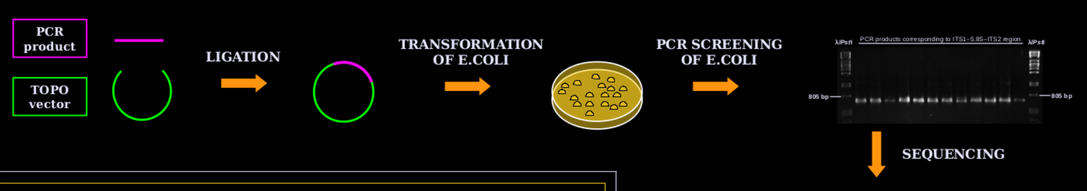
BANANAS: GENOME STRUCTURE, DIVERSITY AND PHYLOGENY

Bananas are staple food and important export commodity in many counties of humid tropics. Traditional taxonomy of *Musa* based on plant phenotype and chromosome number has been questioned. Most of cultivated banana cultivars are seed sterile diploid and polyploid clones originating from natural inter- and intra-specific crosses of two wild diploid species of genus *Musa*: *M. acuminata* (A genome) and *M. balbisiana* (B genome). However, their chromosome constitution is not known due to lack of suitable cytogenetic markers.

Molecular phylogeny of the Musaceae family

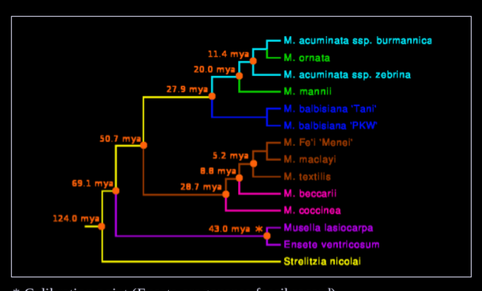
- ITS1-5.8S-ITS2 sequence region analysis as well as single copy genes were selected to reconstruct phylogenetic inference in *Musaceae*

ITS1-5.8S-ITS2 region analysis



Multi gene analysis

- Set of 19 unlinked genes were cloned and sequenced in 13 Musa species
- Phylogenetic inference as well as estimates for divergence times of Musaceae genera were obtained



* Calibration point (Ensete orogonense fossil record)
Dating was done using BEAST v1.4.8.

Diversity of *Musa* – *Musa* Genotyping Centre

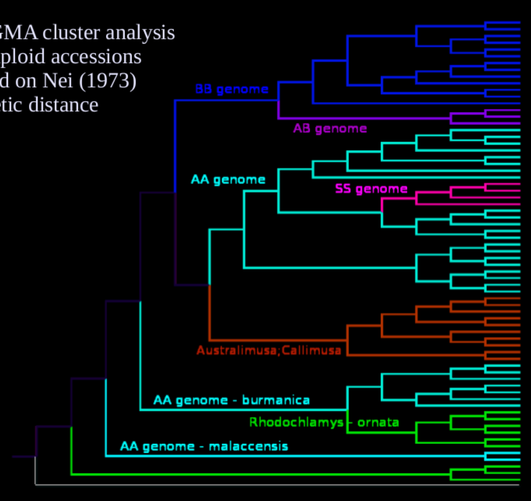
- We have created a standardized SSR genotyping platform (Christelová et al. 2011)
- System is based on 19 microsatellite loci that are scored after the PCR with fluorescently labeled primers
- To facilitate building of the database of electrophoretic profiles, guarantee standard genotyping conditions and reproducibility of results, the genotyping is centralized
- Centralized database of molecular profiles keeps growing with every new sample, resulting in stepwise improvement in the grouping

