Article type : Regular Manuscript

Organellar carbon metabolism is co-ordinated with distinct developmental phases of secondary xylem

Desré Pinard[1,2], Ana Carolina Fierro[3,4], Kathleen Marchal[1,3,4], Alexander A. Myburg[1,2], Eshchar Mizrachi[1,2]

[1] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa; [2] Genomics Research Institute (GRI), University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa; [3] Department of Information Technology, Ghent University – iMinds, Technologiepark 15, Ghent, Belgium; [4] Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium.

Author for correspondence:

Eshchar Mizrachi

Tel: +2712 420 5879

Email: eshchar.mizrachi@fabi.up.ac.za.

ORCIDs

Desré Pinard - https://orcid.org/0000-0001-5600-0139

Kathleen Marchal - https://orcid.org/0000-0002-2169-4588

Alexander A. Myburg - https://orcid.org/0000-0003-0644-5003

Eshchar Mizrachi - https://orcid.org/0000-0002-3126-9593

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/nph.15739

Received: 3 November 2018

Accepted: 5 February 2019

Summary

- Subcellular compartmentation of plant biosynthetic pathways in the mitochondria and plastids requires coordinated regulation of nuclear encoded genes, and the role of these genes has been largely ignored by wood researchers.

- In this study, we constructed a targeted systems genetics coexpression network of xylogenesis in *Eucalyptus* using plastid and mitochondrial carbon metabolic genes and compared the resulting clusters to the aspen xylem developmental series.

- The constructed network clusters reveal the organization of transcriptional modules regulating subcellular metabolic functions in plastids and mitochondria. Overlapping genes between the plastid and mitochondrial networks implicate the common transcriptional regulation of carbon metabolism during xylem secondary growth.

- We show that the central processes of organellar carbon metabolism are distinctly coordinated across the developmental stages of wood formation, and are specifically associated with primary growth and secondary cell wall deposition. We also demonstrate that during xylogenesis, plastid targeted carbon metabolism is partially regulated by the central clock for carbon allocation towards primary and secondary xylem growth, and discuss these networks in the context of previously established associations with wood-related complex traits. This study provides a new resolution into the integration and transcriptional regulation of plastid and mitochondrial localized carbon metabolism during xylogenesis.

Key words

Plastid, mitochondria, xylogenesis, circadian clock, co-expression network, *Eucalyptus*

Introduction

An aspect of wood formation in trees, and secondary xylem growth in general that has not received much attention is the role of plastids and mitochondria, and the subcellular coordination of metabolism in these genome-containing organelles. The majority of plastid and mitochondrial protein coding genes are encoded by the nuclear genome, and are targeted to the organelles after translation, where they are imported by translocase complexes (Wiedemann & Pfanner, 2017; Ling & Jarvis, 2015). During cambial differentiation and primary growth, plastidial *de novo* fatty acid synthesis contributes to cell division and elongation through the supply of phospholipid membrane precursors (Ohlrogge & Browse, 1995). The plastid localized methylerythritol phosphate (MEP) pathway provides precursors for plant hormones gibberellic acid, abscisic acid, cytokinins, brassinosteroids and strigolactones, many of which are involved in cambial differentiation to xylem and phloem, and secondary growth (Banerjee & Sharkey, 2014; Mellerowicz *et al*., 2001). The plastidial shikimate pathway produces tryptophan for the synthesis of auxin (indole-3-acetic acid), which is a primary hormonal regulator of xylem development (Ursache *et al*., 2014). Phenylalanine, also synthesized via the plastidial shikimate pathway, is transported to the cytoplasm/endoplasmic reticulum to be metabolized to cinnamic acid for large-scale monolignol synthesis during secondary growth (Herrmann, 1995), accounting for up to 40% of plant biomass (Tohge *et al*., 2013). The shikimate pathway precursors phosphoenolpyruvate (PEP), and erythrose-4-phosphate (E4P) are derived from sugars that can be used for the synthesis of glucose-6-phosphate (Jensen, 1986), the main sugar phosphate of cellulose synthesis (Taylor, 2008). The shikimate pathway thus represents a central and irreversible carbon flux point between polysaccharide and phenolic biopolymer synthesis during secondary cell wall formation of xylogenesis. Similarly, mitochondrial localized energy metabolism is crucial to xylogenesis (Jacoby *et al*., 2012), as is their involvement in programmed cell death (Van Aken and Van Breusegem 2015; Yu *et al*., 2002), highlighting the importance of these organelles to the process of wood formation.

Currently, state of the art co-expression studies in wood development in trees are those representing expression variation across the spatial development of xylogenesis in aspen (*Populus tremula*) and *Picea abies*. These studies have highlighted the conserved regulation of cambial differentiation into xylem and phloem, secondary cell wall (SCW) deposition, and programmed cell death (Sundell *et al*., 2017; Jokipii-Lukkari *et al*., 2017). With regards to co-expression studies looking at organellar-targeted biology, previous studies have used global, condition independent co-expression networks (Serin *et al*., 2016) to show that plastid genes related to photosynthesis,

abiotic and biotic stress, and circadian rhythm in *Arabidopsis* are coregulated, while mitochondrial-targeted genes are regulated with diverse processes in the cell (Mentzen & Wurtele 2008; Lundquist *et al*., 2012; Majsec *et al*., 2017). These studies have been predominantly performed in photosynthetic *Arabidopsis* tissues containing differentiated chloroplasts. Further, condition independent co-expression analyses are known to miss tissue or condition specific interactions (Serin *et al*., 2016). Therefore, given the importance of plastid and mitochondrial-targeted primary carbon metabolism to xylogenesis, a targeted co-expression network approach would give deeper insight into the roles of plastids and mitochondria during xylogenesis.

A systems genetics analysis of a *Eucalyptus grandis × E. urophylla* F2 (GUxU) interspecific backcross population has shown that variation in expression of primary carbon metabolism genes is significantly associated with wood development and secondary cell wall trait variation (Mizrachi *et al*., 2017). Here, we constructed xylem-specific gene co-expression modules of plastid and mitochondrial carbon metabolism using transcriptomes from this population (Mizrachi *et al*., 2017). Specifically, we asked how wood-forming cells regulate subcellular carbon partitioning in these organelles; to what extent is organellar regulation shared between plastids and mitochondria; and what important biological and biochemical pathways are organelle-specific carbon metabolism linked to? We find that circadian, developmental, and epigenetic regulation of xylogenesis is linked to plastidial carbon metabolism, and that mitochondrial and plastid networks integrate primary metabolism and cellular homeostasis to secondary cell wall formation. Given the importance of these organelles in carbon metabolism for wood development, and the dearth of specific knowledge of organellar biology during xylogenesis, this study provides new insight into the subcellular and temporal regulation of carbon allocation in trees. An understanding of the role of organelles in carbon allocation during secondary growth, and the genetically permissible expression variation of carbon metabolism is of value to tree, wood, and secondary cell wall researchers.

Materials and Methods

Subcellular localization and KEGG pathway prediction

Eucalyptus grandis v.2.0. annotation homologs of *Arabidopsis thaliana* (TAIR10) genes were downloaded from Phytozome v10 (http://phytozome.jgi.doe.gov/pz/portal.html) to infer subcellular localization and KEGG metabolic pathways of *E. grandis* genes. The KEGG database (Kanehisa *et al*., 2008) was used to identify *Eucalyptus* gene identifiers (IDs) from their *A. thaliana* best hits involved in 17 selected carbon metabolic KEGG pathways (Note S1). The SUBA database (Tanz *et al*., 2013) of *Arabidopsis* consensus gene localizations (SUBAcon) was used to assign *Eucalyptus* genes to eleven different subcellular localizations, with an added category of "multiple" targeted genes, if genes were predicted to be localized in more than one location. The number of genes per subcellular compartment was counted for all genes, xylem-expressed genes (FPKM >1 in at least 75% of individuals), and KEGG carbon metabolic genes (Notes **S1**).

Query-based co-expression network construction and graph-based clustering

The data generation of the population-wide transcriptomes used in this study is well described in (Mizrachi *et al*., 2017). Briefly, immature xylem was collected from 156 GUxU F2 interspecific backcross 3-year-old individuals between 09:00 and 16:30 by bark removal and scraping of the inner glutinous layer. The samples were flash frozen in liquid nitrogen and stored at -80□. The samples were then finely ground using a mortar and pestle before shipping to the Beijing Genomics Institute (BGI) for polyA selected mRNA sequencing. Data analysis was performed using the Tuxedo suite, resulting in transcript abundance measurements (Fragments Per Kilobase of transcript per Million - FPKM) for all 36 349 *E. grandis* genes.

