 120.5 | 29.4 | 50 | Zhejiang |
| ZYD04507 | 120.2 | 30.1 | 50 | Zhejiang |
| ZYD04551 | 121.1 | 30.0 | 50 | Zhejiang |
| ZYD04569 | 120.3 | 27.4 | 250 | Zhejiang |
| ZYD04601 | 115.5 | 29.3 | 32 | Jiangxi |
| ZYD04631 | 116.5 | 28.1 | 40 | Jiangxi |
| ZYD04672 | 113.3 | 28.5 | 27 | Hunan |
| ZYD04673 | 111.4 | 29.0 | 33 | Hunan |
| ZYD04680 | 109.3 | 29.3 | 1,200 | Hunan |
| ZYD04684 | 111.3 | 28.5 | 215 | Hunan |
| ZYD04765 | 108.2 | 28.0 | 455 | Guizhou |
| ZYD04783 | 107.3 | 26.2 | 810 | Guizhou |
| ZYD05173 | 100.5 | 27.2 | 2,670 | Yuannan |
| ZYD05704 | 110.5 | 39.4 | 1,000 | Neimenggu |
| ZYD05715 | 110.5 | 39.4 | 1,000 | Neimenggu |
| ZYD05775 | 112.3 | 37.5 | 778 | Shanxi |

*CNSGB* Chinese National Soybean GeneBank

accessions and 1504 *G. max* landraces respectively. The standardization ensured that each locus contributed equally to the average individual size.

To study differences in average SSR allele length between the two species and the clusters within *G. max*, we randomly drew 30 times a sample of 50 landraces for each of the seven model-based clusters, and 30 times a sample of 25 landraces for each of the 20 sub-clusters and calculated the average value. These average standardized allele sizes among seven clusters were compared pairwise with each other and with the average of 30 replications of samples of 50 *G. soja* accessions, using a *t* test. And the average standardized allele sizes among 20 sub-clusters were also compared pairwise with each other and with the average of 30 replications of samples of 25 *G. soja* accessions, using a *t* test.

## Results

### Diversity statistics

Number of alleles, effective number of alleles, observed heterozygosity and expected heterozygosity (gene diversity) of two species in each locus were shown in Table 2. Across all 42 SSR loci, we found fewer alleles (697) in *G. soja* than that in *G. max* (860). The *G. soja* accessions formed a much smaller sample (62 accessions) than *G. max* (1504 landraces). To correct for the difference in sample size between the two species, the mean numbers of distinct and private alleles per locus were analyzed as a function of the sample size for the two species (Fig. 2). With increasing sample size the number of distinct and private allele also increased, but now it is clear that *G. soja* always exhibited a higher value than *G. max*. Moreover, *G. soja* had a much higher mean effective number of alleles (9.8) and gene diversity (0.892) than *G. max* (6.8 and 0.827 respectively). These results illustrated that *G. max* was less genetically diverse than *G. soja*, and that genetic diversity had been lost during domestication of *G. max* from *G. soja*.

### Genetic relationships among sub-clusters of landraces and wild soybean

The NJ tree was generated from pairwise Nei's genetic distances between sub-clusters (Fig. 3). Viewed from *G. soja*, the 20 sub-clusters firstly grouped into two clades, one of which consisted of 11 sub-clusters from south China and the other one consisted of nine sub-clusters from north China. As expected, most of the sub-clusters from one cluster grouped together. Only

