DE GRUYTER
OPEN

Genetic effect in leaf and xylem transcriptome variations among *Eucalyptus urophylla x grandis* hybrids in field conditions

Alexandre Vaillant[1,2], Astrid Honvault[1,2], Stéphanie Bocs[1,2,3], Maryline Summo[1,2,3], Garel Makouanzi[4], Philippe Vigneron[1,4], Jean-Marc Bouvet[1,2]

[1] CIRAD, UMR AGAP, F-34398 Montpellier, France
[2] AGAP, CIRAD, INRA, Montpellier SupAgro, Université Montpellier, Montpellier, France
[3] South Green Bioinformatics Platform, Montpellier, France
[4] Centre de Recherche sur la Durabilité et la Productivité des Plantations Industrielles, CRDPI, BP: 1291, Pointe-Noire, Republic of the Congo.

Corresponding author: Alexandre Vaillant, email: alexandre.vaillant@cirad.fr

Abstract

To assess the genetic and environmental components of gene-expression variation among trees we used RNA-seq technology and *Eucalyptus urophylla x grandis* hybrid clones tested in field conditions. Leaf and xylem transcriptomes of three 20 month old clones differing in terms of growth, repeated in two blocks, were investigated. Transcriptomes were very similar between ramets. The number of expressed genes was significantly (P<0.05) higher in leaf (25,665±634) than in xylem (23,637±1,241). A pairwise clone comparisons approach showed that 4.5 to 14 % of the genes were differentially expressed (false discovery rate [FDR]<0.05) in leaf and 7.1 to 16 % in xylem. An assessment of among clone variance components revealed significant results in leaf and xylem in 3431 (248) genes (at FDR<0.2) and 160 (3) (at FDR<0.05), respectively. These two complementary approaches displayed correlated results. A focus on the phenylpropanoid, cellulose and xylan pathways revealed a large majority of low expressed genes and a few highly expressed ones, with RPKM values ranging from nearly 0 to 600 in leaf and 10,000 in xylem. Out of the 115 genes of these pathways, 45 showed differential expression for at least one pair of genotype, five of which displaying also clone variance components. These preliminary results are promising in evaluating whether gene expression can serve as possible 'intermediate phenotypes' that could improve the accuracy of selection of grossly observable traits.

Keywords: *: cellulose, differential expression, Eucalyptus, forest tree, genetic control, lignin, transcriptome.*

Introduction

Advances in the knowledge and understanding of the genotype-phenotype relationship are fundamental goals for all breeding programs. In populations, the genotype-phenotype relationship is largely addressed by association studies like QTL or, more recently, genome-wide association studies (Huang and Han, 2014) and genomic selection (Lorenz et al., 2011). However, this highly complex relationship has long been like a biological black box. Indeed, from the architecture of the genome, the production of the phenotype is a multi-step process subjected to many endogenous and environmental interactions. Being the main intermediate between genetic makeup and phenotype, transcriptome reflects many mechanisms of the activation and regulation of gene expression such as genetic variation, epigenetics, and environmental factors. In an effort to fill the gap between genotype and phenotype, Jansen and Nap (2001) proposed a joint analysis of genotype and transcript abundance, considered as an intermediate phenotype between DNA variants and phenotypes of interest. The promising field of e-QTL (Druka et al., 2010) has thus emerged within the global concept of genetical genomics. This methodology assumes that transcript abundance is a quantitative trait with a heritable component. The question of identifying the extent to which gene expression is a genetically controlled trait remains crucial. Understanding the genetics of transcript abundance certainly holds a lot of potential in inferring the genetic contributions to complex traits in populations. Soon after the development of microarrays, the genotypic effects of gene expression were investigated first in animals and humans (Brem et al., 2002; Cheung et al., 2003), then in plants (Kirst and Yu, 2007; Druka et al., 2010; Lukens and Downs, 2012). These studies

DOI:10.2478/sg-2018-0008 | CIRAD Centre de Cooperation Internationale en Recherche Agronomique pour le Developpement
edited by the Thünen Institute of Forest Genetics

have highlighted important patterns of heritability and population differentiation in gene expression (Gibson and Weir, 2005; Gilad et al., 2008).

These fundamental advances are now reaching the limits of microarray technology. The recent advances in sequencing of RNA with RNA-seq opened up new opportunities (Martin et al., 2013; Zhao et al., 2014). Compared with hybridization-based transcriptome studies, RNA-seq allows analysis of genome-wide transcription without any prior knowledge of the genome, with higher sensitivity, better dynamic range of detection, lower technical variations and by-transcript quantification. The increasingly reasonable cost of this technology makes it affordable for most studies. However, while RNA-Seq is extensively used in plants to uncover transcriptomic changes between two contrasted growing conditions or developmental stages, analysing differential expression, its potential for studying the genetic determinism of gene expression remains under-explored. RNA-Seq based differential expression between several genotypes or within populations, and its suitability for the determination of the genetic part of gene expression level, especially of plants grown in poorly controlled environments that are actual plantation conditions has been poorly investigated.

We addressed this issue with *Eucalyptus*. This genus has a great economic importance as it is one of the most widely planted trees in the tropical and subtropical regions, for pulp, paper, energy, timber, and possibly biofuel production. The present study focused on the highly valuable *Eucalyptus urophylla* **x** *Eucalyptus grandis* hybrid, being prominent in humid tropics due to its superior adaptation to a humid climate and poor soils, its rapid early growth, its ability to resprout and its wood, which is suitable for solid wood and pulp production (Vigneron and Bouvet, 2001). Indeed, cellulose and lignin constitute respectively 48.6 % and 26.7 % of it biomass (Evtuguin and Pascoal Neto, 2007). Two different tissues that are important for production of lignocellulosic biomass were investigated: leaf, which by ensuring photosynthesis and transpiration provides plants with energy to grow and allows upwards movement of water and minerals, and developing xylem which contributes to wood formation. Aside from its high commercial value, *Eucalyptus* is also a pivotal genus for genomic research in forest trees. From the past 15 years, many genomic resources have been produced, culminating with the recent sequencing of the *Eucalyptus grandis* (Myburg et al., 2014) which provides the research community a reference sequence. Transcriptomics in *Eucalyptus*, including *E. grandis* and its hybrids, has also increased exponentially with many recent studies on different tissues, at different maturity stages and between contrasted conditions (Camargo et al., 2014; Hefer et al., 2015; Liu et al., 2014; Mizrachi et al., 2010; Salazar et al., 2013; Villar et al., 2011; Vining et al., 2014). It generated a wealth of knowledge in terms of annotation and expression features.

