

Capturing genetic diversity in maize populations and inbred lines - Genotyping of composite germplasm set

Principle Investigator: Marilyn Warburton, CIMMYT

Main Collaborators: INRA (France), IITA (Africa), CAAS (China)

Maize composite set

Divided in two parts

- **467 maize populations**

CIMMYT (216), IITA (150), INRA (01)

Genotyping with 50 SSRs by CIMMYT and INRA



- **987 maize inbred lines**

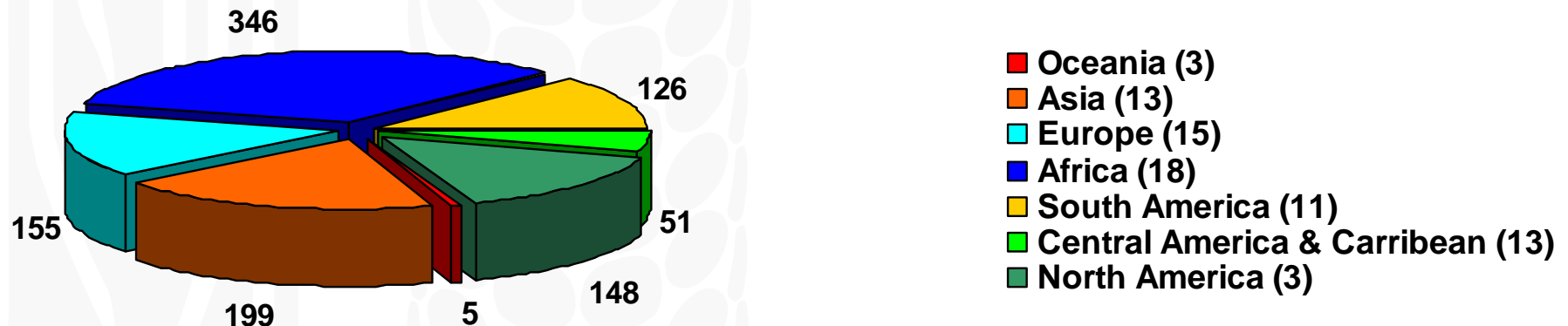
CIMMYT (587), IITA (200), CAAS (200)

Genotyping with 50 SSRs by CIMMYT and CAAS



Global distribution of maize populations

Project	Year of fingerprinting	# SSR	# Populations	Geographical focus
CIMMYT/INRA	2003	24	273	America, Europe
GCP, SP1, tier 1	2004	34	467	Africa, America
GCP, SP1, #14	2006	40 - 50	288	Africa, Asia
Total	2006	24	1028	



Traditional maize races from Central and South America

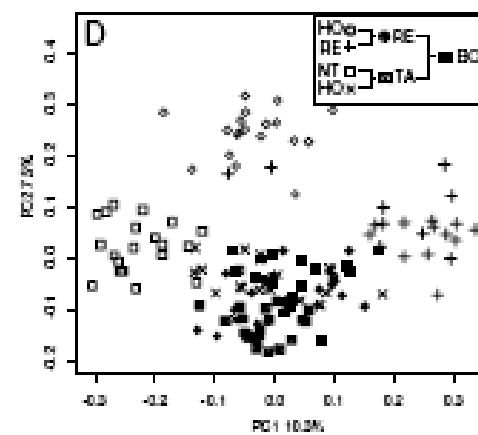
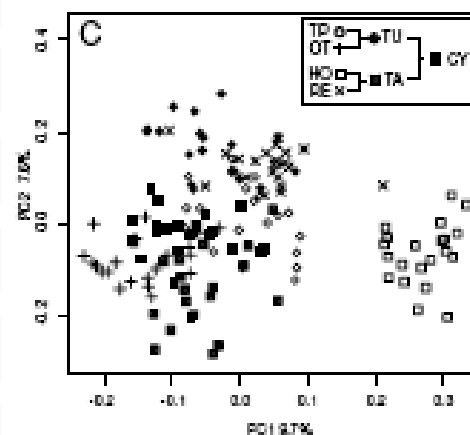
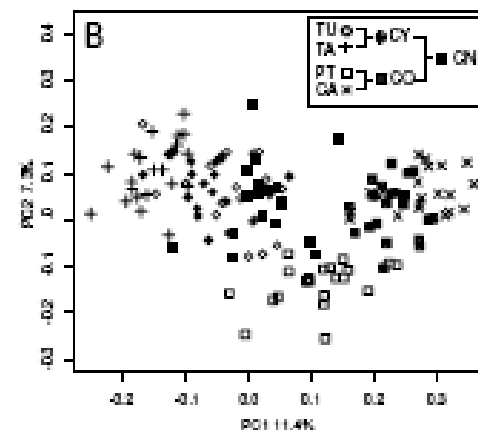
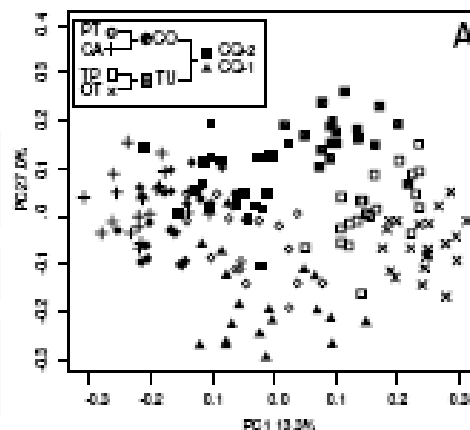
- 400 traditional races from the center of maize diversity in the CIMMYT germplasm bank
- 195 races are in active use *In-situ* and/or *Ex-Situ* (13,663 accessions , 11,200 accessions have been evaluated)

