Systems biology of organellar carbon metabolism in *Eucalyptus* xylogenesis

by

Desré Pinard

Submitted in partial fulfillment of the requirements for the degree

Philosophiae Doctor

In the Faculty of Natural and Agricultural Sciences

Department of Biochemistry, Genetics, and Microbiology

University of Pretoria

Pretoria

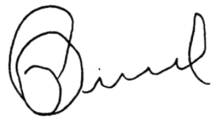
January 2020

Supervisor: Prof. Eshchar Mizrachi

Co-supervisor: Prof. Alexander A. Myburg

Declaration

I, the undersigned, hereby declare that the thesis submitted herewith for the degree Ph.D. to the University of Pretoria contains my own independent work and has not been submitted previously for any degree at any university.



Desré Pinard

28-01-2020

Date

Thesis summary ..... *ix*

Preface ..... *xii*

Acknowledgements ..... *xix*

Chapter 1 ..... *1*

The role of plastids and mitochondria in wood formation *1*

1.1. Introduction **2**

1.2. Organellar targeted metabolism is central to xylogenesis **4**

1.3. The endosymbiotic legacy of organelles in land plants **8**

1.4. Plastid and mitochondrial genome structure **12**

1.5. Transcriptional regulation of organellar-encoded genes **13**

1.6. Non-photosynthetic plastid gene expression **17**

1.7. Post-transcriptional regulation of organellar-encoded genes **20**

1.8. Cytonuclear coordination in land plants **23**

1.9. Conclusion **27**

1.10. References **28**

Chapter 2 ..... *58*

Organellar carbon metabolism is coordinated with distinct developmental phases of secondary xylem **58**

2.1. Summary **59**

2.2. Introduction 60

2.3. Materials and Methods 63

2.3.1. Subcellular localization and KEGG pathway prediction 63

2.3.2. Query-based co-expression network construction and graph-based clustering 63

2.3.3. Overlap of *E. grandis* plastid and mitochondrial cluster genes with the aspen developmental transcriptome series and genes significantly associated with lignocellulosic biomass traits 65

2.3.4. Functional enrichment analysis 66

2.4. Results 68

2.4.1. Subcellular localization and expression of plastid and mitochondrial-targeted genes in *Eucalyptus* 68

2.4.2. Co-expression networks of plastid and mitochondrial carbon metabolic genes 68

2.4.3. Distinct regulatory modules of early and late xylem development in plastid network clusters 70

2.4.4. Plastid localized carbon metabolism is regulated by the circadian rhythm to ensure efficient carbon allocation during xylogenesis.....74

2.4.5. Plastid and mitochondrial network genes overlap with the *Populus tremula* secondary xylem developmental series.....77

2.5. Discussion 80

2.6. Acknowledgements 85

2.7. References 85

2.8. Tables 100

Table 2.1: Predicted subcellular localization of *E. grandis* genes.....100

2.9. Figures 101

Figure 2.1. Network structure and function.....101

Figure 2.2. Overlapping gene analysis between the plastid and mitochondrial targeted carbon metabolic gene co-expression networks....104
Figure 2.3. Circadian gene expression and PEP metabolism 105
Figure 2.4. Network association with lignocellulosic biomass traits....106
Figure 2.5. Xylem developmental series overlap with plastid clusters....108

2.10. Supporting information 110

Figure S2.1. Overview of plastid network construction method. 110
Figure S2.2. Tissue specific gene expression heatmaps of non-overlapping plastid and mitochondrial cluster genes. 111
Figure S2.3. Network properties of the network constructed from shared plastid and mitochondrial query genes. 112
Figure S2.4. Aspen developmental clusters, overlap with plastid clusters, and gene expression of *Eucalyptus* specific genes....114
Table S2.1. Plastid cluster allocation of *Eucalyptus grandis* phenylalanine and monolignol biosynthetic genes. 115
Table S2.2. Plastid cluster allocation of *Eucalyptus grandis* circadian rhythm genes....116
Notes S2.1. Word document of 17 KEGG carbon metabolism pathways used to identify plastid and mitochondrial carbon metabolism genes in *Arabidopsis thaliana*. 117
Notes S2.2. GO enrichment of plastid, mitochondrial, and shared network clusters, and the overlap between plastid clusters and aspen developmental phases. 117
Notes S2.3. 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. 117

Chapter 3 118

The plastid and mitochondrial genomes of *Eucalyptus grandis*.... 118

3.1. Abstract 119

3.2. Background 120

3.3. Results 121

3.3.1. Genome structure and gene content of the *E. grandis* mitochondrial genome 121

3.3.2. Genome structure and gene content of the *E. grandis* plastid genome 124

3.3.3. Post-transcriptional editing in the organellar genomes of *E. grandis* 126

3.3.4. DNA transfer between organellar and nuclear genomes 128

3.3.5. Transcription of NUMT and NUPT genes in *E. grandis* 129

3.4. Discussion 131

3.5. Conclusion 134

3.6. Methods 134

3.6.1. Assembly and annotation of the *E. grandis* organellar genomes 134

3.6.2. Identification and analysis of NUMTs and NUPTs in the *E. grandis* nuclear genome.....136

3.6.3. PolyA-selected mRNA sequencing alignment, quantification, and editing analysis 137

3.7. Abbreviations 138

3.8. Availability of data and material 139

3.9. Acknowledgements 139

3.10. References 139

3.11. Tables 150

Table 3.1. *E. grandis* organellar genome characteristics 150

Table 3.2. *E. grandis* mitochondrial genome gene content 151

3.12. Figures 152

Figure 3.1. Mitochondrial genome of *E. grandis*. 152

Figure 3.2. Plastid genome of *E. grandis*.....154

Figure 3.3. DNA and gene transfer between nuclear and organellar genomes in *E. grandis.* ...156

Figure 3.4. Poly-A selected organellar transcriptomes.....159

3.13. Additional files 160

Figure S3.1. Discordantly mapped read pairs flanking direct repeat 13 of the mitochondrial genome. 160

Figure S3.2. Mitochondrial gene order comparison between *Eucalyptus grandis* and *Lagerstroemia indica*. 161

Figure S3.3. Multiple whole genome alignment of selected land plant mitochondrial genomes. 162

Figure S3.4. Number of predicted C to U editing sites in the mitochondrial and plastid genomes of *E. grandis* 164

Figure S3.5. Sashimi plot of polyA-selected mRNA reads mapped to *Eucgr.E01203* in *E. grandis* mature leaf tissue. 165

Table S3.1. *E. grandis* mitochondrial and plastid genome short repeat elements overview 165

Table S3.2. Amount of *E. grandis* organellar DNA transfer to nuclear chromosomes 165

Table S3.3. Inter-organellar DNA transfers in *E. grandis* show regions of high homology between *E. grandis* plastid and mitochondrial genomes. 165

Table S3.4. Differentially expressed organellar-encoded genes in *E. grandis*. 166

Additional file 3.1. Large mitochondrial repeats. 166

Additional file 3.2. Mitochondrial genome structural variant analysis 166

Additional file 3.3. Predicted and confirmed mitochondrial and plastid transcript editing sites... 166

Additional file 3.4. Inter-organellar DNA transfer 166

Additional file 3.5. Transcription of nuclear genes of organellar origin 167

Chapter 4 168

Transcriptional coordination between nuclear and organellar genomes in carbon source, transport and sink tissues in Eucalyptus grandis 168

4.1. Summary 169

4.2. Introduction 170

4.3. Methods 174

4.3.1. Tissue sample preparation and sequencing 174
4.3.2. Organellar genome copy number estimation 174
4.3.3. Total, polyA, and small RNA transcriptome analysis 175
4.3.4. Organellar transcript editing analysis 177

4.4. Results 178

4.4.1. Plastid and mitochondrial genome copy numbers vary between carbon source, sink, and transport tissues, but are not directly related to organellar transcription 178
4.4.2. Mitochondrial genes are ubiquitously expressed, while plastid genes are differentially regulated across tissues 182
4.4.3. Post-transcriptional editing in *E. grandis* plastid and mitochondrial genomes 184
4.4.4. Tissue specific small RNA footprints highlight nuclear-encoded RNA binding protein regulation of organellar transcriptomes 187
4.4.5. Organellar targeted gene regulation in *E. grandis* reflects tissue-specific metabolic processes 189

4.5. Discussion 195

4.6. Data availability 200

4.7. Acknowledgements & funding 200

4.8. Author contributions 201

4.9. References 201

4.10. Figures 216

Figure 4.1. Relative DNA copy number and total RNA derived reads aligned to the mitochondrial and plastid genomes of *E. grandis*. 216
Figure 4.2. Average tissue specific expression of plastid and mitochondrial encoded genes. 219
Figure 4.3. Plastid gene editing sites in *E. grandis* carbon source, sink and transport tissues 220
Figure 4.4. Tissue-specific abundance of cosRNAs in the *Eucalyptus grandis* plastid genome, and association of differentially abundant plastid cosRNAs and genes. 222

Figure 4.5. Expression of plastid and mitochondrial targeted genes in *E. grandis* mature leaf, immature xylem, and phloem. 225

4.11. Supplementary Tables 226

Table S4.1:. RNA-sequencing mapping statistics 226

4.12. Supplementary Figures 228

Figure S4.1. Plastid and mitochondrial genome copy number relative to diploid nuclear genome copy number in five *E. grandis* tissues at different developmental stages. 228

Figure S4.2. Average read depth per feature in sense and anti-sense orientation in (a.) mitochondria and (b.) plastids. 229

Figure S4.3. Percentage of polyA selected RNA seq reads mapped to the mitochondrial and plastid genomes in ML, IX, and PH. 230

Figure S4.4. Correlation between total and polyA selected RNA seq read counts for *Eucalyptus grandis* plastid and mitochondrial genes in mature leaf, immature xylem, and phloem. 231

Figure S4.5. Summary of identified editing sites in the plastid and mitochondrial genomes of *E. grandis* in mature leaf, immature xylem, and phloem 233

Figure S4.6. Mitochondrial editing and read alignment 234

Figure S4.7. Plastid and mitochondrial cosRNA genomic feature overlap and abundance in *E. grandis* mitochondria 236

Figure S4.8. Biochemical pathway and tissue specific expression of *E. grandis* mitochondrial targeted photorespiration genes. 238

4.13. Supplementary data files 239

Supplementary file 4.1 - *E. grandis* plastid gene expression quantification 239

Supplementary file 4.2 - *E. grandis* mitochondrial gene expression quantification 239

Supplementary file 4.3 - *E. grandis* plastid and mitochondrial differentially expressed genes 239

Supplementary file 4.4 - Plastid and mitochondrial encoded transcript editing sites across tissues in *E. grandis* 240

Supplementary file 4.5 - Polyadenylated reads as a % of total read count in *Eucalyptus grandis* plastid and mitochondrial encoded genes 240

Supplementary file 4.6 - *E. grandis* plastid and mitochondrial cosRNA position, orientation to closest genomic feature, and abundance across tissues 240

Supplementary file 4.7 - *E. grandis* plastid and mitochondrial targeted gene expression across tissues, and gene expression of known and potential plastid targeted RNA binding proteins ... 241

4.14. Supplementary Methods 242

Chapter 5 249

5.1. Concluding remarks 249

5.2. References 260

Thesis summary

Systems biology of organellar carbon metabolism in *Eucalyptus* xylogenesis

Desré Pinard

Supervised by **Prof Eshchar Mizrachi**

Co-supervised by **Prof Alexander A. Myburg**

Submitted in partial fulfillment of the requirements for the degree **Philosophiae Doctor**

Department of Biochemistry, Genetics, and Microbiology

University of Pretoria

Trees play a pivotal role in the global carbon cycle through the fixation and storage of carbon as polysaccharide and phenolic biopolymers in the secondary cell walls of xylem fibre cells (wood). Driven by the need to mitigate climate change, the utilization of wood as a source of renewable lignocellulosic biomass is undergoing a technological renaissance. Fundamental research in understanding and modelling xylogenesis as a developmental program remains crucial, as most carbon in living biomass is present in the secondary growth tissues of terrestrial plants. Although systems biology approaches in poplar, *Eucalyptus* and *Arabidopsis* have made

massive strides in unravelling the complex genetic regulation underpinning xylogenesis, we know little of how the genome-bearing organelles of the cell, the plastids, and mitochondria, are integrated into the system. As the location of many carbon metabolic pathways in the cell, plastids and mitochondria play a crucial role in carbon allocation during wood formation. Despite this, plastid and mitochondrial biology have largely been ignored in the study of xylogenesis, and the vast majority of plant organellar research is focused on photosynthetic tissues.

The work presented in this thesis aimed to provide a basis for understanding the role of plastid and mitochondrial biology during xylogenesis, using *Eucalyptus* as a model. This was achieved through three research objectives: (i.) The analysis of coregulated expression modules of nuclear-encoded plastid and mitochondrial-targeted genes generated from 156 xylem transcriptomes from a *Eucalyptus grandis* × *E. urophylla* F2 interspecific backcross population; (ii.) The assembly and annotation of the plastid and mitochondrial genomes of *E. grandis*, along with analysis of intergenomic DNA transfers and their expression; and (iii.) Comprehensive analysis of the transcriptomes of *E. grandis* plastids and mitochondria using total RNA, polyA-selected RNA, and small RNA sequencing data to understand the regulation of plastids and mitochondria in three tissues representing carbon source (mature leaf), sink (immature xylem) and transport (secondary phloem) tissues.

This research has improved our understanding of carbon allocation during xylogenesis, provided a resource for future studies in *Eucalyptus* organellar biology and is the first to look at the transcriptomes of secondary growth plastids and

mitochondria. Important findings from this work include that the regulation of plastid and mitochondrial metabolism is highly integrated with xylem development and the circadian clock. Plastid specific associations with the central circadian clock and epigenetic regulation show that the central functions of plastid retrograde signalling in leaf and green vascular tissues are conserved in wood formation. Finally, analysis of organellar transcriptomes in multiple tissues has shown for the first time that nuclear-encoded polymerases uniquely drive plastidial gene expression during wood formation and that (as yet unknown) nuclear-encoded RNA-binding proteins may be involved in the active upregulation of selected plastid-encoded genes to facilitate non-photosynthetic metabolism in the plastids of tree sink tissues.

In conclusion, the thesis makes an evidence-based argument for a specific plastid type – the xyloplast – to advance research in the field of wood formation and organellar biology.

Preface

In trees, atmospheric carbon is sequestered in the leaves by photosynthesis and is deposited in large amounts as lignocellulosic biomass in wood. Lignocellulosic biomass is a permanent carbon sink in living plants and is a valuable resource for industrial applications including pulp and paper production, construction timber, biofuels, and cellulose-derived products. Due to a short rotation time and superior fibre qualities, *Eucalyptus* is the most widely planted hardwood crop species with ~20 Mha planted globally and is becoming a research model for wood formation. In South Africa, the forestry industry was valued at ~R30 billion (~$1.9 billion) in 2017, with the production of 10 million tons of biomass, of which 43% was derived from *Eucalyptus* species and their hybrids. The South African government has prioritized economic strategies that utilize biological resources in a sustainable manner, known as the bio-economy, where the forestry industry is a key player. To maximize the yield of woody biomass through selective breeding and biotechnological solutions, a comprehensive understanding of the process of wood formation is essential.

The biological process of wood formation (xylogenesis) in trees is tightly regulated to ensure the efficient partitioning of carbon sequestered during photosynthesis as lignocellulosic biopolymers in wood. In almost all seed plants (gymnosperms and angiosperms - with the notable exception of monocotyledonous angiosperms), vascular cambium cells differentiate to form phloem and xylem, where xylem fibre cells make up the bulk of the wood. The thick secondary cell walls of xylem fibre cells consist of cellulose (40-45%), hemicelluloses (20-25%), and lignin (25-35%), that are

deposited in a coordinated manner, followed by programmed cell death to provide water conductivity and mechanical strength to the stem/trunk.

The regulation of secondary cell wall biosynthesis has been extensively researched to develop a systems biology model of xylogenesis and uncovering a complex hierarchy of transcription factors and metabolic genes that regulate carbon partitioning and metabolism. In the Forest Molecular Genetics research group, the regulation of xylogenesis and the partitioning of carbon between secondary cell wall biopolymers has been studied using next-generation sequencing, molecular biology and population genetics in *Eucalyptus*. Specifically, xylem-specific tissue transcriptomes, metabolomes, and phenotyping of an *E. grandis x E. urophylla* hybrid backcross population have been invaluable in developing a systems genetics model of xylogenesis. In these studies, and the study of wood formation in general, a critical aspect that has been relatively understudied is the spatial partitioning of metabolites *within* the wood fibre cell, and the regulation of subcellular metabolism to facilitate high rates of lignocellulose production.

Notably, plastids and mitochondria play a central role in the spatial regulation of carbon metabolism. During wood formation, carbon is partitioned between sugars and phenolic biopolymers by the plastid localized shikimate pathway. The shikimate pathway produces aromatic amino acids from precursor sugars that can also be used for cellulose and hemicellulose biosynthesis. The aromatic amino acid phenylalanine is utilized in the phenylpropanoid pathway for monolignol production in developing secondary xylem, making the shikimate pathway and plastidial metabolism an

important area of study for xylogenesis. Plastid and mitochondrial localized metabolic pathways are also involved in the production of other amino acids, fatty acids, hormones, co-enzymes, reactive oxygen species, and starch, all of which are integral to xylogenesis. As endosymbiont derived organelles, plastids and mitochondria contain minimal genomes, which encode genes that must be co-ordinately regulated along with the nuclear genes that encode plastid and mitochondrial-targeted proteins. In wood forming tissues, little is known about the regulation of organellar transcription and metabolism, despite their importance to xylogenesis.

To understand the role and transcriptomic regulation of organelles in wood forming tissues, and to lay a comprehensive foundation for the study of organelles during secondary growth and xylogenesis, the research presented here was structured to answer three main questions: (i.) How are nuclear-encoded carbon metabolic genes targeted to plastids and mitochondria regulated and coordinated with the developmental processes of secondary xylem growth in *Eucalyptus*? (ii.) From the sequence and gene content of the *E. grandis* mitochondrial and plastid genomes, do organellar genome transfers result in functional nuclear transcription of organellar proteins? (iii.) Does the organellar transcriptomic landscape in *Eucalyptus* xylem and phloem reflect a unique mechanism of organellar-encoded gene regulation specific to wood formation compared to other non-photosynthetic carbon sink tissues?

The main findings that have emerged from this research are that plastid-targeted genes are transcriptionally coordinated with distinct phases of xylogenesis, and are specifically associated with the circadian clock and epigenetic regulation in contrast to

mitochondrial-targeted genes (Chapter 2). Due to the evolutionary history of endosymbiont genomes in plants, a complete analysis of organellar genome sequences and horizontal genome transfers was necessary before analysis of the organellar transcriptomes. Chapter 3 details the assembly and annotation of the mitochondrial genome of *Eucalyptus*, the revised annotation of the plastid genome, and excludes functional transcription from nuclear copies of organellar genes. In Chapter 4, a comprehensive analysis of nuclear and organellar transcriptomes in *Eucalyptus* leaf, immature xylem, and phloem shows that plastid-encoded gene expression in xylem and phloem is uniquely and actively regulated compared to other heterotrophic plastids. The thesis concludes with motivation for defining a new type of plastid in woody tissues (the xyloplast), and potential applications to wood systems biology derived from the research presented (Chapter 5).

The following peer-reviewed publications and conference presentations have emanated from this PhD work:

Publications in ISI rated Journals

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

Pinard D, Fierro AC, Marchal K, Myburg AA, Mizrachi E. 2019. Organellar carbon metabolism is coordinated with distinct developmental phases of secondary xylem. *New Phytologist*. (**https://doi.org/10.1111/nph.15739**)

Pinard D, Myburg AA, Mizrachi E. 2019. The plastid and mitochondrial genomes of *Eucalyptus grandis*. *BMC Genomics*. **20**:132

Wierzbicki MP, Christie N, Pinard D, Mansfield SD, Mizrachi E, Myburg AA. 2019. A systems genetics analysis in *Eucalyptus* reveals coordination of metabolic pathways associated with xylan modification in wood forming tissues. *New Phytologist*. (**https://doi.org/10.1111/nph.15972**)

Presentations in national (South African) conferences

Mizrachi E, Fierro Gutiérrez AC, Pinard D, Verbeke L, Van der Merwe K, Balkwill CG, Hefer CA, Christie N, Myburg AA. 2014. A systems genetics approach reveals

transcriptional coordination of carbon partitioning and subcellular compartmentation in developing wood. *SA Genetics Society (SAGS) & SA Society for Bioinformatics (SASBi) Joint Congress* (oral presentation).

Pinard D, Behrens D, Myburg AA, Mizrachi E. 2018. Transcriptomics of wood forming plastids in *Eucalyptus grandis*: evidence for a new plastid type in vascular plants. *SA Genetics Society (SAGS) & SA Society for Bioinformatics (SASBi) Joint Congress* (oral presentation).

Presentations in international conferences

Pinard D, Fierro AC, van de Peer Y, Marchal K, Myburg AA, Mizrachi E. 2015. Transcriptional networks of plastid targeted genes affecting carbon metabolism in *Eucalyptus* xylem. *IUFRO Tree Biotechnology Conference - Florence, Italy* (oral presentation).

Pinard D, Behrens D, Myburg AA, Mizrachi E. 2017. Unravelling the complex regulation of xylogenic organelles. *XIX International Botanical Congress - Shenzhen, China* (oral presentation).

Pinard D, Behrens D, Myburg AA, Mizrachi E. 2019. Investigating organelles - Regulation of organellar biology in xylem and phloem. *Plant Vascular Biology Conference - Monterey, USA* (oral presentation).

Pinard D, Behrens D, Myburg AA, Mizrachi E. 2019. Transcriptional and post-transcriptional regulation of *Eucalyptus grandis* organellar gene expression in developing xylem. *IUFRO Tree Biotechnology Conference - Raleigh, USA* (oral presentation).

Abbreviations

ATP : Adenosine triphosphate
AAA : Aromatic amino acid
ACC : Acetyl-CoA carboxylase
AMP : Adenosine monophosphate
AOX : Alternative oxidase
BP : Biological process
bp: base pairs
CDKE1 : Cyclin-dependent kinase E1
CORR : Co-location for redox regulation
cosRNA : Clustered organellar sRNA
E-4-P : Erythrose-4-phosphate
eQTL : Expression Quantitative Trait Loci
FA : Fatty acid
FL : Flower
FPKM : Fragments Per Kilobase of transcript per Million mapped reads
G-6-P : Glucose-6-phosphate
GO : Gene ontology
GUN : Genomes uncoupled
IR : Inverted repeat
IX : Immature xylem
KEGG : Kyoto Encyclopedia of Genes and Genomes
LSC : Large single copy
MC : Mitochondrial cluster
MEcPP : 2-C-methyl-D-erythritol 2,4-cyclodiphosphate
MEP : Methylerythritol 4-phosphate
ML : Mature leaf
NADH : Nicotinamide adenine dinucleotide (NAD) + hydrogen (H)
NBDI : Network-based data integration
NEP : Nuclear-encoded polymerase
NUMT: Nuclear mitochondrial DNA
NUPT: Nuclear plastidial DNA
P-E-P : Phosphoenol pyruvate
PAP : 3′-phosphoadenosine 5′-phosphate
PC : Plastid cluster
PCD : Programmed cell death
pdNAD-MDH : Plastidial NAD-malate dehydrogenase
PEP : Plastid-encoded polymerase
PEPCK : PEP carboxykinase
PH: Phloem
PNPase : Polynucleotide phosphorylase
polyA: polyadenylated

PPDK : pyruvate, orthophosphate dikinase
PPP : Pentose phosphate pathway
PPR : Pentatricopeptide repeat
PPT/CUE1 : Phosphoenolpyruvate-phosphate translocator/Chlorophyll a/b binding protein (CAB) gene underexpressed 1
qPCR : Quantitative polymerase chain reaction
RNAbp : RNA binding protein
ROS : Reactive oxygen species
SAM : S-adenosyl methionine
SCW : Secondary cell wall
siRNA : Small interfering RNA
SNP: Single nucleotide polymorphism
SSC : Small single copy
ST : Shoot tip
SUBA : SUBcellular localization database for Arabidopsis proteins
TAL : Transaldolase
TCA : Tricarboxylic acid
TIC : Translocon at the inner envelope of chloroplasts
TK : Transketolase
TOC : Translocon of the outer envelope of chloroplasts
UTR: Untranslated region
VST: Variance stabilizing transformation
WGD: Whole-genome duplication
WGS: Whole genome sequencing
XRN : Exoribonuclease
YCF : Hypothetical chloroplast open reading frames
YL : Young leaf

Acknowledgements

- Professor Eshchar Mizrachi - Thank you for being an excellent mentor and supervisor, your passion and creativity have inspired me to always aim for excellence. Your trust and faith in me have allowed me to grow my scientific career in ways that I never imagined. Although your five-year plan messages failed to get through to me, your endless curiosity and ability to bring excitement to any topic have not. These qualities are so important to me as a scientist and student, and I cannot thank you enough for fostering them in me with patience and understanding.
- Professor Alexander Myburg - Zander, I am so proud to be a part of what you have built at FMG. I am very grateful for everything you have done for me, and extremely thankful that you have let me pursue the vagaries of my mind with patient guidance and supervision. You always make sure that we are exposed to scientific experts, and by co-workers that have the same vision and passion for science. Thank you for all the work you do to ensure that we experience the very best that academia has to offer. Above all, I will never forget your unwavering passion and support.
- My Belgian collaborators - Thomas; you made being away from home a pleasure; Yves: Your passion and keen eye for the important story is an essential part of this degree. Kathleen: Thank you for all your contributions and support, and for showing me the importance of taking care in doing the small things to allow the science to shine.

