DNA Fingerprinting of Soybean(*Glycine max)* using Microsatellite Markers

Aarti Kumari, Dr. M. K. Rana

Division of Genomic Resources, National Bureau of Plant Genetic Resources (NBPGR), Pusa Campus, New Delhi-110012, India

Abstract- Soybean[Glycine max (L.) Merr], a self pollinating diploid (2n = 40), is one of the valuable and the oldest legume crops grown widely throughtout India. Globally India ranks fifth with an annual production of 6.8 million tonnes from 6.0 million hectare. The molecular characterization was carried out in 87 traditional soybean cultivars of India using Microsatellite marker system. Out of 48 SSR(Simple sequence repeat) primers screened, 11 SSR marker gave very good reproducible banding patterns. A total of 51 reproducible bands/alleles were amplified. The number of alleles detected per locus ranged from 2 to 7 with an average of 4.63 alleles per primer pair. The size of amplified product ranged from 80 bp to 300 bp. A dendrogram was generated by UPGMA(Unweighted pair group method with Arithmetic means) Cluster analysis based on Jaccard's Similarity Coefficients. The dendrogram broadly classified these 87 soybean cultivars into seven major clusters; I, II, III, IV, V, VI and VII. And sample 40 and 23 are uniquely defined. Further efforts are continuing to analyze more number of indigenous soybean cultivars with more number of microsatellite primers to study the nature and extent of genetic variability and interrelationship between the cultivars.

Keywords: Soybean[Glycine max (L.) Merr], DNA profiling, genetic diversity, Microsatellite or SSR markers and PCR analysis.

I. INTRODUCTION

Soybean[*Glycine max* (L.) Merr], a self pollinating diploid (2n = 40), is one of the valuable and the oldest legume crops grown widely throughtout India. Globally India ranks fifth with an annual production of 6.8 million tonnes from 6.0 million hectare. Madhya Pradesh, Maharashtra, Rajasthan, Karnataka and Uttar Pradesh are the major soybean growing states in India (Ali, 2004).

Taxonomically, soybean belongs to the order Fabales, the family Fabaceae, the subfamily Faboidae and the genus *Glycine*. The genus *Glycine* is divided into two subgenera, *Glycine* and *Soja*. The subgenus *Glycine* includes 16 perennial species and the subgenus *Soja* (*Moench*) includes the cultivated soybean, *G. max* (L.) Merril (2n=40) and the two annual species, *Glycine soja Siebold* and *Zucc*. The cultivated species, *G. max*, hybridizes easily with its wild annual relative and most probable progenitor *G. soja* while with the perennial relatives it has low crossability rate (Singh and Hymowitz, 1999).

Soybean is known as the "GOLDEN BEAN" of the 21th Century. The plant is classed as an oilseed rather than a pulse by the UN Food and Agricultural Organization (FAO). Due to very poor cookability on account of inherent presence of trypsin inhibitor, it cannot be utilized as a pulse. It is now the second largest oilseed in India after groundnut. It grows in varied agro-climatic conditions. It has emerged as one of the important commercial crop in many countries. Due to its worldwide popularity, the international trade of soybean is spread globally. Soybean has great potential as an exceptionally nutritive and very rich protein food. It can supply the much needed protein to human diets, because it contains above 40 percent protein of superior quality and all the essential amino acids particularly glycine, tryptophan and lysine, similar to cow's milk and animal proteins. Soybean also contains about tweenty percent oil with an important fatty acid, lecithin and Vitamin A and D. The four percent mineral salts of soybeans are fairly rich in phosphorous and calcium.

Table 1: Nutritional value of edible soybean

COMPONENTS	PERCENTAGE
Proteins	40
Carbohydrates	30
Fibre	05
Lecithins	0.5
Saponins	04
Oil	18-20

Soyabean contain symbiotic bacteria called *Rhizobium* within nodules of their root systems. These bacteria have the special ability of fixing atmospheric nitrogen, molecular nitrogen (N_2) into ammonia (NH_3), thus reducing the demand for nitrogenous fertilizers.

Microsatellites are tandemly repeated nucleotide units of 1 to 6 base pairs and alleles usually differ in the number of repeated units. Among PCR based markers, microsatellite markers are highly polymorphic, more reproducible, co-dominant and well distributed throughout the soybean genome. A random set of these mapped markers providing genome-wide coverage should facilitate an unbiased assay of genetic diversity and thus giving a robust, unambiguous molecular description of soybean cultivars.

II. MATERIALS AND METHODS

Eighty seven soybean *(Glycine max)* cultivars collected from India (National Gene Bank, NBPGR) were used in the present study. Seeds were grown on germination paper under aseptic condition.Total genomic DNA was isolated from young leaf tissue of 3 weeks old seedlings.

Genomic DNA isolation and PCR optimization and amplification

Total genomic DNA was extracted from 3 weeks old seedlings using cetyl trimethyl ammonium bromide (CTAB) DNA extraction method of Saghai-Maroof (1984) with some modifications.

Eighty seven soybean samples (~5 g each) collected were quickly frozen in liquid nitrogen (-196 ^{0}C) and ground to a fine powder with mortar and pestle. The frozen powder was transferred to 50ml centrifuge tubes with 20ml CTAB buffer maintained at 60^{0}C in water bath and DNA was isolated using CTAB method.The purified DNA was quantified by Biowave DNA Flourimeter and adjusted the final concentration of the template to 25-30ng/µl for use in PCR analysis.

