Theor Appl Genet (2003) 106:445–453
DOI 10.1007/s00122-002-1073-3

**J. Abe · D. H. Xu · Y. Suzuki · A. Kanazawa**
**Y. Shimamoto**

# Soybean germplasm pools in Asia revealed by nuclear SSRs

Received: 14 February 2002 / Accepted: 24 May 2002 / Published online: 7 September 2002


**Abstract** Soybean was domesticated in East Asia, where various kinds of landraces have been established as a result of adaptation to different environments and the diversification of food cultures. Asia is thus an important germplasm pool of soybean. In order to evaluate the genetic structure of the Asian soybean population, we analyzed allelic profiles at 20 simple-sequence repeat (SSR) loci of 131 accessions introduced from 14 Asian countries. The SSR loci produced an average of 11.9 alleles and a mean gene diversity of 0.782 in the accessions tested. Quantification theory III analysis and cluster analysis with the UPGMA method clearly separated the Japanese from the Chinese accessions, suggesting that the Japanese and Chinese populations formed different germplasm pools. The Korean accessions were involved in both germplasm pools, whereas most of the accessions from southeast and south/central Asia were derived from the Chinese pool. Relatively high genetic diversity and the absence of region-specific clusters in the southeast and south/central Asian populations suggest that soybean in these areas has been introduced repeatedly and independently from the diverse Chinese germplasm pool. The present study indicates that the two germplasm pools can be used as exotic genetic resources to enlarge the genetic bases of the respective Asian soybean populations.

**Keywords** Soybean · SSR · Diversity · Germplasm pool

## Introduction

Use of exotic genetic resources is of primary importance to enlarge the genetic base in a population of crops. A thorough understanding of the genetic structure in crops is needed for effective use of these exotic genotypes and the construction of core collections. Studies using recently developed molecular techniques have revealed much about the genetic diversity in soybean, *Glycine max* (L.) Merr. subsp. *max* Ohashi, and its wild annual counterpart, *G. max* subsp. *soja* (Sieb. et Zucc.) Ohashi. Extensive studies have been performed mainly on improved cultivars released in North America and their ancestral introductions (Diwan and Cregan 1997; Song et al. 1999; Brown-Guedira et al. 2000; Narvel et al. 2000; Li et al. 2001). The genetic structure of the Asian soybean population, however, still remains unclear despite its usefulness as a genetic resource.

Asian countries have a long history of soybean cultivation, and consequently various kinds of landraces have been established as a result of adaptation to different environments and the diversification of food cultures. Asia is thus a center of diversity of soybean. Genetic diversity and the pattern of variation in the Asian soybean population have been evaluated with seed protein and isozyme loci (Hymowitz and Kaizuma 1979, 1981; Wang et al. 1986; Perry et al. 1991; Griffin and Palmer 1995; Han et al. 1999; Hirata et al. 1999). From an analysis of the Kunitz trypsin inhibitor (*Ti*) and beta-amylase isozyme (*Sp1* = *Amy3*), Hymowitz and Kaizuma (1981) defined seven soybean germplasm pools in Asia: (1) northeast China and the USSR, (2) central and south China, (3) Korea, (4) Japan, (5) Taiwan and south Asia, (6) north India and Nepal and (7) central India. Griffin and Palmer (1995) surveyed 1,005 *G. max* accessions and 258 *G. soja* accessions of the USDA soybean germplasm collection for genotypes at 13 isozyme loci, and found that the *G. max* accessions from southeast Asia stood out from those from the other regions. Hirata et al. (1999) compared the genetic variation at 16 isozyme and seed protein loci of 781 Japanese accessions with the genetic variations of 158 Korean and 94 Chinese accessions. They found a number of region-specific alleles that discriminated Japanese from Chinese accessions. The presence of alleles specific to the Japanese population sug-

Communicated by H.C. Becker

J. Abe (✉) · D.H. Xu · Y. Suzuki · A. Kanazawa · Y. Shimamoto
Laboratory of Plant Genetics and Evolution,
Graduate School of Agriculture, Hokkaido University,
Sapporo 060-8589, Japan
e-mail: jabe@res.agr.hokudai.ac.jp
Fax: +81-11-7064933

gested that the present Japanese soybean population was not solely a subset of the Chinese population (Hirata et al. 1999). Li and Nelson (2001) evaluated the genetic diversity among 120 soybean accessions from Japan, South Korea and China with randomly amplified polymorphic DNAs (RAPDs). They found that the Japanese and South Korean populations were more similar to each other, whereas both were genetically distinct from the Chinese population.

Simple sequence repeats (SSRs) are useful molecular markers in various fields of soybean genetic and breeding research. Their advantages over other kinds of molecular markers are that they are abundant, have a high level of polymorphism, are codominant, can be easily detected with PCR and have known positions in the genome. More than 600 SSRs have been developed and mapped in 20 molecular linkage groups in soybean to-date (Cregan et al. 1999). These SSRs mostly represent polymorphism at single loci and there is no ambiguity in scoring genotypes, in contrast with analyses with RFLP probes that often hybridize to two or more positions in the soybean genome. High levels of polymorphism at SSRs have been reported for both the numbers of alleles per locus and the gene diversity (Maughan et al. 1995; Powell et al. 1996; Diwan and Cregan 1997; Song et al. 1999; Narvel et al. 2000). These values are much higher than those reported for RFLP markers (Keim et al. 1989; Kisha et al. 1998) and isozymes (Perry et al. 1991; Griffin and Palmer 1995; Han et al. 1999; Hirata et al. 1999). Furthermore, use of fluorescent-labeling primers and DNA sequencers make it possible to accurately identify and easily score alleles at a number of loci at once.

To obtain a better understanding of the genetic relationships among populations of different geographical regions in Asia, we evaluated the genetic diversity and pattern of genetic variation in Asian cultivated soybeans by examining the length polymorphism of alleles found in 20 SSR loci from different linkage groups.

