Genetic Diversity Evaluated with Molecular Markers in International and Indonesian Soybean Germplasm

Werner Schenkel and G. Wenzel

Technical University of Munich, Germany

ABSTRACT

To assess future breeding strategies for the development of soybean cultivars tolerant to acid soil a analysis of genetic diversity in Indonesian and international soybean germplasm was performed. Fourty-five of Indonesian, African, European, US, Brazilian and genebank origin were analysed. Genetic diversity calculations were based on data from molecular marker analysis. In total 25 RFLP, 22 RAPD, and 269 AFLP loci were evaluated. Based on genetic distances calculated from binary data, a multidimensional scaling and hierarchical cluster analysis was performed. Multidimensional scaling demonstrated that the Indonesian genepool is separated from all other accessions. This was confirmed by hierarchical cluster analysis which revealed a structure of six clusters within the investigated germplasm. The clusters could be correlated to the origin of genotypes. All but two Indonesian cultivars build a single cluster that contains no foreign genotypes. The other clusters are comprised by accessions of different origin. Nevertheless clusters clearly correspond to origin, if cultivars from southern USA are pooled together with Brazilian cultivars. These results were confirmed by available pedigree information. Seven ancestors contribute 63% of the genes of Indonesian cultivars released since 1974. Eight genotypes cover 74% of the genome of cultivars of southern USA and contribute strongly to the Brazilian genepool. The same genotypes cover only 16% of the genes of northern USA cultivars. Preliminary data of field trials suggest, that genetic similarity to the Indonesian genepool is positively correlated to field performance under Indonesian climatic conditions. Significance of molecular genetic information could be demonstrated and should be used to plan strategies of future breeding programs.

Key words: Soybean germplasm, molecular marker, cluster

INTRODUCTION

Knowledge on the genetic composition and variability of populations of crops is considered increasingly important for planning of breeding strategies and maintenance of genetic diversity in improved cultivars. Information on genetic diversity between genotypes can be used in several ways for planning of breeding strategies:

1. Speeding up cultivar development by selection of similar parents differing in only one or few traits of interest.
2. Selection of parents with wide genetic diversity to maximise transgressive segregation in descendants.
3. Monitoring and maintenance of genetic variability in stock of improved cultivars to keep up adaptability.

4. Selection of genetic similar genotypes from germplasm with improved chance for adaptation to the production system of interest.
5. Accumulation of QTLs by selection of parents with similar performance but high genetic diversity.
6. Core collections can be defined based on maximised genetic variability.
7. Verification of pedigree information.
8. Selection of parental lines for gene maping to maximise number of polymorphisms.

MATERIALS AND METHODS

The objective of the present investigation was to evaluate genetic diversity by molecular techniques within and between genotypes of differing origin (Table 1). The main attention was turned on Indonesian cultivars and other genotypes adapted to tropical and subtropical climate.

For molecular marker analysis DNA was extracted from young leaves of glasshouse grown plants (Keim *et al.*, 1988). Restriction Fragment Length Polymorphism (RFLP) analysis was performed with probes from the USDA-ARS: RFLP map (Shoemaker and Olson, 1993). Random Amplified Polymorphic DNA (RAPD) analysis was performed with decamer random primers with G C content of 60-80%. Amplified Fragment Length Polymorphism (AFLP) analysis was performed according to Zabeau and Vos (1993) and Vos *et al.* (1995). Electrophoretic separation and detection of fragments was performed on an automated DNA sequencer (ABI Prism 377, Perkin Elmer). In total 25 RFLP, 22 RAPD and 269 AFLP polymorphisms could be identified. Banding patterns were scored in binary modus with codominant markers evaluated individually.

Distance coefficients were calculated form contingency matrices, based on formula of Dice (1945):

$$d = 1\text{-}2N_{11}/(2N_{11} + N_{10} + N_{01})$$

Table 1. Origin of analysed germplasm

Indonesia	Europe	Africa	USA	Genebank	Brazil
Tampomas	Kalmit	Samsoy	Gasoy 17	PI159322	Cristalina
Malabar	Dorado	M351	Essex	PI381674	IAC_9
Tambora	Jutro	TGX536-0	Forrest	PI81042	Savanna
Kerinci	Labrador	TGX1448-2	Norchief	PI417021	Doko_RC
Wilis1	Major		Hawkeye	PI416937	IAC_17
Genjah Jepang	Picador		Maple Arrow		
Sindoro	Goldor		Perry		
Slamet			Kent		
B3577			Missoy		
B3578			Biloxi		
B3911			Jogun		
Wilis2					

The resulting distance matrix was basis for Multi Dimensional Scaling (MDS) procedure and hierarchical cluster analysis. Hierarchical cluster analysis was performed by the "average linkage" method (= UPGMA: unweighted pair group method with arithmetic average) (Backhaus *et al.*, 1996) and "minimal variance" or "Ward's" method.