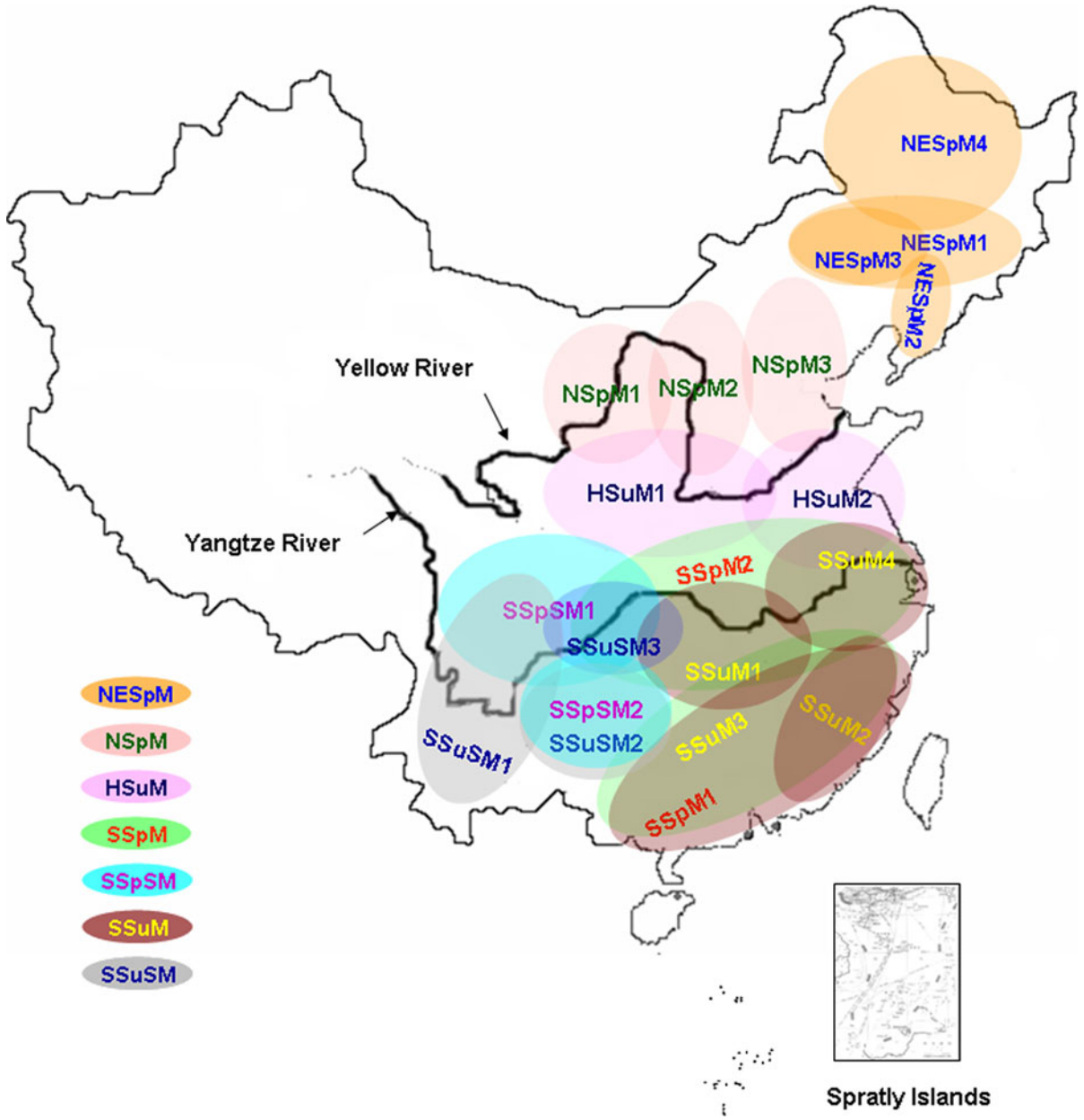


**Fig. 1** The geographical distribution of 20 inferred sub-clusters of soybean landraces in China, which belong to seven inferred clusters including NESpM (NorthEast region, Spring-sowing type, Model-based), NSpM (North region, Spring-sowing type, Model-based), HSuM (Huanghuai (Yellow River) region, Summer-sowing type, Model-based), SSpM (South region, Spring-sowing type, Model-based), SSpSM (South region, Spring-sowing type, Southwest, Model-based), SSuM (South region, Summer-sowing type, Model-based), and SSuSM (South region, Summer-sowing type, Southwest, Model-based)

SSuM4, a summer-sowing sub-cluster from the North of Yangtze River, was grouped with two spring-sowing types from South China. This result suggested that the geographical differentiation was the primary factor for forming genetic differentiation within Chinese landraces, followed by sowing seasons. In the 20 landrace sub-clusters, HSuM1 showed the closest relationship with *G. soja* (Nei's genetic identity, 0.5281), followed by HSuM1 (Nei's genetic identity, 0.5257). And NESpM1 had a most distant relationship

