

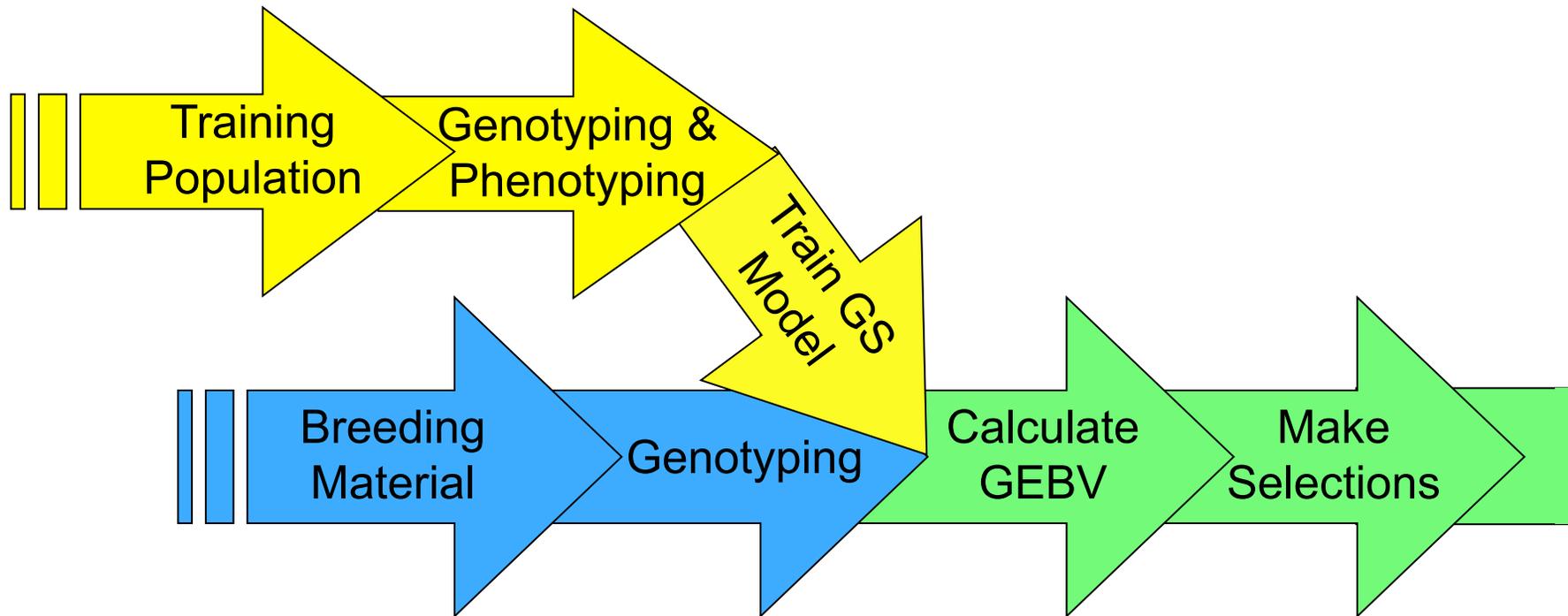
Multiple Trait Genomic Selection (MT-GS) via Multivariate Bayesian Modeling in Plant Breeding

