Shopping in the diversity supermarket; using big data to bring diversity to breeding

Sarah Hearne – CIMMYT
International Maize and Wheat Improvement Centre
Big Data in Agriculture
14th May 2018

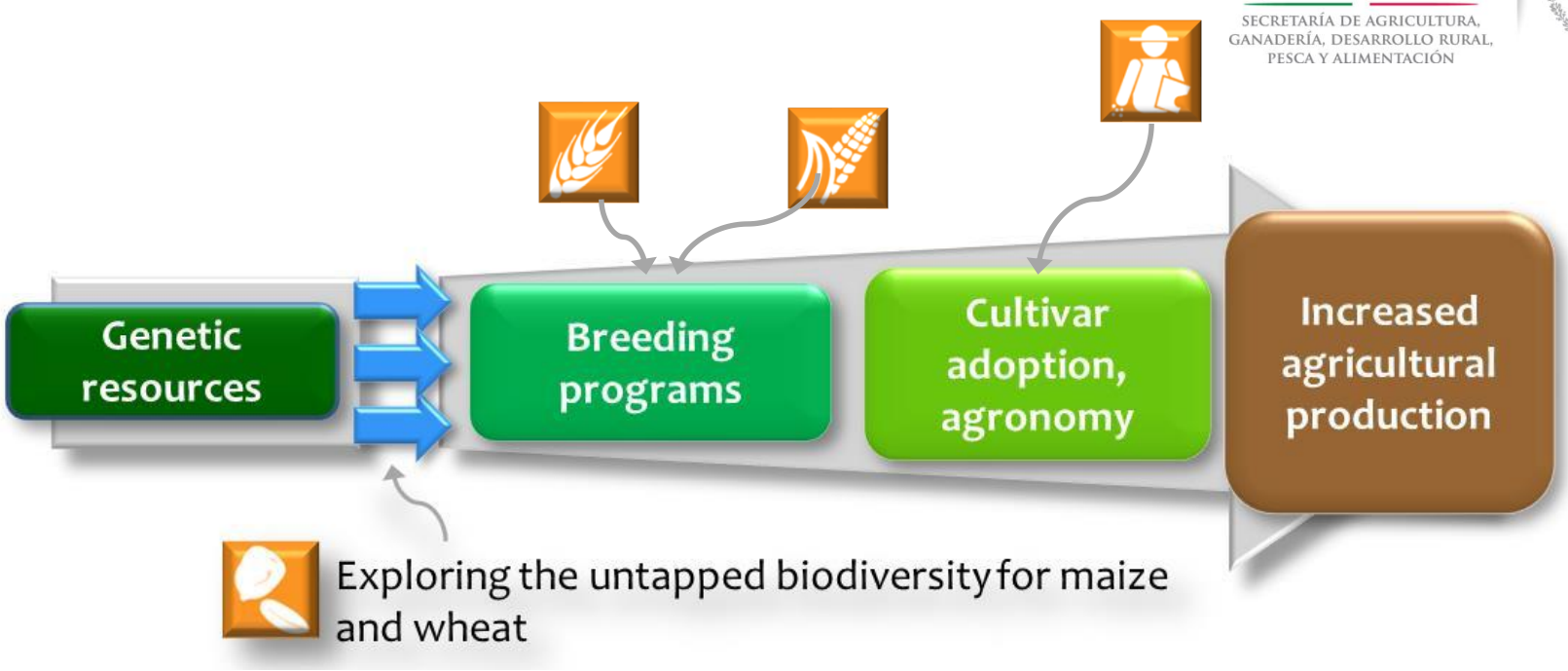
s.hearne@cgiar.org ; cimmyt-mab-seed@cgiar.org

CIMMYT

Seeds of Discovery (SeeD) 2010 (MasAgro Biodiversidad)

MasAgro | Modernización Sustentable de la Agricultura Tradicional

SAGARPA
SECRETARÍA DE AGRICULTURA, GANADERÍA, DESARROLLO RURAL, PESCA Y ALIMENTACIÓN



Genotypic characterization of germplasm banks and public elite germplasm

Phenotyping and Marker-trait associations

Pre-breeding

CIMMYT

Diversity, diversity everywhere but not a drop to drink

Inbred donors and elite lines- basis of breeding in public and private sector

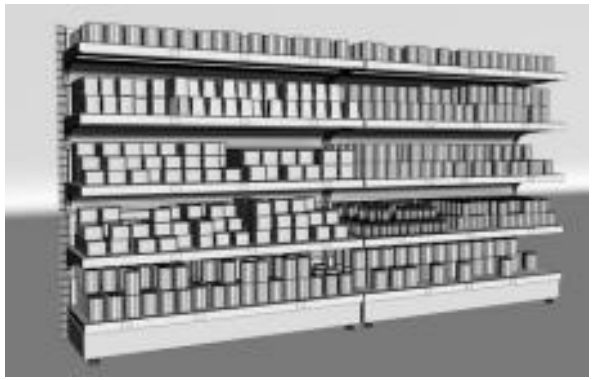
CIMMYT

Barriers to use

Adaptation, linkage drag

CIMMYT

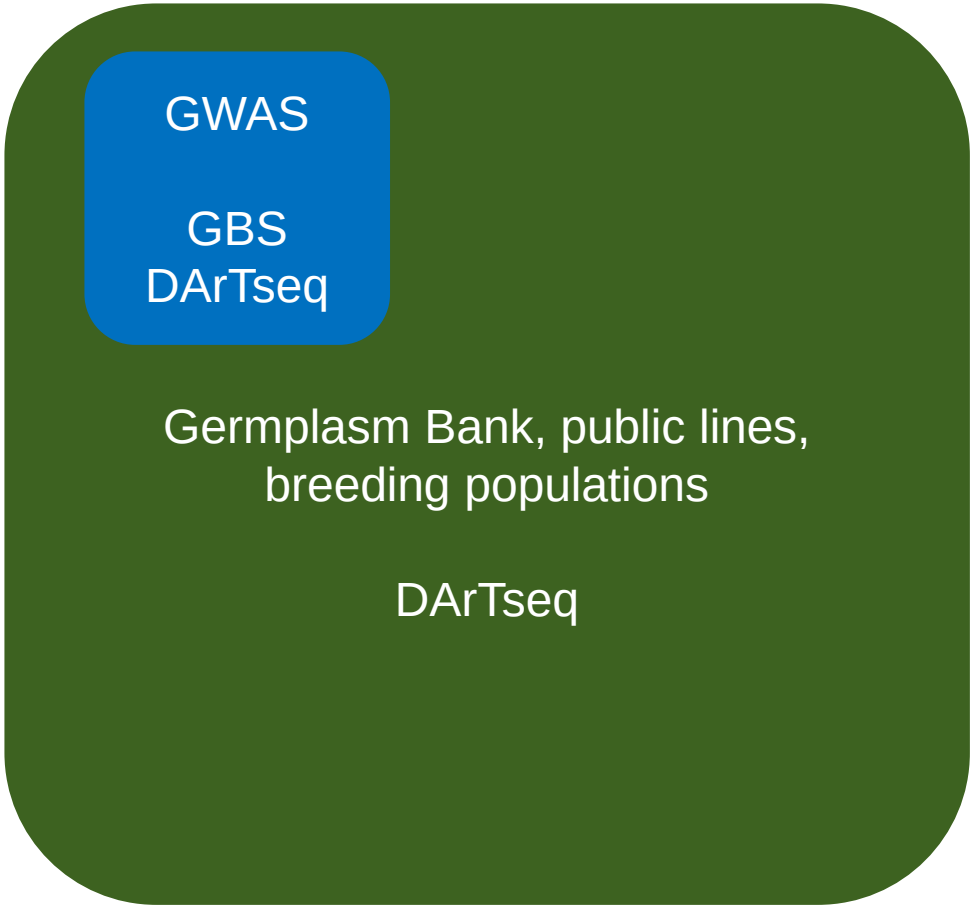
Barriers to use



Where does Big Data fit?

