Plant Breeding Tools and Technologies for Variety Development
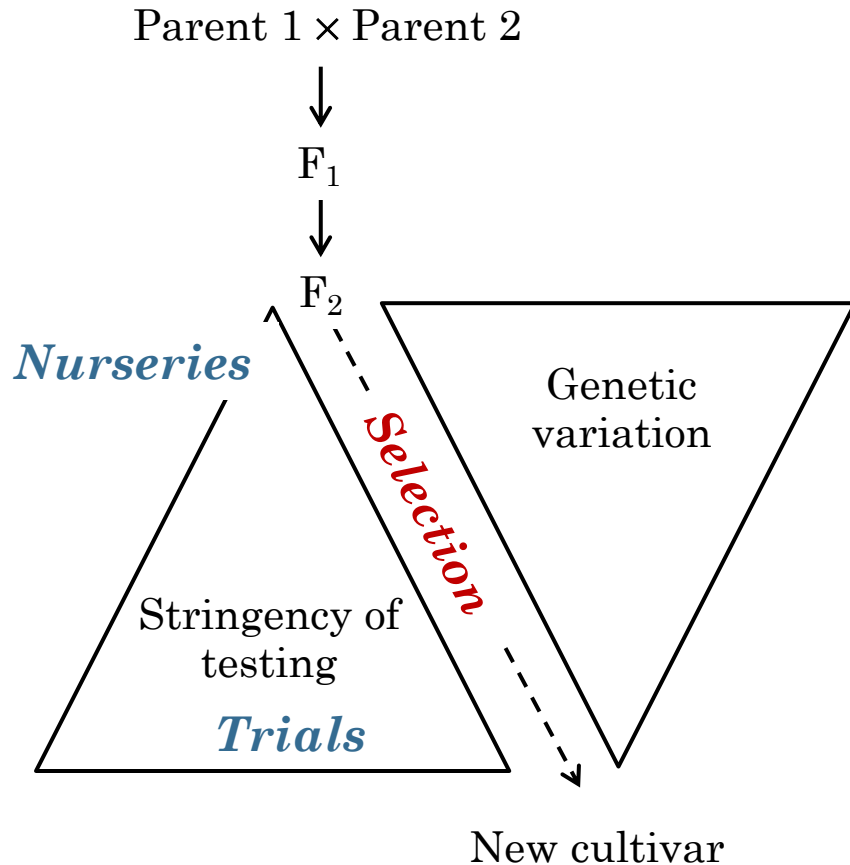

Teosinte

Modern Corn
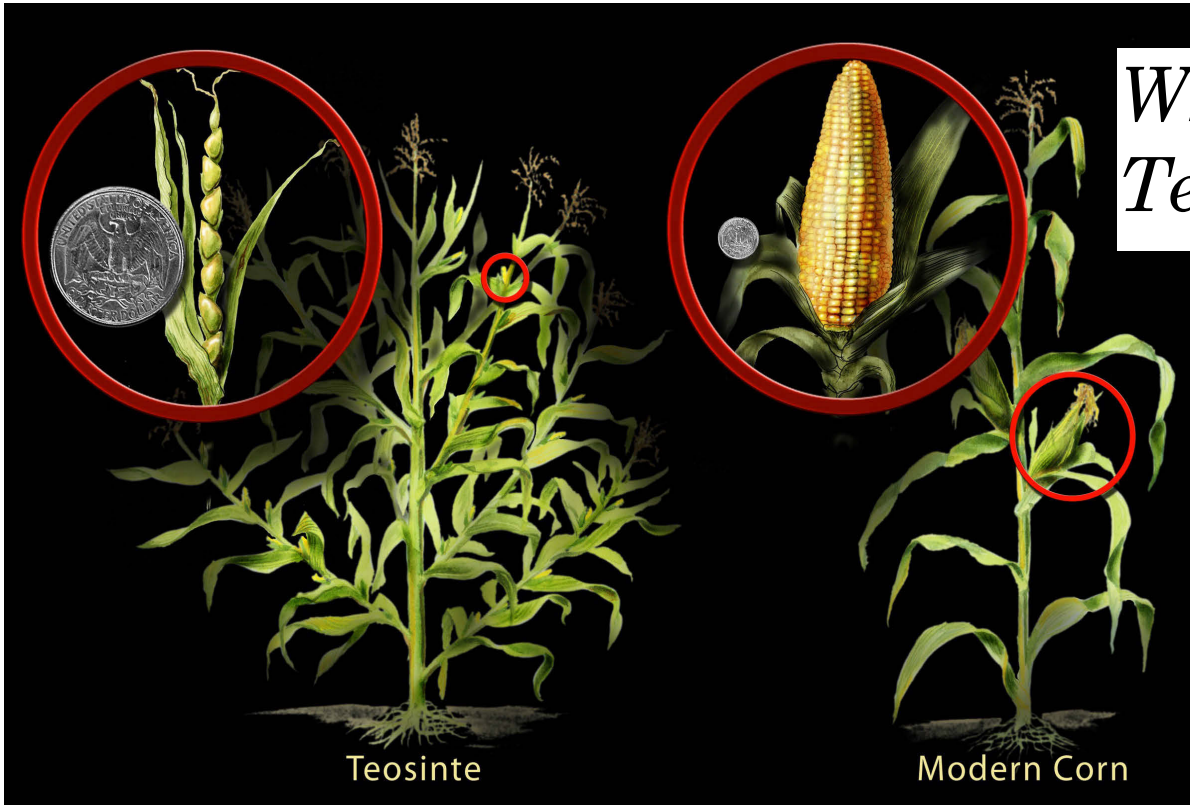
What does Technology look like?

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.

Parent 1 × Parent 2

$\downarrow$

F_1

$\downarrow$

F_2

Nurseries

Genetic variation

Selection

Stringency of testing

Trials

New cultivar

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.

Bernardo, R.(2014) Essentials of Plant Breeding, Chp 1, Stemma Press

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.

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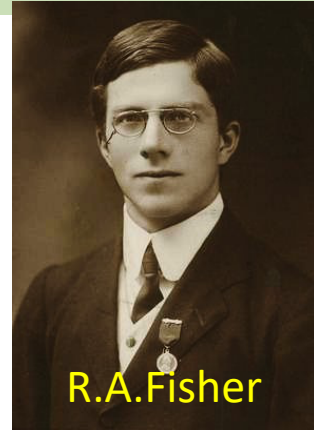
There is always more than one trait of interest, so selection can take several forms:
   one trait at a time
   multiple quantitative traits
   simultaneously: index of traits

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.

