GENETIC DATA IN CONSERVATION PLANNING

Y Jaufeerally-Fakim
University of Mauritius
10 – 13 November 2014

ACP S&T PROGRAMME

ZARI
Zambia Agriculture Research Institute
Researching Soils, Crops and Water in Zambia

Bioversity International

University of Mauritius
www.uom.ac.mu | Reduit, Mauritius

agriculture, forestry & fisheries
Department:
Agriculture, Forestry and Fisheries
REPUBLIC OF SOUTH AFRICA

UNIVERSITY OF BIRMINGHAM

- Types of genetic markers
- Use for diversity assessment
- Studies on *Coffea* in Mauritius
- New technologies for genotyping

Conservation Planning

- Genetic diversity forms an additional level of organisation of individuals/species
- Conservation strategies consider information on ‘evolutionary units ’
- Genetic variation in populations ensures higher evolutionary potential
- Geographical genetic approach for multivariate analysis

Dynamics of molecular evolution

- **Population genetics:** frequencies of alleles change over time under the pressure of various forces
- **Gene pool:** set of all alleles found in a given population
- **Deterministic model:** knowledge of initial conditions can predict changes in allele frequencies
- **Stochastic model:** changes determined through a probabilistic manner

Design

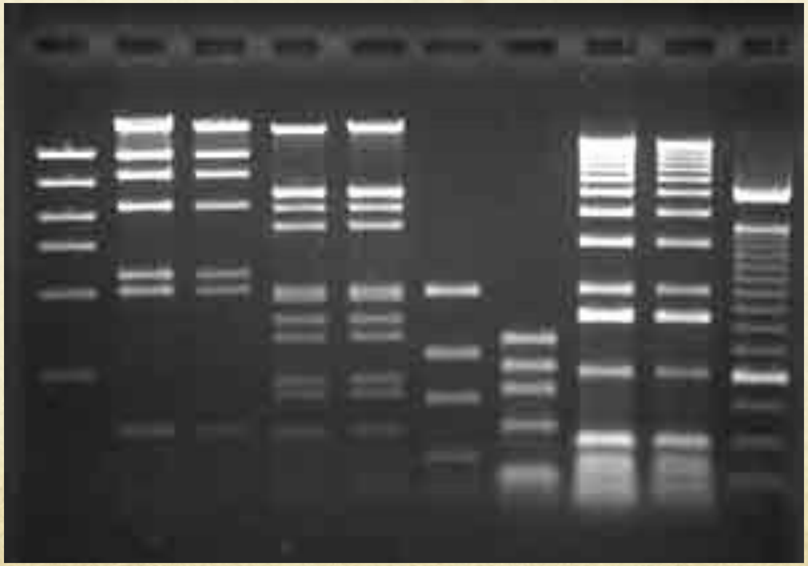
- How many sites to sample (what is the minimum)
- To what extent they represent the overall diversity
- Geographic locations of sites with highest allelic richness

Genetic Data

- Information for diversity, based on DNA sequence differences
- This is highly informative and each nucleotide is considered as one character state
- Can be genic or non-genic sequences
- Can infer evolutionary history

Which loci to use and how to get the data

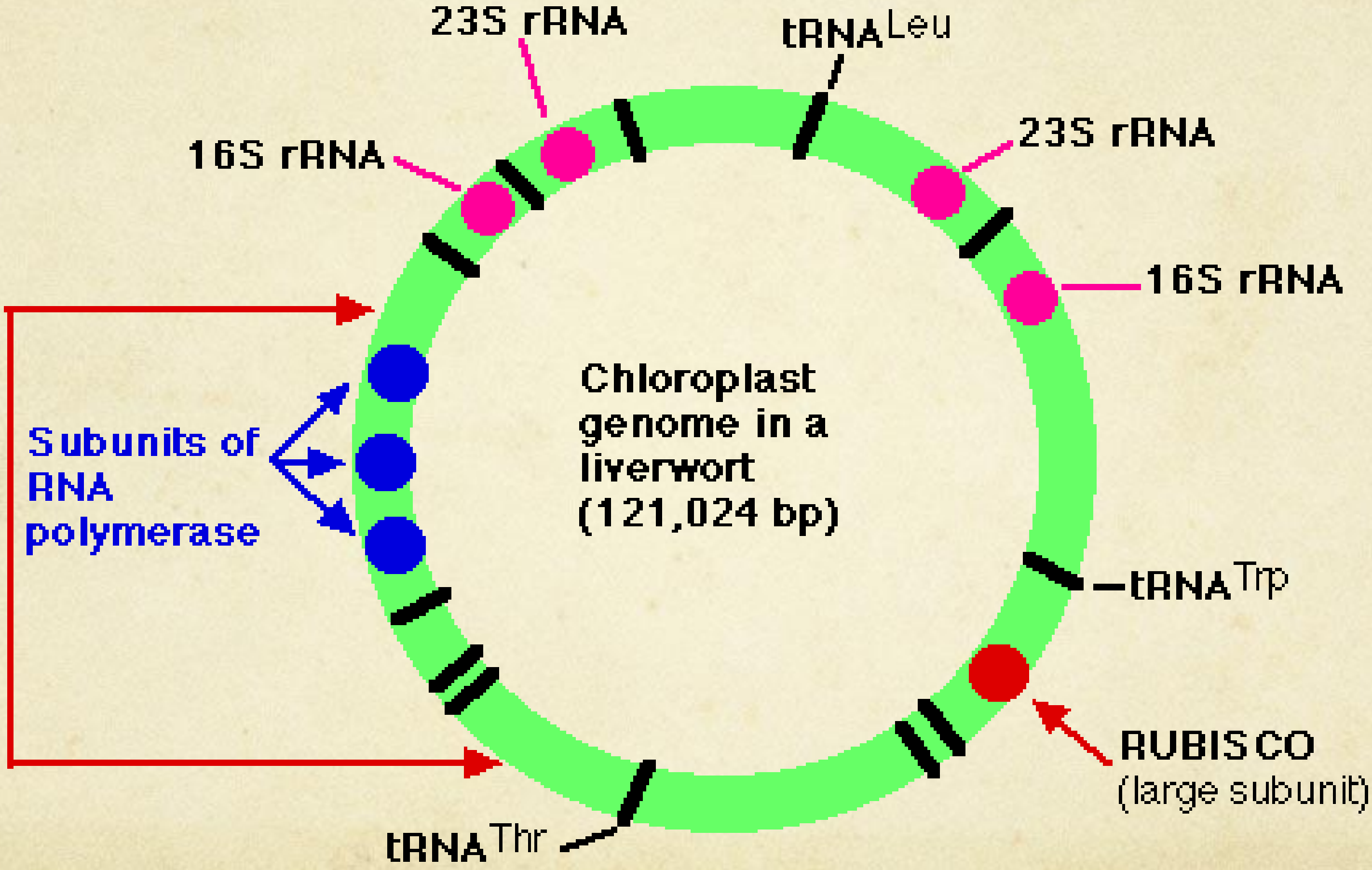
- Wide array of markers available:
    - Length polymorphism eg RFLP
    - PCR Amplification –AFLP, SSR- microsatellites, CAPS