CIMMYT

Genotyping of germplasm



GBS
1M ref genome aligned SNP
Imputation
QC filtered to ~350k

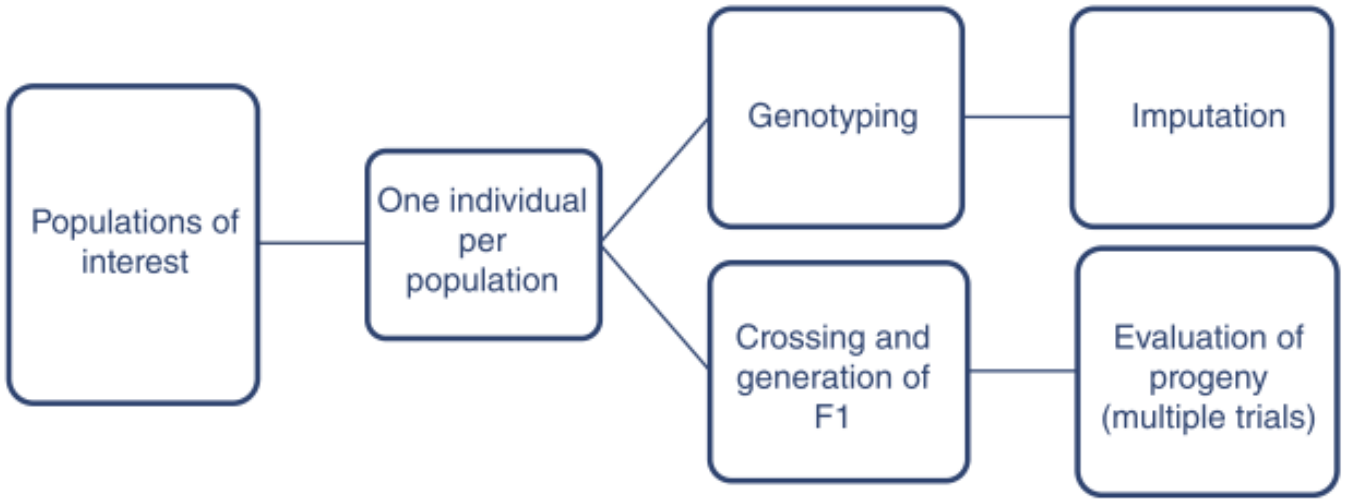
DArT seq
GWAS:- 110k SNP
~50%mapped to ref genome
No imputation

Bank: 1M markers, 650k SNP
filtered to 350-470k SNP (5x)
Allele frequencies
~45% mapped to ref genome

CIMMYT

Marker trait associations - GWAS

F1 association mapping (FOAM)



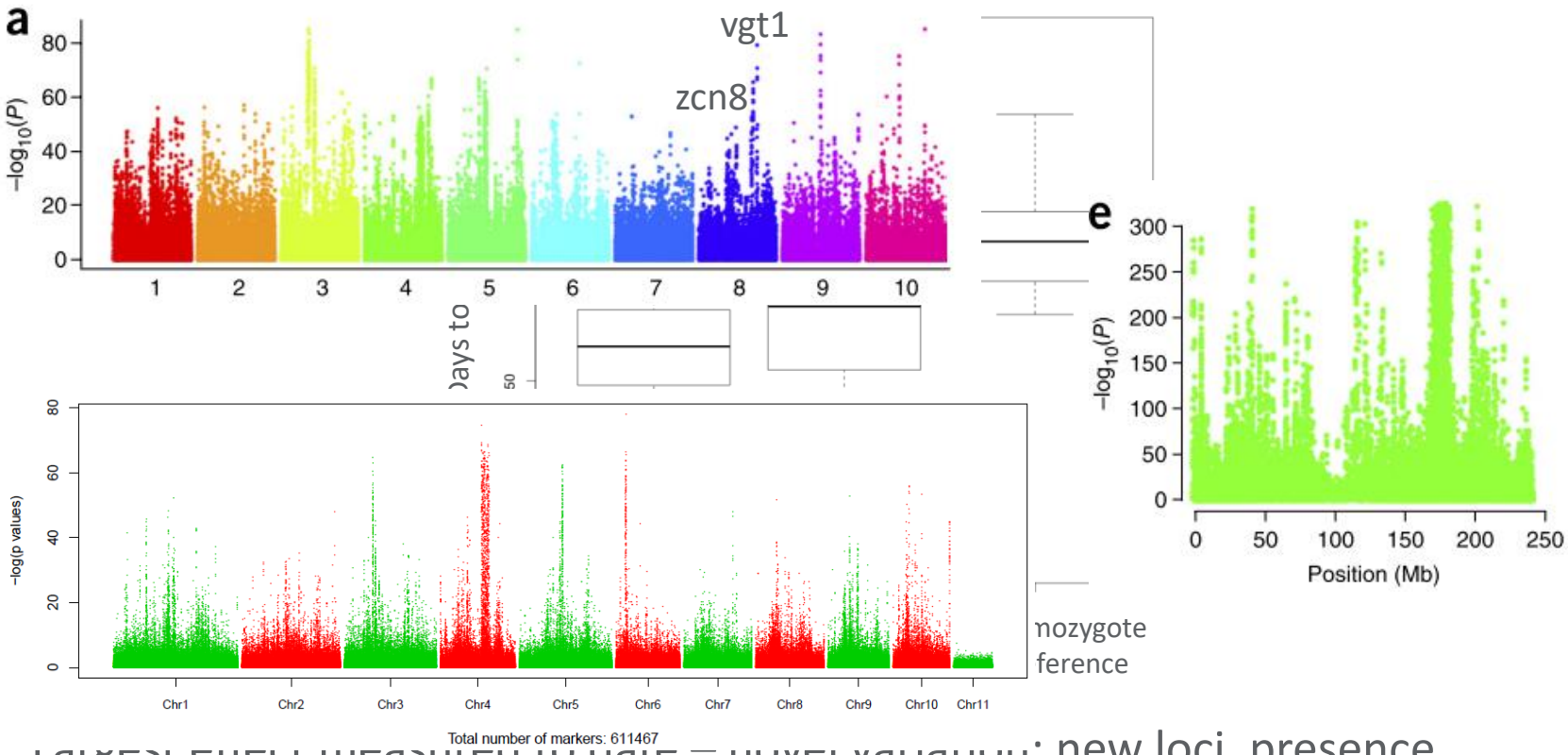
3500 accessions, flowering,plant & ear height…..

