



Generating genomic predictions and understanding mega-environments with real world breeding data

Jeanette Lyerly

Wheat CAP Genomic Selection Workshop
July 2023, Raleigh, NC

NC STATE UNIVERSITY



A little bit of history...

SunGrains - seven university small grain breeding and genetics group

USDA-ARS Eastern Regional Small Grains Genotyping Lab

Collaboration to implement GS in southeastern wheat breeding programs





Training population selection and use of fixed effects to optimize genomic predictions in a historical USA winter wheat panel

J. Martin Sarinelli¹ · J. Paul Murphy¹ · Priyanka Tyagi¹ · James B. Holland^{1,2} · Jerry W. Johnson³ · Mohamed Mergoum³ · Richard E. Mason⁴ · Ali Babar⁸ · Stephen Harrison⁵ · Russell Sutton⁶ · Carl A. Griffey⁷ · Gina Brown-Guedira^{1,2}

Received: 31 January 2018 / Accepted: 7 January 2019 / Published online: 24 January 2019
© The Author(s) 2019

Used cross-validation to evaluate predictive ability in an unbalanced data set of 467 winter wheat genotypes evaluated in 49 environments from 2008 to 2016. (*Gulf-Atlantic Nursery*)

34,095 SNP from GBS

Maximum predictabilities were 0.64 for grain yield, 0.56 for test weight

Results demonstrated the utility of combining unbalanced phenotypic records with genome-wide SNP marker data for predicting the performance of untested genotypes



SunGrains



Mega-environments

- Mega-environments are large areas with similar biotic and abiotic stresses
 - Weather conditions, soil conditions, disease pressures, etc.
- Target our predictions to these broad environments
- Identify lines that may perform well:
 - Across multiple environments
 - Within a given environment



Traits evaluated by breeders and ARS scientists



Grain yield



Test weight



Plant height



Heading date



Powdery mildew



Septoria blotch



Leaf rust



Stripe rust

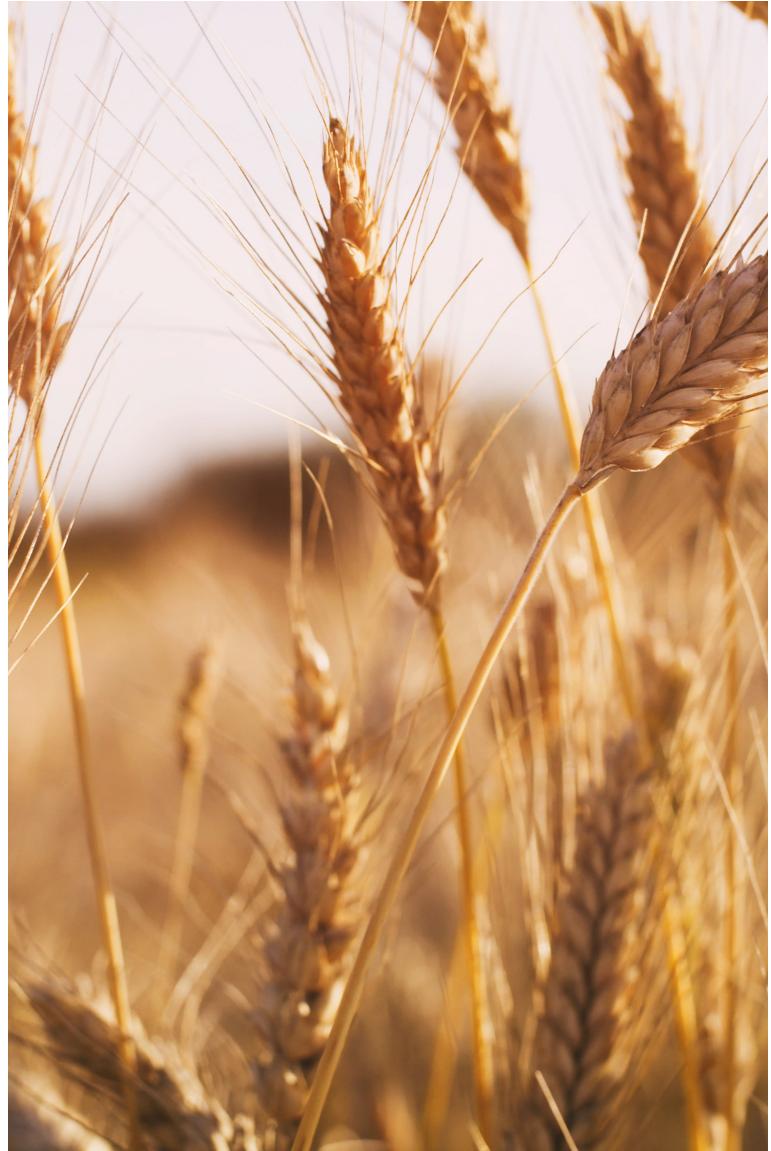


Head scab (FHB)