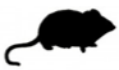
- The Honor's students that I have supervised, especially Daniel and Mike; thank you guys for teaching me that I always have things to learn, and for making teaching a pleasure.
- All the students and staff at the Department of Biochemistry, Genetics, and Microbiology and FABI at the University of Pretoria for the supportive and rewarding work environment. Special thanks to Prof. Christine Maritz Olivier for guiding me as an educator, and Prof. Paulette Bloomer for the support you have given me since Honours.
- Everyone at Bioinformatics, especially Karen, Werner, Alisa, Fourie and Johann for all the support, help, and the best problem-solving discussions.
- Special thanks to Sappi, The University of Pretoria, the Department of Science and Technology, The Department of Trade and Industry, and the National Research Foundation for funding this research.
- I have been exceptionally privileged to have so many excellent mentors and examples of scientific excellence and experience to inspire me. To Nicky, Marja, Vinette, Ouma, and Vicki, thank you for all the help, discussions, and support over the years, you ladies are the cream of the crop, and I will always try to live up to the example you all set on a daily basis.
- Everyone at the Forestry Molecular Genetics Programme - especially Mel, Patience, and Thandeka - thank you for making FMG the best working environment that anyone could ever ask for with a daily commitment to team-work and scientific excellence.
- All my friends for their love and support, I could never have come so far without you guys! Special thanks to the most amazing group of friends and scientists -

Martin, Jono, Karen, Marja, Ritesh, Pooja, Caryn, Kat & Steps, Drew & Gabs - all the discussions, coffee, and beers have been the greatest gift!

- It takes more than a village to complete a PhD, and all my extended family have given me so much love and support - Uncle Richard and Aunty Tersia, Uncle William, Frederik, Erin, Uncle Colin and Aunty Cath, thank you for everything!
- My sister and brother - Charlé and Russell. Thank you so much for all the support and laughs over a lifetime, you two are such an inspiration to me, and I love you so much!
- Mom and Dad - No words can ever express my gratitude for everything you do for me! Thank you for all the love and support, and for being the best examples of hard work, commitment, generosity, and dedication to family I have ever seen. You guys make even the most difficult times easy.

Dedicated to my parents - Colin and Mercia Pinard

Everything I am proud of and value in myself is a direct result of you.



Chapter 1

LITERATURE REVIEW

The role of plastids and mitochondria in wood formation

Desré Pinard[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

This chapter has been prepared in the style of a review manuscript for a peer-reviewed research journal. The review publication Pinard & Mizrachi (2019, Current Opinion in Plant Biology) gives additional detail to this chapter - specifically related to non-photosynthetic plastids - and can be found in the Supplementary Information.

1.1. Introduction

Plastids and mitochondria are semi-autonomous subcellular organelles that evolved through the endosymbiosis of free-living, single-celled organisms into the ancestor of plants (Gould *et al.*, 2008; Gray & Archibald, 2012; Roger *et al.*, 2017), leading to novel metabolic pathways being incorporated into the cell (Martin *et al.*, 2002; Husnik & McCutcheon, 2018). During the process of endosymbiosis, the genome of the endosymbiont is generally vastly reduced, with the majority of genes being transferred to the nuclear genome of the cell (Adams *et al.*, 2002; Timmis *et al.*, 2004). This endosymbiotic relationship results in an expanding repertoire of transporter proteins, regulatory, and signalling mechanisms to ensure the optimal fitness of both the endosymbiont and host (Rand *et al.*, 2004). Plastids and mitochondria are the bioenergetic organelles in plant cells, with their double membranes a testament to their endosymbiotic origin, and the need for redox sub-compartmentation (Allen, 2003). Mitochondria produce the adenosine triphosphate (ATP) necessary for metabolism via the electron transport chain proteins embedded in the inner membrane (Jacoby *et al.*, 2012). Similarly, photosynthetic plastids convert atmospheric carbon dioxide into sugars that form the building blocks of life through the action of membrane-embedded electron transport proteins (Halliwell, 1984). In addition to these fundamental processes, a host of plant metabolic pathways are located in plastids and mitochondria, many of which are essential to secondary growth and the formation of wood (Deschamps *et al.*, 2008; Tohge *et al.*, 2013). These organelle-targeted metabolic pathways have largely been ignored during the process of xylogenesis, where the vast majority of studies do not take the contribution and regulation of these organelles into account.

In trees, secondary growth xylogenesis (wood formation) results in the deposition of lignocellulosic biopolymers in the secondary cell walls (SCWs), specifically of vessel and fibre cells (Plomion *et al*., 2001). These SCWs in wood contain a large amount of sequestered carbon and are important ecologically and economically renewable carbon sinks (Boerjan, 2005; Tollefson, 2017). Recently, systems biology and genetics have emerged as tools to understand and improve the process of xylogenesis in economically important tree crop species such as *Eucalyptus* and *Populus*, which make up a large proportion of planted forests worldwide (Sundell *et al*., 2017; Mizrachi *et al*., 2017; Tuskan *et al*., 2018). Systems biology aims to model organisms as systems by holistically integrating the various levels of metabolism, regulation, and interactions occurring at the cell, tissue, and organ level (Hammer *et al*., 2004; Robinson *et al*., 2014). Given the complex regulation, signalling, and transport mechanisms governing subcellular metabolism in plants, and the critical contribution of plastids and mitochondria to plant evolution, the subcellular compartmentalization of the plant cell should not be ignored when modelling xylogenesis as a system.

Since photosynthesis has been the primary research target of plastid biology, and that mitochondria have been relatively inadequately studied in plants compared to plastids, we know surprisingly little about organelles in secondary growth and wood-forming tissues. Here, the current knowledge of organellar metabolism, evolution, regulation, and integration into xylogenesis in trees is reviewed.

1.2. Organellar targeted metabolism is central to xylogenesis

In angiosperm trees, sucrose produced in the source tissues (leaves) is transported via phloem to the xylem, where secondary cell wall deposition and wood formation occurs (Plomion *et al.*, 2001). Trees can be used as a model system for lignocellulosic biomass production, as they fix a majority of their sequestered carbon as cellulosic (cellulose and hemicellulose) and phenolic (lignin) biopolymers in the secondary cell walls of xylem (permanent sink tissue) vessel and fibre cells (Mellerowicz & Sundberg, 2008). Transported sucrose is used as an energy source in the sink tissue for cellular metabolism and homeostasis, and secondary cell wall biopolymer formation (Koch, 2004). The secondary cell wall is composed of crystalline cellulose microfibrils surrounded by hemicelluloses and lignin (Bailey, 1938). The metabolic processes responsible for carbon allocation and partitioning to secondary cell wall biopolymers such as cellulose and lignin are highly interconnected to primary carbon metabolism, secondary metabolite production, and the homeostasis of individual cells and tissues. A holistic understanding of this network in woody tree species will allow for improved mechanisms of redirecting carbon flux towards cellulose synthesis with minimal disruption of the connected processes.

Xylem is a heterogeneous tissue consisting of vessel, fibre, and parenchyma cells, which have distinct cellular morphologies reflecting their functional roles (Fukuda, 1996). Xylogenesis begins with the differentiation of vascular meristem cambial cells, followed by cellular expansion, SCW deposition, and finally programmed cell death (PCD). The secondary cell walls of xylem fibres and vessels are formed after cell division and elongation have taken place (Kaneda *et al.*, 2010). Cellulose,

hemicellulose, and lignin are synthesized and deposited in a coordinated fashion to form the thick secondary cell walls needed by the tree for mechanical support (fibres) and solute transport (vessels) (Plomion *et al.*, 2001; Chaffey *et al.*, 2002). Xylem vessels and fibres undergo cell death at differing rates, with vessels undergoing rapid cell death to form hollow lignified tubes for solute transport, while fibre cells are characterized by a more controlled, slower partial cell death (Courtois-Moreau *et al.*, 2009; Bollhöner *et al.*, 2012). Ray parenchyma cells are long-living, and often do not undergo PCD for decades, if ever (Bollhöner *et al.*, 2012; Nakaba *et al.*, 2015). Systems-level analyses have been used to great success towards identifying the molecular components and regulatory networks that govern xylogenesis (Ruprecht *et al.*, 2011; Li *et al.*, 2016, Sundell *et al.*, 2017; Mizrachi *et al.*, 2017; Sibout *et al.*, 2017; Tuskan *et al.*, 2018). A confounding factor to modelling these processes at a systems level is the spatial variation of metabolic pathways between cell types, the subcellular compartmentation of pathways within the cell, and temporal variation over days and seasons (do Amaral & Souza, 2017). Our understanding of organellar biology in xylem fibre cells is limited, due to the difficulty in isolating live fibre cells from the heterogeneous cell mixture that characterizes differentiating xylem.

Carbon flux within individual plant cells at the tissue level is tightly controlled not only by transcriptional and protein regulation and modification, but also via partitioning into organelles, subcellular localization, and transporter activity (Sweetlove & Fernie, 2013). Many pathways involved in primary and secondary carbon metabolism in plants are compartmentalized to plastids and mitochondria: including the Calvin cycle, oxidative pentose phosphate pathway, and the shikimate pathway in the plastid, and

the tricarboxylic acid cycle in the mitochondria (Sweetlove & Fernie, 2013; Heinig *et al*., 2013). The organelles in the leaf, flower and root tissues of plants have been well characterized in terms of their morphology, ontology and metabolism (Rolland *et al*., 2018; Wang *et al*., 2018b). In wood forming tissues, there is a general understanding of the contribution of organellar targeted metabolism to secondary cell wall biosynthesis, particularly in the biosynthesis of monolignols from phenylalanine produced by the plastidial shikimate pathway.

The shikimate pathway is responsible for the production of lignin monomer precursor phenylalanine in angiosperms, and tyrosine in grasses (Henkes *et al*., 2001; Rippert *et al*., 2009; Novaes *et al*., 2010; Maeda, 2016). Systems biology analysis of *Arabidopsis* mutants showed that plastid-targeted metabolism is widely affected by perturbing key enzymes in the cytoplasmic phenylpropanoid pathway, indicating feedback between these subcellular compartments (Vanholme *et al*., 2012). Shikimate pathway precursors phosphoenolpyruvate (P-E-P) and erythrose-4-phosphate (E-4-P) are derived from sugars in xylem that can be used in other metabolic pathways, many of which are also plastid localized. P-E-P can be generated from cytosolic or plastidial glycolysis, depending on the presence/absence of plastid-targeted enolase enzyme (Prabhakar *et al*., 2009), from pyruvate by dual localized pyruvate, orthophosphate dikinase (PPDK) (Parsley & Hibberd, 2006), or from oxaloacetate by cytosolic PEPCK (P-E-P carboxykinase) (Leegood & Walker, 2003). Conversely, E-4-P is produced in non-photosynthetic tissues by the plastid localized pentose phosphate pathway (PPP) from glucose-6-phosphate (G-6-P) (Kruger & von Schaewen, 2003). G-6-P is also the main precursor for UDP-glucose, which is the

monomer for cellulose production by the cellulose synthase complex at the plasma membrane (Brett, 2000). Sugars transported via the phloem are therefore irreversibly allocated to the phenylpropanoid pathway via the shikimate pathway in the xylem plastid, making the transport and regulation of sugar metabolites in the plastid a key flux point for secondary cell wall biopolymers (Novaes *et al*., 2010).

Plastid localized metabolism also makes significant contributions to the initial stages of xylem development by the *de novo* synthesis of fatty acids for phospholipid membranes for cell division and elongation (Maréchal *et al*., 1997). Xylogenesis and secondary cell wall formation are regulated by a wide array of phytohormones (Uggla *et* al., 1996; Eriksonn *et al.,* 2000; Didi *et al*., 2015), many of which have precursors that are synthesized in plastids: auxin - derived from tryptophan (Ursache *et al*., 2014); cytokinins, gibberellin, abscisic acid, and strigolactones - derived from the methylerythritol phosphate pathway (Vranová *et al*., 2013); and methyl jasmonate derived from fatty acids (Schmid, 2015). In addition to the aromatic amino acids, Arg, Lys, Thr, Leu, Ile, Val, and His are exclusively produced in plastids and are vital to protein synthesis (Rolland *et al*., 2012). Starch synthesis and storage in the plastid also plays a role in xylogenesis through the supply of energy overnight (Solomon *et al*., 2010), and the gravitropic response in tension wood formation in poplar stems (Gerttula *et al*., 2015).

Mitochondrial-targeted metabolism during xylogenesis is less understood than plastid-targeted biosynthetic pathways, even though they provide energy for xylogenesis via oxidative phosphorylation and the tricarboxylic acid cycle, which also provides amino

acid intermediates and nicotinamide adenine dinucleotide (NAD) + hydrogen (H) (NADH) (Fernie *et al*., 2004; Jacoby *et al*., 2012). Additionally, mitochondria are involved in developmental PCD in *Arabidopsis* tracheary elements (Yu *et al*., 2002; Yao *et al*., 2004; Van Aken, 2018). Direct evidence for an important regulatory role of mitochondrial-targeted glycine metabolism in lignin biosynthesis has been shown in pine compression wood formation, and functions to prevent inhibition of S-adenosyl methionine (SAM) synthesis for the phenylpropanoid pathway (Villalobos *et al*., 2012).

Plastid and mitochondrial-targeted metabolism are therefore critical to all stages of xylogenesis in trees, and the regulation of this metabolism is complicated by the semi-autonomous nature of plastids and mitochondria. Although the localization of metabolic pathways to the organelles during xylogenesis is understood for the most part, how these pathways are regulated between the nuclear and organellar genomes is not.

1.3. The endosymbiotic legacy of organelles in land plants

It is estimated that ~18% of *Arabidopsis thaliana* nuclear genes are derived from the cyanobacterial ancestor of the plastid (Martin *et al*., 2002; Raven & Allen, 2003), and 375 gene families in *A. thaliana* are derived from the α-proteobacterial ancestor of mitochondria (Gabaldón & Huynen, 2005). Recent work has shown that in addition to these ancient transfers of genes that have since been lost from the organellar genes, plastid protein domains are contributing to novel protein emergence of 67 gene families in plants and algae (Méheust *et al*., 2016), along with novel domains derived

from mitochondrial genes (Lama *et al*., 2019). Plant genomes have dynamic evolutionary histories since the endosymbiosis of the organellar ancestors, with some lineages undergoing multiple recent whole-genome duplications (WGD) (Proost *et al*., 2011). Although a large proportion of genes that undergo duplication following WGD are retained for dosage balance or gain novel functions, some organellar targeted genes rapidly return to single-copy status to maintain the tight coordination between the cytonuclear genomes (Woodson & Chory, 2008; De Smet *et al*., 2013). Endosymbiosis in land plants has also resulted in the emergence and expansion of gene families that regulate organellar biology (de Vries *et al*., 2016), and the most predominant gene family in this class are the pentatricopeptide repeat-containing proteins (PPRs) (Fujii & Small, 2011). This gene family is greatly expanded in land plants (> 400 genes), and are involved in the fine-scale regulation of organellar-encoded gene expression (Barkan & Small, 2014). The emergence of novel genes in the nuclear genome via the transfer of organellar DNA is ongoing, and occurs relatively frequently within plant genomes (Michalovova *et al*., 2013).

The process of DNA transfer tends to be lineage-specific and can result in the complete loss of organellar-encoded genes, as they are expressed from the nuclear genome, as seen in the *accD* gene of *Trachelium caeruleum* (Rousseau-Gueutin *et al*., 2013). As organellar-encoded genes are transcribed using bacterial-like polymerases and promoters, successful integration of organellar-encoded genes is presumed to rely on the acquisition of eukaryotic transcription elements (Timmis *et al*., 2004). These eukaryotic transcription elements can be brought close to the organellar gene by insertional effects or deletion of upstream genes, bringing the organellar gene

into the proximity of nuclear promoters (Chen *et al*., 2015). Interestingly, emerging evidence suggests that this may not always be the case, as the promoter of plastid-encoded *16s rrn* has been shown to drive reporter gene expression in the nuclear genome of rice (Wang *et al*., 2014). Despite this ongoing gene transfer and subsequent loss in organellar genomes, there are tight constraints as to which genes may be completely transferred to the nuclear genome in plants. These constraints range from the co-location for redox regulation (CORR) hypothesis of organellar genome retention, which states that organellar genes need to be maintained to minimize the harmful effects of redox by rapid regulation of electron transport genes, and the limited capacity of organellar membrane transporters to import certain proteins (Björkholm *et al*., 2015; Allen & Martin, 2016). An excellent illustration of these hypotheses can be seen in parasitic plant genomes, which have lost the ability to photosynthesize, but maintain a minimal plastid genome (Wolfe *et al*., 1992; Barrett *et al*., 2014).

The ongoing transfer of organellar-encoded genes to the nuclear genome has resulted in interesting research into the mechanism of transfer and subsequent silencing of these genes. Organellar DNA is inserted into the nuclear genome under conditions of stress or embryogenesis, where the organellar genome disintegrates, and fragments pass through to the nucleus and are incorporated by nuclear double-stranded break repair mechanisms (Wang *et al*., 2012; Wang & Timmis, 2013). Organellar DNA tends to be inserted in proximity to transposable element repeat regions, and may span the entire organellar genome, as seen with the *A. thaliana* mitochondrial genome, and the poplar and rice plastid genomes (Noutsos *et al*., 2005; Huang *et al*., 2005; Tuskan *et*

al., 2006; Michalovova *et al*., 2013). As mentioned above, these genes can be expressed from their native promoter, however, in most cases, they undergo rapid mutational decay via 5-methylcytosine hypermutation (Huang *et al*., 2005; Wang *et al*., 2018a). Further, recent DNA insertions are transcriptionally silenced in the nuclear genome, and clues to the mechanisms of this silencing have been found using genome-wide DNA modification data. This research has found that nuclear organellar DNA has a high frequency of nonCG context methylation which represses gene expression (Niederhuth *et al*., 2016). These events are detected and differentiated from the plastid genome, as plastid DNA is methylation resistant, and is used as a control in epigenetic studies in plants (Fojtová *et al*., 2001; Lister *et al*., 2008). The methylation of organellar DNA insertions may be due to their proximity to transposable element regions, where nonCG methylation silences aberrant transcription (Ito & Kakutani, 2014).

Organellar genomes have thus played a large role in the evolution of land plant genomes, and ongoing genome transfers are an interesting source of insight into evolution and endosymbiosis (Sloan, 2015; Bock, 2015). Further, gene transfers may also confound genomics research, especially in phylogenetic analyses where plastid and mitochondrial genes are frequently used to infer evolutionary relationships (Song *et al*., 2008).

1.4. Plastid and mitochondrial genome structure

Across land plants, the structure of plastid genomes is well conserved, while mitochondrial genomes vary widely in their size, organization and gene content (Smith & Keeling, 2015). In general, plastid genomes have a quadripartite structure, with a small single copy (SSC) region and a large single copy (LSC) region flanked by two large inverted repeat (IR) regions, designated as IRA and IRB (Wicke *et al.*, 2011). This structural arrangement is conserved across land plants and algae, suggesting that it was a feature of the cyanobacterial ancestor (Turmel *et al.*, 1999). The IR regions of land plant genomes expand and contract in a lineage-specific manner, leading to differences in the sizes of plastid genomes, which range from 120 kb to 160 kb (excluding non-photosynthetic plants) (Wicke *et al.*, 2011; Zhu *et al.*, 2016). Photosynthetic angiosperm plastid genomes harbour between 100 and 120 unique genes, of which ~80 are protein-coding genes (Jansen *et al.*, 2007). In contrast, angiosperm mitochondrial genomes contain ~30 protein-coding genes, with lineage-specific gene losses relatively common (Adams *et al.*, 2002; Gualberto *et al.*, 2014). Despite their reduced gene content compared to plastids, angiosperm mitochondria are relatively large, ranging from 200 kb up to ~1.6 Mb for the *Cucumis sativus* mitochondrial genome and ~11.3 Mb in *Silene conica* (Alverson *et al.*, 2011; Sloan *et al.*, 2012; Gualberto *et al.*, 2014).

This extreme range in genome sizes for mitochondrial genomes is due to rapidly expanding repeat regions and invasions from nuclear, plastid, and bacterial DNA (Alverson *et al.*, 2011). The expansion of mitochondrial genome size due to repeats and horizontal DNA transfer is a hallmark of land plant mitochondria, seen across

gymnosperms, ferns and lycophytes (Liu *et al*., 2012; Guo *et al*., 2016, 2017). Repeat regions in mitochondria are often responsible for structural variations in genome structure, which lead to sub-stoichiometric genome fragment abundances (Preuten *et al*., 2010; Cole *et al*., 2018). Although mitochondrial genomes are typically represented as single circular molecules, evidence suggests that the "master circle" is an exception in plant mitochondrial genomes; rather, mitochondrial genomes are present as several circular, linear, or branched molecules (Sloan, 2013). It has been hypothesized that the mitochondrial genome master circle may be found specifically in meristematic or embryonic tissues to aid complete transfer during cell division (Woloszynska, 2010). Similarly, plastid genomes can be found as linear molecules, concatemers of plastid DNA, or branched structures (Shaver *et al*., 2008). The heterogeneous structures of organellar genomes in land plants have a specific impact on genome assembly, and especially in the case of mitochondria, may play an important role in gene regulation by the sub-stoichiometric shifting of sub-genomic molecules (Preuten *et al*., 2010; Wicke & Schneeweiss, 2015).

1.5. Transcriptional regulation of organellar-encoded genes

Transcriptional regulation in land plant organelles is relatively complex considering the small number of genes encoded by the organellar genomes, sharing features of prokaryotic and eukaryotic expression systems, along with organelle-specific regulatory mechanisms (Weihe *et al*., 2012). In land plant mitochondria, genes are transcribed from multiple promoters, and no gene-specific transcription factors have been identified as yet (Kuhn *et al*., 2005; Hammani & Giegé, 2014). Despite the

relaxed transcriptional regulation in plant mitochondria, the expression levels of individual mitochondrial genes are distinctly regulated in plants (Fujii *et al*., 2011; Fang *et al*., 2012; Hameed *et al*., 2017). Tissue-specific expression of mitochondrial genes tends to be higher in developing tissues such as buds and young leaves while maintaining a consistent ratio between genes (Hameed *et al*., 2017). In flowers and gametophytes, there is specific up-regulation of *atp1*, which is presumed to reflect the differences in respiratory status between tissues (Kalantidis *et al*., 2002; Hu *et al*., 2006; Fang *et al*., 2012). The factors affecting the tissue-specific upregulation of mitochondrial genes have remained elusive, with the expression of mitochondrial-targeted polymerases shown to be loosely related to transcription, and whole-genome copy number also shown to play a limited role in maize (Hameed *et al*., 2017). Further, for specific mitochondrial genes in *Arabidopsis*, sub-genomic copy number showed variation in non-dividing cells between tissues, and was not specifically related to steady-state transcript abundance or respiratory status (Preuten *et al*., 2010). Analysis of the nuclear-encoded components of the mitochondrial respiratory pathway was found to be tightly coordinated in *Arabidopsis* across multiple conditions (Gonzalez *et al*., 2007; Ng *et al*., 2014a), and were also co-regulated with mitochondrial encoded gene expression to a certain extent (Mentzen & Wurtele, 2008). Transcription rates of different subunits of the same protein complex encoded by the mitochondrial genome have been found to differ in plant species, and the coordination of steady-state protein abundance is believed to be modulated post-transcriptionally and post-translationally (Giegé *et al*., 2005; Okada & Brennicke, 2006; Liere & Börner, 2011; Planchard *et al*., 2018).

