Molecular Analysis of Bambara Groundnut, an Underutilised African Legume Crop as Part of the BAMLINK Project – What Lessons Can We Learn?

S. Mayes[1,5], S.M. Basu[1], O. Molosiwa[1], E.S. Redjeki[1], N.S. Ahmad[1], F. Khan[1], S. Zehra[1], S. Noah[1], K. Mayes[1], J. Roberts[1], F. Stadler[2], F. Massawe[3], A. Kilian[4] and S. Azam-Ali[5]
[1] Plant and Crop Sciences, Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough, Leicestershire, LE12 5RD, United Kingdom
[2] Plant Breeding, Centre of Life and Food Sciences, Technische Universität München, Emil-Ramann-Str. 4, 85354 Freising, Germany
[3] The University of Nottingham, Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia
[4] Diversity Arrays Technology, 1 Wilf Crane Cr., Yarralumla, ACT 2600, Australia
[5] Crops for the Future Research Centre, The University of Nottingham, Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

Keywords: *Vigna subterranea* L. Verc., underutilised crops, domestication, molecular markers, breeding

Abstract

Bambara groundnut (*Vigna subterranea* L. Verdc.) is an underutilised, drought tolerant legume that has the potential to form an important part of Food Security for the coming decades. The challenges facing farmers to produce enough food for the growing world population – particularly that of climatic instability – are well documented and together represent probably one of the biggest challenges humanity has faced. Our extreme reliance on a limited number of staple (often non-indigenous and sometimes also poorly adapted) crops represents a clear vulnerability. This can be partly reduced by the development of alternative crops. These currently underutilised crops often have beneficial characteristics not found in conventional main crops and if these traits address either biotic or abiotic stresses in a sustainable way, then there is the potential for agricultural diversification. There are a number of stumbling blocks to developing such crops, including; poor yields, unadapted crop features, limited processing knowledge, few value-added products, poorly developed transport chains and markets, negative cultural perceptions and little perceived profit margin for commercial breeders. An integrated approach is needed to begin to address these problems. As part of this, we have focused on the application of molecular genetics to Bambara groundnut and the opportunities to exploit knowledge from other species, new technologies and new approaches, to establish a framework for genetic improvement through breeding. We also try to draw out lessons from our work in Bambara groundnut which may be relevant in other underutilised species, to try to contribute to the development of a generic approach (and hopefully, a faster and cheaper approach) to tackling these same questions in other underutilised species. In this paper we ask what is the fundamental information we need about the breeding system of an underutilised species and how could this alter our genetic improvement, using Bambara groundnut as an example.

INTRODUCTION

Developing new crops and expanding the range and level of production and use of existing underutilised crops is a major challenge which could help to generate more resilient agriculture. While this does not underestimate the critical importance of improvements to staple crops, such developments could provide alternatives where climate, demographic or soil nutritional changes prevent current staples from being economically cultivated in future. Bambara groundnut is an indigenous African legume with good drought tolerance which is still grown widely in sub-Saharan Africa, albeit at

low levels. Many such crops exists which have potential to be more utilised than they currently are, but establishing extensive research programmes on a large number of underutilised crops is not feasible, either in terms of capacity or in terms of finance. BAMLINK was an EU FP6 INCO-DEV programme which aimed to evaluate the use of Bambara groundnut at the molecular, eco-physiological, nutritional and end-user levels to try to identify the constraints to enhanced uptake of this crop for a range of uses. We are also attempting to use this crop as an exemplar for other underutilised crops, to understand what we need to know about a crop to be able to make progress with it.

Here, we specifically look at our results for the molecular analysis of this crop to identify what information we need to understand to be able to make focused progress in a generic crop. In Bambara groundnut we have generated within species data, where specific tools were needed for subsequent quality control and breeding and have also applied data across species to try to link this underutilised species with available data from major species. This article focuses on the generation of data within species and how this knowledge has allowed us to map out breeding and genetic improvement options within Bambara groundnut, which could allow a strategic approach to be developed for other underutilised species.