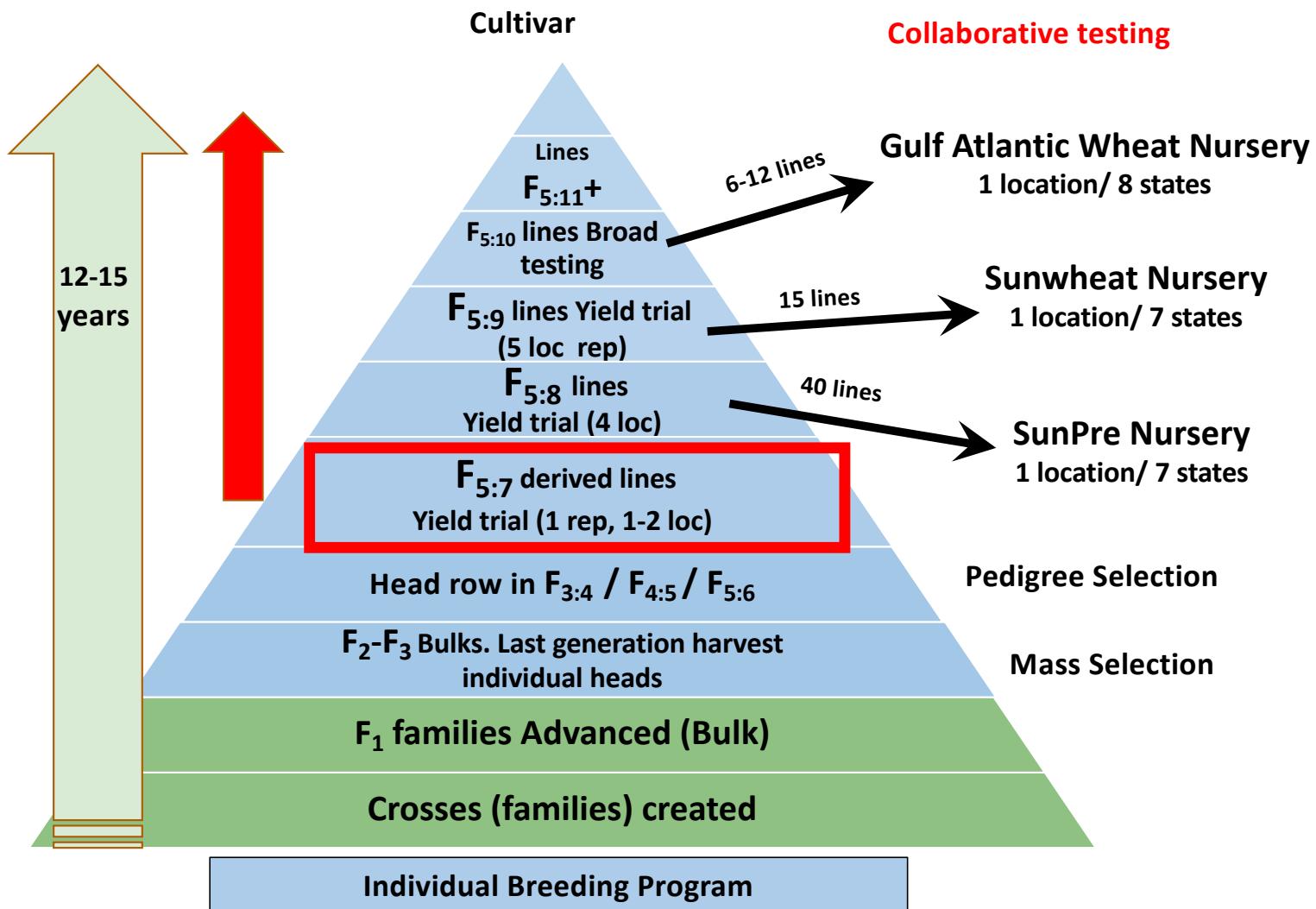
Hessian fly

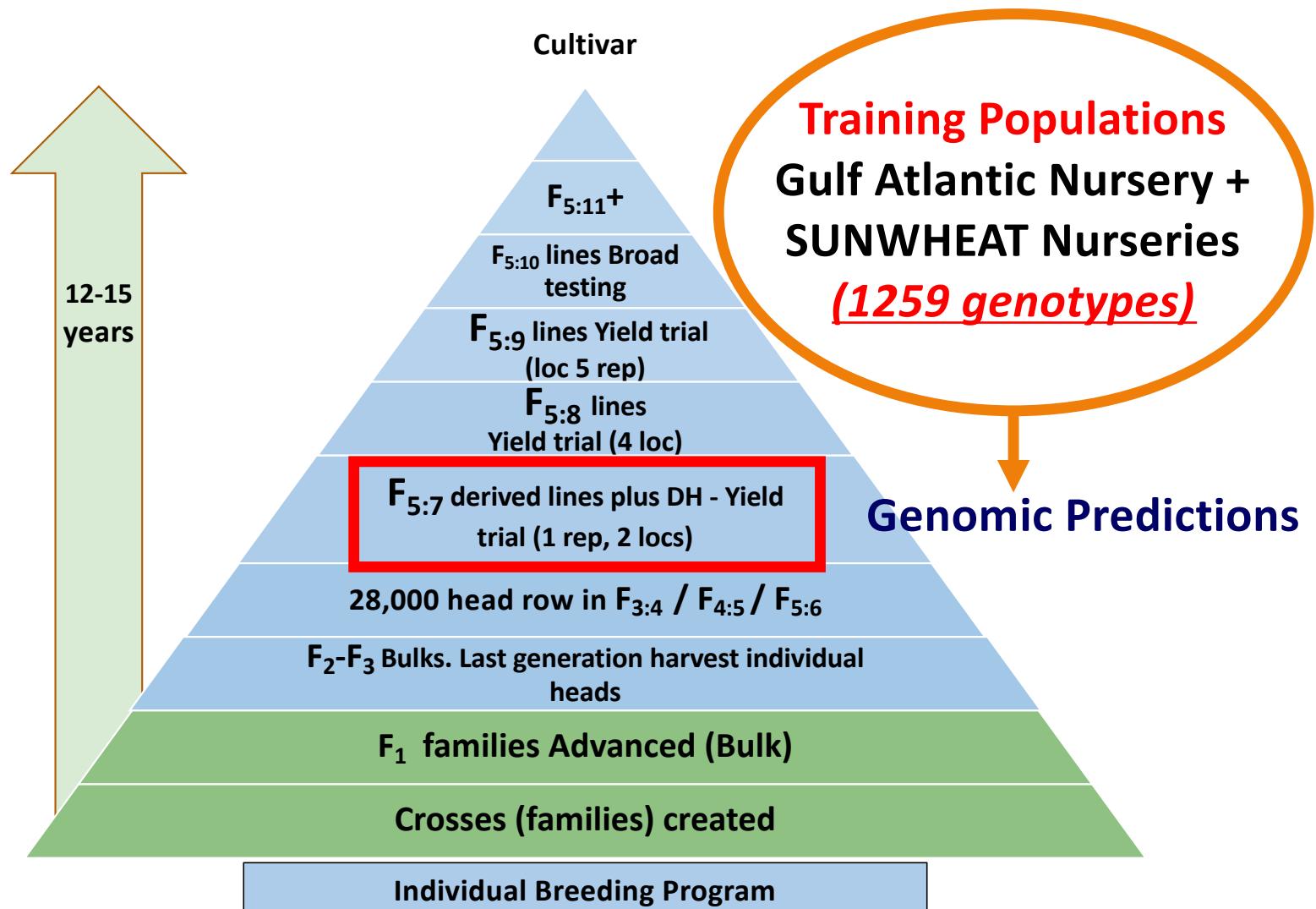




Training population composition: GAWN and Sunwheat

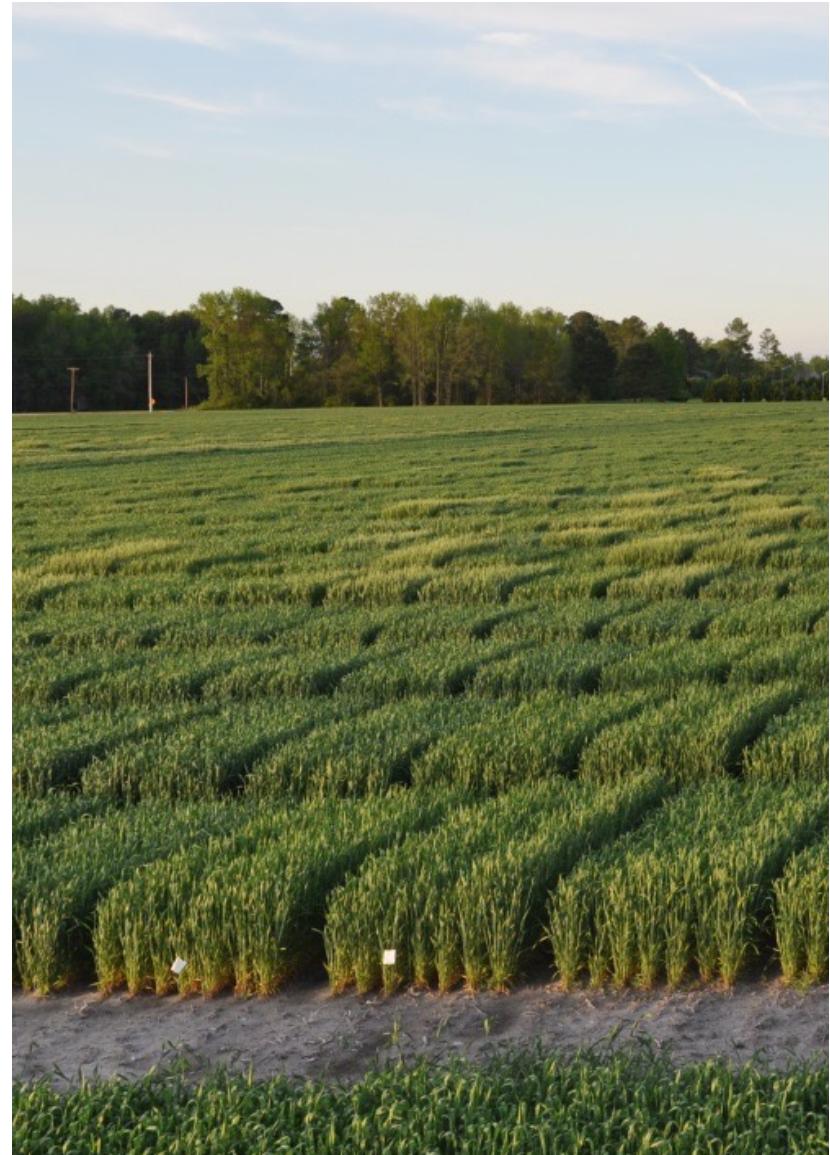
- Historical data from cooperative nurseries
 - Wealth of available data sampled over space and time
 - Lots of locations over many years
 - Very unbalanced
- Leveraging historic data from collaborative testing nurseries saves time and cost
- The training population is improved each year with the addition of new lines and environments