## Materials and methods

### Plant materials

One hundred and thirty one accessions were used in this study (Table 1). These were mostly landraces or pure-line selections, and were selected from 14 Asian countries: China (51), Japan (39), South Korea (6), North Korea (6), far-east Russia (3), Vietnam (3), Thailand (2), Indonesia (4), Myanmar (4), India (3), Bhutan (2), Nepal (3), Pakistan (2) and Kyrgyz (3). The 39 Japanese accessions were selected from each of 12 cultivar groups that were defined based on the variation at 16 isozyme and *Ti* loci in order to cover a large part of the variation observed in 781 Japanese accessions (Hirata et al. 1999). The 51 Chinese accessions were introduced from 19 provinces of four regions: 8 from Northeast China, 16 from Huang-Huai-Hai River Valley area, 20 from Changjiang River Valley area and 7 from Southern China.

### Analysis of SSRs

Twenty tri-nucleotide repeat SSRs that were located in each of 20 molecular linkage groups were used in the present study (Table 2). The forward primers of the 20 SSR loci were labeled with either 6-FAM (blue), HEX (green) or NED (yellow) fluorescent dyes. DNA was extracted from young leaves of a single plant per accession with the method described by Doyle and Doyle (1990). The PCR reaction mixture contained 30 ng of total genome DNA, 0.25 μM of 5′ and 3′ end primers, 200 μM of each dNTP, 0.5 units of *Taq* polymerase (TaKaRa, Japan), and 1 × PCR buffer (10 mM of Tris-HCl, pH 8.3, 50 mM of KCl, 1.5 mM of $MgCl_2$) in a total volume of 20 μl. PCR reactions were performed with a GeneAmp PCR System 9700 (Perkin Elmer/Applied Biosystems, Foster City, USA) using the following program: 32 cycles at 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s. After the amplifications, 1.5 μl of 6-FAM-labeled, 4.0 μl of HEX-labeled and 2.0 μl of NED-labeled PCR products were combined and brought to a total volume of 20 μl by adding distilled water. An aliquot (1.5 μl) of the mixture of PCR products was combined with loading buffer (1.5 μl) containing a ROX (red)-labeled internal size standard (GeneScan-500) and, after denaturation at 95 °C for 5 min, was then loaded and separated by an ABI 377 sequencer (Perkin Elmer/Applied Biosystems, Foster City, USA). GeneScan (version 3.1) software was used to visualize the variants and to estimate their sizes.

### Data analysis

The genetic diversity index ($H$) based on allele frequencies was calculated for each SSR using Nei's unbiased statistics (Nei 1987):

$$H = 1 - \sum P_i^2,$$

where $P_i$ is the frequency of the $i$th allele. The AMOVA analysis (Excoffier et al. 1992) was used to analyze the degree of genetic differentiation between regional groups of accessions. The allelic profile data was subjected to Hayashi's Quantification Theory III analysis (Hayashi 1956) with the software 'Excel Quantification Ver. 1.0' (Esumi, Tokyo, Japan). The analysis is nearly equivalent to the Correspondence Analysis (Benzecri 1992), and was designed to integrate the correlated variations for discontinuous variables into a few vectors, like the Principal Component Analysis for continuous variables. The genetic distance ($D$) between each pair of the 131 accessions was calculated with software 'MSAT' (Minch et al. 1997) by $1 - P$, where $P$ is the proportion of shared alleles for the 20 loci. The genetic distance matrix was subjected to a cluster analysis with the unweighted pair-group method of the arithmetic average (UPGMA) with the PHYLIP software (Felsenstein 1995). A UPGMA dendrogram was also constructed for regional groups of accessions, based on average genetic distances between accessions of different regions.

## Results

### Variation in the Asian soybean population

The length polymorphism of the alleles of 20 SSR loci in 131 Asian accessions was determined (Table 2). For each of the loci, the lengths of the alleles varied in steps of three bases. For all but one of the loci, the lengths were nearly continuously distributed, while for Satt463 the lengths were distributed in two different ranges (114–159 bp and 210–234 bp). Each of the loci had between 8 and 24 alleles (11.9 on average). Satt009 and Satt463 had the highest numbers of alleles (24 and 21, respectively), and Satt228 and Satt253 had the lowest number (8). No amplified product was observed at Satt253 in two accessions from Pakistan (PAK-1 and PAK-2) and one accession from Myanmar (MYA-1). The $H$ values ranged from 0.465 for Satt203 to 0.922 for Satt009, with an average of 0.782.

