UF IFAS
UNIVERSITY of FLORIDA

Getting the Most of my Mixed Models: Applications in Quantitative Genetics and Breeding

Salvador A Gezan
University of Florida, USA
Trigen – Canada

December, 2018

QG ANALYSES

- Understanding of the genetic architecture is critical to any breeding program as it defines the **Breeding Strategy**
- Relevant information includes:
  - *Genetic control (additive, dominance, epistasis).*
  - *Genotype-by-Environment (GxE).*
  - *Genotype-by-Year (GxY).*
  - *Trait-to-trait correlations.*
  - *Temporal correlations.*
  - *Spatial correlations.*
  - *Efficiency of Pedigree- or Molecular-based analyses.*
- All of these require parameters estimated by Linear Mixed Models.

UF IFAS UNIVERSITY of FLORIDA

OUTLINE

- Spatial Analyses
- Pedigree Information
- Optimal Design of Field Experiments
- Genomic Selection (Strawberry)
- Challenges with Genomic Selection

LINEAR MIXED MODELS

- **Linear Mixed Models** extend the linear model by allowing a more flexible specification of the errors (and other random factors). Hence, it allows for a different type of inference and also allows to incorporate *correlation* and *heterogeneous variances* between the observations.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e} \qquad E\begin{bmatrix}\mathbf{b}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix} \qquad Var\begin{bmatrix}\mathbf{b}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$$

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$$

$$Var(\mathbf{y}) = \mathbf{V} = \mathbf{V}(\boldsymbol{\theta}) = \mathbf{V}(\mathbf{y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

$$\mathbf{b} \sim \text{MVN}(0, \mathbf{G}) \quad \text{and} \quad \mathbf{e} \sim \text{MVN}(0, \mathbf{R})$$

UF IFAS
UNIVERSITY of FLORIDA

EARLY SELECTION TRIALS

Spatial Analyses
Pedigree Information

SPATIAL ANALYSIS

- It corresponds to an extension to the single vector repeated measures analysis.
- Incorporates information from physical positions (x and y coordinates).

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e}$$

$\mathbf{a} \sim \text{MVN}(0, \mathbf{G})$, $\mathbf{G} = \sigma_a^2\mathbf{I}$ or $\mathbf{G} = \sigma_a^2\mathbf{A}$

$\mathbf{e} \sim \text{MVN}(0, \mathbf{R})$, $\mathbf{R} = \sigma_e^2\mathbf{\Sigma}_x \otimes \mathbf{\Sigma}_y$

AR1⊗AR1

$\text{Var}(e_{ij}) = \sigma_e^2$

$\text{Cov}(e_{ij}, e_{i'j'}) = \sigma_e^2 \rho_x^{|hx|} \rho_y^{|hy|}$

AR1⊗AR1 + η

$\text{Var}(e_{ij}) = \sigma_e^2 + \sigma_{ms}^2$

$\text{Cov}(e_{ij}, e_{i'j'}) = \sigma_e^2 \rho_x^{|hx|} \rho_y^{|hy|}$

$$\mathbf{R} = \sigma_e^2 \begin{bmatrix} 1 & \rho_x^1 & \rho_x^2 & \rho_x^3 \\ \rho_x^1 & 1 & \rho_x^1 & \rho_x^2 \\ \rho_x^2 & \rho_x^1 & 1 & \rho_x^1 \\ \rho_x^3 & \rho_x^2 & \rho_x^1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & \rho_y^1 & \rho_y^2 & \rho_y^3 \\ \rho_y^1 & 1 & \rho_y^1 & \rho_y^2 \\ \rho_y^2 & \rho_y^1 & 1 & \rho_y^1 \\ \rho_y^3 & \rho_y^2 & \rho_y^1 & 1 \end{bmatrix} + \sigma_m^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

AUGMENTED DESIGNS (AD)

- Field experiments that allows testing several hundreds of genotypes with little or no replication.
- Most treatments (with the exception of controls or checks) have a **single** replication.

11	C2	24	112	23	69	C1	96	22	6	34	C1
85	101	48	C1	28	7	89	60	C2	108	74	56
47	C1	10	43	C2	16	52	5	38	33	C2	93
65	111	64	100	81	104	C2	78	C1	113	21	106
12	C2	44	68	42	C1	97	17	32	73	C1	35
25	C1	27	C2	15	88	29	4	53	C2	55	75
102	84	1	49	C1	61	70	C2	18	95	37	C1
46	86	C2	63	2	51	79	39	59	92	C2	57
66	13	C1	82	41	98	C2	90	C1	77	20	36
C1	45	83	87	C2	62	3	30	72	54	105	76
26	C2	9	14	50	8	40	C1	31	19	C2	C1
110	103	67	C1	99	80	C2	71	91	58	109	94

OBJECTIVE

- ✓ Evaluate the performance of augmented designs (AD) and double replication designs (DR) in an array of genetical and spatial scenarios.
- ✓ Compare the effects of different levels of spatial correlation (with and nugget) on the estimation of genetic parameters.
- ✓ Compare traditional against spatial analyses.
- ✓ Evaluate the ‘benefits’ of incorporation pedigree information into the model.

UF IFAS UNIVERSITY of FLORIDA

MATERIALS AND METHODS

- Simulation of a field: 1,024 plots (64 rows x 16 columns)
- Incorporation of surface with correlated errors (AR1) with and without nugget.
- Base heritability: $H^2 = 0.50$
- Values for ρ_x and ρ_y varied from 0.02 and 0.98 and $|\rho_x - \rho_y| < 0.85$
- Nugget (η) ranged between 30% and 70%.
- 500 simulations of each scenario.

Simulations for Clonal Value Estimation

$$\boldsymbol{y = 1\mu + Z_1 b + Z_2 g + e}$$

Designs	Proportion Control Plots	# Plots		# Genotypes		# Blocks
		Control	Test	Control	Test	
AD6.25	6.25%	64	960	4	960	8
AD12.5	12.5%	128	896	4	896	8
AD25	25%	256	768	4	768	8
DR	0%	0	512	0	512	2

STATISTICAL ANALYSES

Models Fitted

M1: No-spatial analysis

M2: Spatial without Nugget - AR1⊗AR1

M3: Spatial with Nugget - AR1⊗AR1+ η

Goodness-of-fit Statistics

- *CorPc:* correlation between true and predicted clonal value
- H_c^2 broad-sense heritability (H_c^2);
- logREML: Log Likelihood (LogREML);
- $\mathrm{H^2_{PEV}} = 1 - \frac{\overline{PEV}}{\sigma_g^2}$ and $\mathrm{h^2_{PEV}} = 1 - \frac{\overline{PEV}}{\sigma_a^2}$
- SEF: Selection Efficiency
- PM: True Genetic Gain

