Exploring and Exploiting Genetic Diversity in Sorghum

David Jordan
Graeme Hammer
Queensland Alliance for Agriculture and Food Innovation
Emma Mace
Department of Agriculture Fisheries and Forestry
Structure of the Presentation

1. GxExM framework for crop improvement
2. Sorghum BCNAM population
   • Motivation and history
   • Description of the resource
   • The concept of NAM
   • Example mapping a quantitative trait
Our Challenge

• To contribute to improving yield and managing risk in a rain fed crop grown in a highly variable climate

> 10t/ha

> 2 t/ha
We are trying to identify favourable combinations of varieties and management practices in a complex system where the resources available to search for these combinations are limited.
The type of environment experienced by a crop grown at particular location varies from season to season.

We can’t change this!
Many different management systems are possible (planting dates, fertilizers, spacing populations…….)

The value of a particular system depends on the genotypes grown in it and the pattern of environments experienced by it.

We can change them and select the best!
Genotypes (varieties)

Plant Breeding Research

- Many different varieties are possible
- The value of a particular genotype depends on the management system it is grown in and the pattern of environments experienced by it.

We can change them and select the best!
Genomic and crop physiology research
• Genes do not act in isolation.
• Different combinations of genes can result in the same phenotype.
• Interaction between genes occurs at different scales within networks and between sub-traits to produce integrative traits (eg yield)
• The value of a particular gene in an ExM depends on the other genes it is combined with

We can modify networks and select the best!
Constraint imposed by interaction (G x E x M)

- Resource constraints limit our capacity to investigate large numbers of combinations
  - Agronomists evaluate different types of management using a limited range of varieties (ExM)
  - Plant breeders evaluate large numbers of varieties using a limited number of management systems
  - Genomic researchers and trait biologists are only able to look at gene networks in a limited number of genetic backgrounds, can change only small numbers of genes in networks and often use systems that have little linkage to real environments, management systems or varieties
Disconnect between the development of management systems and genotypes

- Changes in management systems may permit or compel changes in varieties
- The move to skip row sorghum systems may alter the sorts of genotypes that can be grown
Disconnect between the development of genotypes and management systems

- Changes in varieties may permit or compel changes in management systems.
- Development of midge resistant sorghum allowed greater flexibility in sorghum planting time.
Disconnect between detailed trait biology and genetics research and performance in real world systems

- Changes in particular traits may have positive or negative impacts on yield depending on the context (GxExM)
Context Dependencies

The large number of permutations and combinations of G,M & E means that plant breeders, agronomists and trait biologists are forced to work on components of the system isolation and to ignore most interactions.

The interactions turn out to be very important.
US Maize yields have more than doubled in 70 years.

Progress of 2.5% per year

This has been due to a combination changes in management systems and genetics.
How do we change the game?

• Employ an integrated approach using “linking technologies” to build bridges across scales and disciplines.
Crop simulation modelling with APSIM gives us the capacity to:

- Understand the types of environments experienced by crops in a target environment and their frequency
- Predict the performance of specific genotypes in various environments and management combinations
- Explore GxExM combinations in silico before investigating the “best bets”
- Estimate the impact of changes in particular genes or traits in different environments and management combinations
- Dissect complex traits into component traits that may be more amenable to breeding and mapping
Crop simulation Modelling

What if questions

- What type of architecture would work best in my current environments and management systems?
- What combination of variation in root angle and row spacing would give the best yields on average at a particular location? What happens if I plant earlier?
- What is the likely variation (risk) associated with growing the best combination?
Linking Technologies

Genome resources, high throughput mapping comparative genomics

- Allow us to accurately locate genes for particular traits
- Allows us to determine which version of a particular gene is carried by a particular genotype
- Allows us to identify the frequency of particular genes in breeding populations and germplasm collections
- Allows us to link genotypes used in trait biology experiments or other breeding programs to genotypes grown in our breeding program
- Allows us to dissect complex traits into component traits that may be more amenable to breeding
Whole genome marker scans and genomic technologies