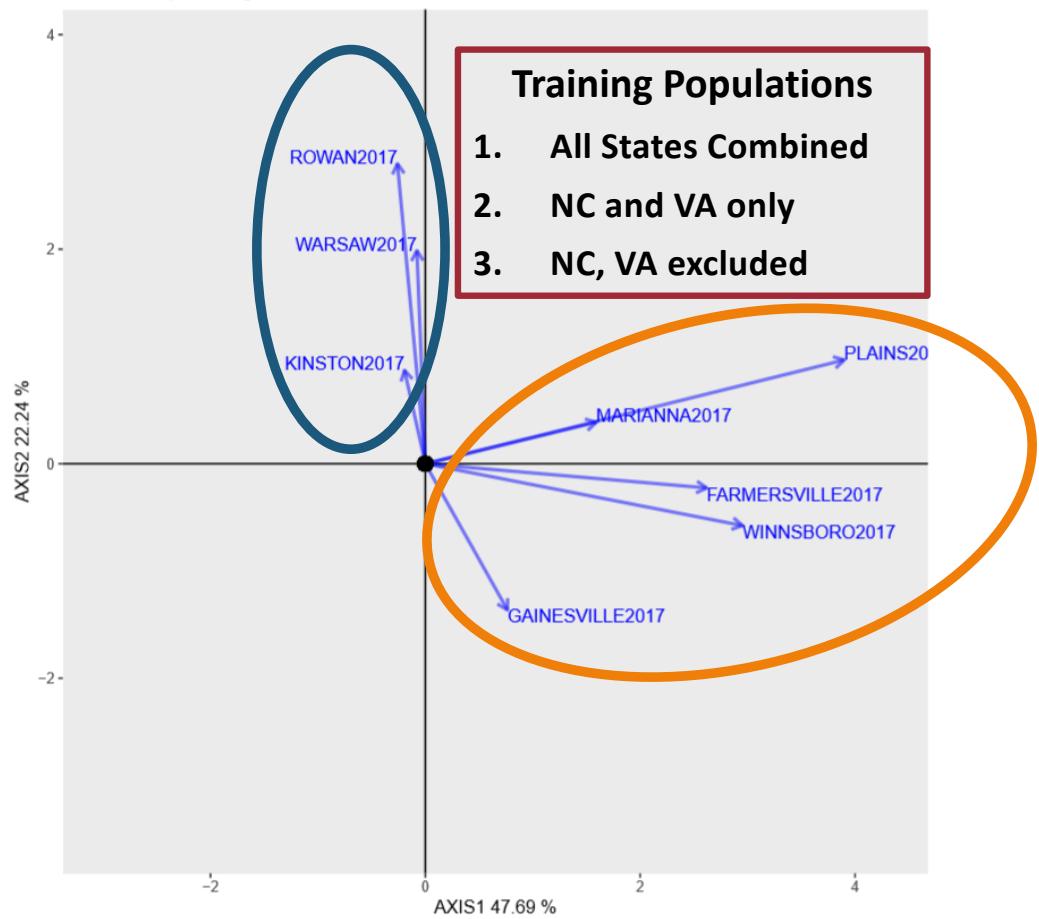


Training populations

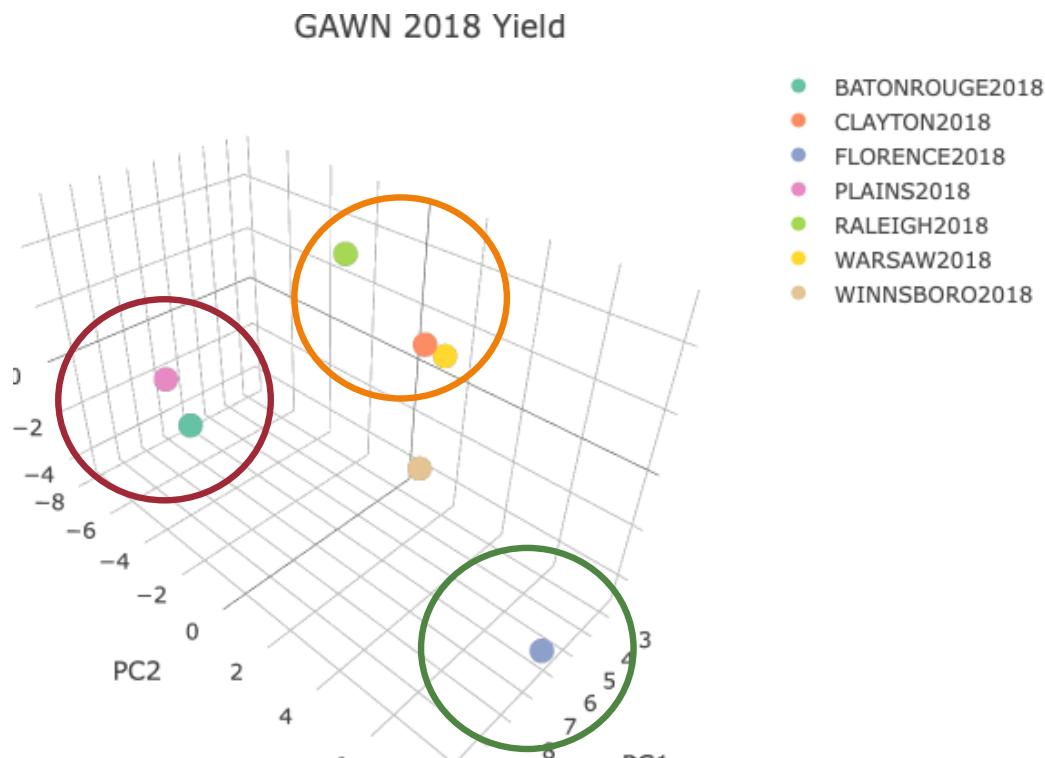
- Training Populations
 - Combined/whole region training population for disease/insect traits (powdery mildew, stripe rust, Hessian fly)
 - More specialized regional training populations for yield, test weight, heading date, and height
- Methods for selecting locations
 - Preliminary data analysis
 - PC plots (yield)
 - Biplots (yield)
 - Mean check performance
 - Cluster analysis