A major difference between plastid and mitochondrial gene expression is the differential regulation of plastid genes according to plastid type (Cardi *et al*., 2012; Liebers *et al*., 2017). Plastid-encoded genes have functions beyond photosynthesis and transcription, whereas mitochondrial encoded genes are solely involved in cellular respiration and gene expression. Plastid genes can be classified into three general functional categories: i. Photosynthetic subunit genes; ii. Gene expression machinery; and iii. Other functions (Barkan, 2011). As a testament to their prokaryotic origin, most plastid genes are transcribed from operons as polycistronic units that are subsequently processed by endonucleases (Mullet, 1993; Bollenbach *et al*., 2007). Unlike bacterial operons, plastid genes can be dynamically translated from both poly- and mono-cistronic transcripts (Barkan, 1988). Plastid genes are transcribed by two evolutionarily diverse RNA polymerases, the prokaryotic-type RNA polymerase, referred to as the plastid-encoded polymerase (PEP), and two phage-type RNA polymerases, known as nuclear-encoded polymerase (NEP) (Hajdukiewicz *et al*., 1997).

In *Arabidopsis*, three types of NEPs are found, one that is specific to plastids (rpoTp), one that is specific to mitochondria (rpoTm), and one that is dual-localized (rpoTmp) (Liere & Börner, 2007; Swiatecka-Hagenbruch *et al*., 2008). NEPs are generally highly active in the early stages of plastid development, and especially in etioplasts, where they act to transcribe the plastid-encoded subunits of PEP (*rpoA, B, C1,* and *C2*) and other non-photosynthetic plastid-encoded genes (Liebers *et al*., 2017). As chloroplasts develop, PEP takes over as the predominant polymerase, transcribing the photosystem genes needed for photosynthesis, in conjunction with nuclear-encoded

sigma factors that confer increased activity and specificity to PEP gene expression (Schweer, 2010; Lerbs-Mache, 2011). Sigma factors are plastid specific transcription factors that have evolved from the eubacterial ancestor, are present in bacteria and green algae, and are expanded in land plants (Isono *et al*., 1997; Feklístov *et al*., 2014). Through their action of specific binding and association with PEP, they enhance the expression of photosynthetic genes (Feklístov *et al*., 2014; Chi *et al*., 2015). Sigma factors are not essential for the basal transcription of photosynthetic genes in the plastid genome, as determined in cis-element truncation studies for the *psbA* promoter sequence (Link, 1984). The utilization of PEP and NEP in plastid transcription are not as defined as they may seem, as NEP promoters are found for almost all plastid-encoded genes, and these genes can be transcribed at low levels in PEP deficient plastids (Legen *et al*., 2002; Swiatecka-Hagenbruch *et al*., 2007).

As such, a general classification system for plastid-encoded genes defines polymerase usage - Class I: Photosystem genes are transcribed by PEP; Class II: ATP synthase genes and expression system genes are transcribed by NEP and PEP; Class III: *ycf2*, *accD*, and the *rpoB, C1, C2* operon are exclusively transcribed by NEP (Liebers *et al*., 2017). The division of labour between NEP and PEP creates a mechanism for the plant cell to specifically control photosynthetic gene expression while maintaining the essential functions of plastids in non-photosynthetic tissues. Almost all research regarding the expression of non-photosynthetic plastid genes has been generated from photosynthetic mutants, or chemically induced plastid dysfunction studies (Zhelyazkova *et al*., 2012b; Tseng *et al*., 2013). These studies have been crucial to identifying the components of chloroplast development and

function in general, but are not representative of non-photosynthetic plastids due to the pleiotropic effects of reduced photosynthetic capacity (Hernández-Verdeja & Strand, 2018). In non-photosynthetic tissues, chloroplast development is actively repressed by nuclear-encoded CONSTITUTIVE PHOTOMORPHOGENIC 1 (COP1), phytochrome interacting proteins, and auxin signalling mediated by auxin response factors (ARF) ARF7, ARF19, and IAA14 (Kobayashi *et al*., 2012; Hernández-Verdeja & Strand, 2018). Previous studies of plastid-encoded gene expression have shown that photosystem genes are downregulated in non-photosynthetic tissues compared to photosynthetic tissues and that the expression of NEP expressed genes are also repressed, but to a lesser extent (Kahlau & Bock, 2008; Valkov *et al*., 2009).

1.6. Non-photosynthetic plastid gene expression

The two studies by Kahlau and Bock (2008), and Valkov *et al* (2009). are some of the only studies that have examined tissue-specific plastid gene expression between non-photosynthetic and photosynthetic tissues. During the development of tomato fruit plastids, photosynthetic genes are massively down-regulated compared to mature leaf, and the expression of most NEP expressed genes is reduced. As they transition from green fruit to mature red fruit, the NEP driven expression of lipid biosynthetic gene *accD* increases to levels comparable to leaf tissue (Kahlau & Bock, 2008). In potato tubers compared to leaf, *clpP*, PEP and rRNA genes were the least down-regulated, along with *ycf1, ycf2,* and *ycf15* (which at the time had no known functions). Together, these two studies have led to the model of non-green plastid gene expression, in which PEP expressed photosystem genes (Class I) are very lowly

expressed, genetic system genes (Class II) are expressed at a basal level to maintain plastid gene expression of Class III genes (Zhang *et al*., 2012b; Liebers *et al*., 2017). The Class III genes that are critical in non-green tissues from these studies are *clpP*, needed to maintain protein integrity of plastid and nuclear-encoded proteins, and *accD*, encoding a subunit of acetyl-CoA carboxylase (ACC), which catalyzes the first committed step of *de novo* fatty-acid synthesis (Kahlau & Bock, 2008; Valkov *et al*., 2009). The NEP expressed Class III genes are therefore distinctly regulated from photosynthetic genes, enabling their expression in non-photosynthetic tissues where they perform critical functions (Cho *et al*., 2009).

In the time since the tomato and potato non-green plastid transcriptomes were reported, potential functions of *ycf1* and *ycf2* as components of the plastid translocon at the inner envelope of chloroplasts (TIC) system have emerged (Kikuchi *et al*., 2013, 2018). The function of *ycf1* (also known as *tic214*) as an essential TIC component has been questioned due to its loss in grasses (de Vries *et al*., 2015), and by the import of nuclear-encoded proteins in plastid translation deficient spectinomycin treated *Arabidopsis* (Bölter & Soll, 2017). Rather, Bölter & Soll (2017) suggest that *ycf1* is involved in fatty acid biosynthesis ACCase complex assembly due to the upregulation of nuclear-encoded *acc2* in spectinomycin treated plastids. This is however not conclusive evidence, as spectinomycin treatment will also reduce the translation of the plastid-encoded *accD* gene, and the upregulation of nuclear-encoded components of *de novo* fatty acid biosynthesis cannot, therefore, be directly related to YCF1 protein deficiency. More contradictory evidence for the function of *ycf1* is found in the proposed stabilization of the *ycf1* transcript by nuclear-encoded protein PBR1, where

knockdown of *pbr1* results in severe deficiencies in photosystem I assembly (Yang *et al*., 2016). In this research, PBR1 was found to bind strongly to the 5' UTR of *ycf1*, and the *pbr1* mutants had decreased YCF1 protein levels. Conversely, a more recent study using RNA immunoprecipitation (RIP) sequencing found that PBR1 (called BSF in the publication) rather bound the transcripts of *petA*, *petB*, and *petD*, which would explain the photosystem I defects of the PBR1 mutants; and did not identify interactions between PBR1/BSF and *ycf1* transcripts (Jiang *et al*., 2019). Evidence for a protein import function of YCF1 is supported by the copurification of YCF1 with the translocon of the outer envelope of chloroplasts (TOC), and pre-proteins imported via the inner membrane (Kikuchi *et al*., 2013; Nakai, 2015). Additionally, recent mass-spectrometry analysis of plastid envelope protein enrichment compared to crude cell lysate showed with high-confidence that YCF1 is exclusively found at the plastid envelope (Bouchnak *et al*., 2019).

Recently, a function for YCF2 in the TIC protein import AAA ATPase motor complex has been proposed by the same research group that identified *ycf1* as *tic214* (Kikuchi *et al*., 2018). Similar to *ycf1*, *ycf2* has been lost from the plastid genomes of monocots (Drescher *et al*., 2000), and is also located in the plastid envelope (Bouchnak *et al*., 2019). Phylogenetic analysis of *ycf1* and *ycf2* has shown that these two plastid-encoded genes may have evolved contemporaneously during the evolution of vascular plants (de Vries *et al*., 2017). In tobacco chloroplasts, haemagglutinin epitope-tagged YCF2 was used to co-purify the proteins present in an envelope 2-MD protein complex, finding that five inactivated ftsH proteins are a part of the complex, along with plastidial NAD-malate dehydrogenase (pdNAD-MDH), which physically

interacts with the TIC complex (Kikuchi *et al*., 2018). pdNAD-MDH is a moonlighting protein that functions enzymatically to regenerate NADH by the reduction of oxaloacetate to malate in non-photosynthetic plastids, and functions in the biogenesis and maturation of chloroplasts, where mutants had multiple defects in photosynthetic rate and carbohydrate metabolism (Beeler *et al*., 2014; Schreier *et al*., 2018). The non-enzymatic domain of pdNAD-MDH functions in the stabilization of FtsH12, which is a component of the YCF2-TIC complex, and mutants show severe defects in prolamellar body formation (Schreier *et al*., 2018). The prolamellar body formation defects in pdNAD-MDH knock-down mutants, and the hypothesis that a specific TIC complex exists for non-photosynthetic protein import (Nakai, 2018), suggests that YCF2 may have a specific role in protein import for heterotrophic plastid metabolism. More evidence for a non-photosynthetic function for YCF2 (a.k.a. ORF2280) is its upregulation in developing tomato fruit chromoplasts (Richards *et al*., 1991), and high protein levels in fruit and flower chromoplasts (Richards *et al*., 1994).

Non-photosynthetic plastids may, therefore, be specifically transcriptionally regulated to facilitate metabolism in heterotrophic tissues, rather than functioning purely to maintain the plastid as a subcellular compartment for metabolism.

1.7. Post-transcriptional regulation of organellar-encoded genes

In land plant organellar genomes, several post-transcriptional mechanisms have evolved to regulate transcript processing. Unlike eukaryotic cytosolic mRNAs, polyA tailing of organellar transcripts by polyA polymerase-like enzymes and polynucleotide

phosphorylase (PNPase) leads to transcript degradation and affects a small proportion of transcripts (Walter *et al*., 2002; Zimmer *et al*., 2009; Rorbach *et al*., 2014). Similar to bacterial genes, the 5' and 3' ends of organellar transcripts are processed from polycistronic transcripts by endo- and exonucleases, and at the 3' end by PNPase (Stern & Goldschmidt-Clermont, 2010; Hammani & Giegé, 2014). These processes are mediated by short inverted repeats forming hairpins at the 3' ends of the transcript, and by the binding of nuclear-encoded RNA binding proteins that impede nucleases (Dombrowski *et al*., 1997; Stoppel & Meurer, 2012; Zhelyazkova *et al*., 2012a). Nuclear-encoded RNA binding proteins, especially the PPR proteins of the helical-repeat protein superfamily, are a widespread, conserved mechanism of organellar RNA stabilization in land plants (Ruwe *et al*., 2016; Legen *et al*., 2018). Binding signatures of these proteins can be identified by small RNA footprints whose differential accumulation can be related to the expression of the related RNA binding proteins, as demonstrated in *Arabidopsis* mitochondria and plastids (Haïli *et al*., 2013; Ruwe *et al*., 2016).

In seed plant organelles, transcripts are edited via cytidine deamination reactions, resulting in post-transcriptional C to U editing, predominantly in the mitochondria (~300 to 600 sites), and less so in plastids (~20 to 60 sites) (Ichinose & Sugita, 2016). Transcript editing in organelles is believed to be a correction mechanism that compensates for T to U mutations in organellar DNA that prevent thymine dimer formation, restoring the conserved coding sequence (Covello & Gray, 1993; Fujii & Small, 2011; Budar & Mireau, 2018). These RNA edits almost always cause non-synonymous amino acid changes in the resulting protein and can create or abolish

start and stop codons or splice sites, change protein secondary structure, or cause degradation of the transcript (Bock, 2000; Tang & Luo, 2018). Organellar editing sites can be completely or partially edited (Mower & Palmer, 2006), tissue-specific (Bock *et al*., 1993; Picardi *et al*., 2010), and can have a large effect on plant viability and protein function (Sasaki *et al*., 2001; Tang & Luo, 2018; Kong *et al*., 2019). Partial editing and tissue-specific regulation of editing is due to the number of nuclear-encoded editing factors present in the editosome that can finely modulate the editing extent at multiple sites (Andrés-Colás *et al*., 2017; Takenaka *et al*., 2018). Amongst these editing factors are the PPR family of proteins, which have evolved and expanded alongside organellar transcript editing (Barkan & Small, 2014; Schallenberg-Rüdinger & Knoop, 2016). Analysis of tissue-specific editing in *Vitis vinifera* showed that 71% of tissue preferential mitochondrial editing sites are found in leaves (Picardi *et al*., 2010), suggesting that post-transcriptional editing may be coordinated between organelles in photosynthetic tissues.

After 5' and 3' end processing and editing, both plastid and mitochondrial transcripts undergo splicing of type I and type II introns (Stern & Goldschmidt-Clermont, 2010; Hammani & Giegé, 2014). As with end processing, this process is performed by nuclear-encoded RNA binding proteins such as PPR proteins and maturases (Stern & Goldschmidt-Clermont, 2010; Hammani & Giegé, 2014). Splicing of organellar transcripts can be tissue-specific, with the defining example being *ndhB* in potato tubers, which is almost completely unspliced compared to leaves (Valkov *et al*., 2009). Together with the tissue-specific editing of *ndhB* in potato tubers and tomato fruit chromoplasts, and other examples of tissue-specific post-transcriptional regulation of

organellar genes, heterogeneous populations of plastid proteins are found between tissues, leading to functional differences in central metabolism (Kahlau & Bock, 2008; Valkov *et al.*, 2009).

1.8. Cytonuclear coordination in land plants

In land plants, transcriptional regulation must be coordinated between the nuclear and organellar genomes to ensure organelle function, and response to biotic and abiotic stress (Woodson & Chory, 2008; Nomura *et al.*, 2012; Sun & Guo, 2016). Two mechanisms exist to regulate nuclear and organellar gene expression: anterograde signalling from the nucleus to the organelles, and retrograde signalling from the organelles to the nucleus (Pogson *et al.*, 2008; Ng *et al.*, 2014a). Additionally, inter-organellar communication between the plastid and the mitochondria relay the functional status of these organelles to each other, adjusting organellar and nuclear transcription (Liao *et al.*, 2016; Alber & Vanlerberghe, 2018). Anterograde signalling plays an important role in the tissue- and environment-specific regulation of plastids in land plants, and regulates plastid differentiation as seen in the induction of amyloplast development from proplastids by cytokinin in tobacco BY-2 cells (Miyazawa *et al.*, 2002). Reactive oxygen species (ROS) perform a central role in organellar retrograde signalling pathways and interorganellar signalling (Mignolet-Spruyt *et al.*, 2016). Unlike plastid retrograde signalling, where a variety of signalling molecules have been characterized, mitochondrial retrograde signalling has not been definitively associated with any potential signalling molecules other than ROS and Ca2+ (Schwarzländer & Finkemeier, 2013; Kmiecik *et al.*, 2016; de Souza *et al.*,

2017). In mitochondria, a component of the respiratory chain, alternative oxidase (AOX) is induced by stress and serves to limit ROS production, and increases the expression of nuclear-encoded ROS scavenging enzymes (Vanlerberghe, 2013). Mitochondrial retrograde signalling is not regulated in a tissue-specific manner and is closely related to the functional state of plastids and general stress response (Van Aken & Whelan, 2012).

In plastids, retrograde signalling is mediated by signalling molecules derived from multiple plastid localized metabolic pathways that exert operational (response to environment) and biogenic control (developmental regulation) on nuclear-encoded genes (Pogson *et al*., 2008). Tetrapyrrole intermediates heme and Mg-ProtoIX are some of the best characterized retrograde signalling molecules, discovered from forward genetics screens of the genomes uncoupled (gun) mutants which regulate photosynthesis-associated nuclear genes (Susek *et al*., 1993; Rodermel & Park, 2003). The methylerythritol 4-phosphate (MEP) pathway intermediate 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcPP) is induced by wounding and abiotic stress and activates the expression of nuclear stress response genes, including unfolded protein response genes, potentially through interaction with histone-like proteins (Xiao *et al*., 2012; Walley *et al*., 2015). MEcPP has recently been shown to regulate auxin abundance and transport in *Arabidopsis*, implicating retrograde signalling in primary growth and development, and similar mechanisms may also regulate other tissue-specific auxin pathways in plants (Jiang *et al*., 2018). The phosphonucleotide 3′-phosphoadenosine 5′-phosphate (PAP) is induced by high light and drought stress and forms part of the SAL1-PAP-XRN retrograde signalling pathway (Estavillo *et al*.,

2011; Phua *et al*., 2018). Plastidial PAP level is decreased by SAL1 dephosphorylation of PAP to Adenosine monophosphate (AMP), and during the stress response, SAL1 is inhibited by oxidative stress, causing to PAP accumulate and translocate to the nucleus (Estavillo *et al*., 2011; Chan *et al*., 2016). In the nucleus, PAP inhibits 5' to 3' exoribonucleases (XRNs) and positively regulates the expression of stress-response genes (Estavillo *et al*., 2011; Crisp *et al*., 2018).

Organellar retrograde signalling in the plastid and mitochondria are tightly linked via several pathways. The first is the regulation of nuclear gene expression by abscisic acid insensitive 4 (ABI4), which responds to Mg-ProtoIX and GUN1 signalling, and activates AOX expression during the mitochondrial stress response (Giraud *et al*., 2009). Additionally, ABI4 responds to sugar signalling, where high sugar levels cause ABI4 to repress photosynthetic gene expression similar to the repression of photosynthetic genes by Mg-ProtoIX (Rook *et al*., 2006). Meanwhile, genomes uncoupled 1 (GUN1) is a PPR protein that relays information about organellar-encoded gene expression to the nucleus via ABI4, causing repression of photosynthetic genes due to plastid dysfunction (Koussevitzky *et al*., 2007). Another pathway involved in the convergence of organellar signalling is the cyclin-dependent kinase E1 (CDKE1) pathway, which regulates nuclear-encoded photosynthetic gene and *aox* expression in response to developmental changes and environmental stimulus in a pathway independent of ABI4 (Blanco *et al*., 2014). The metabolic pathway of *de novo* fatty acid biosynthesis has also been linked to inter-organellar retrograde signalling, as the enoyl-ACP reductase gene *mosaic death1* (*mod1*) mutant causes programmed cell death mediated by mitochondrial ROS production (Mou *et*

al., 2000; Wu *et al*., 2015). The *mod1* phenotype can be suppressed by mutants in the mitochondrial respiratory redox pathway complex I subunits, which is regulated in part by nuclear PRR protein SOM42 (Wu *et al*., 2015).

Organellar retrograde signalling pathways play a role in the regulation of metabolism by the plant central circadian clock via the MEcPP, PAP, and sugar signalling pathways (Jones, 2018). In Arabidopsis, ~37% of nuclear genes are regulated by the circadian clock (Ferrari *et al*., 2019), including the plastid-targeted carbon metabolic pathways of starch biosynthesis and degradation, and the MEP pathway (Vranová *et al*., 2013; Flis *et al*., 2019). PAP accumulation has recently been shown to extend the circadian period, integrating stress response and the central clock in plants (Litthauer *et al*., 2018). Further, plastid stress response and retrograde signalling can have wide-ranging effects on the circadian clock and the epigenetic landscape of the nuclear genome. In *Arabidopsis*, the protein MutS HOMOLOGUE 1 (MSH1) is targeted to the plastid and mitochondria, and the plastid isoform is depleted during the stress response (Xu *et al*., 2011; Virdi *et al*., 2016). Disruption of *msh1* leads to extensive changes in the transcription and translation of nuclear genes, including core circadian clock genes, and heritable epigenetic changes that enhance plant growth and stress responses (Virdi *et al*., 2015; Beltrán *et al*., 2018). MSH1 is preferentially localized to a specific subtype of plastid, termed the sensory chloroplast, that is found in the vascular tissues of *Arabidopsis* leaves (Beltrán *et al*., 2018). The role of the sensory chloroplast highlights the need to clearly define the functions of specific plastid types, as tissue-specific modulation of these responses may decrease unwanted pleiotropic effects for future biotechnological applications (Beltrán *et al*., 2018).

The coordination of nuclear and organellar transcription via anterograde and retrograde signalling has evolved since the endosymbiotic events leading to the plastid and mitochondria in plants. A consequence of this is that hybrid plants may exhibit cytonuclear incompatibility, where the nuclear and organellar genomes have variants that alter coordination between them (Case *et al.*, 2016; Park *et al.*, 2017). Research in recombinant inbred lines of *Arabidopsis* has shown that organellar genome variants can have a large effect on primary metabolism and adaptive fitness traits (Joseph *et al.*, 2015; Roux *et al.*, 2016). In economically important plant species, organellar genome variation and parent of origin effects need to be fully understood to effectively maximize the yield of hybrid species (Zhang *et al.*, 2012a; Ng *et al.*, 2014b).

1.9. Conclusion

In this review, we show that while we know little of plastids and mitochondria in wood formation, they may play a large role in xylogenesis through the evolution of plant genomes, the subcellular compartmentation of important carbon metabolic pathways, and potentially, the coordination between the nuclear and organellar genomes via signalling pathways. Understanding how the regulation of organellar metabolism is integrated into the process of xylogenesis will address a large gap in our understanding of wood formation. Even though organellar genome sequences for tree species are readily available, we have no idea how organellar-encoded genes are regulated during wood formation. Considering the large impact that organellar genome variation and gene expression can have on whole-plant function, understanding how

this could impact plant biomass production in vascular plants has large ecological and economic implications. In the era of next-generation sequencing technologies, studying organellar genomics and transcriptomics in tree species should pave the way for biotechnological applications to increase yield, and for uncovering unique regulation of organelles, which are amongst the most well-studied systems in biology.

1.10. References

Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences* **99**: 9905–9912.

Alber NA, Vanlerberghe GC. 2018. Signalling interactions between mitochondria and chloroplasts in *Nicotiana tabacum* leaf. *Physiologia Plantarum*. doi:10.1111/ppl.12879.

Allen JF. 2003. The function of genomes in bioenergetic organelles. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **358**: 19–37.

Allen JF, Martin WF. 2016. Why have organelles retained genomes? *Cell Systems* **2**: 70–72.

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. *The Plant Cell* **23**: 2499–2513.

do Amaral MN, Souza GM. 2017. The challenge to translate OMICS data to whole plant physiology: The context matters. *Frontiers in Plant Science* **8**: 2146.

Andrés-Colás N, Zhu Q, Takenaka M, De Rybel B, Weijers D, Van Der Straeten D. 2017. Multiple PPR protein interactions are involved in the RNA editing system in *Arabidopsis* mitochondria and plastids. *Proceedings of the National Academy of Sciences* **114**: 8883–8888.

Bailey IW. 1938. Cell Wall Structure of Higher Plants. *Industrial and Engineering Chemistry* **30**: 40–47.

Barkan A. 1988. Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. *The EMBO Journal* **7**: 2637–2644.

Barkan A. 2011. Expression of plastid genes: Organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiology* **155**: 1520–1532.

Barkan A, Small I. 2014. Pentatricopeptide repeat proteins in plants. *Annual Review of Plant Biology.* **65**: 415-442.

Barrett CF, Freudenstein JV, Li J, Mayfield-Jones DR, Perez L, Pires JC, Santos C. 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* **31**: 3095–3112.

Beeler S, Liu H-C, Stadler M, Schreier T, Eicke S, Lue W-L, Truernit E, Zeeman SC, Chen J, Kötting O. 2014. Plastidial NAD-dependent malate dehydrogenase is critical for embryo development and heterotrophic metabolism in *Arabidopsis*. *Plant Physiology* **164**: 1175–1190.

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

Björkholm P, Harish A, Hagström E, Ernst AM, Andersson SGE. 2015. Mitochondrial genomes are retained by selective constraints on protein targeting. *Proceedings of the National Academy of Sciences* **112**: 10154–10161.

Blanco E, Guinea-Díaz M, Whelan J, Strand Å. 2014. Interaction between plastid and mitochondrial retrograde signalling pathways during changes to plastid redox status. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* **369**: 1-8

Bock R. 2000. Sense from nonsense: How the genetic information of chloroplasts is altered by RNA editing. *Biochimie* **82**: 549–557.

