# **Genome Wide Assessment of Maize Genebank Diversity; Synthesis of Next Generation Technologies and GIS Based Approaches**

Sarah Hearne<sup>1\*</sup>, Jorge Franco<sup>2</sup>, Jiafa Chen<sup>1</sup>, Carolina Sansaloni<sup>1</sup>, Cesar Petroli<sup>1</sup>, Andrzej Kilian<sup>3</sup>, Kai Sonder<sup>1</sup>. 1 CIMMYT, 2 Universidad de la Republica Uruguay, 3 Diversity Arrays Technology Pty Ltd (DArT PL). \* shearne@cgiar.org

# Introduction

The Seeds of Discovery Initiative (SeeD), endeavours to explore native genetic resources of maize and wheat and make available to breeders useful genetic variation not present in modern breeding pools. As part of the efforts on maize we have over the last three years developed new genotyping methodologies targeted to exploration of genetic diversity in allogamous populations and have now characterised the entire international maize genebank collection held at CIMMYT using this method. In addition, we have worked to expand and add value to the passport data available using GIS based analysis. Using both genotypic and GIS based data we have begun to explore and understand the genetic diversity and genetic space of the CIMMYT maize genebank collections and relate that to the current elite breeding materials developed at CIMMYT and elsewhere. Initial analysis of 20,000 maize accessions offer exciting opportunities for targeted data mining and phenotypic evaluation of high value accessions. Here we present preliminary analysis with a focus on drought.

#### Broad analysis

Multidimensional scaling of genetic distance (Fig. 1), indicates a clear separation of improved materials from the landraces present in the genebank along dimension 1(Fig1,a). Landraces cover the broadest swathe of space with the 88 teosinte accessions present forming a discrete band across dimension 1 towards the lower part of dimension 2 (obscured in the orientation of the displayed diagram). Classification of landraces into highland tropical and lowland/midaltitude tropical adaptation indicates clear genetic separation based on adaptation (Fig1, b). Geographic patterning is also present but harder to differentiate given the large number of countries represented. Figure 1, c. There is co-location of Mexican highland materials within the lower half of dimension 2, close to teosintes with South American highland accessions being concentrated in a different location in the upper half of dimension 2. This is in broad agreement with previous studies. Brazilian, Argentinian and Uruguayan derived accessions also locate in distinct areas of the genetic space presented.

CIMN

International Maize and Wheat Improvement Center

# Methods and Results

Genotypic characterisation

All accessions within the International Maize Genebank housed at CIMMYT are being genotyped using a composite genotyping by sequencing (GbS) approach specifically developed for the assessment of heterogeneous populations.

- DNA from a composite of 30 individuals per population is used for GbS employing Illumina HiSeq platforms.
- Resulting sequence reads are clustered to an evolving internal reference of clusters.
- SNP are identified and the number of sequence read variants per SNP are determined.
- SNP "count" data is then converted into a frequency of each allele, per locus, per sample and is filtered to remove all variants which fall below a 5X coverage.
- Statistical analysis was conducted in R with multidimensional scaling results visualised in CurlyWhirly. Analysis presented here is based on 21098 accessions.

## GIS based selection of drought prone accessions

Curation of accession geolocation data and extraction of long term climate variables for collection sites was conducted (http://seedsofdiscovery.org) covering 17353 accessions in the genebank collection. Selection of accessions from target environments prone to drought stress was based on long term mean aridity index (a ratio of average precipitation to potential evapotranspiration, AI) using the following criteria:

• Identification of accessions that came from locations with a mean AI over the growing season above 0.3

## Drought based analysis

Visualisation of the six clusters of drought prone accessions shows separation into some adaptation based clusters. Cluster 2 is formed from a mix of highland and subtropical/tropical Mexican accessions, cluster 3 is formed predominantly from Mexican highland accessions while cluster 4 is formed from south American highland materials, (Fig 2, a and b).



Figure 2. Multidimensional scaling of Modified Rogers distance between 21098 genebank accessions with 9954 accessions from drought prone environments colored according to cluster identity. Two orientations of the figure are presented rotating on dimension 2 with dimension one coming out (a) and going into the page (b)

- (non-irrigated)
- Identification of accessions with a mean monthly AI during flowering of <0.5 during one or all of the

three months of flowering and grain fill. 9954 accessions were identified as coming from environments with long term propensity for potential drought stress.

### Clustering and analysis of selected accessions

Genetic distance based clustering of accessions was conducted resulting in a definition of 6 distinct clusters based on a combination of Calinski-Harabasz criterion, entropy and height of separation in a dendrogram representation. ANOVA and Tukey tests were conducted to assess the influence of cluster on AI across each flowering month and 6 month growing period using R.







Figure 1. Multidimensional scaling of Modified Rogers distance between 21098 genebank accessions. Points are colored to represent a) Germplasm type, b) Adaptation and c) Country of origin

Analysis of the climatic conditions found for accessions within each cluster indicates some clear differentiation of both genetic and abiotic characters. Clusters 2, 3 and 4, all containing a high proportion of highland material have significantly lower AI than other groups in at least one of the periods evaluated (table 1).

#### Table 1.

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).

Group	Period				# acc
	FL Month 1	FL Month 2	FL Month 3	6 month growing	
1	1.926 c	1.923 b	2.498 d	1.885 e	2767
2	1.433 b	1.379 a	1.603 a	1.134 a	2279
3	1.459 b	1.486 a	1.776 b	1.249 b	1335
4	1.260 a	1.458 a	1.485 a	1.099 a	781
5	2.005 c,d	2.241 c	1.945 c	1.713 d	1211
6	2.136 d	2.213 c	2.009 c	1.592 c	1581

## Summary

As may be expected elite germplasm is genetically distinct from landrace accessions held in genebanks. Genotyping coupled with cost effective GIS analysis can aid selection of accessions for phenotypic analysis for target traits and future work on allele discovery. We present an example where clear geographic, adaptive and climatic differentiation of accessions can be seen in genotypic data offering opportunities for mini-core collection formation and selection footprint analysis. Interestingly highland accessions, materials which can be challenging to work with, offer a potentially untapped source of novel alleles for drought tolerance; their distance from elite materials is high and further analysis of data is planned to explore this further.

To date we have conducted preliminary analysis based on drought, high temperature stress and also have started to use race classification data present in passport documentation for accessions. We have used distance based design to sample from within clusters to form nested mini-core collections of accessions to enable breeders and researchers interested in further phenotypic analysis to better access broad trait targeted diversity from the genebank.









Modernización Sustentable de la Agricultura Tradiciona Este programa es público, ajeno a cualquier partido político. Queda prohibido el uso para fines distintos a los establecidos en el programa