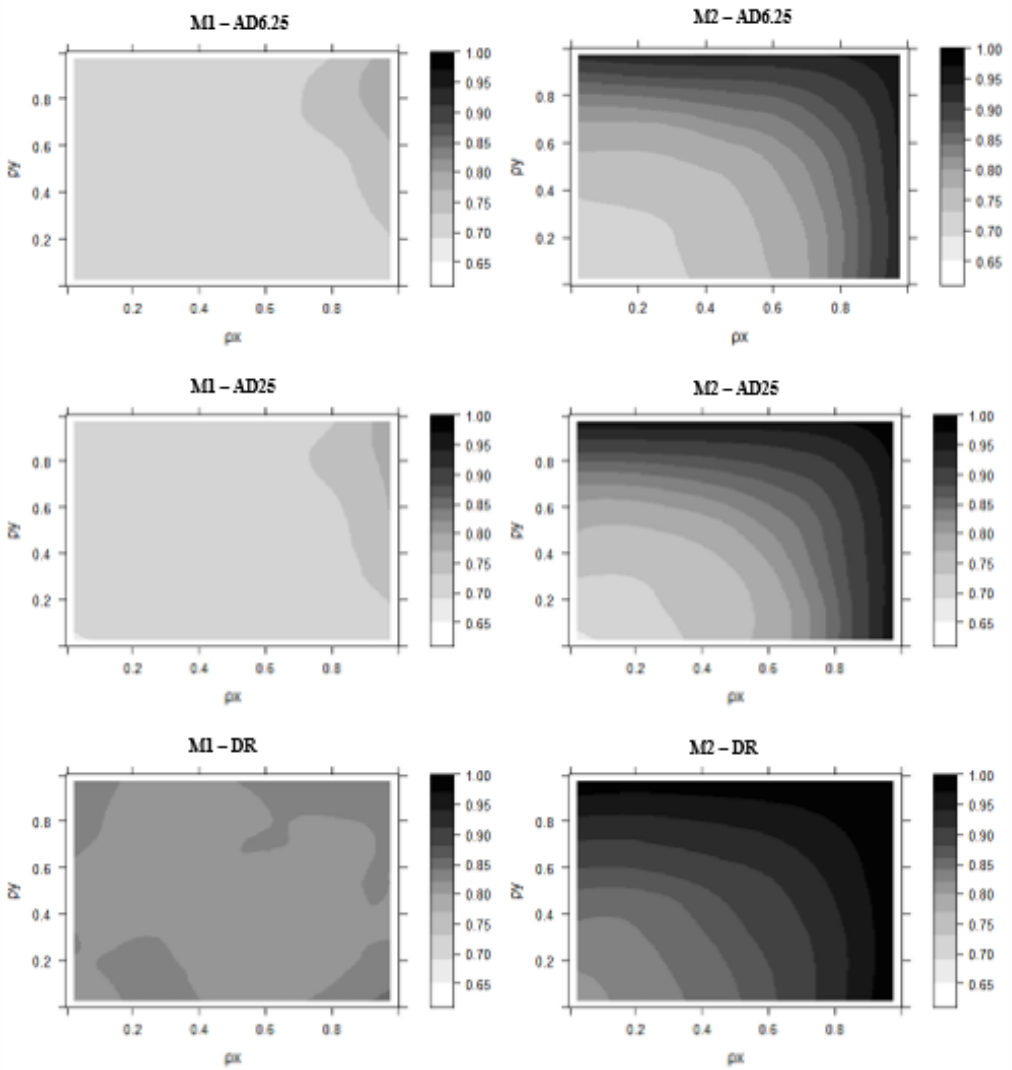
UF IFAS UNIVERSITY of FLORIDA

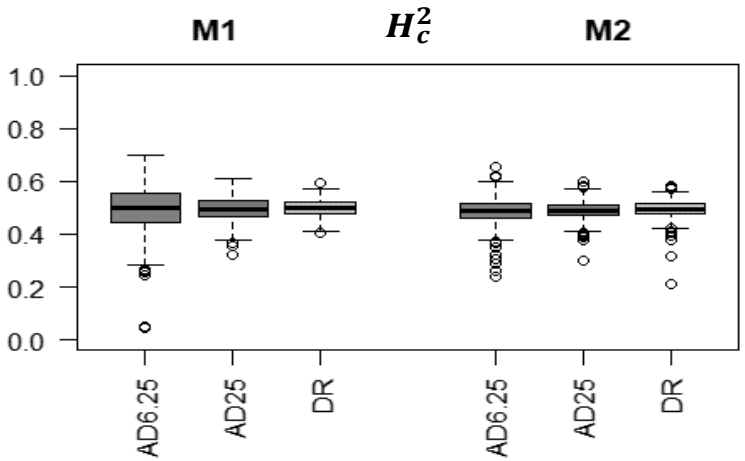
RESULTS

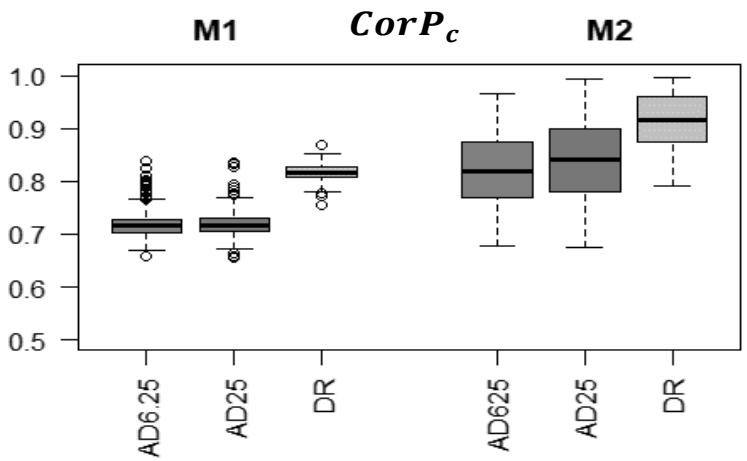
Sites with Nugget							
Designs	**Models**	**logREML**	**H^2_c**	**H^2_{PEV}**	**$CorP_c$**	**SEF_{40}**	**PM_{40}**
AD6.25	**M1**	-490.32	0.503	0.507	0.711	71.1%	11.52
	M2	-465.42	0.552	**0.612**	0.736	73.6%	11.57
	M3	**-460.15**	0.499	0.544	**0.740**	**74.0%**	**11.58**
AD12.5	**M1**	-470.13	0.501	0.504	0.713	71.1%	11.50
	M2	-443.35	0.533	**0.590**	0.737	73.4%	11.55
	M3	**-436.42**	0.501	0.548	**0.743**	**74.0%**	**11.56**
AD25	**M1**	-425.56	0.494	0.498	0.711	71.0%	11.45
	M2	-395.66	0.514	**0.567**	0.738	73.7%	11.51
	M3	**-386.28**	0.493	0.542	**0.744**	**74.3%**	**11.52**
DR	**M1**	-437.55	0.499	0.660	0.817	81.8%	11.76
	M2	-396.98	0.506	**0.713**	0.840	83.8%	11.81
	M3	**-384.62**	0.497	0.705	**0.845**	**84.6%**	**11.82**

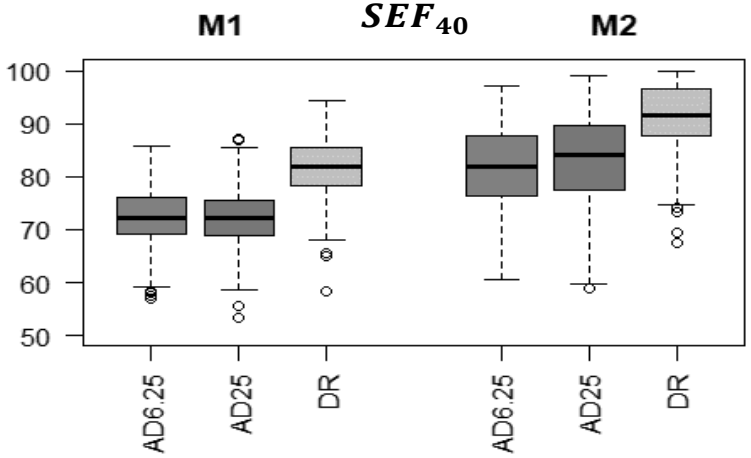
RESULTS

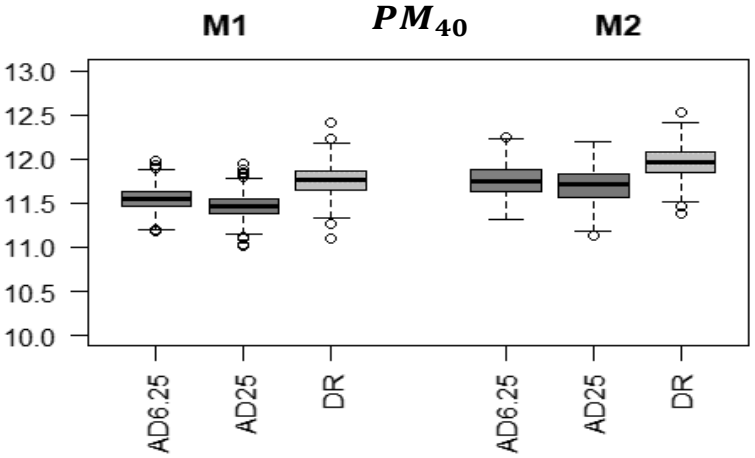
CorPc




M1
H_c^2
M2
1.0
0.8
0.6
0.4
0.2
0.0
AD6.25
AD25
DR
AD6.25
AD25
DR

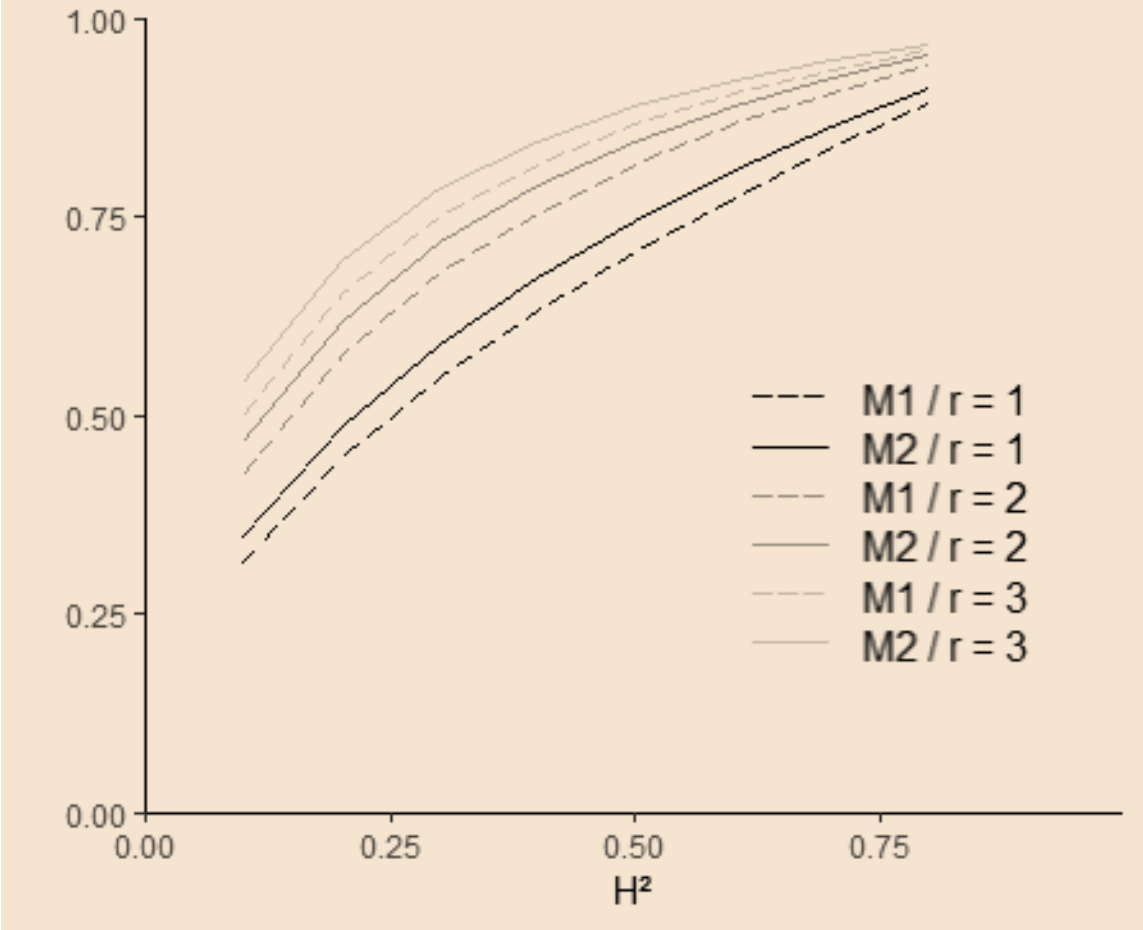

M1
$CorP_c$
M2
1.0
0.9
0.8
0.7
0.6
0.5
AD6.25
AD25
DR
AD625
AD25
DR


M1
SEF_{40}
M2
100
90
80
70
60
50
AD6.25
AD25
DR
AD6.25
AD25
DR


M1
PM_{40}
M2
13.0
12.5
12.0
11.5
11.0
10.5
10.0
AD6.25
AD25
DR
AD6.25
AD25
DR

EXTENSION TO MORE REPLICATIONS

$CorP_c$

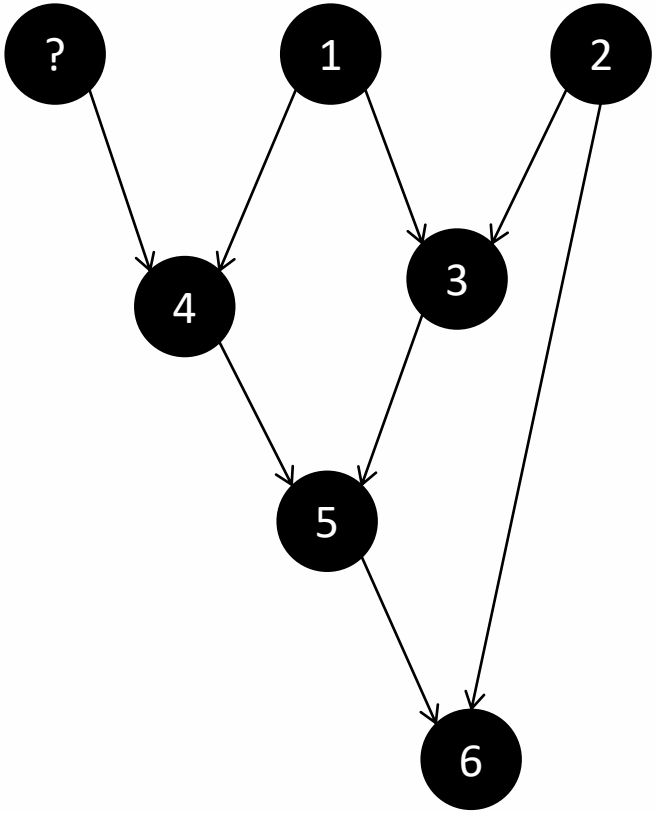


INCORPORATING PEDIGREE

- Why worry about the pedigree in genetic analyses?
  - Statistically, random genetic effects (i.e. BLUPs) are not independent and their matrix of correlations or co-variances (**G** or **A**) needs to be specified.
  - Genetically, it is important to consider information about relatives as they will share some alleles, and therefore their response is correlated.
- How to incorporate this information?
  - *Genetic relationships* can be calculated using genetic theory (expected values) or molecular information (e.g. SNPs), and included into the linear mixed model by specifying a pedigree file,

