

APPLYING MOLECULAR GENETICS TO UNDERUTILISED SPECIES – PROBLEMS AND OPPORTUNITIES

SEAN MAYES^{1,2*}, WAI KUAN HO^{1,2}, PRESIDOR KENDABIE², HUI HUI CHAI², SIISE ALIYU^{1,2}, ARYO FELDMAN¹, RAZLIN AZMAN HALIMI¹, FESTO MASSAWE² and SAYED AZAM-ALI¹

¹*Crops for the Future, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia*

²*The University of Nottingham, Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia*

*Email: sean.mayes@cffresearch.org

ABSTRACT

Molecular markers represent an important tool for marker-assisted breeding in major crop plant breeding programmes. Applying molecular genetics to underutilised and minor crop species is more challenging as the funds available to research and develop such crops are often severely limited. Bambara groundnut is an underutilised African legume crop with good drought tolerance. It is also grown at low levels in Southeast Asia. In this review we examine some of the applications of DNA markers and illustrate their value in bambara groundnut.

Key words: underutilised crops, bambara groundnut, *Vigna subterranea* (L) Verdc., marker-assisted selection, molecular genetics, fingerprinting, DNA markers

INTRODUCTION

The demands on the world food supply are increasing due to increasing population, the need to reduce agricultural inputs and the effects of climate change (e.g. Foley *et al.*, 2011). The current agricultural paradigm focused on the intensive cultivation of a very limited number of crop species will not be sufficient and where these crops are planted inappropriately they will begin to fail more frequently due to weather volatility. To increase agricultural resilience more crops need to be developed from the estimated 7000 species that have previously been used by humankind (Haran, 1992). How do we decide which ones to focus on? One approach is identifying species which have trait values beyond those of equivalent major crops, such as drought tolerance, nutritional content or which act as complementary crops within the agricultural system. Alongside narrowing the candidate crop list, we also need to broaden our approach, through sequence- and trait-based comparative links between species, so that we work on a complex of species, rather than single species in isolation. Legumes are a key part of low input agricultural systems, contributing nitrogen to the soil and also

producing high levels of protein for human diets. Bambara groundnut is a drought tolerance African legume which can serve as an exemplar crop, allowing lessons to be learned and generic approaches developed.

The breeding system and genetic variation

The natural breeding system of a plant ultimately determines the basic structure of the genetic (and, to some extent, trait) variation within a breeding population. ‘Structured’ breeding (in preference to farmer trait-based selection; Allard, 1960) is little more than a century old, so the ‘natural’ breeding history is likely to have had by far the greatest influence on the genetic composition of the species.

Understanding domestication itself is nevertheless important. If the ancestor species and the domesticated species are still sexually compatible then there is the potential to reach back into the wild material to introgress new genes absent from the current domesticated germplasm. A controlled cross between wild and domesticated bambara groundnut (*Vigna subterranea*; Azam-Ali *et al.*, 2001) suggested that the wild extreme spreading type was the ‘natural’ state with the bunched types a selection made during domestication (Basu *et al.*, 2007a).

* To whom correspondence should be addressed.

The existence of domestication bottlenecks, founder effects as crops are introduced to new growing regions (e.g. Alpern 1992; Alpern 2008), differing mutation rates and different generation cycle times affect the intrinsic genetic variation available. These have all then been acted upon at the phenotypic level by farmer selection over long periods of time (Harlan 1992; Gepts 2004; Doebley *et al.*, 2006). Many underutilised species (Padulosi *et al.*, 2002) exist essentially as landraces – collections of genotypes which have been selected ‘in field’ over long periods to be adapted to the growing environment and to farmer requirements for similar growth duration and harvest times. Understanding these factors is critical in making the best choices for the genetic improvement of a particular underutilised species; Figure 1; (Sompta *et al.*, 2011; Olukolu *et al.*, 2012; Aliyu and Massawe, 2013; Molosiwa *et al.*, 2015).

However, even more basic to these is the ability to confirm at each stage of research or breeding development that what we think we have is really what we have. Quality control in research and breeding programmes is often based on trust (for

received material), on assumptions (e.g. literature reports on average outcrossing rates) and on the believe that everyone carries out the Standard Operating Procedure (SOP) perfectly, if it exists. These assumptions lead to major errors in both breeding and research. Molecular markers can and should be used to test these assumptions at every stage (e.g. in oil palm and date palm; Mayes *et al.*, 1996; Al-Ruquaisi *et al.*, 2008; Wong *et al.*, 2015); Figure 2. A microsatellite analysis of 123 individual plants from bambara groundnut landraces revealed very low levels of heterozygosity ($H_o = 0.02$; $H_e = 0.47$; Molosiwa *et al.*, 2015) which essentially means that inbred lines (unselected cultivars) can be developed simply by selecting seed from a single plant. Because of the results of these marker analyses, we have adopted a ‘single plant’ accession of a landrace to represent a unique genotype, presenting characters selected in that landrace by farmers. This also facilitates genetic analysis, by removing the intrinsic genetic variation of the landrace and provides parental genetically uniform lines for controlled cross breeding, to combine desirable traits, some of which will be intrinsic (such