Studies of genetic control of gene expression have already be conducted on *Eucalyptus* with microarrays (Kirst et al., 2004; Kirst et al., 2005; Kirst and Yu, 2007) using eQTL and based on a limited number of genes. Although those studies have demonstrated that gene expression is controlled by genotype through significant QTLs, they did not separate the genetic the environmental components in gene expression variation. Based on an experimental design mimicking the environmental conditions of eucalyptus plantations, our objective was to assess the magnitude of the genetic part in the control of gene expression in the whole genome and within the lignin, cellulose, and glycan pathways. From RNA-Seq assessment of transcript abundance, we used a combined approach of differential expression tests and dissection of the determinants of gene expression levels. To this end, we first compared the transcriptome profiles between biological replicates (i.e. ramets on eucalyptus clones) through correlation testing between read counts. Then, we performed the pairwise differential expression tests classically used to identify the differentially expressed genes in RNA-Seq studies. Lastly, we examined the variance component related to the genotype effect in a linear mixed modelling of the total variance of transcript abundance.

Materials and Methods

Plant material and growing conditions

Trees were selected from a genetic field trial established in the humid tropical conditions of the Republic of the Congo, east of Pointe-Noire (11°59' 21" E, 4°45' 51" S). It consists in 69 full-sib families of the hybrid *Eucalyptus urophylla* (as female) crossed with *Eucalyptus grandis* (as male) tested for their growth performances (Makouanzi et al., 2015). All trees were grown under the same environmental and cultivation conditions, but were subject to micro-environmental variations that exist under field conditions. Three clones, G198, G204 and G309, were selected from three different families, on the basis of their contrasted growth during the juvenile stage. For each clone, two ramets were sampled in two different blocks, so that they had approximately the same height as a proxy for similar growth trajectories. They constitute two biological replicates (1 and 2). Different variables related to growth, leaf morphology and wood chemical content were measured: mean, coefficient of variation and heritability were calculated for each trait (Supplementary Table 1), as described by Makouanzi et al. (2015). Twenty months after planting, samples for RNA-seq analyses were collected. They consisted of mature leaves (L) and immature xylem (X). Mature leaves were collected on branches located at the bottom of the first third of the crown. Developing xylem was collected scratching the stem after bark removal, 20 cm above the ground. In total, 12 samples were collected, corresponding to all combinations of genotype (G), replicate (1 or 2) and tissue (L or X).

RNA-seq, mapping and annotation

Total RNA was isolated from all samples using PureLink® Plant RNA Reagent (Invitrogen, USA) followed by an additional purification step using the RNeasy Plant Mini Kit (Qiagen, USA), according to the manufacturer's specifications. Total RNA purity and concentration were determined using a BioSpec-mini spectrophotometer (Shimadzu, Japan) equipped with a

Hellma TrayCell (1 mm optical path). The integrity of RNA molecules was checked using the Agilent RNA 6000 Nano kit on a 2100 Bioanalyzer (Agilent, USA). Messenger RNA molecules were purified from 3 µg of total RNA, and used to construct 12 tagged random primed cDNA libraries. Libraries were multiplexed in 2 pools and sequenced on 2 Illumina HiSeq 2000 lanes, generating 180,000,000 single reads of 50 bases per pool. Pre-processing of RNA-Seq data, mapping and counts were performed using Galaxy (Goecks et al., 2010). Adapters were removed using Cutadapt (Martin, 2011) and quality of reads was evaluated by FastQC (Andrews, 2010). Reads were then filtered by minimum read lengths (35b) and by minimum quality scores (Phred-score ≥30) with Filter FastQ (Blankenberg et al., 2010). They were then mapped on *Eucalyptus grandis* exome v1.1, downloaded from Phytozome 10 (Goodstein et al., 2012; http://www.phytozome.net). Read alignments were generated with Bowtie for Illumina (v1.1.2), allowing 1 possible location per read in the genome, with 2 possible mismatches, as we were mapping a hybrid genome. As quantitation of expression at transcript level was not accurate enough with the sequencing technology, transcript counts across all isoforms were summed to compute abundance at gene level. Gene annotations were retrieved from Uniprot Knowledgebase, available as the *Eucalyptus grandis* reference proteome (Proteome ID: UP000030711).

Genome-wide analysis of differential expression

Analyses were performed using R system (version 3.0.3) (R Development Core Team 2014) and its dedicated software package EdgeR (Robinson et al., 2010). To conduct appropriate statistical tests, genes that cumulated fewer than 50 counts for all libraries were discarded. Counts were then normalized using the relative log expression (RLE) method, which has proved to be one of the most efficient (Kvam et al., 2012). Repeatability of transcriptome profiles between replicates was explored examining correlations between libraries. A first comparison was made on pseudocounts ($\log_2$(counts + 1)) producing a heatmap of the distance matrix with MixOmics Package (Lê Cao et al., 2009) of R Software. Then, similarities were investigated in normalized libraries by multidimensional scaling analysis of $\log_2$ fold changes ($\log_2$FC), using MDS plot from edgeR.