GGE Biplot to assist training population development



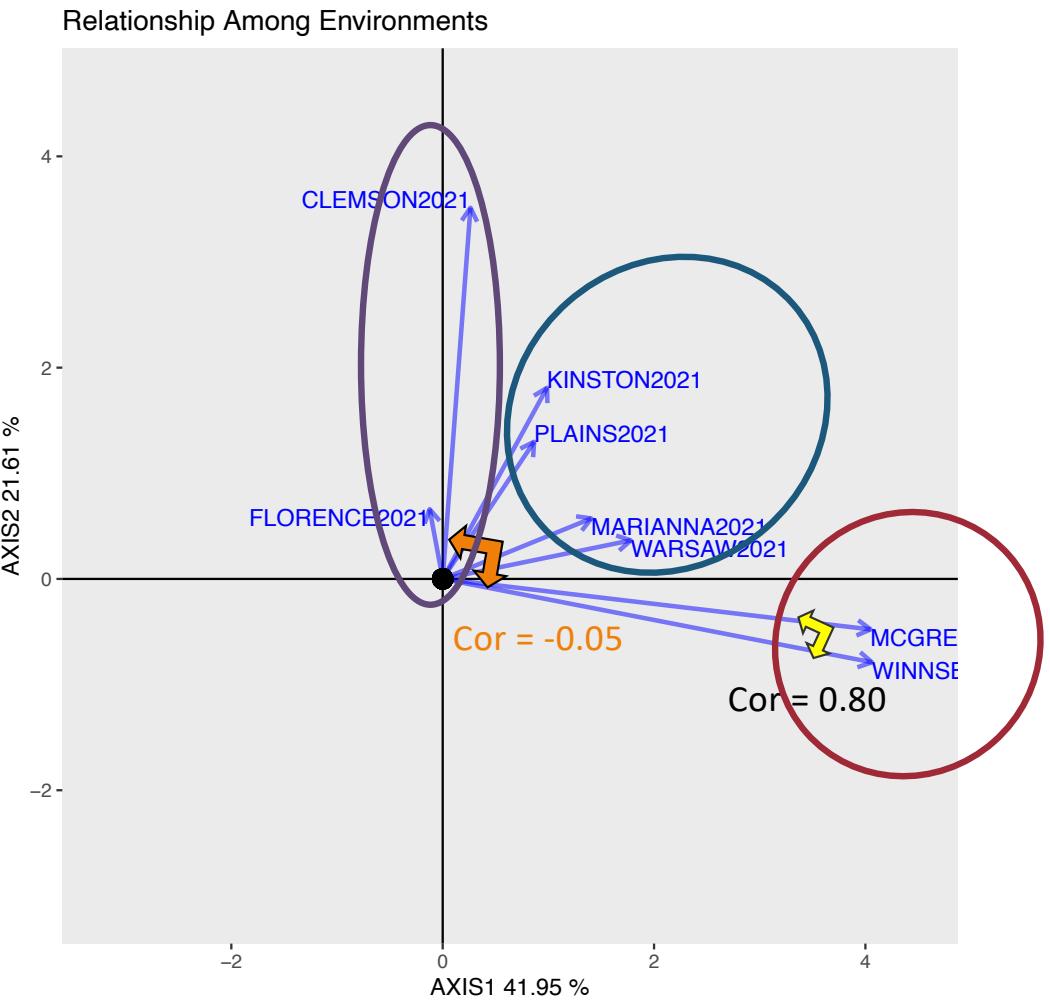
Example PC plot – GAWN 2018 Yield



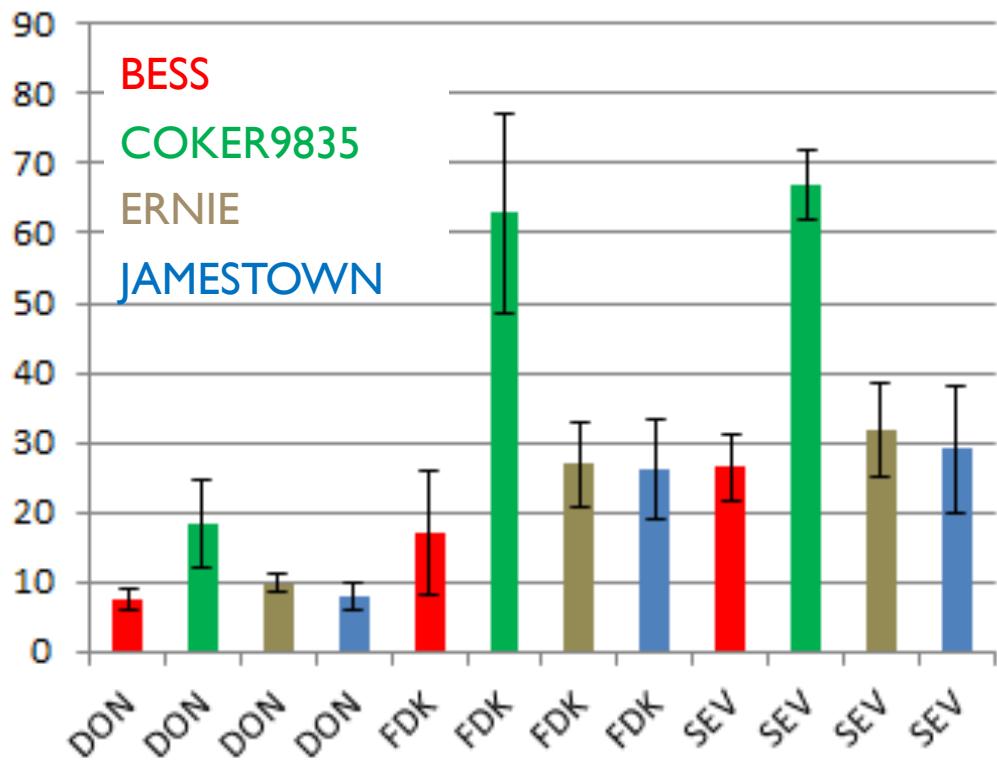
- NC and VA locations tend to cluster together
- Pattern of NC and VA grouping separately from other locations is common across nurseries and years
- 3D image can help resolve questions about groupings for environments

Example biplot – GAWN 2021 Yield

- General trend for NC and VA, may also include some locations from AR and GA
- Length of the vector – discriminating ability
- Angle of the vectors – correlations among the environments



