Theor Appl Genet (2004) 109:352–360
DOI 10.1007/s00122-004-1634-8

ORIGINAL PAPER

X. W. Wang · A. Kaga · N. Tomooka · D. A. Vaughan

The development of SSR markers by a new method in plants and their application to gene flow studies in azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]

Received: 31 May 2003 / Accepted: 9 February 2004 / Published online: 10 March 2004

Abstract To gain a better understanding of wild and weedy azuki population structures in relation to the cultigens we have developed simple sequence repeat (SSR) markers based on a new methodology for plant material. In the azuki bean genome, the number of $(AG)_n$ and $(AC)_n$ motif loci per haploid genome has been estimated to be 3,500 and 2,100, respectively, indicating that $(AG)_n$ motifs are a rich source of markers. We constructed a $(AG)_n$-SSR-enriched library in azuki bean in order to obtain a comprehensive range of SSR markers efficiently. The method applied in this study resulted in a 116-fold enrichment over the non-enriched genomic library, with a high percentage (98%) of successful single-locus amplification by the primer pairs designed. Consequently, this method can be applied to construct SSR-enriched libraries suitable for large-scale sequencing. We obtained 255 unique sequences from an $(AG)_n$-enriched library for azuki bean. Fifty primer pairs were designed and screened against five populations of wild azuki bean. Among these five populations, one population from Bato town, Tochigi prefecture, Japan, showed greater polymorphism using these primers than the others and was therefore chosen for the in-depth study. The genotypes of 20 individuals were investigated using eight of the SSR primers developed. The genetic relationships among individuals revealed a complex spatial pattern of population structure. Although azuki bean is considered to be a predominantly self-pollinating species, 3 of the 20 individuals tested in the population showed heterozygous genotypes, indicating outcrossing. Allele size and DNA sequence in each of the 20 individuals were compared with those of landraces and released cultivars of azuki bean. Plants in part of the population had many alleles of the same size and with the same sequence as those in cultivated azuki bean, suggesting that gene flow from the cultigen to wild plants has occurred in this population. Unintentional transgene escape from azuki could therefore occur when transgenic azuki is grown in areas where its wild and weedy relatives occur. The approach used here could be applied to biosafety monitoring of transgenic azuki bean.

Communicated by C. Möllers

X.W. Wang and A. Kaga contributed equally to the results presented in this paper.

X. W. Wang · A. Kaga (✉) · N. Tomooka · D. A. Vaughan
Crop Evolutionary Dynamics Team,
National Institute of Agrobiological Sciences,
Kannondai 2-1-2, Tsukuba, 305-8602 Ibaraki, Japan
e-mail: kaga@nias.affrc.go.jp

Introduction

Azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi] is an important grain legume in East Asia. In Japan, azuki bean is the second most economically important grain legume—after soybean—and it is cultivated throughout the country except for Okinawa prefecture (Tomooka et al. 2002). The putative wild progenitor, *Vigna angularis* var. *nipponensis*, and weedy forms are distributed throughout Japan except for Hokkaido prefecture. The weedy form shows an intermediate phenotype between the cultivated and wild azuki bean for several traits (Yamaguchi and Nikuma 1996). Cultivated, wild and weedy populations of azuki bean in Japan constitute a crop complex (Vaughan et al. 2004). Populations consisting of mixtures of wild and weedy plants are referred to as inter-breeding complexes (Beebe et al. 1997) or complex populations (Tomooka et al. 2002), and such populations of azuki bean can be found in Japan. The distribution of various types of natural azuki populations in Japan suggests that Japan is a center of diversity for azuki bean. However, a lack of markers that can identify heterozygous genotypes in this legume crop has limited our progress in understanding the genetic diversity and gene flow among components of the azuki bean complex.

Simple sequence repeats (SSRs) are abundant and dispersed throughout all of the eukaryotic genomes (Morgante et al. 2002). The frequency of SSRs in plant genomes is estimated as one in every 6–7 kb based on the information that can be found in public sequence databases (Cardle et al. 2000). SSRs are thus an abundant

resource in the genome and have a high level of allelic diversity; consequently, they are frequently used as genetic markers in plant genetics studies (reviews in Powell et al. 1996; McCouch et al. 1997). The co-dominant nature and allelic polymorphism revealed by SSR markers has provided detailed information on genetic structure (Bonnin et al. 2001; Li et al. 2000) and gene flow (Konuma et al. 2000) in natural plant populations.

Despite the advantages of SSR markers, their development is time-consuming for plant species for which there is little DNA sequence information in the public databases. The public DNA sequence data for the genus *Vigna* is very limited, therefore, it is necessary to find SSRs and obtain the sequences around those SSRs in order to design primer pairs. The relatively low efficiency of SSR identification in genomic libraries using traditional hybridization has led scientists to seek more efficient protocols, such as enrichment methods (reviewed in Zane et al. 2002; Squirrell et al. 2003). Most of the enrichment methods recently reported for plants are modifications based on the method of duplex-hybridization-based formation. In mammals, enrichment based on oligo-primed second-strand synthesis has been reported to be highly efficient for AC motif discovery and has been used for the large-scale discovery of microsatellite repeat sequences (Ostrander et al. 1992; Takahashi et al. 1996).

The objectives of the investigation reported here were: (1) to construct an enriched SSR library for azuki bean; (2) to evaluate the oligo-primed second-strand synthesis enrichment method with respect to large-scale discovery of SSRs in azuki bean; (3) to determine whether the SSRs identified can be used to analyze the intra-population genetic diversity of wild/weedy azuki bean; (4) to determine whether the SSRs identified can be used to analyze gene flow in the azuki bean complex.