Primer screening

Already available 48 SSR primers were used for screening in soybean. The primers were screened on the basis of reproducible and scoreable amplification for the analysis of all the eighty seven genotypes of soyabean. For screening six genotypes of soybean were used for SSR analysis. These fourty eight primers were screened and the best 11 primers were selected for DNA profiling.

Polymerase chain reaction (PCR) amplifications were carried out in a reaction volume of 20µl containing; 3µl of 30ng/µl of template DNA with reaction mixture consisted of 2µl of 10X PCR buffer, 2.4µl of 3mM $MgCl_2$, 0.2µl of 0.25mM $dNTP_s$, 1.6µl of 0.4µM of each primer, 0.2µl of 5Unit/µl of *Taq* DNA polymerase and then adjusting the volume with sterilize deioniozed water(ddH_2O). All the PCR reaction components were obtained from Sigma.The amplifications were performed using a Biometra T Professional Thermo Cycler for 40 cycles. All the SSR amplified products were resolved by electrophoresis using 4% MetaPhor Agarose gel for 2.5h in 1X TAE buffer (pH 8.0) at 100V, stained with EtBr and photographed using the Gel Documentation System (SYNGENE).

Data Scoring and Analysis

The resolved DNA bands were documented and processed for data aanalysis. The clear, unambiguous and reproducible bands present across the DNA samples were scored as '1' and absence of bands were recorded as '0' for each primer allele cultivar combination. Homology of bands based on the distance of migration of amplified DNA fragments according to their molecular weights in the gel was determined. Molecular weight of the bands were estimated by using 50, 100bp DNA ladder of HIMEDIA Gene RulerTm as standard. The genetic relationship among the 87 soybean cultivars was analyzed from the combined 0/1 matrix data of SSR profiles using the software program NTSYSpc version 1.70 (Exeter software, New York) (Rohlf,1993) and the Jaccard's similarity coefficient and unweighted pair group method using arithmetic average(UPGMA). The resultant matrix data was also analyzed for principal component analysis(PCA) using the same software and a dendrogram based on UPGMA was generated. The dendrogram obtained was then used for cluster analysis. Most informative primers were selected based on their extent of polymorphism. The data obtained by scoring the SSR profiles with different primers individually as well as collectively were subjected to the construction of similarity matrix Jaccard's coefficients. The similarity values were used for cluster analysis. These computations were performed using the same software.

III. RESULTS AND DISCUSSION

Total Genomic DNA Extraction

The Cetyl trimethyl ammonium bromide (CTAB) DNA extraction method of Saghai-Maroof (1984) was used with some modification and yielded sufficient quantity of good quality DNA which was amenable to PCR amplification.

Primer Screening and Profiling

Six soybean varieties were used for screening **(Fig. 1.1a, 1.1b)** in soybean using 48 SSR primer pairs in order to identify the suitable ones for diversity analysis. Out of 48 SSR primer pairs, only 14 were able to amplify DNA in soybean. Eleven of these 14 were chosen for PCR amplification to study further genetic diversity among 87 soybean genotypes **(Fig. 2,3, 4).**

Fig. 1.1a Primer screening using soybean SSR primers in 6 soybean genotypes

M 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6

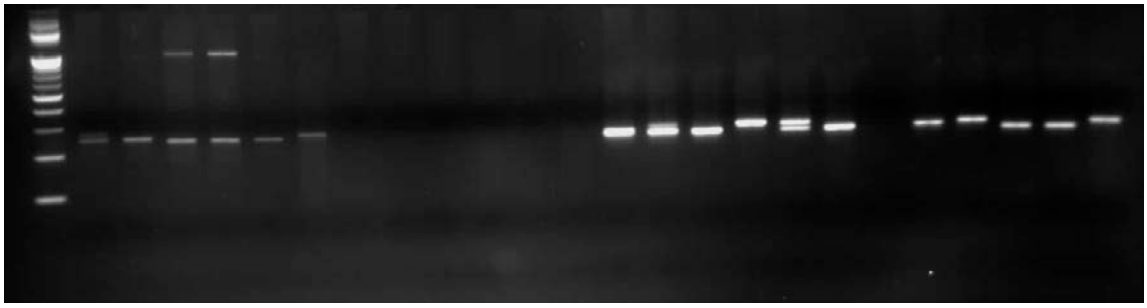
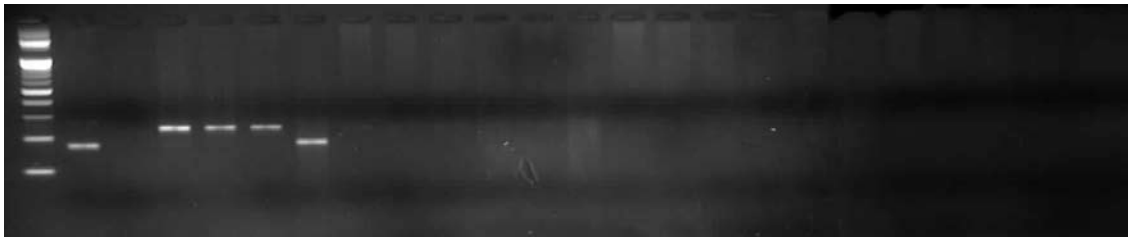