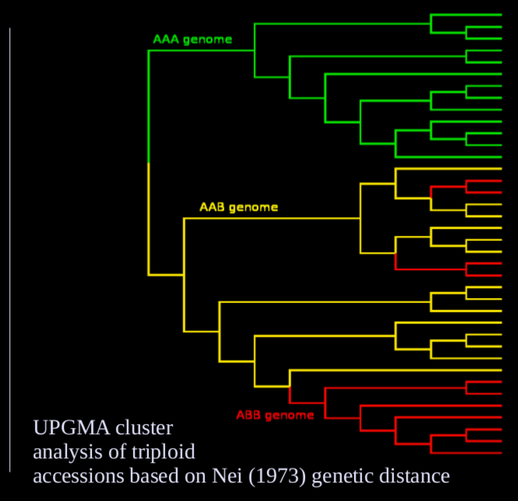
Experimental Design

- Ploidy level estimation using flow cytometry
- Genomic DNA isolation; PCR with fluorescently labeled primers
- Capillary electrophoresis of resulting fragments (ABI 3730xl)
- Call for alleles and cluster analysis
- ITS analysis if needed

Partec PAS flow cytometre — 3730*xl* DNA analyzer — Call for alleles using GeneMarker (Softgenetics) followed by manual check — Cluster analysis

UPGMA cluster analysis of diploid accessions based on Nei (1973) genetic distance





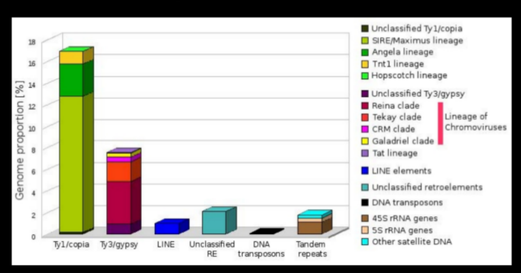
UPGMA cluster analysis of triploid accessions based on Nei (1973) genetic distance

Sequencing banana genome

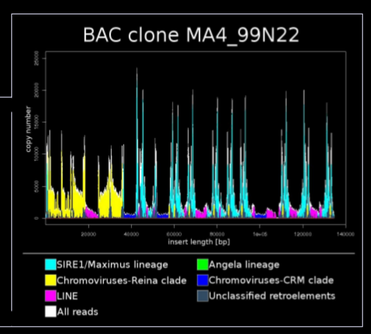
- We are mainly interested in repetitive part of nuclear genome
- With the aims to characterize repeats in the genomes of Musaceae and identify the contribution of repeats to genome diversity we are using NextGen sequencing technologies

454 sequencing of *M. acuminata* 'Calcutta 4'

- 454 sequencing using FLX system resulted in 477,699 reads with average length of 206 nucleotides (almost 15% of 'Calcutta 4' genome)
- Reads were assembled using tgicl software (reads clustering was done using tclust; cap3 was used to assemble the reads within clusters)
- Most of all repeats were assembled and characterized
- Main repeat families were mapped onto meta[hase chromosomes of 'Calcutta 4'
- Database of *Musa* repetitive elements were created and used during *Musa* sequencing project



Verification of repetitive database created during the project. Plot represent genomic copy numbers of idividual inserted regions in the BAC clone (X axis) calculated from numbers of similarity hits to 454 reads databases.

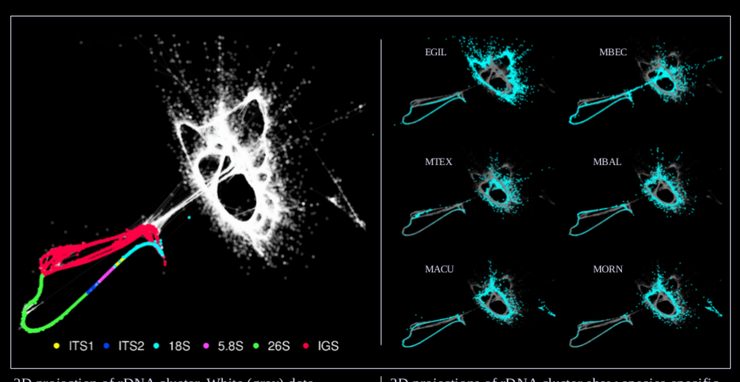


Comparative analysis of 6 species from Musaceae family using 454 sequencing and graph-based clustering

- Six species of the family Musaceae were sequenced using 454 Titanium platform:

  Ensete gilletii ITC 1389 (EGIL)
  Musa beccarii ITC 1070 (MBEC)
  Musa textilis ITC 0539 (MTEX)
  Musa balbisiana 'PKW' (MBAL)
  Musa acuminata ITC 0249 (MACU)
  Musa ornata ITC 0637 (MORN)

- Reads were clustered using graph-based clustering and graph layouts were calculated using Fruchteman and Reingold algorithm (Novák et al. 2010)

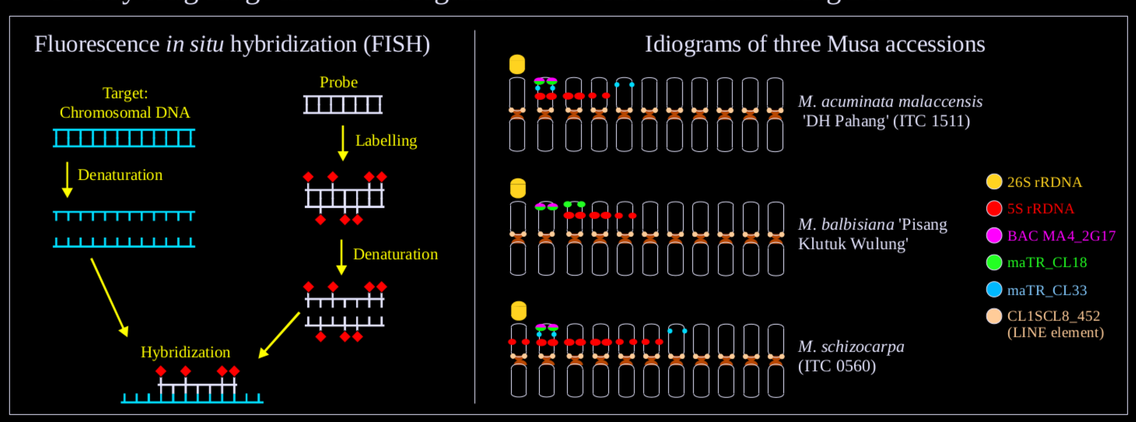


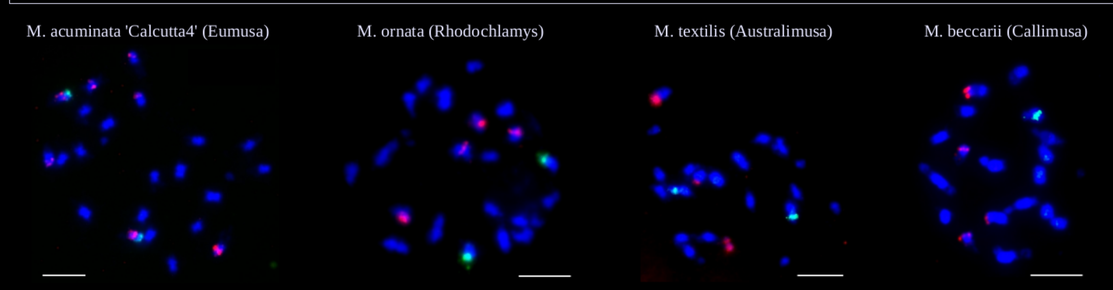
2D projection of rDNA cluster. White (grey) dots correspond to reads with no similarity to rDNA. Individual reads are represented by nodes and those of them with identity at least 80% along 110 nt are connected by lines.

2D projections of rDNA cluster show species-specific differences: reads are colored based on the species. Reads derived from the species are light-blue, remaining reads are grey.

Genome structure and organization

- Relatively small genome size 1C ~ 600 Mb
- Mitotic chromosomes show little morphological differences
- Need for specific cytogenetic markers
- Major part of the genome constitutes for repetitive DNA sequences which can be used to study long range molecular organization of chromosomes using FISH





FISH with the probes for 5S rRNA genes (red) and single copy BAC clone MA4_2G17 (green). Chromosomes were counterstained with DAPI (blue). Bar = 5 um.

The Centre of Plant Structural and Functional Genomics

Institute of Experimental Botany AS CR, v. v. i.