MATERIALS AND METHODS

Plant Materials

All Bambara groundnut genotypes analysed were drawn from either the collection at the International Institute for Tropical Agriculture (IITA, Ibadan, Nigeria) or from the Nottingham University stock collections (Sutton Bonington Campus, Nottingham University, UK). The 24 standard landrace accessions (single genotypes) used for characterising the SSR markers are given in Table 1. Fifty individual seed of UniSwaRed and S19-3 landraces, grown in the Tropical Crops Research Unit (TCRU) glasshouses in 2007, were leaf sampled. DNA was extracted from individual genotypes using standard approaches and all 100 samples were screened against three SSR markers from Basu et al. (2007). Microsatellites were labelled, amplified and analysed as described below. The Correspondence Analysis was generated using the defaults in MVSP v.3 (Kovachs Computing Services) and the first two axes plotted.

Development of Microsatellites for Bambara Groundnut

Three approaches were used to develop within species microsatellite markers:

1. Development of a genomic fragment microsatellite-enriched library, essentially according to Basu et al. (2007). Sanger sequencing of clones containing repeat sequences was used to allow the construction of primers flanking the microsatellite repeat sequences.
2. The amplified PCR products from the adaptor bearing insert of the library from 1) was sequenced using part of a $1/16^{th}$ Roche 454 pyrosequencing run (9 species libraries with single base changes in the adaptors were mixed before sequencing; Eurofins MWG).
3. A plate of 454 pyrosequencing (Titanium reagents) from the leaf transcriptome of a single plant of the Namibian accession S19-3 (50 days after sowing; 28°C/23°C cycle with 12 hours photoperiod, planted directly into a sandy loam) was produced. The data were assembled (Newbler, by Deep Seq, University of Nottingham) and microsatellite repeats within the transcriptome had primer pairs designed to them, were possible.

In each of the three cases, the MISA.pl script (IPK, Gartersleben, Germany; pgrc.pik-garterleben.de/misa/misa.html) was used to characterise the microsatellites present. Primers were designed using the Primer3 software (frodo.wi.mit.edu) including a minimum primer length of 24 nucleotides and optimal length of 27. For all primers an M13 extension was added to the forward primer of the pair to allow third primer labelling (Schuelke, 2000), with a dye-labelled (Well-RED blue; Sigma-Aldrich) to be incorporated allowing fragment size assessment on a Beckmann CEQ 8000. PCR

reactions were according to standard protocols, with initial assessment of amplification using a 15°C annealing temperature gradient, from 45 to 60°C, with a 3 minute pre-denaturation, followed by a 1 minute denaturation at 94°C, 1 minute annealing (45-60°C) and a 2 minute extension phase at 72°C for 35 cycles, with a final 10 minutes at 72°C. A mixture of DNA from six accessions was used for the optimisation PCR step. A total of 24 accessions (single plant genotypes) were used for polymorphism screening. These were selected to represent the available genetic diversity in Bambara groundnut from previous work with selection based on the AFLP analysis conducted by Singrün and Schenkel (2004).

Scoring and Data Analysis

Fluorescently labelled PCR products were run according to the manufacturer's instructions on a Beckmann CEQ 8000, with size determination and allele calling by manual inspection of the microsatellite traces across the 24 accessions. Data were scored on an allele size basis into a Microsoft Office Excel sheet (Office Suite 2007), before analysis using PowerMarker v3.25.

RESULTS

Developing Microsatellites for Bambara Groundnut

The three phases of microsatellite development gave the following results, based on the same 24 individual genotype accessions:

1. 625 colonies identified as containing potential microsatellite repeats were Sanger sequenced and 151 contained clear repeats. 112 primer sets could be designed (74%) and approximately half of these produced clear single PCR products (37%) and 1/3rd of these were also polymorphic within the 24 genotypes (12% recovery).
2. The uncloned insert of the same library was 454 pyrosequenced as part of a mix of libraries. This generated 5443 sequences with an average length of 312 bp, containing 1559 repeats. Designing primer pairs to 143 sequences produced a recovery of 45 polymorphic microsatellites (31% recovery).
3. Screening of a 454 leaf transcriptome using stringent conditions (di>10 repeats, tri>6 repeats) identified 351 potential sequences for primer design, 68 primer pairs gave 43 clear amplification products and 29 polymorphic markers (46% recovery).

