The Genome of the Rubber Tree (*Hevea brasiliensis*)

Marc Seguin
- Cirad, Mixed Research Unit – Plant Development and Genetic Improvement (dap), Team: Structure of Genetic Resources

1- Taxonomy & phylogenetic relationships

2- Genome status – What we know
  - ploidy
  - genome size

3- Genome Sequencing Projects

IRRDB Workshop, Montpellier, 3 June 2009

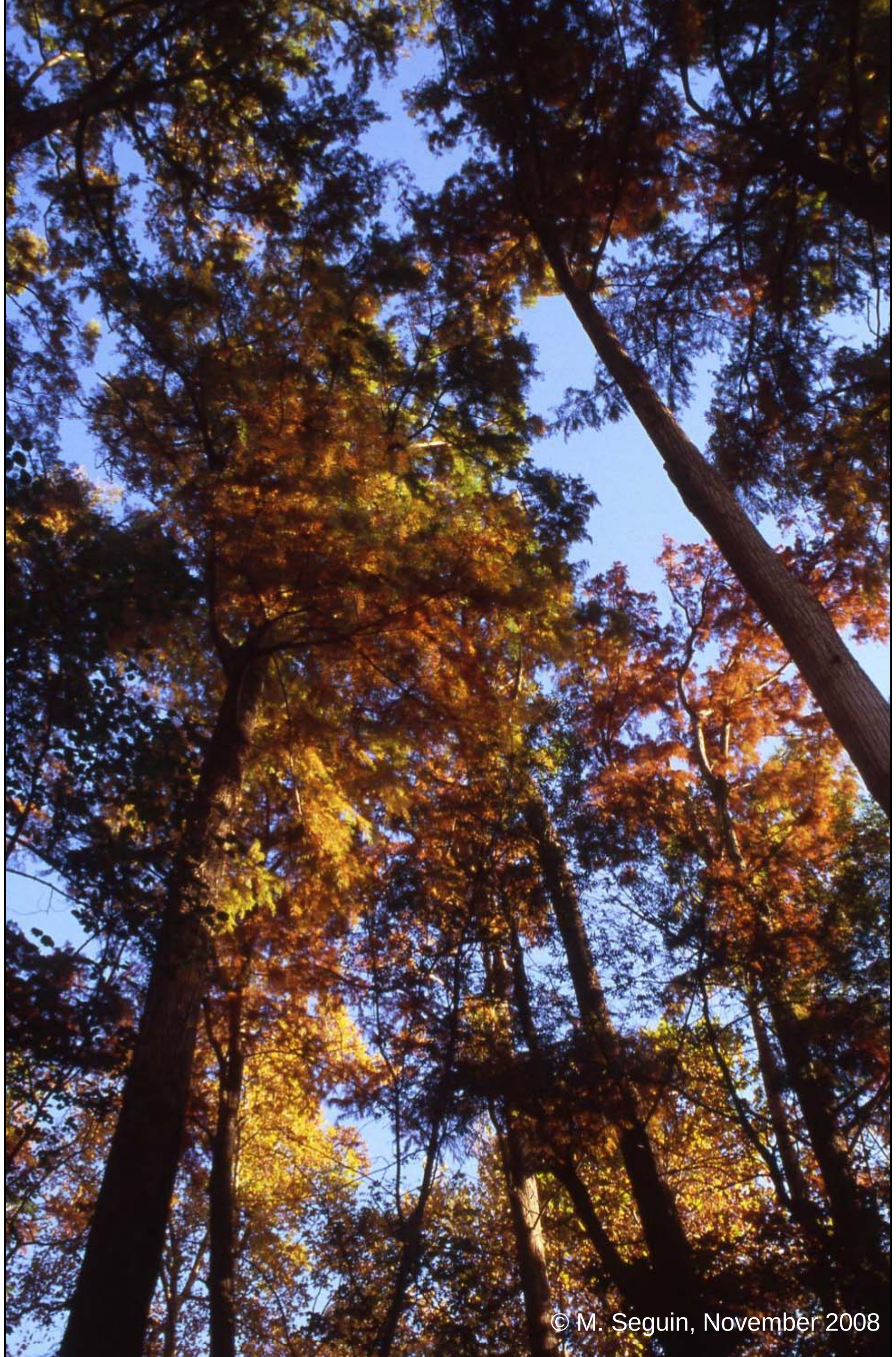
The Genome of the Rubber Tree (*Hevea brasiliensis*)

Marc Seguin
- Cirad, Mixed Research Unit – Plant Development and Genetic Improvement (dap), Team: Structure of Genetic Resources

1- Taxonomy & phylogenetic relationships

2- Genome status – What we know
    - ploidy
    - genome size

3- Genome Sequencing Projects

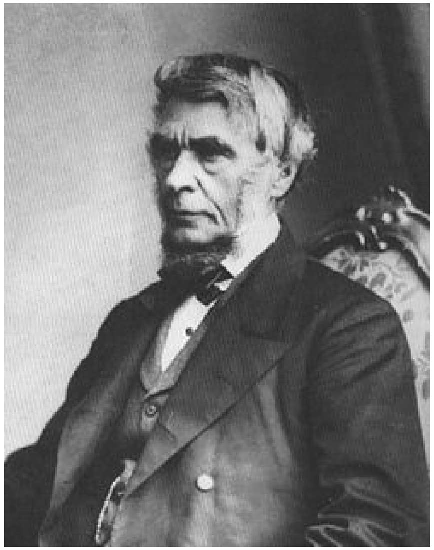


© M. Seguin, November 2008

Taxodium distichum (L.) Rich.
Red cypress
Lez river, La Valette, Montpellier

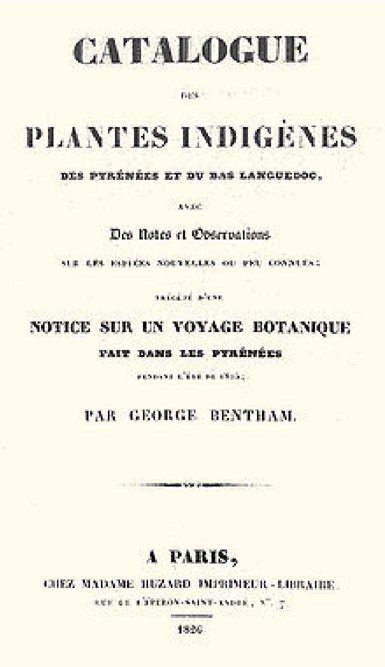
Introduced, among other species, by **George Bentham** who lived nearby Montpellier (1820 – 1835)

domaine de Restinclières, Prades le Lez, Hérault, France



George Bentham (1800 – 1884), Botanist & Taxonomist

1826:
"Catalogue of Indigenous Plants of Montpellier surrounding regions (Bas Languedoc and Pyrénées)"



CATALOGUE

DES

PLANTES INDIGÈNES

DES PYRÉNÉES ET DU BAS LANGUEDOC,

AVEC

Des Notes et Observations

SUR LES ESPÈCES NOUVELLES OU PEU CONNUES;

PRÉCÉDÉ D'UNE

NOTICE SUR UN VOYAGE BOTANIQUE

FAIT DANS LES PYRÉNÉES

PENDANT L'ÉTÉ DE [illegible];

PAR GEORGE BENTHAM.

