New Phytologist Supporting Information Notes S1

Comparative interrogation of the developing xylem transcriptomes of two wood-forming species: *Populus trichocarpa* and *Eucalyptus grandis*

Authors: Charles A. Hefer, Eshchar Mizrachi, Alexander A. Myburg, Carl J. Douglas and Shawn D. Mansfield

Article acceptance date: 04 December 2014

Notes S1

Between 233 million and 234 million *Eucalyptus* xylem and leaf transcriptome derived reads were used in characterizing gene expression. The *Populus* dataset consisted of 213 million xylem and 831 million young leaf reads aligned to the reference transcriptome (Table A). Gene expression values were estimated using the latest publically available genome sequences and annotated gene models of the respective species available at Phytozome (http://www.phytozome.org). Reads were aligned to the target genomes using Tophat version 2.0.8 and differential gene expression analysis performed with Cuffdiff version 2.1.1 (see the Materials and Methods section in the main paper). Reads from the same tissue were used as biological repeats, and the Fragment per kilobase per million mapped reads (FPKM) values normalized with the geometric normalization methods. The gene expression (FPKM) values between samples were distributed equally between replicates (Fig. A), with more expressed genes detected at higher levels in the xylem samples in both species.

Between tissues, FPKM values were correlated at levels between 0.98 (Pearson) between the replicated *Eucalyptus* xylem (xylem_1 and xylem_2) and leaf (leaf_1 and leaf_2) samples, and 0.95 between the unrelated *Populus* samples (Tables B, C). Overall, higher levels of FPKM correlation were observed between the *Eucalyptus* samples. As expected, the gene expression patterns were more similar between tissues than between individual trees, and clustering of the

expression patterns identified the higher similarity between the *Eucalyptus* replicate samples and the unrelated individual, compared to the unrelated poplar samples (Fig. B).

