PhD: Mapping QTLs for Resistance to Downy Mildew in Pearl Millet

Centre for Arid Zones Studies, University of Wales, UK
Cambridge Laboratory, Norwich, UK
ICRISAT, India

Funded by the Overseas Development Administration (ODA)
Genetics is reborn with molecular markers

Mendelian Factors Underlying Quantitative Traits in Tomato: Comparison Across Species, Generations, and Environments

Andrew H. Paterson, Susan Damon, John D. Hewitt, Daniel Zamir, Haim D. Rabinowitch, Stephen E. Lincoln, Eric S. Lander, and Steven D. Tanksley

ABSTRACT

As part of ongoing studies regarding the genetic basis of quantitative variation in phenotype, we...
Pearl millet & downy mildew

- Staple food crop of semi-arid tropics
- Late 1960s male sterile lines developed
- 1971: first of several devastating downy mildew epidemics
- Annual yield reduced from 13 to 3 mill. metric tonnes
- Little know about genetics – quantitative
Origins of pathogen populations used in project

- Senegal
- Niger
- Nigeria
- India
Glasshouse seedling screens for pearl millet downy mildew

In Wales against Indian and African pathogen populations

In ICRISAT, India against local pathogen populations
LGD x ICMP 85410

Map distance (cM)

India, Niger, Nigeria, Senegal
The same QTLs are effective under differing disease pressures...
Resistance – QTLs of major and minor effect?

- Minor QTLs due to the effects of modifier genes?
- Minor QTLs due to race-specific genes to pathogen avirulences at low frequency in the pathogen population
Generating single zoospore isolates

Newly germinated seedlings

Inoculate root with droplet containing '1' zoospore

Incubate overnight

Lift agar with seedlings into pot & cover

Look for infected plants after two weeks
Nigerian single zoospore isolate-1

LG1&2

LG3

LG4

LG5

LG6

LG7

Nigerian pathogen population

Nigerian single zoospore isolate-1
Downy mildew ‘susceptible’ phenotypes

‘Tall/chlorotic’

‘Stunted’
Nigerian single zoospore isolate - 2

LG1&2

LG3

LG4

LG5

LG6

LG7

% downy mildew  % tall / chlorotic  % stunted
Nigerian single zoospore isolate - 2

Phenotypes of parental genotypes at the QTL

<table>
<thead>
<tr>
<th>Phenotypic Measure</th>
<th>AA (sus par)</th>
<th>BB (res par)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% DM</td>
<td>32.8</td>
<td>25.0</td>
</tr>
<tr>
<td>% DM tall</td>
<td>13.0</td>
<td>0.0</td>
</tr>
<tr>
<td>% stunted</td>
<td>8.0</td>
<td>28.6</td>
</tr>
</tbody>
</table>
Main Project Delivery....

• Greater food security to ~2 million people, who grew the previously popular but DM susceptible variety HHB 67
Mapping Resistance to 
*Phytophthora infestans* in 
Tomato

University of California Davis
Causes the serious potato and tomato disease **late blight**

In potato, race–specific R genes lasts 2–3 years

Breeding programs now concentrate on selecting for ‘general’ or ‘field’ resistance
  ◦ Generally quantitative and governed by many genes

**Resistance found in wild species** eg for tomato *L. hirsutum* (S. habrochaites)
Crosses and phenotyping

$L. \text{hirsutum} \times L. \text{esculentum}$

Phenotyped with detached leaf & 4 isolates, whole plants (in growth cabinet) and field screens

$L. \text{hirsutum lines found to be resistant to over 100 isolates; Mike Coffey, UC Riverside}$
QTLs were of small effect and environmentally variable

- Many QTLs mapped to plant architecture traits especially determinacy, flowering, canopy
One QTL associated with droplet dispersal

All of these leaves have been inoculated with a droplet of inoculum

*L. esculentum*

Droplet has not dispersed – leaf becomes susceptible

*L. hirsutum*

Droplet rapidly disperses – leaf is resistant
Droplet dispersal due to trichome differences?

Plant Glandular Trichomes as Targets for Breeding or Engineering of Resistance to Herbivores
NILS generated with combinations of 4 QTL
2013 update:
Linkage Relationships Among Multiple QTL for Horticultural Traits and Late Blight (*P. infestans*) Resistance Introgressed from Wild Tomato *Solanum habrochaites*
J. Erron Haggard, Emily B. Johnson, and Dina A. St. Clair

- QTL colocated or were linked to QTL having negative effects on horticultural traits, such as plant height, plant shape, maturity, yield, and fruit size
Pioneer Hi-bred and Syngenta
Evolution of Marker Systems

Molecular Marker Systems

Isozyme

RFLP

SSR

SNP

85 90 95 00 05
Markers being effectively used for major gene effects/ transgene introgression
Selecting minimum marker sets for variety discrimination and seed purity

- Travelling Salesman Problem (TSP): a mathematical problem in which one tries to find the shortest route that passes through each of a set of points once and only once.
Selected the minimum number of SNPs that could discriminate EVERY inbred in a set

15 SNP markers could discriminate among 400 inbreds

![Graph showing the relationship between the number of markers and the number of inbreds. The x-axis represents the number of inbreds ranging from 0 to 500, and the y-axis represents the number of markers ranging from 0 to 18. The graph shows a positive trend, indicating that as the number of inbreds increases, the number of markers also increases.]
Comparison of isozymes and SNPs

- Standard set of 15 isozymes

- 16 SNPs selected with TSP
  - At least 1 on each chromosome

- SNPs had higher level of missing data (2%) compared with isozymes (0.8%)
SNPs had 16 X resolution score compared with isozymes

8 inbreds compared to a reference set of 438 inbreds

- SNPs maintain power of discriminatory levels in the face of missing data and even mis-scored data
Using Molecular Markers to Aid in the Assessment of Essentially Derived Varieties in Maize

A Joint Project by ASTA (American Seed Trade Association) and UFS (French Breeders Association), and under the umbrella of ISF (International Seed Federation)
What is an Essentially Derived Variety?