From the original 68 primers, a further 10 clear amplified products were also obtained which were larger than the products predicted from coding sequence, suggesting that the products spanned introns.

A PowerMarker analysis of 23 of the 29 SSRs derived from coding sequence (method 3; Table 2) shows major allele frequencies (MAF) ranging from 0.25 to 0.96 (average=0.68), suggesting a reasonably even distribution of the different alleles across the 24 genotypes for some markers. The number of different genotypes identified from 24 accessions ranges from 2 to 11 (average=4), the number of alleles ranging from 2 to 10 (average=3.8) and polymorphic information content (PIC) values for individual markers ranging from 0.08 to 0.8 (average=0.4).

Interestingly, the average heterozygosity observed across 24 genotypes (single accessions from landraces) suggests that the level of heterozygosity is actually very low (average=0.02). This is actually lower than the standard 5% expected in a number of temperate inbred crops and suggests that we can treat seed derived from a single plant of a natural landrace as an inbred line (i.e., an unselected cultivar).

Population Structure between Bambara Groundnut Landraces

Many underutilised crop species exist as landraces i.e., collections of non-identical genotypes.

If the species is naturally out-pollinating, then the majority of genetic variation would be expected to exist between populations. If inbreeding, then a landrace is likely to consist of a series of inbred lines which differ from each other, as is the case for Bambara

groundnut (Basu et al., 2007). Figure 1 presents a UPGMA tree based on Nei and Li's genetic distance constructed from the data for 65 SSRs screened against the standard 24 genotypes in Table 1.

Population Structure within Bambara Groundnut Landraces

From Figure 2 it is clear that the 50 individual seed from the UniSwaRed landrace are genetically more variable than the 50 seed derived from the S19-3 landrace, for these collections. This is also reflected morphologically in the UniSwaRed landrace, with a narrow leaf morphology variant present (approximately 10% of plants).

Comparisons between the Bambara Leaf Transcriptome and Other Species in the Databases

Table 3 presents a Blast2Go analysis using the assembled gene models of the Bambara groundnut 454 Pyrosequenced leaf transcriptome to search the available databases (NCBI). The results clearly show that the appropriate cross-species gene and genome model for Bambara groundnut is soybean (*Glycine max*). From this, the use of the Affymetrix (or other microarray) GeneChips would be the best choice for Bambara groundnut work. In addition, it should also be possible to overlay the transcriptome of Bambara groundnut onto the soybean genome (initially for the leaf transcriptome).

DISCUSSION

Underutilised crops have the potential to contribute to food security in the future. This could be through the increased uptake of minor crops, diversification of diet for nutritional security or the development of high value products providing purchasing power to small-scale farmers. However, many underutilised crops are underutilised for a reason and, unless a coherent and integrated approach to tackle restrictions on further uptake is made, encouraging more extensive use of these crops is likely to fail. One fundamental requirement to translate research results into impact for farmers is the existence of an integrated breeding programme. Breeding programmes depend upon the available germplasm and the breeding system of the crop species. Molecular markers can help to elucidate these processes and identify the best options for building a breeding programme.

In this paper we have described the development and characterisation of in-species microsatellite markers through three different methods and their application to a set of genotypes which represents the breadth of the available germplasm as well as 50 individual genotypes from two landraces.

CONCLUSIONS

- Bambara groundnut is highly inbreeding, with a residual heterozygosity within single individuals of H_o=0.02. Previously, the potential for developing new cultivars through single seed descent from existing landraces has been suggested (Massawe et al., 2005). The current result suggests that we can treat single seed derived from a single plant of a landrace as an unselected variety, without the need for further inbreeding. We have adopted this approach for our experimental and crossing programmes.
- Many underutilised crops will exist as landraces, rather than pure line cultivars. An analysis of variation present can help in the identification of suitable contrasting parental accessions for crossing and breeding improvement or the potential of genetic improvement of a landrace through direct selection of individual plants within that landrace. The source of S19-3 used here is genetically narrow. This is useful for experimental analysis as genetic variation is minimised, but suggests that there may be limited progress possible through selection of individual genotypes within S19-3. However, UniSwa Red has a far wider genetic base (and some degree of leaf morphological variation) and selection of individual genotypes from within this landrace could make greater breeding progress than in S19-3.
- Using resources available in model or major crop species is important to allow progress

to be made in minor and underutilised crop species. The development of an initial Bambara groundnut transcriptome has allowed the most appropriate gene and genome model to be identified – soybean in this case.