34 trials

Romero et al 2017, Nature Genetics

CIMMYT

GWAS - flowering

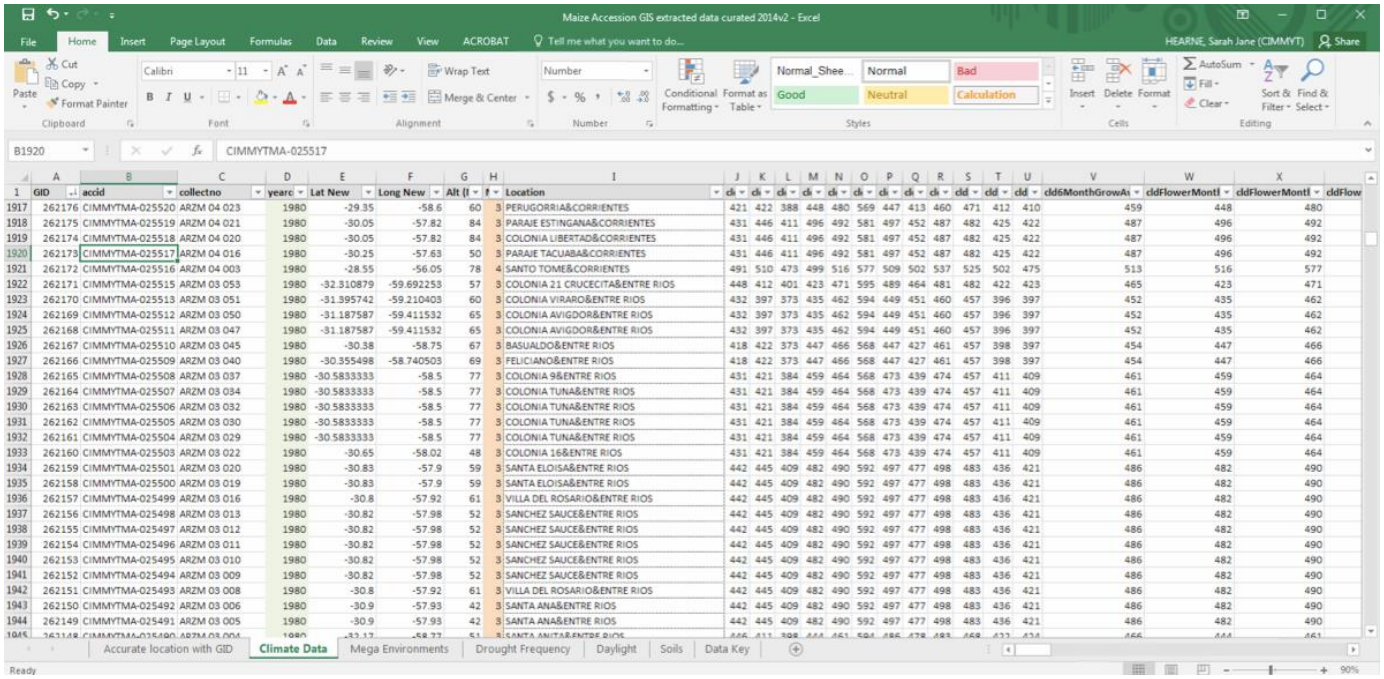


Largest effect measured to date – novel variation; new loci, presence in elite tropical material (CMLs) varies– ~1000 genes

Romero et al 2017, Chen, Palacios, Willcox, Burgueno, Hearne unpublished

Extending and expanding germplasm bank data

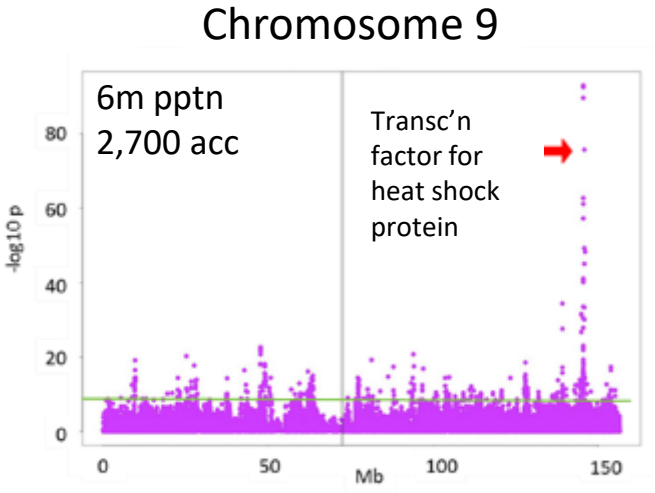
- ~17500 landrace accessions with collection site “data”
- 2y curation effort
- Derive GIS climate variables from collection sites

