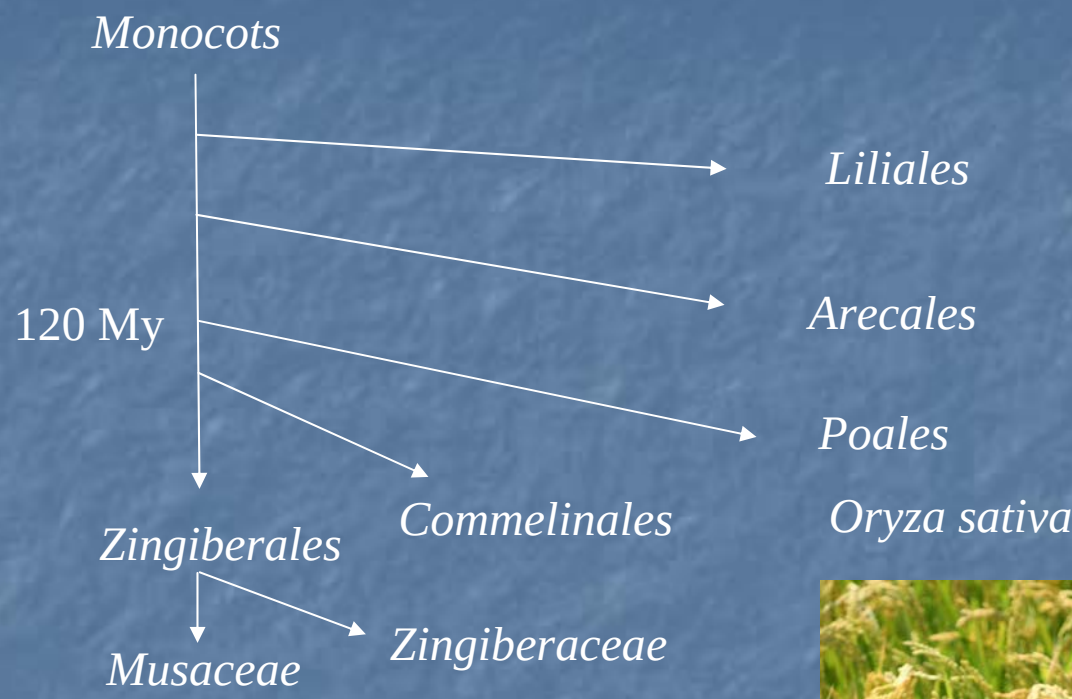
Musa BAC sequencing project

Franc-Christophe BAURENS
CIRAD - UMR DAP
Research team : Genome's Structure and Evolution

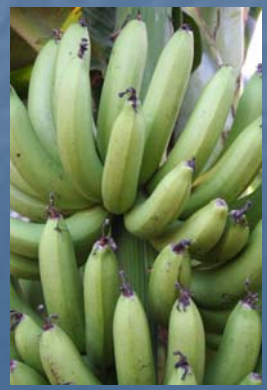
Introduction
Musa-Rice synteny
Musa A-B synteny
Haplotype structure within Musa genus
- RGA cluster in *Musa balbisiana*
- Gene rich region in *Musa acuminata*

Conclusion and prospects

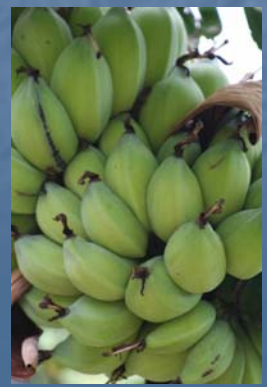
Lineages of rice and banana in Monocot evolution