Example

Pedigree of a group of individuals



P-BLUP

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e}$$

$\mathbf{a} \sim \text{MVN}(0, \mathbf{G})$, $\mathbf{G} = \sigma_a^2\mathbf{A}$

$\mathbf{e} \sim \text{MVN}(0, \mathbf{R})$, $\mathbf{R} = \sigma^2\mathbf{I}_n$

$$\mathbf{A} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & 4 & 5 & 6 \end{array} \\ \begin{bmatrix} 1.00 & 0.00 & 0.50 & 0.50 & 0.50 & 0.25 \\ & 1.00 & 0.50 & 0.00 & 0.25 & 0.625 \\ & & 1.00 & 0.25 & 0.625 & 0.563 \\ & & & 1.00 & 0.625 & 0.313 \\ & & & & 1.125 & 0.688 \\ & & & & & 1.125 \end{bmatrix} \end{array}$$

UF IFAS UNIVERSITY of FLORIDA

MATERIALS AND METHODS

Simulations for Breeding Value Estimation

- Same field conditions as before: 1,024 plots (64 rows x 16 columns)
- Incorporation of pedigree information.
- Breeding values (circular diallel with 42 parents – 64 families)
- 2 scenarios: E1 ($h^2 = 0.40$, $d^2 = 0.05$ and $i^2 = 0.05$) and
  E2 ($h^2 = 0.20$, $d^2 = 0.15$ and $i^2 = 0.15$)

Designs	# Plots		#Family		# Blocks
	Control	Test	Families	Clones per family	
AD6.25	64	960	64	15	8
AD25	256	768	64	12	8
DR	0	512	64	8	2

$$\boldsymbol{y = 1\mu + X\beta + Z_1 b + Z_2 a + Z_3 f + Z_4 c + e}$$

UF IFAS UNIVERSITY of FLORIDA

RESULTS

Scenario E1											
Designs	**Models**	**logREML**	**h^2**	**d^2**	**i^2**	**d^2+i^2**	**H^2_c**	**h^2_{PEV}**	**$CorP_a$**	**SEF_{40}**	**PM_{40}**
AD6.25	**M1**	-413.477	0.403	-	-	0.102	0.505	0.569	0.748	73.9%	11.56
	M2	-379.623	0.401	-	-	0.165	0.566	0.592	0.760	75.1%	**11.58**
	M3	**-372.732**	0.402	-	-	0.098	0.500	**0.593**	**0.762**	**75.3%**	**11.58**
AD25	**M1**	-366.454	0.400	-	-	0.096	0.496	0.560	0.742	72.9%	11.47
	M2	-327.722	0.399	-	-	0.116	0.515	0.583	0.754	74.1%	**11.49**
	M3	**-316.959**	0.400	-	-	0.092	0.491	**0.587**	**0.757**	**74.5%**	**11.49**
DR	**M1**	-386.214	0.392	0.062	0.043	0.105	0.497	0.607	0.788	77.7%	11.66
	M2	-341.244	0.392	0.061	0.053	0.114	0.506	0.630	0.801	79.1%	**11.69**
	M3	**-328.288**	0.392	0.060	0.045	0.105	0.496	**0.635**	**0.804**	**79.5%**	**11.69**
Scenario E2											
Designs	**Models**	**logREML**	**h^2**	**d^2**	**i^2**	**d^2+i^2**	**H^2_c**	**h^2_{PEV}**	**$CorP_c$**	**SEF_{40}**	**PM_{40}**
AD6.25	**M1**	-453.743	0.251	-	-	0.250	0.501	0.472	0.627	61.7%	11.31
	M2	-425.080	0.250	-	-	0.313	0.562	0.488	0.635	62.6%	11.32
	M3	**-418.964**	0.252	-	-	0.248	0.501	**0.491**	**0.636**	**62.8%**	**11.33**
AD25	**M1**	-399.221	0.246	-	-	0.250	0.496	0.455	0.616	60.6%	11.23
	M2	-364.351	0.243	-	-	0.274	0.517	0.470	0.626	61.5%	**11.25**
	M3	**-354.502**	0.245	-	-	0.250	0.495	**0.474**	**0.628**	**61.7%**	**11.25**
DR	**M1**	-413.458	0.205	0.148	0.146	0.294	0.499	0.405	0.648	63.5%	11.38
	M2	-370.898	0.203	0.147	0.155	0.302	0.506	0.416	0.656	64.3%	11.39
	M3	**-358.246**	0.203	0.147	0.146	0.293	0.497	**0.420**	**0.658**	**64.6%**	**11.40**

UF IFAS
UNIVERSITY of FLORIDA

‘OPTIMAL’ DESIGN OF FIELD EXPERIMENTS

Spatial Correlations
Pedigree Information

OBJECTIVES

1. To develop and evaluate statistical methods and computational procedures to generate improved randomized complete block (RCB) designs using mixed models.
2. To evaluate the efficiency of proposed search algorithms to generate improved designs using mixed models.
3. To develop and evaluate statistical and computational procedures to generate improved unbalanced (non-orthogonal) designs: unequally-replicated, incomplete block (IB) and unreplicated designs.

Mixed Models (RCBD)

$$\begin{aligned} \mathbf{y} &= \mathbf{1}\mu + \mathbf{W}\gamma + \mathbf{Z}\mathbf{g} + \mathbf{e} \quad (1) \\ &= \mathbf{X}\beta + \mathbf{Z}\mathbf{g} + \mathbf{e}, \quad \text{where} \\ \mathbf{X} &= \begin{bmatrix} \mathbf{1} & \mathbf{W} \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \mu \\ \gamma \end{bmatrix} \end{aligned}$$

- $\mathbf{y}$: response obs.
- $\mathbf{1}$: a column of ones
- μ: overall mean
- $\mathbf{W}$: design matrix of fixed effects
- γ: a vector of fixed effects
- $\mathbf{X}$: design matrix of fixed effects

continued ...

- β: a vector of fixed effects
- $\mathbf{Z}$: design matrix of random effects
- $\mathbf{g}$: a vector of random effects
- $\mathbf{g} \sim N(0, \mathbf{G})$, where $\mathbf{G} = \sigma_g^2 \mathbf{A}$
- $\mathbf{e} \sim N(0, \mathbf{R})$, where $\mathbf{R} = \sigma_e^2 \theta$, θ is a spatial correlation matrix,
- $\mathbf{R} = \sigma_e^2 \Sigma_r(\rho_r) \otimes \Sigma_c(\rho_c)$
- assume: $\mathbf{g} \perp \mathbf{e}$

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z}+\hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z}+\hat{\mathbf{G}}^{-1} \end{bmatrix}^{-} \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}^{-} \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix} \tag{3}$$

$$= \begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix} = \begin{bmatrix} (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \\ \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y}-\mathbf{X}[\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}) \end{bmatrix}$$

$$\begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \\ \hat{\mathbf{G}}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\hat{\beta}) \end{bmatrix} \tag{4}$$

where by using theorem 8.5.11 from Harville (1997), Hooks et al (2009) showed that,

$$\mathbf{C}^{22} = \mathbf{M}(\Omega) = Var(\hat{\mathbf{g}}-\mathbf{g}) = (\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z}+\hat{\mathbf{G}}^{-1}-\mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z})^{-1} \tag{5}$$

OPTIMALITY CRITERIA

- Let **M** be a variance covariance matrix of treatment effects, for a design Ω.
- (John and Williams, 1995; Butler et al, 2008; Kuhfeld, 2010)

$A_{opt} = min\{trace(\mathbf{M}(\Omega))\}$ or $A_{opt} = min\{trace(\Sigma_g(\Omega))\}$

  - minimize the average variance of the treatment effects

- $D_{opt} = min\{|\mathbf{M}(\Omega)|\}, \quad \{ \text{ for } \quad |\mathbf{M}(\Omega)| \neq 0.$ or

$min|\Sigma_g(\Omega)|$

  - minimize the generalized variance or the volume of an ellipsoid described by $\mathbf{M}(\Omega)$

SIMULATION SCENARIOS

- RCB scenarios: $\Omega_A^{(30)}$, $\Omega_D^{(30)}$, with six blocks,
- $\Omega_A^{(196)}$, with 16 blocks and four blocks,
- Spatial correlations, ρ, of 0.0, 0.3, 0.6, 0.9
- Genetic relatedness: independent, half-sib and full-sib families.
- Simple pairwise algorithm
- Narrow-sense heritabilies, h^2, of 0.1, 0.3, 0.6
- Blocks (fixed effects) and treatments (random effects)
- Residual errors modeled using a 2-dimensional AR1:
  $\mathbf{R} = \sigma_e^2 \boldsymbol{\Sigma}_r(\rho_r) \otimes \boldsymbol{\Sigma}_c(\rho_c)$
- Data simulated for prediction and estimation of treatment effects

RESULTS: RCBD

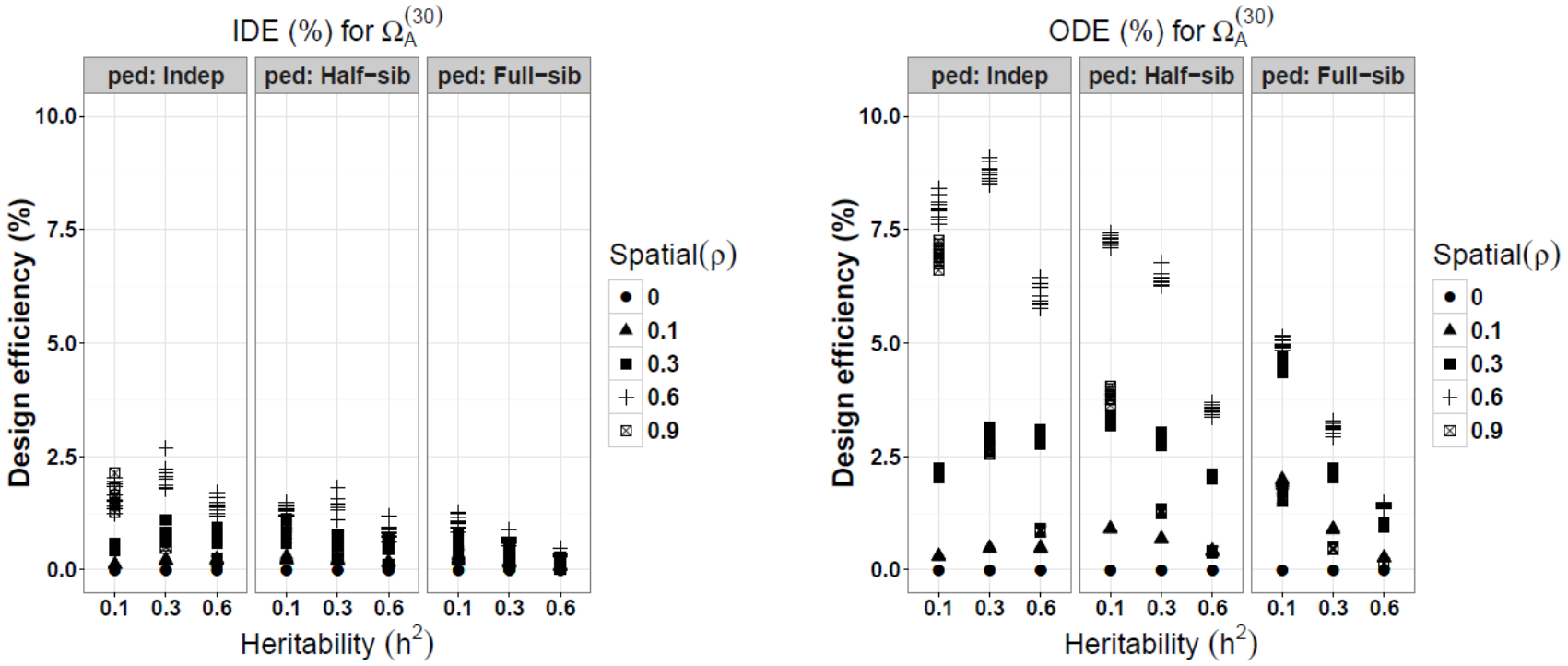


Figure 1: Initial design efficiencies and overall design efficiencies for $\Omega_A^{(30)}$, with $\lambda = 10$, $m = 100$ and $p = 5,000$ iterations.