Target Loci - Sequencing

- Chloroplastic DNA:
  - trnL–F region (the trnL intron and the trnL – trnF IGS), (transfer RNA gene)
  - rpl16 intron (ribosomal protein)
  - accD – psa1 (plastid DNA) IGS (Acetyl CoA carboxylase subunit D and Photosystem I)
  - rbcL (ribulose bisphosphate carboxylase)
- ITS (nuclear encoded internal transcribed spacer)

Chloroplast genome



Nucleotide Sequence polymorphism

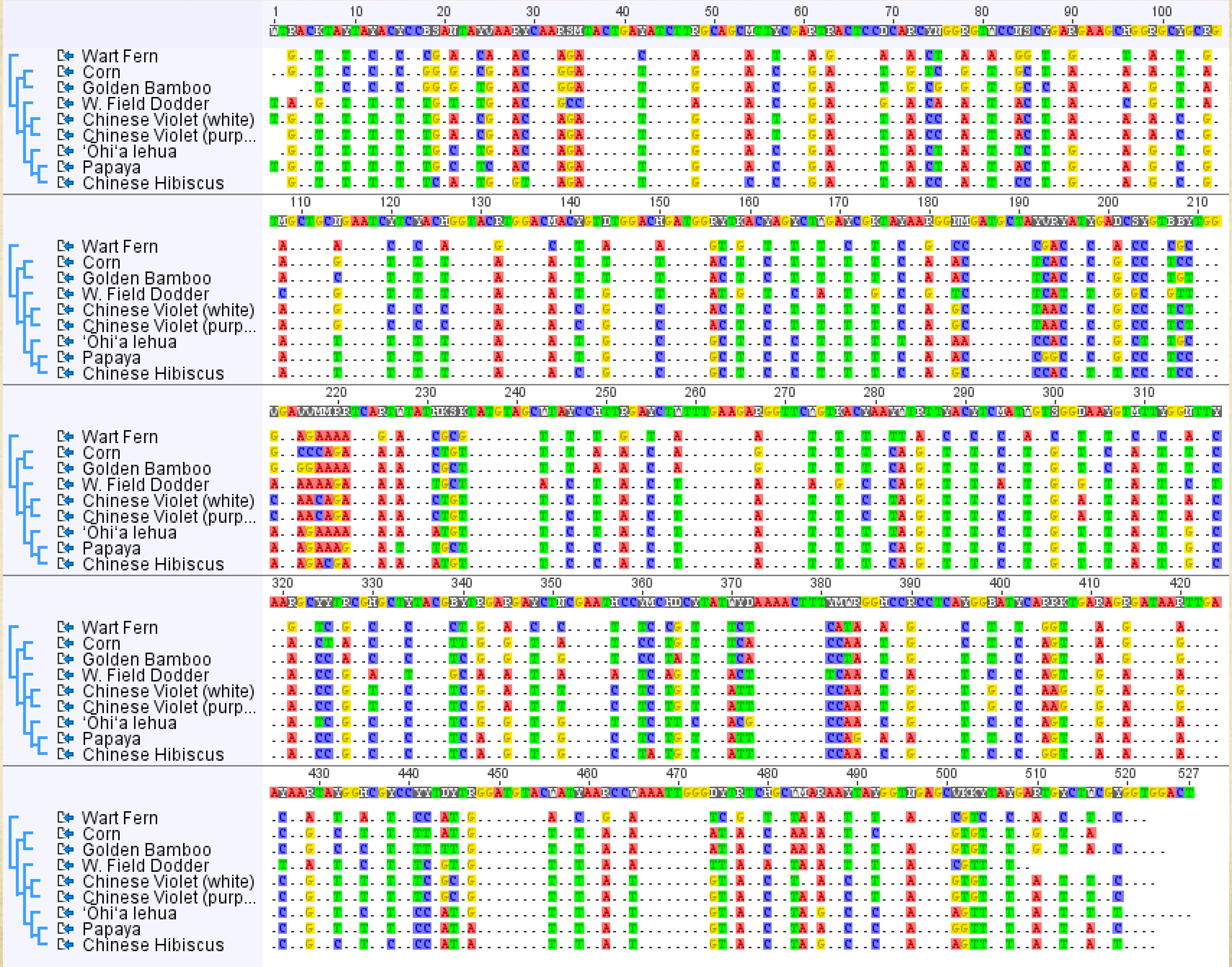
A targetted locus

...ATGGCTGGATAC...
...TACCGACCTATG...