Bock R. 2015. Witnessing genome evolution: Experimental reconstruction of endosymbiotic and horizontal gene transfer. *Annual Review of Genetics.* **51**: 1-22.

Bock R, Hagemann R, Kossel H, Kudla J. 1993. Tissue- and stage-specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids- a new regulatory mechanism? *Molecular Genetics and Genomics* **240**: 238–244.

Boerjan W. 2005. Biotechnology and the domestication of forest trees. *Current Opinion in Biotechnology* **16**: 159–166.

Bollenbach TJ, Schuster G, Portnoy V, Stern DB. 2007. Processing, degradation, and polyadenylation of chloroplast transcripts. In: *Bock R. eds. Cell and Molecular Biology of Plastids. Topics in Current Genetics, vol 19.* Springer, Berlin, Heidelberg: 175–211.

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

Bölter B, Soll J. 2017. Ycf1/Tic214 is not essential for the accumulation of plastid proteins. *Molecular Plant* **10**: 219–221.

Bouchnak I, Brugière S, Moyet L, Le Gall S, Salvi D, Kuntz M, Tardif M, Rolland N. 2019. Unravelling hidden components of the chloroplast envelope proteome: opportunities and limits of better MS sensitivity. *Molecular & Cellular Proteomics.* doi.org/10.1074/mcp.RA118.000988.

Brett CT. 2000. Cellulose microfibrils in plants: biosynthesis, deposition, and integration into the cell wall. *International Review of Cytology* **199**: 161–199.

Budar F, Mireau H. 2018. The cross-talk between genomes: How co-evolution shaped plant mitochondrial gene expression. *Annual Plant Reviews*. doi.org/10.1002/9781119312994.apr0545

Cardi T, Giegé P, Kahlau S, Scotti N. 2012. Expression profiling of organellar genes. In: Bock R, Knoop V, eds. *Genomics of Chloroplasts and Mitochondria*. Dordrecht: Springer Netherlands: 323–355.

Case AL, Finseth FR, Barr CM, Fishman L. 2016. Selfish evolution of cytonuclear hybrid incompatibility in *Mimulus*. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* **283**. doi.org/10.1098/rspb.2016.1493.

Chaffey N, Cholewa E, Regan S, Sundberg B. 2002. Secondary xylem development in *Arabidopsis*: A model for wood formation. *Physiologia Plantarum* **114**: 594–600.

Chan KX, Mabbitt PD, Phua SY, Mueller JW, Nisar N, Gigolashvili T, Stroeher E, Grassl J, Arlt W, Estavillo GM, *et al*. 2016. Sensing and signalling of oxidative stress

in chloroplasts by inactivation of the SAL1 phosphoadenosine phosphatase. *Proceedings of the National Academy of Sciences* **113**: E4567–76.

Chen H, Yu Y, Chen X, Zhang Z, Gong C, Li J, Wang A. 2015. Plastid DNA insertions in plant nuclear genomes: the sites, abundance and ages, and a predicted promoter analysis. *Functional & Integrative Genomics* **15**: 131–139.

Cho WK, Geimer S, Meurer J. 2009. Cluster analysis and comparison of various chloroplast transcriptomes and genes in *Arabidopsis thaliana*. *DNA research* **16**: 31–44.

Cole LW, Guo W, Mower JP, Palmer JD. 2018. High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. *Molecular Biology and Evolution*. **35**: 2773–2785.

Courtois-Moreau CL, Pesquet E, Sjödin A, Muñiz L, Bollhöner B, Kaneda M, Samuels L, Jansson S, Tuominen H. 2009. A unique program for cell death in xylem fibres of *Populus* stem. *The Plant Journal* **58**: 260–274.

Covello PS, Gray MW. 1993. On the evolution of RNA editing. *Trends in Genetics* **9**: 265–268.

Crisp PA, Smith AB, Ganguly DR, Murray KD, Eichten SR, Millar AA, Pogson BJ. 2018. RNA Polymerase II read-through promotes expression of neighboring genes in SAL1-PAP-XRN retrograde signalling. *Plant Physiology* **178**: 1614–1630.

Deschamps P, Colleoni C, Nakamura Y, Suzuki E, Putaux J-L, Buléon A, Haebel S, Ritte G, Steup M, Falcón LI, *et al*. 2008. Metabolic symbiosis and the birth of the plant kingdom. *Molecular Biology and Evolution* **25**: 536–548.

De Smet R, Adams KL, Vandepoele K, Van Montagu MCE, Maere S, Van de Peer Y. 2013. Convergent gene loss following gene and genome duplications creates

single-copy families in flowering plants. *Proceedings of the National Academy of Sciences* **110**: 2898–2903.

Didi V, Jackson P, Hejátko J. 2015. Hormonal regulation of secondary cell wall formation. *Journal of Experimental Botany* **66**: 5015–5027.

Dombrowski S, Brennicke A, Binder S. 1997. 3'-inverted repeats in plant mitochondrial mRNAs are processing signals rather than transcription terminators. *The EMBO Journal* **16**: 5069–5076.

Drescher A, Ruf S, Calsa T, Carrer H, Bock R. 2000. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *The Plant Journal* **22**: 97–104.

Eriksson ME, Israelsson M, Olsson O, Moritz T. 2000. Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. *Nature Biotechnology* **18**: 784–788.

Estavillo GM, Crisp PA, Pornsiriwong W, Wirtz M, Collinge D, Carrie C, Giraud E, Whelan J, David P, Javot H, *et al*. 2011. Evidence for a SAL1-PAP chloroplast retrograde pathway that functions in drought and high light signalling in *Arabidopsis*. *The Plant Cell* **23**: 3992–4012.

Fang Y, Wu H, Zhang T, Yang M, Yin Y, Pan L, Yu X, Zhang X, Hu S, Al-Mssallem IS, *et al*. 2012. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera L.*) mitochondrial genome. *PLoS one* **7**: e37164.

Fernie AR, Carrari F, Sweetlove LJ. 2004. Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Current Opinion in Plant Biology* **7**: 254–261.

Ferrari C, Proost S, Janowski M, Becker J, Nikoloski Z, Bhattacharya D, Price D, Tohge T, Bar-Even A, Fernie A, *et al.* 2019. Kingdom-wide comparison reveals the evolution of diurnal gene expression in Archaeplastida. *Nature Communications* **10**: 737.

Flis A, Mengin V, Ivakov AA, Mugford ST, Hubberten H-M, Encke B, Krohn N, Höhne M, Feil R, Hoefgen R, *et al.* 2019. Multiple circadian clock outputs regulate diel turnover of carbon and nitrogen reserves. *Plant, Cell & Environment* **42**: 549–573.

Fojtová M, Kovarík A, Matyásek R. 2001. Cytosine methylation of plastid genome in higher plants. Fact or artefact? *Plant Science* **160**: 585–593.

Fujii S, Small I. 2011. The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytologist* **191**: 37–47.

Fujii S, Toda T, Kikuchi S, Suzuki R, Yokoyama K, Tsuchida H, Yano K, Toriyama K. 2011. Transcriptome map of plant mitochondria reveals islands of unexpected transcribed regions. *BMC Genomics* **12**: 279.

Fukuda H. 1996. Xylogenesis: Initiation, progression, and cell death. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 299–325.

Gabaldón T, Huynen MA. 2005. Lineage-specific gene loss following mitochondrial endosymbiosis and its potential for function prediction in eukaryotes. *Bioinformatics* **21:** 144–150.

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.

Giegé P, Sweetlove LJ, Cognat V, Leaver CJ. 2005. Coordination of nuclear and mitochondrial genome expression during mitochondrial biogenesis in *Arabidopsis*. *The Plant Cell* **17**: 1497–1512.

Giraud E, Van Aken O, Ho LHM, Whelan J. 2009. The transcription factor ABI4 is a regulator of mitochondrial retrograde expression of ALTERNATIVE OXIDASE1a. *Plant Physiology* **150**: 1286–1296.

Gonzalez DH, Welchen E, Attallah CV, Comelli RN, Mufarrege EF. 2007. Transcriptional coordination of the biogenesis of the oxidative phosphorylation machinery in plants: Coordination of mitochondrial biogenesis**.** *The Plant Journal* **51**: 105–116.

Gould SB, Waller RF, McFadden GI. 2008. Plastid evolution. *Annual Review of Plant Biology* **59**: 491–517.

Gray MW, Archibald JM. 2012. Origins of mitochondria and plastids. In: Bock R, Knoop V, eds. *Genomics of Chloroplasts and Mitochondria*. Dordrecht: Springer Netherlands, 1–30.

Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-Lotfi F, Dietrich A. 2014. The plant mitochondrial genome: dynamics and maintenance. *Biochimie* **100**: 107–120.

Guo W, Grewe F, Fan W, Young GJ, Knoop V. 2016. Ginkgo and Welwitschia mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Molecular Biology and Evolution* **33:** 1448–1460.

Guo W, Zhu A, Fan W, Mower JP. 2017. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. *New Phytologist* **213**: 391–403.

Haïli N, Arnal N, Quadrado M, Amiar S, Tcherkez G, Dahan J, Briozzo P, Colas des Francs-Small C, Vrielynck N, Mireau H. 2013. The pentatricopeptide repeat MTSF1 protein stabilizes the nad4 mRNA in *Arabidopsis* mitochondria. *Nucleic Acids Research* **41**: 6650–6663.

Hajdukiewicz PTJ, Allison LA, Maliga P. 1997. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. *The EMBO Journal* **16**: 4041–4048.

Halliwell B. 1984. Chloroplast metabolism: The structure and function of chloroplasts in green leaf cells. Oxford: Clarendon Press.

Hameed MW, Juszczak I, Bock R, van Dongen JT. 2017. Comparison of mitochondrial gene expression and polysome loading in different tobacco tissues. *Plant Methods* **13**: 112-120.

Hammani K, Giegé P. 2014. RNA metabolism in plant mitochondria. *Trends in Plant Science* **19**: 380–389.

Hammer GL, Sinclair TR, Chapman SC, van Oosterom E. 2004. On systems thinking, systems biology, and the *in silico* plant. *Plant Physiology* **134**: 909–911.

Heinig U, Gutensohn M, Dudareva N, Aharoni A. 2013. The challenges of cellular compartmentalization in plant metabolic engineering. *Current Opinion in Biotechnology* **24**: 239–246.

Henkes S, Sonnewald U, Badur R, Flachmann R, Stitt M. 2001. A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *The Plant Cell* **13**: 535–551.

Hernández-Verdeja T, Strand Å. 2018. Retrograde signals navigate the path to chloroplast development. *Plant Physiology* **176**: 967–976.

Huang CY, Grünheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiology* **138**: 1723–1733.

Husnik F, McCutcheon JP. 2018. Functional horizontal gene transfer from bacteria to eukaryotes. *Nature Reviews Microbiology* **16**: 67–79.

Hu Z, Tong Z, Wei J, Yi H, Deng X. 2006. Mitochondrial gene expression in stamens is differentially regulated during male gametogenesis in *Citrus unshiu*. *The Journal of Horticultural Science & Biotechnology* **81**: 565–569.

Ichinose M, Sugita M. 2016. RNA editing and its molecular mechanism in plant organelles. *Genes* **8**: 231-239.

Ito H, Kakutani T. 2014. Control of transposable elements in *Arabidopsis thaliana*. *Chromosome Research* **22**: 217–223.

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

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, *et al.* 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences* **104**: 19369–19374.

Jiang J, Chai X, Manavski N, Williams-Carrier R, He B, Brachmann A, Ji D, Ouyang M, Liu Y, Barkan A, *et al.* 2019. An RNA-chaperone-like protein plays critical

roles in chloroplast mRNA stability and translation in *Arabidopsis* and maize. *The Plant Cell* **31**: 1308–1327.

Jiang J, Rodriguez-Furlan C, Wang J-Z, de Souza A, Ke H, Pasternak T, Lasok H, Ditengou FA, Palme K, Dehesh K. 2018. Interplay of the two ancient metabolites auxin and MEcPP regulates adaptive growth. *Nature Communications* **9**: 2262-2268.

Jones MA. 2018. Retrograde signalling as an informant of circadian timing. *New Phytologist* **221**: 1749-1753.

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

Kahlau S, Bock R. 2008. Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: chromoplast gene expression largely serves the production of a single protein. *The Plant Cell* **20**: 856–874.

Kalantidis K, Wilson ZA, Mulligan BJ. **2002**. Mitochondrial gene expression in stamens is differentially regulated during male gametogenesis in *Arabidopsis*. *Sexual Plant Reproduction* **14**: 299–304.

Kaneda M, Rensing K, Samuels L. **2010**. Secondary cell wall deposition in developing secondary xylem of poplar. *Journal of Integrative Plant Biology* **52**: 234–243.

Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bédard J, Hirabayashi-Ishioka Y, *et al.* 2018. A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *The Plant Cell* **30**: 2677–2703.

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. **2013**. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **339**: 571–574.

Kmiecik P, Leonardelli M, Teige M. **2016**. Novel connections in plant organellar signalling link different stress responses and signalling pathways. *Journal of Experimental Botany* **136**. 321-329.

Kobayashi K, Baba S, Obayashi T, Sato M, Toyooka K, Keränen M, Aro E-M, Fukaki H, Ohta H, Sugimoto K, *et al.* 2012. Regulation of root greening by light and auxin/cytokinin signalling in *Arabidopsis*. *The Plant Cell* **24**: 1081–1095.

Koch K. **2004**. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Current Opinion in Plant Biology* **7**: 235–246.

Kong X, Liu D, Zheng J, Khan A, Li B, Diao Y, Zhou R. **2019**. RNA editing analysis of ATP synthase genes in the cotton cytoplasmic male sterile line H276A. *Biological Research* **52**: 6.

Koussevitzky S, Nott A, Mockler TC, Hong F, Sachetto-Martins G, Surpin M, Lim J, Mittler R, Chory J. **2007**. Signals from chloroplasts converge to regulate nuclear gene expression. *Science* **316**: 715–719.

Kruger NJ, von Schaewen A. **2003**. The oxidative pentose phosphate pathway: structure and organisation. *Current Opinion in Plant Biology* **6**: 236–246.

Kuhn K, Weihe A, Borner T. **2005**. Multiple promoters are a common feature of mitochondrial genes in *Arabidopsis*. *Nucleic Acids Research* **33**: 337–346.

Lama S, Broda M, Abbas Z, Vaneechoutte D, Belt K, Säll T, Vandepoele K, Van Aken O. **2019**. Neofunctionalization of mitochondrial proteins and incorporation into signalling networks in plants. *Molecular biology and evolution* **36**: 974–989.

Leegood RC, Walker RP. **2003**. Regulation and roles of phosphoenolpyruvate carboxykinase in plants. *Archives of Biochemistry and Biophysics* **414**: 204–210.

Legen J, Kemp S, Krause K, Profanter B, Herrmann RG, Maier RM. **2002**. Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *The Plant Journal* **31**: 171–188.

Legen J, Ruf S, Kroop X, Wang G, Barkan A, Bock R, Schmitz-Linneweber C. **2018**. Stabilization and translation of synthetic operon-derived mRNAs in chloroplasts by sequences representing PPR protein-binding sites. *The Plant Journal* **94**: 8–21.

Lerbs-Mache S. **2011**. Function of plastid sigma factors in higher plants: Regulation of gene expression or just preservation of constitutive transcription? *Plant Molecular Biology* **76**: 235–249.

Liao J-C, Hsieh W-Y, Tseng C-C, Hsieh M-H. **2016**. Dysfunctional chloroplasts up-regulate the expression of mitochondrial genes in *Arabidopsis* seedlings. *Photosynthesis Research* **127**: 151–159.

Liebers M, Grübler B, Chevalier F, Lerbs-Mache S, Merendino L, Blanvillain R, Pfannschmidt T. **2017**. Regulatory shifts in plastid transcription play a key role in morphological conversions of plastids during plant development. *Frontiers in Plant Science* **8**: 23-32.

Liere K, Börner T. **2007**. Transcription and transcriptional regulation in plastids. In: Bock R, ed. *Cell and Molecular Biology of Plastids* Springer Berlin Heidelberg, 121–174.

Liere K, Börner T. **2011**. Transcription in plant mitochondria. In: Kempken F, ed. *Plant Mitochondria*. New York, NY: Springer New York, 85–105.

Li Z, Omranian N, Neumetzler L, Wang T, Herter T, Usadel B, Demura T, Giavalisco P, Nikoloski Z, Persson S. **2016**. A transcriptional and metabolic framework for secondary wall formation in *Arabidopsis*. *Plant Physiology* **172**: 1334–1351.

Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. **2008**. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**: 523–536.

Litthauer S, Chan KX, Jones MA. **2018**. 3'-Phosphoadenosine 5'-Phosphate accumulation delays the circadian system. *Plant Physiology* **176**: 3120–3135.

Liu Y, Wang B, Cui P, Li L, Xue J-Y, Yu J, Qiu Y-L. **2012**. The mitochondrial genome of the lycophyte *Huperzia squarrosa*: the most archaic form in vascular plants. *PLoS one* **7**: e35168.

Maeda HA. **2016**. Lignin biosynthesis: Tyrosine shortcut in grasses. *Nature Plants* **2**: 16080.

Maréchal E, Block MA, Dorne A-J, Douce R, Joyard J. **1997**. Lipid synthesis and metabolism in the plastid envelope. *Physiologia Plantarum* **100**: 65–77.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. **2002**. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proceedings of the National Academy of Sciences* **99**: 12246–12251.

Méheust R, Zelzion E, Bhattacharya D, Lopez P, Bapteste E. **2016**. Protein networks identify novel symbiogenetic genes resulting from plastid endosymbiosis. *Proceedings of the National Academy of Sciences* **113**: 3579–3584.

Mellerowicz EJ, Sundberg B. **2008**. Wood cell walls: biosynthesis, developmental dynamics and their implications for wood properties. *Current Opinion in Plant Biology* **11**: 293–300.

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

Michalovova M, Vyskot B, Kejnovsky E. **2013**. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. *Heredity* **111**: 314–320.

Mignolet-Spruyt L, Xu E, Idänheimo N, Hoeberichts FA, Mühlenbock P, Brosché M, Van Breusegem F, Kangasjärvi J. **2016**. Spreading the news: subcellular and organellar reactive oxygen species production and signalling. *Journal of Experimental Botany* **67**: 3831–3844.

Miyazawa Y, Kato H, Muranaka T, Yoshida S. **2002**. Amyloplast formation in cultured tobacco BY-2 cells requires a high cytokinin content. *Plant & Cell Physiology* **43**: 1534–1541.

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.

Mou Z, He Y, Dai Y, Liu X, Li J. **2000**. Deficiency in fatty acid synthase leads to premature cell death and dramatic alterations in plant morphology. *The Plant Cell* **12**: 405–418.

Mower JP, Palmer JD. **2006**. Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. *Molecular Genetics and Genomics* **276**: 285–293.

Mullet JE. **1993**. Dynamic regulation of chloroplast transcription. *Plant Physiology* **103**: 309–313.

Nakaba S, Takata N, Yoshida M, Funada R. **2015**. Continuous expression of genes for xylem cysteine peptidases in long-lived ray parenchyma cells in *Populus*. *Plant Biotechnology* **32**: 21–29.

Nakai M. **2015**. YCF1: A green TIC: Response to the de Vries *et al*. Commentary. *The Plant Cell* **27**: 1834–1838.

Nakai M. **2018**. New perspectives on chloroplast protein import. *Plant & Cell Physiology* **59**: 1111–1119.

Ng S, De Clercq I, Van Aken O, Law SR, Ivanova A, Willems P, Giraud E, Van Breusegem F, Whelan J. **2014a**. Anterograde and retrograde regulation of nuclear genes encoding mitochondrial proteins during growth, development, and stress. *Molecular Plant* **7**: 1075–1093.

Ng DW-K, Miller M, Yu HH, Huang T-Y, Kim E-D, Lu J, Xie Q, McClung CR, Chen ZJ. **2014b**. A role for CHH methylation in the parent-of-origin effect on altered circadian rhythms and biomass heterosis in *Arabidopsis* intraspecific hybrids. *The Plant Cell* **26**: 2430–2440.

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, *et al.* 2016. Widespread natural variation of DNA methylation within angiosperms. *Genome Biology* **17**: 194-206.

Nomura H, Komori T, Uemura S, Kanda Y, Shimotani K, Nakai K, Furuichi T, Takebayashi K, Sugimoto T, Sano S, *et al.* 2012. Chloroplast-mediated activation of plant immune signalling in *Arabidopsis*. *Nature communications* **3**: 926.

Noutsos C, Richly E, Leister D. **2005**. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Research* **15**: 616–628.

Novaes E, Kirst M, Chiang V, Winter-Sederoff H, Sederoff R. **2010**. Lignin and biomass: a negative correlation for wood formation and lignin content in trees. *Plant Physiology* **154**: 555–561.

Okada S, Brennicke A. **2006**. Transcript levels in plant mitochondria show a tight homeostasis during day and night. *Molecular Genetics and Genomics* **276**: 71–78.

Park S, Ruhlman TA, Weng M-L, Hajrah NH, Sabir JSM, Jansen RK. **2017**. Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. *Genome Biology and Evolution* **9**: 1766–1780.

Parsley K, Hibberd JM. **2006**. The *Arabidopsis* PPDK gene is transcribed from two promoters to produce differentially expressed transcripts responsible for cytosolic and plastidic proteins. *Plant Molecular Biology* **62**: 339–349.

Phua SY, Yan D, Chan KX, Estavillo GM, Nambara E, Pogson BJ. **2018**. The *Arabidopsis* SAL1-PAP Pathway: A case study for integrating chloroplast retrograde, light and hormonal signalling in modulating plant growth and development? *Frontiers in Plant Science* **9**: 1171-1178.

Picardi E, Horner DS, Chiara M, Schiavon R. **2010**. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Research* **35:** 4755–4767.

Planchard N, Bertin P, Quadrado M, Dargel-Graffin C, Hatin I, Namy O, Mireau H. **2018**. The translational landscape of *Arabidopsis* mitochondria. *Nucleic Acids Research* **46**: 6218–6228.

Plomion C, Stokes A, Leprovost G. **2001**. Wood formation in trees. *Plant Physiology* **127**: 1513–1523.

Pogson BJ, Woo NS, Förster B, Small ID. **2008**. Plastid signalling to the nucleus and beyond. *Trends in Plant Science* **13**: 602–609.

Prabhakar V, Löttgert T, Gigolashvili T, Bell K, Flügge U-I, Häusler RE. **2009**. Molecular and functional characterization of the plastid-localized phosphoenolpyruvate enolase (ENO1) from *Arabidopsis thaliana*. *FEBS Letters* **583**: 983–991.

Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Börner T. **2010**. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *The Plant Journal* **64**: 948–959.

Proost S, Pattyn P, Gerats T, Van de Peer Y. **2011**. Journey through the past: 150 million years of plant genome evolution. *The Plant Journal* **66**: 58–65.

Rand DM, Haney RA, Fry AJ. **2004**. Cytonuclear coevolution: The genomics of cooperation. *Trends in Ecology & Evolution* **19**: 645–653.

Raven JA., Allen JF. **2003**. Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biology* **4**: 209-215.

Richards CM, Hardison RC, Boyer CD. **1994**. Expression of the large plastid gene, ORF2280, in tomato fruits and flowers. *Current Genetics* **26**: 494–496.

Richards CM, Hinman SB, Boyer CD, Hardison RC. **1991**. Survey of plastid RNA abundance during tomato fruit ripening: the amounts of RNA from the ORF 2280 region increase in chromoplasts. *Plant Molecular Biology* **17**: 1179–1188.

Rippert P, Puyaubert J, Grisollet D, Derrier L, Matringe M. **2009**. Tyrosine and phenylalanine are synthesized within the plastids in *Arabidopsis*. *Plant Physiology* **149**: 1251–1260.

Robinson SW, Fernandes M, Husi H. **2014**. Current advances in systems and integrative biology. *CSBJ* **11**: 35–46.

Rodermel S, Park S. **2003**. Pathways of intracellular communication: Tetrapyrroles and plastid-to-nucleus signalling. *BioEssays* **25**: 631–636.

Roger AJ, Muñoz-Gómez SA, Kamikawa R. **2017**. The origin and diversification of mitochondria. *Current Biology* **27**: R1177–R1192.

Rolland N, Bouchnak I, Moyet L, Salvi D, Kuntz M. **2018**. The main functions of plastids. *Methods in Molecular Biology* **1829**: 73–85.

Rolland N, Curien G, Finazzi G, Kuntz M, Maréchal E, Matringe M, Ravanel S, Seigneurin-Berny D. **2012**. The biosynthetic capacities of the plastids and integration between cytoplasmic and chloroplast processes. *Annual Review of Genetics* **46**: 233–264.