	Races	Regeneration data	Active In-Situ	Active Ex-Situ
Fingerprints	137	118	61	86
Fingerprints (>1)	42	31	19	32
Total	400	235	108	144

- ▶ Represent all active races
- ▶ Include more populations per race

Reif et al. 2005

- 24 maize races from Mexico (13-21 accessions/race)
- 25 SSRs
- Genetic variation within races
Gene diversity: 0.41 - 0.55
Fixation index: 0.09 - 0.47
- Conformation of racial complexes (Long and narrow-eared races, Northwestern's, low to medium, and high elevation)
- Support of the relationships proposed by Wellhausen (1952)



GCP Competitive Project #14

First year activities:

- Planning meeting in Nairobi (04/05)
 - Formalize communications procedures, lab. procedures, contracts, training, etc.
 - Decision on SSRs
- Create list of accessions (10/05)
- First data analyses workshop (12/05)
- Extraction and exchange of DNA (12/05)
- Begin fingerprinting (01/06)



Fingerprinting of maize populations

Bulk method (Dubreuil et al. 2005)

- Creation of bulks with equal amounts of DNA from 15 plants
- Fingerprinting using highly optimized SSRs
 - insensitive to extra bands (stutter bands, + A bands)
 - no competitive effects among alleles
 - high correlation coefficient between expected and estimated allele frequencies (≥ 0.85)