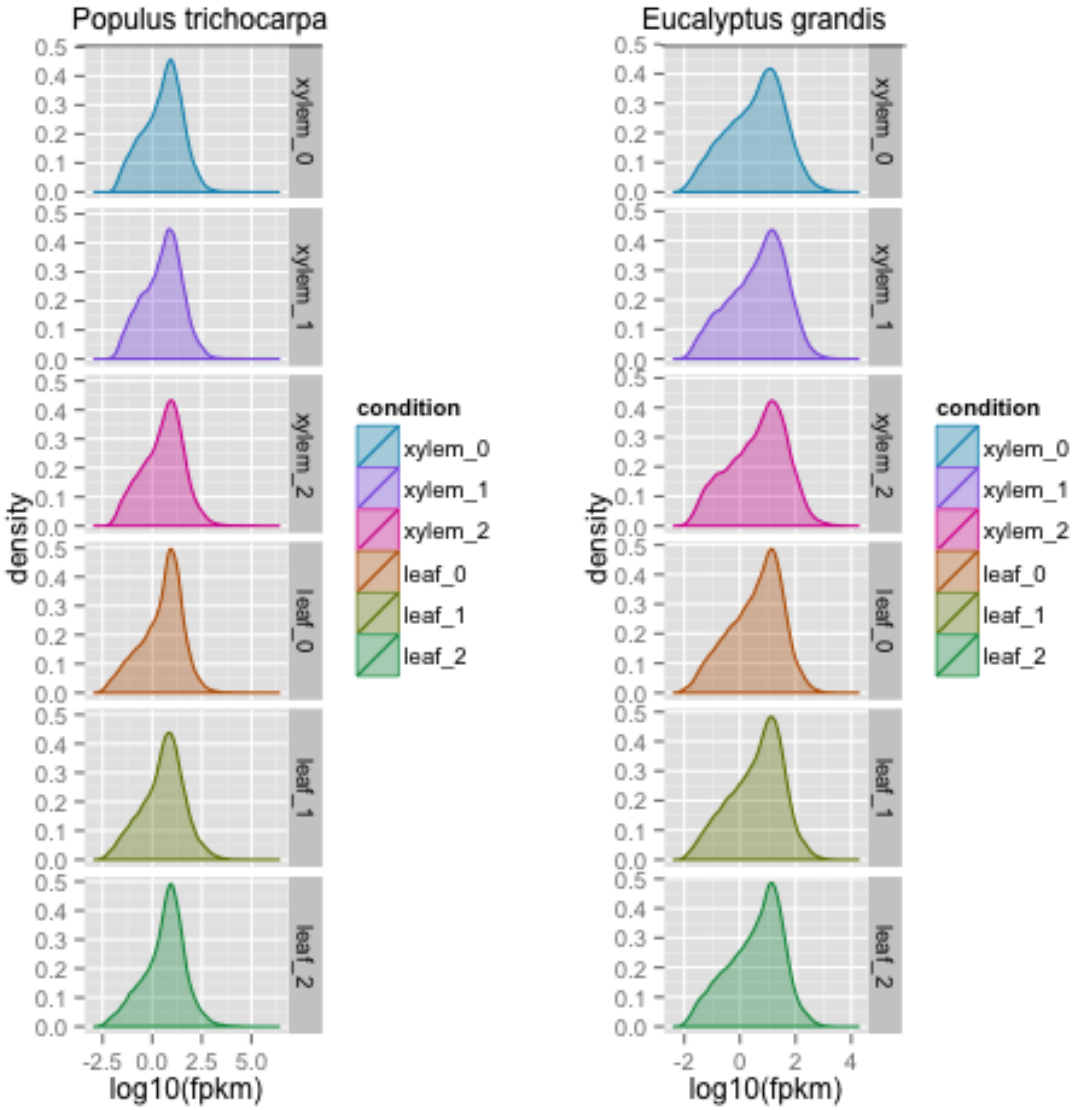

Fig. A The distribution of FPKM value for among replicated samples. Higher FPKM values (*x*-axis) were detected in leaf samples compared to the xylem samples in both species.

Table A The number of reads sequenced, and aligned to the respective reference genome for each species. The number of xylem poplar reads varied from 60-90 million reads, and the number of poplar leaf tissue reads ranged from 268-282 million reads. The eucalypt leaf and xylem datasets contained between 76 and 82 million reads.

Species	Tissue	Analysis ID	Sequenced Reads	Aligned Reads	Read length (bp)
P. trichocarpa	Developing xylem	xylem_0 (PT005)	61,051,832	60,912,577	50
		xylem_1 (PT006)	60,282,704	59,550,097	50
		xylem_2 (PT010)	94,133,182	92,724,774	50
	Young leaf	leaf_0 (PT0033)	291,794,824	282,661,729	51
		leaf_1 (PT0034)	274,305,990	268,676,385	51
		leaf_2 (PT0035)	286,661,360	280,035,480	51
E. grandis	Developing xylem	xylem_0 (EU002)	81,288,188	78,778,342	76
		xylem_1 (EU004)	81,252,238	78,716,555	76
		xylem_2 (EU006)	76,726,784	75,579,005	76
	Young leaf	leaf_0 (EU001)	77,351,452	76,234,013	76
		leaf_1 (EU003)	81,206,588	78,711,877	76
		leaf_2 (EU005)	82,424,374	79,851,862	76

Table B Pearson correlation of gene expression values (FPKM) between sequenced *Eucalyptus* samples. The highest correlation values were observed between xylem_1 and xylem_2 (0.986) and leaf_1 and leaf_2 (0.985). These samples were biological replicates of the same ramet. Lower correlation values were observed between leaf_0 and xylem_0, a genetically distinct individual and the duplicated samples.

Tissue	xylem_0	xylem_1	xylem_2	leaf_0	leaf_1	leaf_2
xylem_0	1.000					
xylem_1	0.943	1.000				
xylem_2	0.942	0.986	1.000			
leaf_0	0.784	0.779	0.774	1.000		
leaf_1	0.755	0.791	0.789	0.948	1.000	
leaf_2	0.765	0.801	0.799	0.946	0.985	1.000

Table C Pearson correlation of gene expression values (FPKM) between sequenced *Populus* samples. The highest correlation of gene expression values was observed between xylem samples, ranging from 0.948-0.963. Gene expression values in leaf tissue showed lower levels of correlation (between 0.907 and 0.919).