A PARIS,

CHEZ MADAME HUZARD IMPRIMEUR-LIBRAIRE,

RUE DE L'ÉPERON-SAINT-ANDRÉ, N°. 7.

1826

Hevea taxonomy:

Hevea spruceana Benth. 1865
Hevea pauciflora (Benth.) Mull.Arg. 1935

Hevea benthamiana Mull.Arg. 1865 Bentham's hevea

And others:

Micrandra Benth. 1854, close to *Hevea* genus

Nicotiana benthamiana Domin. 1868
….

rubber tree species

During XIXth century, several tree species were used for rubber production and named **‘rubber tree’**, distinguished by their geographic origin:

- **Maniçoba** (*Manihot glaziovii,* Euphorbiaceae)
‘Ceara’ rubber tree

- **Rubber fig** (*Ficus elastica,* Moraceae)
‘India’ rubber tree

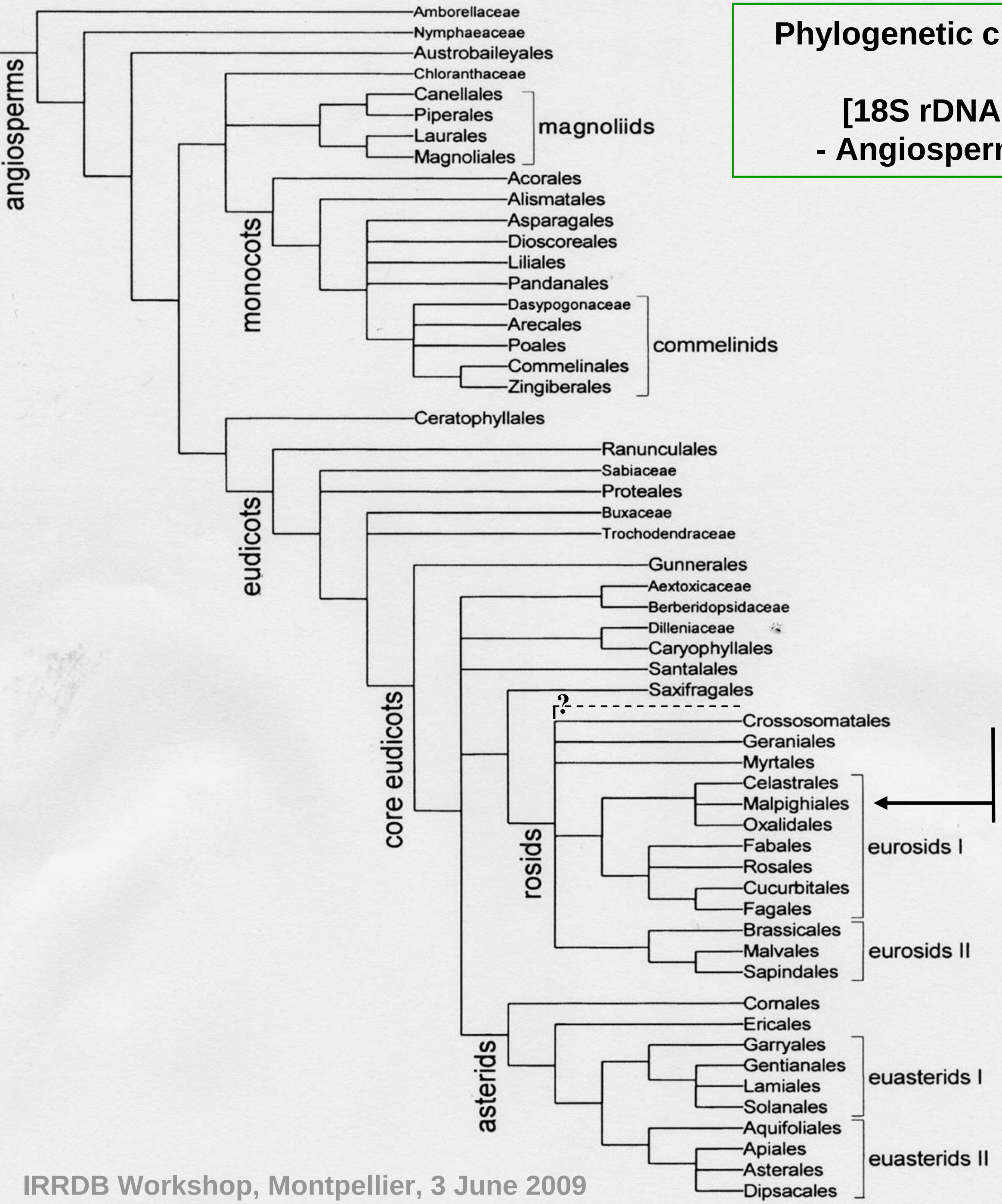
- **Hevea** (*Hevea brasiliensis,* Euphorbiaceae)
‘Para’ rubber tree

…

Today, **rubber tree** refers to:

Hevea brasiliensis (Willd. ex A.Juss.) Müll.Arg.