Fig. 1 Probabilistic description of the distribution of yield
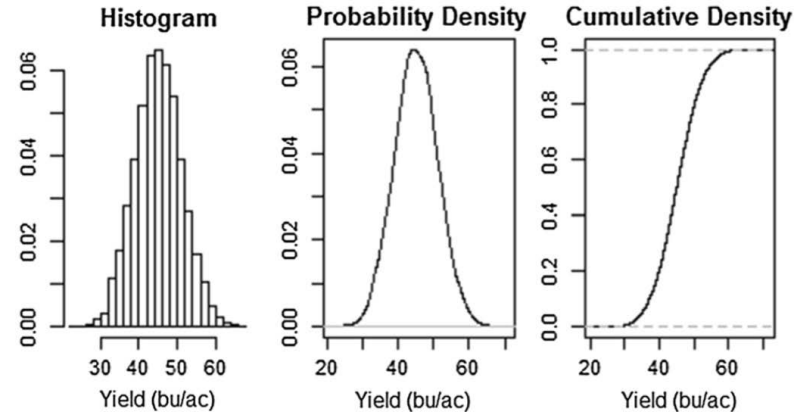
Example Research Questions:

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

*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

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \omega_k + \varepsilon_{ijkl}$$

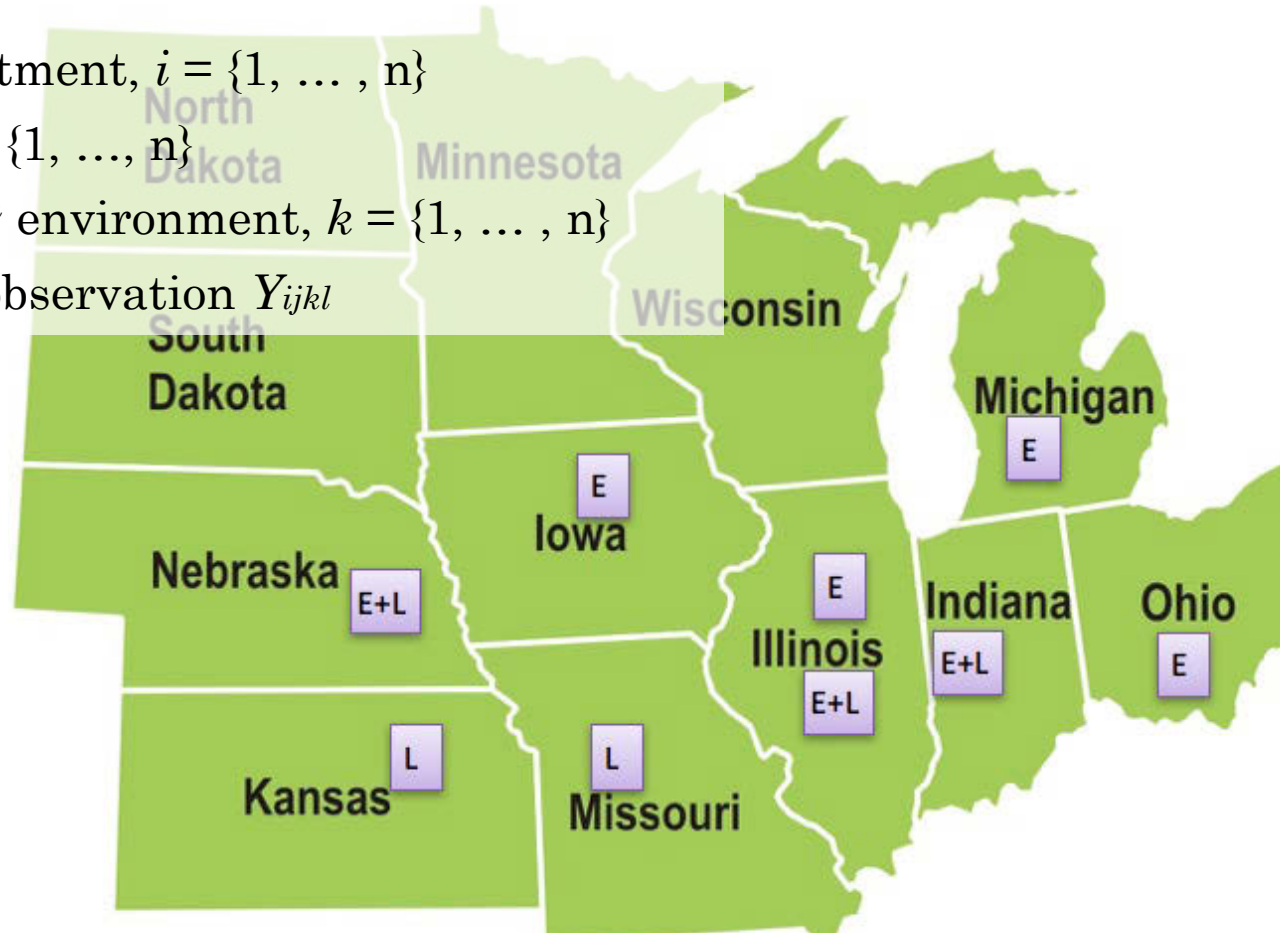Y_{ijkl} is the yield observation of plot $ijkl$

μ is the intercept

α_i is the ith effect of treatment, $i = \{1, \ldots, n\}$

β_j is the block effect, $j = \{1, \ldots, n\}$

ω_k is the effect of the kth environment, $k = \{1, \ldots, n\}$

ε 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
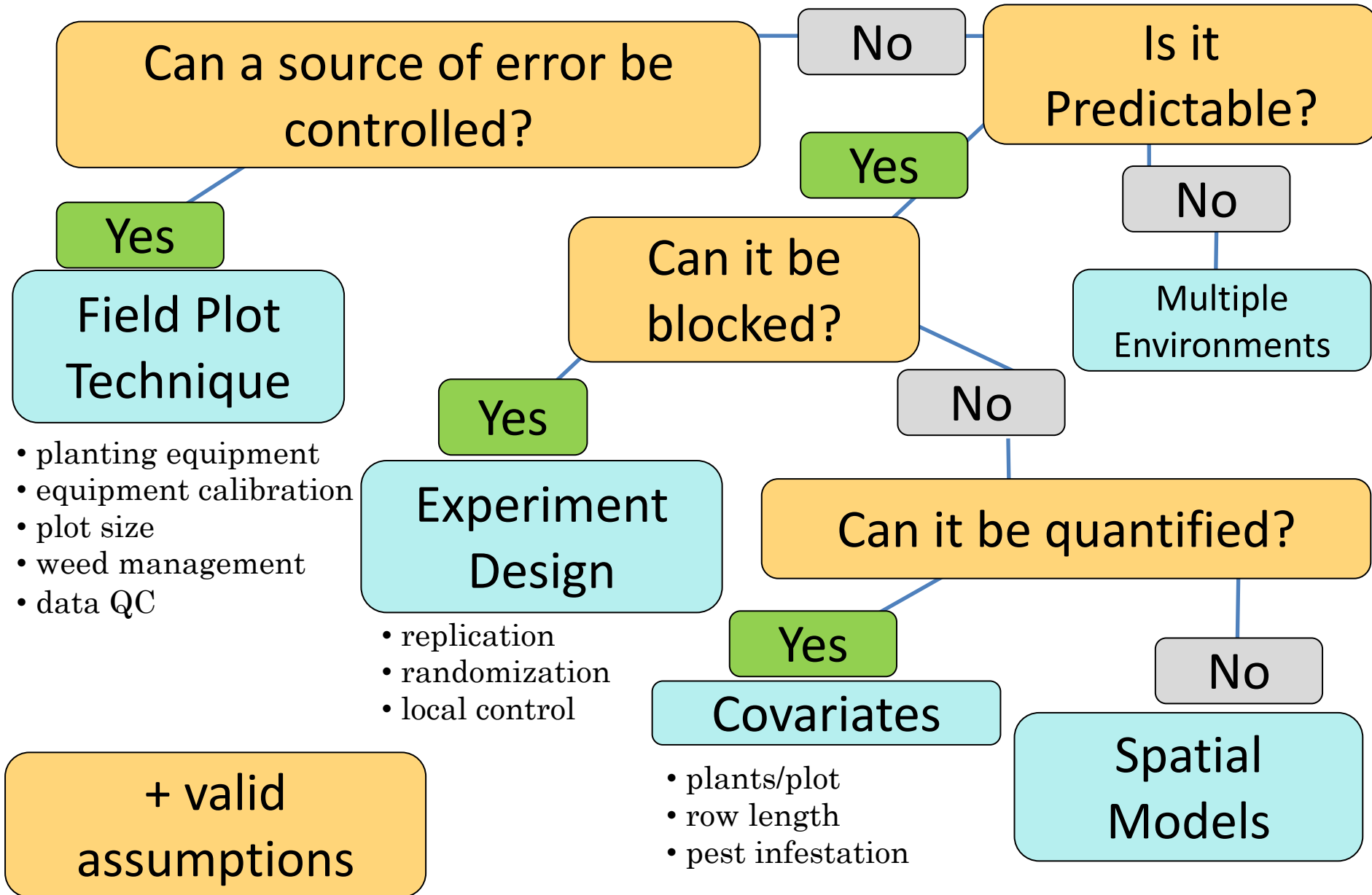
Special Plot Planting and Harvesting Equipment

