From genebank to field- leveraging genomics to identify and bring novel native variation to breeding pools.

**Coordinated by:**

CIMMYT<br>International Maize and Wheat Improvement Center

**Funded by:**

SAGARPA
SECRETARÍA DE AGRICULTURA, GANADERÍA, DESARROLLO RURAL, PESCA Y ALIMENTACIÓN
Potentially valuable genetic variation, the raw material for crop improvement, remains untapped on genebank shelves, at a time when challenges to crop production are unprecedented.
Systematically identify & mobilize novel, beneficial genetic variation into breeding programs to accelerate and strengthen genetic gains.
• Identify underutilized accessions of interest
• Find patterns in genome characteristic of beneficial adaptation: **Selection imprints**
• New **Heterotic** patterns (maize)
• ID rare but beneficial genome recombination

• Novel, beneficial **alleles** and **donor germplasm** identified
• **Markers** linked to genes that control priority traits

• **Bridging germplasm** (breeder-friendly lines and populations) enriched for novel, high-value alleles for stress tolerance, pest resistance and higher nutritional value available to breeders worldwide

• Toolkit to enable rapid adoption and accelerated breeding using bridging germplasm linked genetic elements that control priority traits

• Capacity to enable research and adoption of outputs from SeeD within the breeding community
Genebanks = Supermarket
GbS = configurable platform

# markers vs. # samples → GbS ‘flavors’

- **Cornell** University (*ApeKI*): Lots of markers (~850K markers; SNP only), ~50% missing data, imputation desirable → maize AM

- **DDArTseq** (Diversity Arrays Technology; PstI): Fewer markers (both SNP & ‘PAV’), higher coverage, lower error rates, less missing data, dynamic “reference” not dependent on B73 (re-seq 28 accessions and DH Langebio and NRGENE)
  - **DNA pools** of 30 plants per accession → allele frequencies within accessions from SNPs

Sarah Hearne, DArT, Cesar Petroli
• GbS of entire CIMMYT maize genebank (>27,500) completed

• Initial analysis of 21,000 accessions:
  – 1.2m SNP loci in total - raw
  – Mean 980k loci per accession (~20% missing data)
  – 317k loci with coverage ≥5X

• Also genotyping breeding materials (donors) and ex-PVP lines for comparison

• Mini haplotyping
Missing ingredients

10k accessions and 540 CML

Molecular atlas
ID accessions of interest

ID New sources of high value characters and alleles

Combine data resources

• Drought: 9954 landraces come from environments with long-term propensity for drought during flowering (Low AI)

• Genetic analysis using GbS data indicates these landraces cluster into six main groups

Sarah Hearne, Jorge Franco, Kai Sonder
• The six groups come from significantly different environments: All dry but some much drier than others - indicating some genotypic patterning - adaptation

LS Means of AI across clusters identified for three months of flowering and the entire 6 month growing period. The effect of cluster on AI for all three flowering months and over the growing period was significant (p<0.0001). Entries with the same letter code do not differ significantly (Tukey multiple comparison test <0.01 following ANOVA).
Selection footprints / selection sweep

- Look for selection motifs

Molecular atlas
SeeD Maize GWAS

Three adaptation zones:
- Lowland Tropical
- Subtropical
- Highland Tropical

36 Latin American countries

GbS Cornell and DArTseq methods
- Maximize marker density (Cornell)
- Enable identification of heterozygote regions (DArTseq)

Novel alleles and allele donors

Accession 1

Accession 4,500

34 trials

Tester

Tester

36 Latin American countries

Highland
Subtropical
Tropical
Accurate, field-based phenotyping is the main bottleneck

<table>
<thead>
<tr>
<th>Traits</th>
<th>Maize</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abiotic stresses</strong></td>
<td>heat</td>
</tr>
<tr>
<td></td>
<td>drought</td>
</tr>
<tr>
<td></td>
<td>low N</td>
</tr>
<tr>
<td><strong>Biotic stresses</strong></td>
<td>tar spot, ear rot, stalk rot, <em>Turcicum</em>, <em>Cercospora</em></td>
</tr>
<tr>
<td><strong>Grain quality and nutritional quality</strong></td>
<td>hardness, starch, oil, protein content, amino acids, phenolics, vitamin A</td>
</tr>
</tbody>
</table>

Flowering time, Yield, Yield components

• **Maize**: 800,000 data points from 35 trials across 14 locations

Martha Willcox, Sam Traschel, George Mahuku, Natalia Palacios, Dan Jeffers, Juan Burgueno
GWAS in SeeD POC: Flowering time