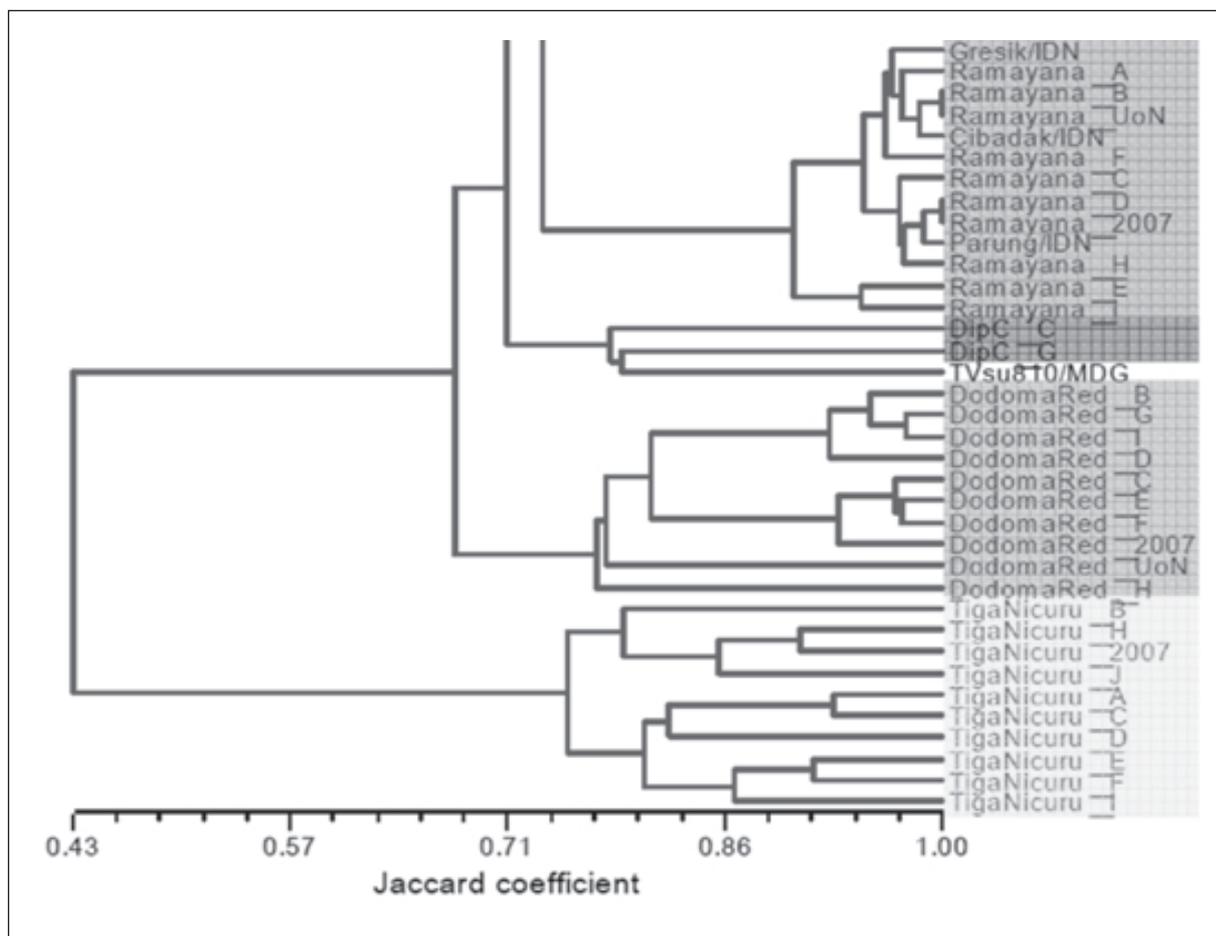


Fig. 1. Genetic diversity within landraces of bambara groundnut. Partial UPGMA dendrogram presenting the genetic similarity of a series of single seed (genotype) samples from landraces; based on 296 DAiT Array markers (from Stadler, 2009).

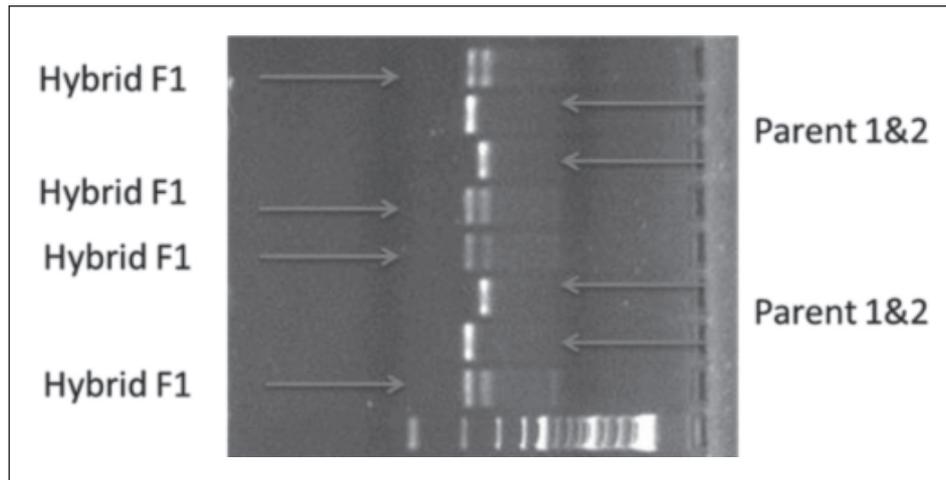


Fig. 2. Quality control in research and breeding programmes is critical. Here microsatellites pre-screened to show large size differences between parental lines are used to test the endosperm of putative hybrids to confirm successful crosses. SSR were ‘mined’ from sequence data as described in Basu *et al.* (2007a) and Molosiwa *et al.* (2015).