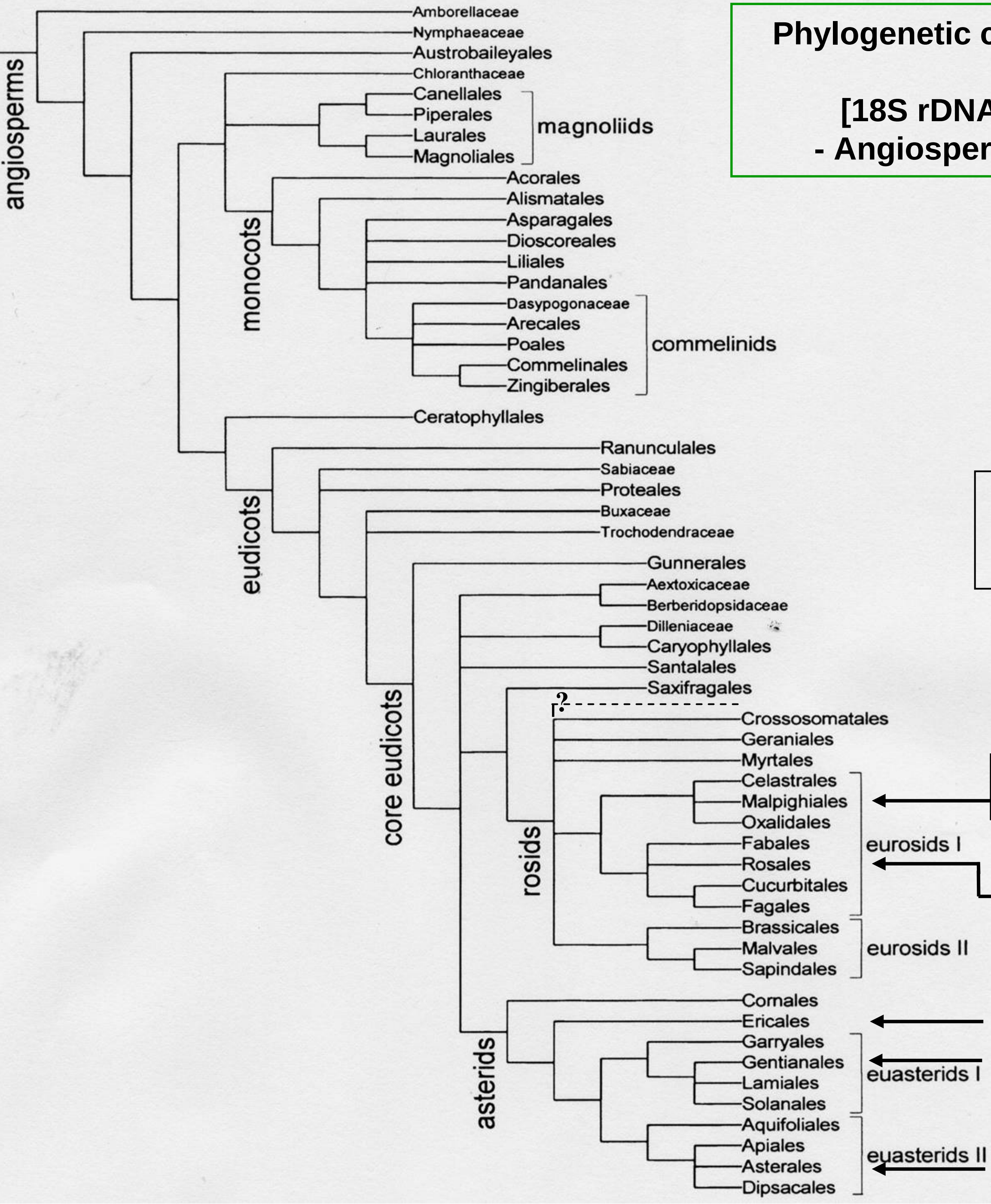
Phylogenetic classification of orders and families of flowering plants
[18S rDNA, *rbcL*, *atpB* sequences analysis]
- Angiosperm Phylogeny Group (APGII, 2003) -



Euphorbiaceae: ***Hevea***

Phylogenetic classification of orders and families of flowering plants
[18S rDNA, *rbcL*, *atpB* sequences analysis]
- Angiosperm Phylogeny Group (APGII, 2003) -

Rubber producing species: phylogenetic relationships



Euphorbiaceae: ***Hevea***,
Manihot glaziovii **(Maniçoba)**

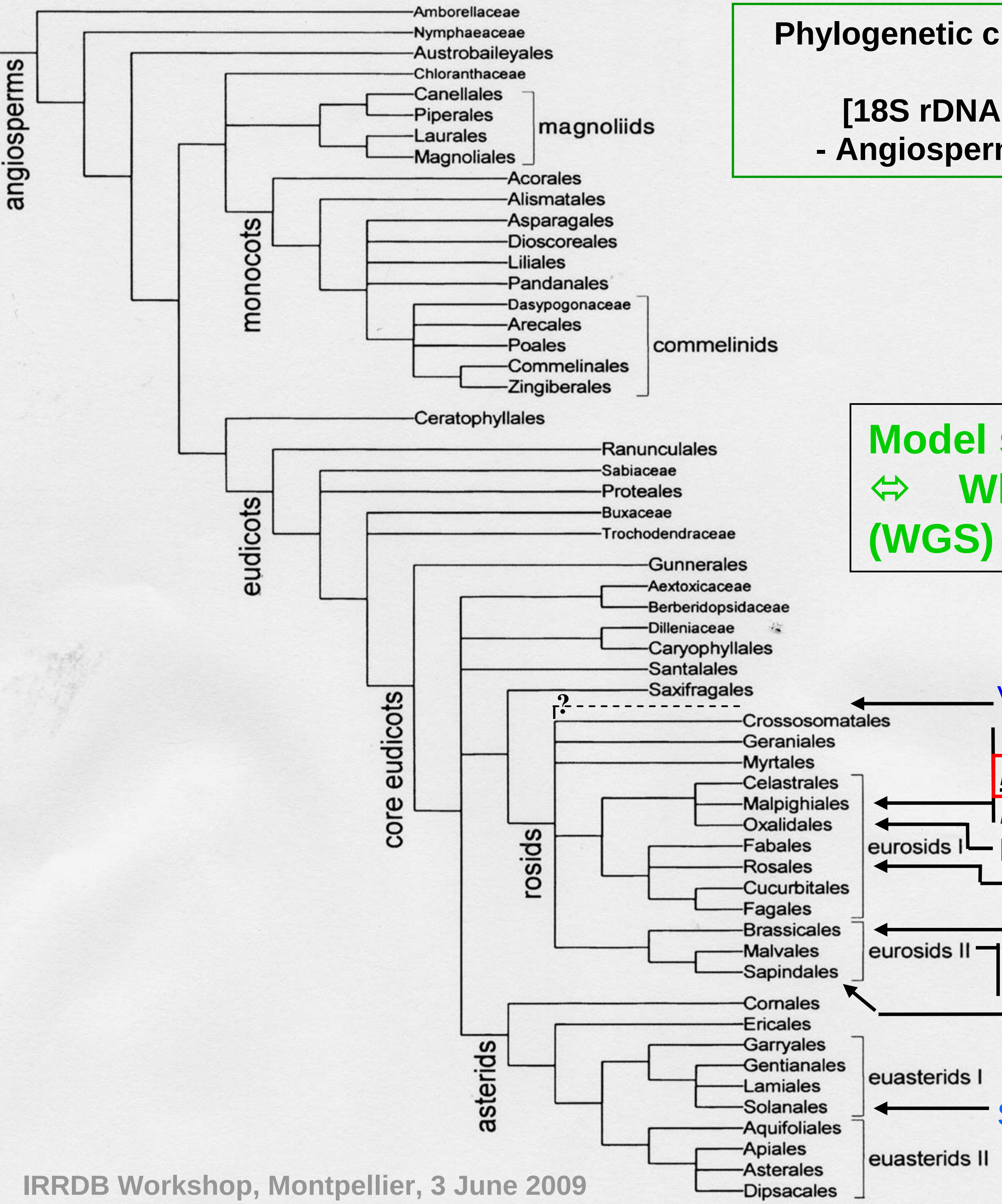
Moraceae: ***Ficus, Castilloa***

Sapotaceae: **balata, gutta-percha, chicle**

Apocynaceae:
Landolphia**,** ***Cryprostegia*****,** ***Funtumia***

Asteraceae:
guayule, ***Taraxacum kok-saghyz***

Phylogenetic classification of orders and families of flowering plants
[18S rDNA, *rbcL*, *atpB* sequences analysis]
- Angiosperm Phylogeny Group (APGII, 2003) -



