A multi-species, low-cost, genome-wide genotyping platform to support molecular breeding in small grains

Jason Fiedler
Research Plant Geneticist
USDA-ARS Small Grains
Genotyping Laboratory
Fargo, ND

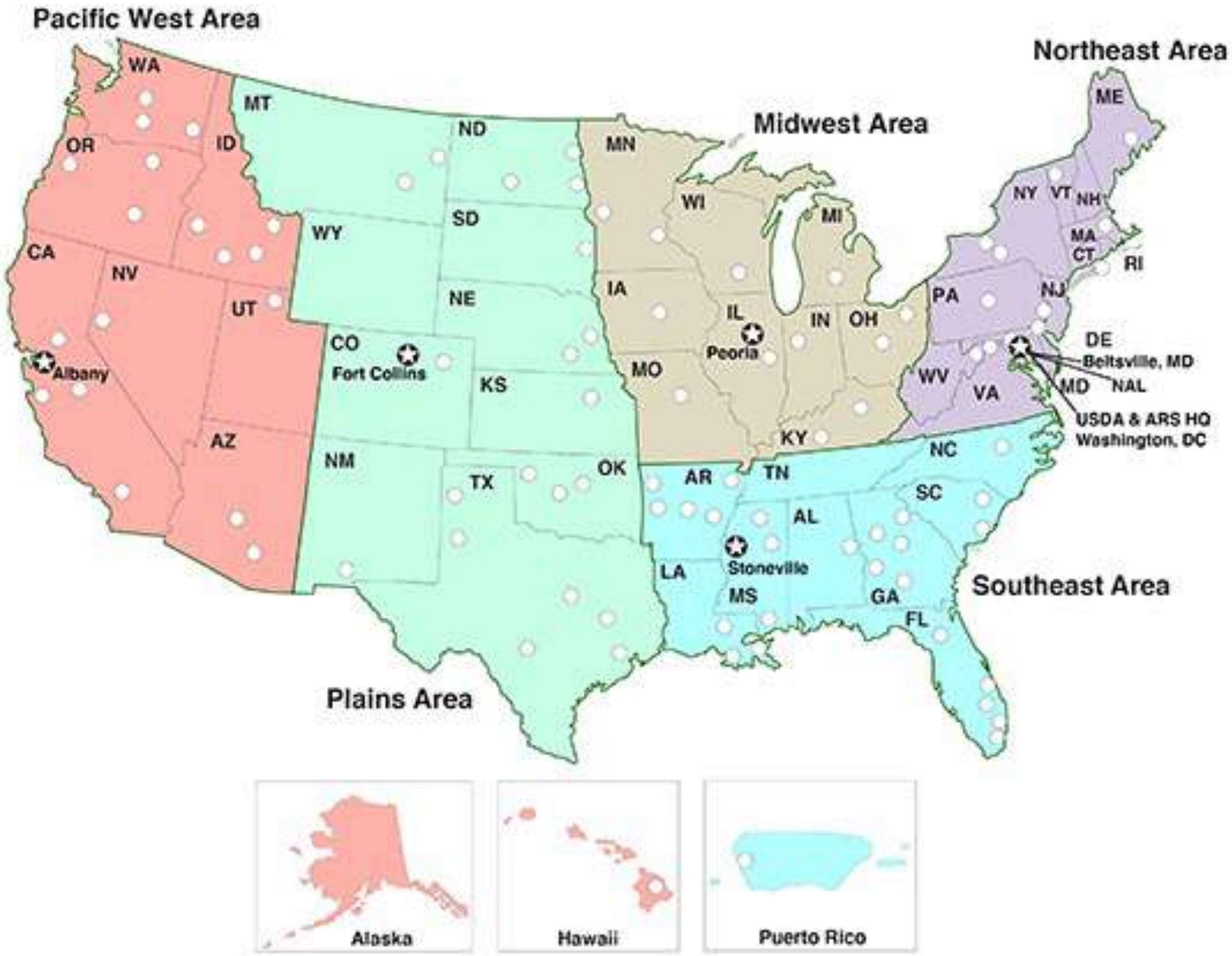
IWGSC Webinar
January 26th, 2023

The USDA-Agriculture Research Service (ARS)

- The in-house research agency within the US Department of Agriculture.

- ~ 7,400 employees
- > 90 locations
- > 1 Billion $US budget

- Use science and technology to tackle agricultural problems.



Problem: breeding is hard!

Yield?

Pest resistance?

Disease resistance?

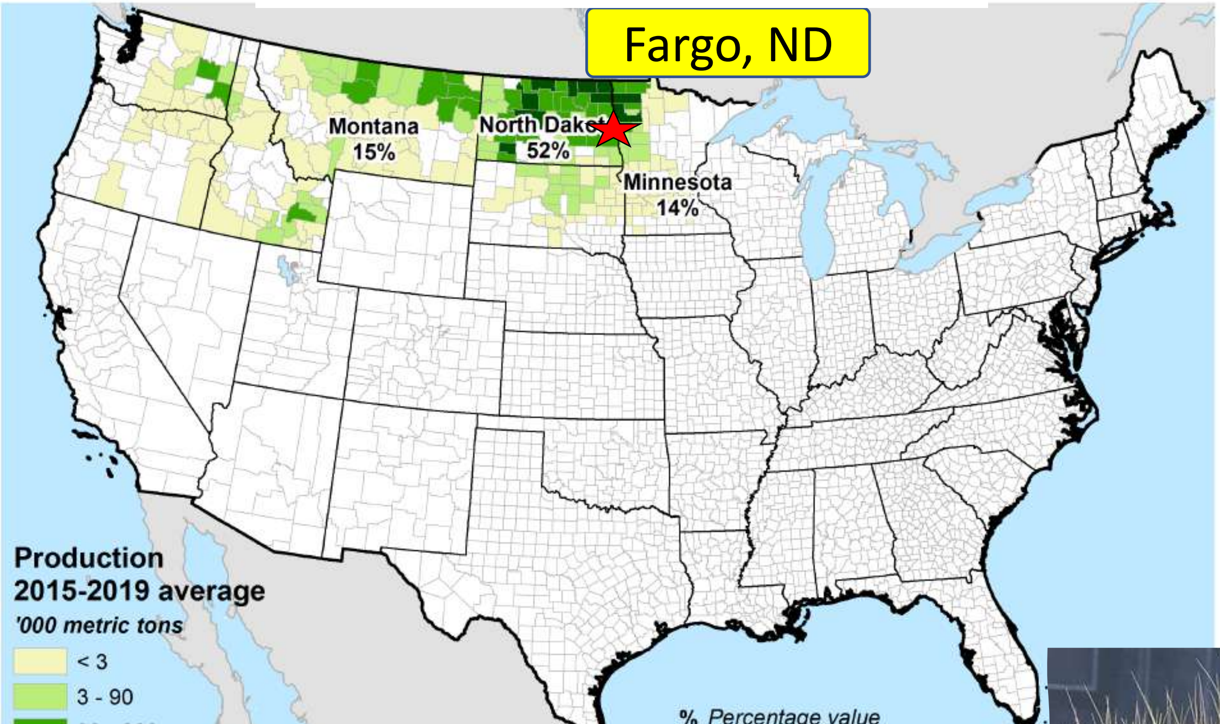
Wide Adaptation?

Quality?

Genomics can speed up the process!

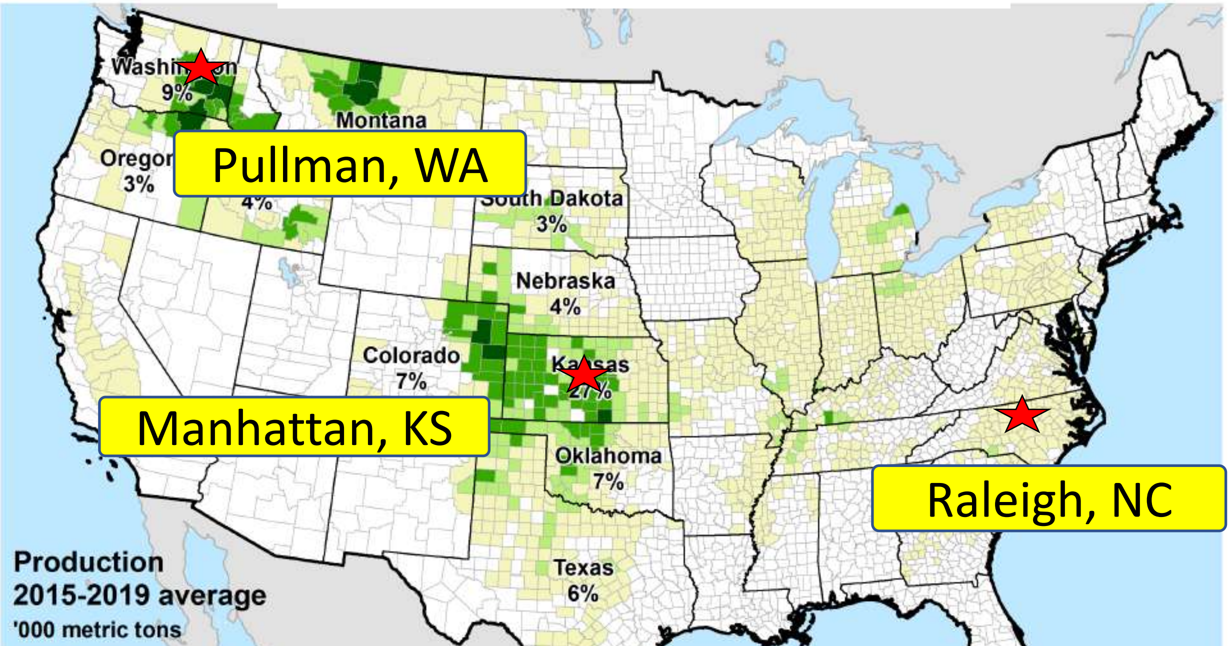
The Regional Small Grains Genotyping Labs

Spring Wheat Production



Fargo, ND

Winter Wheat Production



Pullman, WA

Manhattan, KS

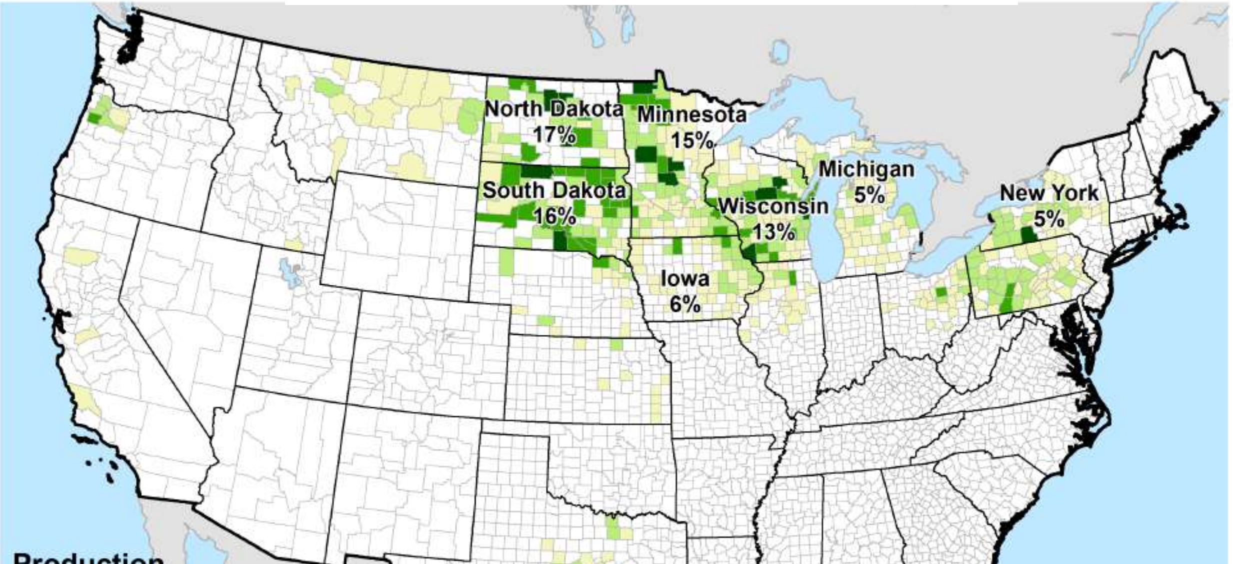
Raleigh, NC

USDA
United States Department of Agriculture
Agricultural Research Service

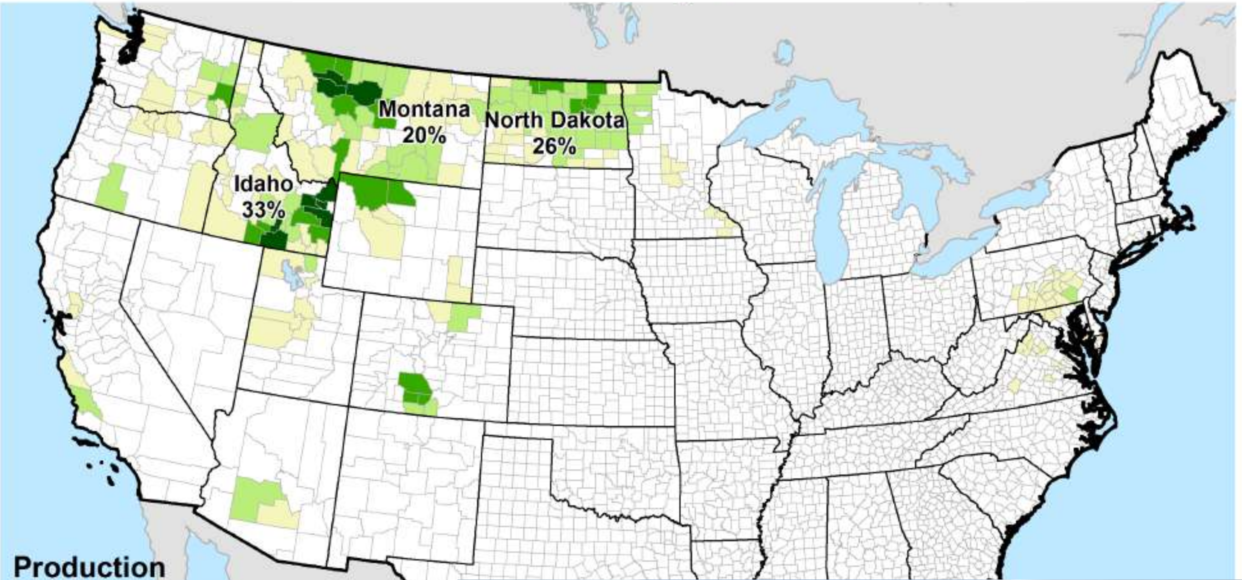
Edward T. Schafer
Agricultural Research Center

The Fargo Lab also supports Oat & Barley

Oat Production



Barley Production



USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

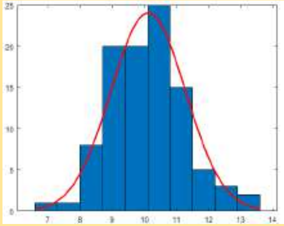
Bridge research and breeding through genomics

Breeding Programs

Crossing

Multi-location trials

Selection



New varieties

High yield

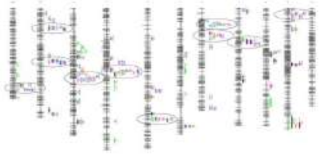
Superior quality

Disease resistant

USDA Small Grains Genotyping (USGG) Labs

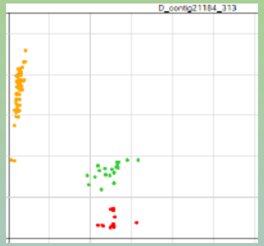
Basic/Applied Research

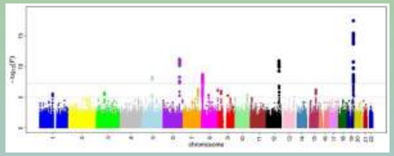
Germplasm development



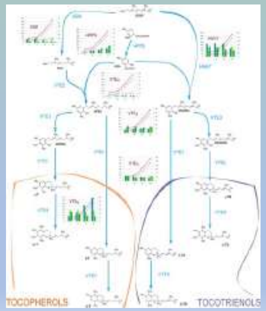
QTL identification

Marker development

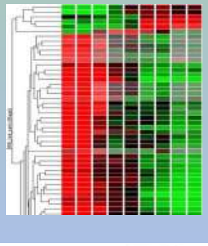




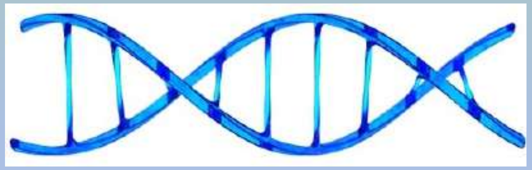
Pathway dissection



Gene discovery



Genome sequences



Karyotyping

USDA
United States Department of Agricult
Agricultural Research Servi

Edward T. Schafer
Agricultural Research Center

Deploying molecular markers in a breeding program

Make crosses

Inbreed/expand

Genotype with diagnostic markers

Predicted traits: disease, quality, etc

Disease/ Preliminary trials

Multi-location year trials

Evaluate quality traits

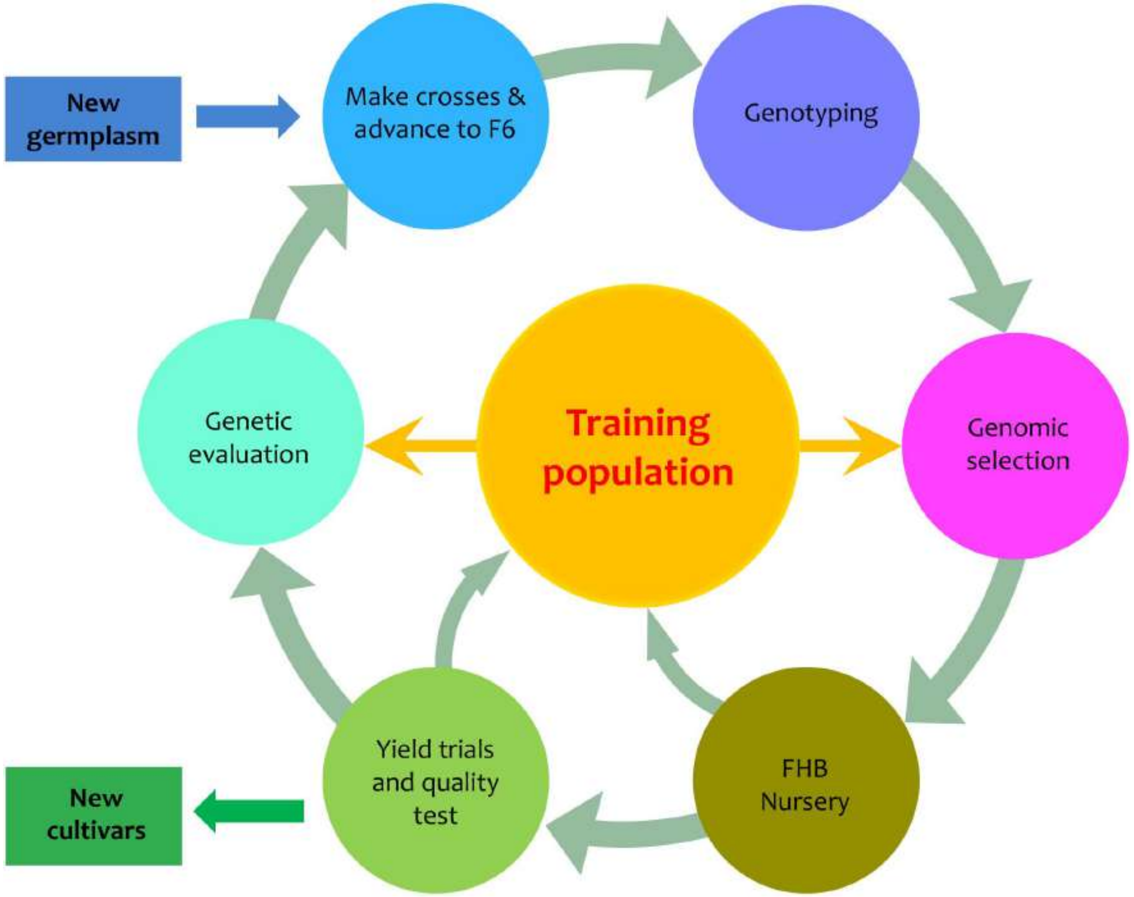
Kumar, et al. PLOS ONE. 2019

**Edward T. Schafer
Agricultural Research Center**

Genomic Selection (GS) is a valuable new tool

- Use genome-wide molecular markers to predict plant performance.
- Quantitative traits
- Commonly used at the F_4 / F_5 generation before multi-location year trials.

