



**Genotyping of Global Cassava  
Genetic Resources;  
Determination of LD and  
Association Mapping for Dry  
Matter Content**

**Projects #02j and #08**

**CIAT, IITA, DArT PL,  
EMBRAPA, IPGRI**

# Cassava Objectives in SP1

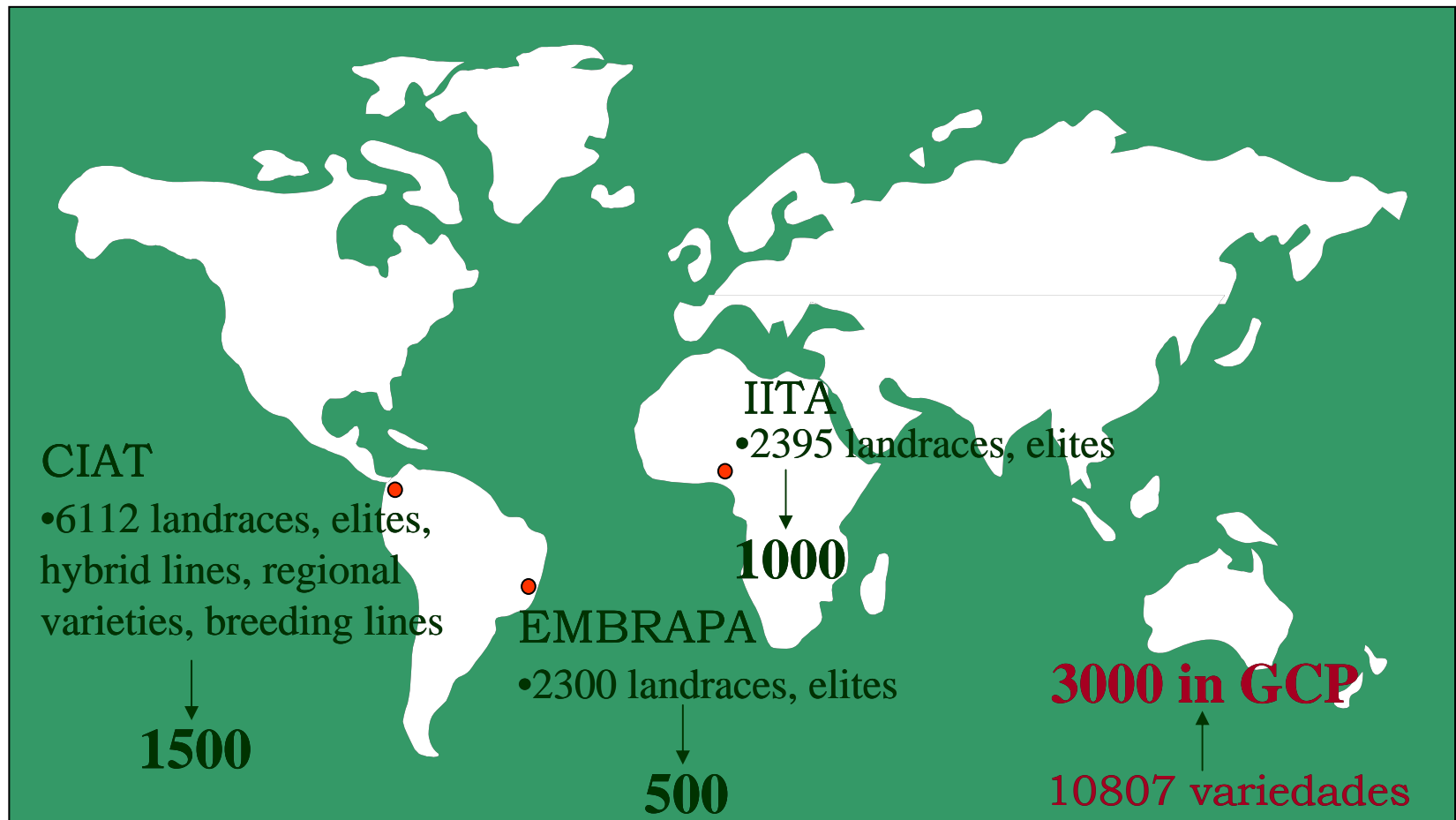
The background of the slide features a large, semi-transparent image of a cassava plant with its characteristic thick, tuberous roots and large, heart-shaped leaves. In the lower right corner, there is a smaller, semi-transparent image of a person wearing a wide-brimmed hat and smiling broadly, suggesting a positive agricultural or community context.

## Year one

- To build a "composite set" representing the diversity of cassava species from the genebanks (3000 accessions)
- To determine the structural genetic diversity of the composite set using molecular markers (DArTs/SSR)
- To select a "reference sample" for advanced characterization and association mapping studies

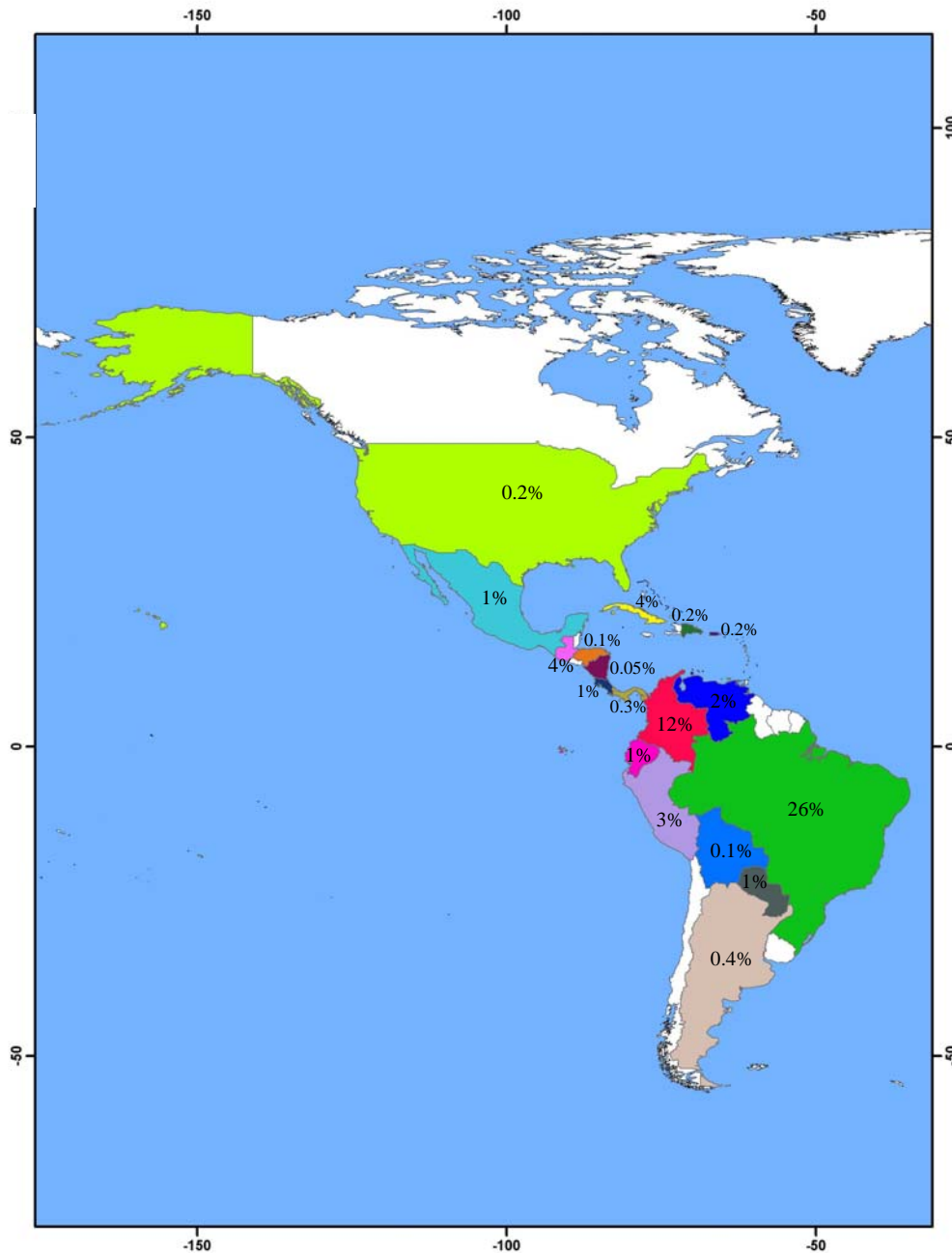
# Genetic Resources of Cassava-A Rich Heritage

