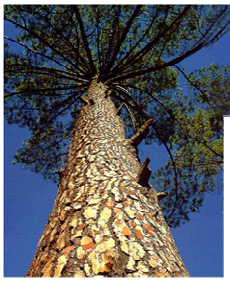


SustainPine

Genomic tools in maritime PINE for enhanced biomass production and SUSTAINable forest management

A.Rupps[1], K. Zoglauer[1], B. Vornam[2], R. Finkeldey[2], M.-T. Cervera[3], A. Guevara[3], I. Aranda[3], C. Díaz-Sala[4], D. Abarca[4], I. Arrillaga[5], C. Miguel[6], J. De Vega-Bartol[6], M. Tonelli[6], A. Matos[6], C. Plomion[7], L. Bouffier[7], E. Chancerel[7], I. Lesur[7], J.B. Lamy[7], F. Lagane[7], M.-A. Lelu-Walter[8], C. Teyssier[8], M.C. Lesage-Descauses[8], C. Le Metté[8], P. Label[9], S. Debille[10], F. Canlet[10], L. Harvengt[10], J.-F. Trontin[10], F.J. Lario[11], M. Gonzalo Claros[12], C. Avila[12] and F.M. Cánovas[12]

1:Humboldt-Univ., Berlin, Germany; 2:Univ. Goettingen, Germany; 3:CIFOR-INIA, Madrid, Spain: 4:Univ. Alcala, Madrid, Spain; 5:Univ. Valencia, Spain; 6:IBET, Oeiras, Portugal; 7:INRA, Pierroton, France; 8:INRA, Orleans, France; 9:INRA, Aubière cedex, France; 10:FCBA, Pierroton, France; 11:TRAGSA, Ourense, Spain; 12:Univ. Malaga, Spain

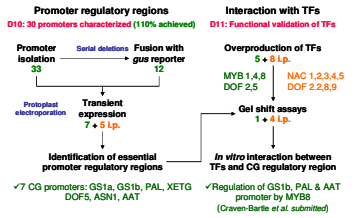
Introduction: The forest sector has to cope with the progressive impact of the global climate change, increasing wood demand, and sustainable biomass economy. The project focuses on maritime pine (*Pinus pinaster* Ait), the most advanced conifer model species for genomic research in Europe and the most widely planted species in France, Spain and Portugal.

General aims

- **Application of the latest technologies towards the identification of key genes determining adaptive traits in conifers.**
- **New knowledge for the European forest-based-sector that promotes the use of renewable resources in the context of global climate change.**
- **Recommendations for marker-based strategies to maintain the competitiveness of forest industries by improving biomass and wood productivity of conifer forests even under stress conditions.**

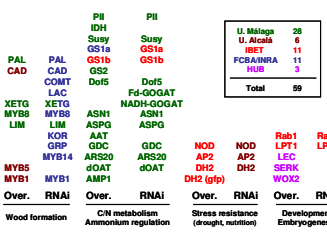
Promoter and ORF functional study

Promoter functional analysis & interaction with transcription factors

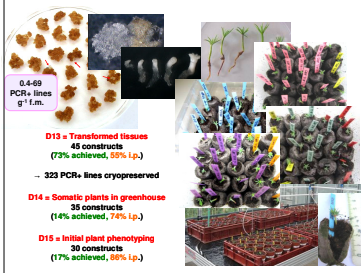


59 available constructs from 36 CG

D12 = 50 constructs (118% achieved)

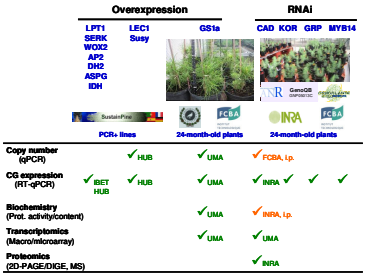


Transgenesis and plant regeneration



Molecular phenotyping

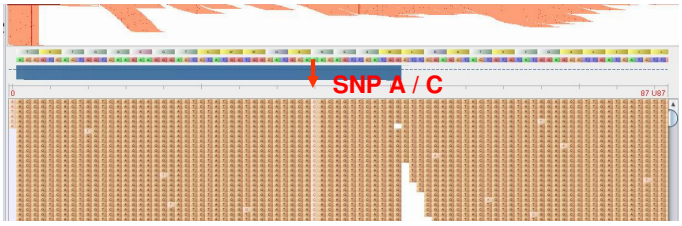
D16 = data available for 10 constructs



➪ Regulation of key genes involved in the phenylalanine pathway by MYB transcription factors has been described (submitted to Plant J. by Craven-Bartle *et al.*). So far, more than 300 transgenic lines were obtained and cryopreserved. Plantlet regeneration is ongoing. Molecular phenotyping of both, newly generated lines and previously produced plants started for 14 gene constructs including transcriptomics and/or proteomics of upregulated GS1a and successfully downregulated CAD.

