Plant Science into Practice

Genetic Diversity within CIAT's Cassava Germplasm Collection

Bruno A. Santos[1], Mohamed Abdelhalim[1], Pradeep Ruperao[1], David Marshall[2], Peter Wenzl[3] & Sarah Dyer [1]

1 NIAB, Cambridge, UK
2 Scotland's Rural College, Edinburgh, UK
3 Centro Internacional de Agricultura Tropical, Palmira, Colombia

Background

Cassava is a globally important food crop and staple for >500 million people in Africa, Asia and South America, and is also used for feed, starch and biofuel. The natural diversity contained within CIAT's genebank collection may provide novel sources of biotic stress resistance and useful alleles linked to traits of interest.

CIAT's cassava collection

CIAT's Genetic Resources Program houses 6,643 accessions of cassava (*Manihot esculenta*) and its wild relatives. Relatively little is known about the accessions within the genebank, aside from their collection information (fig 1).

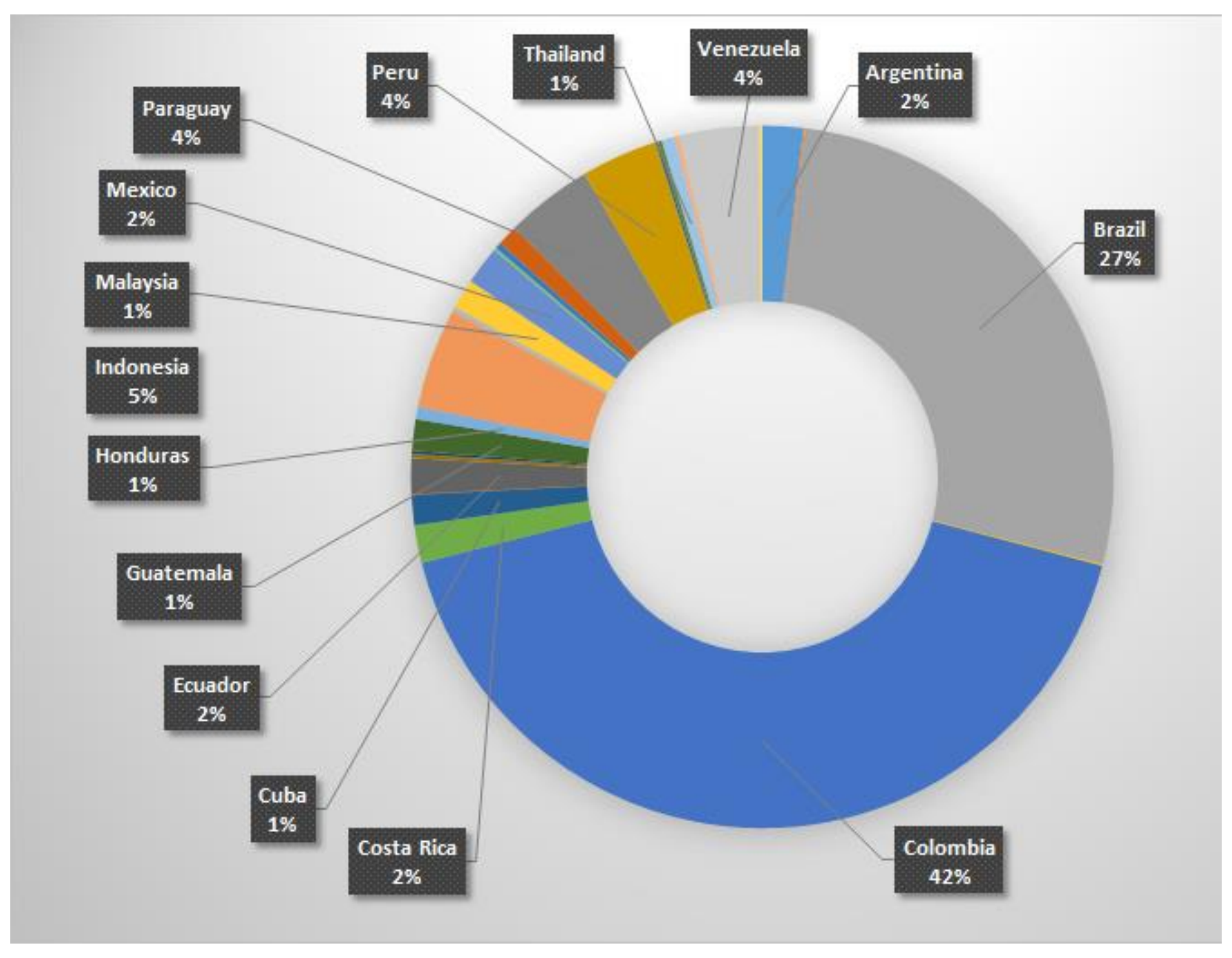


Figure 1: Accessions genotyped grouped by country of origin

Genotyping 4,000 accessions

The cassava genome is diploid and ~770Mb in size. Cassava suffers from in-breeding depression and is clonally propagated, as such the samples are expected to be highly heterozygous. We have **genotyped 4,074 accessions** from the collection using **DArT-Seq**. The data set also includes 178 sample replicates and 725 DArT technical replicates.