RESULTS

Genetic distances of cultivars were subjected a MDS analysis for optimal scaling in two dimensions to represent distances coefficients graphical (Figure 1). The two dimensional model resulted in a residual sum of squares (RSQ) of 0.70. This means that 70% of the variance of scaled distances is accounted for by the actual genetic distance. By using a three dimensional model this proportion could be raised to 80%. Further

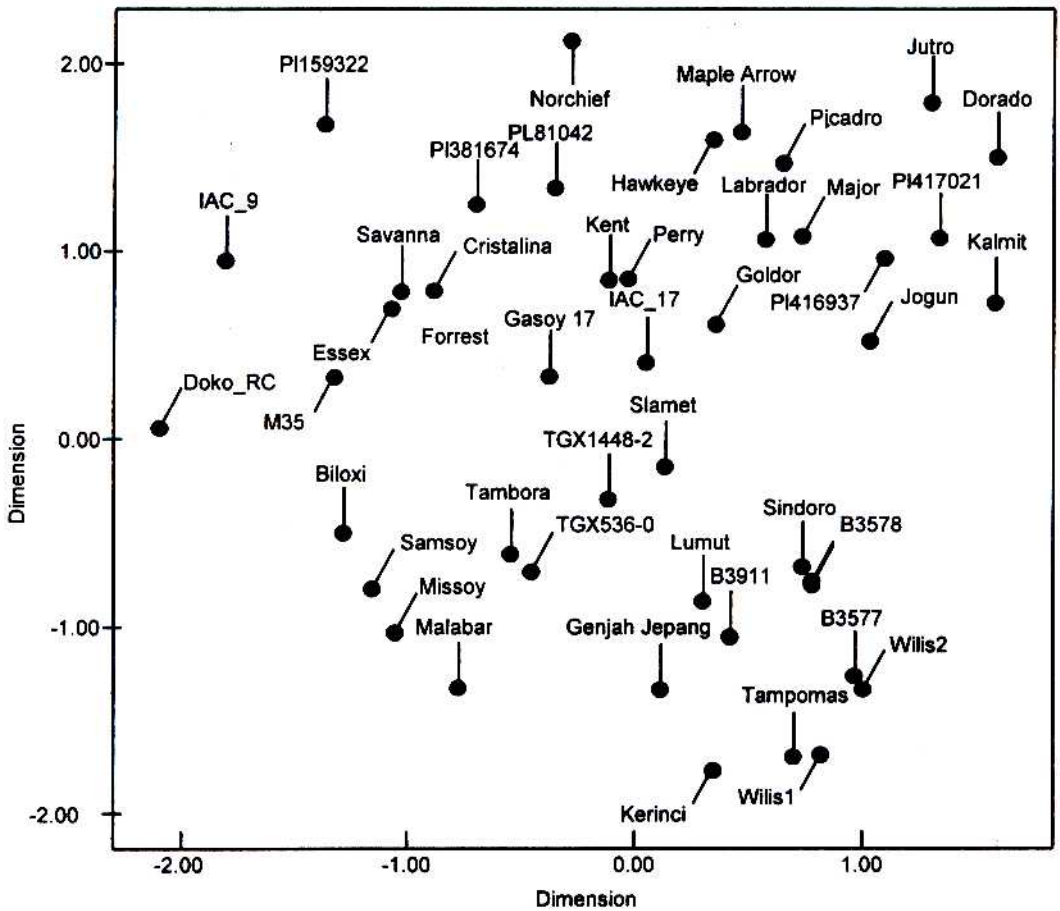


Figure 1. Two dimensional map of genetic distances obtained by MDS procedure

increase of dimensions did not substantially improve the representation of the distances. As can be seen in Figure 1 all Indonesian cultivars despite Malabar, Tambora, and Slamet are displayed in the lower right part of the map in a relatively dense cluster. This cluster is not mixed with cultivars of any other origin. This implies a rather distinct genepool of Indonesian soybean germplasm. All other origins are more or less mixed, however certain trends are visible. The European, US, and Brazilian genotypes are mostly restricted to the upper part of the map. In the upper part of the map European cultivars can be found on the right side whereas Brazilian cultivars are to be found on the left side. Genotypes of US origin seem to be most dispersed.

