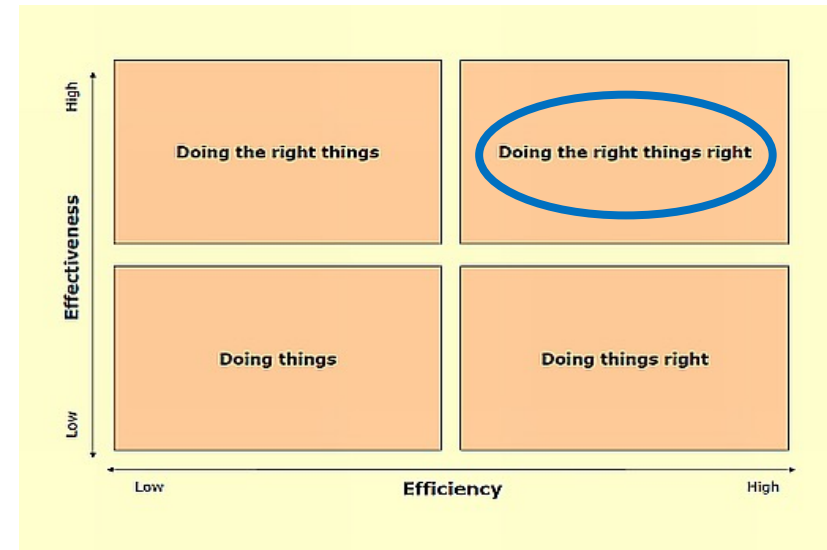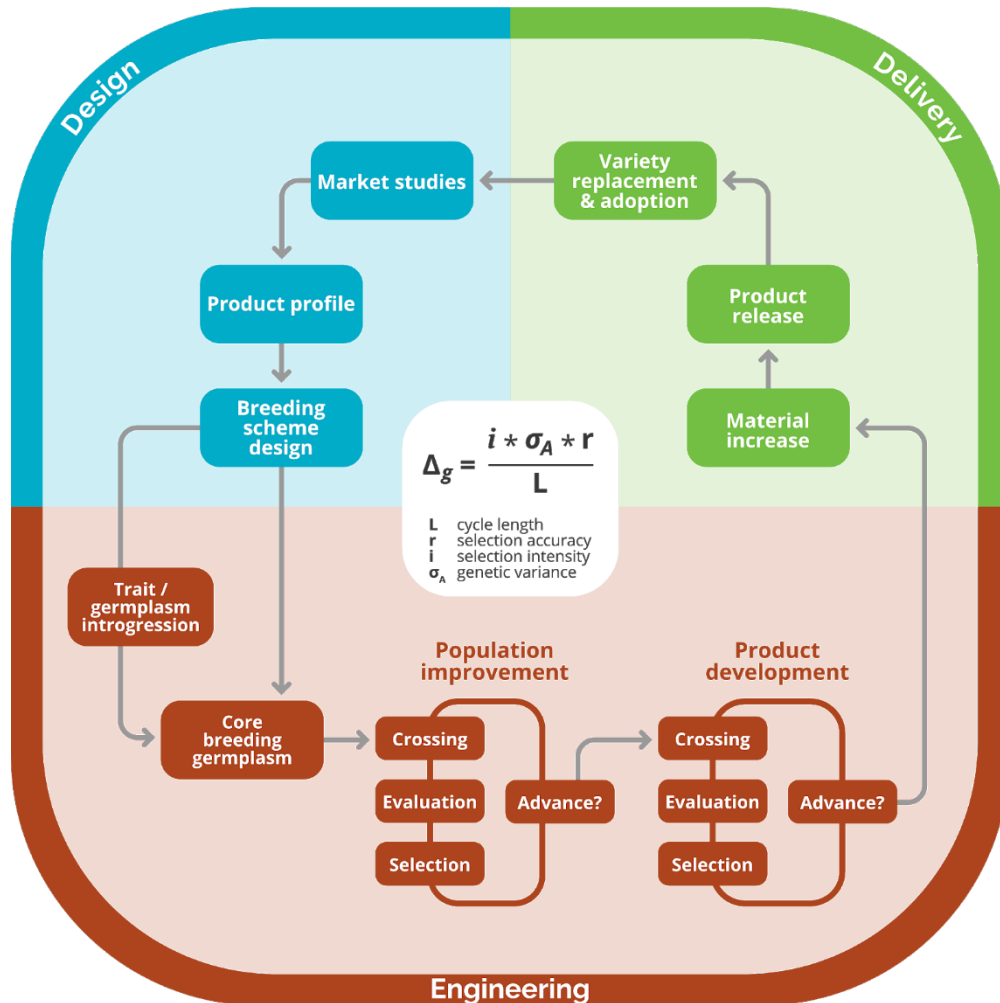# 'Packaging' & recombining genebank