

Future-proof genetic characterization strategies for genebank collections

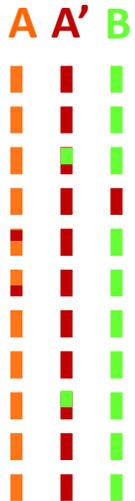
→ Clonal crops with complex genomes

Angélique D'Hont
CIRAD, Montpellier
France

For activities and members of our team see : <https://umr-agap.cirad.fr/en/recherche/equipes-scientifiques/structure-et-evolution-des-genomes/contexte-et-enjeux>

→ Clonal crops with complex genomes

Banana



$2n = 2x, 3x, 4x$

$x = 500 \text{ à } 600 \text{ Mb}$



Sugarcane

S. off. *S. spont.*

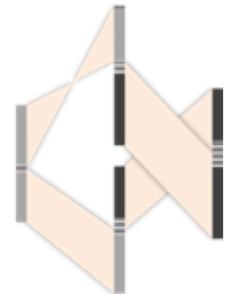


$2n = \text{c.a. } 12x = \text{c.a. } 115$

$x = 950 \text{ Mb}$



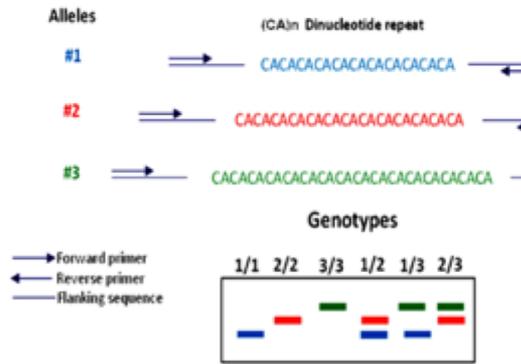
- Polyploidy
- Aneuploidy
- Inter-specificity
- Translocations/ inversions



→ Yam, Sweet potato, ...

Marker systems:

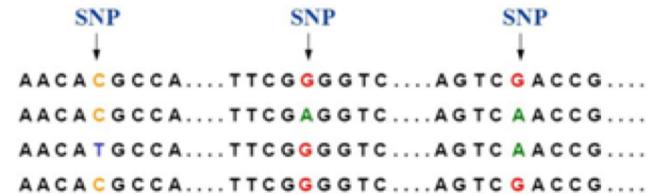
Single Sequence Repeat/microsatellite



For **control of identity/paternity**

Limited for diversity study

Single Nucleotide Polymorphism



Next Generation Sequencing → generate several hundred thousand to millions of SNPs to can be mapped on a **genome reference sequence assembly**