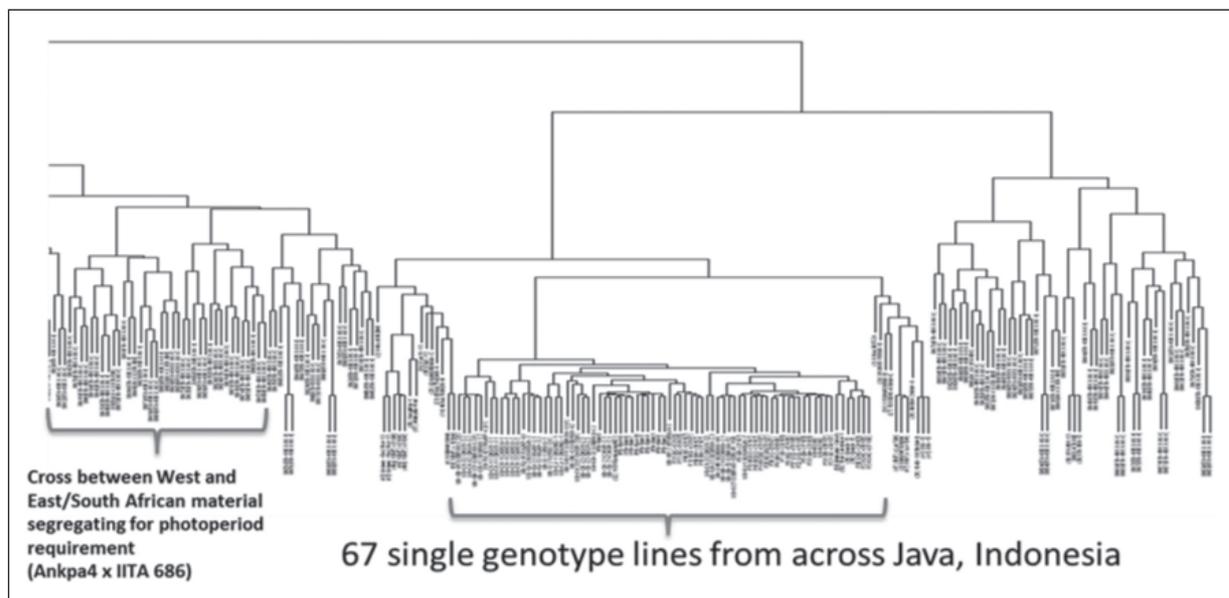


Fig. 3. Diversity Array Technology (DART) Seq analysis of single genotypes derived from bambara groundnut landraces and controlled crosses showing a partial UPGMA tree; based on 3,542 SNPs, 5,780 Silico DART. Indonesian material appears to be genetically narrower than the African samples analysed.

as reduced photoperiod requirement for pod-set) and some end-user and farmer defined (such as shorter cooking times). The landrace/accession that the genotype line was derived from is known so it is possible to look for a number of lines with greater trait value alleles by re-screening the original landrace; Figure 3. This approach could be used to exclude the poorer performing lines from the landrace to increase the trait value and develop a ‘selected landrace’, while retaining most of the genetic diversity. This may be a critical part of the resilience characteristics shown by some underutilised species.

Breeding is essentially a numbers game and generating large numbers of individuals in controlled crosses for underutilised species can be difficult. Given the ‘plant seed descent’ approach above, multiple parental plants can be crossed and seed combined. A cross evaluating photoperiod requirement for pod development between an Ankpa 4 genotype (high sensitivity) and IITA 686 (less sensitive) contained 1,300 segregating F_2 seed derived from four crosses between genotype selected lines (Kendabie, 2015).

Understanding the distribution of genetic diversity is important to determine breeding options.

Analysis in bambara groundnut suggests that there is a genetic disconnect between West African and Southern + East African material, which raises the possibility of there being a major and a minor centre of domestication. Selecting parental lines derived from different genetic groups, but similar agroecologies potential allows the introduction of substantial genetic and trait variation into controlled crosses without disrupting the major complexes for adaptation to that environment; Figure 4. In addition, the introduction to the island of Java (Indonesia; probably during the slave trade, with marker analysis suggesting East Africa as the origin; Figure 3) of bambara groundnut may have come from a limited number of accessions, producing a founder effect which could limit breeding progress for specific desirable traits such as early maturity. African accessions potentially offer genetic variation for such traits (Sri Redjeki, 2014).

The genetic tools

The advent of next generation sequencing (NGS; Varshney *et al.*, 2009) on a number of different technology platforms has revolutionised sequence-based data generation. This has led to an increasing number of plant and crop genomes being made available publically, with oil palm being one of the more recent (Singh *et al.*, 2013). However, generating sufficient sequence coverage is only the start, with medium to large sized diploid genomes being difficult to assemble due to short sequence

read lengths and with islands of repetitive DNA which cannot be bridged even with a high coverage of short reads. BAC end sequencing, restriction fragment BAC or PAC fingerprinting and PacBio (www.pacificbiosciences.com) and Nanopore (www.nanoporetech.com) sequencing can alleviate some of these problems, but at a financial cost. One important tool involves anchoring sequence scaffolds using a genetic map, with markers localised to specific genetic positions and their physical equivalent on the scaffolds. However, marker density and limited numbers of recombination events underlying the genetic maps complicate this. For polyploidy genomes, these problems are even more severe and heterozygosity further complicates genome assembly.

Bambara groundnut is on the target list of the African Orphan Crop Consortium (AOCC; www.africanorphancrops.org) genome sequencing effort and a homozygous line from a Zimbabwean variety has been identified for sequencing after extensive SSR genotyping to identify a homozygous and homogeneous line. A community-based call for 100 lines to re-sequence is also underway; BamNewtork (www.bambara-groundnut.org).