Yi Jia

Postdoctoral Associate

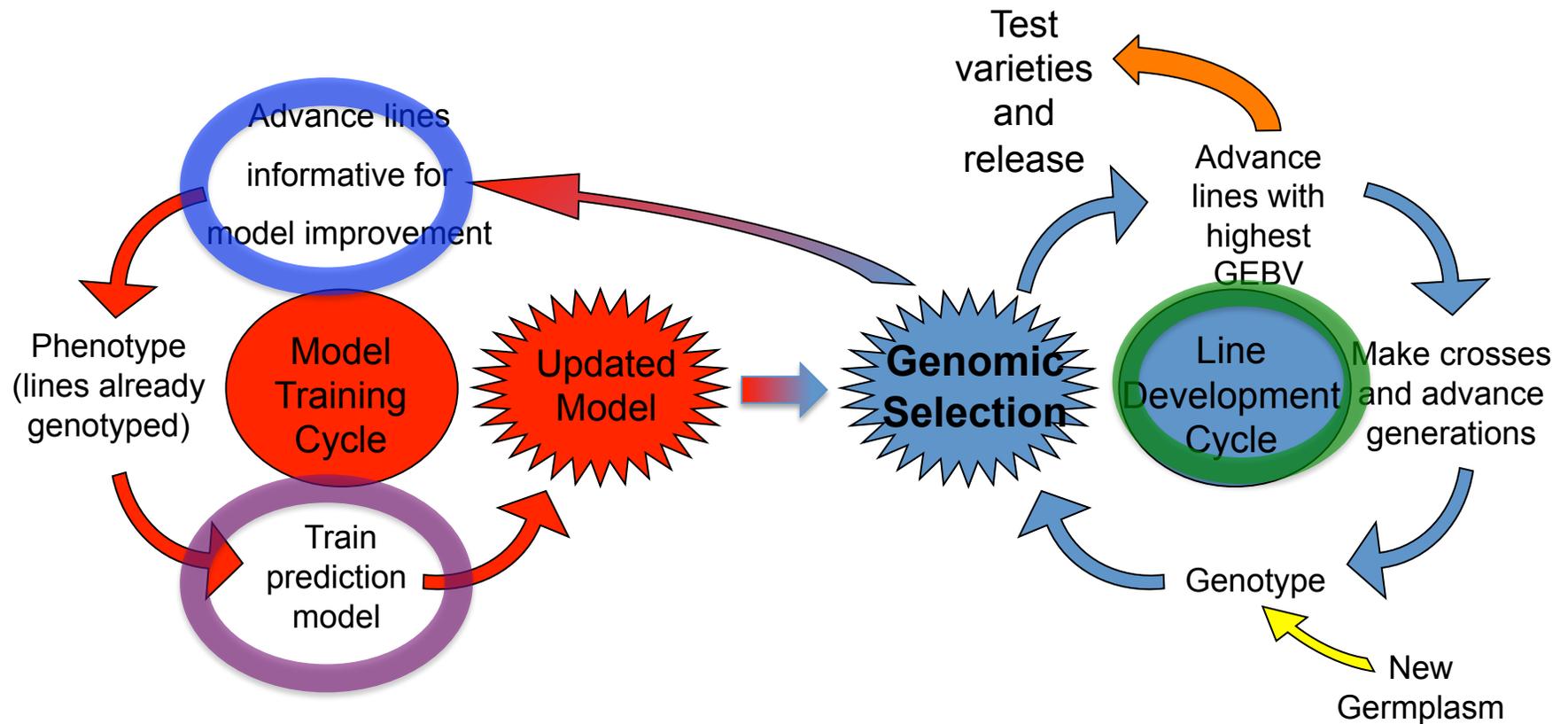
Plant Breeding and Genetics, Cornell University