**Table 1** Soybean accessions used in this study

| Name or collection site of accession | Code | Name or collection site of accession | Code | Name or collection site of accession | Code |
|---|---|---|---|---|---|
| China (CHN) | | Japan (JPN) | | South Korea | |
| Northeast China | | Northern Japan | | Inje/South Korea | SK-1 |
| | | | | Kanghwa/South Korea | SK-2 |
| Badajia | NEC-1 | Aobata | NJ-1 | Kunsan/South Korea | SK-3 |
| Baihualudadou | NEC-2 | Asahi 60 | NJ-2 | Kurye/South Korea | SK-4 |
| Baihuamoshidou | NEC-3 | Asajiro | NJ-3 | Sosan/South Korea | SK-5 |
| Baoandadou | NEC-4 | Banechigo | NJ-4 | Ulchin/South korea | SK-6 |
| Bianchadou | NEC-5 | Bikuni | NJ-5 | | |
| Lindiansuoyiling | NEC-6 | Bonjiro 55 | NJ-6 | Indonesia | |
| Silihuang | NEC-7 | Bukoumame | NJ-7 | | |
| Yuhuizhen | NEC-8 | Chashouryu | NJ-8 | Bromo/Indonesia | INDO-1 |
| | | Gohamame | NJ-9 | Dieng/Indonesia | INDO-2 |
| Huang-Huai-Hai River Valley area | | Higanmame | NJ-10 | Lumajing Bewok/Indonesia | INDO-3 |
| | | Ishikarishiro 1 | NJ-11 | Nakon/Indonesia | INDO-4 |
| Fengshouhuang | HHH-1 | Itachi | NJ-12 | | |
| Guangshanwenshutianedan | HHH-2 | Karihatakiya | NJ-13 | Vietnam | |
| Heihuangdou | HHH-3 | Kimusume | NJ-14 | | |
| Heiyaohuangdou | HHH-4 | Kurosakigake | NJ-15 | Camxuyen/Vietnam | VIE-1 |
| Jindou 1 | HHH-5 | Miyashiro | NJ-16 | Caobang/Vietnam | VIE-2 |
| Kaifengguozhuangqingdou | HHH-6 | Mizukuguri | NJ-17 | Laocai/Vietnam | VIE-3 |
| Luhuangdou | HHH-7 | Nagaokatairyu | NJ-18 | | |
| Peixiandabaijiao | HHH-8 | Nanbu | NJ-19 | Thailand | |
| Raoshangun | HHH-9 | Oyachi 2 | NJ-20 | | |
| Runanpingdingshi | HHH-10 | Rokujunichimame | NJ-21 | Changmai/Thailand | THA-1 |
| Shenchiniuyanheidou | HHH-11 | Shirohanasai 1 | NJ-22 | Maehongson/Thailand | THA-2 |
| Suidaohuang | HHH-12 | Shirohakkoku | NJ-23 | | |
| Wuxiliuyueku | HHH-13 | Shiroturunoko | NJ-24 | Myanmar | |
| Yanchengqisinidou | HHH-14 | | | | |
| Yixianheidou | HHH-15 | Southern Japan | | Pindaya/Myanmar | MYA-1 |
| Yuzipihuangdou | HHH-16 | | | Piyinmana/Myanmar | MYA-2 |
| | | Akawase | SJ-1 | Taunggyi/Myanmar | MYA-3 |
| Changjiang River Valley area | | Chamameshouryu | SJ-2 | Thakhilek/Myanmar | MYA-4 |
| | | Chuteppou | SJ-3 | | |
| Anyuelulanzi | CR-1 | Hachigatudaizu | SJ-4 | Bhutan | |
| Chengbuliuyuehuang | CR-2 | Hiroshimakurodaizu | SJ-5 | | |
| Guiximayiwo | CR-3 | Iyodaizu | SJ-6 | Morgal/Bhutan | BH-1 |
| Hangzhouwuyuebai | CR-4 | Kairyogionbou | SJ-7 | Thimphu/Bhutan | BH-2 |
| Hengshanhongdou | CR-5 | Kairyoshirome | SJ-8 | | |
| Hengyangheidou | CR-6 | Misaodaizu | SJ-9 | Neparl | |
| Honghuliuyuebao | CR-7 | Ohitaakidaizu 1 | SJ-10 | | |
| Huangposhanzibai | CR-8 | Tamanishiki | SJ-11 | Gadlang/Neparl | NEP-1 |
| Hunanniumaohuang | CR-9 | Tanbakuro | SJ-12 | Kakani/Neparl | NEP-2 |
| Matangdadou | CR-10 | Wasekin | SJ-13 | Panchmano/Neparl | NEP-3 |
| Shangraoheiyangdou | CR-11 | Yahagidaizu | SJ-14 | | |
| Shichengqingpidou | CR-12 | Yukikorogashi | SJ-15 | India | |
| Taogedadou | CR-13 | | | | |
| Wanxiandazaohuang | CR-14 | Fareast Russia | | New Delhi (PI20983A) | IND-1 |
| Weiyuanqiyuehuang | CR-15 | | | Madha Predesh (PI307882) | IND-2 |
| Xianjuxiaomaodou | CR-16 | Mestnaya Salatnaya | FR-1 | Madha Predesh (PI374198) | IND-3 |
| Yichangheihuangdou | CR-17 | Ussuriiskya154 | FR-2 | | |
| Yongkangtiejiangdou | CR-18 | Vir-1043 | FR-3 | Pakistan | |
| Yunlianqizhuandou | CR-19 | | | | |
| Yunmengheihuangdou | CR-20 | North Korea | | Col/Pak/1989/IBPGR/2296(3) | PAK-1 |
| | | | | Col/Pak/1989/IBPGR/2319(1) | PAK-2 |
| Southern China | | GL1738 | NK-1 | | |
| | | GL2617 | NK-2 | Kyrgiz | |
| Dalidadou | SC-1 | GL2620 | NK-3 | | |
| Dapudalihuang | SC-2 | GL2629 | NK-4 | Kyrgz 1 | KYR-1 |
| Heibiqing | SC-3 | GL2679 | NK-5 | Kyrgz 3 | KYR-2 |
| Jianyangdadou | SC-4 | GL2683 | NK-6 | Kyrgz 5 | KYR-3 |
| Jinghongdadou | SC-5 | | | | |
| Qingyuanxiaoqingdou | SC-6 | | | | |
| Sanmingaijiaoqing | SC-7 | | | | |

**Table 2** Number of variants, allele size range and gene diversity at 20 SSR loci

| Locus | Linkage group | No. of alleles | Allele size range (bp) | Genetic diversity index (*H*) |
|---|---|---|---|---|
| Satt236 | A1 | 11 | 202–241 | 0.725 |
| Satt228 | A2 | 8 | 215–254 | 0.619 |
| Satt197 | B1 | 14 | 135–204 | 0.839 |
| Satt063 | B2 | 10 | 106–163 | 0.816 |
| Satt180 | C1 | 12 | 216–279 | 0.810 |
| Satt363 | C2 | 12 | 229–276 | 0.786 |
| Satt203 | D1a+Q | 11 | 171–213 | 0.465 |
| Satt600 | D1b+W | 10 | 158–221 | 0.815 |
| Satt002 | D2 | 10 | 111–147 | 0.836 |
| Satt045 | E | 11 | 125–155 | 0.890 |
| Satt030 | F | 11 | 137–170 | 0.843 |
| Satt038 | G | 9 | 157–199 | 0.824 |
| Satt253 | H | 8 | 133–154 | 0.786 |
| Satt292 | I | 12 | 220–259 | 0.731 |
| Satt431 | J | 13 | 190–238 | 0.876 |
| Satt001 | K | 9 | 82–124 | 0.739 |
| Satt156 | L | 10 | 190–232 | 0.752 |
| Satt463 | M | 21 | 114–234 | 0.863 |
| Satt009 | N | 24 | 163–253 | 0.922 |
| Satt262 | O | 12 | 228–282 | 0.734 |
| Average | | 11.9 | | 0.782 |