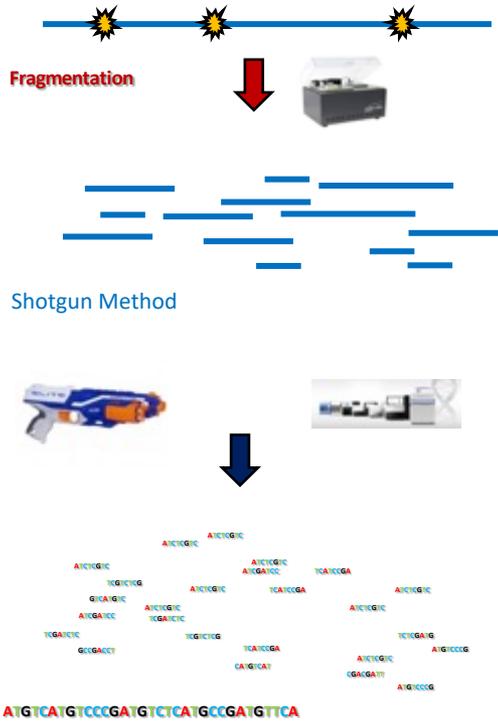
→ Allow high resolution in the analyses

For diversity analysis

For phylogenomic analysis
to go beyond phylogenetic trees

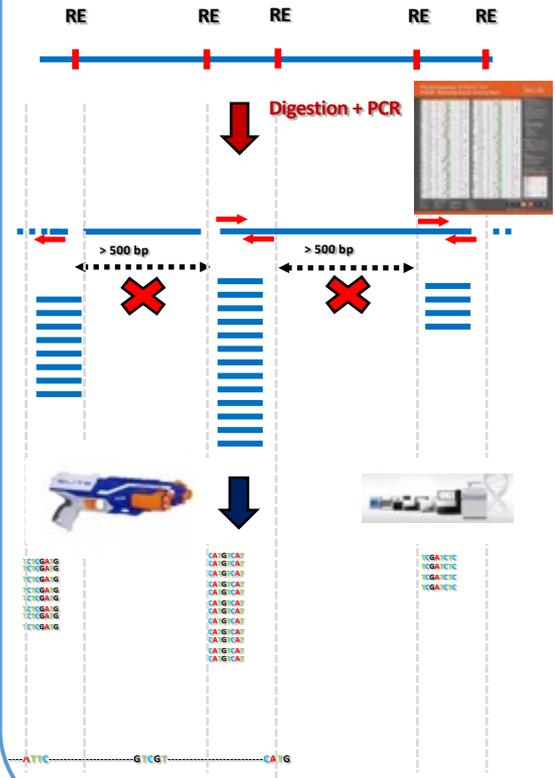
SNP genotyping through sequencing

WGS



Whole genome sequencing

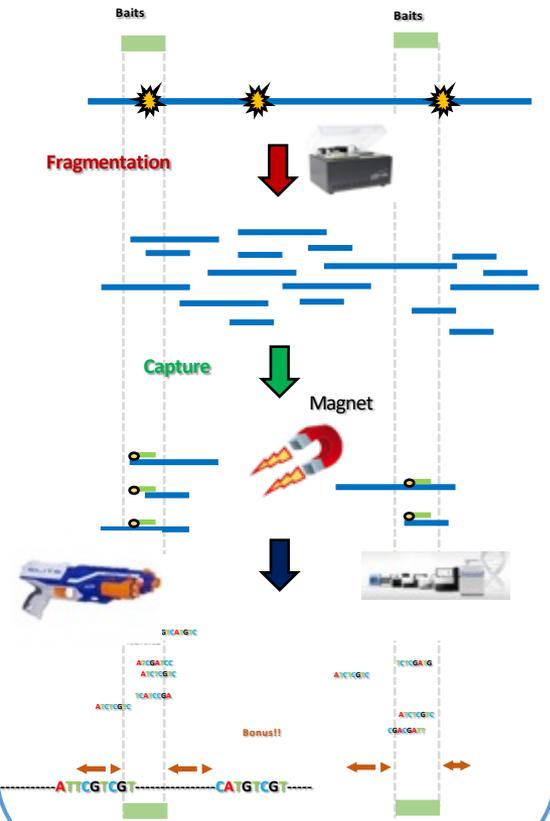
GBS (with various genome reduction methodologies)



Partial genome sequencing

Cheaper, lower bioinformatic requirements

Targeted-GBS



Interspecific genome structure of banana cultivars

From hundred thousand to millions of SNPs

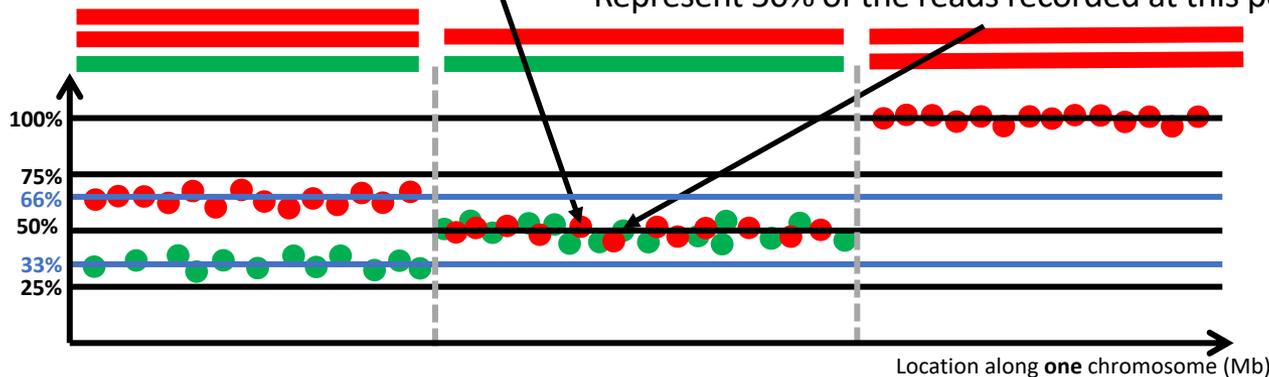
Identification of diagnostic alleles with presence/absence in ancestral representatives