diversity and associated data for more effective use
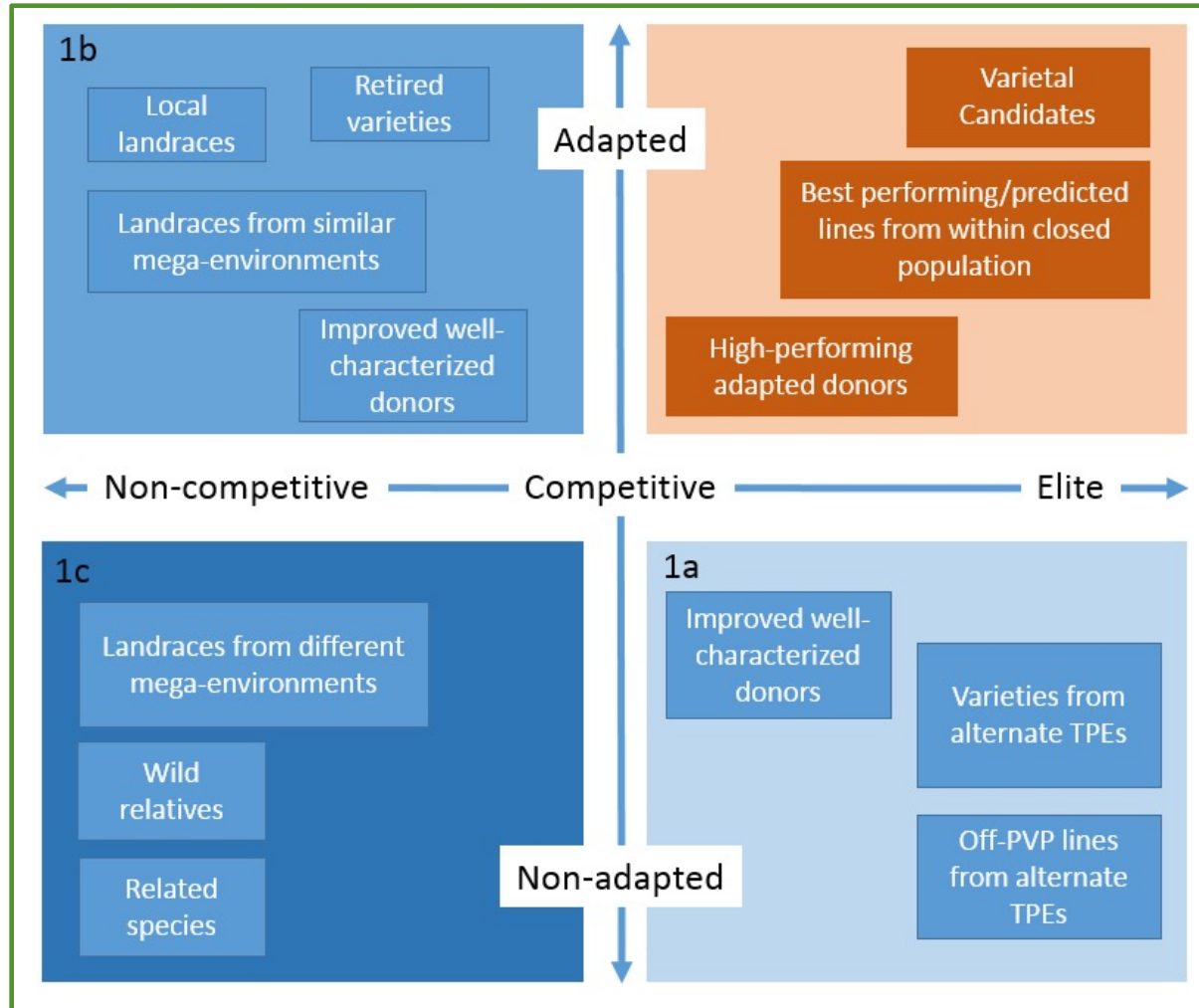
Sarah Hearne 9th July 2021