Building the “composite set” of 3000 accessions from genebanks collections around the world



# AMERICAS (18)

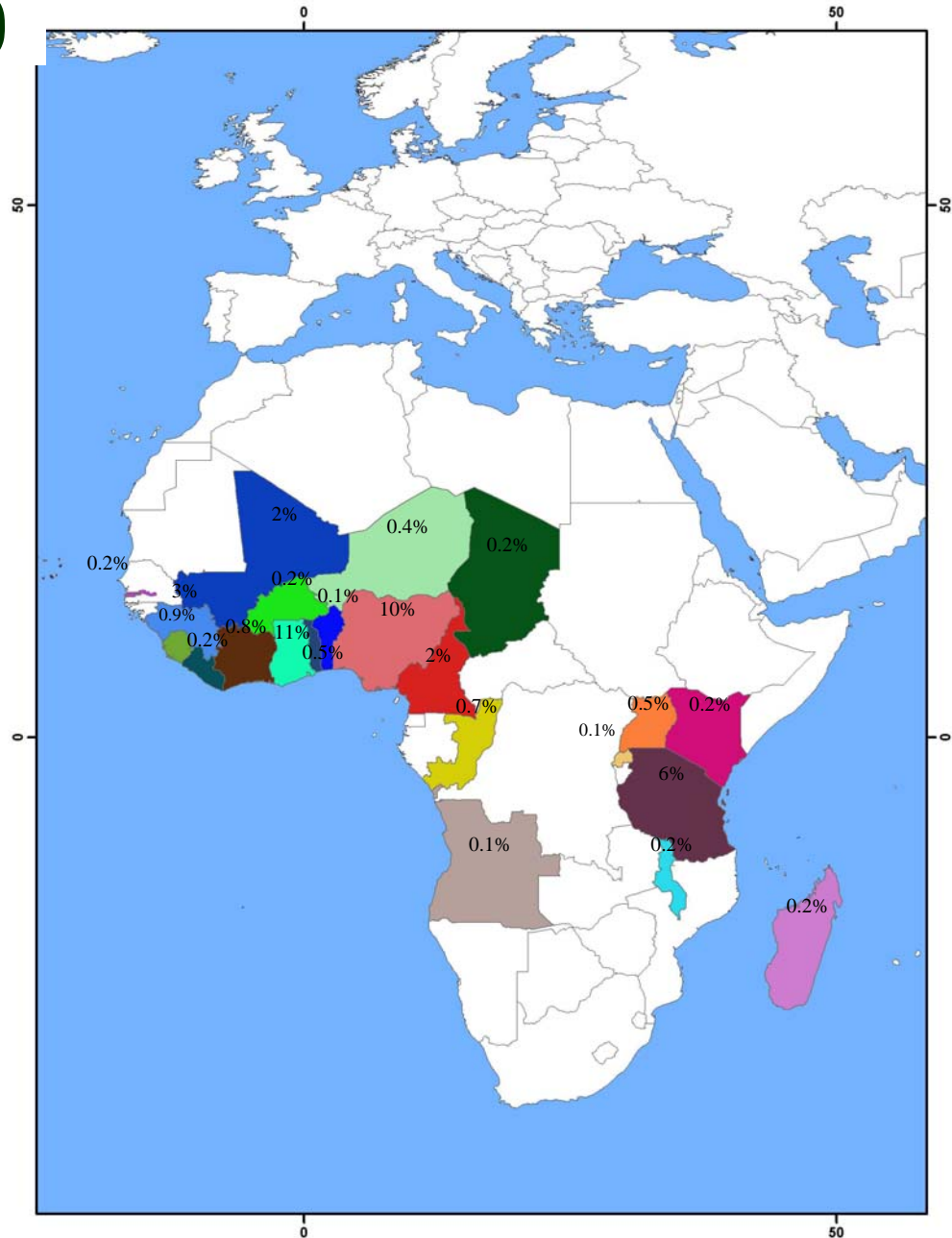
- Argentina
- Bolivia
- Brazil
- Colombia
- Costa Rica
- Cuba
- Dominican Republic
- Ecuador
- Guatemala
- Honduras
- Mexico
- Nicaragua
- Panama
- Paraguay
- Peru
- Puerto Rico
- United States
- Venezuela



