GCP Sub-Programme 1: 2006 Commissioned Research Project #3f

Molecular characterization of groundnut (*Arachis hypogaea* L.) composite collection

Project PI: HD Upadhyaya, ICRISAT

Project Team: ICRISAT – R Bhattacharjee, DA Hoisington, S Chandra, RK Varshney and S Singh
EMBRAPA – JFM Valls, MC Moretzsohn, S Leal-Bertioli and Patricia Guimarães
UCB – David Bertioli

Groundnut [*Arachis hypogaea* (L.)]

- Cultivated groundnut is a tetraploid (2n = 40) and highly self-pollinated crop.
- Primary center of origin is the Chaco region between southern Bolivia and northwestern Argentina.
- Important oilseed crop grown in 109 countries on 25.2 million ha area (FAOSTAT, 2005).
- Over two-thirds of global production occurs in seasonally rainfed regions.

Groundnut composite collection

- The groundnut composite collection was developed considering the phenotypic diversity present in the entire collection held at ICRISAT genebank and EMBRAPA, Brazil (Upadhyaya et al., 2005).
- The composite collection consists of accessions from ICRISAT comprising of mini-core (Upadhyaya, et al., 2002), comparator mini-core and trait-based accessions along with 52 accessions of 14 wild *Arachis* species, and accessions from EMBRAPA, Brazil.

Genetic diversity in composite collection

Plant material

At ICRISAT, DNA extracted from 916 accessions following a high-throughput procedure and quantified to a working concentration of 5ng/µl.

Selection of SSR markers

- At ICRISAT, 20 SSR markers, available in public domain (Ferguson et al., 2004), initially selected and pre-screened on 184 mini core accessions from which 10 polymorphic markers identified.
- Eleven SSR markers from EMBRAPA (Moretzsohn et al. 2005) also included to fingerprint the composite collection.

Molecular characterization

- The PCR components of all 21 SSR markers optimized following Taguchi method (Taguchi, 1986) described in Cobb and Clarkson (1994).
- Fluorescent-based multiplex genotyping system used to generate five multiplexes of four markers each.
- Capillary electrophoresis with an automated system (ABI 3700) to separate the amplified PCR products.
- SSR fragment sizes called to two decimal places using Genotyper v3.7 software.
- Groundnut being a tetraploid crop species, more than two peaks (or alleles) of almost equal size or 80% of the highest peak observed.
- Peaks also observed at two different locations/positions (Fig. 1).
- All the peaks, which were of equal height or 80% of the highest peak, recorded in a raw data file.
- A criteria/macro developed to remove peaks that were less than 80% in peak height compared to highest peak and data reduced to two-peak situation (raw data file being retained for future reference).
- Observed allelic data binned into discrete genetic units using Allelobin program (http://www.icrisat.org/gt-bt/biometrics.htm) developed at ICRISAT based on least squares algorithm of Idury and Cardon (1997).
- Markers produced allele sizes expected based on repeat motif of each of the SSR markers.
- Less than 5% missing data (i.e. marker x genotype) in the dataset.

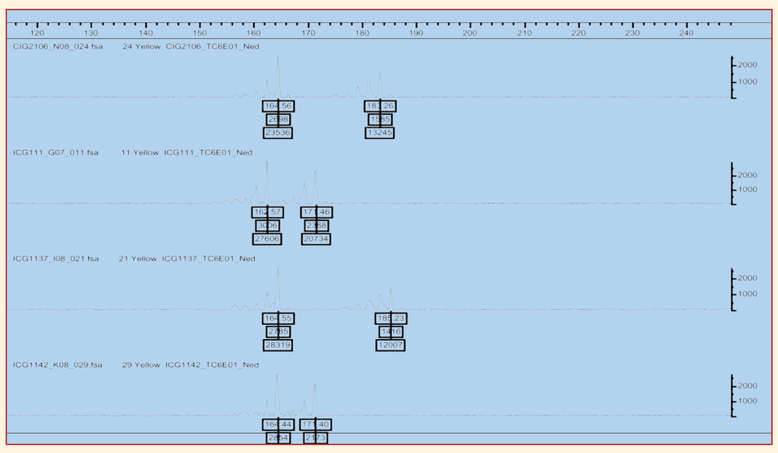

Fig. 1. A typical GeneScan Snapshot.

Data analysis

- DARwin 5.0 Structure program (Perrier et al., 2003) used to determine population structure of the composite collection.
- Sixty accessions with high missing values excluded from data analysis.
- Principal coordinate analysis done considering taxonomical classification of *Arachis*, i.e. at the level of two subspecies (*hypogaea* and *fastigiata*) and six botanical varieties (*hypogaea*, *fastigiata*, *vulgaris*, *peruviana*, *aequatoriana* and *hirusta*), and the wild species.
- Cervus 2.0 software used to calculate allele frequencies and PIC values.

Results

- Statistical analysis detected a total of 491 alleles, ranging from 5 (7H6) to 46 (5D5) with a mean of 23.4 alleles per locus (Table 1).
- The mean Polymorphic Information Content (PIC) value was 0.796 (ranging from 0.483 to 0.923) (Table 1).
- The regional distribution (Table 2) showed that number of alleles per locus ranged from 3.3 in accessions from Oceania to 22.1 in accessions from South America.

Table 1. SSR primers used in the study with information on their repeat units, quality index, number of alleles and PIC values.