Differences among check varieties

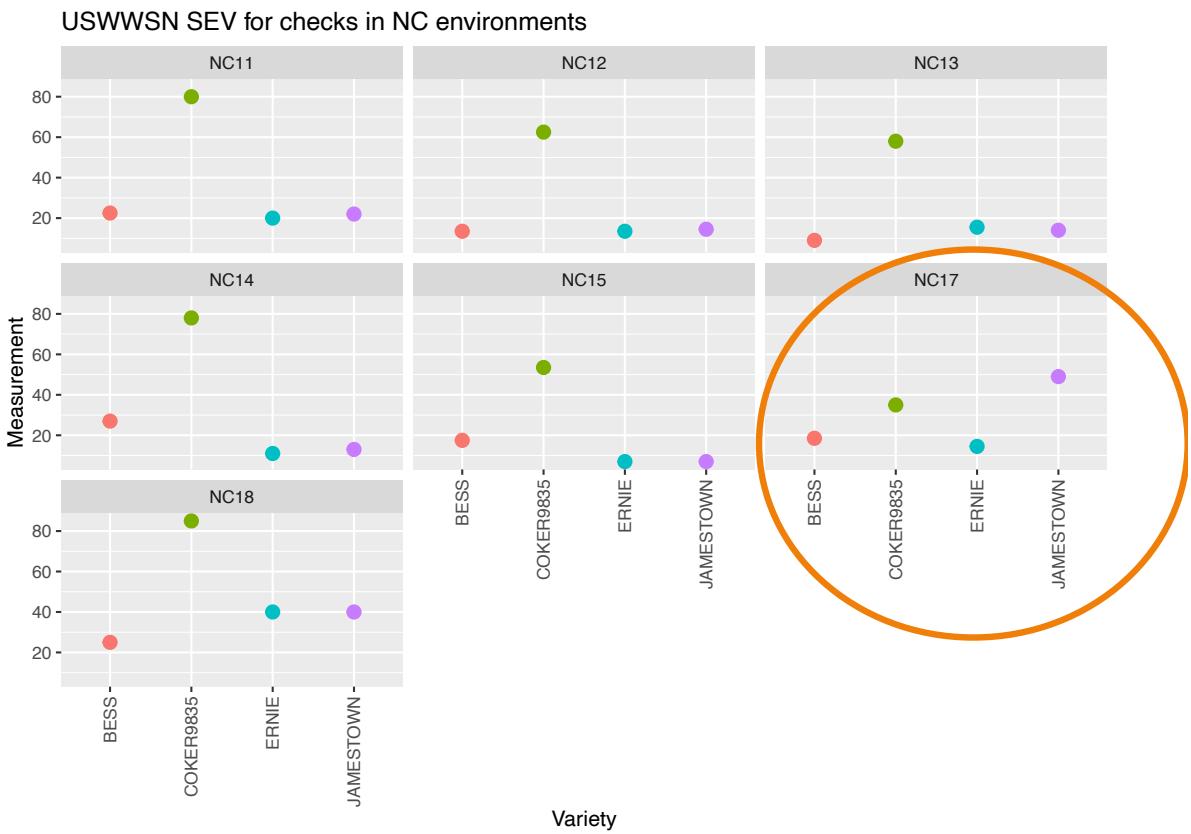


Example plot of means of check varieties

- Look at differences between the means of checks in the Uniform Southern Scab Nursery
- Way to look at the data to see how check varieties performed – tells us something about disease pressure in the field

FHB checks in NC locations

- FHB SEV
- Example of performance of checks in different environments
- Atypical or low disease pressure environments may be less informative for training



Clustering for selecting environments

- Checks are repeated over years
- Clustering to partition environments
- Select which environments to include



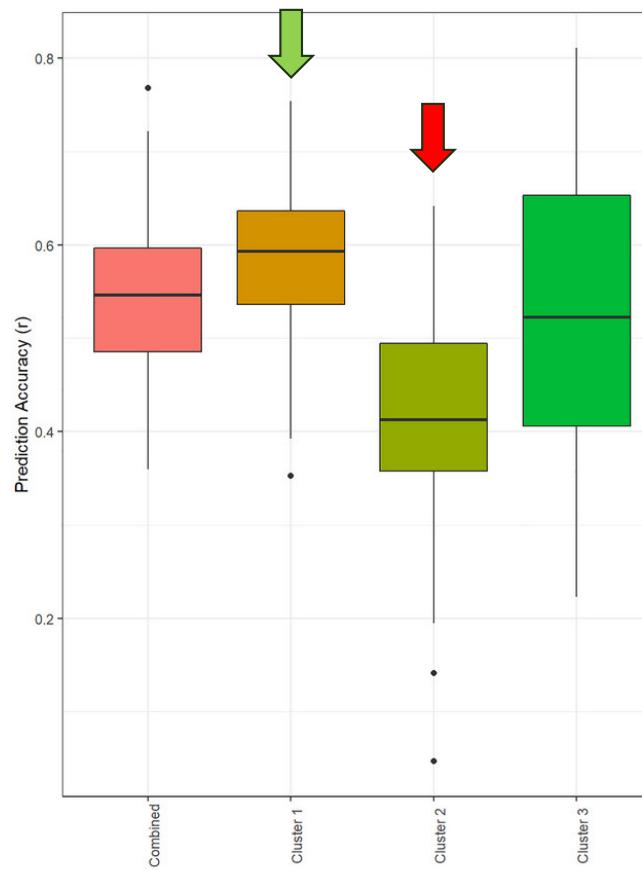
ORIGINAL ARTICLE | Open Access | CC BY

Utilization of a publicly available diversity panel in genomic prediction of *Fusarium* head blight resistance traits in wheat

Zachary J. Winn , Jeanette H. Lyerly , Gina Brown-Guedira, Joseph P. Murphy, Richard Esten Mason