RESULTS: UR and IB Designs

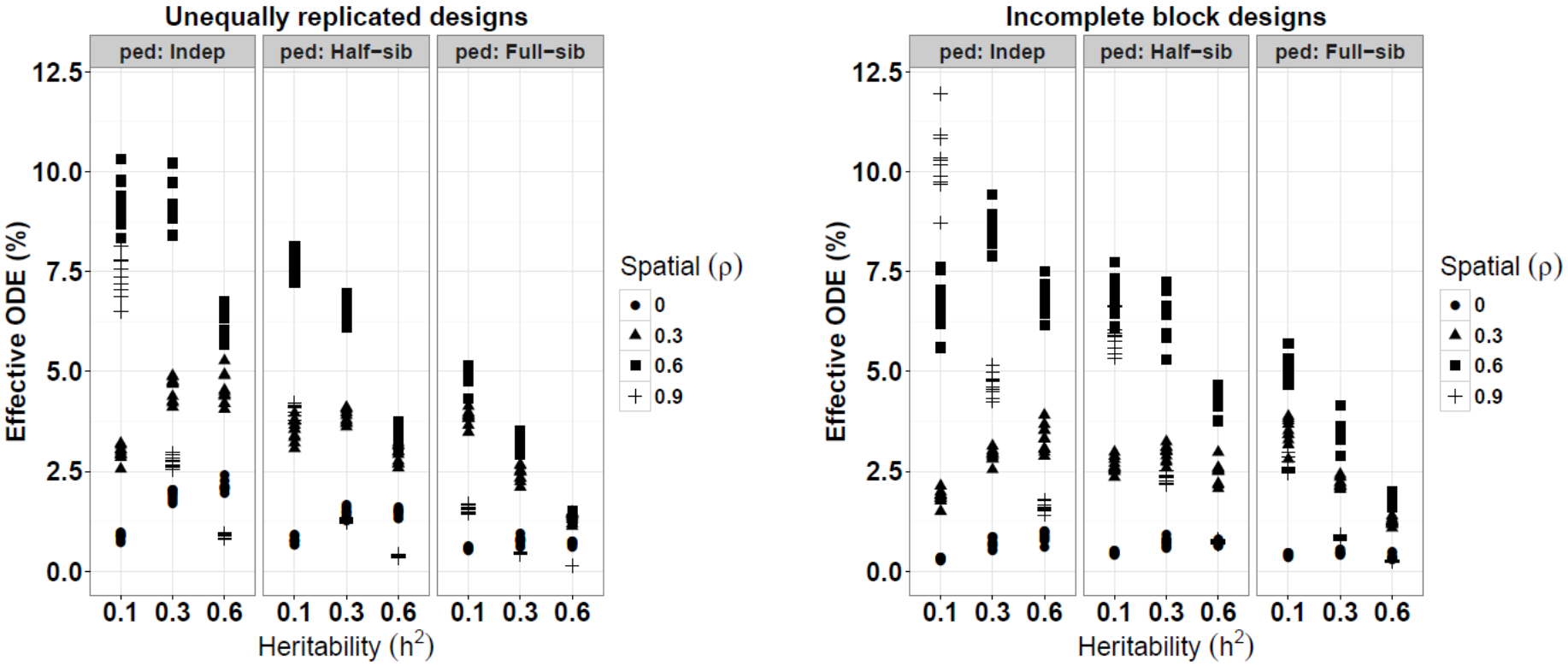


Figure 6: Effective overall design efficiencies for $\Omega_A^{(30)}$ unequally replicated and incomplete block designs.

RESULTS: AUGMENTED Designs

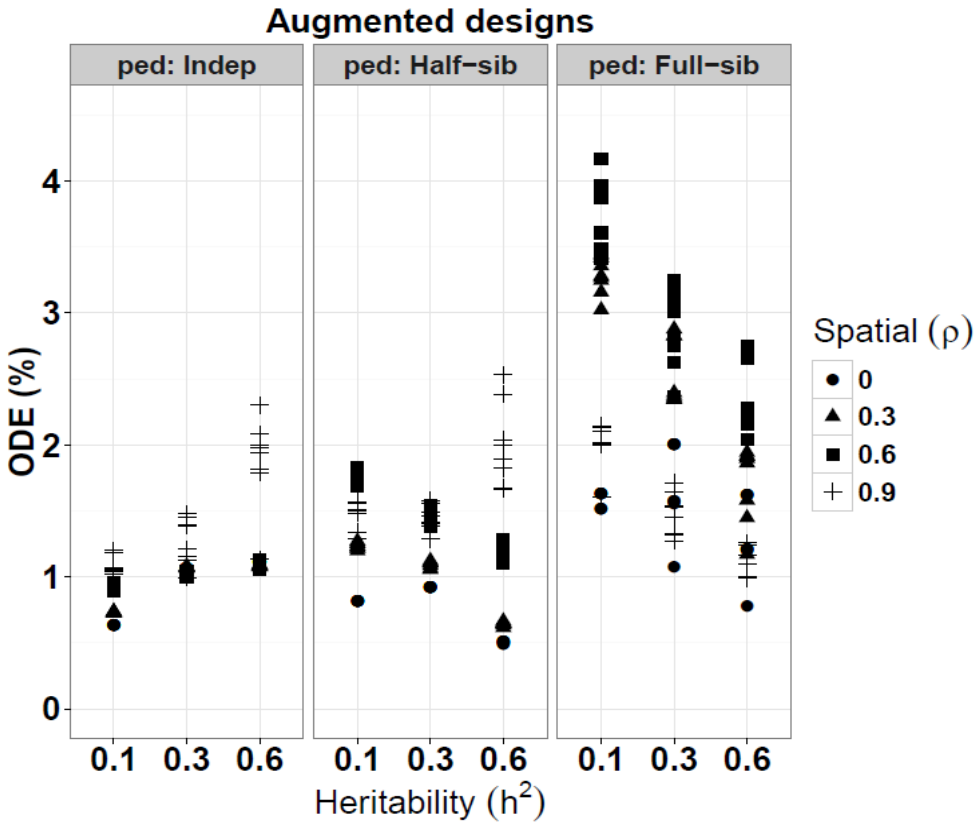


Figure 7: Overall design efficiencies for augmented designs based on $\Omega_A^{(492)}$ unreplicated test treatments and three replicated controls.

UF IFAS
UNIVERSITY of FLORIDA

GENOMIC SELECTION

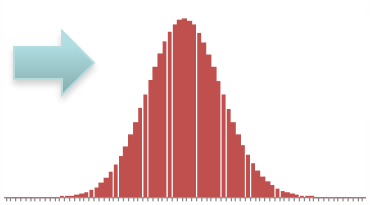
Animal Model
Molecular Information
(Pedigree Information)

UF IFAS UNIVERSITY of FLORIDA

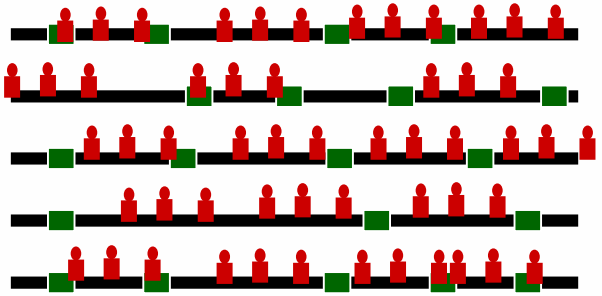
GENOMIC SELECTION

- Construct prediction models using the current breeding population phenotype and molecular markers capturing most of the quantitative variation.

Quantitative phenotypic data



Genotypic data



Breeding Value (BV) + Molecular Markers

Model construction: $a_i = \mu + \sum_{j=1}^{p} x_{ij}\beta_j + e_i$

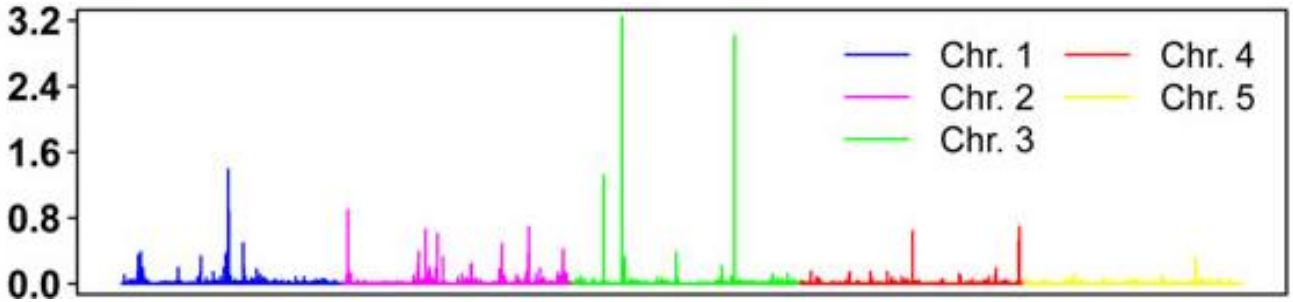
$$\mathbf{a} = \mathbf{X\beta} + \mathbf{e}$$

GENOMIC SELECTION

- Future individuals are genotyped, marker information is used as input on prediction models to select superior genotypes in the next cycles:

$$\hat{\mathbf{a}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

$$\mathbf{X} = \begin{array}{c} \\ g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{bmatrix} 1 & 0 & 1 & 2 \\ 2 & 2 & 0 & 2 \\ 2 & 1 & 1 & 0 \\ 0 & 2 & 2 & 1 \\ 0 & 0 & 2 & 1 \end{bmatrix} \end{array} \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.24 \\ 0.02 \\ -0.08 \\ 0.14 \end{bmatrix} \qquad \hat{\mathbf{a}} = \begin{bmatrix} 0.44 \\ 0.80 \\ 0.42 \\ 0.02 \\ -0.02 \end{bmatrix}$$



