Genomic exploration and use of genebank collections for breeding: Early insights and experiences from Seeds of Discovery-Maize.

Sarah Hearne, Terry Molnar, Martha Willcox, Jorge Franco, Carolina Sansaloni, Cesar Petroli, Juan Burgueño, Jose Crossa, Alberto Romero, Charles Chen, John Hickey, Gregor Gorjanc, Janez Jenko, Andrzej Kilian, David Marshall, Ed Buckler, Peter Wenzl, Kevin Pixley + many more!
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.

Genebanks should NOT be museums. Genetic variation is the basic ingredient of all plant breeding. Genebanks should be a source of high-value genetic diversity to meet tomorrow’s challenges.
Systematically **identify & mobilize novel, beneficial genetic variation** into breeding programs to **accelerate and strengthen genetic gains**
Maize modules

1. Molecular atlas
   - 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

2. Novel alleles and allele donors
   - Novel, beneficial alleles and donor germplasm identified
   - Markers linked to genes that control priority traits

3. “Bridging” germplasm
   - 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

4. Information management
   - Toolkit to enable rapid adoption and accelerated breeding using bridging germplasm linked genetic elements that control priority traits

5. Capacity-strengthening
   - Capacity to enable research and adoption of outputs from SeeD within the breeding community
Molecular atlas
Genebanks - supermarket

Common label - genotypic data
• Genotyped whole MGB
• Composite genotyping GbS (30 individuals per accession in one sample)
• GIS derived information
GBS of entire CIMMYT maize genebank (>27,500) completed end 2014

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

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

<table>
<thead>
<tr>
<th>Group</th>
<th>Month 1</th>
<th>Month 2</th>
<th>Month 3</th>
<th>6 month growing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.926b</td>
<td>1.923b</td>
<td>2.498d</td>
<td>1.885e</td>
</tr>
<tr>
<td>2</td>
<td>1.433b</td>
<td>1.379b</td>
<td>1.603a</td>
<td>1.134a</td>
</tr>
<tr>
<td>3</td>
<td>1.459b</td>
<td>1.486a</td>
<td>1.776b</td>
<td>1.249b</td>
</tr>
<tr>
<td>4</td>
<td>1.260a</td>
<td>1.458a</td>
<td>1.485a</td>
<td>1.099a</td>
</tr>
<tr>
<td>5</td>
<td>2.005c,d</td>
<td>2.241c</td>
<td>1.945c</td>
<td>1.713d</td>
</tr>
<tr>
<td>6</td>
<td>2.136d</td>
<td>2.213c</td>
<td>2.009c</td>
<td>1.592c</td>
</tr>
</tbody>
</table>
SeeD Maize GWAS

Accession 1

Novel alleles and allele donors

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)

34 trials
Breeder-oriented cores

- $300k 2y
- $150k 2y
- $72k 1y
- $30k 1 season
- $14k 1 season
Selection footprints / selection sweep

- Look for of selection motifs
Novel alleles and allele donors
Upstream

Breeding-oriented

Main emphasis:
Mobilize novel alleles for complex traits into breeding programs

Genetically complex traits
[heat/drought tolerance]

Genetically simple traits
[some diseases]

‘Low-hanging fruits’ for breeding

Seek collaborations to mine data for basic research

Prioritization of traits

Novel alleles and allele donors
SeeD Maize GWAS

Accession 1

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

36 Latin American countries

34 trials

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

Novel alleles and allele donors
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, 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, Zn</td>
</tr>
</tbody>
</table>

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

Novel alleles and allele donors

Alleles not present in elite germplasm

Pyramiding alleles: ID best accession sources
Thank you