# AFRICA (23)

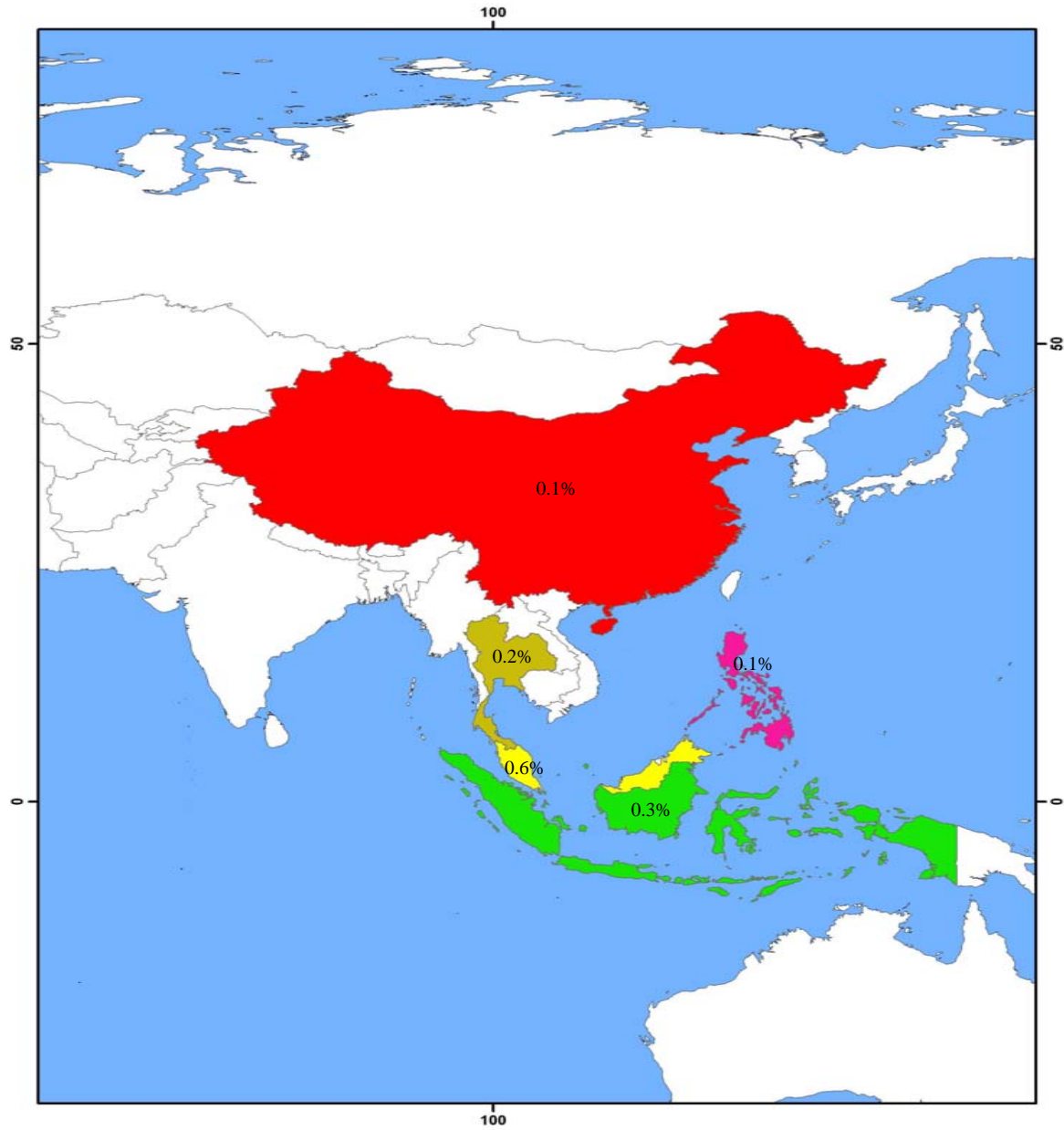
## AFRICA

- Angola
- Benin
- Burkina Faso
- Cameroon
- Cape Verde
- Chad
- Congo
- Gambia
- Ghana
- Guinea
- Ivory Coast
- Kenya
- Liberia
- Madagascar
- Malawi
- Mali
- Niger
- Nigeria
- Rwanda
- Sierra Leone
- Tanzania
- Togo
- Uganda



# ASIA (6)

- China
- Fiji
- Indonesia
- Malaysia
- Philippines
- Thailand



# Structural Diversity in Cassava

The background of the slide features a large, semi-transparent image of a cassava plant with its characteristic thick, brown tubers and green leaves. In the lower right corner, there is a smaller, semi-transparent image of a person wearing a wide-brimmed hat and smiling broadly, looking upwards.

- A pilot study to compare DArT and SSR Markers in structural characterization of 436 accessions
- A diversity study to infer the population structure of 3000 cassava accessions using 36 SSR markers

# Pilot Studies to Assess Marker Systems in Analyzing Structural Diversity

## SSRs – 36 SSR Markers (CIAT)

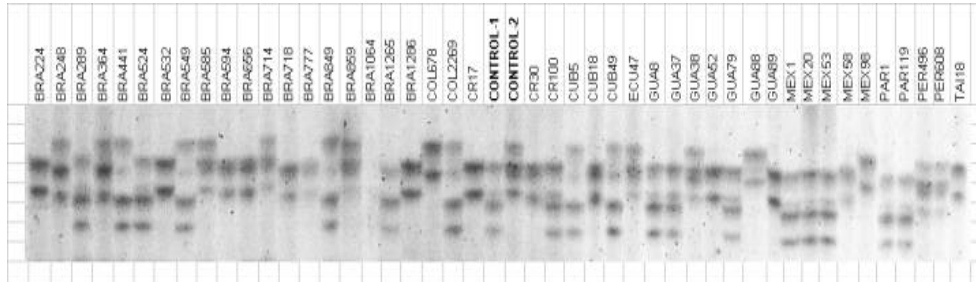
- Codominant
- Highly polymorphic
- Development costs  
~US\$36000 per 36 markers
- Low Throughput
  
- Cost of getting genetic information /locus / genotype: US\$0.5/data point

## DArTs: Diversity Array of ~1000 clones (DArT P/L)

- Dominant
- Moderately polymorphic
- Development costs  
~ US\$ 20000 per chip (1000 polymorphic clones)
- High Throughput
- Cost of getting genetic information /locus / genotype: US\$0.025

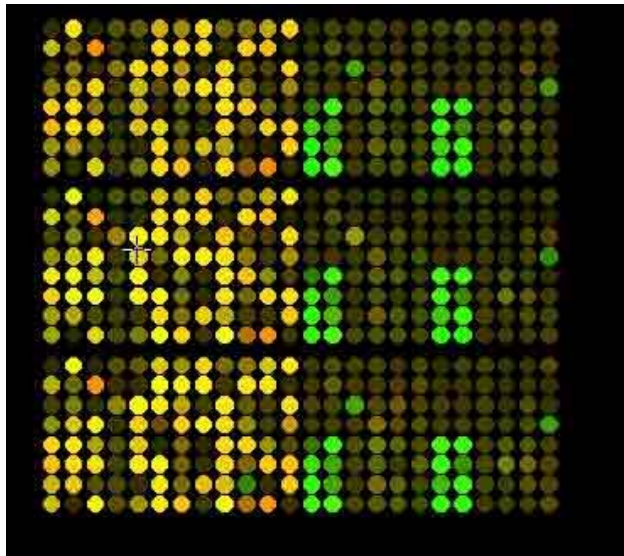
# Pilot Study Results

## SSR



Presence of alleles/locus/genotype

## DArT



Binary scores for genotypes analysed at polymorphic spots

1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1
2	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0
3	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
4	0	0	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1
5	1	1	0	0	1	1	1	1	0	0	0	0	1	1	0	0	0	0
6	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	0	0	1
8	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	0	0	0	0	0	1	1	1	1	1	0	0	1	1
14	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
17	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
18	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
19	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
20	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
21	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
22	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1
23	1	1	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0
24	1	1	0	0	1	1	1	1	0	0	1	1	0	0	0	0	0	0
.	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1
.	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1