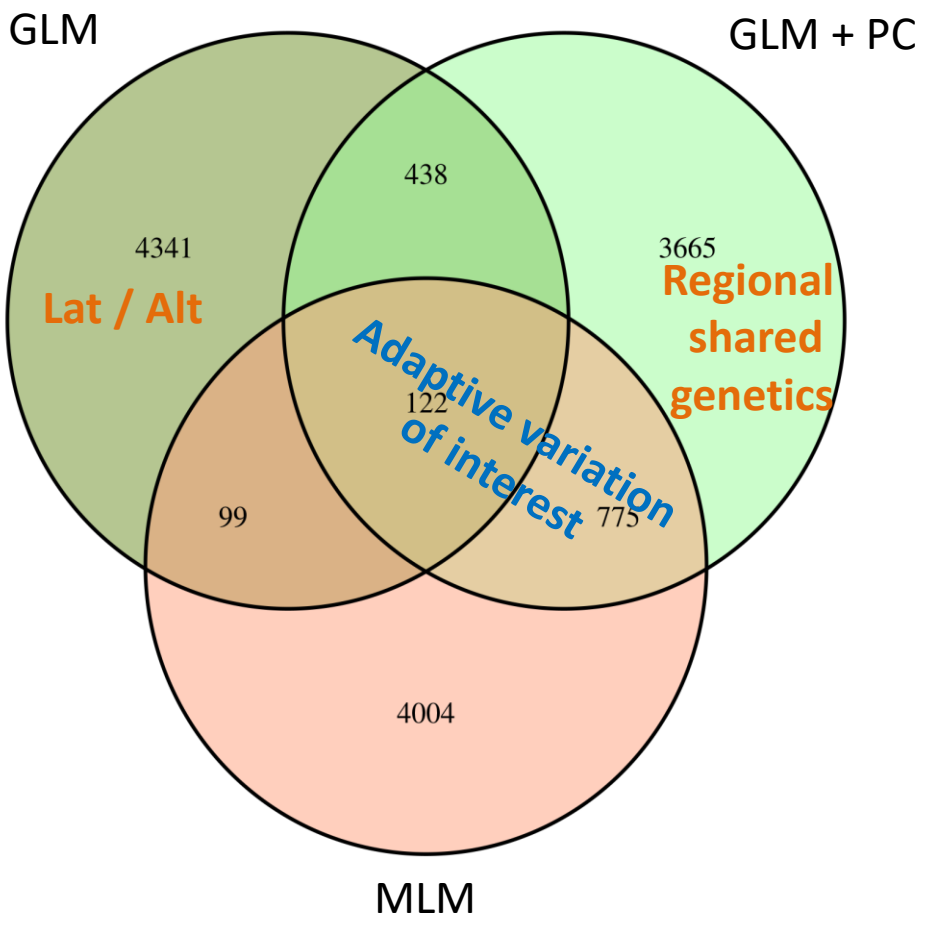


Kai Sonder

CIMMYT

Environmental GWAS



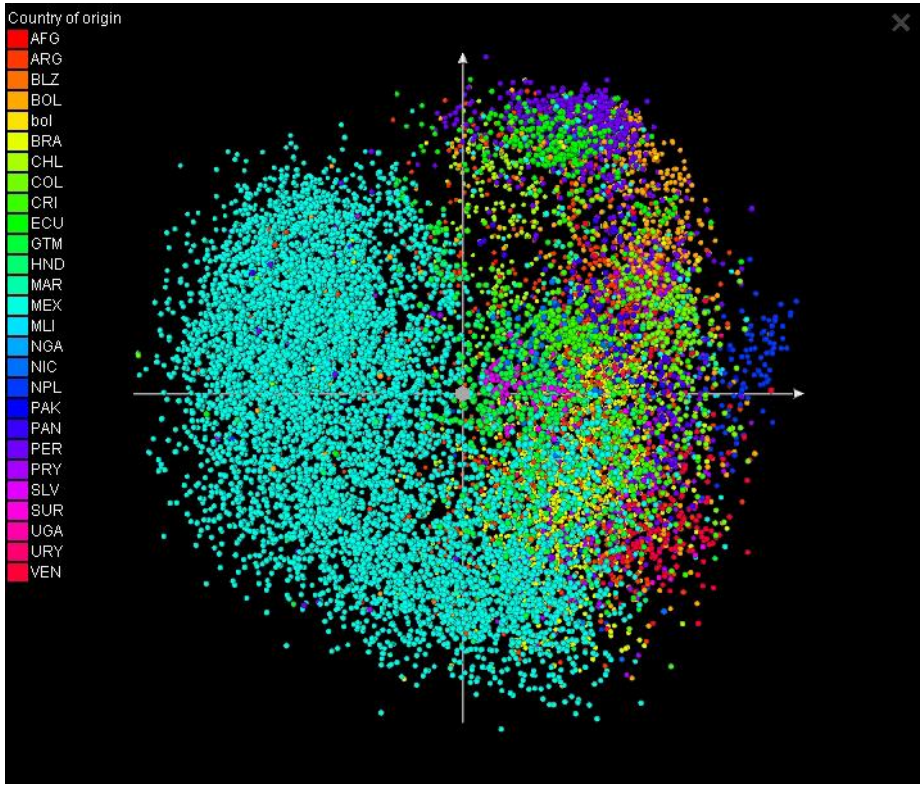


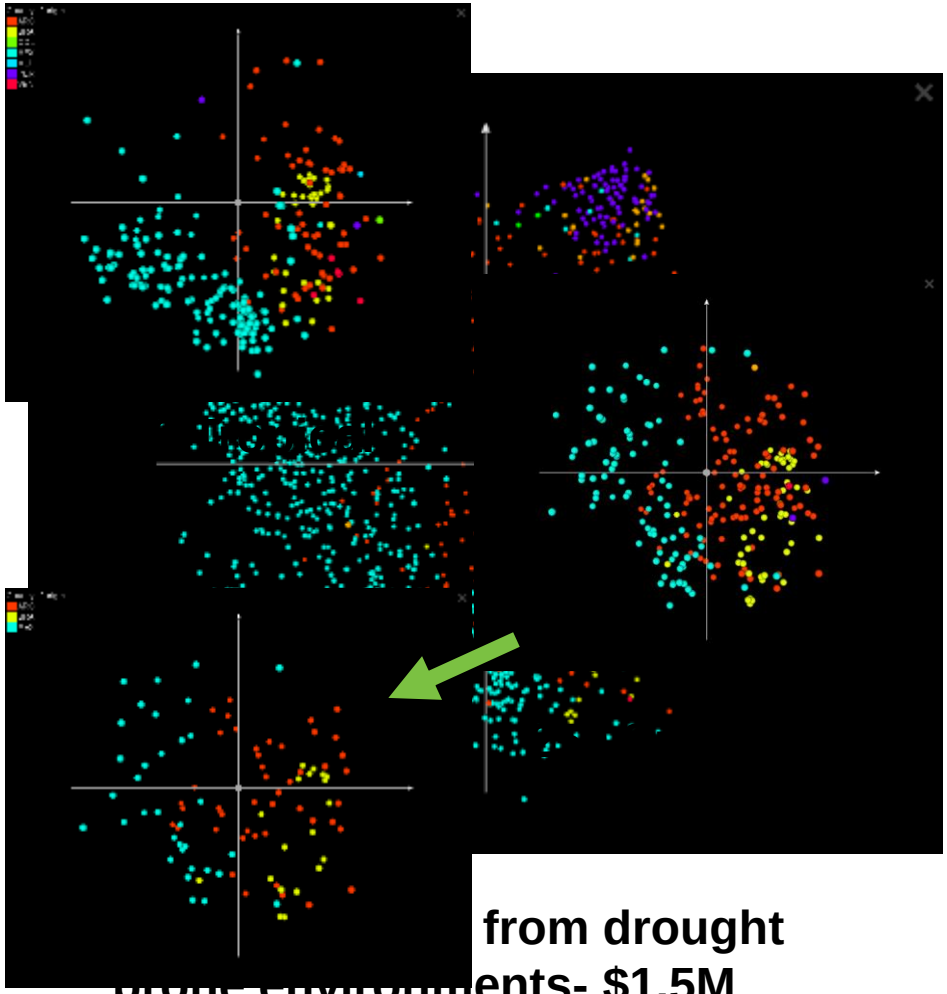
Alberto Romero, Arcadio Valdes- unpublished, HuiHui Li, Jorge Franco, Jose Crossa

CIMMYT

Molecular and environmental diversity – shift from cores to breeder panels

MDS euclidian genetic distances





15,384 landraces

from drought

prone environments- $1.5M

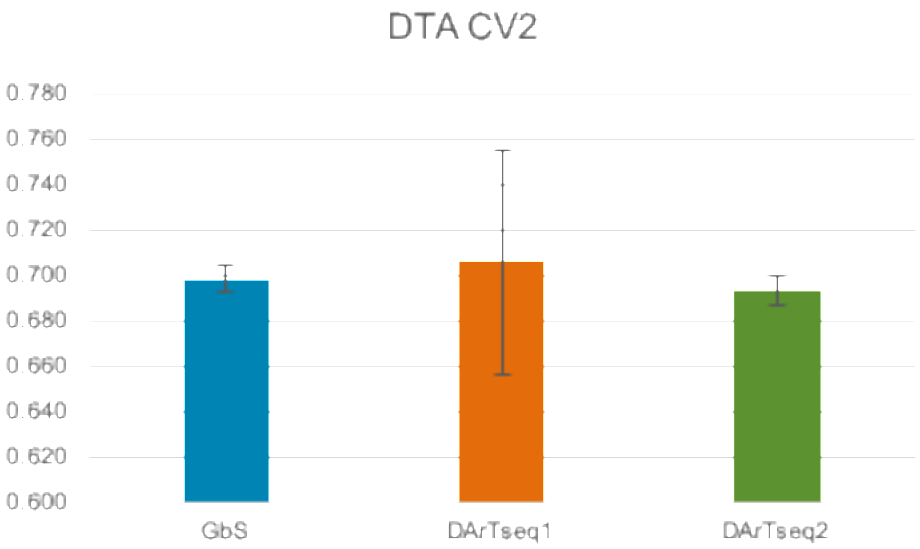
100 Sub-tropical

CIMMYT

Smarter selection?

