Thousands of crops and very little money – how can marker technology be applied to underutilised species?
Crops for the Future Research Centre (CFFRC)

How do we define underutilised crop?

Practically, an underutilised crop is one that does not have a major (CGIAR?) research Centre with a clear research mandate.

We also need to generate **quantitative and comparative** data with the major crops where it does not exist.
BamYIELD: Bambara groundnut as an exemplar crop for Africa and Asia
BiomassPLUS: Novel biomass crops for sustainable renewable energy and high value products
SystemPLUS: Diversifying agricultural systems using underutilised crops and cropping systems
CropBASE: World’s first web-based knowledge and decision-support system for underutilised crops
FishPLUS: Novel plant products to increase the nutritional value of aquaculture feeds
FoodPLUS: Diversification of the food basket for enhanced community nutrition and health
CFFRCPLUS/CFFRC DTC – University of Nottingham (UK/Malaysia/China)

A post-graduate (MRes and PhD) research studentship programme for an initial 300 research years

**Round 1** complete and 43 studentships approved (38 PhD and 5 MRes; 65 supervisors across five continents)

**Round 2** Deadline in 2013, a further 40 PhD and 20 MRes expected
BamYIELD - using an exemplar species to develop generic approaches

Translation of information from major crop and plant model species to Bambara groundnut is becoming feasible, but the methodologies are not fully developed.

The balance needed between within Bambara groundnut analysis and the value of translational analysis is unclear and often untested.

The absence of a coordinated international trials programme for Bambara groundnut prevents the development of new material and the utilisation of gained research knowledge.

The potential of Bambara groundnut from a product, market and socio-economic perspective has yet to be fully investigated.
African genetic diversity; Stadler
Indonesian genetic diversity; Endah Redjeki
Genetic mapping and domestication; Nariman Ahmad
Bam1-006; MAGIC breeding; Aliyu Slise
Photoperiod requirement; Kendabie
Bam1-009 Fertility in bambara groundnut; Bhavya Dhanaraj

Heat tolerance; Al-Shareef
Cold tolerance; Noah
Bam1-001; productivity and N\textsubscript{2} fixation; Mukht Musa
Bam1-003; adaptation to different environments; Philip Cleasby
Bam1-008; Canopy modelling for ideotypes; Josie Dodd

Drought response expression;
Florian Stadler
Responses to drought;
Yusuf Muhamamd
Genetical genomics of drought;
Hui Hui Chai

Overlaying the BG transcriptome onto Soybean (MSc); Faraz Khan
Linking genetic maps of BG with physical maps of Soybean; Hui Hui Chai
Bam1-004 Machine Learning to detect drought tolerance pathways; Venkata Suresh
Bam1-007 Combining multiple data types; Faraz Khan

Translation projects

Recent completed projects
Current on-going projects
BamYield agreed projects
International trial programme:

Thailand
Malaysia
Indonesia

Ghana
Niger/Niger (IITA)
Botswana
South Africa

Crossing, controlled environment glasshouses
UK
## Applications of markers in NUS – an example focused around Bambara groundnut

<table>
<thead>
<tr>
<th>TRANSCRIPTOME SEQ; MISA.pl script results</th>
<th>Oil palm</th>
<th>Bambara groundnut</th>
<th>Lablab</th>
<th>Winged Bean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of sequences examined:</td>
<td>62,337</td>
<td>51,272</td>
<td>49,395</td>
<td>84,892</td>
</tr>
<tr>
<td>Total size of examined sequences (bp):</td>
<td>34,596,101</td>
<td>34,513,803</td>
<td>34,862,628</td>
<td>55,972,781</td>
</tr>
<tr>
<td>Total number of identified SSRs:</td>
<td>1,787</td>
<td>1,187</td>
<td>4,465</td>
<td>3,857</td>
</tr>
<tr>
<td>Number of SSR containing sequences:</td>
<td>1,641</td>
<td>1,110</td>
<td>3,900</td>
<td>3,561</td>
</tr>
<tr>
<td>Number of sequences containing more than 1 SSR:</td>
<td>99</td>
<td>61</td>
<td>489</td>
<td>269</td>
</tr>
<tr>
<td>Number of SSRs present in compound formation:</td>
<td>126</td>
<td>68</td>
<td>243</td>
<td>152</td>
</tr>
<tr>
<td>Developed polymorphic microsatellites</td>
<td>n/a</td>
<td>103</td>
<td>50+</td>
<td>to do</td>
</tr>
</tbody>
</table>

Number of gene models: 62,337, 51,272, 49,395, 83,084 (Newbler assembly from tissue: Mesocarp, Leaf, Leaf, Leaf/shoot/flowers)

**Quality control**

- Hybrid F1
- Parent 1&2

**Simple agarose system**

Confirmation of hybrids

**Horizontal acrylamide system**

Fingerprinting and quality control, anchoring map markers,

**Breeding programme based QC needed**
Genetic diversity

20 SSRs amplified from 35 landraces (3 accessions per landrace); UPGMA tree

Average heterozygosity for single genotype accessions from 121 landraces = 0.01

Develop ‘single genotype’ accession lines (aim 400-500)
Clear differentiation between West and South+East African accessions

SSR-based genetic diversity analysis (PCA)

Crosses to introduce new variation could be made between agro-ecologically matched accessions but from different breeding groups
What is the relationship between genetic and morphological/trait diversity estimates for selection of breeding parent?