Model species for *Hevea* genomics ⬄ Whole Genome Sequencing (WGS) achieved or in progress

Vitaceae: grape

Euphorbiaceae: *Hevea*,
***Manihot esculenta* (cassava),**
***Ricinus communis;* Salicaceae: poplar**

Fabaceae: *Medicago truncatula*

Rosaceae: apple tree

Brassicaceae: *Arabidopsis*

Malvaceae: cotton, cocoa

Rutaceae: *Citrus*

Solanaceae: tomato

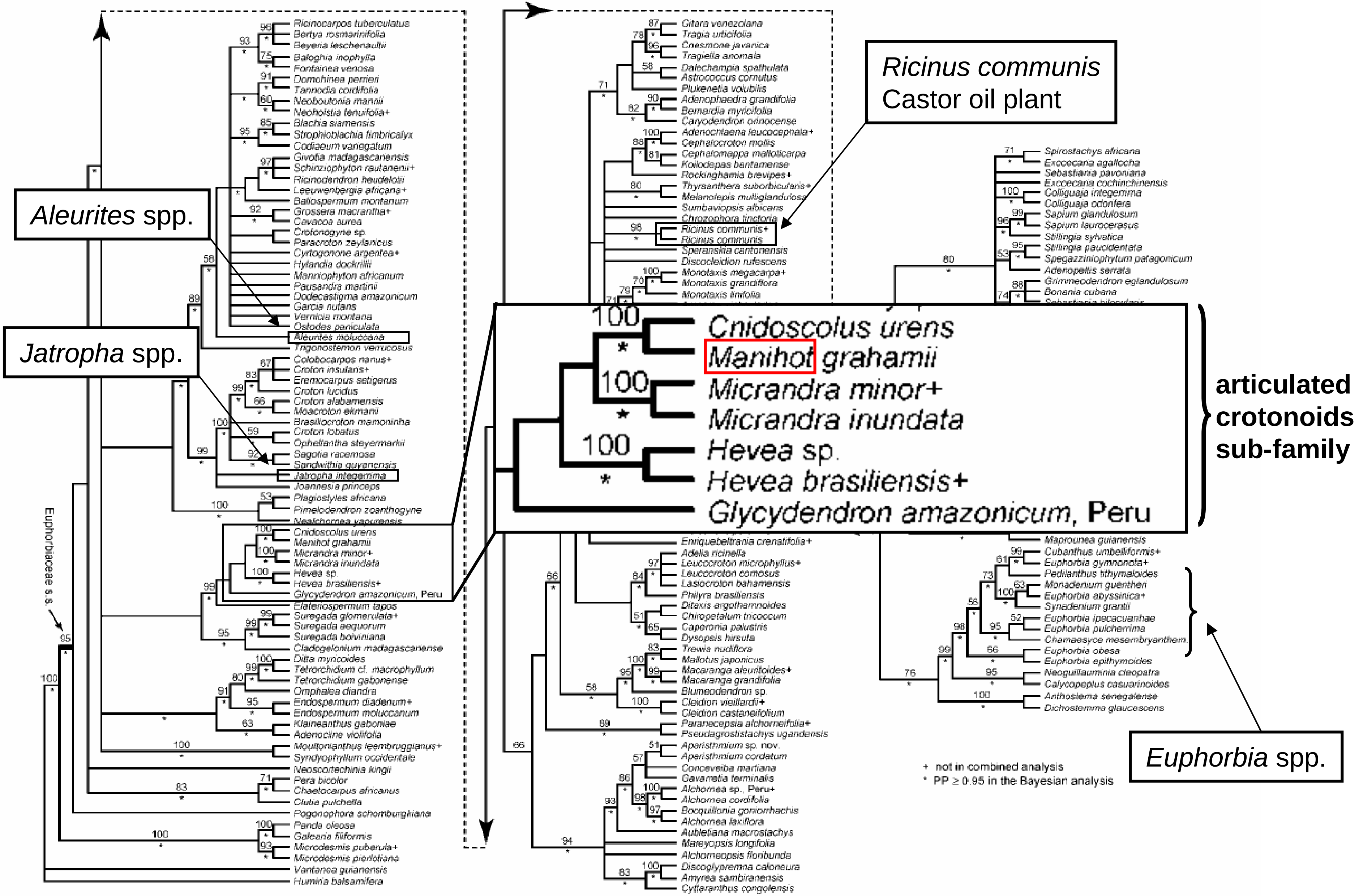


Fig. 1. Strict consensus of 1.0 × 10^6 minimal length trees (tree length = 2358, CI = 0.306, RI = 0.706) resulting from parsimony analysis of *rbcL* data for Euphorbiaceae s.s. The numbers above the branches are bootstrap percentages ≥50%; indicated below branches are posterior probabilites ≥0.95 in the Bayesian analyses.

The Genome of the Rubber Tree (*Hevea brasiliensis*)