GS with landraces

Look in GWAS panel



Marker	Before QC	After QC
GBS	(1M) 352048	335319
DArTseq1	110698	40880
DArTseq2	94644	**5027**

Gorjanc *et al* 2016 BMC Genomics

Crossa and Hearne unpublished

CIMMYT

MORE?

Phenotypes

Genotypes

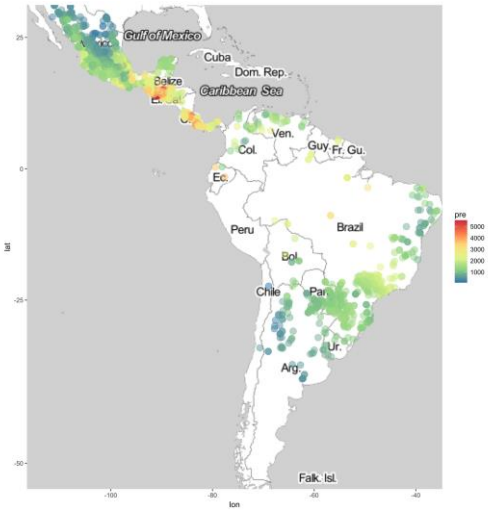
Environments

Environmental GS

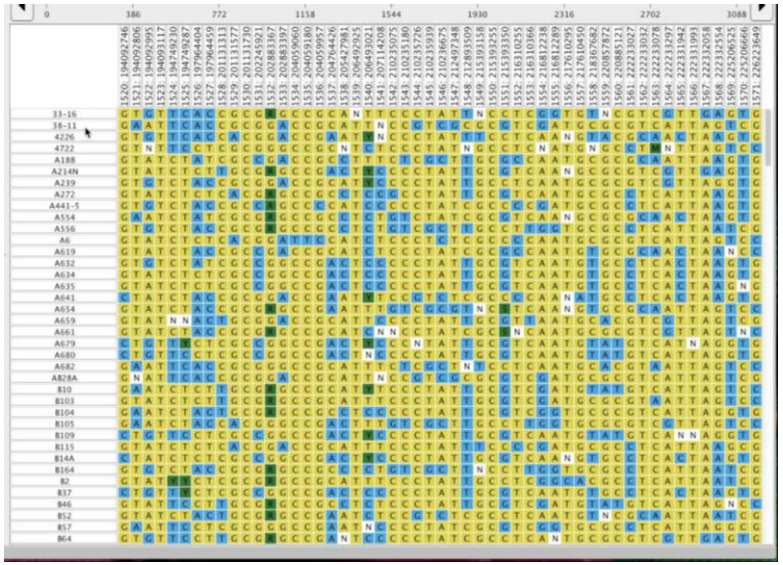
Predict landrace collection site environments?

Genotype Matrix Genotype Effects

Landrace Origin Climatic Data (precip|frost|temp|…) = $\mu + X_{Genotype} B_{Genotype} + error$



=

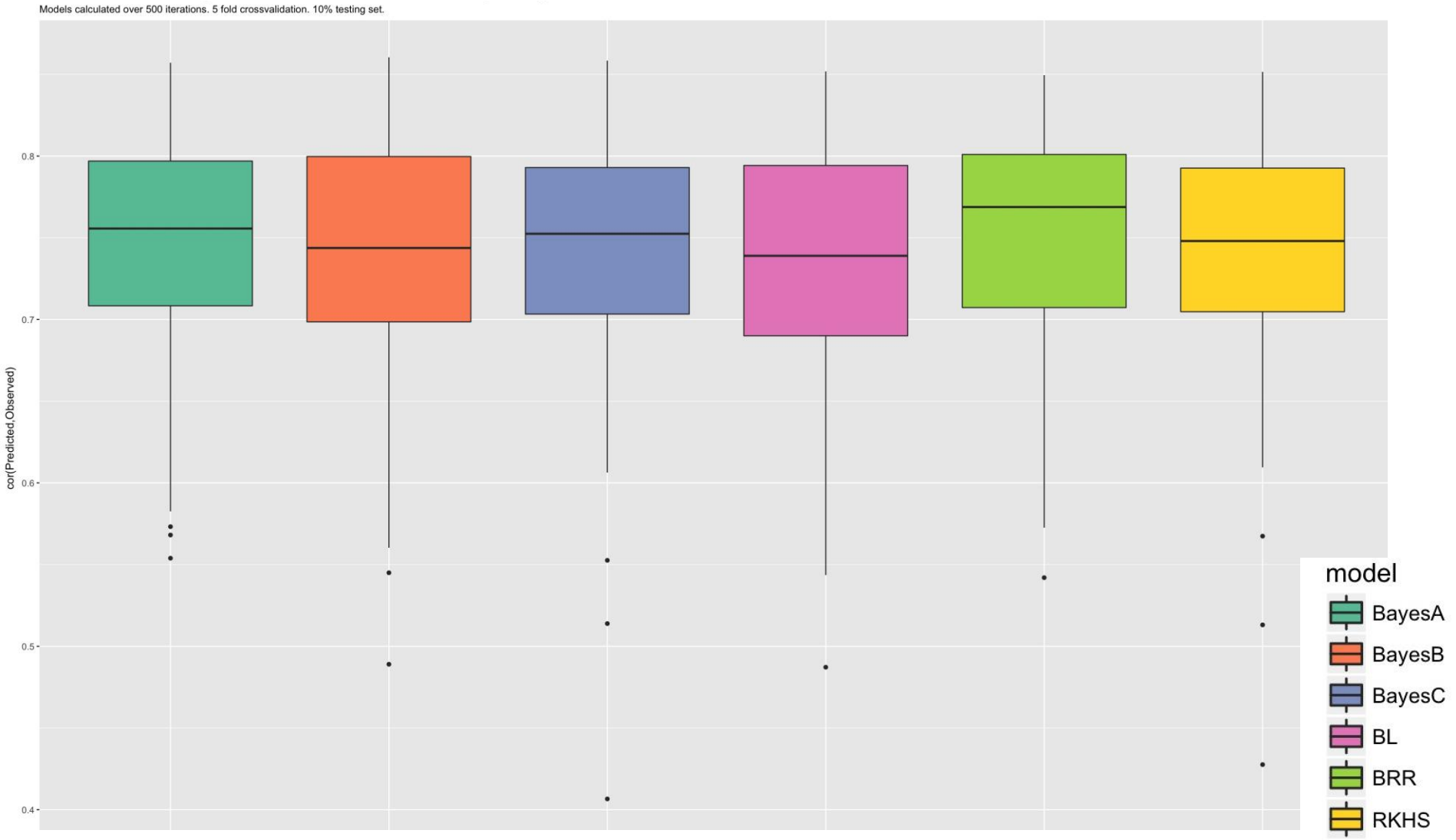


Arcadio Valdes, 2700 acc GWAS panel,120k markers (single SNP 50bp window) imputed GBS