Special Plot Planting and Harvesting Equipment

Reducing Experimental Error (Noise)

Can a source of error be controlled?

No

Is it Predictable?

Yes

Field Plot Technique

- planting equipment
- equipment calibration
- plot size
- weed management
- data QC

Can it be blocked?

Yes

No

Is it Predictable?

No

Multiple Environments

Experiment Design

- replication
- randomization
- local control

Can it be quantified?

Yes

Covariates

- plants/plot
- row length
- pest infestation

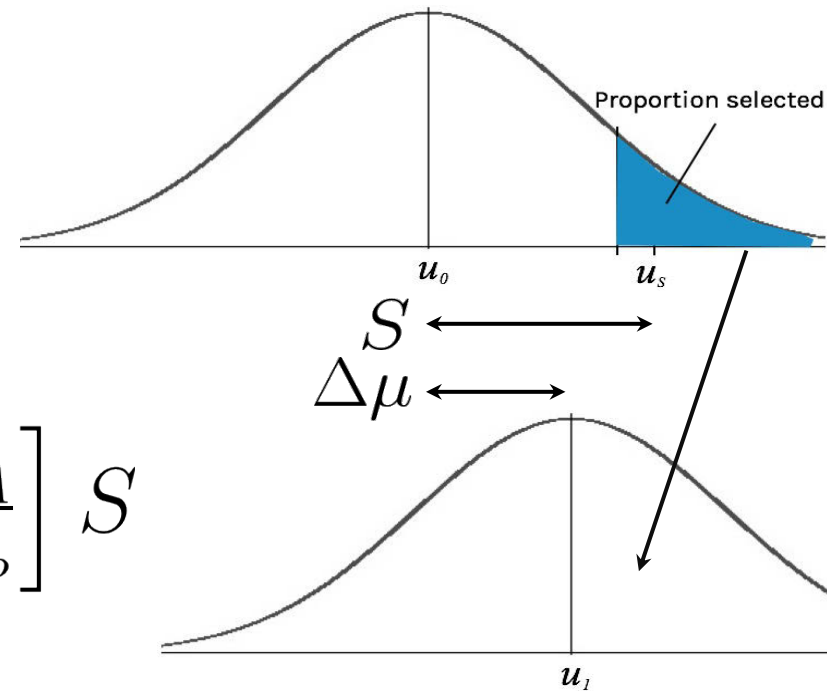**No**

Spatial Models

+ valid assumptions

Genetic Gain

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

The breeders' equation

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

Proportion selected

u_0 u_s

S

$\Delta \mu$

u_1

$$\Delta G = \frac{i r \sigma_A}{T}$$

Where i is the standardized selection differential
r is the selection accuracy
σ_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
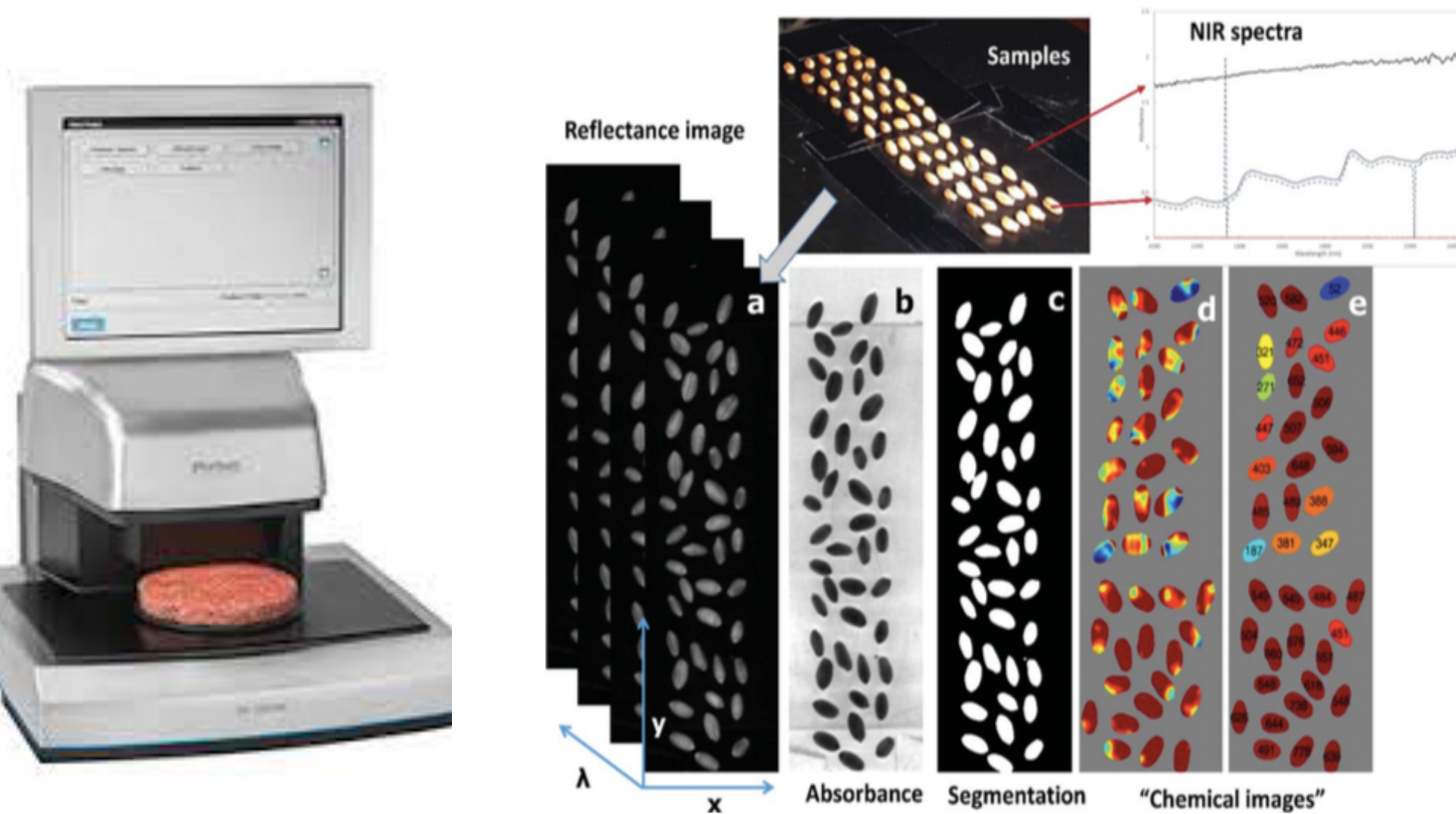