Musa acuminata



Musa balbisiana

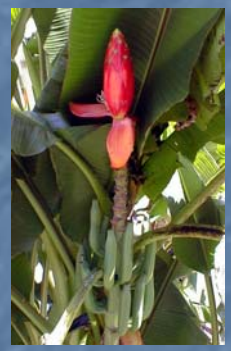


Genus *Musa*

Section *Australimusa* x=10

Musa textilis
Genome TT

Section *Callimusa* x=10

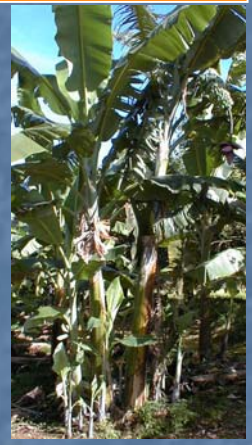


Musa beccarii

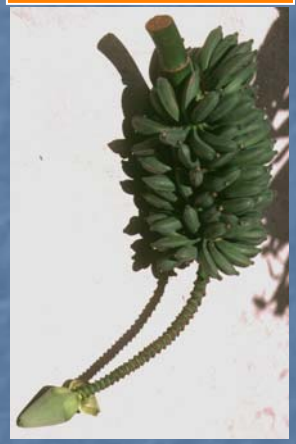
Section *Eumusa* x=11

Species *acuminata*

Species *balbisiana*

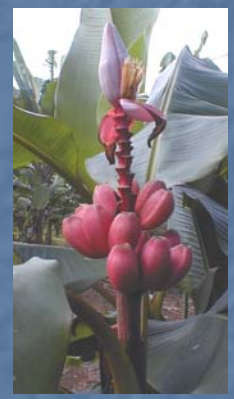


Species *schizocarpa*



Genome AA **Genome BB** **Genome SS**

Section *Rhodochlamys* x=11



Musa velutina

Most of the cultivated clones are derived from these two species

BAC ressources in *Musa*

M. acuminata spp. *burmannicoides* Calcutta 4

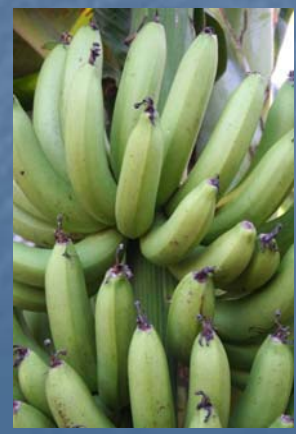
M. acuminata spp. *malaccensis* Pahang HD

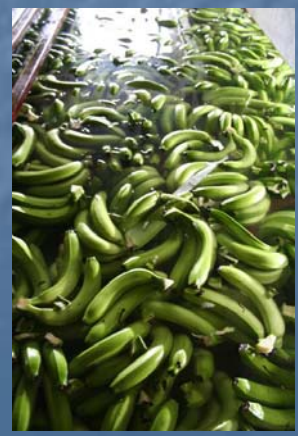
AA

M. balbisiana Pisang Klutuk Wulung (PKW) BB

Musa cv. AAA Cavendish « Grande naine » AAA

Musa cv. AA Tuu Gia AA





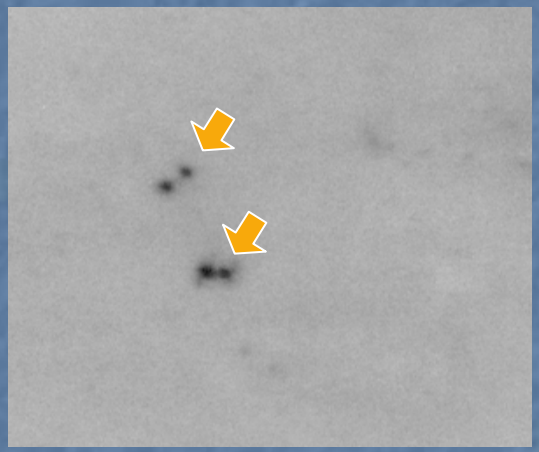
Musa BAC library screening

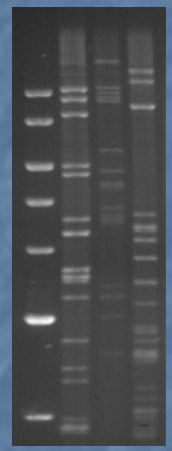
Probes have been used to screen Musa BAC library high density filters.
Positive BAC clones were identified, fingerprinted and grouped into contigs and sequenced BAC sequences are annotated.

BAC Filter hybridization

BAC Fingerprint

Re-Hybridization
with selected cDNA probes





Selection and sequencing

Sequence annotation (Automatic gene prediction and manual curation)