Materials and methods

Plant materials and DNA extraction

Vigna angularis cv. Erimoshouzu was used to construct the genomic library. Five natural populations of wild azuki (var. *nipponensis*) were selected from different locations in Japan (Tomooka et al. 2001). Bulk DNA isolated from 12–17 seedlings per population according to the method of Draper and Scott (1988) was used to evaluate the allelic variation in these five populations (Table 1, nos. 1–5). In order to evaluate gene introgression from cultivated azuki bean, we used 20 individuals of the Bato town population, Tochigi prefecture (Table 1, no. 1), 15 landraces originating from Tochigi prefecture (Table 1, nos. 9–23) and two released varieties (Table 1, nos. 7 and 8). DNA was extracted from leaves of individual plants of the Bato town population immediately following their collection from the natural habitat according to the method mentioned above. DNA of the landraces and released varieties was isolated from a single seed according to the method of Kang et al. (1998).

Genomic library construction

Insert fragments were prepared as described by Birren et al. (1997). In brief, 2.5 μg genomic DNA was digested with 35 U/ml DNase I (TaKaRa, Japan) in the presence of manganese chloride. After removing DNase I using Micropure-EZ (Millipore, Billerica, Mass.), we repaired the ends of the fragments with 5 U T4 DNA polymerase (TaKaRa) and 5 U Klenow fragment (TaKaRa) and phosphorylated them with 1 U T4 polynucleotide kinase (TaKaRa). The fragments were separated on 2% NuSieve GTG low-melting

Table 1 Populations of *Vigna angularis*

Number	Name[a]	Population type	Longitude/latitude	Prefecture	Application
1	Bato population	Complex	N36°46'18''/E140°12'23''	Tochigi	Polymorphism survey, population analysis
2	Tatebe population	Complex	N34°52'55''/E133°54'07''	Okayama	Polymorphism survey
3	Koge population	Complex	N35°25'17''/E134°15'39''	Tottori	Polymorphism survey
4	Miwa population	Wild	N35°26'54''/E134°12'11''	Tottori	Polymorphism survey
5	Tsukuba population	Wild	N36°12'11''/E140°09'46''	Ibaraki	Polymorphism survey
6	Erimoshouzu	Released variety	-	Hokkaido	Library construction
7	Tochigi enba No.1	Released variety	-	Tochigi	Comparison of allelic diversity
8	Tochigi No.1	Released variety	-	Tochigi	Comparison of allelic diversity
9	ACC0284	Landrace	N36°33'/E139°41'	Tochigi	Comparison of allelic diversity
10	ACC1645	Landrace	-	Tochigi	Comparison of allelic diversity
11	ACC2016	Landrace	-	Tochigi	Comparison of allelic diversity
12	ACC2047	Landrace	-	Tochigi	Comparison of allelic diversity
13	ACC2326	Landrace	N36°42'/E139°49'	Tochigi	Comparison of allelic diversity
14	ACC2327	Landrace	N36°41'/E139°49'	Tochigi	Comparison of allelic diversity
15	ACC2328	Landrace	N36°46'/E139°47'	Tochigi	Comparison of allelic diversity
16	ACC2329	Landrace	N37°00'/E139°44'	Tochigi	Comparison of allelic diversity
17	ACC2330	Landrace	N37°00'/E139°44'	Tochigi	Comparison of allelic diversity
18	ACC2331	Landrace	N36°53'/E139°41'	Tochigi	Comparison of allelic diversity
19	ACC2332	Landrace	N36°59'/E139°39'	Tochigi	Comparison of allelic diversity
20	ACC2333	Landrace	N36°51'/E139°43'	Tochigi	Comparison of allelic diversity
21	ACC2334	Landrace	N36°48'/E139°43'	Tochigi	Comparison of allelic diversity
22	JP048748	Landrace	-	Tochigi	Comparison of allelic diversity
23	JP110705	Landrace	N36°46'18''/E140°12'23''	Tochigi	Comparison of allelic diversity

[a] Accessions with the prefix ACC or JP were obtained from the Hokkaido Prefectural Tokachi Agricultural Experiment Station and the Ministry of Agriculture, Forestry and Fisheries (MAFF) Genebank, respectively

agarose gel (BioWhitake Molecular Applications, Rockland, Me.). In order to remove the shorter fragments completely, a gel slice containing 500 to 700 bp long fragments was cut out and embedded into 2% low-melting agarose gel after the direction of the gel slice had been reversed. Following size-fractionation, a gel segment was again sliced out at the position where fragments were concentrated, embedded into 1% low-melting agarose gel and subjected once again to size-fractionation. The fragments were then recovered from the low-melting agarose gel with GELase (Epicentre technologies, USA). About 200 ng of the fragments was ligated to 100 ng of *Eco*RV-digested and dephosphorylated pBluescript SK (+) vector (Stratagene, La Jolla, Calif.) using Ligation Kit ver. 2 (TaKaRa) at 16°C for 30 min, then the ligase was inactivated at 68°C for 10 min. The ligation reaction was transformed into TG1-competent cells (Stratagene) using Gene pulser II (Bio-Rad, Hercules, Calif.) under the conditions recommended by the manufacturer. Transformed cells were re-suspended in SOC medium and incubated at 37°C for 30 min. They were then plated on LB plates containing 50 μg/ml ampicillin, 80 μg/ml X-gal and 0.25 m*M* IPTG.

Production of pooled single-stranded DNA

After confirming the insert size and background at the test transformation above, the remaining ligation reaction was transformed again. The procedure of Vieira and Messing (1987) was used to produce pooled single-stranded DNA in vivo. In brief, the transformed cells were incubated in NZY broth, instead of SOC medium, with shaking at 37°C for 1 h and selected with 50 μg/ml ampicillin for an additional hour. The culture was superinfected with VCSM13 helper phage (Stratagene) at 37°C for 20 min without shaking and then grown in 100 ml Terrific Broth containing 75 μg/ml ampicillin and 70 μg/ml kanamycin with shaking at 37°C for 5 h. The phagemid particles were pelleted with PEG and extracted with phenol/chloroform to obtain ssDNA.

