Breeding Consortiums: Using Genomic Selection to Increase Regional Impact

Paul Murphy, North Carolina State University

Clay Sneller, The Ohio State University

Gina Brown-Guedira, USDA-ARS

Jeanette Lyerly, North Carolina State University

Brian Ward, The Ohio State University

Carlos Ignacio, The Ohio State University

Soft Red Winter Wheat "Consortiums"
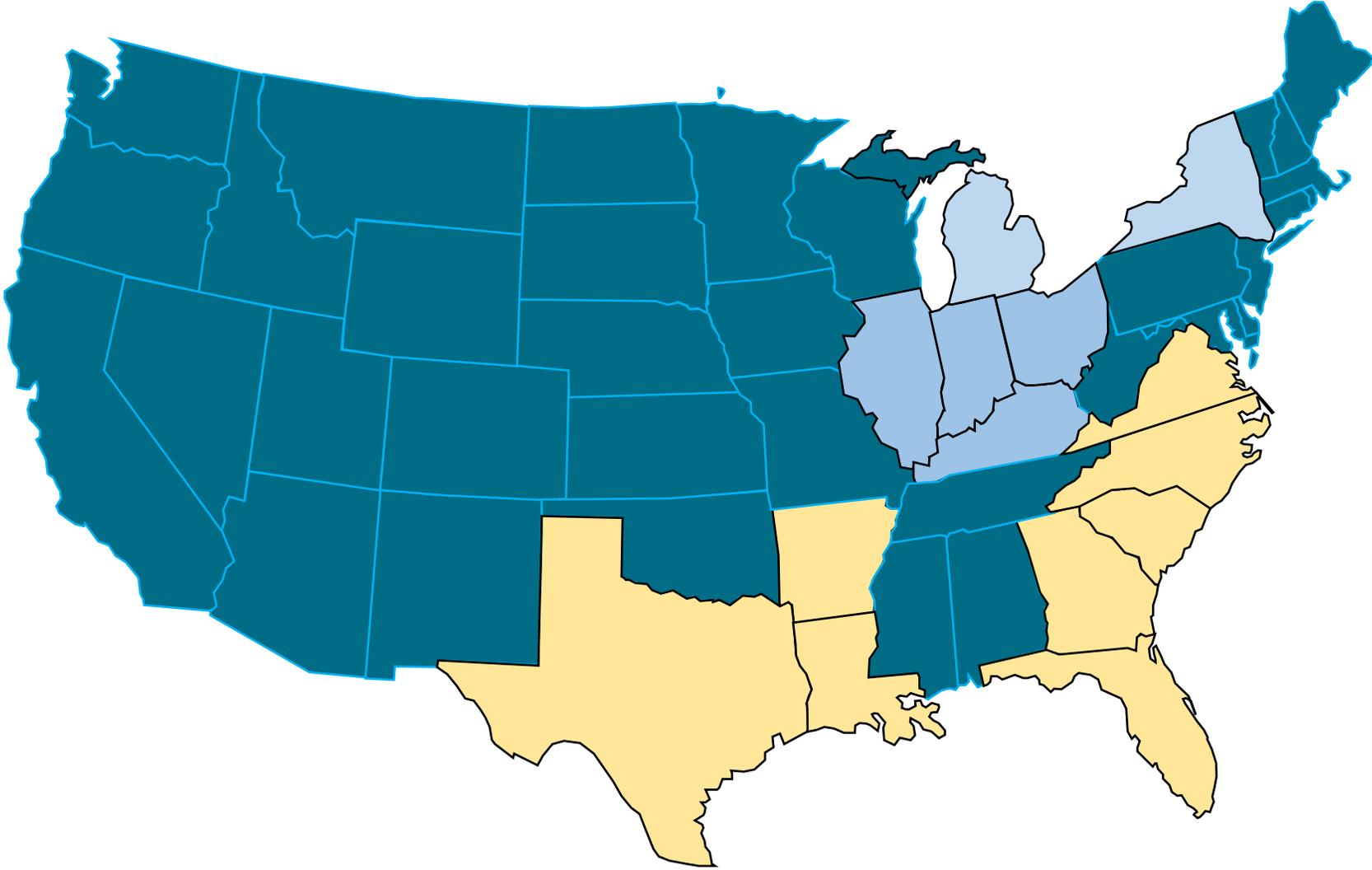
NWW-CP
MI, NY, IL, IN, OH, KY

NIFA
IL, IN, OH, KY

Sun Grains
TX, LA, AR, GA, FL, SC, NC, VA

Potential Benefits of a GS Consortium

- ***Increase Effective Size of Individual Programs:*** Predict the local value of all breeding lines from all stages of testing from each cooperator
- ***Increase size of "cooperative trials":*** Accomplished the goals of uniform trials without extensive phenotyping and apply it to all stages of testing.
- ***Predict Traits Values for Traits you Did Not Phenotype:*** Allow each breeder to predict the value of their germplasm for traits they do not assay
- ***Exploit GEI by incorporating marker by environment interactions*** (MEI)
- ***Understand the relative genetic diversity and genetic value of regional germplasm***.

Requirements for Consortium Success

1. **Germplasm among the member programs must be related**
2. **Germplasm from each member offers value to the other members**
3. **Development of optimal breeding and testing schemes**
4. A common, affordable marker platform
5. Communication among members
6. Common data base for storing phenotypic and genotypic data
7. Skills is GS analyses
8. Coordinator for organizing samples, data files, and executing analyses across programs, and possibly within some programs
9. Funds for genotyping

Experience of the Sun Grains Consortium

Evaluated predictive ability of an unbalanced data set of 467 winter wheat genotypes Grown in 49 environments from 2008 to 2016. (*Gulf-Atlantic Nursery*)

34,095 SNP from GBS

Maximum predictabilities were 0.64 for grain yield.