Binary matrix

# Pilot Study: Statistical Analysis

- **Principal Coordinate Analysis (PCoA)**

  - Binary Matrix

  - Similarity Matrix

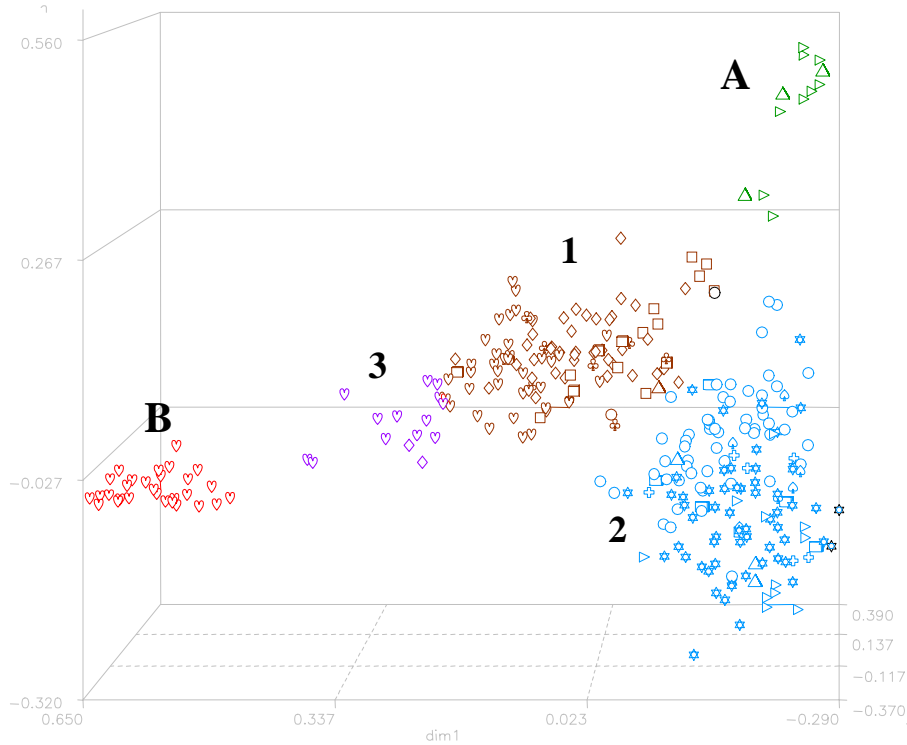
  - Jaccard's* similarity Index

- **Clustering analysis**

- **Similarity between and within the groups defined by PCoA**

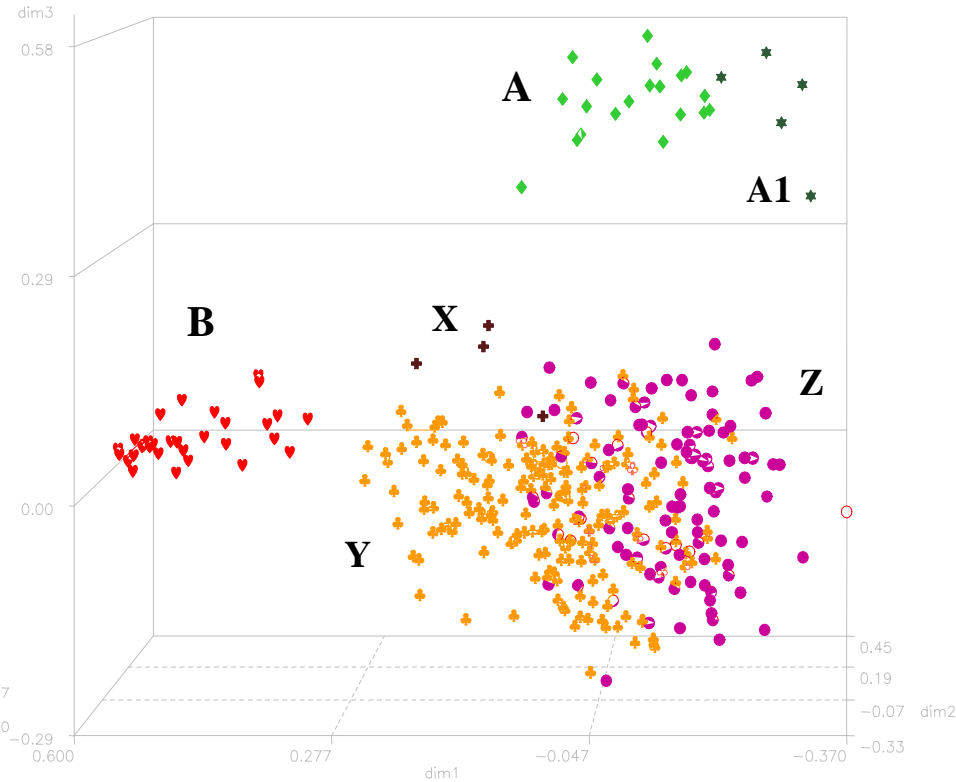
# Pilot Study PCoA

## SSR



GROUP	COUNTRY	FREQUENCY (%)
1	CAMEROUN/ NIGERIA	37.1/37.9
<b>B</b>	NIGERIA	100
<b>A</b>	GUATEMALA	80
2	BRASIL	86.7
3	NIGERIA	86.6

## DArT



GROUP	COUNTRY	FREQUENCY (%)
Z	GHANA	12
X	NIGERIA	100
Y	BRASIL	23
<b>A</b>	GUATEMALA	80
<b>A1</b>	GUATEMALA	40
<b>B</b>	NIGERIA	93

# **Cassava Genotyping of 3000 Accessions Using 36 SSR markers**

## **CIAT-EMBRAPA**

- 2580 genotypes  
evaluated with 22 of 22  
SSR markers
- Raw SSR data set:
  - 2494 genotypes (80-  
100% complete data)
  - N. of alleles: 6-16

## **IITA**

- 2575 genotypes  
evaluated with 8 of 14  
SSR markers
- Raw SSR data set:
  - 1974 genotypes (50-  
100% complete data)
  - N. of alleles: 6-24