Calculation of the proportion of reads supporting alleles from an origin in studied accession

SNP assigned to **B** genome
Represent 50% of the reads recorded at this position

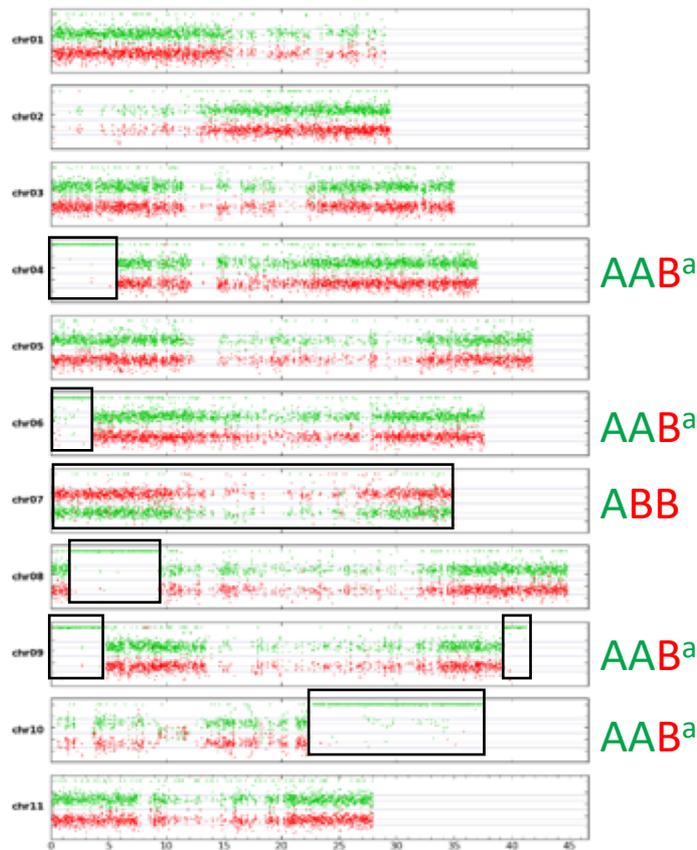
Proportion of reads corresponding to the SNP

SNP assigned to **A** genome
Represent 50% of the reads recorded at this position



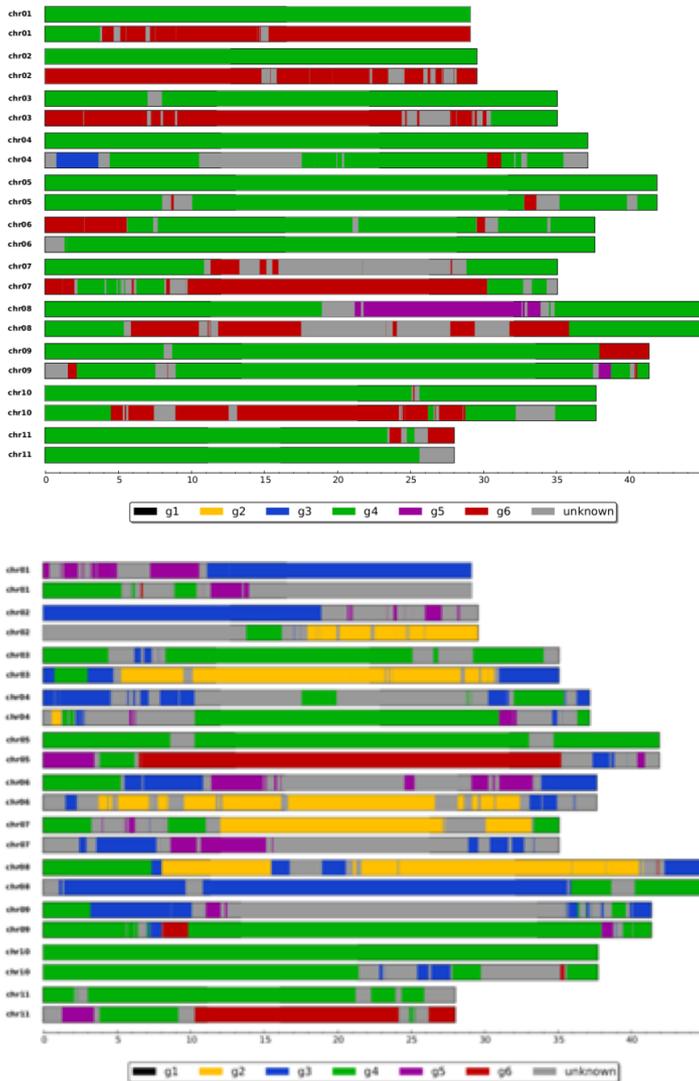
Interspecific genome structure of banana cultivars