The differential expression of genes between the three different genotypes at the two different tissue levels was studied using the EdgeR package. Given our experimental design, we used the generalized linear model (GLM), which takes into consideration the relationship between mean and variance for read counts (McCarthy et al., 2012). A set of six conditions (3 genotypes x 2 tissues) was defined. Conditions were compared pairwise according to contrasts based on genotypes for each kind of tissue. Differential expression was determined for each gene using the GLM likelihood ratio test, which fits negative binomial GLMs with the Cox-Reid dispersion estimates (McCarthy et al., 2012). Genes were considered as differentially expressed (DE) at false discovery rate (Benjamini and Hochberg, 1995) of $p < 0.05$.

Detection and assessment of the genetic control of gene expression

Although the number of genotypes was very small we used the concept of heritability to complete the among genotype differential expression analysis by assessing the ratio of the variance among clones to the total variance. To estimate properly the variance components we first normalize the raw data by $\log_2$ transforming the number of reads using the voom function of the Limma R package (Ritchie et al., 2015). A linear mixed model was then implemented with the clone as the random effect following a normal distribution $N \sim (0, \sigma^2 c Id)$ where $\sigma^2 c$ is the among-clone variance, which is supposed to model the total genetic variance, and a residual effect corresponding to environment, $N \sim (0, \sigma^2 e\ Id)$ where $\sigma^2 e$ is the environmental variance and Id the identity matrix.

The variance component estimation was done using the ASReml version 3 package (Gilmour et al., 2006) implemented in R software (R Development Core Team 2014). The variance component ratio (VCR) was defined as $VCR = \sigma^2 c/(\sigma^2 c + \sigma^2 e)$, based on broad sense heritability concept (Falconer and McKay, 1996). Standard errors of estimates were calculated with a delta method function in the 'car' package in R (R Development Core Team, 2014). A statistical test to decide if $\sigma^2 c$, and in consequence VCR, was different from zero was done using the likelihood ratio test (Neyman and Pearson, 1993). To avoid the Type I error due to the numerous tests the false discovery rate (FDR) correction was used to define the threshold of significance.

Genes of the secondary cell wall formation pathways

The list of genes involved in the phenylpropanoid pathway was taken from Myburg et al. (2014), with a focus on the core lignification toolbox as defined by Carocha et al. (2015). Genes of the cellulose and xylan pathways were selected as described previously by Hefer et al. (2015). Differences of expression of genes between genotypes were examined comparing their relative expression levels (FPKM) and analysing their individual results from the genome-wide differential expression tests described above. Estimates of the genetic control of gene expression were retrieved from the global analysis described in this specific paragraph.

Results and discussion

Libraries and detection of expressed genes

The number of reads per library varied from 41,808,259 (G309-1-L) to 9,680,836 (G204-1-X) (Supplementary Figure 1a), of which 70.68 to 80.86 % were used for mapping. Out of the 36,376 genes present in the reference transcriptome, 23,445 (G204-1-X) to 29,970 (G198-2-L) were expressed in samples, with an average of 27,432.2 (Supplementary Figure 1b). The filtered number of expressed genes was significantly higher in leaf than in xylem, with values of 25,665±634 and 23,637±1,241, respectively ($P<0.05$). Comparison of the number of reads in

libraries with the number of detected genes across all the samples showed that efficiency in detection of expressed genes was not correlated with library size, as shown by the sample G204-1-X for which the number of detected genes was similar to the other samples despite a substantially reduced number of reads. Removal of low expressed genes (sum of reads for the twelve samples ≤50) had a moderate effect on the total number of genes detected, as 86.93 % (G-198-2-L) to 95.33 % (G204-1-X) were retained. This suggested that whatever the library size, generated reads were evenly distributed across the transcriptome, and that the estimation of gene expression levels between samples should not be much affected by variations in the library size once normalized.

Repeatability between replicates

The multidimensional scaling plot of $\log_2$ fold-change normalized values (Figure 1) showed that biological replicates were very similar in terms of global gene expression. However, repeatability of transcript abundance within genotypes is not the same for every couple of replicates. The main difference is related to tissues, as repeats were closely correlated in leaf, but much less in xylem. This tissue-specific pattern could be due to a greater heterogeneity of the mix of cell populations collected in xylem between replicates (López de Heredia and Vázquez-Poletti, 2016), where the active cell proliferation and the cellular differentiation can produce variations in transcriptome profiles.

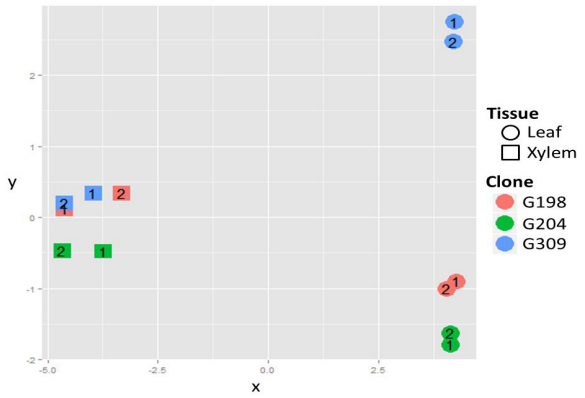

Figure 1
Similarity assessment of transcriptome expression profiles between samples by multidimensional scaling computed from normalized tables of counts of each library. Distances on the plot represent the two leading $\log_2$ fold-change differences between samples

Genotypic effect on transcript abundance between genotypes

Correlation analyses revealed some differences between clones as illustrated by the multidimensional scaling plot of $\log_2$ fold-change (Figure 1). As seen above, those differences seemed to be a function of tissue. In xylem, transcriptome profiles are much more identical than in leaves, the small existing differences between all of them being explained as much by repeats as by genotypes. In leaf, the global gene expressions are more distinct and genotype driven, with a marked distinction between G309 and the other clones G198 and G204.

