



Identifying Loci and Genomic Prediction Models for PHS Tolerance in Northeast Soft Wheat Breeding Programs

Shantel A. Martinez

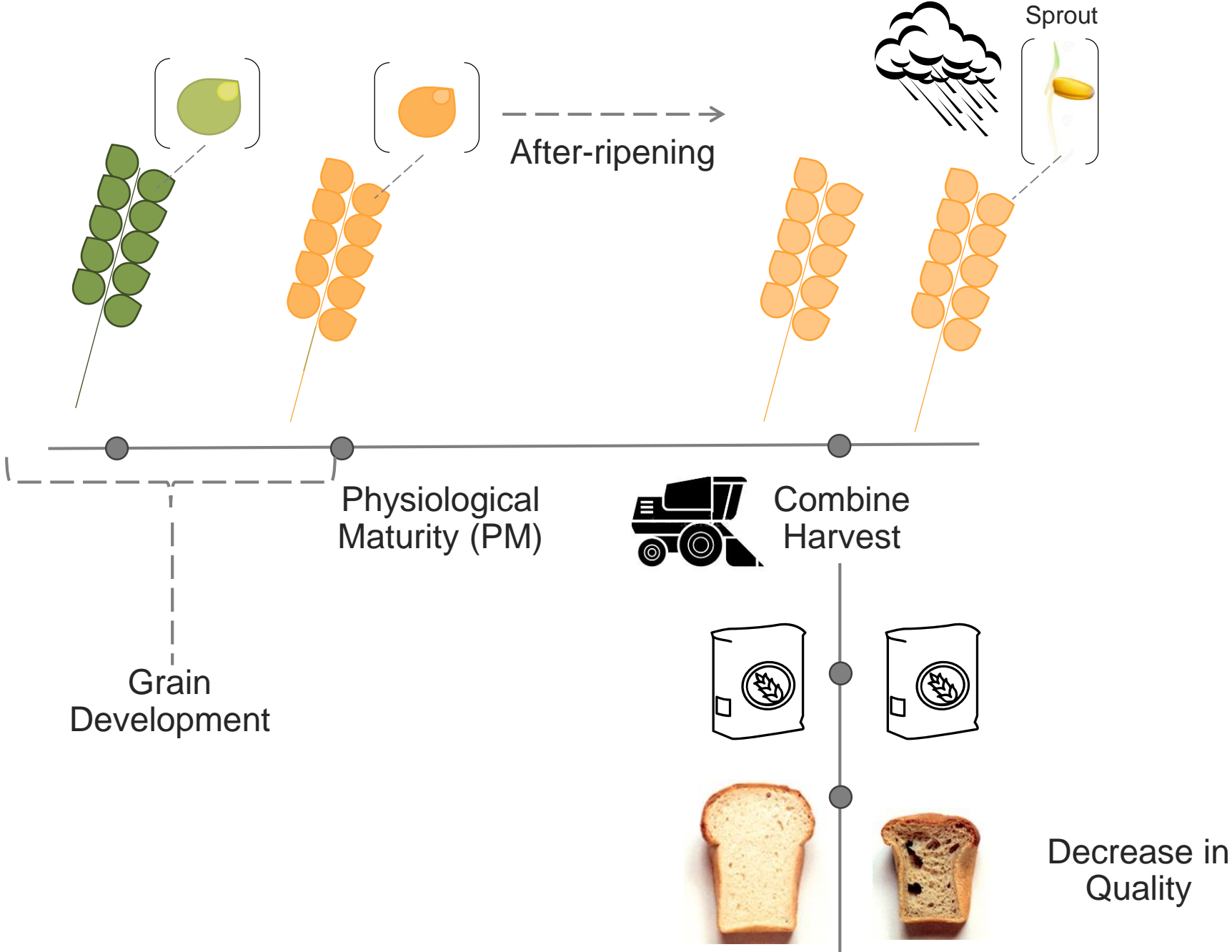
Mark E. Sorrells

Cornell University

ASA/CSSA/SSSA

Nov 11th, 2019

Preharvest Sprouting



The Northeast U.S. is not alone

England

Tobias Barber
@ekte_Toby
Following
I guess this is what you'd call pre-harvest sprouting #Harvest18



Jim Thompson
@jimt_farmer
Follow
Think rain has stopped play #wheatharvest18
@AllpressF @LumleySean @coostiebarrey @chrisbettinson2



Kansas

Kyler Millershaski
@Shaski92
Follow
I'm always happy to have rain, but not the view I want during #WheatHarvest18 #kswx



3:36 PM - 22 Jun 2018
5 Retweets 37 Likes

Pacific Northwest



Nebraska

Chris Cu11an
@ChrisCullan1
Follow
On the edge. #wheatharvest18



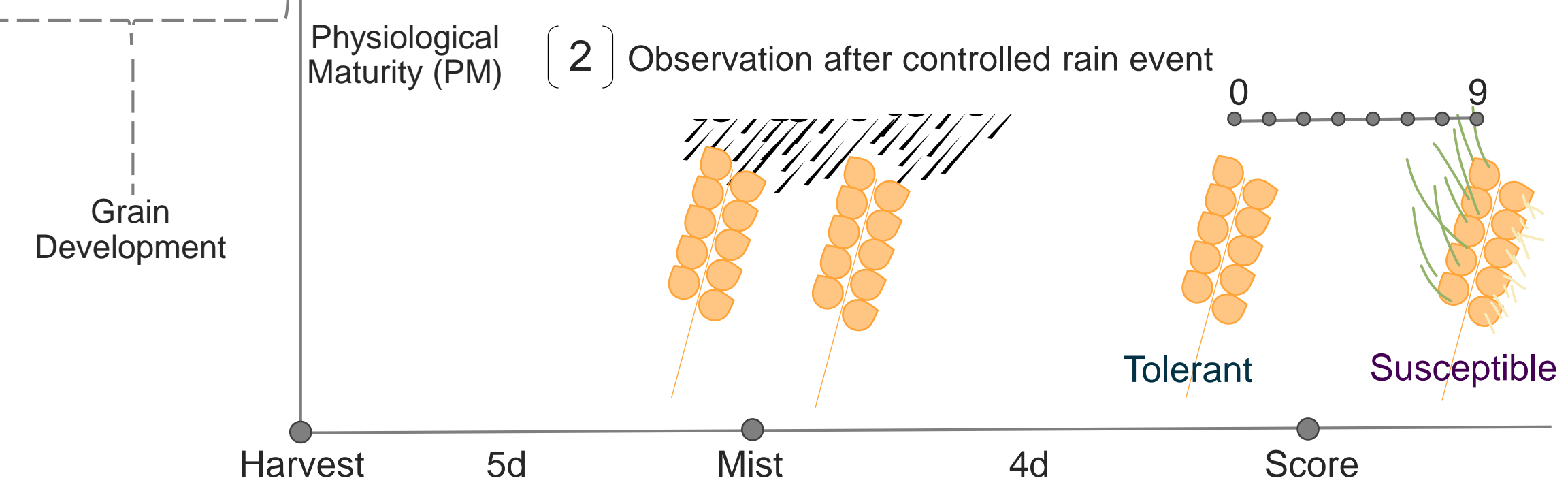
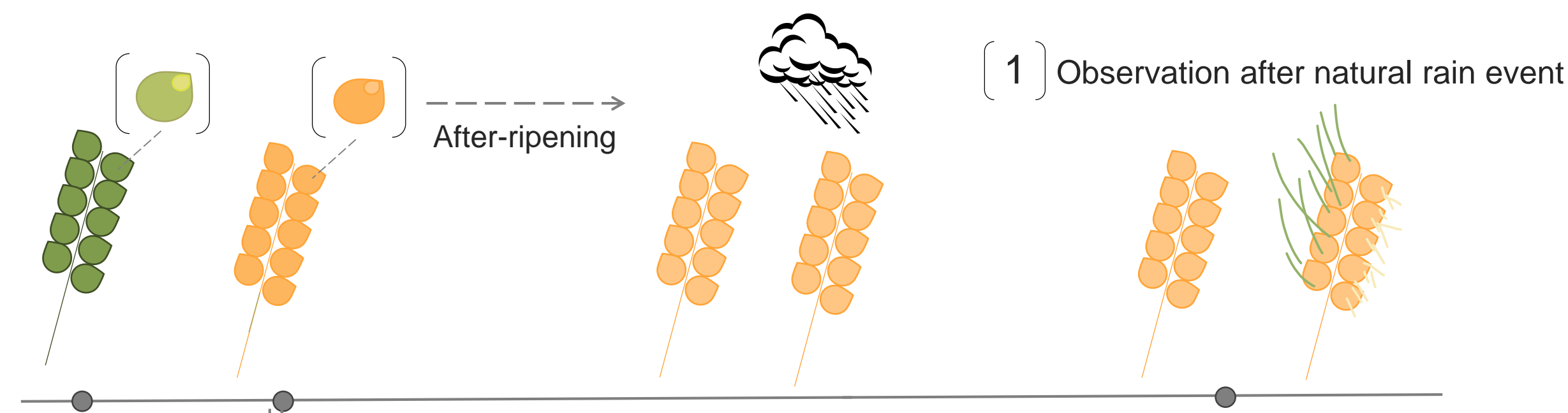
5:06 PM - 12 Jul 2018
1 Retweet 40 Likes

