Using many genes for selection and cross prediction in plant breeding

Howard Eagles
Karen Cane
Russell Eastwood
Gil Hollamby
Haydn Kuchel
Meiqin Lu
Peter Martin
Peter Sharp
Neil Vallance
Marie Appelbee
Molecular Plant Breeding CRC
Project: Funded by GRDC

• Collaborators
  - University of Adelaide
  - Department of Primary Industries Victoria
  - Australian Grain Technologies
  - LongReach Plant Breeders
  - South Australian Research and Development Institute
  - New South Wales Department of Primary Industries (outside collaborator)
  - Agriculture Western Australia
  - University of Sydney (outside collaborator)
Why genes?

1. Longer lasting than cultivars
   - CIMMYT wheats have the same dwarfing genes as in Dr Borlaug’s time
   - Breeding program a failure if cultivars don’t change
   - Allele (variant) frequencies of genes change

2. Mendelian segregation provides predictions for specific crosses
Why cross prediction?

- Improves the efficiency of breeding

- A gain from 3% to 6% of crosses producing a cultivar is a large gain, even though 94% still unsuccessful
Genes and genotypes

• Quantitative genetical theory is built on hypothetical genes

• In practice, for self-pollinated crops calculations are made with inbred lines with unknown genes

• Molecular advances are greatly enhancing the identification of alleles of real genes
Genotypic values

Expected value of a phenotype in the target population of environments

\[ P = G + E + GE \]

P is the observed value (a number)

G is the genotypic value, usually cultivars plus breeding lines, but, can be genes (a number)

E is the environmental deviation

GE is the genotype-by-environment interaction
Calculation of genotypic values for genes
Self-fertilising species like wheat and rice

• Usual methods
  - Precise genetic stocks (isolines)
  - Doubled haploids
  - Random inbred lines

• Alternative
  - Plant breeding data
Strengths and weaknesses
Usual method

- **Strengths**
  - Expect unbiased estimates
  - Computationally simple

- **Weaknesses**
  - Usually small sample of germplasm and target environment
  - Project needs to be fully funded
Plant breeding data

• Strengths
  - Large sample of germplasm and environments
  - Always relevant to the breeding program
  - Shared costs with the breeding program

• Weaknesses
  - Potential for biased estimates
  - Computationally difficult
AGT Wheat Breeding Selection Sites 2006

178,000 yield plots at 53 locations across Australia in 2006
Calculations

- Focus on genes, lines transient carriers
- As an old plant breeder, difficult concept for me
- Not difficult for animal breeders
- REML, genes fixed, lines or relationship matrices random
- Inclusion of both genes and lines to remove bias
My approach to statistics

• Principle: I use mathematical procedures that work, not necessarily those currently promoted. Often from different areas of research.

• My test is the accuracy of the predictions.

• Not internally generated standard errors.

• An age-old controversy in science eg. Richard Feynman versus many theoretical physicists.
Where did my ideas come from?

Oscar Kempthorne
Grain quality and yield

- Grain quality important in itself
- Different quality types in Australian wheat from Australian Soft to Prime Hard
- Reduces selection pressure possible for grain yield
- Therefore, a reduction in the selection pressure for grain quality allows an increase for grain yield
Wheat grain quality

• 6 glutenin genes (many alleles)
• 2 puroindoline genes (function as 1 with 3 alleles)
• 1 serpin gene on 5B (2 alleles)

• Affect
  - Grain hardness
  - Milling yield
  - Dough water absorption
  - Dough development time
  - Dough strength
  - Dough extensibility
The serpin gene

- Serpins are about 4% of endosperm protein.

- Could form disulfide bridges: potential to affect quality traits.

- Protein + and – known from a gene on 5B (Peter Sharp’s contribution).

- Worked with Peter to track + and – protein through 100 years of Australian breeding.
Marker development

• We had DNA from >1000 varieties and breeding lines with inter-connected pedigrees

• We told Peter we wanted a diagnostic, DNA-based marker for the + and – alleles

• Chong-mei Dong in Peter’s lab produced the marker in about 3 months

• Consistent across Australian germplasm
Population approach to marker development

- No doubled haploids
- No QTLs
- Directly from protein to marker
- Later checked with Trident /Molineux DH population (it was correct)
Genotypic values of the serpin

- 899 lines classified for the serpin + 6 glutenin genes + puroindolines + VPM1

- Used lines not mixed with reliable pedigrees

- Segregation matched pedigrees

- 545 environments (site-year combinations)

- REML analysis with relationship matrix to reduce bias
ICIS used to calculate relationship matrices: Graham McLaren’s help essential. Now with the Generation Challenge Program
Serpin genotypic values

<table>
<thead>
<tr>
<th>Srp5B</th>
<th>Flour protein</th>
<th>Rmax</th>
<th>Extens.</th>
<th>Dough dev. time</th>
<th>Water absorp.</th>
<th>Milling yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>11.1</td>
<td>321</td>
<td>20.17</td>
<td>4.65</td>
<td>62.3</td>
<td>74.85</td>
</tr>
<tr>
<td>b</td>
<td>11.1</td>
<td>316</td>
<td>20.01</td>
<td>4.70</td>
<td>62.3</td>
<td>74.46*</td>
</tr>
<tr>
<td>sed</td>
<td>0.1</td>
<td>6</td>
<td>0.12</td>
<td>0.1</td>
<td>0.2</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Serpin genotypic values

• + allele positive for milling yield

• Consistent across time (1982-95, 1996-2001, 2002-06)

• Consistent across flour protein levels

• Consistent for Trident/Molineux DH
Origin of serpin null

- From Indian parents used by Farrer

- Later shown to be linked to *Vrn-B1* (major adaptation gene)

- Pleiotropic effect of *Vrn-B1*? But Trident/Molineux not segregating for *Vrn-B1*.