Seed Counter

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 28. Wire-cut cookie test (AACCI 10-53) parameters by Mondelez

Entry	Wire Cut Cookie Evaluation (10-53)*						
	Dough Firmness (g)	Dough Stickiness (g)	Cookie Stack Ht (cm x4)	Cookie Width (cm x4)	Cookie Length (cm x4)	Weight Loss %	Calculated Final Moisture %
VA07W-415	141		4.26	30.0	30.2	13.2	4.5
VA09W-75	175		4.56	29.3	29.0	12.5	5.1
VA09W-73	155		4.46	30.5	30.5	13.1	4.6
VA09W-188WS	156		4.65	28.7	28.7	12.2	5.4
Shirley (ck)	143		4.45	30.7	30.9	12.8	4.8
SY 547	146		4.32	30.1	29.2	12.6	5.1
F0014	149		4.24	31.5	30.8	12.9	4.7
F0039	120	stickier	4.04	32.5	32.3	13.6	4.1
F0065	119	stickier	4.39	31.3	31.3	12.6	5.1
D8006	128	stickier	4.04	32.2	32.0	13.3	4.4
Caledonia (ck)	112	stickier	4.10	32.5	32.4	13.5	4.2

*Red, yellow & green highlights indicate poor, marginal and good quality parameters, respectively.
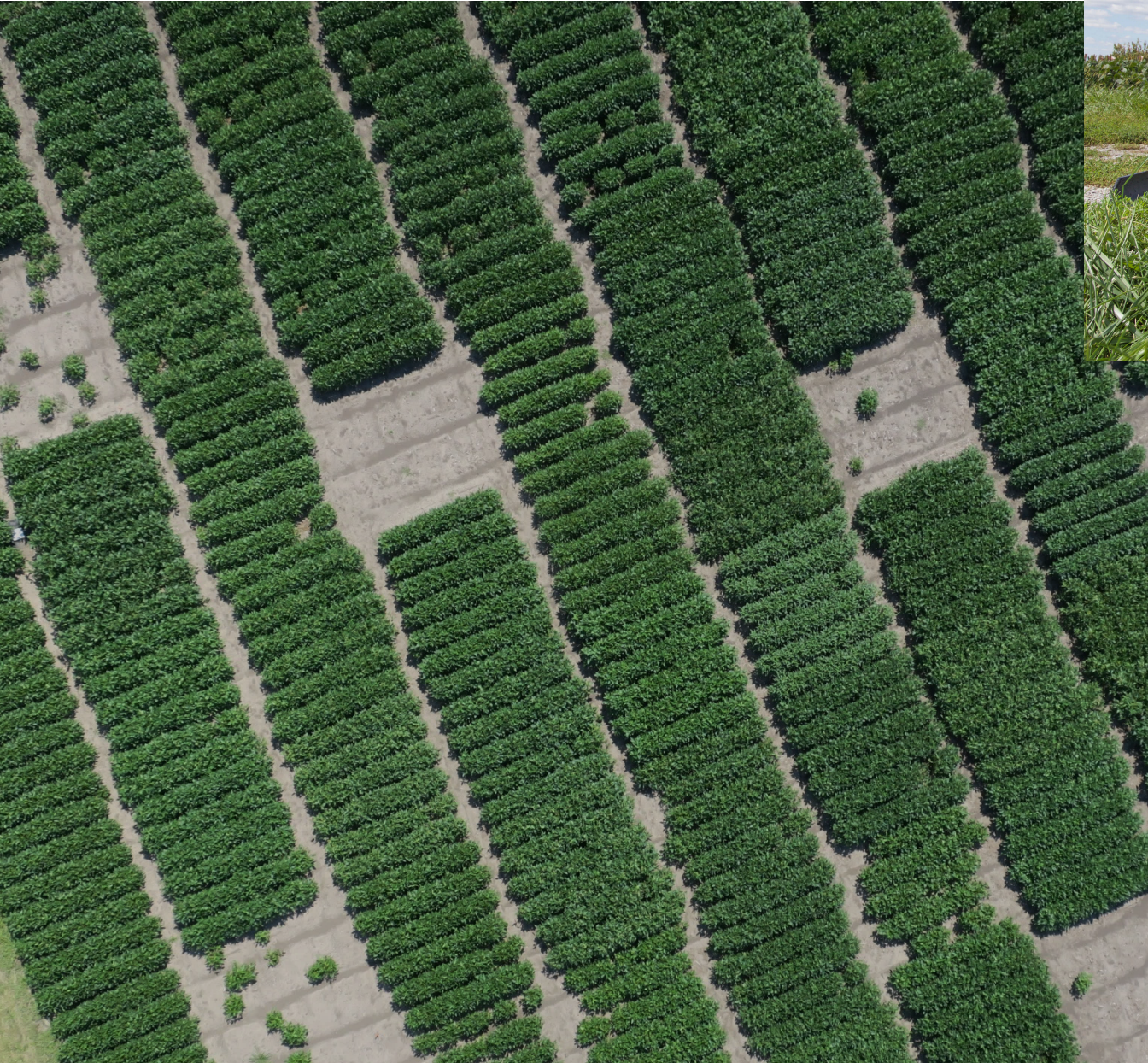

