Developing XSpecies Approaches for Genomics and Transcriptomics – Using Resources Developed in Major Species for Research in Bambara Groundnut

Hui Hui Chai[1], Hung-Ming Lai[1], Hui Guo[1], S.M. Basu[1], F. Massawe[2], S. Azam-Ali[3], N. Graham[1], M. Broadley[1], S. May[1] and S. Mayes[1,3]

[1] Plant and Crop Sciences, Biosciences, University of Nottingham, Sutton Bonington Campus, LE12 5RD, United Kingdom

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

[3] Crops for the Future Research Centre, The University of Nottingham, Malaysia Campus, Jalan Broga, 43500 Semenyih, Selangor Darul Ehsan, Malaysia

Keywords: underutilised crop, Affymetrix microarrays, PCR primers, molecular markers

Abstract

Bambara groundnut (*Vigna subterranea* L. Verdc.) is a legume mainly grown by subsistence and small-scale farmers in sub-Saharan Africa. This underutilised crop has nutritious and protein-rich seeds, can grow in poor soils and is more tolerant to drought compared to similar major crop species. Like Bambara groundnut, there are many minor and underutilised crops which have not undergone extensive research yet and for which few genomic resources currently exist. We aim to investigate ways to use resources that have already been developed in major species for crop improvement programs in minor and underutilised species. Affymetrix microarrays are powerful and popular tools for genomic and transcriptomic studies, but they require extensive species-specific sequence information to design. Thus, to save time and reduce cost, one option is to use an XSpecies microarray. Here we have used the XSpecies approach to detect genomic differences between two parents of a Bambara groundnut controlled cross between VSSP11 (low stem number per plant) and DipC (high stem number per plant), in order to attempt to develop molecular markers that are linked to the gene(s) controlling stem number as a test case. Bambara groundnut DNA derived from two F_2 bulks that are constructed for high and low stem number, respectively, were hybridised separately to the *Arabidopsis* ATH1 and the *Medicago truncatula* Affymetrix GeneChips. PCR primers were designed from array probe sequences based on hybridisation signals visualised in the PIGEONS.v1 software. Differences detected between the two parents and the F_2 bulks in silico are being tested in the lab to determine whether this is a useful way to generate markers to traits controlled by limited numbers of genes in underutilised and minor crop species. Results to date are presented.

INTRODUCTION

Bambara groundnut is an indigenous African legume widely grown in sub-Saharan Africa by subsistence and small-scale farmers. It represents an important source of additional protein to a large proportion of the population in these poor countries. Bambara groundnut is valued for its protein-rich seeds, drought tolerance, adaptability to poor soils and resistance to pests and diseases. It consists of two botanical forms: var. *spontanea* comprising the wild forms, restricted to Cameroon, and var. *subterranea* comprising domesticated forms, which can be found in most of the tropical areas, especially in sub-Saharan Africa. As there are no established cultivars, most of the marginal and subsistence farmers in Africa grow locally adapted landraces which can result in poor and/or unpredictable yields. A cross was made between wild (VSSP1; 'spreading') and domesticated (DipC; 'bunched') Bambara groundnut landraces. The F_1 hybrid was self-pollinated and the resultant F_2 population was used to study a range of agronomic and domestication traits, including days to emergence, days to flowering,

internode length at harvest and number of stems per plant number as reported in Basu et al. (2007). The authors suggested that a possible 1:2:1 segregation pattern is observed in the F_2 population, consistent with a χ^2 test (P=0.03<5.99; 2 df) for the stem number trait, indicating that phenotypic variation in the cross for this trait is controlled by a single locus. Together with the trait 'stem length', these two traits account for the majority of the morphological differences between the wild and domesticated parents. Thus it is a good candidate for testing when molecular approaches can be applied to further investigate the genetic basis of such simple traits and to develop potential markers for selection.

Molecular approaches have been widely used for the genetic improvement of crop plants. Nearly 100 SSR markers have been developed for fingerprinting, diversity analysis and mapping in Bambara groundnut. DArT microarray technology, Affymetrix XSpecies microarrays, QTL analysis of 'wide' and 'narrow' crosses and the development of Massively Parallel Signature Sequencing have been investigated for the genetic improvement of Bambara groundnut. This paper reports results to date from the XSpecies approach.