These observations were used as guidance for the subsequent hierarchical cluster analysis. The two most frequently used methods "UPGMA" and "Ward's" for clustering genotypes were executed on the distance data and results displayed as dendrograms.

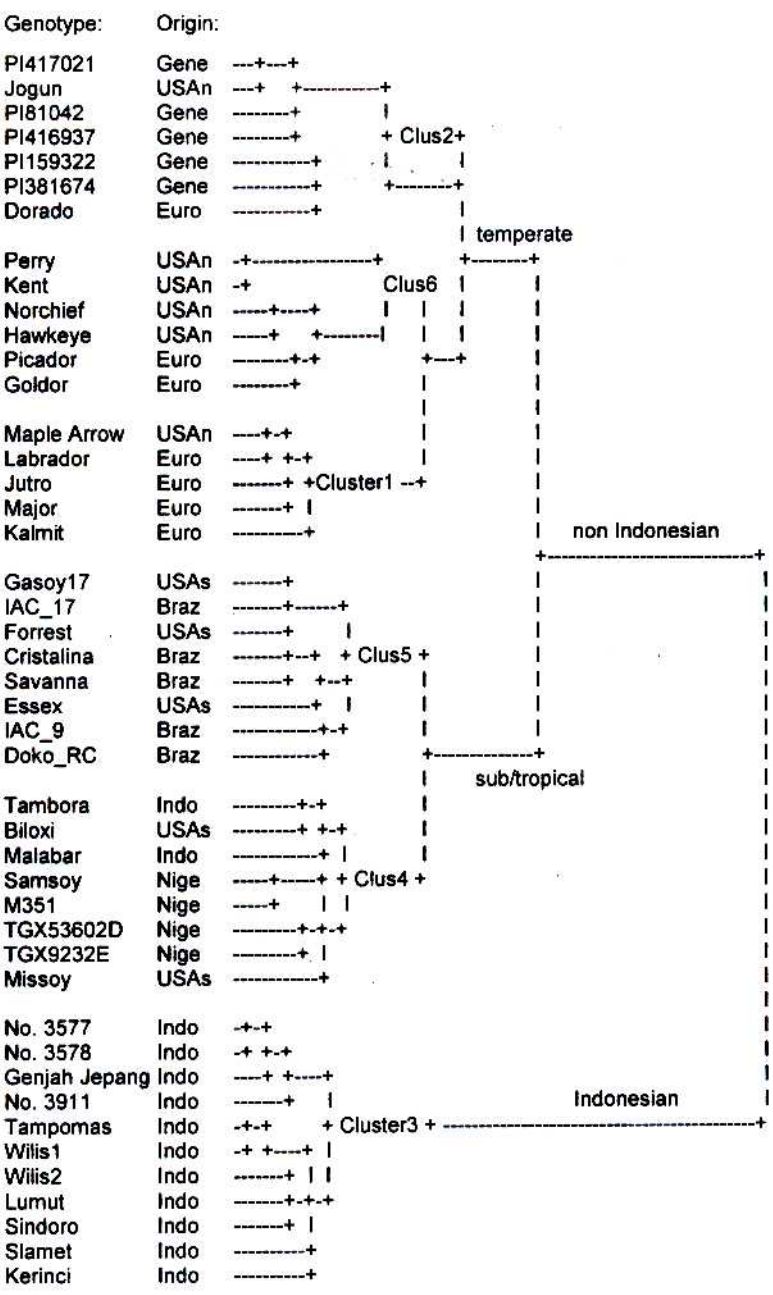
UPGMA method typicaly combines distinct individuals rather late to any cluster and is therefor useful for identification of outliers that are difficult to group with other genotypes.

In contrast Ward's method tends to build homogenous cluster sizes and incorporate outliers early in the clustering procedure.