Enrichment of clones containing SSRs

The enrichment procedure used followed that of Takahashi et al. (1996) with slight modifications. The complementary strand of ssDNA was extended with *Taq* polymerase from a 5′- phosphorylated $(AG)_{12}$ primer. A 500 ng aliquot of pooled ssDNA was primed with 2 pmol 5′- phosphorylated $(AG)_{12}$ primer in 100 μl reaction mixture containing 20 m*M* dNTP and 1× *ExTaq* buffer. The mixture was denatured first for 20 s at 90°C followed by 2 min at 72°C, 1 U *ExTaq* (TaKaRa) was added and the mixture was incubated for 15 min at 72°C. The product was extracted with phenol/chloroform and treated with T4 Ligase (TaKaRa) for 30 min at 16°C to fill the gap between the 5′ and 3′ end of the complement strand. The ligation mixture was incubated with 40 U of mungbean nuclease (TaKaRa) for 2 h to digest the remaining ssDNA. After extraction with phenol/chloroform and precipitation with ethanol, the DNA pellet was dissolved in 20 μl H_2O. Three microliters of the solution was transformed into DH5α.

Colony hybridization

White colonies were transferred into 384-well microtiter plates containing LB freezing medium [36 m*M* K_2HPO_4, 13.2 m*M* KH_2PO_4, 1.7 m*M* tri-sodium citrate dihydrate, 0.4 m*M* $MgSO_4$, 6.8 m*M* $(NH_4)_2SO_4$, 4.4% (v/v) glycerol] and 100 μg/ml ampicillin and incubated at 37°C for approximately 20–24 h. The colonies were replicated onto a Hybond N^+ membrane (Amersham Pharmacia Biotech, UK) on LB-Amp agar plates using 384-pin replicators (Genetix, UK) and incubated at 37°C overnight. The membrane was treated on Whatman 3 mm paper saturated with 10% sodium dodecyl sulfate (SDS) for 5 min, denaturation buffer (0.5 *M* NaOH, 1.5 *M* NaCl) for 10 min and neutralization buffer (0.5 M Tris-HCl, 1.5 M NaCl) twice for 5 min each time. The membrane was washed vigorously in 2× SSC for 5 min, baked for 2 h at 80°C and soaked in 2× SSC containing 2 mg/ml Proteinase K (Nakalai tesque, Japan) for 3 h at 37°C. Colony hybridization was carried out using the Biotin Luminescent Detection kit according to the manufacturers instructions (Roche Molecular Biochemicals, Germany). The membrane was hybridized in hybridization buffer (5× SSC, 1% blocking reagent, 0.1% *N*-lauroylsarcosine, 0.02% SDS) with 50 pmol/ml 5′-biotin-labeled $(AC)_{12}$ or $(AG)_{12}$ oligonucleotide at 65°C overnight. The membranes were washed twice at 65°C for 15 min in washing buffer (0.5× SSC, 0.1% SDS).

Sequence analysis

The plasmid DNA of the positive colonies was prepared using MultiScreen 96-well filter plates (Millipore). The DNA sequence was determined using the Thermo Sequenase II Dye Terminator Cycle Sequencing kit (Amersham Pharmacia Biotech) and either the T3 or T7 primer on an ABI PRISM 373 Genetic Analyzer (Applied Biosystems, Foster City, Calif.). The sequences were analyzed using the BIOEDIT program (Hall 1999). Primer pairs that will amplify a 100 to 200 bp fragment were designed based on the sequence flanking the SSR. PCR fragments were purified using MultiScreen PCR plates (Millipore) and used as template DNA for direct sequencing with a single primer designed to amplify the fragment.

Detection of SSR loci

Amplification was performed in a 10-μl volume containing 0.4 ng of genomic DNA, 1 U KOD-plus (*Thermococcus kodakaraensis*, KOD1 strain) DNA polymerase (TOYOBO, Japan), 1× KOD-plus PCR buffer, 0.2 m*M* dNTPs, 1 m*M* $MgSO_4$ and 5 pmol of forward and reverse primers. The PCR was carried out in a GeneAmp PCR system 9700 (Applied Biosystems) programmed as follows: 2 min at 94°C, followed by 30 cycles of 15 s at 94°C, 15 s at 60°C and 15 s at 68°C. The 5′-end of the reverse primer of CEDG007 and CEDG008 was labeled with 5-FAM, CEDG015 and CEDG024 with VIC, CEDG026 and CEDG029 with NED and CEDG033 and CEDG043 with PET (Applied Biosystems), respectively. One microliter of PCR product diluted ten-fold with water was mixed with 9 μl of Hi-Di formamide containing 0.1 μl GeneScan 500 LIZ size standard (Applied Biosystems). The heat-denatured products were run on an ABI Prism 3100 Genetic Analyser, and the allele data were analyzed with GENEMAPPER (Applied Biosystems).

Population analysis

Nei et al.'s (1983) D_A distances between individuals were calculated with the POPULATIONS ver. 1.2.23 program (Langella 2000). An unrooted phenogram showing relationships between individuals based on D_A distance was constructed using the neighbour-joining method (Saitou and Nei 1987). The bootstrap values were calculated from 1,000 times re-sampling.

Results

Non-enriched library

A genomic library was constructed with pBluescript II using DNase I-digested DNA fragments from cultivated azuki bean (cv. Erimoshouzu). The cloning efficiency was approximately 8.1×10^5 colonies/μg of vector DNA. About 85% of the transformed colonies contained inserts. The inserted fragments ranged in size from 400 bp to 700 bp, with an average size of 500 bp. In order to check the frequency of $(AG)_n$ and $(AC)_n$ motifs in the azuki bean genome, we transferred the white colonies to 384-well

plates; this is referred to as the non-enriched library. Among the 1,536 colonies of the non-enriched genomic library, five (0.33%) and three (0.20%) colonies revealed a positive signal when hybridized with the $(AG)_{12}$ and $(AC)_{12}$ oligonucleotide probes, respectively. The $(AG)_n$ motifs were more frequent than the AC repeats in azuki bean. Consequently, we constructed an $(AG)_n$-enriched library for azuki bean.