We received 75,548 raw SNP calls and 74,524 presence/absence variations. We applied the following filters: Removed samples with failed replicates, wild samples, and samples with >15% missing data; removed loci with <80% call rate or <98% reproducibility (calculated using 725 DArT technical replicates) and removed monomorphic loci.

Reason	Samples removed
Failed replicates	22
Wild samples	51
>15% missing data	47

Reason	Loci removed
<80% call rate	11,824
<98% reproducibility	3,166
Monomorphic loci	27,850

Generating **32,708 high confidence loci for 4,835** samples (3,979 accessions).

Principal Coordinate Analysis

Using R package labdsv we performed a PCO using a distance matrix generated from the genotyping data using DArT-R (fig 2).

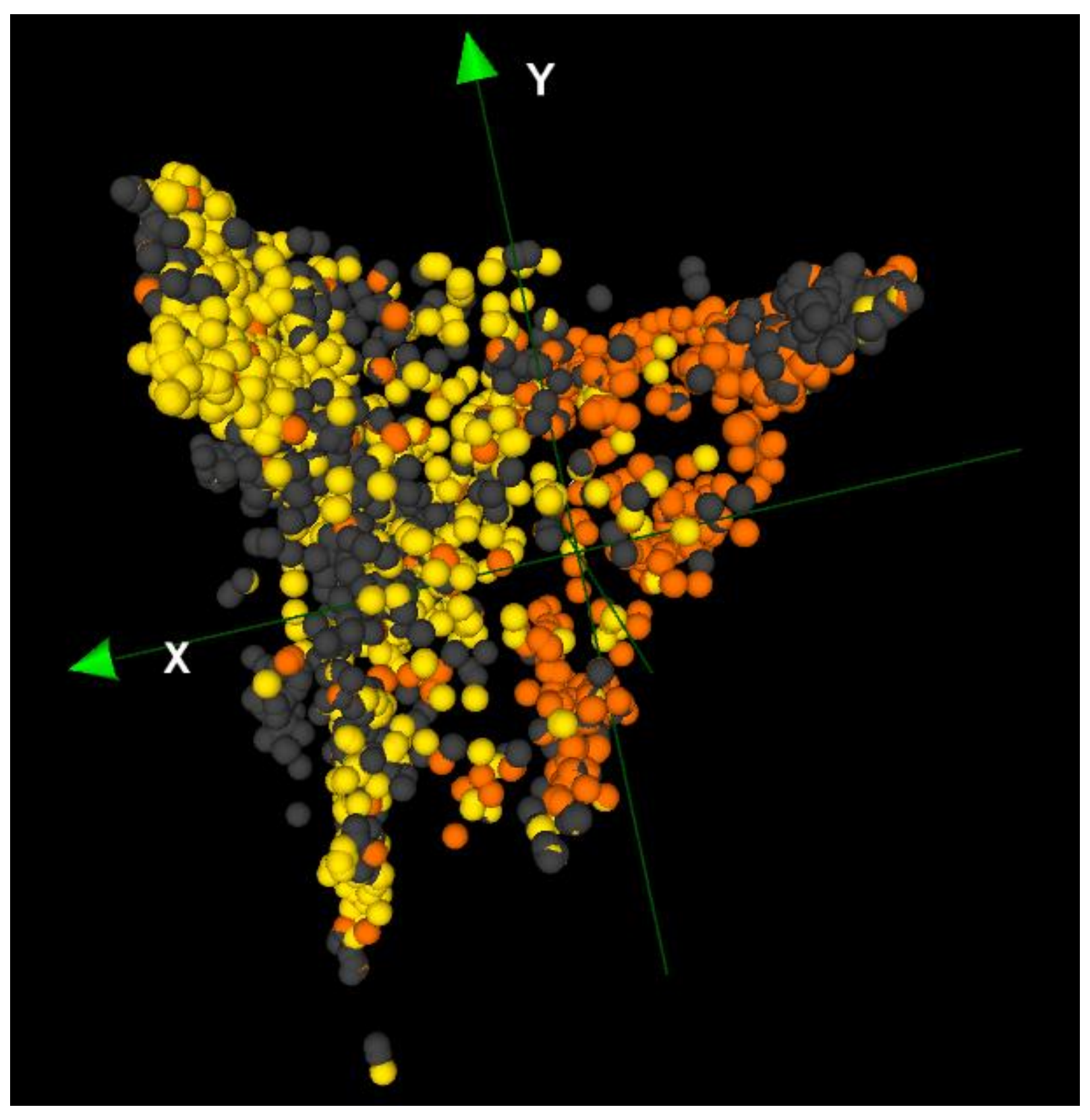