# Statistical Analysis

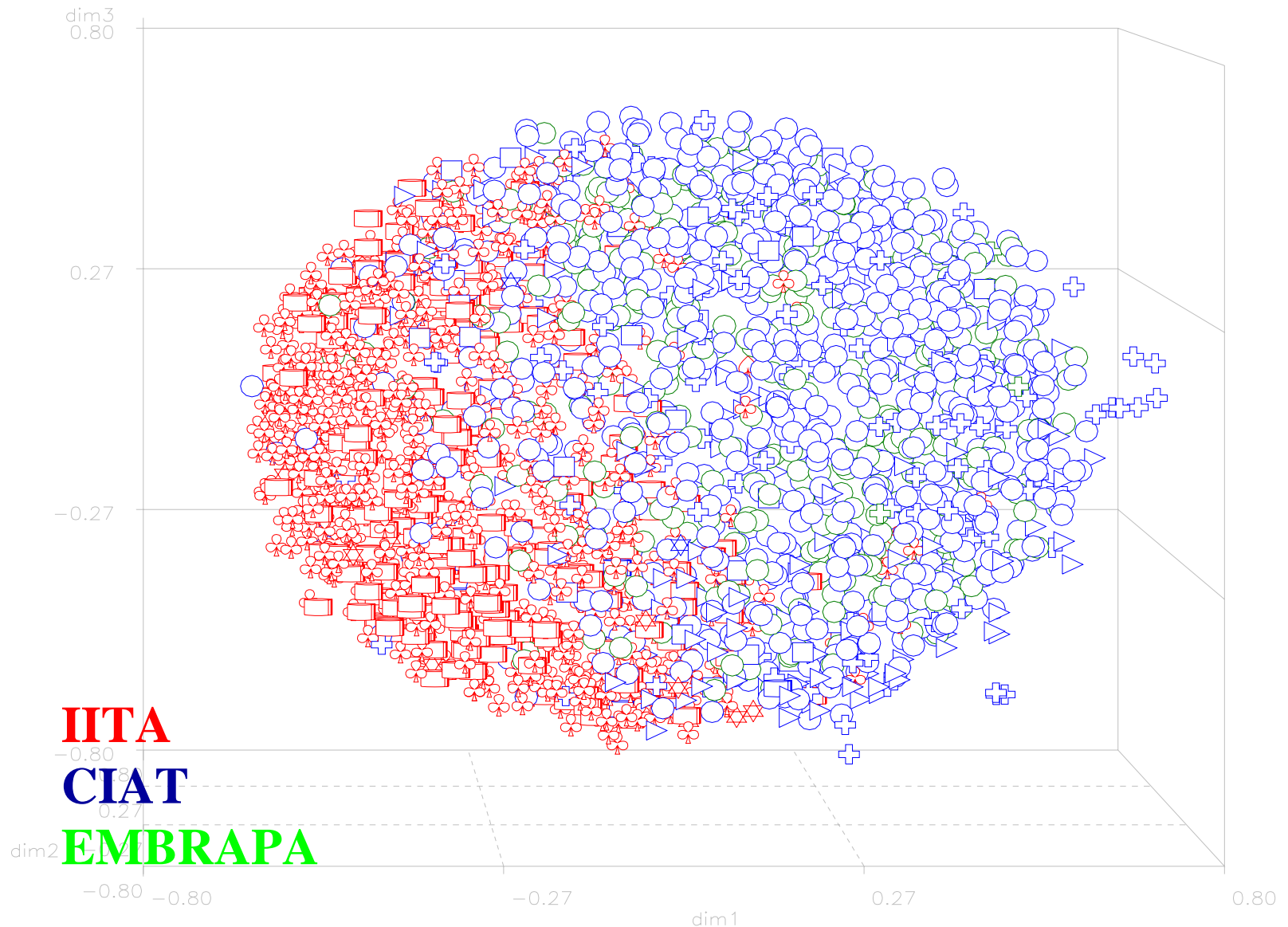
- **Principal Coordinate Analysis (PCoA): based on individuals**

Based on Jaccard's similarity index

- **Cluster Analysis: based on countries**

- **UPGMA**, using Nei's (1972) genetic identity and genetic distance

# PCoA of Global Cassava Germplasm

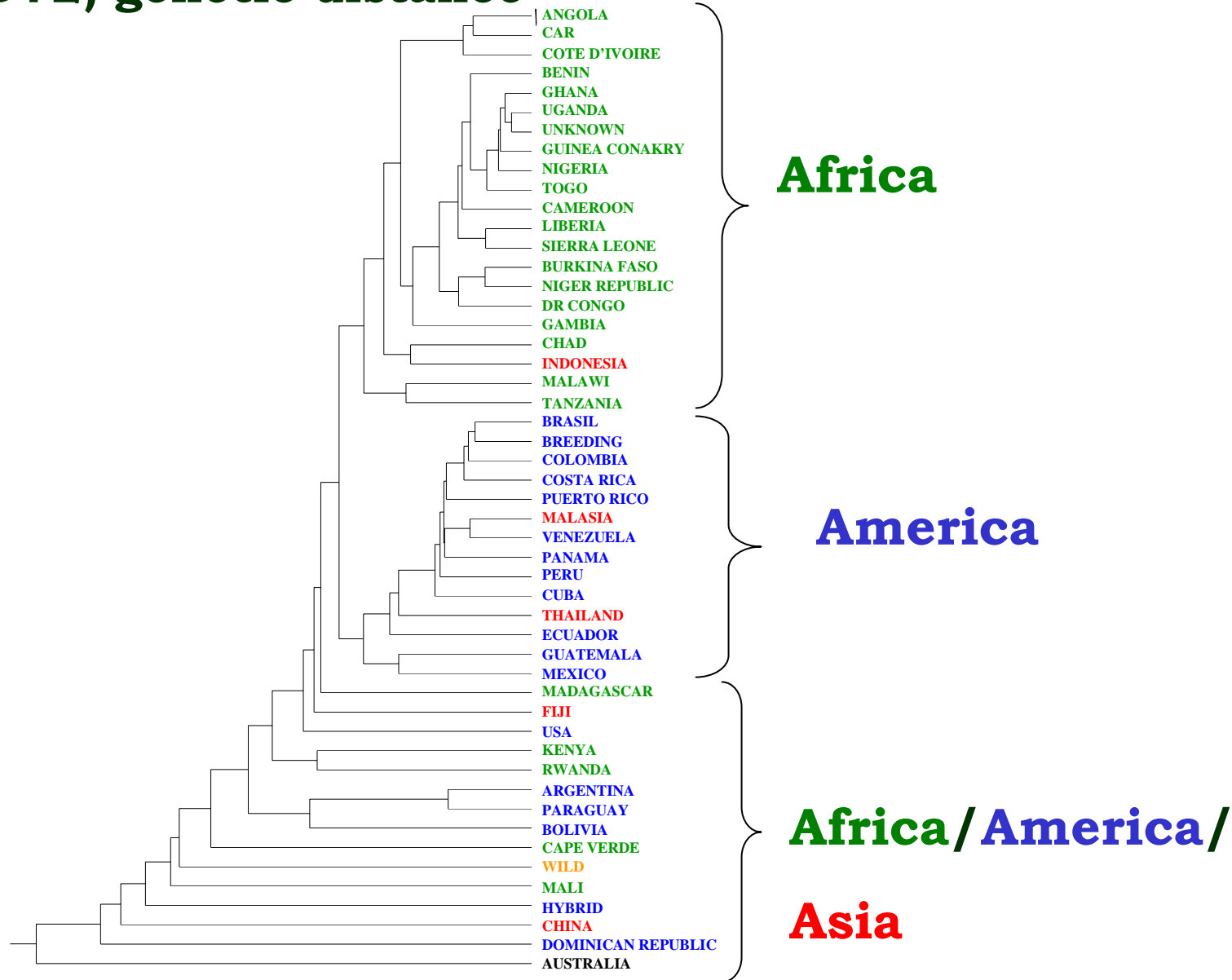






# UPGMA Cluster Analysis (based on countries)

Nei's (1972) genetic distance



# Summary

A background image showing a person wearing a wide-brimmed hat, holding a large, thick, brown cassava root. The person is smiling and looking upwards. The image is semi-transparent and serves as a backdrop for the text.

- **The results agree with previous diversity studies in Cassava (structure is largely explained by the accession's origin)**
- **The additional structure within American and African could be due to introgression from wild relatives, mutation, selection, or small sample sizes**
- **Genotyping with 36 SSR markers needs to be completed to better define structure**

# What Next?

- **CIAT-EMBRAPA:** 420 genotypes are being processed to fill missing data (end of September 2005)
- **IITA:** 6 SSR markers remain to complete the genotyping (end of September 2005)
- **CIAT:** To finish the analysis of the complete data set
  - To analyze the data set from IITA based on previous diversity studies to reduce the allele number per locus
  - PCoA of 3000 accessions evaluated with 36 SSR markers to define the fine structural diversity
  - To calculate the similarity between and within the clusters

# What Next?

The background of the slide features a large, semi-transparent image of a cassava plant with its characteristic thick, brown roots and green leaves. In the lower right corner, there is a smaller, semi-transparent image of a person wearing a wide-brimmed hat and smiling broadly, looking upwards.