Differential expression analyses between pairs of clones revealed that 4.5 to 14 % of the genes were DE in leaf and 7.1 to 16 % in xylem. An increased FDR stringency (p<0.001) still reported hundreds of DE genes showing the reliability of DE genes discovery in this study despite the small number of genotypes and biological repeats. Combining the pairwise differential expression results, the proportion of genes commonly differentially expressed between the 3 clones was dramatically reduced. Yet 0.7 % of the expressed genes in leaf and 0.6 % in xylem had significantly different levels of expression in each genotype (Figure 2). This shows that some genes have a higher potential of variation in expression within a population. By contrast, the number of commonly non-DE genes remained high, being 79.2 % in leaf and 79.1 % in xylem, which means that most non-DE genes are the same ones among all genotypes. In total, 20,984±43 genes, that is about 58 % of *E. grandis* identified genes, were evenly expressed in the three *E. urophylla x E. grandis* hybrid genomes. The much higher number of non-DE genes compared to DE genes is consistent with findings of Thavamanikumar et al. (2014) in xylem of *E. nitens*, and may be at least partly attributable to housekeeping genes that are constitutive genes required for the maintenance of basic cellular function and are expressed at relatively constant levels.

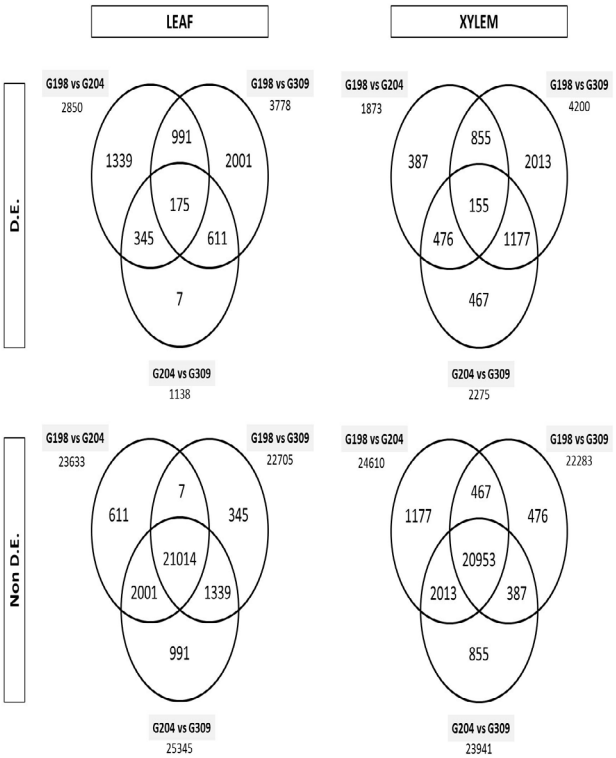

Figure 2
Number of genes differentially expressed (DE) and non-differentially expressed (Non DE) between the three clones for leaf and xylem

Nevertheless such a low proportion of DE genes in our experiment was not expected given the natural differential expression existing between the *urophylla* and *grandis* parental species (Salazar et al., 2013), the disturbance in the regulation of gene expression often observed in interspecific hybrids (Hegarty et al., 2008) notably concerning the trans-regulatory mechanisms, the possible allele-specific expression mechanism (Song, 2016), and the open field cultivation conditions.

Variance component analysis showed that the value of the ratio of the variance among clones to the total variance (VCR) displayed roughly similar distributions in leaf and xylem (Figure 3a and b). A large number of genes (around 7,000 in leaf and 11,000 in xylem) presented a value smaller than 0.1. Then the frequencies dropped to about 1,000 genes and then increased slightly with variance values. At FDR 5 %, values of VCR higher than 0.996 in leaf and 0.999 in xylem could be considered as significantly different from zero. It represents 160 genes in leaf and 3 in xylem (Table 1). As expected, the number of genes increased when loosening threshold stringency (FDR <0.10 and <0.20), but this trend was much more marked for leaf. The lower number of genes showing significant VCR in xylem (Table 1) was consistent with the mean values of VCR for all genes: the mean value was smaller for xylem with marked standard error (0.35±0.42), compared with leaf (0.48±0.23). This lower value and higher standard error for xylem is congruous with the lower repeatability of transcriptome profiles for xylem in all libraries noted above. As previously hypothesised it could be due to a higher heterogeneity in cell populations in samples. Nevertheless, phenotypic analyses conducted on leaf and xylem of the 3 clones used in this study showed that wood chemical variables presented non-significant VCR compared with leaf morphology and mineral content (Supplementary Table 1).

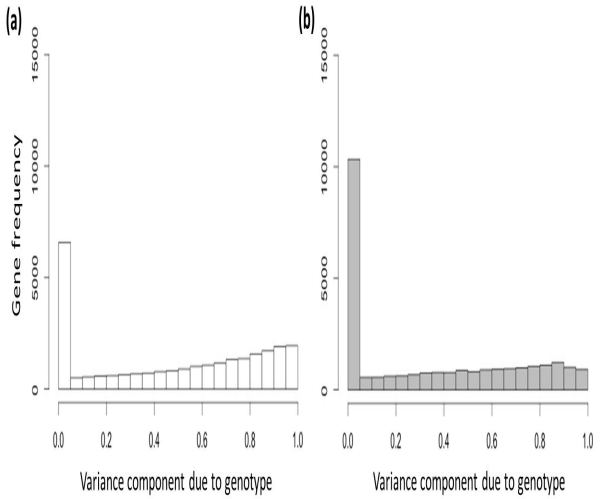

Figure 3
Relationship between gene expression and the variance component ratio in transcript abundance (VCR) for leaf (a) and xylem (b)

Table 1
Variance component ratio in transcript abundance (VCR), and number of significant genes for leaf and xylem, according to different calculation methods of probability (p-value associated to the likelihood ratio test (LRT), false discovery rate (FDR)) and different thresholds

Tissue	VCR	number of significant genes	probability	
			method	threshold
Leaf	0.934	5050	p-value LRT	0.05
	0.955	3431	FDR	0.2
	0.985	1022	FDR	0.1
	0.996	160	FDR	0.05
Xylem	0.929	2705	p-value LRT	0.05
	0.99	248	FDR	0.2
	1	3	FDR	0.1
	1	3	FDR	0.05

This trend was confirmed extending this analysis to all the 1480 clones of this field experiment (Makouanzi, 2015). This argues for a biological origin of the difference in the variance component due to genotype in transcript abundance between leaf and xylem, which could be attributed either to a smaller genetic variance, or to a broader environmental variance in xylem gene expression. This similar pattern of variation between these complex traits and the transcript level of genes reinforces the assumption that, even though correlation between transcript and protein abundance is disputed (Vélez-Bermúdez and Schmidt, 2014), gene expression could be related to some traits in *Eucalyptus* and that the transcriptome is a promising intermediate phenotype to discriminate genotypes.