CIMMYT

Trait discovery and deployment within the breeding pipelines

Germplasm management: repackaging

Sourcing new alleles for demanded traits without disrupting the elite germplasm pool

Segmented trait discovery and deployment developing new elite donors

New alleles in adapted elite donors used in Core breeding

Discovery and Deployment Investment portfolio

Excellence in Breeding Platform

CGIAR

Trait Development and Deployment Responsibilities

BANKS

Collect, curate, describe, and **make available broad genetic diversity** following international treaties and national laws.
Broad characterization of accessions providing **high value information and knowledge** enabling more effective and efficient selection of germplasm "sets" and alleles of potential high value

TRAIT TEAMS

Define most appropriate **strategy/strategies** to meet the trait needs of core breeding
Develop specific **trait knowledge** through targeted characterization to facilitate trait deployment and effective use of donors in elite breeding
Develop intermediary germplasm products and **elite trait donors** to deploy in core breeding

CORE TEAMS

Define **trait needs**
Define most tractable **elite** backgrounds
Define stage gate **metrics** to move germplasm from TD&D to core population improvement

Excellence in Breeding Platform

CGIAR

Accelerated Breeding Initiative

Work Packages

- **Re-Focused Products**: Restructured breeding pipelines targeting defined market segments and product profiles.

- **Re-Focused organization**: Specialized breeding teams working to stage-gated processes for product development.

- **Transformed NARES and SMEs**: Varieties developed through transformed partnerships between CGIAR, NARES and SMEs.

- **Trait discovery and deployment**: Novel genes for high value traits identified and introgressed into highly elite parental lines.

- **Population improvement and variety validation**: Varieties developed by programs optimized for use of technology, traits, environments and safety.

Partners

- NARES
- Breeding companies
- Regulatory agencies (e.g. KEPHIS)
- Technology providers
- Universities and R&D organizations
- NGOs

Outputs

- **Improved varieties** targeting farmers' and women's needs, nutrition and climate adaptation and where possible climate change mitigation.

- **Elite parental material**: Novel alleles aggregated faster in elite genetic backgrounds.

- **Trait discovery**: Trait targeted research draws on genetic resources to support breeding for high-value traits.

- **Partnership model**: Breeding networks that transform the role and capacity of NARES and SMEs.

- **One CGIAR**: A more agile and effective breeding organization targeting low-income farmers needs.

Enhanced effectiveness of targeted trait discovery and deployment

Focus on current and future product profile demands

Working with discovery partners to explore, adapt, validate, optimize and deploy new discovery and deployment approaches

Genomics, Phenomics, Analytics collaborations

2022

CGIAR

CIMMYT MR

'Packaging' & recombining genebank diversity and associated data for more effective use

A working example

SeeD

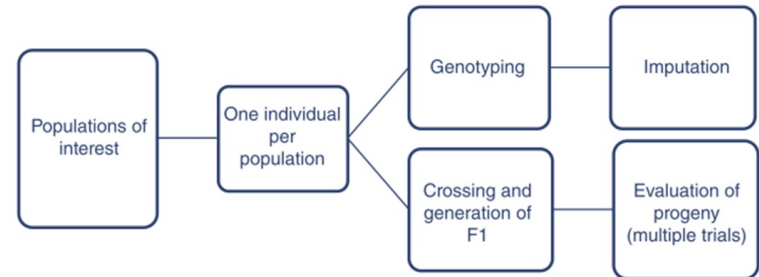"Unlock black box genetic diversity"