Genomic Selection (GS) Evaluates Alleles Instead of Lines



Meuwissen et al. 2001 Genetics

Research Topics in Genomic Selection

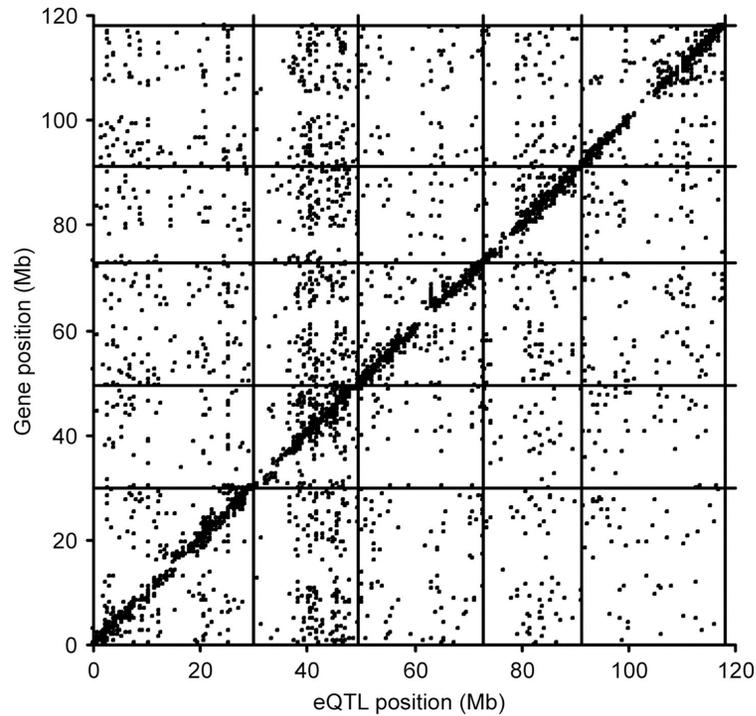


New Cultivar Evaluated for Multiple Traits

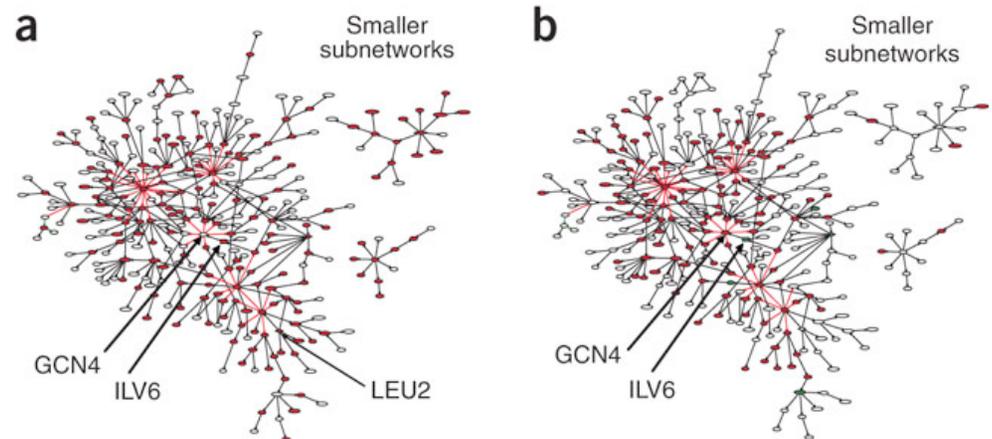


- Yield (Grain, biomass)
- Quality (Flavor, color, shape, nutrient)
- Stress resistance (Biotic and abiotic stresses)

Genetic Correlation is Pervasive



Keurentjes et al. PNAS 2007



Zhu et al. Nature Genetics 2008

Univariate Linear GS Model

$$y_i = u + \sum_{j=1}^p X_{ij} \alpha_j + e_i \quad i = 1, \dots, n \text{ individual}; \quad j = 1, \dots, p \text{ marker};$$

Methods	Variable selection?	Marker variance
BLUP*	No	Identical
Bayes-A	No	Different
Bayes-B	Yes	Different
Bayes-CPi	Yes	Identical/0

Meuwissen et al. 2001 Genetics
Habier et al. 2011 BMC Bioinformatics

Single Trait (ST) BayesA

Sampling model $y_i = u + \sum_{j=1}^p X_{ij} \alpha_j + e_i$

Prior distribution $\alpha_j \sim N(0, \sigma_j^2)$ $\sigma_j^2 \sim \text{Invchisq}(v, s)$
 $v = 4.012$ $s = 0.002$

Posterior distribution $(\sigma_j^2 \mid y, \text{ELSE}) \sim \text{Invchisq}(v + 1, s + \alpha_j^2)$

Meuwissen et al. 2001 Genetics

The prior can dominate and cause bias for the posterior with only one more degree of freedom.

Gianola et al. 2009 Genetics

Full-Bayesian ST-BayesA

Sampling model $y_i = u + \sum_{j=1}^p X_{ij} \alpha_j + e_i$

Prior distribution $\alpha_j \sim N(0, \sigma_j^2)$ $\sigma_j^2 \sim \text{Invchisq}(v, s)$
 ~~$v = 4.012$ $s = 0.002$~~

Posterior distribution $(\sigma_j^2 \mid y, \text{ELSE}) \sim \text{Invchisq}(v + 1, s + \alpha_j^2)$

Joint posterior likelihood $P(u, \alpha_i, \sigma_j, \sigma_e, v \text{ and } s \mid y) \sim p(y \mid u, \alpha_i, \sigma_e) * \Pi(p(\alpha_i \mid \sigma_j)) * \Pi p(\sigma_j \mid v, s) * p(v, s)$

Estimating the v and s via Metropolis Algorithm

Full Bayesian MT-BayesA

Sampling model $y_i = \mu + \sum_{j=1}^p X_j \alpha_j + e_i$ $y_i = [y_{i1}, \dots, y_{im}]$
 $\alpha_j = [\alpha_{j1}, \dots, \alpha_{jm}]$

Priors $\alpha_j \sim N_q(0, \Sigma_j)$ $\Sigma_j \sim \text{Inv - Wishart}(v, S)$

Posterior $(\Sigma_j | y, ELSE) \sim \text{Inv - Wishart}(v + 1, s + \alpha_j^T \alpha_j)$

Joint posterior likelihood $P(\mu, \alpha_i, \Sigma_j, \Sigma_e, v \text{ and } S | y) \sim p(y | \mu, \alpha_i, \Sigma_e) * \Pi(p(\alpha_i | \Sigma_j)) * \Pi p(\Sigma_j | v, S) * p(v, S)$

Estimating the v and S via Metropolis Algorithm

ST-BayesC π

Sampling model

$$y_i = \mu + \sum_{j=1}^p \delta_j X_{ij} \alpha_j + e_i \quad (\alpha_j | \pi, \sigma^2) \begin{cases} N(0, \sigma^2) & \text{prob. } (1 - \pi) \\ 0 & \text{prob. } \pi \end{cases}$$

