"PLANT-KBBE SUSTAINPINE" Genomic tools in maritime Pine for enhanced biomass production and sustainable forest management

Andrea Rupps, Kurt Zoglauer, Vornam Barbara, Reiner Finkeldey, María-Teresa Cervera, Antonio Guevara, Ismael Aranda, Carmen Díaz-Sala, Isabel Arillaga, Célia Miguel, et al.

“PLANT-KBBE 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], I. Arrillaga[5], C. Miguel[6], J. De Vega-Bartol[6], R. Santos[6], C. Plomion[7], L. Bouffier[7], C. Lepoittevin[7], E. Chancere[7], I. Lesur[7], J.B. Lamy[7], F. Lagane[7], M.-A. Lelu-Walter[8], P. Label[8], C. Teyssier[8], S. Debille, F. Canlet, L. Harvengt[9], J.-F. Trontin[9], F.J. Lario[10], C.D.M. Gonzalo[11], C. Avila [11] and F.M. Cánovas[11]
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:FCBA, Nangis, France; 10:TRAGSA, Ourense, Spain; 11:Univ. Malaga, Spain

The forest sector has to cope with progressive impact of the global climate change in a context of increasing economic competition among industrial forest areas. The aim of this project is the identification of key genes determining adaptive traits in conifers, which are crucial for forest productivity, conservation and management. The investigations are focused 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. Results on this species would be easily transferred to closely related *Pinus* species and other economical and/or environmentally important gymnosperm species.
Within the project more than 80.000 new maritime pine ESTs were identified. Furthermore a collection of at least 6.000 pine full-length complementary DNAs for functional genomics studies was established to facilitate protein identification as well as isolation and functional analysis of gene promoters. Out of these 1.000 candidate genes involved in growth, wood formation as well as water and nutritional stress were selected for expression analysis. Candidate genes involved in drought response and with relevance for growth were tested in 10 provenances. Out of the analyzed genes 23% showed differential expression. 33 candidate genes were selected for transcript localization analysis by *in situ* hybridization. These genes are of importance for nitrogen/carbon metabolism, photosynthesis, embryogenesis, wood formation or drought stress.
For promoter functional analysis a set of 33 pine promoter sequences from selected candidate genes (CG) has been isolated and *in silico* analysis has been performed. Regulatory regions of seven promoters (GS1a, GS1b, PAL, XET, DOF5, ASN1, AAT) were identified by serial deletions and subsequent GUS-fusions showing their transient expression behaviour in pine protoplasts. To demonstrate the *in vitro* interaction of the promoter regulatory regions with selected transcription factors (TFs), recombinant proteins were produced (achieved for 5 TFs) and analysed using gel shift assays. Interactions between MYB8 TF and promoter regulatory regions (GS1b, PAL, AAT) have been described.
An extended cDNA functional study (36 CGs, 60 constructs) is performed through *Agrobacterium*-mediated genetic transformation of a pine embryogenic line with over-expression and/or RNAi silencing vectors. Production of transgenic lines and cryopreservation is well progressing for 20 RNAi and 29 over-expression constructs. Depending on the selection procedure, transformation yield has been estimated in the range 0.4–69 lines g^{-1} f.m. tissue (PCR screening). High rate of escapes were detected when low selective pressure was applied (0.1 mg l^{-1} PPT) and alternative procedures have been proposed to increase selection efficiency. After in vitro regeneration the initial survival, growth and phenotype characterization of the transgenic plants will be conciliated with the molecular and biochemical characterization (transgene expression and insertion copy numbers, protein activity/content, transcriptomics/proteomics).
For the construction of high density functional genetic maps, clones derived from a controlled cross were selected and vegetatively propagated. So far, a unigene for maritime pine (PineContigv2) was constructed and SNP-arrays (12k infinium bead array) were designed for different applications: QTL mapping, ultra-high linkage mapping, association mapping and genomic prediction. The genetic architecture of traits derived from microdensitometry (density and radial growth) was carried out in two large pedigrees. On the other hand, the French breeding populations (700 elite trees from which BLUPs are available) were genotyped which is a first step towards the implementation of molecular breeding. A consensus map was established, bringing together the information from 3 genetic maps (4 additional maps will be integrated) using orthologous markers (mainly SSRs and SNPs shared among the maps). Comparative mapping with other high density conifer genetic maps (*Pinus taeda* and *Picea glauca*) will be carried out based on orthologous markers.

For analysis of natural variability a total of about 50.000 nucleotides were analysed by Sanger sequencing in order to create reference sequences on the genomic level. Nucleotide diversity of all identified genes was studied in two test samples using the ABI SOLiD and Illumina platforms. A preliminary analysis of the results showed that the average sequence coverage for the Illumina platform was higher than for the SOLiD platform. Therefore, equimolar amounts of 56 PCR products of all investigated samples will be analysed by paired-end sequencing on a HiSeq-Illumina platform.
QTL detection was carried out based on three generation inbred (500 F2s) and outbred (200 G2s) pedigrees comprising 12-15 year-old-trees genotyped by AFLP and SNP markers. Phenotyping of wood quality, biomass production and drought tolerance was performed by determination of appropriate traits. Wood formation is strongly affected by climate variation. The genetic basis of this plasticity is currently investigated. Specific and shared QTLs for early and late wood were detected, as the results from multi-environment QTL analysis combine different years. For association mapping, 700 elite trees of the breeding programme (first and second generation) were genotyped using an Infinium SNP array generated within this project. 3000 SNPs were kept for ongoing statistical analysis including classical gene-trait associations as well as predictive modelling for genomic prediction.
All results will be annotated and connected in a reference web-interfaced database the final version of the SustainPineDB v2.2 is about to be completed (http://www.scbi.uma.es/sustainpine). Herein, a *P. pinaster* functional network including genes involved in growth, development and response to stress which is based on the *Populus trichocarpa* network was established. Furthermore a tool to predict the function from the sequence based on annotations and protein-protein interactions is projected.