- Efficient usage requires a low-cost genome-wide genotyping platform.
    - Reproducible/Durable



Tsai, HY., et al. *Sci Rep*. 2020; Dong, et al. PLOS ONE. 2018; Fiedler, J.D., et al. *Plant Genome*. 2017; Campbell, M.T., et al. *Frontiers in Genet*. 2021

The USGG labs are developing low-cost platforms

Genotyping by Multiplex Sequencing (GMS)
(Ruff, et al 2020)

2,500 targets (wheat)
1,500 targets (barley)

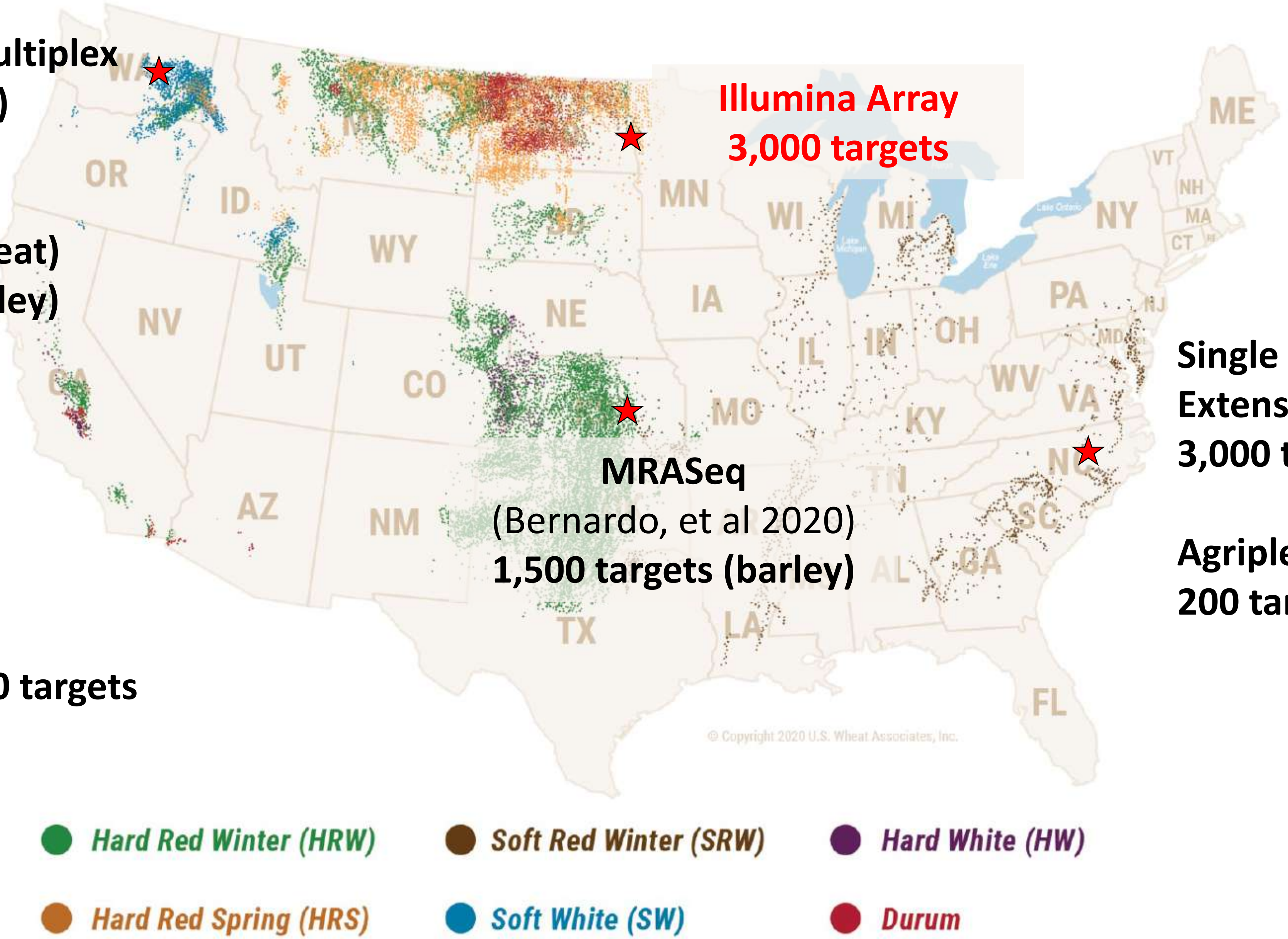
Illumina Array
3,000 targets

MRASeq
(Bernardo, et al 2020)
1,500 targets (barley)

Single Primer Extension (Allegro)
3,000 targets

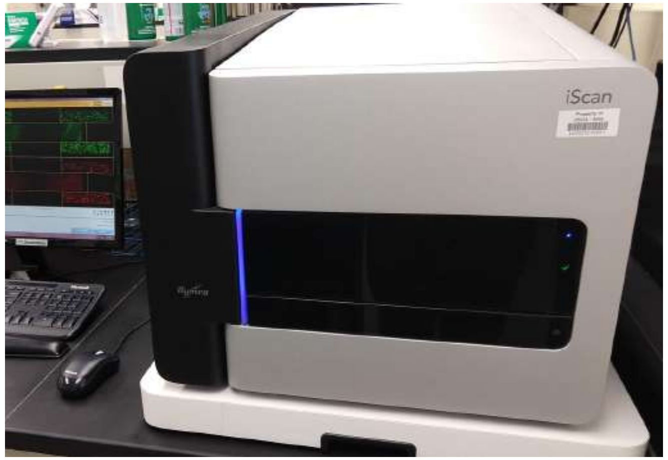
Agriplex
200 targets

AgriSeq
US panel of 5,000 targets (2023)



Custom multi-species Illumina iSelect array

- The Fargo Lab has a lot of experience with the Illumina arrays.
  - Run all the small grains arrays
  - Durable
  - Fast
  - Robust
- Work with different species
  - AgriBio Wheat/Barley 40K
- Qijian Song at the USDA-ARS, Beltsville (2021)
- ~3,000 SNPs in 4 crops



USDA-SoyWheOatBar-3K

- SNP selection
- Technical validation

1) Do the array SNPs match the source SNPs?
2) Can we use imputation to increase the # of SNPs (harmonization)?
3) Can dual hybridization be used to decrease cost?
4) Is the array useful?

USDA-3K

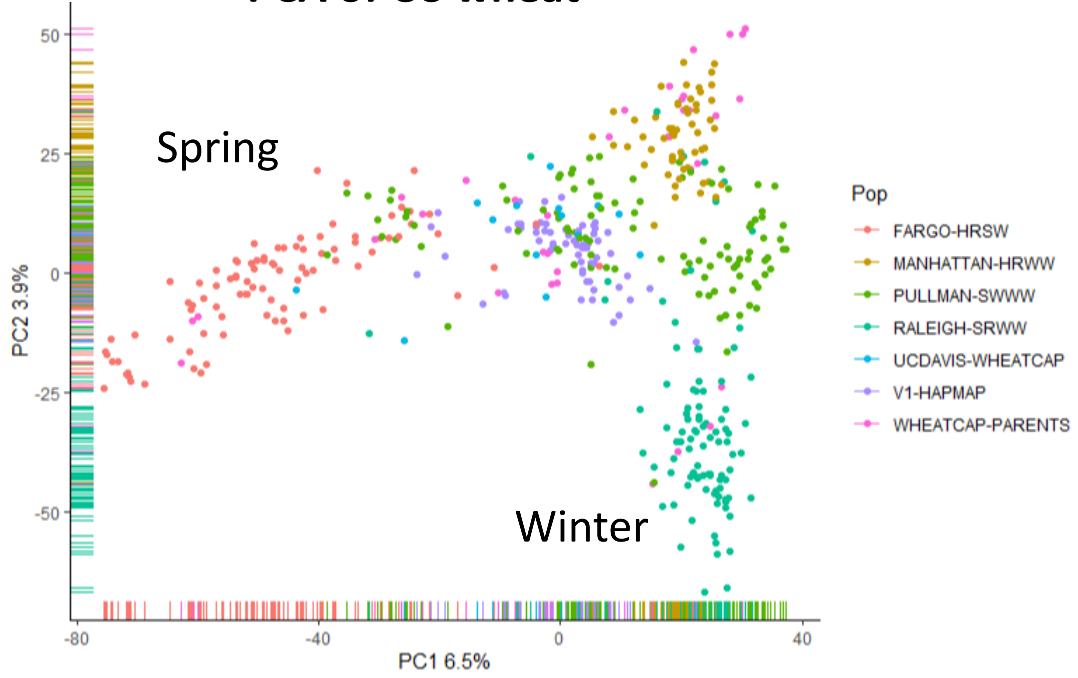
Edward T. Schafer Agricultural Research Center

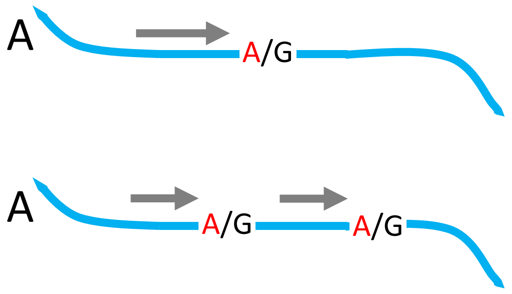
SNP/probe selection

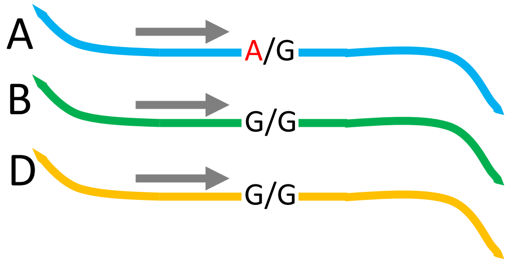
- Barley – 15,446 breeding lines 50K/9K data
- Oat – 7,344 breeding lines 6K data,~1,000 GBS data, PepsiCO array SNPs.
- Wheat – 384 exome capture (Eduard Akhunov, Brian Ward)

1) Cluster analysis to reproduce individual programs/market classes.
2) Interpolate genetic position, bin genome into 0.5 cM or 1 cM regions.
3) Pick one SNP per region (ranked by technical performance, maf, and number of clusters with maf > 0.05)
4) Fill large physical regions
5) Add known informative markers (existing MAS assays)
6) Replace probes that cross-hybridize and/or align multiple times to the genome > 1

PCA of US wheat





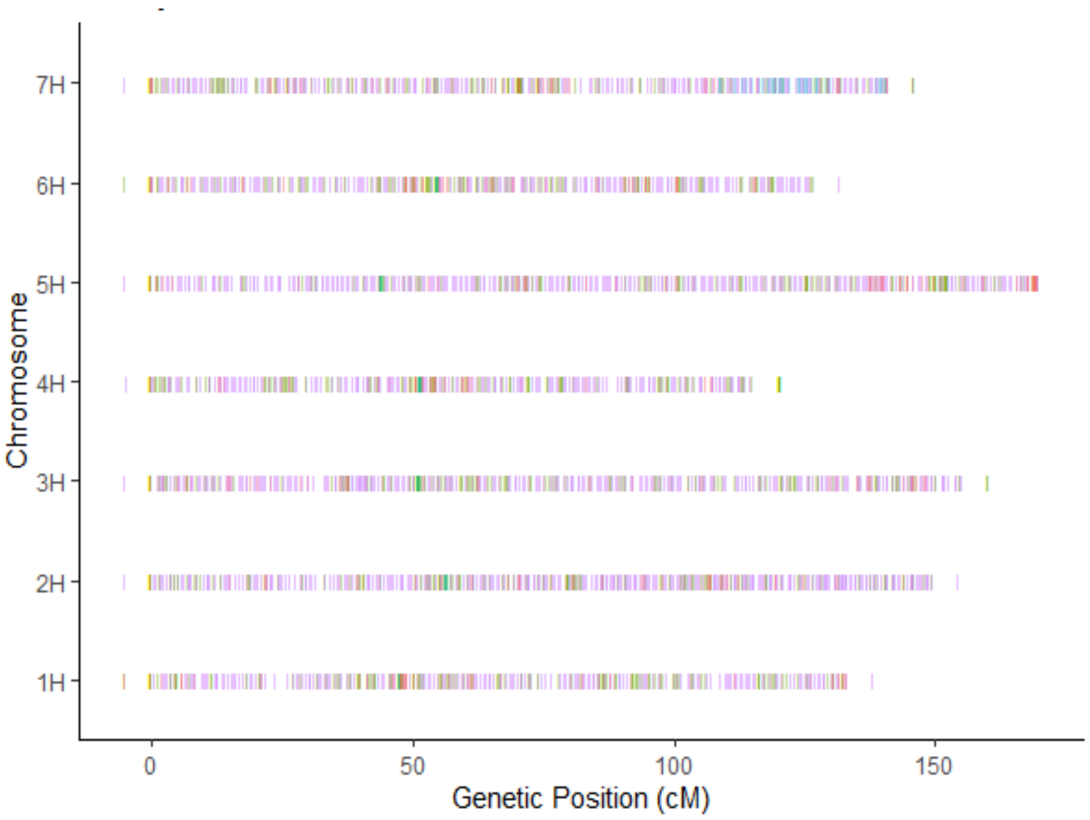


United States Department of Agriculture
Agricultural Research Service

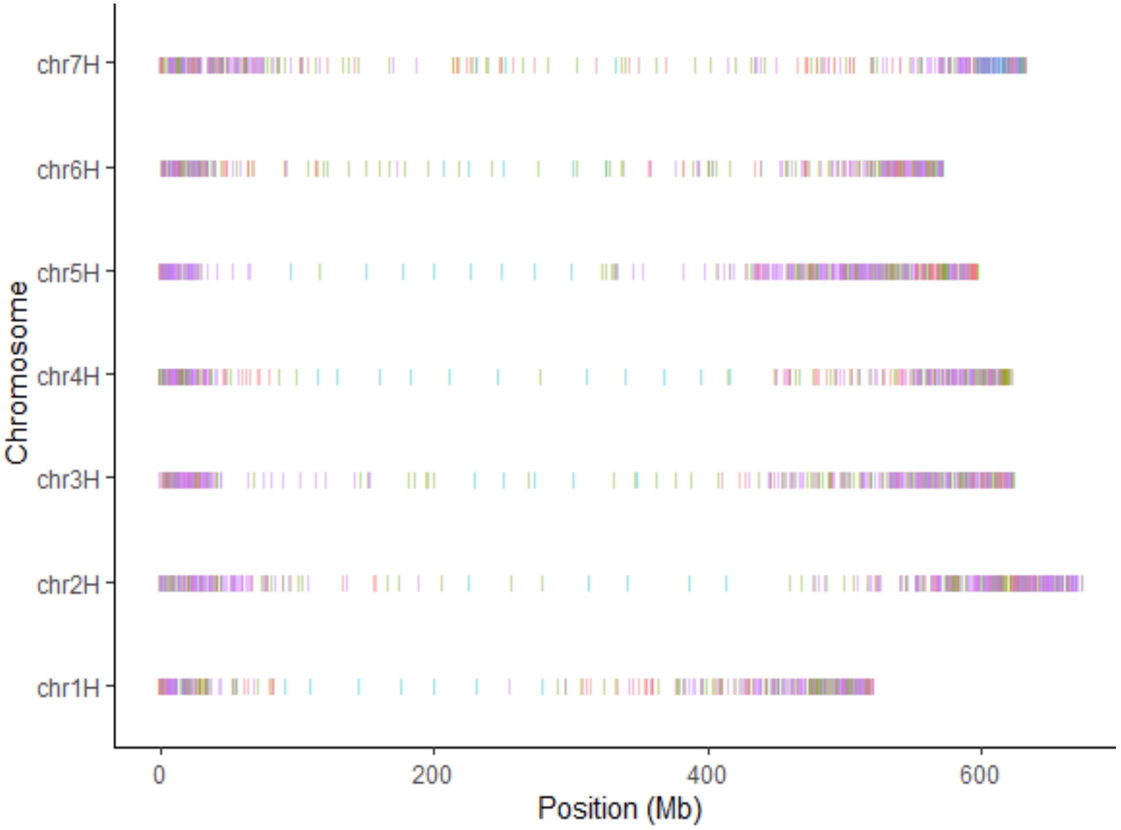
Edward T. Schafer
Agricultural Research Center

3,089 SNPs for barley

Genetic position on consensus map (Muñoz-Amatraín & Moscou et al 2011)



Physical position on Morex v3



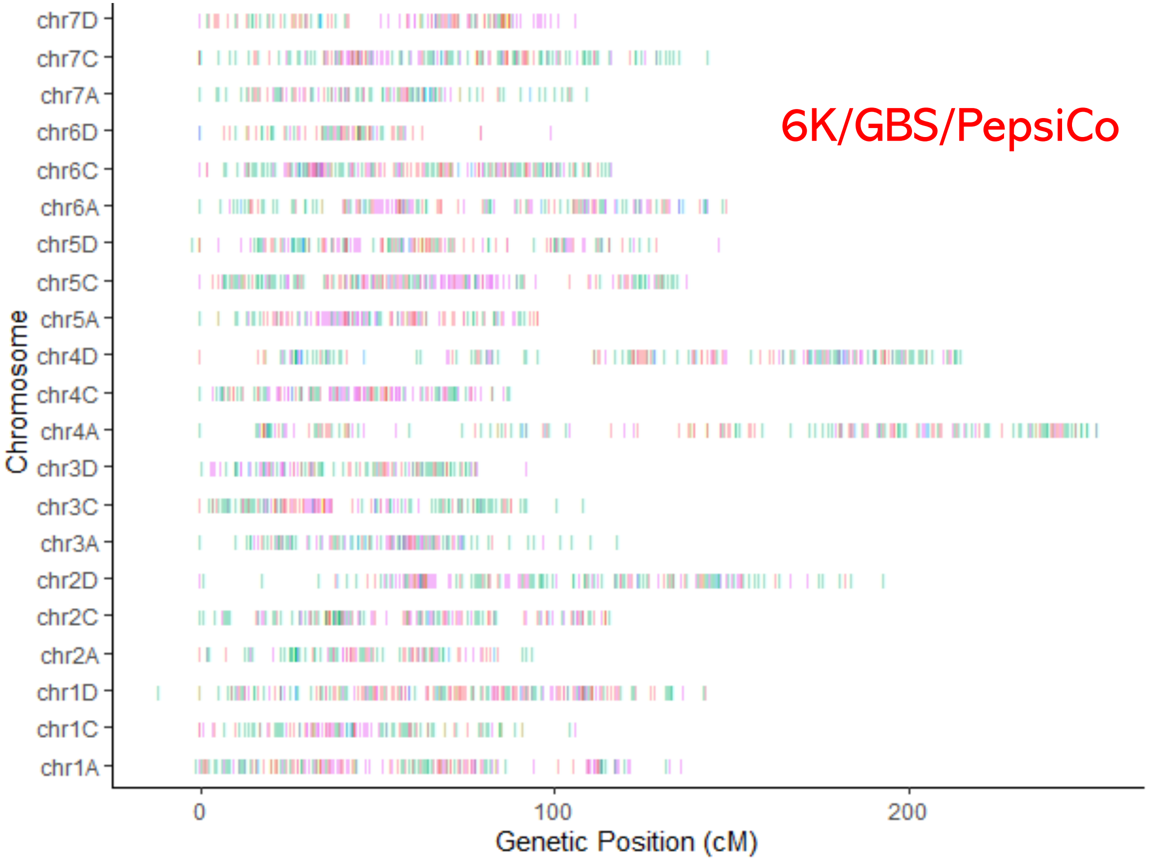
208 known informative markers

United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

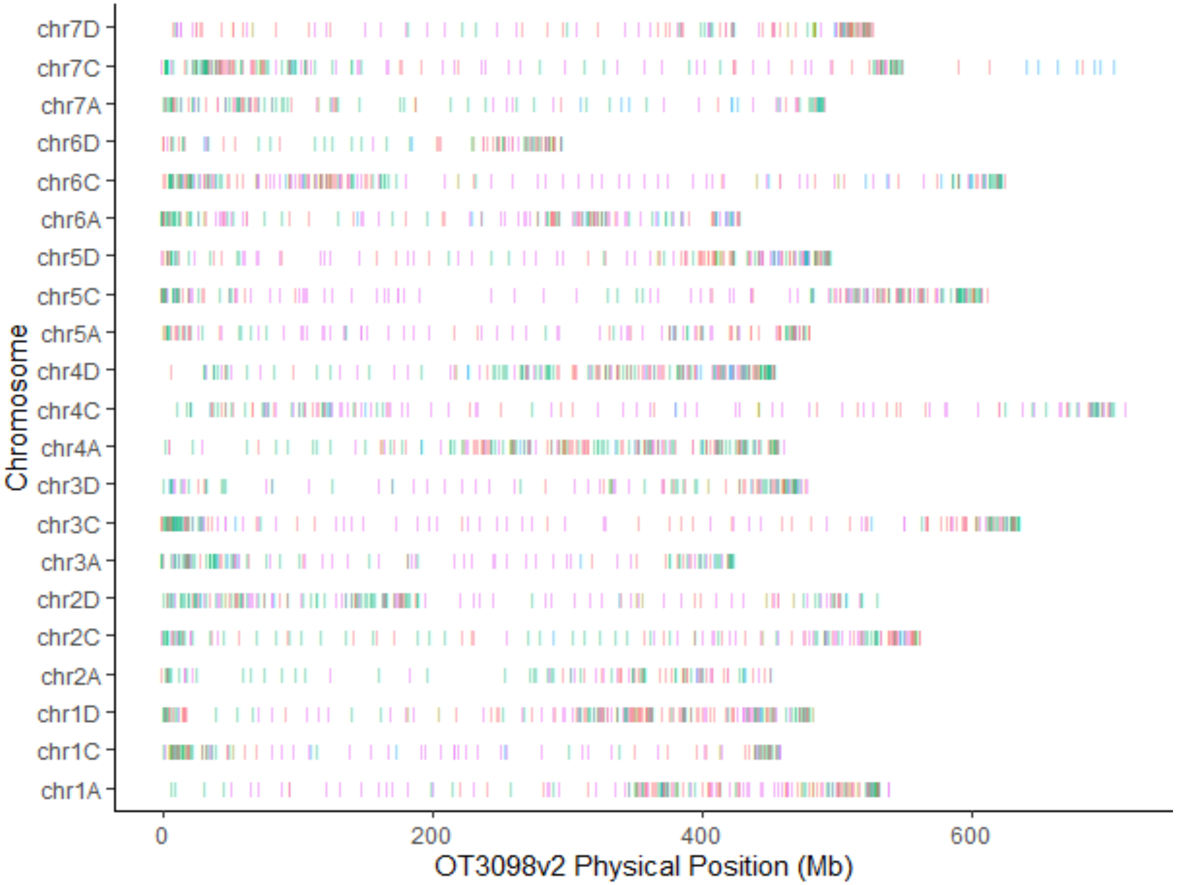
2,989 SNPs for oat

Genetic position on consensus map (Chaffin, et al 2016)