- Also in some European wheats. Farrer wheats major parents of Pacific NW USA and Japanese wheats; maybe there also.
Nature of WSZ1a null allele

WSZ1a gene cloned by PCR from normal “a” allele, and null “b” allele

Amino Acid sequence alignment (from a.a. 163 of published one)

<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z49890 - WSZ1a</td>
<td>FGSIDNTKL VLANLYFGK AWTKT</td>
<td>EQFDSY TNDYFYYLD GS</td>
<td>5QTFMMS</td>
<td></td>
</tr>
<tr>
<td>PCR - a</td>
<td>FGSIDNTKL VLANLYFGK AWTKT</td>
<td>EQFDSY TNDYFYYLD GS</td>
<td>5QTFMMS</td>
<td></td>
</tr>
<tr>
<td>PCR - b</td>
<td>FGSIDNTKL VLANLYFGK AWTKT</td>
<td>EQFDSY TNDYFYYLD GS</td>
<td>5QTFMMS</td>
<td></td>
</tr>
</tbody>
</table>

Number of SNPs

“a” allele is as published

One SNP gives premature “stop” in “b” allele
Cane et al.
Aust J Agric Res 2008
59: 883-890

“The effects on grain quality traits of a grain serpin protein and the VPM1 segment in southern Australian wheat breeding”

NO POPULATIONS DEVELOPED – used accumulated breeders’ data to determine effects of important alleles
Glu, Pin and serpin genes

- Have estimates for 10368 possible combinations of alleles (genotypes) for Rmax, extensibility, dough development time, water absorption
- Can utilise in the Cross Predictor
- **Being used for this purpose**
- Can utilise for marker assisted selection with progeny, but currently expensive
• 8 genes  ->  1 genotypic value per trait

• No reason this can’t be extended to 9, 10, 11 or more genes!

• Just need large datasets to predict them with reasonable accuracy
Current work with vernalisation, photoperiod and height genes

• Have diagnostic markers for \textit{Vrn-A1}, \textit{Vrn-B1}, \textit{Vrn-D1}, \textit{Ppd-D1}, \textit{Rht1}, \textit{Rht2}

• Assembled data from 128 trials 1986 to 2009

• Karen has typed > 1000 lines for these genes

• Must have accurate pedigrees, not mixed, genes match pedigrees
Locations of trials

- Trials at Roseworthy, Walpeup, near Horsham, Wagga Wagga and Temora
- Heading dates and grain yields from all 128 trials
- Daily meteorological data from all trials
Genotypes in Australian wheat

- 24 genotypes (combinations of *Vrn-A1* (3 alleles), *Vrn-B1*, *Vrn-D1* and *Ppd-D1*) possible, but 4 combinations predominate in spring wheats.

1. *Ppd-D1a, Vrn-A1a, Vrn-B1v, Vrn-D1v* (Cook, Condor types)
Winter and facultative wheats

- Winter wheats all \textit{Vrn-A1v, Vrn-B1v, Vrn-D1v} (expected)

- Oxley (facultative wheat) \textit{Vrn-A1v, Vrn-B1v, Vrn-D1v}, but does it have \textit{Vrn-D4}? Kenji Kato’s linked marker suggests it does.

- Suggestion of recombination with the linked \textit{Vrn-D4} marker in Kite family.
Suggestions from typing

• Photoperiod insensitive allele not as predominant as expected (sensitive allele in major cultivars like Spear, Frame, Yitpi)

• Also in modern low latitude cultivars like Baxter, Wylie and Sunlin

• Second Vrn-A1b allele in Farrer cultivars, but not Dundee. Vrn-A1b still being released (eg Yenda 2007)

• Substantial match to regional adaptation and sowing date recommendation

• Paper published
Preliminary conclusions

- Vrn-A1a and Vrn-D1a are a stronger spring alleles than Vrn-B1a
- Vrn-A1a responds differently to vernalising temperatures than Vrn-D1a
- Substituting 1 further spring allele when 2 are present has the smallest effect
- Therefore, some epistasis, but not complete
- However, vernalisation saturation for winter wheats reached in 36 – 51 days in these 128 experiments (all vernalised)
Predictions

• As with genes affecting grain quality, best predictions are made with combinations of genes

• With our very large data-sets, we can do this

• Extends the types of crosses feasible, eg hard wheats can be used to a much greater extent to breed soft wheats
# Simple example

<table>
<thead>
<tr>
<th>Cultivar</th>
<th>Vrn-A1</th>
<th>Vrn-B1</th>
<th>Vrn-D1</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condor type (Parent1)</td>
<td>a</td>
<td>v</td>
<td>v</td>
<td>Spring</td>
</tr>
<tr>
<td>Dollarbird (Parent2)</td>
<td>v</td>
<td>a</td>
<td>a</td>
<td>Spring</td>
</tr>
<tr>
<td>EGA Wedgetail (progeny)</td>
<td>v</td>
<td>v</td>
<td>v</td>
<td>Winter</td>
</tr>
</tbody>
</table>
Don’t have

• Diagnostic marker for *Ppd-B1*. Segregating on Trident/Molineux, almost certainly important.

• Diagnostic marker for *Vrn-D4*, in Gabo, almost certainly important for facultative wheats like Oxley and Vasco.

• Kenji Kato is working on them, likely to be successful.
CROSS PREDICTION

- Major genes, alone or in combination wherever possible. Current CrossPredictor could be extended from genes affecting quality to include genes affecting adaptation

- Breeding values for unknown genes