Until the development of NGS, generating the genetic data was the major constraint to the population-based dissection of traits for marker-assisted breeding. The generation of large sequence datasets from expressed sequences (transcriptomes) or the genome allows the identification of markers for genetic analysis. SSRs can be easily mined from

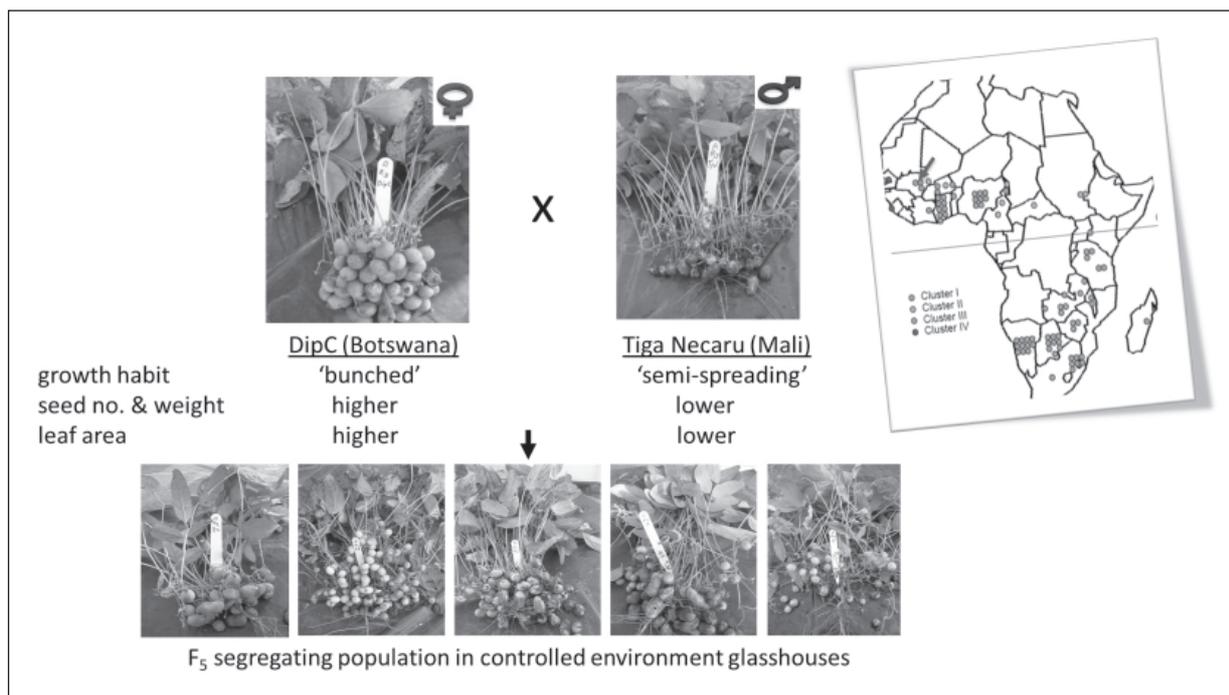


Fig. 4. Genetic diversity analysis can identify genetically distinct individuals from similar environments, allowing the introduction of new diversity while improving the chances of maintaining adaptation to the target environment.

the transcriptome and SNPs also provisionally identified. For underutilised species where an extensive investment in developing molecular genetics may not be possible, the development of Genotype-by-Sequence (GbS) approaches as exemplified by Diversity Arrays Technology (DArT; DArT Pty Ltd, Canberra; www.diversityarrays.com) allows the generation of both SNP and presence/absence (silico DArT) markers in crop genomes without previous information. GbS allows detailed diversity analyses, construction of genetic maps and QTL trait analyses in controlled cross populations;

Figure 5. However, for practical application in breeding programmes, these multi-locus marker systems must be convertible to simple, cheap and clear tests, particularly where there are no taxonomically close genomes sequences; Figure 6.

However, where taxonomically close genomes do exist, the ability to develop positional links between the species under investigation and data rich species (preferably model plants or major crops) can allow conserved synteny analysis and the potential to translate existing positional information on traits in the more studied species to the underutilised

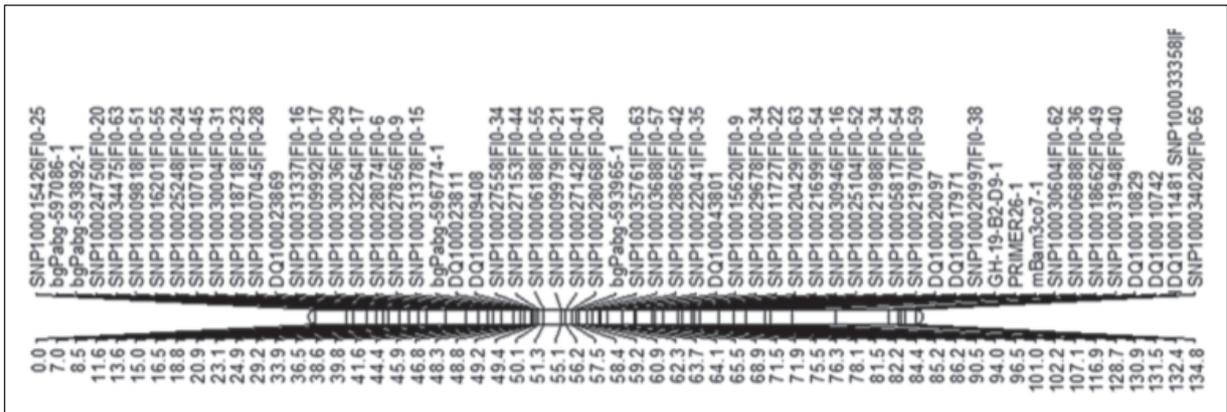


Fig. 5. Construction of genetic maps using DArT Seq markers, showing a linkage group with a mixture of dominant silico DArT and co-dominant SNP markers and transcriptome mined SSR; raw data contained 2996 SNP and over 8000 presence/absence markers.