First published: 17 May 2023 | <https://doi.org/10.1002/tpg2.20353>

Cross validation results for DON



Forward validation population	Trait	Training population	Prediction accuracy (r)
USWWSN20	FDK	Combined	0.62
		Cluster 1	0.70
		Cluster 2	0.37
USWWSN21	DON	Combined	0.67
		Cluster 1	0.77
		Cluster 2	0.52
		Cluster 3	0.39
USWWSN21	FDK	Combined	0.53
		Cluster 1	0.58
		Cluster 2	0.50
USWWSN21	DON	Combined	0.44
		Cluster 1	0.53
		Cluster 2	0.41
		Cluster 3	0.38

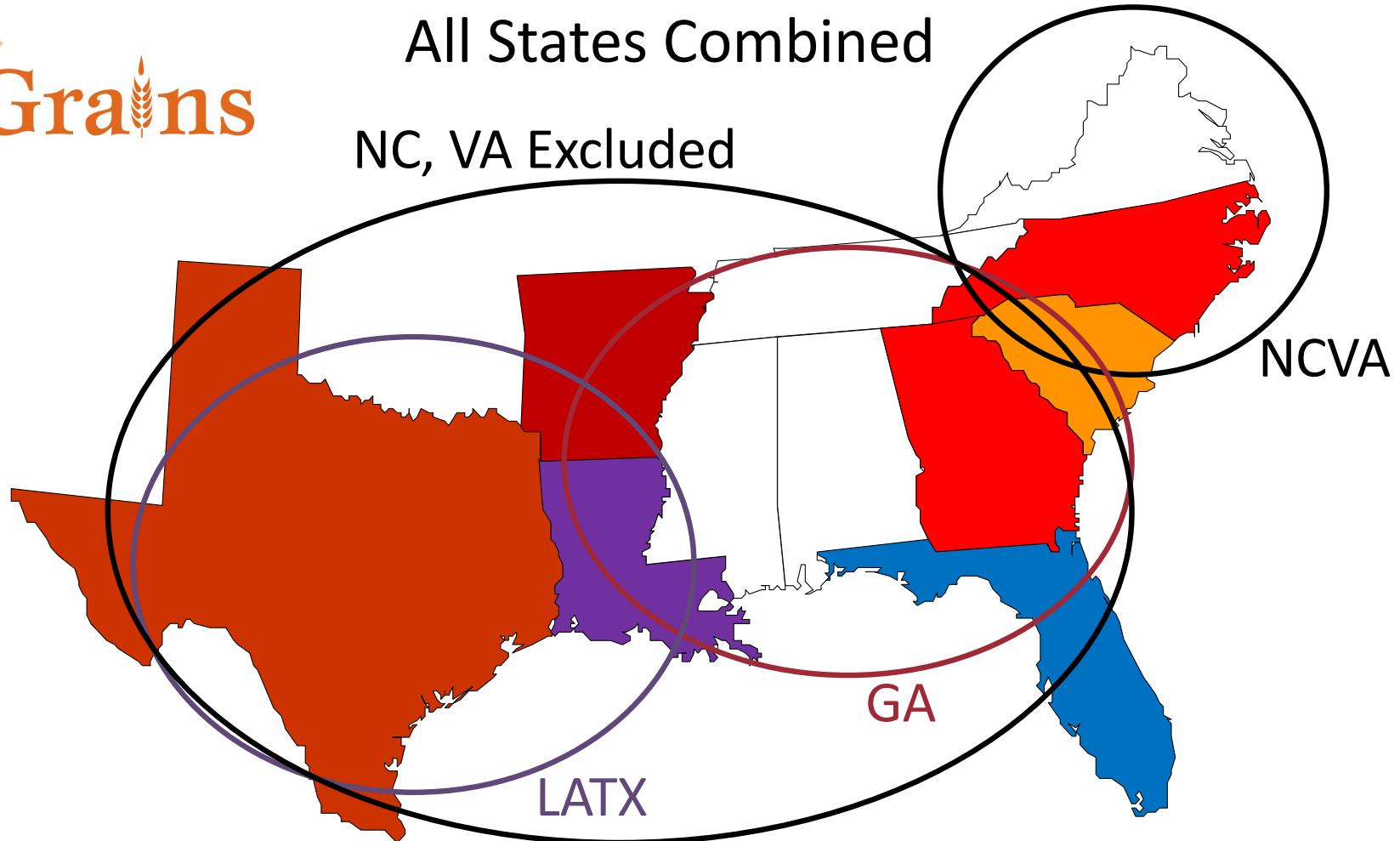
SunGrains training populations

	Regional Training Populations + FHB Training Population	Lines from the GAWN + Sunwheat
Year 1: 2016-2017	3 Regional + 1 FHB	623 from years 2008-2016
Year 2: 2017-2018	4 Regional + 1 FHB	761 from years 2008-2017
Year 3: 2018-2019	5 Regional + 1 FHB	862 from years 2008-2018
Year 4: 2019-2020	5 Regional + 1 FHB	989 from years 2008-2019
Year 5: 2020-2021	5 Regional + 1 FHB	1122 from years 2008-2020
Year 6: 2021-2022	5 Regional + 1 FHB	1259 from years 2008-2021