Lyerly, Brown-Guedira, Murphy

Mason

Griffey

Boyles

Sutton

Ibrahim

Harrison

Babar
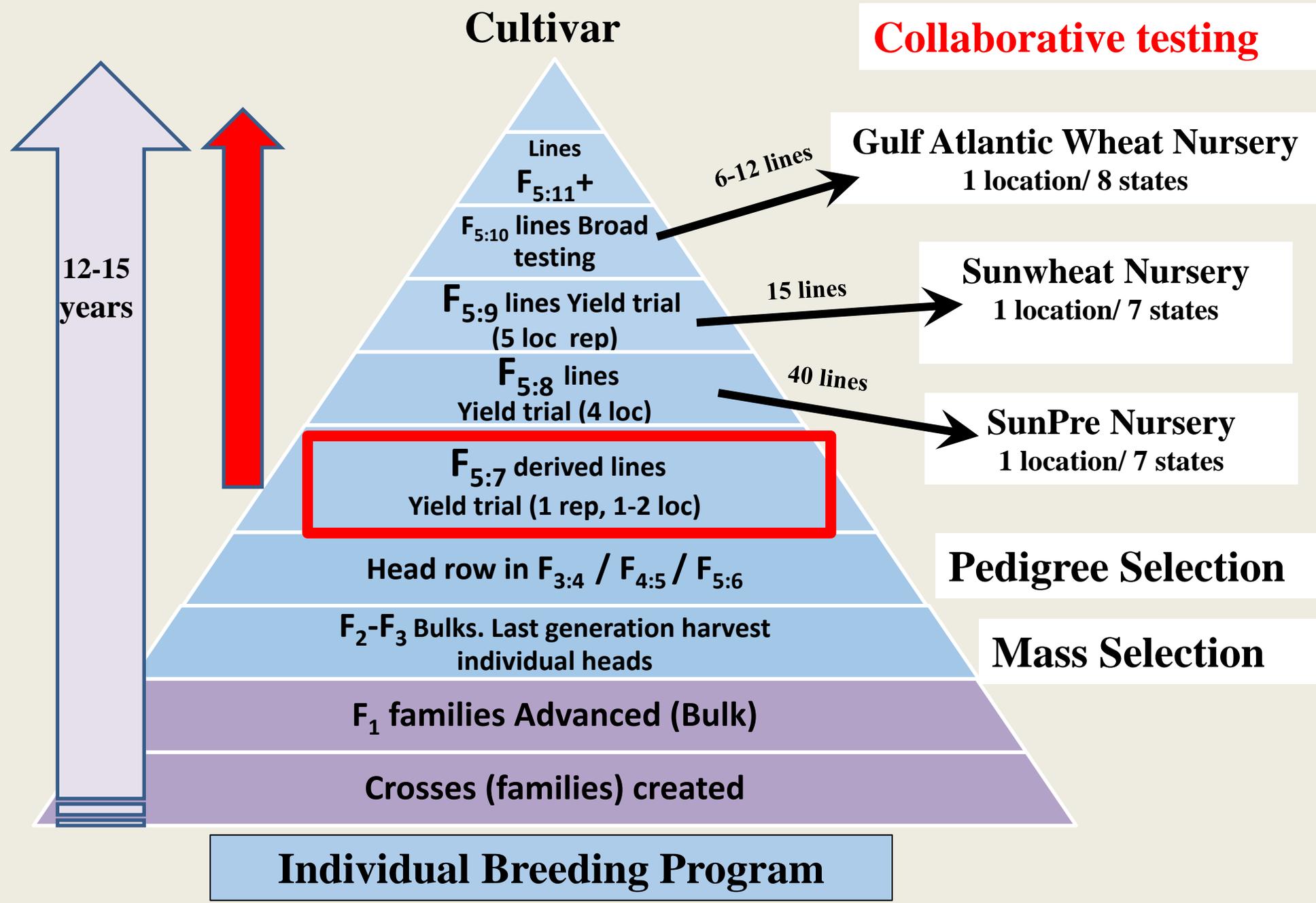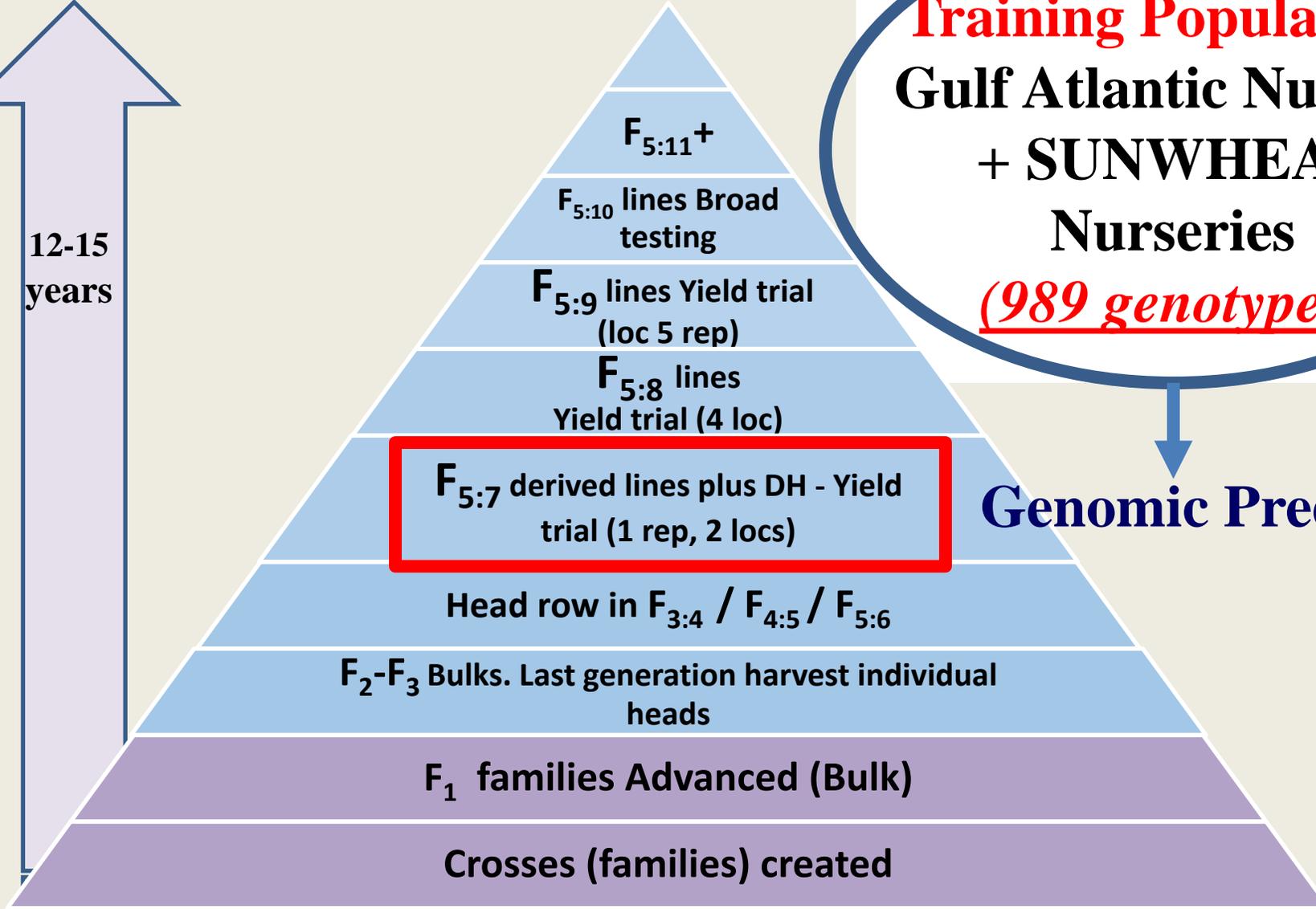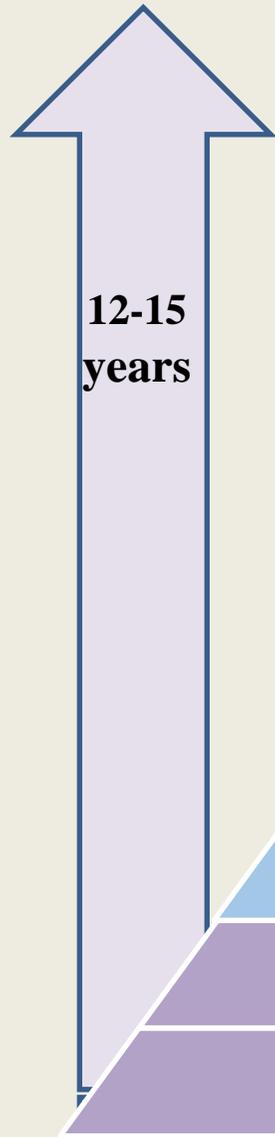
Mergoum, Johnson