- To calculate parameters of diversity for different countries and identify niches of rich diversity
- To use a sub-sampling methodology to select accessions looking for association mapping with superior alleles for different traits
- To validate Cassava DArT array at a gene-level resolution by making a high-density survey of diversity among some breeding lines



**COMMISSIONED PROJECT**

**Population structure, phenotypic  
information and association  
studies in long-generation crops**

**Cassava update**

**SP1C2**



# Specific Objectives

1. Compilation of phenotypic data from regional trials of a selected group of 110 varieties and 25 breeding lines from databases of the cassava program
2. Determination of LD in the cassava genome via genotyping the above materials with SSR markers on a genome-wide basis, 200 SSR markers, every 10cM
3. Association mapping of dry matter content using candidate starch genes and the SSCP-SNP marker evaluation of the 135 selected varieties

# Selection of Plant Materials

## 30 years of cassava breeding

❖ Started with great genetic variation: more than 2000 genotypes from the Neotropics

## Large number of elite lines

❖ 800 breeding lines, elite parents and landraces evaluated over several years in multi-locations and replicated regional trials

## Selection for DMC

200 elite or breeding lines having at least 4 years of replicated DMC data



# Dry Matter Content

- ❖ Correlated with dry matter yield
- ❖ Determines starch content (85-90% is starch)
- ❖ A key selection parameter in cassava breeding



# Preliminary Determination of LD

**117 accessions included in SP1 (year one)**  
**out of 800 in the breeding program**

(42 landraces-elite / 75 advanced breeding lines)

**Pilot study**

436 accessions x 36  
SSR markers

**57 accessions**

9 landraces-elite / 48  
breeding lines

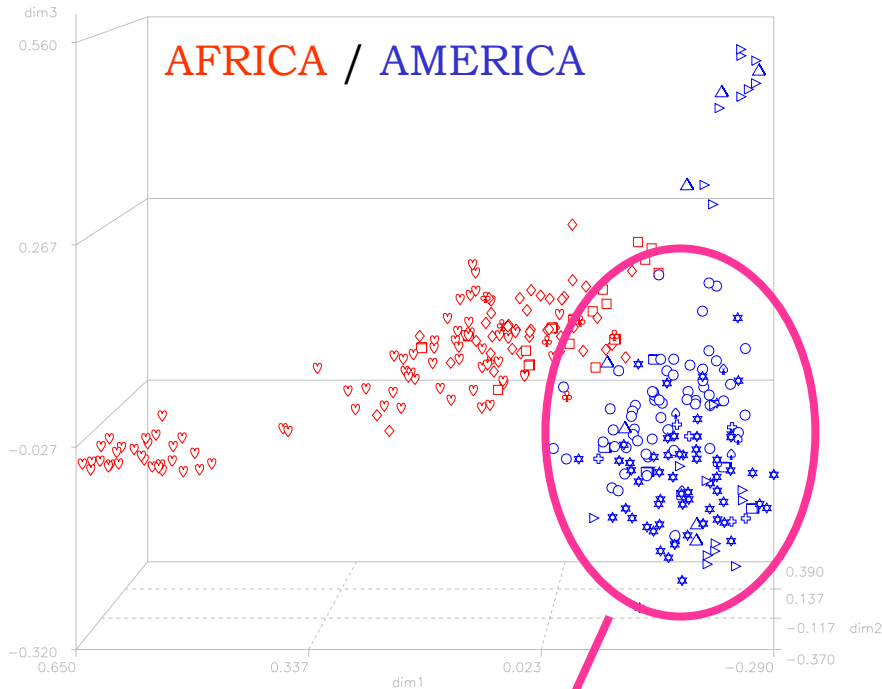
**Diversity GCP**

2494 accessions x  
22 SSR markers

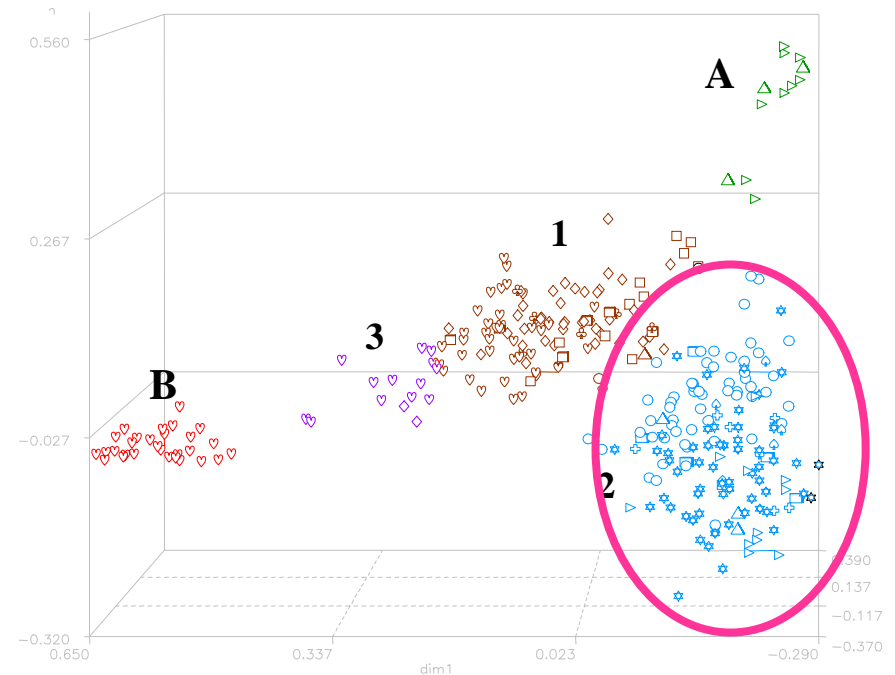
117 accessions

42 landraces-Elite / 75  
breeding lines

# Preliminary Determination of LD



**57 accessions  
(282 alleles)**



GROUP	COUNTRY	FREQUENCY (%)
1	CAMEROUN/ NIGERIA	37.1/37.9
<b>B</b>	<b>NIGERIA</b>	100
<b>A</b>	<b>GUATEMALA</b>	80
2	BRASIL	86.7
3	NIGERIA	86.6