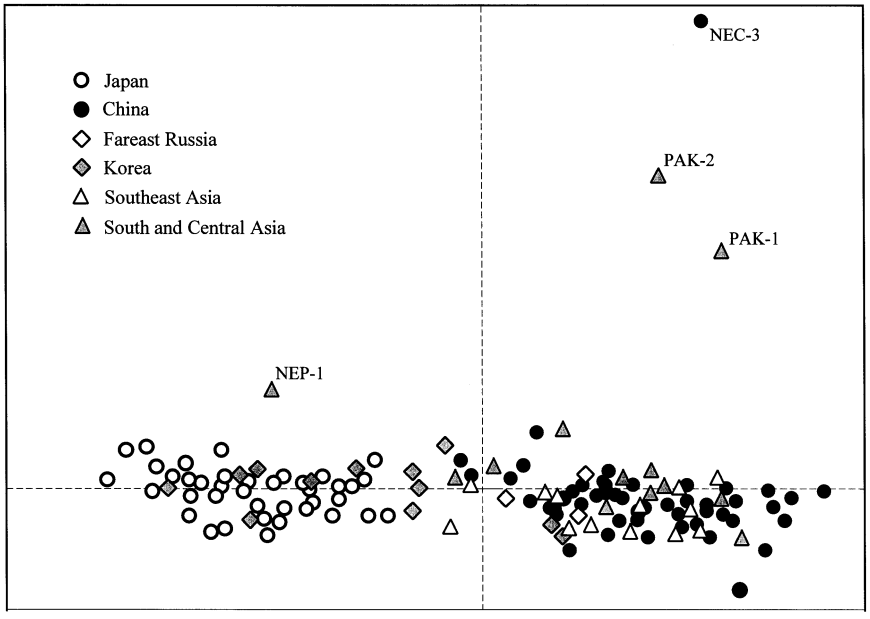


**Fig. 1** Distribution of Asian soybean accessions on the first-dimension (horizontal axis) and the second-dimension (vertical axis) obtained from Hayashi's Quantification Theory III for the variation at 20 SSR loci

The 20 SSR loci produced unique allelic profiles for all of the 131 accessions tested. The *D* value had an average of 0.784, indicating that for any pair of the accessions, an average of 78.4% of the loci had different alleles. The smallest *D* (0.05) was observed between two Japanese accessions, 'Rokujunichimame'(NJ-21) and 'Shirohanasai 1'(NJ-22), both of which were classified into the same cultivar group based on the variation at 16 isozyme and *Ti* loci (Hirata et al. 1999).

## Relatedness of soybean accessions in Asia

A scatter diagram of the 131 accessions on two-dimensions obtained from Hayashi's Quantification Theory III is presented in Fig. 1. The first-dimension (horizontal axis), which contributed 39.8% of the whole variation, completely separated the 39 Japanese accessions from the 51 Chinese accessions. Of the 12 Korean accessions, six overlapped the Japanese accessions, two overlapped the Chinese accessions, and four fell between both groups of accessions. Except for a Nepalese accession (NEP-1), all of the accessions from far-east Russia, Southeast Asia and south/central Asia were scattered, overlapping the Chinese accessions. In the second-

