Genotyping to Facilitate Genomic Selection

Wheat CAP Kick-off

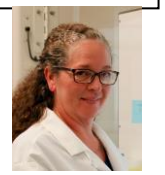
Dec 15, 2022

Small Grains Genotyping

Deven See

Jason Fiedler

Guihua Bai



Gina Brown-Guedira

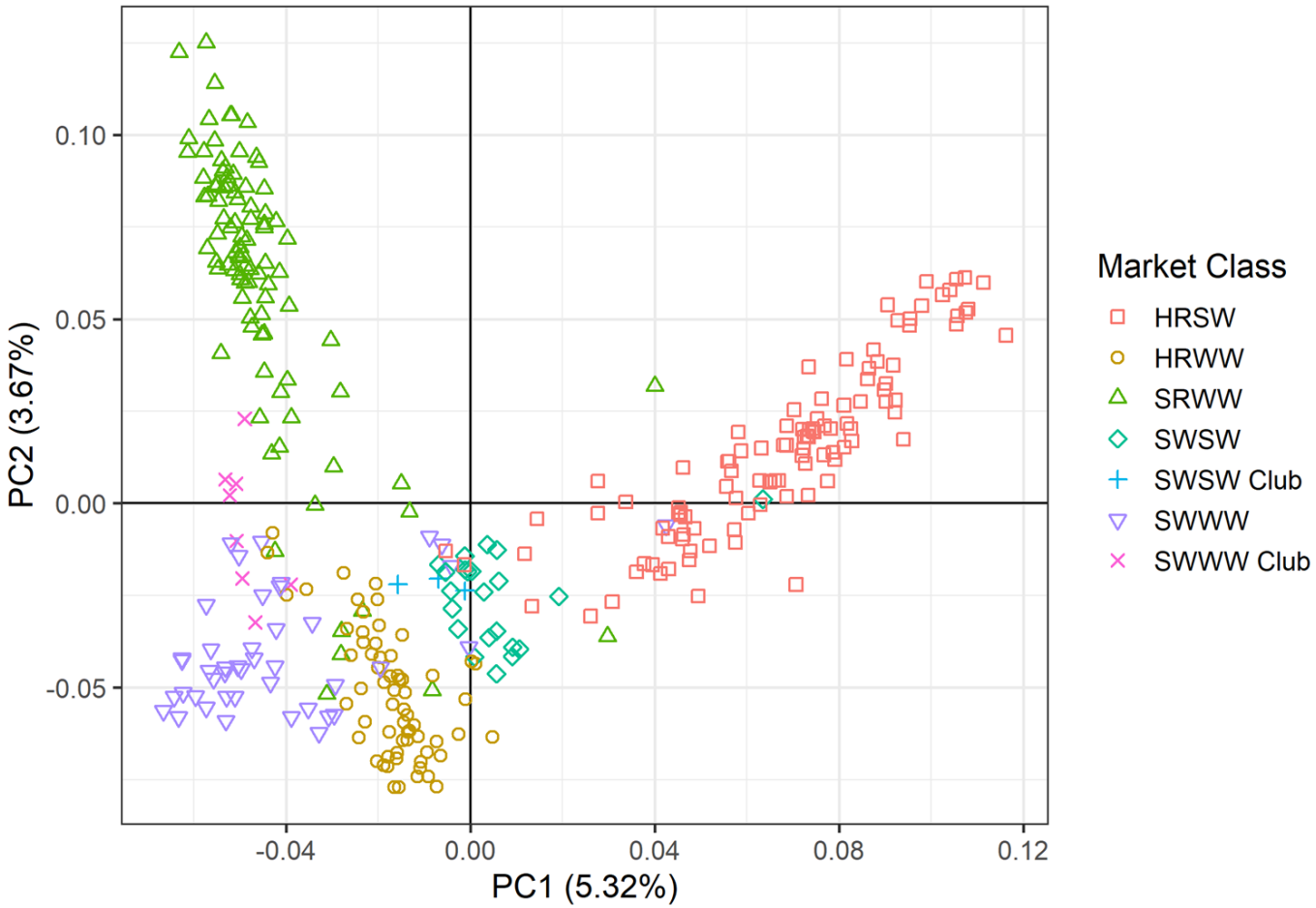
USDA AGRICULTURAL RESEARCH SERVICE

Development of targeted genotyping platforms

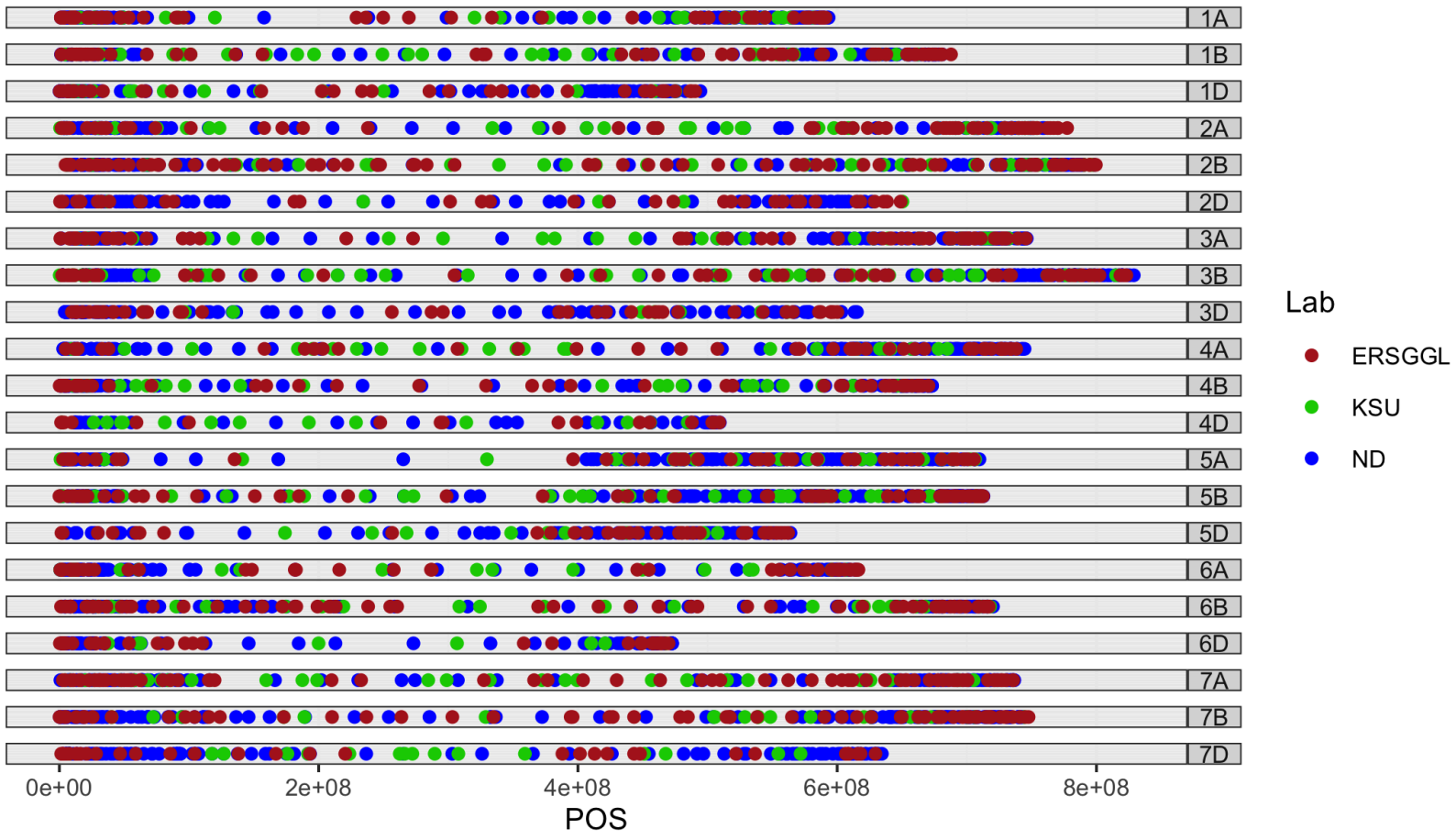
- Most genomic selection based on GBS – IP issues
- Research has shown that smaller numbers of SNP can be used for Genomic Selection
- Would like to have consistent data sets across germplasm
- Target genes, QTL regions and genome wide markers with a same technology
- Sufficient read depth to identify heterozygotes or copy number variants
- Simplify bioinformatics pipeline
- Simplify data storage

Collaborative genotyping of North American Germplasm

- Exome capture sequencing
- All U.S. wheat breeding regions for ~420 accessions total
- Genotypes suggested to genotyping labs by breeders
- Variants in gene space feed into Practical Haplotype Graph
- Variants selected for targeted genotyping