Musa BAC sequencing project

Introduction

Musa-Rice synteny

Musa A-B synteny

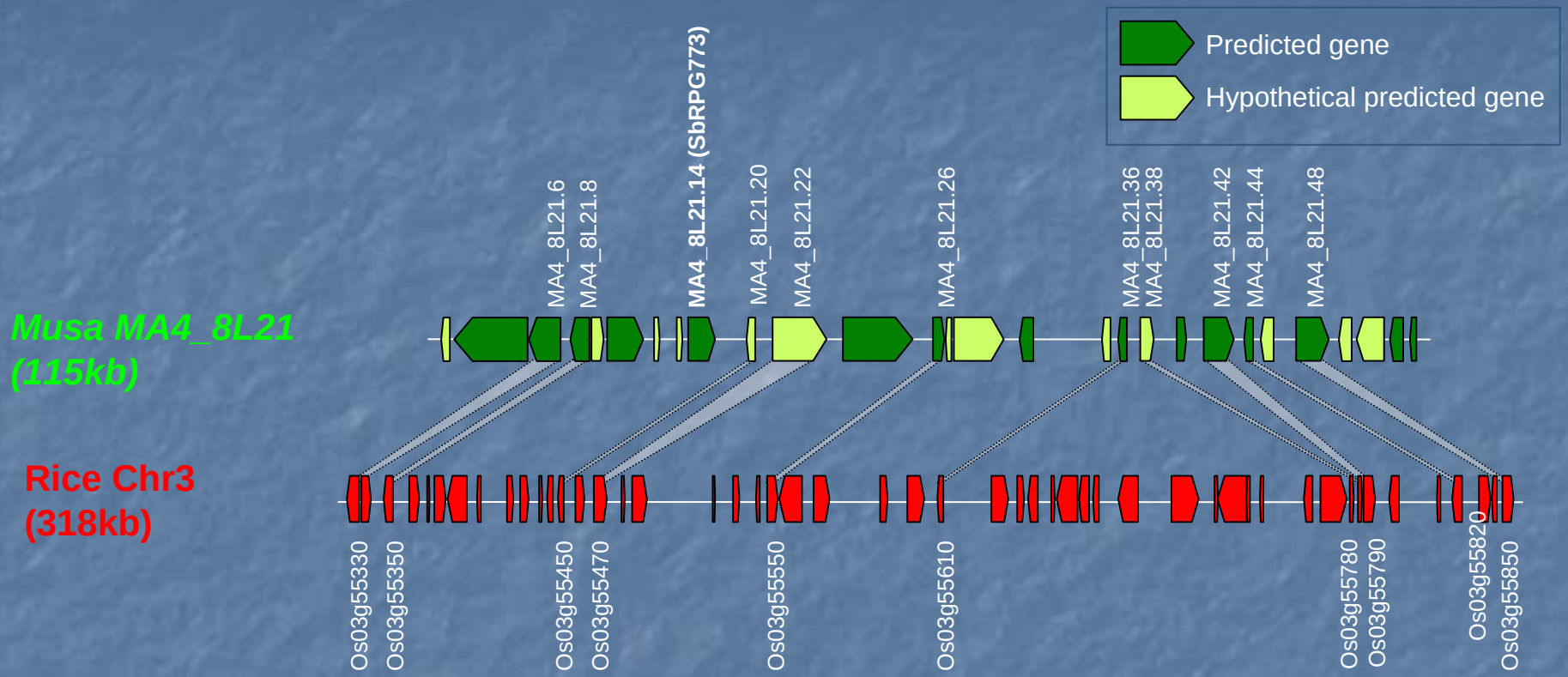
Haplotype structure within Musa genus

- RGA cluster in *Musa balbisiana*
- Gene rich region in *Musa acuminata*

Conclusion and prospects

Example of *Musa*-Rice syntenic region

Lescot et al. BMC Genomics 2007

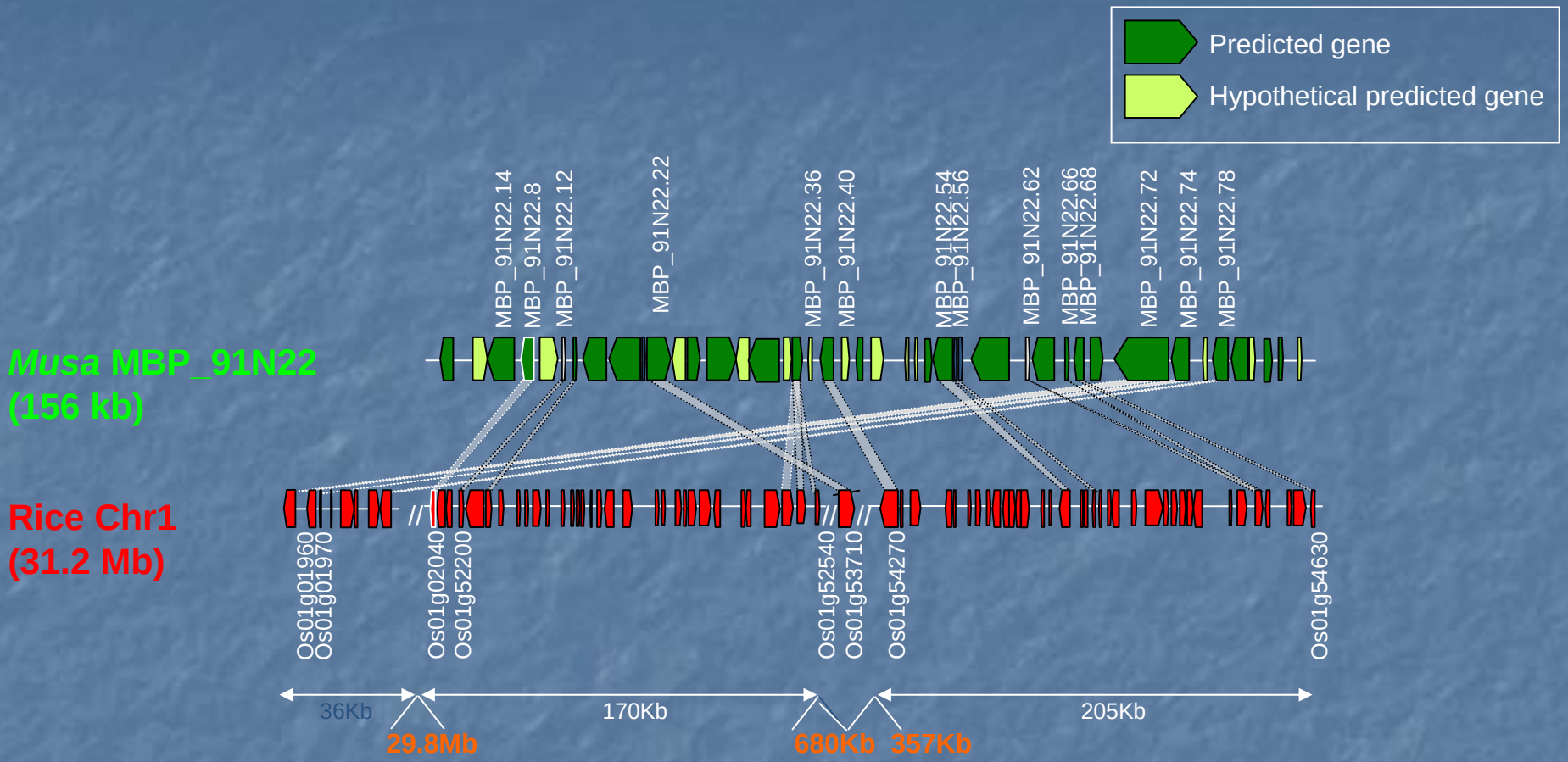


---> 10 genes in common

---> same order and orientation

---> many genes without orthologs, much more numerous in rice