$(AG)_n$-enriched library

Approximately 6 μg of pooled ssDNA was recovered from phage particles that were secreted from 2.4×10^5 TG1 cells of the genomic library using the VCSM13 helper phage. The complementary strand of ssDNA containing the $(CT)_n$ motif was extended with *Taq* polymerase from a 5′-phosphorylated $(AG)_{12}$ primer in the pooled ssDNA. Following enzymatic elimination of ssDNA without extension, the resulting pooled dsDNA containing the $(AG)_n$ motif was transformed into DH5α cells; this is referred to as the $(AG)_n$-enriched library. A total of 5,338 white colonies (95.6% of the total) were obtained from approximately 500 ng pooled ssDNA. Among the 1,536 clones tested, 588 (38.3%) showed a positive signal when hybridized with the $(AG)_{12}$ oligonucleotide probes. Of the 277 positive clones sequenced, 273 (98.6%) had at least one $(AG)_n$ motif while the other four had no motif. Two hundred and forty clones contained unique sequences flanking the motif. However, the remaining 33 clones were duplicated more than once in the enriched library; 13 clones were found twice, one was found three times and one was found four times. The 33 clones were classified into 15 unique clones. In total, 255 unique sequences were obtained from 240 clones and the 15 unique clones. The sequences flanking the $(AG)_n$ motif could be used to design primers for 223 clones. The remaining 32 sequences were not suitable for designing primers because either the repeats were too close to the cloning site or the flanking sequence had inadequate GC-content.

Characterization of (AG)n motif sequences

Perfect $(AG)_n$ motifs were predominant in the enriched library. According to the system of Weber (1990), the 255 unique sequences could be classified into three categories: 148 perfect $(AG)_n$ motifs (58.0%), 39 imperfect $(AG)_n$ motifs (15.3%) and 68 compound motifs (26.7%). Among the compound sequences, $(AG)_n$ motifs accompanied by an $(AT)_n$ motif were found in 40 sequences (58.8%). The number of (AG) repeat units (n) in the 255 unique sequences varied between 4 and 63. The majority of $(AG)_n$ motifs showed between 10 and 20 repeat units (Fig. 1).

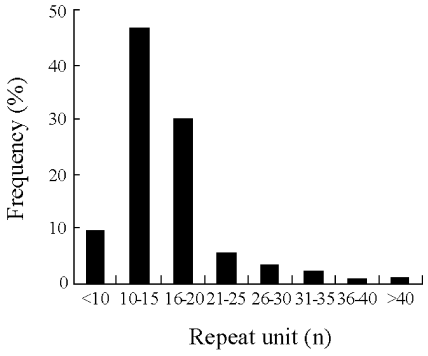

Fig. 1 Number of repeats in the 255 unique SSR motif sequences from the $(AG)_n$-enriched library

Application of SSR primers in azuki bean population analysis

Population and primer screening

From the 223 sequences available for primer design, 50 primer pairs were designed with 50% GC content. Most of primer pairs amplified single fragments of anticipated sequence length, with the exception of one primer pair that did not produce a PCR product even when various PCR conditions were tested. Five natural populations of *V. angularis* var. *nipponensis*, each represented by DNA bulked from 12–19 individuals, were screened with the 49 primer pairs (Table 1, nos.1–5). Among the primer pairs evaluated, eight primer pairs showed clear differences in the number of alleles between complex and wild populations (Table 2). These marker loci were named CEDG007, CEDG008, CEDG015, CEDG024, CEDG026, CEDG029, CEDG033 and CEDG043. Detailed informa-

Table 2 Allelic variation at eight SSR loci in five natural populations of the azuki bean complex

Name	Type	*n*[a]	Number of alleles at each locus								
			CEDG007	CEDG008	CEDG015	CEDG024	CEDG026	CEDG029	CEDG033	CEDG043	Average
Bato population	Complex	17	3	4	3	3	4	2	3	2	3.0
Koge population	Complex	15	3	4	4	3	4	3	4	2	3.4
Tatebe population	Complex	12	4	3	3	3	5	4	3	1	3.3
Miwa population	Wild	15	1	2	2	1	1	2	1	1	1.4
Tsukuba population	Wild	16	1	1	1	1	1	1	2	1	1.1
Total no. of alleles/ locus for all populations			8	9	10	5	11	8	11	4	8.3

[a] The number of individuals bulked in the population for DNA extraction

Table 3 Primer sequences, repeat units, PCR product sizes and accession numbers of the SSR loci

Locus	Repeat	Primers (5′-3′)	Product sizes[a] (bp)	Genbank accession no.
CEDG007	$(AG)_{16}$	GTGCAGCCACTACATGAATG GAAGTTGACACTCATCCACC	123	AB128069
CEDG008	$(AG)_{26}$	AGGCGAGGTTTCGTTTCAAG GCCCATATTTTTACGCCCAC	119	AB128079
CEDG015	$(AG)_{27}$	CCCGATGAACGCTAATGCTG CGCCAAAGGAAACGCAGAAC	211	AB128093
CEDG024	$(AG)_{18}$	CATCTTCCTCACCTGCATTC TTTGGTGAAGATGACAGCCC	140	AB128100
CEDG026	$(AG)_{16}$	TCAGCAATCACTCATGTGGG TGGGACAAACCTCATGGTTG	162	AB128113
CEDG029	$(AG)_{33}$	GATTGCTTTTAGCAGAGGGC GAAGAAACCCATCTCGATCC	185	AB128120
CEDG033	$(AG)_{41}$	GTGAGGTAGCTATGTAGCAC ACTGGACCGACAAGAGTAAG	131	AB128136
CEDG043	$(AG)_{14}$	AGGATTGTGGTTGGTGCATG ACTATTTCCAACCTGCTGGG	159	AB128143