- Maize and wheat
- Focus on characterization and targeted use of germplasm bank accessions for trait discovery and deployment
  - Genomic characterisation
  - Phenotypic characterisation
  - Pre-breeding
  - Capacity development
- Multi million USD resourcing from the Mexican government
- Multi-stakeholder
- Started late 2010 (formally 2011)

GWAS

"Classical" GWAS using landraces

- 4500 landraces and composites Breeders core collection
- Evaluation of ~3500 testcrosses for multiple traits across 36 separate trials
- Flowering, Tar Spot, Drought, Heat, Low N…
- Replicate haplotype not accession
- Cost of phenotyping ~$2.5m USD (consumables)
- Cost of genotyping ~$350k

Romero et al 2017

Martha Willcox, Juan Burgueño, Sarah Hearne, Charles Chen, Ed Buckler, Alberto Romero, Gary Atlin, Felix San Vicente, Sam Trachsel, Peter Wenzl, INIFAP, Multinational and Mexican private sector, Mexican Universities
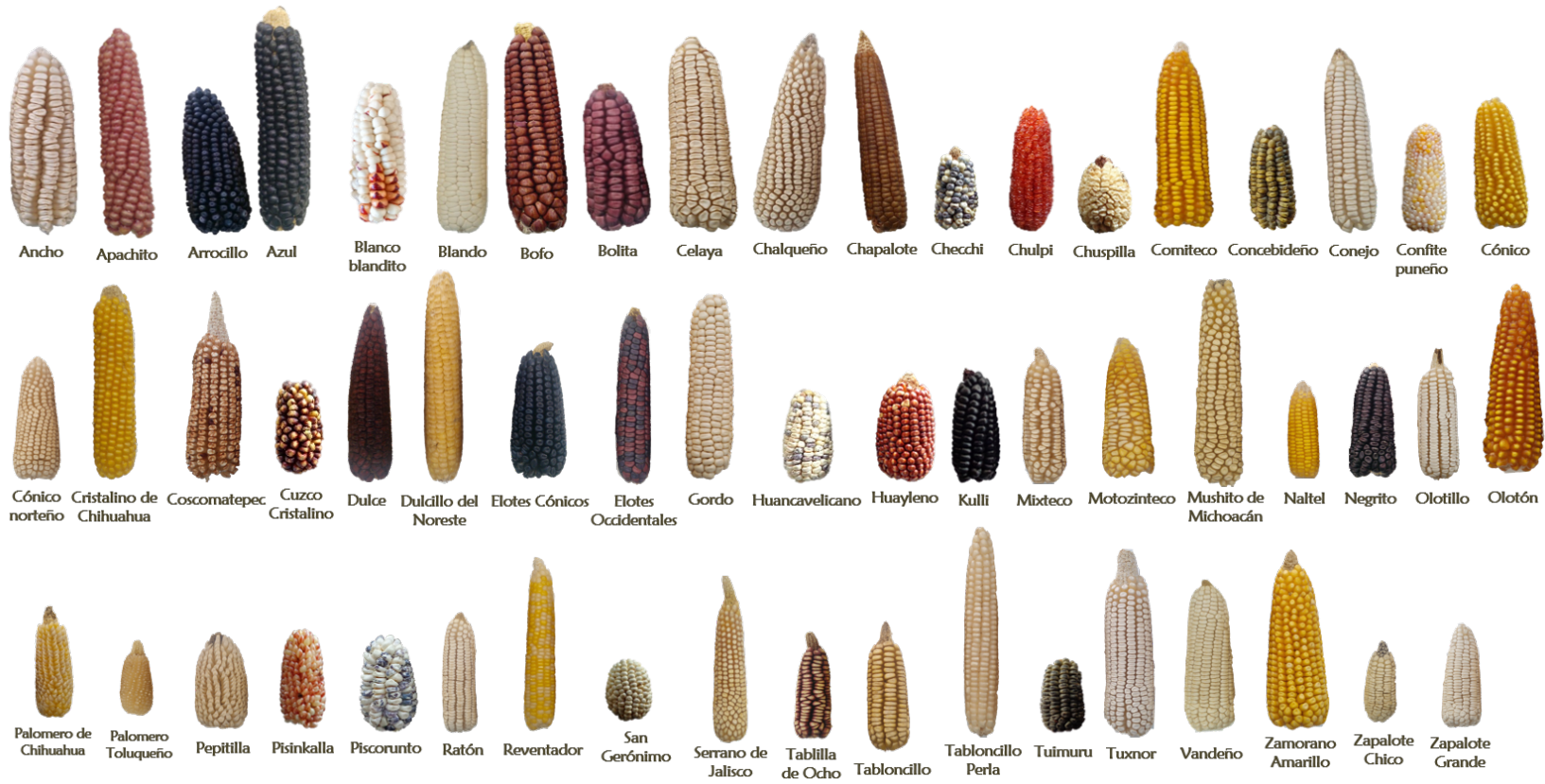
GWAS