Sixteen Affymetrix microarrays are now available for plant species (Affymetrix, 2011) and more are coming in the near future. Microarrays have become a powerful and popular tool for genomic and transcriptomic studies. However, they requires extensive sequence and genome information to design. XSpecies microarrays have emerged, as they allow the identification of oligonucleotide targets of one species by hybridising nucleic acids onto the Affymetrix microarray derived from another species which is currently available, such as *Arabidopsis*, soybean or *Medicago*. The XSpecies approach is a combination of physical and bioinformatics methods, with an appropriate analysis after the hybridisation required in order to generate a valid result. A program, known as 'Photographically InteGrated En-suite for the OligoNucleotides Screen' (PIGEONS), was developed and used to investigate the individual oligonucleotides underlying genomic cross-species studies (Lai, 2009). There are many minor and underutilised crops like Bambara groundnut that have not undergone the extensive research yet needed to generate dedicated microarrays. XSpecies microarray hence provides an alternative option for crop improvement programmes in minor and underutilised crops through the use of genomic resources that have already been developed in major crops.

In this paper, we aim to use the XSpecies approach to detect genomic differences between two Bambara groundnut accessions (VSSP11 and DipC) in order to try to develop molecular markers that are linked to the gene(s) controlling stem number. The genomic resources from model plant, *Arabidopsis*, and model legume species, *Medicago*, will be adopted for research and breeding work in Bambara groundnut.

MATERIALS AND METHODS

Plant Materials

Two parents of a Bambara groundnut controlled cross between VSSP11 (*V. subterranea* var. *spontanea*) from Cameroon and DipC (*V. subterranea* var. *subterranea*) from western Botswana as well as an F_2 population, which was subsequently constructed into DNA bulks for high and low stem number, were used in this study.

Genomic DNA Extraction

Genomic DNA was extracted from Bambara groundnut leaf tissues using the GenElute Plant Genomic DNA Miniprep Kit (Sigma, UK). The concentration of DNA (ng/µl) was estimated using the Nanodrop 1000 (Thermo Scientific, USA) associated with ND-1000 V 3.7.0 software.

XSpecies Analysis

DNA extracted from Bambara groundnut was sent to the NASC Affymetrix chip service, Plant and Crop Sciences, Sutton Bonington, for cross hybridisation with Affymetrix *Arabidopsis* ATH1 and *Medicago truncatula* GeneChips.

PIGEONS Software

PIGEONS was used to analyse the CEL files that resulted from the XSpecies experiments in order to generate a candidate list for potential probe sets that gave high signal strength as well as those that showed differences in signal between VSSP11 and DipC. The two parents VSSP11 and DipC were selected as 'parent1' and 'parent2' while F_2 offspring constructed into bulks for 'high' and 'low' stems per plant were selected for analysis.

RESULTS AND DISCUSSION

Threshold Selection

The threshold boundary for the XSpecies analysis result from the hybridisation of Bambara groundnut DNA on Affymetrix *Arabidopsis* ATH1 GeneChip is between 80 to 160. A threshold value of 120 was chosen as it is in the middle of the suggested threshold boundaries, with a relatively high probe set retention rate, 93.75%, and 3.1 for the ratio of average probe pairs to probe sets (Fig. 1). For the results of the hybridisation on the Affymetrix *Medicago truncatula* GeneChip, the threshold value of 100, with probe set retention of 88.28% and ratio of average probe pairs to probe sets, which is 3.2, was selected (not shown). The *Medicago* chip also represents a number of genes derived from non-*Medicago* species, such as nodulation bacteria, which may account for the lower retention, despite *Medicago truncatula* and Bambara groundnut both being legumes. Threshold boundaries are useful for giving an idea of which threshold should be chosen for analysis, as well as the number of probe pairs, probe sets and ratio of average probe sets to probe pairs. It is suggested that when the threshold is increased from 0 to 1000, probe pairs are lost rapidly although entire probe sets which represent transcripts are lost relatively slowly, as only a minimum of one PM probe is required to retain a probe set (Hammond et al., 2005). It is essential to remove bad probe sets and/or probe pairs which have low threshold value to increase the specificity of cross-hybridisation.

Potential Probe Set Identification

The potential probe sets were then analysed in two stages. Firstly, at the defined threshold value, with fold change set as 2 for 'P' and 1.5 for 'F_2', potential probe sets which have high value of PM signals from Affymetrix array as well as large differences in signal between the two parents or the two F_2 bulks are identified. As an example, using data from cross hybridisation of Bambara groundnut on Affymetrix *Medicago truncatula* GeneChip, the probe set, Mtr. 25972.1.S1_at, obtained PM signals above 1000 as well as high signal difference between the parents and the F_2 bulks (Fig. 2). Secondly, an effort will be made to design PCR primers from array probe sequences on the basis of hybridisation signals that have been observed in PIGEONS software. In silico detection of the differences between the two parents and the F_2 bulks will be tested in the lab to determine whether the polymorphism is either genuine or due to background noise. Good candidates will show strong amplification and hopefully polymorphism in the region where a signal difference was detected in PIGEONS. Such differences can then be used to map putative markers for stem number. An example is primers designed from 262850_at which is suspected to be the GAI gene in *Arabidopsis* (Fig. 3).