Physical Distribution of SNP targets



Criteria: Genome distribution and minor allele frequency

Illumina platform

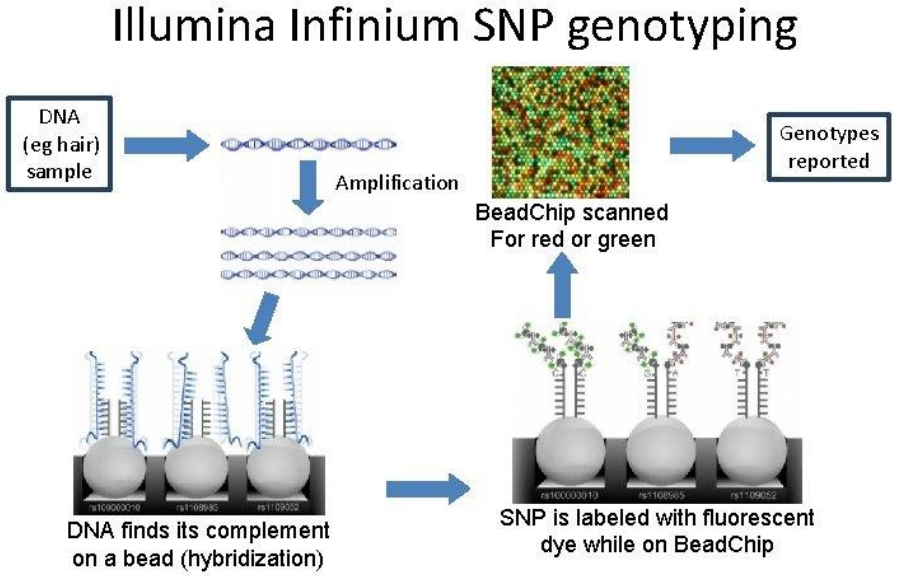
"Work horse" of the previous CAP projects
Fargo lab has equipment and experience required for high-throughput

Multi-species SNP array – wheat, barley, oat, soybean
3,000 SNP each
$14/sample
Reduces costs with multi-species simultaneously

Commitment of 25,000 samples

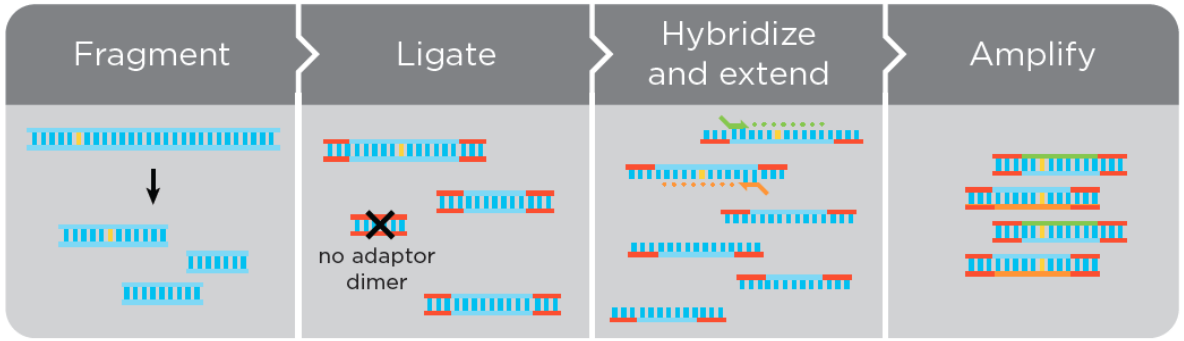
Low missing data
Consistent data sets
Reliable calling of heterozygotes

Significant up-front costs – equipment and sample commitment
Locked into large numbers of fixed arrays



Single Primer Enrichment Technology (SPET)

Annals of Botany, 2019. **124**: 543-551



Tested in NC lab

Reagents marketed by NuGen (Tecan) as Allegro
Technology behind LGC SeqSNP service

Can target up to 100K probes
5K probes - $9.22, 10K probes $10.71
Using 0.4 reaction volume – cuts reagent cost in half $4.70 to $5.36/ sample
Best done with automated liquid handling
Sequencing cost depends on multiplex level and coverage
1536-plex with 100x coverage ~$2/sample
Total cost range $8-12