Priors

$$\alpha_j \sim N(0, \sigma^2) \quad \sigma^2 \sim \text{Invchisq}(v, s)$$

Posteriors

$$(\sigma^2 | y, \text{ELSE}) \sim \text{Invchisq}(v + m, s + \sum \alpha_j^2)$$

$$(\pi | y, \text{ELSE}) \sim \text{beta}(p - m + 1, m + 1)$$

MT-BayesC π

Sampling model

$$y_i = u + \sum_{j=1}^p \delta_j X_j \alpha_j + e_i \quad (\alpha_j | \pi, \Sigma_\alpha) \begin{cases} \sim N(0, \Sigma_\alpha) & \text{prob. } (1 - \pi) \\ 0 & \text{prob. } \pi \end{cases}$$

$$y_i = [y_{i1}, \dots, y_{im}] \quad \alpha_j = [\alpha_{j1}, \dots, \alpha_{jm}]$$

Priors

$$\alpha_j \sim N_q(0, \Sigma_\alpha)$$

$$\Sigma_\alpha \sim \text{Inv - Wishart}(v, S)$$

Posteriors

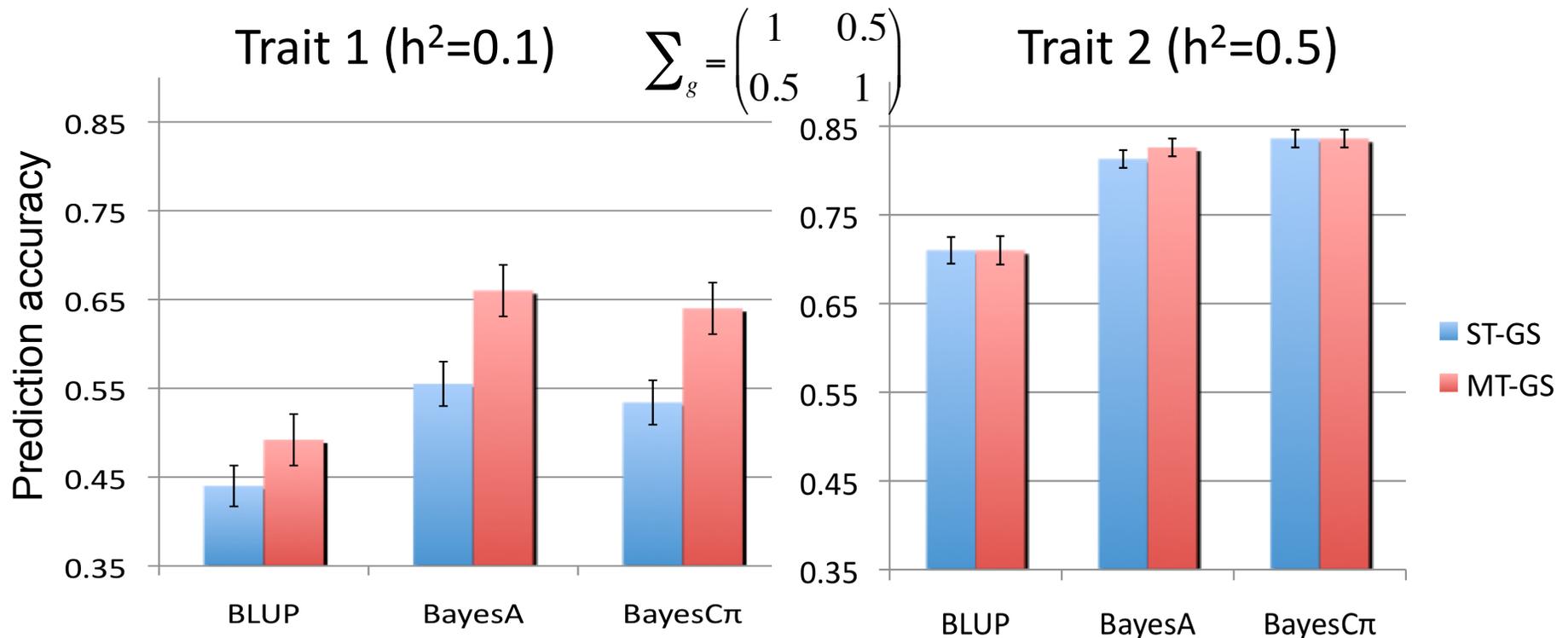
$$(\Sigma_\alpha | y, \text{ELSE}) \sim \text{Inv - Wishart}(v + m, S + \alpha^T \alpha)$$

$$(\pi | y, \text{ELSE}) \sim \text{beta}(p - m + 1, m + 1)$$

Data Simulation for Model Testing

Parameter	Value	Parameter	Value
Effective pop size	50	Total Traits	2
Total Individuals	500	Total QTL	20
Total Markers	2,000	h^2 (Trait1)	0.1
minMAF	0.01	h^2 (Trait2)	0.5
Replication	24	Variance	$\Sigma_g = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$