Example of *Musa*-Rice syntenic region



---> 3 + 10 genes in common

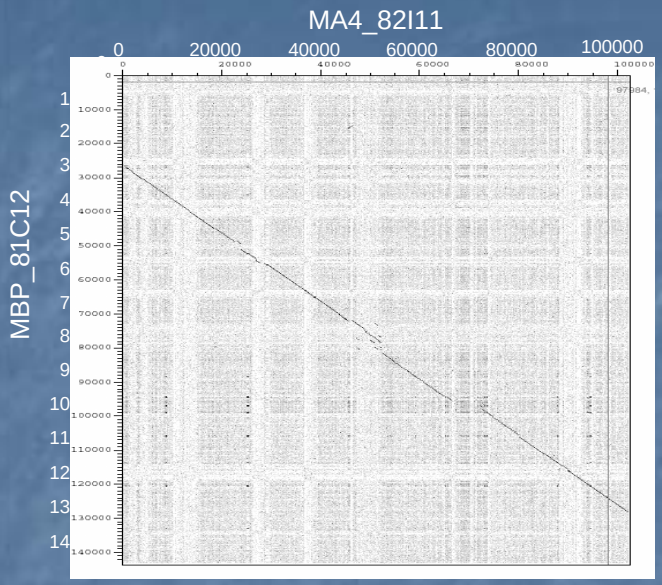
---> two separate rice regions + inversion

---> many genes without orthologs, much more numerous in rice

Musa BAC sequencing project

Introduction
Musa-Rice synteny
Musa A-B synteny
Haplotype structure within Musa genus
    RGA cluster in *Musa balbisiana*
    Gene rich region in *Musa acuminata*
Conclusion and prospects

Synteny between *M. acuminata* and *M. balbisiana* (4 BAC, 150Kb)