Low cost of entry – as few as 192 samples
Iterative assay design process

Molecular Inversion Probes (MIPs)

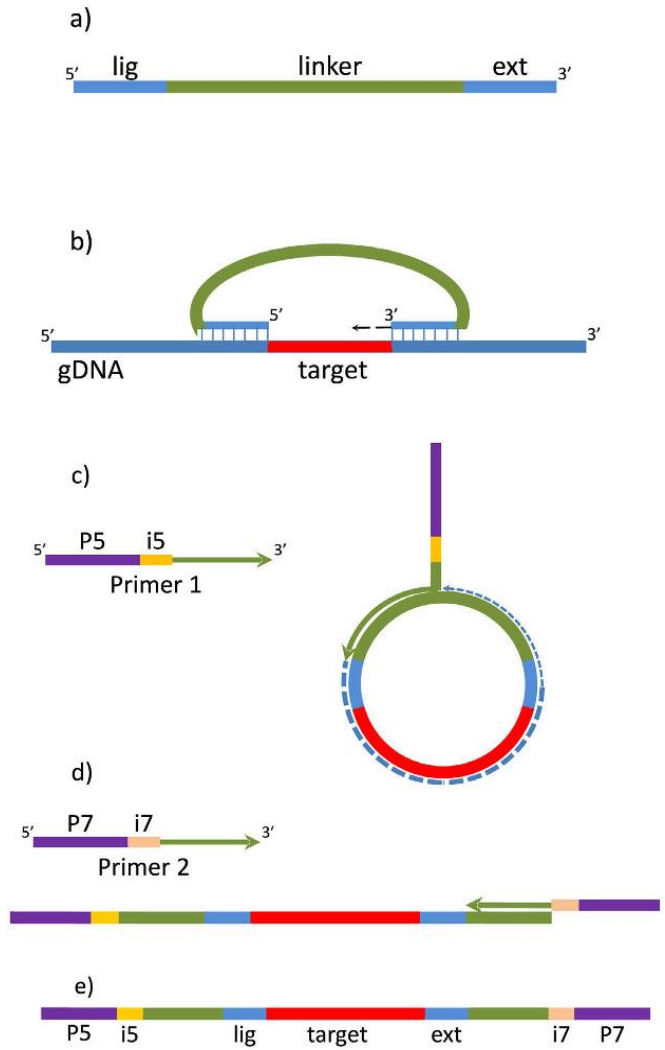
Scientific Reports 6:24051 DOI: 10.1038/srep24051

Akhunov lab is testing

Recovered 287 SNPs from a set of 363 SNPs (79%)
Estimated $10/sample using company reagents
Could significantly reduce costs with optimization of in-house reagents

DArTAG - MIPs based platfom

- Diversity Array Technology handles DNA extraction, library prep, sequencing - $11/sample
- Second round of design for CIMMYT, ~2000 quality variants (~50%), Good GS accuracy
- Some difficulty in agreement with calls for gene/QTL markers



Genotyping by Multiplex Sequencing (GMS)

PLoS ONE 15(5): e0229207. https://doi.org/10.1371/journal.pone.0229207

Developed at WA Genotyping lab

Homemade multiplexed assay

2,100 targets amplified in 8 PCR pools
SNP calling pipeline based on alignment to targets
Targets selected from 90K SNP
Based on genome distribution and MAF in mostly PNW material
Includes ~40 markers for agronomic, quality and disease resistance loci

Estimated cost is $10/sample

Multiple Restriction Amplicon Sequencing (MRASeq)

Plant Biotechnology Journal, 2020. **18**: 254-265

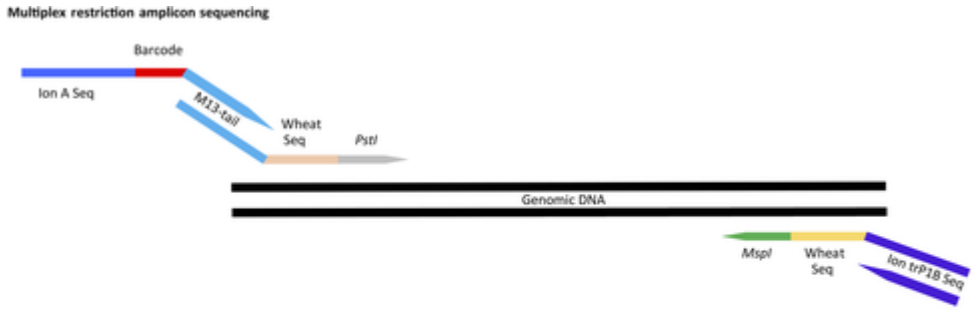
Developed at Kansas Genotyping Lab
Two-Step PCR-based protocol that targets same sites as GBS – avoids IP issues associated with GBS