Plantain 'AAB' (French Clair)



➤ Mosaic genome
= a few generations/meioses
since the interspecific hybridization

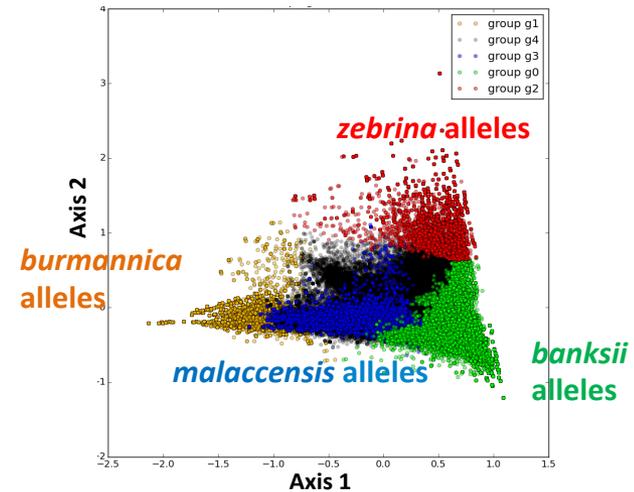
Inter-sub-specific genome structure of banana cultivars



Identification of diagnostic alleles through multivariate analysis on ancestral representatives and clustering



Calculating haplotype number “probabilities” given expected and observed values



- Complex mosaic genomes
= several generations/meioses
= several ancestral contributors

Applications of genetic/phylogenomic characterization

For Genebank management :

- Detection of gaps in collections
- Prioritization for conservation (in collection or natural area to be protected)
- Detection of in-vitro induced aneuploids (off-type)

But cannot differentiate clonally diversified cultivars such as the various types of Plantain

For Direct use (e.g.):

- Selection of genetically diverse materials for farmers testing
- (in the future) Selection of materials holding markers associated to desirable traits

For research (e.g.) :

- Evolutionary studies, phylogeography
- Help understand origin of cultivars and elucidate the domestication processes
- Detect large structural variations (translocations/inversions) and analyze their impact on chromosome segregation