---> very high colinearity

---> sequence identity

- 82.9 % genomic
- 96.0 % gene

---> diverged 4.6 Mya

Musa acuminata

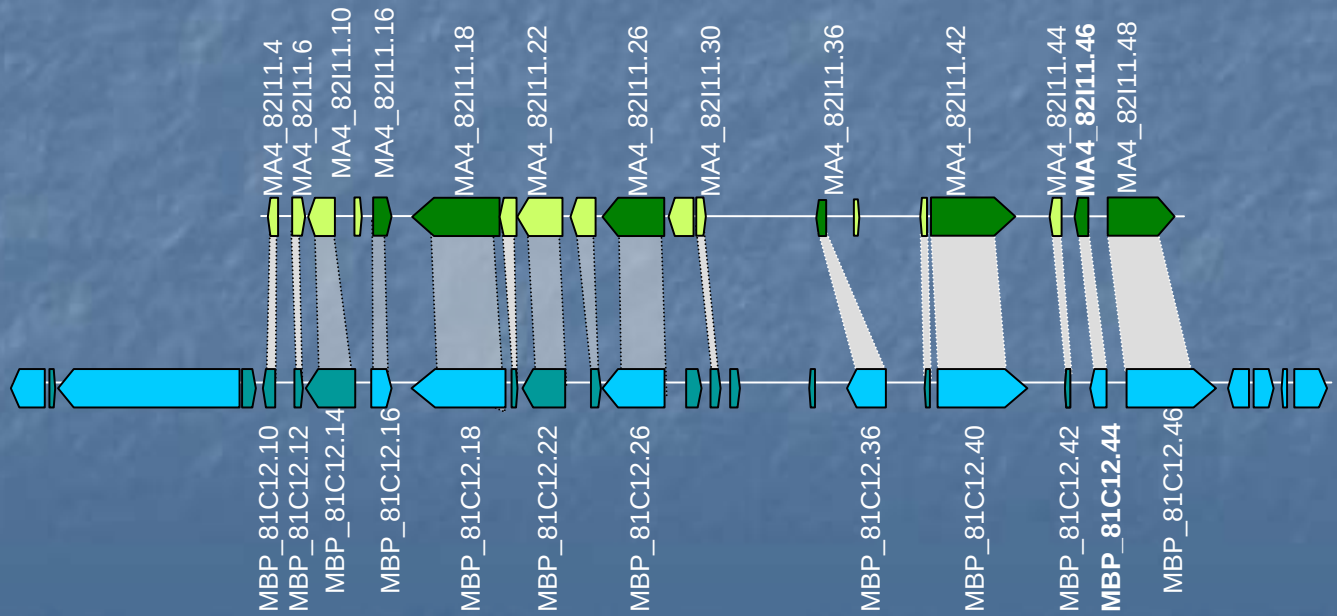
MA4-82I11

(103Kb)

Musa balbisiana

MBP-81C12

(144Kb)



Musa BAC sequencing project

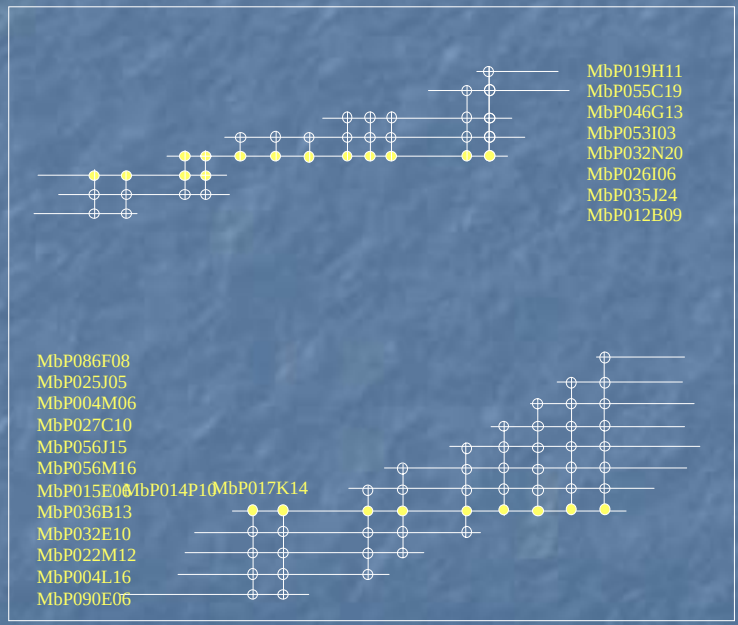
Introduction
Musa-Rice synteny
Musa A-B synteny
Haplotype structure within Musa genus
    RGA cluster in *Musa balbisiana*
    Gene rich region in *Musa acuminata*
Conclusion and prospects

Analysis of the MARGA08 gene cluster in *Musa balbisiana*

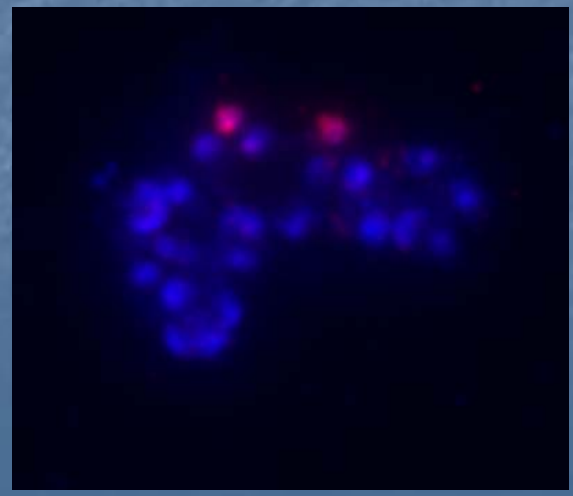
Baurens, Sidibe Bocs et al. In prep

Musa balbisiana
Clone PKW
2n = 2x = 22
B genome

The 630 pb probe MARGA08 (a resistance gene analog; Miller et al. 2008)
has been used to screen *Musa balbisiana* BAC library.
Twenty two positive BAC clones were identified, fingerprinted and grouped into two contigs