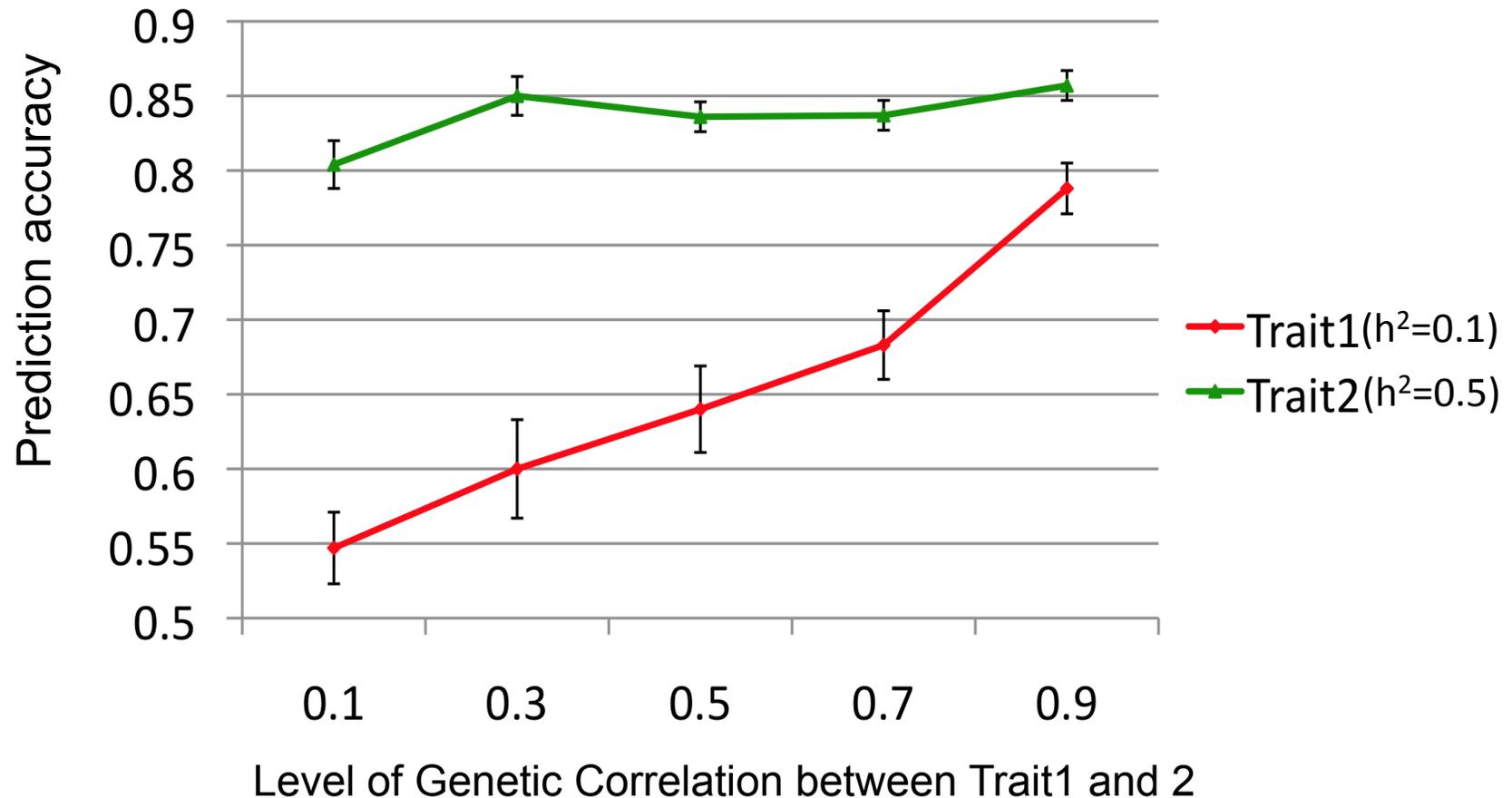
ST vs. MT-BayesGS under Standard Simulation