Trait Biology
Root angle varies in sorghum germplasm

Genome resources maps, markers, genes

Breeding program
Genotypes and phenotypic data

Questions we can answer

• Are the genes for root architecture segregating in my breeding program?
• Am I selecting for particular root architecture?
• Are there other sources of the trait I should look at?
• What impact does a particular gene for root architecture have in a specific environment?
• Does variation in these genes have different effects in early flowering compared to late flowering genotypes?
Sorghum Core Breeding Program
Sorghum Core Breeding
Program Background

- Long term pre-breeding program (~70 years) (QLD gov now QLD gov & UQ QAAFI)
- Emphasis of the program has changed over time
  3. Germplasm enhancement, parent development and integrated trait discovery (1995 to the present)
Sorghum Core Breeding Program Background

• **Main traits**: midge resistance, stay-green drought, lodging, grain yield

• **Secondary traits**: grain quality, ergot resistance, small effort in forage and biomass

• **History of success**
  – 100% of Australian sorghum hybrids genetics from the program,
  – >80% of hybrid seed sold currently has 1 parent from the program
  – 1300 lines licensed to the private sector
  – Germplasm from the program is used widely in commercial breeding programs internationally
An Integrated set of Genetic Resources and Information Resources for Trait Dissection

Databases
phenotypes, environment characterisation, pedigrees, marker genotypes, sequence data

EMS Mutant Population
Breeding Populations

Diversity Panel

Sequenced Genomes

Mapping Populations

NAM Population

Elite × Exotics
Diversified Gene Pool/ BCNAM
Motivation and History
Genetic Diversity Problem in Australian Sorghum

Sample of elite MR SG DAFF Germplasm (1990+)

Grain Sorghum Females (B lines)
Grain Sorghum Males R lines
DAFF B lines
DAFF R lines
Causes of Diversity Loss

• Selection for quantitative traits (MR, SG, other agronomic)
• Small effective population size in breeding populations
Causes of Diversity Loss

- Selection for quantitative traits (MR, SG, other agronomic)
- Small effective population size in breeding populations

Figure 1. The Relationship Between Midge Resistance And Genetic Distance Between 25 Sorghum Hybrids And A Standard Resistant Hybrid (MR51).

Figure 2 The Relationship Between Midge Resistance And RFLP Heterozygosity In 26 Sorghum Hybrids.

How do we increase the genetic diversity of elite germplasm while retaining adaptive traits?
Context Dependency

• There is no absolute relative merit value for a gene or trait or genotype

• The value of a gene always depends on the context (genetic, environment, management)

• Context dependencies occur at all scales and generate major challenges for crop improvement programs (eg epistasis, heterosis, GxExM)
Example 1 Influence of Tillering and Maturity on Yield at a Single Location
Example 2 Heterotic Pools

- Selection within hybrid breeding programs generates strong genetic structure.
- Genetic context can be important to consider when designing populations to dissect complex traits with a view to improving hybrid performance.

**Diversity Analysis of DAFF/QAAFI Germplasm ~2010**

<table>
<thead>
<tr>
<th>Axis</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16.86</td>
</tr>
<tr>
<td>2</td>
<td>10.89</td>
</tr>
</tbody>
</table>

Factorial analysis: Axes 1 / 2

**Female parents**

**Male parents**
### Example 3 Relationship between Stay-green and Yield

**Male Testers x Environment**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R931945-2-2</td>
<td>R986087-2-4-1</td>
<td>R940386</td>
<td>R993396</td>
<td>R995248</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R931945-2-2</td>
<td>R986087-2-4-1</td>
<td>R940386</td>
<td>R993396</td>
<td>R995248</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R931945-2-2</td>
<td>R986087-2-4-1</td>
<td>R940386</td>
<td>R993396</td>
<td>R995248</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R931945-2-2</td>
<td>R986087-2-4-1</td>
<td>R940386</td>
<td>R993396</td>
<td>R995248</td>
</tr>
</tbody>
</table>