Fig. 1.1b Primer screening using soybean SSR primers in 6 soybean genotypes

M 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 1 2 3 4 5 6

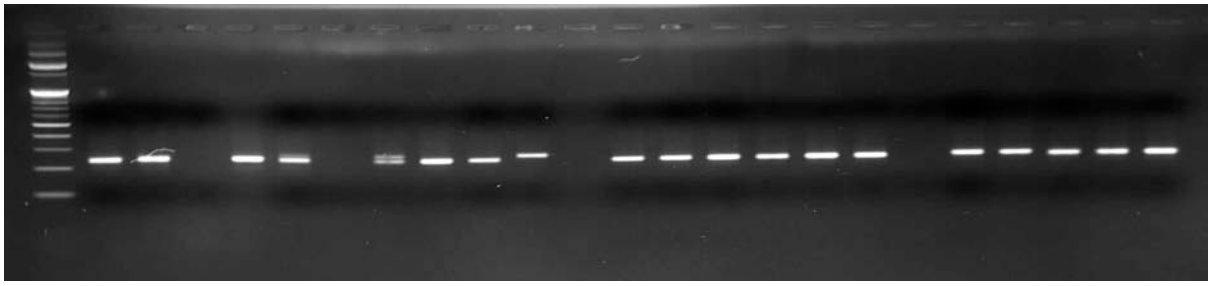


M = Molecular weight marker

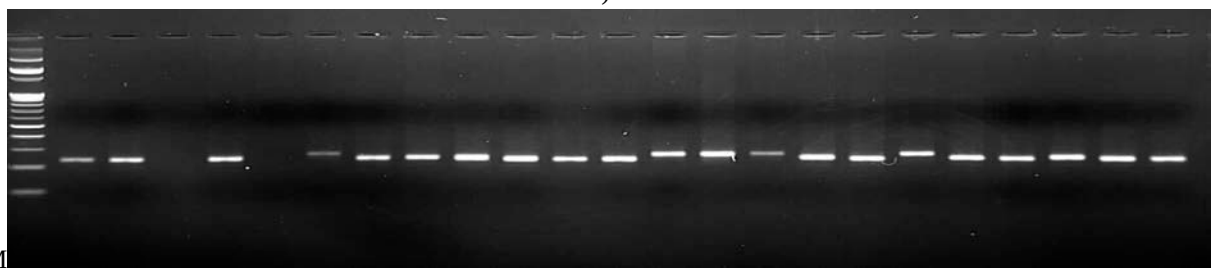
Numbers above (1 to 6) represent different samples used for primer screening

Fig. 2 Microsatellite profile of 87 soybean genotypes employing primer GLY SATT 300

M

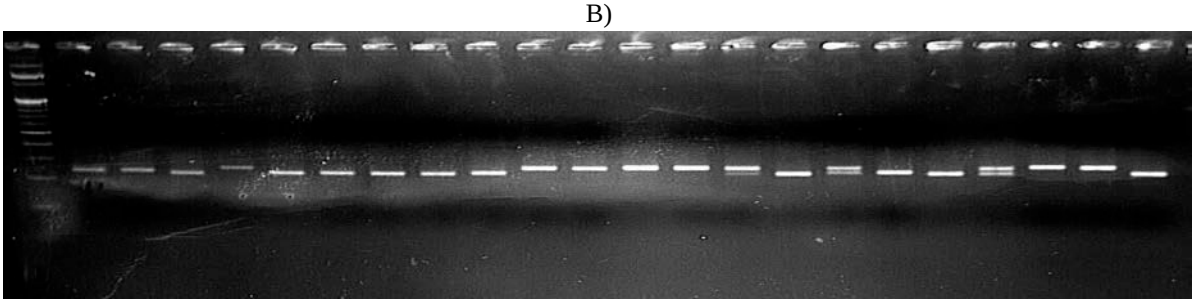


A)



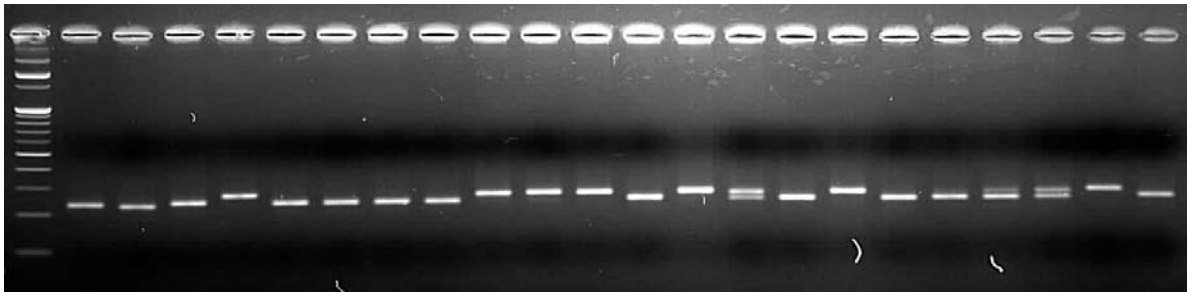
M

B)



M C)