To identify the co-expression network of carbon metabolism localized to the plastid and mitochondria during xylogenesis, 152 plastid-targeted genes and 65 mitochondrial-targeted genes from 17 KEGG primary carbon metabolic pathways (defined above) were used as query lists for network generation. First, the query gene to all gene Pearson correlation matrix was constructed and was then filtered by joint likelihood score to ensure reciprocal co-expression between a query gene and its associated network genes. A joint likelihood score for a query gene and a selected cluster gene depends on the distribution of correlation values for a query gene, and on the distribution of correlation values for the selected gene against all other genes. Thus, the joint

likelihood score is $f(Z_i,Z_j = \sqrt{Z_i^2 + Z_j^2}$, where Z_i and Z_j are the z-scores of the correlation from the marginal distributions for query gene *i* and selected gene *j* (Faith *et al*., 2007). To avoid clusters of genes that are specific to only one query, the co-expression network was filtered to keep genes correlated to at least two query genes, and query genes were retained if they were correlated to at least fifty other genes. Secondly, a weighted co-expression network was generated in which edges represent correlations between queries and their associated network genes. The multilevel community detection method was used to sub-divide the network into consensus clusters based on the number and weights of shared edges within clusters, where clusters have more shared edges with higher weights (Djidjev, 2006). A third network of overlapping query genes that were found in the resulting plastid and mitochondrial networks was constructed as above. For more details, see Fig. **S1**. Tissue specific gene expression of *E. grandis* v2.0 genes was retrieved from previous publications (Mizrachi *et al*., 2010; Myburg *et* al., 2014; Vining *et al*., 2015) to analyse the tissue specific expression profiles of plastid and mitochondrial cluster genes.

Overlap of *E. grandis* plastid and mitochondrial cluster genes with the aspen developmental transcriptome series and genes significantly associated with lignocellulosic biomass traits

In order to place the plastid and mitochondrial cluster genes in a developmental context, the number of genes which overlap with the defined developmental phases (a to h) in aspen (*Populus tremula*) were counted (Sundell *et al*., 2017). First, the *E. grandis* orthologs of *Populus trichocarpa* genes were identified using the PLAZA database by downloading all orthologous gene families (PLAZA file = genefamily_data.orth.csv) and cross-referencing the gene family IDs between *E. grandis* and *P. trichocarpa* using the merge function in R (version 3.2.2) (Proost *et al*., 2014). Then, overlapping genes per plastid network cluster based on *E. grandis* gene IDs were identified using the merge function in R (version 3.2.2). For the final visualization and counting of the number of overlaps in each cluster and developmental series, any genes that had duplicated *E. grandis* gene IDs within the same overlap were filtered. This was done to standardize the number of genes for the next analysis, GO biological process (GO-BP) enrichment. Additionally, to identify more distantly related homologs, the procedure for identifying orthologs was repeated using the broader homolog gene families (PLAZA file = genefamily_data.hom.csv). In order to reduce the complexity of the comparisons stemming from large duplicated homologous gene families in *E. grandis* and *P. trichocarpa*, only orthologs were used in the overlap and developmental series analysis. The overlapping genes were visualized in Tableau (Tableau Software). The developmental series

expression profiles for clusters with large overlaps were visualized using ggplot2 (version 3.1.0 (Wickham, 2009)) in R (version 3.2.2) using the expression values of the aspen developmental series taken from Sundell *et al*., 2017.

Plastid cluster genes that were significantly associated with 13 lignocellulosic biomass traits were identified by downloading the curated list of 1 597 *E. grandis* genes and cross-referencing the list with the plastid cluster genes (Mizrachi *et al*., 2017). The number of plastid cluster genes associated with each trait was visualized using Tableau (Tableau Software).

Functional enrichment analysis

Functional annotation of consensus clusters was done using GO and pathway enrichment (). Gene annotation for GO-BPs were obtained from TAIR10 (www.arabidopsis.org). The GO gene annotation was transferred from *Arabidopsis* to their corresponding *E. grandis* genes as determined above (Myburg *et al*., 2014). Functional enrichment of was calculated using Fisher's exact test, and P-values were corrected for multiple testing using the Benjamini and Hochberg method (Benjamini & Hochberg, 1995). A corrected P-value of < 0.05 was used as a threshold to select enriched terms or pathways. The enrichment analysis was repeated without the query genes to determine which GO-BP terms were enriched due to the presence of the query genes. This was done to ensure that the query genes were not biasing the GO-BP term enrichment for the clusters. GO-BP terms that are not enriched without the query genes are indicated with * in Notes **S2.** Shared plastid and mitochondrial genes were used for an additional enrichment per overlapping gene set (Fig. **2 (see later)** ; Notes **S2**). GO-BP enrichment of plastid clusters with the *P. tremula* developmental phases was done as above, using *E. grandis* gene IDs, and *E. grandis* as the background genome for enrichment.

Genes that were present in both plastid and mitochondrial clusters were identified, and the results were visualized using a Sankey diagram generated with Google Developers charts (Zhu, 2012). The MapMan annotation for each gene in the networks were sourced from the PLAZA database for plant comparative genomics (Proost *et al*., 2014), along with the subcellular localization of cluster genes previously retrieved from SUBAcon (Tanz *et al*., 2013). MapMan was used to assess the known functions of the genes within each cluster in order to interpret the biological relevance of the networks (Usadel *et al*., 2009). Network properties and architecture were evaluated using Cytoscape

and NetworkAnalyzer (Shannon *et al*., 2003), after exporting the network adjacency matrix using the igraph package (Csardi & Nepusz, 2006). Unless otherwise stated, all data processing and analysis was done using R (version 3.2.2).

Results

Subcellular localization and expression of plastid and mitochondrial-targeted genes in *Eucalyptus*

To obtain a baseline understanding of the subcellular localization of primary carbon metabolism in *Eucalyptus*, the predicted subcellular localization (SUBAcon) and biochemical functions (KEGG) of annotated *Arabidopsis* homologs for all *E. grandis v2* proteins were examined (Myburg *et al*. 2014; Tanz *et al*., 2013). Of the 36 349 annotated *E. grandis* proteins, 7,6% and 5,5% are predicted to be targeted to the plastid and mitochondria, respectively (Table **1**). 747 genes in 17 KEGG metabolic pathways representing primary carbon metabolism (Aharoni & Galili, 2011) are expressed in the immature xylem of the GUxU interspecific backcross population (FPKM > 1 in > 75% of individuals) (Table **1**). The predicted localization of these genes shows that the majority of primary carbon metabolism in xylem is attributed to cytosolic proteins (33%), with ~20% localizing to the plastid and ~8% to the mitochondria (Table **1**). Thus, almost a third of all KEGG annotated enzymatic reactions relating to primary carbon metabolism are predicted to occur in the plastid and mitochondria during xylogenesis.

Co-expression networks of plastid and mitochondrial carbon metabolic genes

The immature xylem RNA-sequencing data generated in Mizrachi *et al*., 2017 was sampled by removing the bark and scraping the xylem tissue, a method which inherently samples all cell types and the spectrum of development from the cambial layer to the later stages of xylem before programmed cell death (PCD). This outer glutinous layer comprises of developing xylem compared to mature xylem tissue, which predominantly consists of cells that have already undergone PCD (see (Mizrachi *et al*., 2010) for more detail on sampling). In order to determine the biological pathways that are coordinated with carbon metabolism in plastids and mitochondria during xylogenesis in these samples, the identified organelle-targeted carbon metabolic genes were used as query genes for co-expression network construction (Magwene & Kim, 2004). Briefly, all genes that were expressed in at least 75% of the individuals in the population were correlated to the plastid and

mitochondrial query genes. Genes that had a Pearson correlation of > 0.6 to at least two query genes were retained. In order to minimize biological noise stemming from unconnected or lowly expressed genes, query genes were retained if they were correlated to at least fifty other genes, and had a joint likelihood of reciprocal correlation threshold of 3.8 based on row Z-score for the final network construction (See Materials and methods and Fig. **S1**). The multilevel community detection algorithm was used to define co-expression clusters, using weighted shared edges to identify clusters (Djidjev, 2006). The constructed plastid query gene network for the population contained 47 query genes, and 2,500 co-expressed genes (hereafter referred to as cluster genes) in seven clusters, representing 11.8% of all xylem-expressed genes (Fig. **1a, b**). The mitochondrial query gene network contained 23 query genes, and 2,194 cluster genes in five clusters, representing 10.3% of all xylem-expressed genes (Fig. **1 c, d**).

For the seven Plastid Clusters (PC1 to PC7) and five Mitochondrial Clusters (MC1 to MC5), GO biological process (GO-BP) term enrichment analysis was used to identify the distinct biological processes associated with carbon metabolism targeted to plastids and mitochondria during *Eucalyptus* xylogenesis (Fig. **1b, d**; Notes **S1**). Between these two networks, there is a 36% overlap of shared genes (Fig. **2a, b**), and clusters with large overlaps were assigned the same colour for comparison. The majority of the overlapping genes were between PC1 and MC4 (red) (126 genes), PC3 (green) and MC3 (dark green) and MC5 (light green) (265 and 358 genes, respectively), and PC7 and MC2 (138 genes) (Fig. **2a**). The clusters with little to no overlap are potential plastid and mitochondria specific clusters, and those with large overlaps and GO enriched terms in common represent biological processes that are transcriptionally coordinated between organelles (Fig. **2c**). A network of the shared query genes between the plastid and mitochondrial networks showed that the overlap of PC3 - MC3 and MC5 and PC7 - MC2 is highly coordinated in the population, while the overlap of PC1 - MC4 is likely due to organelle specific regulation (discussed below) (Figure **S2**). All genes and cluster membership can be found in Notes **S3.**

Distinct regulatory modules of early and late xylem development in plastid network clusters

Transcription during xylem formation is marked by three main developmental phase transitions: cambial meristem differentiation and primary growth, followed by the later stage of secondary cell wall biosynthesis and deposition, and finally programmed cell death to form fully functional xylem fiber cells. In the plastid network, primary growth and secondary cell wall synthesis are separated

into two distinct clusters, PC1 (early) and PC3 (late). PC3 and mitochondrial clusters MC3 and MC5 have the largest overlap of 687 shared genes between plastid and mitochondrial clusters, with shared enriched GO-BP terms for xylem development and secondary cell wall biopolymer synthesis, and no unique GO-BP terms (Fig. **2c;** Notes **S2**). Further, the shared secondary cell wall genes are highly and specifically expressed in immature xylem compared to five other *E. grandis* tissue specific transcriptomic datasets (Fig. **2b**) (Vining *et al*., 2015; Mizrachi *et al*., 2010). The large overlap of xylem specific genes between the plastid and mitochondrial networks highlight the strength of xylem as a carbon sink during wood formation, where the utilization of metabolite precursors must be shared between cellular compartments to ensure efficient SCW biosynthesis.