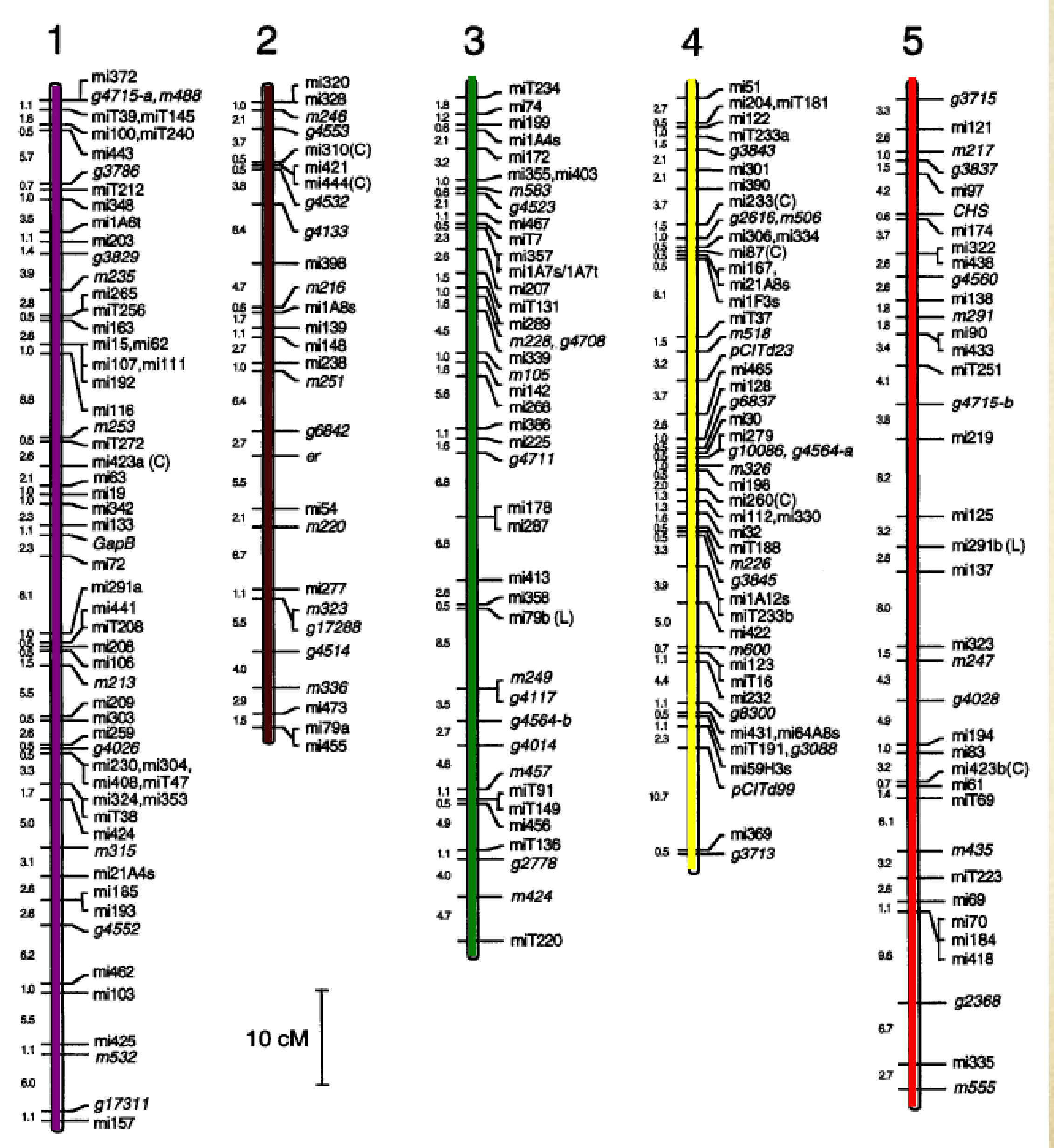
a different allele

...ATGGCCGGATAC...
...TACCGGCCTATG...

SNPs:
Single nucleotide polymorphisms


Wart Fern
Corn
Golden Bamboo
W. Field Dodder
Chinese Violet (white)
Chinese Violet (purp...
'Ōhi'a lehua
Papaya
Chinese Hibiscus
1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
470
480
490
500
510
520
527

Genetic Map



Each chromosome is one linkage group.