GxY=10

Conv
Conv GS
PYT GS
Head GS

Year
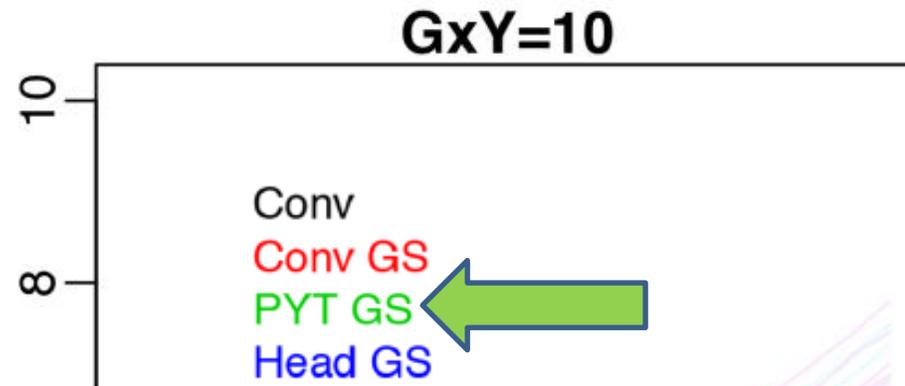
Test-flying GS
Running a conventional breeding program in parallel with GS.
Selections made on phenotypic results not predictions.
But paying attention to the genomic predictions of the lines being advanced.

Gaynor et al. 2016 Crop Sci. 56: 2372-86
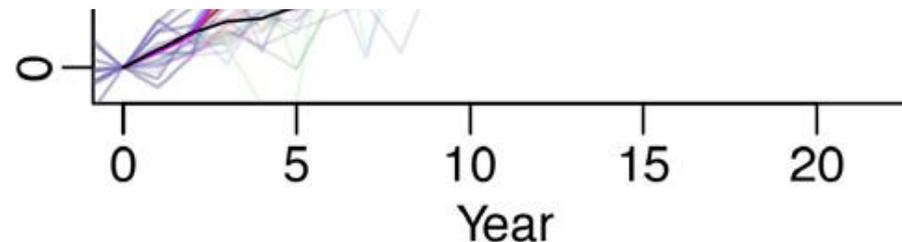
Timeline for SunGrains GS

Aug	Complete project plans; finalize number of samples per breeder; plant seeds
Sept-Oct	Collect tissue; send samples to Raleigh; begin DNA extractions
Nov-Dec	Prepare libraries for sequencing; submit all samples to the NCSU Genomic Sciences Lab; compile phenotype data
Jan-Feb	Biplots analysis; update training populations; receive sequence data; sequence data processing and QC
Mar-Apr	Generate first GEBVs for breeders; GWAS for covariates
May-June	Additional data analysis and testing; report further results to breeders
July-Aug	Receive field data; evaluate field data vs predictions (continuous as new data arrives); run PopVar and send results

GGE Biplot to Assist Training Population Development / Enhancement
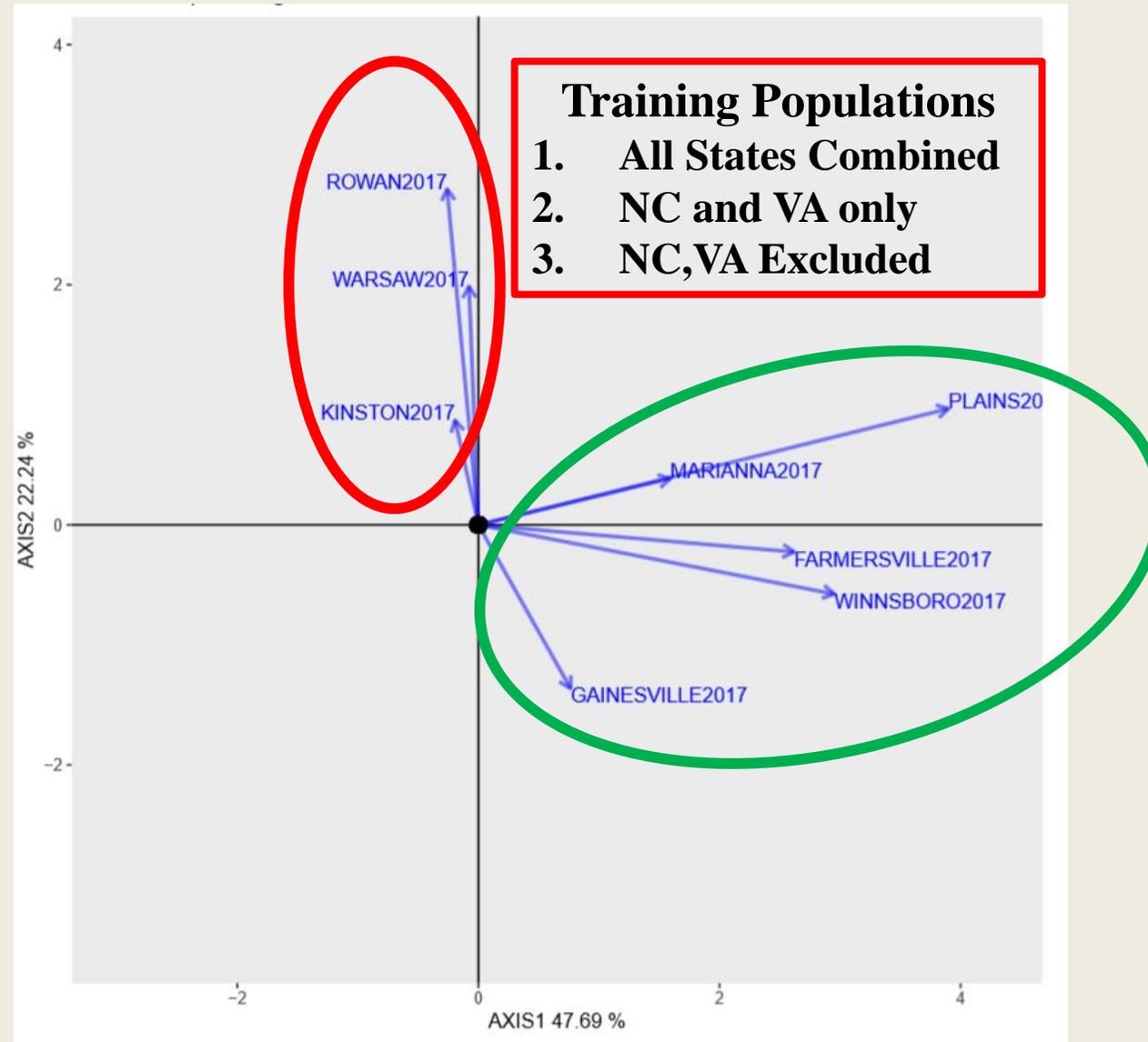
Training populations