Association at known loci can provide insight into statistical power

500k SNP, imputation with Beagle & FILLIN

GWAS were performed per trial using BLUPs for days to silking using naïve GLM and GLM+Q. Non parametric meta-analysis

There are markers with significant association at and close to Vgt1 and ZCN8

POC: we can perform GWAS in the SeeD panel

Alberto Romero, Ed Buckler, Sarah Hearne
Signal in a new locus on chromosome 4

- Structural rearrangement-Inv4m locus
- Previously reported inversion in teosinte and highland maize (Hufford *et al*, 2013; Pyhäjärvi *et al*, 2013)
- Introgression with potential selective advantage
Inv4m locus 4 has an additive effect on days to anthesis

Inv4m haplotype clustering

Days to anthesis by cluster

Largest effect on flowering documented to date
Novel alleles and allele donors

- 20 x enrichment for centromeric regions.
- 6Mb suspected inversion in Chr3
- Centromere Chr 5 (heterosis)
- Chr6 12Mb region encompassing centromere
Tar spot disease complex

Phyllachora maydis infection, followed by Monographella maydis and Coniothyrium phyllachorae.
Novel alleles and allele donors

Tar spot disease complex

Alleles not present in CMLs, validated

Martha Willcox, Charles Chen, Juan Burgueno
Novel alleles and allele donors

Drought

Alleles not present in elite germplasm

Pyramiding alleles: ID best accession sources

Sarah Hearne, Martha Willcox, Juan Burgueno
MAIZE: grain protein content

Charles Chen, Natalia Palacios
MAIZE phenotypic diversity: per-se

- Per-se phenotyping **difficult**
  - Accessions = populations
  - Lodging
  - Three adaptation zones
- **Drought, heat, anthocyanin**
- Used AI as proxy for selection for drought
- Evaluation of 700 Lowland tropical and sub-tropical materials
- 100 BiP populations formed
- GIS analyses = good complement/substitute for expensive phenotyping

Terry Molnar, Martha Willcox, Natalia Palacios, Sarah Hearne, Kai Sonder
Bridging” germplasm

Maize ‘bridging germplasm’

Useful novel alleles & haplotypes

Early generation lines & pools enriched for favorable alleles

...using multiple strategies defined by trait complexity and breeder needs (desired input germplasm, demand for new sources)

<table>
<thead>
<tr>
<th>Breeder demand</th>
<th>Trait complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Monogenic (1-3)</td>
</tr>
<tr>
<td>Urgent</td>
<td>DH from landrace &amp; landrace / line crosses, selfing</td>
</tr>
<tr>
<td>Medium-term</td>
<td>MABC</td>
</tr>
<tr>
<td>Long-term</td>
<td>MABC &amp; GS</td>
</tr>
<tr>
<td></td>
<td>Oligogenic (4-10)</td>
</tr>
<tr>
<td>Urgent</td>
<td>DH from landrace &amp; landrace / line crosses, selfing</td>
</tr>
<tr>
<td>Medium-term</td>
<td>MARS &amp; prediction index</td>
</tr>
<tr>
<td>Long-term</td>
<td>MARS, prediction index &amp; GS</td>
</tr>
<tr>
<td></td>
<td>Polygenic (&gt;10)</td>
</tr>
<tr>
<td>Urgent</td>
<td>GS with MABC for BC1S1 development</td>
</tr>
<tr>
<td>Medium-term</td>
<td>GS with MABC for BC1S2 development</td>
</tr>
<tr>
<td>Long-term</td>
<td>GS with MABC for BC1S2 development</td>
</tr>
</tbody>
</table>

Terry Molnar, Sarah Hearne, Jiafa Chen
Pre-breeding

6 GS populations- broad accession based synthetics- drought, heat, low N

100 lowland and sub-tropical accessions ID for drought – BC1S1 (CML) –selection of 50 pops data for GS

2 accessions for Tar Spot BC to favorable CMLs – ID additional alleles and use in MABC

Blue maize – 8 accessions BC pops

ID accessions Cercospora, Turcicum, PVA, protein, oil, Fusarium stalk rot.

• DH- works but put in inducible background first

Terry Molnar, Sarah Hearne, Jiafa Chen
Swim in data (not drown), collaborate and enhance capacity

- Data release - Germinate, Dataverse
- Protocols, software, scripts
- Analysis
  - QTL; GWAS, Bi-P
  - “Global” diversity
  - Core sets
  - Selection footprints
- Germplasm
  - Pre-bred germplasm with novel high value traits - disease, drought, heat, quality and associated marker information – from 2016
- Collaborations - functional genomics

Kate Dreher, Sarah Hearne, JHI
Thank you