D8006   F0039   VA09W-75

(magenta)	poor quality
(yellow)	marginal quality
(green)	good quality

Imaging With UAS

Days After Planting

25 28 32 38 43 50 53   25 28 32 38 43 50 53

Percentage Canopy Coverage

GB$_{28}$
GB$_{35}$
GB$_{42}$
GB$_{50}$
GB$_{56}$
GB$_{65}$
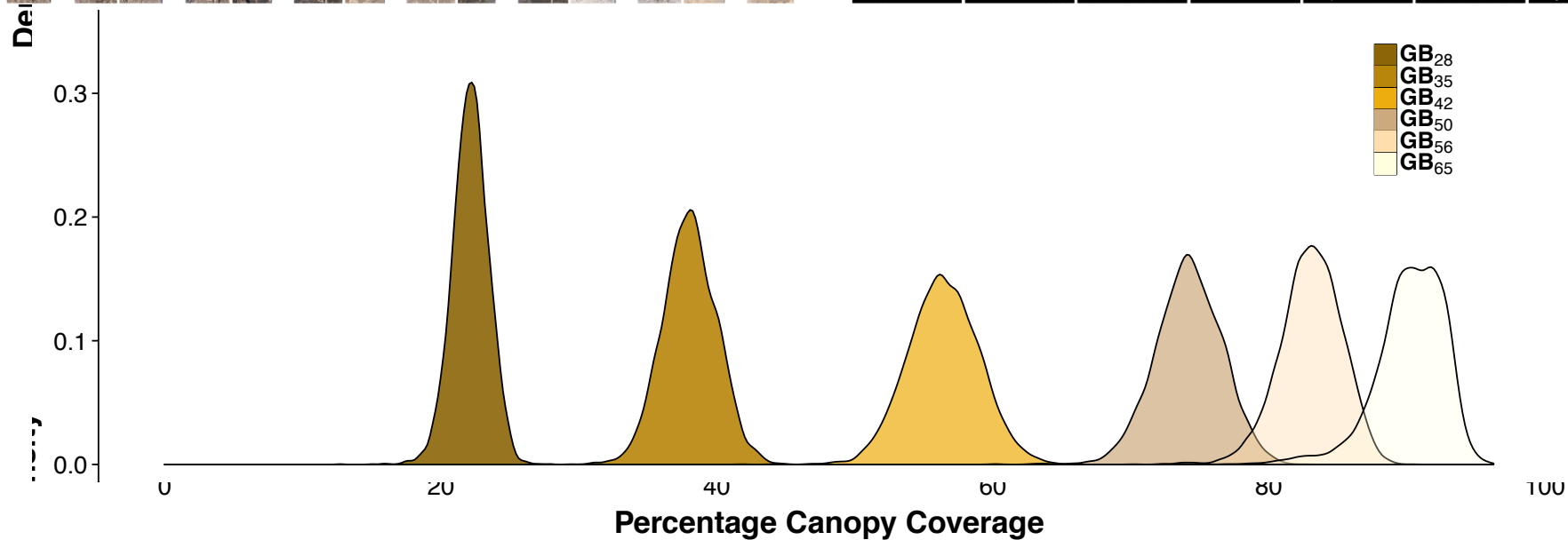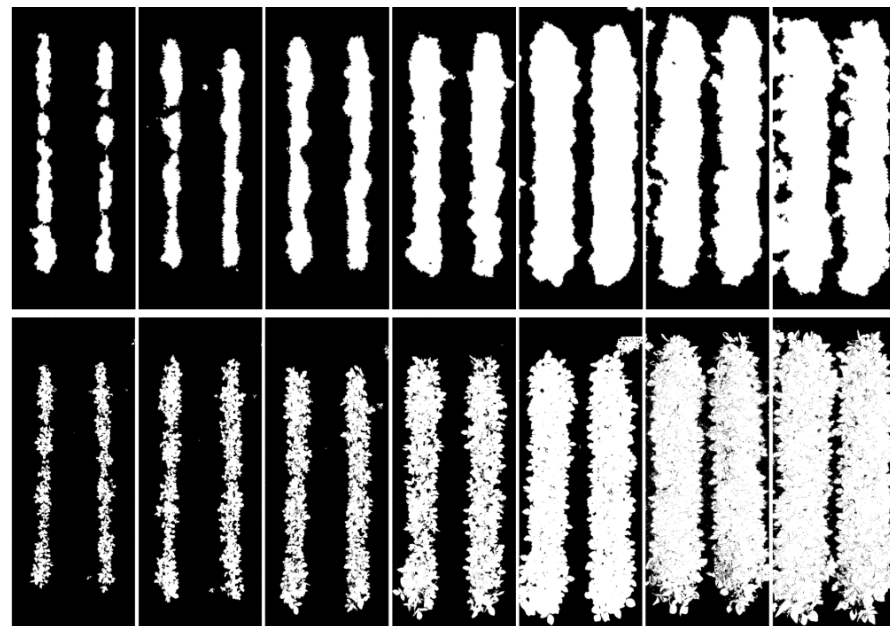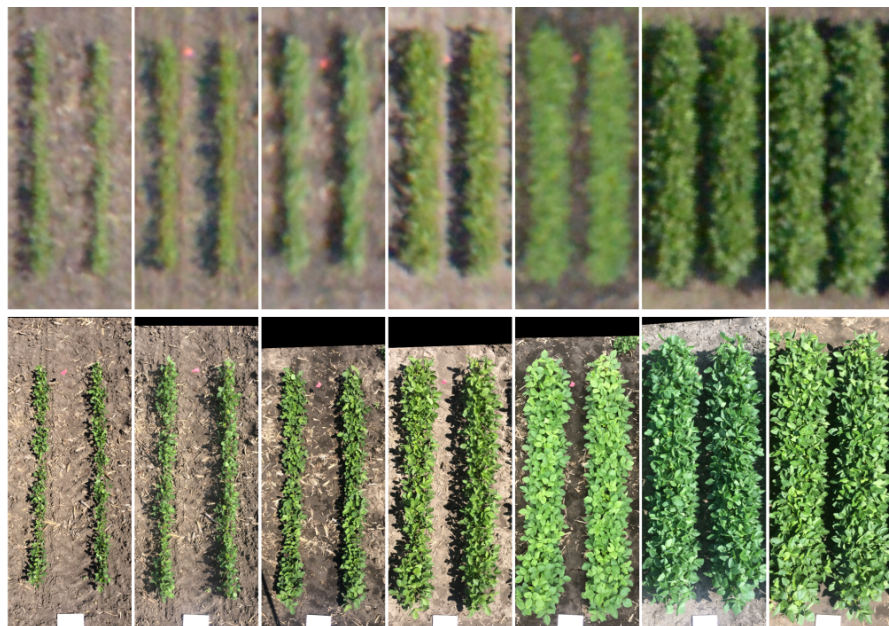
Soybean Progeny Row Selection Experiment

F. Moreira *et al.*

Year 3 Advanced Yield Trials

| AYT early 17	AYT late 17	Rank	AYT early 18	AYT late 18
Yield, Yield|ACC	Yield|ACC	1	Yield	Yield, Yield|ACC
Yield, Yield|ACC	Yield|ACC	2	Yield, Yield|ACC	Yield|ACC
ACC, Yield	ACC, Yield|ACC	3	Yield	Yield, Yield|ACC
Yield, Yield|ACC	ACC	4	Yield, Yield|ACC	Yield|ACC
ACC, Yield	Yield|ACC	5	Yield	Yield
Yield|ACC	ACC, Yield, Yield|ACC	6	ACC, Yield	Yield, Yield|ACC
Yield, Yield|ACC	ACC, Yield, Yield|ACC	7	ACC	Yield, Yield|ACC
Yield, Yield|ACC	Yield|ACC	8	Yield, Yield|ACC	Yield
ACC, Yield, Yield|ACC	ACC, Yield|ACC	9	Yield, Yield|ACC	Yield|ACC
Yield, Yield|ACC	Yield	10	Yield|ACC	Yield|ACC
9	3	Yield **26**	8	6
8	8	Yield|ACC **29**	5	8
3	5	ACC **10**	2	0