[a] Allele sizes were determined using an ABI Prism 3100 Genetic Analyzer

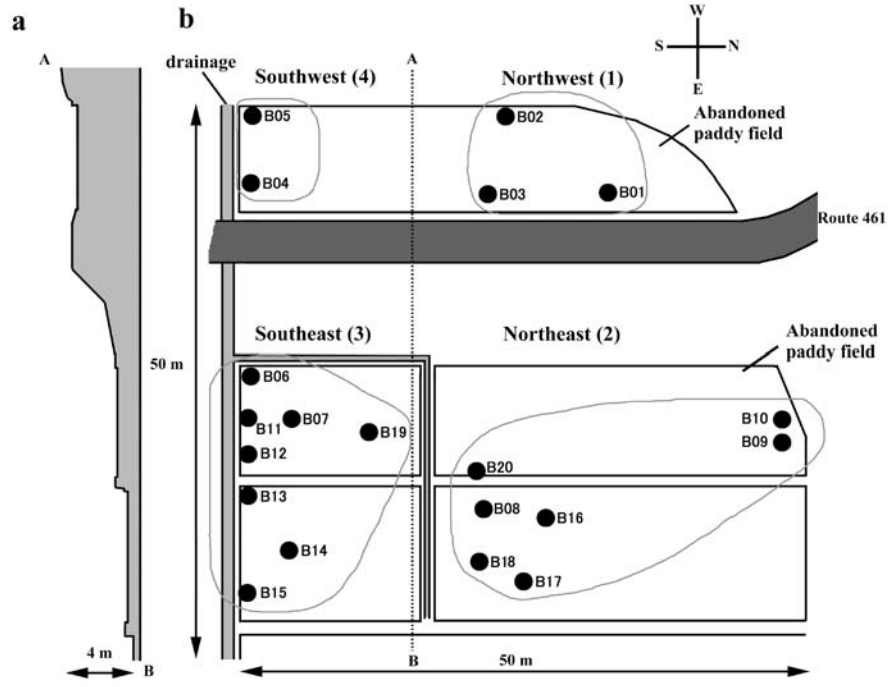

Fig. 2 Transect A–B (**a**) and overhead view (**b**) of the collection site at Bato town, Tochigi prefecture, Japan. A *black dot* by the *number* indicates the location of the individual collected

tion on the SSR motif, primer sequences and the expected PCR product size at each locus is listed in Table 3. The average number of alleles in the complex populations (3.0–3.4) was higher than that found in wild populations (1.1–1.4).

Analysis of one population at Bato town

As an example of the use of SSR markers in azuki bean population studies, one population from Bato town was selected for further analysis. The geographical features of the site and the position of the 20 individuals sampled (so that they reflect population density) are depicted in Fig. 2. This site in Bato town consists of abandoned terraced fields on the west side of the Mumo River valley. The ends of transects A (to the west) and B (to the east) (Fig. 2a) correspond to the lower slope of the valley side and the riverside, respectively. The terraced paddy fields where the azuki complex population was growing had been fallow for about 10 years. The site area is approximately 50×50 m and dissected by Route 461 (Fig. 2b). The difference in elevation between the highest and lowest position that individuals were collected is approximately 4 m (Fig. 2a).

The genotype of 20 individuals was determined using eight SSR primer pairs (Table 4). A total of 33 alleles with an average of four alleles per locus were found in this population. Sequencing revealed that the differences in allele sizes were due to the number of AG repeats.

Table 4 Sizes of alleles detected at eight SSR loci in 20 individuals from the Bato town complex population and cultivars of azuki bean

Location in Bato population	Plant number	Allele sizes[a] (bp)							
		CEDG007	CEDG008	CEDG015	CEDG024	CEDG026	CEDG029	CEDG033	CEDG043
Northwest (1)	B01, B02, B03	137:137	139:139	231:231	138:138	186:186	165:165	129:129	147:147
Northeast (2)	B08, B09, B10, B16, B17, B18, B20	139:139	141:141	231:231	138:138	184:184	165:165	129:129	147:147
Southeast (3)	B15	139:139	109:109	253:253	138:138	162:162	179:179	129:129	147:147
	B06	119:119	113:113	215:215	130:130	166:166	179:179	129:129	159:159
	B07	121:121	113:113	**215:253**	130:130	166:166	179:179	131:131	159:159
	B14	121:121	109:109	253:253	**130:140**	162:162	179:179	147:147	159:159
	B11, B12, B19	121:121	115:115	253:253	130:130	162:162	179:179	131:131	159:159
	B13	121:121	115:115	253:253	140:140	162:162	179:179	131:131	159:159
Southwest (4)	B05	121:121	**115:125**[b]	**213:253**	**128:140**	**162:204**	**179:187**	**105:131**	**159:163**
	B04	121:121	125:125	213:213	128:128	204:204	187:187	105:105	163:163
Allele size range of Bato population		119–139	109–141	213–253	128–140	162–204	165–187	105–147	147–163
Allele size range of cultivars		117–123	101–125	197–209	130–142	156–168	159–189	103–157	157–163

[a] Allele sizes were determined using an ABI Prism 3100 Genetic Analyzer
[b] Numbers in bold represent heterozygous loci

Among them, the CEDG008 locus showed the highest allelic variation, with six alleles being recognized among the 20 individuals. The genetic relationship of individuals based on the pairwise D_A distance partially reflects the spatial distribution of subpopulations (Fig. 3). This population comprises four groups of individuals with each group corresponding to four areas: northwest (1) and southwest (4) above Route 461, and northeast (2) and southeast (3) below Route 461 (Fig. 2b). The genotypes of individuals that are distributed in the same area tend to be similar. The highest number of individuals sharing the same genotype were all those in the northeast (2) of the population (Table 4). Individuals in the northwest (1) and northeast (2) had common alleles or similar-sized alleles, but all these alleles were uncommon or absent in the other two areas. One individual, B05, in the southwest (4) was heterozygous for almost all loci. The genotype of this individual can be explained by an allelic combination of the adjacent individual, B04, and individual B13 in the southeast (3) of the population. Heterozygosity was found in individuals B07 at the CEDG015 locus and B14 at the CEDG024 locus in the southeast (3). In contrast to individuals in the north of the population, those in the southeast (3) possessed more alleles than other groups. The genotypes of those individuals consist of a combination of two or three alleles specific to this area.

