

# *Eucalyptus*: sequencing a global tree genome for energy, fiber, and wood



Alexander A. Myburg, **Dario Grattapaglia**, Gerald A. Tuskan,  
Jeremy Schmutz, Daniel S. Rokhsar, Kerrie Barry, Jim Bristow,  
The Eucalyptus Genome Network (EUCAGEN)

## GENOMICS OF ENERGY & ENVIRONMENT

U.S. DEPARTMENT OF ENERGY  
JOINT GENOME INSTITUTE (JGI)

THIRD ANNUAL  
USER MEETING

MARRIOTT HOTEL  
WALNUT CREEK, CALIFORNIA  
MARCH 26-28, 2008



# EUCAGEN - Eucalyptus Genome Network

*Azores June 2007*



~130 scientists in 18 countries

## SEQUENCING OF THE *EUCALYPTUS* GENOME: A PROPOSAL TO DOE-JGI



Prepared by

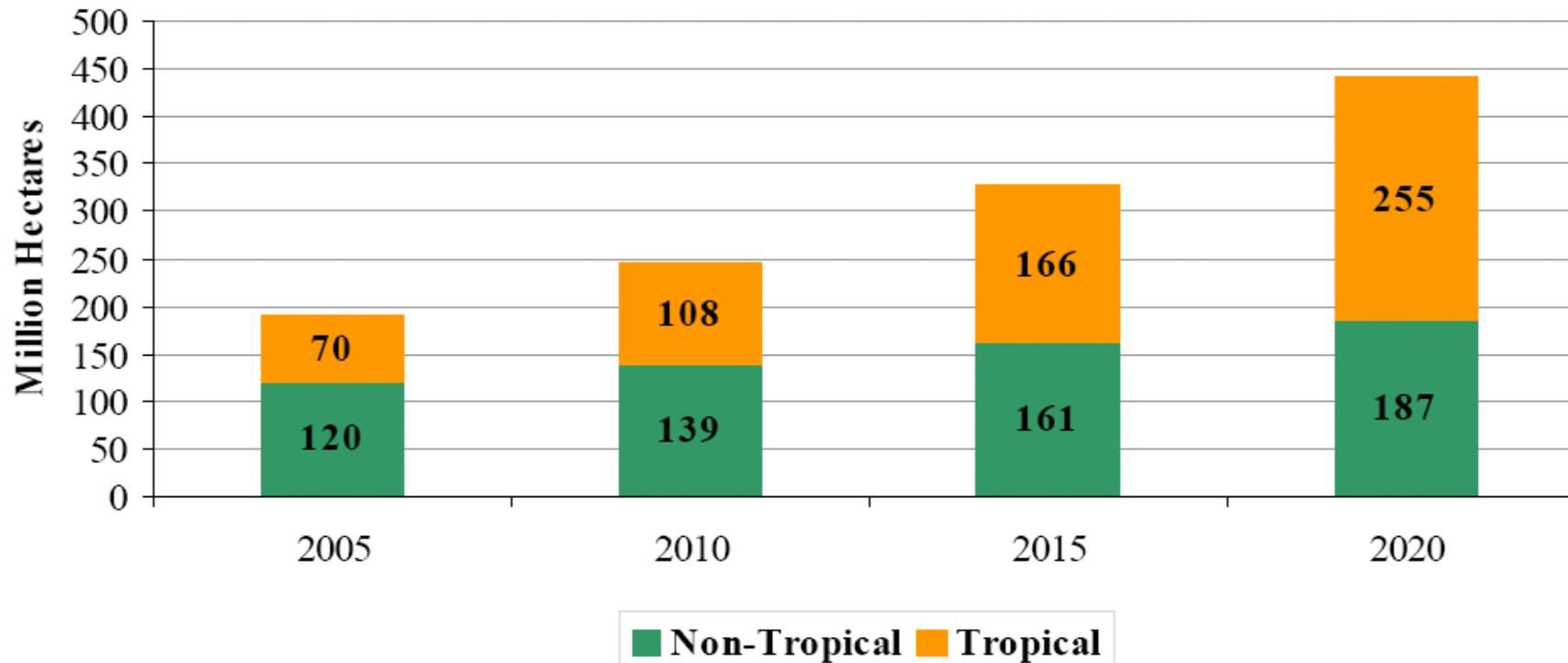
Zander Myburg, Dario Grattapaglia, Brad Potts, Carlos Labate, Gerd Bossinger,  
Margaret Byrne, Rena Vallancourt, Ron Sederoff, Simon Southerton

and other members of

the *Eucalyptus* Genome Network  
(EUCAGEN)

18 June 2006

# Projected world forest plantation area



- ✓ Eucalyptus is currently the most productive and main planted species in the tropics
- ✓ Key species for pulp, energy and increasing use for solid wood
- ✓ By 2020 Eucalyptus pulp will have over 60% of the international trade of short fibre pulp
- ✓ Expected larger share in the sawnwood and plywood industry, taking market shares of tropical timber from natural forests

# Why *Eucalyptus* ?

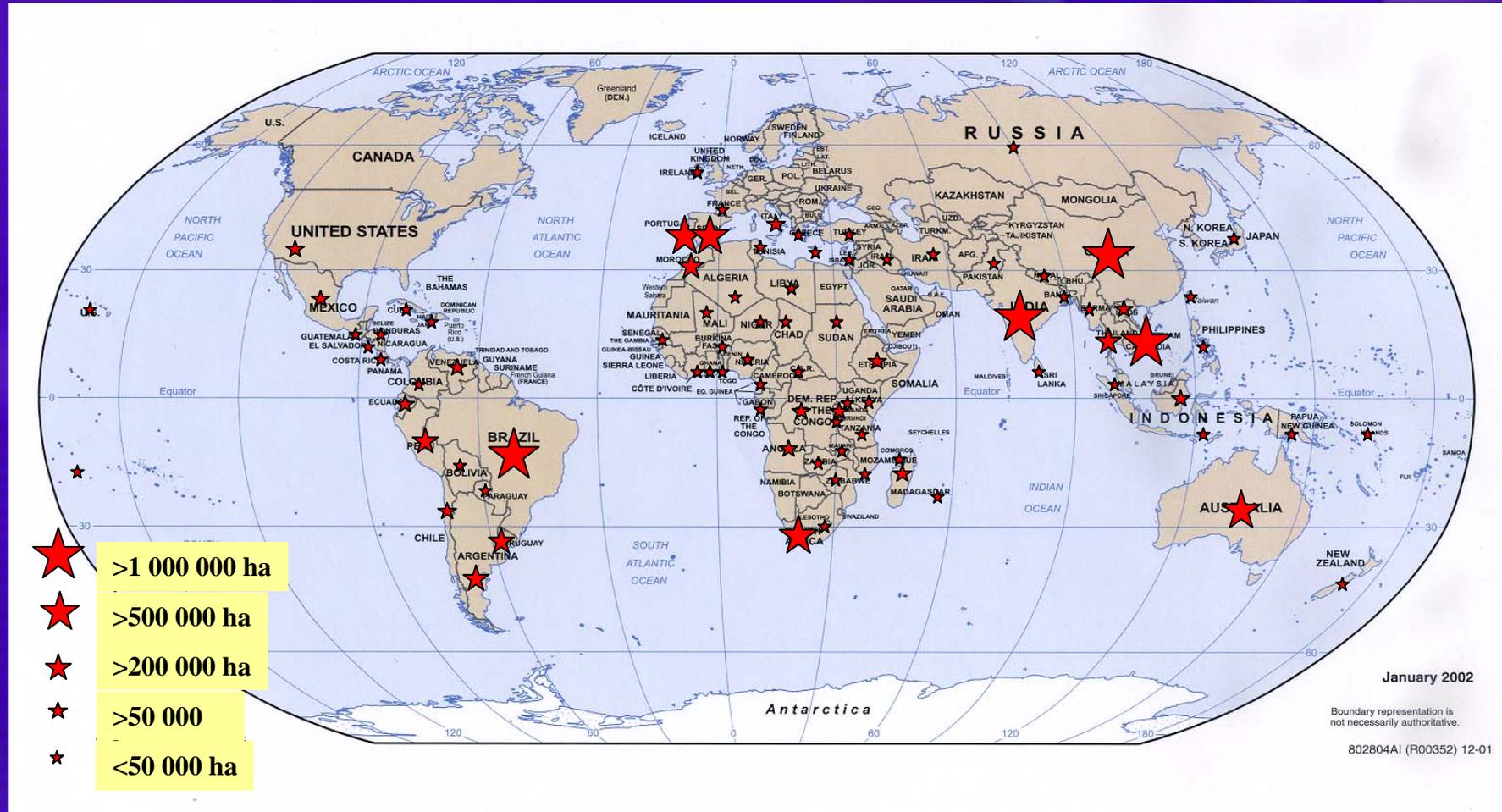
- ✓ **Target forest crop for the "Genomes to Energy" sequencing program**
  - ✓ Global tree with superior growth, adaptability, and wood properties
- ✓ **Good basic genomic resources**
  - ✓ Genetic maps, ~ 1 million expressed gene sequences
- ✓ **Large and diverse scientific community worldwide**
  - ✓ Strong breeding, industrial and academic research programs
- ✓ **A second forest tree genome for comparative plant genomics**
  - ✓ Comparative genomics with poplar, rice, Arabidopsis etc.
- ✓ **Solid wood, pulp, paper and bioenergy production in the continental USA**
  - ✓ Emphasis on cold tolerance

## Natural distribution of *Eucalyptus* species



*Eucalyptus* occur naturally from sea level to the alpine tree line, from high rainfall to semi-arid zones and from the tropics to latitudes as high as 43° south

# *Eucalyptus*: a global tree , ~20 million ha of planted forests



## Eucalypts are the world's tallest trees

*The world's tallest tree ever recorded was a fallen *Eucalyptus regnans* tree measured at **133 meters (400 feet)** at Watts River, Victoria in 1872.*

*The tree had a broken top and the entire tree was estimated to **have once been over 152 meters tall (~500 feet)**. However there is some doubt.....*



## *Eucalyptus* facts sheet

- ✓ Key feature of the majority of eucalypts is the fusion of either the petals and/or sepals to form an operculum from which the eucalypts derive their name
- ✓ **From the Greek eu, “well,” and calyptos, “covered,”**
- ✓ Eucalypt lineage extends back to the Late Cretaceous ~ 70 million years ago
- ✓ Long-lived, evergreen species belonging to
- ✓ Division: *Magnoliophyta*
  - ✓ Class: *Magnoliopsida*
  - ✓ Order: *Myrtales*
  - ✓ Family: *Myrtaceae*
  - ✓ Genus: *Eucalyptus*
- ✓ Over 740 species have been described
- ✓ Eucalypts are dominant or co-dominant in almost all vegetation types where they occur
- ✓ Keystone species for ecological studies
- ✓ Generally sclerophyllous and adapted to low nutrient soils



## *Eucalyptus* – 13 subgenera; 40 sections

Subgenus *Symphyomyrtus* - Most of the 29 commercially planted species

| Pryor & Johnson's subgenera/genera | Brooker's subgenera               | No. of species | Examples of well-known forestry species   |
|------------------------------------|-----------------------------------|----------------|---|
| <i>Angophora</i> (genus)           | <i>Angophora</i> <sup>a</sup>     | 7              |   |
| <i>Blakella</i>                    | <i>Blakella</i> <sup>a</sup>      | 15             |   |
| <i>Corymbia</i>                    | <i>Corymbia</i> <sup>a</sup>      | 67             | <i>C. torrelliana</i> , <i>C. citridora</i> , <i>C. variegata</i><br><i>C. maculata</i>   |
| <i>Eudesmia</i>                    | <i>Eudesmia</i>                   | 19             |   |
| <i>Gaubaea</i>                     | <i>Acerosa</i>                    | 1              |   |
| <i>Gaubaea</i>                     | <i>Cuboidea</i>                   | 1              |   |
| <i>Idiogenes</i>                   | <i>Idiogenes</i>                  | 1              | <i>E. doeziana</i>  |
| <i>Monocalyptus</i>                | <i>Primitiva</i>                  | 1              |   |
| <i>Monocalyptus</i>                | <i>Eucalyptus</i>                 | 110            | <i>E. regnans</i> , <i>E. delegatensis</i> , <i>E. obliqua</i> ,<br><i>E. marginata</i> , <i>E. fastigata</i>   |
| <i>Symphyomyrtus</i>               | <i>Cruciformes</i>                | 1              | <i>E. gulffoylei</i>  |
| <i>Symphyomyrtus</i>               | <i>Alveolata</i>                  | 1              | <i>E. microcorys</i>  |
| <i>Symphyomyrtus</i>               | <i>Symphyomyrtus</i>              | 474            | <i>E. camaldulensis</i> , <i>E. exserta</i> ,<br><i>E. globulus</i> , <i>E. grandis</i> , <i>E. nitens</i> ,<br><i>E. paniculata</i> , <i>E. robusta</i> , <i>E. saligna</i> ,<br><i>E. tereticornis</i> , <i>E. urophylla</i> ,<br><i>E. viminalis</i> |
| <i>Telocalyptus</i>                | <i>Minutifructus</i> <sup>b</sup> | 4              | <i>E. deglupta</i>  |

Monocotyledons

Dicotyledons

*O. sativa*

*P. trichocarpa*

*V. vinifera*

*A. thaliana*

Eurosids I

Eurosids II

Flowering plants

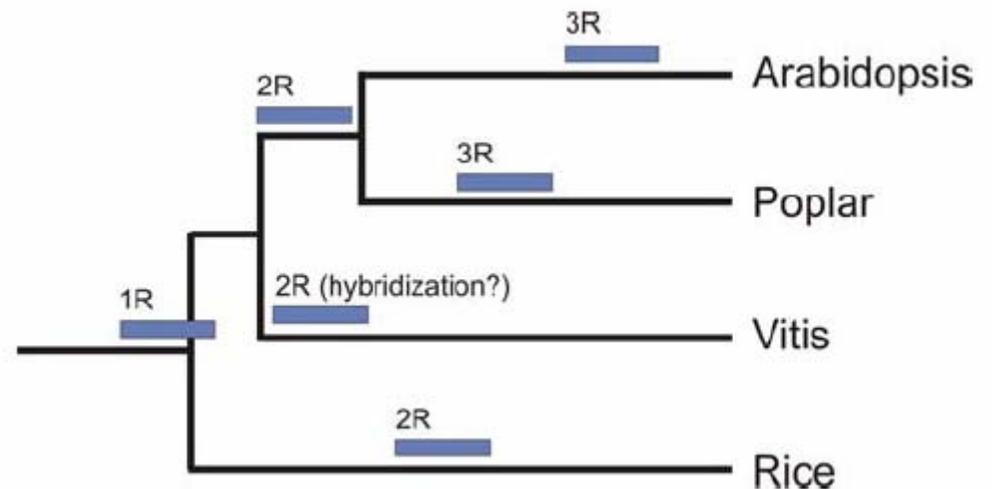
Formation of the palaeo-hexaploid genome