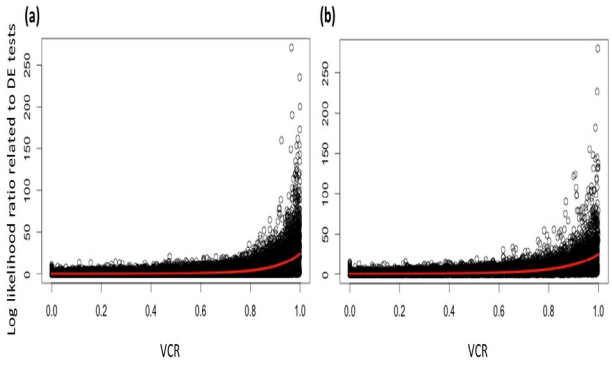

Figure 4
Relationship between the variance component ratio (VCR) and the log likelihood ratio related to differential expression test (DE) for leaf (a) and xylem (b). The red line is the curve fitting the data calculated with the non-linear model $y=a*\exp(b*x^2)$, (nls2 R package)

By contrast, no relationship was detected between estimates of VCR and gene position along chromosomes or with gene expression levels, for both tissues. This last result differs from the findings of Yang et al. (2014) who noted that lower heritability estimates (h^2<0.2) were more likely to occur in genes with low expression levels. This absence of correlation in our study may be attributed to the limited sample size.

In our study, because there was very small number of genotypes, both DE and variance ratio methods were conducted to address the genetic control of the gene expression. In order to verify the consistency between these two complementary approaches, one considering clone as a fixed effect, the other as a random effect, the relationship between clone variance component and DE test likelihood ratios was analysed. A non-linear relationship was observed that was significantly fitted by an exponential model for both leaf and xylem (Figure 4 a and b). This analysis reinforces the conclusion that differentially expressed genes among genotypes could result from a genetic control of gene expression.

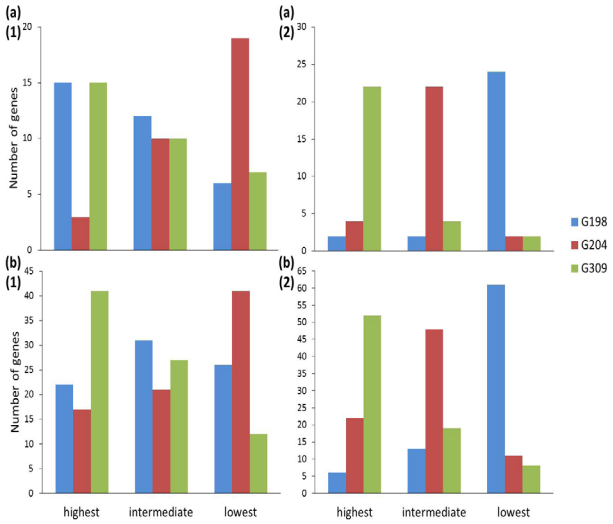

Figure 5
Classes of clone expression based on the ranking of the level of expression of their genes, in the phenylpropanoid pathway (a), and the cellulose and xylan pathways (b), for leaf (1) and xylem (2)

Gene expression and genotypic effects within pathways of secondary cell wall formation

Expression levels in both tissues were highly heterogeneous in the phenylpropanoid, cellulose and xylan pathways, with RPKM values ranging from nearly 0 to 600 in leaf and 10,000 in xylem. But they displayed the same global trend, consisting in a large majority of very low to moderately low expressed genes, and a few highly expressed ones (Supplementary Figures 2 and 3).

Some differences of expression between clones were highlighted by ranking genotypes according to the level of expression of their genes (highest, intermediate and lowest). In xylem, the three genotypes exhibited a marked specific profile, whereas in leaf patterns were less pronounced (Figure 5). Those distinctions mainly relied on small differences in terms of level of expression, $\log_2$FC values between genotypes being ≤1 for 75 % of genes in the phenylpropanoid pathway and 89 % in the combined cellulose and xylan pathways. Nevertheless some genes displayed $\log_2$FC ≥2. It occurred in 6 % of genes in the lignin pathway, and in 1 % of genes of the combined cellulose and xylan pathways. Filtering these RPKM-based differences with differential expression tests, significant differential expression (at FDR<0.05) was detected in 24 pairs of genes tested out of 214 (Supplementary Figure 2) in the phenylpropanoid pathway and in 40 cases out of 486 in the cellulose and xylan pathways (Supplementary Figure 3), 80 % of which exhibited a $\log_2$FC ≥1.5. The number of genes displaying differential expression, the specific homologs within gene families and the enzymatic functions concerned differed according to tissue and pathways.

In leaf, differential expression was found in 19 genes. Eight belonged to the phenylpropanoid pathway (Supplementary Figure 2a), involving 6 enzymatic functions: PAL (1), C4H (2), F5H (2), COMT (1), CAD (2), and especially in HCT (1, 2 and 3). The cellulose pathway was rather evenly mobilized among genotypes (Supplementary Figure 3a). Log_2FC values were mostly close to zero and no DE genes were found in the SUSY and CESA families. Only one gene of the alternative D-glucose route displayed significant DE, which was, however, based on a moderate level of expression and a rather low $\log_2$FC difference. The 10 other DE genes found in the xylan pathway, GATL and DUF231 being the 2 most represented functions.