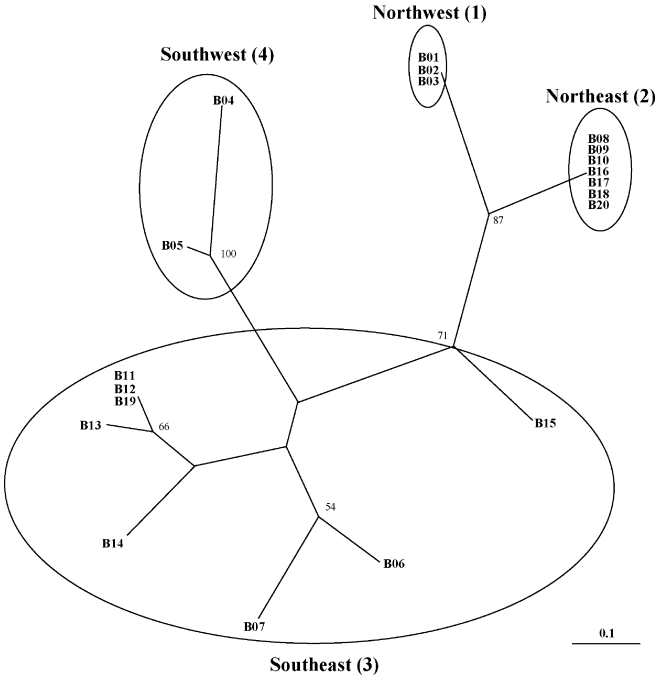

Fig. 3 An unrooted phenogram based on D_A distance showing the genetic relationship between individuals of the Bato town population. The *number* at the node is a bootstrap value higher than 50%

Comparison of allelic diversity in the complex population at Bato town with cultigens

The allele sizes of individuals from the Bato town population were compared with alleles from cultivated azuki, 15 landraces and two released varieties that were collected in the same prefecture as Bato town. There was no nucleotide substitution, insertion or deletion in the sequences flanking the $(AG)_n$ motif for all alleles in both the Bato town population and cultigens. The sequences appear as accession numbers AB128066–AB128148 in DDBJ and Genbank. The differences in allele sizes were due to the number of AG repeat. Of the loci whose alleles showed a greater range with respect to size in plants of the

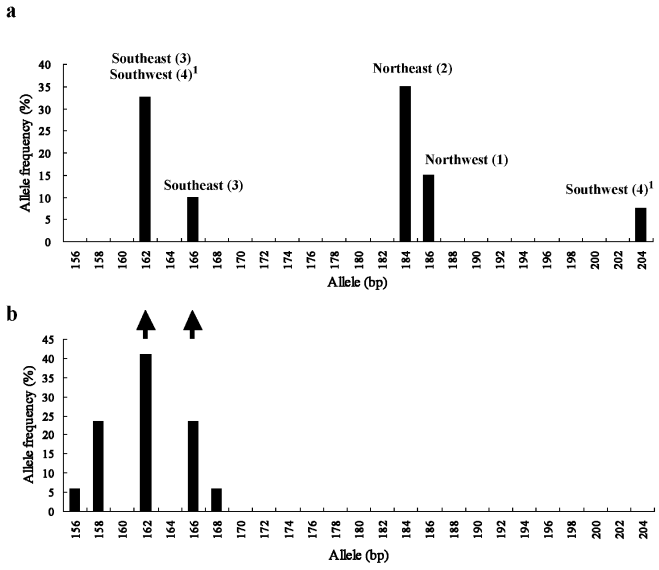

Fig. 4 Proportion of different alleles for microsatellite locus (CEDG026) in: **a** 20 individuals from the Bato town complex azuki population, **b** 17 cultivars of azuki bean. *Superscript 1* indicates one individual was heterozygous with the 162-bp and 204-bp alleles. An *arrow* indicates alleles of the same size found in the complex azuki population and cultivated azuki bean

Bato town population than in the cultigens, alleles 162 bp and 166 bp at the CEDG026 locus (Fig. 4), alleles 119 bp and 121 bp at the CEDG007 locus, alleles 109 bp, 113 bp and 115 bp at the CEDG008 locus and allele 159 bp at the CEDG043 locus were found to be same as alleles from the cultigens (Table 4). All of these alleles having the same size as those in the cultigens were found in plants in the southeast (3) and one plant in the southwest (4) of the population. No alleles of sizes similar to those in cultivated azuki were found in individuals from the other two northern groups.

Discussion

SSR library construction

Construction of a genomic library or SSR-enriched library has been the principal means of discovering SSRs in eukaryotic genomes for which public DNA sequence data is lacking. The DNA sequence information available in Genbank (May 23, 2003) for azuki bean (*Vigna angularis*) consists of only 77 sequences. Therefore, we constructed an SSR-enriched library in azuki bean in order to efficiently and comprehensively track down SSRs. In the present study the enrichment method for large-scale discovery of SSRs was used. To our knowledge, this is the first report of SSR enrichment based on oligo-primed second-strand synthesis in plants. In broad outlines the construction of the SSR-enriched library consists of four steps: (1) construction of a genomic library, (2) production of a ssDNA pool from the genomic library, (3) complementary second-strand synthesis on the ssDNA pool using the SSR motif primer and (4) enzymatic elimination of ssDNA from the synthesized dsDNA pool.

The genomic library was constructed using a method for shotgun sequencing (Birren et al. 1997), and the average size of the resulting insert fragments was 500 bp. The range of insert size—between 400 bp and 700 bp—was convenient for obtaining the insert sequence to enable reading from one end. Furthermore, the strict size-fractionation using a gel decreased contamination with small or jointed fragments through blunt-end ligation in the genomic library. This procedure resulted in a high-percentage success (98%) of single-locus amplification.