# Linkage Disequilibrium

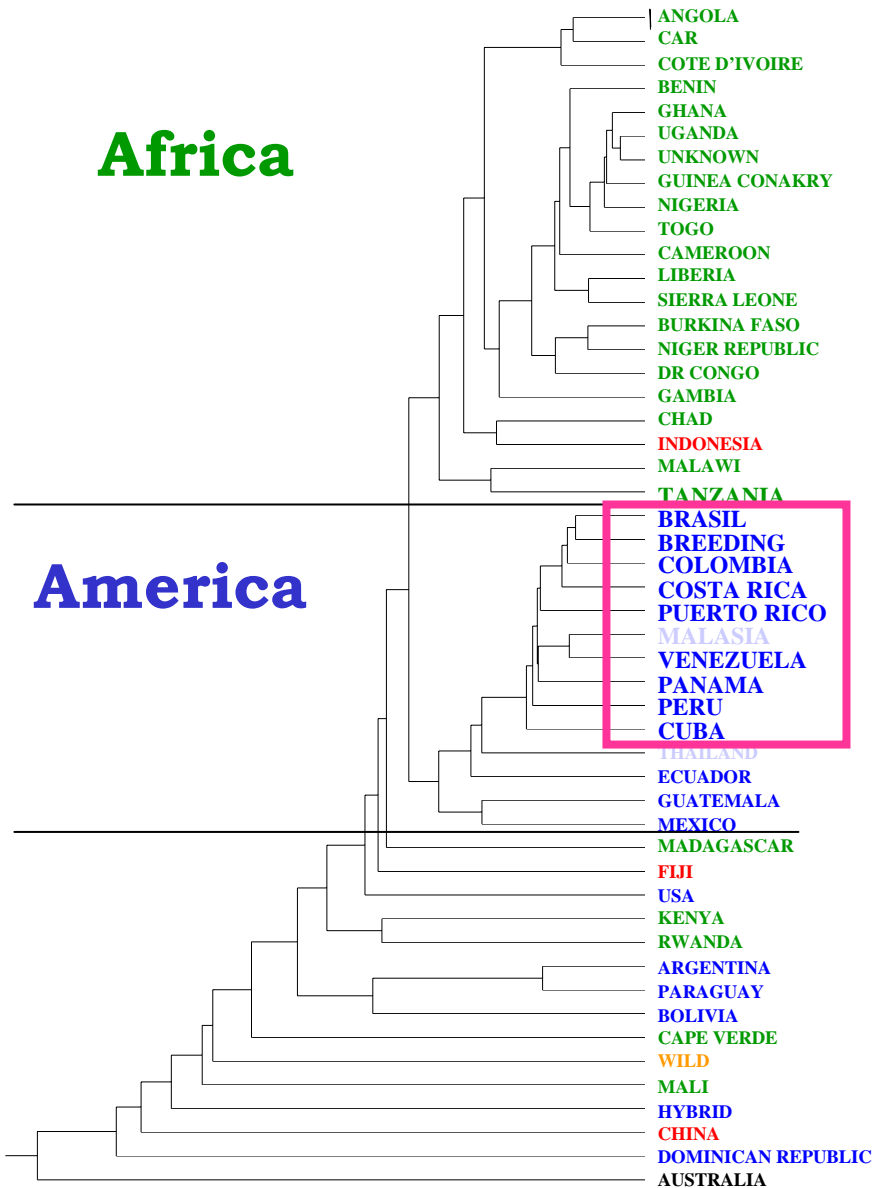
## 57 accessions from pilot study

[Locus	Allele]	-	[Locus	Allele]	Burrows	Correlation	Chisq*	Prob.
[SSRY4	A	-	[SSRY181	F	0.0106	0.3169	4.42	0.0356
[SSRY9	D	-	[SSRY59	B	0.0228	0.3211	4.95	0.0261
[SSRY12	C	-	[SSRY21	E	0.0329	0.2830	4.25	0.0394
[SSRY12	D	-	[SSRY52	D	0.0710	0.4845	12.68	0.0004
[SSRY12	F	-	[SSRY52	F	0.0550	0.5000	13.50	0.0002
[SSRY12	F	-	[SSRY169	D	0.0308	0.3116	5.15	0.0233
[SSRY19	H	-	[SSRY181	F	0.0048	0.3499	6.12	0.0133
[SSRY20	F	-	[SSRY52	C	0.0179	0.3070	4.62	0.0316
[SSRY20	F	-	[SSRY100	A	0.0132	0.3191	4.99	0.0255
[SSRY20	A	-	[SSRY110	D	0.0061	0.5000	10.00	0.0016
[SSRY20	A	-	[SSRY161	J	0.0059	0.5000	10.25	0.0014
[SSRY20	A	-	[SSRY177	H	0.0064	0.3486	4.50	0.0340
[SSRY20	H	-	[SSRY181	F	0.0106	0.4128	7.67	0.0056
[SSRY21	D	-	[SSRY64	D	0.0268	0.3164	5.30	0.0213
[SSRY51	G	-	[SSRY59	G	0.0221	0.3382	5.60	0.0179
[SSRY52	F	-	[SSRY169	D	0.0308	0.3116	5.15	0.0233
[SSRY59	D	-	[SSRY100	D	0.0689	0.2924	4.19	0.0407
[SSRY59	F	-	[SSRY171	D	0.0102	0.3181	4.66	0.0309
[SSRY59	I	-	[SSRY181	F	0.0101	0.3176	4.54	0.0331
[SSRY64	E	-	[SSRY110	A	0.0055	0.3493	5.25	0.0220
[SSRY64	D	-	[SSRY179	D	0.0197	0.3055	4.39	0.0362
[SSRY82	H	-	[SSRY105	G	0.0176	0.2874	3.96	0.0465
[SSRY100	A	-	[SSRY169	D	0.0167	0.3021	4.84	0.0279
[SSRY100	A	-	[SSRY171	F	0.0203	0.2817	4.05	0.0442
[SSRY100	E	-	[SSRY161	I	0.0305	0.4371	8.79	0.0030
[SSRY105	A	-	[SSRY103	G	0.0189	0.3426	5.40	0.0202
[SSRY110	D	-	[SSRY161	J	0.0069	0.5000	8.75	0.0031
[SSRY110	D	-	[SSRY177	H	0.0073	0.3478	3.87	0.0491
[SSRY110	C	-	[SSRY179	D	0.0131	0.3715	4.97	0.0258
[SSRY169	B	-	[SSRY171	E	0.0133	0.2901	4.29	0.0383
[SSRY169	C	-	[SSRY171	G	0.0089	0.3014	4.63	0.0314
[SSRY171	D	-	[SSRY179	D	0.0203	0.3199	4.71	0.0300
[SSRY161	J	-	[SSRY177	H	0.0068	0.5000	9.00	0.0027
[SSRY161	J	-	[SSRY182	B	0.0107	0.3045	4.08	0.0434
[SSRY177	H	-	[SSRY182	B	0.0229	0.3639	5.03	0.0249

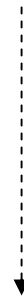
Significant  
LD in 35  
allele  
pairs

The number of significant ( $P < 0.0500$ ) linkage disequilibria (LD) = 35.  
\* All Chisquare tests have one degree of freedom.