"Classical" GWAS using landraces

- It works if done at appropriate scale
  - Powerful discovery approach
  - Provides some good data to ID breeding starts
  - ID variation not present in elite materials
- Filtering- missing and MAF, matters (MAF of QTL 2%)
- It is a lot of work for the ROI
  - Landraces are not a walk in the park to work with
    - Phenology
    - Adaptation per se
    - Assays are not always appropriate
- Polygenic-"simple" >40 QTL

- Much of the interesting "novel" variation is rare across the GWAS panel but is at high frequency/fixed in some accessions how do we better leverage this?
- Can we sample MORE?
- Can we better target/select "best bet" germplasm and phenotype wisely?
- How do we best conduct trait introgression?

CGIAR

CIMMYT MR

Make X0,000 years of evolution work for us: Data

Re-think GWAS

## Extending and expanding germplasm bank data

- Genotyped the whole collection CIMMYT, IIITA

- ~17300 landrace accessions with collection site "data" – center of domestication and centers of diversity

- ~Two years interns checking and curating these data – mapping locs, review collection records

- Extract short, medium and long-term monthly and annual climate variables from collection sites- T mean, T min, T max, precipitation, AI, soil pH, salinity

- Derive growing season and flowering & grain fill climate metrics

Global Surface Temperature

⬇ DOWNLOAD DATA

GLOBAL LAND-OCEAN TEMPERATURE INDEX

Data source: NASA's Goddard Institute for Space Studies (GISS) This trend agrees with other global temperature records provided by the U.S. National Climatic Data Center, the Japanese Meteorological Agency and the Met Office Hadley Centre / Climatic Research Unit in the U.K. Credit: NASA/GISS
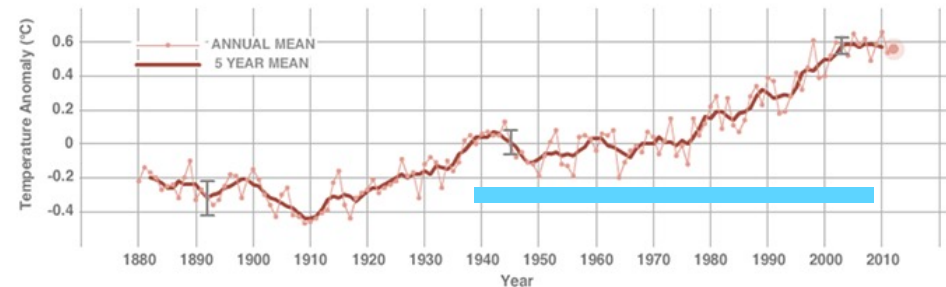
WorldClim - Global Climate Data