The large differential distribution of mono-, di-, tri-, and tetranucleotide repeats between human and *Arabidopsis* has been clarified (Cardle et al. 2000; Katti et al. 2001). With respect to the frequency of dinucleotide repeats in plants, there are many more AT/TA motif loci than AC/CA, AG/GA and CG/GC motif loci. In the present study, the frequencies of AC/CA and AG/GA dinucleotide repeats were examined by means of colony hybridization of the non-enriched azuki bean genomic library. In the azuki bean genome, the $(AG)_n$ motifs were found to be more abundant than $(AC)_n$ motifs, which is in agreement with the general finding in most plants (Powell et al. 1996; McCouch et al. 1997; Cardle et al. 2000). If we assume that the frequency of the motifs obtained from the non-enriched library with an average insert size of 500 bp is correct, the number of $(AG)_n$ and $(AC)_n$ motif loci per 1 Mbp is estimated as 6.51 and 3.91, respectively. In comparison with *Arabidopsis* (Katti et al. 2001), this suggests that the $(AG)_n$ motif loci in azuki bean might be threefold less frequent and the $(AC)_n$ motifs threefold more frequent. Assuming that the genome size of azuki bean is 539 Mb (Bennett and Leitch 2001), the number of $(AG)_n$ and $(AC)_n$ motif loci per haploid genome can be estimated to be 3,500 and 2,100, respectively, indicating that these motifs are rich marker resources in the azuki bean genome.

The modified method that we adopted resulted in a 116-fold enrichment over the non-enriched genomic library. The positive clones from colony hybridization contained at least one SSR motif (98.6%). Thus, this method has the potential to be applied to the construction of an SSR-enriched library suitable for large-scale sequencing. To completely cover information in the haploid azuki genome, ssDNA from at least five independent genomic libraries are needed since the ssDNA obtained here in the 2.4×10^5 genomic library corresponds to approximately one-fifth of the haploid genome of azuki.

A potential drawback to most enrichment procedures is the amplification step of the library. This step is necessary in order to compensate for the loss of informative sequences during the enrichment process, but it may alter the contents because redundant clones can be produced during this step. Just such redundant clones have been found in enrichment libraries (Rallo et al. 2000; Jones et al. 2001; Kölliker et al. 2001; Squirrell et al. 2003), and this may result in wasted effort during the large-scale

sequencing of clones. In contrast to the PCR amplification step in other enrichment methods, our method amplified the whole genomic library in vivo at the ssDNA preparation step. When ssDNA was prepared following an overnight culture with *Escherichia coli*, a protocol based on the original one of Vieira and Messing (1987), redundant clones were prominent, about 60%. The overnight culture period might be more advantageous for redundant clones that grow at a higher speed. By shortening the culture period (5 h) we were able to reduce the frequency of redundant clones to about 7% in the present study. Further modification of the culture period is expected to lead to a decrease in the number of redundant clones in the SSR-enriched library.

Population analysis

Azuki bean is one of the few crops that constitute a crop complex with its wild form in Japan. The crop complex is composed of three forms—cultivated, wild and weedy. There are many populations that have continuous phenotypic variation between wild and cultivated forms. Such populations are classified as complex populations (Tomooka et al. 2002). Amplified fragment length polymorphism (Xu et al. 2000a) and random amplified polymorphic DNA (RAPD) analyses (Xu et al. 2000b) have shown that a higher level of genetic variation can be found in complex populations than in either wild or weedy populations. In the present study, the average number of alleles in the complex populations (3.0–3.4) was higher than that in wild populations (1.1–1.4). These data strongly suggest that introgression is the cause of higher genetic variation in complex populations, which naturally leads to the assumption that some individuals in complex populations are offspring from natural outcrossing between cultivated and wild forms. Although there is no precise information on the level of outcrossing in azuki, insects such as the carpenter bee (*Xyclocapa appendiculata*) are thought to contribute to temporal gene flow in azuki populations (Tomooka et al. 2002). Based on RAPD analysis, the Bato town population studied was found to possess the same level of intra-population variation as other complex populations analyzed from across Japan (Xu et al. 2000b). While the typical erect phenotype of the cultigens has not been found in plants of the Bato town population, variations in color of the stem, seed coat and pod are conspicuous (Weerasekera et al. 2004). To clarify the origin of such phenotypic variation, the individuals were analyzed using SSR markers developed by the method described here.

The availability of SSR markers enabled us to describe the genetic structure of the population. A total of 33 alleles were detected with eight loci in this small population (50×50 m), which demonstrates the high discriminating power of SSR markers. The Bato town population consists of four groups of individuals (Table 4). The genetic structure differed more between the north and south ends of the population than between the east (lower) to west (upper) ends (Fig. 2). This was unexpected as it seems that genetic variation is not primarily affected by the population being transected by a road. The pattern of variation observed is best explained by gene flow from cultivated azuki bean, which is grown in the area, to wild plants in the southeast and/or southwest of the population (Fig. 4). The most frequent alleles of individuals in the southeast (3) of the Bato population were indistinguishable from those found in local landraces and released varieties. In addition, seed color segregation was observed in individuals growing in that area (3) of the population (Weerasekera et al. 2004). This suggests that individuals in the southeast (3) of the population are offspring from natural outcrossing between cultivated and wild azuki. The segregation of hybrids between wild and cultivated azuki—and the possible subsequent gene flow among the wild plants—would explain both the microsatellite and morphological variation observed.

This is the first report of outcrossing events in natural populations of azuki bean, which has always been considered to be a self-pollinating plant. Our detection of heterozygotes at multiple loci provides direct evidence of outcrossing. The ecological risks of unintentional transgene escape into natural populations from genetically modified plants by gene flow have become a matter of public concern in many countries. Our data suggest that the release of transgenic azuki bean produced for insect resistance (Ishimoto et al. 1996) carries with it the risk of unintentional transgene escape into the wild population at locations where complex or weedy populations are found. In actual fact, the estimation of practical transgene dispersal in predominantly self-pollinating crops for which insects contribute to the gene flow will be difficult in controlled trial fields because interaction between many environmental and biological factors are involved. Furthermore, it is necessary to develop evaluation methods for studying transgene dispersal over many years so that fitness in practical field conditions can be assessed. Natural complex populations, such as the one analyzed here, may represent useful ‘model’ populations to provide practical and long-term transgene risk information.