# Preliminary Determination of LD



**(2494 accessions x 22 SSR)**



**117 accessions  
(212 alleles)**

# Linkage disequilibrium

## 117 accessions from GCP

[Locus	Allele]	-	[Locus	Allele]	Burrows	Correlation	Chisq*	Prob.
[SSRY12	E	-	[SSRY20	D	0.0023	0.2452	6.37	0.0116
[SSRY12	A	-	[SSRY52	A	0.0608	0.2383	6.13	0.0133
[SSRY12	C	-	[SSRY52	C	0.0321	0.3138	10.64	0.0011
[SSRY12	D	-	[SSRY52	D	0.0858	0.4970	26.67	0.0000
[SSRY12	E	-	[SSRY52	E	0.0023	0.3319	13.37	0.0003
[SSRY12	F	-	[SSRY52	F	0.0374	0.4866	25.57	0.0000
[SSRY12	B	-	[SSRY69	D	0.0286	0.2202	5.04	0.0247
[SSRY12	F	-	[SSRY82	I	0.0043	0.2360	5.96	0.0146
[SSRY12	E	-	[SSRY151	I	0.0022	0.2453	6.50	0.0108
[SSRY19	K	-	[SSRY21	I	0.0048	0.5059	26.62	0.0000
[SSRY19	K	-	[SSRY32	E	0.0023	0.2451	6.31	0.0120
[SSRY19	D	-	[SSRY59	A	0.0047	0.1991	4.04	0.0444
[SSRY19	L	-	[SSRY82	C	0.0024	0.3318	12.87	0.0003
[SSRY19	D	-	[SSRY100	I	0.0095	0.3206	10.48	0.0012
[SSRY19	K	-	[SSRY100	E	0.0024	0.5000	25.50	0.0000
[SSRY19	K	-	[SSRY108	B	0.0047	0.2886	8.66	0.0032
[SSRY19	K	-	[SSRY110	F	0.0025	0.2447	5.75	0.0163
[SSRY19	K	-	[SSRY171	H	0.0024	0.3518	12.37	0.0004
[SSRY20	B	-	[SSRY38	E	0.0023	0.2860	8.83	0.0030
[SSRY20	O	-	[SSRY51	B	0.0023	0.2452	6.37	0.0116
[SSRY20	K	-	[SSRY59	J	0.0089	0.2393	5.84	0.0157
[SSRY20	O	-	[SSRY59	A	0.0072	0.4776	23.26	0.0000
[SSRY20	P	-	[SSRY64	I	0.0044	0.2278	5.44	0.0197
[SSRY20	F	-	[SSRY82	J	0.0105	0.2239	5.11	0.0237
[SSRY20	F	-	[SSRY106	A	0.0169	0.2268	5.40	0.0201
[SSRY20	F	-	[SSRY108	C	0.0356	0.2139	4.67	0.0307
[SSRY20	H	-	[SSRY108	F	0.0521	0.2098	4.49	0.0341
[SSRY20	B	-	[SSRY110	F	0.0026	0.3518	11.62	0.0007
[SSRY20	K	-	[SSRY110	F	0.0026	0.2117	4.21	0.0401
[SSRY20	B	-	[SSRY151	K	0.0024	0.3518	12.75	0.0004
[SSRY20	D	-	[SSRY151	E	0.0023	0.2450	6.19	0.0129
[SSRY20	E	-	[SSRY151	K	0.0046	0.2186	4.92	0.0265
[SSRY20	M	-	[SSRY155	E	0.0112	0.1929	3.91	0.0481
[SSRY20	D	-	[SSRY169	F	0.0045	0.2606	7.06	0.0079
[SSRY21	I	-	[SSRY32	E	0.0047	0.3360	13.18	0.0003
[SSRY21	I	-	[SSRY59	I	0.0093	0.2269	5.15	0.0232
[SSRY21	F	-	[SSRY64	H	0.0139	0.2310	5.60	0.0179
[SSRY21	J	-	[SSRY69	K	0.0047	0.2027	4.15	0.0416
[SSRY21	E	-	[SSRY82	A	0.0146	0.2641	7.11	0.0076
[SSRY21	J	-	[SSRY82	C	0.0048	0.3576	13.05	0.0003
[SSRY21	J	-	[SSRY100	J	0.0094	0.2018	4.07	0.0436
[SSRY21	I	-	[SSRY106	H	0.0091	0.2250	5.32	0.0211
[SSRY21	E	-	[SSRY110	F	0.0053	0.3654	12.42	0.0004
[SSRY21	E	-	[SSRY151	A	0.0150	0.2063	4.39	0.0363
[SSRY21	I	-	[SSRY151	I	0.0087	0.2078	4.45	0.0350

Significant  
LD in 92  
allele  
pairs

The number of significant ( $P < 0.0500$ ) linkage disequilibria (LD) = 92.

\* All chisquare tests have one degree of freedom.



# Determination of LD in the Cassava Genome

Selection of 200 SSR markers evenly distributed in the genome

(Would need to generate SSR markers for some regions)



Evaluate the 200 elite/breeding lines with 200 SSR markers



Estimation of LD across the cassava genome

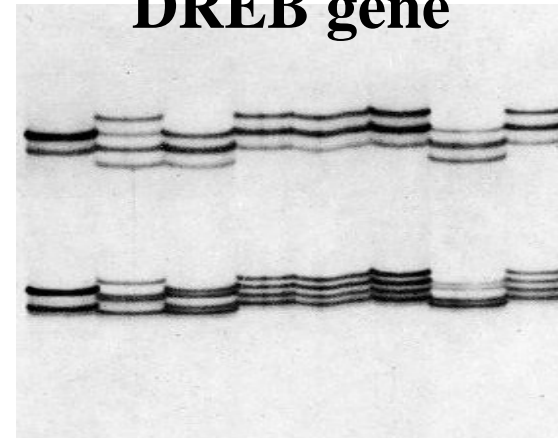
# Future Perspectives

- ❖ Is the sub-sampling methodology appropriate?
- ❖ Is the total sample size adequate?
- ❖ What is a most appropriate methodology or software to calculate LD?
- ❖ Do we need a physical map of clonally propagated crops via fingerprinting of BAC clones to accelerate LD mapping for traits of unknown genes?

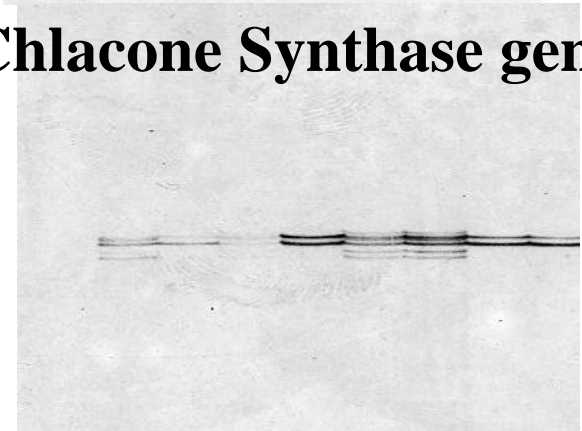
# Association Mapping of DMC

- ❖ 64 starch biosynthesis genes as candidate genes
- ❖ Evaluation of the 200 elite/breeding lines with the candidate genes as SSCP-SNP markers
- ❖ Association mapping using 'Structure' to determine genes linked to DMC

**DREB gene**



**Chlacone Synthase gene**



# Acknowledgements

- **Genetic Resources Unit, CIAT**

Graciela Mafla

Daniel Debouck

- **Cassava Genetics, CIAT**

Cesar Ospina

Jaime Marin

Charles Buitrago

Wilson Castelblanco

Luis Guillermo

Adriana Alzate

Samira Moreno

Martin Fregene

- **SB2 Agrobiodiversity Project, CIAT**

Myriam Duque

- **IPGRI, Cali**

Carmen De Vicente

- **IITA, Ibadan**

Yvonne Lokko

Mahalaskmi Visvatan

- **DArT PL**

Peter Wenzel

Andrzej Killian

- **Field Crop Research Station, Rayong**

Prapit Wongtiem

- **EMBRAPA**

Claudia Fortes Ferreira

Alfredo Alves