Canada

Fermes Chauvin Farms Ltd.
@MoeChauvin
Follow
#wheatharvest18 started here at fermeschauvinfarms.com in StoneyPoint. Decent yields for no rain ... #OntAg #AgMoreThanEver #goodineverygrain



Farms.com

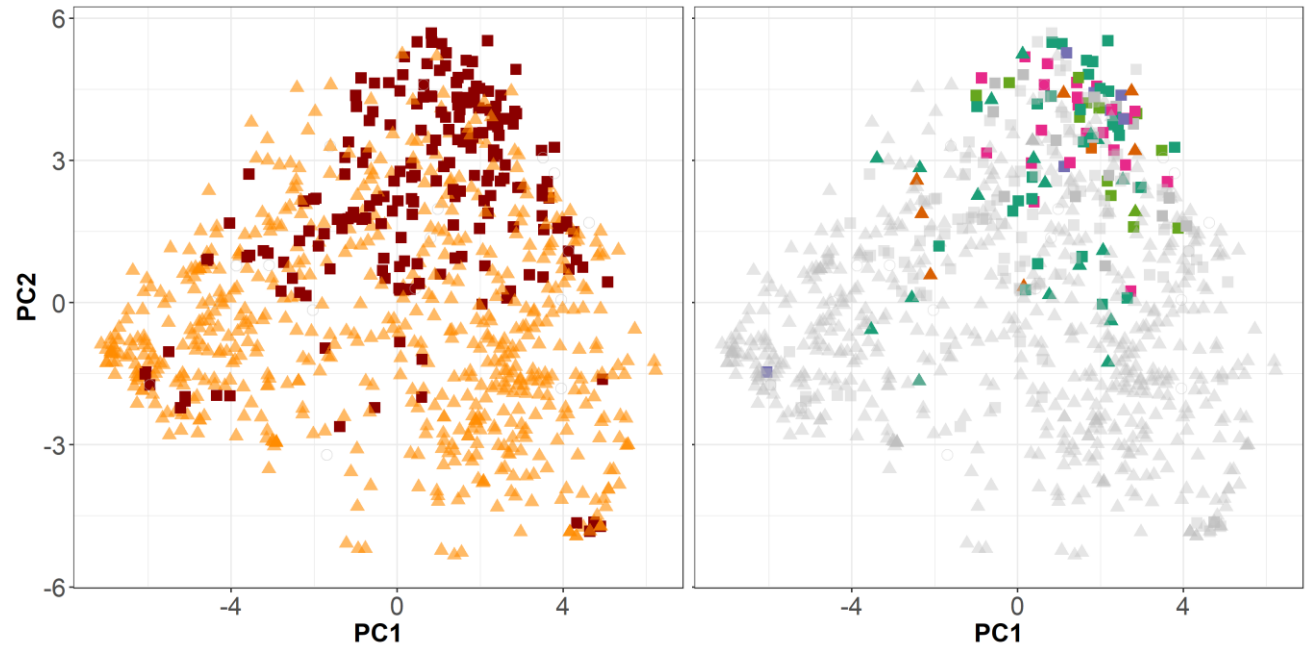
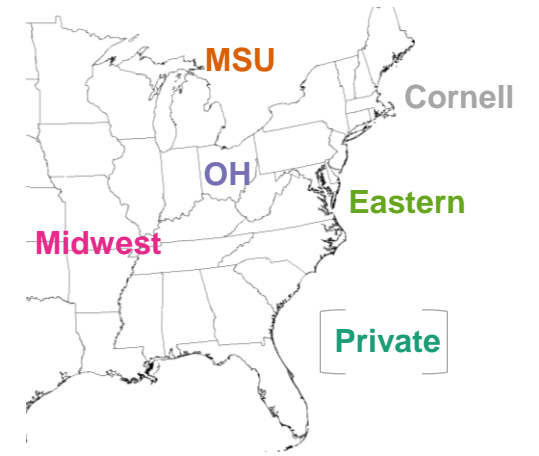




Wheat has quite a bit of genetic variation

Year	Location	White	Red	Total (Comb)
2008	Ketola	109	11	120
	Snyder	109	11	120
2009	McGowen	109	11	120
2010	Helfer	149	59	208
	Ketola	157	59	216
	Snyder	153	59	212
2011	Helfer	198	76	274
	Ketola	206	69	275
	McGowen	208	66	274
2012	Helfer	183	80	263
	Ketola	194	78	272
	Snyder	187	74	261
2013	Helfer	52	29	81
	Ketola	56	25	81
	McGowen	106	23	129
2014	Helfer	128	85	213
	Ketola	128	85	213
	Snyder	128	85	213
2015	Helfer	218	160	378
	Ketola	192	138	330
	McGowen	195	136	331
2016	Helfer	181	192	373
	Ketola	154	185	339
	Snyder	175	184	359
2017	Ketola	199	139	338
	Snyder	199	139	338

CNL Master Nursery



Mixed Model

Observations/Phenotype

$$y = \mu + X\beta_{loc} + X\beta_{yr} + X\beta_{harv} + Zg + \varepsilon$$

population mean

loc: location

yr: year

design matrix

harv: sampling date

design matrix

e: experimental error

g: BLUP

y_{obs}

fixed effects

random effects

$$y_{obs} \sim \text{Loc} + \text{Yr} + \text{HarvDate} + (1|\text{variety})$$

Genomic Prediction

Observations/Phenotype

$$\mathbf{u} \sim N(0, \mathbf{K}\sigma_g^2)$$

\mathbf{u} is the vector of breeding values with estimated relationship matrix \mathbf{K}