**Table 2** Description of the population-genetic statistics of 42 SSR loci for the *G. soja* and *G. max* dataset

| Locus | Linkage group | cM Position in LG | Core motif | *G. soja* (62 accessions) | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | No. of alleles | Size range | Effective number of alleles | Observed heterozygosity | Expected heterozygosity | Average standardized allele sizes |
| Sat_099 | L | 78.23 | $(AT)_{25}$ | 22 | 223–287 | 16.7 | 0.137 | 0.948 | −0.707 |
| Sat_112 | E | 8.67 | $(AT)_8(GA)_{18}$ | 23 | 297–377 | 14.4 | 0.094 | 0.94 | 0.136 |
| Satt002 | D2 | 47.73 | $(ATT)_{25}$ | 15 | 102–163 | 10.2 | 0.016 | 0.91 | −0.230 |
| Satt005 | D1b + W | 75.29 | $(ATT)_{19}$ | 19 | 118–180 | 15.6 | 0.017 | 0.944 | −1.224 |
| Satt012 | G | 66.55 | $(ATT)_{19}$ | 23 | 104–181 | 16.0 | 0.000 | 0.946 | 0.011 |
| Satt022 | N | 102.06 | $(ATT)_{17}$ | 12 | 159–193 | 8.5 | 0.000 | 0.889 | 0.339 |
| Satt130 | C2 | 23.1 | $(ATT)_{14}$ | 14 | 212–328 | 8.9 | 0.089 | 0.896 | 0.041 |
| Satt146 | F | 1.92 | $(ATT)_{17}$ | 11 | 280–312 | 5.5 | 0.083 | 0.826 | −0.838 |
| Satt168 | B2 | 55.2 | $(ATT)_{16}$ | 18 | 183–256 | 7.3 | 0.179 | 0.87 | −0.644 |
| Satt180 | C1 | 127.77 | $(ATT)_{16}$ | 16 | 210–261 | 6.4 | 0.113 | 0.85 | −1.297 |
| Satt184 | D1a + Q | 17.52 | $(ATT)_{13}$ | 16 | 143–209 | 5.8 | 0.000 | 0.835 | 1.209 |
| Satt194 | C1 | 26.35 | $(ATT)_{23}$ | 13 | 200–251 | 10.6 | 0.086 | 0.914 | −0.803 |
| Satt197 | B1 | 46.39 | $(ATT)_{20}$ | 17 | 135–219 | 4.7 | 0.117 | 0.794 | −0.728 |
| Satt216 | D1b + W | 9.8 | $(ATT)_{19}$ | 24 | 140–240 | 9.8 | 0.016 | 0.905 | −0.453 |
| Satt226 | D2 | 85.15 | $(ATT)_{18}$ | 20 | 299–398 | 11.9 | 0.033 | 0.924 | 0.988 |
| Satt230 | E | 71.31 | $(ATT)_{15}$ | 15 | 204–269 | 5.9 | 0.018 | 0.838 | −0.153 |
| Satt236 | A1 | 93.2 | $(ATT)_{19}$ | 14 | 196–245 | 9.7 | 0.066 | 0.905 | −0.202 |
| Satt239 | I | 36.94 | $(ATT)_{22}$ | 16 | 145–192 | 11.8 | 0.125 | 0.923 | −2.578 |
| Satt243 | O | 119.5 | $(ATT)_{17}$ | 19 | 170–234 | 11.4 | 0.035 | 0.921 | −0.512 |
| Satt267 | D1a + Q | 57.34 | $(ATT)_{16}$ | 12 | 217–256 | 6.6 | 0.017 | 0.856 | −0.025 |
| Satt268 | E | 44.27 | $(ATT)_{17}$ | 15 | 203–251 | 9.0 | 0.033 | 0.896 | 0.189 |
| Satt279 | H | 68.5 | $(ATT)_{28}$ | 15 | 153–195 | 10.3 | 0.000 | 0.911 | −1.856 |
| Satt281 | C2 | 40.3 | $(ATT)_{19}$ | 21 | 164–241 | 14.7 | 0.018 | 0.941 | −0.170 |
| Satt307 | C2 | 121.27 | $(ATT)_{12}$ | 17 | 143–204 | 9.2 | 0.000 | 0.899 | −0.324 |
| Satt308 | M | 130.8 | $(ATT)_{21}$ | 16 | 118–180 | 10.1 | 0.000 | 0.909 | −0.798 |
| Satt309 | G | 4.53 | $(ATT)_{13}$ | 8 | 127–154 | 4.2 | 0.048 | 0.766 | 0.687 |
| Satt345 | O | 59.43 | $(ATT)_{27}$ | 12 | 198–273 | 4.6 | 0.000 | 0.788 | −0.709 |
| Satt346 | M | 112.8 | $(ATT)_{17}$ | 13 | 186–227 | 6.6 | 0.000 | 0.856 | 0.000 |
| Satt352 | G | 50.53 | $(ATT)_{18}$ | 14 | 151–201 | 7.7 | 0.051 | 0.877 | −2.259 |
| Satt414 | J | 37.04 | $(ATT)_{23}$ | 21 | 246–324 | 12.2 | 0.036 | 0.927 | −0.951 |
| Satt429 | A2 | 162 | $(ATT)_{25}$ | 13 | 228–290 | 8.0 | 0.104 | 0.883 | −0.400 |
| Satt431 | G | 78.57 | $(ATT)_{21}$ | 17 | 179–235 | 10.8 | 0.148 | 0.915 | −0.952 |
| Satt442 | H | 46.95 | $(ATT)_{35}$ | 22 | 211–284 | 14.7 | 0.000 | 0.94 | −0.811 |
| Satt487 | O | 9.53 | $(ATT)_{22}$ | 13 | 182–226 | 8.9 | 0.018 | 0.896 | −0.105 |
| Satt530 | N | 32.85 | $(ATT)_{12}$ | 17 | 190–267 | 11.2 | 0.016 | 0.918 | 0.293 |
| Satt556 | B2 | 73.21 | $(ATT)_{14}$ | 16 | 157–227 | 8.5 | 0.109 | 0.89 | −0.113 |
| Satt571 | I | 18.5 | $(ATT)_{14}$ | 15 | 123–184 | 9.5 | 0.000 | 0.902 | 1.217 |
| Satt586 | F | 3.63 | $(ATT)_{19}$ | 16 | 183–241 | 8.3 | 0.050 | 0.887 | 0.591 |
| Satt588 | K | 117.02 | $(ATT)_{18}$ $(AT)_{10}(CT)_{14}$ | 17 | 103–177 | 10.6 | 0.000 | 0.913 | −1.371 |
| Satt590 | M | 7.84 | $(ATT)_{26}$ | 25 | 253–249 | 13.9 | 0.017 | 0.936 | −0.519 |
| Satt596 | J | 39.64 | $(ATT)_{17}$ | 13 | 214–266 | 6.3 | 0.000 | 0.847 | −0.762 |
| Sct189 | I | 113.8 | $(CT)_{17}$ | 22 | 153–214 | 14.9 | 0.017 | 0.941 | −0.104 |
| Mean | | | | 16.6 | | 9.8 | 0.045 | 0.892 | −0.707 |
| SD | | | | 3.9 | | 3.3 | | 0.045 | |