Free climate data for ecological modeling and GIS

GIS data is "virtually" free – 2% of the cost of the GWAS phenotyping

Kai Sonder

EnvGWAS concept

Change the response variable

GWAS

EnvGWAS

Genotype + Phenotype

Genotype + Collection Site Environment
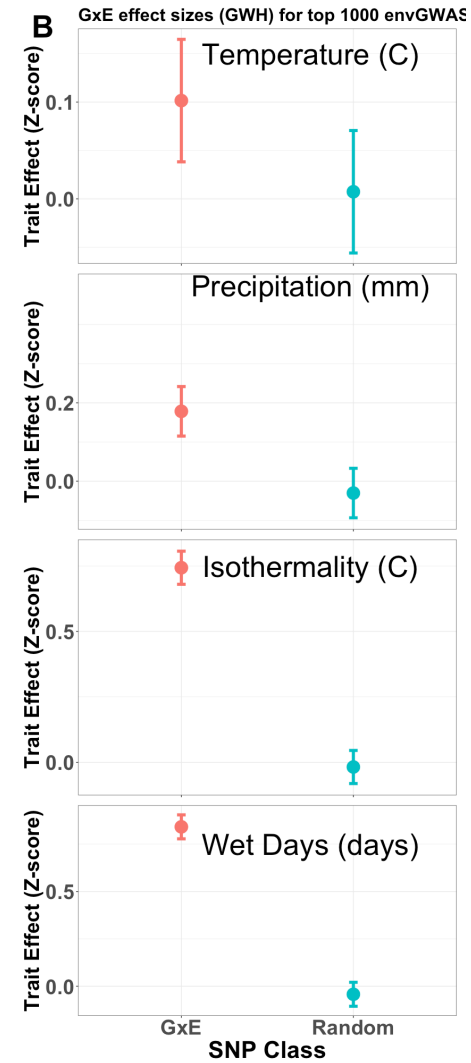
Analysis

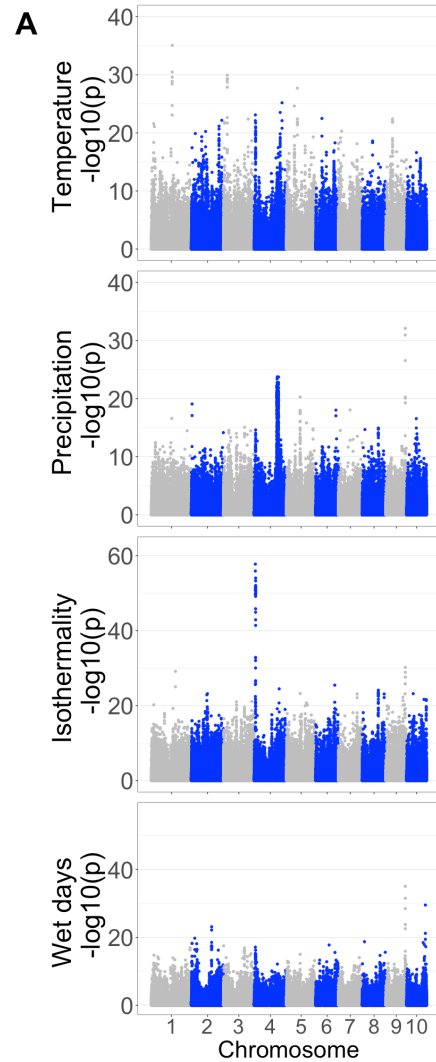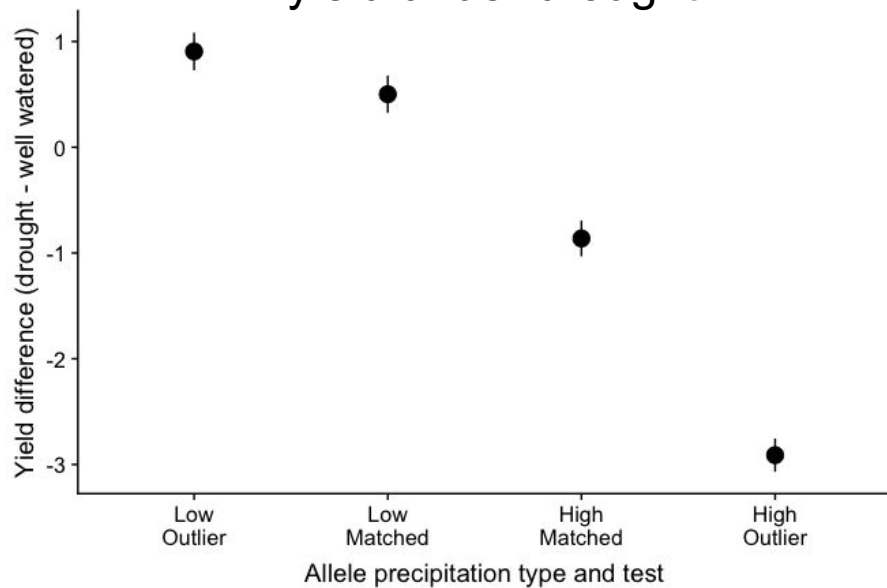Analysis