In xylem, differential expression affected only 3 enzymatic functions out of the 11 ones that constitute the phenylpropanoid pathway (Supplementary Figure 2b). Interestingly, 5 of the 8 DE genes belonged to the phenylalanine ammonia lyase (PAL) family and one corresponded to a cinnamate 4-hydroxylase (C4H), which catalyzes the first and second common steps of the phenylpropanoid pathway respectively. The two other DE genes coded for some hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferases (HCT). It is noteworthy that the lignin specific PAL3, C4H2 and HCT5, which also display high levels of expression and $\log_2$FC values between 1 and 2, were among those DE genes. The cellulose and xylan pathways exhibited 22 DE genes that were distributed in most families (Supplementary Figure 3b). Among them, sucrose synthase (SUSY), one of the two vital enzymes able to mobilize sucrose in plant pathways exhibited differential expression in 3 homologs out of the 7 known ones. The values of expression levels and $\log_2$FC associated with these statistically significant DE genes made these results particularly trustworthy. By contrast, the large and key cellulose synthase (CESA) family was poorly represented, as only one homolog displayed some differential expression.

The variance component due to genotype in transcript abundance within these secondary cell wall formation pathways followed the same pattern as for the totality of genes (Figure 6), as well as the same range of variation. Mean VCR of

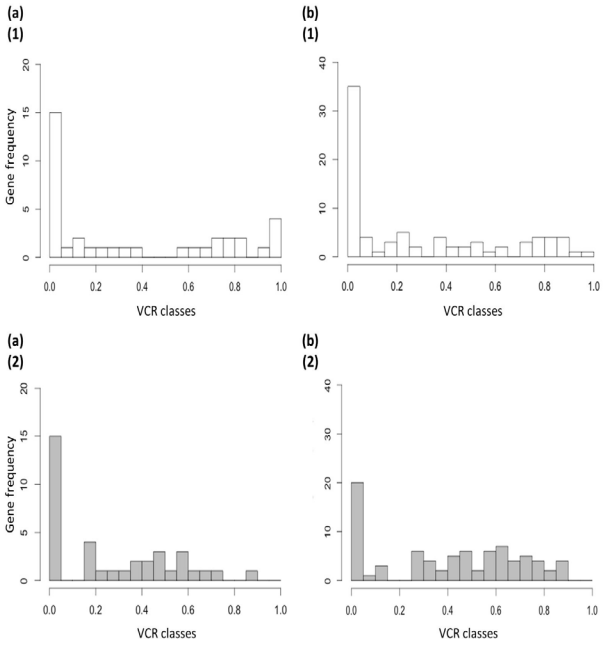

Figure 6
Distribution of frequencies of the variance component ratio in transcript abundance (VCR) for leaf (1) and xylem (2) in the phenylpropanoid (a) and the cellulose and xylan pathways (b)

phenylpropanoid genes was higher in leaf (VCR =0.37) than in xylem (VCR =0.28), as for the totality of genes, but the opposite result was seen for cellulose and xylan (VCR =0.29 in leaf and VCR =0.40 in xylem). This result suggest that the genetic control in gene expression within these pathways were not related to the genes constituting these particular pathways but to regulation ones. Significant VCR (at FDR<0.20) was found in the level of expression of 5 genes of these lignin, cellulose, and xylan pathways, despite the small number of significant genes discovered at the whole transcriptome level. Only leaf tissue was concerned. Four of these 5 genes belonged to the phenylpropanoid pathway (PAL1, HCT2, HCT3, F5H2), but no lignin-specific homolog was involved. DUF231/Eucgr.J00985 was the only gene of the cellulose and xylan pathways. All these 5 genes displayed significant differential expression between one or two couples of genotypes. The high associated $\log_2$FC values and levels of expression supported a biological significance of these differential expressions, apart from the low expressed PAL1.

Conclusion

Both differential expression and variance component ratios (related to the broad sense heritability) succeeded in detecting some genetic control in the levels of gene expression between three *Eucalyptus* clones. This result was in line with previous studies on *Eucalyptus* using eQTL approaches (Kirst et al., 2004; Kirst et al., 2005; Kirst and Yu, 2007). However, our RNA-seq based approaches brought new elements to these earlier findings, exploring the whole transcriptome without a priori instead of a restricted set of preselected genes, and separating the genetic from the environmental part in gene expression that were confounded in eQTL studies. Detection of genetic control in gene expression was not really expected given the poorly controlled experimental layout characterised by strong micro-environmental effects on trees and field sampling conditions, and our limited experimental set-up. Nevertheless, this result is particularly sound as the majority of genes were found commonly non-differentially expressed while transcriptome replicates were repeatable. A detailed analysis of the secondary cell wall formation pathways detected some of the genes susceptible to genetic control in the key lignin and xylan pathways, more particularly in leaf tissue.

Acknowledgements

Field experiments were conducted in the CRDPI in the Congo and molecular analyses at the Cirad laboratories in Montpellier, France. We thank our colleagues in the Congo for their valuable help in sampling.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Cambridge: Babraham Institute Bioinformatics [online]. To be found at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> [quoted 18 June 2016]

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Roy Statist Soc Ser B (Methodological) 57(1):289–300

Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy Team (2010) Manipulation of FASTQ data with Galaxy. Bioinformatics 26(14):1783–1785. https://doi.org/10.1093/bioinformatics/btq281

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. Science 296(5568):752–755. https://doi.org/10.1126/science.1069516

Camargo EL, Nascimento LC, Soler M, Salazar MM, Lepikson-Neto J, Marques WL, Alves A, Teixeira PJ, Mieczkowski P, Carazzolle MF, Martinez Y, Deckmann AC, Rodrigues JC, Grima-Pettenati J, Pereira GA (2014) Contrasting nitrogen fertilization treatments impact xylem gene expression and secondary cell wall lignification in Eucalyptus. BMC Plant Biol 14:256 https://doi.org/10.1186/s12870-014-0256-9