Sample	xylem_0	xylem_1	xylem_2	leaf_0	leaf_1	leaf_2
xylem_0	1.000					
xylem_1	0.963	1.000				
xylem_2	0.948	0.954	1.000			
leaf_0	0.723	0.735	0.762	1.000		
leaf_1	0.719	0.728	0.740	0.919	1.000	
leaf_2	0.677	0.697	0.700	0.907	0.916	1.000

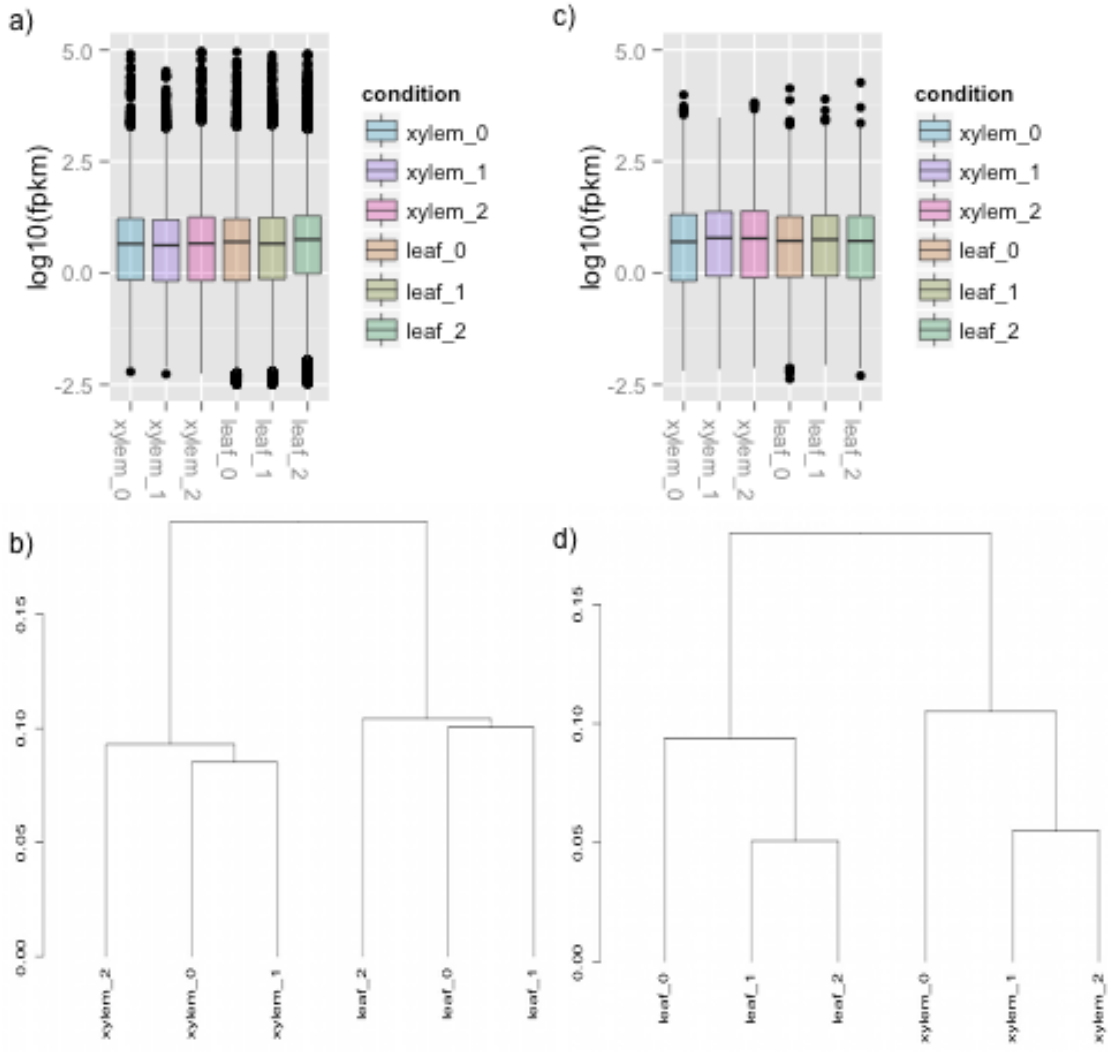

Fig. B Gene expression values were more similar between tissues than between individuals sampled. Higher expression values were reported within the leaf samples compared to the xylem samples (a and c). As expected, FPKM values were more similar between tissues than between individuals, and the similarity of the biological replicated samples in the *Eucalyptus* samples can clearly be identified (d), compared to the larger differences between *Populus* samples (b).