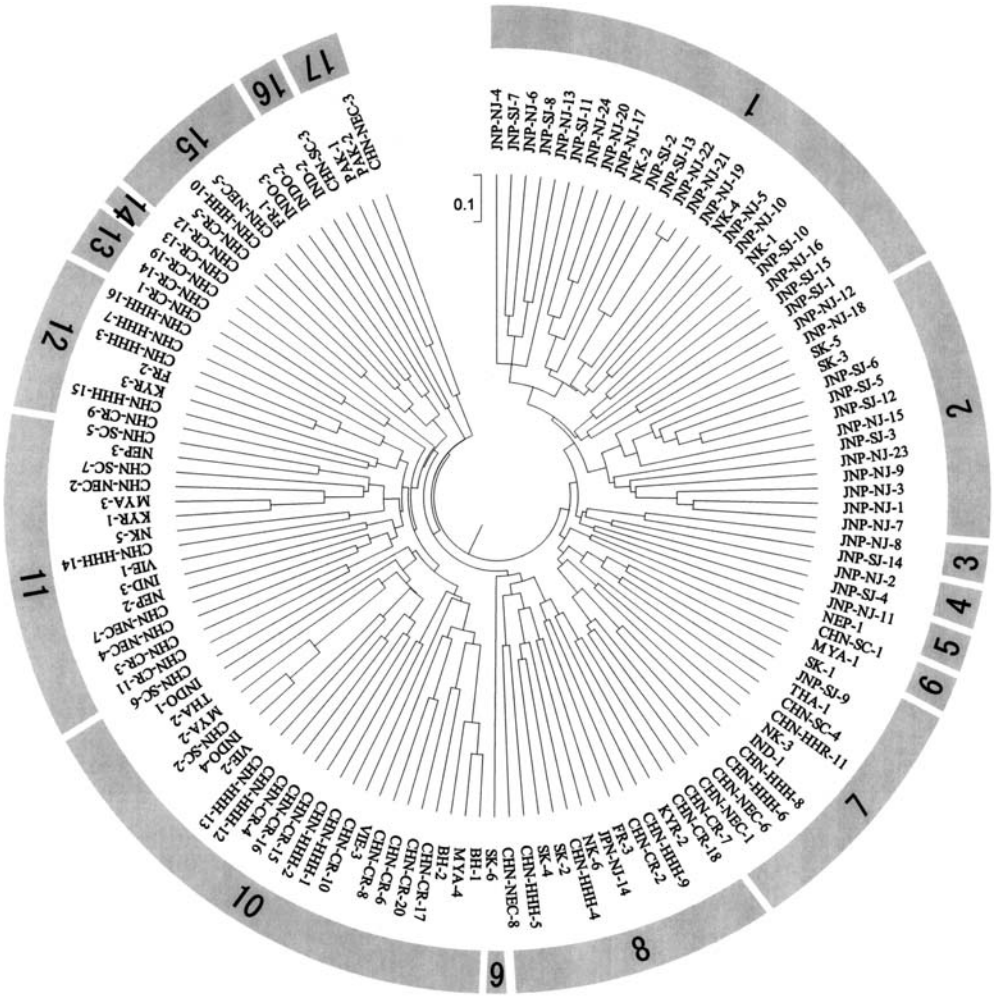


**Fig. 2** A UPGMA dendrogram representing phenetic relationships among 131 Asian soybean accessions. The dendrogram was constructed based on the genetic distances (*D*) between accessions

dimension (vertical axis), which contributed 28.3% of the whole variation, one Chinese accession 'Baihuamoshidou' (NEC-3) and two accessions from Pakistan (PAK-1 and PAK-2) stood out from the majority of accessions.

A cluster analysis with the UPGMA also clearly separated the Japanese and Chinese accessions (Fig. 2). The cluster analysis assigned the 131 accessions into two groups, A (cluster #1 to #7) and B (cluster #8 to #17). Except for 'Kimusume' (NJ-14) and 'Misaodaizu' (SJ-9), the Japanese accessions were included in five clusters (#1 to #5) of Group A. The Chinese accessions were scattered in two clusters (#6 and #7) of Group A and nine clusters (#8, #10 to #17) of Group B. No accession from China was contained in the four Japan-specific clusters (#1, #2, #3 and #4). One Japanese accession, NJ-14, was grouped with a Korean accession (NK-6), and was included in cluster #8. Another Japanese accession, SJ-9, was grouped with a Korean accession (SK-1), and was included in cluster #7. Both of these clusters were dominated by the Chinese accessions. The Korean accessions were included in two clusters specific to the Japanese accessions (cluster #1 and #2), three clusters specific to the Chinese accessions (cluster #7, #8 and #11), and an outlier (cluster #9). Except for NEP-1, all of the accessions from far-east Russia, Southeast Asia and south/central Asia were scattered widely in the Chinese clusters. Except for two accessions from Bhutan, the accessions from southeast and south/central Asian countries did not form their own clusters. The relationship of the five regional populations is also reflected in a UPGMA dendrogram based on the average genetic distances between accessions of the different populations (Fig. 3).

## Genetic differentiation between Japanese and Chinese soybean populations

Since multivariate analyses clearly separated the Japanese accessions from the Chinese accessions, comparisons were made between the Japanese and Chinese populations to identify alleles specific to each of the two populations (Table 3). Of the 212 alleles detected in the two populations, 44 were specific to the Japanese population, and 71 were specific to the Chinese population.

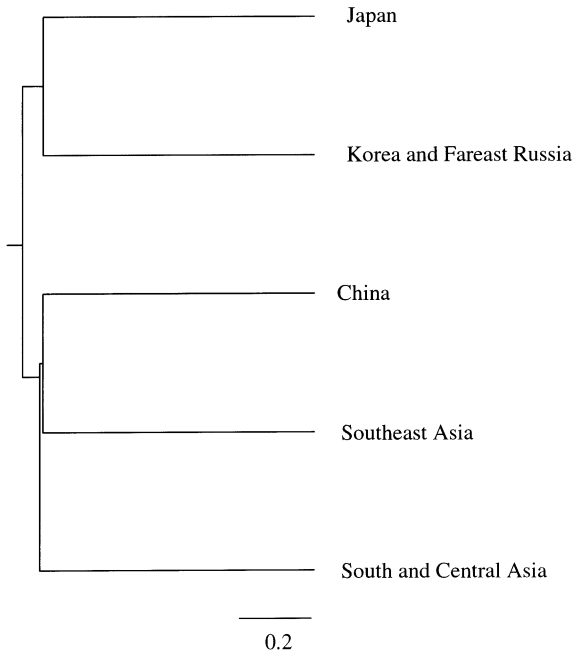


**Fig. 3** A UPGMA dendrogram representing relationships among five regional populations in Asia. The dendrogram was constructed based on the average genetic distances (*D*) between accessions of different populations

Most of the alleles specific to each population occurred at frequencies of 0.25 or less. Three alleles in the Japanese population (241 bp at Satt363, 221 bp at Satt600, 255 bp at Satt262) and five alleles in the Chinese population (106 bp at Satt063, 158 bp at Satt600, 175 bp and 178 bp at Satt038, 126 bp at Satt463) occurred at frequencies of more than 0.26. The AMOVA test for each of the SSR loci indicated that, of the total variation, the percentage of variation attributed to differences between the two populations ranged from 0.9% at Satt292 to 28.4% at Satt156, with an average of 12.2% (Table 3). In addition to the population-specific alleles, a number of alleles showed large differences in their frequencies between the two populations.