**Table 2** continued

| Locus | Linkage group | cM Position in LG | Core motif | *G. max* (1863 accessions)* | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | No. of alleles | Size range | Effective number of alleles | Observed heterozygosity | Expected heterozygosity | Average standardized allele sizes |
| Sat_099 | L | 78.23 | $(AT)_{25}$ | 36 | 206–299 | 11.50 | 0.021 | 0.9144 | 0.080 |
| Sat_112 | E | 8.67 | $(AT)_8(GA)_{18}$ | 25 | 306–360 | 7.33 | 0.010 | 0.872 | −0.043 |
| Satt002 | D2 | 47.73 | $(ATT)_{25}$ | 13 | 102–154 | 3.64 | 0.018 | 0.721 | 0.001 |
| Satt005 | D1b + W | 75.29 | $(ATT)_{19}$ | 26 | 123–198 | 9.31 | 0.017 | 0.8917 | 0.076 |
| Satt012 | G | 66.55 | $(ATT)_{19}$ | 28 | 101–190 | 7.73 | 0.043 | 0.8662 | −0.016 |
| Satt022 | N | 102.06 | $(ATT)_{17}$ | 17 | 150–196 | 5.27 | 0.010 | 0.8116 | −0.018 |
| Satt130 | C2 | 23.1 | $(ATT)_{14}$ | 30 | 229–396 | 4.25 | 0.003 | 0.7409 | 0.013 |
| Satt146 | F | 1.92 | $(ATT)_{17}$ | 24 | 277–348 | 7.92 | 0.008 | 0.8831 | 0.095 |
| Satt168 | B2 | 55.2 | $(ATT)_{16}$ | 17 | 195–263 | 5.52 | 0.013 | 0.8207 | 0.024 |
| Satt180 | C1 | 127.77 | $(ATT)_{16}$ | 18 | 210–282 | 5.51 | 0.006 | 0.8163 | 0.034 |
| Satt184 | D1a + Q | 17.52 | $(ATT)_{13}$ | 14 | 133–190 | 7.21 | 0.011 | 0.8581 | −0.095 |
| Satt194 | C1 | 26.35 | $(ATT)_{23}$ | 16 | 202–253 | 3.90 | 0.008 | 0.7517 | 0.097 |
| Satt197 | B1 | 46.39 | $(ATT)_{20}$ | 22 | 130–207 | 7.04 | 0.037 | 0.8626 | −0.038 |
| Satt216 | D1b + W | 9.8 | $(ATT)_{19}$ | 28 | 135–247 | 7.93 | 0.020 | 0.8834 | 0.024 |
| Satt226 | D2 | 85.15 | $(ATT)_{18}$ | 22 | 290–371 | 9.69 | 0.007 | 0.8979 | −0.082 |
| Satt230 | E | 71.31 | $(ATT)_{15}$ | 10 | 213–245 | 2.27 | 0.001 | 0.5741 | −0.015 |
| Satt236 | A1 | 93.2 | $(ATT)_{19}$ | 20 | 196–257 | 3.86 | 0.006 | 0.7537 | 0.032 |
| Satt239 | I | 36.94 | $(ATT)_{22}$ | 17 | 151–201 | 5.74 | 0.007 | 0.8361 | 0.025 |
| Satt243 | O | 119.5 | $(ATT)_{17}$ | 18 | 190–240 | 3.91 | 0.020 | 0.7466 | −0.064 |
| Satt267 | D1a + Q | 57.34 | $(ATT)_{16}$ | 13 | 207–253 | 3.55 | 0.031 | 0.7179 | −0.173 |
| Satt268 | E | 44.27 | $(ATT)_{17}$ | 19 | 203–268 | 5.77 | 0.022 | 0.8261 | −0.051 |
| Satt279 | H | 68.5 | $(ATT)_{28}$ | 21 | 150–219 | 6.34 | 0.007 | 0.8497 | 0.122 |
| Satt281 | C2 | 40.3 | $(ATT)_{19}$ | 41 | 158–295 | 14.34 | 0.012 | 0.93 | 0.054 |
| Satt307 | C2 | 121.27 | $(ATT)_{12}$ | 17 | 146–207 | 5.52 | 0.015 | 0.8186 | −0.036 |
| Satt308 | M | 130.8 | $(ATT)_{21}$ | 27 | 123–219 | 10.08 | 0.020 | 0.9043 | −0.018 |
| Satt309 | G | 4.53 | $(ATT)_{13}$ | 6 | 123–148 | 2.68 | 0.010 | 0.624 | 0.023 |
| Satt345 | O | 59.43 | $(ATT)_{27}$ | 20 | 195–258 | 8.01 | 0.011 | 0.8795 | 0.022 |
| Satt346 | M | 112.8 | $(ATT)_{17}$ | 12 | 184–220 | 6.99 | 0.005 | 0.8511 | −0.034 |
| Satt352 | G | 50.53 | $(ATT)_{18}$ | 17 | 151–198 | 5.96 | 0.007 | 0.8327 | 0.028 |
| Satt414 | J | 37.04 | $(ATT)_{23}$ | 28 | 246–339 | 6.67 | 0.021 | 0.8493 | −0.010 |
| Satt429 | A2 | 162 | $(ATT)_{25}$ | 22 | 215–293 | 9.70 | 0.010 | 0.8909 | 0.065 |
| Satt431 | G | 78.57 | $(ATT)_{21}$ | 22 | 184–260 | 8.60 | 0.015 | 0.89 | −0.046 |
| Satt442 | H | 46.95 | $(ATT)_{35}$ | 21 | 213–276 | 9.88 | 0.018 | 0.9011 | 0.021 |
| Satt487 | O | 9.53 | $(ATT)_{22}$ | 13 | 184–220 | 6.06 | 0.011 | 0.8316 | −0.017 |
| Satt530 | N | 32.85 | $(ATT)_{12}$ | 14 | 211–251 | 7.04 | 0.019 | 0.8617 | −0.006 |
| Satt556 | B2 | 73.21 | $(ATT)_{14}$ | 25 | 150–230 | 2.40 | 0.008 | 0.6117 | −0.014 |
| Satt571 | I | 18.5 | $(ATT)_{14}$ | 15 | 127–177 | 3.96 | 0.007 | 0.7526 | −0.035 |
| Satt586 | F | 3.63 | $(ATT)_{19}$ | 15 | 177–237 | 7.69 | 0.012 | 0.8694 | −0.055 |
| Satt588 | K | 117.02 | $(ATT)_{18}$ $(AT)_{10}(CT)_{14}$ | 19 | 116–185 | 7.52 | 0.075 | 0.859 | 0.019 |
| Satt590 | M | 7.84 | $(ATT)_{26}$ | 28 | 253–352 | 8.87 | 0.011 | 0.8922 | −0.018 |
| Satt596 | J | 39.64 | $(ATT)_{17}$ | 21 | 233–298 | 7.38 | 0.014 | 0.8723 | −0.032 |
| Sct189 | I | 113.8 | $(CT)_{17}$ | 23 | 155–206 | 10.80 | 0.009 | 0.9047 | 0.035 |
| Mean | | | | 20.5 | | 6.8 | 0.015 | 0.827 | 0.080 |
| SD | | | | 6.8 | | 2.6 | | 0.083 | |

*Note that the data of 1863 accessions was from the previous study (Li et al. 2008, Theor Appl Genet)

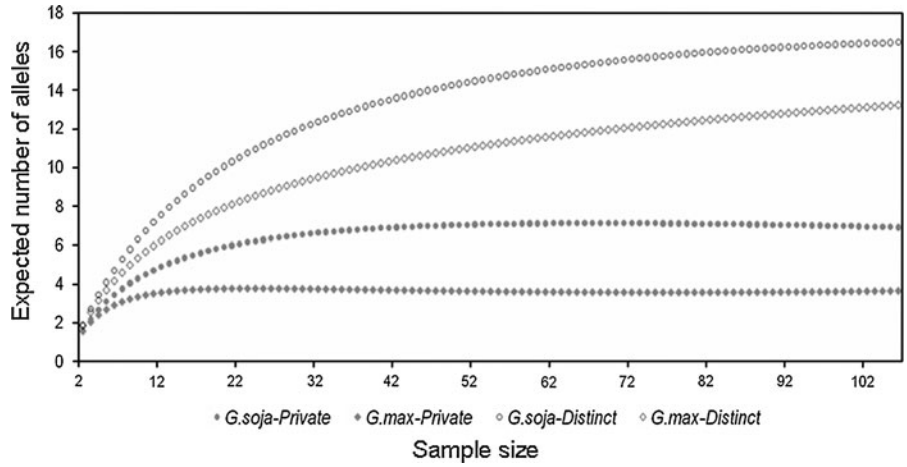


**Fig. 2** The mean expected number of distinct and private alleles per locus as a function of standardized sample size for two species (*G. soja* and *G. max*)