In PC3, nine out of 18 query genes are aromatic amino acid (AAA) synthesis genes, and eight out of nine query genes for MC3 and MC5 are serine/glycine metabolic genes, showing that amino acid metabolism, and specifically AAA synthesis is highly co-regulated with SCW biosynthesis. In addition to all known *E. grandis* SCW *CELLULOSE SYNTHASE* (*CESA*) genes and all but one of the monolignol biosynthetic genes, PC3, MC3, and MC5 include known SCW formation regulators in the MYB and NAC families (Hussey *et al*., 2013) (Notes **S3**; Table **S1**). Out of 174 putative monolignol genes in the *E. grandis* genome, 18 directly involved with SCW formation are included in the networks, showing the specificity of the analysis (Carocha *et al*., 2015). The shikimate pathway for monolignol precursor phenylalanine is almost completely represented in PC3, highlighting the intercellular coordination of monolignol synthesis between the plastid and cytosol (Table **S1**). Additionally, the mitochondrial genes involved in the metabolism of glycine in MC3 and MC5 are proposed to modulate S-adenosylmethionine levels for lignin accumulation during SCW formation, showing that lignin synthesis is a major driver of plastid and mitochondrial inter-organellar coordination in xylem (Villalobos *et al*., 2012).

A key element in the coordination of cellular metabolism during xylogenesis is the inter- and intra-cellular transport of metabolites (Linka & Weber, 2010). In PC3, we identified 52 predicted transporters, of which 33 are predicted to be localized to the plasma membrane, including *SUCROSE TRANSPORTER (SUT) 1* and *3* genes, linking sucrose transport from the leaves via phloem tissue to SCW formation in xylem (Kühn & Grof, 2010) (Notes **S3**). Also present in PC3, are two known plastid localized transporters: *PPT/CUE1*, which transports shikimate pathway precursor phosphoenolpyruvate (PEP) or triose phosphate (TP) into the plastid in heterotrophic tissues; and

inorganic phosphate transporter *PHT4;2* (Staehr *et al*., 2014; Streatfield *et al*., 1999; Irigoyen *et al*., 2011) (Notes **S3**). These plastid localized transporters may represent key nodes for mediating metabolic flux during SCW biosynthesis and deposition.

PC1 (557 genes) and MC4 (219 genes), which have an overlap of 126 genes (Fig. **2a**), have enriched GO-BP terms related to cell identity, elongation and primary cell wall deposition (red in Fig. **1**; Fig. **2**), with all three *E. grandis* primary cell wall (PCW) *CESA* genes present (Notes **S2**; Notes **S3)** (Myburg *et al*., 2014). Lipid biosynthetic genes were the primary plastid query genes in this cluster, highlighting the requirement for phospholipid membranes in dividing cells (Kwok & Wong, 2005), while MC4 query genes were a combination of genes from the tricarboxylic acid (TCA) cycle and the photorespiratory pathway (Notes **S3**). In the early phases of xylogenesis, auxin concentration gradients in the cambium are known to be a primary signal for xylem fate specification (Milhinhos & Miguel, 2013). PC1 contains an *Arabidopsis PIN-FORMED1 (PIN1)* auxin transporter homolog, AUXIN RESPONSE FACTOR (ARF) gene *ARF7/ARF19*, and *SUPPRESSOR OF AUXIN RESISTANCE 1* and *3*. Related to cambial auxin distribution is *SLK2* (PC1, MC4), which has been shown to regulate auxin distribution in *Arabidopsis* shoot apical meristems (Lee *et al*., 2014). In PC1, HD-ZIPIII transcription factor *REVOLUTA* functions in response to auxin to induce xylem fiber cell formation (Talbert *et al*., 1995). Other regulators of xylogenesis found in PC1 include *LHW*, which regulates procambial cell division (reviewed in De Rybel *et al*., 2015). PC1 and MC4 therefore represent the structural metabolic genes and transcription factors regulating the early phases of xylem development in *E. grandis*.

GO-BP terms that are enriched in PC1 but not MC4 are those representing regulatory processes associated with epigenetic modifications and non-coding RNA (ncRNA). PC1 contains genes for DNA methylation maintenance methyltransferases *MET1* and *CMT3*, which act to maintain CG and CHG DNA methylation for transcriptional gene silencing at the DNA level (Cao *et al*., 2003). The DNA methylation small interfering RNA (siRNA) pathway in plants is mediated by RDR proteins, of which *RDR1* and *RDR2* genes are both found in PC1 (Yang *et al*., 2016). There are multiple chromatin remodeling proteins in PC1, such as *DDM1, CHR5, CHR11,* and *CHR42*, along with histone methyltransferases, particularly for Histone H3 lysine 4 (H3K4) (Lippman *et al*., 2004; Shen *et al*., 2015; Huanca-Mamani *et al*., 2005; Narlikar *et al*., 2013). *SDG8* histone H3 methyltransferase (PC1) regulates the expression of brassinosteroid signaling, light response, carbon metabolic genes (Li *et*

al., 2015). *SDG8* has previously been shown to affect flowering time, shoot branching, and carotenoid composition in *Arabidopsis*, and is linked here to the hormonal regulation of early xylem development (Wang *et al*., 2014). Post-transcriptional gene silencing is a widespread regulatory mechanism in plants that is mediated by non-coding RNAs. The genes *AGO1* and *AGO4* in PC1, in conjunction with the *DICER-LIKE 2*, produce 21 nt long tasi-RNAs and 24 nt long siRNAs (Wang *et al*., 2011), which have previously been shown to regulate tissue-specific gene expression in plants. The overlap between PC1 and MC4 is not conserved in the network of shared genes (Fig. **S2**) and may reflect the specific association of plastid query genes with epigenetic regulation that indirectly regulates mitochondrial targeted metabolism. The implication that plastids are more closely associated with chromatin architecture, remodeling, and non-coding RNA regulation during xylogenesis than mitochondria, makes the 54 genes of unknown function in PC1 particularly attractive for further study (Notes **S3**).

Plastid localized carbon metabolism is regulated by the circadian rhythm to ensure efficient carbon allocation during xylogenesis

The immature xylem samples used in this study were collected from early morning to late afternoon, resulting in the time-dependent perturbation of some *Eucalyptus* genes, which ended up in distinct clusters (Table **S2**). The plastid network clusters revealed the circadian control of gene expression and primary metabolite utilization; PC6 (366 genes, dark blue), and PC7 (581 genes, purple) (Fig. **1a, b**) contain genes of evening/afternoon and morning circadian clock module respectively (Fig. **3a;** Table **S2**). These clusters therefore integrate the circadian clock with starch metabolism, energy production, fatty acid (FA) biosynthesis, and protein and nucleotide metabolism, the primary carbon metabolic processes necessary for xylogenesis (Fig. **1b**, Table **S2**). PC6 contains homologs of the evening clock module in *Arabidopsis*, including the central clock loop gene *TOC1*, and evening loop genes *ELF-4a*, *ELF3* and *LUX*, and little overlap with any mitochondrial clusters (Fig. **2a;** Table **S2**) (Müller *et al*., 2014). *TET2* and *TET3* genes in PC6 affect many biological processes, and have gene family members that interact with *TOC1*, which are relatively understudied in plants (Wang *et al*., 2014; Reimann *et al*., 2017). Circadian fluctuations of starch metabolism are well studied, and the plastid network identified starch biosynthetic genes associated with the evening clock genes in PC6 (ADP-glucose pyrophosphorylase and Starch synthase 3) (Ball *et al*., 1998).

PC7 contains genes associated with the morning module of the circadian clock, along with enriched GO-BP terms for central metabolite production, including nucleotides, FAs, UDP-sugars, acetyl-CoA and glycolysis, which overlap with MC2 (Fig. **2a, c**; Notes **S2**). The shared network shows that the driver of this overlap is glycolysis, translation, and nucleotide synthesis (Notes **S3**), while the circadian genes are specific to PC7. The central morning clock gene *LHY* (*A. thaliana CCA1*), and several morning clock associated genes from the *LNK* and *RVE* gene families are found in PC7 (Table **S2**). PC7 therefore represents morning-associated energy homeostasis linked to the mitochondrial TCA cycle and production of primary metabolites, along with the degradation of transitory starch by *beta amylase-1* (*BAM1*). There are several potential transcriptional regulators in PC7 that act in conjunction with the primary circadian regulators to ensure diel metabolic regulation. Amongst these is *MYB3*, which has been shown to act as a negative regulator of phenylpropanoid biosynthesis in *Arabidopsis* through interaction with *LNK1* and *LNK2* in PC7 (Zhou *et al*., 2017). Phenylpropanoid biosynthesis is known to peak before dawn (Harmer *et al*., 2000), and these results show that this mechanism of morning repression of phenylpropanoid biosynthesis is conserved in secondary xylem. Interestingly, the genes of the plastid localized FA synthesis pathway were split between PC1 and PC7, with the majority of the initial synthesis pathway represented in PC7 (Fig. **3b**). FA biosynthesis and the shikimate pathway share a metabolic intermediate in phosphoenolpyruvate (PEP), which can be used to generate pyruvate for FA synthesis by plastidial pyruvate kinase (also in PC7) (Nishida, 2004). The distinct cluster membership of the shikimate pathway (PC3), FA synthesis and circadian regulators (PC7) suggest that FA biosynthesis and the shikimate pathway are temporally regulated to ensure efficient allocation of PEP during xylogenesis for primary growth and lignin formation (Fig. **3b**).