**Table 3** Frequencies of the two most-often occurring alleles at each of 20 SSR loci in the Japanese and Chinese populations. Bold numbers indicate population-specific alleles

| Locus | Japanese population | | Chinese population | | % Total[d] |
|---|---|---|---|---|---|
| | N[a] | Predominant alleles[b] | N[a] | Predominant alleles[b] | |
| Satt236 | 6 (1[c]) | 226 (36), 223 (27) | 7 (2[c]) | 226 (57), 229 (24) | 8.7 |
| Satt228 | 3 (1) | 218 (69), 215 (28) | 6 (4) | 215 (51), 218 (27) | 17.8 |
| Satt197 | 8 (1) | 135 (49), 186 (13) | 11 (4) | 183 (35), **180 (18)** | 14.3 |
| Satt063 | 6 (1) | 130 (46), 127 (21) | 8 (3) | **106 (37)**,145 (27) | 20.1 |
| Satt180 | 5 (0) | 216 (63), 261 (17) | 9 (4) | 261 (35), 267 (20) | 17.7 |
| Satt363 | 9 (5) | **241 (28)**, 262 (21) | 8 (4) | 259 (47), 262 (18) | 10.1 |
| Satt203 | 6 (3) | 183 (77) | 7 (4) | 183 (75) | 1.2 |
| Satt600 | 6 (4) | 218 (28), **221 (28)** | 6 (4) | **158 (39)**, 218 (23) | 15.1 |
| Satt002 | 7 (1) | 144 (41), 141 (26) | 7 (1) | 126 (35), **132 (18)** | 13.3 |
| Satt045 | 10 (1) | 125 (23), 137 (18) | 10 (1) | 134 (33), 140 (29) | 9.2 |
| Satt030 | 7 (1) | 161 (46), 164 (23) | 9 (3) | 167 (27), 152 (18), 164 (18) | 10.3 |
| Satt038 | 5 (1) | 157 (40), 172 (42) | 6 (2) | **175 (33)**, **178 (33)** | 22.4 |
| Satt253 | 4 (0) | 154 (46), 151 (21), 148 (21) | 6 (2) | 139 (37), 148 (29) | 12.7 |
| Satt292 | 5 (1) | 223 (49), 253 (28) | 8 (4) | 223 (45), 253 (24) | 0.9 |
| Satt431 | 8 (3) | 226 (33), 202 (28) | 9 (4) | 205 (20), **232 (18)** | 10.2 |
| Satt001 | 4 (1) | 109 (56), **121 (22)** | 8 (5) | 109 (39), 112 (22) | 7.0 |
| Satt156 | 5 (1) | 208 (67), 205 (26) | 9 (5) | 223 (43), 205 (24) | 28.4 |
| Satt463 | 13 (5) | 138 (21), 129 (15) | 14 (6) | 138 (33), **126 (26)** | 6.5 |
| Satt009 | 17 (9) | 217 (23), 220 (15) | 13 (5) | **163 (24)**, 220 (15) | 5.8 |
| Satt262 | 7 (4) | 246 (46), **255 (31)** | 7 (4) | 246 (39), 340 (35) | 12.1 |

[a] Number of alleles observed in each population
[b] Numbers in parentheses indicate the frequencies of each allele
[c] Numbers in parentheses indicate the number of alleles specific to the population
[d] Percentage of the total molecular variance contributed by the variance between the two populations

**Table 4** Genetic diversity statistics for five regional populations of Asian soybean

| Population | No. of alleles (*A*) | | Genetic diversity index (*H*) | | Average distance between accessions (*D*) | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| China (51[a]) | 8.4 | 2.1 | 0.740 | 0.095 | 0.755 | 0.101 |
| Japan (39) | 7.1 | 3.3 | 0.689 | 0.138 | 0.707 | 0.111 |
| Korea and Far-east Russia (15) | 6.0 | 1.7 | 0.720 | 0.148 | 0.771 | 0.101 |
| Southeast Asia (13) | 5.1 | 1.8 | 0.692 | 0.113 | 0.750 | 0.107 |
| South and Central Asia (13) | 5.5 | 1.3 | 0.685 | 0.098 | 0.742 | 0.120 |

[a] Numbers in parentheses indicate the number of accessions tested

### Genetic diversity in different regional populations

The genetic diversities of the five regional populations are compared in Table 4. The Chinese population showed the highest *A* and *H* values, while the Korean and far-east Russian population had the highest *D* value. No considerable differences in genetic diversity were found among the different populations although the numbers of the Chinese and Japanese accessions that were tested were greater than the numbers tested in the other three populations. Thus, each of the different soybean populations in Asia had a relatively high genetic diversity.

## Discussion

### Genetic diversity in the Asian soybean population

The present analysis detected a high level of length polymorphism at 20 SSR loci in the 131 Asian soybean accessions tested. Any pair of accessions, on average, had different alleles at about three-quarters of the 20 loci tested. The observed diversity among the Asian accessions is thus sufficient to evaluate their genetic relatedness, although the variation among individuals within accessions was not evaluated in this study.

The genetic diversity observed among the Asian accessions is much higher than the values reported in North American improved cultivars and ancestral plant introductions (Diwan and Cregan 1997; Song et al. 1999; Narvel et al. 2000). Of the 20 loci tested in this study, five (Satt002, 045, 038, 001, 009) were examined by Diwan and Cregan (1997), seven (Satt236, 197, 180, 002, 038, 259, 009) were examined by Song et al. (1999) and five (Satt063, 002, 045, 001, 009) were examined by Narvel et al. (2000). In the above five loci examined by Diwan and Cregan (1997), they reported finding a total of 44 alleles in 35 ancestral introductions and 22 alleles in 36 commercial cultivars, while we found a total of 63 alleles in our 131 Asian accessions. In the above seven loci examined by Song et al. (1999), they reported finding a total of 53 alleles in 35 ancestral introductions and 66 elite cultivars, while we found a total of 89 alleles. In the above five loci examined by Narvel et al. (2000), they reported finding a total of 29 alleles in 40 plant introductions and 21 alleles in 39 elite cultivars, while we found a total of 64 alleles. The present study thus revealed that there is a high level of genetic diversity in the Asian soybean population that has not been utilized in the North American population.