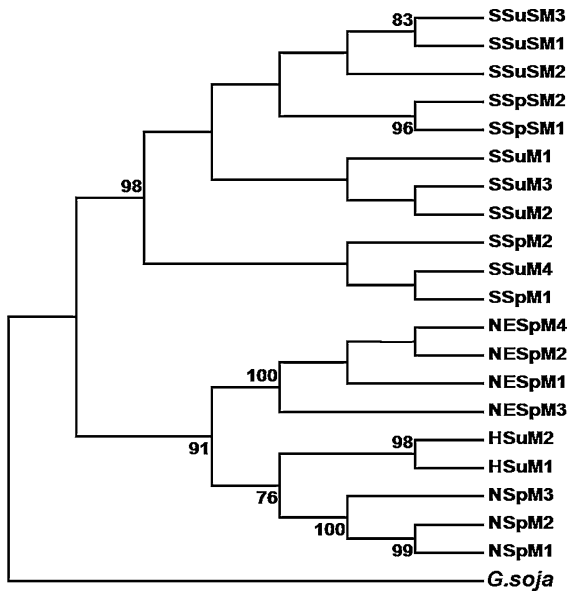


**Fig. 3** Shared allele distance-based Neighbour-joining tree of sub-clusters of soybean landraces (*G. max*) in China rooted with *G. soja* based on 42 SSR loci with bootstrap support values (>80 %, based on 1,000 bootstraps). Sample codes are described in the text

with *G. soja* since the Nei's genetic identity between them (0.3585) was the lowest.

## Directional evolution of SSR allele size in soybean

It has been widely accepted that *G. max* originated from its wild relative *G. soja*. To test whether directional evolution of SSR size occurred in soybean, the average standardized allele lengths were compared between the sample of 50 landraces and the sample of 50 wild accessions which were randomly drawn 30 times from 1504 landraces and 62 *G. soja* accessions (which were collected across China) respectively, using a set of 42 SSR loci. The average standardized allele length across all loci in the whole dataset was zero, as a result of the standardization. The difference between the *G. max* landraces group (0.009) and the *G. soja* group (−0.406) was around 0.4 units, which was highly significant (*t* test, $P = 8.63 \times 10^{-58}$). When we looked at each of accessions, the distribution of average SSR standardized allele size overlapped between *G. soja* and *G. max* populations. In the *G. max* population, 18 landraces were discovered with smaller allele size than the average size of *G. soja* (−0.406). It was worth noting that these accessions were associated with some ancestral traits, such as small seed size, black seed color, twinning or semi-twinning habit. Among them, 15 were from Huanghuai Region and the others were from Northeast China. When we looked at each of the SSR markers separately, 30 out of the 42 showed a clear increase in standardized allele size between the two species, while eight (Satt022, Satt130, Satt184, Satt226, Satt268, Satt309, Satt346, Satt530, Satt571 and Satt586) were significant (*t* test, $P < 0.05$) smaller in *G. max* compared to *G. soja* and the other four (Sat_112, Satt012, Satt130 and Satt346) were not significant at 0.05 level.

For detecting the domestication location of cultivated soybean, the average standardized allele sizes were analyzed for each of the clusters of landraces, taking *G. soja* group as control. At the cluster level, significant ($P < 0.001$) differences in SSR allele size are present between most pairs of clusters (Table 3). NSpM and HSuM, both located along the Yellow

**Table 3** Difference in average standardized allele sizes for pairs of clusters

| Clusters | Average standardized allele size[a] | | Difference in average standardized allele sizes for pairs of clusters | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | All accessions | 50 Randomly sampled accessions | NESpM | NSpM | HSuM | SSpM | SSpSM | SSuM | SSuSM |
| *G. soja* | −0.404 | −0.406 | 0.412* | 0.291* | 0.363* | 0.439* | 0.457* | 0.482* | 0.428* |
| NESpM | 0.008 | 0.013 | | 0.121* | 0.049* | 0.027 NS | 0.044 NS | 0.070* | 0.016 NS |
| NSpM | −0.113 | −0.121 | | | 0.072* | 0.148* | 0.166* | 0.191* | 0.137* |
| HSuM | −0.041 | −0.042 | | | | 0.076* | 0.094* | 0.119* | 0.065* |
| SSpM | 0.035 | 0.037 | | | | | 0.017 NS | 0.042 NS | 0.011 NS |
| SSpSM | 0.052 | 0.047 | | | | | | 0.025 NS | 0.028 NS |
| SSuM | 0.078 | 0.076 | | | | | | | 0.053* |
| SSuSM | 0.024 | 0.024 | | | | | | | |

[a] Note that the average standardized allele size across all alleles and all loci using all landraces and accessions in the study, is zero