*Eucalyptus*: an early diverging lineage of the Rosid group of plants

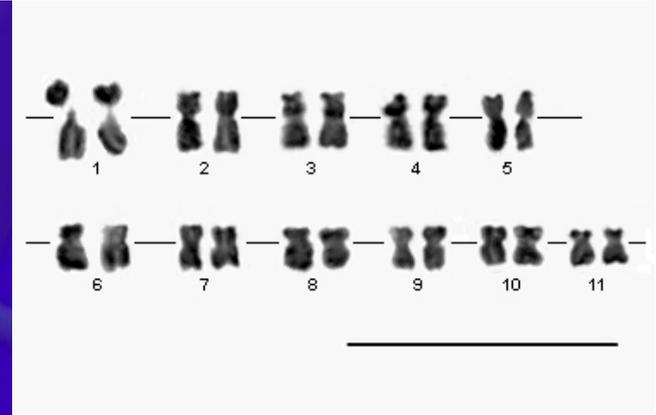
# The *Eucalyptus* genome may help understand the pattern and timing of genome duplications in the Eudicots

Velasco et al. 2007 Based on a heterozygous *Vitis* genome sequence and whole genome comparative analysis proposed a chronology of genome duplications in the different taxa

Jaillon et al. 2007 Based on an inbred *Vitis* genome sequence and whole genome comparative analysis proposed an ancestral hexaploidization in angiosperms



# Some preliminary *Eucalyptus* genome facts



- ✓ The estimated genome size of *Eucalyptus grandis* ~640 Mbp
- ✓ Diploid species with a haploid chromosome number of  $n = 11$
- ✓ Preferentially outcrossing mating system (> 90%) and recent domestication,
- ✓ Sample sequencing of 3 Mbp of sheared *E. grandis* DNA shotgun sequence
  - ✓ GC content of 40.15%
  - ✓ Gene regions composed of 38.95% introns and 45.51% exons
  - ✓ 1.4% were located in transposons, distributed in 310 interspersed repetitive elements
  - ✓ 299 were retroelements, mainly LTRs with 244 copia-like and 52 gypsy-like elements
  - ✓ Total of 1636 low complexity sequences and 987 microsatellites
  - ✓ Estimated frequency of one microsatellite every 3 kb

# Eucalypts in tropical sites accumulate biomass 4X faster than conifers in temperate and 2X in tropical sites

*Pinus taeda* (Loblolly pine)

15 m<sup>3</sup>/ha/yr (US) 30 m<sup>3</sup>/ha/yr (Brazil)



*Eucalyptus grandis*

25 m<sup>3</sup>/ha/yr (Portugal) 50-60 m<sup>3</sup>/ha/yr (Brazil)



Mechanisms by which planted forests and their wood products can displace the consumption of fossil fuels and make a contribution to reducing net CO<sub>2</sub> emission

- ✓ Wood products may replace fossil fuels such as mineral coal that traditionally have been used for energy production
- ✓ Wood products may replace materials that require high energy levels to be manufactured

# Drink the best and drive the rest

Brazil's sugar-cane ethanol industry is the world's best and able to get better, says **Emma Marris**.

- ✓ Today 11% of Brazilian cars are bi-fuel
- ✓ Expected 32% by 2010



- ✓ CO<sub>2</sub> sequestration of Eucalyptus: 10 tons of CO<sub>2</sub>/hectare/year from planting to harvesting
- ✓ Up to 14 CO<sub>2</sub>/hectare/year in fast-growing tropical plantations
- ✓ Positive net carbon balance even when computing production of CO<sub>2</sub> used for energy from charcoal or as pulp and paper
- ✓ Present estimates of sequestration costs are \$100 to \$300/ton of CO<sub>2</sub> emissions avoided



# Forest based industrial products in Brazil

| Product         | UN<br>1000     | Production | Domestic<br>Consumption | Source                 |                       |
|-----------------|----------------|------------|-------------------------|------------------------|-----------------------|
|                 |                |            |                         | Planted <sup>(1)</sup> | Native <sup>(2)</sup> |
| Pulp            | t              | 8.020      | 5.020                   | 100%                   | -                     |
| Paper           | t              | 7.800      | 6.879                   | 100%                   | -                     |
| Charcoal        | mdc            | 26.200     | 26.200                  | 68%                    | 32%                   |
| Sawnwood        | m <sup>3</sup> | 22.300     | 20.000                  | 35%                    | 65%                   |
| Plywood         | m <sup>3</sup> | 2.600      | 900                     | 60%                    | 40%                   |
| MDF             | m <sup>3</sup> | 845        | 716                     | 100%                   | -                     |
| Particle Board  | m <sup>3</sup> | 1.800      | 1.800                   | 100%                   | -                     |
| OSB             | m <sup>3</sup> | 90         | 80                      | 100%                   | -                     |
| Fiberboard      | m <sup>3</sup> | 507        | 295                     | 100%                   | -                     |
| EGP             | m <sup>3</sup> | 285        | 220                     | 100%                   | -                     |
| Mouldings       | m <sup>3</sup> | 490        | 50                      | 40%                    | 60%                   |
| Doors           | um             | 6.300      | 4.700                   | 70%                    | 30%                   |
| Floors          | M <sup>2</sup> | 22,50      | 15,20                   | 50%                    | 50%                   |
| Blocks / Blanks | m <sup>3</sup> | 430        | 360                     | 100%                   | -                     |

(1) Industrial roundwood: 110 millions m<sup>3</sup>

(2) Industrial roundwood: 66 millions m<sup>3</sup>

# *Eucalyptus* industrial forests in Brazil

- ✓ Fast growing eucalypt forests
  - ✓ ~4 million hectares, 40% of the intensively world planted *Eucalyptus* area
  - ✓ Main supplier of woody biomass with specific wood properties for several industrial activities ( pulp and paper, steel, solid wood products)
  - ✓ Average productivities of 45 m<sup>3</sup>/ha/yr
- ✓ Main competitiveness factors:
  - ✓ Genetic material: adaptability and variability
  - ✓ Advanced breeding programs for wood quality
  - ✓ Clonal forests – high selection intensity
  - ✓ Silvicultural practices
  - ✓ Growing critical mass in R&D both in industries and universities





Edição Abril  
edição 2023 - ano 51 - nº 11  
20 de março de 2023

# veja

www.veja.com.br

## ESPECIAL AMAZÔNIA

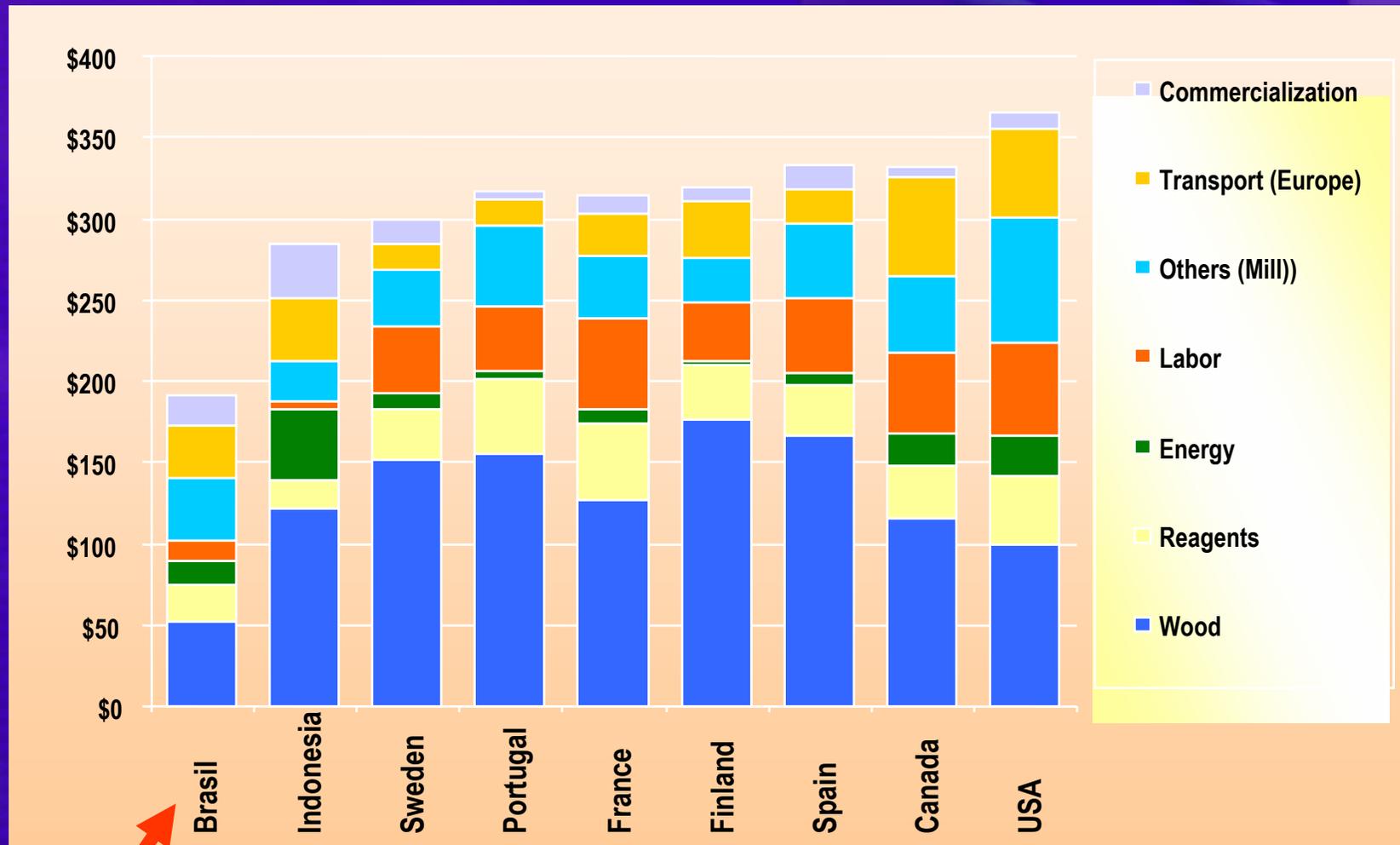


**A VERDADE SOBRE AS  
QUEIMADAS E O RITMO  
DO DESMATAMENTO**

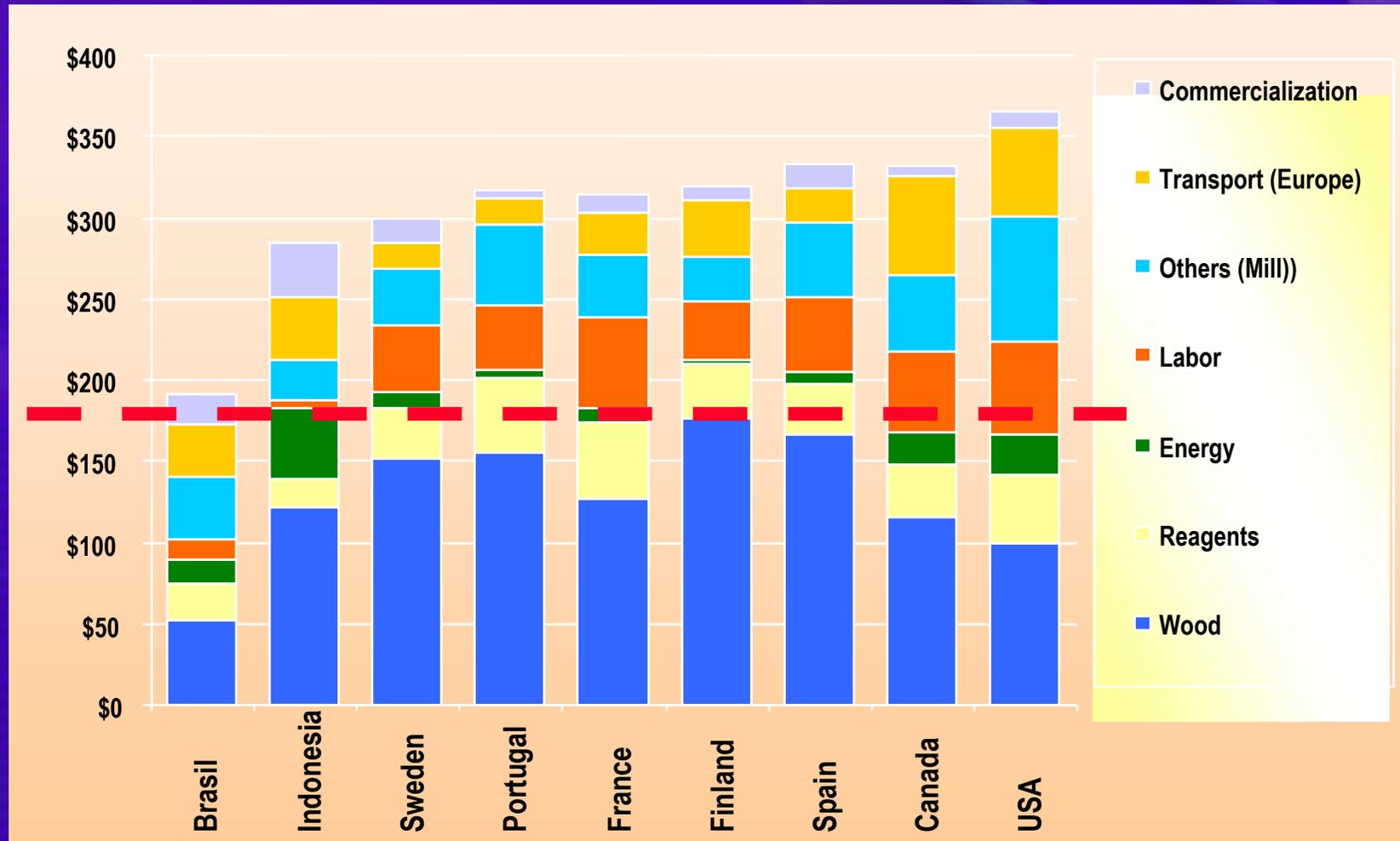
**POR QUE A EXPLORAÇÃO  
ECONÔMICA NÃO PRECISA  
SER CRIMINOSA**