Acknowledgements We thank Dr. S. Fujita, Hokkaido Prefectural Tokachi Agricultural Experiment Station, for providing us with seed materials. This research was performed while the first author held a Science and Technology Agency (STA) of Japan fellowship. The support from the Global Environment Research Fund of the Japanese Ministry of the Environment to AK is acknowledged.

References

Beebe S, Toro O, Gonzalez, Chacon MI, Debouck DG. (1997) Wild-weed-crop complexes of common bean (*Phaseolus vulgaris* L., Fabaceae) in Andes of Peru and Colombia, and their implications for conservation and breeding. Genet Res Crop Evol 44:73–91

Bennett MD, Leitch IJ (2001) Plant DNA C-values database (release 1.0, September 2001) http://www.rbgkew.org.uk/cval/homepage.html

Birren B, Green ED, Klapholz S, Myers RM, Roskams J (1997) Genome analysis. Cold Spring Harbor Laboratory press, Cold Spring Harbor, pp 424–430

Bonnin I, Ronfort J, Wozniak F, Olivieri I (2001) Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae). Mol Ecol 10:1371–1383

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Draper J, Scott R (1988) The isolation of plant nucleic acids. In: Draper J, Scott R, Armitage P, Walden R (eds) Plant genetic transformation and gene expression. Blackwell Scientific Publ, London, pp 212–214

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp 41:95–98

Ishimoto M, Sato T, Chrispeels MJ, Kitamura K (1996) Bruchid resistance of transgenic azuki bean expressing seed α-amylase inhibitor of common bean. Entomol Exp Appl 79:309–315

Jones ES, Dupal MP, Kölliker R, Drayton MC, Forster JW (2001) Development and characterization of simple sequence repeat (SSR) markers for perennial ryegrass (*Lolium perenne* L.). Theor Appl Genet 102:405–415

Kang HW, Cho YG, Yoon UH, Eun MY (1998) A rapid DNA extraction method for RFLP and PCR analysis from a single dry seed. Plant Mol Biol Rep 16:1–9

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequence. Mol Biol Evol 18:1161–1167

Kölliker R, Jones ES, Drayton MC, Dupal MP, Forster JW (2001) Development and characterization of simple sequence repeat (SSR) markers for white clover (*Trifolium repens* L.). Theor Appl Genet 102:416–424

Konuma A, Tsumura Y, Lee CT, Lee SL, Okuda T (2000) Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. Mol Ecol 9:1843–1852

Langella O (2000) POPULATIONS 1.2.23: Population genetic software. Department of populations, genetics and evolution, CNRS, France

Li YC, Röder MS, Fahima T, Kirzhner VM, Beiles A, Korol AB, Nevo E (2000) Natural selection causing microsatellite divergence in wild emmer wheat at the ecologically variable microsite at Ammiad, Israel. Theor Appl Genet 100:985–999

McCouch SR, Chen X, Panaud O, Temnykh S, Xu Y, Cho YG, Huang N, Ishii T, Blair M (1997) Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant Mol Biol 35:89–99

Morgante, M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. Nat Genet 30:194–200

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170

Ostrander EA, Jong PM, Rine J, Duyk G (1992) Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. Proc Natl Acad Sci USA 89:3419–3423

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Rallo R, Dorado G, Martin A (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europacea* L.). Theor Appl Genet 101:984–989

Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Squirrell, J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plant? Mol Ecol 12:1339–1348

Takahashi H, Nirasawa K, Furukawa T (1996) An efficient method to clone chicken microsatellite repeat sequences. Jpn Poult Sci 33:292–299

Tomooka N, Kaga A, Vaughan DA, Kobayashi N, Yoshida T, Nobori T, Komatsuzaki T, Akiba M, Omizu T, Taguchi T, Pickersgill B (2001) Monitoring and collecting of the azuki bean complex (*Vigna angularis*) gene pool in Tottori, Okayama, Ibaraki and Tochigi prefectures, Japan, 2000. Annual report on exploration and introduction of plant genetic resources. NIAR 17:17–32

Tomooka N, Vaughan DA, Moss H, Maxted N (2002) The Asian *Vigna*. *Vigna* subgenus *Ceratotropis* genetic resources. Kluwer, Dordrecht

Vaughan DA, Tomooka N, Kaga A (2004) Azuki beans. In: Singh RJ, Jauhar PP (eds) Genetic resources, chromosome engineering and crop improvement. 2. Grain legumes. CRC Press, Boca Rotan (in press)

Vieira J, Messing J (1987) Production of single-stranded plasmid DNA. Methods Enzymol 153:3–11

Weber ZL (1990) Informativeness of human $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524–530

Weerasekera D, Wang XW, Zong XX, Han OK, Taguchi T, Tomiyama H, Kaga A, Tomooka N, Vaughan DA (2004) Monitoring and analysis of *Vigna angularis* complex populations at Batou, Tochigi Prefecture. In: Jayasuriya AHM, Vaughan DA (eds) Conservation and use of wild relatives of crops. Proc Joint Dept Agric Sri Lanka and Natl Inst Agrobiol Sci Jpn Workshop. Department of Agriculture, Peradeniya, Sri Lanka (in press)

Xu RQ, Tomooka N, Vaughan DA (2000a) AFLP markers for characterizing the azuki bean complex. Crop Sci 48:808–815

Xu RQ, Tomooka N, Vaughan DA, Doi K (2000b). The *Vigna angularis* complex: genetic variation and relationships revealed by RAPD analysis and their implications for in-situ conservation and domestication. Genet Res Crop Evol 47:123–134

Yamaguchi H, Nikuma Y (1996) Biometric analysis on classification of weed, wild and cultivated azuki beans. Weed Res 41:55–62

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1–16