Carocha V, Soler M, Hefer C, Cassan-Wang H, Fevereiro P, Myburg AA, Paiva JA, Grima-Pettenati J (2015) Genome-wide analysis of the lignin toolbox of Eucalyptus grandis. New Phytol 206(4):1297–1313. https://doi.org/10.1111/nph.13313

Cheung VG, Jen KY, Weber T, Morley M, Devlin JL, et al. (2003) Genetics of quantitative variation in human gene expression. Cold Spring Harbor Symp Quant Biol 68:403–407. https://doi.org/10.1101/sqb.2003.68.403

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21(18):3674–3676. https://doi.org/10.1093/bioinformatics/bti610

Druka A, Potokina E, Luo Z, Jiang N, Chen X, Kearsey M, Waugh R (2010) Expression quantitative trait loci analysis in plants. Plant Biotechnol J 8(1):10–27. https://doi.org/10.1111/j.1467-7652.2009.00460.x

Evtuguin DV, Pascoal Neto C (2007) Recent advances in eucalyptus wood chemistry: Structural features through the prism of technological response. In: 3th International Colloquium on Eucalyptus Pulp. Belo Horizonte, Brasil

Falconer DS, Mackay Longman TFC (1996) Introduction to Quantitative Genetics, 4th ed. Harlow, UK: Pearson United Kingdom, 480p, ISBN 9780582243026

Gibson G, Weir B (2005) The quantitative genetics of transcription. Trends Genet 21(11): 616-623. https://doi.org/10.1016/j.tig.2005.08.010

Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet 24(8):408–415. https://doi.org/10.1016/j.tig.2008.06.001

Gilmour AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R (2006) ASReml, User Guide. Release 2.0. VSN International Ltd: Hemel Hempstead, UK

Goecks J, Nekrutenko A, Taylor J, Galaxy Team (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol 11(8):R86. https://doi.org/10.1186/gb-2010-11-8-r86

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40(D1): D1178–D1186 https://doi.org/10.1093/nar/gkr944

Harikrishnan SL, Pucholt P, Berlin S (2015) Sequence and gene expression evolution of paralogous genes in willows. Sci Rep 5:18662. https://doi.org/10.1038/srep18662

Hefer CA, Mizrachi E, Myburg AA, Douglas CJ, Mansfield SD (2015) Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: populus trichocarpa and Eucalyptus grandis. New Phytol 206(4):1391–1405. https://doi.org/10.1111/nph.13277

Hegarty MJ, Barker GLA, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ (2008) Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in Senecio. Philos Trans R Soc B 363(1506):3055–3069. https://doi.org/10.1098/rstb.2008.0080

Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. Annu Rev Plant Biol 65:531–551. https://doi.org/10.1146/annurev-arplant-050213-035715

Jansen RC, Nap JP (2001) Genetical genomics: the added value from segregation. Trends Genet 17(7):388–391. https://doi.org/10.1016/s0168-9525(01)02310-1

Kirst M, Myburg AA, De León JPG, Kirst ME, Scott J, Sederoff RR (2004) Coordinated Genetic Regulation of Growth and Lignin Revealed by Quantitative Trait Locus Analysis of cDNA Microarray Data in an Interspecific Backcross of Eucalyptus. Plant Physiol 135(4):2368–2378. https://doi.org/10.1104/pp.103.037960

Kirst M, Basten CJ, Myburg AA, Zeng ZB, Sederoff RR (2005) Genetic Architecture of Transcript-Level Variation in Differentiating Xylem of a Eucalyptus Hybrid. Genetics 169(4):2295–2303 . https://doi.org/10.1534/genetics.104.039198

Kirst M, Yu Q (2007) Genetical genomics: successes and prospects in plants. In: Varshney RK, Tuberosa R (ed) Genomics-Assisted Crop Improvement. Vol 1: Genomics Approaches and Platforms. Dordrecht, Netherlands: Springer, pp 245–265, ISBN 9781402062940. https://doi.org/10.1007/978-1-4020-6295-7_11

Kvam VM, Liu P, Si Y (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. Am J Bot 99(2):248–256. https://doi.org/10.3732/ajb.1100340

Lê Cao, K-A., González I. and Déjean S. (2009) integrOmics: an R package to unravel relationships between two omics data sets. Bioinformatics 25(21):2855-2856. NOTE: the package 'integrOmics' has been renamed to 'mixOmics' . https://doi.org/10.1093/bioinformatics/btp515

Liu Y, Jiang Y, Lan J, Zou Y, Gao J (2014) Comparative Transcriptomic Analysis of the Response to Cold Acclimation in Eucalyptus dunnii. PLoS One 9(11):e113091. https://doi.org/10.1371/journal.pone.0113091

López de Heredia U, Vázquez-Poletti JL (2016) RNA-seq analysis in forest tree species: bioinformatic problems and solutions. Tree Genet. Genomes 12(2):30. https://doi.org/10.1007/s11295-016-0995-x

Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells MK, Jannink JL (2011) Genomic selection in plant breeding: knowledge and prospects. Adv Agron 110:77–123. https://doi.org/10.1016/b978-0-12-385531-2.00002-5

Lukens L, Downs G (2012) Bioinformatics Techniques for Understanding and Analyzing Tree Gene Expression Data. In: Schnell RJ, Priyadarshan PM (eds) Genomics of Tree Crops. Heidelberg, Germany: Springer, pp17–38. https://doi.org/10.1007/978-1-4614-0920-5_2

Makouanzi GC, (2015) Composantes de la variance phénotypique et de l'interaction GxE de la croissance et des traits écophysiologiques de l'Eucalyptus urophylla x Eucalyptus grandis. Dissertation. University of Marien Gouabi