**Yield (t/ha)**

-2 -1 0 1 2

**LSN**

-2 -1 0 1 2

### Environment

#### Male Tester
Example 3 Relationship between Stay-green and Yield
Male Testers x Environment

Same environment different male testers
Example 3 Relationship between Stay-green and Yield
Male Testers x Environment

Same male different environments
Diversified Gene Pool/BCNAM
Design Parameters For Diversification Activity

1. We needed to increase genetic diversity by crossing to our elite lines with exotic material

2. We needed to maintain existing adaptive complexes (height, maturity, stay-green midge resistance etc)

3. We needed to be able to evaluate genes in a genetic context that was relevant to our environments, management systems and populations

4. We wanted to develop material that could be used in breeding relatively quickly

5. We wanted to generate populations where marker information could be used to add value and contribute to biological understanding
Description of the Resource
Diversified Gene Pool - Breeding & Testing

In each generation select for height and maturity but aim for minimal other selection.

Produce ~ 60 testcross hybrids for each pedigree
Evaluate across TPE sites

New RP last 5 years

Why Backcross

- Lines are theoretically 75% on average elite parent if no selection
  - reality was ~80%
  - can recover enough lines with appropriate height and maturity
  - elite alleles for important traits are at high frequency

- Simplifies the genetics and helps identify genes that contribute to yield in an elite background
  - contribution of exotic x exotic epistasis is reduced
  - easy to generate isolines if required

Box plot of the distribution of the percentage of non recurrent parent genome present in a sample of lines from 9 of the populations
Example

S. arundinaceum

F1

Backcross progeny
Selection of Germplasm

• A range of strategies were used to choose the exotic germplasm including:
  
  – visual phenotypic diversity,
  – phenotypic extremes from published or unpublished studies
  – elite lines from breeding programs in other countries
  – geographic diversity
  – racial diversity
  – fertile wild species.
36 of these parents have been resequenced.

Black Dots CIRAD core collection, Red dots mostly BCNAM NRPs.
Evaluation of populations

Part of an ongoing pre-breeding activity

• Each year we evaluate 20 populations in multi-environment trials (yield flowering time, height, head shape, stay-green…….)

• 15 new populations and 5 previous populations (allows analysis as a single experiment)

• Environment characterisation to allow simulation modelling in more recent trials
Height of BCNAM Populations Compared With The Reference Parent (F1 hybrid data)
Maturity of BCNAM Populations Compared With The Reference Parent (F1 hybrid data)
Yield of BCNAM Populations Compared With The Reference Parent in two different environment types (F1 hybrid data)
The Resource

- > 4000 lines
- >100 exotic parents evaluated
- DArT marker data on ~1300 lines (plan to GBS 1000+)
- Data from more than >50 trials (>2M phenotypic data points)
- Some trials have EC for crop simulation modelling
- The reference parent, the female testers and 26 exotic donors have been re-sequenced
- >500 lines licensed to commercial companies

Concept of NAM
Conventional QTL Analysis

• Linkage analysis uses recent recombination between two different plant lines to identify QTL associated with a trait

Advantages
• High levels of LD means that few markers are required to get genome wide coverage
• High level of allele replication means that statistical power is high

Disadvantages
• Low levels of resolution due to limited number of events (high LD)
• Alleles that can be identified are limited to those present in the populations (ie 2 alleles at each locus and only a proportion of loci are polymorphic)
Association Mapping

• Association mapping uses historic recombination and is performed by scanning a genome for SNPs in linkage disequilibrium with a trait of interest.

• Advantages
  – Low levels of LD mean that genes can be mapped with high resolution
  – Association mapping can sample a large number polymorphic loci and multiple alleles at each loci

• Disadvantages
  – Only QTL contrasting between the parents can be detected
  – False associations are common and may be associated with populations structure
  – Genetic background effects may be large and compromise phenotyping and detection
  – Power to detect is low for alleles that are rare
Nested Association Mapping

• Nested association mapping combines the power of linkage analysis with the high resolution of association mapping Yu et al. (2008).