Models can calculate \mathbf{K} differently:

Ridge Regression
Gaussian Kernel

fixed effects

random effects

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{WZu} + \boldsymbol{\varepsilon}$$

\mathbf{u} : GEBV

$\boldsymbol{\varepsilon}$: experimental error

design matrix

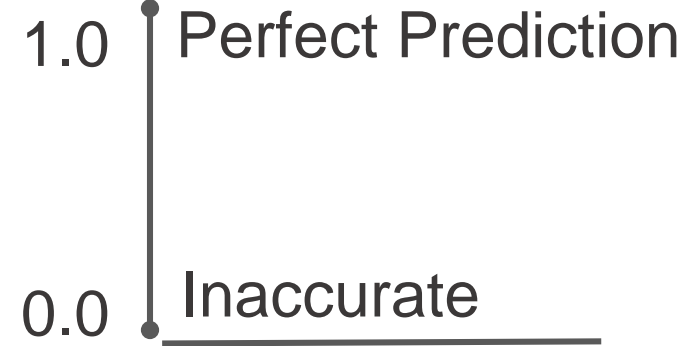
design matrix

$$y_{\text{obs}} \sim \text{Loc} + \text{Yr} + \text{HarvDate} + (1|\text{variety})$$

Genomic Prediction

Prediction Accuracy (PA)

$$\text{cor} (y_{\text{obs}} , \text{GEBVs}_{\text{model}})$$
$$\text{cor} (\mathbf{y}_{\text{obs_test}} , \mathbf{GEBVs}_{\text{test}})$$

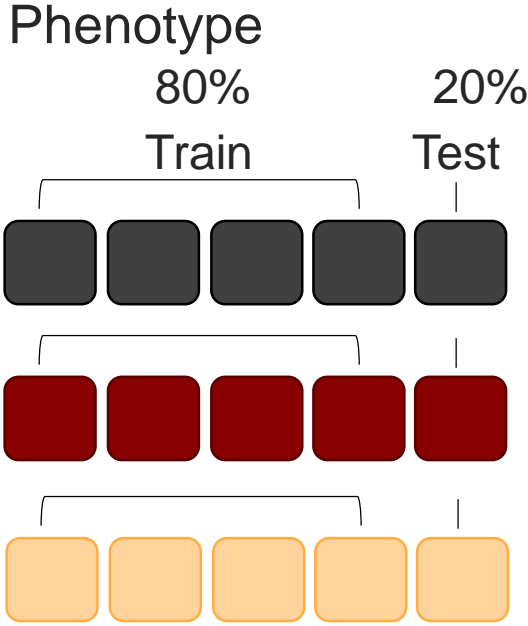


Prediction Model

- Gaussian kernel
- Ridge regression

Output:
GEBVs_{train+test}

Five-fold CV



Environment	total n
All Env	1287
Both	369
White	904

Genomic Prediction

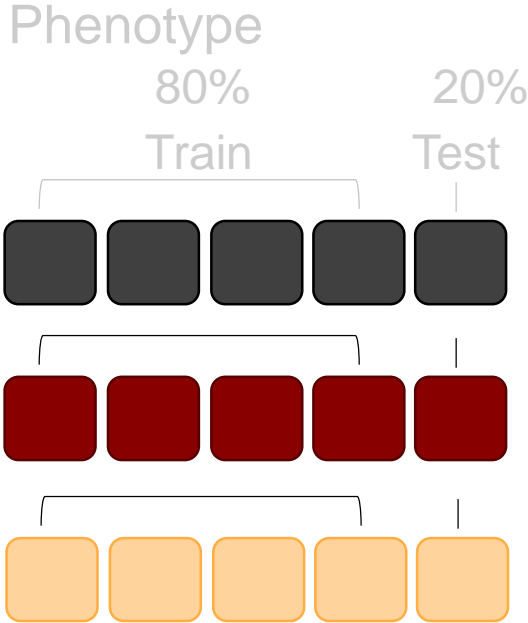
Prediction Accuracy (PA)

Prediction Model

$cor(y_{obs}, GEBVs_{model})$
 $cor(y_{obs_test}, GEBVs_{test})$

- Gaussian kernel
- Ridge regression

Output:
GEBVs_{train+test}

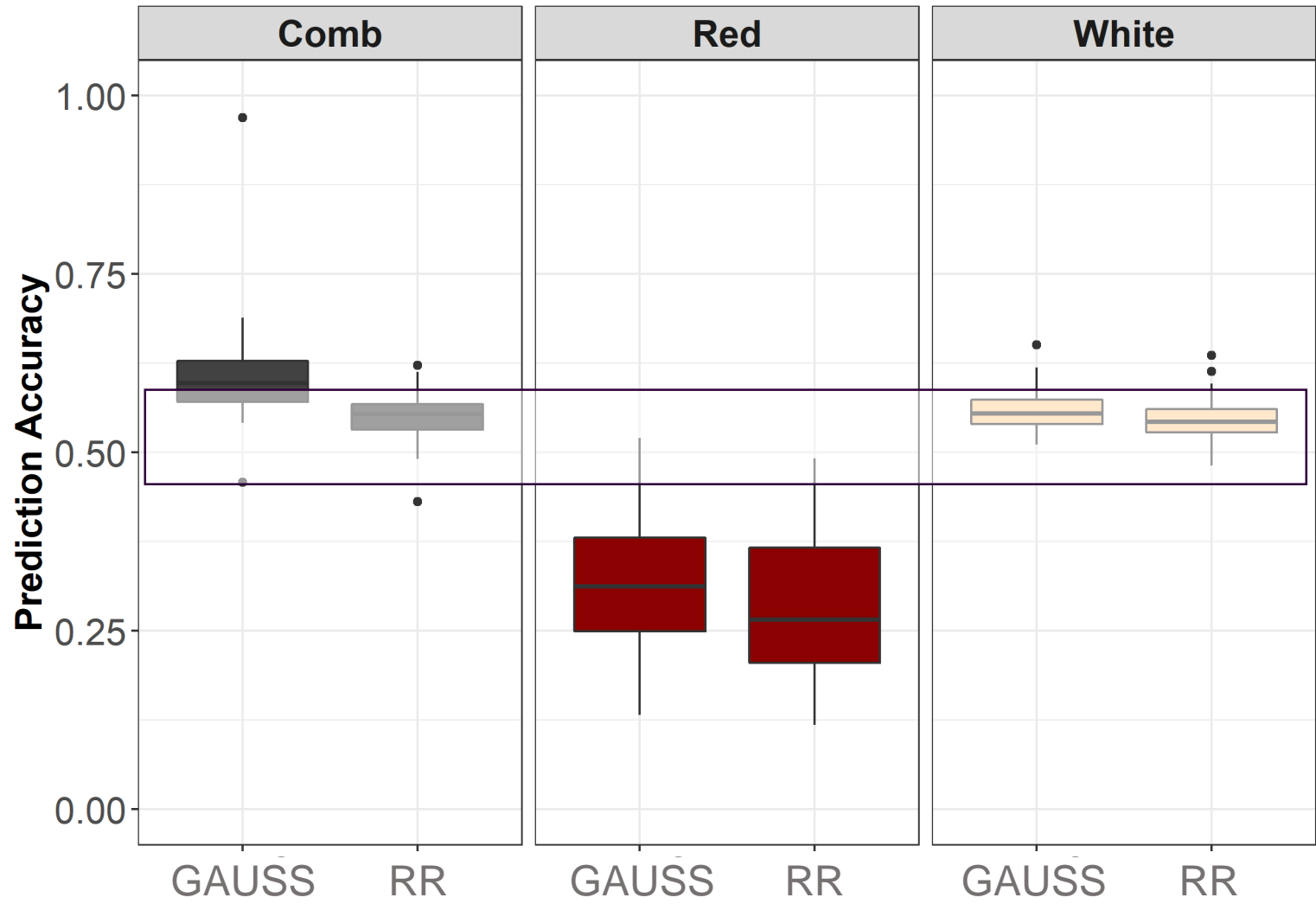


Both $y_{obs_both} \sim Loc + Yr + HarvDate + KColor + (1|variety)$

Red $y_{obs} \sim Loc + Yr + HarvDate + (1|variety)$

White $y_{obs} \sim Loc + Yr + HarvDate + (1|variety)$

A comparison of Ridge Regression and Gaussian kernel models across seed coat colors