6K/GBS/PepsiCo

Physical position on OT3098 v2



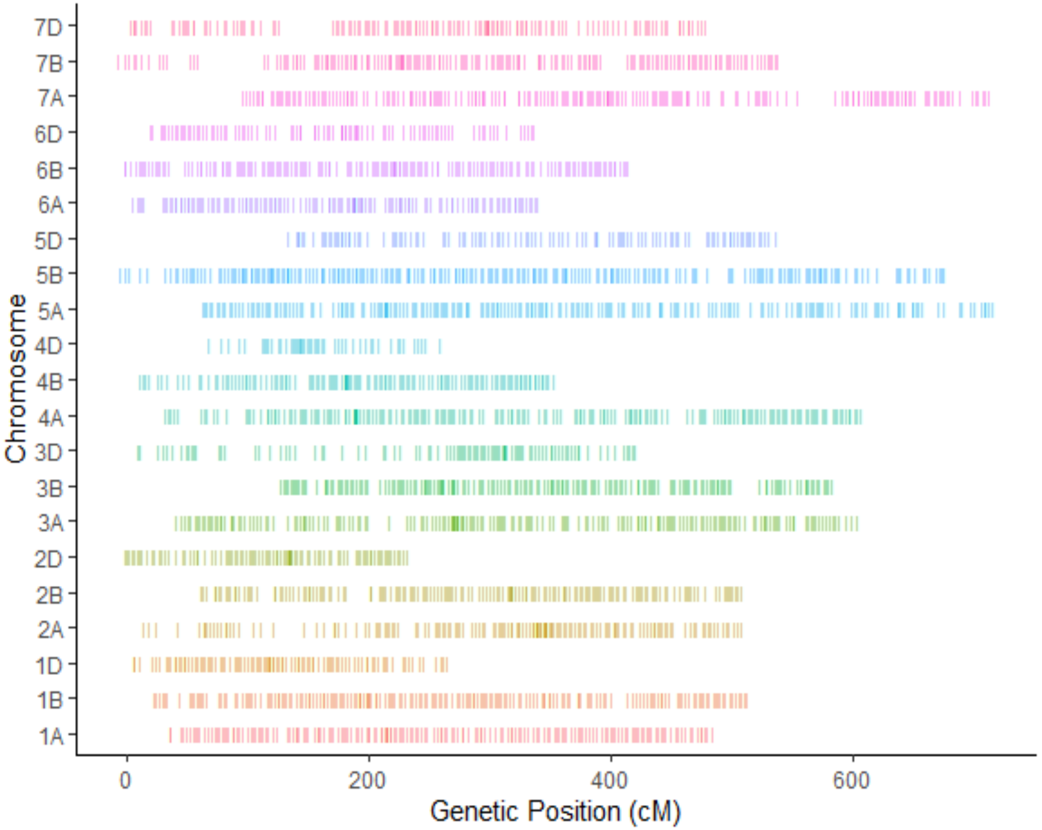
91 known informative markers

USDA
United States Department of Agriculture
Agricultural Research Service

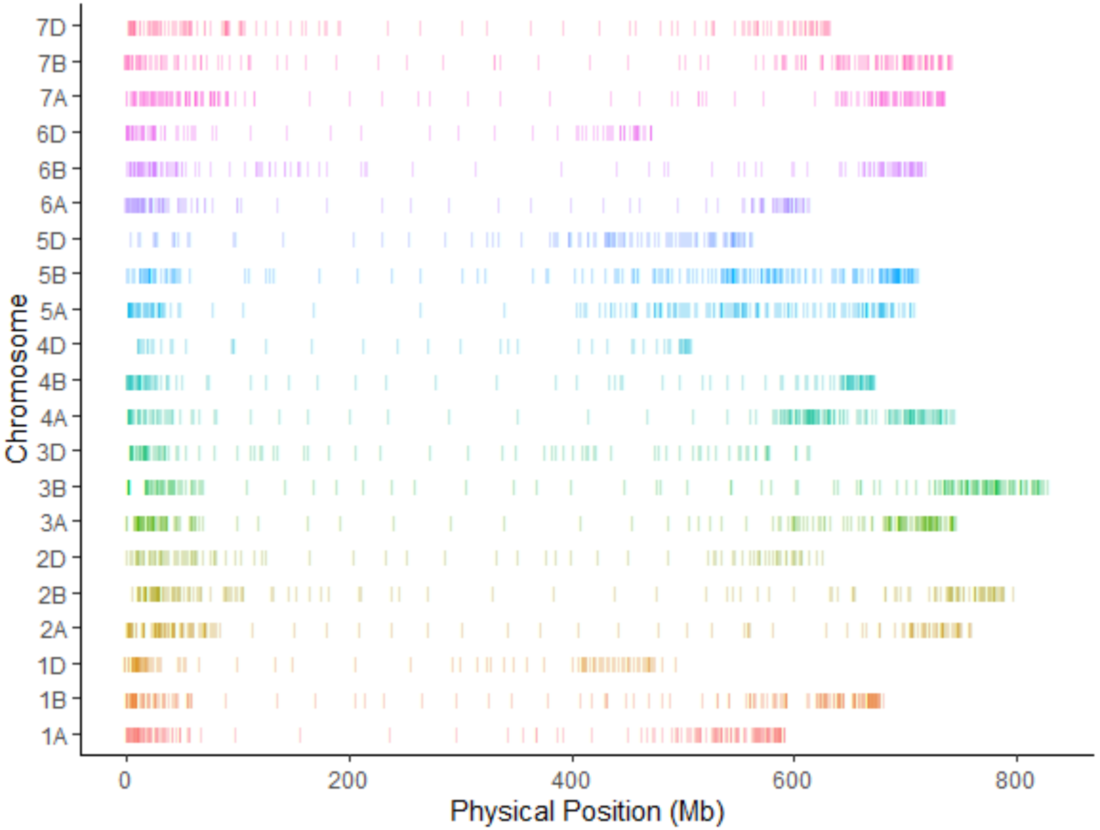
Edward T. Schafer
Agricultural Research Center

3,422 SNPs for wheat

Genetic position on consensus map (Wang, et al 2014)



Physical position on Refseq v2



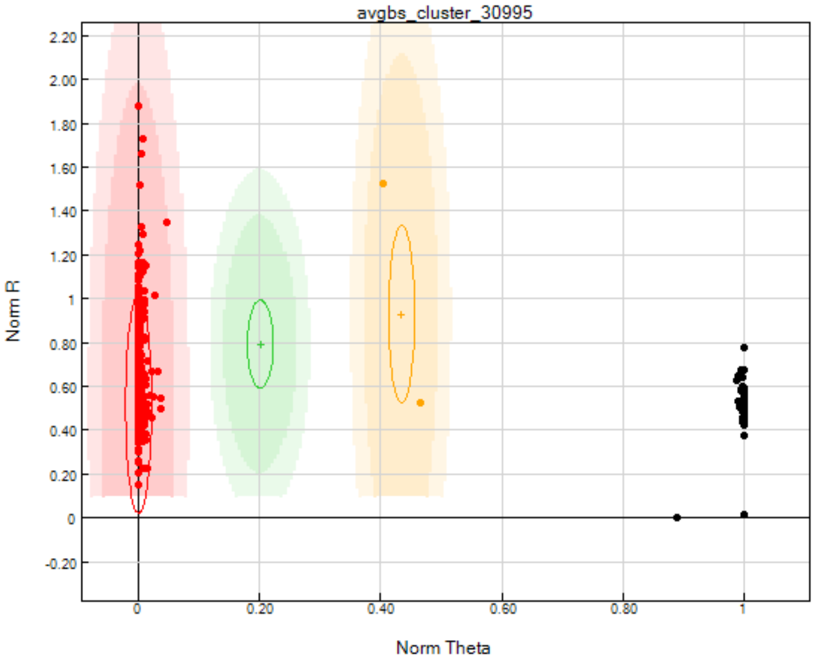
273 known informative markers

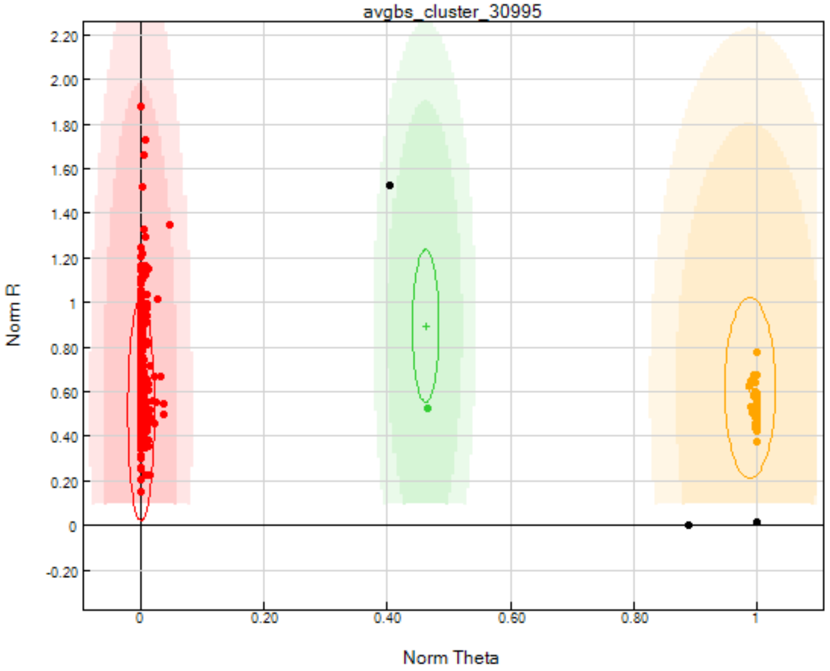
United States Department of Agriculture
Agricultural Research Service

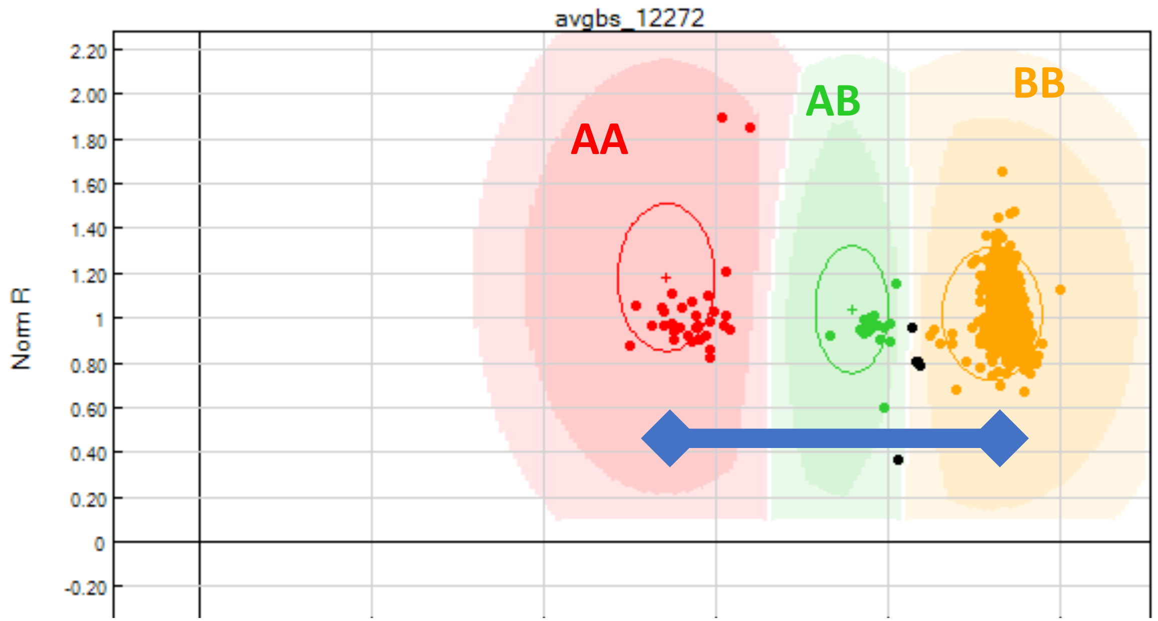
Edward T. Schafer
Agricultural Research Center

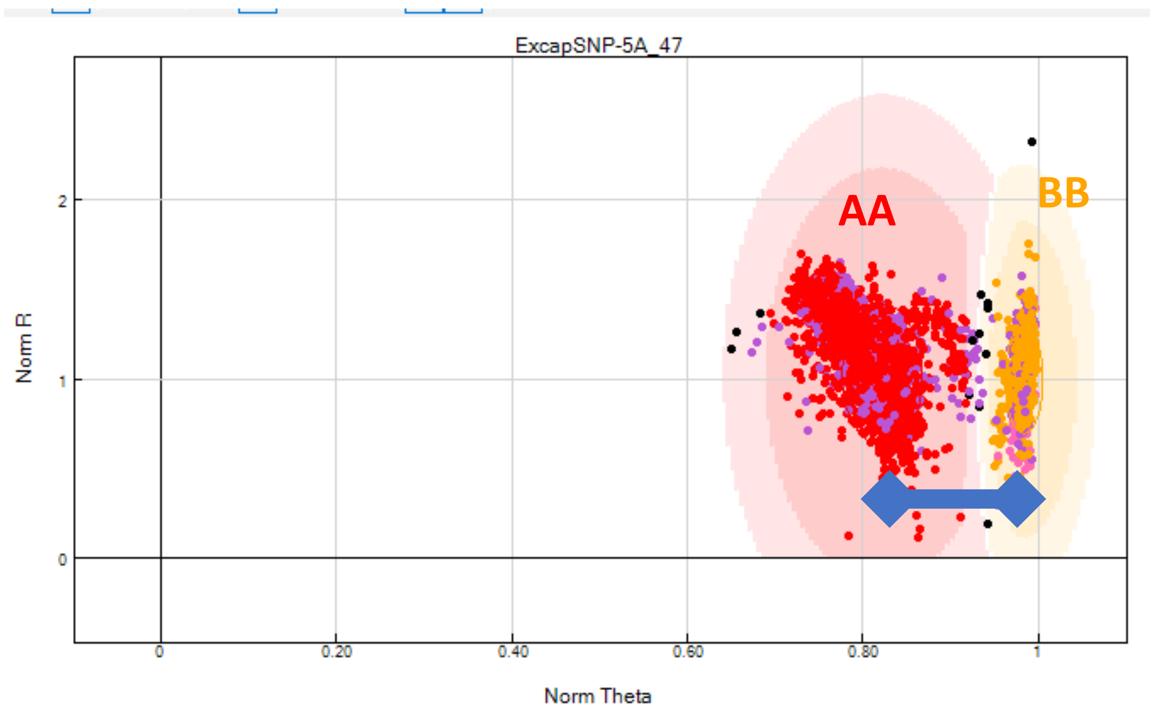
Validation Sets

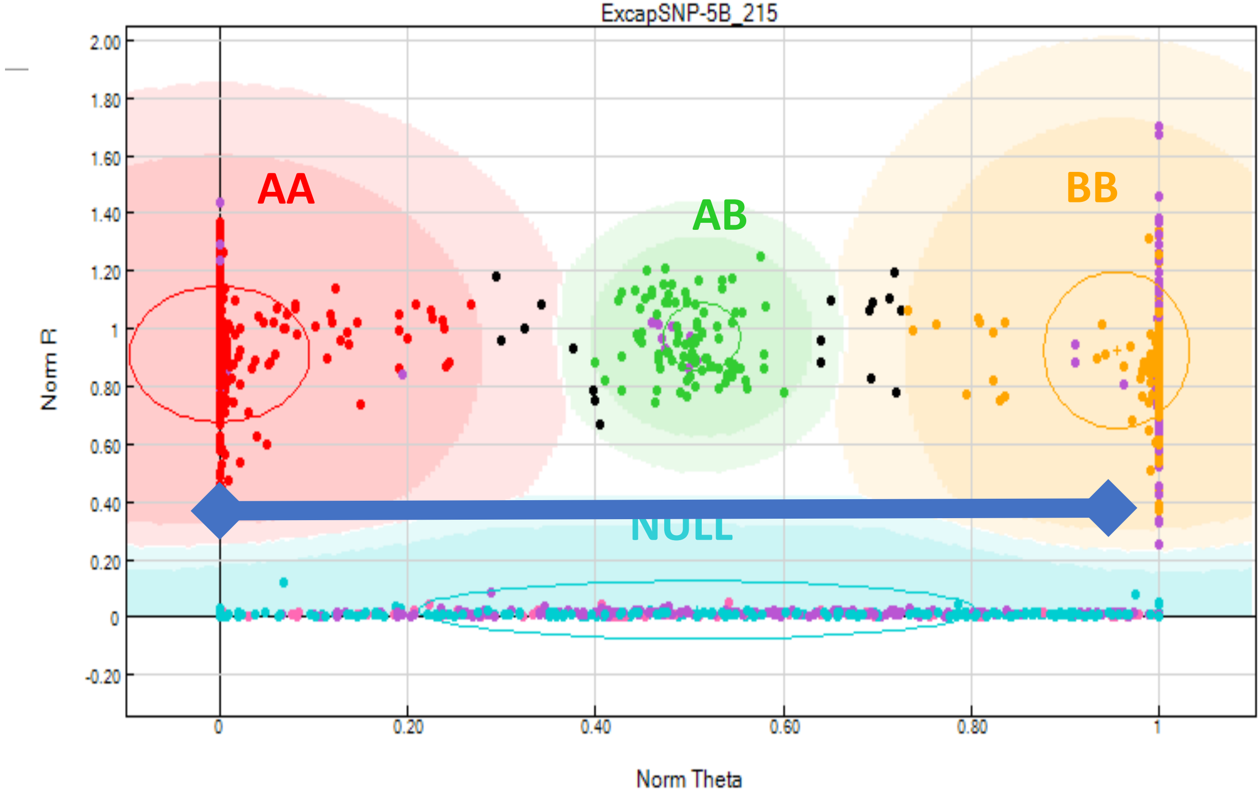
- Barley: 2,273 | Oat: 3,071 | Wheat: 5,397 (Spring, Winter, Durum)
- Manually checked/manipulated SNPs to better represent the data
  - Optimized positions used on all outgoing projects
- Conversion to VCF/Hapmap