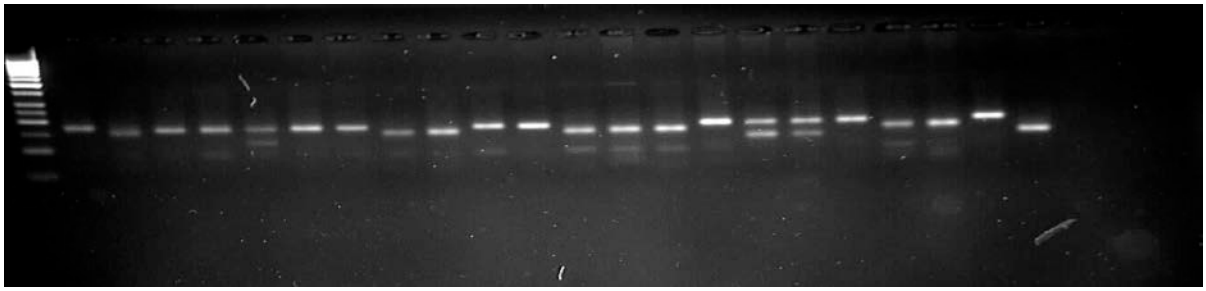
M



D)

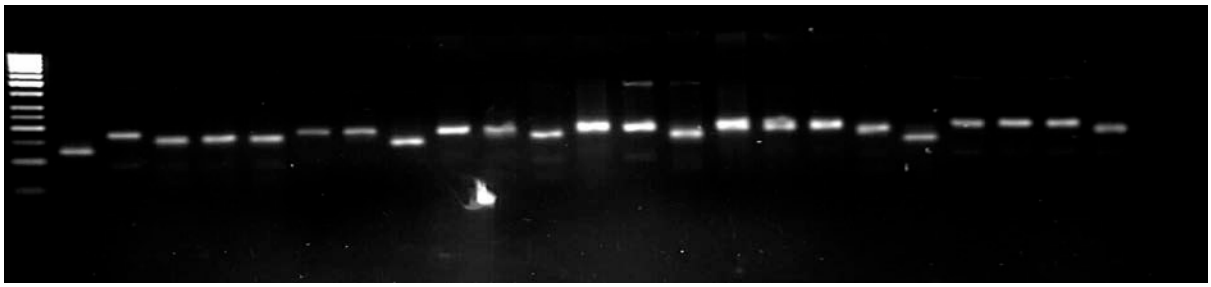
Fig. 3 Microsatellite profile of 87 soybean genotypes employing primer GLY SATT 588

M



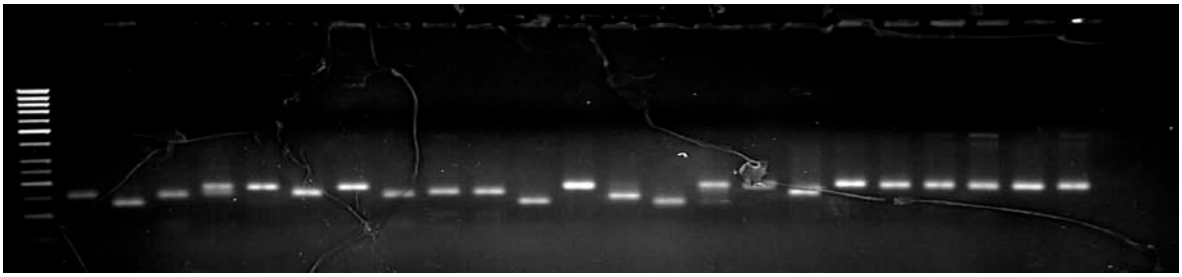
A)

M



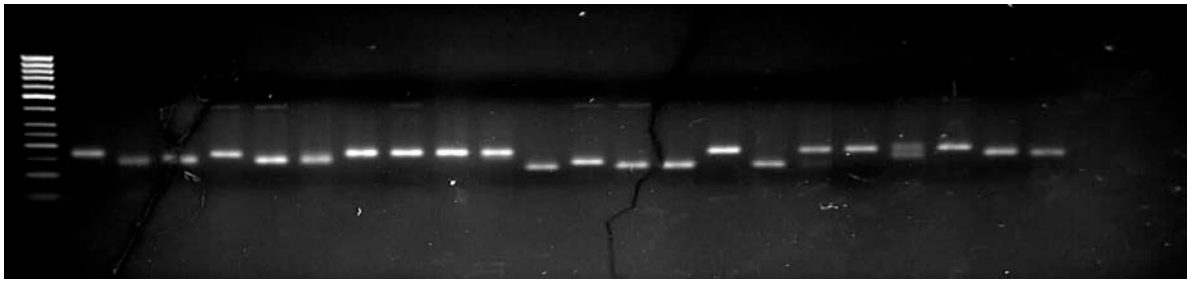
B)

M



C)

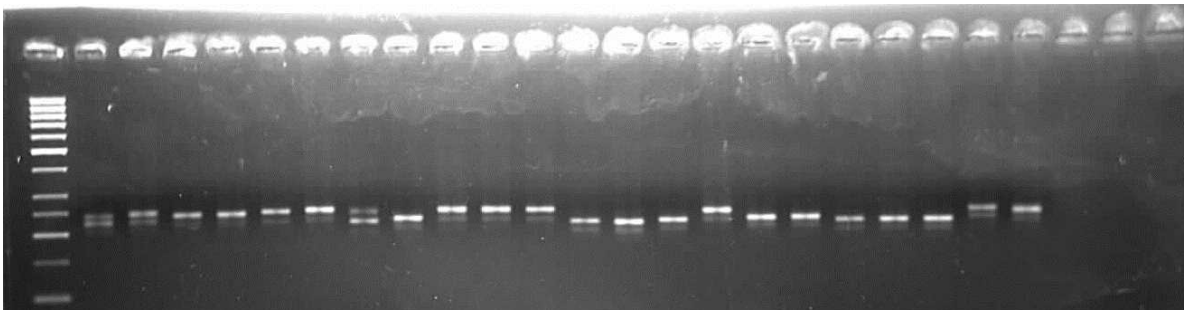
M



D)