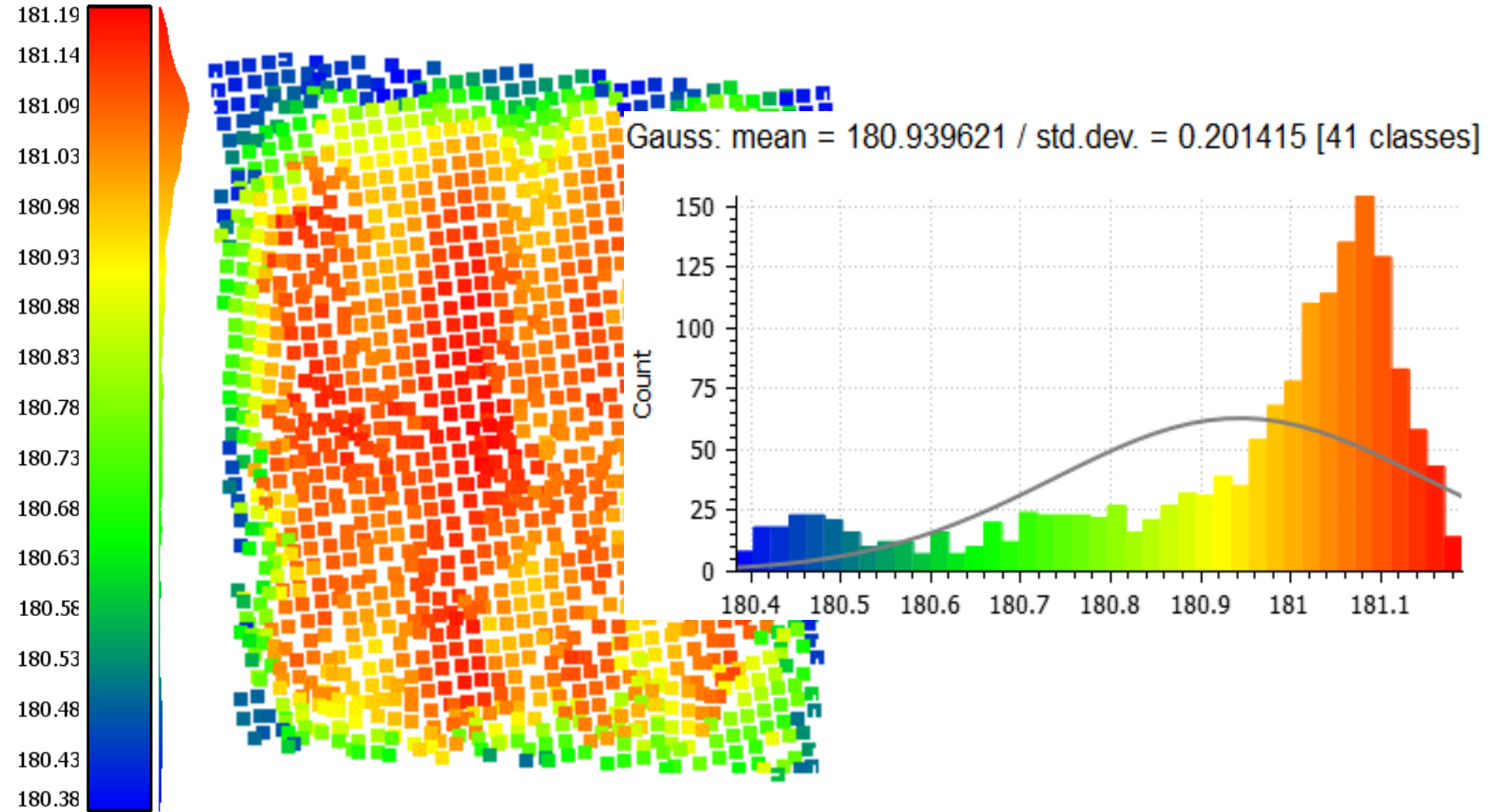# Mesh from Point Cloud by Delaunay triangulation
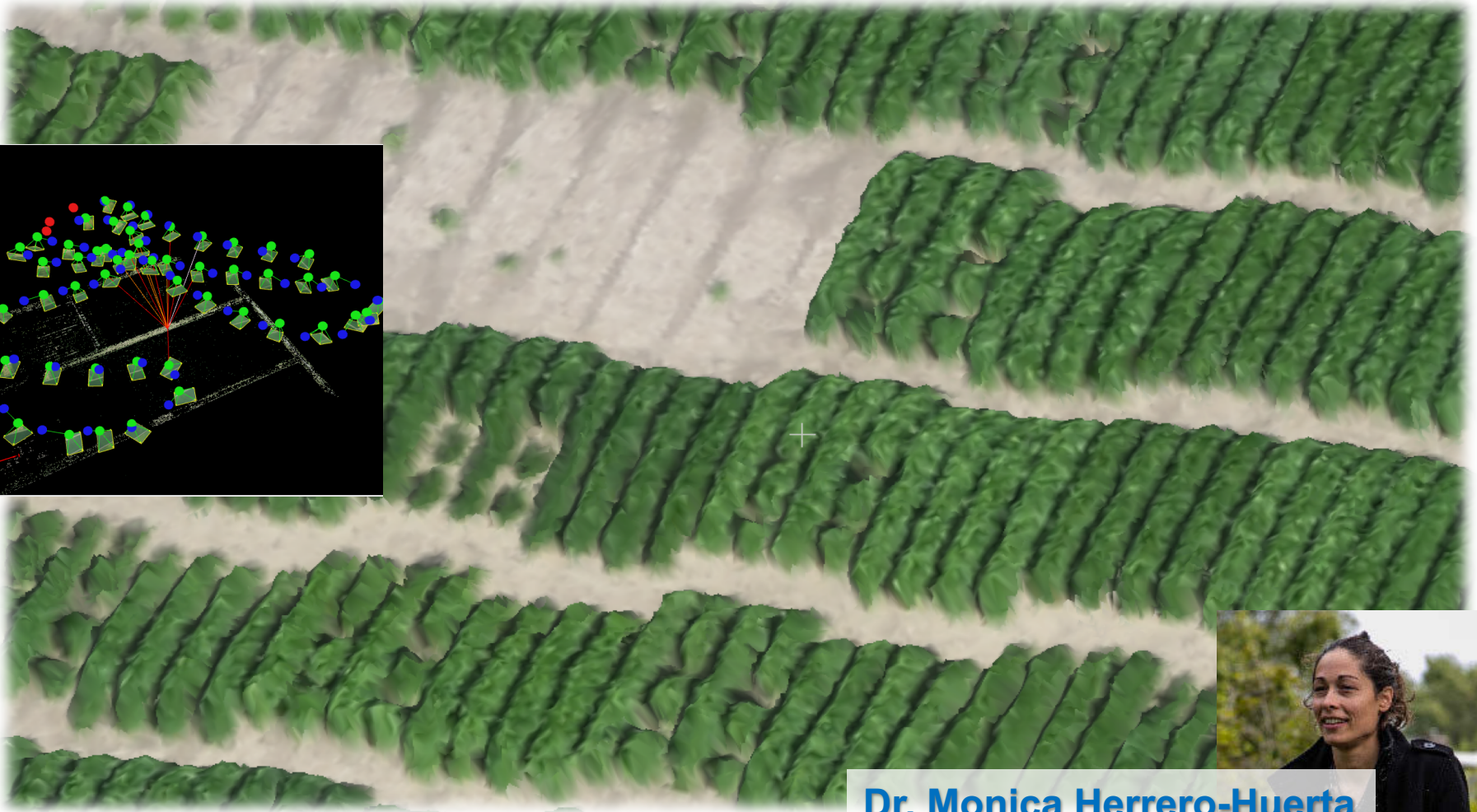
Dr. Monica Herrero-Huerta
**Rainey Lab**