A variety that is predominantly derived from a first variety while retaining the expression of the essential characteristics of the first variety that has Plant Variety Protection.

- What are the ‘essential characteristics’?
- Having a marker set can help define thresholds
Recommend 3072 SNPs as a first step to assessing the likelihood of essential derivation

Defined the set of 3072 SNPs

Recommended distance thresholds of:

- <91% ‘green’ no essential derivation
- ≥91% ‘orange’ potential EDV; reversal of burden of proof
- ≥ 95% ‘red’ definite EDV (UFS) or strong indication (ASTA) of EDV
Next steps – Markers for PVP Protection

- Can markers replace or supplement PVP characteristics?
Markers being effectively used for major gene effects and transgene introgression: *What about other QTLs? Too much of a good thing...?*

- Multiple QTLs, pleiotropic effects of QTLs
- Too many to manage in a breeding program

Genomic selection  
Gene modelling

Hypothesis + tilling/gene editing + ecotilling, gene editing
Genomic selection: can we sequence breeding lines?

Cost per Genome

- $100M
- $10M
- $1M
- $100K
- $10K
- $1K


Moore's Law

GbS

- Genomic DNA
- ApeKI site
- ApeKI site
- ApeCI site
- Digestion with ApeKI
- Sticky end
- ApeKI sticky end
- ApeKI sticky end
- Forward adaptor
- Unique DNA barcode for each sample
- Reverse adaptor
- Digested and barcoded DNA samples
- Pool DNA
- PCR amplify
- Illumina sequencing
- ApeKI restriction site
- Barcode
- DNA sequence

NIH National Human Genome Research Institute
genome.gov/sequencingcosts
Evolution of marker systems

- No. of markers
- No. of samples

- Single plex eg KASP
- Med density chip
- High density chip
- Seq
Evolution of marker systems

- Single plex eg KASP
- Med density chip
- GbS?
## Comparison of chips and GbS

<table>
<thead>
<tr>
<th>Chips</th>
<th>GbS</th>
</tr>
</thead>
</table>
| - Existing technology  
  ◦ costs stabilize/increase  
  ◦ Instrumentation outdated/not supported |
| - High development costs |
| - Ascertainment bias |
| - Hard to capture haplotypes |
| - Next generation technology  
  ◦ costs decrease |
| - Limited development costs |
| - Limited ascertainment bias |
| - Local haplotypes captured |
Reality: comparison of chips and GbS

- New bead systems
- Higher volumes

Costs

- Library time/cost
- Processing costs for data volume and imputation
- FTO

Chips vs. GbS
Reality: comparison of chips and GbS

Costs
- New bead systems
- High volumes

Chips

• Library time/cost
• Processing costs for data volume and imputation
• FTO

Press release – KeyGene’s SBG patent upheld by the USPTO after ex parte reexamination

March 3, 2016

KeyGene announced today that its United States patent (US 8,815,512) protecting Sequence-Based Genotyping (SBG) technology was upheld by the United States Patent and Trademark Office, following ex parte reexamination initiated by Cornell University. The patent is part of KeyGene’s global dominant patent portfolio protecting methods for simultaneous polymorphism discovery and genotyping, including SBG, GBS, RAD, ddRAD and related methods. Michiel van Eijk, Ph.D., CSO of KeyGene:
How many SNPs are needed?

European maize
Zhao et al. TAG 2007
Pressures on industry

Commodity prices – maize

- Collapse in S. American markets
- Slowing growth world-wide
- Pressure to reduce research and production costs
## Cost versus effectiveness of a field plot versus marker data

<table>
<thead>
<tr>
<th>Lower cost</th>
<th>Higher cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>High G x E</td>
<td>No G x E</td>
</tr>
</tbody>
</table>

![Image of field plot]

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>B</th>
<th>A</th>
<th>A</th>
<th>A</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>
Cost versus effectiveness of a field plot versus marker data

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>-</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Table: 

- A: Marker A
- B: Marker B

Diagram: 

- Field plot with markers A and B
- Comparison chart
Evolution of marker systems

- **GbS**
  - High-density chip
- **Med density chip**
- **NGG?**
- **Single plex eg KASP**

- **No. of samples**
- **No. of markers**
Next Generation Genotyping

- Targeted sequencing
  - Primers/probes designed to specific sequences
- Samples are barcoded
- Mostly amplicon or ligase based
- Affymetrics, Keygene, DArT, Illumina etc
Amplicon sequencing
eg Thermofisher ampliseq

Barcode samples with unique sequence
Genotyping becomes a logistics issue

- Sampling
- Coordinated batching of samples
- Shipping logistics
- Permits and documentation
- Data management
- Automated analysis

2 weeks
Evolution of markers: IBD

- Single plex eg KASP
- New computational methods haplotypes/IBD
- GbS
  - High density chip
  - Med density chip
- No. of markers
- No. of samples
- IBD segment
Future of Molecular Breeding

- Simplified
- Process driven
- Data management pipelines, decision support systems
- Improvements in sampling
- Continuous improvements made through genomic selection, gene modelling, ecotilling and gene editing