	Year 1: 2016-2017	Year 2: 2017-2018	Year 3: 2018-2019	Year 4: 2019-2020
Regional Training Populations + FHB Training Population	3 Regional + 1 FHB	4 Regional + 1 FHB	5 Regional + 1 FHB	5 Regional + 1 FHB
Lines from the GAWN + Sunwheat	623 from years 2008-2016	761 from years 2008-2017	862 from years 2008-2018	989 from years 2008-2019
Lines from the Uniform Southern Scab Nursery	285 from years 2011-2016	247 from years 2011-2017	292 from years 2011-2018	320 lines from 2011-2019

Training population optimization/modification testing

- Add additional regional training populations
- Use SNPs associated with major genes as covariates:
  - VrnA1, H13, Yr17, Lr18, etc.
- Adjust training population composition:
  - Use algorithm to select genotypes for training based on a list of genotypes you want to predict
- Adjust training population size:
  - Include 300, 350, 400, or 500 genotypes
- Add selections to training data:
  - Use NCVA as a test case
- Select SNPs via GWAS to use in prediction:
  - Significance thresholds 0.01, 0.05, 0.10. 0.15

HF Percent Infested
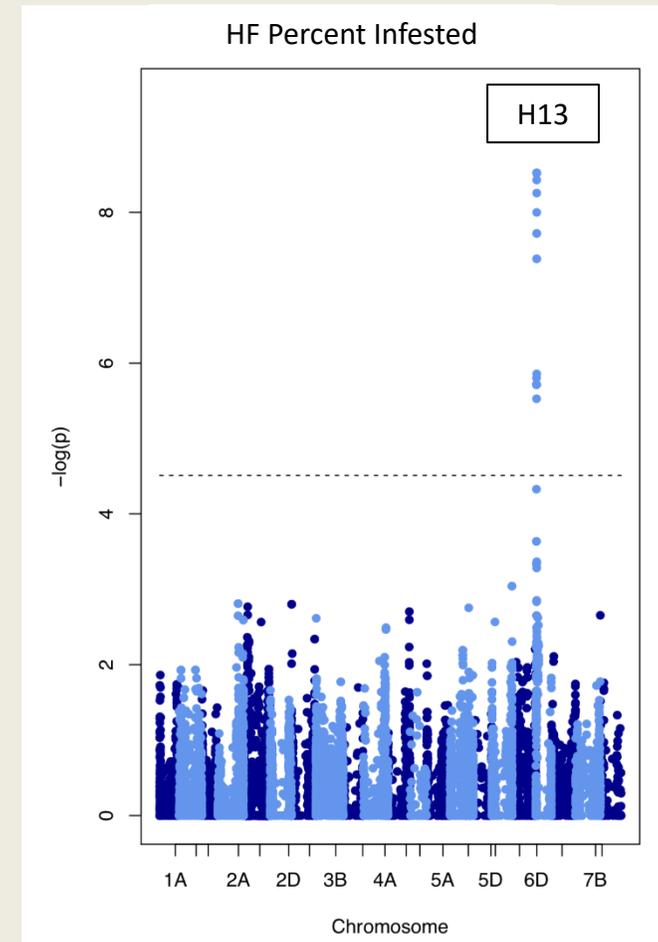
Sunwheat: 2020 Correlation results for yield

Correlations between observed and predicted values for Sunwheat 2020 for yield for each regional training population

Test Location	Trait: Yield Training Population				
	Comb	GA	LATX	NCVA	noNCVA
AR	0.24	0.33	0.34	-0.10	0.38
GA	0.32	0.37	0.21	0.17	0.32
NC	-0.03	-0.18	-0.08	0.31	-0.11
LA	-0.15	-0.18	-0.15	-0.03	-0.10
SC	0.36	0.36	0.29	0.16	0.39
TX	0.09	0.11	0.07	0.04	0.07
Mean over Locs	0.29	0.29	0.25	0.15	0.32
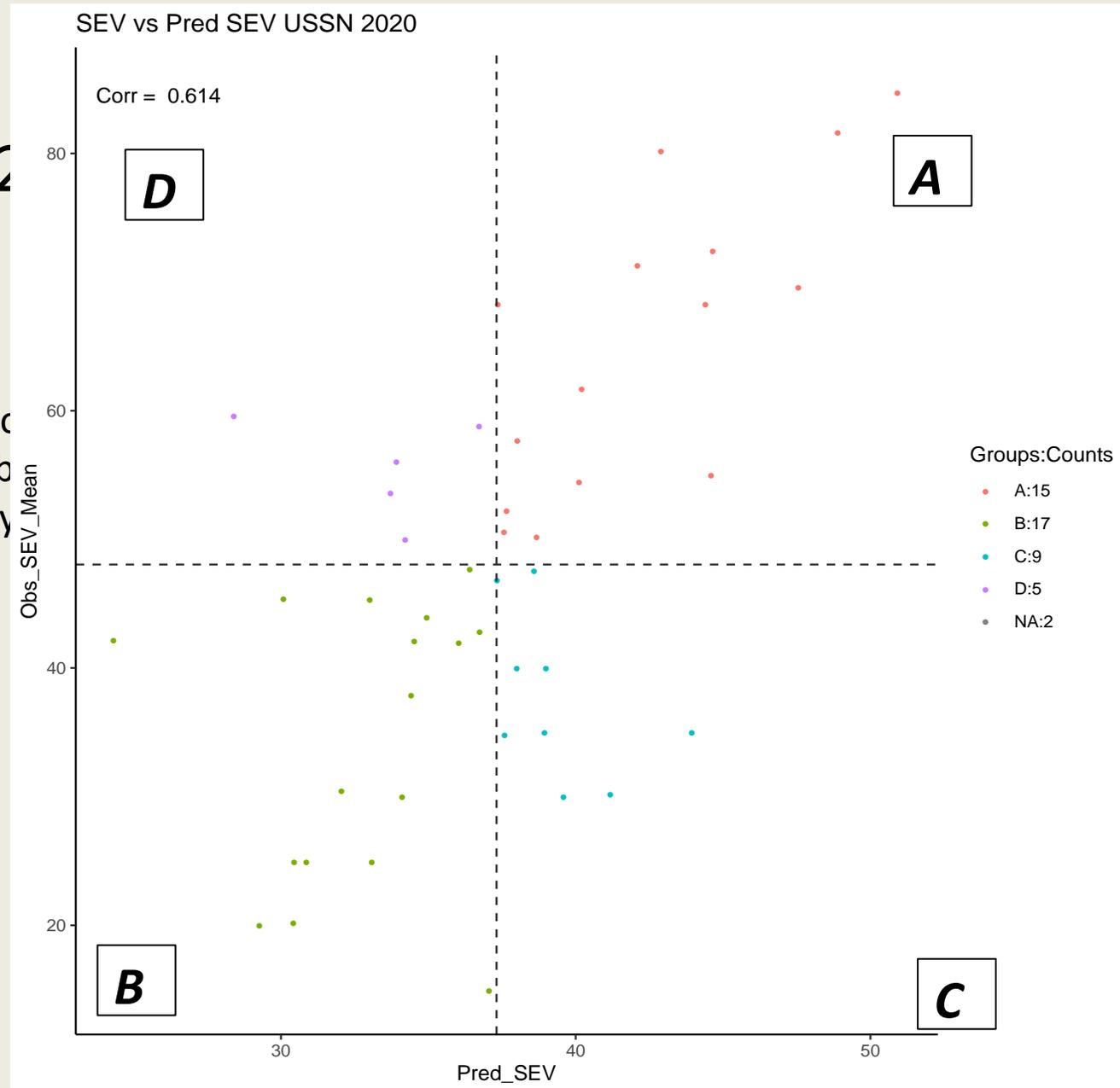
NCSU Wheat Advanced Test WAT Entries 2021
Accuracies of Yield Predictions?