*NS* not significant ($P > 0.001$); * $P < 0.001$

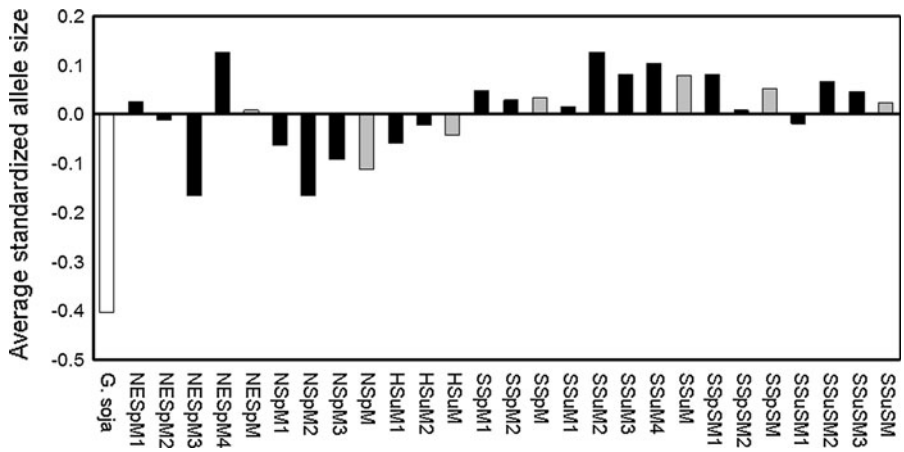


**Fig. 4** The average standardized allele sizes of 20 cultivated sub-clusters, seven clusters and wild *G. soja* population. Note that the overall average of the standardized allele length across all alleles, loci, and samples, is zero. *Empty bar*, *G. soja*; *Grey bar*, inferred clusters; *Black bar*, inferred sub-clusters

River valley (Fig. 1), had the smallest average standardized allele size, which was significantly lower than that of any other cluster ($P < 0.001$). Their allele sizes are therefore closest to the value found in *G. soja*. Moreover, NSpM had a significant ($P < 0.01$) smaller allele size than HSuM. HSuM may then be considered as the derived cluster, although the genetic diversity of HsuM is higher than that of NSpM. NESpM had the third smallest allele size, significantly smaller than SSuM ($P < 0.001$) but not significantly smaller than the other three clusters from South China. The difference between any pair of the three southern clusters SSuM, SSpM and SSuSM was not significant (at $P = 0.001$).

When looking at the sub-clusters, a comparable pattern can be discerned (Fig. 4). The average standardized allele sizes of all of four sub-clusters from Yellow River valley were negative and closest to the value of *G. soja*. In contrast to these sub-clusters from central China, most (90.9 %) of the sub-clusters from South China exhibited a positive allele size, and hence a large difference with that of *G. soja*. In Northeast China, two of the four sub-clusters had a negative allele size and the other two were positive. Among all 20 sub-clusters, NSpM2, which includes landraces mainly distributed along the middle regions of the Yellow River, exhibited the smallest allele size (−0.169), followed by NESpM3 (−0.166), NSpM3 (−0.095), NSpM1 (−0.062), HSuM2 (−0.060) and HSuM1 (−0.027). It is worth noting that most of the NSpM2 landraces display traits considered ancestral, such as black (63 %) or brown (15 %) seed coat and

small seed size (74 %). This phenomenon was also observed in the landraces from NESpM3.

Seed dormancy is a crucial adaptive trait, as it prevents sprouting in wild soybean until the growth conditions become favorable. Therefore, there is only one sowing type, spring, in *G. soja*. Compared to wild soybean, cultivars can easily break up dormancy via reduction of germination inhibitors and there are at lease three sowing types, spring, summer and autumn, in *G. max*. For investigating the evolution of ecotypes, the average allele size of spring-sowing type, summer-sowing type and autumn-sowing type in landraces were analyzed. The *t* test analysis showed that the difference among these three types were significant ($P < 0.01$). The average allele size of spring-sowing types (−0.036) is smaller than that in the summer-sowing types (0.048). SSuM2, the one and only autumn-sowing type sub-cluster, exhibited the largest average allele size of all sub-clusters (0.133). This result is consistent with the hypothesis that the spring-sowing type was the original state from which summer-sowing types developed, while the autumn-sowing type is the most derived group, as it is derived from the summer-sowing type.

## Discussion

### Differences in standardized allele lengths and the most ancestral populations

The standardized allele length in *G. max* (0.009) was much larger than in *G. soja* (−0.406; Table 3). Also among clusters and sub-clusters of *G. max* landraces there were large differences. This may be an example of the directional evolution of SSRs towards larger alleles (Amos and Rubinsztein 1996), which has been used in investigations of the origin of various species since it was first observed in a study between humans and nonhuman primates (Rubinsztein et al. 1995). Some scientists thought that this tendency was an artifact, caused by the selection criteria during marker discovery, resulting in so-called ascertainment bias (Ellegren et al. 1995; Neff and Gross 2001; Forbes 1995). Indeed, ascertainment bias plays a role in many studies comparing diversity across taxa, as markers that have been developed by screening one of the taxa for polymorphic markers, are unlikely to be just as polymorphic in all other taxa. This is true for SNPs as well, as these are identified against a large set of reference sequence information from one taxon (Sun et al. 2009). However, further study demonstrated that directional evolution was the main reason for the differences in allele size in SSRs even when ascertainment bias was taken into account (Amos et al. 2003).

Although SSR have all been developed and used for both cultivated and wild soybean, it could be argued that at least part of the difference in length could still be due to ascertainment bias. Two argument against this are (1) that there are also significant differences in allele length among sub-clusters within *G. max*, and (2) that the largest allele size (0.133) was found in a sub-cluster of *G. max* (SSuM2) that contains the autumn-type soybean landraces, which are thought to be secondarily derived from summer-sowing types. On the other hand, even if there is some ascertainment bias, the length difference still provide information on the genetic distance from the derived species *G. max* to the ancestral, more distantly related *G. soja*, and landrace sub-clusters having smaller alleles (e.g., NSpM2 with −0.169) would indeed be more closely related to *G. soja*.