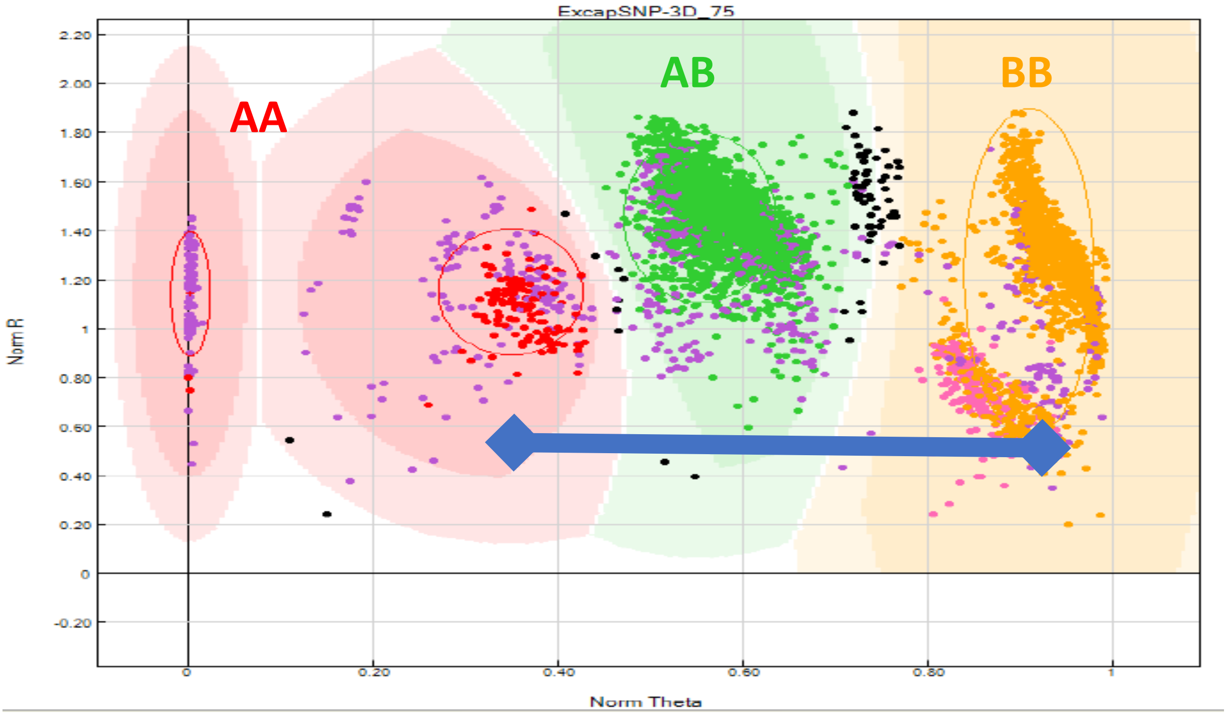




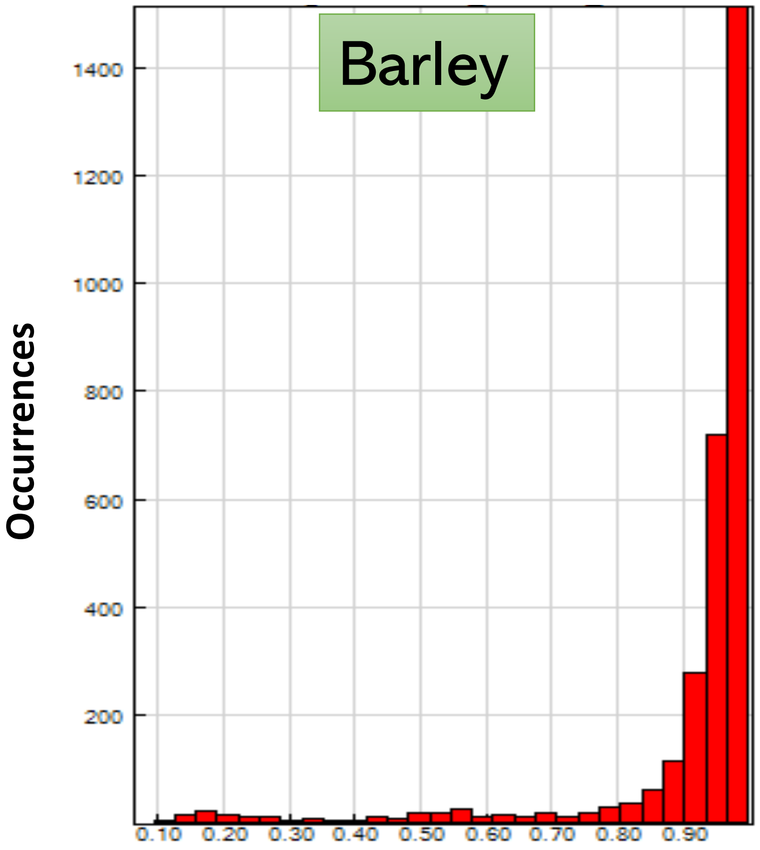

avgbs_12272
AA
AB
BB
Norm R
2.20
2.00
1.80
1.60
1.40
1.20
1
0.80
0.60
0.40
0.20
0
-0.20

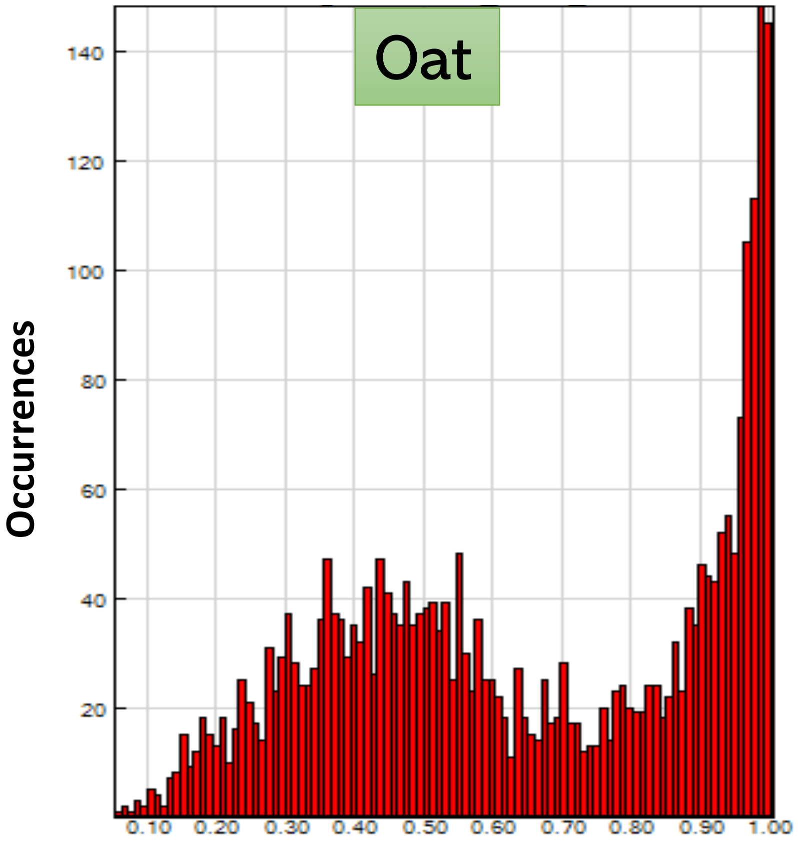

ExcapSNP-5A_47
AA
BB
Norm R
2
1
0
0
0.20
0.40
0.60
0.80
1
Norm Theta

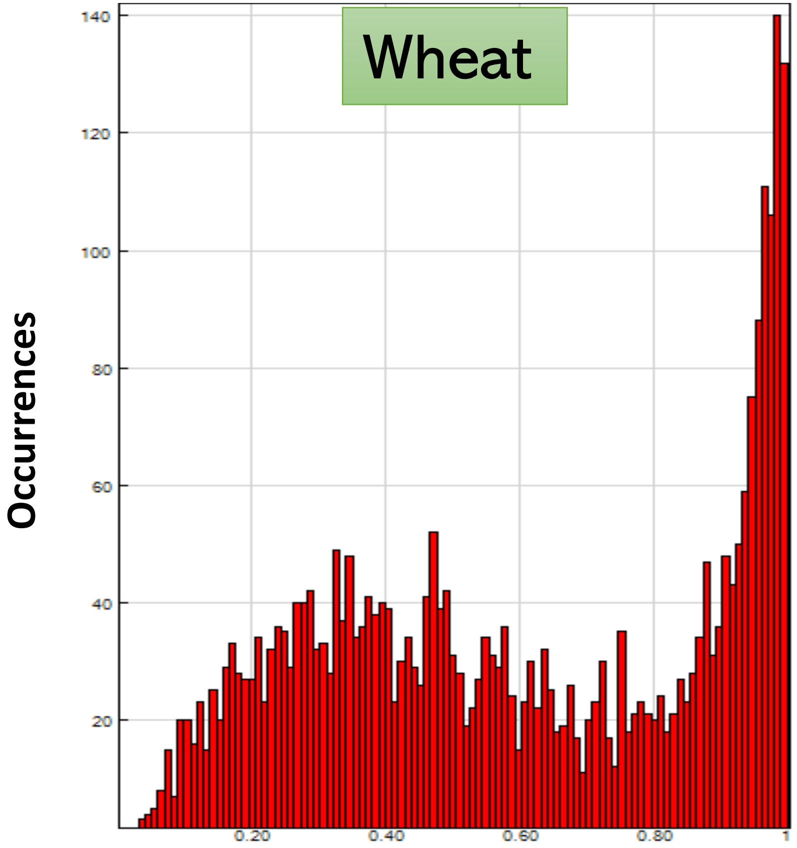

ExcapSNP-5B_215
AA
AB
BB
NULL
Norm R
2.00
1.80
1.60
1.40
1.20
1
0.80
0.60
0.40
0.20
0
-0.20
0
0.20
0.40
0.60
0.80
1
Norm Theta


ExcapSNP-3D_75
AA
AB
BB
Norm R
2.20
2.00
1.80
1.60
1.40
1.20
1
0.80
0.60
0.40
0.20
0
-0.20
0
0.20
0.40
0.60
0.80
1
Norm Theta

Distance between AA and BB Cluster

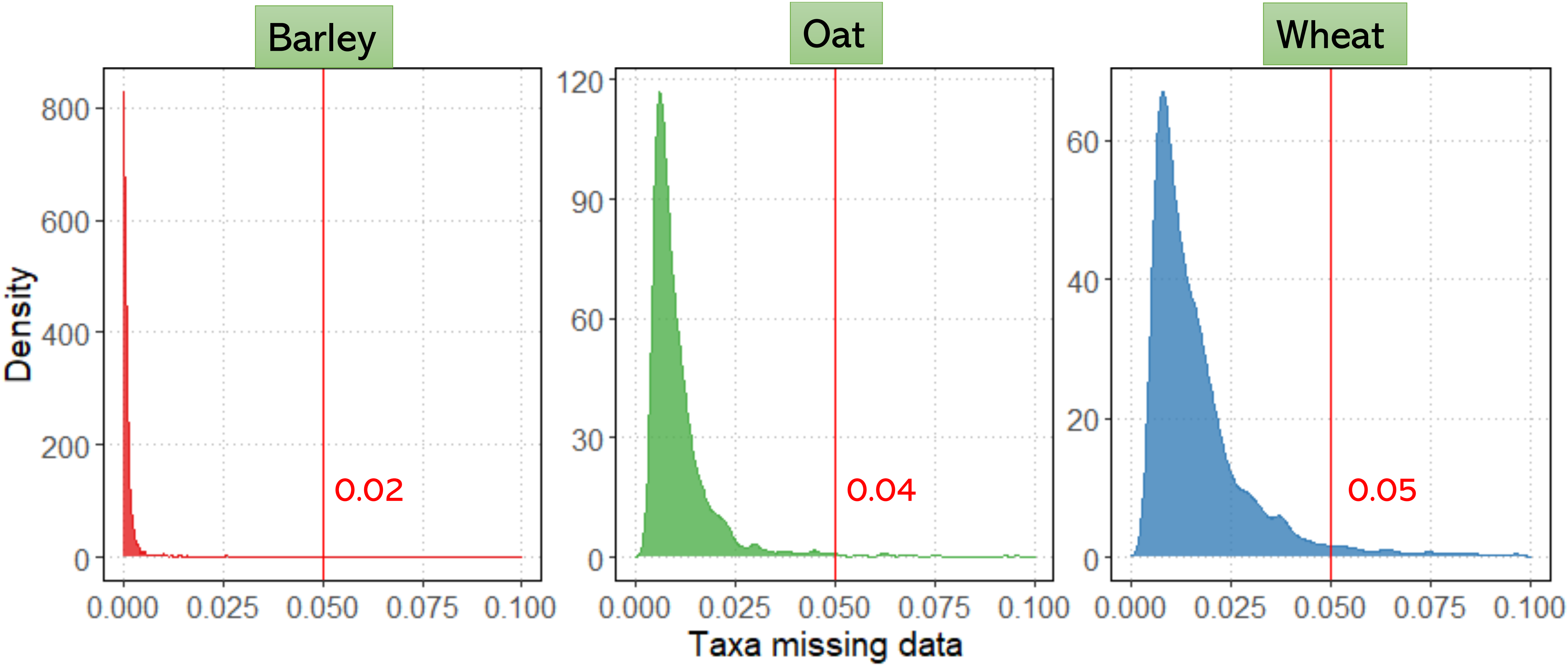




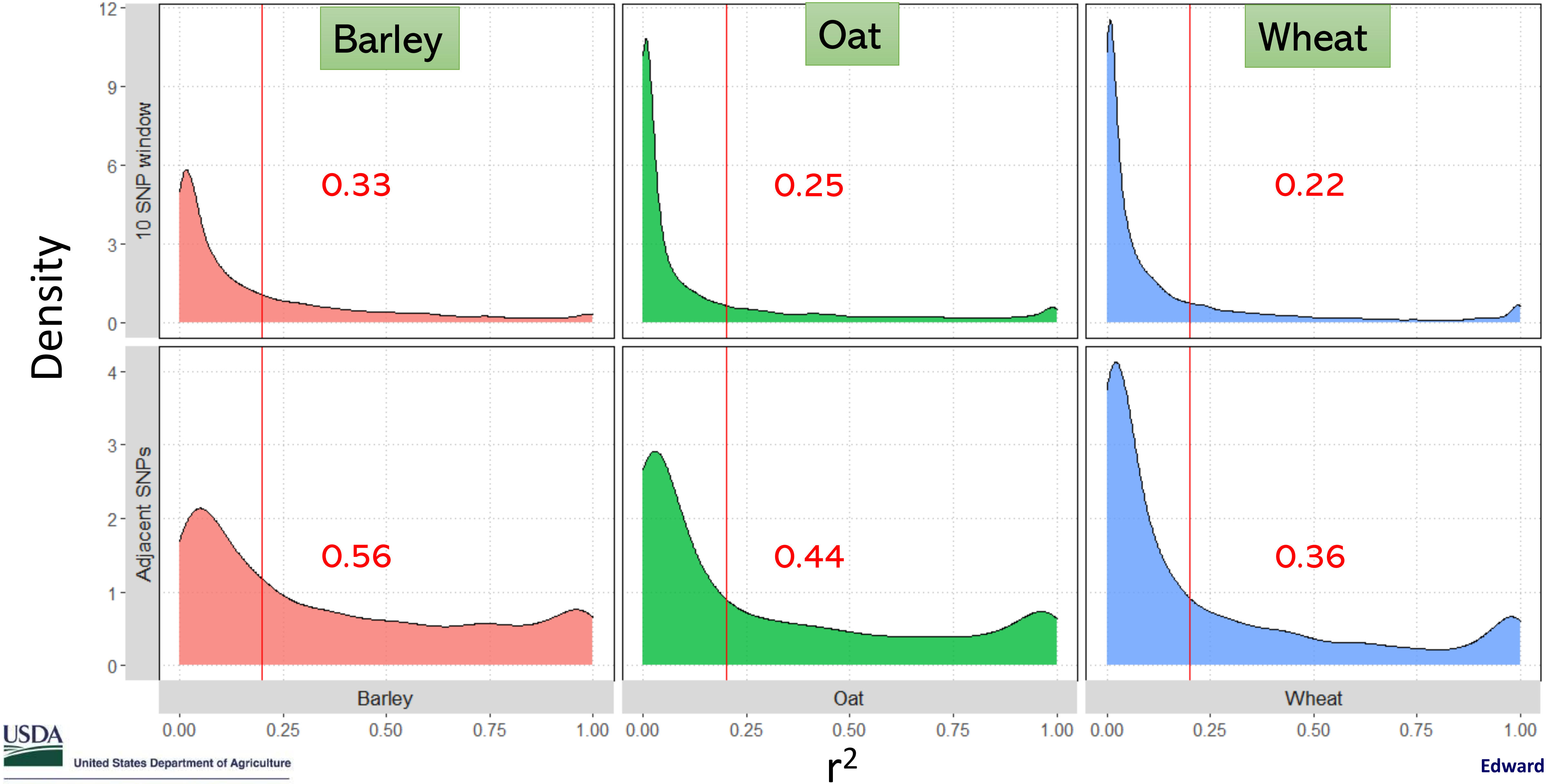


Call Frequency

Excellent call rates



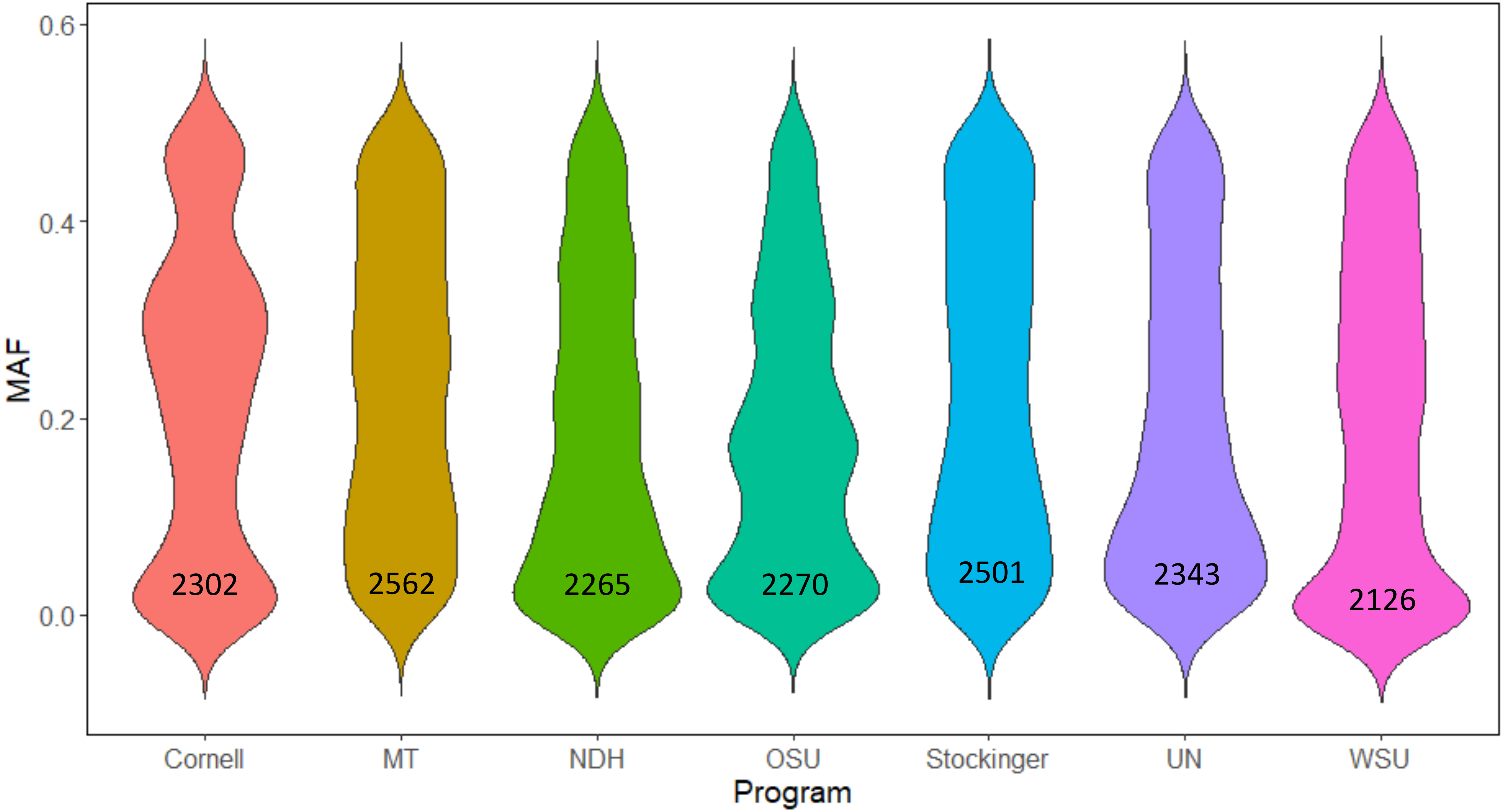
Linkage disequilibrium distributions at local scales



USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Allele frequency distribution by program - barley