This significance of the relationship between FA synthesis and the shikimate pathway is illustrated by plastid and mitochondrial network genes associated with lignocellulosic biomass traits in the GUxU interspecific backcross population (Mizrachi *et al*., 2017). The network-based data integration method (NBDI) used a combined analysis of transcriptome, eQTL, metabolite, and metabolic pathway data to identify a highly curated list of *Eucalyptus* genes that are significantly associated with 13 biomass traits (Mizrachi *et al*., 2017). Of the 1 597 NBDI genes, 35% and 30% are found the plastid and mitochondrial networks respectively, with the vast majority of genes associated with traits found in PC3 - MC3 and MC5 (Fig. **4**). The clear exception is PC7 genes associated with lignin content (45 genes), which includes the nuclear encoded plastidial acetyl-CoA carboxylase 1 (*Eucgr.B01425*) and PEP carboxykinase 1 (*Eucgr.I00628*) (Fig. **4**; Notes **S3**). Therefore, despite PC7

genes not being as highly or specifically expressed in xylem as PC3 genes (Fig. **2b**; Fig. **S3a**), the temporal regulation of carbon allocation plays a significant role in *Eucalyptus* wood formation.

Plastid and mitochondrial network genes overlap with the *Populus tremula* secondary xylem developmental series

The *Populus tremula* (aspen) xylem developmental series provides high resolution transcriptomic profiles of the three main transcriptional reprogramming events of wood formation (Fig. **S4a**). The three transcriptomic transitions are further separated into 16 smaller clusters (a - h), based on the peak of transcription over the developmental series (see Fig. **S4a** for representative expression profiles across aspen clusters). The plastid network cluster genes can be placed in a developmental context using the Aspen xylem developmental series, in addition to the underlying genetic and circadian rhythm dependent gene expression variation across the population (Sundell *et al*., 2017). Of the 3 423 *E. grandis* genes in the plastid and mitochondrial networks, 2 684 had orthologous *P. trichocarpa* gene IDs representing 5 225 unique aspen genes (Notes **S3**). Of the 739 genes in *E. grandis* clusters with no Aspen cluster overlap, 490 have PLAZA v3 gene family homologs in *P. trichocarpa*, and 20 are unique to *E. grandis* (Notes **S3**). All 20 have no annotated PFAM domains, and one is highly and specifically expressed in *E. grandis* immature xylem compared to five other tissues and may be a good candidate for experimental analysis in the future (*Eucgr.C00653* - PC3 & MC5) (Fig. S3).

Overlapping genes between aspen developmental stage clusters and *E. grandis* plastid networks show that the majority of overlap is between PC1, PC3, and PC7 with the corresponding developmental clusters in aspen (e1, e2, f, g1, g2) (Fig. **5**; Fig. **S4**). The expression profiles of the overlapping genes compared to all other aspen cluster genes shows that the PC genes are representative of the average profile per aspen cluster (Fig. **5)**. To identify the shared biological processes between the clusters and developmental stages, GO BP enrichment was performed using the overlapping genes with *E. grandis* as a genomic background, after removing duplicate genes within a specific overlap. The results show that xylem and SCW specific terms are enriched between PC3 and developmental stages g1 and g2 (Notes **S2**). In the overlap between PC1 and developmental stage f, terms related to meristem initiation and maintenance and epigenetic regulation are found, showing that these processes are highly conserved between the two species (Notes **S2**). Interestingly, there are several genes related to xylem PCD in PC3, such as *bifunctional endonuclease*

(*BFN1*), and *xylem cysteine peptidase 1* (*XCP1*) and *metacaspase 9* (*MC9*) (Farage-Barhom *et al*., 2008; Van Hautegem *et al*., 2015). In the aspen clusters, these genes peak in late xylogenesis, but are expressed at a lower level across the series (clusters d1 and d2 - Fig. **S4**). This suggests that although PCD may occur specifically in mature xylem, the expression of these genes is coordinated in the earlier stages of secondary growth.

The aspen wood samples were collected in between 10:00 and 12:00 in the morning (Personal communication – T. Hvidsten & N. Street), and accordingly, PC6 (afternoon/evening module) had no significant GO BP terms in any of the overlapping genes, while PC7 (morning module) shared significant GO BP terms with early developmental phases e1 and f (Fig. **5**; Notes **S2**). These terms are related to purine and pyrimidine metabolism, translation, cell wall organization and cell wall modification (Notes **S2**). Together, these results provide further evidence that central metabolism for nucleotide synthesis and translation are indeed taking place in the morning, during the initial phases of xylogenesis. The clock genes represent the distinct resolution of this biology in the plastid-network, as the clock genes are found in PC6 and 7 but are spread across the developmental phases in the aspen dataset (Table **S2**).

This comparison further highlights the utility of these networks in understanding primary metabolism during wood formation, as well as providing new context to genes whose expression has less structure in developmental networks due to temporal variation in gene expression.

Discussion

Although many of the metabolic, hormonal and transcriptional components of xylogenesis have been identified (Ruprecht *et al*., 2011; Mizrachi *et al*., 2017; Milhinhos & Miguel 2013), their integration and system-level regulatory architecture across the bioenergetic organelles (plastids and mitochondria) is not very well understood. In this study, xylem transcriptomes from 156 GUxU were used to identify co-expression networks and network clusters associated with plastid and mitochondrial carbon metabolism. The co-expression networks in this study are a direct result of transcript abundance variation due to genetic variation between individuals, and thus represent the underlying genetic regulation in xylem tissue (Mizrachi & Myburg 2016). These networks show that plastid and mitochondrial carbon metabolic pathway genes are central to, and co-expressed with, the distinct developmental processes of xylogenesis (Fig. **1**, Fig. **5**). The results show that these

transcriptional regulatory events are genetically "hard-wired" (Mizrachi & Myburg 2016) and correspond well to the spatial developmental phases of xylogenesis (Fig. **5**) (Sundell *et al*., 2017). Specifically, the plastid-targeted carbon metabolic network clusters provide insight into the subcellular and diurnal coordination of metabolic pathways during xylogenesis (Fig. **3**). Genes present in both plastid and mitochondrial network clusters highlight the regulation of plastid and mitochondrial carbon utilization during secondary cell wall biosynthesis and their impact on lignocellulosic biomass traits (Fig. **4**). The combined networks represent only ~16% of *Eucalyptus* detectable xylem-expressed genes, making these networks a fine-scale representation of the biological processes and genes important for secondary xylem development.

The plastid network clusters clearly discriminate the programs of cell differentiation, expansion and elongation during early xylogenesis, and the SCW biosynthetic processes in PC1 and PC3, respectively. PC1 is enriched in GO terms related to epigenetic modifications such as non-coding RNAs, histone remodeling, and DNA methylation, which are not found in the mitochondrial network. Epigenetic regulation during initial xylem development in *Eucalyptus* is linked to plastidial carbon metabolism via histone modifying genes such as *SDG8*, and its regulation of strigolactone, which in conjunction with auxin signaling, regulates secondary cambial growth (Li *et al*., 2015; Cazzonelli *et al*., 2009; Agusti *et al*., 2011). Epigenetic modifications and early xylogenesis may also be directly connected by plastid localized metabolites such as acetyl-CoA and S-adenosylmethionine (SAM), that are involved in primary metabolism, and histone and chromatin modifications (Chen *et al.,* 2016; Oliver *et al*., 2009; Bouvier *et al*., 2006). Further, plastidial PEP can be used for a variety of metabolic pathways, and PEP importer *CUE1* has been previously associated with transcriptional gene silencing via histone modifications (Shen *et al*., 2009). The division of fatty acid biosynthesis (PC1, PC7) and the shikimate pathway (PC3) may be a regulatory mechanism for ensuring that PEP utilization is balanced with the carbon requirements of primary growth and secondary cell wall phenylpropanoids during secondary growth (Fig. **3**) (Joyard *et al*. 2010; Staehr *et al*. 2014). In *Arabidopsis* seedlings, *ACCase* mutations in the first committed step of fatty acid biosynthesis increased flux towards protein synthesis, suggesting a regulatory mechanism balancing the flux of carbon skeletons in plastids (Chen *et al*., 2009). Although plastidial signaling has been shown to affect epigenetic modifications and cause widespread changes in nuclear gene regulatory networks in photosynthetic tissues (Virdi *et al*., 2016; Beltrán *et al*., 2018), these results show that the plastid-specific association of epigenetic regulatory mechanisms and primary development of xylogenesis warrant further attention.