*Eucalyptus*  
substitution forests  
to provide energy  
and wood products  
that would  
otherwise come  
from the native  
tropical forest

# Production and commercialization costs of short fiber pulp in several countries in 2002 (US\$/ton)



# Production and commercialization costs of short fiber pulp in several countries in 2002 (US\$/ton)





- ✓ Eucalyptus can be managed for high quality wood
- ✓ Certified, renewable wood
- ✓ Substitution to native tropical species
- ✓ Great way to fix carbon

## “Green” steel with *Eucalyptus* charcoal



### **MINERAL COAL X EUCALYPT CHARCOAL IN STEEL PRODUCTION**

- ✓ **Mineral coal: generates 1.65 ton of CO<sub>2</sub>, sequesters 1.54 ton of O<sub>2</sub> and generates 7 kg of SO<sub>2</sub> per ton of steel**
- ✓ **Eucalypt charcoal: sequesters 16.3 tons of CO<sub>2</sub> and regenerates 11.9 tons of O<sub>2</sub> from plantation through harvesting to steel production**

# Breeding technologies for *Eucalyptus*



*Controlled crosses*



*Progeny trials*



*Selection of elite trees*



*Molecular markers in support of operational breeding and production forestry*



*High productivity clonal forests*

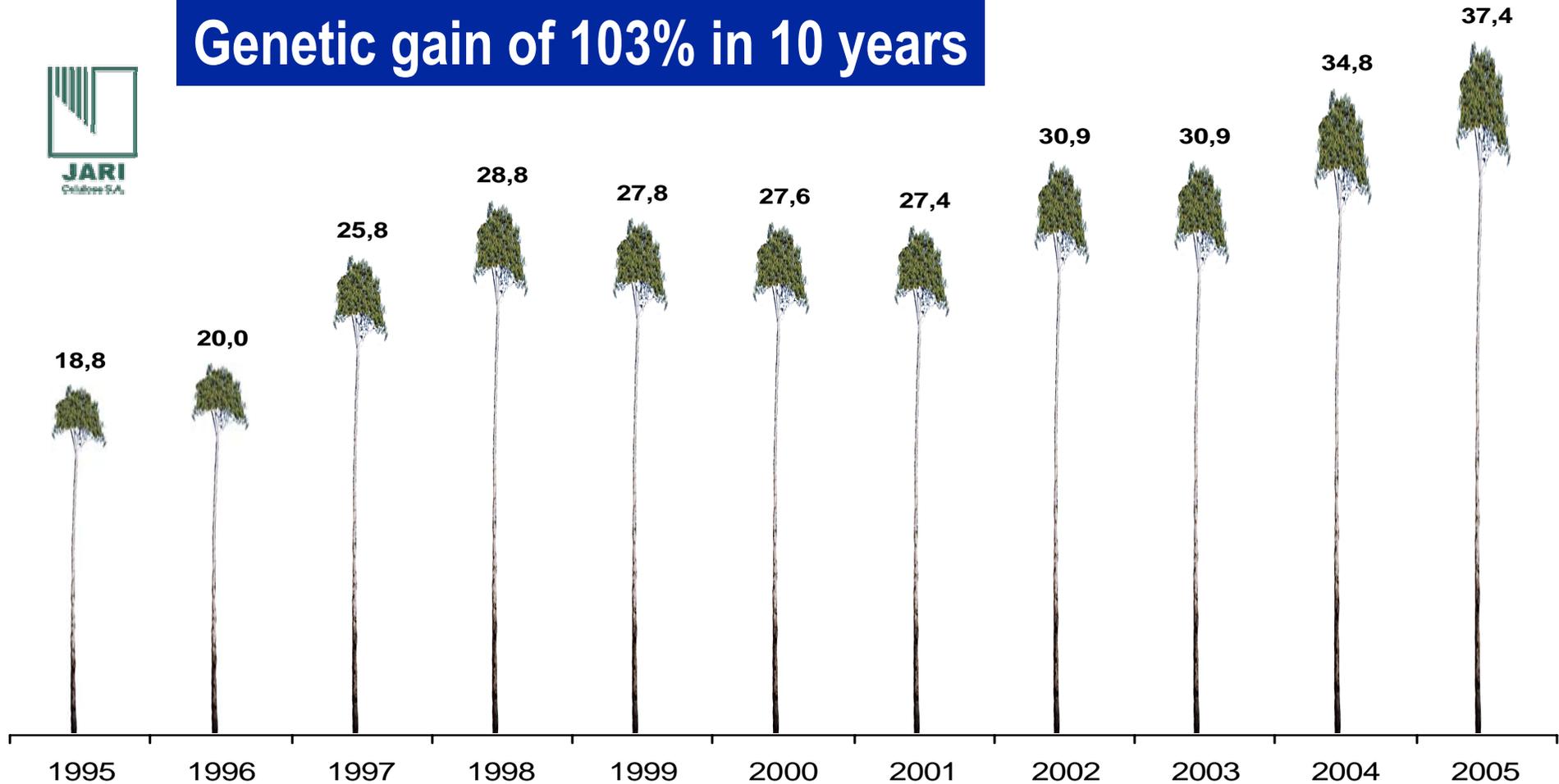


*Cloning by microcutting*

# Breeding & clonal propagation have been very successful in increasing growth rates

Example (Jari Celulose Co. on equatorial sites)

**Genetic gain of 103% in 10 years**



# The tree is the real pulp mill

- ✓ Increased forest productivity:  $\text{m}^3$  wood/hectare/year
- ✓ Reduction in wood specific consumption for pulp production:  $\text{m}^3$  of wood/ton of cellulose pulp
- ✓ Combination of large wood volume per hectare with low specific consumption of wood per ton of pulp: reduced industrial costs
- ✓ *Breeding is key to competitiveness*
- ✓ *Genomics will be one more tool available to the breeder*



# Economic value of volume growth x pulp yield in *Eucalyptus*

- ✓ Wood density and pulp yield are highly correlated
- ✓ Wood density and pulp yield are 3X more important than volume growth in the final cost of cellulose pulp
- ✓ Gains of 10% in wood specific consumption (ex. reducing from 3.6 to 3.24 m<sup>3</sup>/ton pulp) corresponds to a gain of 30% in volume growth (ex. going from 35 to 50 m<sup>3</sup>/ha/yr)
- ✓ Gains are made along the whole production chain (harvesting, transportation, yearly mill efficiency etc.) as these make up 2/3 of the final cost of pulp production

(SOURCE: Borralho et al. 1993)

## Interspecific variation for wood quality traits in *Eucalyptus*

| <b>ESPÉCIE</b>         | <b>AMI<br/>m<sup>3</sup>/ha/<br/>ano</b> | <b>PULP<br/>YIELD<br/>(%)</b> | <b>DENS<br/>(g/cm<sup>3</sup>)</b> | <b>S5<br/>(%)</b> | <b>SPEC.<br/>CONS.<br/>(m<sup>3</sup>/ton)</b> | <b>LIGNIN<br/>(%)</b> |
|------------------------|--|-------------------------------|------------------------------------|-------------------|--|-----------------------|
| <i>E. grandis</i>      | 42                                       | 50                            | 0,42                               | 10                | 4,7  | 25,5                  |
| <i>E. tereticornis</i> | 25                                       | 46                            | 0,56                               | 9                 | 3,8  | 30                    |
| <i>E. saligna</i>      | 36                                       | 49                            | 0,46                               | 9                 | 4,4  | 26                    |
| <i>E. urophylla</i>    | 30                                       | 49                            | 0,48                               | 11                | 4,2  | 28                    |
| <i>E. globulus</i>     | 25                                       | 54                            | 0,55                               | 15                | 3,3  | 22                    |

# *Eucalyptus globulus:* *Paradigm of wood quality for pulp*

- ✓ Best combination of wood properties for pulp and paper
- ✓ Requires ~25% less wood per ton of cellulose
  - ✓ *E. grandis*: 3.89 m<sup>3</sup>/ton of cellulose
  - ✓ *E. globulus*: 2.98 m<sup>3</sup>/ton of cellulose
- ✓ Longer and thicker fiber
- ✓ Larger holocellulose content
- ✓ Better fiber flexibility
- ✓ Basic density around 550 kg/m<sup>3</sup>
- ✓ Requires less energy in the industrial process
- ✓ Better fiber preservation during the pulping process

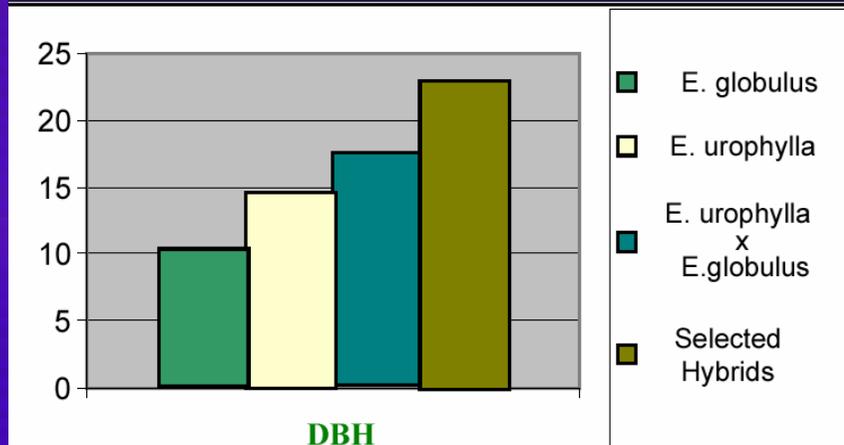


# Wood properties and adaptation to the tropics

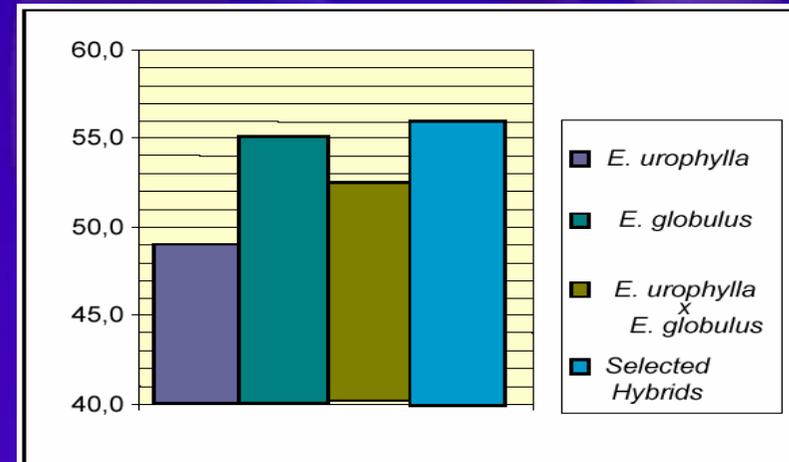
- ✓ *E. globulus* is adapted to latitudes below 33°S and clearly unadapted to less than 28°S
- ✓ Evidences exist that wood anatomy co-evolves with latitude and tropicality in response to high demands of transpiration and possibly disease resistance (e.g. Jansen et al. 2004; Noshiro and Baas 2000; Swenson and Enquist 2007),
- ✓ This variation seems to be related mostly to vessel structure and less to fibers, the most relevant elements to wood quality
- ✓ Good perspectives of introgressing the exceptional wood traits of *E. globulus* into tropical Eucs



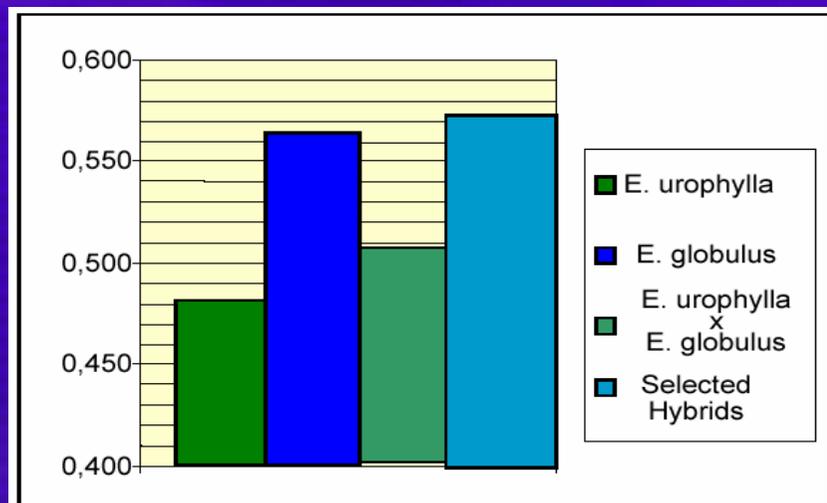
# Introgressing *E. globulus* genes into tropical *Eucalyptus*: gene action for wood traits is mostly additive