Background noise is one of the major problems when XSpecies approach is used for molecular work in Bambara groundnut. Noise can result from the inefficient hybridisation of certain transcripts to the probes on the array due to sequence polymorphisms between two different species (Hammond et al., 2005). To reduce noise, several approaches could be adopted. One of them is to increase the number of individual plants within the bulks constructed for low stem and high stem number per plant. The original sequences of probe sets can be obtained from Affymetrix website. Sequences can then be screened via TblastX with databases and sequences that are currently available to look for sequences that align with Bambara groundnut. Primers are designed using aligned sequences from Bambara groundnut for more specific PCR amplification. For

species that are separated by large genetic distances, only the more conserved genes may be directly comparable.

CONCLUSION

Using resources developed in major crop species for research in minor and underutilised crop species is one way to develop genomic resources in species where sequence information is still limited. The recent introduction of high-throughput DNA sequencing instruments, so-called "Deep Sequencing" technologies for "next-generation sequencing" have the potential to identify polymorphisms between individuals, for gene expression analysis and mutation discovery (Mardis, 2008). Deep Sequencing can be used to produce extensive sequences which will allow the development of markers as well. These approaches are comparatively inexpensive and allow the production of millions of DNA sequence reads per instrument run. It is thus suggested that the combination of the use of deep sequencing and XSpecies may accelerate the generation of markers to traits controlled by limited numbers of genes in underutilised and minor crop species.

ACKNOWLEDGEMENTS

Thanks to Zoe Phillips at NASC for array work.

Literature Cited

Affymetrix. 2011. Affymetrix tools for genomic research. http://www.affymetrix.com (accessed 25/05/2011).

Basu, S., Mayes, S., Davey, M., Roberts, J.A., Azam-Ali, S.N., Mithen, R. and Pasquet, R.S. 2007. Inheritance of 'domestication' traits in bambara groundnut (*Vigna subterranea* (L.) Verdc.). Euphytica 157:59-68.

Hammond, J.P., Broadley, M.R., Craigon, D.J., Higgins, J., Emmerson, Z.F., Townsend, H.J., White, P.J. and May, S.T. 2005. Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. Plant Methods 1:10-18

Lai, H.M. 2009. PIGEONS: an innovative bioinformatics software to across species investigation in minor crops based on the XSpecies microarray technologies. M.Res. Thesis. The University of Nottingham.

Mardis, E.R. 2008. The impact of next-generation sequencing technology on genetics. Trends in Genetics 24:133-141.