1- Taxonomy & phylogenetic relationships

2- Genome status – What we know
- ploidy
- genome size

3- Genome Sequencing Projects

The *Hevea* genome is diploid

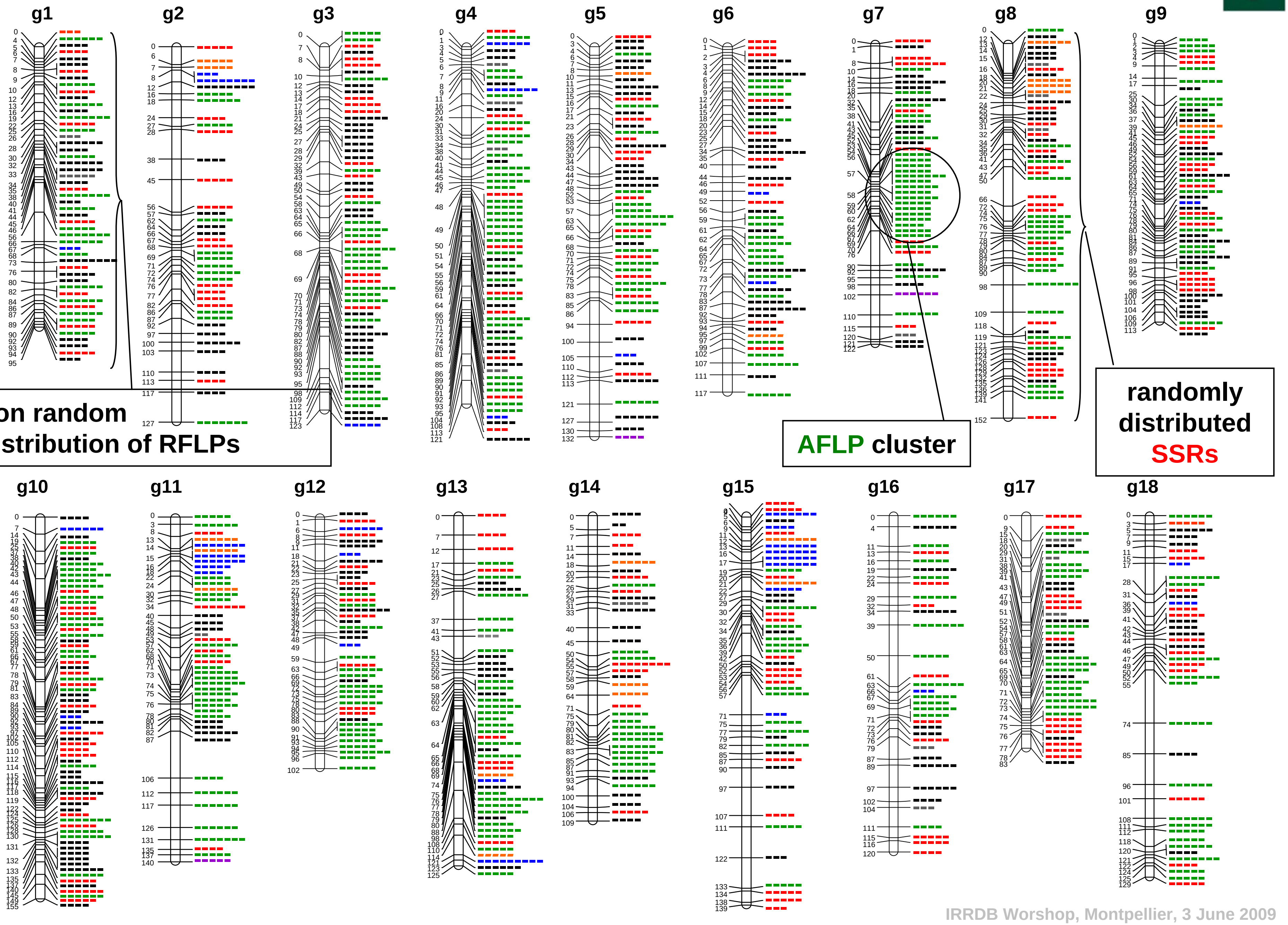
- The 10 *Hevea* species (*Hevea* spp.) have 2n = 36 chromosomes
- Based on cytogenetics studies, there has been a suggestion of allotetraploidy (2n = 4x = 36), but with unknown hypothetical diploid ancestral species

- Nevertheless, Lespinasse *et al.* (2000) observed only disomic segregation for the hundreds of markers analysed in mapping work, demontrating the <u>diploid structure of *H brasiliensis* genome</u>
- Experimental crosses revealed <u>no genetic barriers</u> and hevea interspecific hybrids may occur naturally. In addition, genetic mapping using a *H. brasiliensis x H benthamiana* hybrid as progenitor (RO38, Lespinasse et al 2000) revealed a complete <u>colinearity of the genome of these 2 species</u> and no chromosome rearrangement were suspected (few distorted segregation observed).