Figure 2: 3D Curlywhirly [1] plot of PCO genotype distances. Samples from Colombia (yellow) and Brazil (orange) are coloured, samples from all other countries are shown in grey

Of the 32,708 loci, 95% (31,207 loci) align to the cassava reference genome (at least one allele, fig 3).

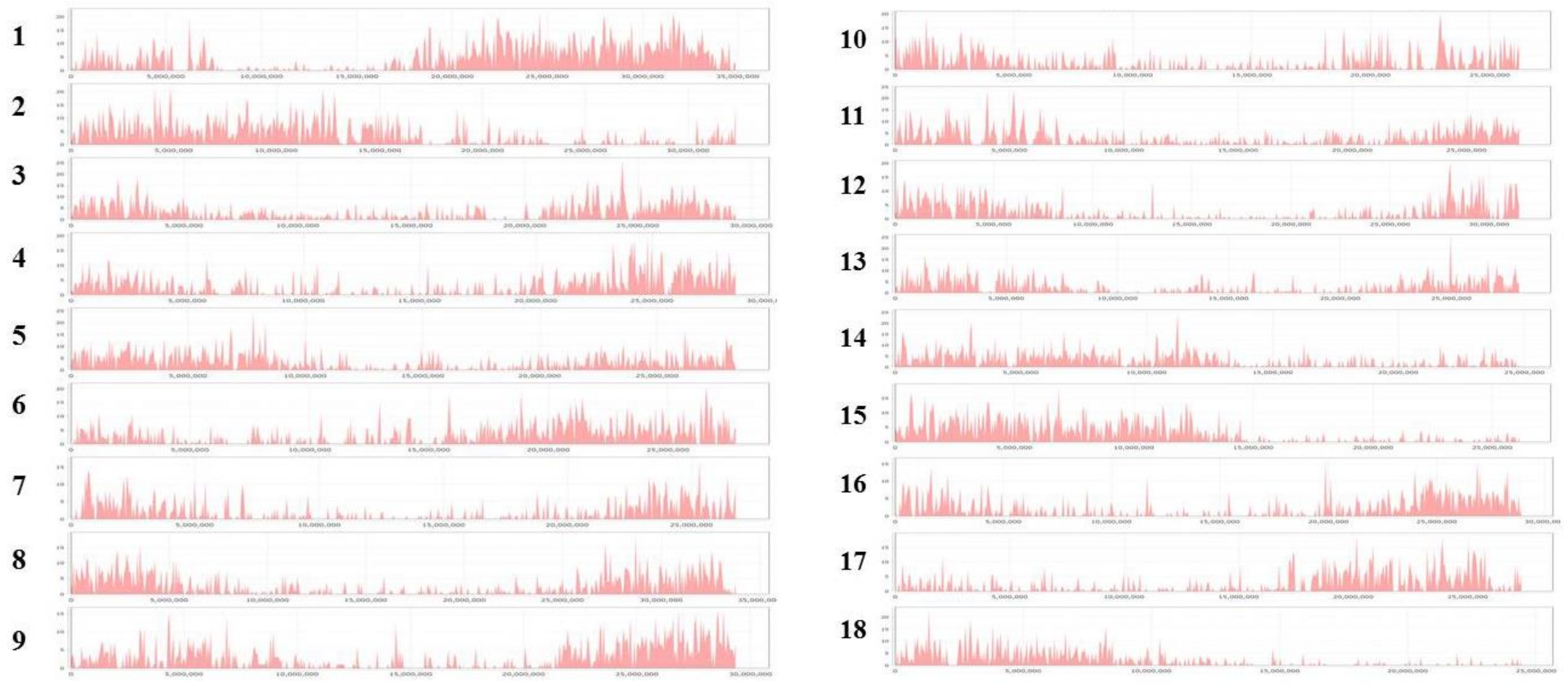


Figure 3: IGV [2] plot showing density of loci aligned to cassava reference v6.1 [3]

Whole genome sequencing

We have selected 25 samples for further investigation based on known responses to pests (thrips, whitefly and greenmites), use as parental lines for breeding, frequency of requests from the genebank and genetic diversity using the genotyping data. These genetically diverse individuals are being whole genome sequenced to allow us to explore their genomic diversity more fully (fig 4).