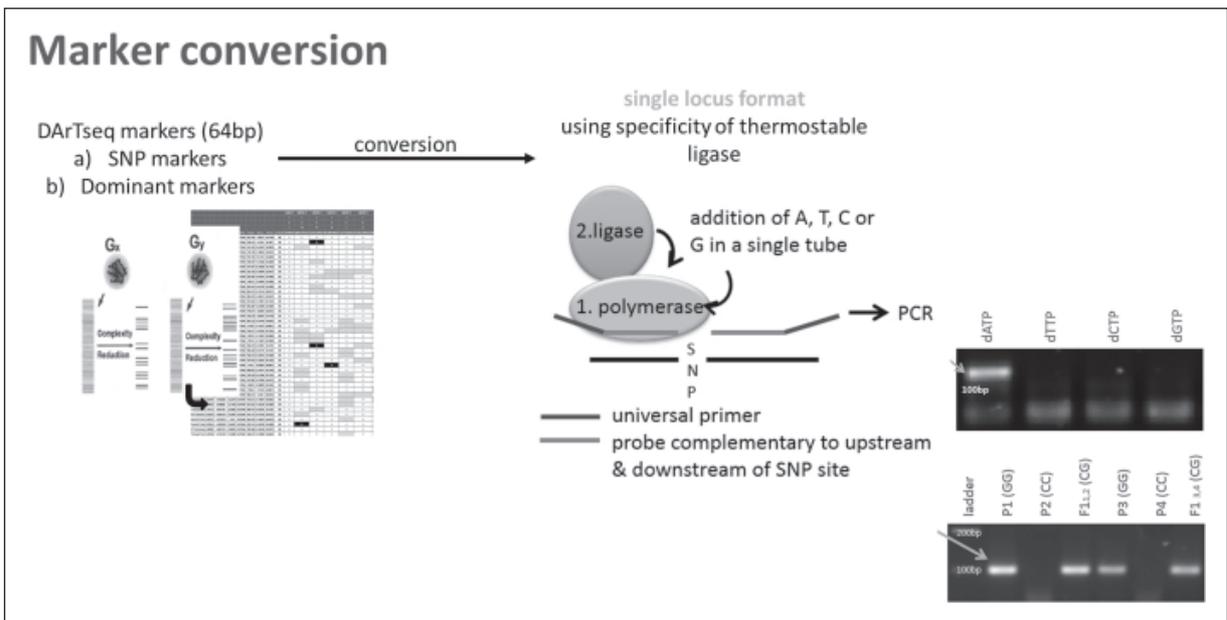


Fig. 6. High density marker development can identify loci for selection within breeding programmes, but the ability to convert flanking markers into a simple format is important for practical application. Here conversion of markers flanking a region of interest is attempted using the DArT Seq 64bp sequence tag and a primer extension-ligation test. Of the four loci tested, two could be converted to a simple agarose gel format, when the SNP difference was located within the central 12 bases of the 64bp tag, when two 21 nucleotide flanking oligonucleotide probes were used. Given the high density of markers which can be mapped with DArT Seq, the number of available flanking markers that should convert is unlikely to be a constraint.

species of study. In bambara groundnut, the segregating marker sequence tags (64bp) from the Dip C x Tiga Necaru cross were screened against the common bean (*Phaseolus vulgaris*) genome for unique, high sequence homology hits (95% minimum) using CLC Genomic Workbench v6 (Schmutz *et al.*, 2014). Those markers with such hits were used to construct the genetic map and the genetic marker order was then compared directly with the common bean physical genome; Figure 7. As the technology leads to an increase in tag length (currently around 130 bp for DArT Seq) the taxonomic distance this approach can bridge will also increase.

While not perfect, this approach potentially allows the location of genetic effects identified within bambara groundnut to be compared with the equivalent physical region in common bean or other taxonomically close species. Common bean is around 5 million years diverged from bambara groundnut. For greater taxonomical distances, the gene models associated with the DArT Seq markers (via the transcriptome used to mine SSRs) can be used to link gene orthologues (as with soybean in Figure 8, at roughly 20 million years divergence with a tetraploid genome. The links to common bean were used to extract additional sequence

surrounding the 64bp bambara groundnut tag from the common bean genome (approximately 200 bp in total) and the hybrid sequence blasted to the soybean genome.

Comparison of the DArT seq 64bp bambara groundnut tags with the transcriptome of bambara groundnut suggests that around a third of tags correspond to an expressed sequence. This is likely to be due to the use of the methylation sensitive *Pst*I restriction endonuclease to create the genomic representations for DArT seq analysis.

The high density markers generated by GbS also have the potential to be tested in Association Genetic studies and the development of a panel of 500 single genotypes representing the available germplasm of bambara groundnut is underway in collaboration with the International Institute for Tropical Agriculture (IITA), coupled with standardised trait recording (IPGRI/IITA/BAMNET, 2000). In the longer term, approaches such as Genomic Selection may be relevant, but for the moment there has been so little structured breeding in bambara groundnut that it should be possible to make major progress through selection from landrace and bi-parental crosses. One problem (or advantage) of bambara groundnut is that most of the accessions