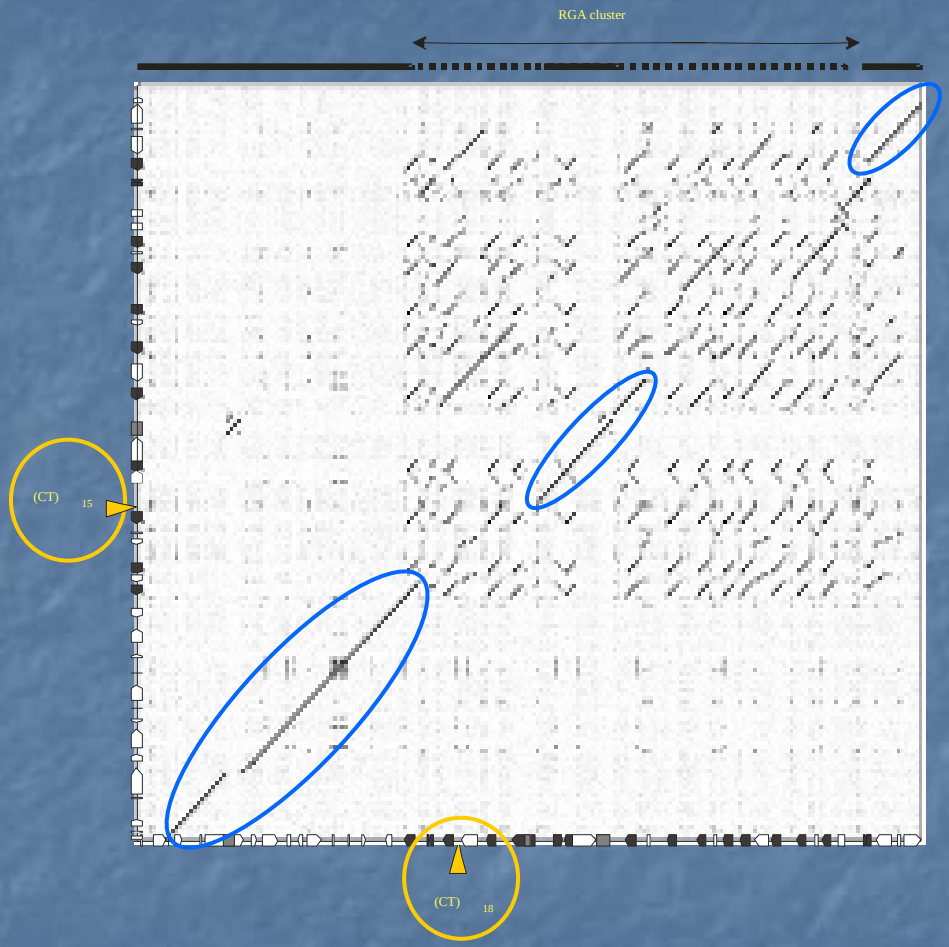
BAC FISH reveal two chromosomes



BAC contigs are haplotypes of a single locus.