=> ***Hevea* genus can be considered as diploid with 2n = 2x = 36.**

Current status of the *Hevea* reference map (PB260xRO38): RFLPs, AFLPs and SSRs

CIRAD



non random distribution of RFLPs

AFLP cluster

randomly distributed SSRs

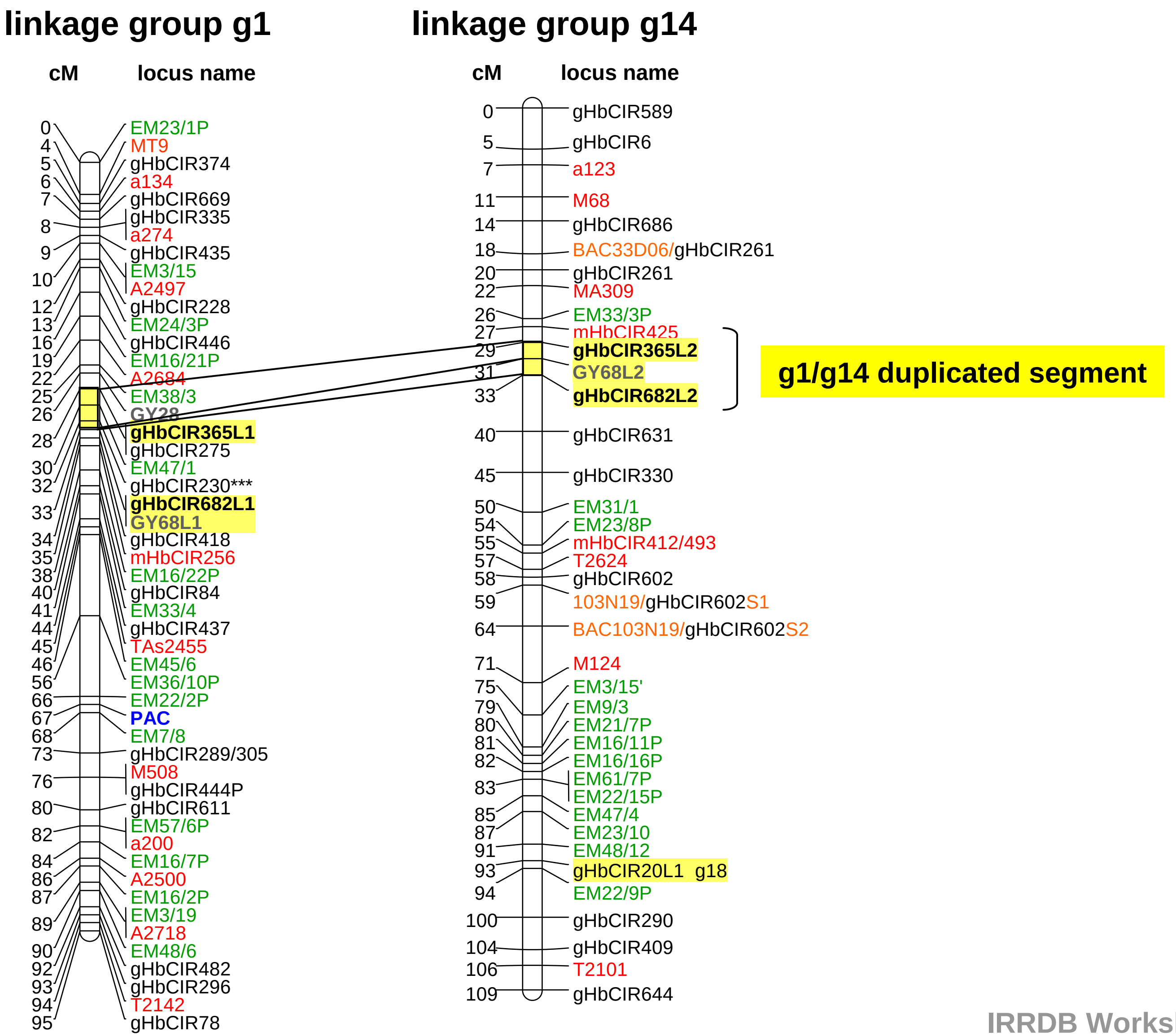
Duplications in *Hevea* genome

- Southern hybridization using genomic single copy probes in RFLP (Restriction Fragment Length Polymorphism) mapping revealed <u>10 duplicated chromosome segments</u> located on 9 chromosomes
- But, there is apparently no pairing between homeologous chromosomes during meiosis, and <u>inheritance of duplicated loci remains disomic</u>

=> *Hevea* genome can be considered a segmental polyploid with only disomic inheritance

Duplicated chromosomal segments in rubber tree genome revealed by RFLP markers [PB260xRO38 map]

(from Lespinasse et al 2000)

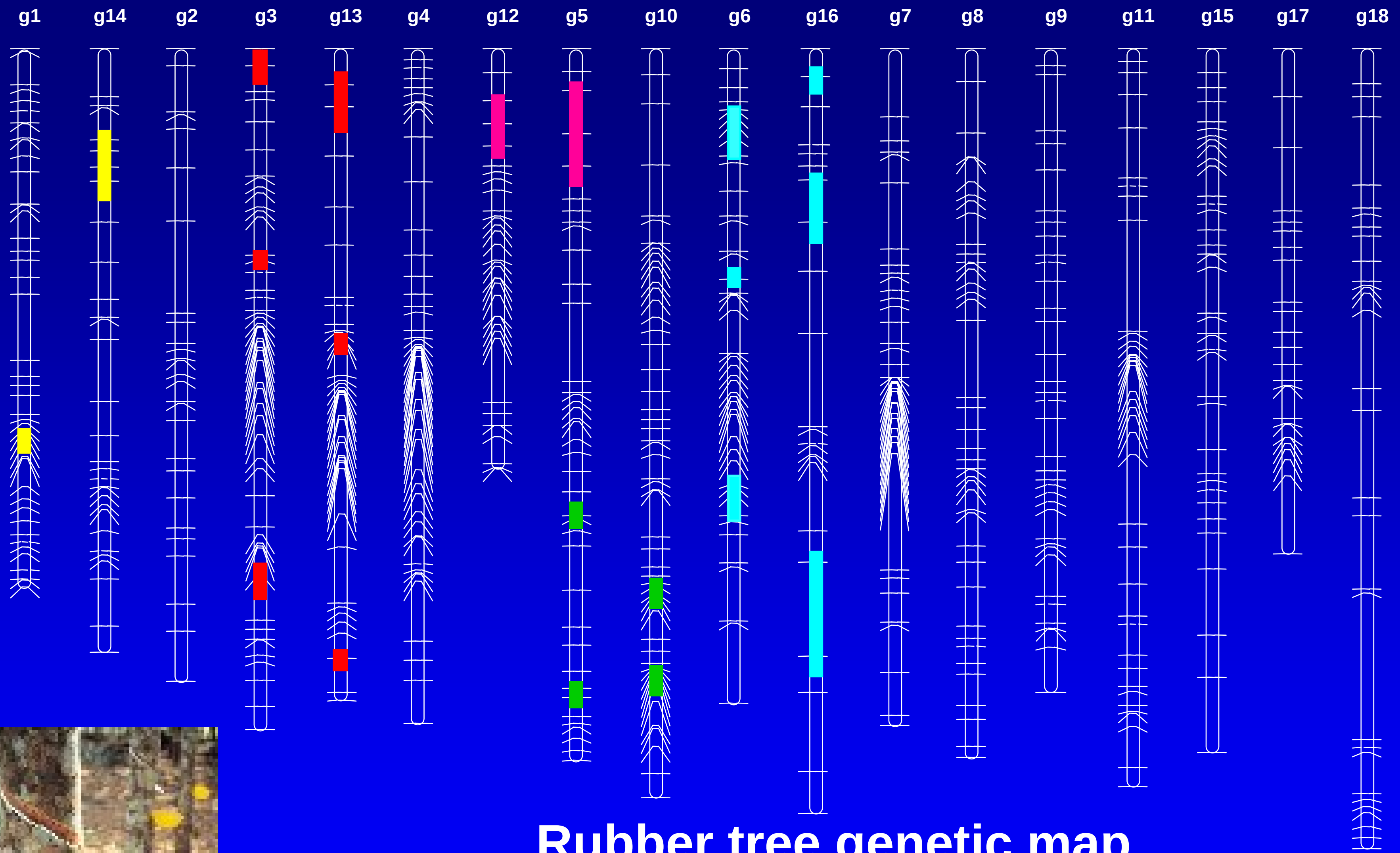


SSR

AFLP

RFLP

PAC: Isozyme locus



Rubber tree genetic map
717 loci (RFLP, AFLP, microsatellites, isozymes)
1C = 2000 Mb, 2n = 36, 2230 cM

Coloured segments: duplicated chromosome regions **Lespinasse *et al.* (2000a)**

Similarity between rubber tree and cassava genomes

Interestingly, the *Manihot* genus has a similar cytogenetic background:

- All the 100 *Manihot* spp. have 2n = 36 chromosomes
- Cytogenetic studies gave evidence of allo-tetraploidy for at least cassava (*M. esculenta)* and maniçoba (*M. glaziovii*) – but with disomic inheritance of genetic markers even in the case of duplicated loci or chromosomal segments.

Similarity between rubber tree and cassava genomes

Interestingly, the *Manihot* genus has a similar cytogenetic background:

- All the 100 *Manihot* spp. have 2n = 36 chromosomes
- Cytogenetic studies gave evidence of allo-tetraploidy for at least cassava (*M. esculenta)* and maniçoba (*M. glaziovii*) – but with disomic inheritance of genetic markers even in the case of duplicated loci or chromosomal segments.