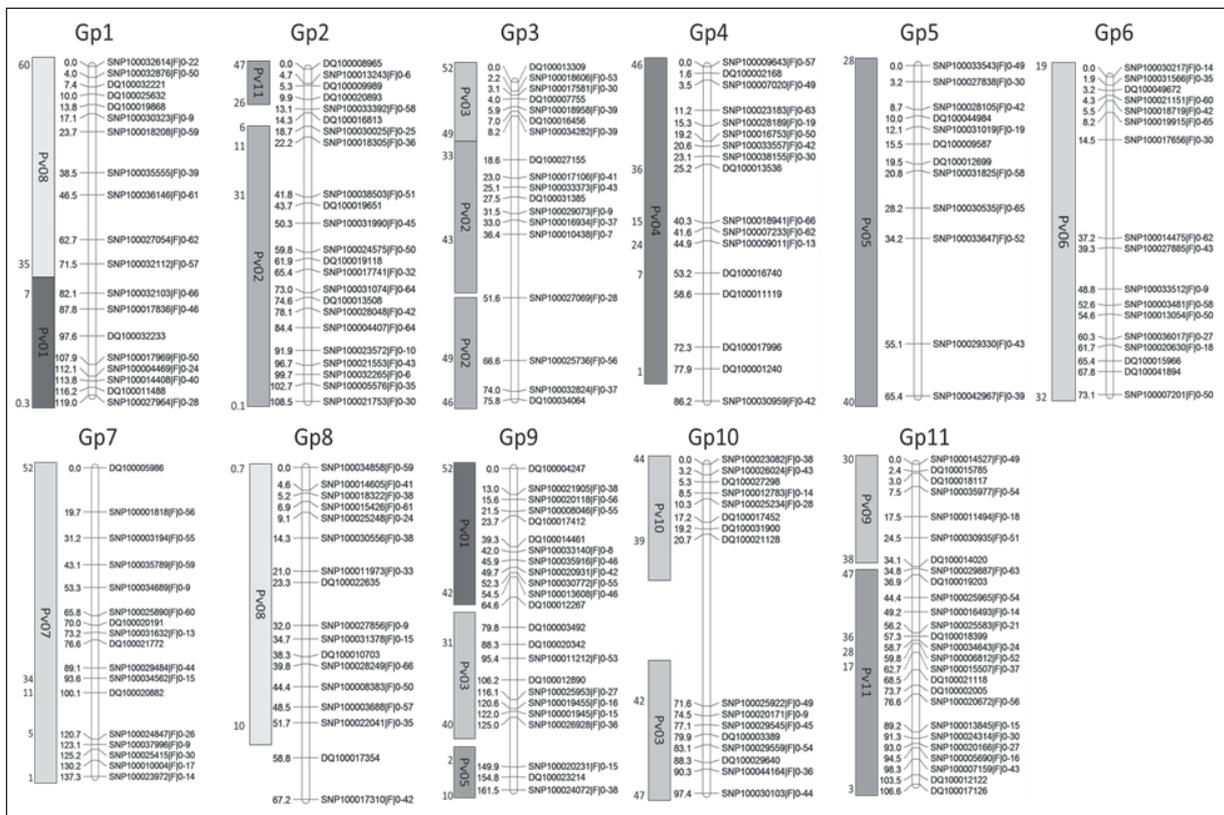


Fig. 7. Using DArT Seq marker sequence tags to cross-locate markers from a spaced genetic map in bambara groundnut onto the common bean physical genome; 171 spaced markers; 7.8cM/marker; 1341cM total. At a macro level, significant conserved synteny exists, with the 11 genetic linkage groups in bambara groundnut translating to 19 physical genome blocks in common bean.

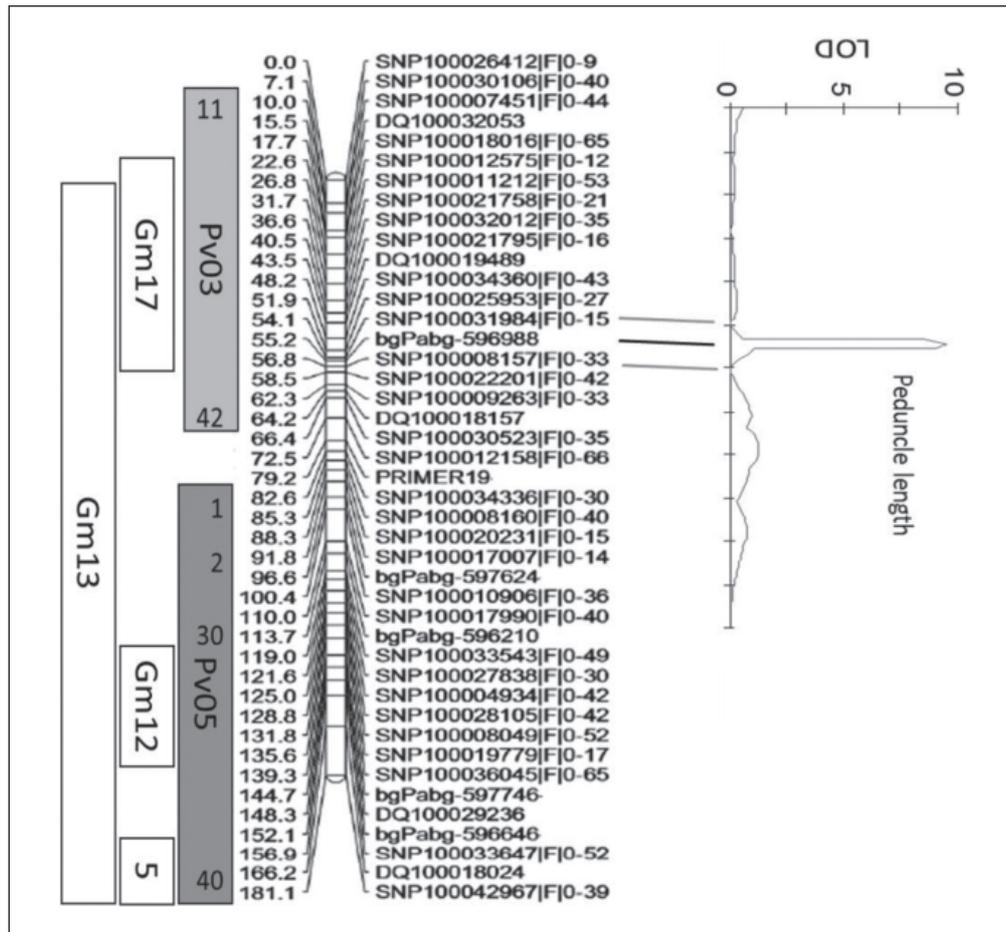


Fig. 8. Identification of a QTL for peduncle length in a bambara groundnut population and cross-location to common bean and soybean to produce a potential candidate gene list.