CIMMYT

Environmental GS

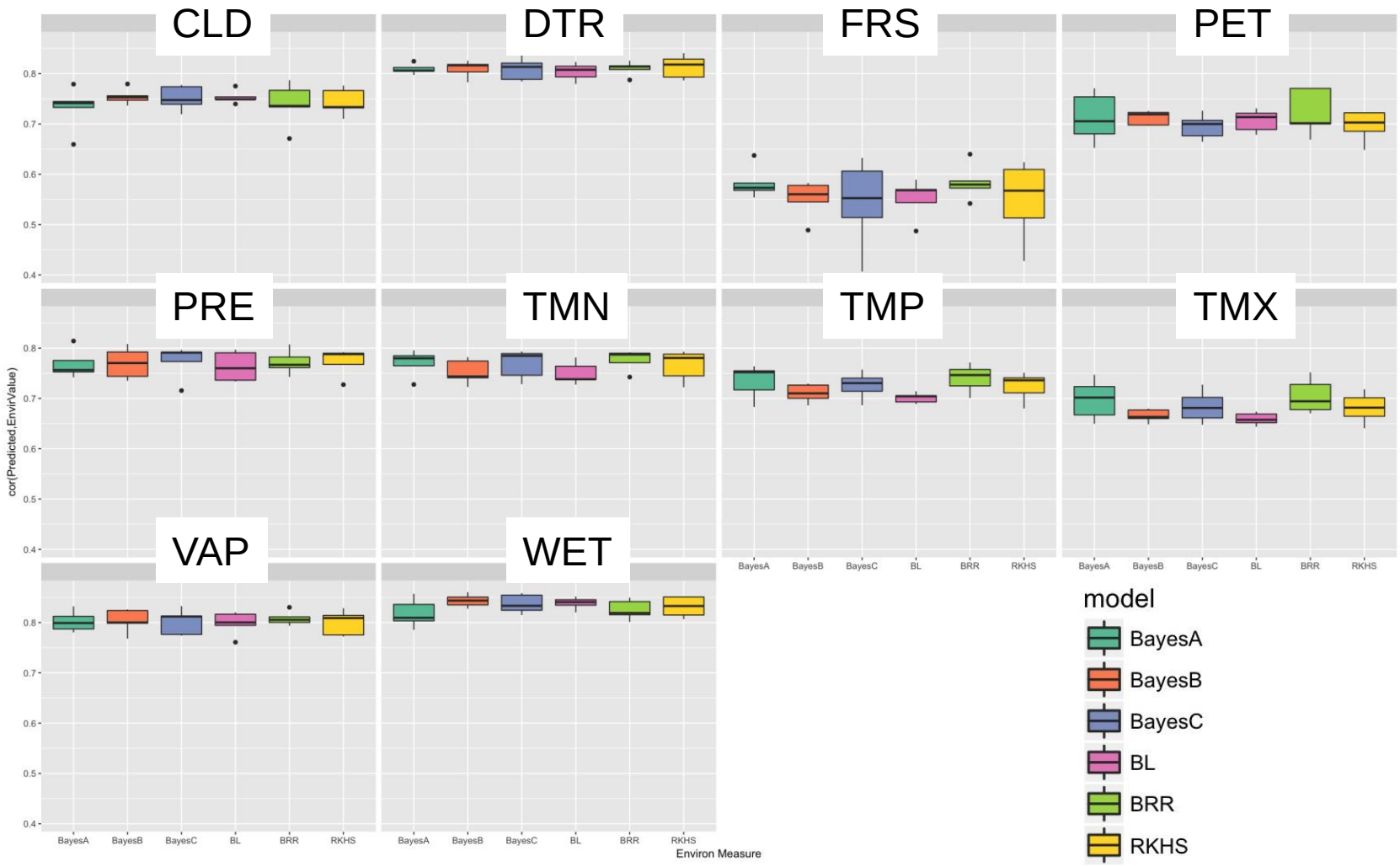
Landrace origin environment



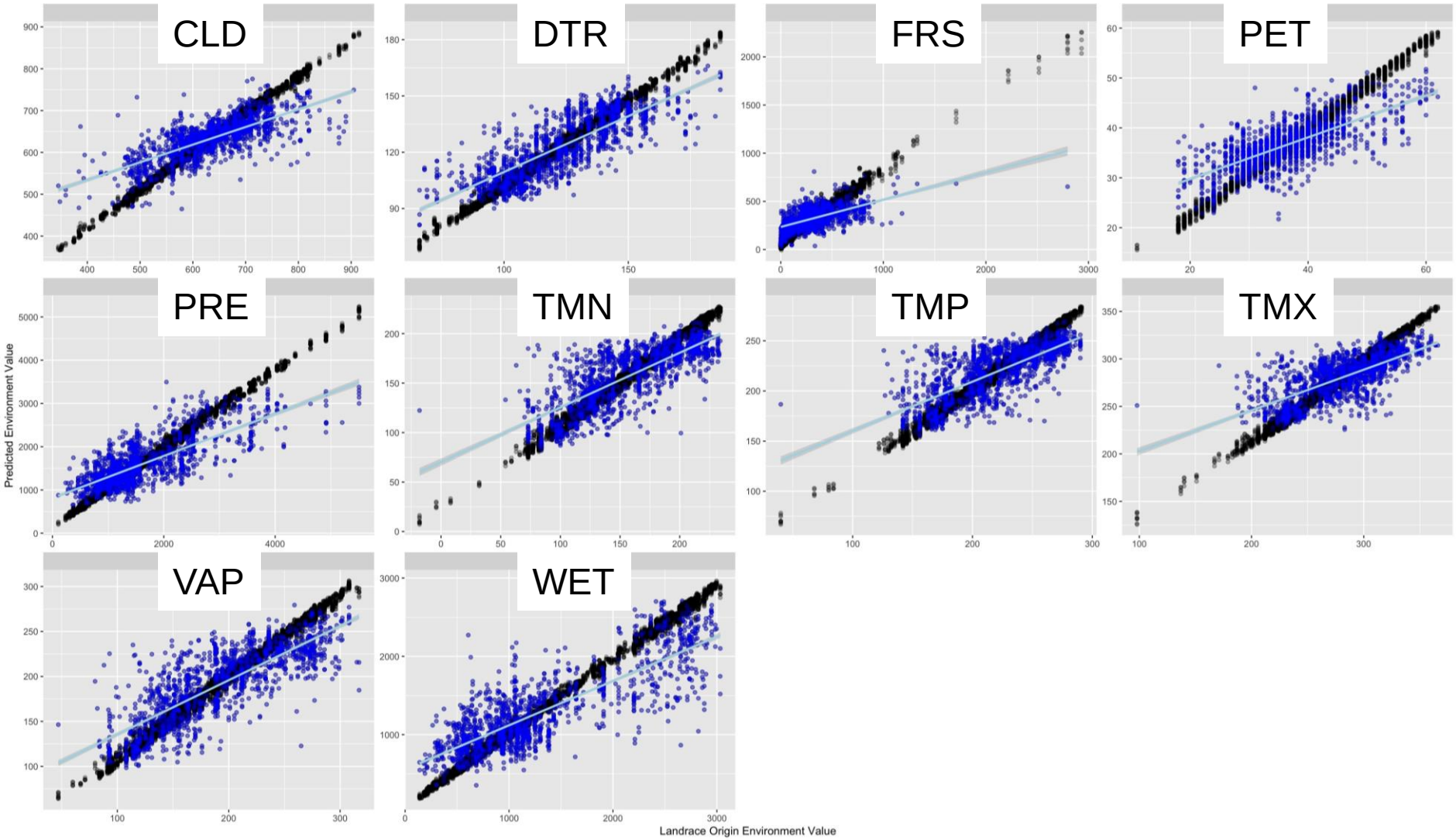
Models calculated over 500 iterations. 5 fold crossvalidation. 10% testing set.