Higher number of markers gives better resolution

Incorporating genetic information for conservation planning of California oak

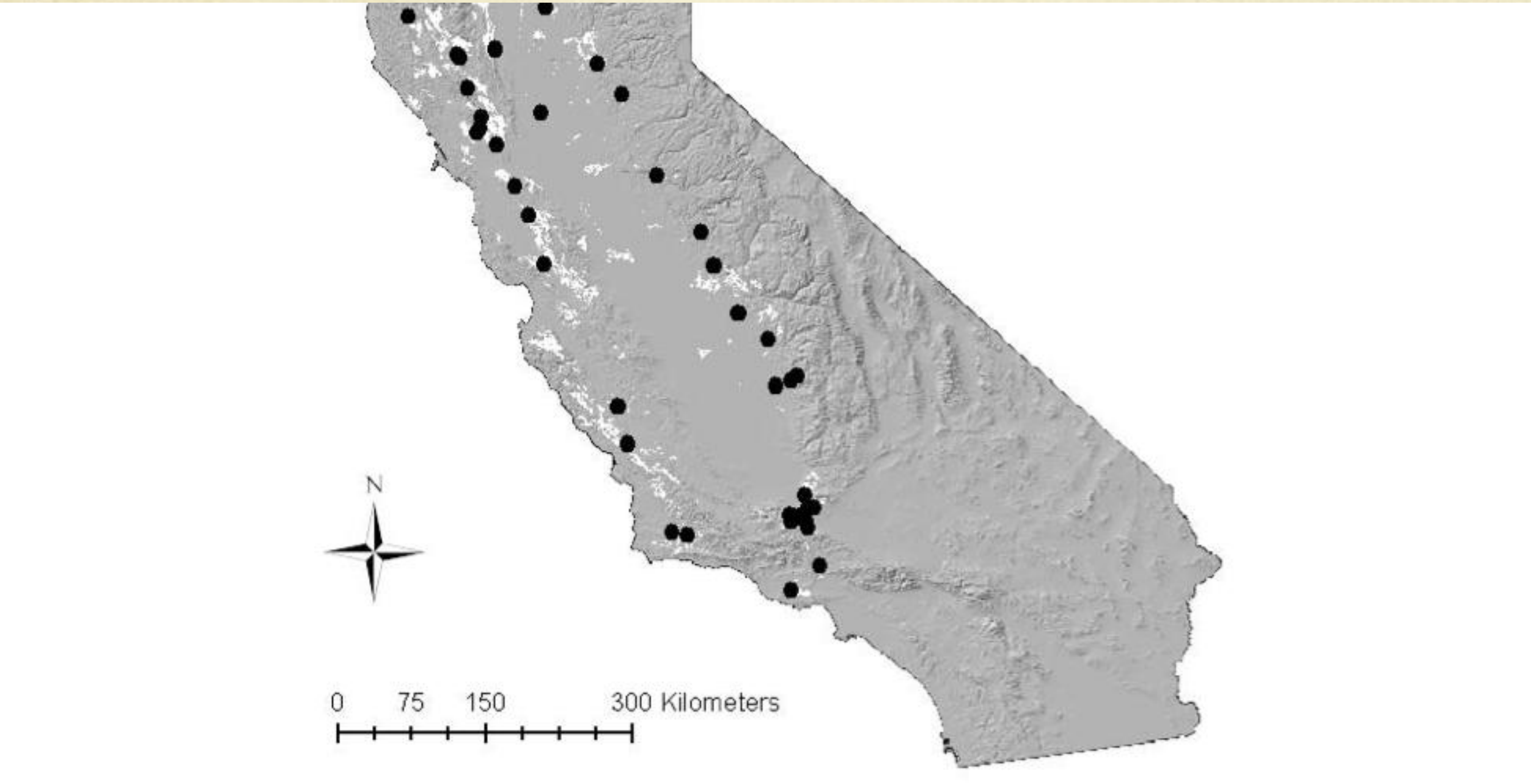


Figure 1—Location of 37 sampling sites for California valley oak (*Quercus lobata*).

Sork et al 2006

- 37 populations sampled = 113 individuals;

- Chloroplastic data: 6 chloroplastic microsatellites = 22 alleles
- Nuclear data: 7 microsatellites = 78 alleles