Rook F, Hadingham SA, Li Y, Bevan MW. **2006**. Sugar and ABA response pathways and the control of gene expression. *Plant, Cell & Environment* **29**: 426–434.

Rorbach J, Bobrowicz A, Pearce S, Minczuk M. **2014**. Polyadenylation in bacteria and organelles. *Methods in Molecular Biology* **1125**: 211–227.

Rousseau-Gueutin M, Huang X, Higginson E, Ayliffe M, Day A, Timmis JN. **2013**. Potential functional replacement of the plastidic acetyl-CoA carboxylase subunit (accD) gene by recent transfers to the nucleus in some angiosperm lineages. *Plant Physiology* **161**: 1918–1929.

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.

Ruwe H, Wang G, Gusewski S, Schmitz-linneweber C. **2016**. Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Research*: 1–12.

Sasaki Y, Kozaki A, Ohmori A, Iguchi H, Nagano Y. **2001**. Chloroplast RNA editing required for functional acetyl-CoA carboxylase in plants. *The Journal of Biological Chemistry* **276**: 3937–3940.

Schallenberg-Rüdinger M, Knoop V. **2016**. Coevolution of organelle RNA editing and nuclear specificity factors in early land plants. *Advances in Botanical Research* **78**: 1–57.

Schmid KM. **2015**. Lipid metabolism in plants. In. Ridgway ND, McLeod RS, eds. *Biochemistry of Lipids, Lipoproteins and Membranes*. Elsevier. 113-147.

Schreier TB, Cléry A, Schläfli M, Galbier F, Stadler M, Demarsy E, Albertini D, Maier BA, Kessler F, Hörtensteiner S, *et al.* 2018. Plastidial NAD-dependent malate

dehydrogenase: A moonlighting protein involved in early chloroplast development through its interaction with an FtsH12-FtsHi protease complex. *The Plant Cell* **30**: 1745–1769.

Schwarzländer M, Finkemeier I. **2013**. Mitochondrial energy and redox signalling in plants. *Antioxidants & Redox Signalling* **18**: 2122–2144.

Schweer J. **2010**. Plant sigma factors come of age: Flexible transcription factor network for regulated plastid gene expression. *Endocytobiosis and Cell Research* **20:** 1-12.

Shaver JM, Oldenburg DJ, Bendich AJ. **2008**. The structure of chloroplast DNA molecules and the effects of light on the amount of chloroplast DNA during development in *Medicago truncatula*. *Plant Physiology* **146**: 1064–1074.

Sibout R, Proost S, Hansen BO, Vaid N, Giorgi FM, Ho-Yue-Kuang S, Legée F, Cézart L, Bouchabké-Coussa O, Soulhat C, *et al.* 2017. Expression atlas and comparative coexpression network analyses reveal important genes involved in the formation of lignified cell wall in *Brachypodium distachyon*. *New Phytologist* **215**: 1009–1025.

Sloan DB. **2013**. One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytologist* **200**: 978–985.

Sloan DB. **2015**. Using plants to elucidate the mechanisms of cytonuclear co-evolution. *New Phytologist* **205**: 1040–1046.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. **2012**. Rapid evolution of enormous, multichromosomal genomes in

flowering plant mitochondria with exceptionally high mutation rates. *PLoS biology* **10**: 1–17.

Smith DR, Keeling PJ. **2015**. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proceedings of the National Academy of Sciences* **112**: 10177–10184.

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.

Song H, Buhay JE, Whiting MF, Crandall KA. **2008**. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences* **105**: 13486–13491.

de Souza A, Wang J-Z, Dehesh K. **2017**. Retrograde signals: Integrators of interorganellar communication and orchestrators of plant development. *Annual Review of Plant Biology* **68**: 85–108.

Stern DB, Goldschmidt-Clermont M. **2010**. Chloroplast RNA metabolism. *Annual Review of Plant Biology* **61:** 125-155.

Stoppel R, Meurer J. **2012**. The cutting crew- ribonucleases are key players in the control of plastid gene expression. *Journal of Experimental Botany* **63**: 1663–1673.

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.

Sun A-Z, Guo F-Q. **2016**. Chloroplast retrograde regulation of heat stress responses in plants. *Frontiers in Plant Science* **7**: 398-405.

Susek RE, Ausubel FM, Chory J. **1993**. Signal transduction mutants of *Arabidopsis* uncouple nuclear CAB and RBCS gene expression from chloroplast development. *Cell* **74**: 787–799.

Sweetlove LJ, Fernie AR. **2013**. The spatial organization of metabolism within the plant cell. *Annual Review of Plant Biology* **64**: 723–746.

Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K, Börner T. **2008**. Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplast genes is compensated by a second phage-type RNA polymerase. *Nucleic Acids Research* **36**: 785–792.

Swiatecka-Hagenbruch M, Liere K, Börner T. **2007**. High diversity of plastidial promoters in *Arabidopsis thaliana*. *Molecular Genetics and Genomics* **277**: 725–734.

Takenaka M, Jörg A, Burger M, Haag S. **2018**. Requirement of various protein combinations for each C-to-U RNA editosome in plant organelles. In: Cruz-Reyes J, Gray MW, eds. *RNA Metabolism in Mitochondria*. Cham: Springer International Publishing, 223–249.

Tang W, Luo C. **2018**. Molecular and functional diversity of RNA editing in plant mitochondria. *Molecular Biotechnology* **60**: 935–945.

Timmis JN, Ayliffe MA, Huang CY, Martin W. **2004**. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Reviews Genetics* **5**: 123–135.

Tohge T, Watanabe M, Hoefgen R, Fernie AR. 2013. Shikimate and phenylalanine biosynthesis in the green lineage. *Frontiers in Plant Science* **4:** 62–73.

Tollefson J. **2017**. The wooden skyscrapers that could help to cool the planet. *Nature* **545**: 280–282.

Tseng C-C, Lee C-J, Chung Y-T, Sung T-Y, Hsieh M-H. **2013**. Differential regulation of *Arabidopsis* plastid gene expression and RNA editing in non-photosynthetic tissues. *Plant Molecular Biology* **82**: 375–392.

Turmel M, Otis C, Lemieux C. **1999**. The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: Insights into the architecture of ancestral chloroplast genomes. *Proceedings of the National Academy of Sciences* **96**: 10248–10253.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

Tuskan GA, Groover AT, Schmutz J, DiFazio SP, Myburg AA, Grattapaglia D, Smart LB, Yin T, Aury J-M, Kremer A, *et al.* 2018. Hardwood tree genomics: Unlocking woody plant biology. *Frontiers in Plant Science* **9**: 1799-1804.

Uggla C, Moritz T, Sandberg G, Sundberg B. 1996. Auxin as a positional signal in pattern formation in plants. Proceedings of the National Academy of Sciences **93**: 9282–9286.

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.

Valkov VT, Scotti N, Kahlau S, Maclean D, Grillo S, Gray JC, Bock R, Cardi T. **2009**. Genome-wide analysis of plastid gene expression in potato leaf chloroplasts

and tuber amyloplasts: transcriptional and posttranscriptional control. *Plant Physiology* **150**: 2030–2044.

Van Aken O. **2018**. Mitochondria and cell death. In: *Roberts JA. eds. Plant Mitochondria. Annual Plant Reviews, vol 50.* Wiley, Hoboken, New Jersey: 343-372.

Van Aken O, Whelan J. **2012**. Comparison of transcriptional changes to chloroplast and mitochondrial perturbations reveals common and specific responses in *Arabidopsis*. *Frontiers in Plant Science* **3**: 1–18.

Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W. **2012**. A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *The Plant Cell* **24**: 3506–3529.

Vanlerberghe GC. **2013**. Alternative oxidase: a mitochondrial respiratory pathway to maintain metabolic and signalling homeostasis during abiotic and biotic stress in plants. *International Journal of Molecular Sciences* **14**: 6805–6847.

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**: 100-113.

Virdi KS, Laurie JD, Xu Y-Z, Yu J, Shao M-R, Sanchez R, Kundariya H, Wang D, Riethoven J-JM, Wamboldt Y, *et al.* 2015. *Arabidopsis* MSH1 mutation alters the epigenome and produces heritable changes in plant growth. *Nature Communications* **6**: 6386-6392.

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.

Vranová E, Coman D, Gruissem W. **2013**. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annual Review of Plant Biology* **64**: 665–700.

de Vries J, Archibald JM, Gould SB. **2017**. The carboxy terminus of YCF1 contains a motif conserved throughout> 500 Myr of Streptophyte evolution. *Genome Biology and Evolution* **9:** 473–479.

de Vries J, Sousa FL, Bölter B, Soll J, Gould SB. **2015**. YCF1: A green TIC? *The Plant Cell* **27**: 1827–1833.

de Vries J, Stanton A, Archibald JM, Gould SB. **2016**. Streptophyte terrestrialization in light of plastid evolution. *Trends in Plant Science* **21**: 467–476.

Walley J, Xiao Y, Wang J-Z, Baidoo EE, Keasling JD, Shen Z, Briggs SP, Dehesh K. **2015**. Plastid-produced interorgannellar stress signal MEcPP potentiates induction of the unfolded protein response in endoplasmic reticulum. *Proceedings of the National Academy of Sciences* **112**: 6212–6217.

Walter M, Kilian J, Kudla J. **2002**. PNPase activity determines the efficiency of mRNA 3'-end processing, the degradation of tRNA and the extent of polyadenylation in chloroplasts. *The EMBO Journal* **21**: 6905–6914.

Wang D, Gu J, David R, Wang Z, Yang S, Searle IR, Zhu J-K, Timmis JN. **2018a**. Experimental reconstruction of double-stranded break repair-mediated plastid DNA insertion into the tobacco nucleus. *The Plant Journal* **93**: 227–234.

Wang D, Lloyd AH, Timmis JN. **2012**. Environmental stress increases the entry of cytoplasmic organellar DNA into the nucleus in plants. *Proceedings of the National Academy of Sciences* **109**: 2444–2448.

Wang D, Qu Z, Adelson DL, Zhu J-K, Timmis JN. **2014**. Transcription of nuclear organellar DNA in a model plant system. *Genome Biology and Evolution* **6**: 1327–1334.

Wang D, Timmis JN. **2013**. Cytoplasmic organelle DNA preferentially inserts into open chromatin. *Genome Biology and Evolution* **5**: 1060–1064.

Wang W-Q, Wang Y, Zhang Q, Møller IM, Song S-Q. **2018b**. Changes in the mitochondrial proteome of developing maize seed embryos. *Physiologia Plantarum* **163**: 552–572.

Weihe A, Liere K, Börner T. **2012**. Transcription and transcription regulation in chloroplasts and mitochondria of higher plants. In: Bullerwell CE, ed. *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression*. Berlin, Heidelberg: Springer Berlin Heidelberg, 297–325.

Wicke S, Schneeweiss GM. **2015**. Next-generation organellar genomics: Potentials and pitfalls of high-throughput technologies for molecular evolutionary studies and plant systematics. In: Hörandl E, Appelhans M, eds. *Next Generation Sequencing in Plant Systematics*. International Association for Plant Taxonomy. DOI: 10.14630/000002.

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. **2011**. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76**: 273–297.

Wolfe KH, Morden CW, Palmer JD. **1992**. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences* **89**: 10648–10652.

Woloszynska M. **2010**. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. *Journal of Experimental Botany* **61**: 657–671.

Woodson JD, Chory J. **2008**. Coordination of gene expression between organellar and nuclear genomes. *Nature Reviews Genetics* **9**: 383–395.

Wu J, Sun Y, Zhao Y, Zhang J, Luo L, Li M, Wang J, Yu H, Liu G, Yang L, *et al.* 2015. Deficient plastidic fatty acid synthesis triggers cell death by modulating mitochondrial reactive oxygen species. *Cell Research* **25**: 621–633.

Xiao Y, Savchenko T, Baidoo EEK, Chehab WE, Hayden DM, Tolstikov V, Corwin JA, Kliebenstein DJ, Keasling JD, Dehesh K. **2012**. Retrograde signalling by the plastidial metabolite MEcPP regulates expression of nuclear stress-response genes. *Cell* **149**: 1525–1535.

Xu Y-Z, Arrieta-Montiel MP, Virdi KS, de Paula WBM, Widhalm JR, Basset GJ, Davila JI, Elthon TE, Elowsky CG, Sato SJ, *et al.* 2011. MutS HOMOLOG1 is a nucleoid protein that alters mitochondrial and plastid properties and plant response to high light. *The Plant Cell* **23**: 3428–3441.

Yang X-F, Wang Y-T, Chen S-T, Li J-K, Shen H-T, Guo F-Q. **2016**. PBR1 selectively controls biogenesis of photosynthetic complexes by modulating translation of the large chloroplast gene *ycf1* in *Arabidopsis*. *Cell discovery* **2**: 16003.

Yao N, Eisfelder BJ, Marvin J, Greenberg JT. **2004**. The mitochondrion- an organelle commonly involved in programmed cell death in *Arabidopsis thaliana*. *The Plant Journal* **40**: 596–610.

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.

Zhang B, Carrie C, Ivanova A, Narsai R, Murcha MW, Duncan O, Wang Y, Law SR, Albrecht V, Pogson B, *et al.* 2012a. LETM proteins play a role in the accumulation of mitochondrially encoded proteins in *Arabidopsis thaliana* and *AtLETM2* displays parent of origin effects. *The Journal of Biological Chemistry* **287**: 41757–41773.

Zhang J, Ruf S, Hasse C, Childs L, Scharff LB. **2012b**. Identification of cis-elements conferring high levels of gene expression in non-green plastids. *The Plant Journal* **72**: 115–128.

Zhelyazkova P, Hammani K, Rojas M, Voelker R, Vargas-Suárez M, Börner T, Barkan A. **2012a**. Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. *Nucleic Acids Research* **40**: 3092–3105.

Zhelyazkova P, Sharma CM, Forstner KU, Liere K, Vogel J, Borner T. **2012b**. The primary transcriptome of barley chloroplasts: Numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *The Plant Cell* **24**: 123–136.

Zhu A, Guo W, Gupta S, Fan W, Mower JP. **2016**. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytologist* **209**: 1747–1756.

Zimmer SL, Schein A, Zipor G, Stern DB, Schuster G. **2009**. Polyadenylation in *Arabidopsis* and *Chlamydomonas* organelles: the input of nucleotidyltransferases, poly(A) polymerases and polynucleotide phosphorylase. *The Plant Journal* **59**: 88–99.

Chapter 2

Organellar carbon metabolism is coordinated 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

This chapter has been prepared in the format of an original research publication for a peer-reviewed research journal. 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, generated the figures, 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. This manuscript has been published in New Phytologist under the same title.

2.1. 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 co-expression 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.

2.2. 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 (P-E-P), and erythrose-4-phosphate (E-4-P) 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 (PCD) (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, and of general interest to plant biologists.

2.3. Materials and Methods

2.3.1. 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 Kyoto Encyclopedia of Genes and Genomes (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 **S2.1**.). The SUBcellular localization database for Arabidopsis proteins (SUBA) (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 **S2.1**).

2.3.2. 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°C. 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 $f(Z_i, Z_j) = (Z_i^2 + Z_j^2)^{1/2}$, 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. **S2.1**. 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.

2.3.3. 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).

2.3.4. Functional enrichment analysis

Functional annotation of consensus clusters was done using GO-BP and KEGG pathway enrichment (Kanehisa *et al.*, 2008). 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 network clusters 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 **S2.2.** Shared plastid and mitochondrial genes were used for an additional enrichment per overlapping gene set (Fig. **2.2**; Notes **S2.2**). 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).

2.4. Results

2.4.1. 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 **2.1**). 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 **2.1**). 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 **2.1**). Thus, almost a third of all KEGG annotated enzymes relating to primary carbon metabolism are predicted to occur in the plastid and mitochondria during xylogenesis.

2.4.2. 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. **S2.1**). 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. **2.1a, 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. **2.1c, 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. **2.1b, d**; Notes **S1**). Between these two networks, there is a 36% overlap of shared genes (Fig. **2.2a, b**), and clusters with large overlaps were assigned the same colour for comparison. Most 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. **2.2a**). 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. **2.2c**). 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 **S2.2**). All genes and cluster membership can be found in Notes **S2.3.**

2.4.3. 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 fibre 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. **2.2c;** Notes **S2.2**). 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. **2.2b**) (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 **S2.3**; Table **S2.1**). 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 **S2.1**). 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 **S2.3**). 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* in *Arabidopsis* (Staehr *et al*., 2014; Streatfield *et al*., 1999; Irigoyen *et al*., 2011) (Notes **S2.3**). 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. **2.1**; Fig. **2.2**), with all three *E. grandis* primary cell wall (PCW) *CESA* genes present (Notes **S2.2**; Notes **S2.3)** (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 **S2.3**). 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 fibre 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 remodelling 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 signalling, 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. **S2.2**) 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, remodelling, and non-coding RNA regulation during xylogenesis than mitochondria, makes the 54 genes of unknown function in PC1 particularly attractive for further study (Notes **S2.3**).

2.4.4. 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 **S2.2**). 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. **2.1a, b**) contain genes of evening/afternoon and morning circadian clock module respectively (Fig. **2.3a;** Table **S2.2**). 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. **2.1b**, Table **S2.2**). PC6 contains homologs of the evening clock module in *Arabidopsis*, including the central clock loop gene *TIMING OF CAB EXPRESSION 1* (*TOC1*), and evening loop genes *EARLY FLOWERING* (*ELF*) *ELF-4a*, *ELF-3* and *LUX*, and little overlap with any mitochondrial clusters (Fig. **2.2a;** Table **S2.2**) (Müller *et al*., 2014). *TETRASPANIN* (*TET*) *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. **2.2a, c**; Notes **S2.2**). The shared network shows that the driver of this overlap is glycolysis, translation, and nucleotide synthesis (Notes **S2.3**), while the circadian genes are specific to PC7. The central morning clock gene *CIRCADIAN CLOCK-ASSOCIATED 1 (CCA1)/LATE ELONGATED HYPOCOTYL (LHY)*, and several morning clock associated genes from the *NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED*

GENES (*LNK)* and *REVEILLE (RVE)* gene families are found in PC7 (Table **S2.2**). 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. **2.3b**). FA biosynthesis and the shikimate pathway share a metabolic intermediate in phosphoenolpyruvate (P-E-P), 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 P-E-P during xylogenesis for primary growth and lignin formation (Fig. **2.3b**).

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. **2.4**). 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 P-E-P carboxykinase 1 (*Eucgr.I00628*) (Fig. **2.4**; Notes **S2.3**). Therefore, despite PC7 genes not being as highly or specifically expressed in xylem as PC3 genes (Fig. **2.2b**; Fig. **S2.3a**), the temporal regulation of carbon allocation plays a significant role in *Eucalyptus* wood formation.

2.4.5. 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. **S2.4a**). 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. **S2.4a** 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 **S2.3**). 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 **S2.3**). 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. **S2.3**).

Overlapping genes between aspen developmental stage clusters and *E. grandis* plastid networks show that the greatest overlap is between PC1, PC3, and PC7 with the corresponding developmental clusters in aspen (e1, e2, f, g1, g2) (Fig. **2.5**; Fig. **S2.4**). 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. **2.5)**. 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 **S2.2**). 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 **S2.2**). 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. **S2.4**). 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. **2.5**; Notes **S2.2**). These terms are related to purine and pyrimidine metabolism, translation, cell wall organization and cell wall modification (Notes **S2.2**). 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 **S2.2**).

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.

2.5. 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. **2.1**, Fig. **2.5**). 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. **2.5**) (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. **2.3**). Genes present in both plastid and mitochondrial network clusters (PC and MC) highlight the regulation of plastid and mitochondrial carbon utilization during secondary cell wall biosynthesis and their impact on lignocellulosic biomass traits (Fig. **2.4**). 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 remodelling, 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 signalling, 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 P-E-P can be used for a variety of metabolic pathways, and P-E-P 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 P-E-P utilization is balanced with the carbon requirements of primary growth and secondary cell wall phenylpropanoids during secondary growth (Fig. **2.3**) (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 signalling 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 signalling 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. **2.3**, Notes **S2.2**). 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 *Arabidopsis* 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 **S2.2**). 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. **2.5**).

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. **2.4**). 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 suggest plastid associated epigenetic modification during early xylogenesis, which is potentially regulated by retrograde signalling. 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.

2.6. 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.

2.7. 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 signalling 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 signalling 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. **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 signalling 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-remodelling 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 defence. *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-remodelling 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*. *Development* **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. *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

2.8. Tables

Table 2.1: Predicted subcellular localization of *E. grandis* genes

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 (Notes **S2.1**).

2.9. Figures

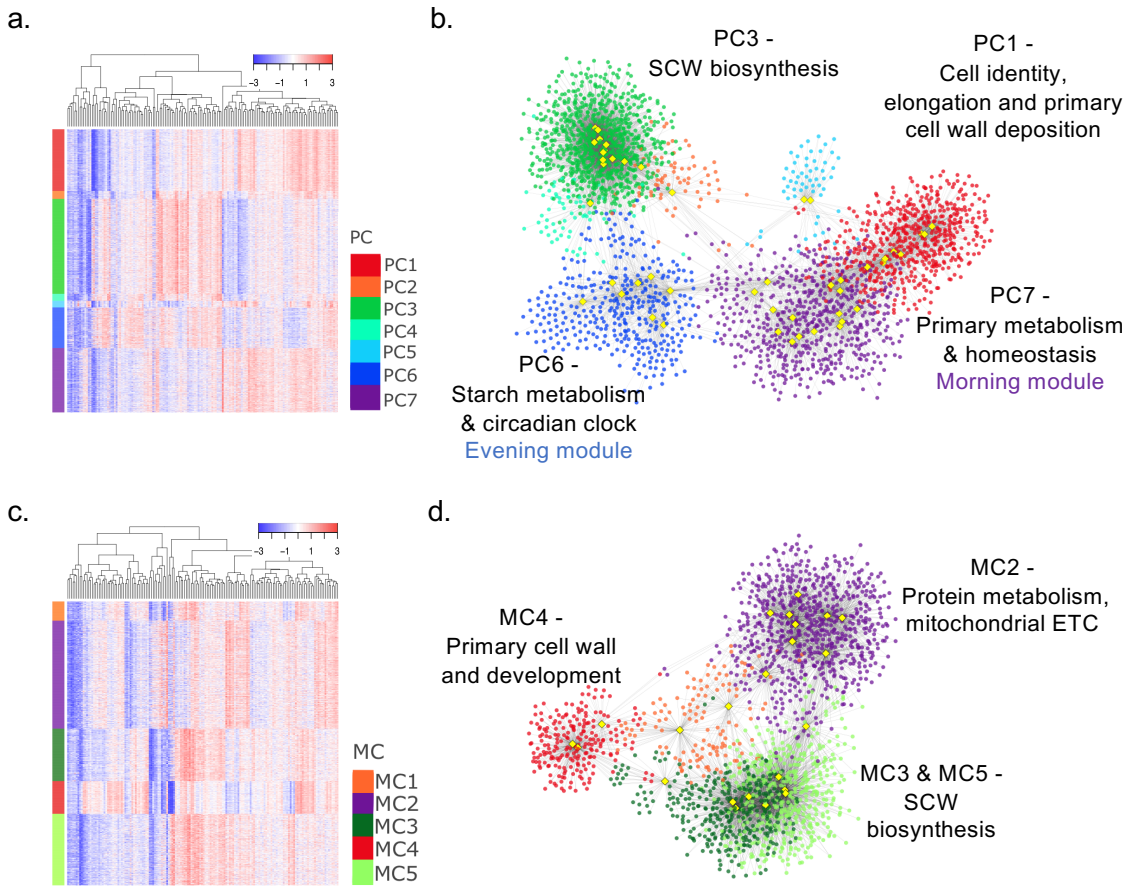


Figure 2.1. Network structure and function

Eucalyptus grandis x *E. urophylla* x *E. urophylla* (GUxU) backcross population plastid and mitochondrial network structure and function.

a. Heatmap of network clusters for the GUxU backcross population plastid query co-expression 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 to 7 are shown in the sidebar as follows: PC1 - red, PC2 - orange, PC3 - green, PC4 - turquoise, PC5 – light blue, PC6 – blue, and PC7 - purple. Mean centred and rescaled log10 FPKM Z-scores are shown from blue (low) to red (high) with the colour key in the top right.

b. Plastid co-expression network with clusters and query genes shown. Clusters are colour coded as in the heatmap.

c. Heatmap of network clusters of the mitochondrial query co-expression network. Mitochondrial clusters (MC) 1 to 5 are shown in the sidebar as follows: MC1 - orange, MC2 - purple, MC3 - dark green, MC4 - red, and MC5 - light green.

d. Mitochondrial co-expression 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 (SCW - secondary cell wall). Network layout is Cytoscape perfuse force directed layout, with correlation as distance between nodes.