Herrero-Huerta and Rainey (2019) High Throughput Phenotyping of
Physiological Growth Dynamics from UAS-based 3d Modeling In Soybean

3D Processing: Height Point Clouds

Gauss: mean = 180.939621 / std.dev. = 0.201415 [41 classes]

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
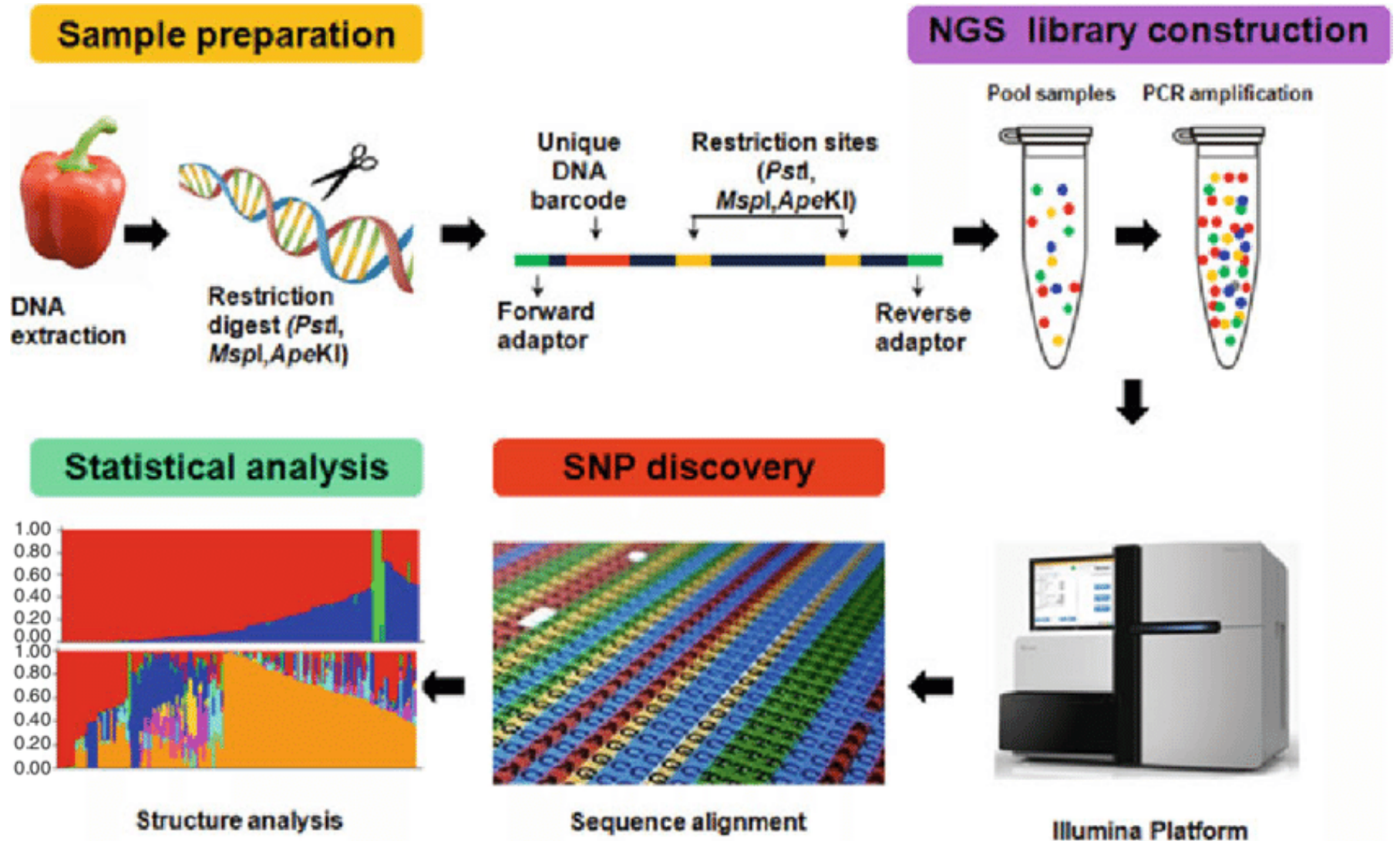
Genomic Selection/ Genomic Prediction/ Genomewide Selection

