Plant Breeding Tools and Technologies for Variety Development

What does Technology look like?

Dr. Katy Martin Rainey
Soybean Genetics and Breeding
Agronomy Department
Purdue University
Overview of the Breeding Process

1. A breeder, along with others, determines the objectives.
2. The breeder assembles variation that would meet the defined objectives.
3. The breeder selects among different types of plant progeny.
4. New and improved varieties are disseminated to producers.

**Selection** programs are characterized by an interplay between the amount of genetic variation and the stringency of selection.

**Selection** is based on precision ranking from phenotypes.

Only a small fraction of lines developed in breeding programs are released as varieties.
3. The breeder selects among different types of plant progeny

Fisher’s *infinitesimal model*, which forms the basis of the principles of plant breeding, states that an infinite number of stochastic processes (referring to genes) control the observed phenotype.

*Phenotypes* are what we observe that is determined by genetics; the data collected to represent a *trait*.

*Quantitative traits* are polygenic or interact with the environment, or both. Nongenetic effects contribute to a blurring of the differences among phenotypic classes.

There is always more than one trait of interest, so selection can take several forms:
- one trait at a time
- multiple quantitative traits
- simultaneously: index of traits

**Genetic Architecture:**

\[ P = G + E \]

Three metrics to *select* quantitative traits:
- phenotypic value
- genetic value
- breeding value

*Selecting* on G requires kinship information from pedigrees, genome-wide markers, or both.
Most Agronomic Solutions are Determined Experimentally

The production environment is simulated and treatment effects are isolated statistically using generalized linear models or linear mixed models.

Response variables generally follow normal distributions.

**Example Research Questions:**

- Which variety out of many is the best?
- When should irrigation be applied?

![Histogram](image1.png) **Fig. 1** Probabilistic description of the distribution of yield

These questions can only be assessed in a field experiment, where it is essential to control all other variables except those that are under consideration.

**Fundamental Unit:** Field plot with variable treatments.

Field plots simulate the production environment.
Genetic Yield Potential is Assessed from Replicated Trials

\[ Y_{ijkl} = \mu + \alpha_i + \beta_j + \omega_k + \varepsilon_{ijkl} \]

\( Y_{ijkl} \) is the yield observation of plot \( ijkl \)
\( \mu \) is the intercept
\( \alpha_i \) is the \( i^{th} \) effect of treatment, \( i = \{1, \ldots, n\} \)
\( \beta_j \) is the block effect, \( j = \{1, \ldots, n\} \)
\( \omega_k \) is the effect of the \( k^{th} \) environment, \( k = \{1, \ldots, n\} \)
\( \varepsilon \) is the residual of the observation \( Y_{ijkl} \)

**GENETIC YIELD POTENTIAL**: the yield of optimal or ideal genotypes in a target environment, across major growing regions and environments
Special Plot Planting and Harvesting Equipment
Special Plot Planting and Harvesting Equipment
Reducing Experimental Error (Noise)

Can a source of error be controlled?
- No
- Yes
  - Field Plot Technique
    - planting equipment
    - equipment calibration
    - plot size
    - weed management
    - data QC

Is it Predictable?
- No
- Yes
  - Multiple Environments

Can it be blocked?
- No
- Yes
  - Experiment Design
    - replication
    - randomization
    - local control

Can it be quantified?
- No
- Yes
  - Covariates
    - plants/plot
    - row length
    - pest infestation

+ valid assumptions
Genetic Gain

Genetic gain is the amount of increase in performance that is achieved through genetic improvement programs.