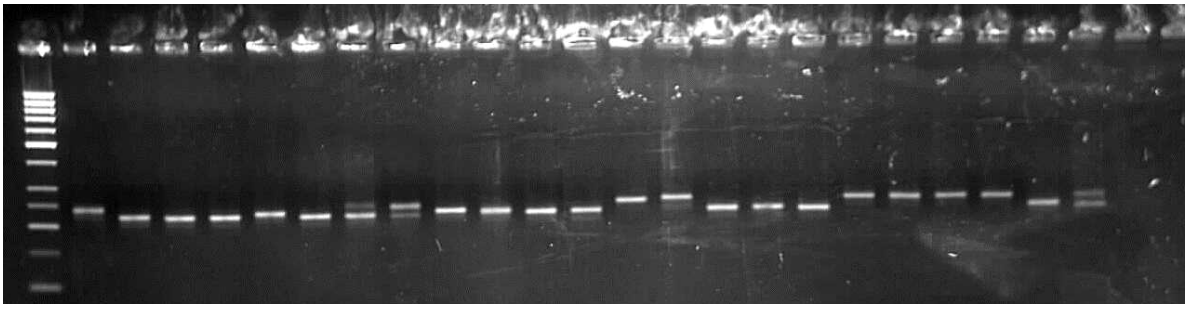
Fig. 4 Microsatellite profile of 87 soybean genotypes employing primer GLY SATT 530

M

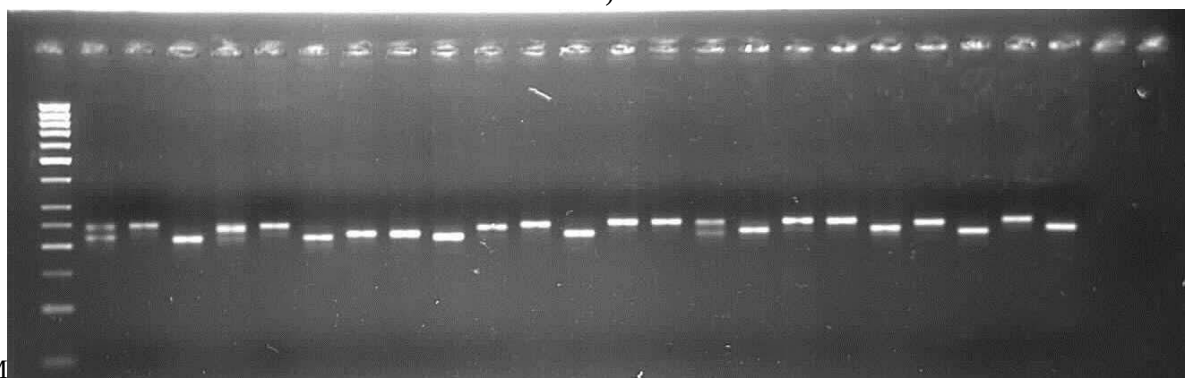


A)

M

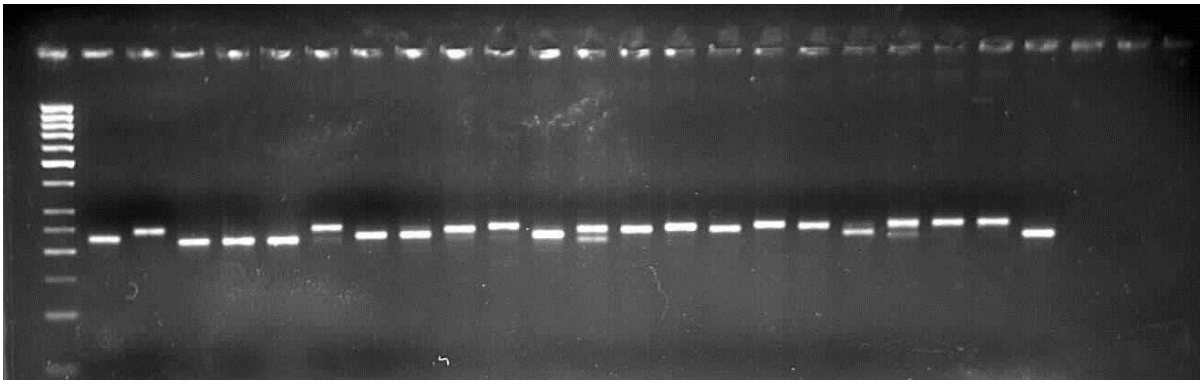


B)



M C)

M



D)

M = Molecular weight marker (100 bp DNA ladder)

A) SOYBEAN CULTIVAR 1 to 22

B) SOYBEAN CULTIVAR 22 to 44

C) SOYBEAN CULTIVAR 44 to 66

D) SOYBEAN CULTIVAR 66 to 87

Band Statistics (DNA fingerprint analysis)

Eleven microsatellite primer pairs, at least one from each chromosome were used for molecular marker analysis of 87 soybean cultivars. A total of fifty one different reproducible bands were amplified all being polymorphic. The number of bands per primer ranged from 2 to 7 with an average of 4.63 bands per primer. The size of amplified products ranged from 80 bp (GLY SATT 114) to 300 bp (GLY SATT 277). Maximum numbers of polymorphic bands (7) were obtained with the primers GLY SATT 530, GLY SATT 114 and GLY SATT 277 **(Table 2)**.