Low heritable Trait Gets Information from High heritable Trait

Heritability		GS Accuracy (MT-BayesC π)	
Trait1	Trait2	Trait1	Trait2
0.1	0.1	0.51 \pm 0.03	0.52 \pm 0.02
0.1	0.5	0.64 \pm 0.03	0.84 \pm 0.01
0.1	0.8	0.67 \pm 0.03	0.91 \pm 0.01

Genetic Correlation Benefit the Low Heritable Trait More

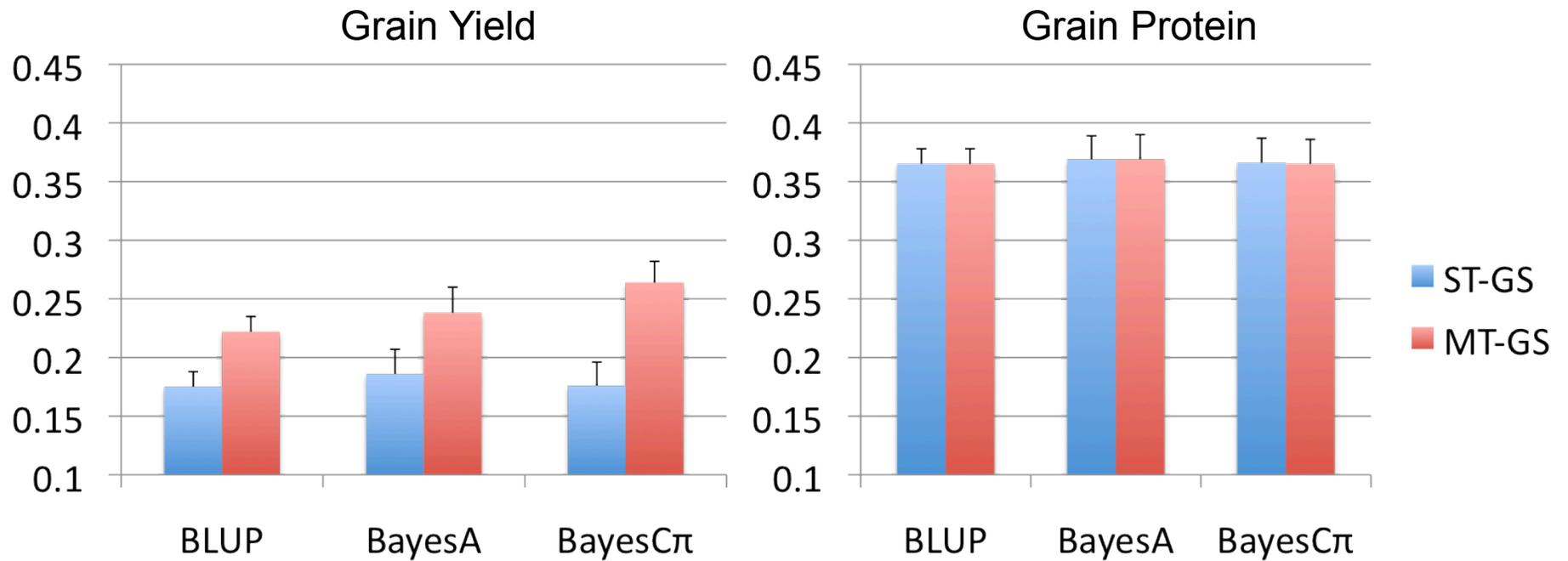


Multi-family Wheat Breeding data

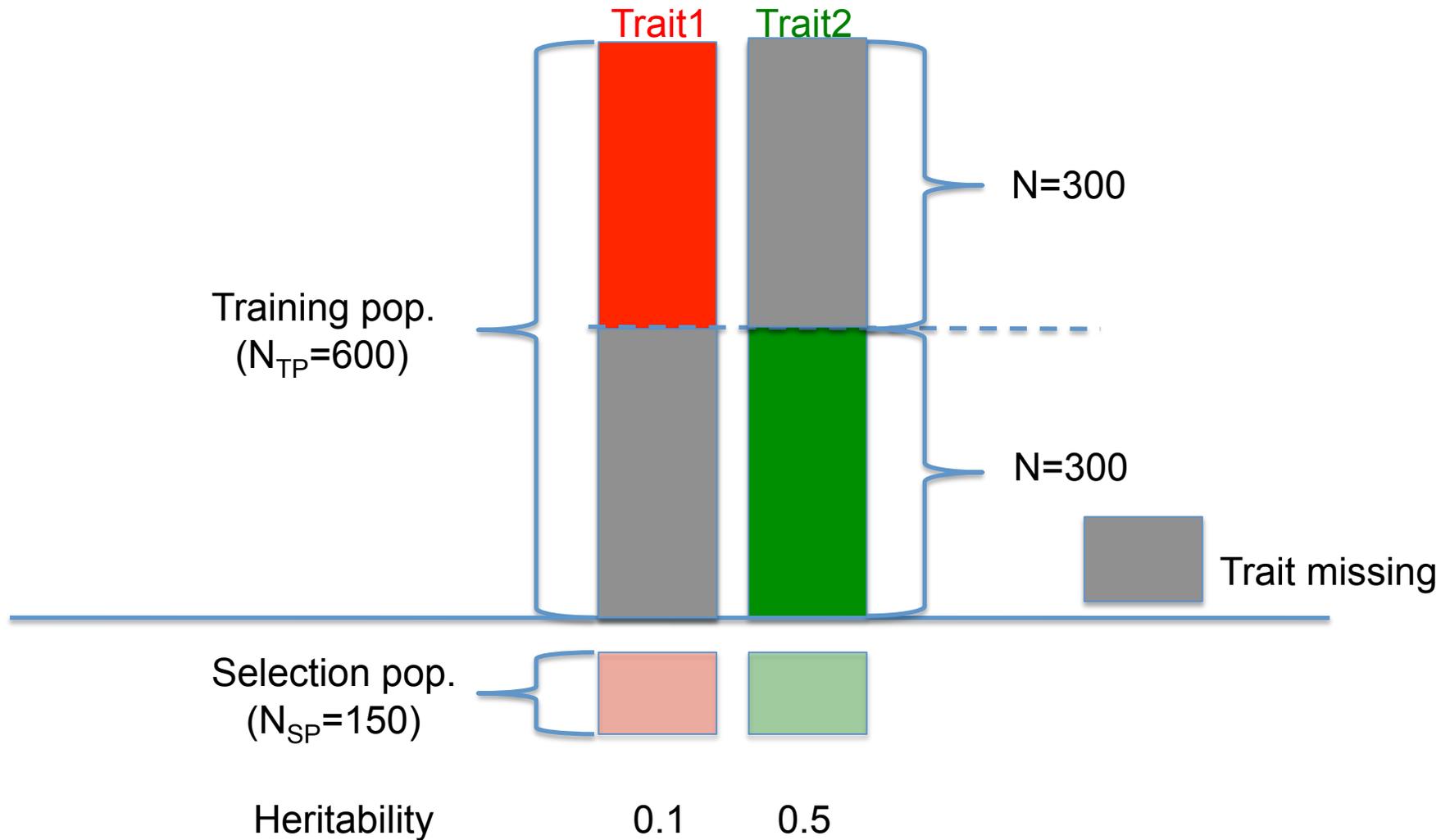
- 374 F₅ breeding lines
- 1,158 DArT markers
- Grain yield and grain protein from 2008 and 2009