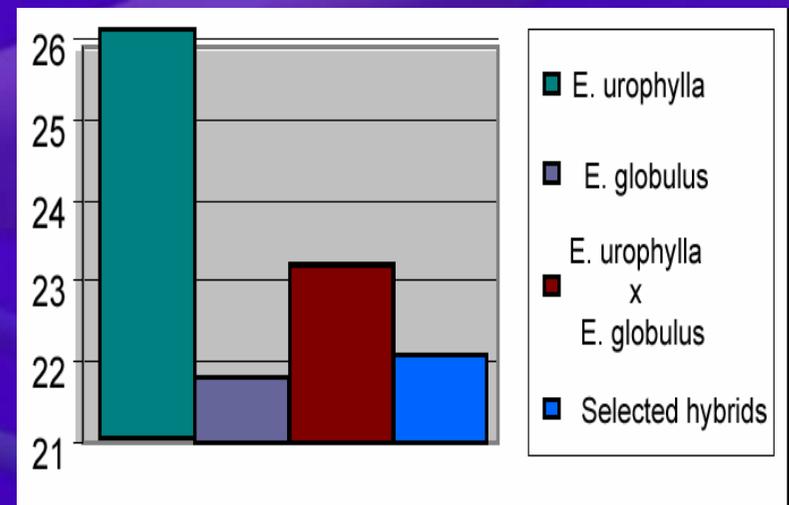
*Volume growth*



*Cellulose yield*



*Wood specific gravity*



*Total lignin content*

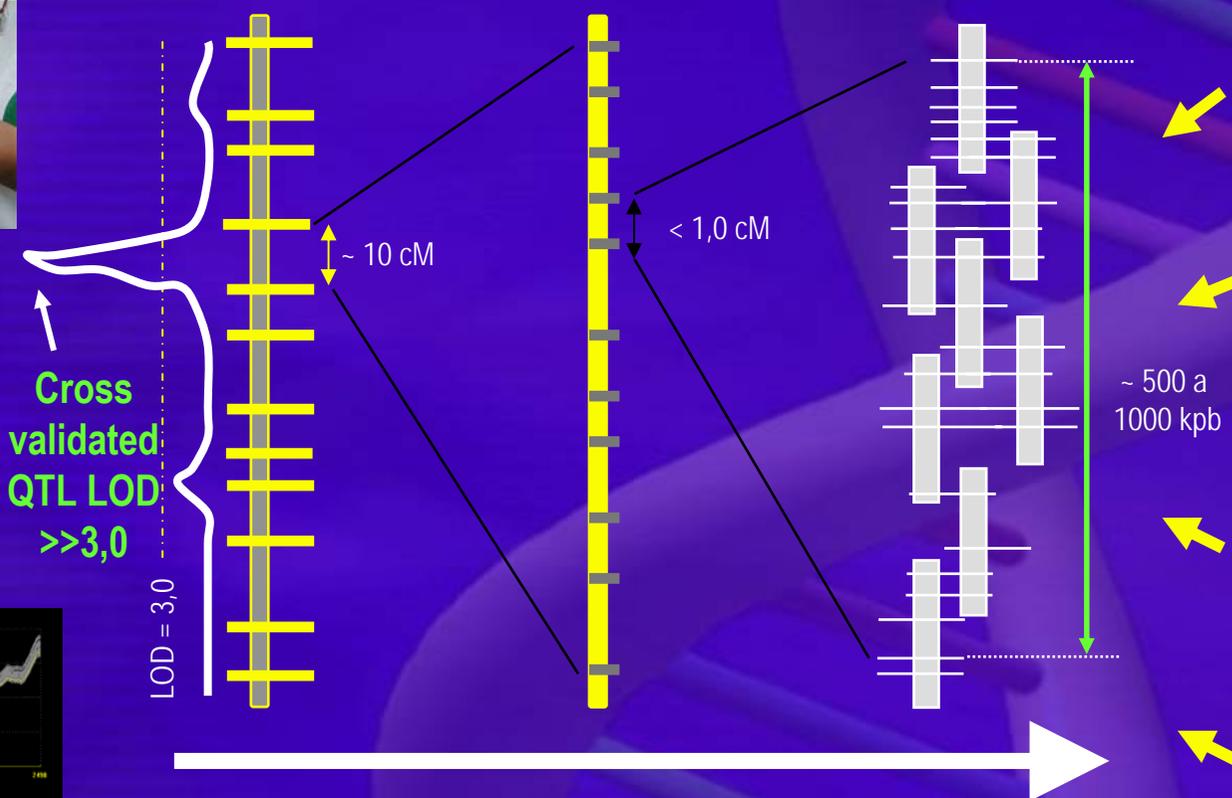
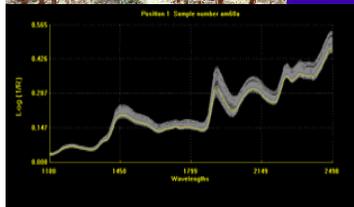
Large scale phenotyping in segregating populations

Genetic map of transferable markers and QTLs

High resolution genetic map (SNPs)

Minimum tiling path physical map (BAC *Fingerprinting*)

Identification and validation of target genes



Genetic mapping of candidate genes

Association mapping (*whole genome scan*, candidate genes)

Full genome sequence assembly

Functional tests by transformation (complementation, superexpression, interference)

Forward genomics approach from phenotypes to genes

# Participating institutions in the GENOLYPTUS project (Brazilian Network of Eucalyptus Genome Research)



# Mating design used to generate segregating families to understand wood properties and adaptation and map QTLs

|   | <i>E. grandis</i><br>Atherton<br>Aracruz | <i>E. urophylla</i><br>Timor<br>Cenibra | <i>E. globulus</i><br>K-Riocell      | <i>E. dunni</i><br>K-Riocell        | <i>E. camaldulensis</i><br>V-Mannesmann | <i>E. uro. x E. glob.</i><br>K-Riocell  |
|---|--|---|--------------------------------------|-------------------------------------|---|---|
| <i>E. grandis</i><br>Coffs Harb. VCP    | G1 x G2<br>(est. VCP x pól. AR)          | G1 x U2<br>(est. VCP x pól. CE)         | G1 x GL2<br>(est. VCP x pól. K-R)    | G1 x D2<br>(est. VCP x pól. K-R)    |   | G1 x (UxGL)<br>(est. VCP x pól. K-R)    |
| <i>E. urophylla</i><br>(Flores) IP      | U1 x G2<br>(est. IP x pól. AR)           | U1 x U2<br>(est. IP x pól. CE)          | U1 x GL2<br>(est. IP x pól. K-R)     | U1 x D2<br>(est. IP x pól. K-R)     | U1 x C2<br>(est. IP x pól. V-M)         | U1 x (UxGL)<br>(est. IP x pól. K-R)     |
| <i>E. globulus</i><br>K-Riocell         | G2 x GL1<br>(est. AR x pól. K-R)         | U2 x GL1<br>(est. CE x pól. K-R)        |                                      |                                     | C2 x GL1<br>(est. V-M x pól. K-R)       |   |
| <i>E. dunni</i><br>Rigesa               | D1 x G2<br>(est. RG x pól. AR)           | D1 x U2<br>(est. RG x pól. CE)          | D1 x GL2<br>(est. RG x pól. K-R)     | D1 x D2<br>(est. RG x pól. K-R)     |   | D1 x (UxGL)<br>(est. RG x pól. K-R)     |
| <i>E. camaldulensis</i><br>V-Mannesmann | G2 x C1<br>(est. AR x pól. V-M)          | U2 x C1<br>(est. CE x pól. V-M)         | C1 x GL2<br>(est. V-M x pól. K-R)    | C1 x D2<br>(est. V-M x pól. K-R)    | C1 x C2<br>(est. V-M x pól. V-M)        | C1 x (UxGL)<br>(est. V-M x pól. K-R)    |
| <i>E. gran. x E. dunni</i><br>K-Riocell | (GxD) x G2<br>(est. K-R x pól. AR)       | (GxD) x U2<br>(est. K-R x pól. CE)      | (GxD) x GL2<br>(est. K-R x pól. K-R) | (GxD) x D2<br>(est. K-R x pól. K-R) | (GxD) x C2<br>(est. K-R x pól. V-M)     | (GxD) x (UxGL)<br>(est. K-R x pól. K-R) |

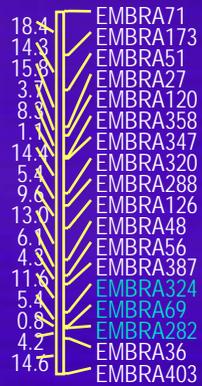


Crosses made with elite parents from different companies

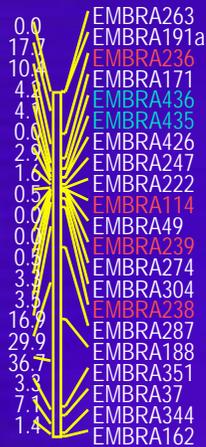


*(E. grandis x E. dunnii) X (E. urophylla x E. globulus)* hybrid  
Age 36 months

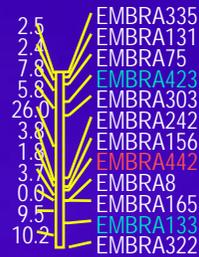
# Reference genetic map based on ~300 microsatellites and growing...



Group 1



Group 2



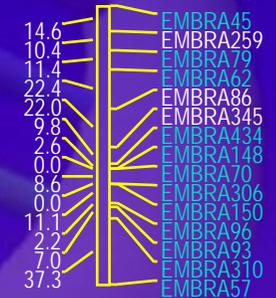
Group 3



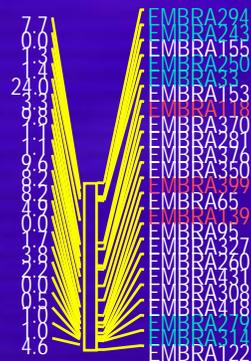
Group 4



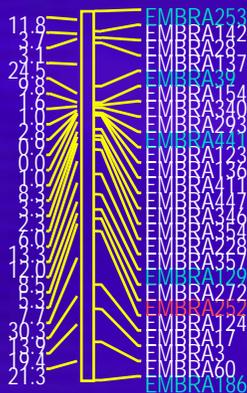
Group 5



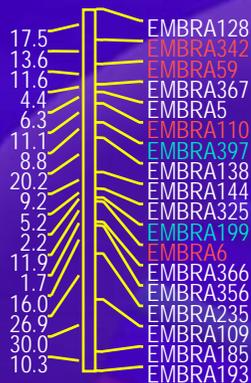
Group 6



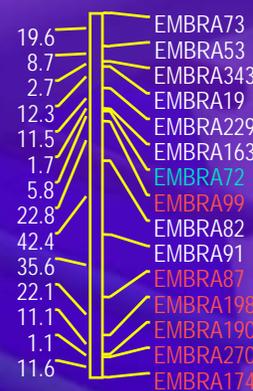
Group 7



Group 8



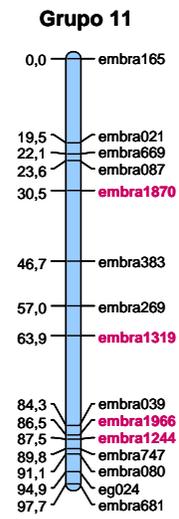
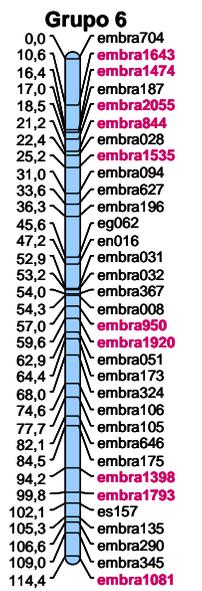
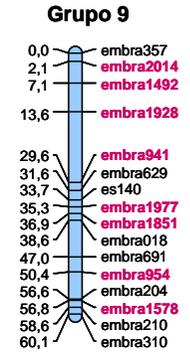
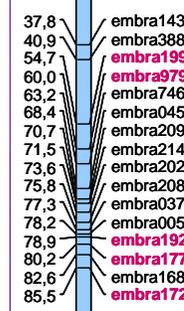
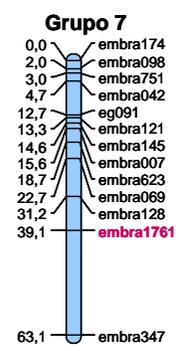
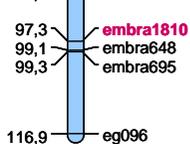
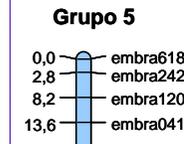
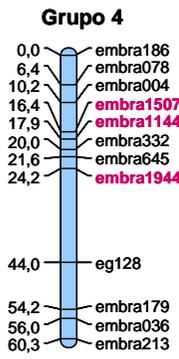
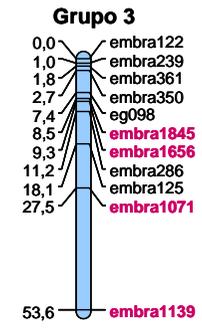
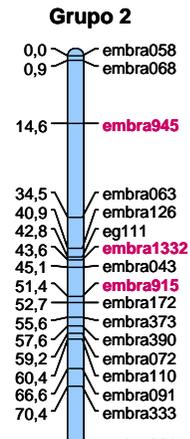
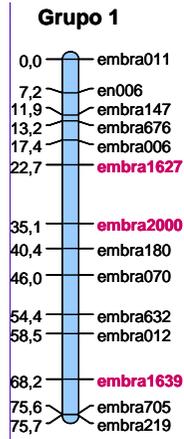
Group 9