Previous PHS GP studies **KC**

Heffner et al. 2011 White

PA 0.52 – 0.53

Heslot et al. 2013 White

PA 0.47 – 0.57

Moore et al. 2017 White

PA 0.49 – 0.59 **& Red**

GAUSS: Gaussian kernel model

RR: Ridge Regression model

What does an accuracy of 0.6 even mean?

Accuracy of an phenotypic estimate (without genetic data)

$$PA = \text{cor}(\text{Obs_EnvA}, \text{Obs_EnvB})$$

`cor(Obs2008, Obs2009)`

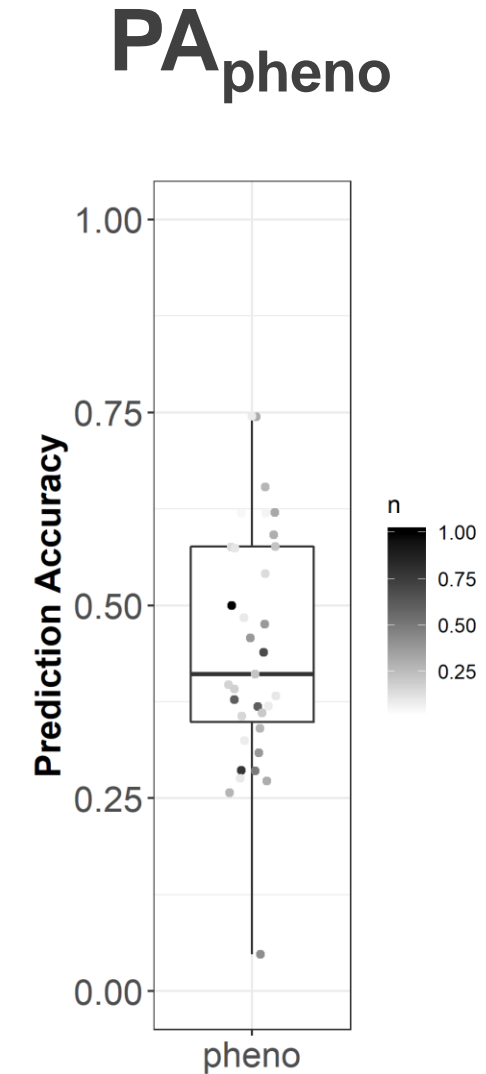
▪

▪

▪

`cor(Obs2015, Obs2017)`

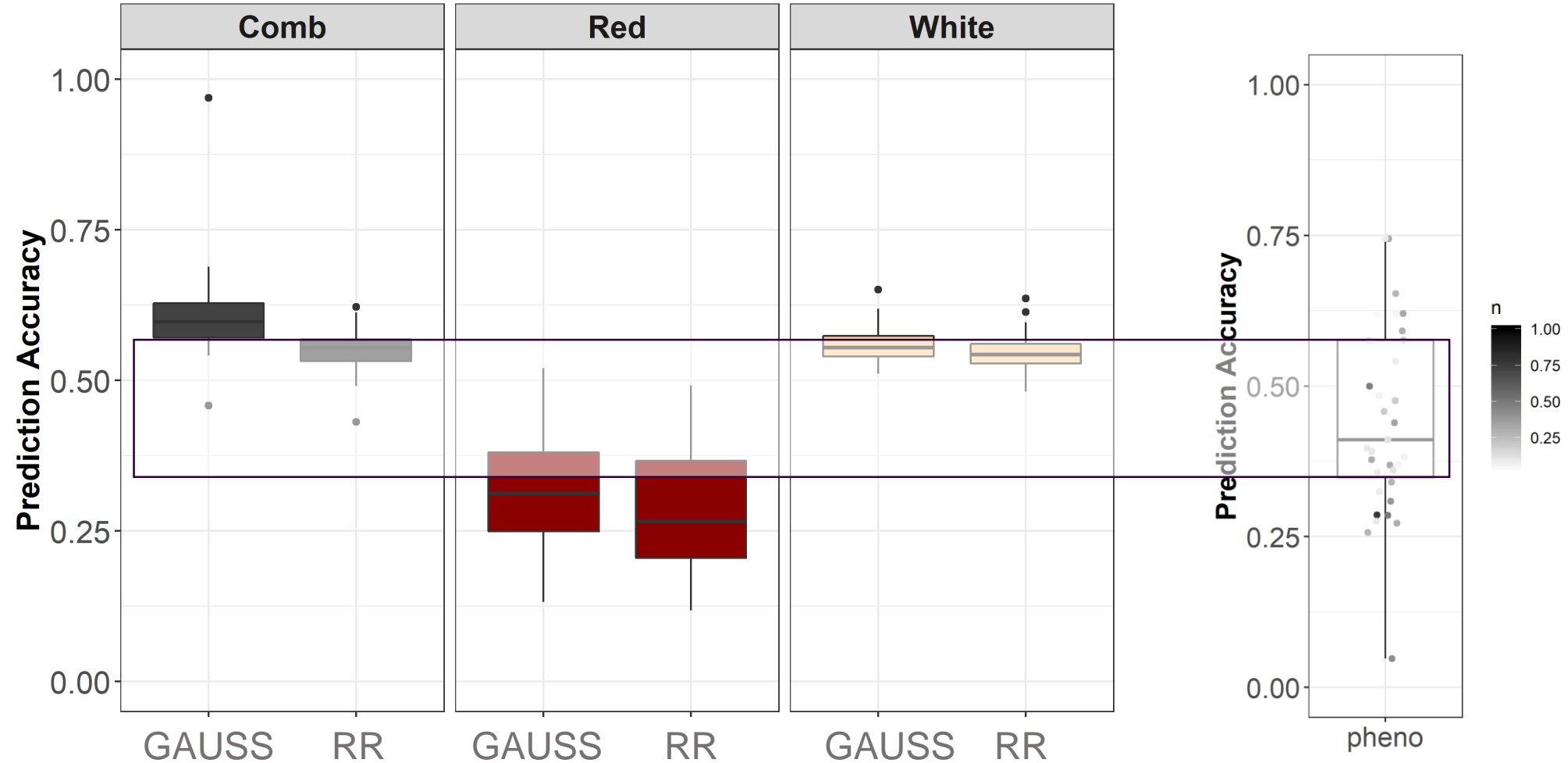
Average PA
0.45



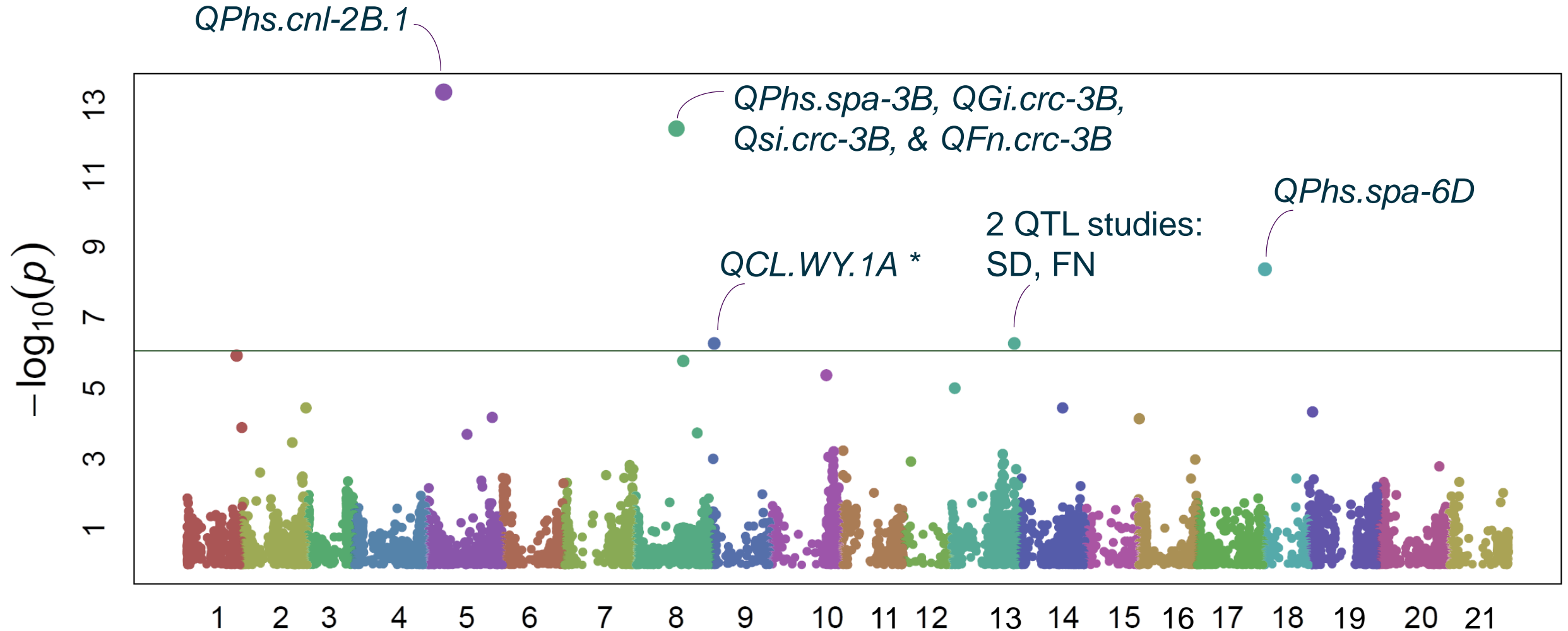
The correlation (PA) from one environment to another environment is comparable, to genomic prediction accuracies.