In photosynthetic tissues, the circadian clock is entrained by light and temperature, resulting in the diurnal fluctuation of ~30% of *Arabidopsis* transcripts (Covington *et al*., 2008). Tissue-specific circadian clocks are present in plants, and these clocks are decentralized, and fluctuate with different cues (Shimizu *et al*., 2015). In non-photosynthetic tissues, the clock is entrained in the absence of environmental cues by the availability of sucrose transported from photosynthetically active source tissues, also known as "metabolic dawn" (Endo *et al*., 2014; Haydon *et al*., 2013). There is a strong association between sugar availability and signaling with cell division and expansion (Van Dingenen *et al*., 2016), and in *Populus* stem tissues cambial differentiation into xylem has shown to be associated with the circadian clock (Edwards *et al*., 2018). Plastid-targeted carbon metabolism in xylem is coordinated with the circadian clock, and morning-associated PC7 suggests that primary metabolites, such as nucleotides, UDP sugars, and acetyl coA are produced in the morning in preparation for SCW deposition (Fig. **3**, Notes **S2**). In line with previous studies, PC6 shows that starch metabolism occurs later in the day in association with evening clock components such as *LUX*, *ELF3*, and *PRR5* (Solomon *et al*., 2010; Wijnen & Young, 2006). The onset of metabolic dawn entrains the central circadian clock in xylem via the release of *PRR7* repression of *CCA1* (*LHY* - PC7), leading to the repression of evening clock gene *PRR1* (*TOC1* - PC6) (Müller *et al*., 2014; Haydon *et al*., 2013). Although *PRR7* (PC6 - evening module) expression peaks in the morning in photosynthetic tissue (Farré *et al*., 2005), it is known to be delayed in roots, and has an evening peak in *Populus* stems (James *et al*., 2008; Edwards *et al*., 2018), and this decoupling of clock components is also found in developing xylem in *Eucalyptus* (Table **S2**). The aspen developmental series show that the morning-associated genes in PC7 overlap with the initial phases of xylogenesis (aspen developmental phases e1, e2, and f) (Fig. **5**).

Other than the regulation of starch biosynthesis in plastids, little is known about the circadian regulation of plastids in xylem (Solomon *et al*., 2010). Here, several links between plastid biology and the xylem circadian clock are reported. Specifically, *GPT2*, a glucose-6-phosphate plastid membrane transporter which has a positive effect on cell growth in *Arabidopsis* is found in PC7 and is known to be induced by sugar availability (Kunz *et al*., 2010; Van Dingenen *et al*., 2016). Perhaps *GPT2* is induced by metabolic dawn in xylem, leading to the activation of primary growth processes such as *de novo* fatty acid biosynthesis, while phenylpropanoid metabolism is repressed by the interaction of *LNK1* and *LNK2* with *MYB3* (Fig. **3**) (Zhou *et al*., 2017). These regulatory mechanisms are further validated by the significant trait association of PC7 genes with lignin content in wood (Fig. **4**). Taken together, these results show that fine-scale regulation of carbon metabolism during xylogenesis acts

to spatially and temporally separate metabolite precursors for the major phases of wood formation via epigenetic, transcriptional, and metabolite transporter regulation.

The analysis of xylem plastid type and function is complicated by the diversity of cell types present in wood and the developmental changes associated with xylogenesis. For instance, xylem tracheary element plastids are degraded during programmed cell death, while parenchyma cells contain amyloplasts and can live for decades (Spicer & Holbrook, 2007; Bollhöner *et al*., 2012). Although xylem plastids may be amyloplasts producing phenylalanine in addition to starch storage, but previous research has shown that starch filled grana in *Eucalyptus elaeophora* differentiating ray parenchyma plastids are replaced by phenolic biopolymers (Wardrop & Cronshaw 1962). Amyloplasts in specialized cells of secondary phloem of *Populus* are established during tension wood formation, acting as gravity perceiving statoliths, and were not observed in secondary xylem (Gerttula *et al*., 2015). The discovery of the vanilla fruit cell phenyloplasts, which accumulate phenyl glucoside in re-differentiated chloroplasts (Brillouet *et al*., 2014), shows that plastids previously classified as leucoplasts, such as developing xylem plastids (xyloplasts), may be specifically adapted to the metabolic needs of a wide array of cell types (Pinard & Mizrachi 2018).

In the effort to improve important biomass crop species, organellar variation and the resulting cytonuclear coordination should not be ignored, given the large effect that organellar genome variation can have on plant fitness and primary metabolism (Roux *et al*., 2016; Joseph *et al*., 2015; Langridge & Fleury, 2011; Kersten *et al*., 2016). Novel insights from this work include the plastid specific temporal regulation of carbon allocation, and epigenetic modification during early xylogenesis, which is potentially regulated by retrograde signaling. The conservation of regulatory networks linked to organellar biology in the developmental stages of xylogenesis between *Populus* and *Eucalyptus* should encourage future research into the role of organelles in wood formation. The research presented here shows that organelle specific biology should be considered when aiming to improve carbon flux towards increased biomass production in lignocellulosic biomass-producing species.

Acknowledgements

This work was supported by the Department of Science and Technology (Strategic Grant for the Eucalyptus Genomics Platform) and National Research Foundation of South Africa (Bioinformatics and Functional Genomics Programme, Grants 86936 and 97911 to A.A.M.), Sappi South Africa and the Technology and Human Resources for Industry Programme (Grant 80118) through the Forest Molecular Genetics Programme at the University of Pretoria (to A.A.M.), and the Ghent University Multidisciplinary Research Partnership from nucleotides to networks (Project 01MR0410W to K.M.). DP was supported by the National Research Foundation of South Africa Scarce Skills grant. The authors have no conflicts of interest to declare.

Author contributions

E.M. is the lead investigator and conceived of the study. A.C.F. wrote the network generation and clustering script. D.P. performed the data analysis and wrote the article with E.M. A.A.M. was a co-investigator for the population and tissue-specific transcriptome analysis and edited the manuscript. K.M. was a co- investigator for the data analysis and edited the manuscript. All authors have read and commented on the article.

References

Agusti J, Herold S, Schwarz M, Sanchez P, Ljung K, Dun EA, Brewer PB, Beveridge CA, Sieberer T, Sehr EM, *et al.* 2011. Strigolactone signaling is required for auxin-dependent stimulation of secondary growth in plants. *Proceedings of the National Academy of Sciences* **108**: 20242–20247.

Aharoni A, Galili G. **2011**. Metabolic engineering of the plant primary-secondary metabolism interface. *Current Opinion in Biotechnology* **22**: 239–244.

Banerjee A, Sharkey TD. **2014**. Methylerythritol 4-phosphate (MEP) pathway metabolic regulation. *Natural product reports* **31**: 1043–1055.

Beltrán J, Wamboldt Y, Sanchez R, LaBrant EW, Kundariya H, Virdi KS, Elowsky C, Mackenzie SA. **2018**. Specialized plastids trigger tissue-specific signaling for systemic stress response in plants. *Plant Physiology* **178**: 672–683.

Benjamini Y, Hochberg Y. **1995**. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology* **57**: 289–300.

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. **2008**. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **P10008**.

Bollhöner B, Prestele J, Tuominen H. **2012**. Xylem cell death: emerging understanding of regulation and function. *Journal of Experimental Botany* **63**: 1081–1094.

Bouvier F, Linka N, Isner J-C, Mutterer J, Weber APM, Camara B. **2006**. *Arabidopsis* SAMT1 defines a plastid transporter regulating plastid biogenesis and plant development. *The Plant Cell* **18**: 3088–3105.

Brillouet JM, Romieu C, Schoefs B, Solymosi K, Cheynier V, Fulcrand H, Verdeil JL, Conéjéro G. **2013**. The tannosome is an organelle forming condensed tannins in the chlorophyllous organs of *Tracheophyta*. *Annals of Botany* **112**: 1003–1014.

Brillouet JM, Verdeil JL, Odoux E, Lartaud M, Grisoni M, Conéjéro G. **2014**. Phenol homeostasis is ensured in vanilla fruit by storage under solid form in a new chloroplast-derived organelle, the phenyloplast. *Journal of Experimental Botany* **65**: 2427–2435.

Cao X, Aufsatz W, Zilberman D, Mette MF, Huang MS, Matzke M, Jacobsen SE. **2003**. Role of the DRM and CMT3 methyltransferases in RNA-directed DNA methylation. *Current biology* **13**: 2212–2217.

Carocha V, Hefer C, Cassan-Wang H, Fevereiro P, Alexander A, Paiva JAP, Grima-pettenati J, Grima-Pettenati. **2015**. Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *The New Phytologist* **206:** 1297–1313.

Cazzonelli CI, Cuttriss AJ, Cossetto SB, Pye W, Crisp P, Whelan J, Finnegan EJ, Turnbull C, Pogson BJ. **2009**. Regulation of carotenoid composition and shoot branching in *Arabidopsis* by a chromatin modifying histone methyltransferase, SDG8. *The Plant Cell* **21**: 39–53.

Cheng S-H, Willmann MR, Chen H-C, Sheen J. **2002**. Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiology* **129**: 469–485.

Chen M, Mooney BP, Hajduch M, Joshi T, Zhou M, Xu D, Thelen JJ. **2009**. System analysis of an *Arabidopsis* mutant altered in de novo fatty acid synthesis reveals diverse changes in seed composition and metabolism. *Plant Physiology* **150**: 27–41.