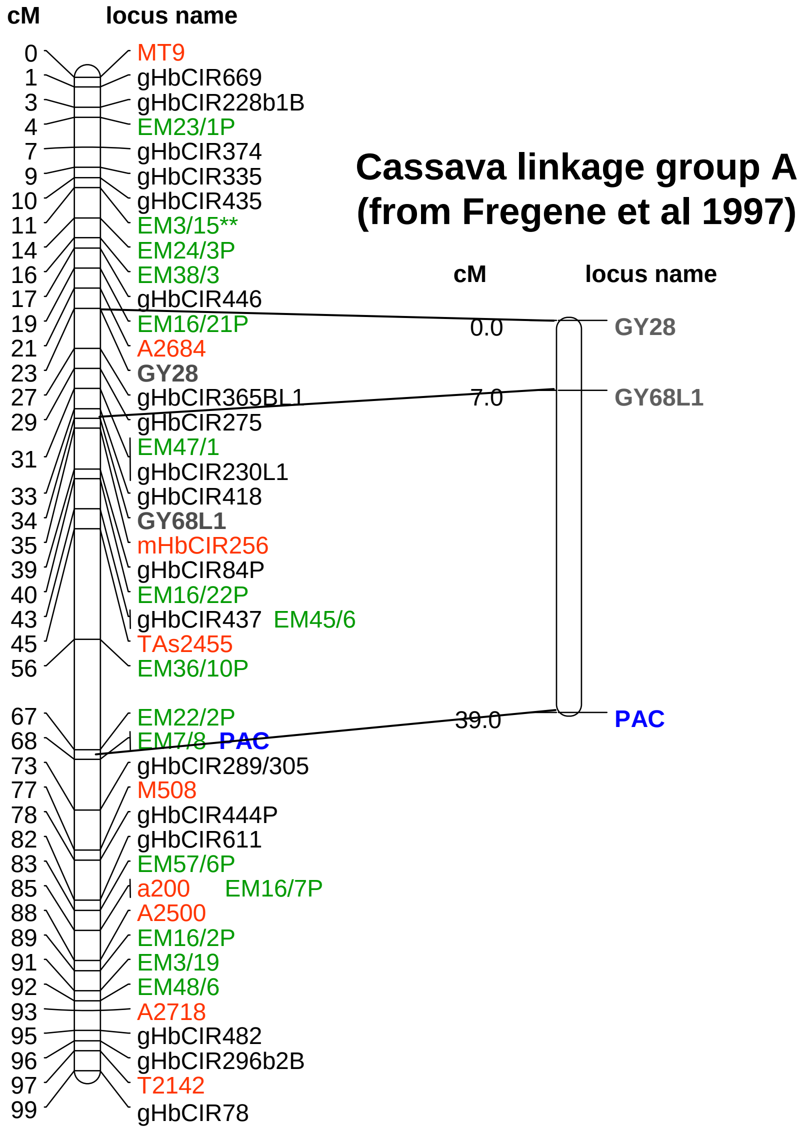
=> *Hevea* and *Manihot* can be considered as diploids for genetic studies, or as segmental polyploids for structural genomics

=> for *Hevea* the origin of the observed duplications is not known: 1) trace of ancient polyploidization event, or 2) more recent chromosome segments duplications (?)

Colinearity between rubber tree and cassava genomes

Rubber tree linkage group g1 (from Lespinasse et al 2000)

Cassava linkage group A (from Fregene et al 1997)



-18/26 cassava genomic probes revealed 17 loci in rubber tree map

- 8 loci mapped in common revealed colinearity between the 2 genomes

SSR

AFLP

RFLP / *Hevea* genomic probes

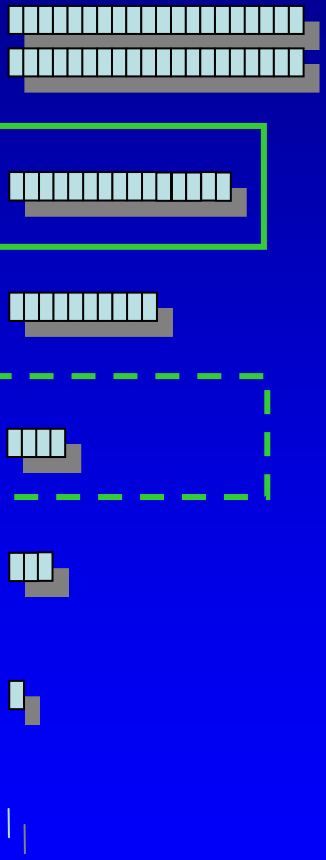
RFLP / Cassava genomic probes (GY#)

PAC: Phosphatase Acid Isozyme locus

M. Seguin, pers. Comm.

Genome size of several species

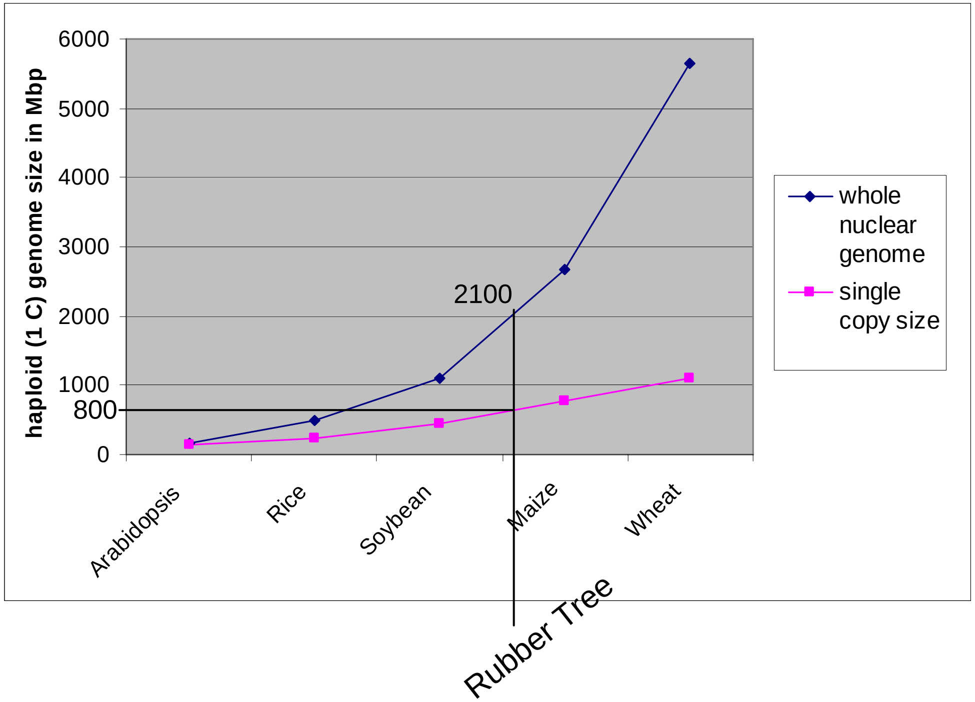
Rubber tree	2.1 10^9pb
Cassava	7.8 10^8pb
Poplar	4.8 10^8pb
Castor bean	2.2 10^8pb
Arabidopsis	1.5 10^8pb
M. grisea (fungus)	4 10^7pb
E. coli	5 10^6pb