Group 10



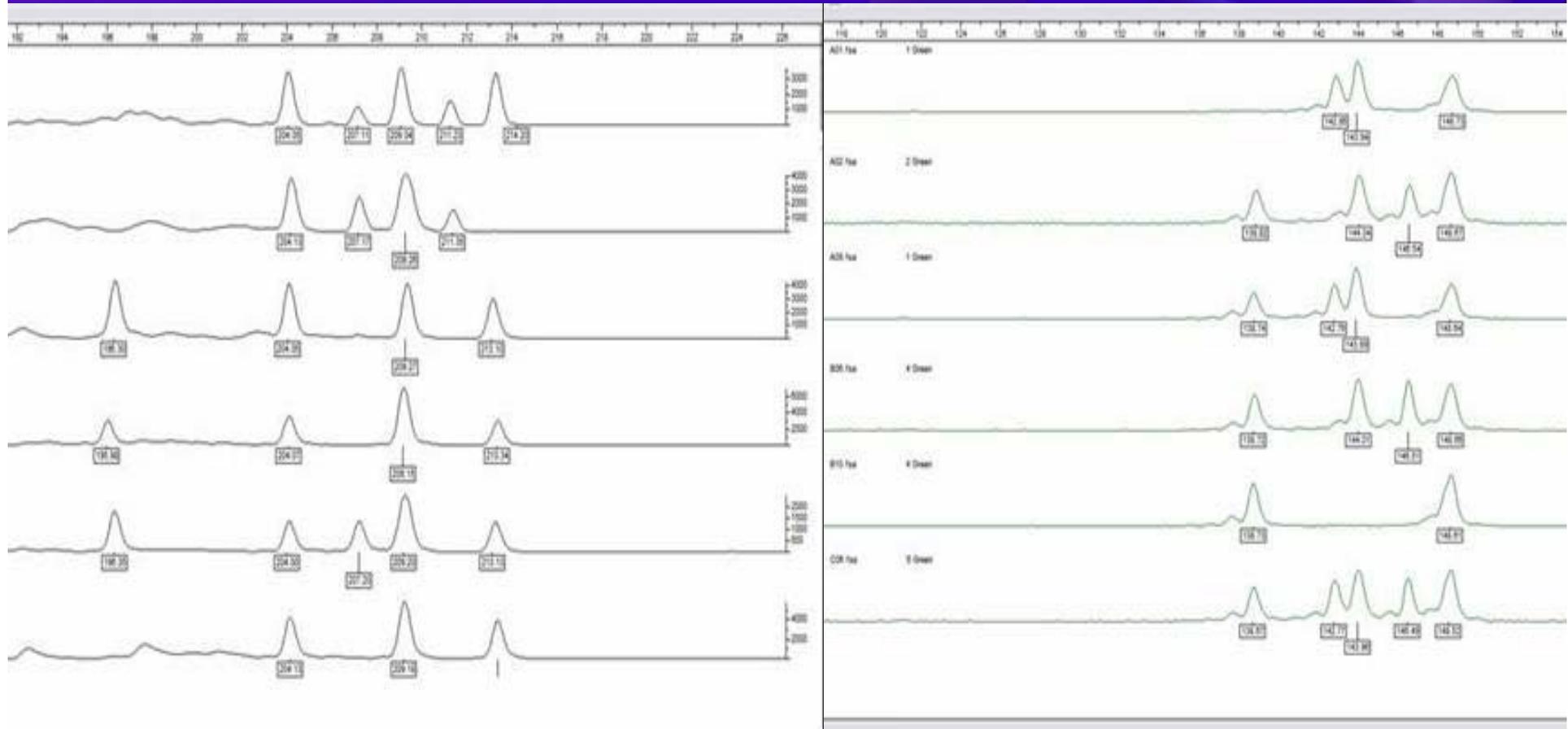
Group 11



| LOCO SSR  | GENE COM MAIOR SIMILARIDADE NA ANALISE DE BLAST  |
|-----------|--|
| EMBRA915  | DNA binding protein [Elaeis oleifera]  |
| EMBRA941  | Skp1 [Medicago sativa]   |
| EMBRA954  | HB2 homeodomain protein [Populus tremula x Populus tremuloides]                                |
| EMBRA1081 | zinc finger (C3HC4-type RING finger) protein family [Arabidopsis thaliana]                     |
| EMBRA1244 | LIM domain protein PLIM1 [Nicotiana tabacum]   |
| EMBRA1319 | KH domain/zinc finger protein [Arabidopsis thaliana]   |
| EMBRA1398 | NOI protein, nitrate-induced [Arabidopsis thaliana]  |
| EMBRA1428 | wound induced protein kinase [Nicotiana tabacum]   |
| EMBRA1474 | senescence-associated protein -related [Arabidopsis thaliana]                                  |
| EMBRA1627 | glycosyltransferase family 43 [Arabidopsis thaliana]   |
| EMBRA1639 | PIN1-like auxin transport protein [Populus tremula x Populus tremuloides]                      |
| EMBRA1722 | transfactor-related protein [Arabidopsis thaliana]   |
| EMBRA1770 | RHO GDP-dissociation inhibitor 1 -related [Arabidopsis thaliana]                               |
| EMBRA1829 | histone H2A, putative [Arabidopsis thaliana]   |
| EMBRA1870 | auxin-induced (indole-3-acetic acid induced) protein, putative (SAUR_e) [Arabidopsis thaliana] |
| EMBRA1928 | Heat Shock factor protein HSF24 [Lycopersicon peruvianum]                                      |
| EMBRA1944 | rac family GTP-binding protein (ARAC9) [Arabidopsis thaliana]                                  |
| EMBRA1990 | calcium-binding EF-hand family protein [Arabidopsis thaliana]                                  |

*EST derived microsatellite mapping  
EMBRA 800 to 1999  
~130 mapped to date*

*Microsatellite duplicated loci are observed at a rate of ~20% suggesting significant and possibly recent duplication in some genome regions*



## Existing EST resources from the EUCAGEN community

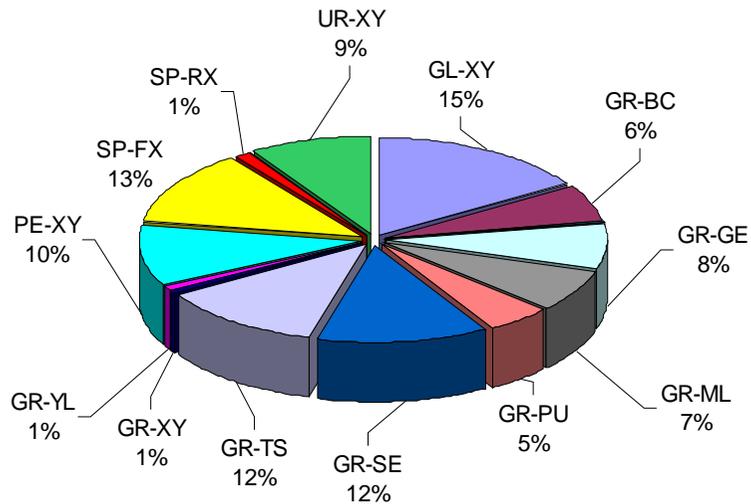
### EST in Genbank

- ✓ 24,698 ESTs as of Dec 31, 2007
- ✓ French, Chilean and US projects

### GENOLYPTUS EST database (Brazil)

- ✓ 4 species ; 11 libraries
- ✓ 124,851 reads
- ✓ 21,442 tentative consensi
- ✓ To be published in 2007

### Genolyptus libraries



### 454 Sequencing

*Novaes et al. 2007 University of Florida*

*Pool of 21 E. grandis trees*

*1,024,251 reads*

*148.4 Mbp*

*29,000 tentative consensi*

*23,742 SNPs*

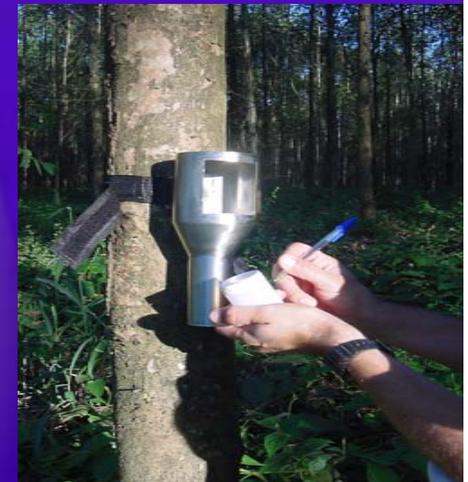
➤ *Submitted manuscript*

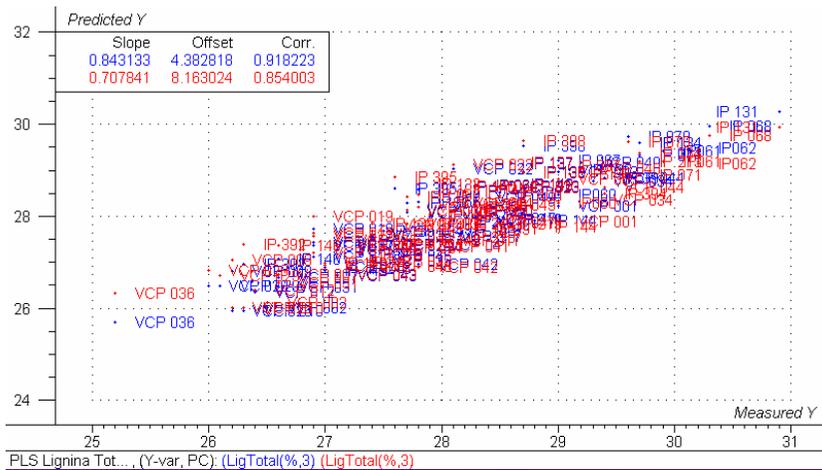
### ARBORGEN EST database (US)

- ✓ *Several species*
- ✓ *~ 400,000 ESTs*
- ✓ *Private but to be released by an MTA*