Primer	Repeat Unit	Quality Index	No. of Accessions Genotyped	No. of Alleles	PIC Values
1B9	(GA)19	0.21	840	26	0.84
2D12B	(TAA)16	0.43	830	21	0.88
7H6	(CTT)12	0.26	849	5	0.48
8E12	(TTG)6 (TAA)15	0.22	844	20	0.68
13E9	(TAA)16	0.24	851	14	0.64
5D5	(GA)32	0.34	825	46	0.92
15C12	(TAA)28	0.30	794	22	0.84
17E3	(CTT)15	0.24	827	13	0.76
18C5	(TAA)23	0.18	842	14	0.80
19B1	(AG)2	0.25	837	28	0.69
TC1A02	(TC)35	0.49	841	27	0.90
TC4F12	(CT)23	0.20	840	20	0.89
TC6E01	(GA)22	0.47	851	25	0.89
TC6H03	(AG)21	0.47	812	36	0.90
TC11H06	(AG)34	0.30	817	27	0.92
TC1E01	(GA)29	0.17	830	23	0.77
TC11A04	(CT)16 + (CT)33	0.22	843	23	0.89
TC7H11	(AG)18	0.47	822	29	0.86
TC9F10	(AG)31	0.34	825	28	0.86
TC3E02	(CT)26 + (CA)7 + (CA)5	0.29	830	22	0.62
TC6G09	(CT)18	0.25	855	22	0.72

Table 2. Number of alleles recorded per SSR marker in different geographical regions.

Region	Accessions	No. of alleles Locus
South America	173	22.1
North America	98	13.0
Caribbean	5	4.4
South Asia	311	15.6
East Asia	49	10.8
South-East Asia	45	10.7
Central Asia	7	5.9
Middle-East	20	9.8
West Africa	71	11.3
South Africa	48	10.9
East Africa	31	10.3
Central Africa	9	6.8
Europe	4	4.5
Oceania	3	3.3
Unknown	42	11.6

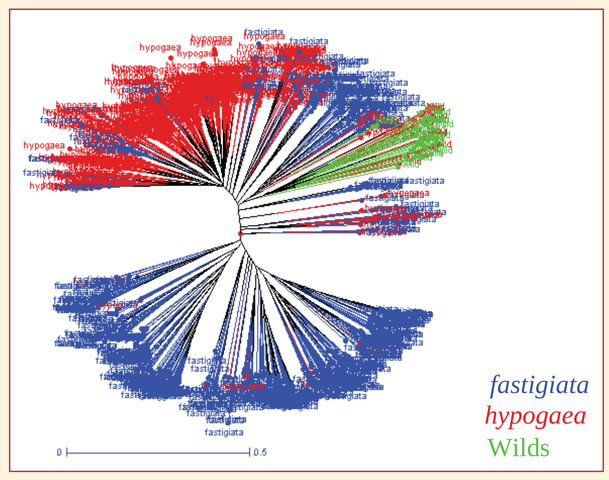

Fig. 2. Tree diagram of 856 accessions with 21 SSR markers at subspecies level.

- Phylogenetic tree, constructed at subspecies level, revealed that both *hypogaea* and *fastigiata* accessions formed distinct clusters (Fig. 2). However, some *fastigiata* accessions grouped with *hypogaea* types mainly due to common geographical origin.
- The accessions belonging to 14 wild *Arachis* species grouped together in a separate cluster, but close to *hypogaea* accessions (Fig. 2).

Future plan

- Further analysis of data in progress to fully understand the genetic diversity and population structure of the composite collection.
- Results from genotypic data will be used to identify a reference set of 300 diverse accessions for future use.
- To ascertain the quality and position of SSR markers, these will be checked on 15-20 plants in each of four F_2 populations, whose parents have been included in the composite collection.

References

Cobb BD and **Clarkson JM.** 1994. A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods. Nucleic Acids Research, 22: 3801-3805.

FAOSTAT 2005. http://apps.fao.org/

Ferguson ME, Burow M, Schulze SR, Bramel PJ, Paterson A, Kresovich S and **Mitchell S**. 2004. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theoretical and Applied Genetics, 108:1064-1070.

Idury RM and Cardon LR (1997). A simple method for automated allele binning in microsatellite markers. Genome Research. 7: 1104-1109.

Moretzsohn M, Leoi C, Proite L, Guimara K, Leal-Bertioli PM, Gimenes SCM, Martins MA, Valls WS, Grattapaglia JFM and **Bertioli DJ.** 2005. A microsatellite-based, gene-rich linkage map for the AA genome of Arachis (Fabaceae) Theoretical and Applied Genetics, 111: 1060–1071

Perrier X, Flori A and **Bonnot F.** 2003. Data analysis methods. Pp. 43-76. In: Genetic Diversity of Cultivated Tropical Plants. (Hamon P, Seguin M, Perrier X and Glazmann JC eds.). Enfield, Science Publishers, Montpellier.

Taguchi G. 1986. Introduction to Quality Engineering, Asian Productivity Organization. American Supplier Institute Inc., Dearborn, MI.

Upadhyaya HD, Bramel PJ, Ortiz R and **Singh S**. (2002). Developing a mini core of peanut for utilization of genetic resources. Crop Science, 42: 2150-2156.

Upadhyaya HD, Ranjana Bhattacharjee, Bertioli D, Hoisington D, Pundir RPS, Valls JFM, Chandra S, Singh S, Moretzsohn MC, Leal-Bertioli S and **Guimaraes P**. (2005). Genotyping groundnut (*Arachis hypogaea*) composite collection using SSR markers. A poster presented in the Annual Research Meeting of Generation Challenge Program, 29 Sep-1 October 2005. Rome, Italy.

International Crops Research Institute for the Semi-Arid Tropics
Patancheru 502 324, Andhra Pradesh, India
www.icrisat.org