GBLUP

- GBLUP replaces the *pedigree-based* relationship matrix **A** by the molecular-based relationship matrix $\mathbf{G_A}$
- It is equivalent to RR-BLUP but it can be used for complex data (e.g. MET).

$$y_i = \mu + a_i + e_i$$

$$\hat{\mathbf{a}} = \mathbf{X}\hat{\boldsymbol{\beta}}$$

- If the markers are capturing all genetic variation, then we can assume that:
- If we also assume: $V(\boldsymbol{\beta}) = \mathbf{I}\,\sigma^2_\beta$
- Then we get: $V(\mathbf{a}) = \mathbf{X}\,\mathbf{X}'\sigma^2_m$
- An by scaling: $V(\mathbf{a}) = \mathbf{X}\mathbf{X}'\dfrac{\sigma^2_a}{\sum_i 2\,p_i\,q_i} = \mathbf{G}_{\mathrm{A}}\,\sigma^2_a$

MOLECULAR-BASED RELAT. MATRIX

- The **numerator relationship matrix (A)** is derived from pedigree.
- The **realized relationship matrix (G_A)** is derived from molecular markers.
- G_A is also known as **observed relationship matrix** or **genomic matrix.**

$$\mathbf{A} = \begin{bmatrix} 1 & 0.50 & 0.25 & 0.00 \\ 0.50 & 1 & 0.25 & 0.00 \\ 0.25 & 0.25 & 1 & 0.25 \\ 0.00 & 0.00 & 0.25 & 1 \end{bmatrix} \Rightarrow \mathbf{G}_A = \begin{bmatrix} 0.98 & 0.42 & 0.23 & -0.02 \\ 0.42 & 0.99 & 0.26 & 0.01 \\ 0.23 & 0.26 & 1.03 & 0.20 \\ -0.02 & 0.01 & 0.20 & 0.99 \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Za} + \mathbf{e}$$

PBLUP

$\mathbf{a} \sim MVN(0, \mathbf{G})$, $\mathbf{G} = \sigma_a^2\mathbf{A}$

$\mathbf{e} \sim MVN(0, \mathbf{R})$, $\mathbf{R} = \sigma^2\mathbf{I}_n$

GBLUP

$\mathbf{a} \sim MVN(0, \mathbf{G})$, $\mathbf{G} = \sigma_a^2\mathbf{G}_A$

$\mathbf{e} \sim MVN(0, \mathbf{R})$, $\mathbf{R} = \sigma^2\mathbf{I}_n$

MOLECULAR-BASED RELAT. MATRIX

COMPUTING THE RELATIONSHIP MATRIX

$$\{a_G\}_{jk} = \begin{cases} \frac{1}{M}\sum_i \frac{(g_{ij}-2p_i)(g_{ik}-2p_i)}{2p_i(1-p_i)} & if \quad j \neq k \\ 1+\frac{1}{M}\sum_i \frac{g_{ij}^2-(1+2p_i)g_{ij}+2p_i^2}{2p_i(1-p_i)} & if \quad j = k \end{cases}$$

where,

$\{a_G\}_{jk}$ is the genomic additive relationship coefficient corresponding to individuals j and k,

M is the number of markers,

g_{ij} and g_{ik} are the numeric genotypic values of individual j and k at marker i

p_i is the frequency of allele with numeric value of 1 at marker i.

Coding g_{ij}:

AA → 0, AC → 1, CC→ 2

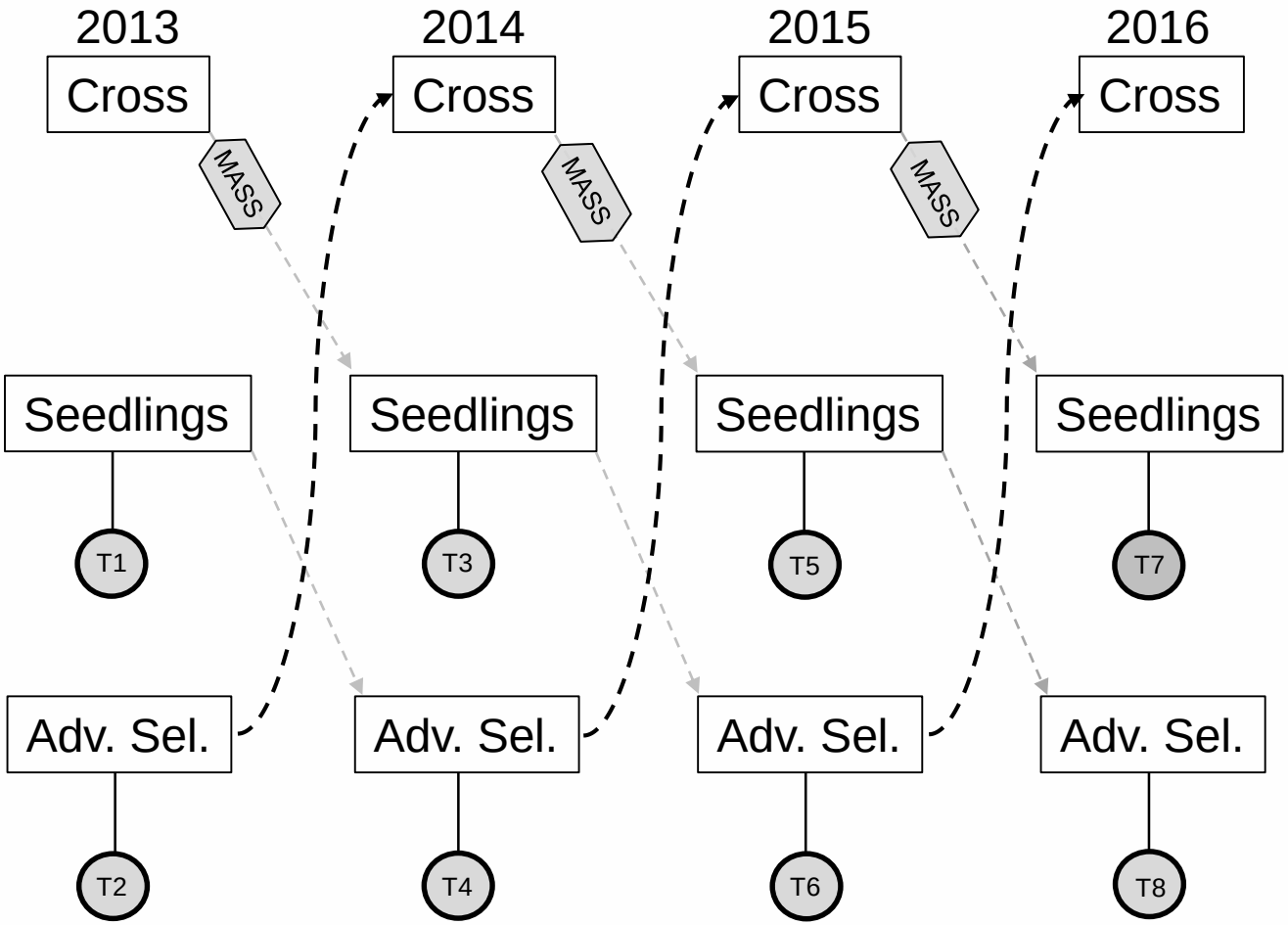
STRAWBERRY BREEDING PROGRAM UNIVERSITY OF FLORIDA

- Established in 1964
- Main traits:
    - Fruit size
    - Total yield
    - Early yield
    - Culling
    - Sugar content (Brix)
    - Disease resistance
- Interest in implementing Genomic Selection

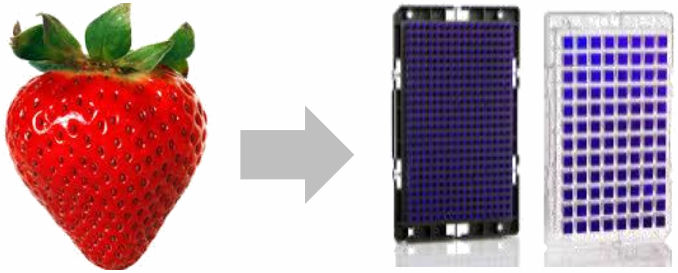
Objectives

- Explore different GS models for some of the most important traits in strawberry.
- Use and evaluate cross-validation and true-validation for prediction.
- Refine a breeding strategy that incorporates GS.
- Evaluate alternatives to improve the predictability of GS.

Breeding Strategy



Genotyping



90K IStraw90 Axiom® SNP array
Affimetrix & RosBREED

(Bassil and Davis et al. 2015)

Marker Quality Control

- 17,479 markers were used in GS
- Markers with MAF < 5% were removed
- Markers with missing values > 5% were removed
- Missing marker data was imputed by using the average allele frequency

Phenotyping

Traits	Measurements (per trial/season) No.	Units
Marketable Yield (TMY)	15	grams
Average Fruit Wt (AWT)	15	grams
Early Marketable Yield (EMY)	~4	grams
Total Culls (TC)	15	%
Brix (SSC)	5	%

Genomic Prediction Methods

- General Model: $y_i = u + \sum_{j=1}^{p} x_{ij}\,\beta_j + e_{ij}$

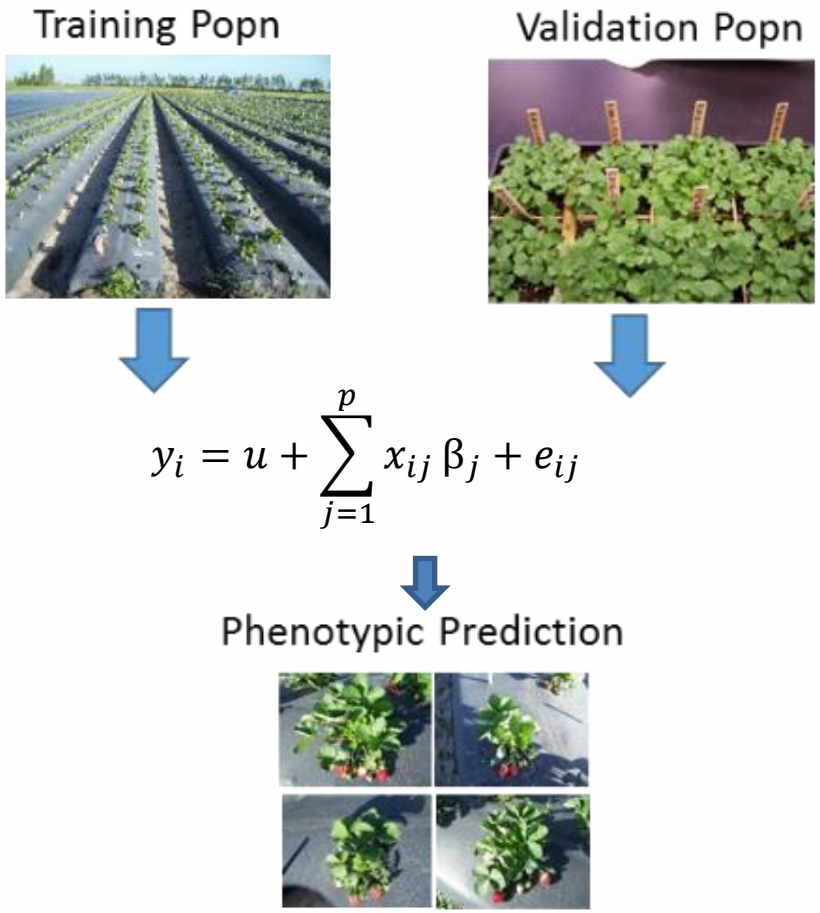
- G-BLUP: assumes marker effects have identical variance
  - $\mathbf{A_g}$ - GenoMatrix (Nazarian and Gezan, 2015), ASReml – R.