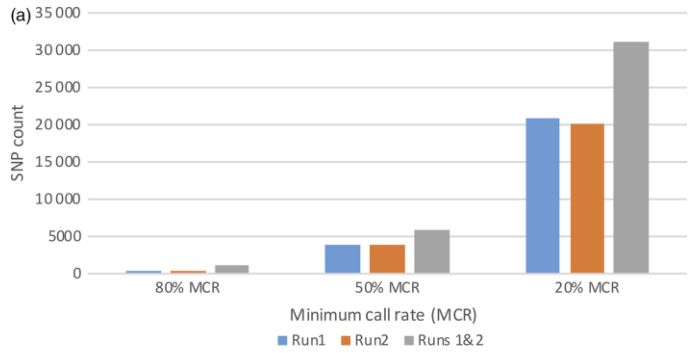
Similar data as GBS –
In 1000 genotype panel, recovered ~10,000 SNP missing data 40%, MAF >0.03
Some issues with excessive heterozygous calls

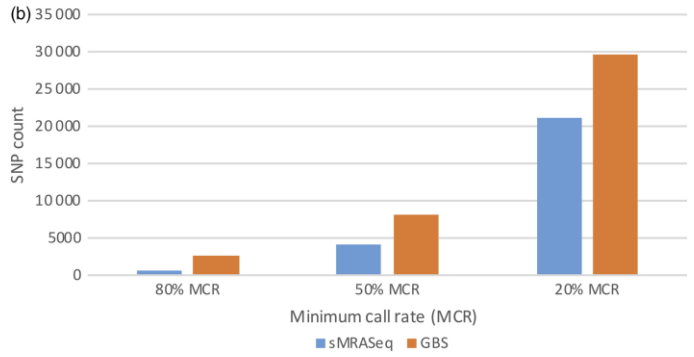
~ $5/sample

Straightforward workflow
Low cost

Missing data requires imputation
Consistency of data across runs
Based on low read depths
Can not reliably score heterozygotes







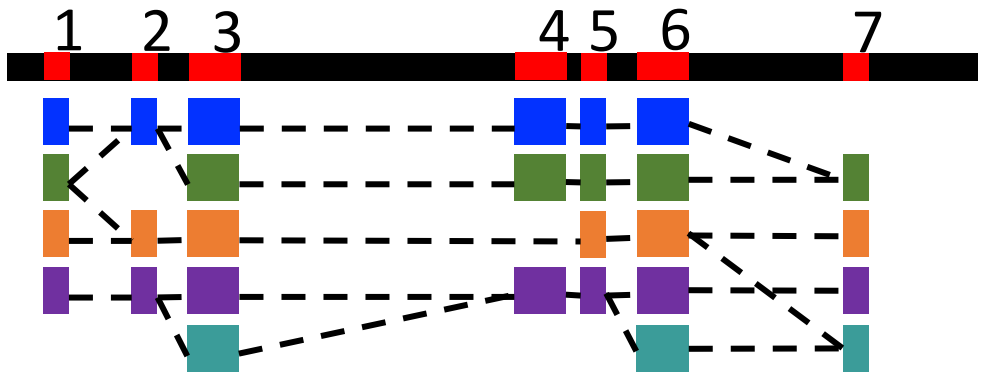
Practical Haplotype Graph - putting it all together