Genetic Diversity and Relationship

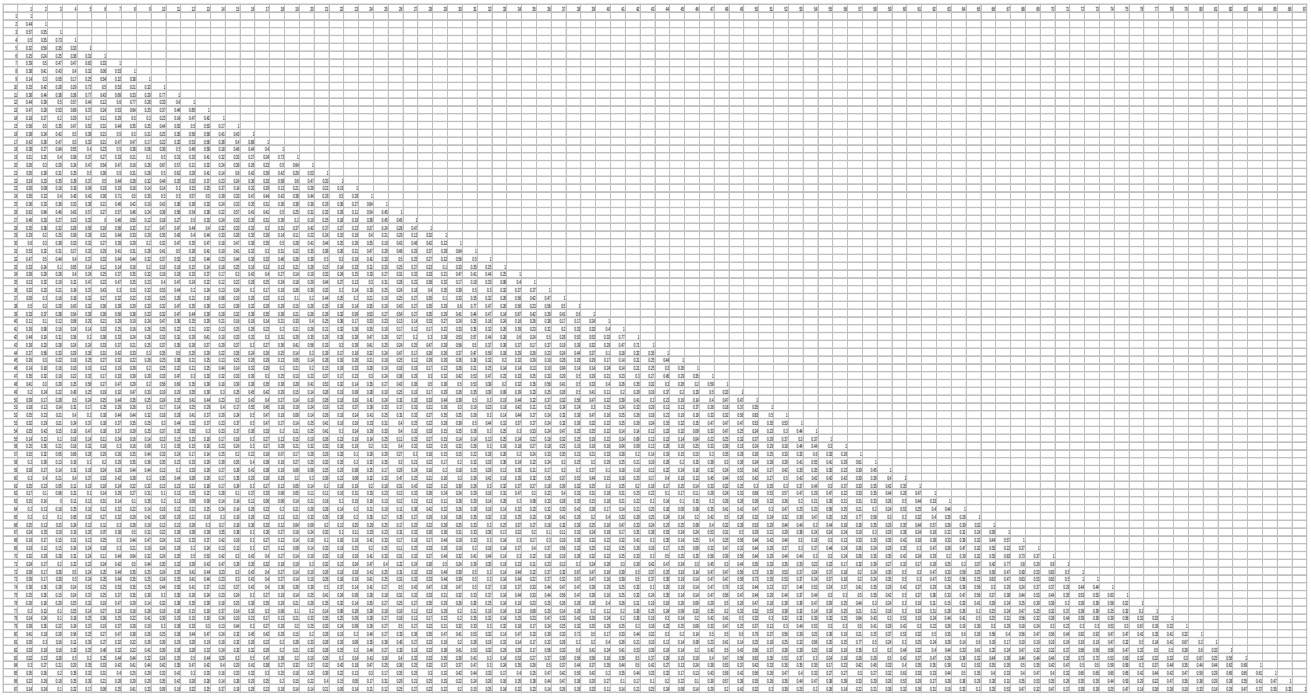
Genetic relationships between soybean genotypes were determined on the basis of Jaccard's pairwise similarity coefficients. The similarity coefficient values obtained are presented in **Table 3**.

Similarity index of pair-wise comparisons estimated on the basis of all the 11 primers ranged from 0.00 to 0.92. The cultivars Pratap Soya (RAUS 5) and MAUS 81 (Shakti) showed highest similarity with JS 95-60 (i.e., 0.92) while 9560 showed least similarity with MACS-450 (i.e., 0.00).

Table 2: Information about SSR primers used and level of polymorphism detected in soybean

Sr. No.	Primer Name	No. of Alleles	Size Range (bp)	No. Of polymorphic allele
1.	GLY SATT (F+R) 267	2	230-240	7
2.	GLY SATT (F+R) 300	2	240-250	6
3.	GLY SATT (F+R) 449	3	250-300	2
4.	GLY SATT (F+R) 277	6	140-300	2
5.	GLY SATT (F+R) 588	6	90-150	3
6.	GLY SATT (F+R) 530	7	200-260	4
7.	GLY SATT (F+R) 114	5	80-120	4
8.	GLY SATT (F+R) 279	3	80-100	7
9.	GLY SATT (F+R) 197	6	110-200	3
10.	GLY SATT (F+R)586	4	190-220	6
11.	GLY SATT (F+R)577	4	90-120	7

