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

Sarah Hearne, Terry Molnar, Martha Willcox, Carolina Sansaloni, Cesar Petroli, Juan Burgueño, Jose Crossa, Alberto Romero, Charles Chen, HuiHui Li, Jorge Franco, Natalia Palacios, Sam Traschel, Kate Dreher, John Hickey, Gregor Gorjanc, Janez Jenko, Andrzej Kilian, David Marshall, Ed Buckler, Peter Wenzl, Kevin Pixley + many more! Coordinated by:

CIMMYT

International Maize and Wheat Improvement Center

SAGARPA

SECRETARÍA DE AGRICULTURA, Ganadería, desarrollo Rural, Pesca y Alimentación

Funded by:



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**





Seeds of Discovery : SeeD



Systematically identify & mobilize novel, beneficial genetic variation into breeding programs to accelerate and strengthen genetic gains

Maize modules

adoption and

using bridging

germplasm linked

accelerated breeding

genetic elements that control priority traits



worldwide

 Capacity to enable research and adoption of outputs from SeeD within the breeding community

Molecular atlas

Genebanks = Supermarket



GbS = configurable platform

СІММҮТ

markers vs. # samples \rightarrow GbS 'flavors'

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

Molecular atlas



Pop type dim2 Collection Composite Group Pools Teosinte Variety dim3 -dim1

MAIZE genetic diversity survey

- 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

Sarah Hearne, Jorge Franco, Jiafa Chen, Cesar Petroli

Missing ingredients





10k accessions and 540 CML

Molecular atlas



ID New sources of high value characters and alleles

Combine data resources

- Drought: 9954 landraces come from environments with longterm 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

ID accessions of interest





ID accessions of interest

 The six groups come from significantly different environments: All dry but some much drier than others- indicating some genotypic patterningadaptation



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 of selection motifs

039(C>T (0)	
网络帕尔特拉亚希尔 减低的 使无论的 计正式算法分析 网络中国的第三日复数	
國際性 가장을 물질 수 있는 것 같아요. 가는 것 같아요. 이는 것 같아요. 이는 것 같아요. 이는 것 같아요. 이는 것 같아요.	
國際領域保護保護和保護和保護保護保護保護保護保護保護保護保護保護保護保護保護保護保護保	
時間 이 가슴 때 같은 것 같아요. 김 가지의 에 관련을 위한 이가 가지 않는 것을 가지 않는 것을 하는 것 같아요.	
	Palomero
http:///information.com/action/acti	
的复数形式 机热力 化增强 计可保持 机动力计算机 计算机 动物 化乙酸乙酸	
医结肠 化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化化	
	- Cónico
精制的法心的 机油酸盐 化结晶 化合物油的 化拉拉拉 法自己的现在分词 计可以定义的 化自己的过去式 化合物化合物	
	Arrocillo
新生活。44.2.2.4.4.4.4.4.4.4.4.4.4.4.4.4.4.4.4.	



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)

Highland Subtropical Tropical





Accurate, field-based phenotyping is the main bottleneck

Traits	Maize	
Abiotic stresses	heat	
	drought	
	low N	
Biotic stresses	tar spot, ear rot, stalk rot, Turcicum, Cercospora	
Grain quality and nutritional quality	hardness, starch, oil, protein content, amino acids, phenolics, vitamin A	

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

chment across models

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

zcn8 vgt1 50 12 2 5 0 1.25e+08 1.30e+08 1.35e+08 Position on chr 8

GWAS significance for days to silking on SeeD



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

Agua Fria 2011, MLM Q+K



CIMMYT



Inv4m locus 4 has an additive effect on days to anthesis



Largest effect on flowering documented to date



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





Tar spot disease complex



Close-up of maize leaf: Tar spot causes lesions, drying and premature leaf death



Phyllachora maydis infection, followed by *Monographella maydis* and *Coniothyrium phyllachorae*.

Novel alleles and allele donors

>) Manifestinisti

Tar spot disease complex



Alleles not present in CMLs, validated

Martha Willcox, Charles Chen, Juan Burgueno



Drought

Alleles not present in elite germplasm



Numero de TC Frecuencia del alelo en las CML

Pyramiding alleles: ID best accession sources



Regiones cromosómicas presentes en las accesiones



/ha)

▲ No. de accesiones

Rendimiento con la región

Rendimiento sin la región

cromosómica o sin datos (kg

cromosómica (kg ha)

Sarah Hearne, Martha Willcox, Juan Burgueno

MAIZE: grain protein content



Charles Chen, Natalia Palacios

Novel alleles and allele donors



"Bridging"

germplasm

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 **ICIMMYT**...

"Bridging"

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)

	Trait complexity			
Breeder demand	Monogenic (1-3)	Oligogenic (4-10)	Polygenic (>10)	
Urgent	DH from landrace & landrace / line crosses, selfing	DH from landrace & landrace / line crosses, selfing	GS with MABC for BC1S1 develop- ment	
Medium- term	MABC	MARS & prediction index	GS with MABC for BC1S2 develop- ment	
Long-term	MABC & GS	MARS, prediction index & GS	GS with MABC for BC1S2 develop- ment	

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

Molecular atlas





Prospects

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



SECRETARIA DE AGRICULTURA, Ganadería, desarrollo rural, Pesca y alimentación