Single copy fraction of plant genome

The nuclear genome of flowering plants varies greatly in size, mainly because of a variable amount of repetitive DNA. The single copy fraction, containing the expressed genes, is less variable.

from Morgante et al, Nature Genet 2002 => genome size of model plants



⟹ **A rough graphical estimate of single copy genome size of rubber tree:**
< 800 Mbp ⇔ < 40 % of whole genome

Comparative genomics for rubber tree is needed

Cassava (*M. esculenta)* is the best model plant species for genomics of rubber tree: small genome size, close phylogenetic relationship with *Hevea*, and consequently a strong genome colinearity is expected between *Hevea* and *Manihot*.

Whole Genome Sequencing of cassava is in progress:

- Donald Danforth Plant Science Center, USA (Dr Claude Fauquet); grants from the U.S. Department of Energy Joint Genome Institute (http://www.ars.usda.gov/IS/pr/2006/060830.htm)

Other model species :

- castor bean (*Ricinus communis*), genetically more distant from *Hevea* but with a very small nuclear genome already sequenced (public availability of sequence from Craig Venter Institute ?)
- poplar (*Populus trichocarpa)*, genome annotation in progress

The Genome of the Rubber Tree (*Hevea brasiliensis*)

1- Taxonomy & phylogenetic relationships

2- Genome status – What we know
    - ploidy
    - genome size

3- Genome Sequencing Projects

Whole Genome Sequencing of the nuclear genome of rubber tree – 1/2

A tremendous task:

- due to genome size, extent of highly repetitive DNA, chromosome segments duplications and … heterozygosity

  <= all rubber tree accessions are highly heterozygous: for example, from 41 to 71 % or the 450 SSR loci were found heterozygous in cultivated rubber tree clones

⇒ A draft WGS would provide thousands of small unarranged contigs of DNA sequences, and progress in physical mapping might be limited; and gene identification and annotation difficult.

⇒ at least one useful output can be expected from this draft sequence: a great number of SSR markers (???)

Whole Genome Sequencing of the nuclear genome of rubber tree - 2/2

Possibilities of action:

- **Concentrate sequencing effort on the single copy fraction of the genome** in order to avoid wasting time and money on repetitive DNA:
  - Large scale EST / cDNA sequencing (-> gene annotation) and re-sequencing (-> SNPs identification)
  - Genomic DNA enriched in single copy genome using for instance Cot hybridization methods…
  - ….
- **Choice of the sequencing / data mining strategy and methods**:
  - GS-FLEX
  - Whole Genome Profiling (Keygene N.V. and Amplicon Express Inc.) => Whole Genome sequence-based physical BAC map
  - **Sequence alignment with cassava** / castor bean / poplar

Whole Genome Sequencing of the nuclear genome of rubber tree - 2/2

Possibilities of action:

- **Concentrate sequencing effort on the single copy fraction of the genome** in order to avoid wasting time and money on repetitive DNA:
  - Large scale EST / cDNA sequencing (-> gene annotation) and re-sequencing (-> SNPs identification)
  - Genomic DNA enriched in single copy genome using for instance Cot hybridization methods…
  - ….
- **Choice of the sequencing / data mining strategy and methods**:
  - GS-FLEX
  - Whole Genome Profiling (Keygene N.V. and Amplicon Express Inc.) => Whole Genome sequence-based physical BAC map
  - **Sequence alignment with cassava** / castor bean / poplar

- **Or wait for next generations of DNA sequencing technologies**

⟹ **The aim of the present IRRDB workshop**

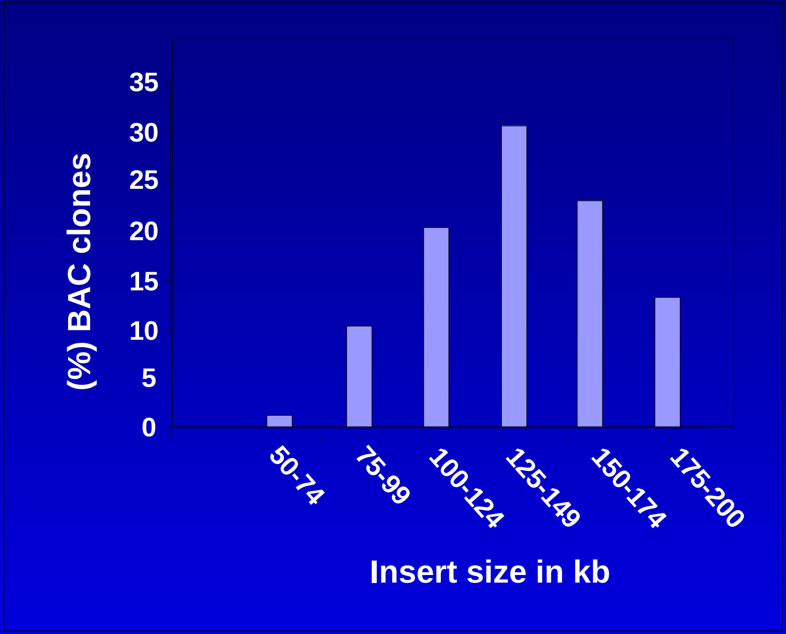
Thank you for your attention

RUBBER TREE (*Hevea* spp.) BAC library

- **Species: (interspecific hybrid)**

H.benthamiana* x *H. brasiliensis

- **Genotype: FX3899 (syn. RO38)**
- **Ploidy: diploid 2n=2x = 36**
- **Genome size: 2.1 10^9 bp / 1C**
- **Scientific interest:**
  - **✓disease resistance (*Microcyclus ulei*)**
  - **✓linkage disequilibrium**



- **Number of BAC clones: 92,160**
- **Average BAC size: 100 Kb**
- **Genome equivalents: 3.5 - 4 X**

from Piffanelli *et al.* 2002

rubber tree taxonomy

Hevea brasiliensis (Willd. ex A.Juss.) Müll.Arg.

Phylogenetic classification of Angiosperms (APGII) :

class : Eudicot
sub-class: Rosids
order: Malpighiales
family: Euphorbiaceae -> 5735 species
sub-family: articulated crotonoids
tribe: Micrandreae
genus: *Hevea* -> 10 species
species: *brasiliensis*