PA_{pred}

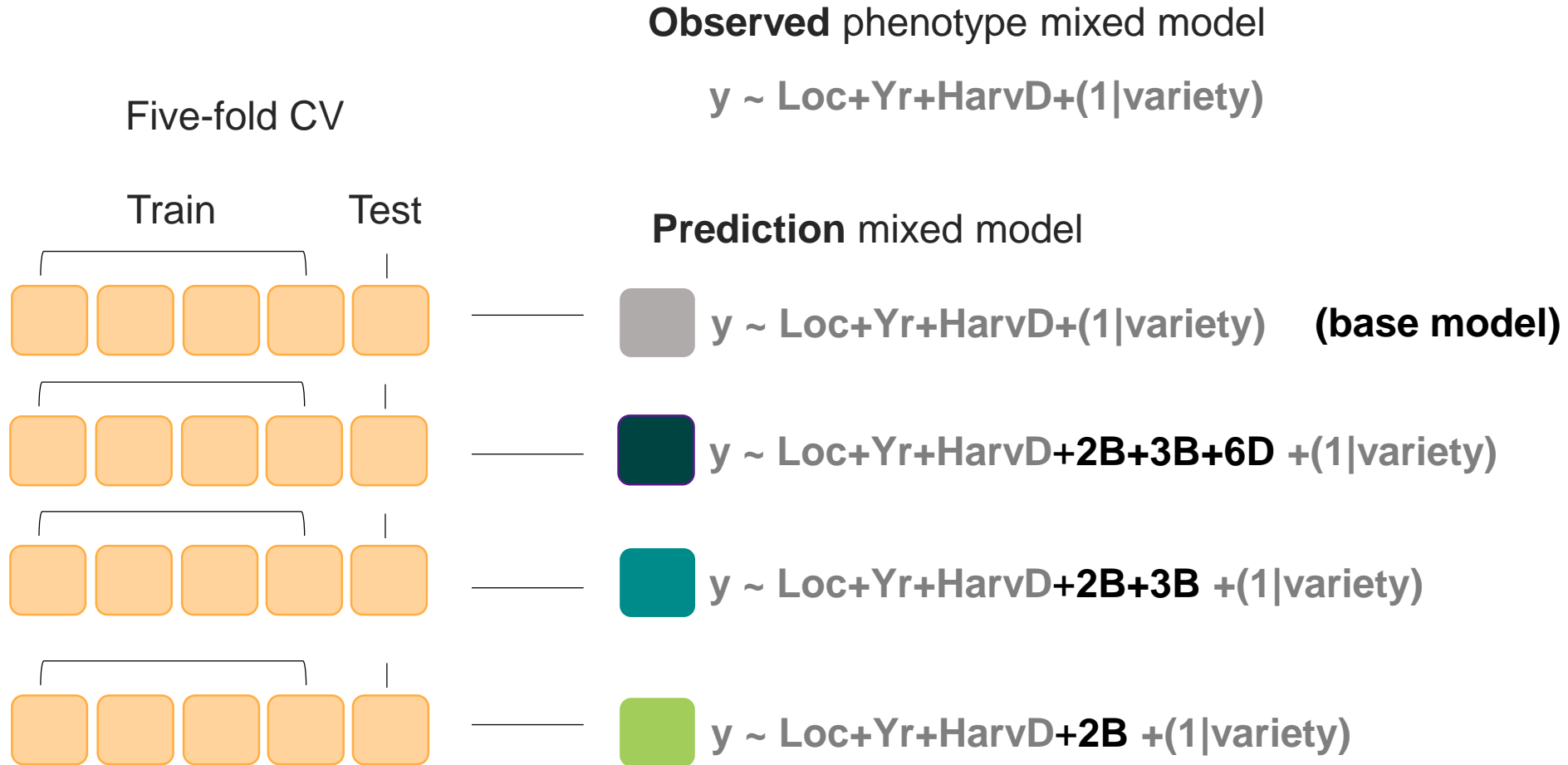
PA_{pheno}



White kernel GWAS found multiple significant loci

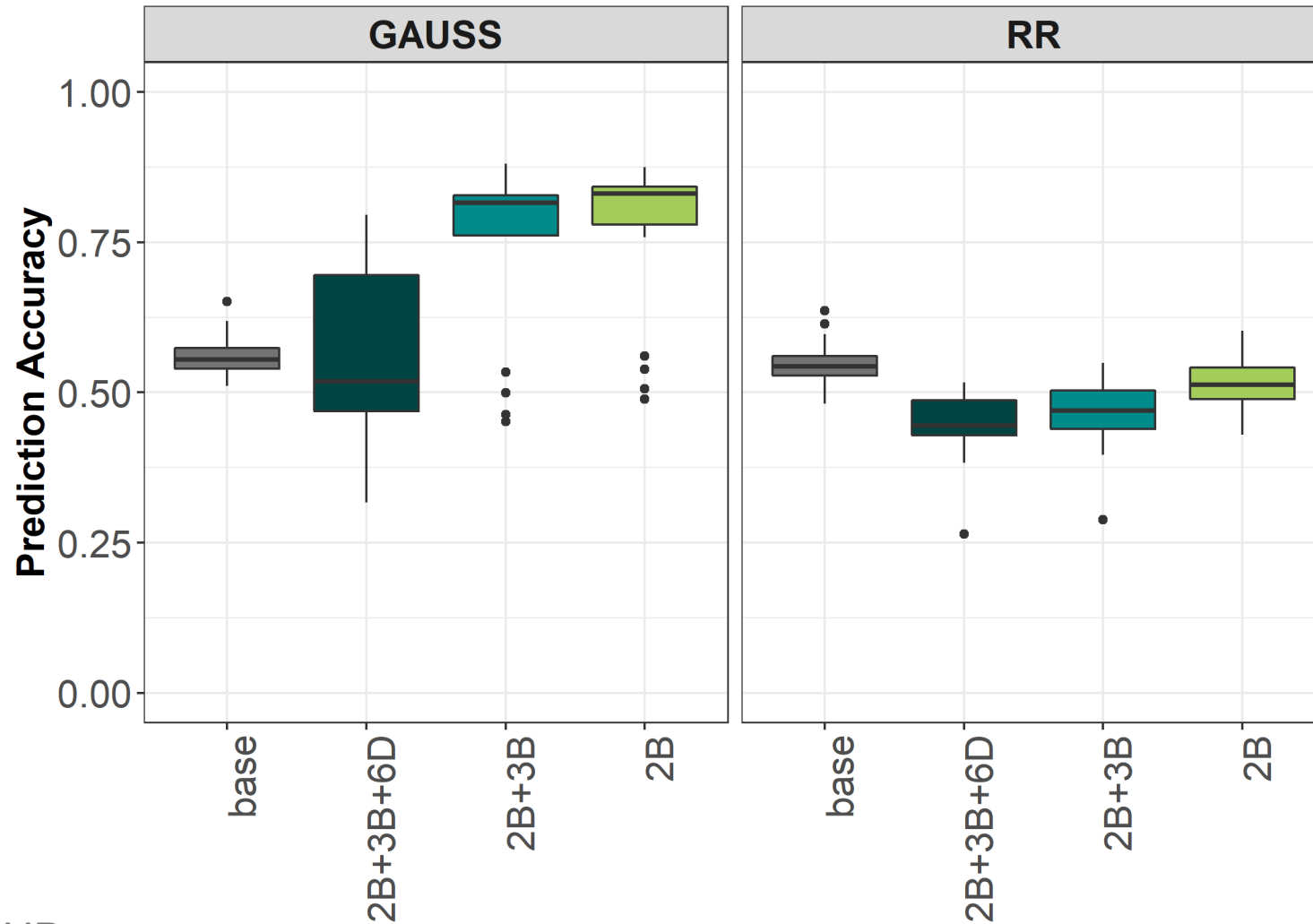


Will associated QTL improve prediction?



$$\text{PA} = \text{cor} (y_{\text{obs_test}} , \text{GEBVs}_{\text{test}})$$

Adding significant QTL as fixed effects seems to improve accuracy for Gaussian kernel model



GAUSS additive + non-additive