### Germplasm pools in the Asian soybean population

Multivariate analyses of the 20 SSR loci clearly allocated the observed diversity into two major groups, a Japanese group and a Chinese group. The distinctness of the Japanese and Chinese accessions is not a result of the sampling bias due to small sample sizes in both populations, because the former accessions were selected to cover a large part of the variation observed at 16 isozyme and seed-protein loci for 781 Japanese accessions (Hirata et al. 1999), and the latter accessions were selected to cover the major soybean cultivation areas in China. The difference in the allelic constitutions of the Japanese and Chinese populations is in good agreement with the findings obtained from analyses of the *Ti* locus (Hymowitz and Kaizuma 1979, 1981; Wang et al. 1986), the isozyme loci (Hirata et al. 1999) and the RAPDs (Li and Nelson 2001). All of these results suggest that the Japanese and Chinese soybean populations have different germplasm pools.

The Chinese accessions did not form clear region-specific clusters, except for cluster #13 which consisted of three accessions from the Changjiang River Valley area. A number of clusters, which consisted of two to four accessions from the same region of China, were formed, but were integrated with the accessions from the different regions into clusters of higher order. Hymowitz and Kaizuma (1981) assumed that two regions of China, northeast China and central and southern China, each represented different germplasm pools, based on the difference in the response to daylength and the usage of soybean, although there was no marked difference in allelic frequencies at *Ti* and *Sp1* loci between both regions. The present results do not support the existence of the two different germplasm pools in the Chinese population proposed by Hymowitz and Kaizuma (1981). There was also no evidence supporting the genetic differentiations among the four regions of China from which Chinese accessions were collected (Table 1).

One accession from Northeast China and two from Pakistan stood out from the rest of the accessions. These are forage or green manure soybeans and possess a number of primitive characters such as a climbing habit and small brown or black seeds. Further, according to Xu et al. (2002), these accessions all had a region-specific chloroplast SSR haplotype, which was predominantly observed in wild soybean plants distributed in the Huang-Huai-Hai River Valley and its surrounding areas. The present analysis revealed the existence of many alleles unique to each of the three accessions that were absent or rarely observed in the other accessions. Different allelic constitutions have also been observed between wild and cultivated soybeans (Maughan et al. 1995; Powell et al. 1996). The above three accessions, each having the region-specific chloroplast haplotype, unique nuclear SSR alleles and primitive phenotypes, are most-likely hybrids between wild and cultivated soybeans.

All of the accessions from Southeast Asian and south/central Asian countries, and far-east Russia, except for a Nepalese accession, were included in the Chinese germplasm pool. The accessions from these areas were thus more similar to the Chinese accessions than the Japanese ones. Except for two accessions from Bhutan, the accessions from Southeast Asia, south/central Asia and far-east Russia did not form their own clusters. Most of

the accessions were classified into different Chinese clusters, independently of their country origins. The relatively high genetic diversities and the absence of their own clusters may therefore indicate that soybean accessions in these areas had been introduced repeatedly and independently from the diverse Chinese germplasm pool.

The Korean accessions, on the other hand, were scattered, overlapping both the Japanese and Chinese germplasm pools. Hymowitz and Kaizuma (1981) reported a geographical cline of allelic frequencies at the *Ti* locus in the Korean population. The *Ti*-b allele is present at a high frequency in the Japanese population, whereas it is rarely observed in the Chinese accessions (Hymowitz and Kaizuma 1979, 1981; Wang et al. 1986). Hymowitz and Kaizuma (1981) found that, within Korea, the frequency of the *Ti*-b allele was low in the districts that were closest to China and high in the districts that were closest to Japan. A similar geographical cline of allelic frequencies was found in some isozyme loci in the South Korean population (Han et al. 1999). An involvement of the Korean accessions in both of the Japanese and Chinese germplasm pools suggests that the Korean population was intermingled with the accessions from the Japanese and Chinese pools. Alternatively, the Korean population might be ancestral to the two populations.

### Origin of the Japanese soybean population

The distinctness of the Japanese and Chinese populations has some implications regarding the origin of the Japanese soybean population. The soybean has been considered to have originated in the eastern half of northern China, and thereafter disseminated into various regions of Asia (Hymowitz and Newell 1981). As suggested by Li and Nelson (2001), one possibility is that the Japanese soybean population was established from a relatively small number of introductions from China that are rare or that have been lost in the present Chinese population. A few founders could have resulted in allelic constitutions that were more uniform and different from the ancestral population. However, the present study and our previous isozyme analysis (Hirata et al. 1999) revealed that there is no marked reduction in the genetic diversity in the Japanese population. A relatively high genetic diversity in the Japanese population does not necessarily support the above hypothesis.

Another possibility is that gene flow from sympatrically growing wild soybean plants may have resulted in the establishment of new alleles that were not present in the ancestral population. Recent organelle DNA analyses have revealed that the present Asian cultivated soybean population possessed different haplotypes for both the chloroplast and mitochondrial genomes, all of which were observed in the wild soybean (Shimamoto 2001; Xu et al. 2002). Some haplotypes observed in the cultivated soybean were distributed regionally and overlapped with the distribution of wild soybean accessions of the same respective haplotypes. The overlapping distribution in haplotypes of organelle genomes between wild and cultivated soybeans strongly suggests that various cultivated forms had originated in various regions of East Asia, including Japan, from independent wild populations or hybrid swarms between wild and primitive soybeans (Xu et al. 2002). Several observations support the hypothesis that cultivated forms were established as a result of human selection from hybrid swarms. These observations include the frequent occurrence of wild soybean near cultivated soybean fields in East Asia, a relatively high rate of cross-pollination observed in natural habitats of wild soybean (Fujita et al. 1997) and the existence of intermediate weedy forms between wild and cultivated soybeans (Sekizuka and Yoshiyama 1960; Zheng and Chen 1980). However, these observations do not necessarily exclude the hypothesis that cultivated forms were established from independent wild populations. The different allelic constitutions at the 20 SSR loci between the Japanese and Chinese populations may reflect modifications due to gene flow from sympatrically growing wild soybean plants after the primitive soybean had been disseminated. A detailed analysis of wild soybean is in progress to determine what roles the gene flow from wild soybean have played on the evolution of cultivated soybean.