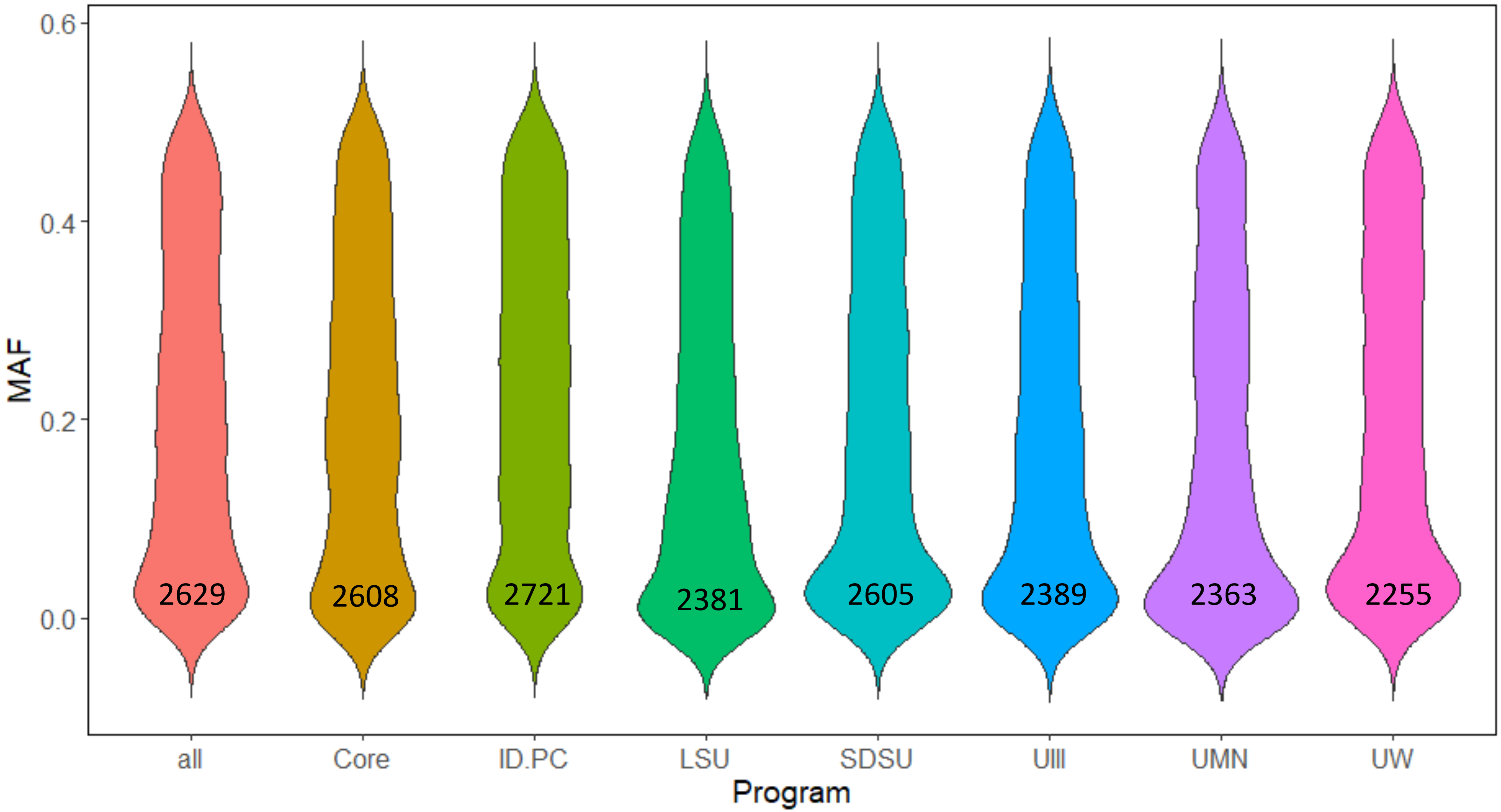


* SNPs MAF >= 0.05

United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Allele frequency distribution by program - oat

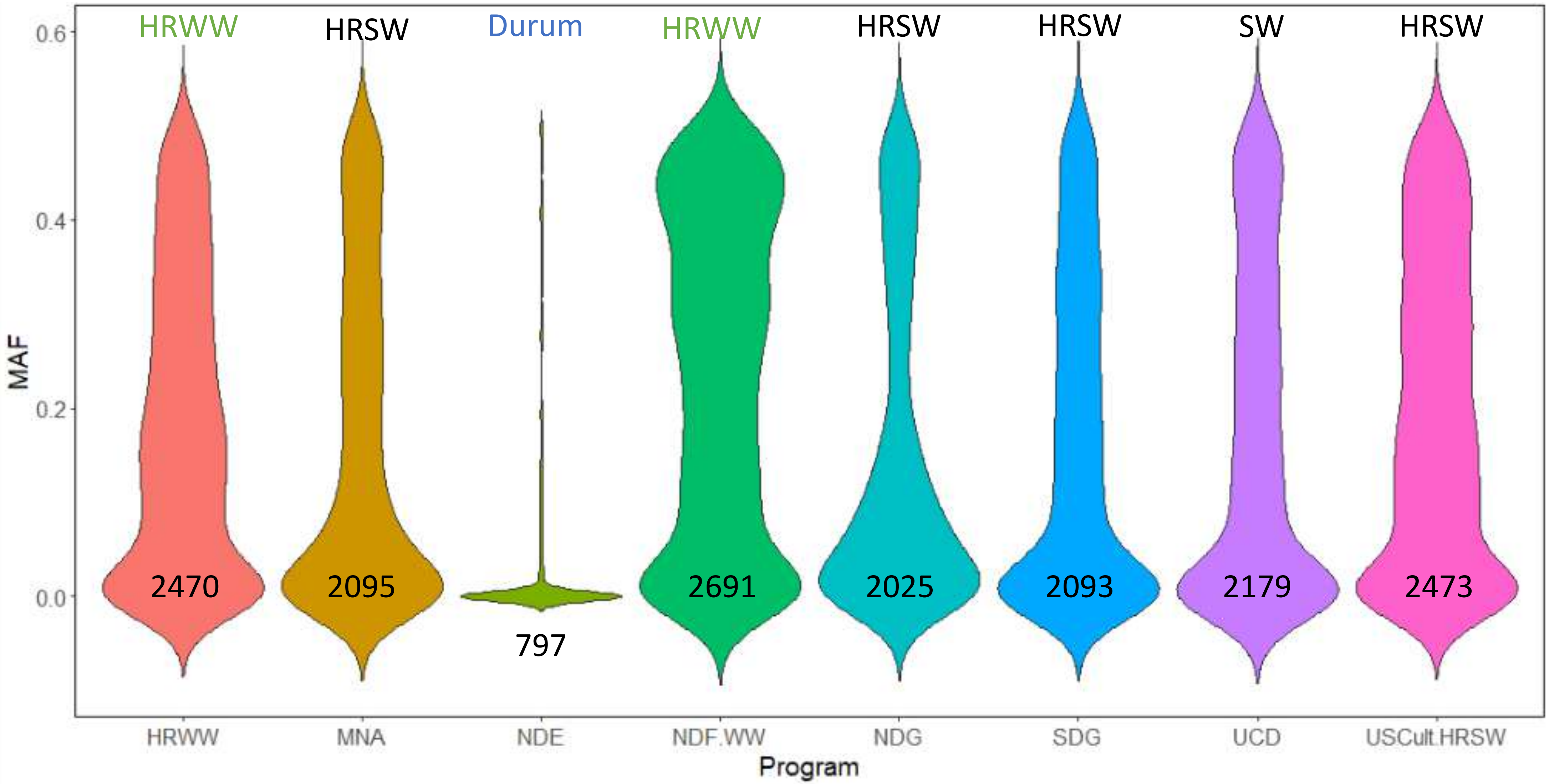


* SNPs MAF >= 0.05

USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Allele frequency distribution by program - wheat

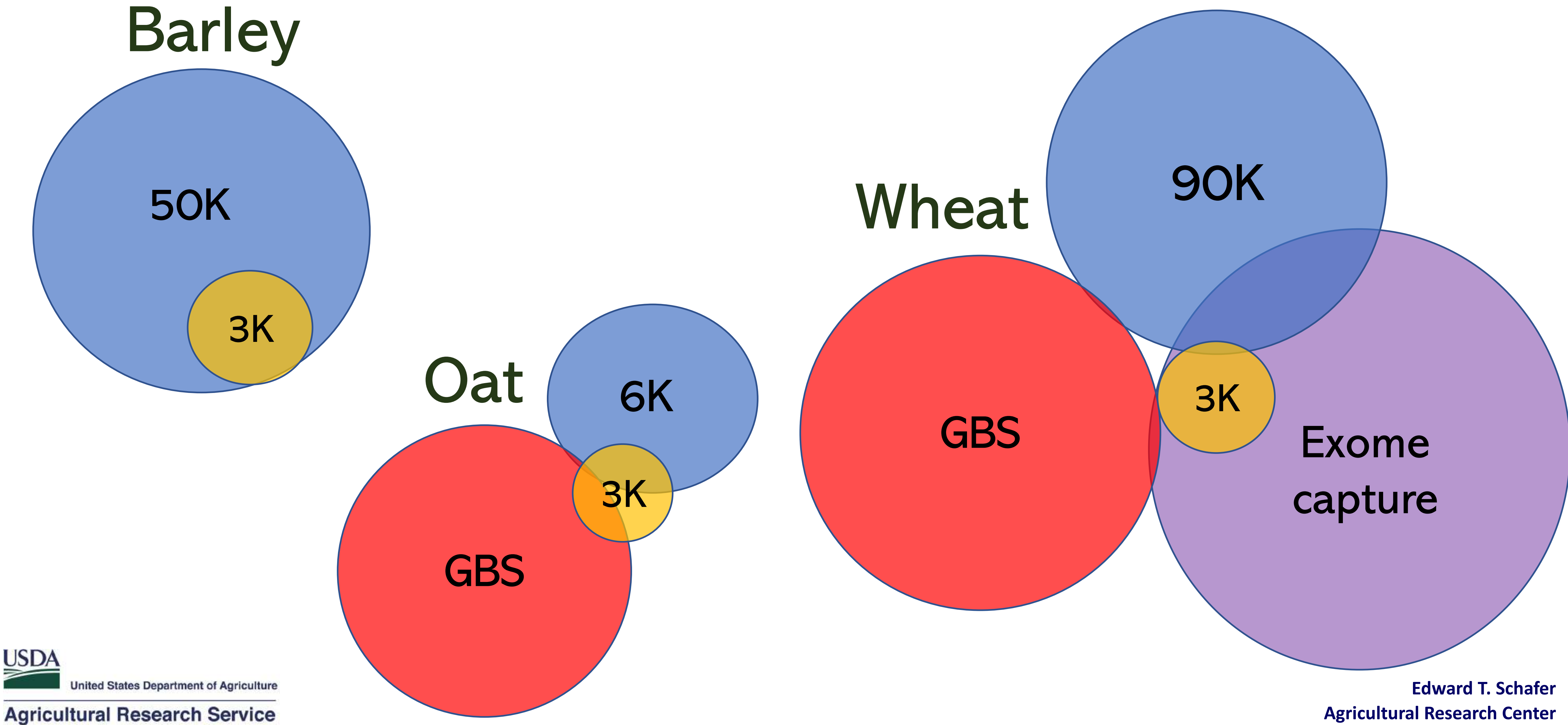


* SNPs MAF >= 0.05

USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

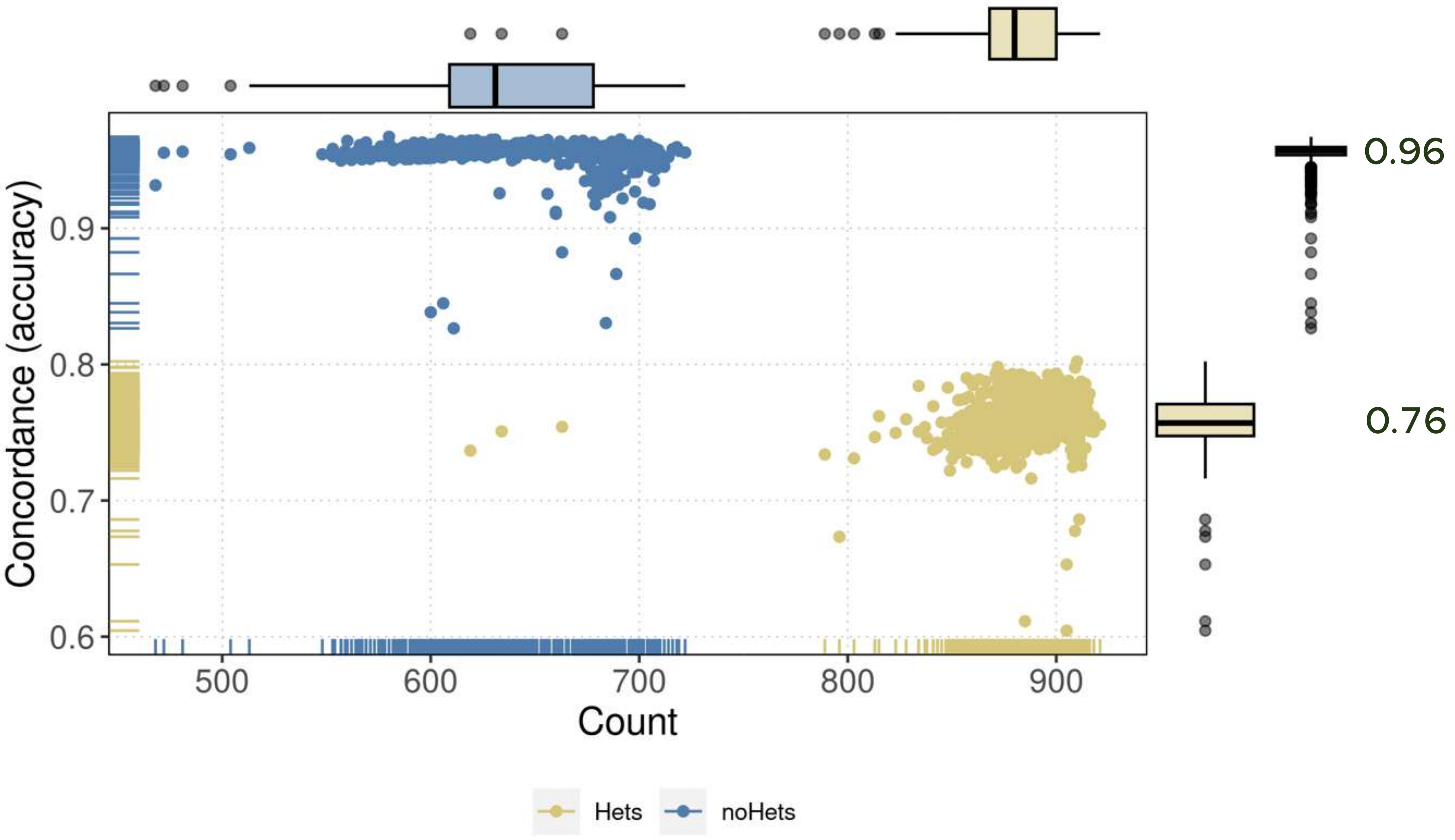
Do the SNPs match the source?



United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

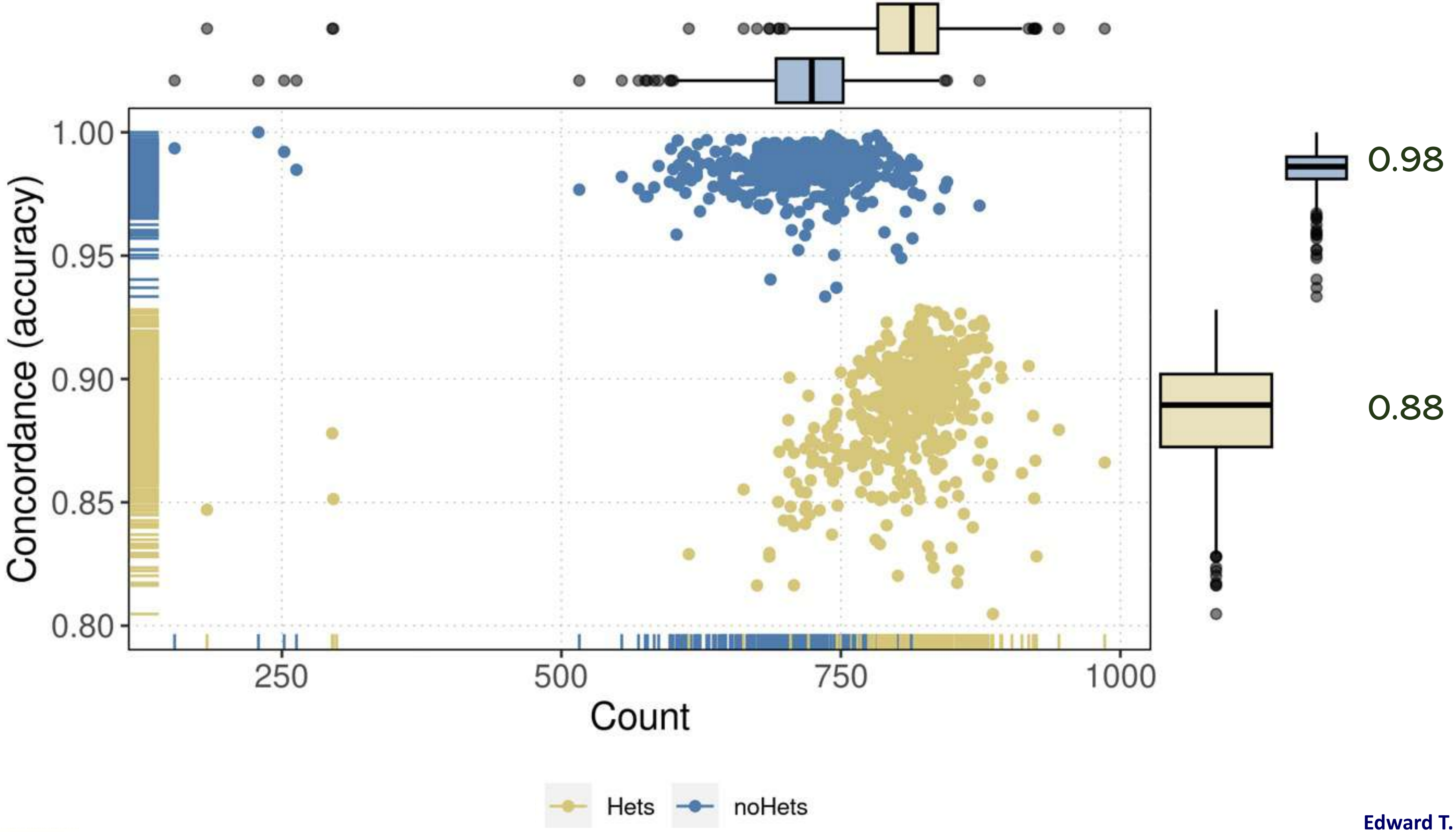
Oat 6K SNPs match 3K



United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

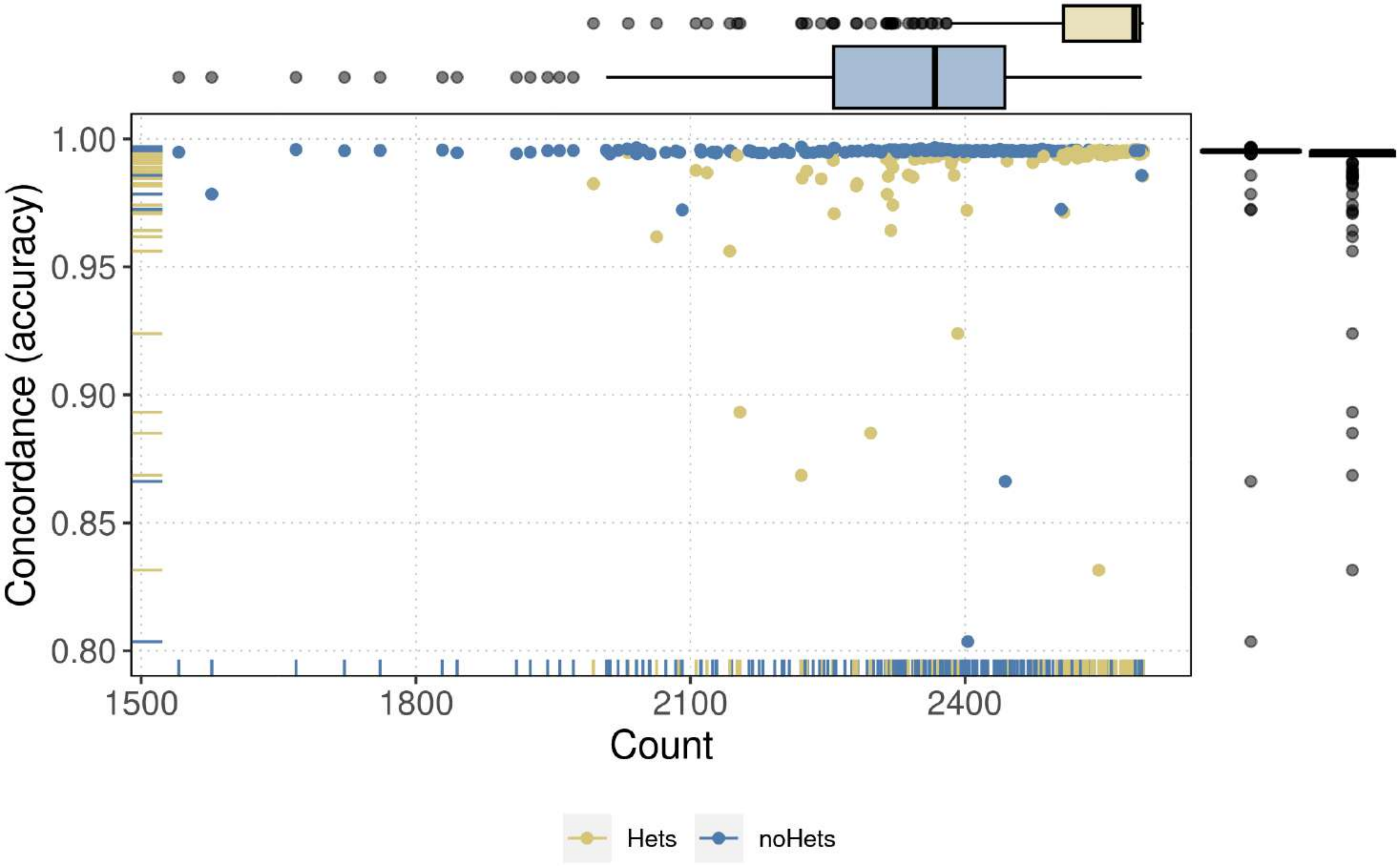
Oat GBS SNPs match 3K



USDA
United States Department
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Barley 50K SNPs match 3K

