Abundance and variation of microsatellite DNA sequences in common beans

Kangfu Yu, Soon J. Park and Vaino Poysa
Agriculture and Agri-Food Canada, Greenhouse and Processing Crops Research Center
Harrow, Ontario, Canada N0R 1G0

Introduction: Microsatellites, also known as simple sequence repeats (SSRs), are among the most variable types of tandem repetitive DNA in mammals and plants (Edwards et al. 1991). They are genomic regions in which a single base pair or a small number of base pairs of DNA (usually less than 6 bp) is repeated multiple times (Rongwen et al. 1995). They have been widely used in the analysis of mammalian (Murray et al. 1994) and plant genomes (Akkaya et al. 1995). There are several advantages with the use of microsatellites as genetic markers: (1) they are codominant, (2) they are PCR based, therefore automation of the process for marker generation and analysis is possible, (3) they are multiallelic and hypervariable, (4) they appear to be randomly and uniformly dispersed throughout eukaryotic genomes (Dietrich et al. 1994), and (5) they are accessible to other research laboratories via published primer sequences (Weber 1989). The objectives of this study were: (1) to survey the abundance and variation of bean microsatellite sequences, (2) to evaluate their potential usefulness as genetic markers for genomic mapping and DNA fingerprinting.

Materials and methods: All of the 203 DNA sequences in the Bean Genes database were retrieved from the GenBank and saved on floppy disks. The resultant sequences were screened for all theoretically possible tandem repeated sequences (4 dimeric, 10 trimeric, and 33 tetrameric) using the program DNASIS. The search criteria included all possible di-nucleotide repeats with $n \geq 6$, all possible tri-nucleotide repeats with $n \geq 5$, and all possible tetra-nucleoride repeats with $n \geq 3$. No 5-nucleotide repeats were searched in this study. PCR primers, between 19 and 29 bp in length flanking each of the repeated sequences, were designed using the computer program GENE RUNNER. Primers were selected to have a Tm $\geq$ 46 ^{0}C at a salt concentration of 50 mM with an expected fragment size between 120 and 250 bp. All of the plant materials were grown in a growth chamber. Twelve common bean lines and one F_6 recombinant inbred population were used to evaluate allelic variation and segregation, respectively. Genomic DNA from young leaves was isolated according to the procedure described by Yu and Pauls (1994). PCR products were [alpha-^{33}P] dATP labeled and separated on a DNA sequencing gel.

Results and discussion: By following the search criteria , 47 potential microsatellite sequences were identified. The number of DNA sequences for each bean species in the database and the number of potential microsatellite sequences identified from each of them are summarized in Table 1. Theoretically, there are 47 (4 dimeric, 10 trimeric, and 33 tetrameric) nucleotide repeat motifs that could be identified in any genome, but in reality only a few are present at high frequency in any given organism. In this study, microsatellites with AT/TA repeats are most prevalent, followed by those with CT/AG repeats. The distributions of di-, tri-, and tetra-nucleotide repeats found in this survey are very similar to most previous studies, which are: (1) all of the di-nucleotide repeats were only found in non-coding regions of a gene such as, in upstream, downstream or intron regions, (2) most of the tri-nucleotide repeats were found in coding regions, and (3) most of the tetra-nucleotide repeats were primarily found in non-coding regions of a gene as well. PCR analysis of 14 of the microsatellite-containing loci revealed that 12 of the 14 primer pairs could produce fragments, or

groups of fragments, which were readily scored. Allelic variation of the 12 loci was surveyed with 12 bean inbred lines representing a diversity of germplasms. Seven of the 12 microsatellite loci were polymorphic and yielded 2-10 alleles. Segregation analysis of the polymorphic loci with a common bean F_6 recombinant inbred population confirmed their Mendelian segregation.

Table 1. Number of DNA sequences from different bean species in the GenBank database and occurrence of microsatellite sequences.

Bean species in database	**Number of entries**	**Total sequences searched (kb)**	**Total SSR sequences**	**% of SSR over entries**	**occurrence of 1 SSR in kb**
Phaseolus	**142**	**202.3**	**40**	**28.2**	**5.6**
vulgaris	127	186.6	40	31.5	4.7
lunatus	6	4.9	-	-	-
acutifolius	5	5.0	-	-	-
coccineus	3	5.0	-	-	-
macalatus	1	0.8	-	-	-
Vigna	**61**	**71.7**	**7**	**11.5**	**10.2**
radiata	29	35.8	5	17.2	7.2
unguiculata	15	9.9	1	6.7	9.9
aconitifolia	9	12.2	1	11.1	12.2
angularis	4	3.8	-	-	-
mungo	4	10.0	-	-	-
Total	**203**	**274**	**47**	**23.2**	**5.8**

References

Akkaya, M.S. et al. 1995. Crop Sci. 35:1439-1445. Dietrich, W.F. et al. 1994. Nature Genet. 7:220-225. Edwards, A. et al. 1991. Am. J. Hum. Genet. 49:746-756. Murray, J.C. et al. 1994. Science, 265:2049-2054. Rongwen, J. et al. 1995. Theor. Appl. Genet. 90:43-48. Weber, J.K. and May, P.E. 1989. Am. J. Hum. Genet. 44:388-397. Yu, K. and Pauls, K.P. 1994. *In* PCR technology: current innovations. *Edited by* H.G. Griffin and A.M. Griffin. CRC press, Boca Raton, Florida. pp. 193-200.