Variants in Ref set
Exome capture
GBS
90K SNP

Broad or regional sets

Training sets
GBS
90K SNP

Practical haplotype graph



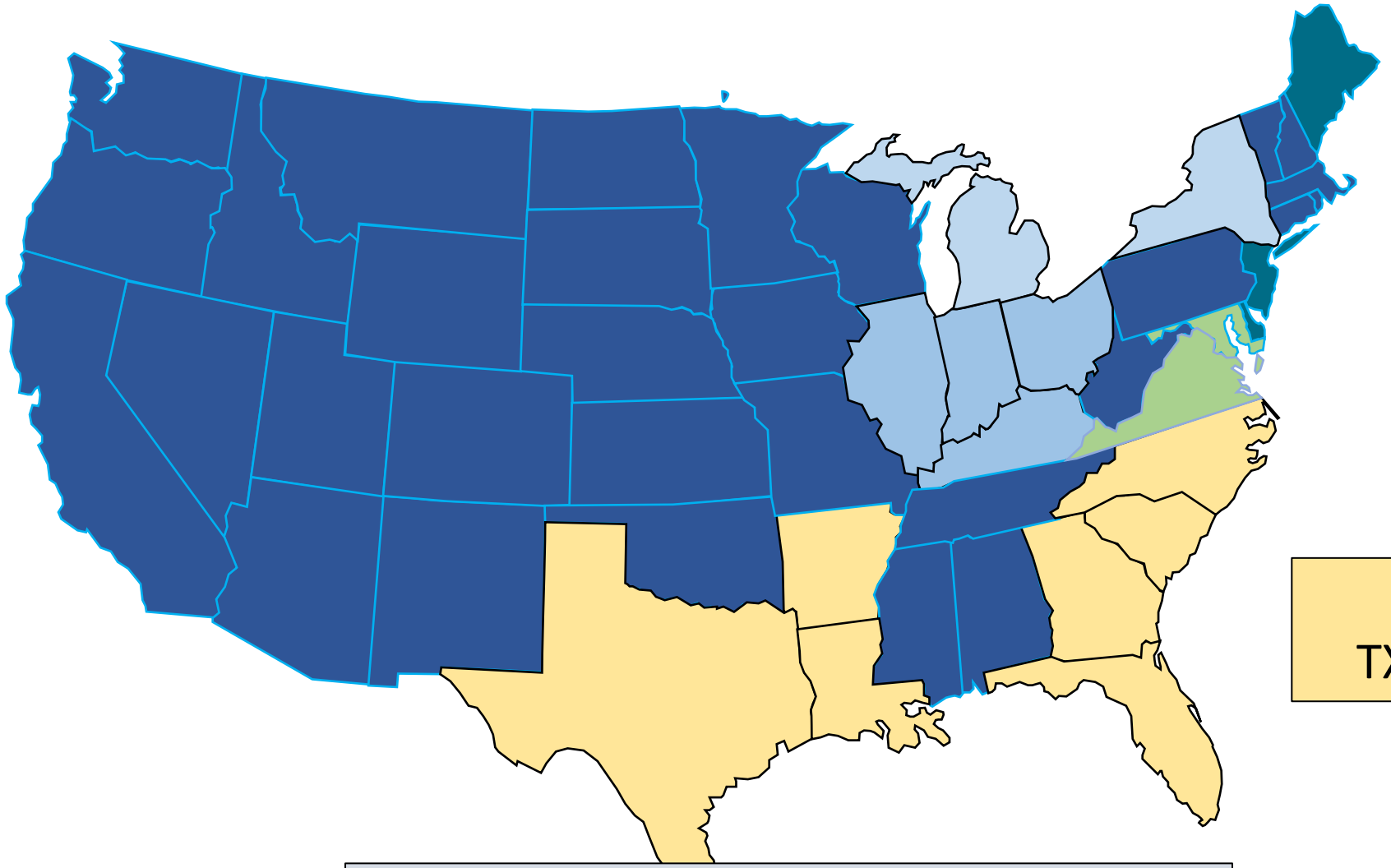
Imputation

Unified data
Target SNP
GBS
90K SNP

Training and selection sets

Selection candidates
Mid-density targets

Lessons from the East



NorGrains
MI, NY, IL, IN, OH, KY

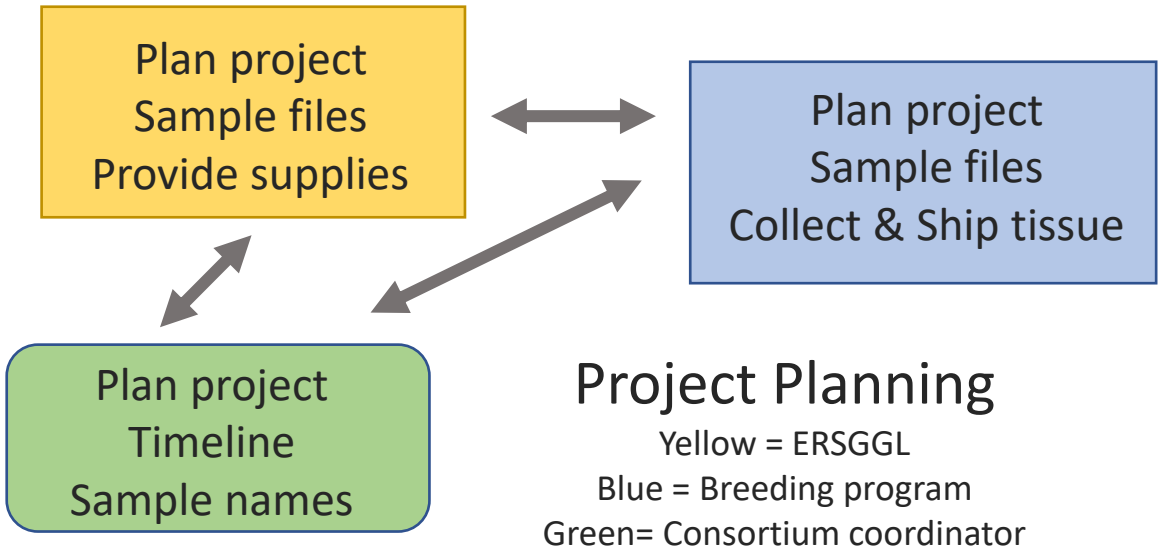
Misfits: VA, MD

SUNGrains
TX, LA, AR, GA, FL, SC, NC

Soft Winter Wheat “Consortiums”

GS with centralized genotyping,
data analysis and collaborative testing

Communication is Key



Groups have regular meetings
small # of people in the room