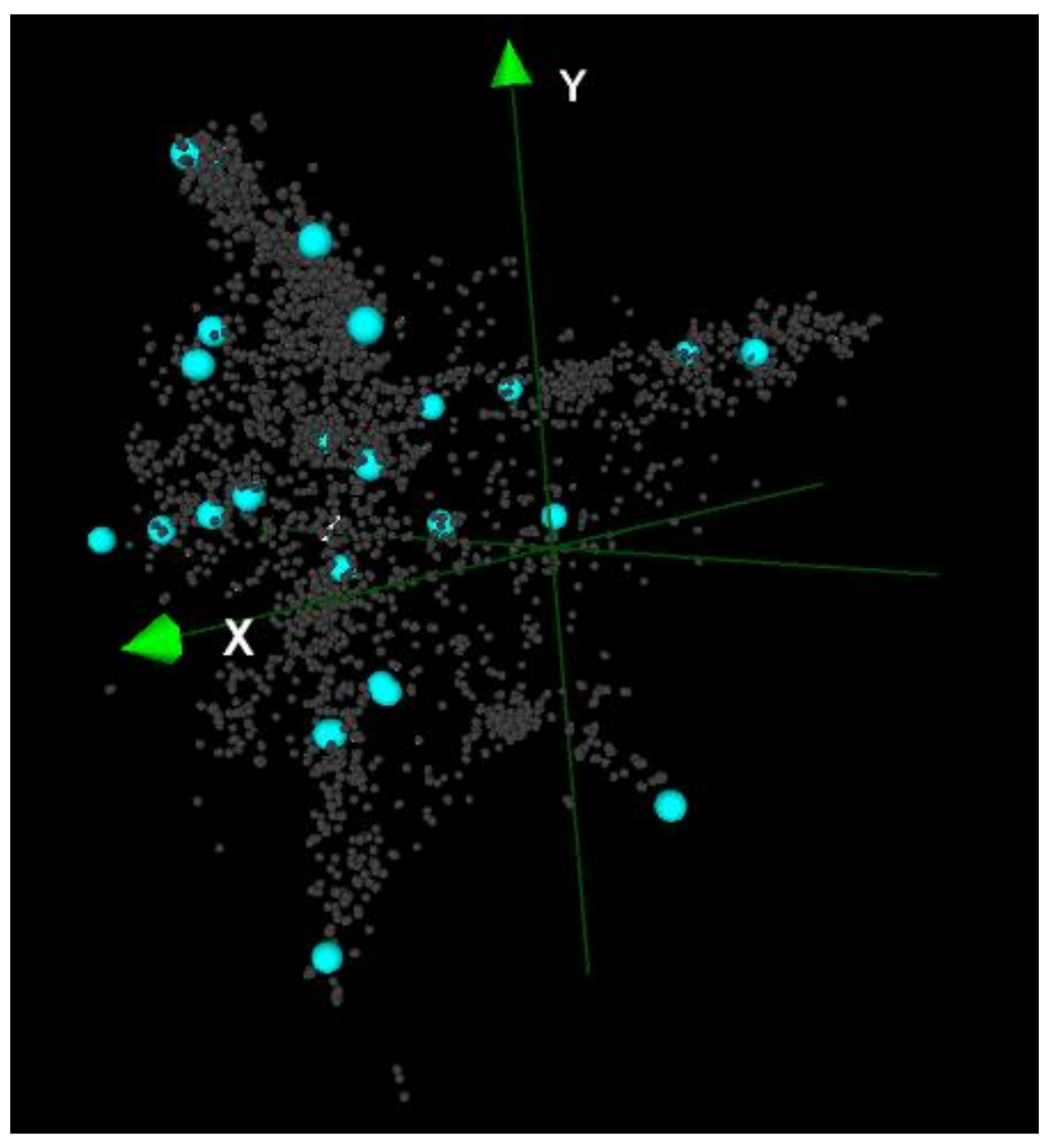
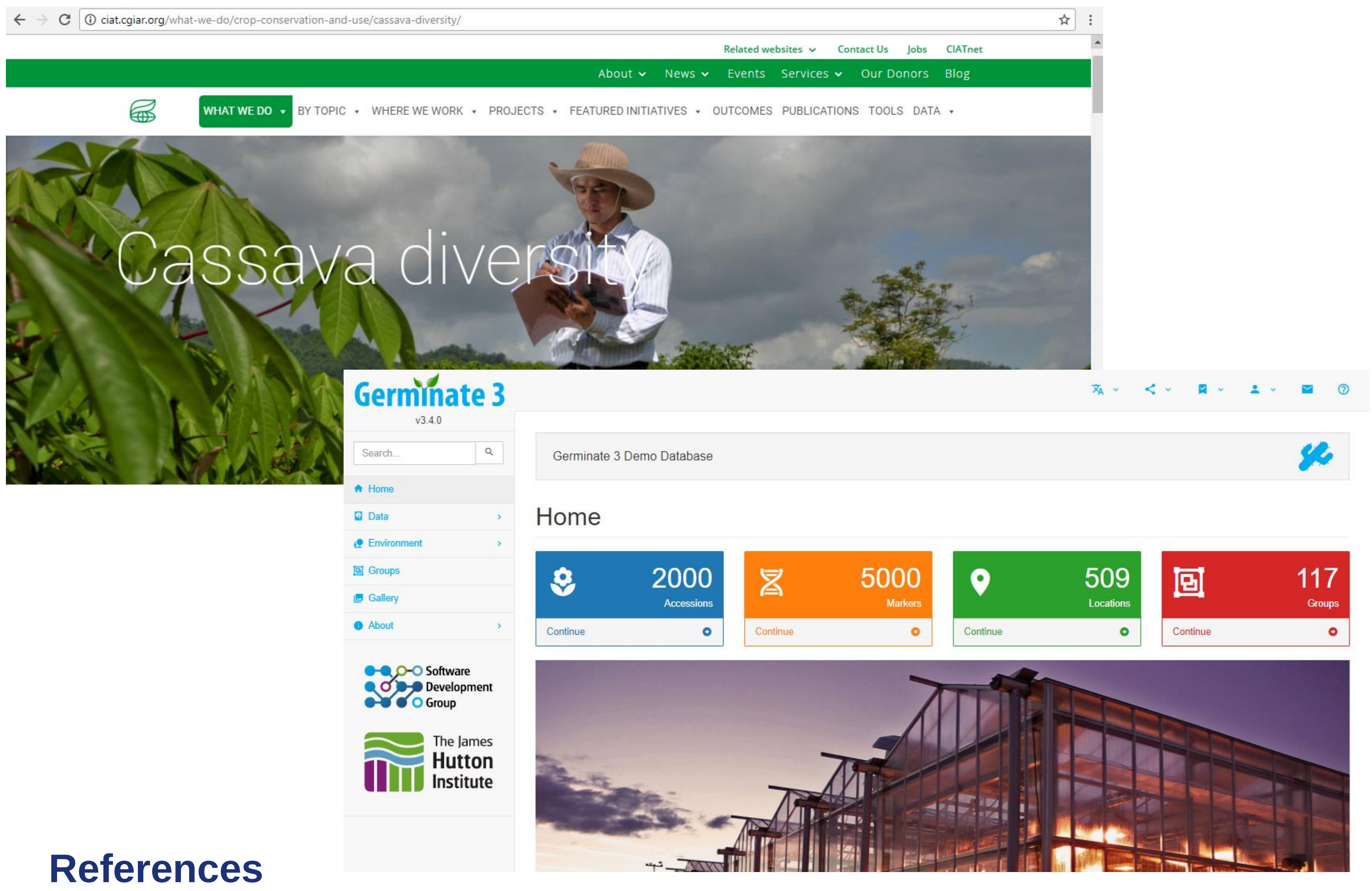


Figure 4: PCO plot showing which samples have been selected for whole genome sequencing in blue, all other samples shown in grey.

Germinate 3 database

All data will be available through a Germinate 3 [4] instance currently under development for CIAT's genetic resources web portal.


ciat.cgiar.org/what-we-do/crop-conservation-and-use/cassava-diversity/

Cassava diversity

Germinate 3 v3.4.0

Germinate 3 Demo Database

Home

2000 Accessions Continue

5000 Markers Continue

509 Locations Continue

117 Groups Continue

References

[1] https://ics.hutton.ac.uk/curlywhirly/
[2] Robinson *et al*. (2011) Nature Biotech. 29:24-26
[3] https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Mesculenta
[4] Shaw *et al.* (2017) Crop Sci. 57:1259-1273