Natural variation

For analysis of natural variability a total of about 50,000 nucleotides were analysed by Sanger sequencing in order to create reference sequences of 50 selected CGs on the genomic level. SNP detection within the population samples was performed using the Illumina NGS platform. For each sample equimolar amounts of all amplicons were pooled and after a sonication step the library was prepared according to the Illumina protocol. After a paired-end sequencing run the data were analysed by aligning the generated sequences to the reference sequences (see Figure below).



Alignment and coverage of sequences generated by an Illumina run of one population sample to the reference sequence of *Laccase*.

➪ So far, depending on the target gene a moderate to high haplotype diversity was found and as expected the nucleotide diversity was higher in non-coding than in coding regions and also higher for synonymous than for non-synonymous sites.

See also Poster PLANT KBBE no. C06

Candidate gene discovery and data-management

The catalog of maritime pine sequences and full-length cDNAs (FLcDNAs) has been extended and has been annotated at http://www.scbi.uma.es/sustainpine/.

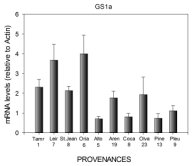
Statistics:

SustainPineDB *v3.0*	
Useful reads	3 877 472
UniGenes	210 513
Annotated	130 944
Full-length	17 815

The current version of the *P. pinaster* transcriptomics database (SustainPineDB *v3.0*) consists of 454 and Illumina libraries that accounted for 5 551 763 reads that were pre-processed and assembled.

Out of these a total of 1000 FLcDNAs were selected assigned to different functional categories: Plant growth, Nitrogen/Carbon metabolism, Photo-synthesis, Wood formation and Environmental Stress.

- Large-scale expression analysis by qRT-PCR of selected candidate genes was performed using ten maritime pine provenances (Tamr: Tamrabta; Leir: Leiria; St.Jean: St.Jeande Monts; Oria; Alto: Alto la Llama; Aren: Arenas de San Pedro; Coca; Olva; Pine: Pineta; Pleu: Pleucadec). The figure shows the expression pattern examplarily for GS1a. The mRNA level is significantly increased in Leiria and Oria genotypes. Water stress samples are increased in progress.

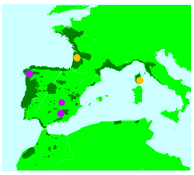


- Validation of CGs by a forestry company using pine microarray has shown that 36 genes are differentially expressed in individuals with superior stem growth and sugests a carbon flux from photosynthesis to lignin.

- More than fifty CGs have been identified and so far the precise distribution of 30 transcripts in different cell-types has been analyzed by *in situ* hybridization.

Genetic mapping: Construction of a composite SNP-based linkage map of *Pinus pinaster*

Re-sequenced CGs (*in vitro* SNPs) / SNP markers for dense mapping by high throughput sequencing mapping populations (GS FLX SNPs) / SNPs screened from ESTs (*in silico* SNPs)



Progenitors from populations covering *Pinus pinaster* natural distribution

NIA progenitors
INRA progenitors

1,536 Golden Gate array | **12K Infinium array** | **384 SNP plex array**

High throughput SNP genotyping (600 trees in total) & segregation analysis

↓

Highly-dense genetic maps mainly based on SNPs:

1,215 SNPs (881 genes) in 12 LGs (linkage groups) for the F2 → 1,660 cM
2,264 SNPs (1,473 genes) in 12 ♀ & 12 ♂ LGs for the F1 → 1,500 cM
1,624 SNPs (1,020 genes) in 16 ♀ & 13 ♂ LGs for F1 Coca14 x Coca15 → 1,132 cM
1,639 SNPs (1,245 genes) in 17 ♀ & 16 ♂ LGs (13 LGs integrated ♀ & ♂ maps) for F1 Gal1056 x Oria6 → 1,545 cM

↓

Software development for map merging (*Lpmerge* available in *R*)

↓

Construction of a composite map by merging 7 individual maps (ongoing)

↓

Comparative mapping with linkage maps of other conifers

Forward genetics: Association between trait variation and SNPs

• Pedigree-based analysis: **500** 12 year-old trees (F2 inbred pedigree)

→ Phenotyping: height growth, radial growth and wood density over 10 years, water-use efficiency, cavitation resistance
→ Genotyping: Goldengate & Mass array technologies
→ Linkage map: 248 SNPs in 12 linkage groups spanning 1,754cM
→ Major QTLs detected for all traits analyzed (e.g. 2 major QTL for WUE accounting for 69% of the variation) and for the first time insights into the dynamic of the genetic architecture of secondary growth

• Population-based analysis: **192** (G0) and **600** (G1) trees of the first and second breeding populations, respectively

→ Phenotyping: BLUPs estimated for height and radial growth, wood density, water-use efficiency based on progeny testing
→ Genotyping : 2,600 SNPs (1,600 SNPs mapped)
→ Analysis of genetic structure and short/long distance linkage disequilibrium
→ Statistical analysis under two different modes (ongoing): association mapping (SNP by SNP) and genomic prediction