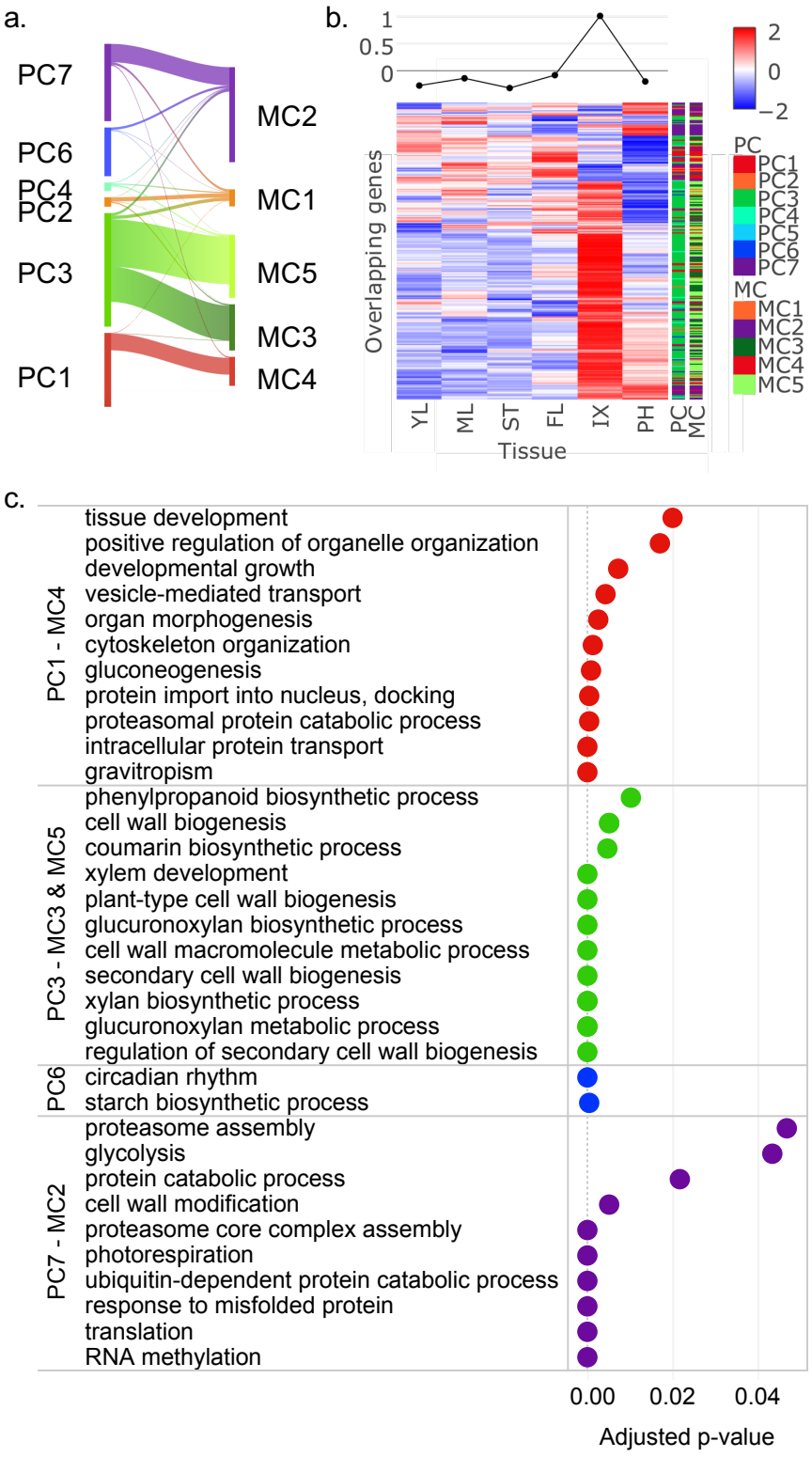

a.
PC7
PC6
PC4
PC2
PC3
PC1
MC2
MC1
MC5
MC3
MC4
b.
1
0.5
0
2
0
−2
Overlapping genes
YL
ML
ST
FL
IX
PH
PC
MC
Tissue
PC
PC1
PC2
PC3
PC4
PC5
PC6
PC7
MC
MC1
MC2
MC3
MC4
MC5
c.
PC1 - MC4
tissue development
positive regulation of organelle organization
developmental growth
vesicle-mediated transport
organ morphogenesis
cytoskeleton organization
gluconeogenesis
protein import into nucleus, docking
proteasomal protein catabolic process
intracellular protein transport
gravitropism
PC3 - MC3 & MC5
phenylpropanoid biosynthetic process
cell wall biogenesis
coumarin biosynthetic process
xylem development
plant-type cell wall biogenesis
glucuronoxylan biosynthetic process
cell wall macromolecule metabolic process
secondary cell wall biogenesis
xylan biosynthetic process
glucuronoxylan metabolic process
regulation of secondary cell wall biogenesis
PC6
circadian rhythm
starch biosynthetic process
PC7 - MC2
proteasome assembly
glycolysis
protein catabolic process
cell wall modification
proteasome core complex assembly
photorespiration
ubiquitin-dependent protein catabolic process
response to misfolded protein
translation
RNA methylation
0.00
0.02
0.04
Adjusted p-value

Figure 2.2. Overlapping gene analysis between the plastid and mitochondrial targeted carbon metabolic gene co-expression 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 *E. grandis* tissues in the plastid and mitochondrial clusters (Vining *et al*. 2015, Myburg *et al*., 2014). The row means centred 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.

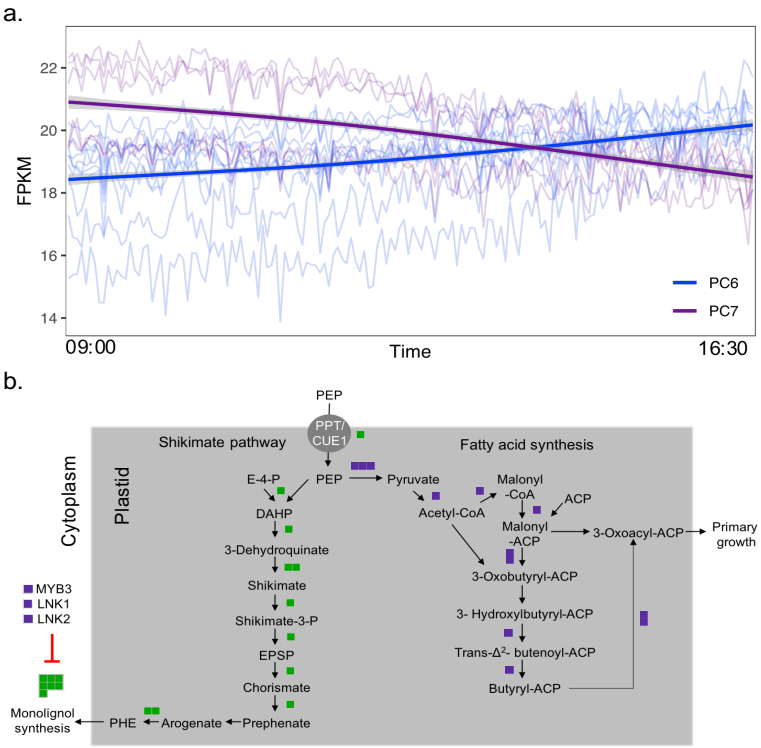


Figure 2.3. Circadian gene expression and PEP metabolism

a. Expression profile of *E. grandis* circadian clock genes across 156 individuals in the plastid network sorted by time of collection between 09:00 to 16:30 in log10 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 (P-E-P/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 Tables

S2.1 and **S2.2** for gene IDs and cluster assignment, and Notes **S2.3** for full annotation and tissue specific expression of these genes. ACP - acyl carrier protein; P-E-P - phosphoenolpyruvate; E-4-P - erythrose-4-phosphate; DAHP - 3-deoxy-D-arabino-heptulosonate 7-phosphate; EPSP - 5-enolpyruvylshikimate 3-phosphate; PHE - phenylalanine.

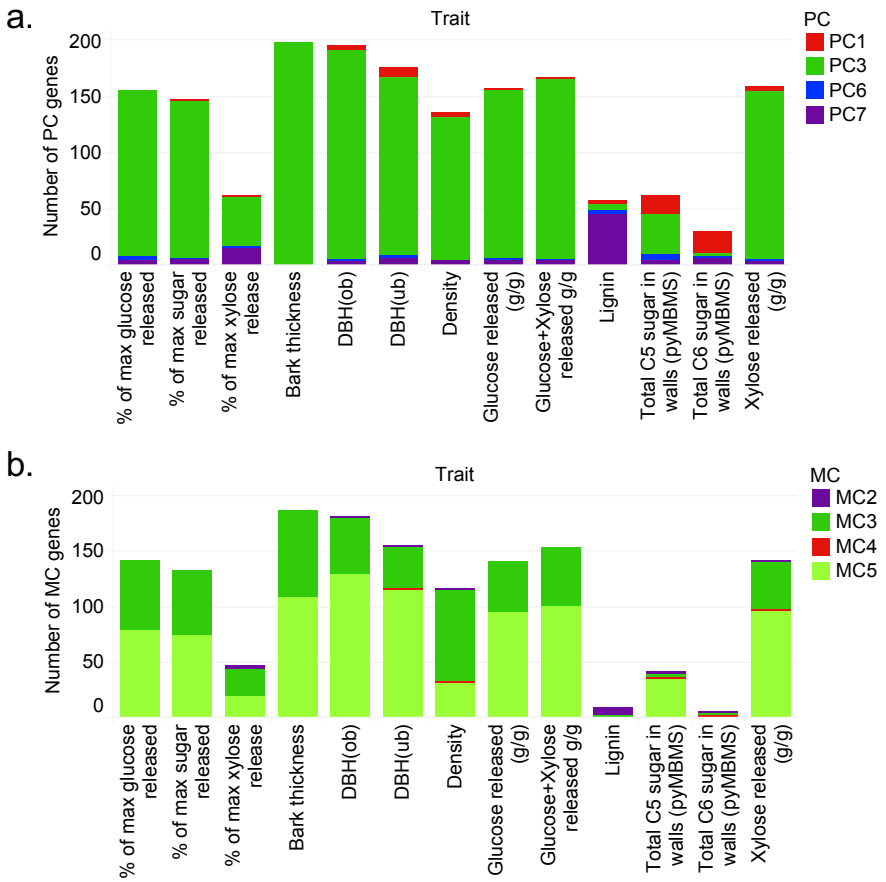


Figure 2.4. Network association with lignocellulosic biomass traits

Number of genes in plastid (**a.**) and mitochondrial (**b.**) 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 x E. urophylla x 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.

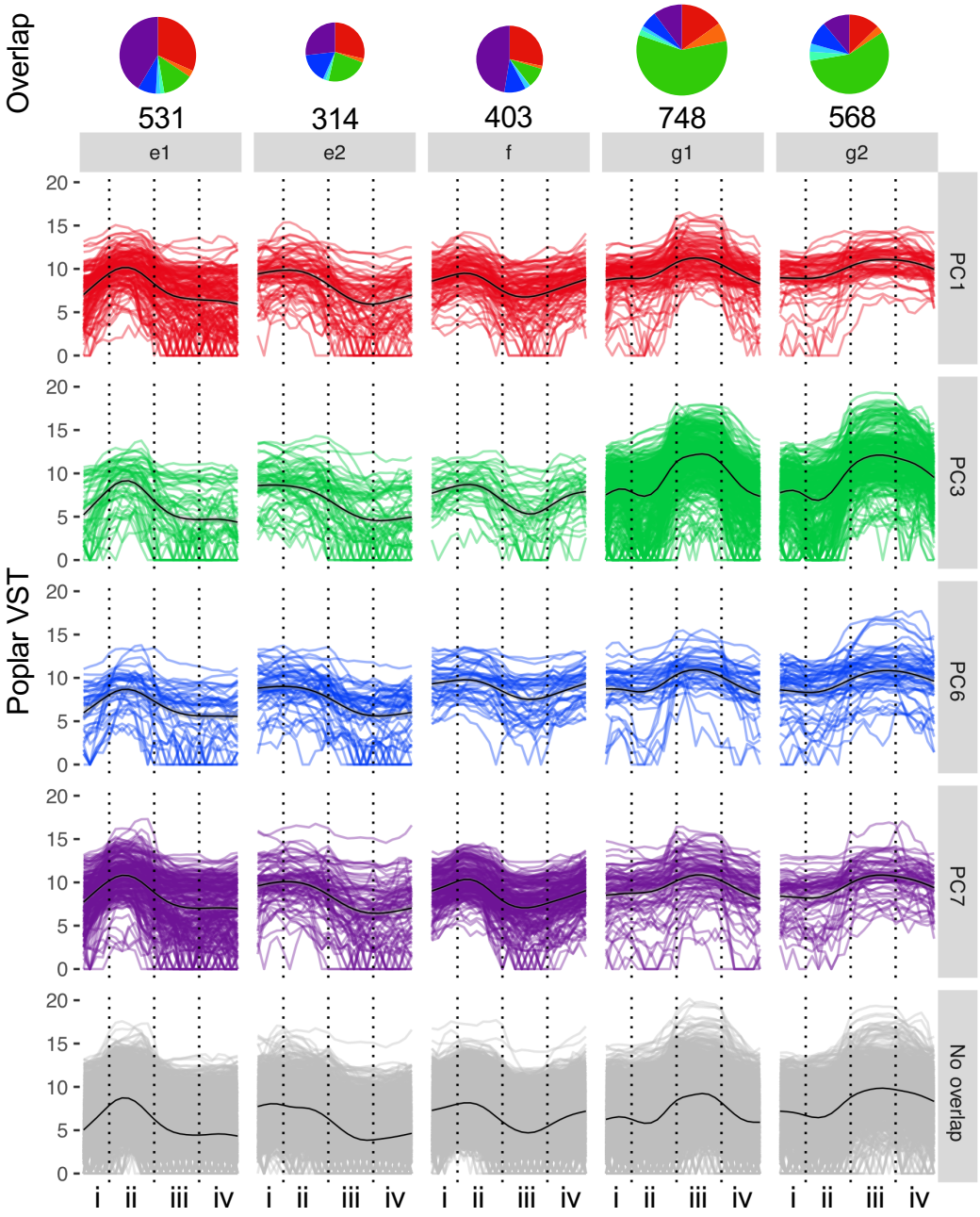


Figure 2.5. Xylem developmental series overlap with plastid clusters

Aspen developmental phase specific expression profiles of genes with orthologs in plastid clusters P1, P3, P6, and P7. The expression of aspen (*P. tremula*) genes over

the three main developmental transitions (cluster sections i to iv) of wood formation representing 25 high spatial resolution samples of wood development were taken from Sundell *et al*., 2017. *E. 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 to 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 Fig. **S2.4.**

2.10. Supporting information

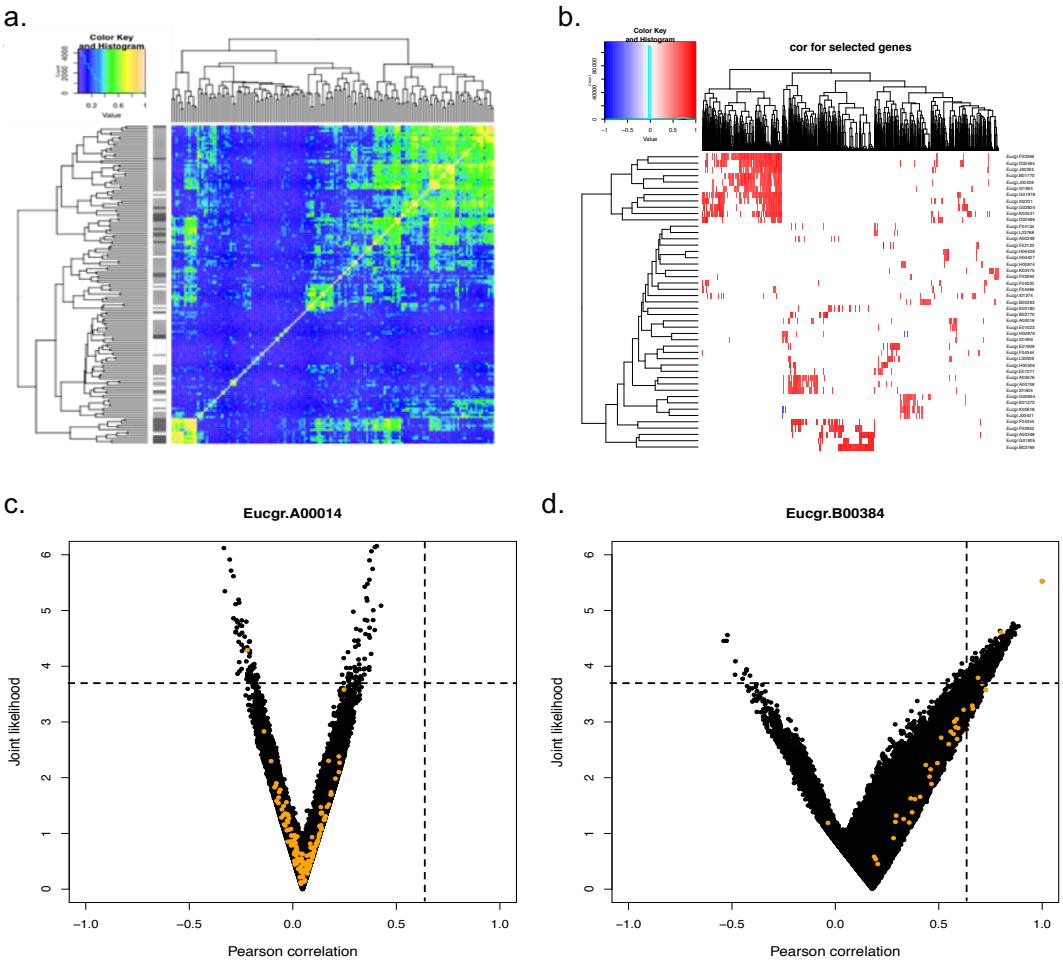


Figure S2.1. Overview of plastid network construction method.

a. Heatmap showing the absolute positive Pearson correlation between *E. grandis* plastid targeted query genes in the *Eucalyptus grandis × E. urophylla* GUxU backcross population. Pearson correlation is shown on a colour scale from low (blue) to high (yellow) as indicated by the colour key in the top left corner. The gray scale bars on the left-hand side of the heatmaps indicate log10 FPKM for each query gene (low- light grey to high- dark grey. The regions of higher correlation show query genes that are co-regulated between individuals. **b.** Heatmap of the adjacency matrix showing the genes that are correlated with plastid query genes, where Pearson correlation is shown from 1 (red) to -1 (blue). **c.**

and **d.** Joint likelihood score (y-axis) and Pearson correlation (x-axis) for two plastid query genes (*Eucgr.A00014* - not retained, and *Eucgr.B00384* - retained) showing the distribution of correlation for all genes. Additional query genes are indicated in orange. The dashed black lines at 3.8 on the y-axis and 0.6 on the x-axis show the selected thresholds for the plastid network, which minimize biological noise.

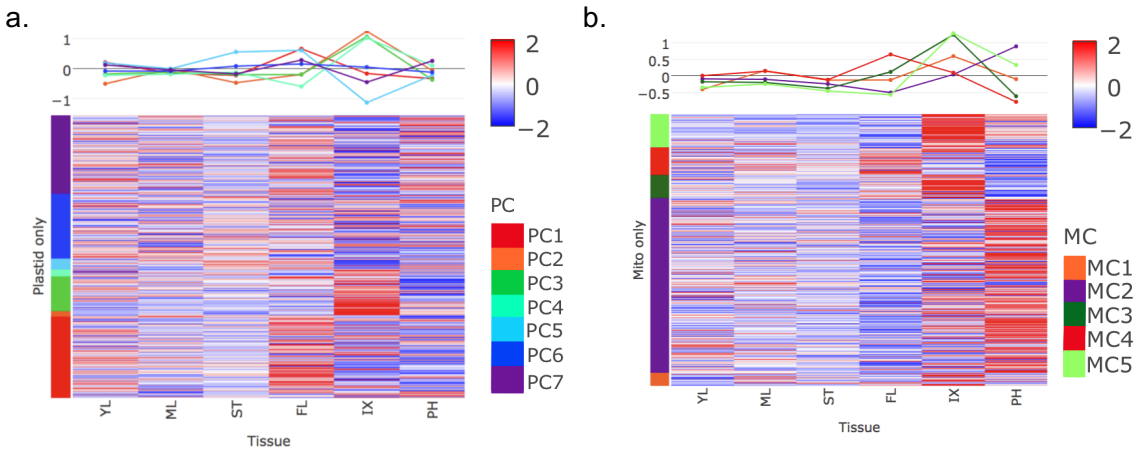


Figure S2.2. Tissue specific gene expression heatmaps of non-overlapping plastid and mitochondrial cluster genes.

Plastid and mitochondrial clusters (PC and MC) are indicated on the left side colour bars, and are coloured as shown in the key and throughout the manuscript. Expression levels are shown from high (red) to low (blue) as indicated by the colour bars in mean row scaled and centred log10 FPKM. The plots above the heatmaps show the average expression per tissue for each cluster.

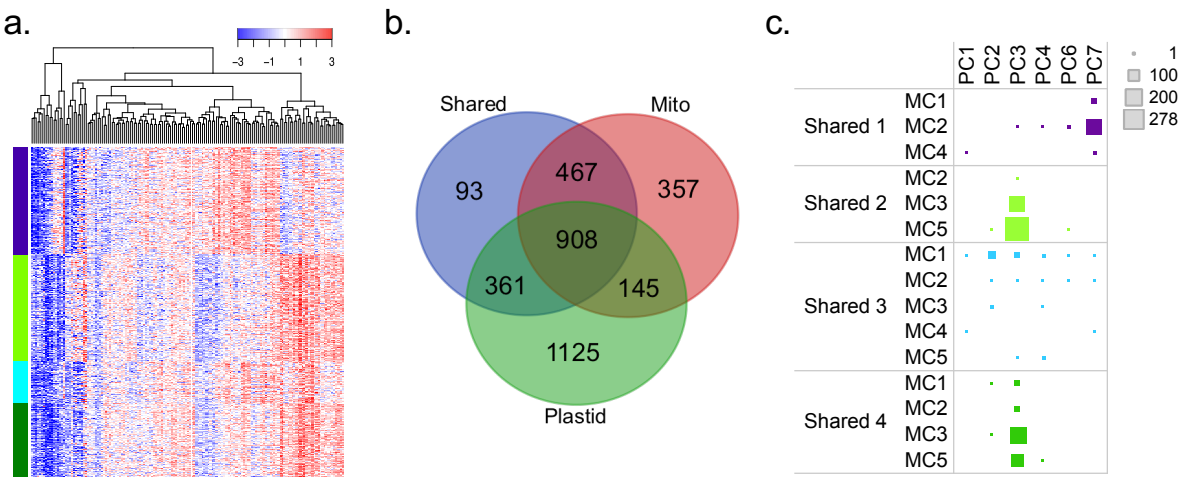
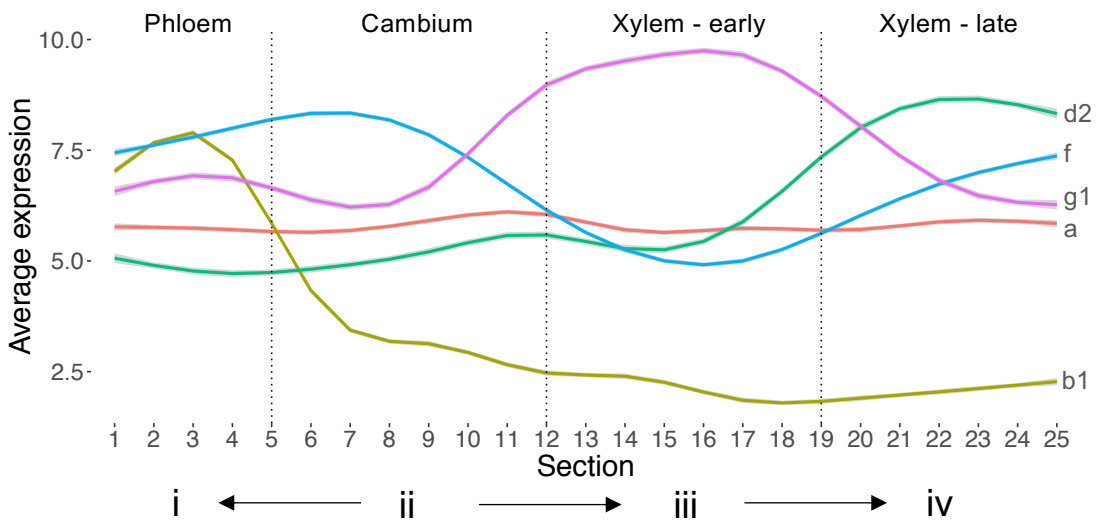


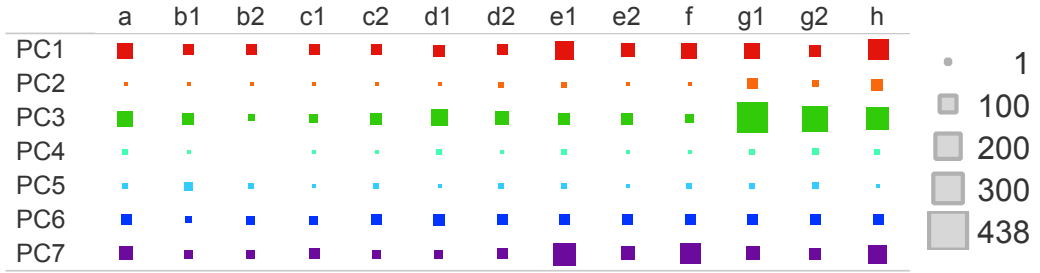
Figure S2.3. Network properties of the network constructed from shared plastid and mitochondrial query genes.

a. Heatmap and cluster assignment of the shared query gene network, where shared clusters (SC) 1 to 4 are coloured according to the largest overlap with plastid and mitochondrial clusters (PC and MC) from blue (low) to high (red) log10 row mean centred and scaled FPKM. **b**. Venn diagram over overlapping genes in the plastid, mitochondrial and shared networks. **c**. Overlapping genes in the shared network per plastid and mitochondrial cluster.

a.



b.



c.