are genetically distinct and have been selected in particular locations for long periods of time. As such, there are no ‘elites’ or the highly structured pedigrees which influence breeders in major crops. Multiple Advanced Generational Intercross (MAGIC) populations might be one way forward to develop material for selection by farmers in multiple environments (Aliyu *et al.*, 2015).

CONCLUSION

Understanding the breeding system, available genetic diversity and effects of domestication in an underutilised species is important to be able to determine the best approach for genetic improvement. Molecular markers allow an evaluation of these to be made and the same markers can also be used to ensure genotype tracing and quality control in the breeding programme.

Translation of trait and genome information via conserved synteny will become increasingly important and also increasingly feasible, as more crop plant genomes are reported. However, there are

perhaps 7,000 plant species which have been used by humankind (Harlan, 1992) and it will still be a long time before all of these are available as sequenced, assembled and annotated genomes.

The challenges we face in the next 40-50 years are immense, as we attempt to both increase overall crop yields and drastically reduce inputs. We are far too reliant on a limited number of crops, many of which have been ‘pampered’ by and selected to respond to intensive agriculture. Staple crops will continue to be critical to future food security but we also need to be exploiting the potential present in many minor crops that have long been grown in low input agricultural systems. These complement the major crops, but also compete for research investment. Markers could allow a more efficient use of research funding to evaluate those with the greatest potential and identify the best breeding approaches available. The yields of both wheat and oil palm have quadrupled in the last century (Calderini and Slafer, 1998; Corley and Lee, 1992). This suggests similar progress can be made with underutilised crops and progress accelerated through the use of molecular markers (Mayes *et al.*, 2011).

However, simply replacing major crops with new major crops will have limited impact in a world where yield potential in major crops continues to increase but where average farm yields in a wide range of crops have stagnated (Foley *et al.*, 2011; Ray *et al.*, 2012). More appropriate use of minor crops with resilience traits and adaptation to poor environments could lead to greater agricultural production overall. In addition, an understanding of the intricate interactions in complex farming systems, rather than monoculture, may improve both resilience and sustainability. Markers have a key role to play in the future for unravelling these issues.

REFERENCES

- Allard, R.W. 1960. Principles of plant breeding. Wiley, New York.
- Aliyu, S. and Massawe, F.J. 2013. Microsatellites based marker molecular analysis of Ghanaian bambara groundnut (*Vigna subterranea* (L.) Verdc.) landraces alongside morphological characterization. *Genetic Resources and Crop Evolution* **60**(2): 777-787
- Aliyu, S., Massawe, F. and Mayes, S. 2015. Beyond landraces: developing improved germplasm resources for underutilized species – a case for Bambara groundnut *Biotechnology and Genetic Engineering Reviews* **30**(2): 127-141
- Alpern, S.B. 1992. The European introduction of crops into West Africa in precolonial times *History in Africa* **19**: 13-43.
- Alpern, S.B. 2008. Exotic plants of West Africa: where they came from and when. *History in Africa* **35**: 63-102.
- Al-Rugaishi, I.A., Davey, M., Alderson, P. and Mayes, S. 2008. Genetic relationships and genotype-tracing in date palms (*Phoenix dactylifera* L.) in Oman, based on microsatellite markers. *Plant Genetic Resources: Characterisation and Utilisation* **6**(1): 70-72.
- Azam-Ali, S.N., Sesay, A., Karikari, S.K., Massawe, F.J., Aguilar-Manjarrez, J., Bannayan, M. and Hampson, K.J. 2001. Assessing the potential of an underutilized crop: a case study using Bambara groundnut. *Experimental Agriculture* **37**: 433-472
- Basu, S., Mayes, S., Davey, M., Roberts, J.A., Azam-Ali, S.N., Mithen, R. and Pasquet, R.S. 2007a. Inheritance of “domestication” traits in bambara groundnut (*Vigna subterranea* (L.) Verdc.) *Euphytica* **157**: 59-68.
- Basu, S., Roberts, J.A., Azam-Ali, S.N. and Mayes, S. 2007b. Development of microsatellite markers for bambara groundnut (*Vigna subterranea* (L.) Verdc.) – an underutilized African legume crop species. *Molecular Ecology Notes* **7**: 1326-1328.
- Calderini, D.F. and Slafer, S.A. 1998. Changes in yield and yield stability in wheat during the 20th century. *Field Crops Research* **57**: 335-347.
- Corley, R.H.V. and Lee, C.H. 1992. The physiological basis for the genetic improvement of oil palm in Malaysia. *Euphytica* **60**: 179-184.
- Doebley, J., Gaut, B.S. and Smith, B.D. 2006. The molecular genetics of crop domestication. *Cell* **127**: 1309-1321.
- Foley, J.A., Ramankutty, N. and Brauman, K.A. 2011. Solutions for a cultivated planet. *Nature* **478**: 337-342.
- Gepts, P. 2004. Plant crop domestication as a long-term selection experiment. In: Janick J (ed) *Plant Breeding Reviews*, **24**(2). John Wiley & Sons, pp 1-44.
- Harlan, J.R. 1992. Crops and man. Madison, Wisconsin, USA: American Society of Agronomy.
- IPGRI, IITA, BAMNET. 2000. Descriptors for Bambara groundnut (*Vigna subterranea*) International Institute of Tropical Agriculture, Ibadan, Nigeria/International Plant Genetic Resources Institute, Rome, Italy.
- Kendabie, P. 2015. Unravelling photoperiod effects on pod-set and pod-filling in Bambara groundnut [*Vigna subterranea* (L.) Verdc.] – a drought tolerant African legume with potential to contribute to global food security. PhD Thesis University of Nottingham, UK.
- Mayes, S., Massawe, F.J., Alderson, P.G., Roberts, J.A., Azam-Ali, S.N. and Hermann, M. 2011. The potential for underutilized crops to improve security of food production. *Journal of Experimental Botany* **63**(3): 1075-1079.
- Mayes, S., James, C.M., Horner, F., Jack P.L. and Corley, R.H.V. 1996. The application of restriction fragment length polymorphism for the genetic fingerprinting of oil palm (*E. guineensis* Jacq). *Molecular Breeding* **2**: 175-180.
- Molosiwa, O.O., Aliyu, S., Stadler, F., Mayes, K., Massawe, F., Kilian, A. and Mayes, S. 2015. SSR marker development, genetic diversity and population structure analysis of Bambara groundnut [*Vigna subterranea* (L.) Verdc.] landraces *Genetic Resources and Crop Evolution* DOI 10.1007/s10722-015-0226-6.
- Olukolu, B.A., Mayes, S., Stadler, F., Ng, N.Q., Fawole, I., Dominique, D., Azam-Ali, S.N., Abbott, A.G. and Kole, C. 2012. Genetic diversity in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as revealed by phenotypic descriptors and DArT marker analysis. *Genetic Resources and Crop Evolution* **59**: 347-358.