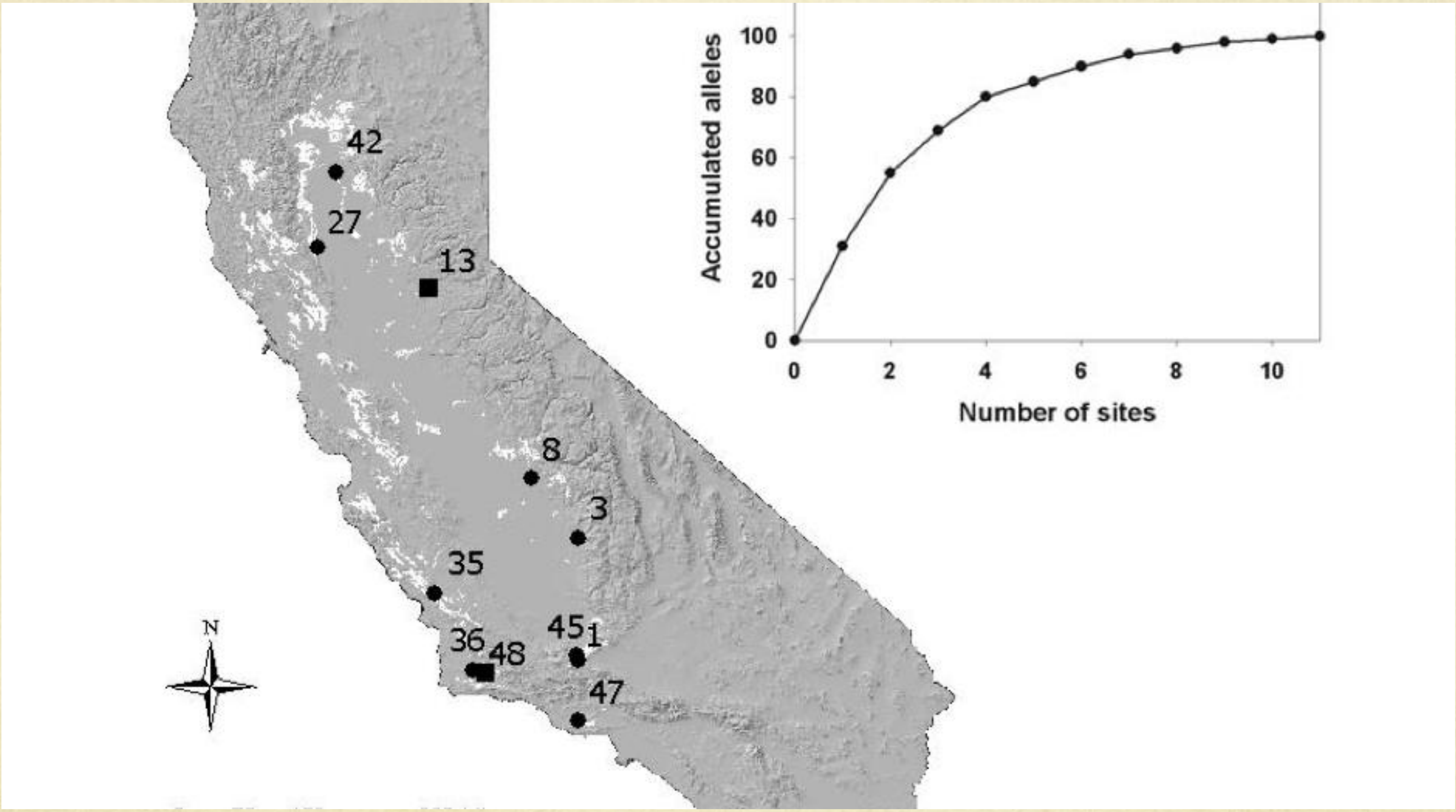
Cp DNA

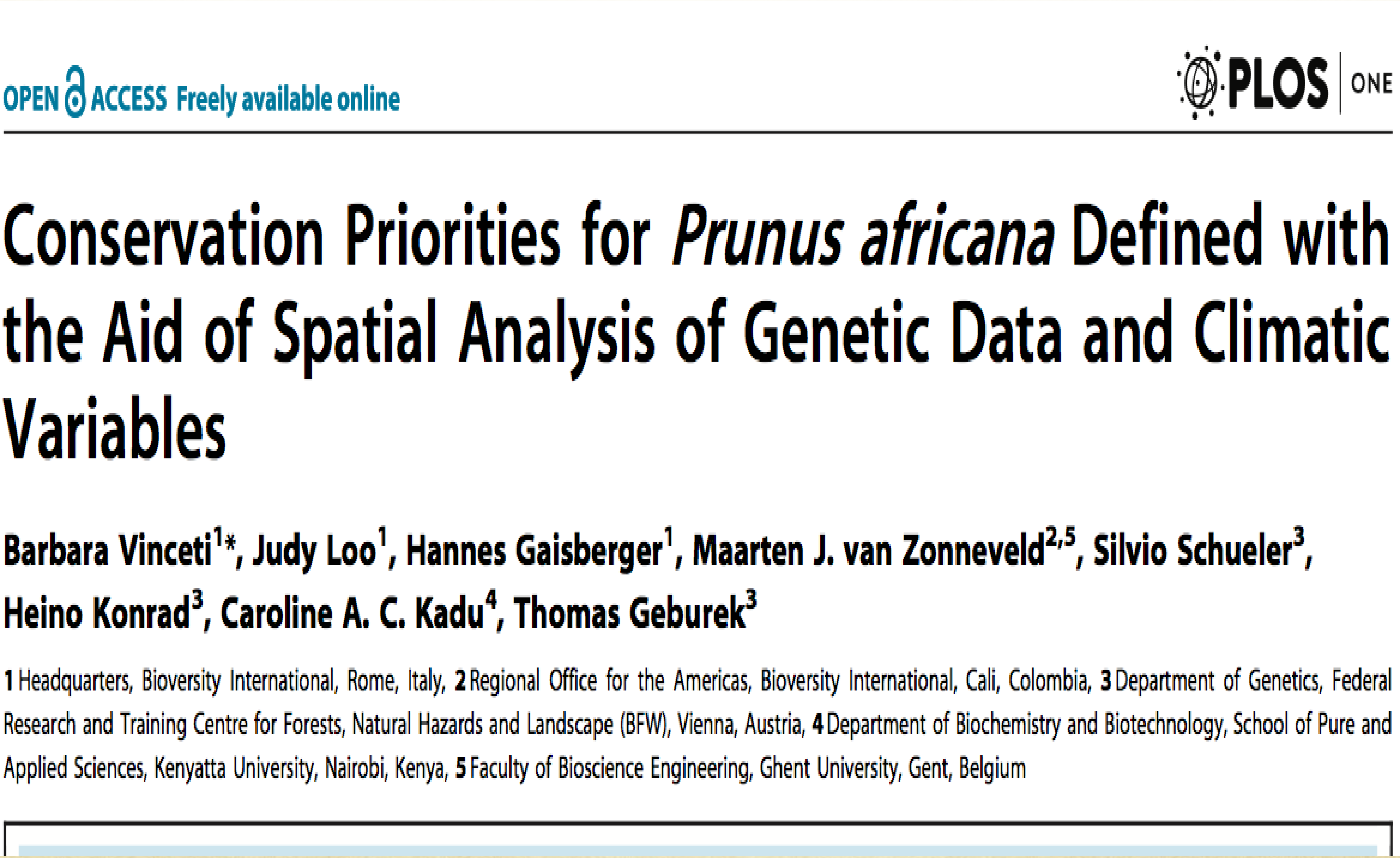
- 32 sites with cpDNA data, (site 31 has the highest allelic richness and represents 12 of 22 alleles.)
- 20 of 22 alleles represented in three of the 32 sites.
- Six sites represent all 22 alleles

nDNA

- 10 sites are required to obtain complete coverage of 78 alleles
- For the reserve design based on the combined markers: 32 populations and 100 total alleles:
- Final design: 11 sites are required and two sites capture more than 50 percent of the alleles

Pooling chloroplast and nuclear microsat alleles





OPEN ACCESS Freely available online

PLOS ONE

Conservation Priorities for *Prunus africana* Defined with the Aid of Spatial Analysis of Genetic Data and Climatic Variables

Barbara Vinceti[1]*, Judy Loo[1], Hannes Gaisberger[1], Maarten J. van Zonneveld[2,5], Silvio Schueler[3], Heino Konrad[3], Caroline A. C. Kadu[4], Thomas Geburek[3]

1 Headquarters, Bioversity International, Rome, Italy, **2** Regional Office for the Americas, Bioversity International, Cali, Colombia, **3** Department of Genetics, Federal Research and Training Centre for Forests, Natural Hazards and Landscape (BFW), Vienna, Austria, **4** Department of Biochemistry and Biotechnology, School of Pure and Applied Sciences, Kenyatta University, Nairobi, Kenya, **5** Faculty of Bioscience Engineering, Ghent University, Gent, Belgium