CIMMYT

Environmental GS



Environmental GS

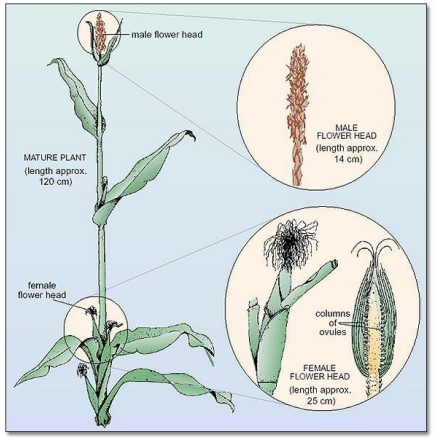


Environmental GS

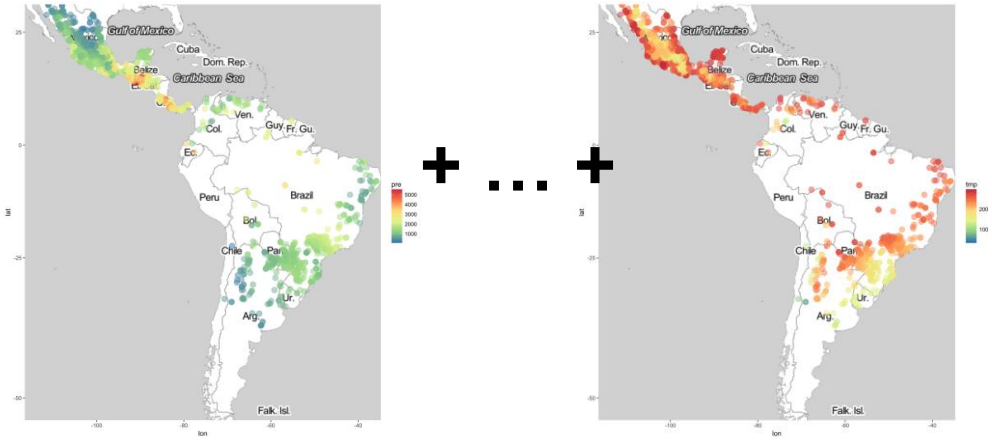
Predicting Flowering Time from Landrace collection site environment?

Field Trial Phenotype (DTA) = Landrace Origin Climatic Data

$$\mu + X_{precip}B_{temp} + X_{temp}B_{temp} + \ldots + X_{env}B_{env} + error$$



=

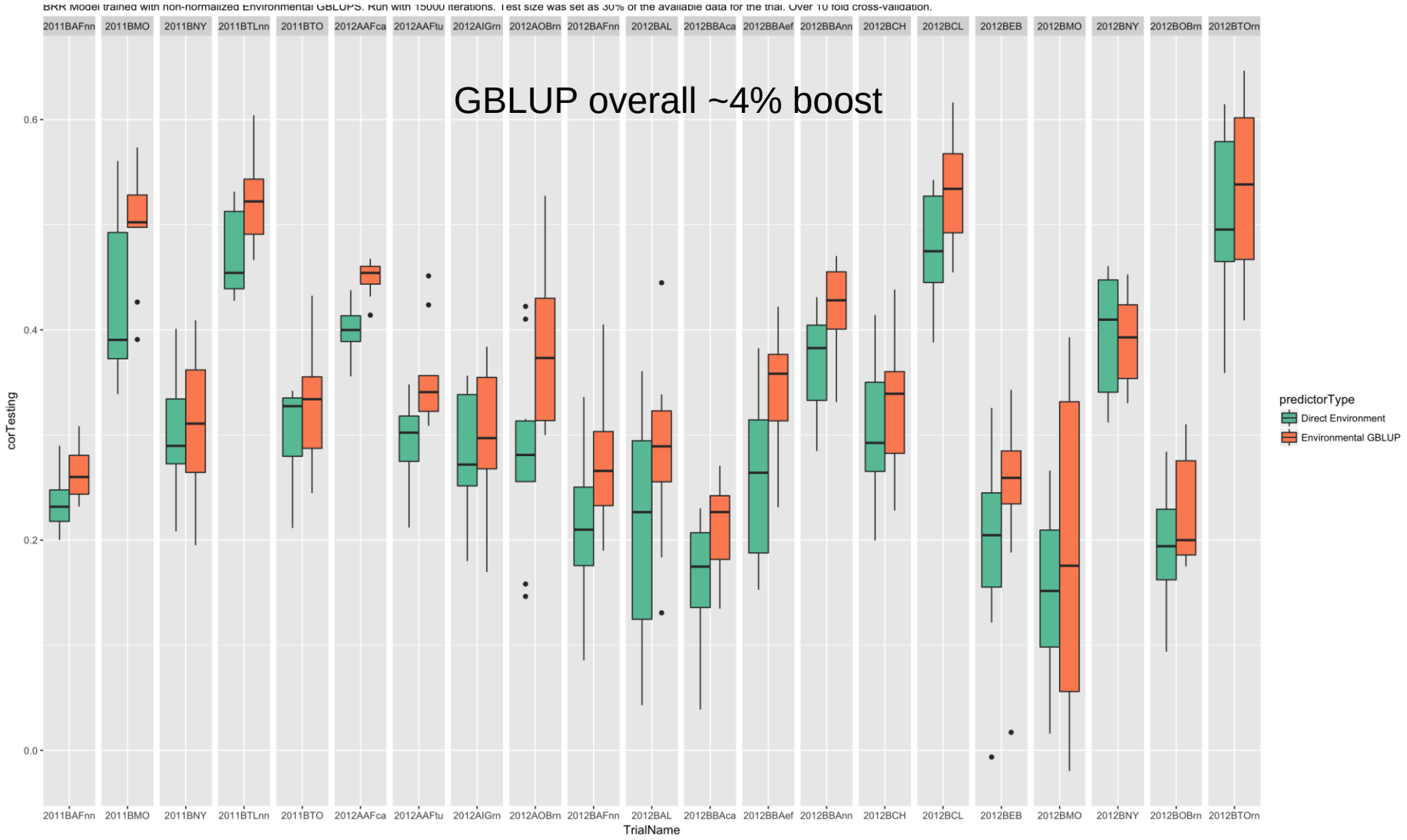


Landrace Origin Climatic Data GBLUP

CIMMYT

Environmental GS

Days to Anthesis predicted using landrace collection site environment- cross trials