Chen Y, Zou T, McCormick S. **2016**. S-Adenosylmethionine Synthetase 3 Is important for pollen tube growth. *Plant Physiology* **172**: 244–253.

Covington MF, Maloof JN, Straume M, Kay SA, Harmer SL. **2008**. Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biology* **9**: R130.

Csardi G, Nepusz T. **2006**. The igraph software package for complex network research. *InterJournal, Complex Systems* **5:** 1-9.

De Rybel B, Mähönen AP, Helariutta Y, Weijers D. **2015**. Plant vascular development: from early specification to differentiation. *Nature Reviews. Molecular Cell Biology* **17**: 30–40.

Djidjev HN. **2006**. A scalable multilevel algorithm for graph clustering and community structure Detection. In: W.Aiello, A.Z.Broder, J.C.M.Janssen, E.E.Milios, eds. *Algorithms and Models for the Web-Graph*. Springer, Berlin, Heidelberg, 117–128.

Edwards KD, Takata N, Johansson M. **2018**. Circadian clock components control daily growth activities by modulating cytokinin levels and cell division□associated gene expression in *Populus* trees. *Plant, Cell & Environment* **41:** 1468-1482.

Endo M, Shimizu H, Nohales MA, Araki T, Kay SA. **2014**. Tissue-specific clocks in *Arabidopsis* show asymmetric coupling. *Nature* **515**: 419-422.

Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. **2007**. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS biology* **5**: e8.

Farage-Barhom S, Burd S, Sonego L, Perl-Treves R, Lers A. **2008**. Expression analysis of the BFN1 nuclease gene promoter during senescence, abscission, and programmed cell death-related processes. *Journal of Experimental Botany* **59**: 3247–3258.

Farré EM, Harmer SL, Harmon FG, Yanovsky MJ, Kay SA. **2005**. Overlapping and distinct roles of PRR7 and PRR9 in the *Arabidopsis* circadian clock. *Current Biology* **15**: 47–54.

Feng K, Liu F, Zou J, Xing G, Deng P, Song W, Tong W, Nie X. **2016**. Genome-wide identification, evolution and co-expression network analysis of mitogen-activated protein kinase kinase kinases in *Brachypodium distachyon*. *Frontiers in Plant Science* **7**: 1400.

Fujii H, Verslues PE, Zhu J-K. **2007**. Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in *Arabidopsis*. *The Plant Cell* **19**: 485–494.

Gerttula S, Zinkgraf M, Muday GK, Lewis DR, Ibatullin FM, Brumer H, Hart F, Mansfield SD, Filkov V, Groover A. **2015**. Transcriptional and hormonal regulation of gravitropism of woody stems in *Populus*. *The Plant Cell* **27**: 2800–2813.

Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA. **2000**. Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. *Science* **290**: 2110–2113.

Haydon MJ, Mielczarek O, Robertson FC, Hubbard KE, Webb AA. **2013**. Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock. *Nature* **502**: 689–692.

Herrmann KM. **1995**. The shikimate pathway as an entry to aromatic secondary metabolism. *Plant physiology* **107**: 7–12.

Huanca-Mamani W, Garcia-Aguilar M, León-Martínez G, Grossniklaus U, Vielle-Calzada J-P. **2005**. CHR11, a chromatin-remodeling factor essential for nuclear proliferation during female gametogenesis in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* **102**: 17231–17236.

Hussey SG, Mizrachi E, Creux NM, Myburg AA. **2013**. Navigating the transcriptional roadmap regulating plant secondary cell wall deposition. *Frontiers in Plant Science* **4**: 1-21.

Irigoyen S, Karlsson PM, Kuruvilla J, Spetea C, Versaw WK. **2011**. The sink-specific plastidic phosphate transporter PHT4;2 influences starch accumulation and leaf size in *Arabidopsis*. *Plant physiology* **157**: 1765–1777.

Jacoby RP, Li L, Huang S, Pong Lee C, Millar AH, Taylor NL. **2012**. Mitochondrial composition, function and stress response in plants. *Journal of Integrative Plant Biology* **54**: 887–906.

James AB, Monreal JA, Nimmo GA, Kelly CL, Herzyk P, Jenkins GI, Nimmo HG. **2008**. The circadian clock in *Arabidopsis* roots is a simplified slave version of the clock in shoots. *Science* **322**: 1832–1835.

Jensen RA. **1986**. The shikimate/arogenate pathway: link between carbohydrate metabolism and secondary metabolism. *Physiologia Plantarum* **66**: 164-168.

Jokipii□Lukkari S, Sundell D, Nilsson O. **2017**. NorWood: a gene expression resource for evo□devo studies of conifer wood development. *New Phytologist* **216**: 482-494.

Jones AM, Assmann SM. **2004**. Plants: the latest model system for G□protein research. *EMBO Reports* **5**: 572-578.

Joseph B, Corwin JA., Kliebenstein DJ. **2015**. Genetic variation in the nuclear and organellar genomes modulates stochastic variation in the metabolome, growth, and defense. *PLoS genetics* **11**: e1004779.

Joyard J, Ferro M, Masselon C, Seigneurin-Berny D, Salvi D, Garin J, Rolland N. **2010**. Chloroplast proteomics highlights the subcellular compartmentation of lipid metabolism. *Progress in Lipid Research* **49**: 128–158.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, *et al.* 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Research* **36**: 480–484.

Kersten B, Faivre Rampant P, Mader M, Le Paslier M-C, Bounon R, Berard A, Vettori C, Schroeder H, Leplé J-C, Fladung M. **2016**. Genome sequences of *Populus tremula* chloroplast and mitochondrion: implications for holistic poplar breeding. *PloS one* **11**: e0147209.

Kühn C, Grof CPL. **2010**. Sucrose transporters of higher plants. *Current Opinion in Plant Biology* **13**: 288–298.

Kunz HH, Häusler RE, Fettke J, Herbst K, Niewiadomski P, Gierth M, Bell K, Steup M, Flügge U-I, Schneider A. **2010**. The role of plastidial glucose-6-phosphate/phosphate translocators in vegetative tissues of *Arabidopsis thaliana* mutants impaired in starch biosynthesis. *Plant Biology* **12**: 115–128.

Kwok ACM, Wong JTY. **2005**. Lipid biosynthesis and its coordination with cell cycle progression. *Plant & Cell Physiology* **46**: 1973–1986.

Langridge P, Fleury D. **2011**. Making the most of 'omics' for crop breeding. *Trends in Biotechnology* **29**: 33–40.

Lee JE, Lampugnani ER, Bacic A, Golz JF. **2014**. SEUSS and SEUSS-LIKE 2 coordinate auxin distribution and KNOXI activity during embryogenesis. *The Plant Journal* **80**: 122–135.

Li Y, Mukherjee I, Thum KE, Tanurdzic M, Katari MS, Obertello M, Edwards MB, McCombie WR, Martienssen RA, Coruzzi GM. **2015**. The histone methyltransferase SDG8 mediates the epigenetic modification of light and carbon responsive genes in plants. *Genome Biology* **16**: 1-15.

Ling Q, Jarvis P. **2015**. Functions of plastid protein import and the ubiquitin–proteasome system in plastid development. *Biochimica et Biophysica Acta- Bioenergetics* **1847**: 939–948.

Linka N, Weber APM. **2010**. Intracellular metabolite transporters in plants. *Molecular Plant* **3**: 21–53.

Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, Richard McCombie W, Lavine K, Mittal V, May B, Kasschau KD, *et al.* 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 471–476.

Lloyd C, Chan J. **2008**. The parallel lives of microtubules and cellulose microfibrils. *Current Opinion in Plant Biology* **11**: 641–646.

Lucas WJ, Groover A, Lichtenberger R, Furuta K, Yadav S-R, Helariutta Y, He X-Q, Fukuda H, Kang J, Brady SM, *et al.* 2013. The plant vascular system: evolution, development and functions. *Journal of Integrative Plant Biology* **55**: 294–388.

Lundquist PK, Poliakov A, Bhuiyan NH, Zybailov B, Sun Q, van Wijk KJ. **2012**. The

functional network of the *Arabidopsis* plastoglobule proteome based on quantitative proteomics and genome-wide coexpression analysis. *Plant Physiology* **158**: 1172–1192.

Magwene PM, Kim J. **2004**. Estimating genomic coexpression networks using first-order conditional independence. *Genome Biology* **5**: R100.

Majsec K, Bhuiyan NH, Sun Q, Kumari S, Kumar V, Ware D, van Wijk KJ. **2017**. The plastid and mitochondrial peptidase network in *Arabidopsis thaliana*: A foundation for testing genetic interactions and functions in organellar proteostasis. *The Plant Cell* **29**: 2687–2710.

Mellerowicz EJ, Baucher M, Sundberg B, Boerjan W. **2001**. Unravelling cell wall formation in the woody dicot stem. *Plant Molecular Biology* **47**: 239–274.

Mentzen WI, Wurtele ES. **2008**. Regulon organization of *Arabidopsis*. *BMC Plant Biology* **8**: 1–22.

Milhinhos A, Miguel CM. **2013**. Hormone interactions in xylem development: A matter of signals. *Plant Cell Reports* **32**: 867–883.

Mizrachi E, Hefer C, 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**: 681–693

Mizrachi E, Myburg AA. **2016**. Systems genetics of wood formation. *Current Opinion in Plant Biology* **30**: 94–100.