Column 1	Column 2
NC14-20369	NC18-17941
NC15-21834	NC18-17944
NC15-21835	NC18-17619
NC11546-14	NC13804-A113
NC16-19288	NC13804-A170
NC11363-25	NC13955-G8
NC16-21185	NC13955-G28
NC12642-81	NC13955-G84
NC12093-10	NC13955-G91
NC12164-97T	NC13955-G92
NC12164-200T	NC13955-G114
NC13211-9	NC13955-G125
NC13202-128	NC13955-G135
NC13207-16	NC13955-G151
NC13220-37	NC13955-G183
NC15V25-20	NC13955-G200
NC15V26-19	NC13810-M29
NC15V41-13	NC13906-W5
NC16-19349	NC13906-W10
NC13206-40	NC13906-W12
NC18-16900	NC13906-W39
NC18-16901	NC13217-W293
NC18-16913	NC13217-W2111
NC18-16920	NC13220-Z2163
NC18-17936	NCVT.D-33

50 Entries

Four in program before GS *(Clear)*

Three were sequencing failures *(Blue)*

Of the remaining 43:

37 in top 50% of Yield Predictions *(Yellow)*

6 in bottom 50% of Yield Predictions *(Red)*

86 percent success rate.

How do Sungrains breeders use GEBVs?

- Selection when phenotypic data is limited – PYT (FHB, SR)
- Aid during in season field evaluation and note taking
- Eliminating low performing lines - bottom 50%
- Selecting regionally and/or locally adapted lines
- Selection in sub-optimal environments
- Fast-track parents for crossing or designer crosses
- Selection for non-target traits

NWWCP & NIFA Consortiums
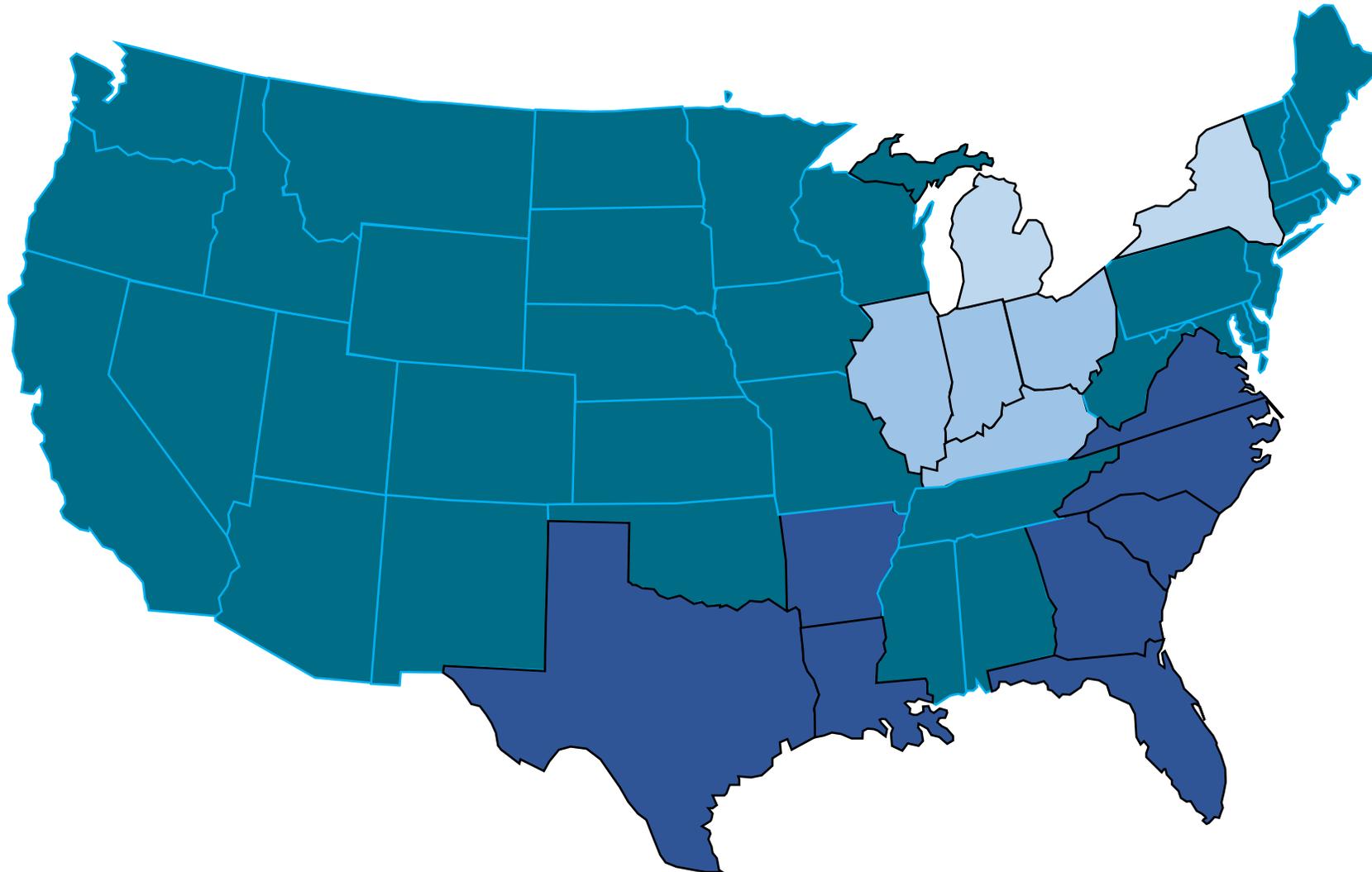
NWW-CP
MI, NY, IL, IN, OH, KY

NIFA
IL, IN, OH, KY

Requirements for Consortium Success

1. Germplasm among the member programs must be related
2. Germplasm from each member offers value to the other members
3. Development of optimal breeding and testing schemes