Figures

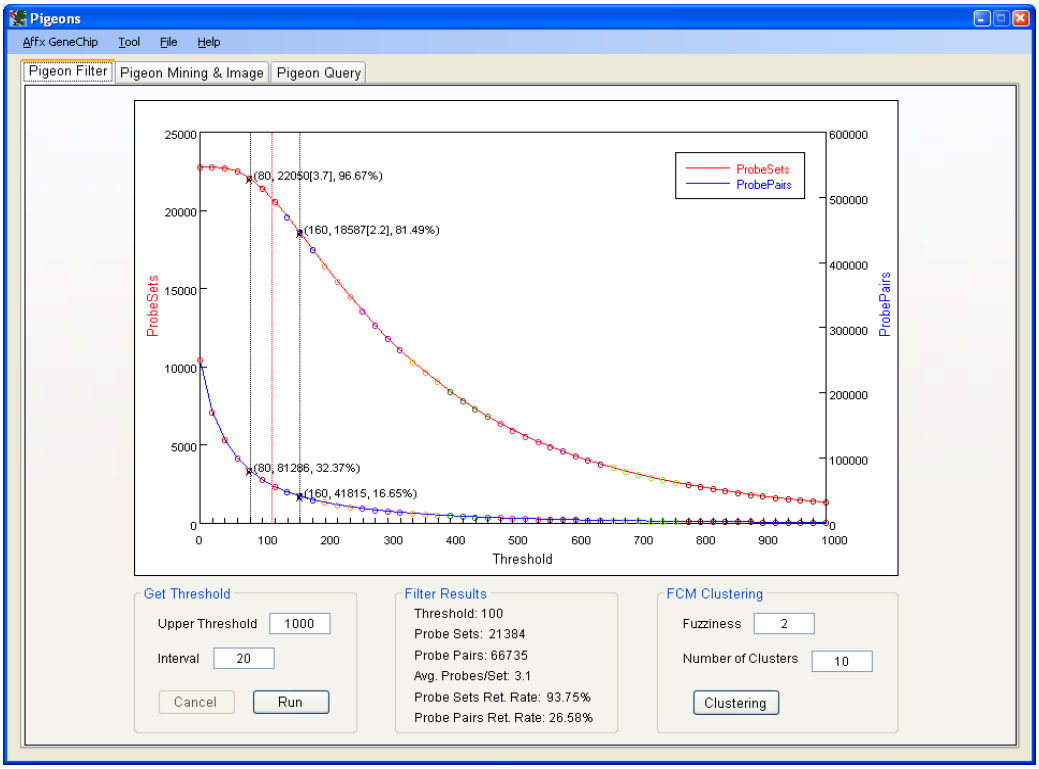

Fig. 1. Threshold boundaries for the XSpecies analysis result from the hybridisation of Bambara groundnut DNA on Affymetrix *Arabidopsis* ATH1 GeneChip.

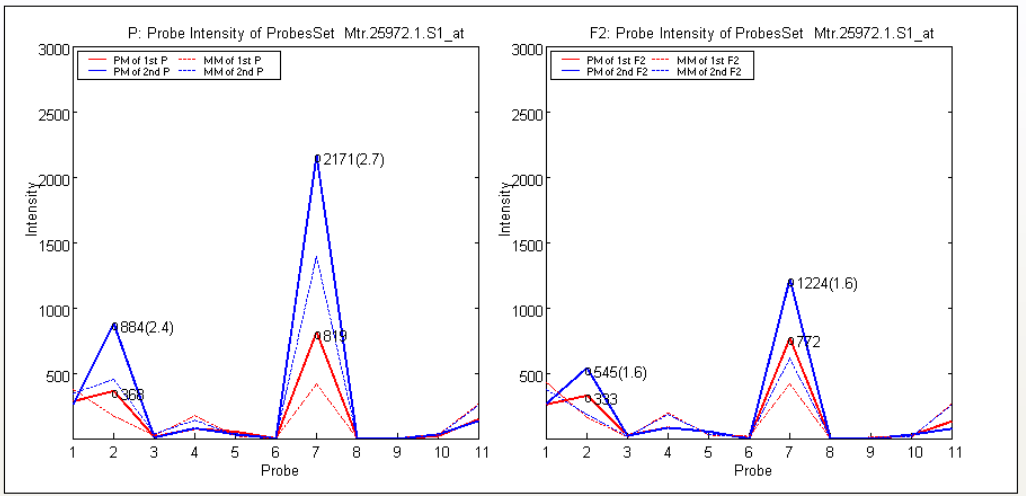

Fig. 2. The potential probe set, Mtr. 25972.1.S1_at, selected from cross hybridisation of Bambara groundnut on Affymetrix *Medicago truncatula* GeneChip.

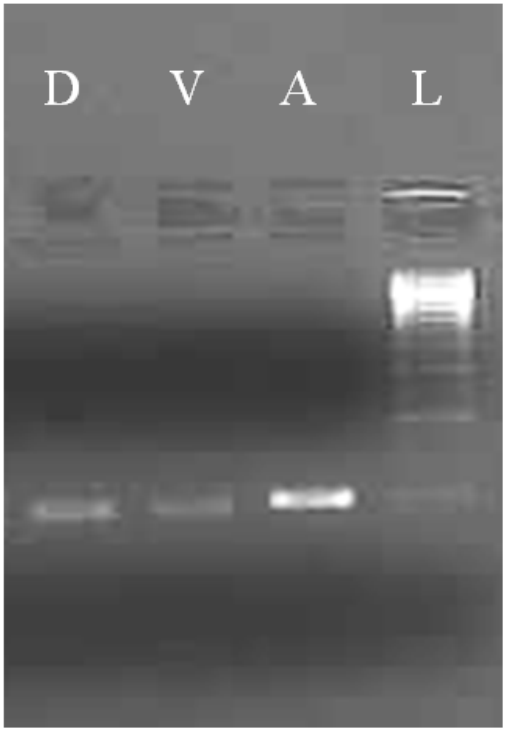

Fig. 3. Primer designed from probe set 262850_at is used to amplify Bambara groundnut DNA. L: 1 kb DNA ladder; D: DipC; V: VSSP11; A: *Arabidopsis* (positive control).