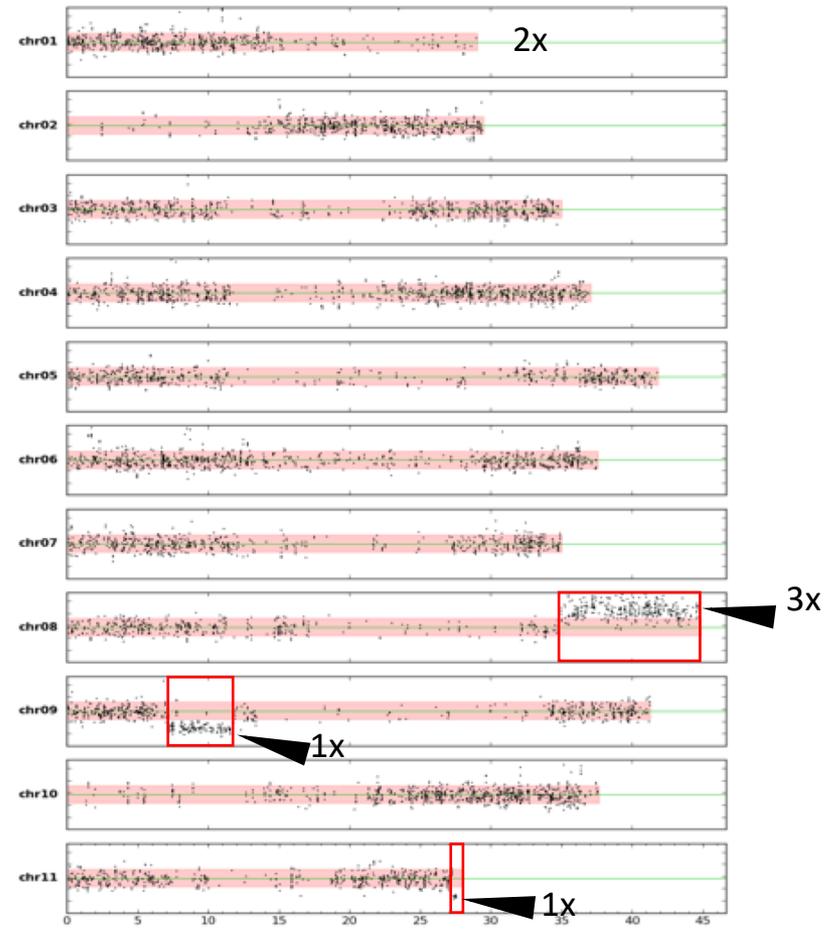
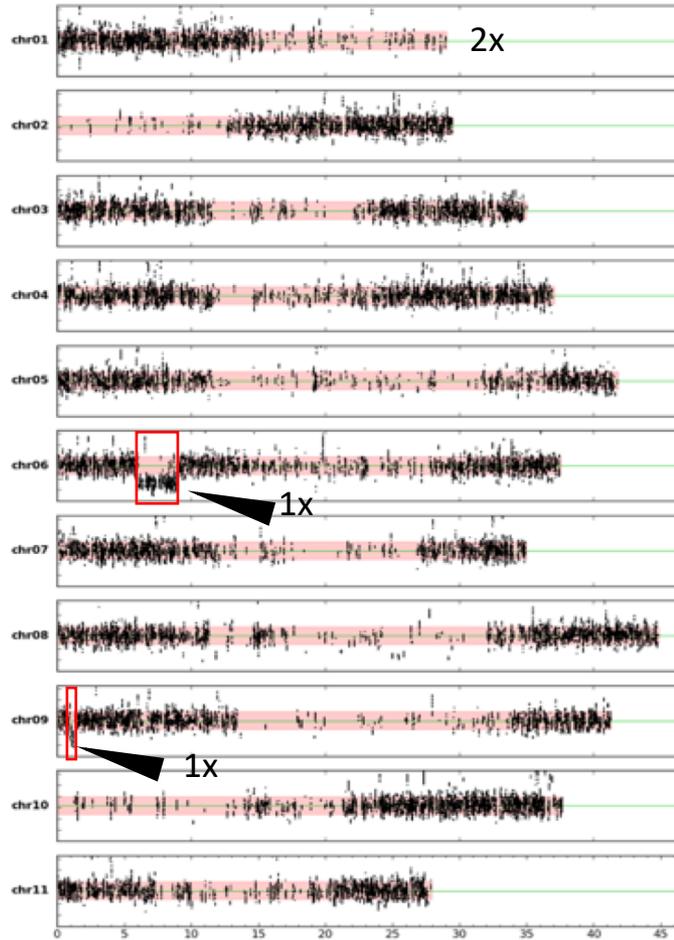
For Breeding (e.g.) :

- Identify the origin of particular traits and gene pools where to find allelic variation for these traits
- Integrate chromosome architecture (translocations/inversions) in choice of parents for QTLs analysis
- Guide germplasm exploitation in breeding programs