1. Relatedness of 10,246 Lines from IN, IL, KY, OH
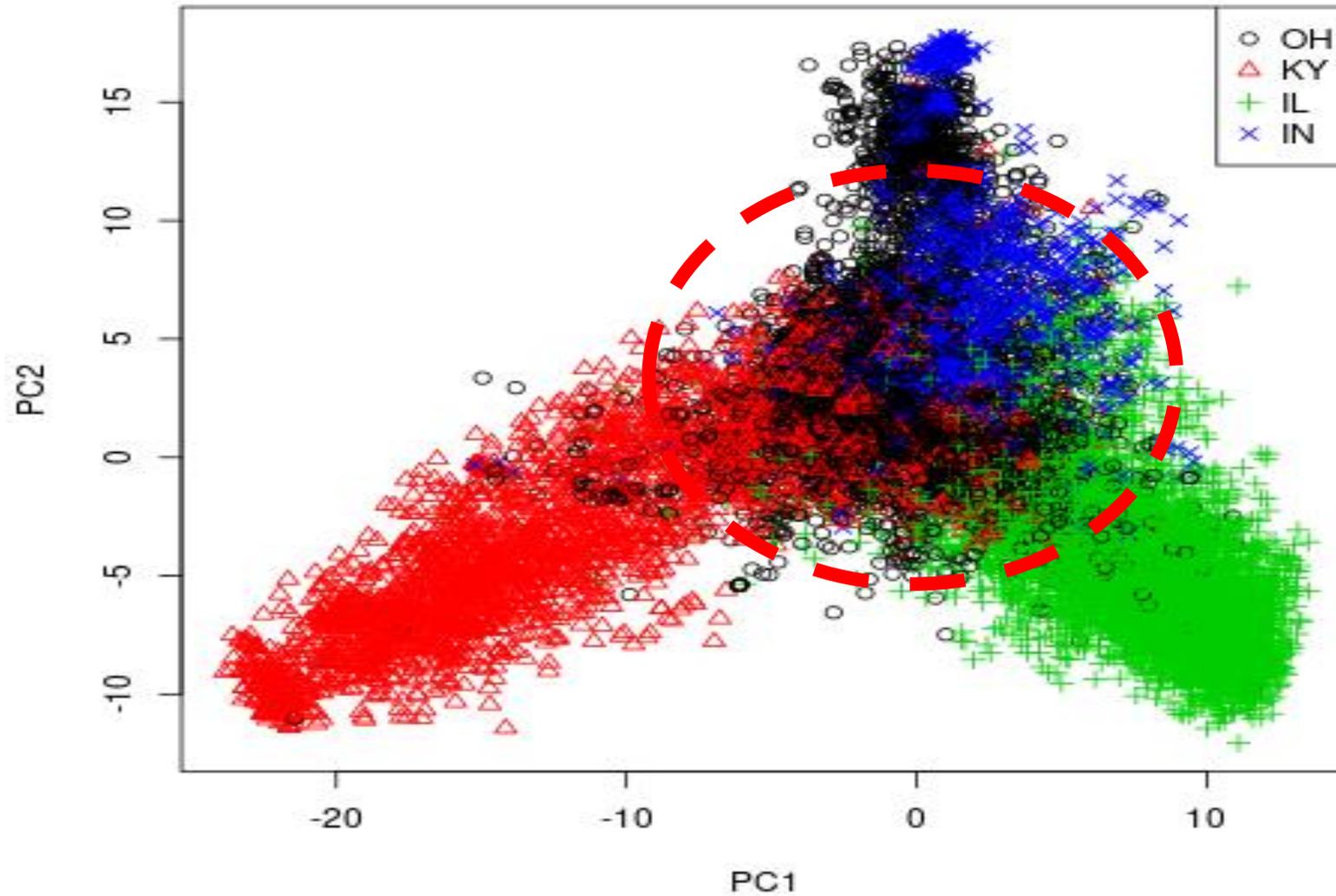
2. Relevance of germplasm between states: Analysis of performance from 5-state trials, 2012-2018

Comparison Type	Testing Location	Origin of Lines	Yield (bu/ac)	GEI (bu/ac)
WITHIN	IL	IL	**82.4**	2.0
BETWEEN	IL	IN	82.2	1.4
BETWEEN	IL	KY	76.9	-1.3
BETWEEN	IL	OH	78.0	-4.1
WITHIN	IN	IN	86.8	-0.5
BETWEEN	IN	IL	**92.8**	5.3
BETWEEN	IN	KY	83.6	-1.4
BETWEEN	IN	OH	87.0	-1.9
WITHIN	KY	KY	73.3	1.6
BETWEEN	KY	IL	69.2	-3.7
BETWEEN	KY	IN	73.9	0.3
BETWEEN	KY	OH	**76.7**	1.8
WITHIN	OH	OH	**71.9**	0.7
BETWEEN	OH	IL	70.1	0.6
BETWEEN	OH	IN	67.8	-1.9
BETWEEN	OH	KY	68.4	1.0

Average GEI of local germplasm within a state = 0.7 bu/ac

Average GEI between states = -0.3 bu/ac

Average superiority of local germplasm within a state = 1.4 bu/ac

GS Accuracy

	Yield	Test Weight	FHB (Trait)
IN	0.44	0.33	0.40 (DON)
IL	0.45		0.58 (DON)
KY	0.51	0.63	
OH	0.63	0.45	0.53 (Index)

Cross-validation accuracy within a program

		Predict Yield in this State			
		OH	IN	IL	KY
Source of Yield Data used in TP	ON		-0.12	0.18	0.11
	IN	0.04		0.01	0.07
	IL	0.10	-0.10		0.17
	KY	0.08	0.01	0.15	

Accuracy between programs

Testing Schemes

Within Ohio, Selection is based on:
OH phenotypic data
Ohio-based GEBVs

	Average # Lines per Program	Total # Lines
Stage-1	950	3800
Stage-2	240	960
Stage-3	35	140
Stage-4	15	60
Total	1240	4960
Coopertive testing	11	45

95% of lines

Greatest Diversity

< 1%, Only "Elite"

3. Proposed Sparse Testing Schemes for Stage-1

Scenario 1	
	Current Testing of Ohio Stage-1 Lines
Test Location	OH Only
Cross 1	9
Cross 2	9
Cross 3	9
Cross 4	9
Cross 5	9
Total	45

Scenario 2			
Sparse Testing of Ohio Stage 1 Lines			
OH	IL	IN	KY
3	2	2	2
3	2	2	2
3	2	2	2
3	2	2	2
3	2	2	2
15	10	10	10

Scenario 3	
Ohio Testing of Stage-1 Lines	
OH Lines	15
IN Lines	10
IL Lines	10
KY Lines	10
Total	45

Stage-1 in Ohio
33% OH Lines

| OH | IL | IN | KY |

Stage-1 in Illinois
33% IL Lines

| IL | OH | IN | KY |

Stage-1 in Indiana
33% IN Lines

| IN | OH | IL | KY |

Stage-1 in Kentucky
33% KY Lines

| KY | OH | IL | IN |
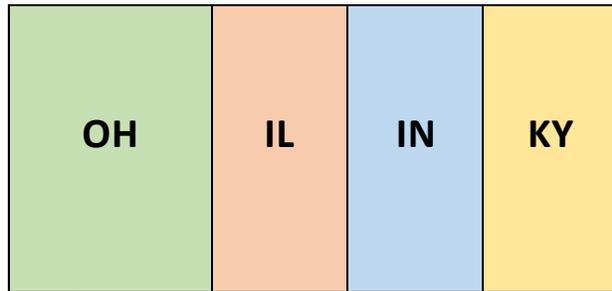
New Stage-1 Plan:

**OSU selects among 900 OH-Tested lines using**

1) **OH phenotypic data**
2) **Ohio-based GEBVs**
3) **GEBVs from 3 other environments & overall envs**
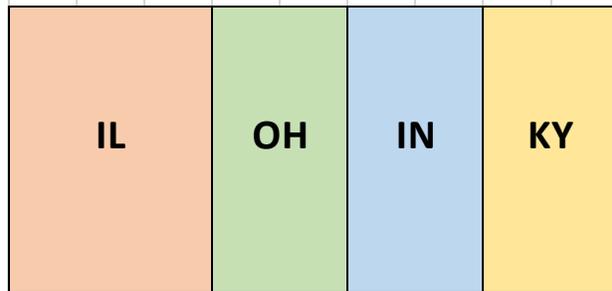
OSU also selects among 2700 other lines using

