A novel pipeline for dissecting genetic architecture of wheat yield responses: linking semi-automated phenotyping and Bayesian network modelling

S. Amer, P. Jackson, R. Casebow, D. O’Sullivan
School of Agriculture, Policy and Development, University of Reading.

Monogram 4th - 6th April 2017
Post genomics era - Phenotyping bottle neck

Accumulated investment by year in K€

Year
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013
- 2014
- 2015

Investment until that year
New investment in that year

Cost per million base pairs

www.phenospex.com
Genetic architecture

Plant population
- Large population
- High allelic variation
- High level of recombination

Phenotyping
- Field trials
- No. of lines
- Selected traits
- Tendency to automation
- Data processing

Genetic modeling
- High density SNP coverage
- Robust model
## MAGIC population

(Multiparent Advanced Generation Inter-Cross)

<table>
<thead>
<tr>
<th>Variety</th>
<th>Trait Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alchemy</td>
<td>Yield, disease resistance, breeding use, soft</td>
</tr>
<tr>
<td>Brompton</td>
<td>Hard feed, 1BL/1RS, OWBM-resistant</td>
</tr>
<tr>
<td>Claire</td>
<td>Soft biscuit/distilling, slow apical development</td>
</tr>
<tr>
<td>Hereward</td>
<td>High-quality benchmark 1 bread-making</td>
</tr>
<tr>
<td>Rialto</td>
<td>Moderate bread-making, 1BL/1RS</td>
</tr>
<tr>
<td>Robigus</td>
<td>Exotic introgression, disease resistance, breeding use, OWBM-resistant, Rht-B1</td>
</tr>
<tr>
<td>Sissons</td>
<td>Bread-making quality, early flowering, Rht-B1</td>
</tr>
<tr>
<td>Xi19</td>
<td>Bread-making quality, facultative type, breeding use</td>
</tr>
</tbody>
</table>

Mackay et al. 2014, *G3: Genes, Genomes, Genetics* 4

LD decay in the MAGIC population
MAGIC population panel

**Field trails:** 2014-15 and 2015-16

**790 RILs + 8 parents + 2 checks**

Skyfall & Kielder.

Plot size: **2*6 m**

2 reps (partially randomized complete block design)

**Target traits**

- Height
- Date of anthesis
- Vegetation indices (GAI, R:FR, NDVI)
- Grain yield
- Post harvest traits:
  - TKW, Grain width-length, spikelet length
  - No. of grains/spike and protein content
Semi-automated phenotyping

Proximal sensing cart

- RTK Positioning
- Infra red logger
- RGB Camera
- Air temperature logger
- PAR sensors
- Battery
- Tough book
Why to bother with pushing a cart?

- Weather resilience
- Resolution
- System Mobility
- Operation simplicity

Bottleneck
Data processing

- Sensors data
- Assign measurements to GPS coordinate
- Image analysis
Phenoharvest - GAI retrieving

Time series images taken throughout the growing season

Seedling   Tillering   repining

Graph showing a time series from November to July with peaks in April and July.
Multispectral indices

NDVI

Crop cover percentage
Clustering according to GAI
Post harvest phenotyping

Grain scan software
Whan et al. plant Methods 2014

Grain length/ width/area
Transgressive segregations

Hereward

Robigus
2014-15 season
Bayesian networks

- A Bayesian network of a set of variables $X = \{X_1, \ldots, X_n\}$ represents a joint probability distribution over those variables.

- It is graphically directed acyclic graphs, whose nodes denote the variables, Edges represent conditional dependencies, so that nodes which are disconnected represent variables which are conditionally independent of each other.

- The basic properties of BNs do not depend on the distributional assumptions of the data.
Bayesian networks: structure and parameters learning (Scutari et al, 2014, Genetics)

1- SI-HITON-PC algorithm to learn the parents and the children of the trait.

2- Drop all the markers that are not in any Markov blanket.

3- Setting the directions of the arcs according to assumptions.

4- Investigating the properties of the resulting models using, in each case, 10 runs of 10-fold cross-validation.

5- Inference: average of 10*10 to estimating the parameters of the averaged BN with RR using the whole data set.

6- 4000 SNPS, phenotypes BLUPS
Bayesian Network (2014-15)

Legend:
- **SNP markers**
- **Traits**
- **Marker to trait**
- **Trait to trait**
Bayesian networks predictive ability

<table>
<thead>
<tr>
<th>Trait</th>
<th>Genetic predictive correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthesis</td>
<td>0.2</td>
</tr>
<tr>
<td>Awns</td>
<td>0.8</td>
</tr>
<tr>
<td>Height</td>
<td>0.413</td>
</tr>
<tr>
<td>Late GAI</td>
<td>0.0923</td>
</tr>
<tr>
<td>Yield</td>
<td>0.13</td>
</tr>
</tbody>
</table>
2016-17 field trial

400 Lines (393 MAGIC RILs + 8 Parents + 32 Kielder plots).

Experimental design: Split plot (2 reps).
Main plot: stress/irrigation, subplot: RILs.

Supplementary irrigation: late April

Target traits:
- Height
- Date of anthesis
- Time series Vegetation indices (GAI, Multi spectral indices)
- Grain yield
- Post harvest traits: TKW, Grain width-length, spikelet length, No. of grains/spike and protein content.
Gauging water availability

PR2 soil profile probe
32 tubes/1m depth

I buttons Hydrochronic
16 at 50 cm depth
Summary

1- BNs provide competitive predictive accuracy for multi-trait modelling.

2- BNs graphical nature is easily interpreted.

3- The pipeline using the Phenocart, followed by data processing and image analysis by phenoharvest provides high quality phenotypic data to be Modelled through Bayesian Networks.

4- BNs as a model and the phenotyping pipeline is promising to dissect genetics wheat responses to water limitation.
ACKNOWLEDGMENTS

Supervisor: Donal O’sullivan

3rd year PhD: Peter Jackson

Crop research unit, University of Reading
Richard Casbow
Caroline Hadley
Liam Doherty
David Mclay