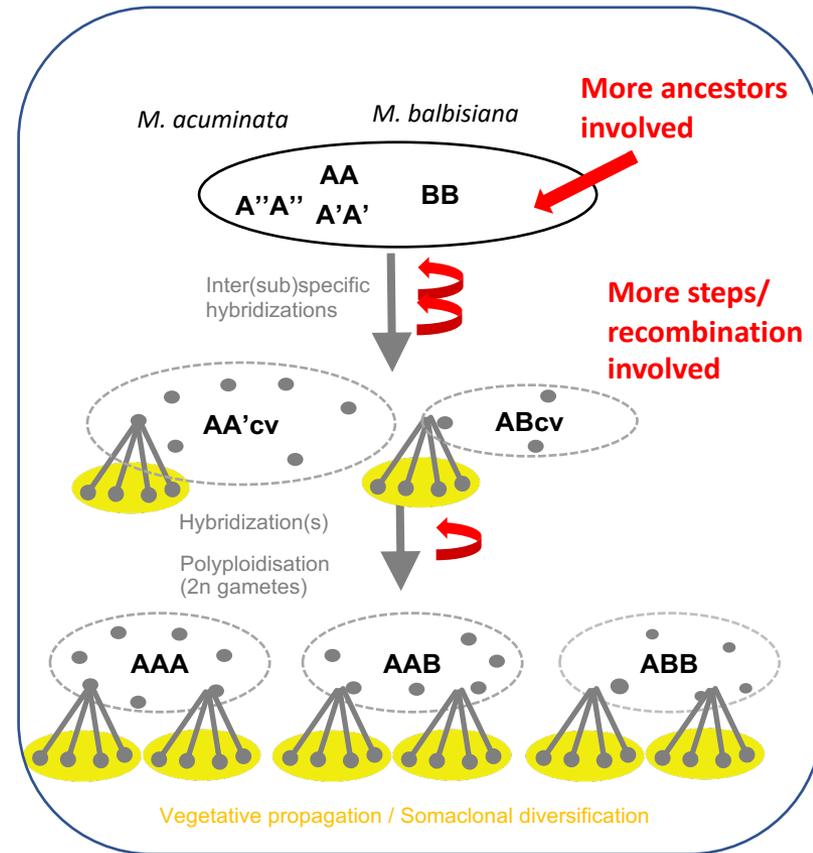


Detect aneuploidy

Base on read coverage



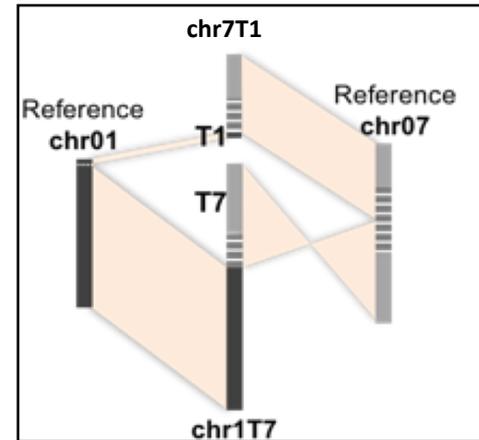
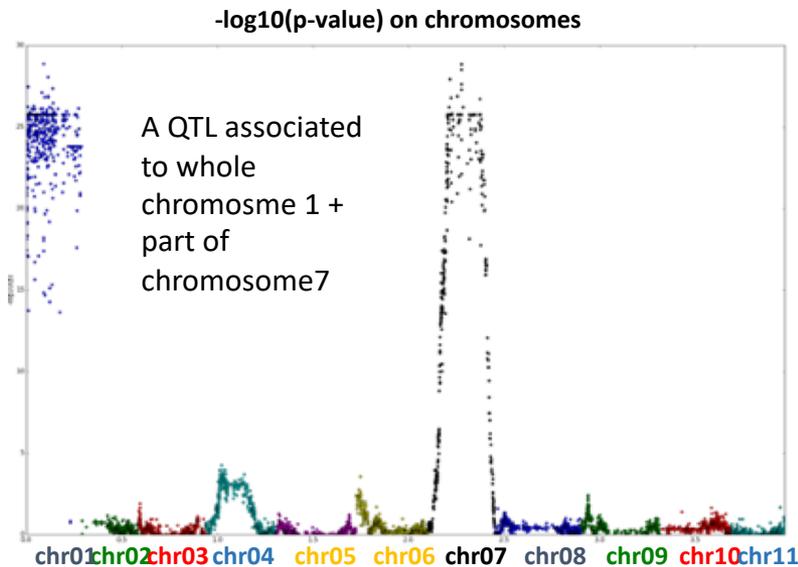
Help understand cultivar origins



- Complex mosaic = several generations/meiosis and ancestral groups involved
- Unknown ancestral groups involved, will help discover pure representatives of these groups
- Shared recombination breakpoints/introgression → common ancestry
 - **Help understand origin of cultivars/ domestication process**