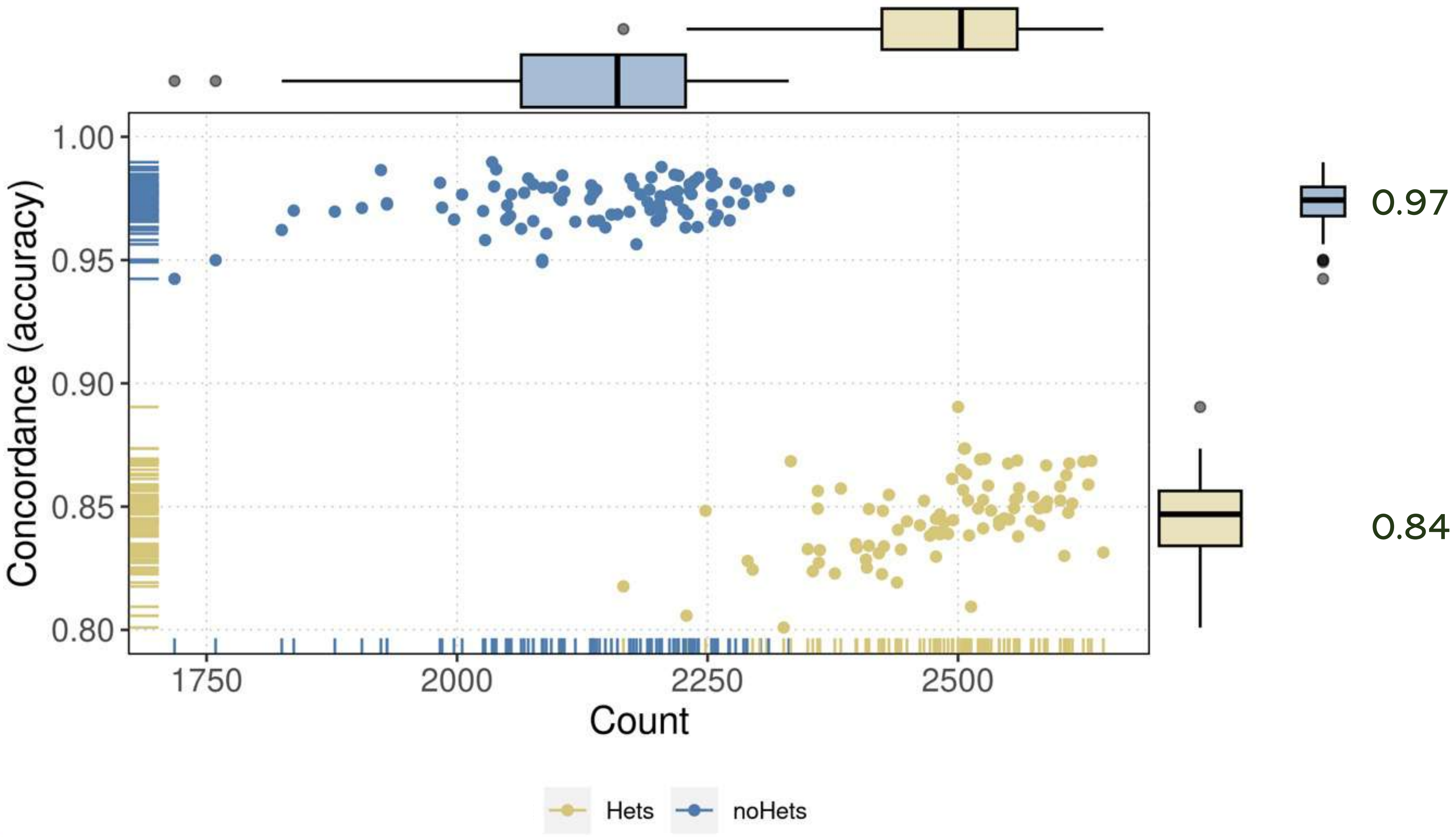


0.995

USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Wheat exome capture SNPs match 3K

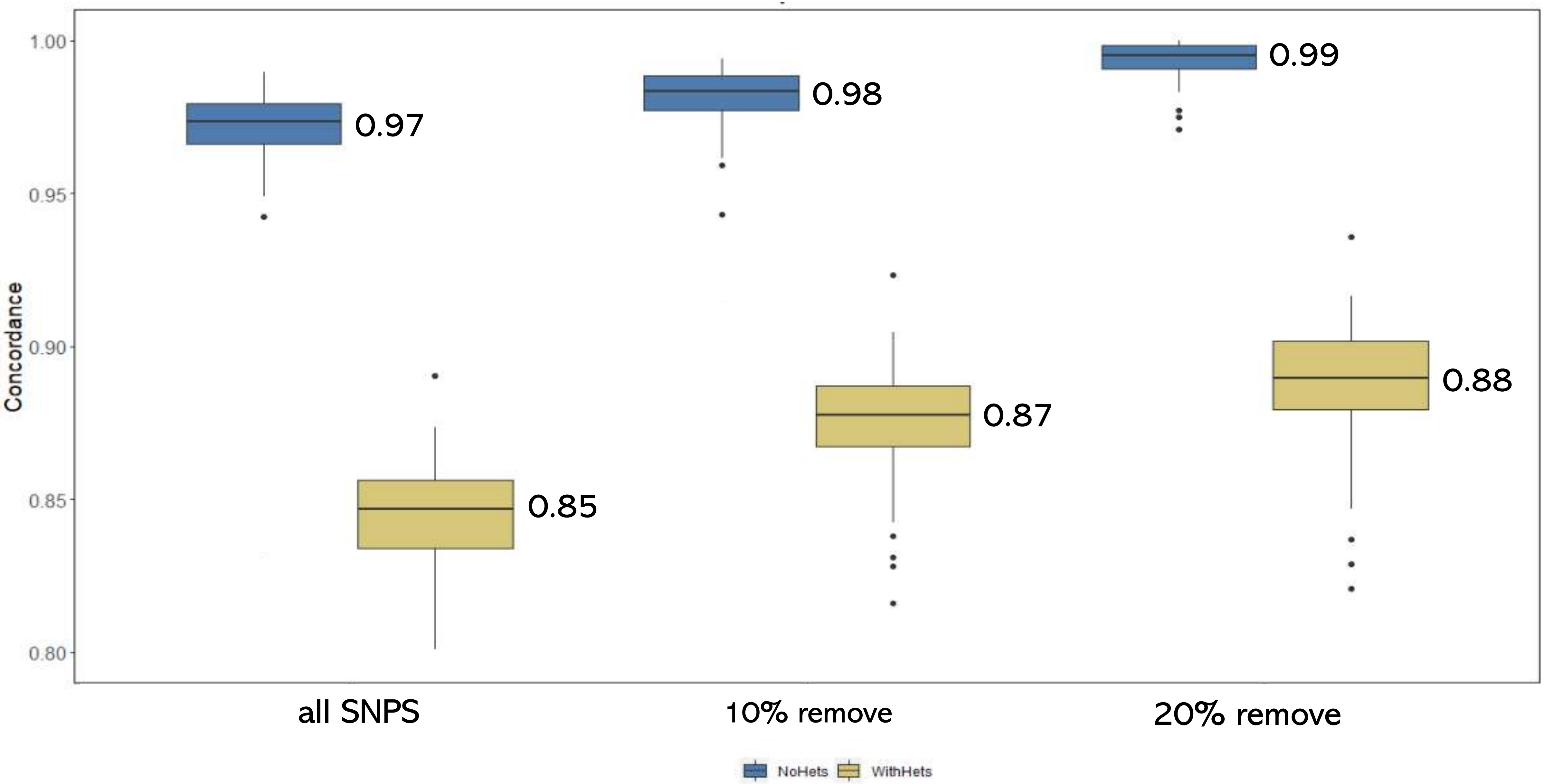


United States Department of Agriculture

Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

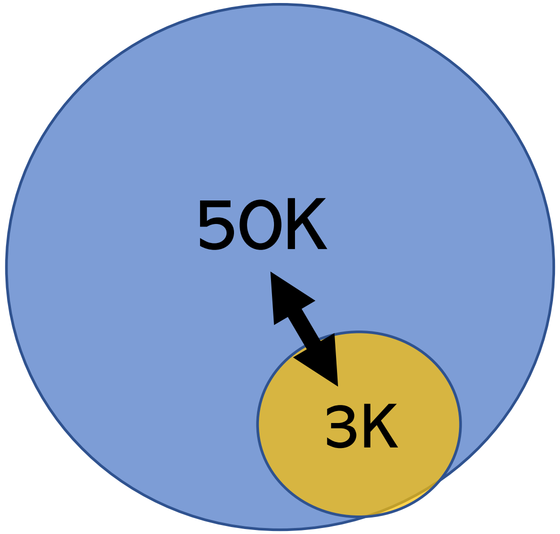
Removing poor performing SNPs increases taxa accuracy



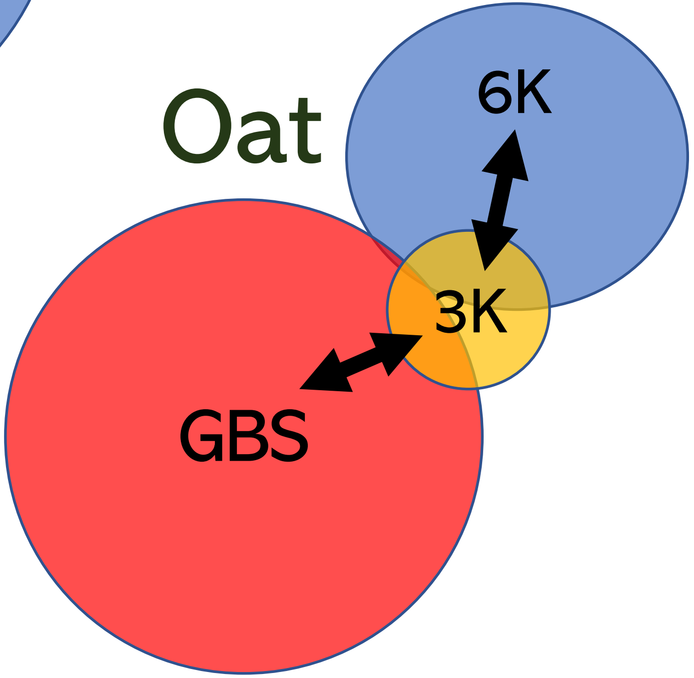
Can we use imputation to *upgrade* the data?