1) **Predicted value in OH**
2) **Observed values from their testing location**
3) **GEBV from each environment & overall envs**
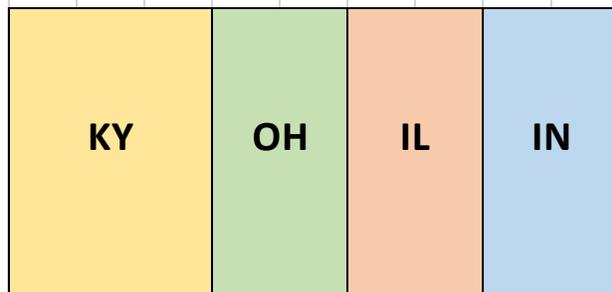
Each program can access 3600 stage-1 lines
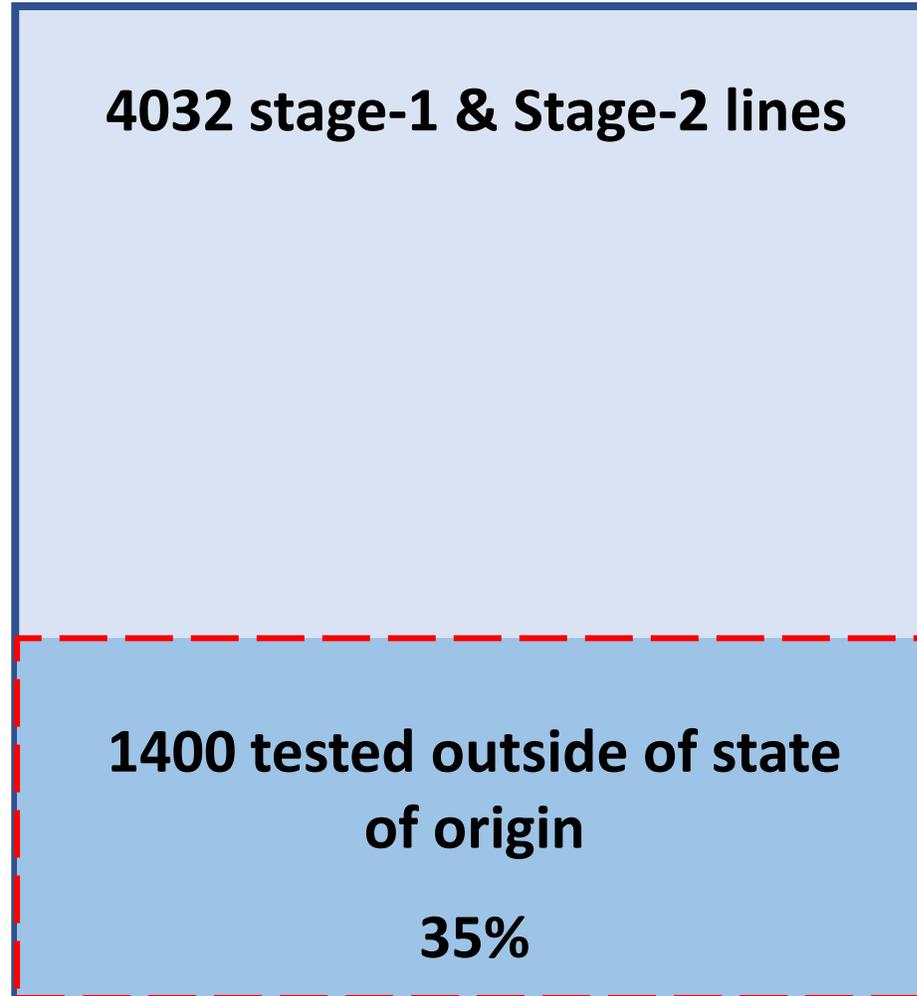
Benefits to Genotyping Lab of working with Consortiums

Brian Ward
Coordinator USWBSI Northern GS

- Coordinator provides point person

- Consortium projects streamline workflow

- Large sample numbers can leverage better pricing

- Mine database of genotypic data and connected to phenotypes

- Collection of DNA samples

Jeanette Lyerly
Coordinator SUNGrains GS

Genotyping pipeline

Project Planning
Yellow = ERSGGL
Blue = Breeding program
Green= Consortium coordinator

Plan project / Sample files / Provide supplies

Plan project / Sample files / Collect & Ship tissue

Plan project / Timeline / Sample names

DNA isolation / Quant & Normalize / Library prep_ID / Assign barcodes

Library preparation / QC

Sequencing lab

Download FastQ files / FastQC / Demultiplex / Create Key Files

Unified dataset / VCF files / Filtering / Imputation

SNP calling: / All Training Pops / Selection Candidates

GEBV

Consortium / Phenotypic data

Individual program / Phenotypic data

Local storage / Compressed FastQ / Key files

Off-site storage / Compressed FastQ / Key files

Development of targeted sequencing platform

- Research has shown that smaller numbers of SNP can be used for Genomic Selection