The breeders’ equation

\[ R = \Delta G = \Delta \mu = h^2 S = \left( \frac{\sigma_A^2}{\sigma_P^2} \right) S \]

\[ \Delta G = \frac{ir \sigma_A}{T} \]

Where \( i \) is the standardized selection differential
\( r \) is the selection accuracy
\( \sigma_A \) is the square root of the additive genetic variance
\( T \) is the length of time to complete one breeding cycle
Tools that improve accuracy

Seed Counter  GPS Data Logger  Bar Coded Tools

FIELD INFORMATICS: automation of data management and quality control

GPS  RFID  GIS
Postharvest Data

Reflectance-based Predictions of Seed Composition
## End-Use Quality

<table>
<thead>
<tr>
<th>Entry</th>
<th>Dough Firmness (g)</th>
<th>Dough Stickiness (g)</th>
<th>Cookie Stack Ht (cm x4)</th>
<th>Cookie Width (cm x4)</th>
<th>Cookie Length (cm x4)</th>
<th>Weight Loss %</th>
<th>Calculated Final Moisture %</th>
</tr>
</thead>
<tbody>
<tr>
<td>VA07W-415</td>
<td>141</td>
<td></td>
<td>4.26</td>
<td>30.0</td>
<td>30.2</td>
<td>13.2</td>
<td>4.5</td>
</tr>
<tr>
<td>VA09W-75</td>
<td>175</td>
<td></td>
<td>4.56</td>
<td>29.3</td>
<td>29.0</td>
<td>12.5</td>
<td>5.1</td>
</tr>
<tr>
<td>VA09W-73</td>
<td>155</td>
<td></td>
<td>4.46</td>
<td>30.5</td>
<td>30.5</td>
<td>13.1</td>
<td>4.6</td>
</tr>
<tr>
<td>VA09W-188WS</td>
<td>156</td>
<td></td>
<td>4.65</td>
<td>28.7</td>
<td>28.7</td>
<td>12.2</td>
<td>5.4</td>
</tr>
<tr>
<td>Shirley (ck)</td>
<td>143</td>
<td></td>
<td>4.45</td>
<td>30.7</td>
<td>30.9</td>
<td>12.8</td>
<td>4.8</td>
</tr>
<tr>
<td>SY 547</td>
<td>146</td>
<td></td>
<td>4.32</td>
<td>30.1</td>
<td>29.2</td>
<td>12.6</td>
<td>5.1</td>
</tr>
<tr>
<td>F0014</td>
<td>149</td>
<td>stickier</td>
<td>4.24</td>
<td>31.5</td>
<td>30.8</td>
<td>12.9</td>
<td>4.7</td>
</tr>
<tr>
<td>F0039</td>
<td>120</td>
<td></td>
<td>4.04</td>
<td>32.5</td>
<td>32.3</td>
<td>13.6</td>
<td>4.1</td>
</tr>
<tr>
<td>F0065</td>
<td>119</td>
<td>stickier</td>
<td>4.39</td>
<td>31.3</td>
<td>31.3</td>
<td>13.6</td>
<td>5.1</td>
</tr>
<tr>
<td>D8006</td>
<td>128</td>
<td>stickier</td>
<td>4.04</td>
<td>32.2</td>
<td>32.0</td>
<td>13.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Caledonia (ck)</td>
<td>112</td>
<td>stickier</td>
<td>4.10</td>
<td>32.5</td>
<td>32.4</td>
<td>13.5</td>
<td>4.2</td>
</tr>
</tbody>
</table>