Mizrachi E, Verbeke L, Christie N, Fierro AC, Mansfield SD, Davis MF, Gjersing E, Tuskan GA, Van Montagu M, Van de Peer Y, *et al.* 2017. Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proceedings of the National Academy of Sciences* **114**: 1195–1200.

Müller LM, Von Korff M, Davis SJ. **2014**. Connections between circadian clocks and carbon metabolism reveal species-specific effects on growth control. *Journal of Experimental Botany* **65**: 2915–2923.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D, *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* **510**: 356–362.

Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. **2013**. Mechanisms and functions of

ATP-dependent chromatin-remodeling enzymes. *Cell* **154**: 490–503.

Nielsen AZ, Mellor SB, Vavitsas K, Wlodarczyk AJ, Gnanasekaran T, Perestrello Ramos H de Jesus M, King BC, Bakowski K, Jensen PE. **2016**. Extending the biosynthetic repertoires of cyanobacteria and chloroplasts. *The Plant Journal* **87**: 87–102.

Oda Y. **2015**. Cortical microtubule rearrangements and cell wall patterning. *Frontiers in Plant Science* **6**: 1–7.

Ohlrogge J, Browse J. **1995**. Lipid biosynthesis. *The Plant Cell* **7**: 957–970.

Oliver DJ, Nikolau BJ, Syrkin Wurtele E. **2009**. Acetyl-CoA - Life at the metabolic nexus. *Plant Science* **176**: 597–601.

Pesaresi P, Schneider A, Kleine T, Leister D. **2007**. Interorganellar communication. *Current Opinion in Plant Biology* **10**: 600–606.

Pinard D, Mizrachi E. 2018. Unsung and understudied: plastids involved in secondary growth. *Current Opinion in Plant Biology* **42**: 30–36.

Proost S, Bel MV, Vaneechoutte D, Peer YVD, Mueller-Roeber B, Vandepoele K. **2014**. PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Research* **43**: 1–8.

Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, Andersson CR, Zhao Y, Ljung K, Harmer SL. **2009**. REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. *Proceedings of the National Academy of Sciences* **106**: 16883–16888.

Reimann R, Kost B, Dettmer J. **2017**. TETRASPANINs in plants. *Frontiers in Plant Science* **8**: 1-10.

Roux F, Mary-Huard T, Barillot E, Wenes E, Botran L, Durand S, Villoutreix R, Martin-Magniette M-L, Camilleri C, Budar F. **2016**. Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. *Proceedings of the National Academy of Sciences* **113**: 3687–3692.

Ruprecht C, Mutwil M, Saxe F, Eder M, Nikoloski Z, Persson S. **2011**. Large-scale co-expression approach to dissect secondary cell wall formation across plant species. *Plant Physiology* **2**: 1–13.

Sánchez-Rodríguez C, Bauer S, Hématy K, Saxe F, Ibáñez AB, Vodermaier V, Konlechner C, Sampathkumar A, Rüggeberg M, Aichinger E, *et al.* 2012. Chitinase-

like1/pom-pom1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in *Arabidopsis*. *The Plant Cell* **24**: 589–607.

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. **2016**. Learning from co-expression networks: Possibilities and challenges. *Frontiers in Plant Science* **7**: 1-18.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. **2003**. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498–2504.

Shen Y, Devic M, Lepiniec L, Zhou DX. **2015**. Chromodomain, helicase and DNA-binding CHD1 protein, CHR5, are involved in establishing active chromatin state of seed maturation genes. *Plant Biotechnology Journal* **13**: 811–820.

Shen J, Ren X, Cao R, Liu J, Gong Z. **2009**. Transcriptional gene silencing mediated by a plastid inner envelope phosphoenolpyruvate/phosphate translocator CUE1 in *Arabidopsis*. *Plant Physiology* **150**: 1990–1996.

Shimizu H, Katayama K, Koto T, Torii K, Araki T, Endo M. **2015**. Decentralized circadian clocks process thermal and photoperiodic cues in specific tissues. *Nature Plants* **1**: 1-6.

Solomon OL, Berger DK, Myburg AA. **2010**. Diurnal and circadian patterns of gene expression in the developing xylem of *Eucalyptus* trees. *South African Journal of Botany* **76**: 425–439.

Spicer R, Holbrook NM. **2007**. Parenchyma cell respiration and survival in secondary xylem: does metabolic activity decline with cell age? *Plant, Cell & Environment* **30**: 934–943.

Staehr P, Löttgert T, Christmann A, Krueger S, Rosar C, Rolčík J, Novák O, Strnad M, Bell K, Weber APM, *et al.* 2014. Reticulate leaves and stunted roots are independent phenotypes pointing at opposite roles of the phosphoenolpyruvate/phosphate translocator defective in *cue1* in the plastids of both organs. *Frontiers in Plant Science* **5**: 1-22.

Streatfield SJ, Weber A, Kinsman EA, Häusler RE, Li J, Post-Beittenmiller D, Kaiser WM, Pyke KA, Flügge UI, Chory J. **1999**. The phosphoenolpyruvate/phosphate translocator is required for phenolic metabolism, palisade cell development, and plastid-dependent nuclear gene expression. *The Plant Cell* **11**: 1609–1622.

Sundell D, Street NR, Kumar M, Mellerowicz EJ, Kucukoglu M, Johnsson C, Kumar V, Mannapperuma C, Delhomme N, Nilsson O, *et al.* 2017. AspWood: High-spatial-resolution transcriptome profiles reveal uncharacterized modularity of wood formation in *Populus*

tremula. *The Plant Cell* **29**: 1585-1604.

Tableau Software. *Tableau Software v10.5* URL https://www.tableau.com/.

Talbert PB, Adler HT, Parks DW, Comai L. **1995**. The REVOLUTA gene is necessary for apical meristem development and for limiting cell divisions in the leaves and stems of *Arabidopsis thaliana*. **121**: 2723–2735.

Tanz SK, Castleden I, Hooper CM, Vacher M, Small I, Millar HA. **2013**. SUBA3: a database for integrating experimentation and prediction to define the SUBcellular location of proteins in *Arabidopsis*. *Nucleic Acids Research* **41**: 1185–1191.

Taylor NG. **2008**. Cellulose biosynthesis and deposition in higher plants. *The New Phytologist* **178**: 239–252.

Tohge T, Watanabe M, Hoefgen R, Fernie AR. **2013**. Shikimate and phenylalanine biosynthesis in the green lineage. *Frontiers in Plant Science* **4**: 1-13.

Ursache R, Miyashima S, Chen Q, Vatén A, Nakajima K, Carlsbecker A, Zhao Y, Helariutta Y, Dettmer J. **2014**. Tryptophan-dependent auxin biosynthesis is required for HD-ZIP III-mediated xylem patterning. *Development* **141**: 1250–1259.

Usadel B, Obayashi T, Mutwil M, Giorgi FM, Bassel GW, Tanimoto M, Chow A, Steinhauser D, Persson S, Provart NJ. **2009**. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant, Cell & Environment* **32**: 1633–1651.

Van Aken O, Van Breusegem F. **2015**. Licensed to kill: mitochondria, chloroplasts, and cell death. *Trends in Plant Science* **20**: 754–766.

Van Dingenen J, De Milde L, Vermeersch M, Maleux K, De Rycke R, De Bruyne M, Storme V, Gonzalez N, Dhondt S, Inzé D. **2016**. Chloroplasts are central players in sugar-induced leaf growth. *Plant Physiology* **171**: 590–605.

Van Hautegem T, Waters AJ, Goodrich J, Nowack MK. **2015**. Only in dying, life: programmed cell death during plant development. *Trends in Plant Science* **20**: 102–113.

Villalobos DP, Díaz-Moreno SM, Said E-SS, Cañas RA, Osuna D, Van Kerckhoven SHE, Bautista R, Claros MG, Cánovas FM, Cantón FR. 2012. Reprogramming of gene expression during compression wood formation in pine: Coordinated modulation of S-adenosylmethionine, lignin and lignan related genes. *BMC plant biology* **12**: 1-17.

Virdi KS, Wamboldt Y, Kundariya H, Laurie JD, Keren I, Kumar KRS, Block A, Basset G, Luebker S, Elowsky C, *et al.* 2016. MSH1 Is a plant organellar DNA binding and thylakoid protein under precise spatial regulation to alter development. *Molecular Plant* **9**: 245–260.

Wang X, Chen J, Xie Z, Liu S, Nolan T, Ye H, Zhang M, Guo H, Schnable PS, Li Z, *et al.* 2014. Histone lysine methyltransferase SDG8 is involved in brassinosteroid-regulated gene expression in *Arabidopsis thaliana*. *Molecular Plant* **7**: 1303–1315.

Wang H, Zhang X, Liu J, Kiba T, Woo J, Ojo T, Hafner M, Tuschl T, Chua N-H, Wang X-J. **2011**. Deep sequencing of small RNAs specifically associated with *Arabidopsis* AGO1 and AGO4 uncovers new AGO functions. *The Plant Journal* **67**: 292–304.

Wardrop AB, Cronshaw J. **1962**. Formation of phenolic substances in the ray parenchyma of angiosperms. *Nature* **193**: 90–92.

Wickham H. **2009.** *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag New York.

Wiedemann N, Pfanner N. **2017**. Mitochondrial machineries for protein import and assembly. *Annual Review of Biochemistry* **86**: 685–714.