# Sample collection for phenotypic analysis

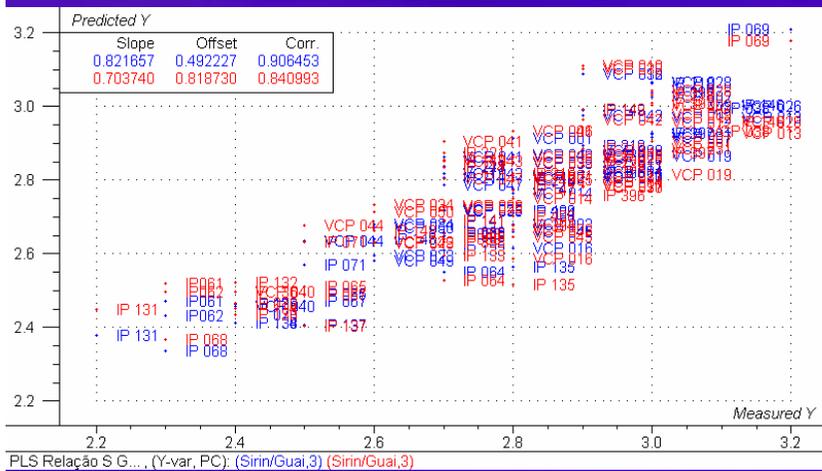
*~35.000 trees sampled in 2006*



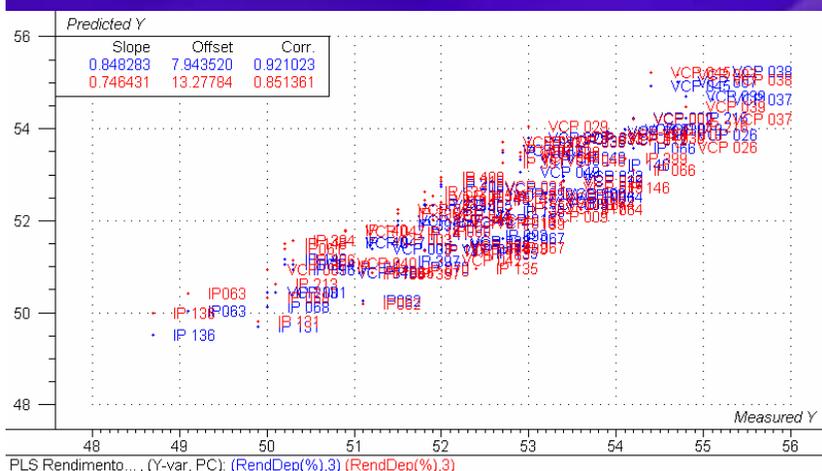


Calibration curves for some main traits

Total lignin



Syringyl/ Guaiacyl lignin ratio

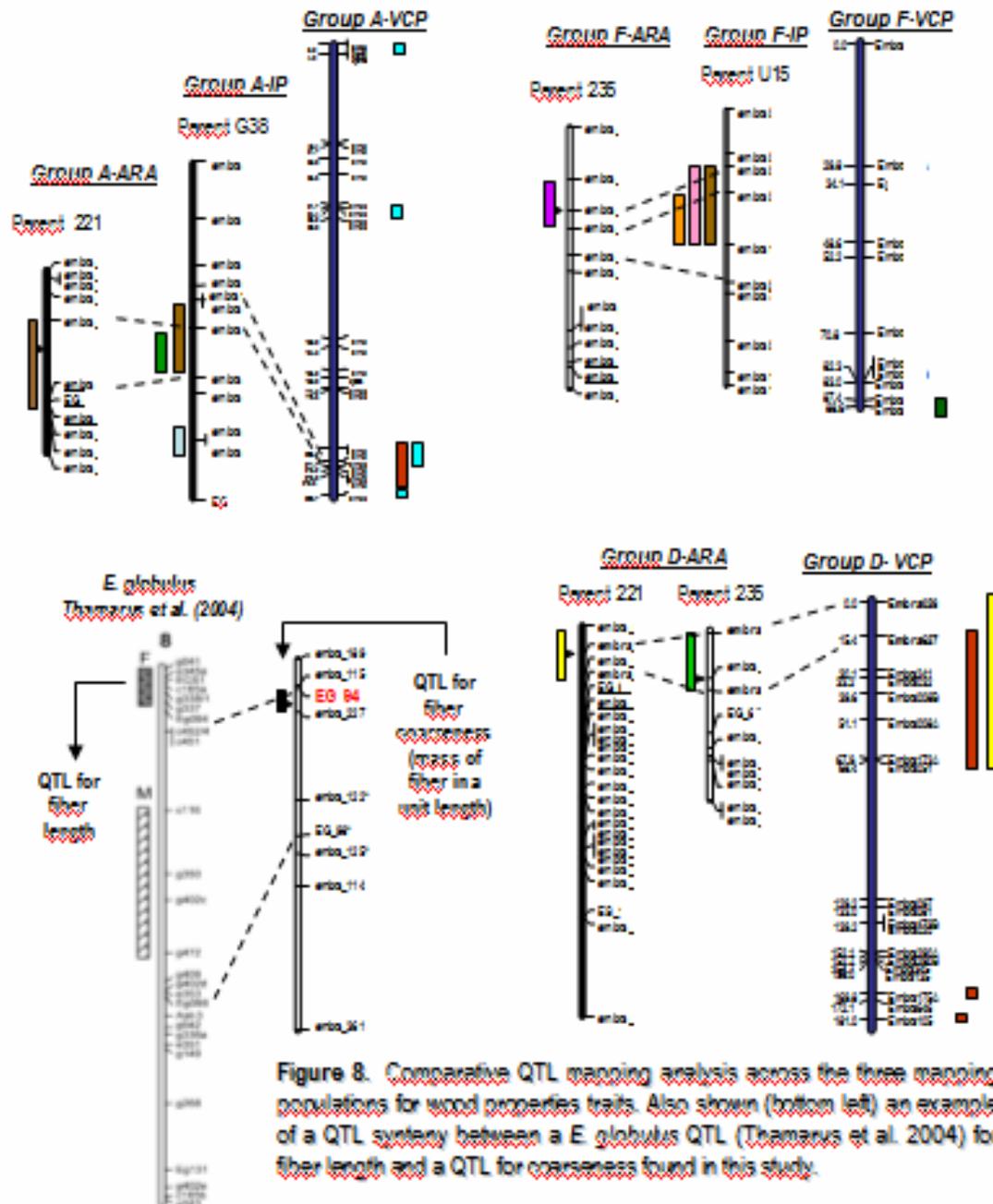


Pulp yield

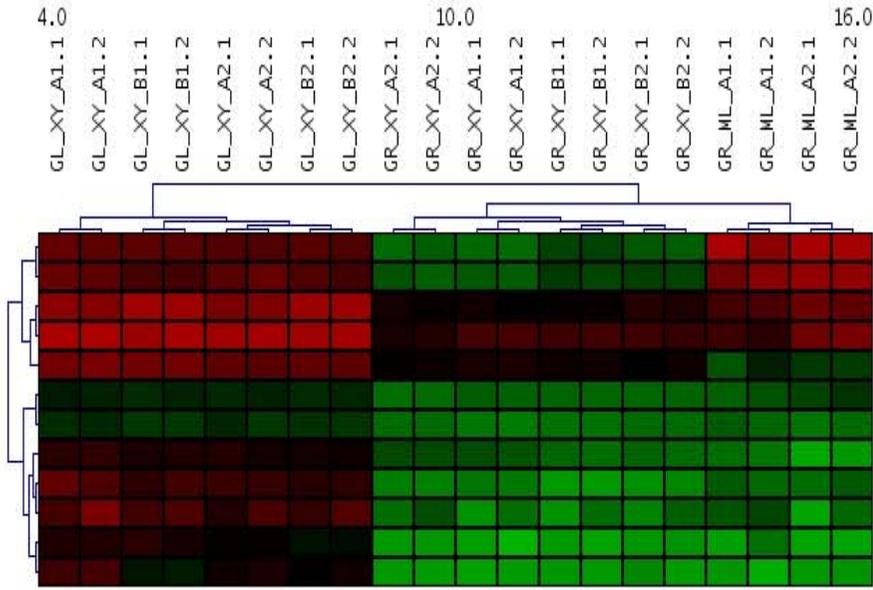
# Comparative QTL mapping for wood properties

*Data from three unrelated families*

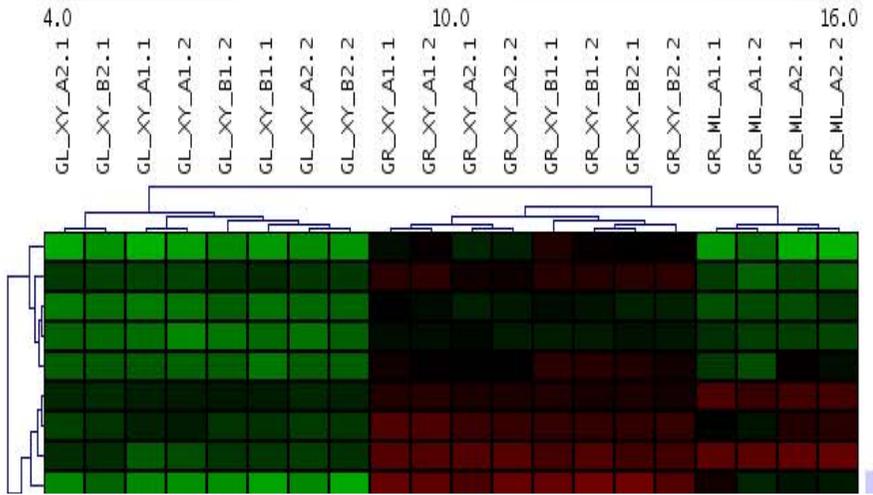
*Currently these QTL targets are being validated in five further unrelated families involving several different parents and species*



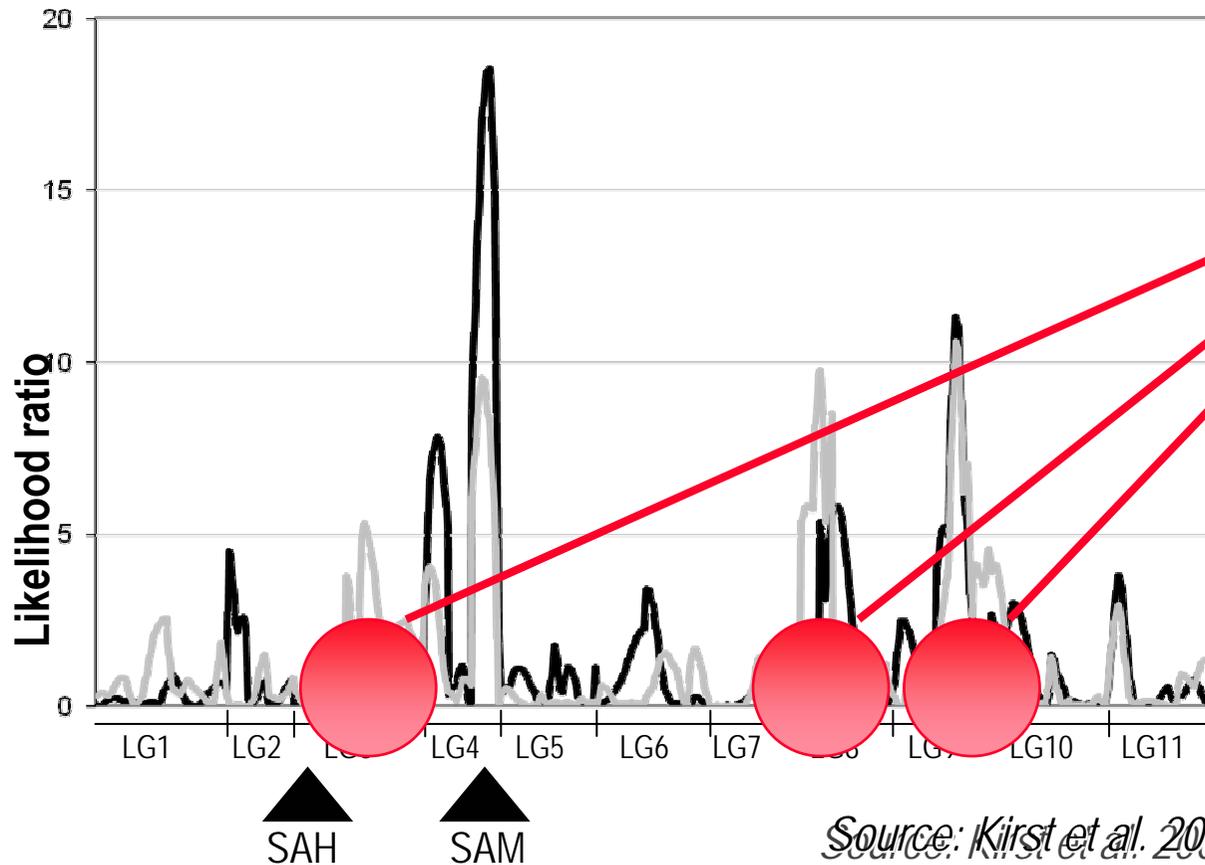
# Interspecific variation: *E. globulus* x *E. grandis*



no hit  
 TIR-NBS disease resistance-like protein [(Populus tomentosa x P. bolleana) x P. tomentosa]  
 unknown protein [Arabidopsis thaliana] gb|AAF79910.1 Contains similarity to SCUTL1  
 putative aquaporin PIP1-2 [Vitis berlandieri x Vitis rupestris]  
 no hit  
 protein binding / ubiquitin-protein ligase/ zinc ion binding [Arabidopsis thaliana]  
 unknown protein [Arabidopsis thaliana] gb|AAM62838.1 Yippee-like protein [Arabidopsis thaliana]  
 unknown protein [Arabidopsis thaliana] gb|AAM97053.1 unknown protein [Arabidopsis thaliana]  
 Protein of unknown function DUF567 [Medicago truncatula]  
 no hit  
 no hit  
 SKP1 component [Medicago truncatula]



no hit  
 no hit  
 unknown protein [Arabidopsis thaliana] gb|AAD10685.1 Hypothetical protein [Arabidopsis thaliana]  
 alpha-mannosidase [Arabidopsis thaliana] gb|AAM47314.1 AT3g26720/MLJ15\_12 [Arabidopsis thaliana]  
 Mal d 1-like [Malus x domestica]  
 nucleotide binding [Arabidopsis thaliana] gb|AAL47352.1 WD-repeat protein-like [Arabidopsis thaliana]  
 Pyridoxal phosphate biosynthetic protein pdxA [Rhizobium sp. NGR234]  
 Similar to gb|X84260 POS5 gene product from Saccharomyces cerevisiae. EST gb|W4387  
 zinc transporter [Eucalyptus grandis]



*Hot spots for  
gene mining*

*To what extent is variation in gene expression associated with variation in complex traits at the phenotypic level?*

QTL mapping experiments are now being expanded to expression QTL mapping experiments to provide a global analysis of *cis* and *trans* eQTLs and identify candidate genes for complex traits

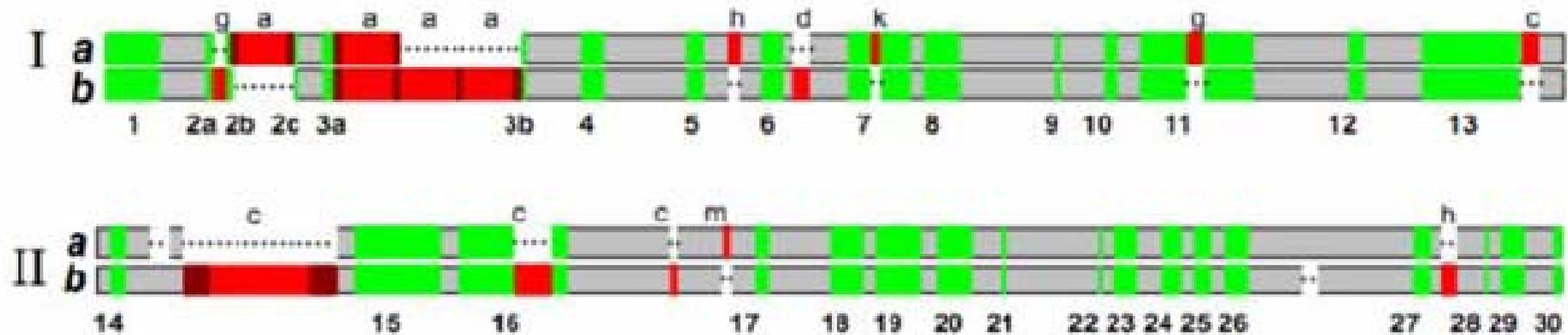


## Application of QTL information for marker assisted selection

*Hybrid vigor at the  
family level*

*Hybrid vigor at the  
individual level*

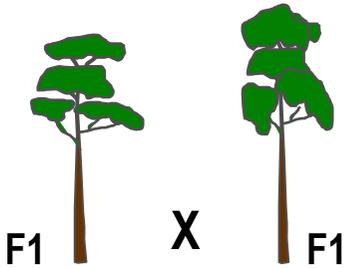




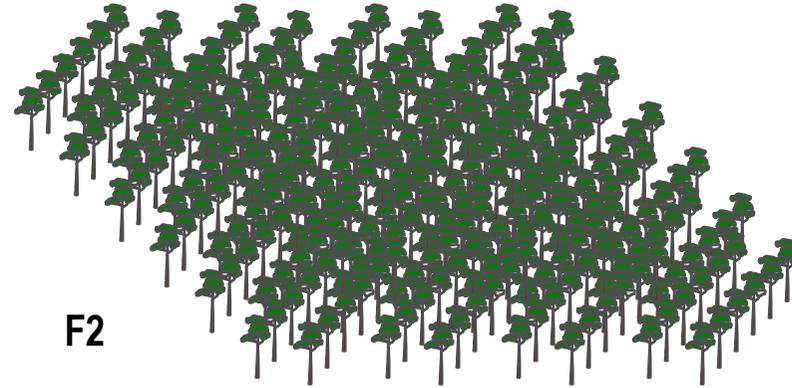
**Gene or pseudo-gene**      **Transposable elements**      **Gaps**

Velasco et al. 2007- structural variation between haplotypes in the *Vitis* genome is associated with transposable elements and gaps

- ✓ Differences in repetitive DNA genome content and distribution supply different genome environments that in turn affect specificity and temporal regulation of gene expression
- ✓ Such differences have been proposed as the explanation to heterotic complementation in maize
- ✓ Structural variation both within and between *Eucalyptus* species should be abundant in and might help understand the molecular basis of hybrid superiority and hybrid segregation



**MAS STAGE**



*Mating between plus hybrid trees to maximize segregation for several traits in the outbred F2*

**F2**

*Deployment of a large number (> 1000) of F2 progeny individuals to maximize probability of generating a recombinant individual with a superior multiple QTL allele content*

**QTL MAPPING STAGE**

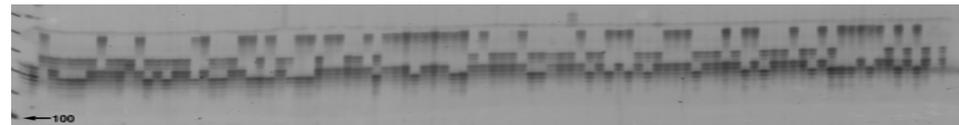
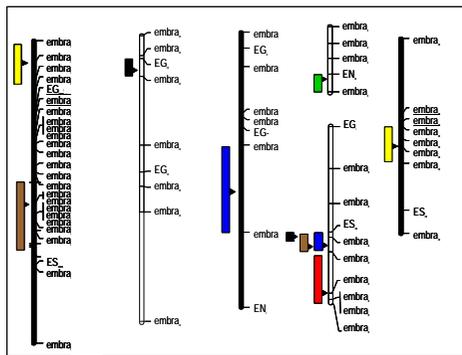