*Red, yellow & green highlights indicate poor, marginal and good quality parameters, respectively.*
Imaging With UAS
<table>
<thead>
<tr>
<th>AYT early 17</th>
<th>AYT late 17</th>
<th>Rank</th>
<th>AYT early 18</th>
<th>AYT late 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>Yield</td>
<td>Yield, Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>2</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield</td>
<td>ACC, Yield, Yield</td>
<td>3</td>
<td>ACC</td>
<td>Yield, Yield</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>4</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield</td>
<td>Yield</td>
<td>5</td>
<td>Yield</td>
<td>Yield</td>
</tr>
<tr>
<td>Yield</td>
<td>ACC</td>
<td>6</td>
<td>ACC, Yield</td>
<td>Yield, Yield</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>7</td>
<td>ACC</td>
<td>Yield, Yield</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>8</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield, Yield</td>
<td>ACC, Yield</td>
<td>9</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>10</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year 3 Advanced Yield Trials</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Yield, Yield</th>
<th>ACC</th>
<th>Yield</th>
<th>Yield, Yield, Yield</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>2</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield</td>
<td>ACC, Yield, Yield</td>
<td>3</td>
<td>ACC</td>
<td>Yield, Yield</td>
</tr>
<tr>
<td>Yield</td>
<td>ACC</td>
<td>4</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield</td>
<td>Yield</td>
<td>5</td>
<td>Yield</td>
<td>Yield</td>
</tr>
<tr>
<td>Yield</td>
<td>ACC</td>
<td>6</td>
<td>ACC, Yield</td>
<td>Yield, Yield</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>7</td>
<td>ACC</td>
<td>Yield, Yield</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>8</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield, Yield</td>
<td>ACC, Yield</td>
<td>9</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>10</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Yield, Yield</th>
<th>ACC</th>
<th>Yield</th>
<th>Yield, Yield, Yield</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yield, Yield</td>
<td>ACC</td>
<td>2</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield</td>
<td>Yield</td>
<td>5</td>
<td>Yield</td>
<td>Yield</td>
</tr>
<tr>
<td>Yield</td>
<td>ACC</td>
<td>8</td>
<td>Yield, Yield, Yield</td>
<td>ACC</td>
</tr>
<tr>
<td>ACC, Yield, Yield</td>
<td>ACC, Yield</td>
<td>9</td>
<td>Yield, Yield</td>
<td>ACC</td>
</tr>
</tbody>
</table>

9 3  Yield 26 8 6
8 8  Yield, ACC 29 5 8
3 5  ACC 10 2 0
Monica Herrero-Huerta, and K. M. Rainey. ‘High Throughput Phenotyping of Physiological Growth Dynamics from UAS-Based 3D Modeling in Soybean’. (2019) ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4213
High Density Molecular Markers

Sample preparation

DNA extraction

Restriction digest (PstI, MspI, ApeKI)

Unique DNA barcode

Forward adaptor

Restriction sites (PstI, MspI, ApeKI)

Reverse adaptor

NGS library construction

Pool samples

PCR amplification

Statistical analysis

SNP discovery

Sequence alignment

Illumina Platform

Structure analysis
Genomic Selection/ Genomic Prediction/ Genomewide Selection


**GENOMIC SELECTION:** Use a large set of random markers directly to perform a marker-based *selection* using the predicted marker effects.

GS is a tool for predicting breeding values for quantitative traits using dense DNA markers throughout the genome. It improves reliability by accounting for the *inheritance of genes with small effects.*
When to use GS

1. When phenotypic selection is ineffective.
2. To increase gain per unit time.
3. For traits that are difficult to measure.
4. For other TPES, to predict performance of lines in environments where they have not been phenotyped (i.e. hort crops).
5. When there are too many candidates to phenotype.
   (prediction of performance in a seedbank).
6. When seed amounts are insufficient (i.e. DH).

Criteria

1. Genomewide-markers must be cheap, easy.
2. Trait inheritance quantitative and generally additive.
3. GS must be cheaper than phenotyping, which is a function of prediction accuracy.
When to use GS

1. When phenotypic selection is ineffective

Fusarium Head Blight (FHB)
Another challenging disease in Barley

Major outbreak in Midwest U.S. in 1993
Mycotoxin deoxynivalenol (DON)
Sources of resistance are unadapted
Quantitatively inherited resistance
Many QTL with small effects
Challenging to phenotype

K. Smith et al.
When to use GS

6. When there are too many candidates to phenotype (prediction of performance in a seedbank)

Schematic overview of the project. Germplasm base, 962-accession reference set, 299-accession training set and prediction extrapolation

Yield potential for 663 untested accessions grouped into eight combinations based on predicted genotypic effects of three other traits. Green rectangles, favourable values; blue rectangles, unfavourable values. HT, plant height; RL, root lodging; and SN, stalk number.