Table 3: Jaccard's similarity estimates between cultivar pairs in soybean



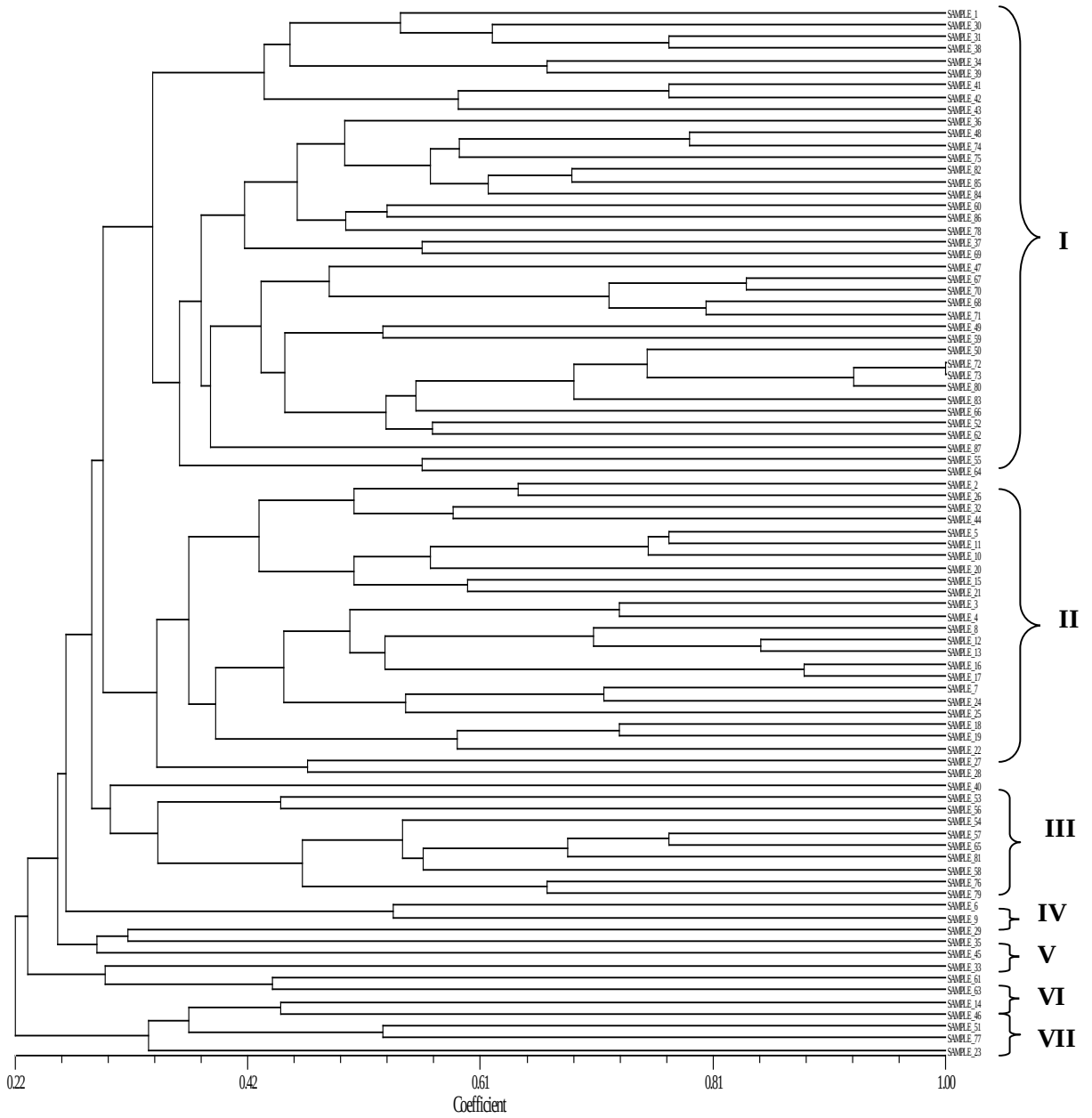


Fig. 5: UPGMA cluster analysis (i.e. dendrogram) of soybean genotypes based on Jaccard's Similarity Coefficient. Major clusters are indicated in the right margin.

Cluster Analysis

A dendrogram was generated by UPGMA cluster analysis based on Jaccard's similarity coefficients **(Fig. 5)**.Cluster analysis based on UPGMA using the program NTSYSpc version 1.70 (Exeter software, New York) (Rohlf,1993) provided a clear resolution of relationship among all the 87 soybean cultivars. The dendrogram broadly classified these soybean cultivars into seven major groups, I, II, III, IV, V, VI and VII.In cluster I there are 2 sub-clusters, sub-cluster IA consist of 9 cultivars (sample 1, 30, 31, 38, 34, 39, 41, 42 and 43) with similarity coefficient of 0.43 and sub-cluster IB consist of 30 cultivars (sample 36, 48, 74, 75, 82, 85, 84, 60, 86, 78, 37, 69, 47, 67, 70, 68, 71, 49, 59, 50, 72, 73, 80, 83, 66, 52, 62, 87, 55 and 64) at similarity coefficient of 0.38.Cluster II also have 3 sub-clusters, sub-cluster IIA consist of 10 cultivars (sample 2, 26, 32, 44, 5, 11, 10, 20, 15 and 21) with similarity coefficient of 0.43, sub-cluster IIB consists of 13 cultivars (sample 3, 4, 8, 12, 13, 16, 17, 7, 24, 25, 18, 19 and 22) with similarity coefficient of 0.39 and sub-cluster IIC consist of 2 cultivars (sample 27 and 28) with similarity coefficient of 0.48. Cluster III have 2 sub-clusters, sub-cluster IIIA consists of 2 cultivars (sample 53 and 56) with similarity coefficient of 0.44 and sub-cluster IIIB consists of 7 cultivars (sample 54, 57, 65, 81, 58, 76 and 79) similarity coefficient of 0.47. In cluster IV there are 2 cultivars (sample 6 and 9) with similarity coefficient of 0.55. In cluster V there are 3 cultivars (sample 29, 35 and 45) with

similarity coefficient of 0.28. Similarly in cluster VI there are 3 cultivar (sample 33, 61 and 63) with similarity coefficient of 0.29 and in cluster VII there are 5 cultivars (sample 14, 46, 51, 77 and 23) with similarity coefficient of 0.33.