Combined genetic and climate data: cpSSR and nSSR

- 7 Chloroplast loci -582 individuals
- 6 nuclear loci – 484 individuals
- 32 populations from 22 countries–georeferenced
- Hierarchical clustering of :
  - 6 nSSR – 147 alleles
  - 7 cpSSR- 19 alleles- 22 haplotypes
- 6 Priority populations in 5 countries identified on basis of haplotype richness and presence of local common alleles

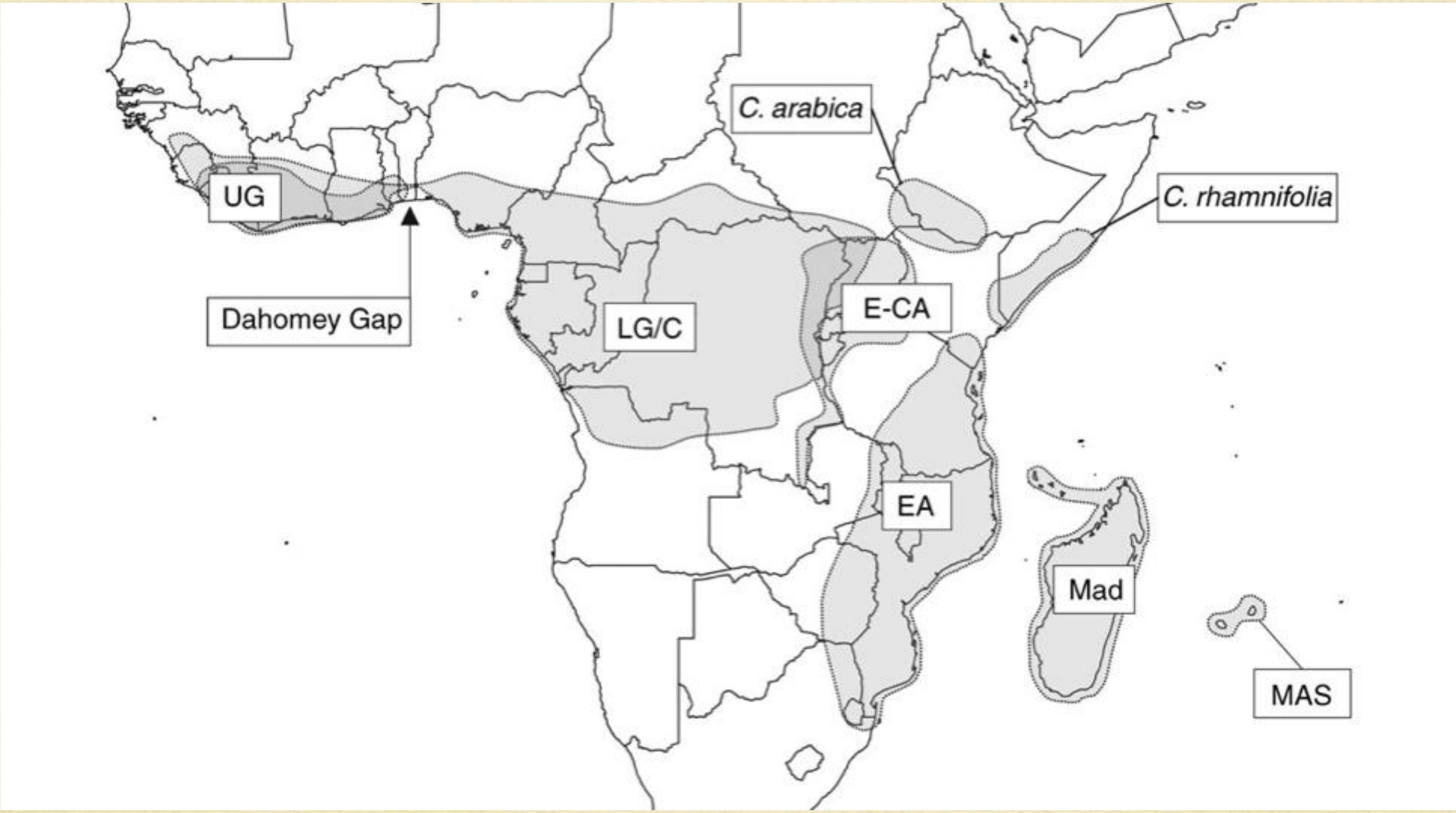
Coffea species from Mascarene

- C *macrocarpa*
- C *mauritianna*
- C *myrtifolia*

Monophyletic

- Maurin et al 2007
- Nowak et al 2011
- Nowak et al 2014

88 Taxa: Chloroplast: trnL, rbcL, accD-psaI And nuclear ITS



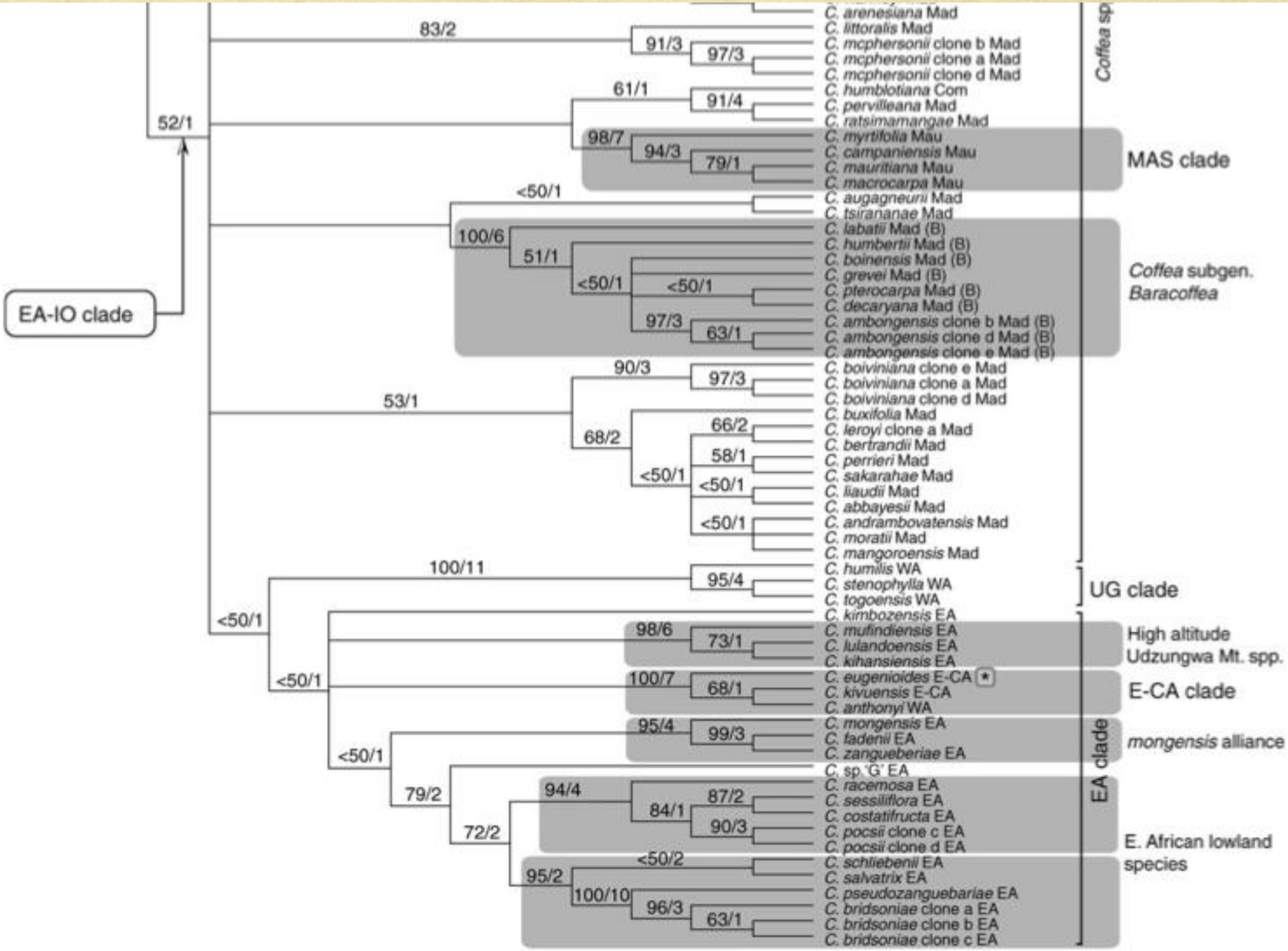


FIG. 4. Strict consensus tree generated from combined molecular (plastid–ITS) analysis. Bootstrap values of >50 % are placed above the branches, followed by Bremer support (decay) values. See Table 1 for species authorities and provenance. EA-IO clade = East Africa-Indian Ocean clade; EA clade = East Africa clade; E-CA clade = East-Central Africa clade; LG/C clade = Lower Guinea/Congolian clade; UG clade = Upper Guinea clade; MAS clade = Mascarene clade. Regions: EA = East Africa; E-CA = East Central Africa; Com = Comoros; In = India; Mad = Madagascar; Mau = Mauritius; WA = West Africa. *Denotes progenitor species for *C. arabica*. (B) = species belonging to *Coffea* subgenus *Baracoffea*.

Self-Incompatibility RNase

- Mechanism of SI: Interaction between the growing pollen tube F- protein and RNase from the style and stigma.
- 13 individuals form 3 populations- SI RNase sequenced
- Allelic diversity of SI Rnases for Mascarene coffea species : seven distinct lineages.

High-resolution methodologies

- Classical markers are limited to a few per run ; need a large number to saturate a map. Techniques are cumbersome and tedious.

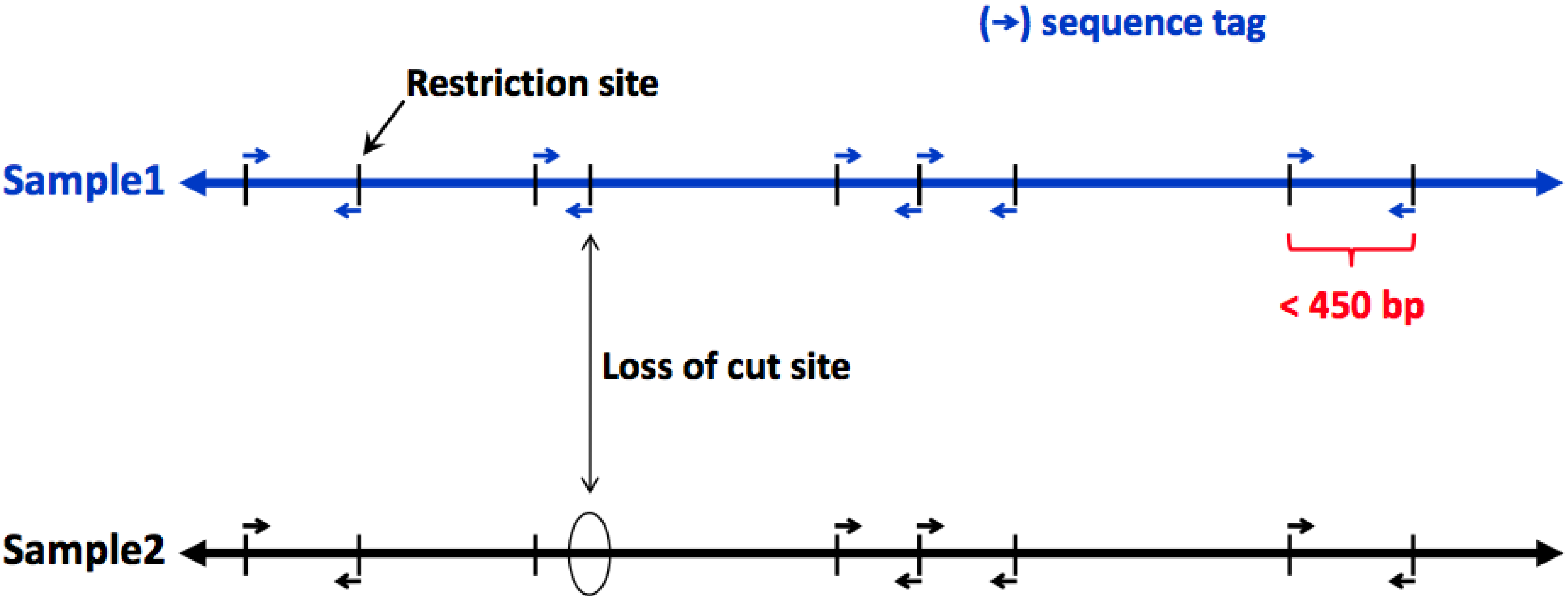
- Next generation sequencing (NGS) methodologies have brought new approaches for diversity studies