**F2**



*QTL mapping for wood properties traits such as lignin, fiber and wood density*

QTL mapping information to be used in MAS



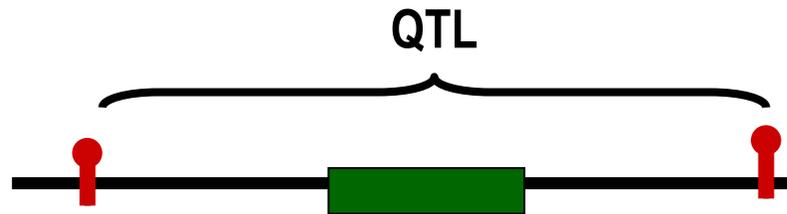
*Genotyping > 1000 seedlings with a small (~ 6 to 10) set of flanking markers for a targeted number of QTL for wood quality traits*



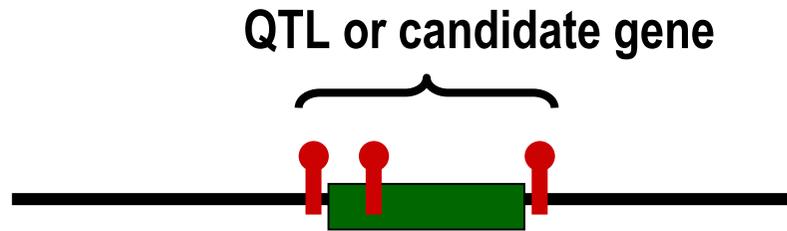
*Early marker assisted selection. Selection intensity is increased by MAS for late expressing traits but number of trees commonly deployed in progeny (~100) test is kept the same, thus allowing large variation to select for other traits such as volume growth, form and branching habit.*



Resolution  
+  
-



**LE markers – Linkage Equilibrium**  
Ex. Microsatellite markers flanking a QTL mapped in a high LD pedigree - Centimorgan resolution,  $\sim 10^5$  a  $10^7$  bp

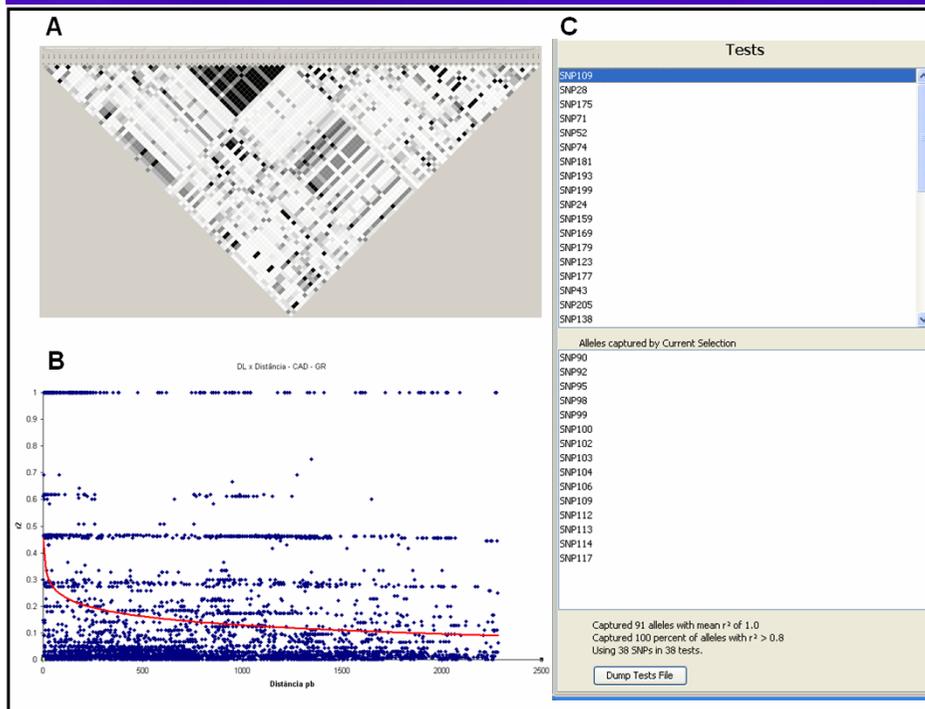
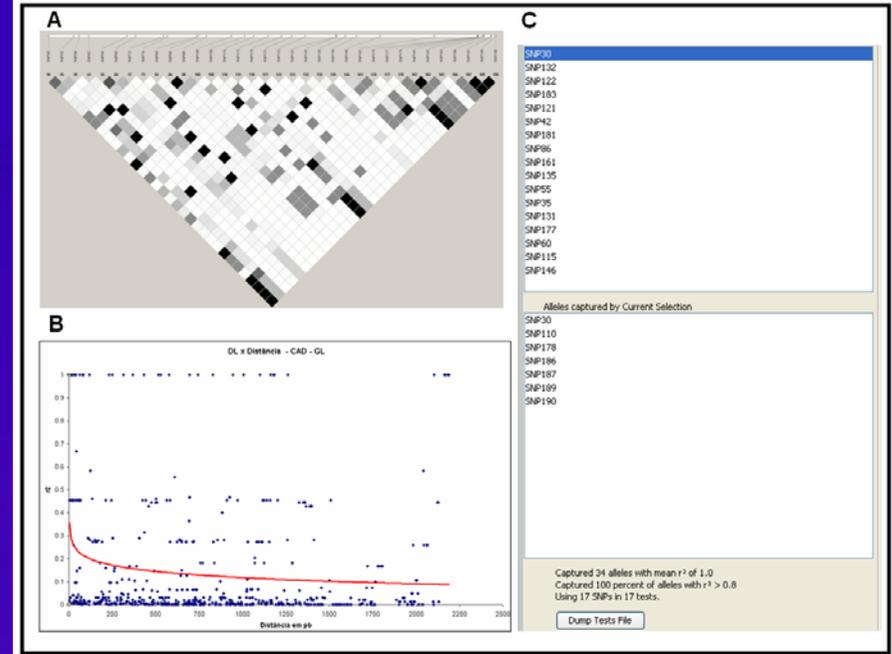
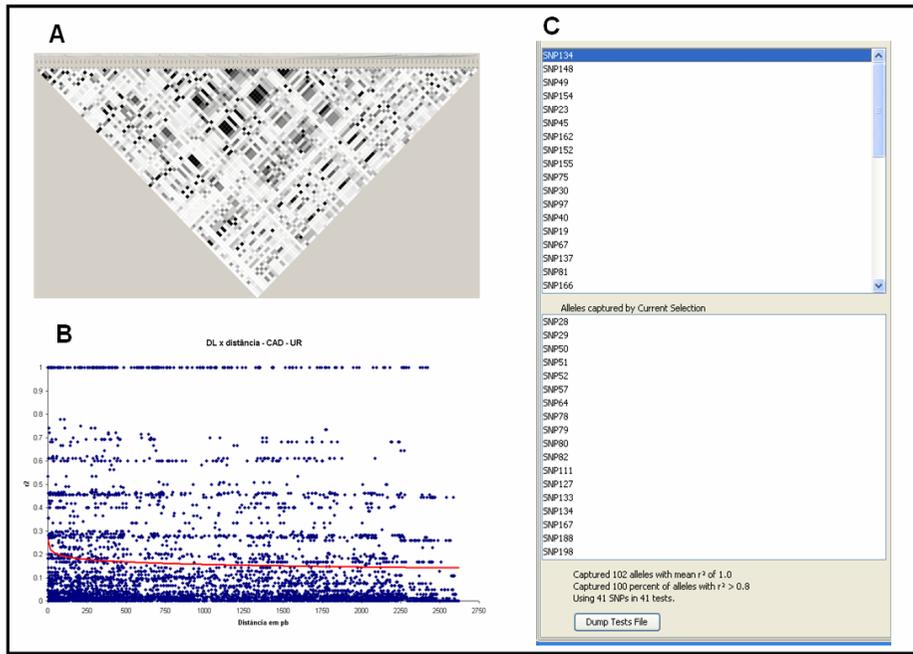


**LD markers – Linkage Disequilibrium**  
Ex. SNP's strongly associated with the QTL or candidate gene - Subcentimorgan resolution  $\sim 10^2$  a  $10^4$  pb

**Gene and exact polymorphism (QTN) identified**



**Direct Markers**  
Ex. Causal SNPs (QTNs) of quantitative variation  
Maximum resolution and identification of exact allele



## Linkage disequilibrium in the CAD gene in three *Eucalyptus* species

- ✓ **LD:** very rapid decay, < 500 bp
- ✓ **NUCLEOTIDE DIVERSITY:** very high, 1 SNP every 70 to 100 bp in coding regions
- ✓ **TAG SNPs:** very little redundancy – still lots of tagSNPs needed to capture all the haplotype variation
- ✓ **SNP VARIATION:** SNPs vary widely across species

# A reference genome will be key for applied genomics in *Eucalyptus* breeding

- ✓ *Genomic resources will be abundant and public*
- ✓ *No need to choose candidate genes*
- ✓ *Cost reduction of genomic methods: SNPs will be typed by simply resequencing whole genomes*
- ✓ *Several genomes will be sequenced for association genetics studies*
- ✓ *Short reads will be mapped onto the reference sequence of *E. grandis**
- ✓ *Main limitation: availability of appropriate structured material, sufficiently replicated across field sites and precisely phenotyped for traits of interest*
- ✓ *Genome-wide selection methods to capture all relevant genetic variation*
- ✓ *Perspectives of applied genomics in *Eucalyptus* are encouraging given the existing variation, the upcoming draft genome and the evolution of better and cheaper genotyping technologies*



**BRASUZ1**  
**BR**azil **SUZ**ano **S1**  
**TARGET REFERENCE GENOME**



***Shinitiro Oda***  
*Suzano breeder who developed and  
selected BRASUZ1*

# Why did we want a more homozygous genome?

- ✓ Eucalyptus is preferentially outcrossing (~90% outcrossing rate) with late acting self incompatibility
- ✓ Heterozygosity throughout the genome is very high
- ✓ Nucleotide variation in *Eucalyptus* is also very high (1 SNP/~100 bp)
- ✓ Difficulties expected for whole-genome shotgun assembly due to high within-individual haplotype variation
- ✓ Humans: Venter genome paper developed methods to assemble the alternative alleles
- ✓ Grape genome project: two draft genomes were published, one from a homozygous line and one from a heterozygous variety

# History of the target tree BRASUZ1

- ✓ **1968:** Seeds from *E. grandis* Coffs Harbor (Australia) were bought by Suzano and a commercial stand was planted in 1968 in São Paulo state
- ✓ **1974:** mass selected trees in this commercial stand for volume and form and collected seeds from them
- ✓ **1975:** establishment of an open pollinated progeny trial with seeds
- ✓ **1979/1980:** best trees selected between and within families in the progeny trial; trees cloned by grafting
- ✓ **1982:** clonal seed orchard established with selected trees
- ✓ **1986:** selfing program of all trees in seed orchard
- ✓ **1990:** surviving S1 seedling established a seedling seed orchard composed exclusively of S1 trees among which 7D now BRASUZ1 is one of them.
- ✓ **2008:** BRASUZ1 is now 18 years old. It has good general combining ability. and is resistant to *Puccinia* rust

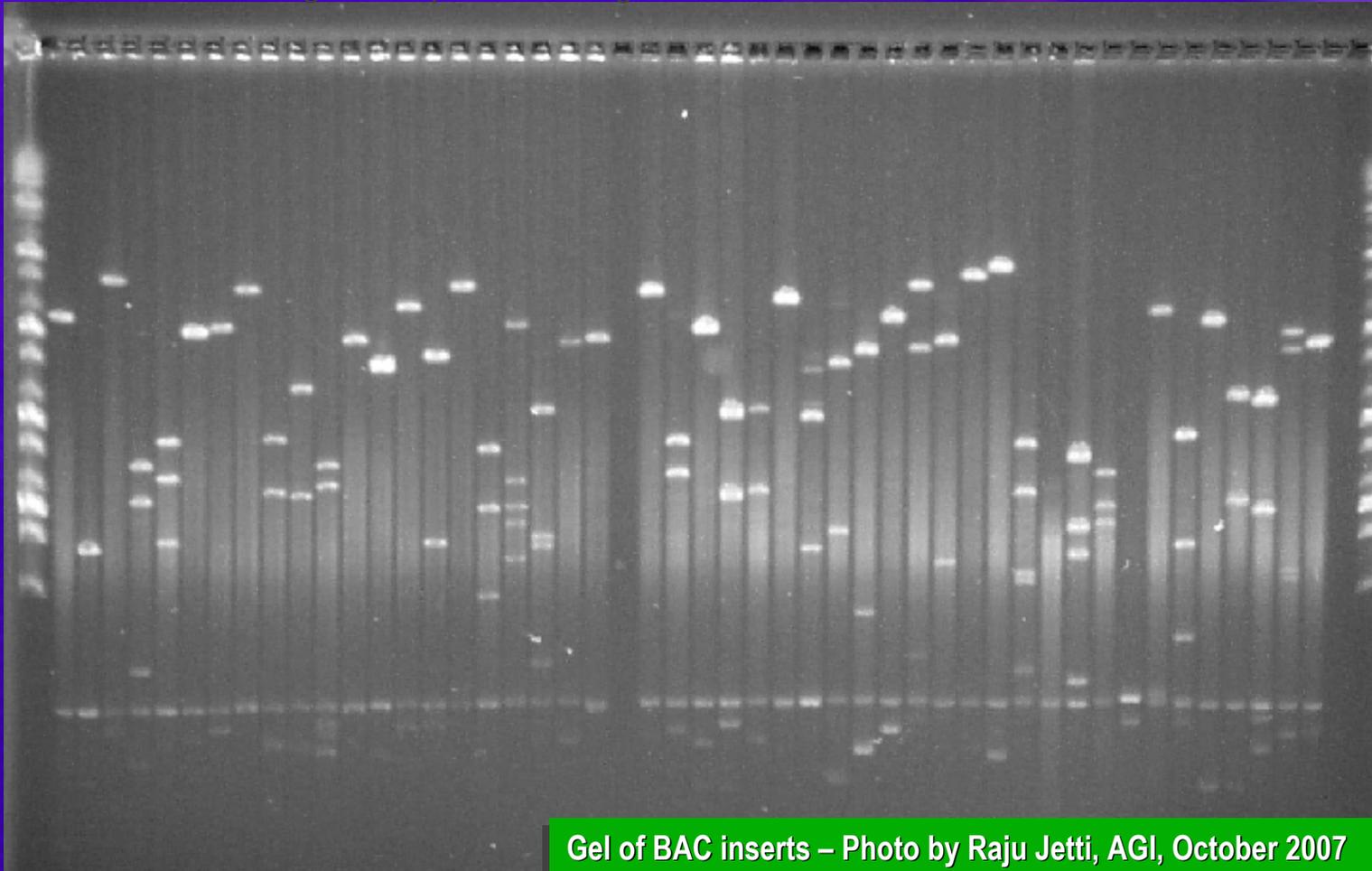


Seedling seed orchard of S1 trees where BRASUZ1 is located in Itapetininga. São Paulo State