24 SSRs were optimized by INRA, 27 SSRs by CIMMYT

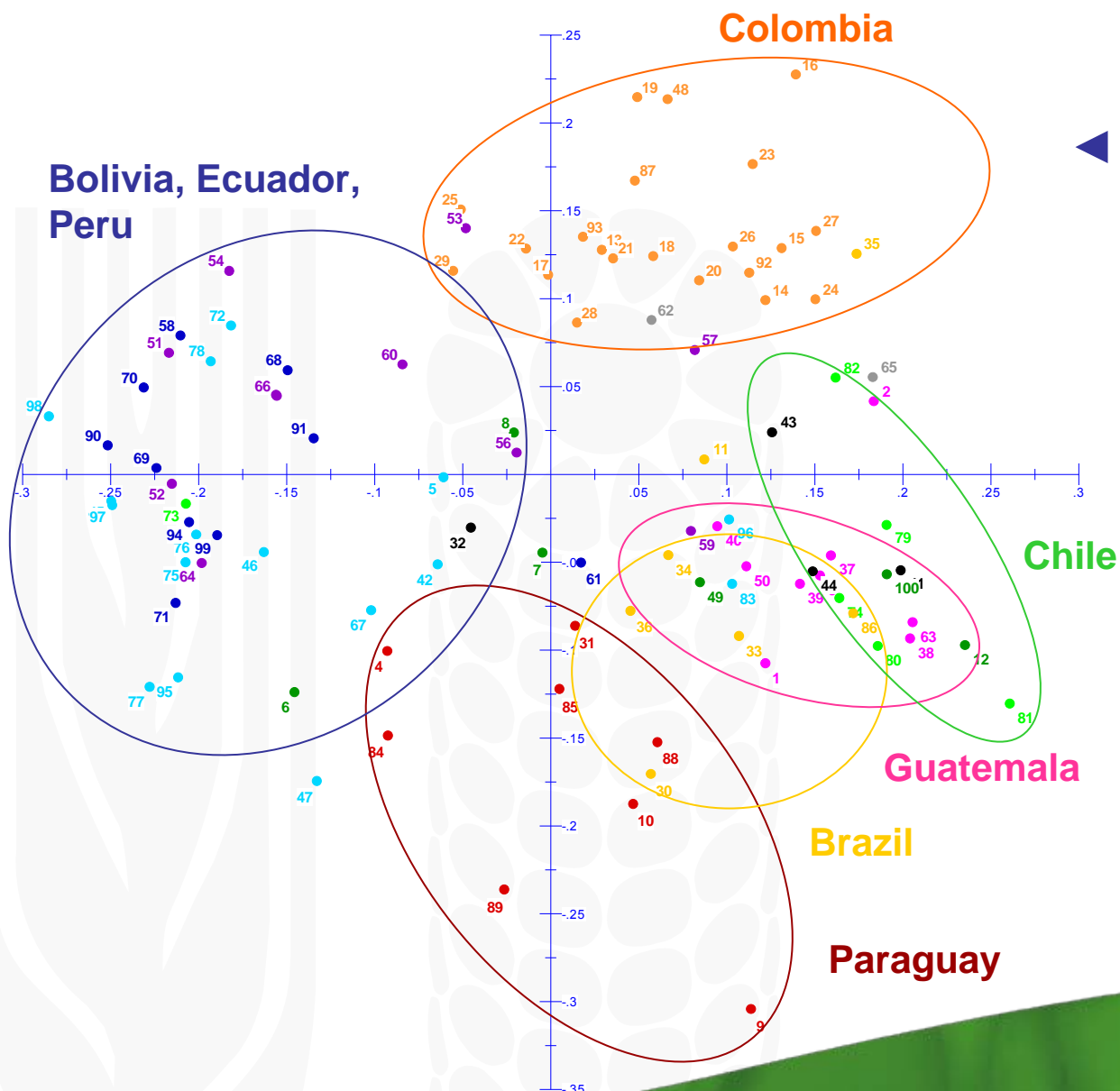
- Estimation of allele frequencies from the peak height of each band in the bulks

Preliminary analyses on 100 populations with 30 SSRs

- Central and South American landraces (Guatemala, Colombia, Ecuador, Peru, Bolivia, Paraguay, Brazil, Uruguay, Chile, Argentina)
- Summary statistics:

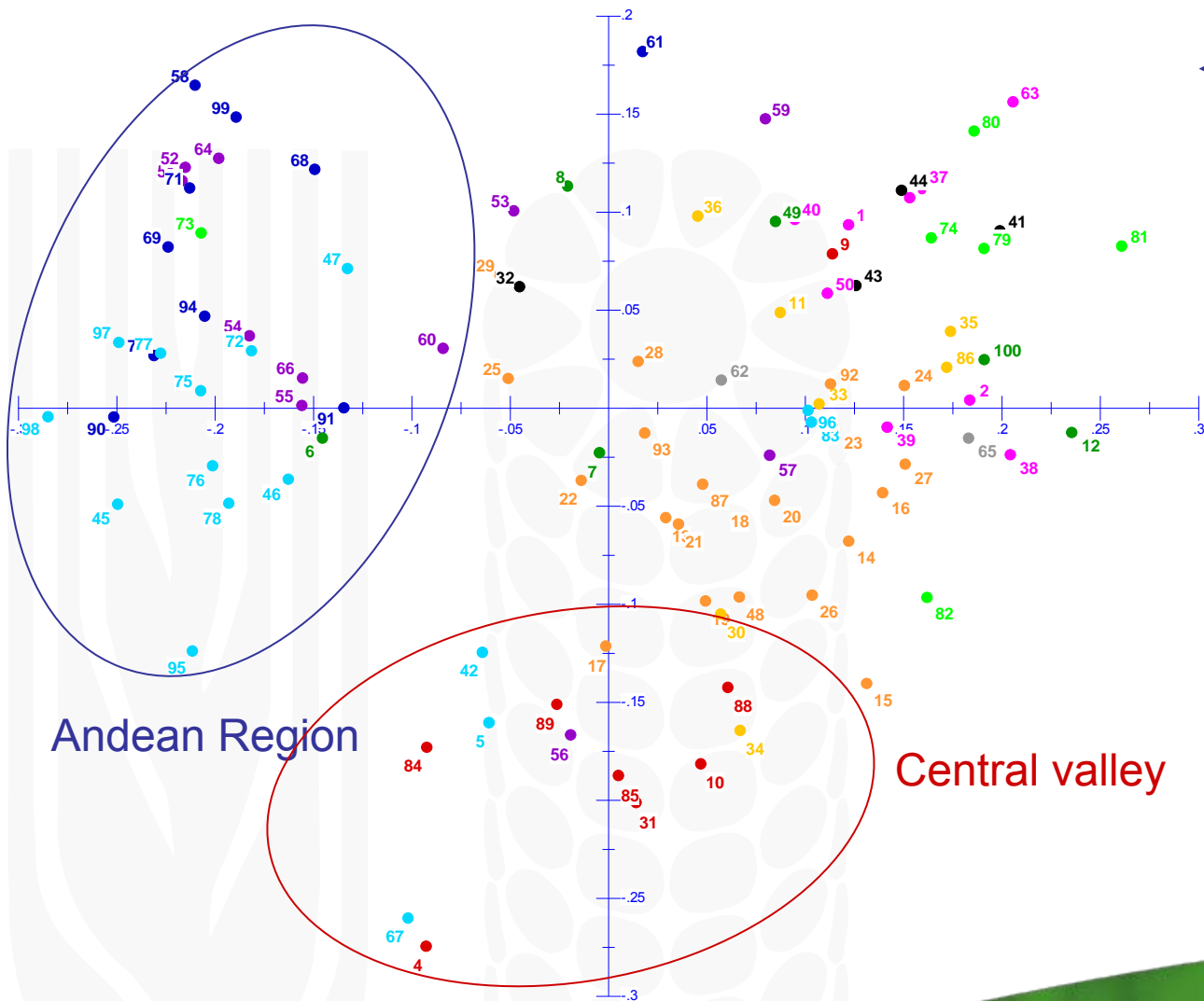
Missing data	12%
Total no. of alleles	245 (8.1)
No. of unique alleles	29 (11.8%)
No. of rare alleles (<5%)	72 (29.4%)
Range of MRD	0.162 (0.017 - 0.517)
PIC	0.59

Factorial analyses



◀ Grouping according the geographic origin

Factorial analyses



◀ Grouping according to absolute altitude

Andean Region

Central valley

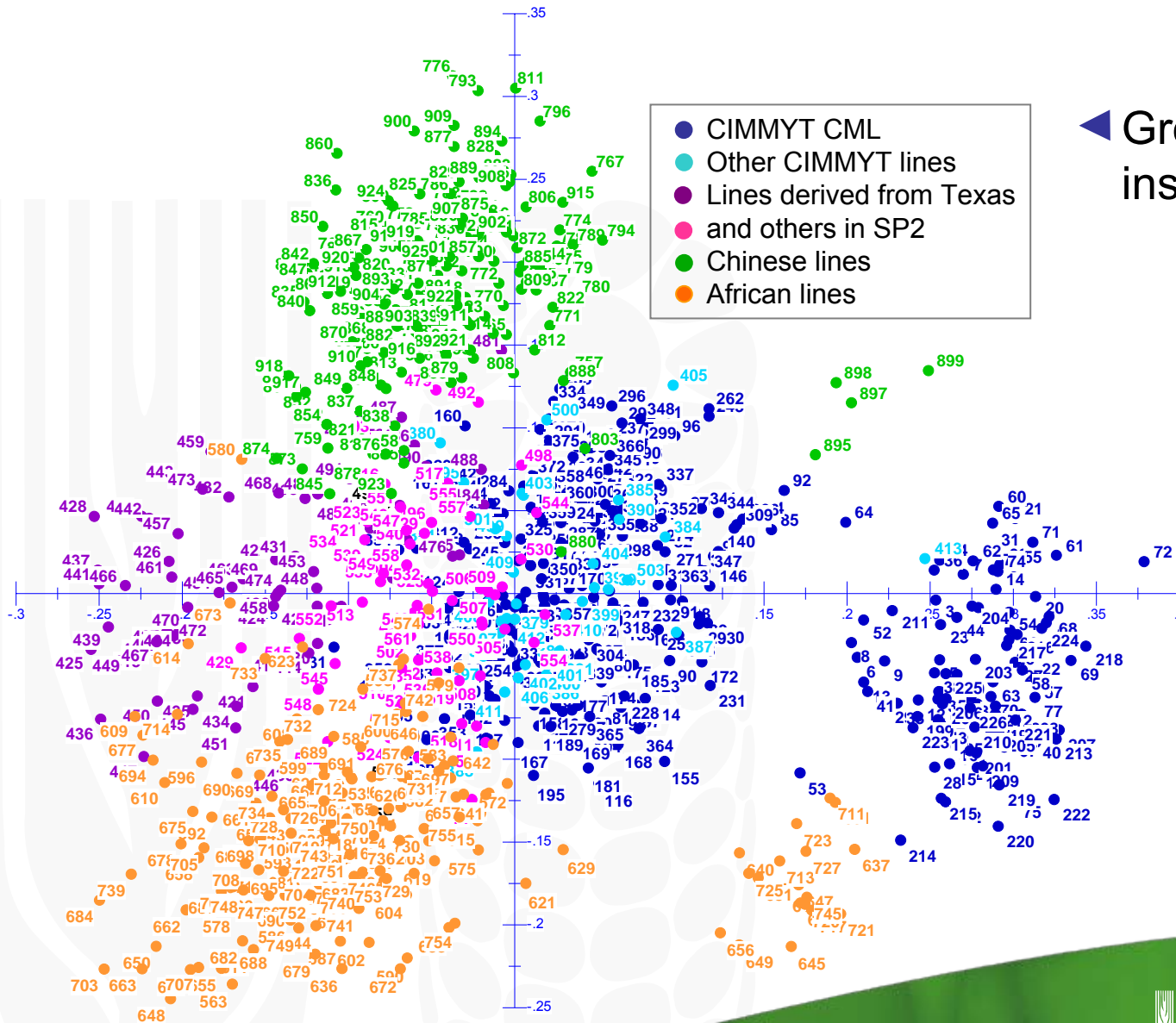
Preliminary analyses on 925 inbred lines with 36 SSRs