A homogenous cluster consisted strictly of Indonesian cultivars is separated form all other genotypes in the two cluster solution (Figure 2). The only Indonesian cultivars outside this cluster are Tambora and Malabar which were already identified as difficult to group by UPGMA method. In the three cluster model the "non Indonesian" group is split to two clusters which can be best described as temperate and tropical/subtropical respectively. A six cluster solution was evaluated to identify association of clusters membership and origin of genotypes. Cluster 3 (Figure 2) is obviously correlated to the Indonesian genepool. All other clusters contain genotypes of different origin, but are dominated by one origin each. For better interpretation the US origin was split to a northern group with cultivars of maturity group IV or less and a southern group with higher maturity groups. By joining the Brazilian and southern US genotypes to a origin of tropical/subtropical America, a statistical significant correlation to Cluster 5 could be obtained. Equally Cluster 2 and genebank, Cluster 6 and northern USA, Cluster 1 and Europe as well as Cluster 4 and Nigeria were correlated significantly.

DISCUSSION

To evaluate the merit of the cluster analysis the obtained information was compared to available pedigree data. A review on the pedigree of Indonesian cultivars released since 1974 (Darman, 1997) revealed that estimated 50% of the genes are provided by the Davros, No. 1682, Shakti and Dempo. This supports the finding of relative low genetic distances within the Indonesian genepool. The distinctness of the Indonesian germplasm from all others may result form fact the most ancestral lines were introduced from Taiwan, whereas US breeding was based on cultivars from mainland



Notes: Braz = Brazil, Euro = Europe, Gene = Genebank, Indo = Indonesia, Nige = Nigeria, USAs = southern USA, USAn = northern USA

Figure 2. Dendrogram according to Ward's clustering method

China. Not all pedigree information could be confirmed by molecular data. Slamet and Sindoro are expected to be closely related since these cultivars are full siblings, but show significant genetic diversity (Figure 1). The cultivar Malabar is expected to have relative close relation to Wilis which is the female parent, but was found to have a rather high genetic distance and was not clustered to the same group. Furthermore two accessions of Wilis which were obtained from different sources proved to be of different genotype, indicating uncertain naming of genotypes during breeding history.

Splitting of the USA cultivars was fully supported by pedigree analysis. It was demonstrated that only seven ancestral lines contributed more than 73% to the southern US cultivars released from 1947-1988, but only 13% to northern US cultivars. Spehar (1995) showed that the pedigree of Brazilian germplasm is strongly based on the same ancestral lines as the southern US cultivars. This supports the fusion of the Brazilian and southern US cultivars to one genepool.

Besides common ancestral lines genetic similarity can be explained by similar genetic constitution due to selection or evolution under analogous climatic conditions.

There are some hints that adaptation to specific growing conditions is correlated to genetic similarity based on marker data. In fieldtrials genotypes of USA, Nigeria, Brazil and Europe were compared to Indonesian cultivars in three locations in Indonesia. Climatical conditions almost completely inhibited reproductive growth of cultivars adapted to temperate climates of Europe and USA. According to preliminary results of fieldtrials the lines TGX 1448-2 and M351 from Nigeria performed best followed by Brazilian cultivars. This corresponds roughly with genetic similarity data. The Nigerian genepool is most closely related to the Indonesian as revealed by MDS and UPGMA clustering, whereas the Brazilian genepool is a little more distinct, but still closer than the European and northern US genepool.

Prediction of climatic adaptation based on molecular information would allow preliminary selection of germplasm material for field testing. This might reduce efforts to find promising genotypes for use in new breeding programs. Before systematic use of information from molecular data, the correlation of genetic similarity and field performance has to be verified in further studies.

REFERENCES

Backhaus, H., B. Erichson, W. Plinke, and R. Weiber. 1996. Multivariate analysen methoden: Eine anwendungsorientierte. Einführung, Springer Verlag, Berlin.

Darman, M.A. 1997. Improved varieties and breeding strategy of soybean in Indonesia. Proceeding of a Workshop held at BPP Technologi. Jakarta, 6-7 August 1996.

Dice, L.R. 1945. Measures of the amount of ecological association between species. Ecology 26:297-302.

Keim, P., R. Groose, and R.C. Shoemaker. 1988. A rapid protocol for isolating soybean DNA. Soybean Genetics Newsl. 15:150-151.

Shoemaker, R.C. and T.C. Olson. 1993. Molecular linkage map of soybean. *In* O'Brien, S.J. (*Ed.*). Genetic Maps: Locus Maps of Complex Genomes. 6th ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY. p. 131-139.

Spehar, C.R. 1995. Diallel analysis for mineral element absorption in tropical adapted soybeans (*Glycine max* (L.) Merrill). TAG 90:707-713.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. Nucl. Acids Res. 23:4407-4414.

Zabeau, M.P. and P. Vos. 1993. European patent application. Publication No. EP 534858.