- Harmonization of different platforms.
- Identify linked markers

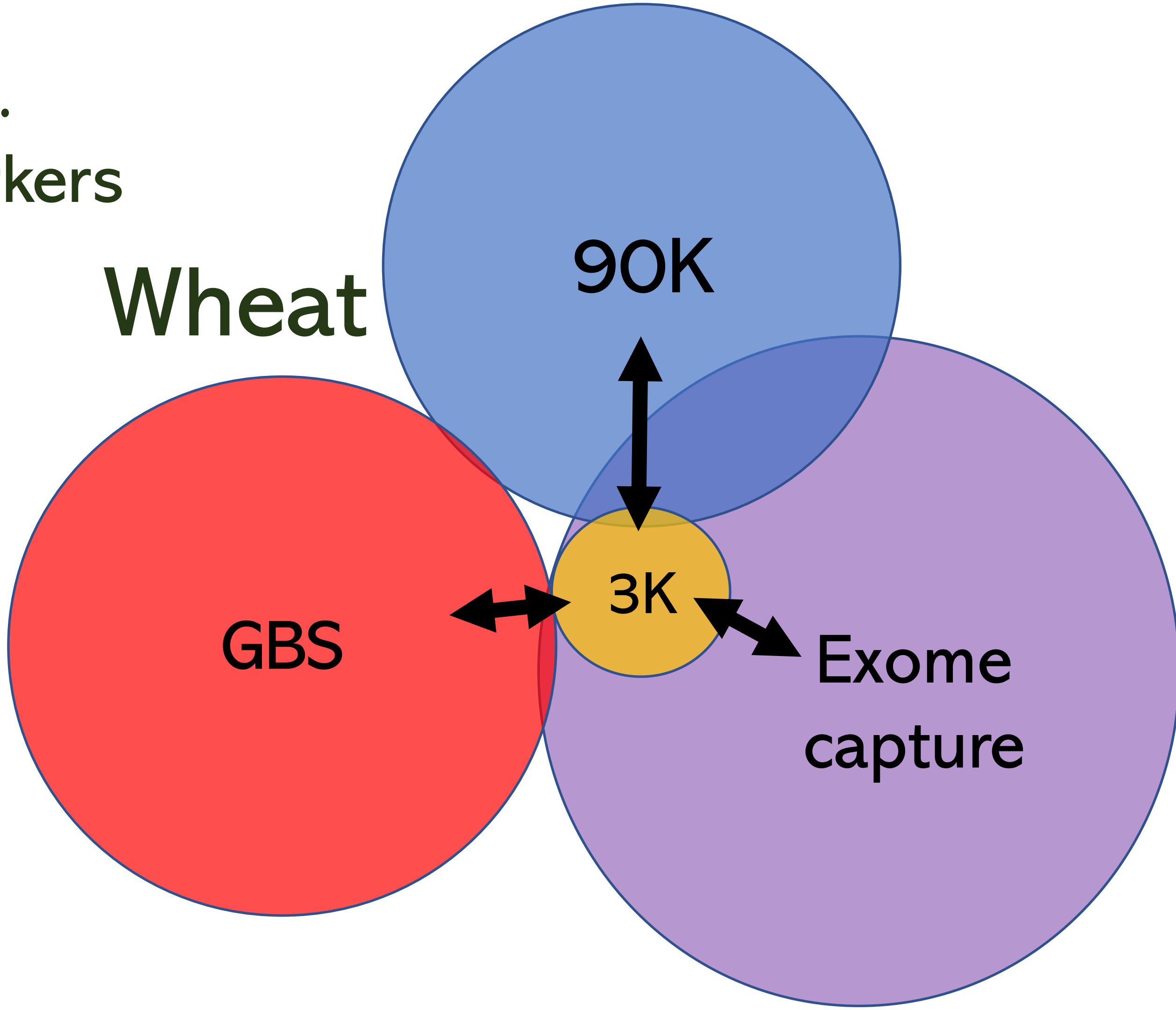
Barley



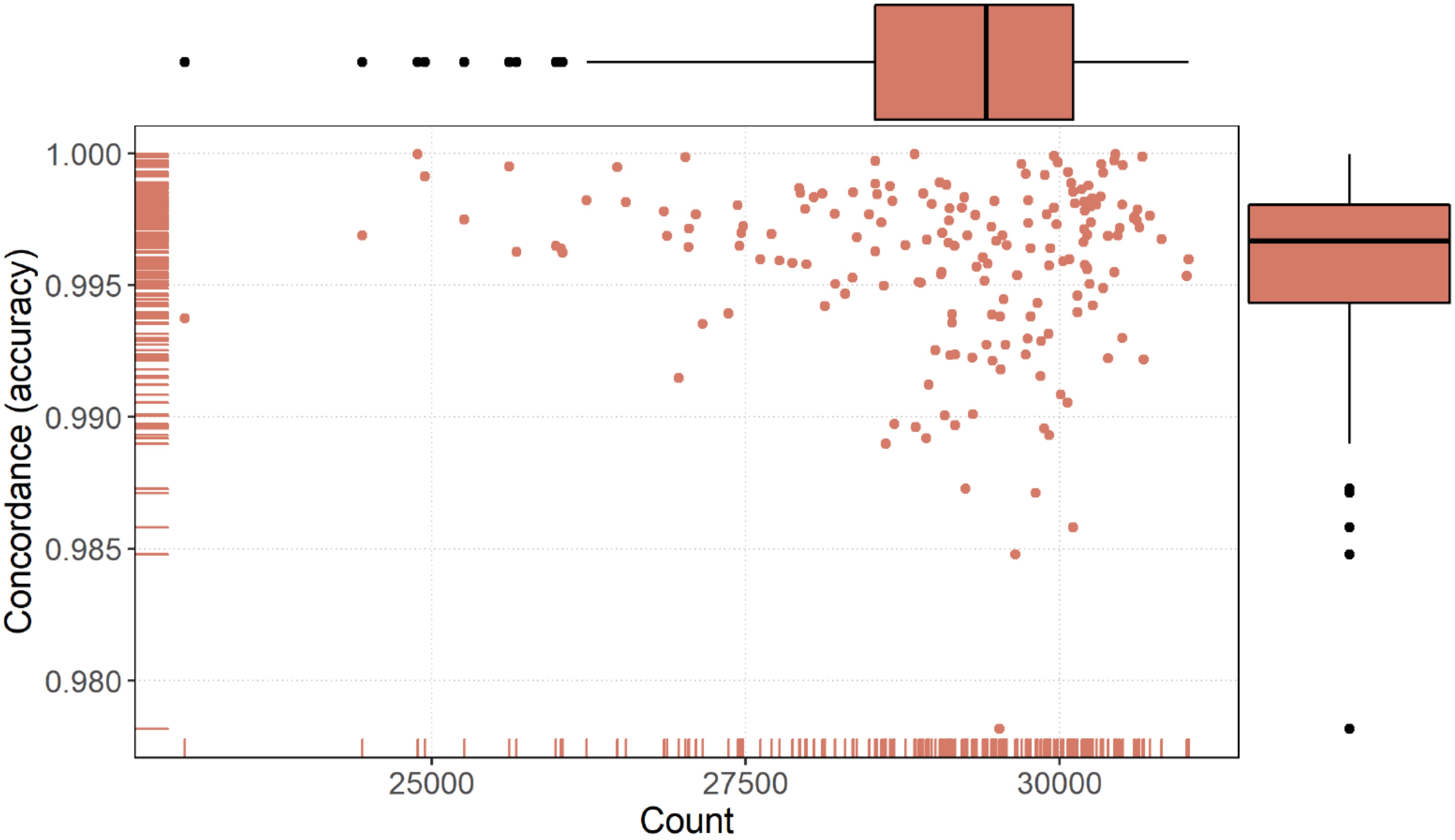
Oat



Wheat



Barley imputed 50K SNPs

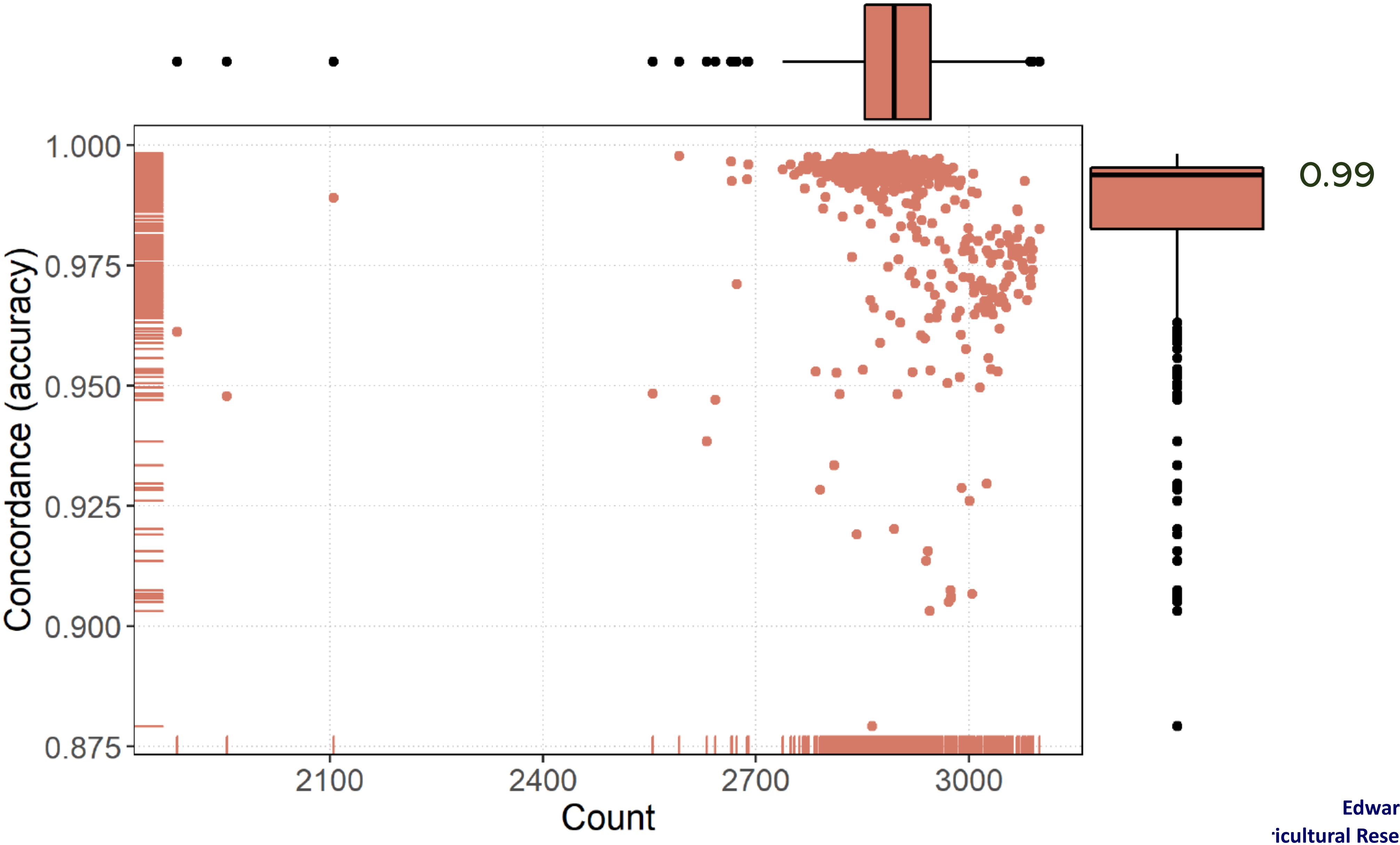


0.997

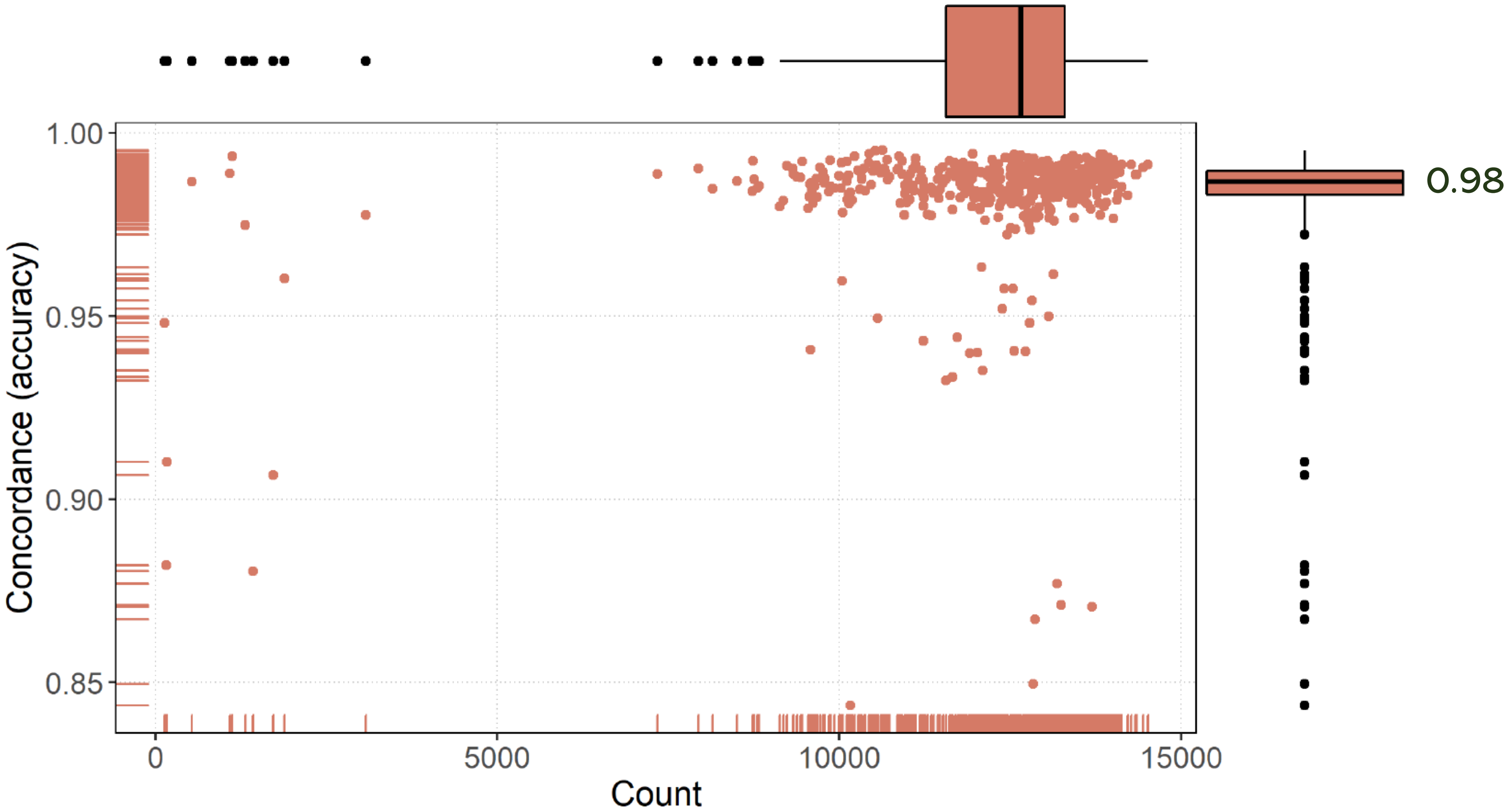
USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Oat imputed 6K SNPs



Oat imputed GBS SNPs



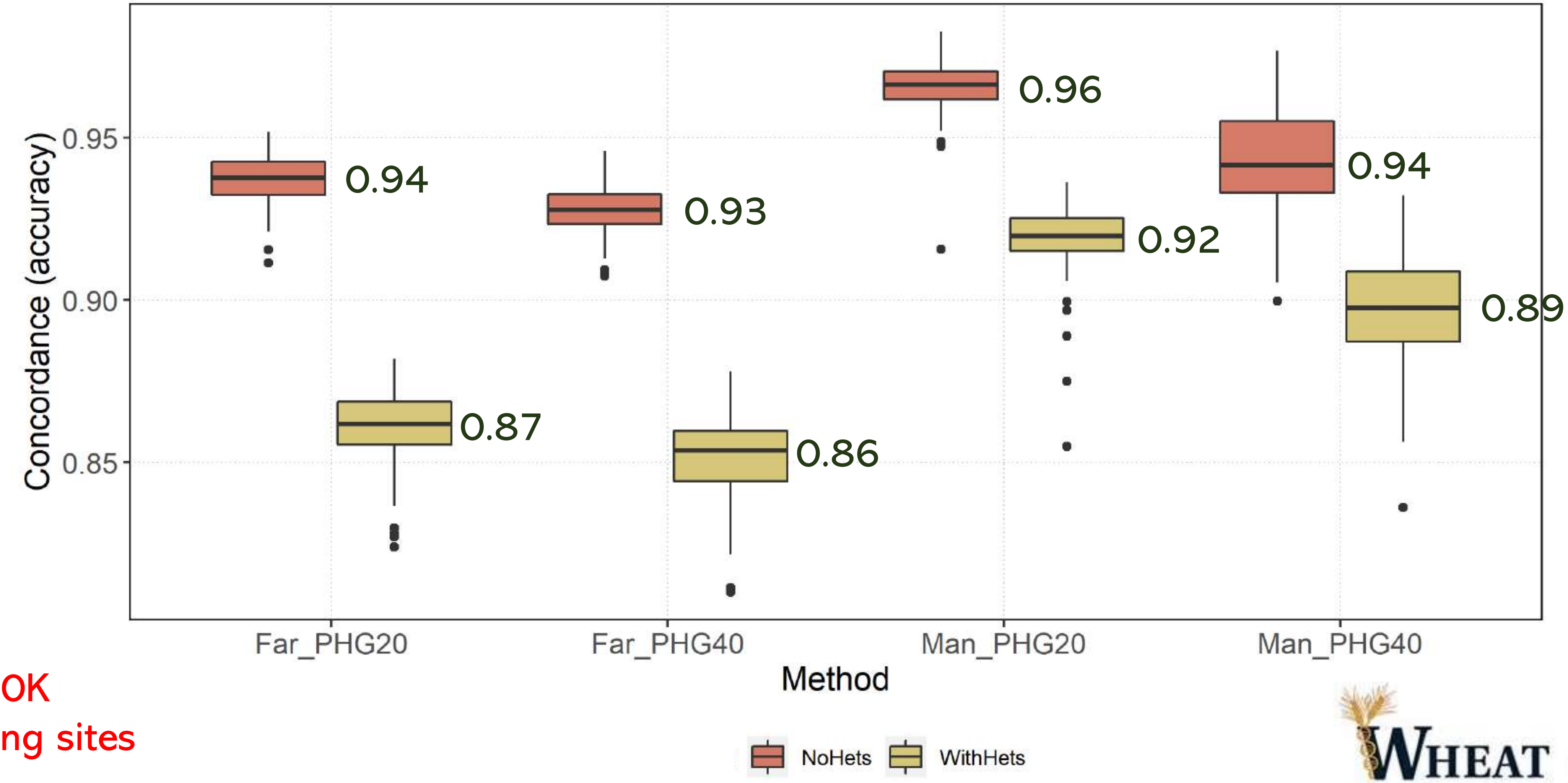
USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Average concordance of imputed wheat data sets

Pop	Imputed platform	Concordance No hets	Count No hets	Concordance With hets	Count With hets
HRSW cultivars/ breeding	exome capture (maf1)	0.957	88,569	0.900	94,293
HRSW cultivars/ breeding	exome capture (top 90%)	0.980	80,032	0.925	84,855
MSU breeding	GBS (maf1)	0.954	14,603	0.836	16,725
UMN breeding	GBS (maf1)	0.960	3,040	0.718	4,064
HRSW cultivars/ NDSU breeding	90K (maf1)	0.931	22,593	0.625	33,776
HRSW cultivars/ NDSU breeding	90K (top 80%)	0.981	18,976	0.695	27,017

Imputation of HRSW with a Practical Haplotype Graph



~300K
segregating sites

Wheat CAP
Coordinated Agricultural Project

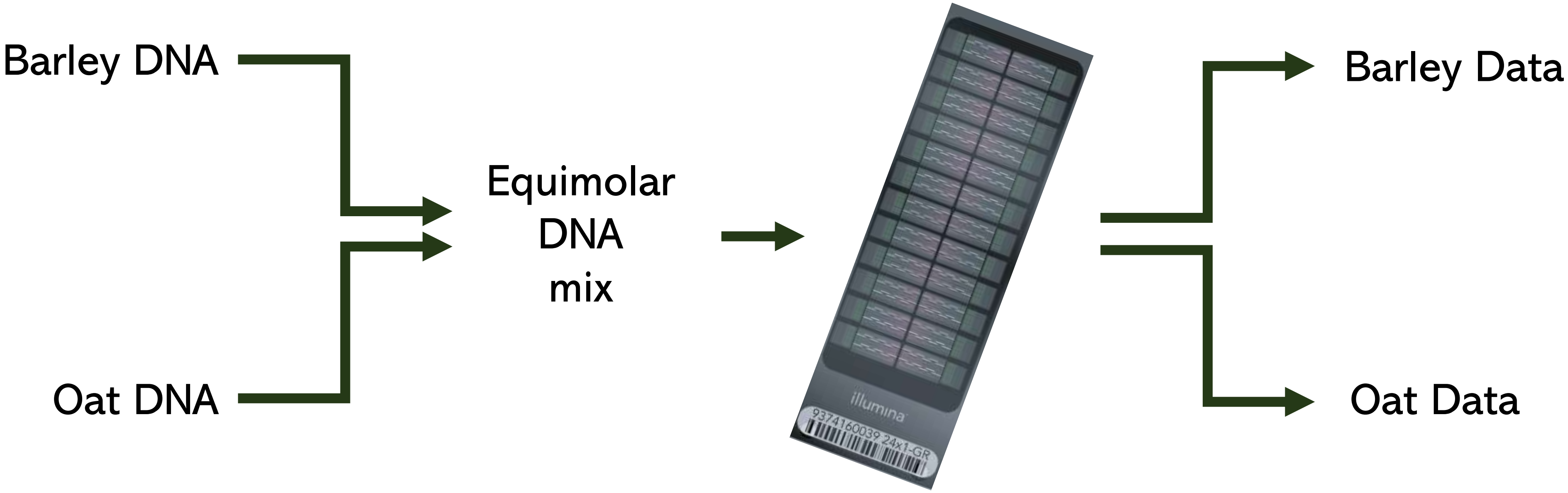
USDA
United States Department of Agriculture
Agricultural Research Service

Bradbury, et al 2022, Bioinformatics
www.maizegenetics.net/

Katie Jordan, Clay Birkett
wheat.triticeaetoolbox.org

Edward T. Schafer
Agricultural Research Center

Does dual-hybridization work?



Barley DNA

Oat DNA

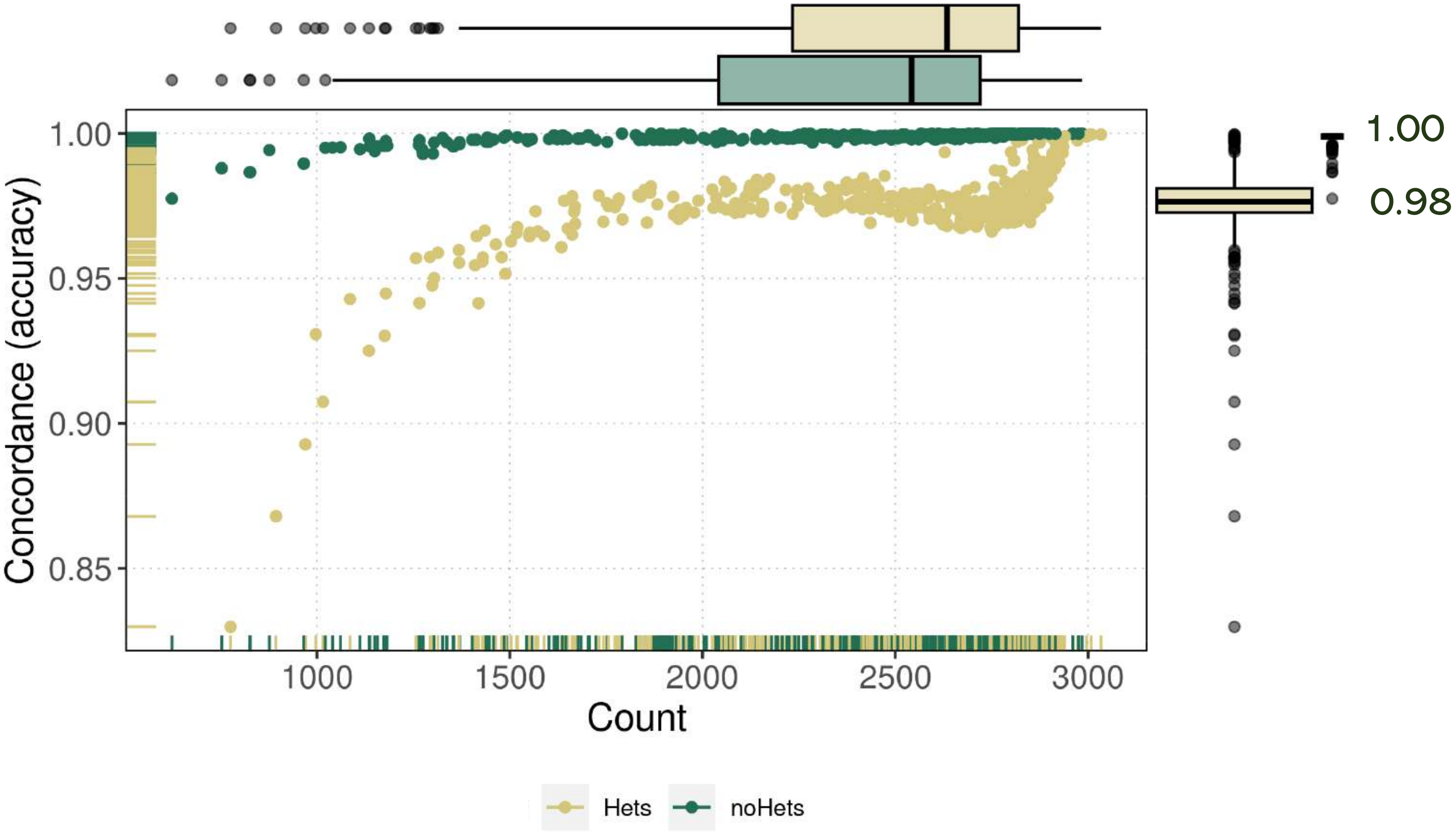
Equimolar
DNA
mix

Barley Data

Oat Data

Keeble-Gagnère, et al 2021, Front. Plant. Sci.