- Bayes B, Bayes C: markers have different variance effects
  - Gaussian mixture distribution of SNP's effects (π, df_β , S_β). BGLR (Perez and de los Campos, 2014)

- Reproducing kernel Hilbert spaces (RKHS):
  - It captures some non-additive effects. BGLR (Perez and de los Campos, 2014)

Genomic Prediction Model

- Predictions made using "training population"
- Predictions checked on "validation population" (independent breeding families)



Comparison of GS Methods

- 5-fold Cross-validation: Test data was randomly divided in 2 sets of data
  - Predictive Ability: $corr(y, \hat{y})$

- True-validation: Training and validation populations are from different tests
  - Predictive Ability: $corr(y, \hat{y})$
  - Efficiency of Selection: $\hat{y}_{inc}/\hat{y}_{com}$
  - Accuracy of selection: $corr(g, \hat{g}) = \frac{corr(y,\hat{y})}{h}$

Predictive Ability (T2-T4)

$$corr(y, \hat{y})$$

Trait	PBLUP	GS Models			
		GBLUP	Bayes B	Bayes C	RKHS
AWT	0.444	0.490	0.494	0.488	0.515
SSC	0.371	0.427	0.438	0.436	0.451
TMY	0.238	0.306	0.353	0.337	0.333
TC	0.139	0.320	0.350	0.352	0.318

AWT: average weight; SSC: soluble solids content;
TMYL total marketable yield; TC: total percent culls.

Prediction Accuracy (T2-T4)

$$corr(g, \hat{g}) = \frac{corr(y, \hat{y})}{h}$$

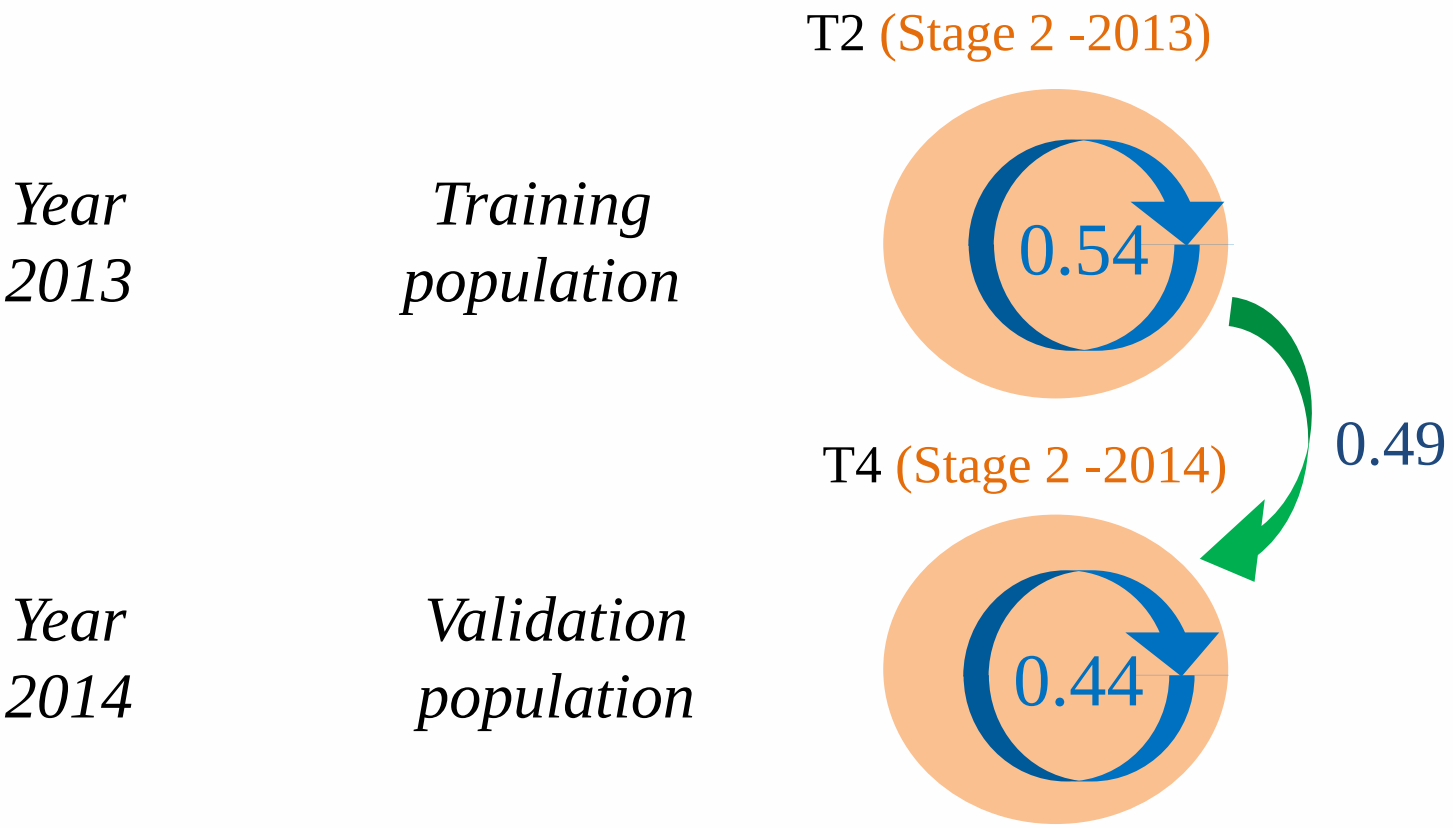
Trait	PBLUP	GS Models			
		GBLUP	Bayes B	Bayes C	RKHS
AWT	0.549	0.606	0.610	0.603	0.636
SSC	0.630	0.725	0.744	0.740	0.766
TMY	0.507	0.652	0.753	0.718	0.710
TC	0.159	0.365	0.400	0.402	0.363

AWT: average weight; SSC: soluble solids content;
TMYL total marketable yield; TC: total percent culls.

Predictive Ability for Cross- and True-Validation of AWT by RKHS

$corr(y, \hat{y})$

Predictive Ability for Cross Validation and True Validation of AWT



$$corr(y, \hat{y})$$

UF IFAS UNIVERSITY of FLORIDA

Conclusions

- The predictive ability of the studied traits was similar across the different GS methods (Bayes B slightly better).
- Prediction accuracy was high (>0.60) for all models and traits except for TC.
- The ability to predict phenotypic performance is linearly related to the heritability of the trait.
- The efficiency of selection was high (>83%) for prediction of T4 based on T2.
- GxY has an important influence on the efficiency of GS
- Thanks LMM!

BETTER GENOMIC SELECTION

Additive + Dominance Model
Multiple-year data

GBLUP with non-additive effects

Additive + Dominance Model

$$y_i = \mu + a_i + d_i + e_i$$

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{d} + \mathbf{e}$$

β vector of fixed effects

a vector of random additive effects (i.e. BV), ~ N(0, $\mathbf{G_A}\sigma^2_a$)

d vector of random dominance effects, ~ N(0, $\mathbf{G_D}\sigma^2_d$)

e vector of random residual effects, ~ N(0, $\mathbf{I}\sigma^2$)

Note:

- The variance-covariance matrix $\mathbf{G_A}$ and $\mathbf{G_D}$ are derived from molecular markers.

GENOMIC MATRICES

- **Additive Relationship Matrix (G_A)**

$$\{a_G\}_{jk} = \begin{cases} \frac{1}{M}\sum_i \frac{(g_{ij}-2p_i)(g_{ik}-2p_i)}{2p_i(1-p_i)} & if \quad j \neq k \\ 1 + \frac{1}{M}\sum_i \frac{g_{ij}^2-(1+2p_i)g_{ij}+2p_i^2}{2p_i(1-p_i)} & if \quad j = k \end{cases}$$

- **Dominance Relationship Matrix (G_D)**

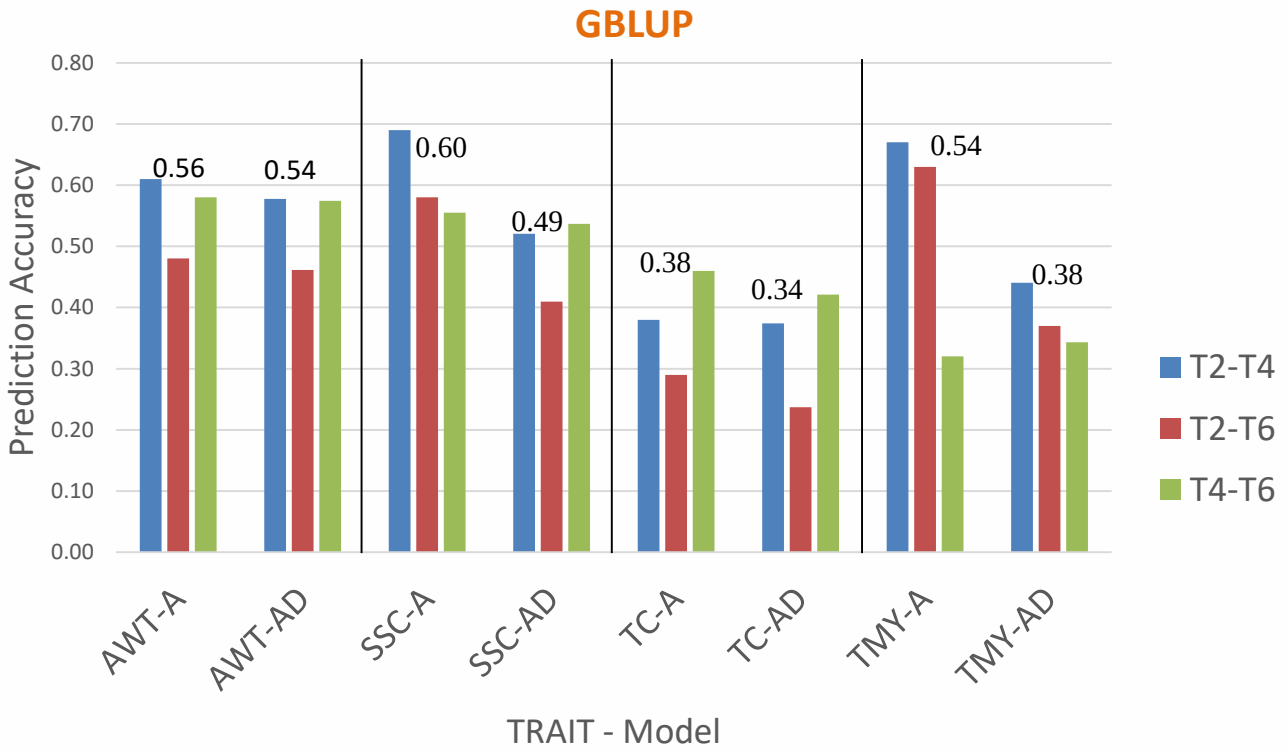
$$\{h\}_{ij} = \begin{cases} -2p_i^2 & if\ g_{ij} = 0 \\ 2p_iq_i & if\ g_{ij} = 1 \\ -2q_i^2 & if\ g_{ij} = 2 \end{cases}$$

$$\{h\}_{ij} = \begin{cases} -2p_iq_i & if\ g_{ij} = 0 \\ 1-2p_iq_i & if\ g_{ij} = 1 \\ -2p_iq_i & if\ g_{ij} = 2 \end{cases}$$

$$D_G = \frac{HH'}{\sum_i (2p_iq_i)^2}$$

$$D_{G^*} = \frac{HH'}{\sum_i 2p_iq_i(1-2p_iq_i)}$$

Prediction Accuracies Using GBLUP and Two Genetic Models (A and AD) on True-Validation