A

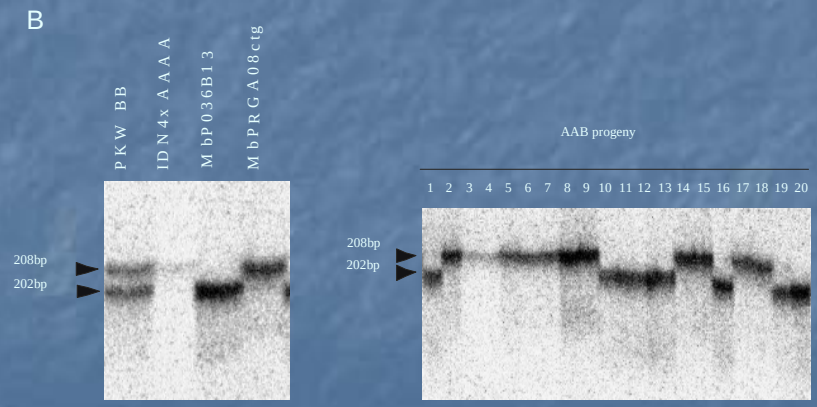
Dotplot analysis of MARGA08 locus



Three collinear and very similar zones between two BACs

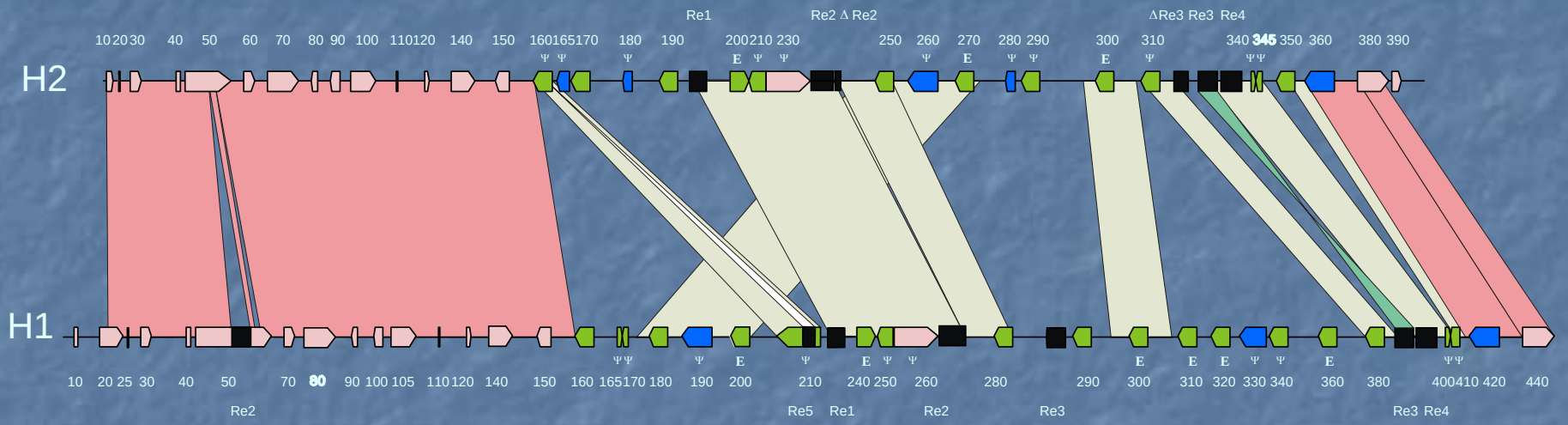
Definition of SSR marker mMaCIR 341

B



1:1 ratio alleles segregation in the 67 tested progenies (χ^2= 0.37).

Fine genomic structure : alleles and/or paralogs



Gene

Repetitive element

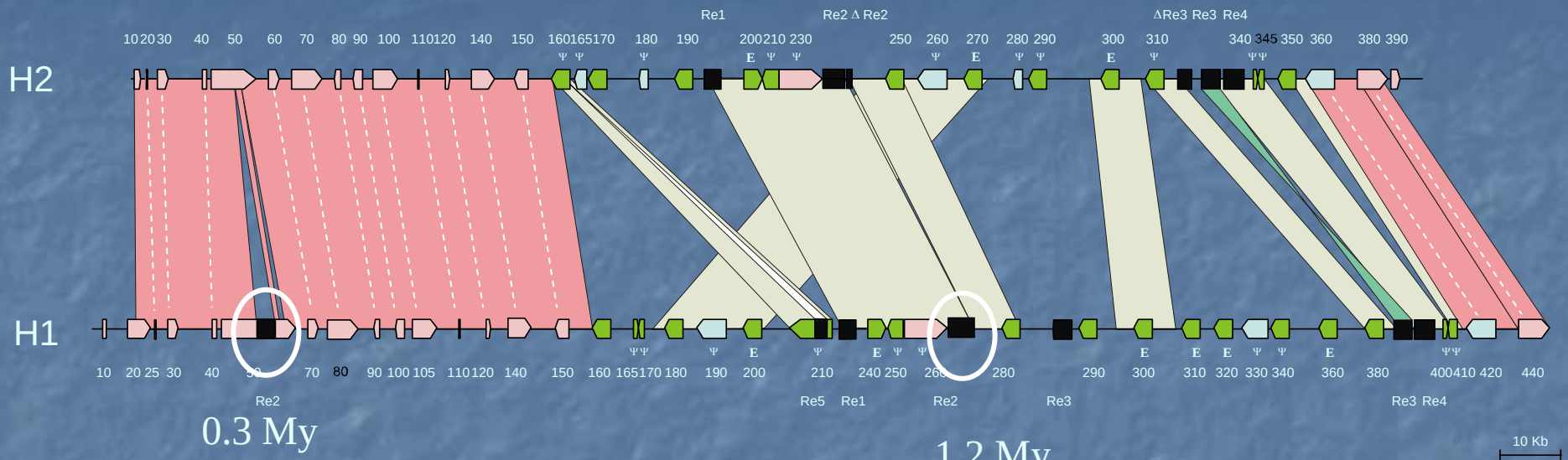
mTERF gene or fragment

RGA

• Flanking sequences: High sequence identity (98 to 99 %) with SNPs, SSR polymorphism. Only two colinearity interruptions due to repetitive element insertion.

•RGA gene cluster: Complex colinear disrupted pattern, some alleles identified, numerous paralogs, each haplotype is composed of different RGA collection