Associations between SNP and traits of interest

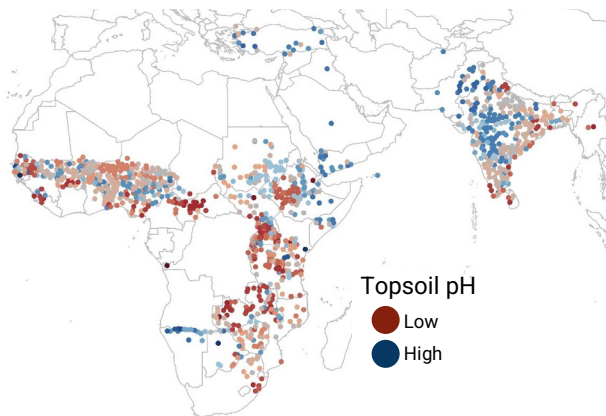Associations between SNP and "variables" of interest

It is effective for tractable traits

Low precipitation candidates enhance yield under drought
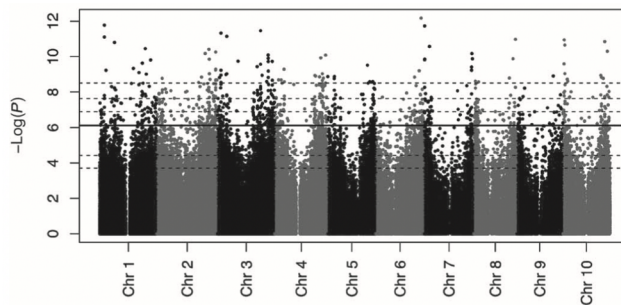
Sorghum- Environment associations predict adaptive traits
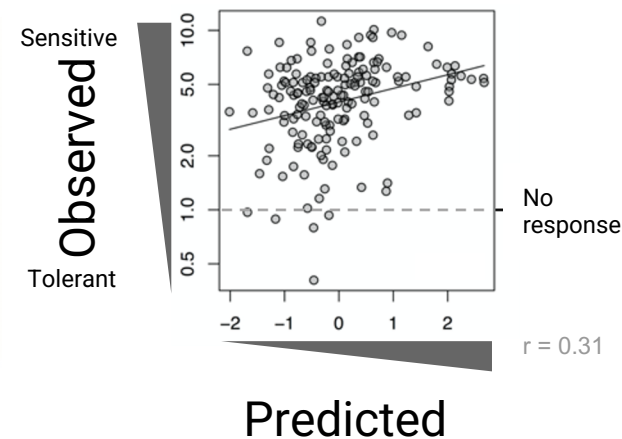
Genome-wide SNPs for georeferenced landraces

Environmental GWAS

Environmental genome-wide prediction (GWP)