$$P = G + E$$
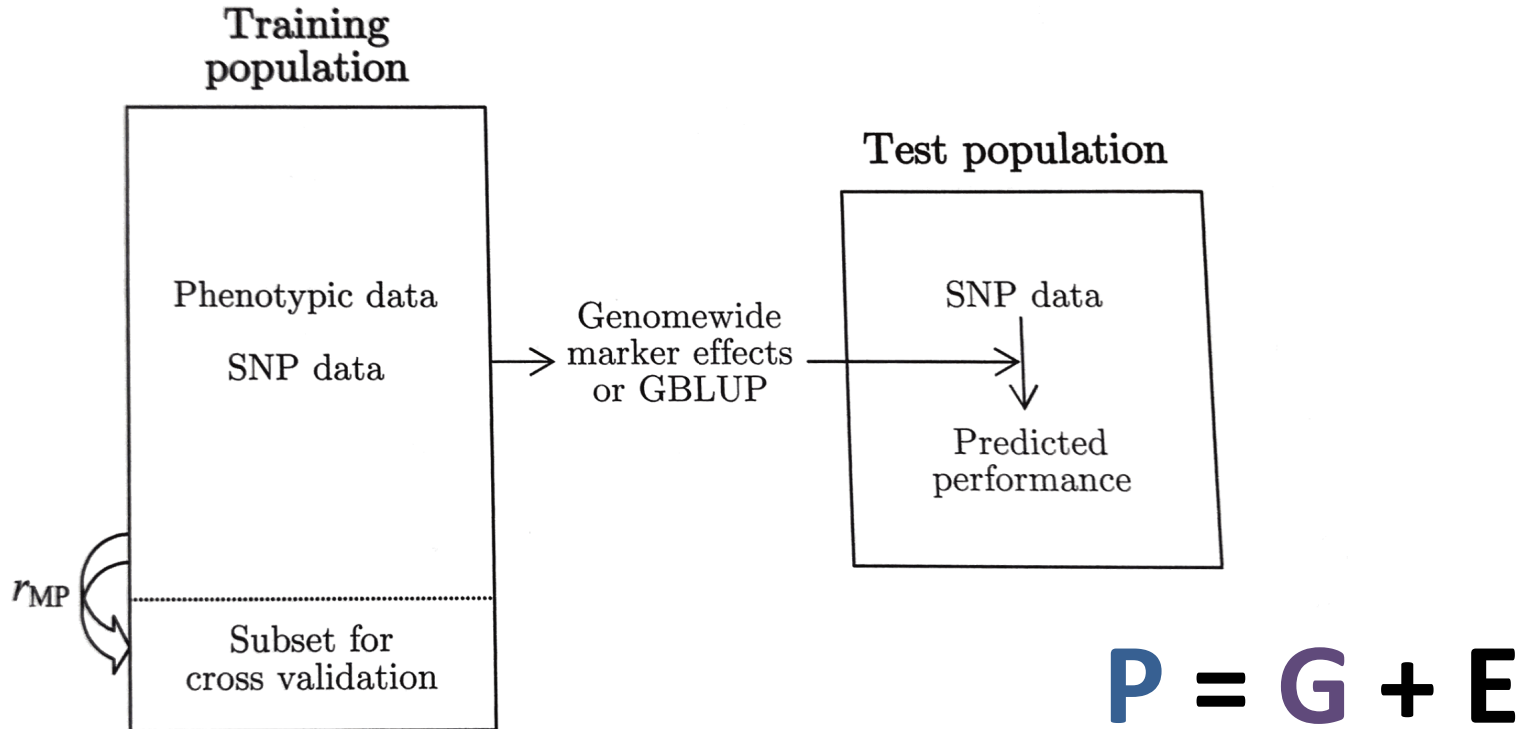
GENOMIC SELECTION: Use a large set of random markers directly to perform a marker-based *selection* using the predicted marker effects.

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

When to use GS

6. When there are too many candidates to phenotype (prediction of performance in a seedbank)
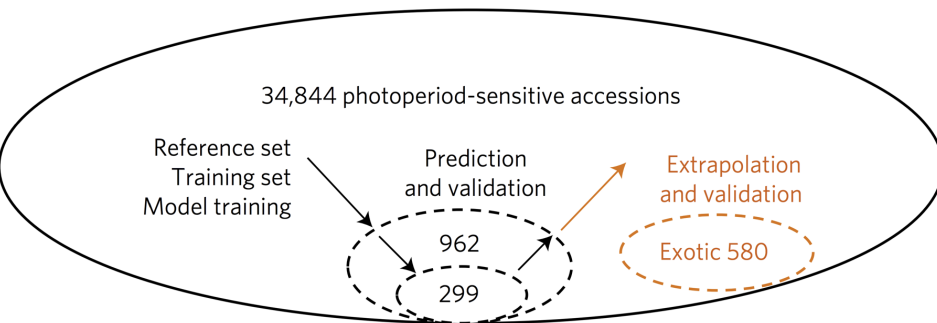
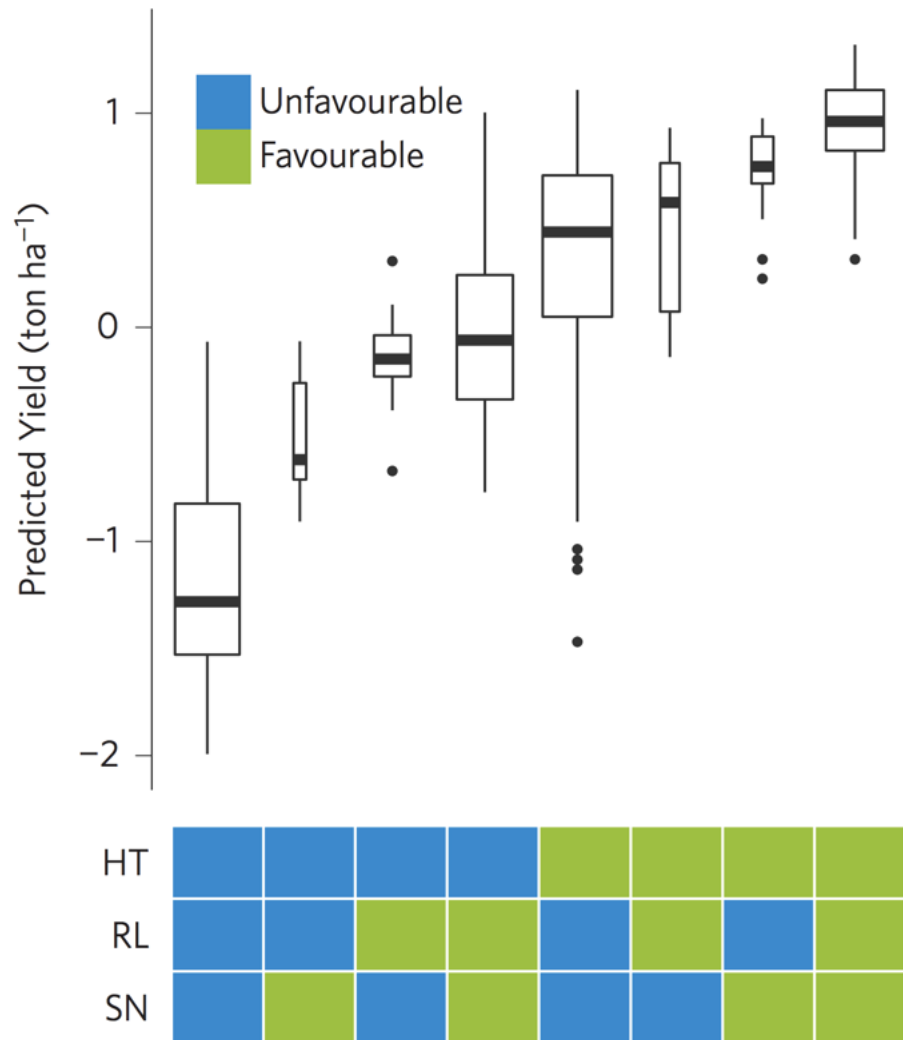Schematic overview of the project. Germplasm base, 962-accession reference set, 299-accession training set and prediction extrapolation

34,844 photoperiod-sensitive accessions

Reference set
Training set
Model training

Prediction and validation

962

299

Extrapolation and validation

Exotic 580

Unfavourable
Favourable

Predicted Yield (ton ha^{-1})

HT
RL
SN

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.

Yu, X., Li, X., Guo, T. *et al.* Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nature Plants* **2,** 16150 (2016) doi:10.1038/nplants.2016.150