- Would like to have consistent data sets across germplasm

- Target genes, QTL regions and genome wide markers with a same technology

- Sufficient read depth to identify heterozygotes or copy number variants

- Simplify bioinformatics pipeline

- Empower labs to work independently
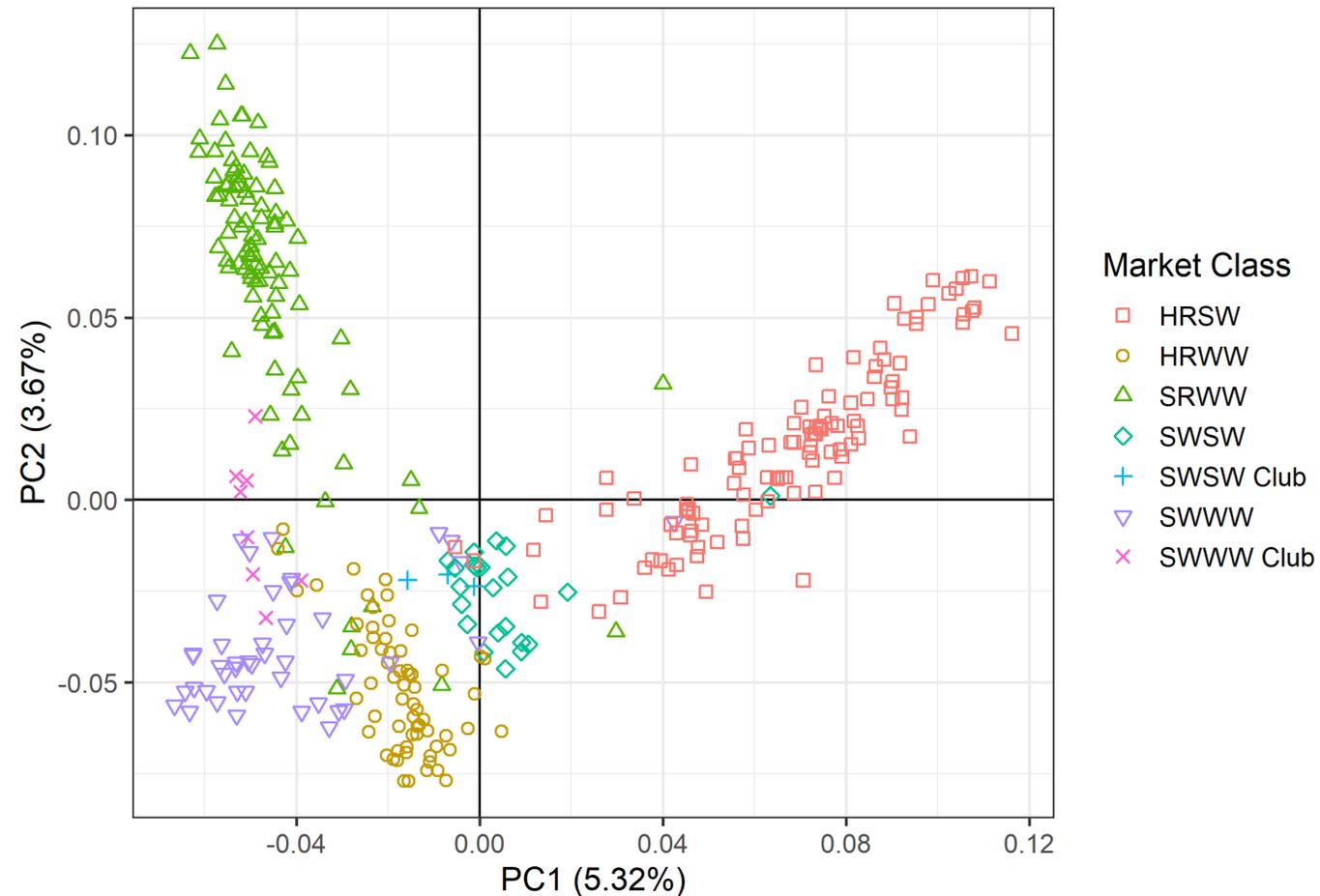
Diversity of North American Wheat based on Exome Capture of ~400 Cultivars
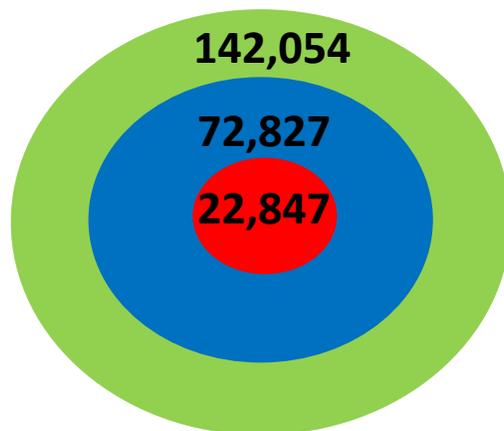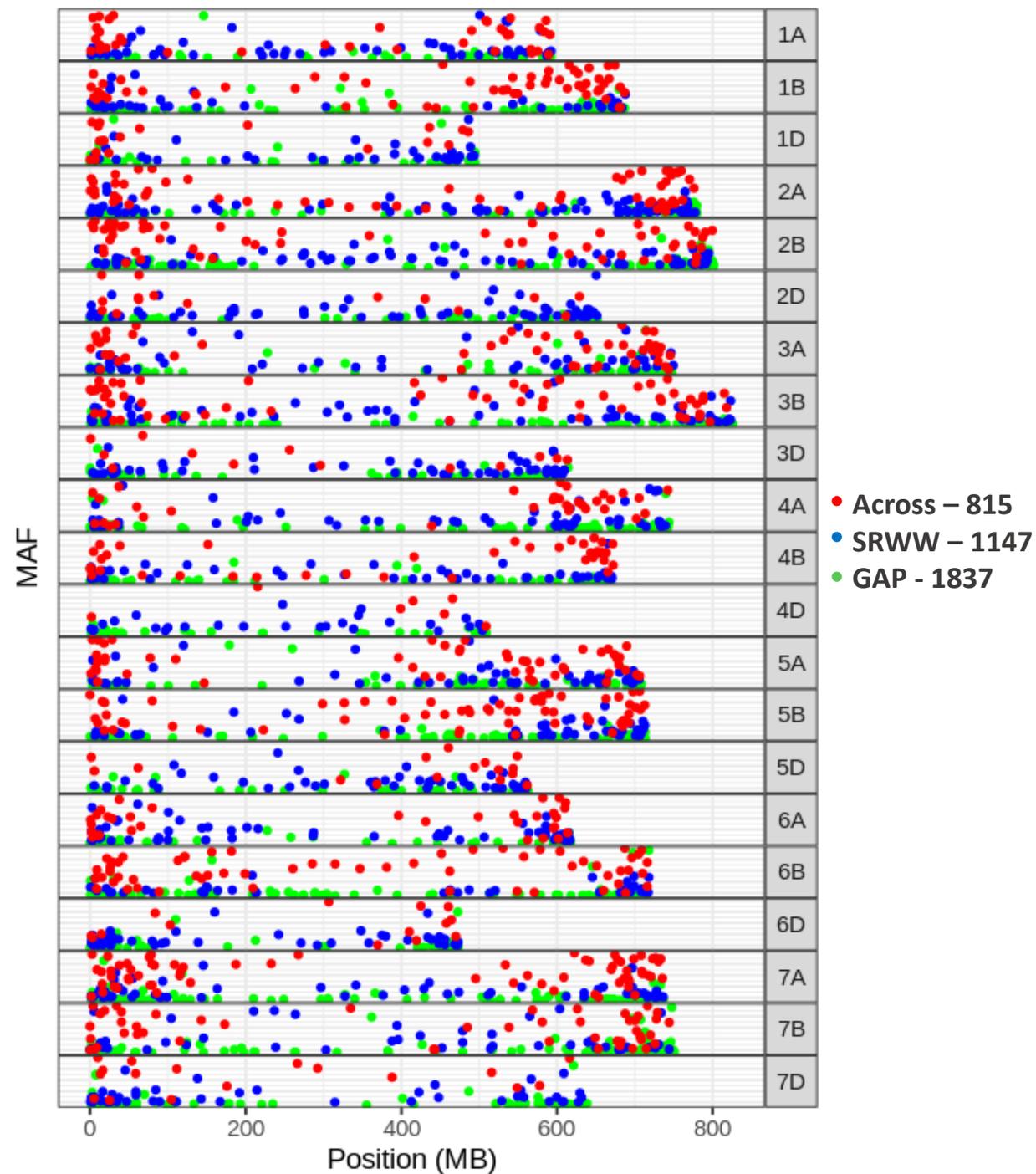
Development of targeted sequencing platform

Data set can be used to select SNP for any targeted genotyping technology
Data can be used to select SNP for different germplasm, breeding programs

Sticking point:
Process for development and design of a target technology is time and labor-intensive or expensive
Tied to large number of samples for approach/design that may not be successful

Need flexibility of iterative design process:
           Add new SNP associated with traits over time
           Replace monomorphic, failed or overabundant targets
           Update based on new information about biology

Allegro Targeted Genotyping – NuGen (Tecan)
           Single Primer Enrichment Technology
           No upfront cost of design process
           Not tied to large numbers of samples
           Disadvantage is cost (~$13-15/sample)

Many Thanks to ……