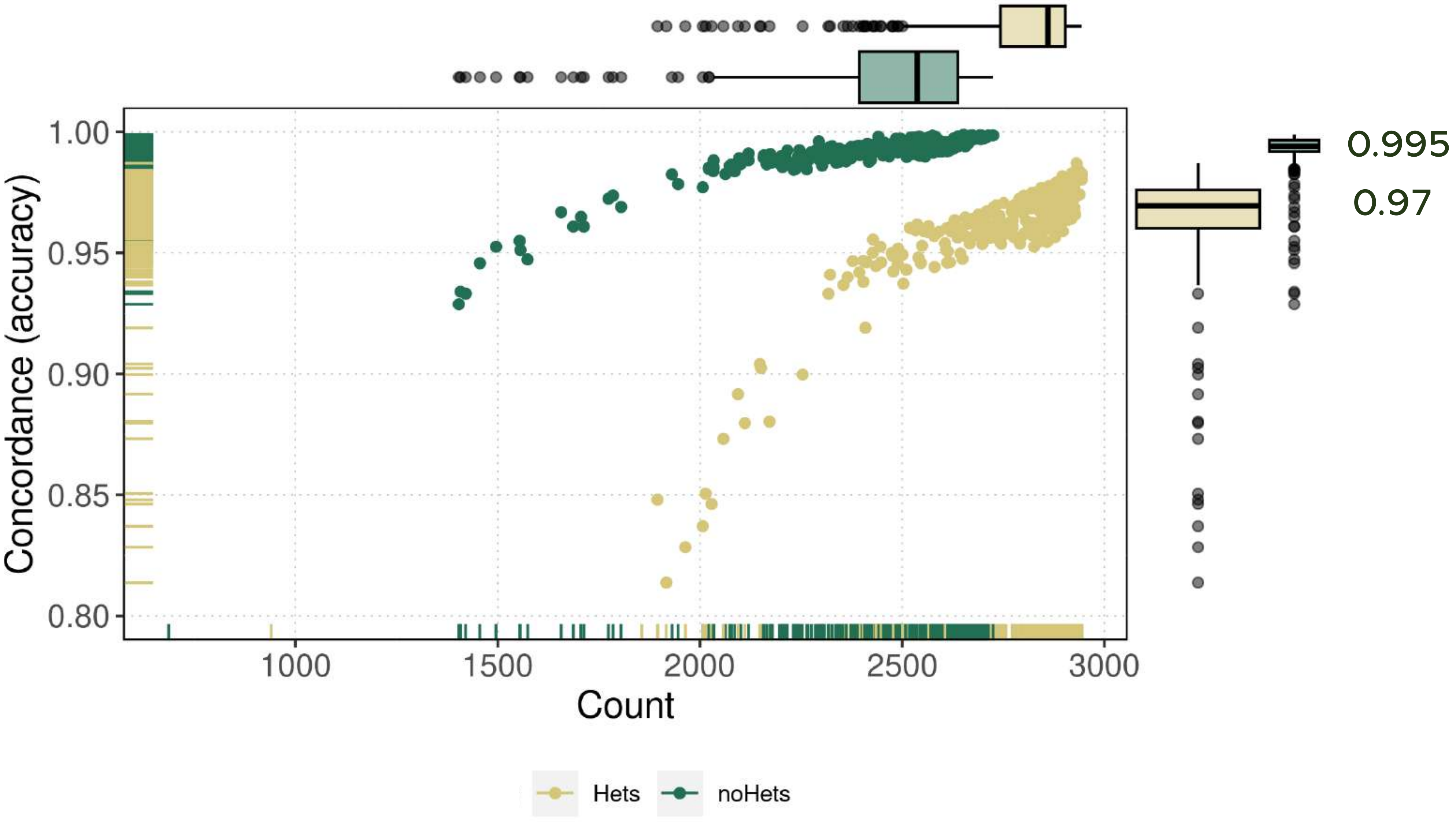
Dual hybridization works well - Barley



USDA
United States Department of Agriculture
Agricultural Research Service

Edward T. Schafer
Agricultural Research Center

Dual-Hyb in Oat – missing data predicts accuracy

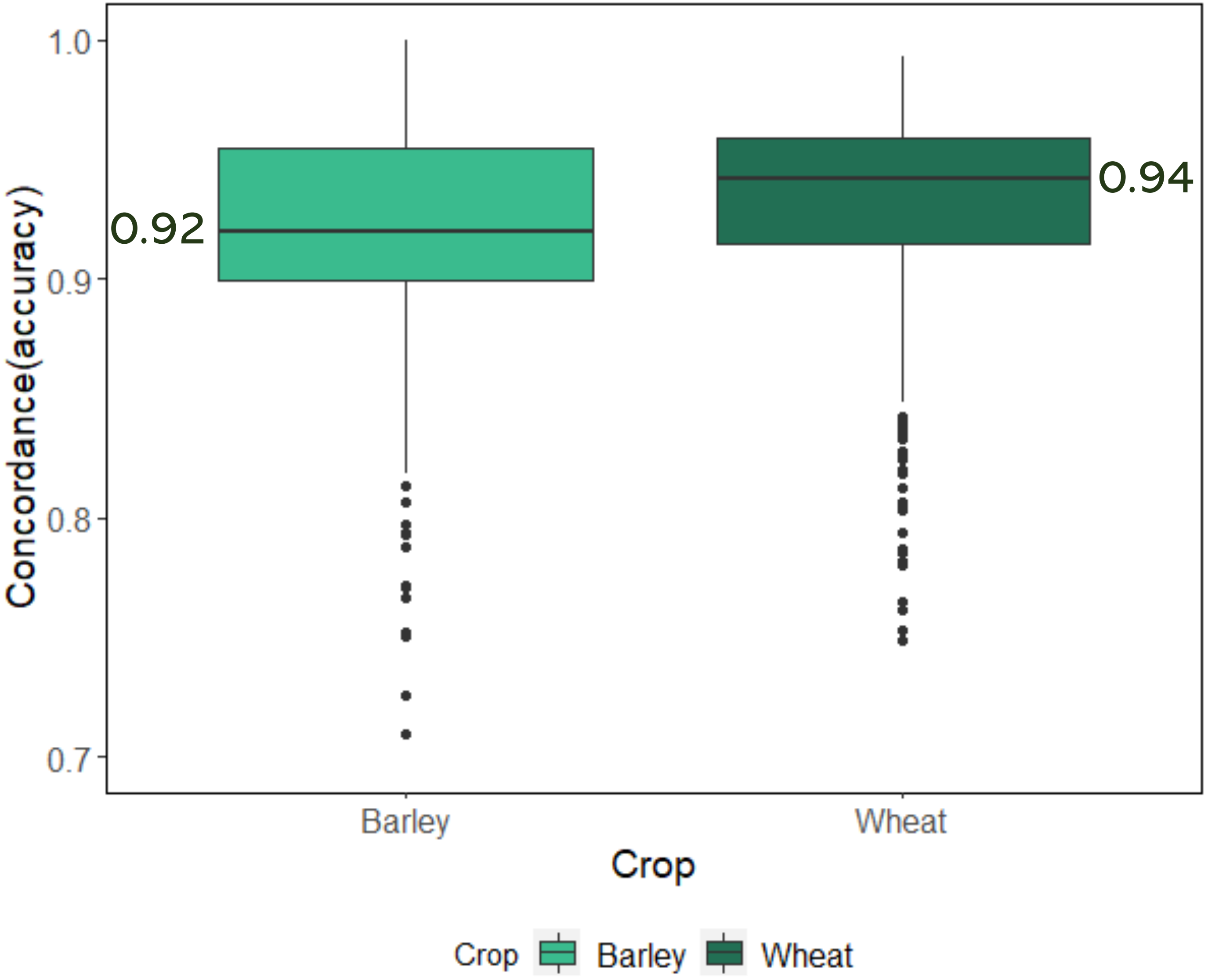


0.995

0.97

Edward T. Schafer
Agricultural Research Center

Wheat/Barley Dual-Hyb on AgriBio 40K array

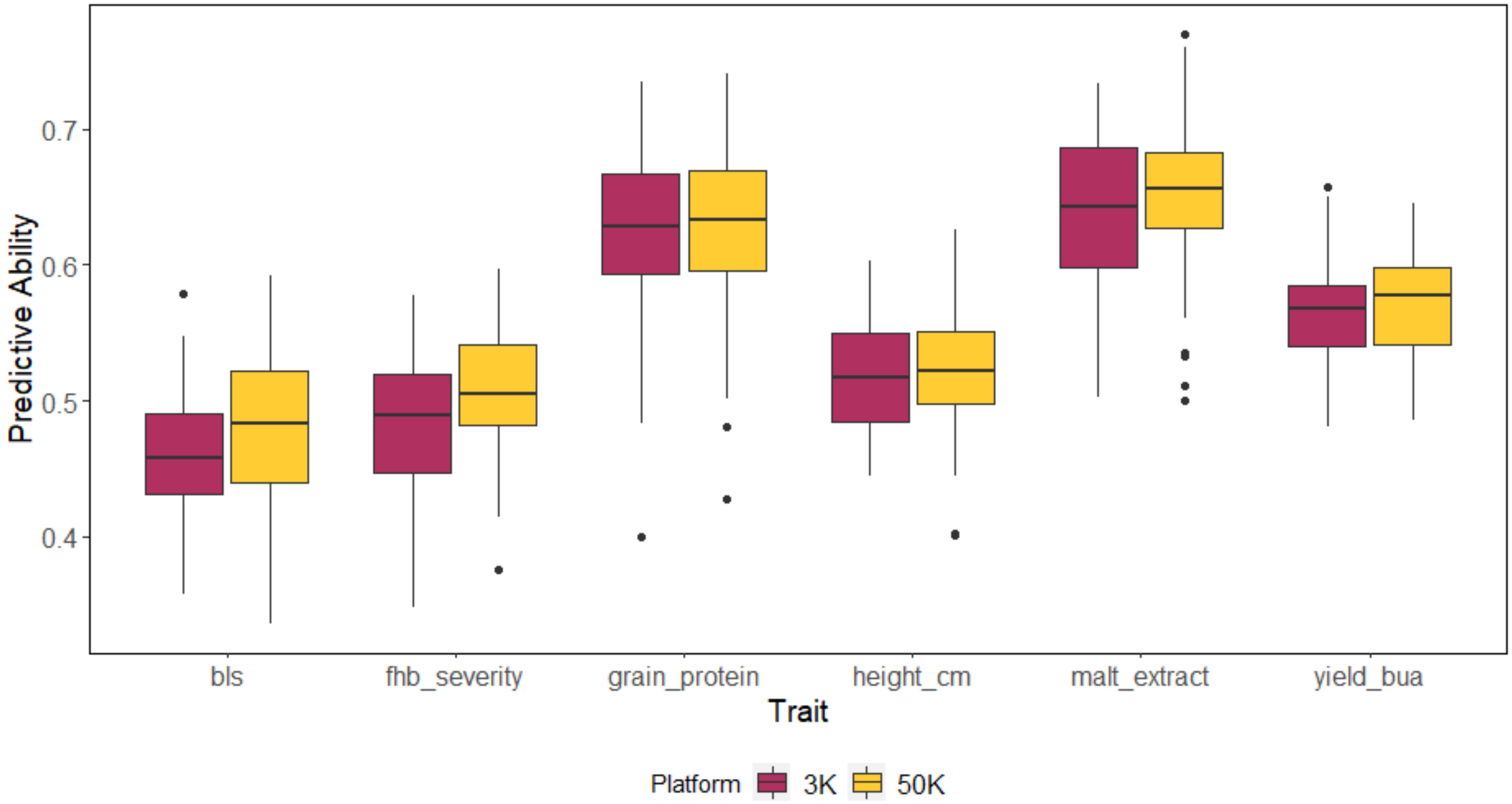


Is the array useful?

- Fingerprinting / QC
- Genomic Selection
- Genetic diversity studies
    - International Oat Nursery

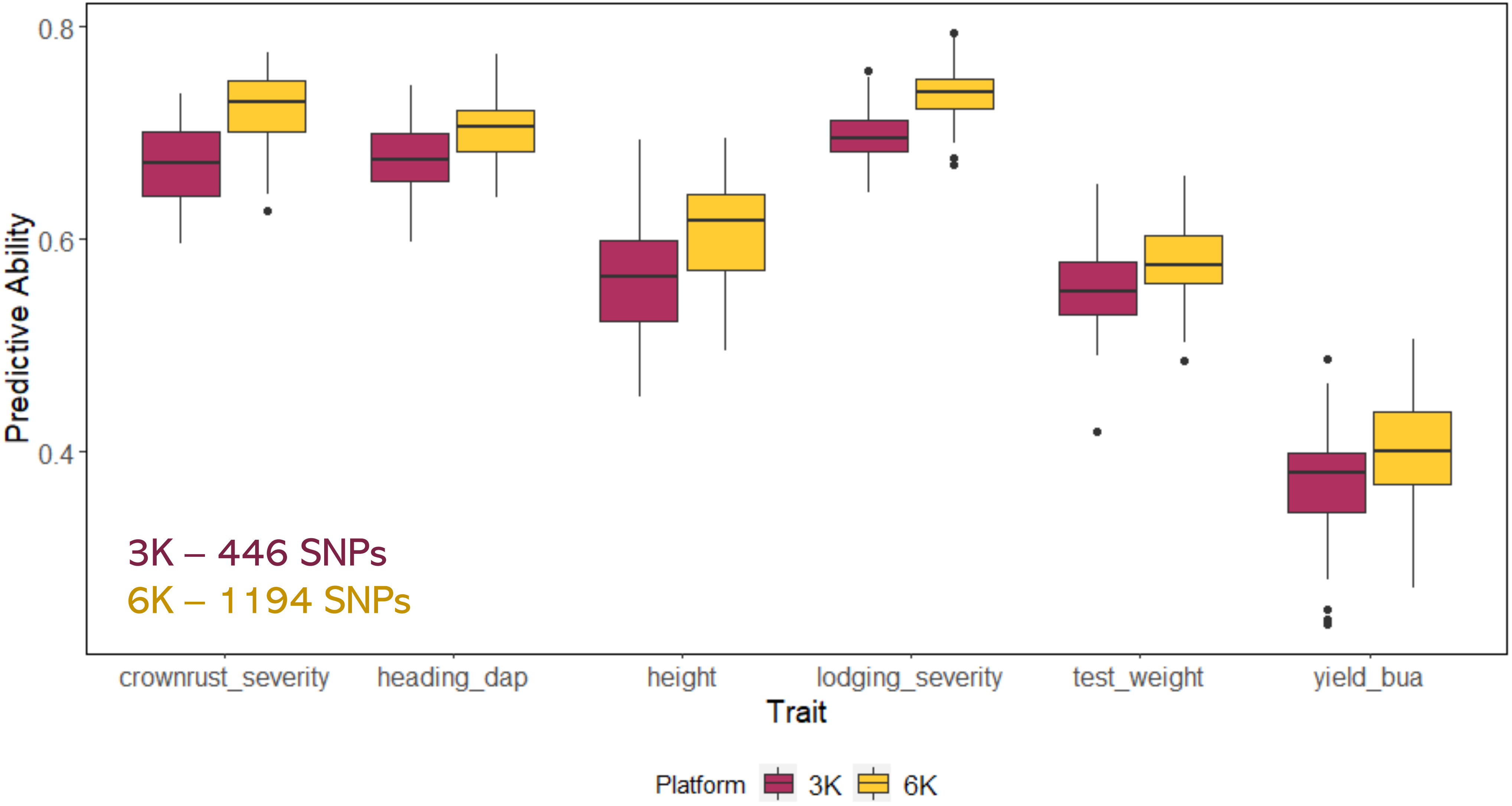
United States Department of Agriculture
Agricultural Research Service

Accuracy of GS is similar between platforms - barley



Karen Beaubien & Kevin Smith, UMN

Accuracy of GS is similar between platforms - oat



3K – 446 SNPs

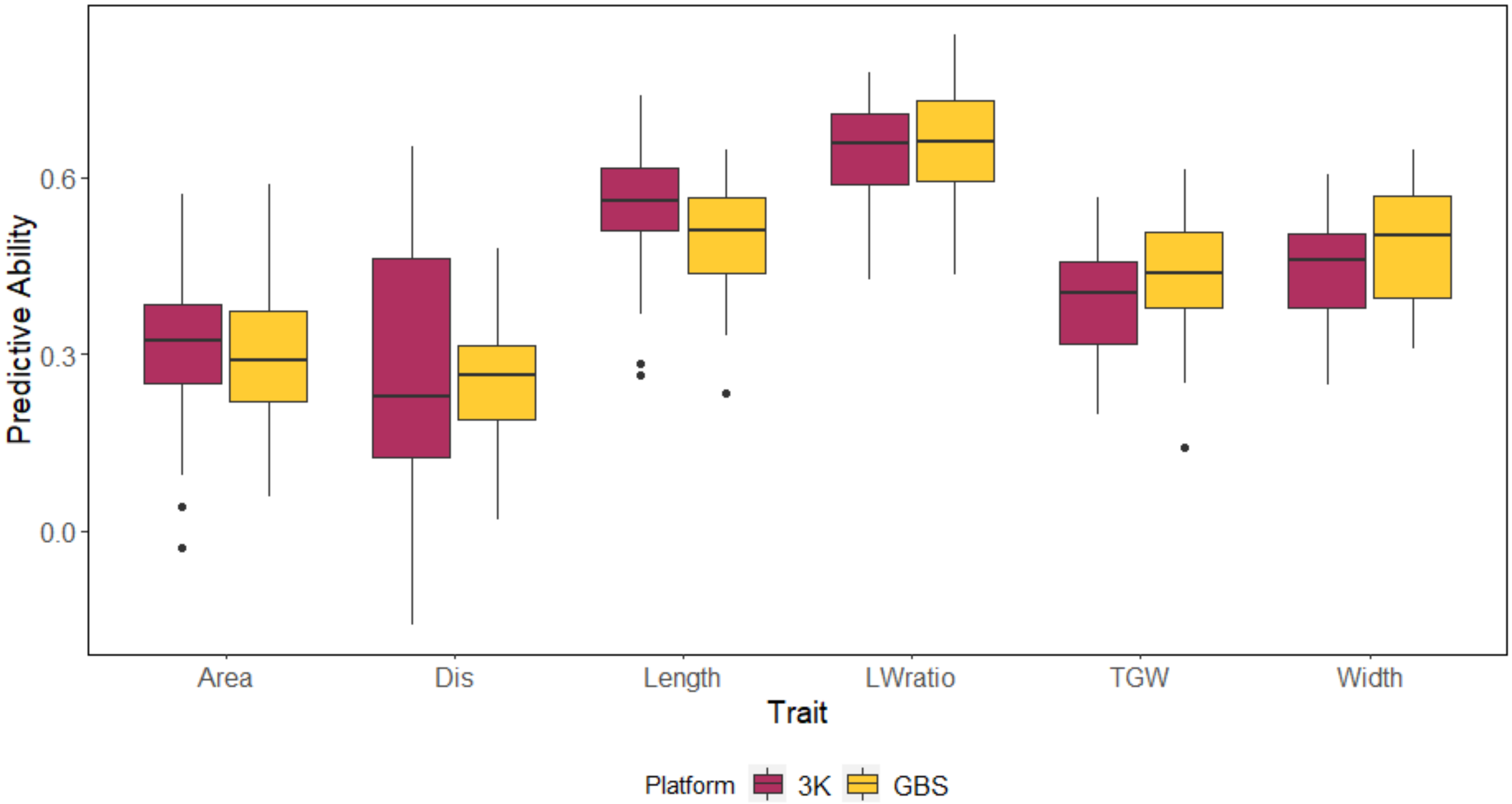
6K – 1194 SNPs

United States Department of Agriculture
Agricultural Research Service

Karen Beaubien & Kevin Smith, UMN

Edward T. Schafer
Agricultural Research Center

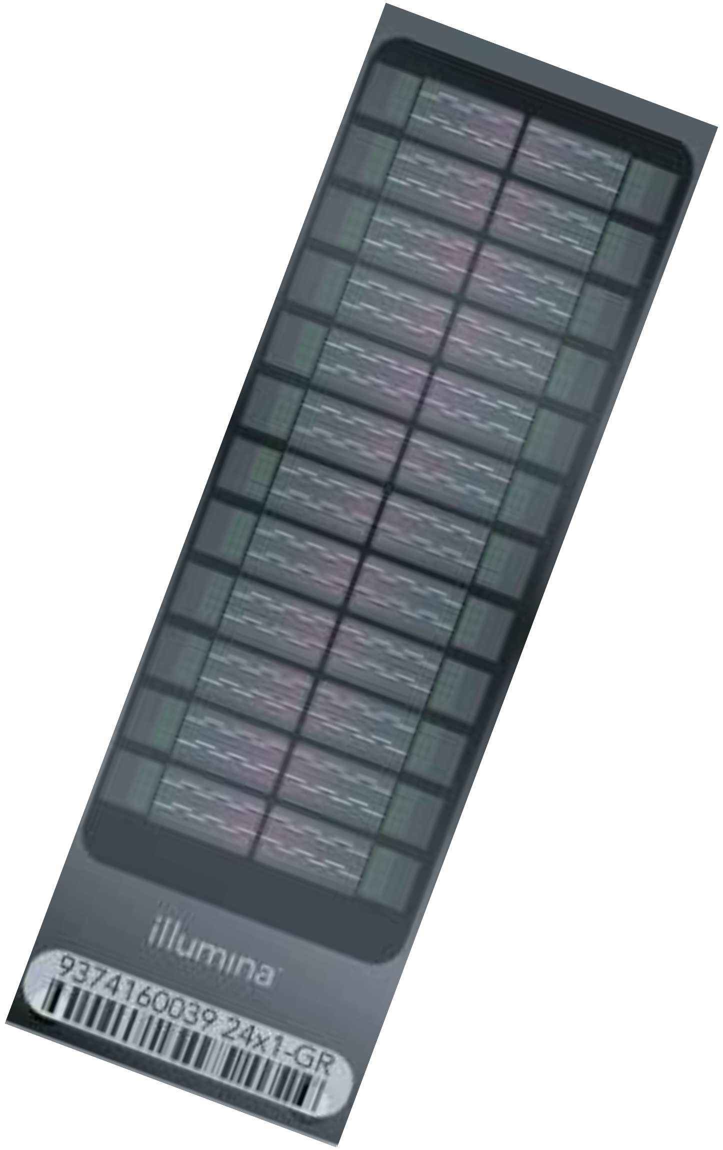
Accuracy of GS is similar between platforms - wheat



Emily Conley & Jim Anderson, UMN

Conclusions

- The USDA-3K array works quite well.
  - Fast, durable, and robust (low risk).
  - Accurately calls the source SNPs.
- Imputation can increase the # of SNPs.
  - Avoid re-genotyping all training germplasm.
  - Wheat Practical Haplotype Graph is ready to be used through T3 now.
- Dual-Hyb can immediately cut the price in half.
  - Verifying that dual mode works with wheat.
  - Tri-mode next
- GS models are similar to higher-density counterparts.



Acknowledgements

USDA-ARS

- Mary Osenga
- Qijian Song
- Rajeev Gupta
- Gina Brown-Guedira
- Deven See
- Guihua Bai
- Bikash Poudel
- Katie Jordan
- Kathy Klos
- Raj Nandety

- Mandy Waters, PepsiCO
- Eduard Akhunov, KSU
- Brian Ward, Land O'Lakes
- Small Grains Community
- Karen Beaubien, UMN
- Kevin Smith, UMN
- Emily Conley, UMN
- Jim Anderson, UMN
- Illumina
  - Matt Cerezo
  - Kahlil Lawless
  - Yuting Bai
  - Brenda LeCuyer

NORTH AMERICAN MILLERS' ASSOCIATION

U.S. Wheat & Barley Scab Initiative

WHEAT CAP
Coordinated Agricultural Project

USDA
United States Department of Agriculture
Agricultural Research Service

Jason.fiedler@usda.gov

USDA-ARS CRIS 3060-2100-038

Edward T. Schafer
Agricultural Research Center