• A range of different genetic designs are possible
  – Reference design where a wide range of exotic parents are crossed to one or more common parents
  – Various subsets of diallele designs, chain crosses or nested designs

• Our diversified gene pool populations are a form of NAM design “BCNAM”
Sequence Data

- To do the association part of NAM you need high density marker data or resequencing

- Australian sorghum re-sequencing partnership consortium (UQ, DAFF-Q, GRDC and BGI)
  - Sequenced 44 genotypes including 26 BC NAM parents
  - Sequence depth 16x-45x (average 22x)
Opportunity to Exploit Diversity

Number of SNPs Across Three Groups of *S. bicolor* Resequenced Genotypes

- **Improved inbreds:** 8% (458865)
- **Wild and weedy:** 34% (1919232)
- **Landraces:** 18% (1026333)
- **7% (942751)**
- **59% of the SNPS in our sample were unique to land races and or wild & weedy**
Example Mapping A Quantitative Trait
QTL Analysis in BCNAM

Elite x Exotics

Centre for Plant Science
Flowering Time

- Flowering time in plants is a complex trait
- Multiple QTL
- Most RIL mapping studies in sorghum find 3-5 QTL
Flowering QTLs in a Standard RIL Population

- Study by Parh 2005
- 130 RILs
- 2 locations
- 3 QTL detected

Mapping Flowering Time in the BCNAM

• Selection removed major effect genes for photoperiod sensitivity

• Phenotypic data from 20 trials 10 locations 24 BCNAM families (~1300 individuals)

• Even though the population had been selected for constrained flowering time we identified 40 QTL for flowering time
  – 3-5 alleles at each locus
  – allele effects were all less than 2 days
  – strong consistency with QTL identified in maize by Buckler et al. (2009)
QTL for Flowering Time

QDTF_NAM_10_61
Structure of Allelic Variation

- The number of QTL segregating in each population varied between 2 and 18 average >10
- Most genotypes have a mixture of early and late alleles
Maize and Sorghum

• Sequence mapped significant markers associated with maize flowering from the Buckler et al (2009)

• 65% of the significant markers identified for flowering in maize NAM were within 10cM of the mid-point of at least one sorghum QTL detected in this study. Many were within 1cM!

Zm Chr1 (50-200cM) SBI-01 Zm Chr5 (0-75cM)

* * *

* * *

* * *

* * *

Zm Chr9 (40-100cM)

Maize and Sorghum

• Sequence mapped significant markers associated with maize flowering from the Buckler et al (2009)

• 65% of the significant markers identified for flowering in maize NAM were within 10cM of the mid-point of at least one sorghum QTL detected in this study. Many were within 1cM!

Zm Chr1 (50-200cM) SBI-01 Zm Chr5 (0-75cM)

* * *

* * *

* * *

* * *

Zm Chr9 (40-100cM)
Maize and Sorghum

Maize and sorghum provide mutual insight

Mace et al (2013) Supermodels; sorghum and maize provide mutual insight into the genetics of flowering time. TAG accepted
Flowering Study
Conclusions

• Identified a large proportion of previously identified sorghum flowering time QTL

• High degree of correspondence between flowering time QTL identified in sorghum and maize NAM populations; (50 out of 76 identified by Buckler et al 2009)

• Evidence for an allelic series at most flowering time QTL
BCNAM Where to Next

• BCNAM will form will be a major element of our future breeding and trait dissection activities

  – Use GBS to increase marker density and impute the sequence data
  – For some traits gene level resolution will allow us to look for sequence variation underlying allelic variation
  – Develop and implement new strategies for mining diversity
  – High throughput phenotyping for trait discovery
  – Integration analysis and extension of information via gene to phenotype modelling
Future Breeding Program

• Opportunity to completely change the way we do plant breeding by using new technologies to enhance integration across disciplines
Acknowledgements

University of Queensland
Graeme Hammer
Erik van Oosterom
Andrew Borrell
Vijaya Singh

Queensland Government
Emma Mace
Alan Cruickshank
Colleen Hunt
Sorghum Breeding Team

CSIRO
Scott Chapman