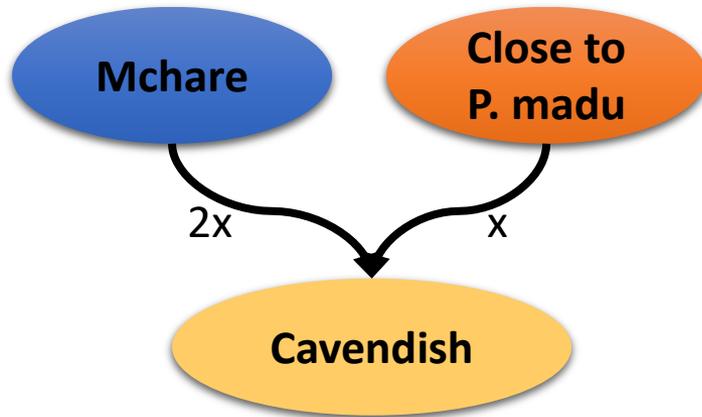
Guide exploitation of germplasm for QTL analysis

- **Translocation/inversion at heterozygous state can prevent recombination**
- Unprecise position of QTL
- Phylogenomic analysis → chose accession structurally homozygote that can recombine

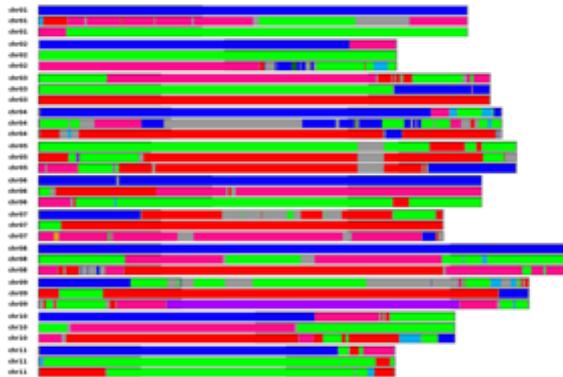


- **Identify the origin of particular traits/QTL and the gene pool where to find allelic variation for these traits**

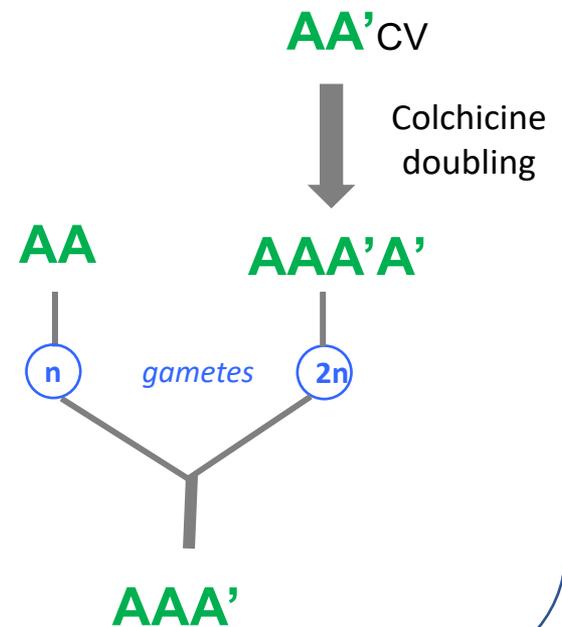
Guide breeding strategies



Grande Naine



Reconstructive breeding



- Identify parents or close relatives of successful cultivars
- to guide choice of parent in breeding programs

Future direction

- **Develop/chose and promote a system to be used by the community working on a crop**
- ➔ **So the results from various labs/collections could be connected**

Ex: for banana

Various systems used so far : WGS, GBS, DArTseq, RADSeq

Test a targeted genotyping by sequencing system, with targets allowing to connect these data

Need to define the optimum number of SNP/targets and coverage to obtain a phylogenomic profile/mosaic with enough resolution

- **High polyploidy → very large genome**

Ex: for sugarcane (genome size 10 Gb)

Test a targeted genotyping by sequencing system, to try find a **good compromise between the number of SNPs obtained & cost & size of bio-informatic file to be manipulated/stored**

Requirements

- High quality sequence data with enough coverage (by haplotype) and enough SNPs along chromosomes
 - WGS : 6 500 000 SNPs, RADseq: 450 000, “GBS”: 250 000 (approximate in banana)
 - WGS : 400€, GBS: 25€, Targeted sequencing: 45€ (approximate prices in banana vary depending on platforms)
 - Bioinformaticists
 - Bioinformatics pipelines for SNP analysis and result visualization
- adapted to the genome characteristic of each crop