POPULATION SIMULATION FOR A+D

- Five chromosomes, each of 100 cMorgan in length, with 1,100 polymorphic variants (1,000 markers and 100 QTLs) distributed across each chromosome.
- A total of 20 subsets were randomly selected from the base population, each containing 100 males and 100 females (**G1** population) to generate a total of 300 full-sib families (10 offspring per family).
- Phenotypic values were assigned to G2 individuals as the summation of population mean (supposed to be 100), additive genetic effects, dominance genetic effects, and residual error.
- Different scenarios were considered:

$$y_i = \mu + a_i + d_i + e_i$$

Scenario	h^2	d^2	σ^2_a	σ^2_d	σ^2_e
1	0.4	0	1	0	1.5
2	0.4	0.1	1	0.25	1.25
3	0.4	0.4	1	1	0.5

UF IFAS UNIVERSITY of FLORIDA

SIMULATION STUDY

Additive, Dominance and Genetic Values Correlations

Model	Additive Values Correlation			Dominance Values Correlation			Genetic Values Correlation		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
Full Datasets									
A	0.756	0.735	0.697	-	-	-	0.756	0.771	0.862
A_G	0.789	0.770	0.728	-	-	-	0.789	0.784	0.817
A+D	0.756	0.736	0.701	-	0.348	0.653	0.755	0.776	0.903
A_G+D	0.787	0.768	0.726	-	0.343	0.650	0.789	0.802	0.906
A_G+D_G	0.788	0.769	0.730	-	0.321	0.592	0.791	0.799	0.881
A_G+D_{G*}	0.788	0.769	0.730	-	0.321	0.592	0.791	0.799	0.881
Partial Datasets									
A	0.736	0.717	0.687	-	-	-	0.736	0.758	0.861
A_G	0.750	0.734	0.703	-	-	-	0.750	0.759	0.827
A+D	0.736	0.717	0.689	-	0.340	0.646	0.734	0.760	0.899
A_G+D	0.748	0.732	0.701	-	0.338	0.644	0.749	0.771	0.901
A_G+D_G	0.748	0.733	0.703	-	0.325	0.610	0.750	0.771	0.886
A_G+D_{G*}	0.748	0.733	0.703	-	0.325	0.610	0.750	0.771	0.886
Validating Datasets									
A	0.586	0.563	0.526	-	-	-	0.586	0.544	0.496
A_G	0.631	0.620	0.588	-	-	-	0.631	0.597	0.551
A+D	0.586	0.563	0.526	-	0.154	0.255	0.585	0.543	0.511
A_G+D	0.629	0.618	0.585	-	0.155	0.259	0.631	0.599	0.569
A_G+D_G	0.630	0.618	0.589	-	0.149	0.247	0.629	0.599	0.568
A_G+D_{G*}	0.629	0.618	0.589	-	0.149	0.247	0.629	0.599	0.568

GS FOR MULTIPLE YEARS

$$\mathbf{y} = \mathbf{X}_1\mathbf{s} + \mathbf{X}_2\boldsymbol{\beta}_s + \mathbf{Z}_1\mathbf{b}_s + \mathbf{Z}_3\mathbf{as} + \mathbf{e}$$

s vector of fixed environment effects (e.g. year or site)

β vector of fixed design effects (e.g. replicates)

β_s vector of fixed design effects within year

b vector of random design effects (e.g. blocks, plots), ~ N(0, $\mathbf{I}\sigma^2_b$)

b_s vector of random design effects within year (e.g. blocks, plots), ~ N(0, **D**)

as vector of random animal effects nested within year , ~ N(0, $\mathbf{G_A}\otimes\mathbf{U}$)

e vector of random residual effects, ~ N(0, $\oplus\mathbf{R}$)

$$\mathbf{U} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 & \rho_{14}\sigma_1\sigma_4 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 & \rho_{24}\sigma_2\sigma_4 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 & \rho_{34}\sigma_3\sigma_4 \\ \rho_{14}\sigma_1\sigma_4 & \rho_{24}\sigma_2\sigma_4 & \rho_{34}\sigma_3\sigma_4 & \sigma_4^2 \end{bmatrix}$$

Combining additional sources of information (to T8)

- GBLUP: - Combine additional year for analyses
- BayesB: - One-to-one predictions (average of indep. predictions)

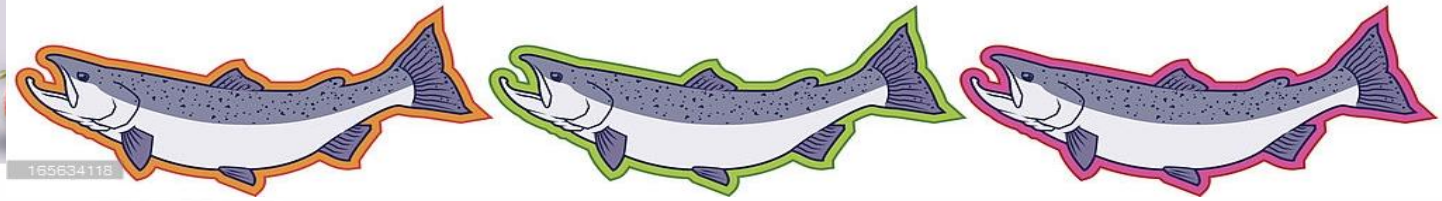
	AWT	EMY	SSC	TC	TMY
T2	0.39	0.15	0.34	0.21	0.18
T4	0.37	0.18	0.36	0.30	0.20
T6	0.36	0.23	0.33	0.19	0.30

	AWT	EMY	SSC	TC	TMY
T2	0.39	0.15	0.34	0.21	0.18
T2+T4	0.41	0.20	0.39	0.29	0.22
T2+T4+T6	0.43	0.23	0.40	0.30	0.28

AWT: average weight; SSC: soluble solids content;
TMYL total marketable yield; TC: total percent culls;
EMY: early marketable yield.

GENOMIC SELECTION IN SALMON

Number of Markers
(parentage chip vs. genomic chip)
Blending Pedigree with Molecular relat.



Atlantic Salmon

Experiment

- 534 Atlantic salmon (offspring). (yield)
- 59 nuclear families
- 90 parents
- Fish challenged with sea lice (*L. salmonis*)

Traits

- Body weight (g)
- Number of lice attached (disease)

The pedigrees of the fish were identified using PIT-tagging, and an adipose fin clip of each fish was collected for DNA extraction.

Molecular Data

- All samples were genotyped using the Affymetrix Axiom 132 K Atlantic salmon SNP chip.
- A final set of 78,362 SNP markers from a total of 624 fish was available.

Tsai, H-Y. et al. (2017). Genes | Genomes | Genetics 6:1313-1326

MATERIAL AND METHODS

- Selecting a random set of markers from the complete set of 78,362 markers (m = 10).
- Number of markers: 100, 400, 700, 3000, 10000, 78362
- Obtain genomic matrix $\mathbf{G_A}$
- Fit an animal model (GBLUP) with $\mathbf{G_A}$ or $\mathbf{A}$.
- Evaluate a blending by using: $\mathbf{G_A}^* = (1 - p)\times\mathbf{G_A} + p\times\mathbf{A}$
- Evaluation using cross-validation.

UF IFAS UNIVERSITY of FLORIDA

RESULTS

Body Weight

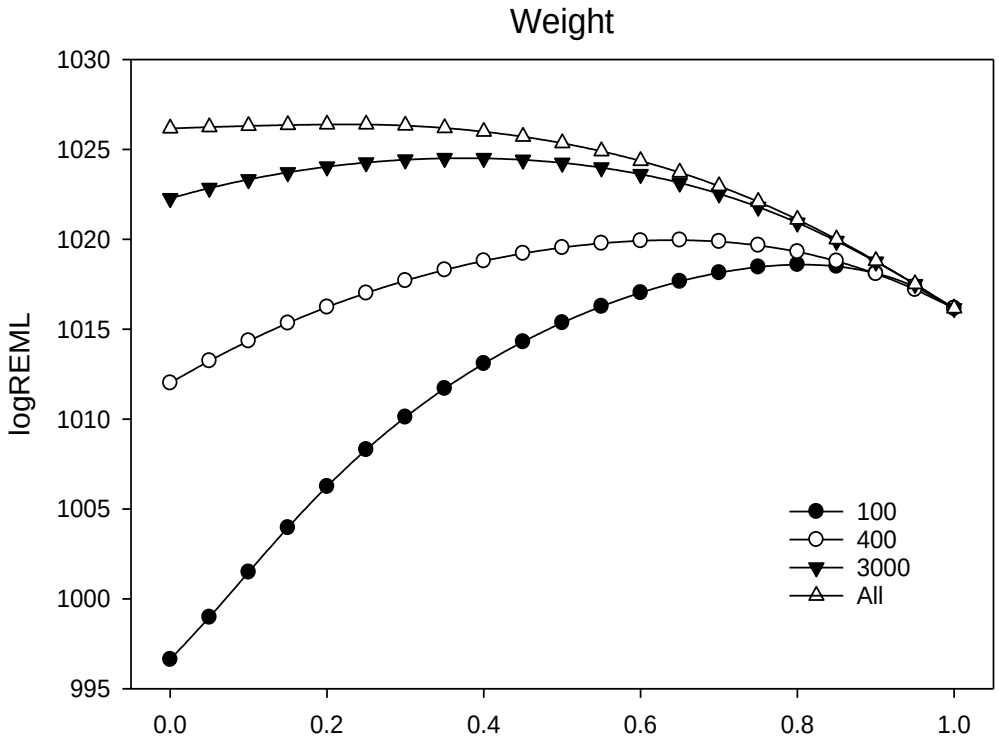
# SNPs	*p**	logREML	h^2	h^2_{CV}	PA	ACC	a_{corr}
-	Ped	1016.2	0.494	0.493	0.397	0.566	0.658
100	0.00	996.6	0.232	0.225	0.268	0.564	0.463
400	0.00	1012.0	0.401	0.398	0.378	**0.602**	0.650
700	0.00	1015.7	0.424	0.427	0.392	0.600	0.683
3,000	0.00	1022.3	0.532	0.520	0.434	0.601	0.745
10,000	0.00	1024.4	0.567	0.550	0.439	0.593	0.758
All	0.00	1025.7	0.581	0.562	0.449	0.599	0.766