RR only additive

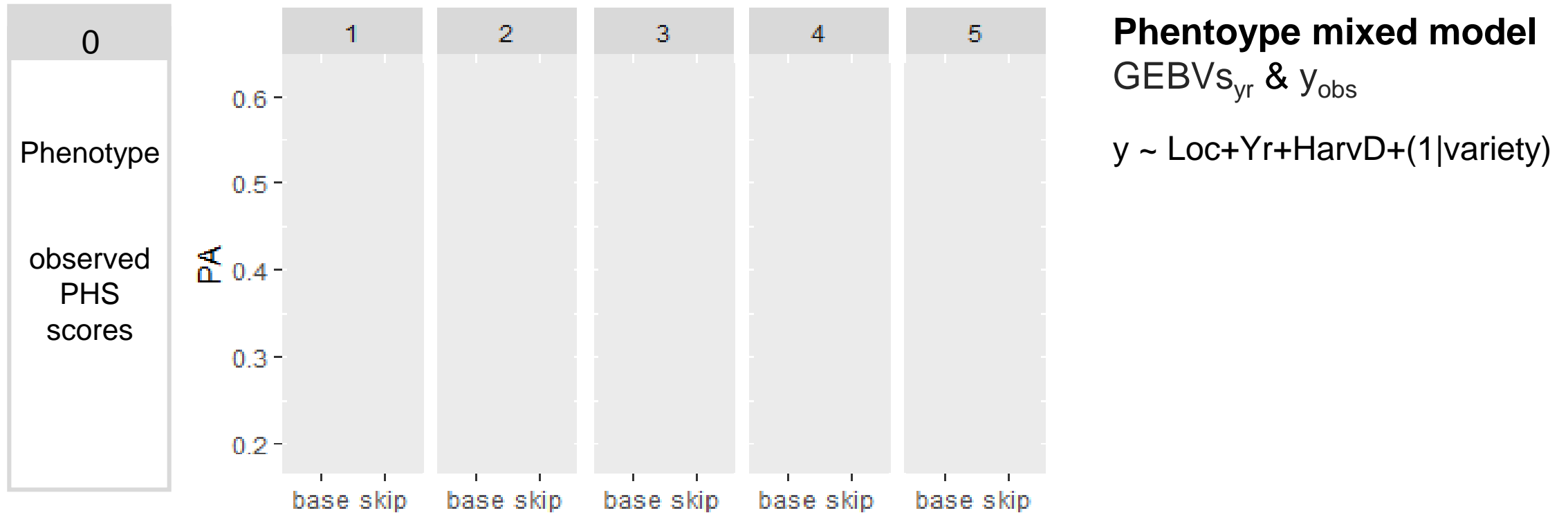
GAUSS vs RR

The real question is, what do you want to use it for?

- Want additive RR for selection
- Want +non-add for prediction

Need to investigate if adding significant QTL markers as a fixed effect is overfitting the model?

How many years between phenotyping can a breeding program skip without losing substantial PA?