Martin LB, Fei Z, Giovannoni JJ, Rose JK (2013) Catalyzing plant science research with RNA-seq. Front Plant Sci 4:1–10. https://doi.org/10.3389/fpls.2013.00066

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1):10–12. https://doi.org/10.14806/ej.17.1.200

McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40(10):4288–4297. https://doi.org/10.1093/nar/gks042

Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA (2010) De novo assembled expressed gene catalog of a fast-growing Eucalyptus tree produced by Illumina mRNA-Seq. BMC Genomics 11:681. https://doi.org/10.1186/1471-2164-11-681

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, Goodstein DM, Dubchak I, Poliakov A, Mizrachi E, Kullan AR, Hussey SG, Pinard D, van der Merwe K, Singh P, van Jaarsveld I, Silva-Junior OB, Togawa RC, Pappas MR, Faria DA, Sansaloni CP, Petroli CD, Yang X, Ranjan P, Tschaplinski TJ, Ye CY, Li T, Sterck L, Vanneste K, Murat F, Soler M, Clemente HS, Saidi N, Cassan-Wang H, Dunand C, Hefer CA, Bornberg-Bauer E, Kersting AR, Vining K, Amarasinghe V, Ranik M, Naithani S, Elser J, Boyd AE, Liston A, Spatafora JW, Dharmwardhana P, Raja R, Sullivan C, Romanel E, Alves-Ferreira M, Külheim C, Foley W, Carocha V, Paiva J, Kudrna D, Brommonschenkel SH, Pasquali G, Byrne M, Rigault P, Tibbits J, Spokevicius A, Jones RC, Steane DA, Vaillancourt RE, Potts BM, Joubert F, Barry K, Pappas GJ, Strauss SH, Jaiswal P, Grima-Pettenati J, Salse J, Van de Peer Y, Rokhsar D, Schmutz J (2014) The genome of Eucalyptus grandis. Nature 510(7505):356–362. https://doi.org/10.1038/nature13308

Neyman J, Pearson ES (1933) On the Problem of the Most Efficient Tests of Statistical Hypotheses. Philos Trans R Soc, A 231(1933):289–337. https://doi.org/10.1098/rsta.1933.0009

R Development Core Team (2014) R: A language and environment for statistical computing [online]. Vienna, Austria: R Foundation for Statistical Computing, to be found at <http://www.R-project.org/> [quoted 12 January 2016]

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43(7):e47. https://doi.org/10.1093/nar/gkv007

Robinson MD, McCarthy DJ, Smyth GK (2010) EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140. https://doi.org/10.1093/bioinformatics/btp616

Salazar MM, Nascimento LC, Camargo EL, Goncalves DC, Neto JL, Marques WL, Teixeira PJ, Mieczkowski P, Mondego JM, Carazzolle MF, Deckmann AC, Pereira GA (2013) Xylem transcription profiles indicate potential metabolic responses for economically relevant characteristics of Eucalyptus species. BMC Genomics 14:201–214. https://doi.org/10.1186/1471-2164-14-201

Shorack GR, Wellner JA (2009) Empirical processes with applications to statistics. Philadelphia, USA: Society for Industrial & Applied Mathematics, 956p. https://doi.org/10.1137/1.9780898719017

Soler M, Camargo ELO, Carocha V, Cassan-Wang H, San Clemente H, Savelli B, Hefer CA, Paiva JA, Myburg AA, Grima-Pettenati J (2015) The Eucalyptus grandis R2R3-MYB transcription factor family: evidence for woody growth-related evolution and function. New Phytol 206(4):1364–1377. https://doi.org/10.1111/nph.13039

Song G, Guo Z, Liu Z, Cheng Q, Qu X, Chen R, Jiang D, Liu C, Wang W, Sun Y, Zhang L, Zhu Y, Yang D (2013) Global RNA sequencing reveals that genotype-dependent allele-specific expression contributes to differential expression in rice F1 hybrids. BMC Plant Biol 13:221. https://doi.org/10.1186/1471-2229-13-221

Thavamanikumar S, Southerton S, Thumma B (2014) RNA-Seq using two populations reveals genes and alleles controlling wood traits and growth in Eucalyptus nitens. PLoS ONE 9(6):e101104. https://doi.org/10.1371/journal.pone.0101104

Vélez-Bermúdez IC, Schmidt W (2014). The conundrum of discordant protein and mRNA expression. Are plants special? Front Plant Sci 5:619. https://doi.org/10.3389/fpls.2014.00619

Vigneron P, Bouvet J-M (2001) Eucalyptus. In: Charrier A, Jacquot M, Hamon S, Nicolas D (eds) Tropical plant breeding. Montpellier: CIRAD-Science Publishers, pp 223–245, ISBN: 978-2-87614-426-2

Villar E, Klopp C, Noirot C, Novaes E, Kirst M, Plomion C, Gion JM (2011) RNA-Seq reveals genotype-specific molecular responses to water deficit in eucalyptus. BMC Genomics 12:538. https://doi.org/10.1186/1471-2164-12-538

Vining KJ, Romanel E, Jones RC, Klocko A, Alves-Ferreira M, Hefer CA, Amarasinghe V, Dharmawardhana P, Naithani S, Ranik M, Wesley-Smith J, Solomon L, Jaiswal P, Myburg AA, Strauss SH (2015) The floral transcriptome of Eucalyptus grandis. New Phytol 206(4):1406–1422. https://doi.org/10.1111/nph.13077

Yandell BS (1997) Practical Data Analysis for Designed Experiments. London, UK: Chapman & Hall, 440p, ISBN 9780412063411

Yang S, Liu Y, Jiang N, Chen J, Leach L, Luo Z, Wang M (2014) Genome-wide eQTLs and heritability for gene expression traits in unrelated individuals. BMC Genomics 15:13. https://doi.org/10.1186/1471-2164-15-13

Zhao S, Fung-Leung WP, Bittner A, Ngo K, Liu X (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One 9(1):e78644 . https://doi.org/10.1371/journal.pone.0078644