- Most importantly drastic reduction in costs

Genotyping by Sequencing (GBS)

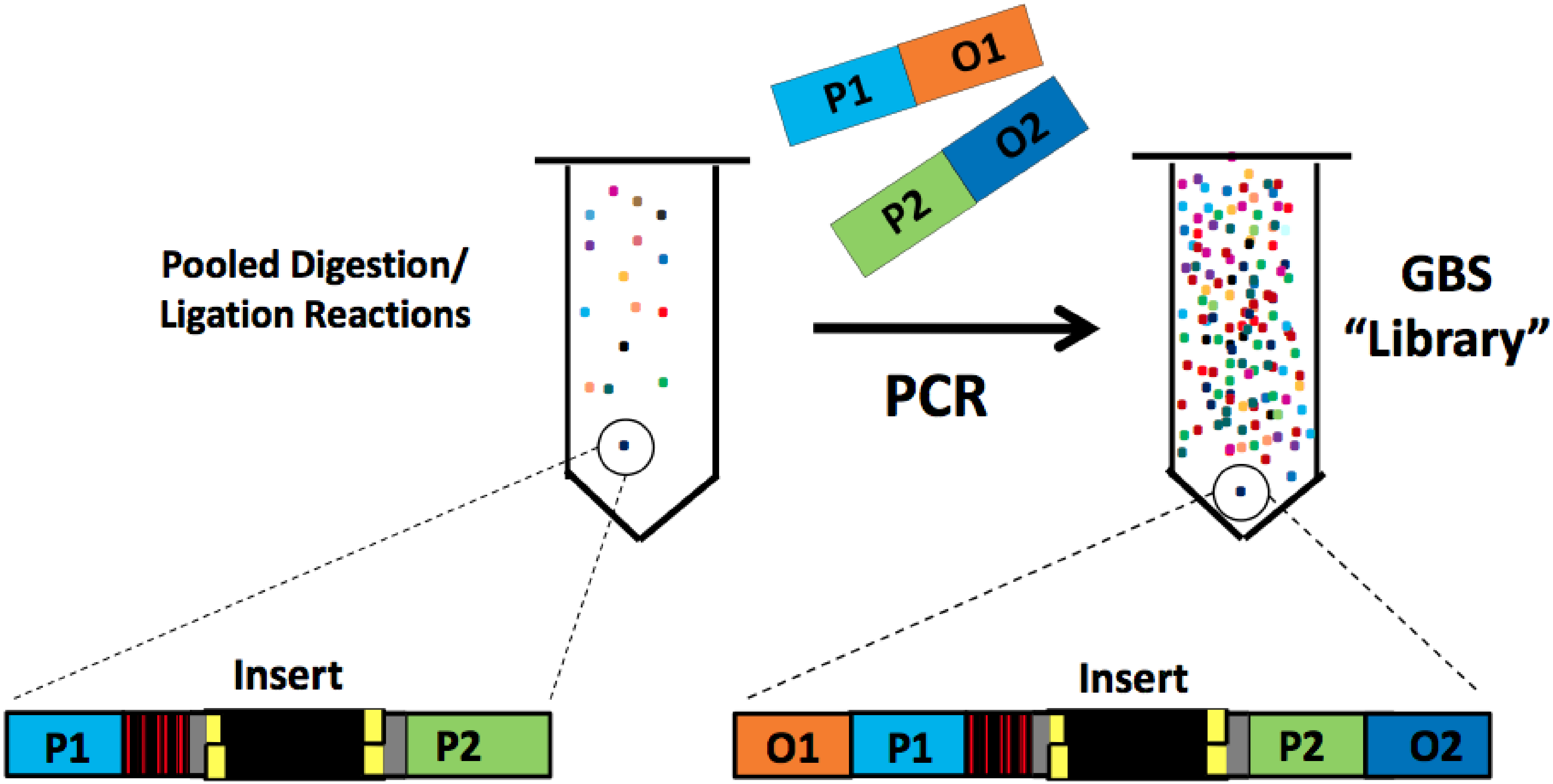
- Uses NGS and a bioinformatics pipeline
- Genome complexity is reduced by using restriction enzyme-cut DNA fragments
- Multiplexing is done through the use of bar-coded adapters

Overview of Genotyping by Sequencing (GBS)



- Focuses NextGen sequencing power to ends of restriction fragments
- Scores both SNPs and presence/absence markers

PRC primers:



Pooled Digestion/
Ligation Reactions

PCR

GBS
"Library"

Insert

Insert

- GBS markers have been shown to be present at regular spacing within the genomes of wheat and barley (Poland et al 2012)
- High rate of marker discovery
- Suitable for biodiversity assessment

GENOME SCIENCE

- First plant genome *Arabidopsis thaliana* in 2000 -125 Mb
- 1000 plant genomes
- 1000 human genome
- Genome 10K project- 10,000 vertebrate species
- Fish T 1K project- BGI- 1000 fish transcriptomes
- i5K project- Baylor college of Medicine: 5000 insect and other arthropod genomes

THANK YOU