Number Sea Lice

# SNPs	*p**	logREML	h^2	h^2_{CV}	PA	ACC	a_{corr}
-	Ped	310.7	0.299	0.301	0.278	0.507	0.741
100	0.00	300.8	0.119	0.123	0.156	0.448	0.522
400	0.00	305.6	0.200	0.201	0.232	0.520	0.708
700	0.00	305.5	0.209	0.206	0.237	0.524	0.740
3,000	0.00	307.5	0.233	0.234	0.248	0.512	0.790
10,000	0.00	308.0	0.242	0.242	0.253	0.515	**0.800**
All	0.00	308.2	0.250	0.260	0.240	0.472	0.798

RESULTS

ML proportions *p*



$$\mathbf{G_A}^* = (1 - p) \times \mathbf{G_A} + p \times \mathbf{A}$$

STATISTICAL CHALLENGES WITH GS

- **Can we have a GS model that predicts total genetic value?**

  g = a + d + i

- Partition genetics into more components from a LMM:

  G = A + D + A#A + A#D + D#D

- Use similarity matrices **S** for continuous and categorical data

  AA → 0, AC → 1, CC→ 2 Additive coding

  AA → 0, AC → 1, CC→ 0 Dominance coding

  AA → X, AC → Y, CC→ Z Categorical coding

- Incorporate some higher order interaction
- Evaluate other Machine Learning methods (e.g. NN)

STATISTICAL CHALLENGES WITH GS

- **What to do with very large genomic G_A matrices?**

- Mixed Model Equations are too dense!! (compared with **A**)
- We need computational and mathematical solutions!

Solution 1

- Use conditional normal distribution to 'update' breeding values.

$$\begin{bmatrix} \boldsymbol{g_x} \\ \boldsymbol{g_y} \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \boldsymbol{G_{xx}} & \boldsymbol{G_{xy}} \\ \boldsymbol{G_{yx}} & \boldsymbol{G_{yy}} \end{bmatrix}\right)$$

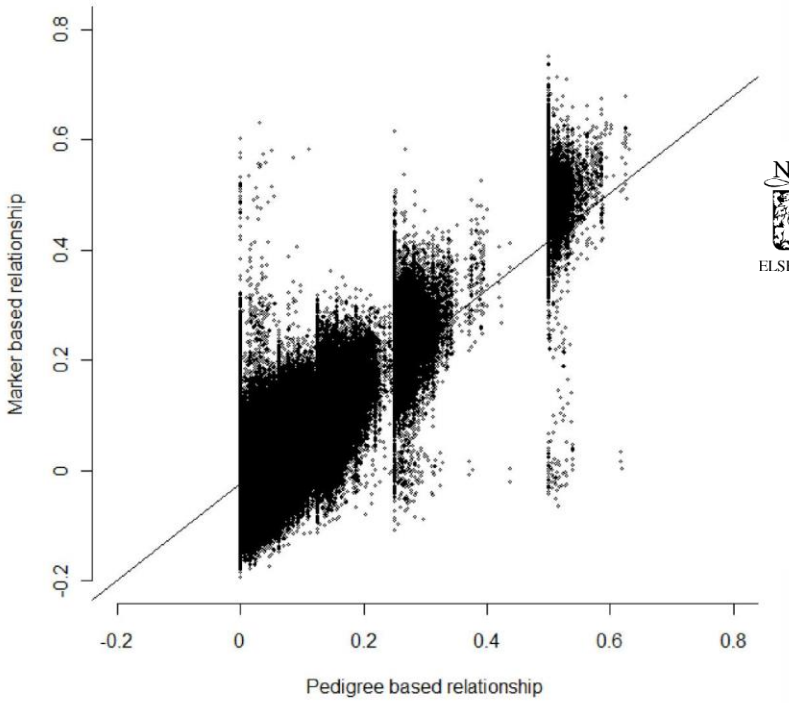
$$E(\boldsymbol{g_x}|\boldsymbol{g_y}) = \begin{pmatrix} \mathbf{1} & \boldsymbol{G_{xy}}\boldsymbol{G_{yy}^{-1}} \end{pmatrix} \begin{pmatrix} \mu_y \\ \boldsymbol{g_y} - \mathbf{1}\mu_y \end{pmatrix}$$

STATISTICAL CHALLENGES WITH GS

- **What to do with very large genomic G_A matrices?**

Solution 2

- Generate sparse $\mathbf{G_A}$ or $\mathbf{G_A^{-1}}$ with minimum loss of information.



ELSEVIER

Applied Numerical Mathematics 30 (1999) 305–340

APPLIED NUMERICAL MATHEMATICS

A comparative study of sparse approximate inverse preconditioners

Michele Benzi [a,*,1], Miroslav Tůma [b,2]

[a] *Los Alamos National Laboratory, MS B256, Los Alamos, NM 87545, USA*

[b] *Institute of Computer Science, Czech Academy of Sciences, 182 07 Prague 8, Czech Republic*

STATISTICAL CHALLENGES WITH GS

- **What to do with very large genomic G_A matrices?**

Solution 3

- Use shrinkage and/or matrix blending on top of sparse options.

Shrinkage (Powell et al. 2010)

$$\{a_{Gadj}\}_{jk} = (1 - \frac{1/M}{var(a)})\,(a_{Gjk} - a_{jk}) + a_{jk}$$

Blending

$$\mathbf{G_A}^* = (1 - p)\times\mathbf{G_A} + p\times\mathbf{A}$$

STATISTICAL CHALLENGES WITH GS

- **What to do with missing values in the molecular matrix?**

Solution

- Imputation?
- Ignore missing values?

$$\{a_G\}_{jk} = \begin{cases} \frac{1}{M}\sum_i \frac{(g_{ij} - 2p_i)(g_{ik} - 2p_i)}{2p_i(1-p_i)} & if \quad j \neq k \\ 1 + \frac{1}{M}\sum_i \frac{g_{ij}^2 - (1 + 2p_i)g_{ij} + 2p_i^2}{2p_i(1-p_i)} & if \quad j = k \end{cases}$$

STATISTICAL CHALLENGES WITH GS

- **How do we combine pedigree- and molecular-based relationship matrices? A vs. G_A**

Solution

- Use of the $\mathbf{H}^{-1}$ matrix instead of the $\mathbf{A}^{-1}$!

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G}^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix}$$

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \tau(\alpha\mathbf{G} + \beta\mathbf{A}_{22})^{-1} - \omega\mathbf{A}_{22}^{-1} \end{bmatrix}$$

- Not as simple, as $\mathbf{A}_{22}$ and $\mathbf{G}^{-1}$ come from different ‘populations’.

STATISTICAL CHALLENGES WITH GS

- **How do we combine pedigree- and molecular-based relationship matrices? A vs. G_A**

STATISTICAL CHALLENGES WITH GS

- **How do we combine pedigree- and molecular-based relationship matrices? A vs. G_A**

Solution

- Some pedigree-based adjustments might be required for p (or 1-p).

$$\{a_G\}_{jk} = \begin{cases} \frac{1}{M}\sum_i \frac{(g_{ij} - 2p_i)(g_{ik} - 2p_i)}{2p_i(1-p_i)} & if \quad j \neq k \\ 1 + \frac{1}{M}\sum_i \frac{g_{ij}^2 - (1 + 2p_i)g_{ij} + 2p_i^2}{2p_i(1-p_i)} & if \quad j = k \end{cases}$$

STATISTICAL CHALLENGES WITH GS

- **What to do with a non-positive definitive G_A matrix?**

Solution 1

- Make it a positive definite (PD) and therefore invertible matrix!
  - Use iterative or non-iterative Bending (minimum accepted eigenvalue)
  - Blend the matrix: $\mathbf{G_A}^* = (1 - p) \times \mathbf{G_A} + p \times \mathbf{A}$

Solution 2

- Absorb singularities within the Mixed Model Equations (ASReml-SA).

A FEW MORE CHALLENGES WITH GS

- **How to obtain the G_A matrix for polyploids?**
- **How do we combine QTL analysis with GS into a single analysis?**
- **Can we really combine different batches of molecular data?**
- **How do we prepare for very- very large molecular matrices?**

QG ANALYSES

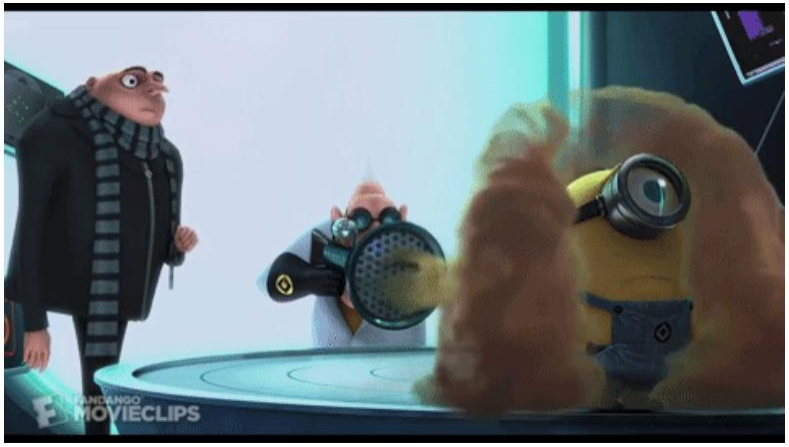
- Understanding of the genetic architecture is critical to any breeding program as it defines the **Breeding Strategy**
- Relevant information includes:
  - *Genetic control (additive, dominance, epistasis).*
  - *Genotype-by-Environment (GxE).*
  - *Genotype-by-Year (GxY).*
  - *Trait-to-trait correlations.*
  - *Temporal correlations.*
  - *Spatial correlations.*
  - *Efficiency of Pedigree- or Molecular-based analyses.*
- All of these require parameters estimated by Linear Mixed Models.

SUMMARY

- LMM applications in breeding has become a reality with lots of benefits.
- These are exciting times to exploit LMMs and make breeding programs more efficient!
- These are very exciting times to incorporate molecular data and develop new tools for many biological systems (plants, animals and humans!)
- Software and computers have allow us to perform many of these analyses in an flexible, fast and accurate way.

ASreml®

UF IFAS
UNIVERSITY of FLORIDA



ASreml®

UF IFAS
UNIVERSITY of FLORIDA

ASreml®

Acknowledgment

sgezan@ufl.edu

- V. Whitaker (UF, Strawberry)
- L.F. Osorio (UF, Strawberry)
- J.C. Motamayor (MARS, Cocoa)
- G. Mustiga (MARS, Cocoa)
- A. Duval (MARS, Cocoa)
- A. Garber (Huntsmann, Salmon)
- B. Swift (Trigen, Salmon)
- V. Passos (UFLA, Forestry)
- L. Mramba (UF, Forestry)
- A. Nazarian (UF, Human)
- I.J. Baracuhi (CTC, Sugarcane)
- G. Peres Silva (CTC, Sugarcane)

SFRC

UF UNIVERSITY of FLORIDA