Materials:

482 CIMMYT lines (376 CML), 78 US-lines (Texas), 170 Chinese lines, 194 African lines,

	Mean	Min.	Max.
Missing data	10%	0.7%	22.8%
No. of alleles	461 (12.5)	3	40
No. of unique alleles	108	-	-
No. of rare alleles	329	-	-
Gene Diversity	0.602	0.307	0.859
PIC	0.57	0.27	0.85

Factorial analyses



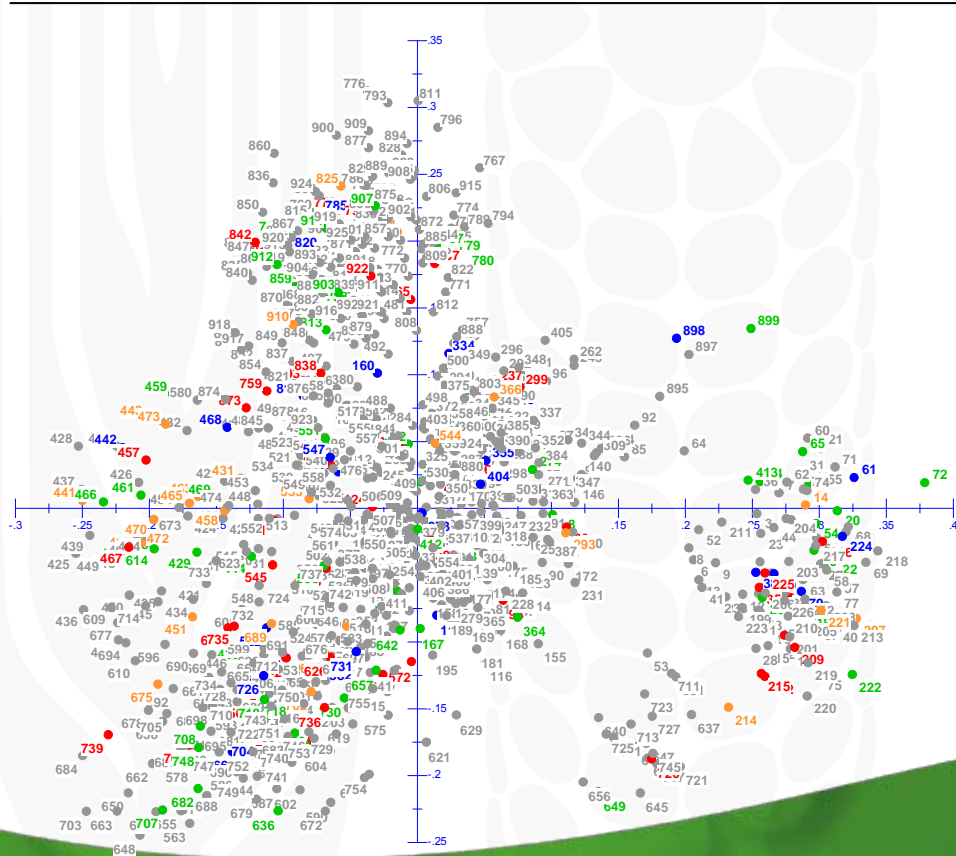
◀ Grouping according to institutional origin