Prediction Accuracy

base: if phenotype every year, and predict the following year

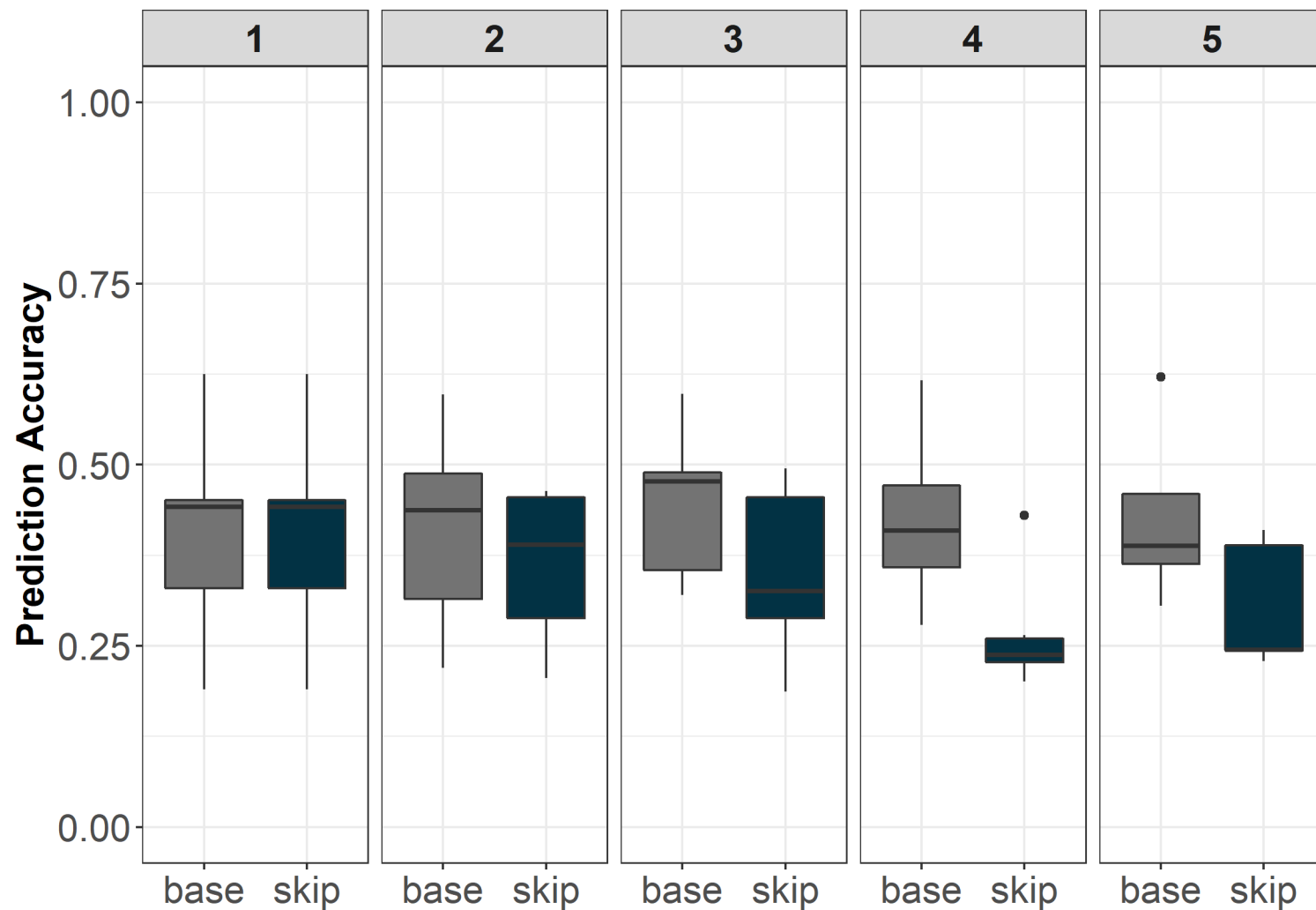
$$PA = \text{cor}(y_{obs_yr0+\dots+yr(N-1)}, GEBVs_{yrN})$$

skip: 1, 2, 3, 4, 5 years between phenotyping

$$PA = \text{cor}(y_{obs_yr0}, GEBVs_{yrN})$$

Ridge Regression
White KC only

How many years between phenotyping can a breeding program skip without losing substantial PA?



It may be possible to phenotype for PHS tolerance every couple of years, without effecting accuracy

Something to consider:

- Increased genetic diversity year to year will reduce ability to accurately calculated GEBVs
- Year to year GEBV calculations tend to have lower PA than GEBV calculated from hundreds of individuals over multiple environments

Ridge Regression
White KC only



Why use genomic prediction for this trait?

Phenotyping occurs at the end of the growth cycle

No “mid-cycle” selections before harvest



Very labor intensive

Precise sampling

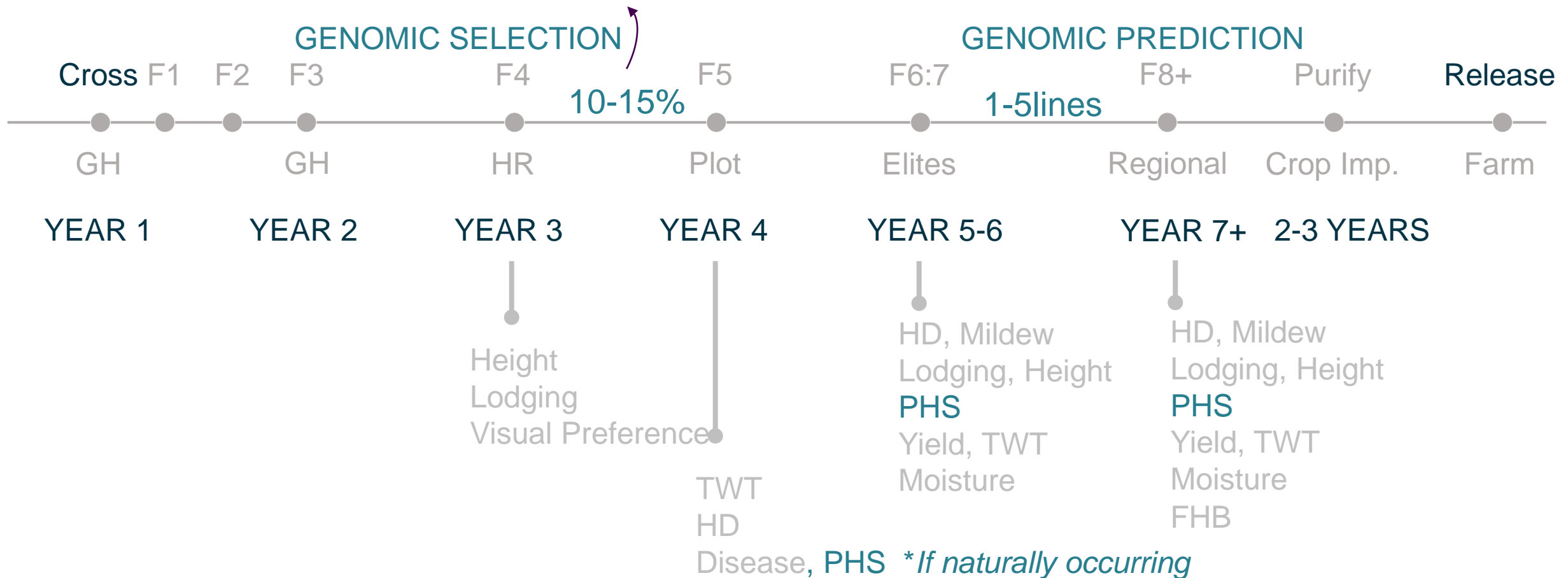


Breeding for PHS tolerance

Still need to confirm selected lines with phenotyping in later generations

GEBVs can help guide selection; less likely to throw out PHS Tolerant lines when applying selection pressure

Assist in predicting PHS tolerance for variety selection, if unable to apply high frequency phenotypic pressure



ACKNOWLEDGMENTS

Cornell Small Grains Group	Data Collection & Support
Jannink & Sorrells Labs	Project Feedback
Daniel Sweeney	Genomic Prediction theory
Stephanie Sjoberg	Published QTL Position