- Padulosi, S., Hodgkin, T., Williams, J.T. and Haq, N. 2002. Underutilised crops: trends, challenges and opportunities in the 21st century. In: Engels JMM, Ramanatha Rao V, Brown AHD, Jackson MT, eds. *Managing plant genetic diversity*, Vol. 30. Rome: IPGRI, 323-338.
- Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C. and Foley, J.A. 2012. Recent Patterns of crop yield growth and stagnation. *Nature Communications* **3**: 1293.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, O., Thelen, J.L., Cheng, J., Xu, D., Hellsten, U., May, G.D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M.K., Sandhu, D., Valliyodan, B., Lindquist, E., Peto, M., Grant, D., Shu, S., Goodstein, D., Barry, D., Futrell-Griggs, M., Abernathy, B., Du, J., Tian, Z., Zhu, L., Gill, N., Joshi, T., Libault, M., Sethuraman, A., Zhang, X.-C., Shinozaki, K., Nguyen, H.-T., Wing, R.A., Cregan, P., Specht, J., Grimwood, G., Rokhsar, D., Stacey, G., Shoemaker, R.C. and Jackson, S.A. 2010. *Nature* **463**: 178-183.
- Schmutz, J., McClean, P.E., Mamidi, S., Wu, G.A., Cannon, S.B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S.M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M.A., Chovatia, M., Gepts, P., Goodstein, D.M., Gonzales, M., Hellsten, U., Hyten, D.L., Jia, G., Kelly, J.D., Kudrna, D., Lee, R., Richard, M.M.S., Miklas, P.N., Osorno, J.N., Rodrigues, J., Thareau, V., Urrea, C.A., Wang, M., Yu, Y., Zhang, M., Wing, R.A., Cregan, P.B., Rokhsar, D.S., Jackson, S.A. 2014. A reference genome for common bean and genome-wide analysis of dual domestications *Nature* **467**: 707-716.
- Singh, R., Ong-Abdullah, M., Low, E.T., Mana, M.M.A., Rosli, R., Nookiah, R., Ooi, L., C-L., Ooi, S.E., Chan, K.L., Halim, M.A., Azizi, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S.W., He, D., Hogan, M., Budiman, M.A., Lee, E.K., DeSalle, R., Kudrna, D., Goicoechea, J.L., Wing, R.A., Wilson, R.K., Fulton, R.S., Ordway, J.M., Martienssen, R.A. and Sambanthamurthi, R. 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds. *Nature* **500**: 335-341.
- Sri Redjeki, E.S. 2014. Thesis submitted in partial fulfilment of the an Master of Philosophy, University of Nottingham.
- Somta, P., Chankaew, S., Rungnoi, O. and Srinives, P. 2011. Genetic diversity of the Bambara groundnut (*Vigna subterranea* (L.) Verdc.) as assessed by SSR markers. *Genome* **54**: 898-910.
- Stadler, F. 2009. Analysis of differential expression under water deficit stress and genetic diversity in Bambara groundnut (*Vigna subterranea* (L.) Verdc.) using novel high throughput technologies. Dissertation, Technische Universitat Munchen (TUM).
- Varshney, R.K., Nayak, S.N., May, G.D. and Jackson, S.A. 2009. Next generation sequencing technologies and their implications for crop genetics and breeding. *Trends in Biotechnology* **27**: 522-530.
- Wong, W.C., Teo, C.J., Wong, C.K., Mayes, S., Rajinder, S. and Soh, A.C. (2015). Development of an effective SSR-based fingerprinting system for commercial planting materials and breeding applications in oil palm. *Journal of Oil Palm Research* **27(2)**: 113-127.