Wijnen H, Young MW. **2006**. Interplay of circadian clocks and metabolic rhythms. *Annual Review of Genetics* **40**: 409–448.

Yang D-L, Zhang G, Tang K, Li J, Yang L, Huang H, Zhang H, Zhu J-K. 2016. Dicer-independent RNA-directed DNA methylation in *Arabidopsis*. *Cell Research* **26**: 66–82.

Yu X-H, Perdue TD, Heimer YM, Jones AM. **2002**. Mitochondrial involvement in tracheary element programmed cell death. *Cell Death and Differentiation* **9**: 189–198.

Zhou M, Zhang K, Sun Z, Yan M, Chen C, Zhang X, Tang Y, Wu Y. **2017**. LNK1 and LNK2 corepressors interact with the MYB3 transcription factor in phenylpropanoid biosynthesis. *Plant Physiology* **174**: 1348–1358.

Tables

Table 1: SUBA predicted subcellular localization of *Eucalyptus grandis* and *Arabidopsis thaliana* genes.

Table 1

Predicted subcellular location of *E. grandis* genes				
Subcellular localization	***E. grandis* genomic**	**% GUxU xylem expressed**	**GUxU xylem KEGG carbon expressed**	**% KEGG carbon expressed**
Cytosol	5 558	17.46%	243	32.53%
Endoplasmic reticulum	928	2.14%	20	2.68%
Extracellular	2 549	4.94%	58	7.76%
Golgi	165	0.73%	11	1.47%
Mitochondrion	2 023	8.04%	62	8.30%
Multiple	3 443	12.57%	113	15.13%
Nucleus	7 094	27.22%	9	1.20%
Peroxisome	627	1.83%	42	5.62%
Plasma membrane	5 649	12.73%	35	4.69%
Plastid	2 767	11.05%	148	19.81%
Vacuole	494	1.29%	6	0.80%

SUBA predicted subcellular localization of *Eucalyptus* genes. Column one shows the subcellular compartment, where “multiple” indicates genes predicted to be targeted to more than one location. *E. grandis* genomic is the predicted location of all genes in the *E. grandis* v.2.0 genome. The next columns show the percentage of genes in the *Eucalyptus grandis* x *E. urophylla* x *E. urophylla* (GUxU) backcross population that are expressed in at least 75% of the individuals in immature xylem, followed by the number and percentage of expressed genes in 17 KEGG primary carbon metabolic pathways (Supporting Information Notes S1).

Fig. 1 *Eucalyptus grandis* × *E. urophylla* × *E. urophylla* (GUxU) backcross population plastid and mitochondrial network structure and function. (a) Heatmap of network clusters for the GUxU backcross population plastid query coexpression network. Rows denote genes in the network, and the columns represent each individual within the population clustered according to the dendrogram. The bars on the left-hand side of the figure show the cluster assignment of genes in the network. Plastid clusters (PC) 1–7 are shown in the sidebar as follows: PC1, red; PC2, orange; PC3, green; PC4, turquoise; PC5, light blue; PC6, blue; PC7, purple. Mean centered and rescaled $\log_{10}$ FPKM Z-scores are shown from blue (low) to red (high) with the colour key in the top right. (b) Plastid coexpression network with clusters and query genes shown. Clusters are colour coded as in the heatmap. (c) Heatmap of network clusters of the mitochondrial query coexpression network. Mitochondrial clusters (MC) 1 to 5 are shown in the sidebar as follows: MC1, orange; MC2, purple; MC3, dark green; MC4, red; MC5, light green. (d) Mitochondrial coexpression network with clusters and query genes shown. Clusters are colour coded as in the heatmap. For the networks, selected query genes are shown as large yellow diamonds, and additional text is the main functional assignment of selected clusters (secondary cell wall, SCW). Network layout is Cytoscape perfuse force directed layout, with correlation as distance between nodes.

Fig. 2 Overlapping gene analysis between the plastid and mitochondrial targeted carbon metabolic gene coexpression networks. (a) Sankey diagram showing overlapping genes between the plastid (left) and mitochondrial (right) network clusters. (b) Heatmap of tissue specific gene expression of overlapping genes in six *Eucalyptus grandis* tissues in the plastid and mitochondrial clusters (Vining *et al*. 2015, Myburg *et al*., 2014). The row means centered and scaled log10 FPKM values for each gene is shown from blue (low) to red (high) across young leaf (YL), mature leaf (ML), shoot tips (ST), flowers (FL), immature xylem (IX), and phloem (PH). The bars on the right-hand side show the cluster assignment of genes in the plastid network (PC) and the mitochondrial network (MC). The plot above the heatmap shows the average expression level in each tissue after normalization. (c) GO biological process term enrichment for genes in the overlaps PC1 - MC4, PC3 - MC3 & MC5, and PC7 - MC2, and for PC6 which had little overlap between plastid and mitochondrial clusters. The FDR adjusted *P*-values for each term are shown on the *x*-axis, and points are coloured to match the clusters.

Fig. 3 (a) Expression profile of *Eucalyptus grandis* circadian clock genes across 156 individuals in the plastid network sorted by time of collection between 09:00 to 16:30 h in $\log_{10}$ fragments per kilobase of transcript per million (FPKM). Genes found in plastid cluster (PC) 6 (evening module) are blue, and PC7 (morning module) are purple. The smooth lines show the average expression profile of all clock genes in each cluster, and the grey shaded area represents the standard deviation of each

cluster. (b) Plastid cluster assignment of genes involved in fatty acid synthesis and the shikimate pathway which can use phosphoenolpyruvate (PEP) as a substrate. The figure was generated using a custom pathway in MapMan (v3.5.1.R2) and shows genes in PC3 (green) and PC7 (purple). See Supporting Information Tables S1 and S2 for gene IDs and cluster assignment, and Notes S3 for full annotation and tissue specific expression of these genes. ACP, acyl carrier protein; PEP, phosphoenolpyruvate; E-4-P, erythrose-4-phosphate; DAHP, 3-deoxy-D-arabino-heptulosonate 7-phosphate; EPSP, 5-enolpyruvylshikimate 3-phosphate; PHE, phenylalanine.

Fig. 4 Number of genes in (a) plastid and (b) mitochondrial carbon metabolic network clusters significantly associated with lignocellulosic biomass traits in Mizrachi *et al*. (2017). Significant trait association to genes was determined by a network-based data integration approach using the same *Eucalyptus grandis × E. urophylla × E. urophylla* (GUxU) backcross population used to construct the plastid and mitochondrial networks. The traits are shown on the *x*-axis, and the number of genes is shown on the *y*-axis as stacked bars coloured according to plastid and mitochondrial clusters (PC and MC). For clarity, only the clusters with large overlaps are shown. DBH (ob), diameter at breast height (over bark); DBH (ub), diameter at breast height (under bark); pyMBMS, pyrolysis molecular beam mass spectrometry.

Fig. 5 Aspen developmental phase specific expression profiles of genes with orthologs in plastid clusters P1, P3, P6, and P7. The expression of aspen (*Populus tremula*) genes over the three main developmental transitions (cluster sections i–iv) of wood formation representing 25 high spatial resolution samples of wood development were taken from Sundell *et al*. (2017). *Eucalyptus grandis* orthologs of *Populus trichocarpa* genes were identified using PLAZA v3, and were used to identify overlapping genes between plastid network clusters (PC1 to PC7), and aspen developmental clusters a–h. The pie charts at the top of the graphs indicate the proportion of unique *E. grandis* genes that overlap aspen clusters e1, e2, f, g1, and g2. The total number of overlapping genes between the two datasets are indicated by the size of the pie chart and the number below, and the colour of the segments represent the plastid clusters as previously defined. For aspen clusters e1, e2, f, g1, and g2, the expression profiles of the genes which do not overlap with plastid clusters are in grey in the bottom panel to show that the plastid cluster overlapping genes do not deviate from other genes in each aspen cluster. The black lines indicate the smoothed average expression profiles of the genes in each panel. For clarity, only overlaps between aspen clusters e1, e2, f, g1, and g2 and plastid clusters PC1, PC3, PC6, and PC7 are shown. Overlap between all clusters and details of the aspen clusters and developmental transitions can be found in Supporting Information Fig. S4.

Supporting information

Figure S1: Tissue specific gene expression heatmaps of non-overlapping plastid and mitochondrial cluster genes.

Figure S2: Network properties of the network constructed from shared plastid and mitochondrial query genes.

Figure S3: Aspen developmental clusters, overlap with plastid clusters, and gene expression of *Eucalyptus* specific genes.

Figure S4: Overview of plastid network construction method

Table S1: Plastid cluster allocation of *Eucalyptus grandis* phenylalanine and monolignol biosynthetic genes.

Table S2: Plastid cluster allocation of *Eucalyptus grandis* circadian rhythm genes.

Notes S1: Word document of 17 KEGG carbon metabolism pathways used to identify plastid and mitochondrial carbon metabolism genes in *Arabidopsis thaliana*.

Notes S2: GO enrichment of plastid, mitochondrial, and shared network clusters, and the overlap between plastid clusters and aspen developmental phases.

Notes S3: Annotation file of all *E. grandis* genes that are in the plastid (PC), mitochondrial (MC), and shared (SC) networks, and *E. grandis* network overlaps with the aspen developmental phases, and aspen circadian clock genes.