Gene ID	PC	MC	YL	ML	ST	FL	IX	PH	Gene Family	PFAM
Eucgr.A02650	3	NA	10.92	0.00	10.90	12.97	20.00	14.05	HOM03D090514	-
Eucgr.B03756	4	1	21.30	22.31	21.18	21.45	23.12	22.41	HOM03D041116	-
Eucgr.C00194	7	NA	19.17	20.23	19.24	18.70	15.52	17.87	HOM03D048287	-
Eucgr.C00604	3	5	17.76	17.28	17.07	17.22	18.98	17.16	HOM03D093347	-
Eucgr.C00653	3	5	17.67	18.71	18.20	18.49	23.90	20.15	HOM03D089123	-
Eucgr.D02145	3	5	15.11	14.74	15.96	16.48	18.75	11.82	HOM03D059402	-
Eucgr.E03883	3	NA	0.00	0.00	0.00	0.00	16.60	0.00	HOM03D043030	-
Eucgr.F02005	3	5	19.52	18.92	18.39	17.94	21.96	19.38	HOM03D028910	-
Eucgr.F03327	4	2	18.75	18.72	18.53	18.55	18.88	18.43	HOM03D078484	-
Eucgr.G00032	3	5	15.47	17.43	16.15	15.98	19.33	15.54	HOM03D106565	-
Eucgr.G01754	3	5	13.35	14.70	13.23	13.45	19.91	16.57	HOM03D097560	-
Eucgr.G02424	4	5	12.75	14.52	13.71	14.62	18.25	15.39	HOM03D058564	-
Eucgr.H04611	6	NA	17.86	18.88	17.47	19.16	18.07	17.35	HOM03D069868	-
Eucgr.I01747	1	NA	15.41	15.99	16.04	17.74	19.60	18.78	HOM03D046085	-
Eucgr.I02604	4	5	15.05	16.75	16.40	15.77	18.19	15.69	HOM03D039729	-
Eucgr.J01188	3	5	16.02	15.10	15.58	13.03	19.10	15.04	HOM03D106008	-
Eucgr.J01205	6	NA	20.62	21.18	21.20	20.56	20.65	21.41	HOM03D036635	-
Eucgr.J02481	4	5	13.33	13.59	12.48	15.49	21.34	18.09	HOM03D042486	-
Eucgr.K00172	1	NA	19.85	19.99	19.94	20.48	19.40	19.41	HOM03D040843	-
Eucgr.K01854	7	2	19.51	19.98	20.16	20.10	19.88	20.71	HOM03D038554	-

Figure S2.4. Aspen developmental clusters, overlap with plastid clusters, and gene expression of *Eucalyptus* specific genes.

a. Average expression profiles of aspen clusters representing the major developmental transitions of xylogenesis where the clusters are defined by the peak of gene expression. Cluster b1 - peak expression in phloem; cluster f - peak expression in cambium; cluster g1 - peak expression in early xylem development; cluster d2 - peak expression in late xylem development; cluster a - no defined peak across the developmental series. The arrows indicate the developmental transitions as defined by Sundell *et al*., 2017. **b.** Overlap of all aspen and plastid clusters (a - h), where the colours are according to plastid cluster membership (PC1 - PC7), and size is the number of unique *Populus trichocarpa* gene IDs. **c**. Expression profile of *E. grandis* genes with no orthologs or homologs in any other plant species taken from PLAZA v3.0. The average log10 FPKM expression of genes is coloured from red (high) to blue (low), and the PLAZA homolog gene family and annotated PFAM domains are indicated. YL - young leaf; ML - mature leaf; ST - shoot tips; FL - flowers; IX - immature xylem; PH - phloem.

Table S2.1. Plastid cluster allocation of *Eucalyptus grandis* phenylalanine and monolignol biosynthetic genes.

***Eucalyptus* plastid cluster genes involved in phenylalanine and monolignol biosynthesis**		
Gene ID	**Gene name**	**Cluster**
Phenylalanine biosynthesis - Shikimate pathway		
Eucgr.F00366	DAHPS	3
Eucgr.L02803	DAHPS	NA
Eucgr.J02467	DHQS	3
Eucgr.B01770	DHQ/SDH	3
Eucgr.H01214	DHQ/SDH	NA
Eucgr.H04427	DHQ/SDH	5
Eucgr.H04428	DHQ/SDH	5
Eucgr.J00263	DHQ/SDH	3
Eucgr.F02664	SK1	NA
Eucgr.I00321	SK1	3
Eucgr.D02465	EPSPS	3
Eucgr.G02824	CS	3
Eucgr.I01994	CM	3
Eucgr.J00167	CM	NA
Eucgr.I00395	AAT	NA
Eucgr.I00398	AAT	NA
Eucgr.I00463	AAT	NA
Monolignol biosynthesis		
Eucgr.A01392	COMT4	3
Eucgr.A01397	COMT1	3
Eucgr.A02188	C3H2	3
Eucgr.A02190	C3H3	3
Eucgr.B02222	CCR3	6
Eucgr.C00065	C4H1	3
Eucgr.C01240	CCR4	5
Eucgr.C02284	4CL1	3
Eucgr.E01119	CAD22	3
Eucgr.F03978	HCT4	3
Eucgr.G01350	CAD2	3
Eucgr.G01417	CCoAOMT2	3
Eucgr.G02848	PAL3	3
Eucgr.G03199	C3H4	6
Eucgr.I01134	CCoAOMT1	3
Eucgr.I01783	CCR9	5
Eucgr.J00907	PAL8	5
Eucgr.J01079	PAL9	3
Eucgr.J01844	C4H2	3
Eucgr.J02393	F5H1	3
Eucgr.J03126	HCT5	3
Eucgr.K00087	4CL2	5
Eucgr.K00951	COMT61	3
Eucgr.F02557	CSE	3

Table S2.2. Plastid cluster allocation of *Eucalyptus grandis* circadian rhythm genes.

***Eucalyptus* circadian rhythm genes in the plastid co-expression network**			
Gene ID	**Gene name**	**Plastid cluster**	**Aspwood developmental phase**
Eucgr.A01405	PRR5	6	c2, b2
Eucgr.A02733	RVE6	6	g2
Eucgr.B00537	PRR1	6	d1
Eucgr.B02313	LUX/NOX	6	b1
Eucgr.C02997	ELF3	6	a
Eucgr.G02098	FKF1	6	c2, d1
Eucgr.J00214	ELF4a	6	d2
Eucgr.J00814	JMJD5	6	f
Eucgr.J02111	PRR7	6	c2
Eucgr.A01184	RVE8	7	c2
Eucgr.D02651	LHY	7	c2
Eucgr.F02906	RVE1	7	a, b2
Eucgr.G00771	LNK1	7	d1
Eucgr.J01191	LNK2a	7	a
Eucgr.K02302	RVE7	7	f

Notes S2.1. Word document of 17 KEGG carbon metabolism pathways used to identify plastid and mitochondrial carbon metabolism genes in *Arabidopsis thaliana*.

Notes S2.2. GO enrichment of plastid (Sheet 1), mitochondrial (Sheet 2), and shared (Sheet 3) network clusters, and the overlap between plastid clusters and aspen developmental phases (Sheet 4).

Notes S2.3. Annotation file of all *E. grandis* genes that are in the plastid (PC - Sheet 1), mitochondrial (MC - Sheet 2), and shared (SC - Sheet 3) networks, and *E. grandis* network overlaps with the aspen developmental phases, and aspen circadian clock genes.

Chapter 3

The plastid and mitochondrial genomes of *Eucalyptus grandis*

D. Pinard[1, 2], A. A. Myburg[1, 2], E. 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

This chapter has been prepared in the format of an original research publication for a peer-reviewed research journal. EM is the lead investigator and conceived of the study with DP. DP performed the data analysis and wrote the manuscript with EM. AAM is the lead investigator for the genome sequence and transcriptome analysis and edited the manuscript. This manuscript has been published in BMC Genomics under the same title.

3.1. Abstract

Background: Land plant organellar genomes have significant impact on metabolism and adaptation, and as such, accurate assembly and annotation of plant organellar genomes is an important tool in understanding the evolutionary history and interactions between these genomes. Intracellular DNA transfer is ongoing between the nuclear and organellar genomes, and can lead to significant genomic variation between, and within, species that impacts downstream analysis of genomes and transcriptomes.

Results: In order to facilitate further studies of cytonuclear interactions in *Eucalyptus*, we report an updated annotation of the *E. grandis* plastid genome, and the second sequenced and annotated mitochondrial genome of the Myrtales, that of *E. grandis*. The 478 813 bp mitochondrial genome shows the conserved protein coding regions and gene order rearrangements typical of land plants. There have been widespread insertions of organellar DNA into the *E. grandis* nuclear genome, which span 141 annotated nuclear genes. Further, we identify predicted editing sites to allow for the discrimination of RNA-sequencing reads between nuclear and organellar gene copies, finding that nuclear copies of organellar genes are not expressed in *E. grandis*.

Conclusions: The implications of organellar DNA transfer to the nucleus are often ignored, despite the insight they can give into the ongoing evolution of plant genomes, and the problems they can cause in many applications of genomics. Future comparisons of the transcription and regulation of organellar genes between *Eucalyptus* genotypes may provide insight to the cytonuclear interactions that impact economically important traits in this widely grown lignocellulosic crop species.

3.2. Background

Plastid and mitochondrial genomes are well studied aspects of land plant genomics, with 2 484 plastid (1 987 “chloroplast”, 506 “plastid”), and 167 mitochondrial genomes on NCBI for land plants as of June 2018, compared to the 141 nuclear genomes completed at the chromosome level [1]. A consequence of the endosymbiotic integration of plastids and mitochondria into plant cells is that the coding potential of their genomes is severely diminished compared to their ancestral genomes [2, 3]. The majority of organellar proteomes are encoded by the nuclear genome of plants, with ±97% of plastid, and ±99% of mitochondrial targeted proteins encoded by the nucleus [4]. Retained protein-coding organellar genes are essential to the metabolic functions of plastids and mitochondria, and variation in organellar genomes impact fitness and metabolism in angiosperms [5–9].

In plants, intracellular DNA transfer results in nuclear plastidial DNAs (NUPTs) and nuclear mitochondrial DNAs (NUMTs), that are still present in the organellar genomes [4]. Phylogenetic analysis of *Arabidopsis* and rice organellar DNA insertions show that large, primary insertions of organellar DNA into the nuclear genomes of plants occur, and these insertions decay over time [10]. The rate and distribution of organellar inserts into the nuclear genome vary between plant species, as do the location and proximity to transposable elements, which rearrange and expand inserted regions [10]. These recent inter-genomic DNA transfers between the nuclear and organellar genomes can result in multiple copies of organellar genes in the nuclear genome, presenting interesting avenues of research into the evolutionary history of plants and

the process of endosymbiosis, as ongoing gene transfer may lead to the loss of the organellar-encoded copy [11].

Key requirements to understanding the impact of organellar genome variation and transcript expression are high-quality annotated genomes, and a catalogue of intracellular genome transfers in order to distinguish between RNA originating from the organellar and nuclear genomes. Since it was sequenced in 2014, the *Eucalyptus* genome has become an important, and highly utilized genome for a variety of biological, ecological, and biotechnological studies [12]. Here, we update the assembly and annotation of the *E. grandis* plastid genome (adding 14 genes) and assemble and annotate the mitochondrial genome of *E. grandis*. We identified recent organellar genome transfers, and potential editing sites that can be used to distinguish transcripts originating from the organellar and nuclear genomes.

3.3. Results

3.3.1. Genome structure and gene content of the *E. grandis* mitochondrial genome

We used mitochondrial genome scaffolds from the Joint Genome Institute assembly of the *E. grandis* nuclear genome to perform a reference-based assembly of the mitochondrial genome from Illumina whole genome sequencing (WGS) data. The assembled mitochondrial genome is a single scaffold of 478 813 bp, with average GC content of 44.8% (Figure 3.1a and b, Table 3.1). The average coverage of the WGS

reads across the mitochondrial genome is ~700, with regions of ten times the average coverage representing overlaps between the plastid and mitochondrial genomes (Figure 3.1b) Repeat elements make up 2.47 % of the *E. grandis* mitochondrial genome, consisting mainly of simple and low complexity repeats (Table 3.1, Table S3.1). We identified 19 direct repeat regions larger than 100 bp in the *E. grandis* mitochondrial genome, the largest of which is 4 210 bp long (Figure 3.1b, Additional file 3.1). Additionally, we identified 11 inverted repeat regions longer than 100 bp in the *E. grandis* mitochondrial genome, the largest of which is 1 352 bp (Figure 3.1b, Additional file 3.1). Due to the fact that we could not assemble a circular mitochondrial genome for *E. grandis* from whole genome sequencing data, we considered that the genome may indeed be present as a linear molecule, or as sub-genomic molecules that arise via recombination of the repeat regions [13]. We did not find evidence of sub-genomic molecules from the depth of coverage across the mitochondrial genome assembly (Figure 3.1b). We used SVDetect to determine if any structural variations exist by filtering the alignment file based on the distance and orientation of aligned reads, along with removing any reads whose mate mapped to the nuclear or plastid genomes [14]. The SVDetect defined breakpoints were cross-referenced with the large repeat regions, and the results suggest that most repeat regions are not mediating mitochondrial sub-genomic molecules, as 8 breakpoints are within 250 bp of a repeat region and are found to predominantly be supported by less than 100 read pairs, with three being supported by 309, 219, and 163 read pairs (Additional file 3.2). Of these, direct repeat 13 shows evidence of repeat mediated structural variation, supported by 219 read pairs (Figure S3.1). Any other alternate conformations present

in the *E. grandis* mitochondrial genome could not be identified using this data and could be further assessed using long-read sequencing in the future.

A total of 39 protein coding genes were annotated in the genome, in addition to 20 annotated tRNA and 4 rRNA genes (Figure 3.1a, Table 3.2). The vast majority of the *E. grandis* mitochondrial genome is non-coding, with ~13% comprising of protein coding regions, and ~6% of introns. The mitochondrial protein coding genes are all single copy genes, with no duplications present (Table 3.2). The *E. grandis* mitochondrial genome does not contain any sequences similar to the ribosomal protein subunit genes *rps11*, *rps8*, and *rpl6*, which have been lost in angiosperms [15, 16]. Short fragments of *rps2* (141 nt) and *rps19* (42 and 69 nt) were found, but no full-length copies of these genes were present. The gene content was similar to other sequenced land plant mitochondrial genomes [17], with no genes exclusively lost in *E. grandis*. Ten *E. grandis* mitochondrial genes contain introns, with three of these, *nad1, nad2,* and *nad5*, being *trans*-spliced (Figure 3.1a, Table 3.2). There are 16 single copy tRNA genes in the *E. grandis* mitochondrial genome, with two copies each of *tRNA-Asn*, *tRNA-Met*, *tRNA-Tyr* and *tRNA-fMet* (Figure 3.1a, Table 3.1, Table 3.2). The mitochondrial genome of *E. grandis* contains four rRNA genes, with two copies of *5S rRNA* present (Figure 3.1a, Table 3.1, Table 3.2).

Recently, the first mitochondrial genome of the order Myrtales was released, that of *Lagerstroemia indica* (NC_035616.1). Compared to the *E. grandis* mitochondrial genome, the 333 948 bp long mitochondrial genome of *L. indica* is smaller, with a higher GC content at 46% compared to 44% in *E. grandis*. Of the annotated *L. indica*

mitochondrial genes, one has been lost in *E. grandis* (*rps19*), while two are not present in *L. indica* (*sdh3* and *rps13*). As is typical of land plant mitochondrial genomes [18], there has been massive re-arrangement of gene order between the two Myrtales families, with the largest block of collinear genes being *sdh4-cox3-atp8* (Figure S3.2). Further, *rpl16* has gained an intron in *L. indica*, which is not present in *E. grandis*. Given the diverse nature of the Myrtales [19], and the frequent rearrangements and gene losses present in mitochondrial genomes [20] (Figure S3.3), the differences between the two families are expected, and can be used in further phylogenetic analyses.

3.3.2. Genome structure and gene content of the *E. grandis* plastid genome

Although the plastid genome of *E. grandis* has been previously reported [21], some discrepancies in gene content exist when compared to other published *Eucalyptus* plastid genomes [22]. *Eucalyptus* plastid genomes typically contain 85 protein coding genes [22], and the available *E. grandis* plastid genome (NC_014570.1) contains 74 annotated protein coding genes [21, 23]. In the assembly reported here, the plastid genome of *E. grandis* was assembled using whole genome sequencing data (as above for the mitochondrial genome) and was subsequently annotated (Figure 3.2a). The assembled plastid genome of *E. grandis* is 160 109 bp long, having the quadripartite structure of most land plant plastid genomes, with two large inverted repeat (IR) regions that are flanked by two single copy (SC) regions (small- SSC and large- LSC) (Figure 3.2a). Coverage of the WGS reads aligned to the assembled plastid genome shows high coverage of 3 500x across the length of the genome (Figure 3.2b, Table

3.1). The high coverage of reads mapped give confidence in the downstream annotation and analysis of the assembled plastid genome.

The *E. grandis* plastid genome consists of 3.05% short repeat elements, the most abundant being simple and low complexity repeats (Table 3.1, Table S3.1) The genome contains 90 genes, which includes six pseudogenes, for a total of 84 protein coding genes (Figure 3.2a, Table 3.1). There are 37 annotated tRNA genes, representing 20 unique tRNAs. Introns are present in eight of the annotated tRNA genes, namely *tRNA-Lys, tRNA-Gly, tRNA-Leu, tRNA-Val, tRNA-Ile* (two copies), and *tRNA-Ala* (2 copies). The 8 rRNA genes in the plastid genome are found in the repeat regions, for a total of four unique rRNA genes. The intron structure of the plastid protein coding genes is highly conserved, with 11 genes containing at least two introns, of these, *ndhB* and *rpl12* are present as duplicates in the IR region. Three of the intron containing genes contain three exons; *ycf3*, *clpP*, and *rps12*. Two exons of *rps12* are present in the IR regions, and are trans-spliced to exon 1 found in the LSC region, as is common in land plants [24, 25]. The only difference in the coding regions of the previously published *Eucalyptus* plastid genomes (excluding the 2011 *E. grandis* plastid genome) is the annotation of *psbL*, which is annotated as a pseudogene, but has a predicted C to U editing site that creates a start codon (Additional file 3.3 - Sheet 2). The creation of a canonical start codon via C to U editing in the *psbL* gene has been well documented in other land plants [26, 27]. Thus, we include *psbL* as a *bona fide* gene in *E. grandis* plastid genome annotation.

3.3.3. Post-transcriptional editing in the organellar genomes of *E. grandis*

Land plant plastid and mitochondrial encoded transcripts are known to undergo extensive post-transcriptional C to U editing, which generally results in non-synonymous amino acids in the protein compared to the DNA and transcript, and can create and abolish start and stop codons [28]. In order to identify potential transcript editing sites in the *E. grandis* plastid and mitochondrial genomes, we predicted editing events using two homology based predictive approaches, PREPACT and PREP-suite [29, 30]. In the *E. grandis* mitochondrial genome, we identified 505 and 470 predicted C to U editing sites for an average of ~13 and ~12 editing sites per gene with PREPACT and PREP-mt respectively (Table 3.1, Figure S3.4a, Additional file 3.3 - Sheet 1). Three of the predicted edits create canonical AUG translational start sites in mitochondrial *nad1A*, *nad4L*, and *rps10*, which have been reported in other plant species [31–33]. Interestingly, mitochondrial *rpl16* is annotated as a pseudogene due to an internal stop codon (TAG). In other plant species, this codon position is encoded as CAG and is post-transcriptionally edited to a stop codon (TAG), leading to a downstream non-canonical start codon (GTG) being used instead [32, 34]. This GTG is conserved in the mitochondrial genome of *E. grandis*, and it may be possible that *rpl16* is not a pseudogene and is translated from the GTG codon.

Plastid protein coding gene transcripts are also post-transcriptionally edited by C to U, although the frequency of editing sites in plastid genomes are drastically lower in land plant plastids compared to mitochondria [35]. In the plastid genome of *E. grandis*, we report 49 predicted C to U editing sites as predicted by PREPACT, using *Arabidopsis thaliana* as reference protein databases, and 53 using PREP-cp (Table 1, Figure

S3.4b, Additional file 3.3 - Sheet 2) [29]. These editing sites exclude sites duplicated in the inverted repeat regions, keeping only the sites found in IRA, as it includes the full length of *ycf1*. These results are standard for the highly conserved plastid genomes of land plants [36, 37].

We found evidence of editing sites in the organellar genomes of *E. grandis* with 24 bulked polyA-selected, paired end transcriptome datasets from eight *E. grandis* tissues (Figure S4a and b). Using REDItools to discriminate between potential variants at the DNA level and true RNA editing sites, we could confirm 377 of the predicted mitochondrial editing sites, and 32 of the predicted plastid editing sites (Figure S3.4 c and d, Additional file 3.3) [38]. These include the predicted start codons of *psbL*, *nad4L*, and *rps10*. REDItools identified 52 mitochondrial and 6 plastid edits not predicted by either PREPACT or PREP-suite, (Figure S3.4c and d), which may be *bona fide* editing sites, or may be due the relatively low cut-offs defined in the analysis (total coverage >10 reads, at least three reads supporting the edit). Further, REDItools identified synonymous editing sites in codon position 1 of plastid and mitochondrial genes, one of which is found in the plastid genome, and six in the mitochondrial genome (Additional file 3.3). Due to the fact that the transcriptome data was prepared from polyA selected RNA, the editing sites identified should be confirmed using total RNA sequencing, as polyadenylated transcripts in organelles are destined for degradation, and do not accurately reflect organellar transcriptomes [39, 40].

3.3.4. DNA transfer between organellar and nuclear genomes

In order to identify transferred DNA between the nuclear and organellar genomes of *E. grandis*, we used BLAST analysis to identify sequences of significant homology between the three genomes. After filtering the BLAST analysis results for sequences longer than 100 bp with e-values < $1x10^{-3}$ and identity > 75%, we found a total of 751 886 bp of plastid origin and 1 256 558 bp of mitochondrial origin the nuclear genome (Figure 3.3, Table S3.2). The nuclear regions of organellar homology are distributed across all chromosomes of the nuclear genome (Figure 3.3, Table S3.2), with the largest proportion found on chromosome 5 for plastid DNA (88 691 bp), and chromosome 8 for mitochondrial DNA (193 727 bp). The mitochondrial genome of *E. grandis* consists of 6% (28 123 bp) chloroplast-like DNA sequences over 18 regions, with transfers ranging from 7 281 bp to 152 bp in length. A single plastid gene, *psbM*, has been transferred and annotated in the *E. grandis* mitochondrial genome. We find that eight tRNA genes in the mitochondrial genome overlap with plastid transferred regions (indicated by ^{pl} in Table 3.2). BLAST analyses of the inter-organellar DNA transfers against all NCBI land plant organellar genomes show that inter-organellar DNA transfers are from the plastid to the mitochondria, and that no mitochondrial to plastid DNA transfer has taken place in *E. grandis* (Table S3.3, Additional file 3.4).