EnvGWAS: topsoil pH

Relative root growth

Topsoil pH
- Low
- High

Sensitive

Observed

Tolerant

No response

r = 0.31

Predicted
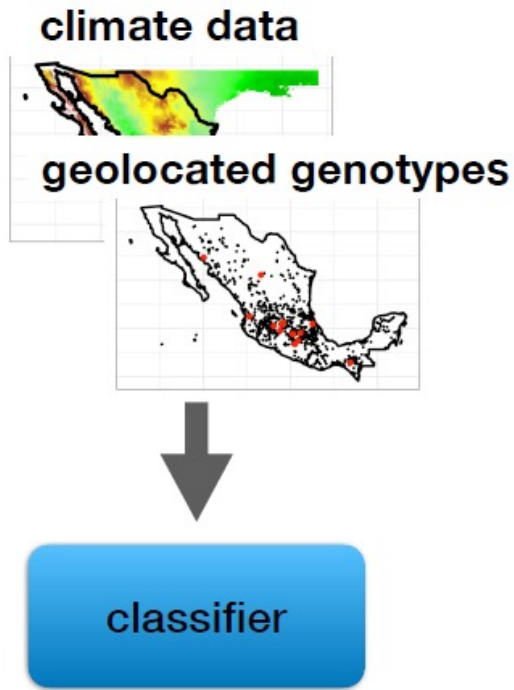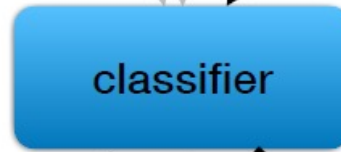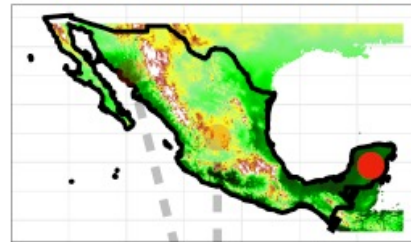
Lasky, ..., Morris 2015
Sci. Adv.

y-axis: root length, control/Al-toxic
x-axis: polygenic score for topsoil pH, z-score

CGIAR

CIMMYT MR

Predictions

predict genotype

climate data

geolocated genotypes

classifier

novel climate data
(location, time)

classifier

Predicted Genotype

A/A, G/A, AT … G/G   A/T, A/A, A/T … C/G   A/T, G/G, A/A … C/C

trial validation

grain weight/ha (Z-score)

distance from predicted genotype

Jeff Ross-Ibarra, Gates et al in prep
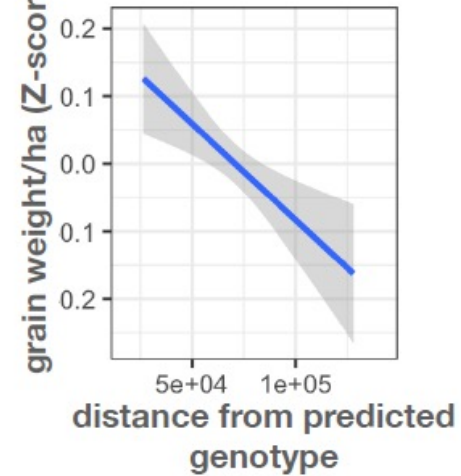
Using new knowledge in deployment

Lists of high potential value variants….what to do with it?

Mechanistic studies?

Discovery
   What? Find something of value
   Why? Understand the thing of value

First focus on the what and translating discovery to something clients can use

Using new knowledge in deployment

Take a germplasm customer focused approach – <u>Breeders</u>, Market Specialists
Be selective- efforts $$$; ensure you can deliver

Validation
- Leverage in-silico approach- functional, molecular physiology transcriptome, mine existing gen and phen data -to define highest value variants
- Phenotype within breeding populations
  - Landrace-based phen not high ROI in deployment for many traits (even disease)
  - Elite based populations
  - Advanced selection multiple backgrounds

Brute force- start big and slash
- Germplasm focus, important mechanistic studies conducted later (no breeder used NAM germplasm)
- Throw the kitchen sink at deployment- cycle time cuts, MAS, GS.

Define stage gates and KPIs
- Peer evaluated, high value products

CIMMYT Q ≡

News > CIMMYT releases its first ever maize genetic resource lines

NEWS **NUTRITION, HEALTH AND FOOD SECURITY**

CIMMYT releases its first ever maize genetic resource lines

The nine new maize lines are especially valuable for breeders seeking drought tolerance or tar spot complex disease resistance.

By CIMMYT

October 26, 2020