Literature Cited

Basu, S., Roberts, S.N., Azam-Ali, S.N. and Mayes, S. 2007. Development of microsatellite markers for bambara groundnut (*Vigna subterranea* L. Verdc.) - an underutilised African legume crop species. Molecular Ecology Notes. 7(6):1326-1328.

Massawe, F.J., Mwale, S.S., Azam-Ali, S.N. and Roberts, J.A. 2005. Breeding in bambara groundnut (*Vigna subterranea* (L.) Verdc.): strategic considerations. African Journal of Biotechnology 4:463-471.

Mayes, S., Basu, S., Murchie, E., Roberts, J.A., Azam-Ali, S.N., Stadler, F., Mohler, V., Wenzel, G., Massawe, F., Kilian, A., Bonin, A., Beena, A. and Sheshshayee, M.S. 2009. BAMLINK – a Cross Disciplinary Programme to Enhance the Role of Bambara Groundnut (*Vigna subterranea* L. Verdc.) for Food Security in Africa and India. Acta Hort. 806:137-150.

Schuelke, M. 2000. An economic method for the fluorescent labelling of PCR fragments. Nature Biotechnology 18:233-234.

Singrün, C. and Schenkel, W. 2004. Fingerprinting of bambara groundnut germplasm with molecular markers. Proceedings of the Int. Symposium on Bambara Groundnut, Botswana College of Agriculture, 8-12 September 2003.

Tables

Table 1. Twenty four standard accessions (single genotype) used to characterise all SSRs markers developed. Selection of landraces was based on Singrün and Schenkel (2004), to try to represent the available germplasm.

Landraces	Origin	Region
DodR	Tanzania (TZA)	East Africa
DodC	Tanzania (TZA)	East Africa
AS17	South Africa (RSA)	Southern Africa
DipC	Botswana (BWA)	Southern Africa
SwaziRed	Swaziland (SWA)	Southern Africa
TicaNicuru	Mali (MLI)	West Africa
Ramayana	Indonesia(IND)	Asia
LunT	Sierra Leone (SLA)	West Africa
Vssp6	Cameroon (CMR)	West Africa
Nav 4	Ghana (GHA)	West Africa
Nav red	Ghana (GHA)	West Africa
Mahenene black	Namibia (NAM)	Southern Africa
S19/3	Namibia (NAM)	Southern Africa
S19-3	Namibia (NAM)	Southern Africa
UniswaRed	Swaziland (SWA)	Southern Africa
SB16 5A	Namibia (NAM)	Southern Africa
AHM968	Namibia (NAM)	Southern Africa
NAM 1761/3	Namibia (NAM)	Southern Africa
Malawi 3	Malawi (MW)	Southern Africa
Tvsu 569	Cameroon (CMR)	West Africa
Tvsu 610	Nigeria (NGA)	West Africa
Tvsu 747	Zambia (ZMB)	Southern Africa
GabC	Botswana (BWA)	Southern Africa
Tvsu 999	Zimbabwe (ZWE)	Southern Africa

Table 2. Powermarker analysis of 23 microsatellites derived from leaf transcriptome data and amplified from the single genotype accessions in Table One. MAF – Major allele Frequency; Geno – number of genotypes identified; Alleles – number of alleles; Het – percentage of heterozygosity identified within individuals; PIC – polymorphic information content.