Selection of “reference samples” according to Franco et al. 2005

- Sampling method: stratified sampling
 - Clustering method: UPGMA based on Cavalli-Sforza and Edwards distance (CED,1967)
 - Allocation method: D - Strategy; sample size of each cluster is drawn, proportional to the genetic distance within each cluster
- Sampling size: 50, 75, 100 genotypes
- Simulation of 50 possible cores for each size, Selection of the core containing highest genetic variation based on the values of CED and heterozygosity

Diversity measures of selected “reference samples”

Index	Size of reference sample			
	987	50	75	100
Gene Diversity	0.588	0.598	0.601	0.600
PIC	0.54	0.55	0.55	0.55
Heterozygosity	0.572	0.625	0.619	0.621
Cavalli-Sforza & Edwards distance	0.728	0.739	0.737	0.736



Reference samples with:

- 50 genotypes
- 75 genotypes
- 100 genotypes
- Genotypes present in more than one reference sample (29)

CIMMYT Fingerprinting Database

CIMMYT



The CIMMYT Maize and Wheat Fingerprinting Database – Version 2 with more data and better tools!
J. Monnard, M. Andrade, G. Davenport, S. Döring-Göckler and M. Warburton

CIMMYT Fingerprinting Database

The CIMMYT Fingerprinting Database includes fingerprinting data on all publicly-released experimental varieties/lines of maize and wheat from CIMMYT and collaborating scientists. The database has two volumes—one for maize data and one for wheat—and was developed using Microsoft Access, a package required to access the database.

Database Content

Several tables and forms make up the database front end. By clicking on "tables" under the object list, you will see a list of primary data tables. All are linked to each other. The data are divided by studies, which are listed in the study table. Two separate tables include genotypes and markers. The actual fingerprinting data is held in two additional tables, one for individuals (Marker_Data_Ind) and one for populations (Marker_Data_Pops). Populations were genotyped in bulk, which allows for more than two alleles per genotype/marker combination.

New features in version 2

A number of new features have been added in version 2. These include capabilities to:

- Generate Error Reports for and between studies.
- Query data by allele, genotype, marker and study (see the query interface below).
- Export a study into the GCP template format.
- Delete a study (useful for maintaining subsets of the data in different databases).

Menu system

The menu system will allow you to access different functionalities of the database. You open it by clicking on "Forms" and then double clicking on "Main" in the selection list. The following screen appears:



Main page



Query interface

Version 2.01:

- New datasets
- Generate error reports
- Query data by allele, genotype, marker and study
- Export a study into the GCP template format

www.genesilico.org/bioinformatics

SSR allele kit

Dr. Jompatong Chaba (National Corn and Sorghum Research Center, Thailand, Dr. Pichet Grudloyma)

- ❖ Screening of 1189 maize inbred lines from CIMMYT, and IITA for 43 SSR markers.
- ❖ Observation of 306 alleles which can be all represented with a total of 123 genotypes.
- ❖ Reamplification of all alleles to verify the alleles.
- ❖ Screening of 467 SSRs from CIMMYT and INRA for 15 SSRS



Acknowledgements

CIMMYT

M. Warburton
S. Taba
J. Franco
L. Diaz
A-L. Gomez
H. Lopez
C. Jampatong

Collaborators

A. Charcosset (INRA)
V. Mahalakshmi (IITA)
Yu Li (CAAS)
Zachary Muthamia (KARI)
S.H. Zhang (China)
B. M. Prasanna (Indian Agri. Research Inst.)
Sutrisno (Indonesian Department of Agri.)
P. Grudloyma (Nakhon Sawon Field Crops
Research Center)
P.- X. Hao (National Maize Research Inst.)
A. Salazar (Philippine Department of Agri.)