### Domestication of cultivated soybean

Both NSpM2 and NESpM3 had the smallest average standardized allele size. NSpM2 is located on the middle regions of the Yellow River, from 34.5 to 40.3°N and from 107.4 to 114.2°E (Fig. 1). This region has always been taken as the origin of rain-fed crops (drought-farm agriculture), one of the eight centers of origin for cultivated plants in the world (Vavilov 1951). It was the first location that was proposed as the domestication place of soybean by Fukuda (1933) who was followed by a number of other scholars (Zhao and Gai 2004; Dong et al. 2004; Hymowitz and Newell 1981; Vavilov 1951; Chang 1989). Genetic diversity analysis has shown that this region was one of the genetic diversity centers of cultivated soybean (Wang et al. 2008; Xu et al. 1985, 1986; Zhou et al. 1998). At sub-cluster level, the highest value was found in HSuM1 (8.1 alleles in 19 randomly taken landraces), which contains landraces (summer-sowing type) largely from the same region along the Yellow River as NSpM2 (Li et al. 2010b).

So both types of evidence locate at the same region along the Yellow River, but not to the same group of

landraces. Normally it is assumed that the group with the highest diversity is the most ancestral one (Vavilov 1951). Conceptually, though, it is difficult to envisage a process in which spring-sowing type *G. soja* is firstly domesticated as summer-sowing type *G. max*, before developing into spring-sowing type *G. max* landraces. The presence of primitive morphological characters, such as black seed color (63 % of NSpM landraces has this color, versus only 18 % of those of HSuM) and small seed weight (67 % of NSpM2 landraces has small seed weight, versus only 17 and 20 % of NSpM1 and NSpM3 respectively), and the presence of the most ancestral allele size, indicate that the spring-sowing types of NSpM2 would be most close to *G. soja*, but if so, either the spring-sowing types have lost genetic diversity since then, or, as the summer-sowing types have a significantly higher gene diversity, these have received diversity through secondary introgression. This result is in agreement with the theory proposed by Harlan (1971) that the secondary centre of origin may sometimes have the highest genetic diversity (Zhao and Gai 2004). Indeed, in maize (van Heerwaarden et al. 2011) and in apple (Cornille et al. 2012) the contribution of a second species was so large that the cultivated crop *Zea mays L.* is genetically more closely related to current-day highland landraces than to lowland *Z. mays* ssp. *parviglumis* Jetis et Doebley from which the crop was domesticated, and current varieties of *Malus domestica* Borkh. are more closely related to the European crabapple *M. sylvestris* (L.) Mill. than to their central Asian progenitor *M. sieversii* (Ledeb.) Roem.

When dealing with landraces, the amount of genetic variation, and the presence or absence of specific ancestral traits, and even the location at which they are grown, may be influenced by the relative use in the course of thousands of years. The summer-sowing type of soybean has become popular probably soon after domestication, as it can be combined with wheat cultivation in winter, so that farmers may have grown summer-sowing type soybean landraces in this region, and also in large areas south of Yellow River, which are very diverse in agro-ecological conditions. Indeed, in the south of China there are patches (sub-clusters) that have nearly as much diversity as NSuM, separated by regions with lower diversity. It may indicate that most genetic diversity was channeled into and maintained in the summer-sowing types, even though they were developed after the initial domestication step. Our results support therefore domestication along the Yellow River, but cannot exclude other regions as landrace cultivation may have moved in intermediate years. We agree with Kim et al. (2010) that more soybean genomes need to be sequenced and compared for better understanding origins of domesticated soybean and the processes underlying domestication. Landraces from NSpM2, that we identified here as having the shortest standardized allele length, and from the spring- and summer-sown sub-clusters with the highest genetic diversity, are good candidates for such resequencing.

**Acknowledgments** This research was supported by the State Key Basic Research and Development Plan of China (973) (No. 2010CB125900 and 2004CB117203), National Key Technologies R&D Program in the 11th Five-Year Plan (No.2006BAD 13B05), the National Natural Science Foundation of China (No.30490251 and 30471096) and the Academy and Institute Foundation for Basic Scientific Research in Institute of Crop Science, Chinese Academy of Agricultural Sciences.

## References

Amos W, Rubinsztein DC (1996) Microsatellites are subject to directional evolution. Nat Genet 12(1):13–14

Amos W, Hutter CM, Schug MD, Aquadro CF (2003) Directional evolution of size coupled with ascertainment bias for variation in drosophila microsatellites. Mol Biol Evol 20:660–662

Anthony F, Combes M, Astorga C, Bertrand B, Graziosi G, Lashermes P (2002) The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. Theor Appl Genet 104(5):894–900

Carter TE, Nelson RL, Sneller CH, Cui ZL (eds) (2004) Genetic diversity in soybean. Soybeans: improvement, production and uses. ASA, CSSA, SSSA, Madison

Chang RZ (1989) Studies on the origin of the cultivated soybean (*Glycine max* (L) Merr.) Chinese J Oil Crops Sci 1:1–6 (in Chinese with English abstract)

Cornille A, Gladieux P, Smulders M, Roldán-Ruiz I, Laurens F, Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang X, Tenaillon M, Giraud T (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet 8(5):e1002703

Dong YS, Zhao LM, Liu B, Wang ZW, Jin ZQ, Sun H (2004) The genetic diversity of cultivated soybean grown in China. Theor Appl Genet 108(5):931–936

Ellegren H, Primmer CR, Sheldon BC (1995) Microsatellite 'evolution': directionality or bias? Nat Genet 11(4): 360–362

Forbes SH (1995) Microsatellite evolution in congeneric mammals: domestic and bighorn sheep. Mol Biol Evol 12(6):1106–1113

Fukuda Y (1933) Cyto-genetical studies on the wild and cultivated Manchurian soy beans (*Glycine* L.). Jpn J Bot 6:489–506

Gai JY, Xu DH, Gao Z, Shimamoto Y, Abe J, Fukushi H, Kitajima S (2000) Studies on the evolutionary relationship among eco-types of *G.max* and *G. soja* in China. Acta Agri Sin 26(5):513–520 (in Chinese with English abstract)

Garris AJ, Tai TH, Coburn J, Kresovich S, McCouch S (2005) Genetic structure and diversity in *Oryza sativa* L. Genetics 169(3):1631–1638

Grassi F, Labra M, Imazio S, Spada A, Sgorbati S, Scienza A, Sala F (2003) Evidence of a secondary grapevine domestication centre detected by SSR analysis. Theor Appl Genet 107(7):1315–1320

Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H (2010) A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. Ann Bot 106(3):505–514

Harlan JR (1971) Agricultural origins: centers and noncenters. Science 174(4008):468

Hymowitz T, Newell CA (1981) Taxonomy of the genus *Glycine*, domestication and uses of soybeans. Econ Bot 35(3):272–288

Kim M, Lee S, Vana K, Kimb T, Jeong S, Choi I, Kim D, Lee Y, Park D, Ma J, Kim W, Kim B, Park S, Lee K, Kim D, Kim K, Shin J, Jang Y, Kim K, Liu W, Chaisan T, Kang Y, Lee Y, Kim K, Moon J, Schmutzh J, Jacksone S, Bhak J, Lee S (2010) Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. Proc Natl Acad Sci USA 107:22032–22037

Li FS (1994) A study on origin and evolution of soybean. Soybean Sci 13(1):61–66 (in Chinese with English abstract)

Li YH, Guan RX, Liu ZH, Ma YS, Wang LX, Li LH, Lin FY, Luan WJ, Chen PY, Yan Z, Guan Y, Zhu L, Ning X, Smulders MJM, Li W, Piao R, Cui YH, Yu ZM, Guan M, Chang RZ, Hou AF, Shi AN, Zhang B, Zhu SL, Qiu LJ (2008) Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. Theor Appl Genet 117:857–871

Li YH, Li W, Zhang C, Yang L, Chang RZ, Gaut BS, Qiu LJ (2010a) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single nucleotide polymorphism loci. New Phytol 188:242–253

Li YH, Smulders MJM, Chang RZ, Qiu LJ (2010b) Analysis of SSRs uncovers hierarchical structure and genetic diversity in Chinese soybean landraces. Agric Sci China 9(12): 1739–1748

Liu MS, Amirkhanian VD (2003) DNA fragment analysis by an affordable multiple-channel capillary electrophoresis system (Short communication). Electrophoresis-Weinheim 24(1/2):93–95

Liu K, Muse S (2005) Power marker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez G (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99(9):6080–6084

Maughan PJ, Saghi Maroof MA, Buss GR (1995) Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome 38(4):715–723

Neff BD, Gross MR (2001) Microsatellite evolution in vertebrates: inference from acdinucleotide repeats. Evolution 55(9):1717–1733

Petit RJ, Deguilloux MF, Chat J, Grivet D, Garnier-Gere P, Vendramin GG (2005) Standardizing for microsatellite length in comparisons of genetic diversity. Mol Ecol 14(3):885–890

Qiu L, Cao Y, Chang R, Zhou X, Wang G, Sun J, Xie H, Zhang B, Li X, Xu Z, Liu L (2003) Establishment of Chinese soybean (*G. max*) core collection I. sampling strategy. Sci Agri Sin 36(12):1442–1449 (in Chinese with English abstract)

Qiu L, Li Y, Guan R, Liu Z, Wang L, Chang R (2009) Establishment, representative testing and research progress of soybean core collection and mini core collection. Acta Agron Sin 35(4):571–579 (in Chinese with English abstract)

Ren TH, Chen F, Zou YT, Jia YH, Zhang HQ, Yan BJ, Ren ZL (2011) Evolutionary trends of microsatellites during the speciation process and phylogenetic relationships within the genus *Secale*. Genome 54:316–326

Rubinsztein DC, Amos W, Leggo J, Goodburn S, Jain S, Li SH, Margolis RL, Ross CA, Ferguson-Smith MA (1995) Microsatellite evolution: evidence for directionality and variation in rate between species. Nat Genet 10(3):337–343

Sun JX, Mullikin JC, Patterson N, Reich DE (2009) Microsatellites are molecular clocks that support accurate inferences about history. Mol Biol Evol 26(5):1017–1027

Szpiech ZA, Jakobsson M, Rosenberg NA (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24(21):2498

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

van Heerwaarden J, Doebley J, Briggs W, Glaubitz J, Goodman M, Gonzalez J, Ross-Ibarra J (2011) Genetic signals of origin, spread, and introgression in a large sample of maize landraces. Proc Natl Acad Sci USA 108:1088–1092

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Translated from the Russian by K. Starrchester. Chron Bot 13:1–364

Vigouroux Y, Matsuoka Y, Doebley J (2003) Directional evolution for microsatellite size in maize. Mol Biol Evol 20(9):1480–1483

Wang LX, Guan RX, Zhang XL, Chang RZ, Qiu LJ (2006) Genetic diversity of Chinese cultivated soybean revealed by SSR markers. Crop Sci 46(3):1032–1038

Wang LX, Guan RX, Li YH, Lin FY, Luan WJ, Li W, Ma YS, Liu ZX, Chang RZ, Qiu LJ (2008) Genetic diversity of Chinese spring soybean germplasm revealed by SSR markers. Plant Breed 127(1):56–61

Xu B, Zhao SW, Zhou SH, Zheng HY, Hu ZA, Wang HX (1985) Seed protein electrophoresis profiles of wild soybean (*G. soja*) in China: the frequencies and geographical distribution of Tia and Sp1 alleles and the hypothesis on the original area of soybean. Soybean Sci 4:7–13 (in Chinese with English abstract)

Xu B, Zheng HY, Lu QH, Zhao SW, Zhou SH, Hu ZA (1986) Three new evidences of the original area of soybean.

Soybean Sci 5(2):123–130 (in Chinese with English abstract)

Xu D, Abe J, Gai J, Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet 105(5):645–653

Zhao TJ, Gai JY (2004) The origin and evolution of cultivated soybean [*Glycine max* (L.) Merr.]. Sci Agri Sin 37 (7):954–962 (in Chinese with English abstract)

Zhou XA, Peng YH, Wang GX, Chang RZ (1998) Preliminary studies on the centers of genetic diversity and origination of cultivated soybean in China. Sci Agri Sin 31(003):37–43 (in Chinese with English abstract)