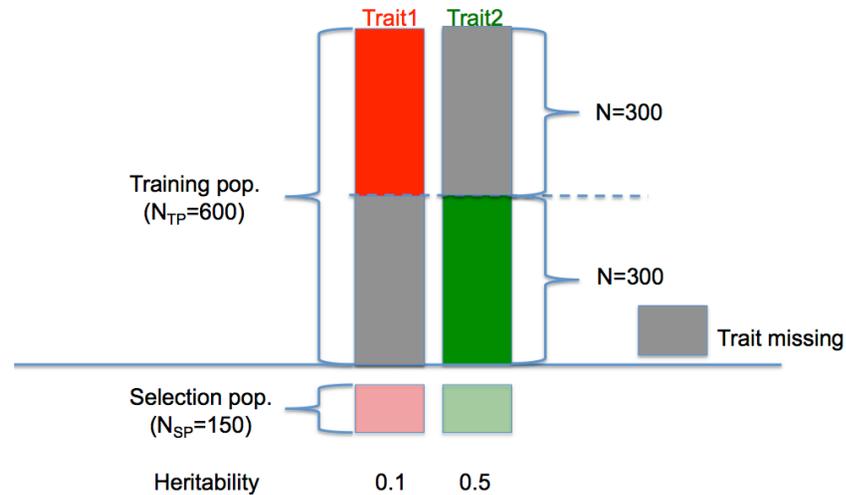
MT-GS for Wheat Breeding Data



Expanding Training Population via MT-GS

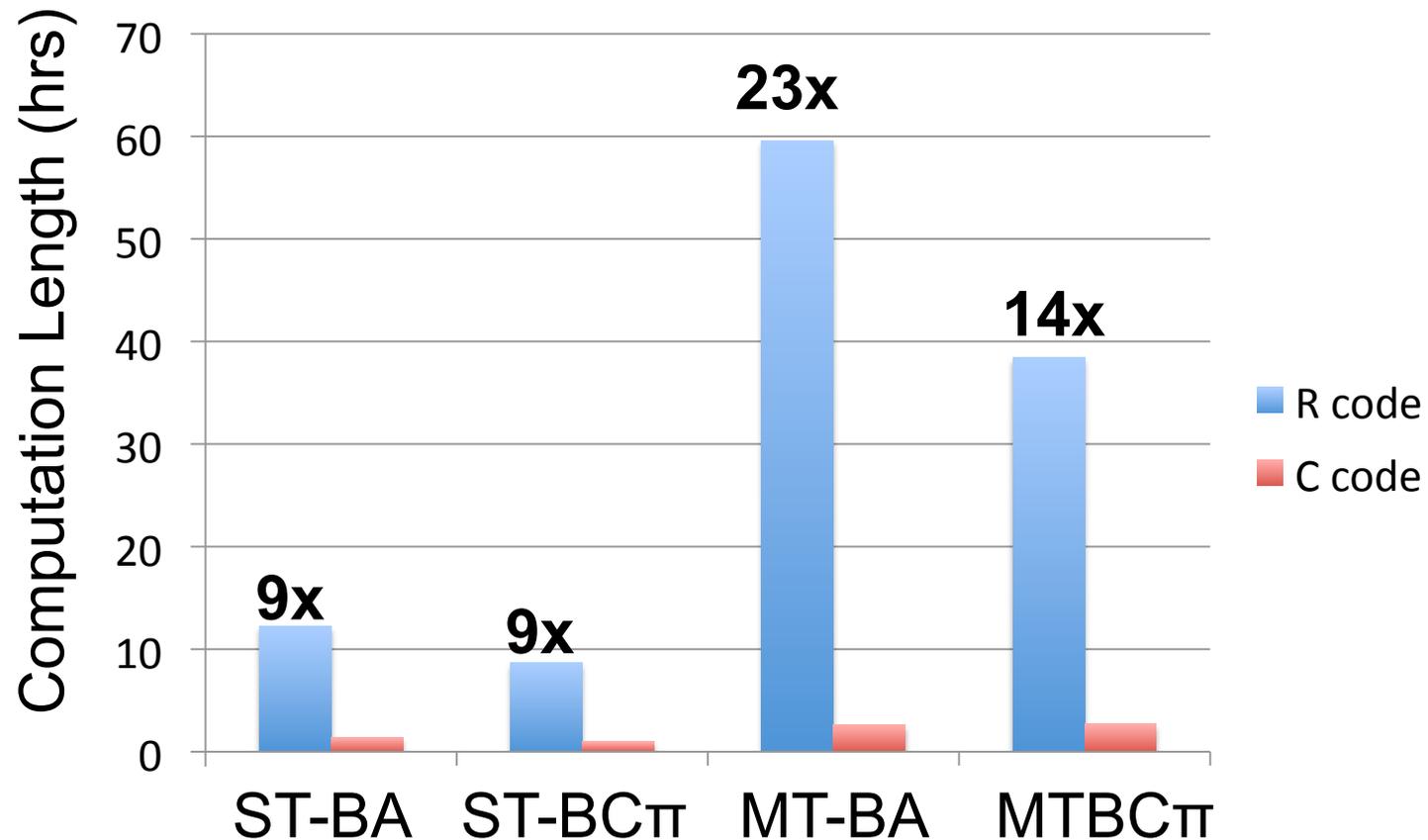


Expanding Training Population via MT-GS



No.	imputation	GS Method	Traits in model	N _{TP}	Accuracy (Trait1)	Accuracy (Trait2)
1	No	ST-BayeCπ	Trait 1	300	0.41±0.02	NA
2	No	ST-BayeCπ	Trait 2	300	NA	0.71±0.01
3	Yes	MT-BayeCπ	Trait 1+2	600	0.50±0.02	0.73±0.01

Running time of R vs C code for 50,000 MCMC iterations



Take-home Messages

- MT-GS exhibits advantages over ST-GS for both simulated and empirical data
- Low heritable trait benefits from the correlated high heritable trait by MT-GS
- MT-GS can expand the training population by phenotype imputations

Acknowledgements

- Jean-Luc Jannink (Cornell)
- Mark Sorrells (Cornell)
- Aaron Lorenz (U of Nebraska)
- Kevin Smith (U of Minnesota)

Cornell group members

- Peter Bradbury
- Nicolas Heslot
- Deniz Akdemir
- Jeffrey Endelman
- Martha Hamblin
- Hsiao-Pei Yang