*Can we sample variation in genes known to be involved in morphology/traits – a ‘meaningful’ genetic diversity index?*
Within species marker data – genotyping by sequence – DArT Seq

Marker Expression data generated using the Soybean Affymetrix microarray
3 linkage groups based on DArT and SSR; QTLs shown; Dip C (Botswana) x Tiga necaru (Mali); 67 lines

Generate dense genetic maps through an outsourced service provider/collaborator

Composite linkage groups, DArT, SSR and DArT Seq.; raw 2996 SNP and 8500 +/- markers
Relate the NUS genetic map to close relatives and model for ‘pseudo’ physical map

Drawing links through cross-species array maps

Drawing links through the DArT seq. markers which link to sequence via 64bp

oil palm – 22 of 30 markers close to trait could be placed uniquely on the same scaffold

Date palm – 49% of DArT Seq Tags had unique hits in genome sequence

Bambara groundnut – 17% of Tags uniquely identified a gene model from the transcriptome
<table>
<thead>
<tr>
<th>Within species tool generation</th>
<th>Cross-species species tool generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Within house’ tool use</td>
<td>‘Within house’ tool use</td>
</tr>
<tr>
<td>‘Service’ tool use</td>
<td>‘Service’ tool use</td>
</tr>
<tr>
<td>Within species tool generation</td>
<td>Cross-species species tool generation</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>‘Within house’ tool use</td>
<td>‘Within house’ tool use</td>
</tr>
<tr>
<td>4. Development of SSRs for F1/breeding system evaluation/breeding programme QC</td>
<td>8. Genetically located gene candidates</td>
</tr>
<tr>
<td>5. Conversion of markers (via gene models/DArT tags?) to single locus for MAB</td>
<td>9. Cross species, genetical genomics (e.g. Affymetrix)</td>
</tr>
<tr>
<td>6. Linkage of in-species genetic maps to model and major crop species</td>
<td></td>
</tr>
<tr>
<td>7. Parental selection and ideotypes</td>
<td></td>
</tr>
<tr>
<td><strong>Within species tool generation</strong></td>
<td><strong>Cross-species species tool generation</strong></td>
</tr>
<tr>
<td>‘Service’ tool use</td>
<td>‘Service’ tool use</td>
</tr>
<tr>
<td>1. Leaf transcriptome sequencing</td>
<td></td>
</tr>
<tr>
<td>2. ‘Genotype by sequencing' diversity and mapping</td>
<td></td>
</tr>
<tr>
<td>3. Other genetic analysis – GWAS, Genomic selection</td>
<td></td>
</tr>
<tr>
<td>10. Next gen sequencing genetical genomics,</td>
<td></td>
</tr>
</tbody>
</table>
Overall conclusions:

1. Biotech has the potential to make a major contribution to NUS, but it should only be used to where there is a clear application and the same cannot be achieved by a good Breeding programme.

2. Some tools and information need to be developed in species, to be robust and allow accurate assessment and QC of breeding and trials programmes.

3. Some cross-species tools may be directly valuable, but an information framework that allows translation of data from major crops and model species is critical for NUS.

The current approach of:

- transcriptome (SSRs-20+; Linkage across species using gene models)
- DArT Seq Markers (Diversity analysis (400?), mapping (200), links to gene models.)

Current costs around $15K per species and is becoming cheaper.
Funding: Government of Malaysia, University of Nottingham (UNiM)
EU, National government scholarships scheme

SSR development: Katie Mayes, Ozie Molosiwa, Nariman Ahmad
DArT Array: Florian Stadler, Andrzej Kilian
DArT Seq: Nariman Ahmad, Andrzej Kilian
DArT Array/SSR diversity: Florian Stadler, Ozie Molosiwa
Mapping: Rakhi Basu, Nariman Ahmad, Hui Hui Chai
QTL analysis: Nariman Ahmad
Genetical Genomics: Hui Hui Chai
Photoperiod and crossing: Presidor Kendabie
LabLab: Eliezah Kamau, Pavithravani, B.V. (Kirkhouse Trust)
Winged Bean: Quin Nee Wong