All States Combined

NC, VA Excluded

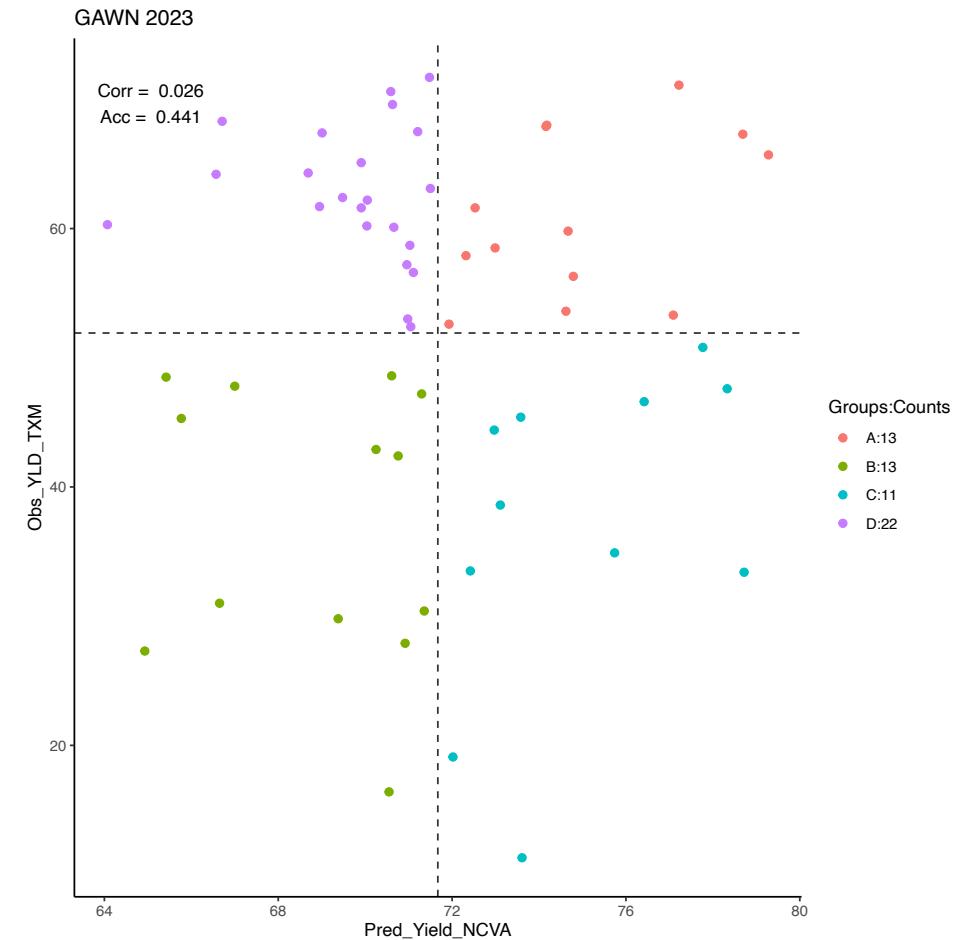
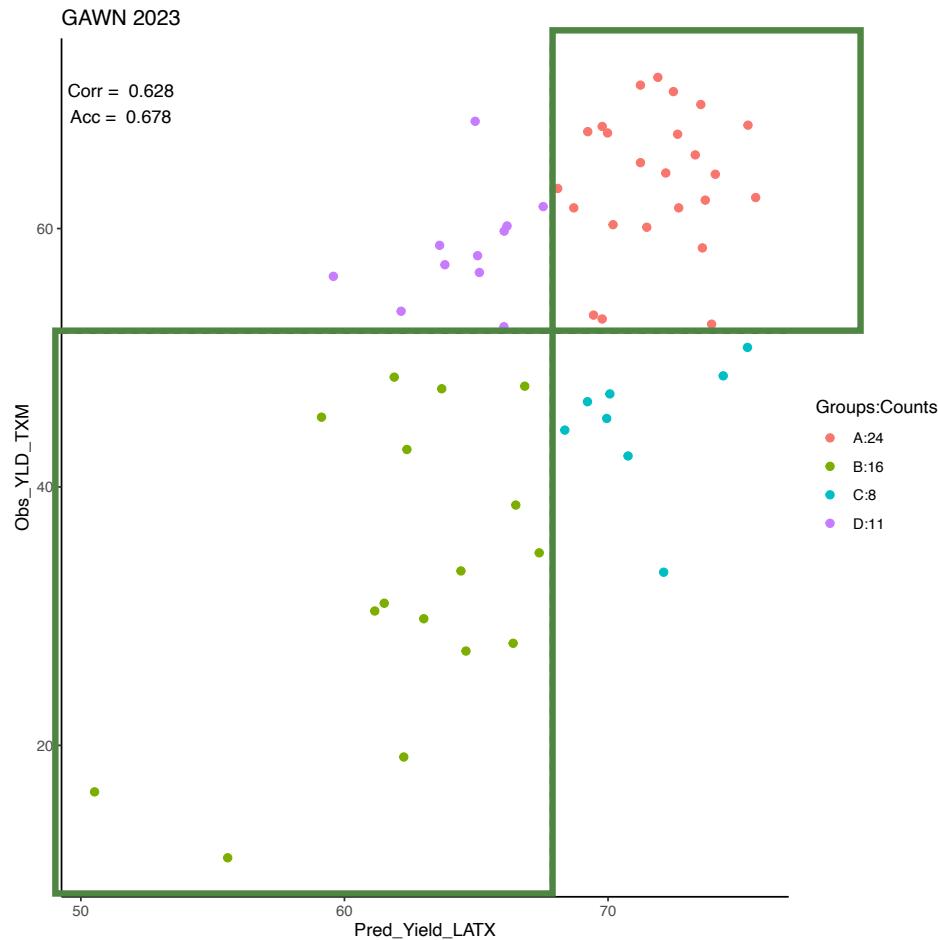


Calculate GEBVs for breeders

- Calculate predictions
 - Forward prediction
 - Cross validation
- Generate predictions for:
 - Observations by state
 - Regional nurseries
 - Update predictions for genotypes that were advanced
- Do our data make sense?
 - Include controls in our data – are they as we expected?
 - Is our cross validation reasonable?



Example results: Observed yield vs predicted yield - TX





Discussion and Hands-on Example

- Multiple ways to characterize mega-environments
 - Climate and soil data
 - Stress index
 - Flowering time/maturity
- Discussion
 - How are other groups characterizing mega-environments?
- Hands-on example