working meetings – decisions get made

Result : Increased diversity of **germplasm and ideas**

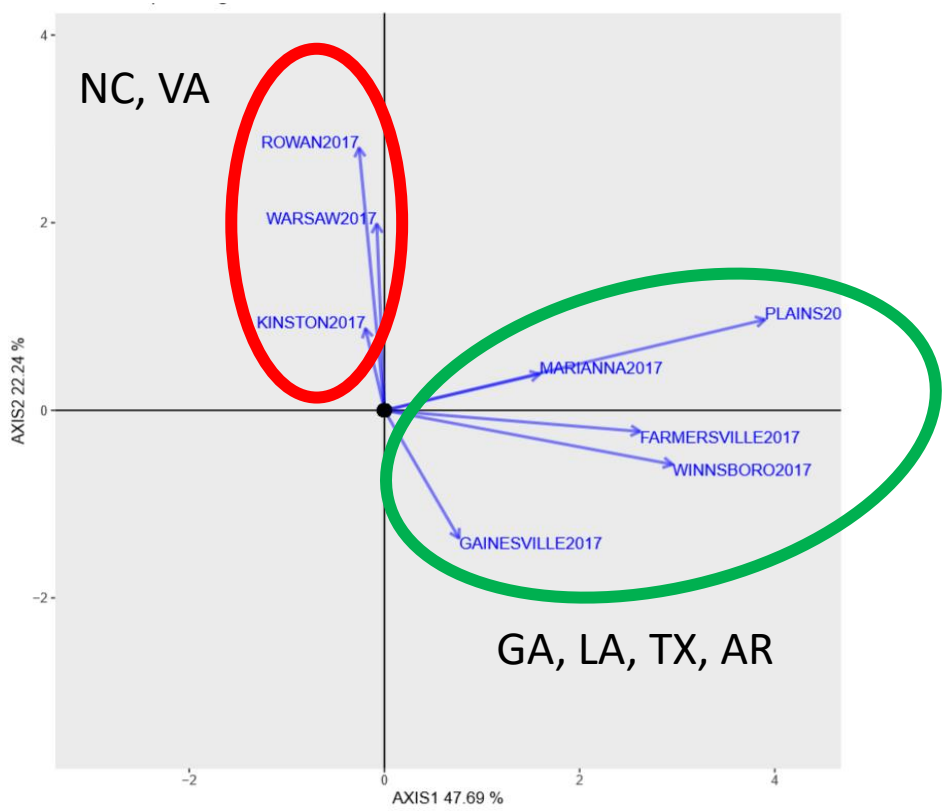
Observations:

Learned about testing environments –
poor accuracy when comparing multi-environment data if environments are not well correlated.
Selection of individual lines with specific/regional adaptation.

GEBV increases confidence to enter parents into crossing earlier.

Know your parents –
Data for trait related markers play role in parent selection.

Moving toward application of GS to replace stage 1 testing.



Challenges of simultaneous development and deployment