Cluster I and II have been identified at similarity coefficient of 0.34. Cluster II and III have been identified at similarity coefficient of 0.30. Cluster III and IV have been identified at similarity coefficient of 0.29. Cluster IV and V have been identified at similarity coefficient of 0.27. Cluster V and IV have been identified at similarity coefficient of 0.25. Cluster IV and V have been identified at similarity coefficient of 0.23. Sample 40 and 23 are uniquely defined. Average Jaccard's similarity was calculated in all soybean cultivars and it was found to be 4.63 %.

All the soybean cultivars used in the present study could be distinguished from each other precisely from each other at the level of 51 polymorphic bands between individuals in pair wise comparison over all the 11 microsatellite primers. In particular, the combination of all the polymorphic alleles obtained with all the 11 primers enabled development of DNA fingerprints of the soybean cultivars, which was diagrammatically represented in **Fig. 5.**

The advent of molecular marker technology facilities estimation of the genetic diversity and determine cultivars identity. Characterization of cultivars using DNA profiling techniques like RAPD and hyperpolymorphic microsatellite markers has been used successfully in several crop species (Fukaoka et al., 1992; Virk *et al*., 1995a; Choudhury *et al*., 2001). Among the several molecular techniques available for the detection of the genetic variability, microsatellite markers detect a high degree of polymorphism due to variation in the number of repeated units. Microsatellite marker system is considered as one of the best molecular markers for soybean because of its co-dominant nature, simple and reproducible banding pattern, high level of polymorphism compared to other systems and also the availability of thousands of such markers on the soybean genome.

DNA profiling was done using eleven microsatellite primer pairs distributed on all the forty chromosomes of soybean genome. We found efficient enough to reveal usable level of DNA polymorphism among chosen genotypes. The number of bands detected by microsatellite primers varied from 2 to 7 with an average of 4.63 bands per primer.

With the availability of ultra dense genetic and molecular linkage maps in soybean, it is now possible to choose locus specific, highly polymorphic and co-dominant markers like Microsatellite to carry out diversity and fingerprinting studies in a more structured way. This is not only studying relationship between the soybean genotypes but the whole soybean genome in perspective.

Genetic diversity analysis of a large number of soybean collections employing DNA profiling would a great deal of effort, time and cost. On the basis of the observations in the present study, it is suggested that microsatellite analysis can be efficiently utilized for this purpose. Moreover, since the markers were chosen from open reading frames of soybean DNA, the levels of diversity exhibited by them are likely to be unbiased and not due to chance. Such specific markers would be of great value to serve DNA fingerprints for the characterization of the genetic resources of soybean for promising traits. Informative microsatellite markers are cost effective and useful in diversity analysis.

However, the present study is still at its infancy. A lot of primers from different chromosomes are required to analyze even much larger germplasm to get a reproducible data that can be used as ready references by soybean breeders, variety registration authority, private agencies, etc. The investigation demonstrated the potential use of SSR analysis for assessment of varietal identification, nature and magnitude of variability, interrelationship, and conservation and for maintaining the distinctiveness of valuable germplasms.

IV. SUMMARY AND CONCLUSIONS

Soybeans have occupied a prime position in Indian society not only because of their high protein contents, but also due to its flavour, palatability and aesthetic value. It has been estimated that more than 87 cultivars of indigenous soybean are grown in various parts of India. Relatively few works on the genetic diversity have been done on a molecular level. Although a wide array of molecular markers are available in soybean, SSR markers are of choice because of its co-dominant, multi-allelic and highly polymorphic nature. An attempt was made to find out the nature and extent of genetic divergence among these cultivars.

In the present study, eleven microsatellite primer pairs were used for evaluating the genetic diversity of 87 traditional soybean cultivars of India. A total of 51 reproducible bands/alleles were amplified. The number of alleles detected per locus ranged from 2 to 7 with an average of 4.63 alleles per primer pair. The size of amplified product ranged from 80 bp to 300 bp.Cluster analysis grouped all the 87 soybean cultivars into seven major groups, I, II, III, IV, V, VI and VII.

Further efforts are continuing to analyze more number of indigenous soybean cultivars with more number of microsatellite primers to study the nature and extent of

genetic variability and interrelationship between the cultivars.

V. ACKNOWLEDGEMENTS

Author gratefully acknowledge to Dr. M.K. Rana, (Senior Scientist), Division of Genomic Resources, National Bureau of Plant Genetic Resources (NBPGR) for his exemplary guidance, monitoring and constant encouragement and Dr. Amit Kumar Singh (Junior Scientist), Dr. Sonika Tiwari (R.A.) and Mr. Deep Singh Rawat and also Mr. Shiv Pratap Chaoudhary, Miss Shipra Deo for the valuable information provided by them for completion of this project. I am grateful for their cooperation during the period of my assignment.

REFERENCES

[1] Ali, N. 2004. Mechanisation of soybean production and post production operations in India. Indian Farm. p:9-15.

[2] Rohlf F.J.(1993). NT SYS-PC. Numerical Taxonomy and multivariate analysis system .ver.1.70. Exeter Publ.Ltd., Setauket, New York.

[3] Singh, R. J and T. Hymowitz. 1999. Soybean genetic resources and crop improvement. Genome, 42: 605-616.

[4] Welsh J., M. McClelland,1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acid Res, 18:7213-7218.

[5] Williams J.G.K., A.R. Kubeilk, K.J. Lival, J.A. Rafalski, V.Tingey, 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acid Res* 18:6531-6535.