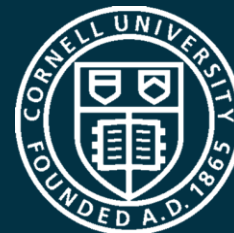
PROJECT FUNDING

Education and Workforce
Development (EWD) Proposal #2017-07109



United States
Department of
Agriculture

National Institute
of Food and
Agriculture



Cornell University

Thank you for your attention

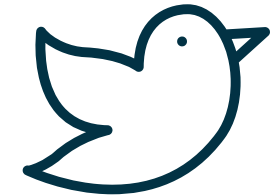
Wed. Poster #1571

LinkedIn

Website

Email

Twitter



shantel-a-martinez

shantel-martinez.github.io

shantel.a.martinez@gmail.com

sam594@cornell.edu

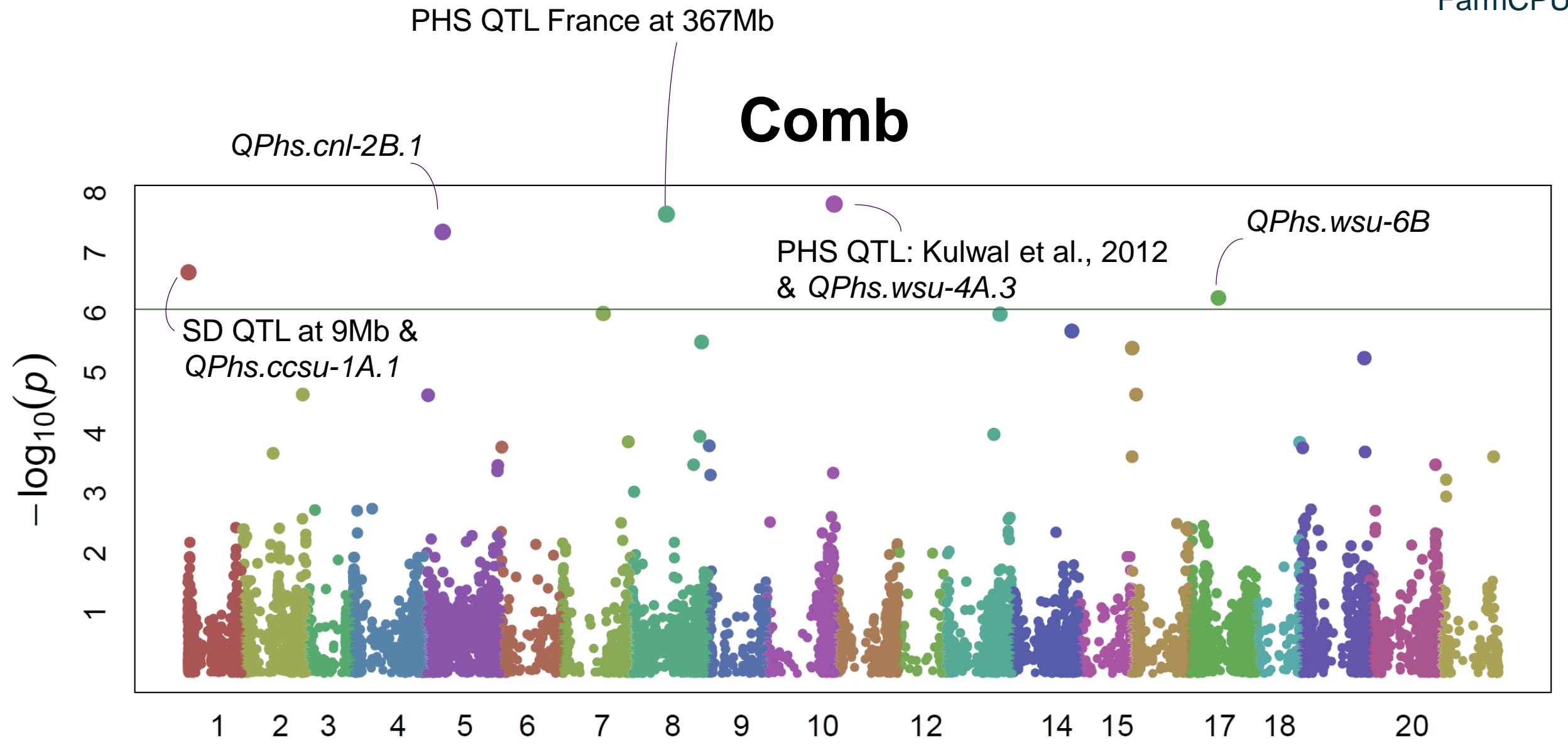
@s_amealia

REFERENCES

- Endelman JB (2011) Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome* 4:250–255. <https://doi.org/10.3835/plantgenome2011.08.0024>
- Fofana B, Humphreys DG, Rasul G, et al (2009) Mapping quantitative trait loci controlling pre-harvest sprouting resistance in a red x white seeded spring wheat cross. *Euphytica* 165:509–521. <https://doi.org/10.1007/s10681-008-9766-6>
- Heffner EL, Jannink J-L, Sorrells ME (2011) Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. *The Plant Genome* 4:65. <https://doi.org/10.3835/plantgenome2010.12.0029>
- Heslot N, Jannink J-L, Sorrells ME (2013) Using Genomic Prediction to Characterize Environments and Optimize Prediction Accuracy in Applied Breeding Data. *Crop Science* 53:921–933. <https://doi.org/10.2135/cropsci2012.07.0420>
- Kumar S, Knox RE, Clarke FR, et al (2015) Maximizing the identification of QTL for pre-harvest sprouting resistance using seed dormancy measures in a white-grained hexaploid wheat population. *Euphytica* 205:287–309. <https://doi.org/10.1007/s10681-015-1460-x>
- Martinez SA, Godoy J, Huang M, et al (2018) Genome-wide Association Mapping for Tolerance to Preharvest Sprouting and Low Falling Numbers in Wheat. *Front Plant Sci* 9:1–16. <https://doi.org/10.3389/fpls.2018.00141>
- Moore JK, Manmathan HK, Anderson VA, et al (2017) Improving Genomic Prediction for Pre-Harvest Sprouting Tolerance in Wheat by Weighting Large-Effect Quantitative Trait Loci. *Crop Science* 0: <https://doi.org/10.2135/cropsci2016.06.0453>
- Munkvold JD, Tanaka J, Benscher D, Sorrells ME (2009) Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. *Theor Appl Genet* 119:1223–1235. <https://doi.org/10.1007/s00122-009-1123-1>

REFERENCES CON'T

- Zanetti S, Winzeler M, Keller M, et al (2000) Genetic analysis of pre-harvest sprouting resistance in a wheat × spelt cross. *Crop Science* 40:1406–1417. <https://doi.org/10.2135/cropsci2000.4051406x>
- Zhang H, Cui F, Wang L, et al (2013) Conditional and unconditional QTL mapping of drought-tolerance-related traits of wheat seedling using two related RIL populations. *J Genet* 92:213–231. <https://doi.org/10.1007/s12041-013-0253-z>
- Zuo J, Lin C-T, Cao H, et al (2019) Genome-wide association study and quantitative trait loci mapping of seed dormancy in common wheat (*Triticum aestivum* L.). *Planta* 250:187–198. <https://doi.org/10.1007/s00425-019-03164-9>



PHS_Blup_Red