Haplotype divergence at MARGA08 gene cluster in *Musa balbisiana*

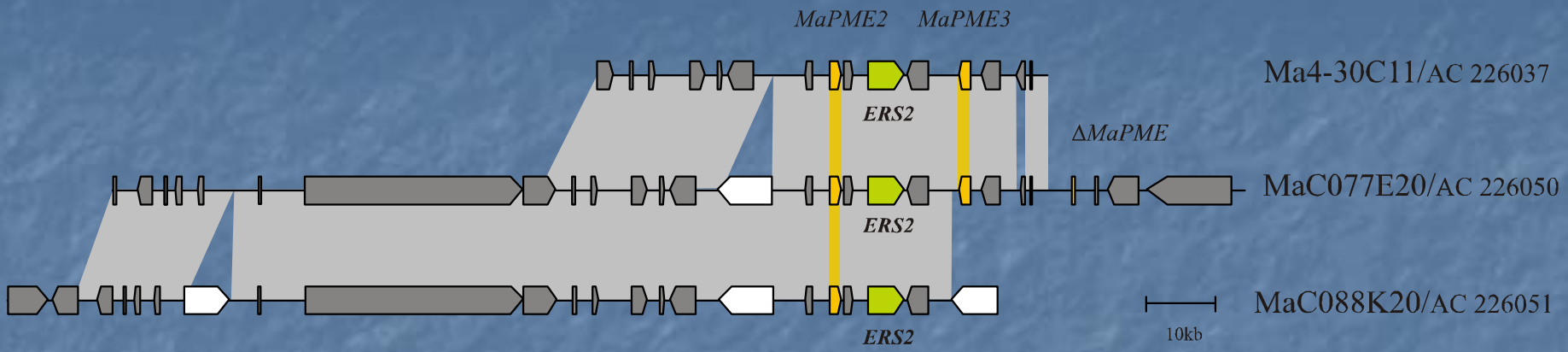


Gene	length	Sd	p-distance	Divergence
Aspartate carbamoyltransferase	519	3	0.0240	**2.67**
Conserved hypothetical protein_1	655	0	0.0000	**0.00**
Chlorophyll synthase	1185	3	0.0103	**1.14**
Hypothetical protein	339	0	0.0000	**0.00**
Plasma membrane ATPase	2865	7	0.0106	**1.18**
Conserved hypothetical protein_2	324	0	0.0000	**0.00**
Conserved hypothetical protein_3	1440	1	0.0029	**0.32**
Conserved hypothetical protein_4	315	2	0.0253	**2.81**
WRKY transcription factor	732	1	0.0055	**0.61**
Conserved hypothetical protein_5	1461	9	0.0246	**2.73**
Mitochondrial transcription termination factor	1833	1	0.0024	**0.27**
Serine threonine protein kinase	1211	0	0.0000	**0.00**
Overall	12879	27	0.0088	**0.98**

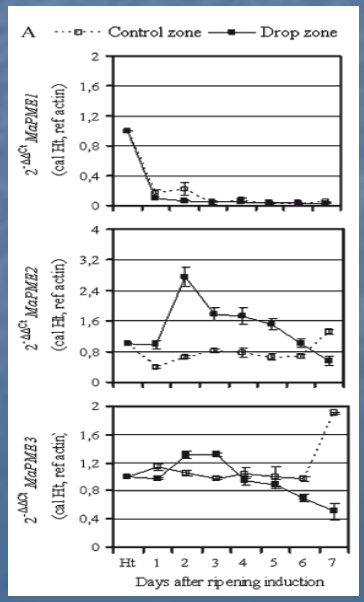
Haplotype divergence : 1My (Musa-Rice: 120 My, Musa A/B: 4.6 My)

Example of gene-rich region in *Musa acuminata*

M'Béguié et al. J.Exp Bot 2009



Three new Musa Pectine methyl esterase (PME) genes and pseudogene discovered



Allows for promoter and expression analysis

Expression patterns of cell wall modifying genes from banana during fruit ripening and in relationship with finger drop

Didier Mbéguié-A-Mbéguié, Olivier Hubert, Franc-Christophe Baurens, Takashi Matsumoto, Marc Chillet, Bernard Fils-Lycaon, and Stéphanie Sidibe-Bocs J.Exp.Bot 2009

Musa BAC sequencing project

Introduction
Musa-Rice synteny
Musa A-B synteny
Haplotype structure within Musa genus
    RGA cluster in *Musa balbisiana*
    Gene rich region in *Musa acuminata*
Conclusion and prospects

Conclusion

- Micro-syntenic regions have persisted over 117 Mya since the lineage of rice and Musa diverged

- However no general micro-synteny conservation and numerous insertions and deletions of genes ---> limited efficiency for tagging genes by comparative mapping

- High level of synteny between *M. acuminata* (Genome A) and *M. balbisiana* (Genome B)
  - Sequencing of one of these genomes will also benefit the other

- Very high degree of conservation for allelic regions except for RGA clusters

-Next step: Sequencing the *Musa* genome
  - strong foundation for studying monocot genome evolution
  - help localizing and cloning genes of agronomic interest for breeding

MusaTract projet (2009-2010)

- The proposal submitted to the French ANR by Genoscope and CIRAD was fully successful
- Project target is doubled haploid Pahang clone (M. acuminata, genome A, Cirad)
- Genome sequencing, deep sequencing of cDNA, automatic annotation and building of a platform for expert annotation are funded
- Anchoring of scaffolds to chromosomes is funded through the development of a high density genetic map

Centre National de Séquençage GENOSCOPE

CIRAD

AGENCE NATIONALE DE LA RECHERCHE ANR

musa genomics


Thanks for your attention