In conclusion, the results obtained in this study indicate that each of the Japanese and Chinese populations form different germplasm pools in soybean. A pedigree analysis of the improved Japanese cultivars demonstrated that only a small percentage of the whole genetic variation in the improved cultivar population came from the Chinese population (Zhou et al. 2000). On the other hand, Cui et al. (2000) demonstrated with pedigree analysis that the improved Chinese cultivars have been mainly bred with the Chinese landrace itself and with North American cultivars. The present results thus suggest that both the Japanese and Chinese populations can be used as exotic genetic resources to enlarge their respective genetic bases, and also those of populations from different countries.

**Acknowledgements** This research was supported by the Japan Society for the Promotion of Science (JSPS) (Grant No. PI98120 and No. 12460142). We are thankful to Dr. H. Knüpffer (Institute of Plant Genetics and Crop Research, Germany), Dr. P.P. Bulakh (Fareast Station of Vavilov Institute, Russia) and Dr. R.L. Nelson (USDA-Agricultural Research Service, USA) for their providing soybean accessions. We also appreciate Professor J.Y. Gai (Nanjing Agricultural University, China) for his great help.

## References

Benzecri JP (1992) Correspondence analysis handbook. Marcel Dekker, New York

Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML (2000) Evaluation of genetic diversity of soybean introductions and north American ancestors using RAPD and SSR markers. Crop Sci 40:815–823

Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, VanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39:1464–1490

Cui Z, Carter TE, Burton JW (2000) Genetic diversity patterns in Chinese soybean cultivars based on coefficient of parentage. Crop Sci 40:1780–1793

Diwan N, Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat markers to assay genetic variation in soybean. Theor Appl Genet 95:723–731

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA data. Genetics 131:479–491

Felsenstein J (1995) PHYLIP (Phylogeny inference package) version 3.57c. University of Washington Press, Seattle

Fujita R, Ohara M, Okazaki K, Shimamoto Y (1997) The extent of natural cross-pollination wild soybean (*Glycine soja*). J Hered 88:124–128

Griffin JD, Palmer RG (1995) Variability of thirteen isozyme loci in the USDA soybean germplasm collections. Crop Sci 35:897–904

Han O, Abe J, Shimamoto Y (1999) Genetic diversity of soybean landraces in Korea. Korean J Crop Sci 44:256–262

Hayashi C (1956) Theory and examples of quantification III (in Japanese). Proc Institute Statistical Math, Tokyo 4:19–30

Hirata T, Abe J, Shimamoto Y (1999) Genetic structure of the Japanese soybean population. Genet Resour Crop Evol 46:441–453

Hymowitz T, Kaizuma N (1979) Dissemination of soybean (*Glycine max*): seed protein electrophoresis profiles among Japanese cultivar. Econ Bot 33:311–319

Hymowitz T, Kaizuma N (1981) Soybean seed protein electrophoresis profiles from 15 Asian countries or regions: hypotheses on paths of dissemination of soybeans from China. Econ Bot 35:10–23

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

Keim P, Shoemaker RC, Palmer RG (1989) Restriction fragment length polymorphism diversity in soybean. Theor Appl Genet 77:786–792

Kisha TJ, Diers BW, Hoyt JM, Sneller CH (1998) Genetic diversity among soybean plant introductions and north American germplasm. Crop Sci 38:1669–1680

Li Z, Nelson RL (2001) Genetic diversity among soybean accessions from three countries measured by RAPDs. Crop Sci 41:1337–1347

Li Z, Qiu L, Thompson JA, Welsh MM, Nelson RL (2001) Molecular genetic analysis of U.S. and Chinese soybean ancestral lines. Crop Sci 41:1330–1336

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

Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL (1997) Microsat (version <WHATEVER>): a computer program for calculating various statistics on microsatellite allele data. WWW: http://hpgl.stanford.edu/projects/microsat/

Narvel JM, Fehr WR, Chu WC, Grant D, Shoemaker RC (2000) Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Sci 40:1452–1458

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Perry MC, McIntosh MS, Stoner AK (1991) Geographical patterns of variation in the USDA soybean germplasm collection. II. Allozyme frequencies. Crop Sci 31:1356–1360

Powell W, Morgante M, Doyle JJ, McNicol JW, Tingey SV, Rafalski AJ (1996) Genepool variation in the genus *Glycine* subgenus *Soja* revealed by polymorphic nuclear and chloroplast microsatellites. Genetics 144:793–803

Sekizuka S, Yoshiyama T (1960) Studies on the native wild grasses for fodder. IV. Crop-scientific studies on wild species of *Glycine soja* in Japan (in Japanese with an English abstract). J Kantou-Tosan Agric Exp Station 15:57–73

Shimamoto Y (2001) Polymorphism and phylogeny of soybean based on chloroplast and mitochondrial DNA analysis. Japan Agric Res Quarterly 35:79–84

Song QJ, Quigley CV, Nelson RL, Carter TE, Boerma HR, Strachan JL, Cregan PB (1999) A selected set of trinucletide simple sequence repeat markers for soybean cultivar identification. Plant Var Seeds 12:207–220

Wang Y, Li F, Chang R (1986) The distribution of the alleles of the *Ti* and *Sp1* loci among Chinese soybean (*G. max*) cultivars by seed protein electrophoresis (in Chinese with an English abstract). Acta Agron Sinica 12:31–37

Xu DH, Abe J, Gai JY, and Shimamoto Y (2002) Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet (in press)

Zheng HY, Chen HD (1980) A preliminary study of the resources of wild soybean in Jilin province (in Chinese with an English abstract). Sci Agric Sinica 13:26–32

Zhou X, Carter TE, Cui Z, Miyazaki S, Burton JW (2000) Genetic base of Japanese soybean cultivars released during 1950 to 1988. Crop Sci 40:1794–1802