Marker	MAF	Geno	Alleles	Het	PIC
D.36186	0.90	3	2	0.04	0.17
D.8148	0.79	3	3	0.00	0.31
D.15508	0.29	7	5	0.08	0.72
D.25551	0.79	2	2	0.00	0.28
D.21950	0.95	2	2	0.00	0.09
D.8387	0.83	3	3	0.00	0.27
D.8999	0.25	9	8	0.08	0.80
D.42026	0.72	3	2	0.04	0.32
D.14265	0.67	3	2	0.04	0.34
D.15619	0.52	4	3	0.04	0.47
D.12522	0.39	5	5	0.00	0.66
D.1006	0.96	2	2	0.00	0.08
D.21310	0.78	3	3	0.00	0.33
D.24269	0.88	3	3	0.00	0.21
D.37053	0.75	2	2	0.00	0.30
D.32937	0.33	6	6	0.00	0.74
D.35497	0.35	11	10	0.04	0.80
D.48339	0.29	7	6	0.05	0.76
D.7215	0.77	3	2	0.04	0.29
D.655	0.96	2	2	0.00	0.08
D.51646	0.36	9	9	0.00	0.79
D.11860	0.94	2	2	0.00	0.10
D.125	0.88	2	2	0.00	0.19

Table 3. Blast2Go analysis based on a 454 Pyrosequencing assembly (full plate, Titanium reagents) of a leaf transcriptome (S19-3 accession) assembled using Newbler. Number of transcripts from Bambara groundnut identifying matches above $1e^{-5}$ and $1e^{-10}$ stringencies. The results show that for any cross-species work, soybean (*Glycine max*) is the most appropriate gene and genome model for Bambara groundnut.

Species	Transcript best matches $1e^{-5}$	Transcript best matches $1e^{-10}$
Glycine max	31,274	29,978
Medicago truncatula	1,871	1,658
Vitis vinifera	435	378
Populus trichocarpa	416	368
Ricinus communis	371	308
Arabidopsis lyrata subsp. *lyrata*	139	102
Vigna radiate	104	101