## Public BAC libraries of BRASUZ1

- ✓ HMW DNA, cpDNA and mtDNA free was prepared in Brazil
- ✓ Delivered to Arizona Genomics Institute in September 2007
- ✓ Genolyptus project funded construction of two 15X BAC libraries at AGI: total 30 X coverage
- ✓ Both 15X library ~145 kb average, were shipped to JGI for BAC end sequencing
- ✓ Sequencing has started at Stanford: expected ~150,000 BES
- ✓ Libraries available for the general public through AGI



Gel of BAC inserts – Photo by Raju Jetti, AGI, October 2007



**Tissue collection for RNA:**

*JGI has planned to generate a large EST collection sepecifically for BRASUZ1 using 454 technology*

# How homozygous is BRASUZ1 ?

- ✓ Whole-genome microsatellite survey would allow estimating the percentage of the genome that went into homozygosity
- ✓ Assumption is that microsatellite heterozygosity is a good surrogate for sequence heterozygosity although mutation rate at microsats are typically  $10^3$  higher
- ✓ This survey would give us a first look at what to expect for haplotype discrimination and assembly challenges

♀♂ **TREE**

**Aa**

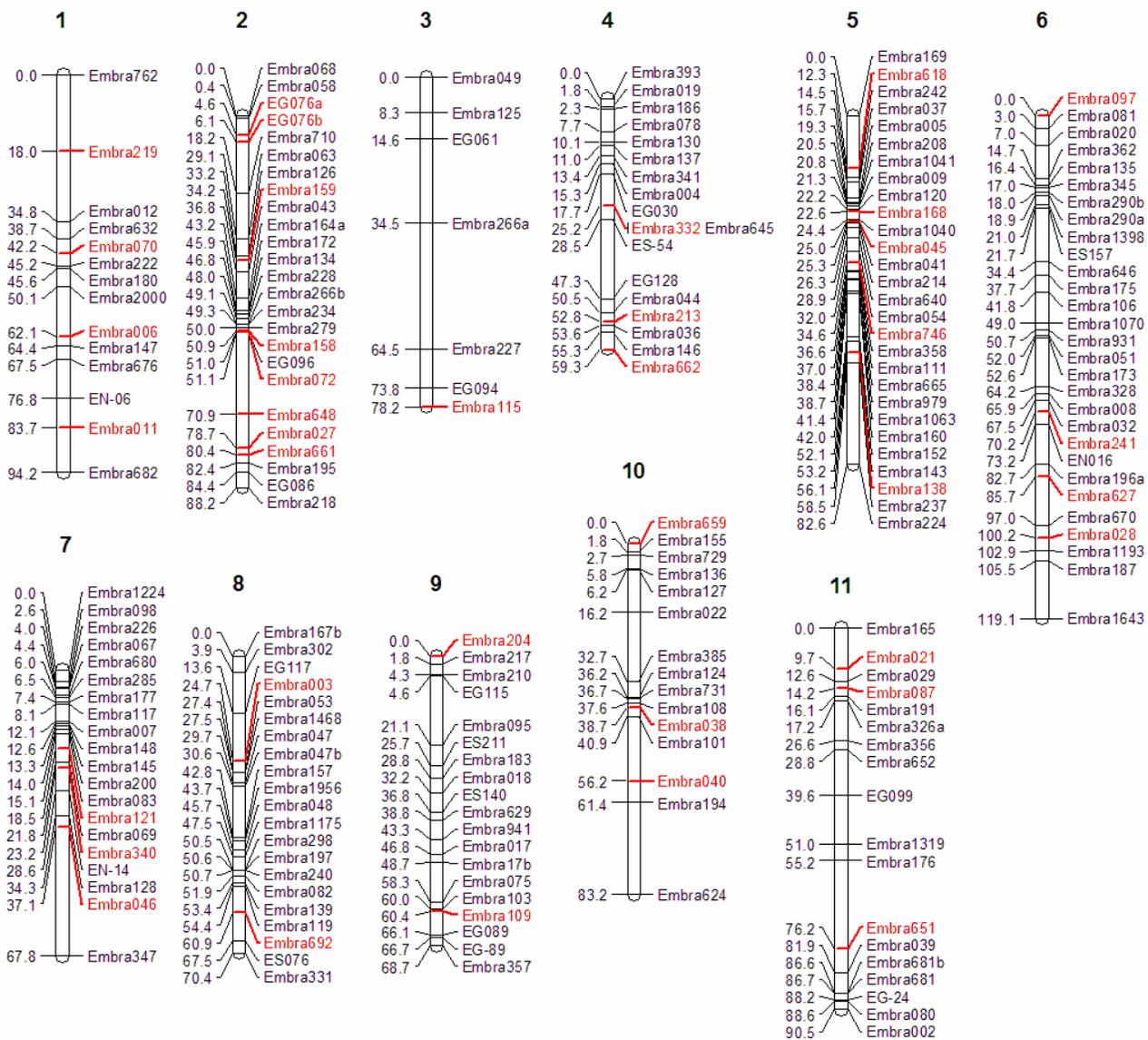
**S1 PROGENY**

**$\frac{1}{4}$  AA**

**$\frac{1}{2}$  Aa**

**$\frac{1}{4}$  aa**

# Selected microsat marker set to cover the genome based on existing maps picking markers at evenly spaced distances in cM



✓ Genotyped 212 microsatellites at an average spacing of 5 to 10 cM for a total of 900 cM

✓ Genotyped BRASUZ1 plus six other S1 full sibs available from the same orchard

## BRASUZ1 increased homozygosity in relation to a regular *Eucalyptus grandis* tree was estimated at 21%

| Linkage group | cM         | # markers  | # inf.markers | Homozygous | Heterozygous | % homozygosity | % adj.homozy. |
|---------------|------------|------------|---------------|------------|--------------|----------------|---------------|
| 1             | 76         | 16         | 13            | 10         | 3            | 0,77           | 0,06          |
| 2             | 117        | 28         | 21            | 3          | 18           | 0,14           | 0,02          |
| 3             | 54         | 14         | 11            | 3          | 8            | 0,27           | 0,02          |
| 4             | 60         | 16         | 13            | 1          | 12           | 0,08           | 0,01          |
| 5             | 110        | 26         | 17            | 2          | 15           | 0,12           | 0,01          |
| 6             | 115        | 30         | 11            | 0          | 11           | 0,00           | 0,00          |
| 7             | 63         | 17         | 14            | 6          | 8            | 0,43           | 0,03          |
| 8             | 77         | 17         | 13            | 2          | 11           | 0,15           | 0,01          |
| 9             | 60         | 18         | 13            | 0          | 13           | 0,00           | 0,00          |
| 10            | 72         | 16         | 10            | 6          | 4            | 0,60           | 0,05          |
| 11            | 98         | 14         | 11            | 0          | 11           | 0,00           | 0,00          |
|               | <b>902</b> | <b>212</b> | <b>147</b>    | <b>33</b>  | <b>114</b>   | <b>0,23</b>    | <b>0,21</b>   |

### BRASUZ1 total homozygosity estimated from microsatellites:

*212 microsats*

*65 already homozygous*

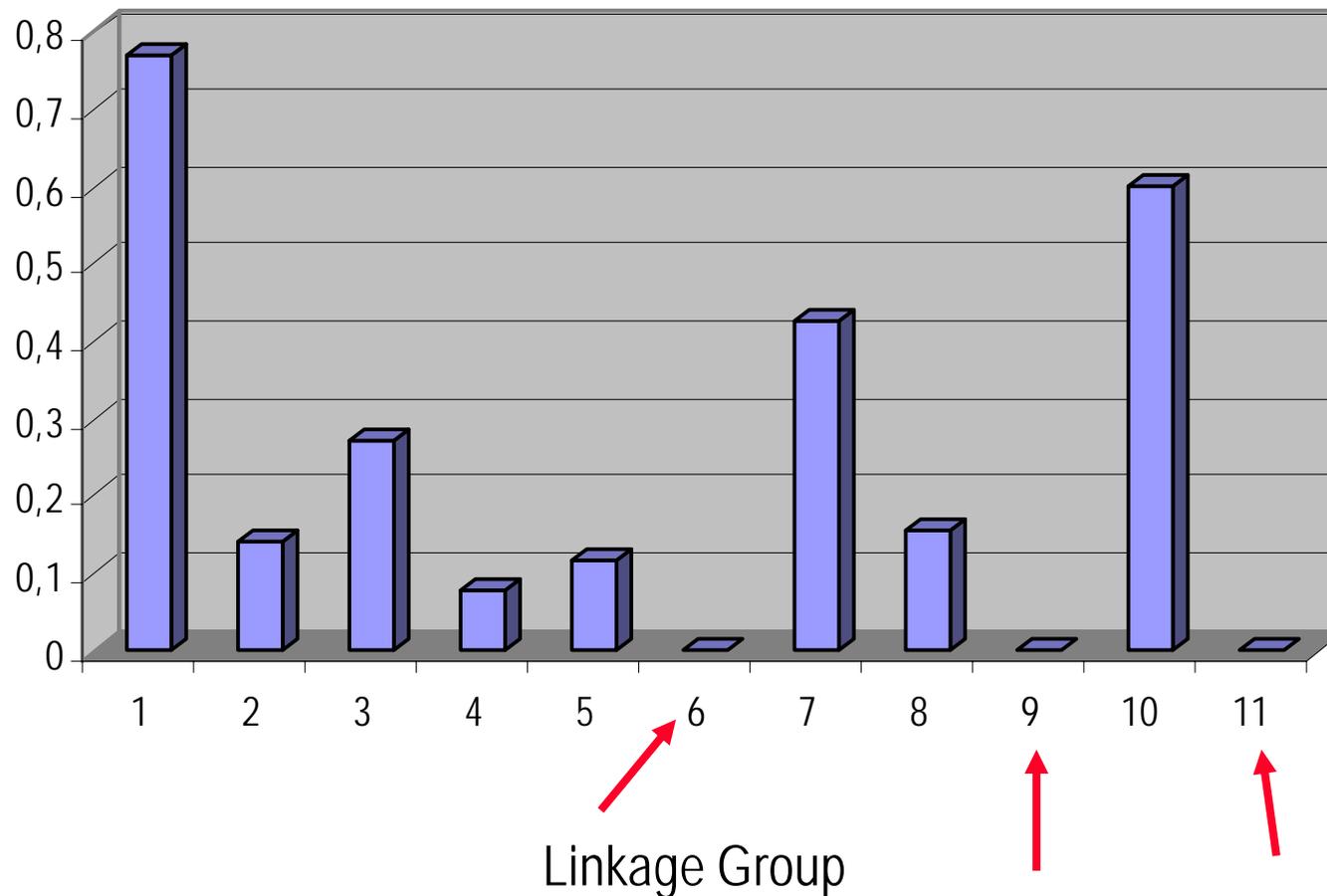
*33 went into homozygosity following selfing*

*98 microsatellites are homozygous*

*98/212 = 46% homozygosity*

# RESULTS: putative homozygosity by linkage group

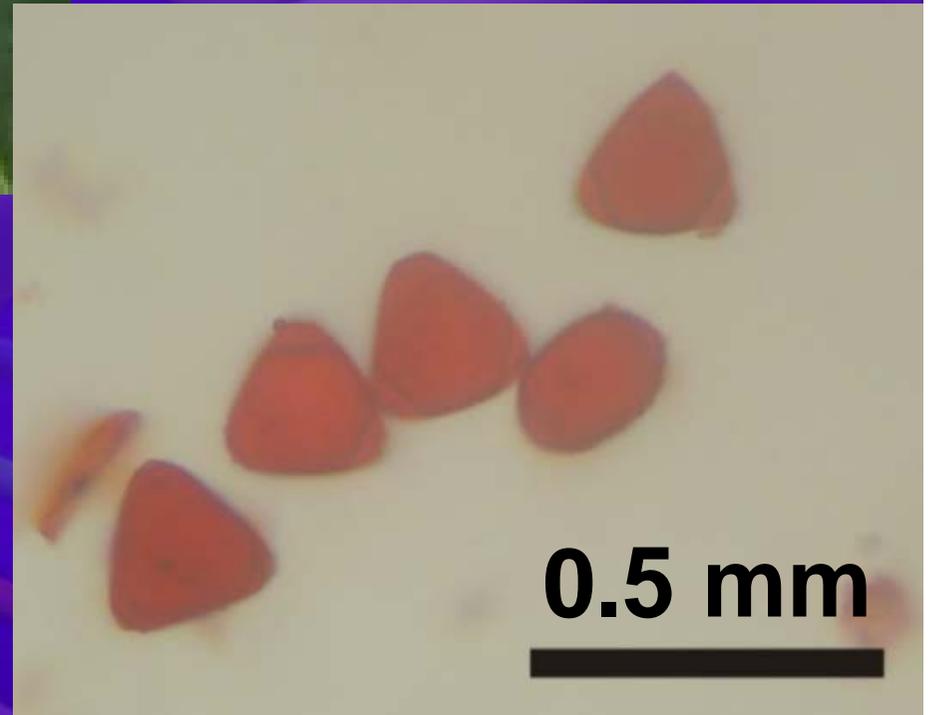
*Significant variation across linkage groups suggests variable tolerance to homozygosity due to variable distribution of genetic load*





Pollen of BRASUZ1 tree has been collected and was used this last flowering season to generate a very large mapping population ( $n > 2500$ )

We have also established isolated single pollen grains in vitro to potentially generate haploid microcallus tissue to be sequenced by UHT and "extract" one of the haplotypes to aid genome assembly





# Acknowledgments

- ✓ EUCAGEN members
- ✓ GENOLYPTUS project members
- ✓ Brazilian Ministry of Science and Technology
  - ✓ FINEP – Fundo Verde Amarelo
  - ✓ CNPq