Extend from GWAS panel to MGB

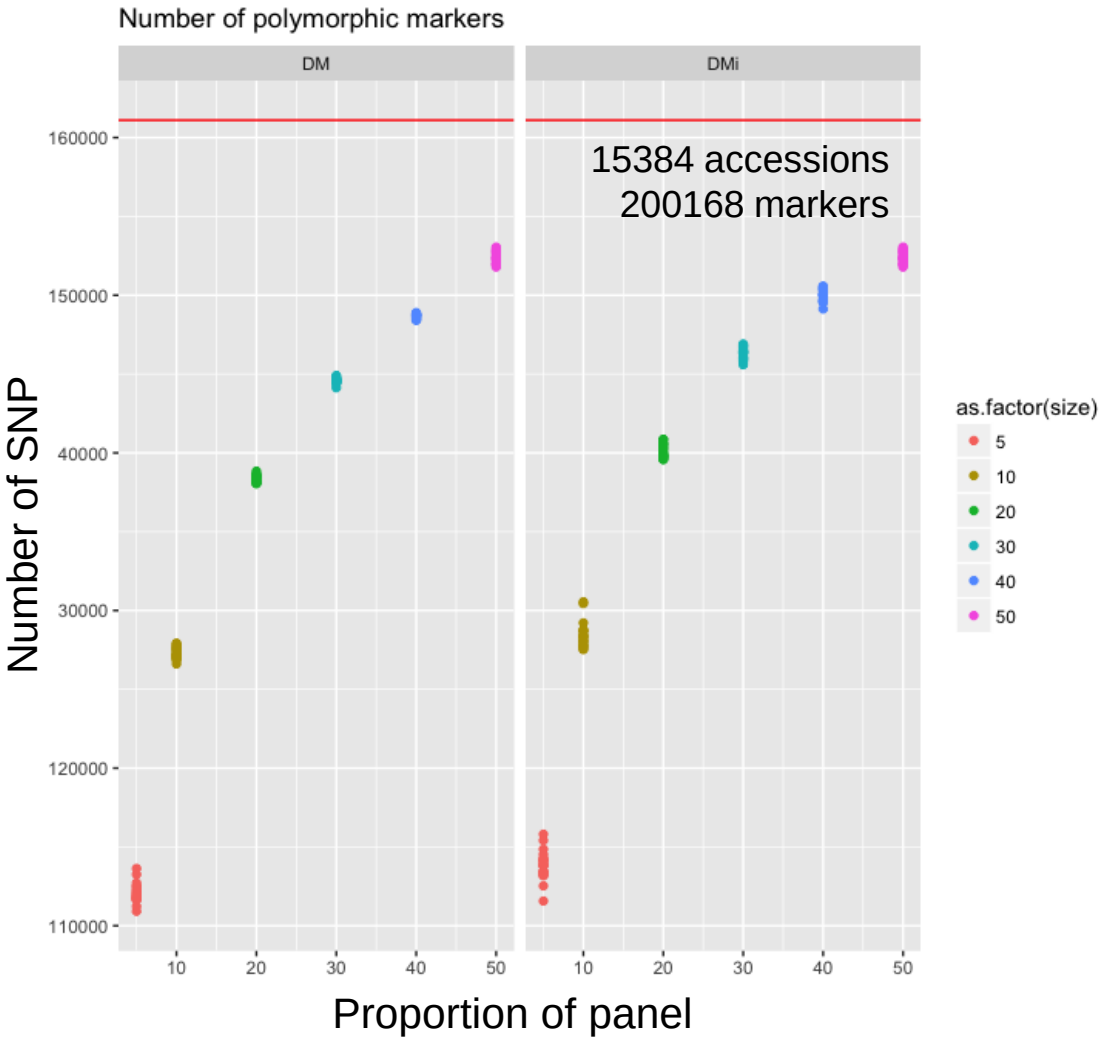
CIMMYT

Where now?

Genotypes
24,500 LR
28,500 MGB

Environments
17,500

Phenotypes
?



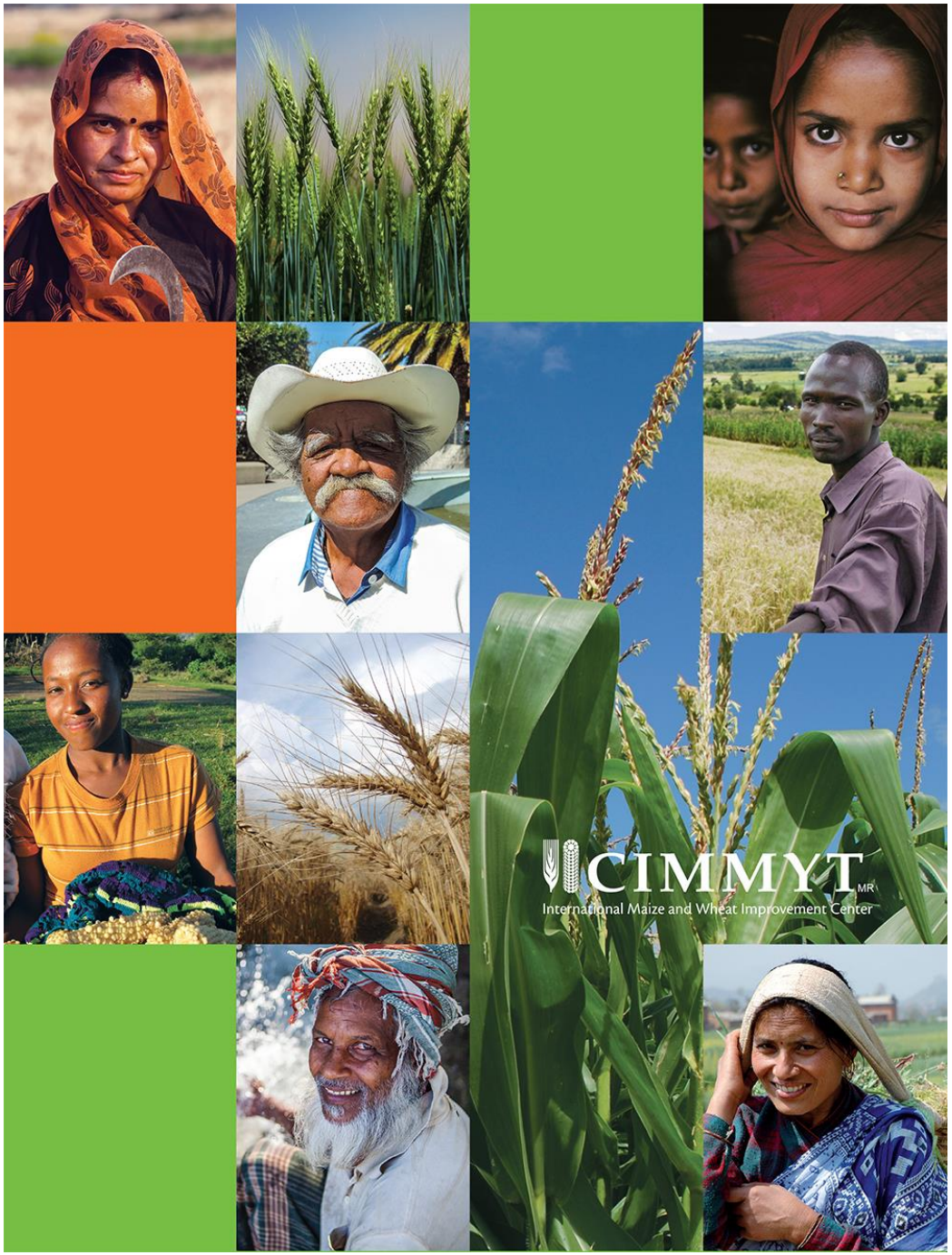
EnvGS: Best bets – highest potential value (soc-econ links)

EnvGWAS: What is potentially novel

CIMMYT

Thank you to fantastic, colleagues, collaborators and students

CIMMYT MR



MasAgro | Modernización Sustentable de la Agricultura Tradicional

SAGARPA
SECRETARÍA DE AGRICULTURA, GANADERÍA, DESARROLLO RURAL, PESCA Y ALIMENTACIÓN

Thank you for your interest!

CGIAR | RESEARCH PROGRAM ON Maize

BBSRC bioscience for the future

www.seedsofdiscovery.org