Figures

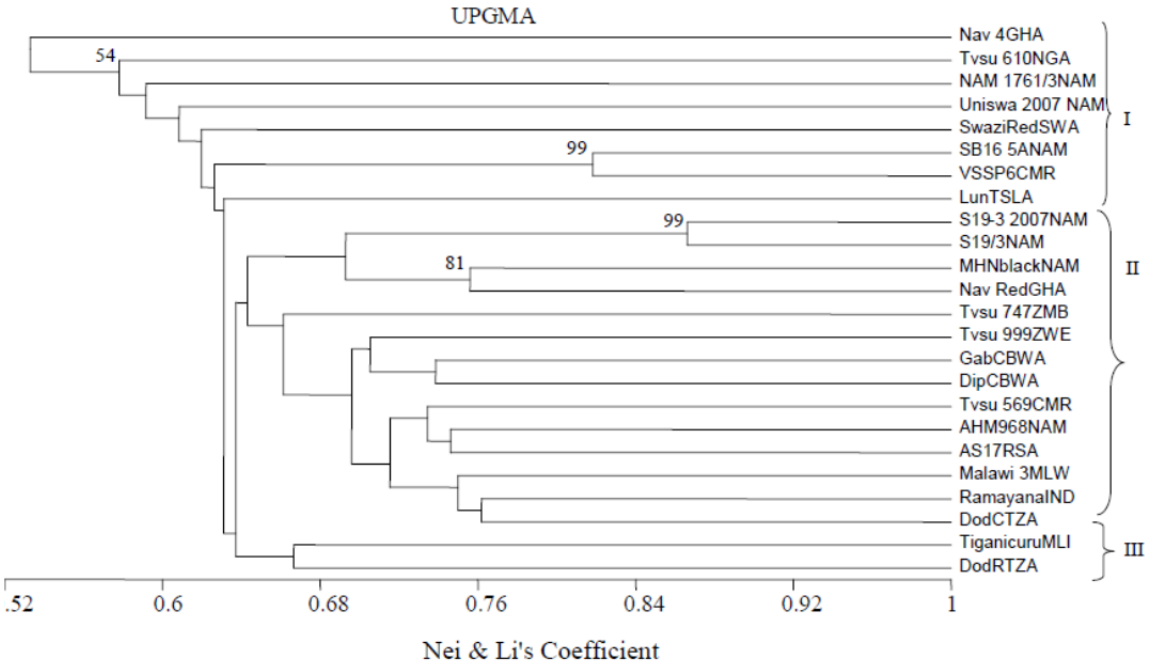
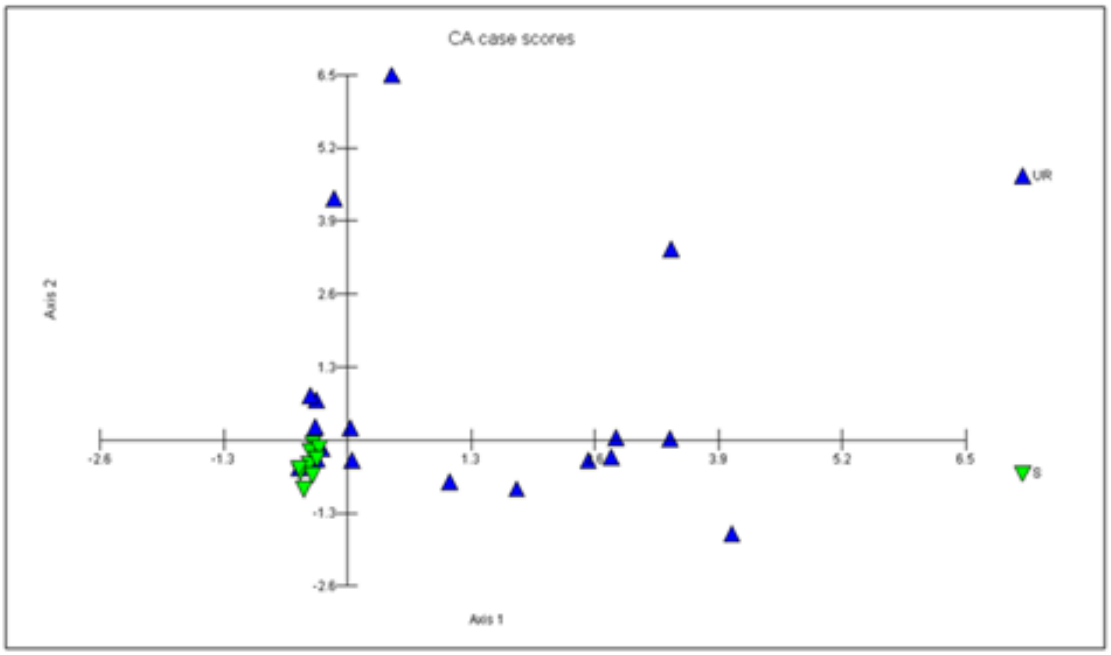


Fig. 1. UPGMA tree based on Nei and Li's coefficient using data from a total of 65 SSR markers derived from Bambara groundnut amplified from the standard 24 genotypes representing the available germplasm, based on Singrün and Schenkel (2004). Bootstrap values are given based on 1000 resamplings.



Cultivar

$V_P = V_G + V_E + V_{GxE}$

Landrace

$V_p = V_{P1} = V_{G1} + V_{E1} + V_{GxE1} +$
$V_{P2} = V_{G2} + V_{E2} + V_{GxE2} +$
$\ldots V_{Pn} = V_{Gn} + V_{En} + V_{GxEn}$

Fig. 2. Principle Component Analysis of 50 individual plants from both UniSwaRed (UR) and S19-3 landraces (S) which have been repeatedly grown over the years at the University of Nottingham in the Tropical Crops Research Unit (TCRU; 2007 harvest). The individual genotypes have been amplified with three pre-selected microsatellite markers. The confounding effect of landrace structure compared to cultivar analysis is presented below the PCA figure, with a cultivar in a perfect case scenario having Phenotypic variance (V_P) composed of Genotypic variance (V_G), Environmental variance (V_E) and the interaction (V_{GxE}). However, because landraces consist of multiple genotypes the dissection of traits becomes significantly more complex.