Transferred DNA between the nuclear and organellar genomes of land plants creates the potential for complete transcript transfer that could be expressed from the nuclear genome [4]. In order to identify full length organellar transcripts in the nuclear genome of *E. grandis*, we used BLAST to align predicted organellar genes to *E. grandis* nuclear genes (> 80% of nuclear or organellar transcript length), and the annotation of the *E.*

grandis v2.0 genome (Figure 3.3, Additional file 3.5). We find 101 nuclear genes that have been transferred from the plastid genome (32 annotated as *A. thaliana* chloroplast genes, and 69 from the BLAST analysis). Further, there are 40 nuclear genes of mitochondrial origin (1 annotated as *A. thaliana* mitochondrial gene and 39 from the BLAST analysis). When genes without annotations in this group are examined for potential homologs in other plant species using the PLAZA database [41], we find that most of these nuclear genes are in fact orphan genes, with no homologs in the nuclear or organellar genomes of other dicot plant species. There are two exceptions, *Eucgr.J01097* and *Eucgr.J02736*, which are members of conserved gene families in plants. The first of these, *Eucgr.J01097,* is a homolog of a mitovirus RNA dependant polymerase [42], which occurs in the nuclear and mitochondrial genomes of 12 other dicot plant species (PLAZA family HOM03D004415) [34]. *Eucgr.J02736* forms part of a gene family that is present in six dicot plant species (PLAZA family HOM03006657) [41]. This gene is likely of plastid origin, as it is also found in the plastid transferred gene set (Additional file 3.5). Mismatches and indels present in the nuclear copies of full length organellar genes will allow for the identification of mRNA-seq reads mapped to the genome that they are expressed from (Additional file 3.5).

3.3.5. Transcription of NUMT and NUPT genes in *E. grandis*

In order to assess whether the *E. grandis* NUMTs and NUPTs identified above are functionally expressed, we aligned polyA-selected reads (from [43, 44]) to the nuclear genome, and compared read counts with the same reads aligned to the nuclear and

organellar genomes (Figure 3.4). To ensure that the reads aligned accurately to the organellar genomes, GSNAP was used with predicted organellar transcript editing sites defined as single nucleotide polymorphisms (SNPs) [45]. Evidence from eight tissues specific datasets revealed that organellar transferred genes in the nuclear genome are not functionally expressed (Figure 3.4). The reads aligning to the nuclear genome (Figure 3.4a) were drastically reduced when mapped to all three genomes simultaneously (Figure 3.4b), and instead, mapped preferentially to the organellar genomes (Figure 3.4c).

Of all the identified genes that are potentially transferred from the organellar genome to the nuclear genome, only one does not have decreased read counts when the polyA mRNA data is aligned to all three *E. grandis* genomes. This gene, *Eucgr.E01203*, was identified as a transferred gene due to its annotation as an *A. thaliana* chloroplast NADH-Ubiquinone/plastoquinone (complex I) protein gene (*ndhB2*). The parameters used in the BLAST analysis above did not identify this gene as an organellar transferred gene, as the CDS of this gene is truncated compared to the organellar *ndhB2* gene (*Eucgr.P00068*), with a length of 228 versus 1 533 nt. Read coverage across this gene in mature leaf tissue shows that the aligned reads do not span the annotated CDS, rather, they are found in the 5' UTR (Figure S3.5). The variance stabilizing transformation (VST) counts of *Eucgr.E01203* are thus unlikely to represent functional gene expression.

Organellar-encoded genes show that the polyA-selected mRNA reads aligned differentially across tissues in *E. grandis*. In general, the plastid and mitochondrial

genes have low numbers of reads aligning across all tissues, with some genes having high numbers of reads aligning in leaf and flower tissues (Figure 3.4c). Compared to the all nuclear-encoded genes, we identified 28 organellar genes with significant polyA-selected read abundance variation between immature xylem and mature leaf tissues (Table S3.4). All 28 of these organellar-encoded genes have decreased polyadenylated transcripts in immature xylem as compared to mature leaf. Of these, only one is a mitochondrial encoded gene, *Eucgr.M00039* (*maturase R*). The plastid differentially polyadenylated genes are predominantly photosystem genes (*psaA, B,* and *J,* and *psbA, B, C, D, E, H, I, J, K, L,* and *T*). The tissue specific nature of the read abundance variation in the photosystem genes specifically shows that these reads are not an artefact of transcripts "escaping" polyA selection based their GC content [46]. Further, we conclude that plastid-encoded photosystem genes are differentially polyadenylated between tissues in *E. grandis,* and that organellar-encoded genes are either not significantly polyadenylated or are lowly expressed in xylem.

3.4. Discussion

Organellar genomes are an important resource for many genomic and biotechnological applications [47], and as such, we aimed to provide a resource of high-quality sequences and annotations for the mitochondrial and plastid genomes of *Eucalyptus grandis*. The genus *Eucalyptus* consists of more than 700 species and their hybrids, many of which are economically and ecologically important [48, 49]. Additionally, *E. grandis* is an emerging model species for the study of xylogenesis [50]. The mitochondrial genome of *E. grandis* the second for the order Myrtales and should

facilitate further studies in the phylogeny of this order [51, 52]. The size of the *E. grandis* mitochondrial genome, GC content, number of coding genes, and predicted RNA editing sites is well within the range of sequenced land plant mitochondrial genomes [53]. The mitochondrial genome of *E. grandis* shares many features with other published land plant mitochondrial genomes, specifically the loss of *rps* and *rpl* subunits [54]. The genome structure of the mitochondrial genome is potentially linear, or present as sub-genomic circles due to the presence of large repeat regions [20, 55]. We could identify one repeat mediated structural variant from the aligned paired end reads, although any of the repeat regions could be involved in alternate conformations of the mitochondrial genome. As we could not confidently detect any other possible structural variants, mitochondrial DNA isolation from meristematic tissues or ovules [13, 55], and long-read sequencing methods may improve the assembly in future [56].

Organellar DNA is surprisingly mobile, and DNA transfers between organellar and nuclear genomes, and between species occur frequently [4], predominantly from the plastid and mitochondria to the nucleus [10], and from the nucleus and plastid to the mitochondria [9]. In many commercially important biomass crop species, large amounts of organellar DNA has been transferred to the nuclear genome. In *Populus trichocarpa* and *Gossypium raimondii*, near complete chloroplast and mitochondrial genomes respectively have been transferred to the nuclear genome [15, 57]. In *E. grandis*, we identified DNA transfers from the organelles to the nucleus, and from the plastid to the mitochondria (Figure 3.3, Table S3.1). Nuclear genes that align to the organellar genomes are gene fragments that have been annotated as complete genes due to the evidence of gene expression resulting from polyadenylated organellar

transcripts. Using next-generation RNA-sequencing, we were able to show that the NUMT and NUPT genes, and nuclear genes which align to the organellar genomes are not functionally expressed from the nuclear genome of *E. grandis* (Figure 3.4). Utilizing a method of SNP aware alignment, using predicted editing sites as SNPs, we show that reads in transferred regions preferentially align to the organellar genomes (Figure 3.4). Further analysis showed that feature counts, especially when they are extremely low, do not accurately reflect transcript expression, but rather fragmented alignment of a few reads across the transcript (Figure S3.5). This analysis allows for the confident alignment of mRNA reads to the three genomes of *E. grandis* for the quantification of organellar transcripts in future experiments.

The analysis of polyA-selected mRNA sequencing read alignment to the organellar genomes has value beyond identifying expressed NUMT and NUPT genes, as organellar genes are polyadenylated as a degradation signal [40, 58, 59]. We find that between mature leaf and immature xylem, the vast majority of differentially polyadenylated genes are photosystem genes from the chloroplast genome (Table S3.4). Photosystem genes are either not expressed, or very lowly expressed in non-photosynthetic tissues such as xylem [60, 61]. Given RNA turnover requirements, and imprecise transcriptional termination, the highly expressed photosystem genes in chloroplasts may lead to the polyadenylation of those transcripts in mature leaf [40]. Additionally, mature leaf chloroplast transcriptomes are differentially regulated compared to those in young leaf [62], and transcript degradation may play a role in this process.

3.5. Conclusion

This work provides a platform for further investigation into the myrtaceae by providing a reference genome and annotations for the mitochondria of *E. grandis*. The organellar genomes can be used in the future to study the transcription of organellar genes, and the tissue specific mechanism of transcriptional regulation by polyadenylation [5, 7]. Further, the co-evolution of nuclear and organellar genomes have been shown to affect hybrid vigour and speciation [63–66], and this work will allow for such studies in *Eucalyptus*, genera in which hybrids are ecologically and industrially important.

3.6. Methods

3.6.1. Assembly and annotation of the *E. grandis* organellar genomes

Paired end, whole genome sequencing reads of a three-year-old *E. grandis* genotype TAG0014 from mature leaf tissue was used in the assembly of the *E. grandis* mitochondrial and plastid genomes (SRP132546). The reads were sequenced by the Beijing Genomics Institute using the Illumina HiSeq 2000 platform. Contigs of mitochondrial origin, identified from the nuclear genome assembly project [12], were used as seed sequences for assembly using MITObim v1.6 using the -- quick flag, and kmer length of 41 [67]. The mitochondrial genome was assessed for circularity using the circules.py script available as a part of MITObim (https://github.com/chrishah/MITObim). SVDetect was used to determine if the WGS reads aligned to the mitochondrial genome assembly using Bowtie 2 showed evidence of alternative genome configurations [14, 68]. As they are mediated by large repeat

regions, alternate configurations of the genome can be identified from discordant read pairs that mapped in the wrong orientation, or at a distance larger or smaller than half the insert size (<250 bp, >750 bp), which were identified using SAMtools v1.3.1 view flag '-F 1294' [14, 69]. To avoid regions which may be artifacts of plastid and nuclear DNA transfer, we further removed all reads which were not uniquely mapped to the mitochondrial genome. The identified SVDetect breakpoints within 250 bp of a mitochondrial repeat region were identified using bedtools v2.27.1 closest [70].

The mitochondrial genome was manually annotated using a combination of homology-based predictions, namely Mitofy [71], MFannot [72], and Geneious v10.0.5 [73]. Similarly, the plastid genome was assembled using NOVOPlasty v1.1 with kmer length of 39 [74], with the previous *E. grandis* plastid genome as seed sequence (NC_014570.1). The plastid genome was manually annotated using DOGMA [75], CpGAVAS [76], Geneious v10.0.5 [73], and MFannot [72].

Transcript editing sites were identified using the PREPACT web server and PREP-suite (Mt and Cp) for both genomes [29, 30]. For PREPACT analysis of the mitochondrial genome, *Arabidopsis thaliana*, *Nicotiana tabacum*, and *Vitis vinifera* was used to identify conserved C to U edits using BLASTx prediction, with stop codons edited if possible and all other parameters kept at default. A predicted editing site was classified as being predicted by PREPACT if it occurred in at least two of the species used for prediction. For the PREPACT plastid genome editing site prediction, *Arabidopsis thaliana* was used as reference protein database for BLASTx prediction, with all other parameters kept at default. For PREP-suite analysis of the plastid and

mitochondrial genomes, a prediction confidence cut-off of 0.5 was used to predict editing sites, with all other parameters at default. Low-complexity repeats were identified in both genomes using RepeatMasker [77], with reference set to *Arabidopsis thaliana*, and all other parameters as default. Large genomic repeats were identified with Unipro UGENE [78], with repeat identity set to > 95%, and repeat length >100 nt. Both genomes were visualized with OrganellarGenomeDRAW [79], and WGS reads were aligned using Bowtie 2 [68] to visualize coverage using the Integrative Genomics Viewer (IGV [80]).

3.6.2. Identification and analysis of NUMTs and NUPTs in the *E. grandis* nuclear genome

BLAST (BLAST 2.3.0+) hits of > 100 bp, e-value > $1x10^{-5}$, and 75% identity were used in the analysis of NUPTs and NUMTs, and inter-organellar genome transfer [81]. Regions originating from the IR regions of the plastid genome were counted once, unless they spanned the SC flanking regions. Inter-organellar DNA transfers were assigned an organelle of origin using a custom BLAST database of all land plant organelles retrieved from GenBank in June 2017 [51]. Results of the DNA transfer analysis outlined above were visualized using Circos v0.69 [82] with transferred regions > 500 bp shown for clarity. In order to identify transferred protein coding genes between the nuclear and organellar genomes, BLAST analysis of full-length transcripts from the organellar genomes to the complete nuclear genome and vice versa was done. Transcripts are considered complete transfers if they covered > 80% of the transcript length in either the nuclear or organellar gene and had > 75% identity

between transcripts. Nuclear genes that are annotated as organellar genes were identified based on their closest *A. thaliana* homolog from the *E. grandis* v2 nuclear genome annotation [12].

3.6.3. PolyA-selected mRNA sequencing alignment, quantification, and editing analysis

PolyA-selected, paired end mRNA sequencing data from eight *E. grandis* tissues (as described in [43, 44]) were aligned to all three *E. grandis* genomes using GSNAP with allowed mismatch set to 1 (gmap v2016-09-23 [45]). Predicted editing sites of the organellar transcripts identified in the annotation step were used as SNP files for GSNAP alignment in order not to bias the alignment towards the nuclear genome. The resulting sam alignment files were converted to bam format using SAMtools view and sorted by position with SAMtools sort (SAMtools v1.3.1 [69]). The sorted bam files were then used to generate raw feature counts using HTSeq-count v0.6.1 [83] with concatenated nuclear and organellar gtf annotation files. DESeq2 v1.8.2 [84], implemented in RStudio v1.0.136 [85], was used to generate variance stabilized transformed (VST) counts and identify differentially expressed genes between immature xylem, phloem, and mature leaf tissue samples. The results were visualized using ggplot2 v2.2.1 [86] in RStudio v1.0.136.

REDItools version 1.0.4 [38] was used to identify editing sites using the aligned polyA selected reads across 8 *E. grandis* tissues. We used the REDItoolDnaRna.py script to ensure that organellar genomic variants were not called as editing sites due to

transferred DNA regions, using the Bowtie 2 [68] genomic DNA alignments to differentiate between DNA variants and RNA editing [38]. The settings used were as follows: predict C to U and G to A edits (for sense and antisense genes, respectively), editing sites must have > 10 reads aligned, with > 3 reads supporting the editing event, minimum per base quality > 25. We then filtered the identified editing sites based on the following parameters: No DNA variants in the site, sense orientation with organellar gene coding regions (C to U for sense genes, and G to A for antisense genes). All tissue samples were bulked, and edits were identified if they were found in any dataset and were in codon position 1 or 2 of the sense strand of plastid and mitochondrial genes.

3.7. Abbreviations

NUPT: nuclear plastidial DNA

NUMT: nuclear mitochondrial DNA

WGS: whole genome sequencing

bp: base pairs

IR: inverted repeat

SC: single copy

UTR: untranslated region

polyA: polyadenylated

VST: variance stabilizing transformation

SNP: single nucleotide polymorphism

3.8. Availability of data and material

FastQ files of the whole genome sequencing of *E. grandis* TAG0014 mature leaf tissue are available at the Sequence Read Archive under the project accession PRJNA433608 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA433608). Mitochondrial and plastid genome sequences and annotations have been submitted to the NCBI Genbank database and can be found under accession number NC_040010.1 for the mitochondria (https://www.ncbi.nlm.nih.gov/genome/browse/?report=5#!/organelles/NC_040010), and MG925369.1 for the plastid (https://www.ncbi.nlm.nih.gov/nuccore/MG925369).

3.9. Acknowledgements

We thank Ms. Karen van der Merwe for generating the Circos diagram used in this study.

3.10. References

1. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, *et al*. GenBank. Nucleic Acids Res. 2018;46:D41–7.

2. Stiller JW. Plastid endosymbiosis, genome evolution and the origin of green plants. Trends Plant Sci. 2007;12:391–6.

3. Gualberto JM, Mileshina D, Wallet C, Niazi AK, Weber-lot F, Dietrich A. The plant mitochondrial genome : Dynamics and maintenance. BMC Genomics. 2014;100:107–20.

4. Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol. 2009;60:115–38.

5. Joseph B, Corwin JA, Züst T, Li B, Iravani M, Schaepman-Strub G, *et al*. Hierarchical nuclear and cytoplasmic genetic architectures for plant growth and defence within *Arabidopsis*. Plant Cell. 2013;25 June:1929–45.

6. Bock D, Andrew RL, Rieseberg LH. On the adaptive value of cytoplasmic genomes in plants. Mol Ecol. 2014;23:4899–911.

7. Budar F, Roux F. The role of organelle genomes in plant adaptation. Plant Signal Behav. 2016;2324 February. doi:10.4161/psb.6.5.14524.

8. Wright AF, Murphy MP, Turnbull DM. Do organellar genomes function as long-term redox damage sensors? Trends Genet. 2009;25:253–61.

9. Kubo T, Newton KJ. Angiosperm mitochondrial genomes and mutations. Mitochondrion. 2008;8:5–14.

10. Michalovova M, Vyskot B, Kejnovsky E. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. Heredity . 2013;111:314–20.

11. Rockenbach K, Havird JC, Monroe JG, Triant DA, Taylor DR, Sloan DB. Positive selection in rapidly evolving plastid-nuclear enzyme complexes. Genetics. 2016;204:1507–22.

12. Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, *et al*. The genome of *Eucalyptus grandis*. Nature. 2014;510:356–62.

13. Sloan DB. One ring to rule them all? Genome sequencing provides new insights into the “master circle” model of plant mitochondrial DNA structure. New Phytol. 2013;200:978–85.

14. Zeitouni B, Boeva V, Janoueix-Lerosey I, Loeillet S, Legoix-Né P, Nicolas A, *et al*. SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. Bioinformatics. 2010;26:1895–6.

15. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, *et al*. Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. Biomed Res Int. 2016;2016:5040598.

16. Ye N, Wang X, Li J, Bi C, Xu Y, Wu D, *et al*. Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. PeerJ. 2017;5:e3148.

17. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, *et al*. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS biol. 2012;10:1–17.

18. Park S, Ruhlman TA, Sabir JSM, Mutwakil MHZ, Baeshen MN, Sabir MJ, *et al*. Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. BMC Genomics. 2014;15:405.

19. Berger BA, Kriebel R, Spalink D, Sytsma KJ. Divergence times, historical biogeography, and shifts in speciation rates of Myrtales. Mol Phylogenet Evol. 2016;95:116–36.

20. Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. Annu Rev Plant Biol. 2017;68:225–52.

21. Paiva JAP, Prat E, Vautrin S, Santos MD, San-Clemente H, Brommonschenkel S, *et al*. Advancing *Eucalyptus* genomics: identification and sequencing of lignin

biosynthesis genes from deep-coverage BAC libraries. BMC Genomics. 2011;12:137–50.

22. Bayly MJ, Rigault P, Spokevicius A, Ladiges PY, Ades PK, Anderson C, *et al*. Molecular Phylogenetics and evolution chloroplast genome analysis of australian eucalypts – *Eucalyptus, Corymbia, Angophora, Allosyncarpia* and *Stockwellia* (Myrtaceae). Mol Phylogenet Evol. 2013;69:704–16.

23. Steane DA. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). DNA Res. 2005;12:215–20.

24. Hildebrand M, Hallick RB. Trans-splicing in chloroplasts: the *rps12* loci of *Nicotiana tabacum*. PNAS. 1988;85: 372-376.

25. Schmitz-Linneweber C, Williams-Carrier RE, Williams-Voelker PM, Kroeger TS, Vichas A, Barkan A. A pentatricopeptide repeat protein facilitates the trans-splicing of the maize chloroplast *rps12* pre-mRNA. Plant Cell. 2006;18:2650–63.

26. Bock R, Hagemann R, Kossel H, Kudla J. Tissue- and stage-specific modulation of RNA editing of the psbF and psbL transcript from spinach plastids- A new regulatory mechanism? Mol Genet Genomics. 1993;240:238–44.

27. Kudla J, Gabor L, Metzlaff M, Hagemann R, Kossel H. RNA editing in tobacco chloroplasts leads to the formation of a translatable psbL mRNA by a C to U substitution within the initiation codon. Embo. 1992;1:1099–103.

28. Ichinose M, Sugita M. RNA editing and its molecular mechanism in plant organelles. Genes . 2016;8:1-15

29. Lenz H, Knoop V. PREPACT 2.0: Predicting C-to-U and U-to-C RNA editing in organelle genome sequences with multiple references and curated RNA editing annotation. Bioinform Biol Insights. 2013;7:1–19.

30. Mower JP. The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res. 2009;37 Web Server issue:W253–9.

31. Park S, Grewe F, Zhu A, Ruhlman TA., Sabir J, Mower JP, *et al*. Dynamic evolution of Geranium mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol. 2015. doi:10.1111/nph.13467.

32. Picardi E, Horner DS, Chiara M, Schiavon R, Valle G, Pesole G. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. Nucleic Acids Res. 2010;38:4755–67.

33. Jackman SD, Warren RL, Gibb EA, Vandervalk BP, Mohamadi H, Chu J, *et al*. Organellar genomes of white spruce (*Picea glauca*): Assembly and annotation. Genome Biol Evol. 2015;8:29–41.

34. Bock H, Brennicke A, Schuster W. Rps3 and rpl16 genes do not overlap in *Oenothera* mitochondria: GTG as a potential translation initiation codon in plant mitochondria? Plant Mol Biol. 1994;24:811–8.

35. Takenaka M, Zehrmann A, Verbitskiy D, Härtel B, Brennicke A. RNA editing in plants and its evolution. Annu Rev Genet. 2013;47:335–52.

36. Freyer R, Kiefer-Meyer M-C, Kossel H. Occurrence of plastid RNA editing in all major lineages of land plants. PNAS. 1997;94:6285–6290.

37. Tseng C-C, Lee C-J, Chung Y-T, Sung T-Y, Hsieh M-H. Differential regulation of *Arabidopsis* plastid gene expression and RNA editing in non-photosynthetic tissues. Plant Mol Biol. 2013;82:375–92.

38. Picardi E, D'Erchia AM, Montalvo A, Pesole G. Using REDItools to detect RNA editing events in NGS datasets. Curr Protoc Bioinformatics. 2015;49:12.12.1–15.

39. Levy S, Schuster G. Polyadenylation and degradation of RNA in the mitochondria. Biochem Soc Trans. 2016;44:1475–82.

40. Lisitsky I, Schuster G. Polyadenylation and Degradation of mRNA in the Chloroplast. In: The Chloroplast: From Molecular Biology to Biotechnology. Springer, Dordrecht; 1999. p. 85–90.

41. Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, *et al*. PLAZA: a comparative genomics resource to study gene and genome evolution in plants. Plant Cell. 2009;21:3718–31.

42. Hong Y, Cole TE, Brasier CM, Buck KW. Evolutionary relationships among putative RNA-dependent RNA polymerases encoded by a mitochondrial virus-like RNA in the dutch elm disease fungus, *Ophiostoma novo-ulmi*, by other viruses and virus-like RNAs and by the *Arabidopsis* mitochondrial genome. Virology. 1998;246:158–69.

43. Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA. De novo assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. BMC Genomics. 2010;11:681–93.

44. Vining KJ, Romanel E, Jones RC, Klocko A, Alves-Ferreira M, Hefer CA, *et al*. The floral transcriptome of *Eucalyptus grandis*. New Phytol. 2015;206:1406–22.

45. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. Methods Mol Biol. 2016;1418:283–334.

46. Smith DR. RNA-Seq data: a goldmine for organelle research. Brief Funct Genomics. 2013;12:454–6.

47. Bock R. Engineering plastid genomes: methods, tools, and applications in basic research and biotechnology. Annu Rev Plant Biol. 2015;66:211–41.

48. Rozefelds AC. *Eucalyptus* phylogeny and history: a brief summary. TASFORESTS-HOBART-. 1996;8:15–26.

49. Folk RA, Mandel JR, Freudenstein JV. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of Angiosperms. Syst Biol. 2017;66:320–37.

50. Grattapaglia D, Vaillancourt RE, Shepherd M, Thumma BR, Foley W, Külheim C, *et al*. Progress in Myrtaceae genetics and genomics: *Eucalyptus* as the pivotal genus. Tree Genet Genomes. 2012;8:463–508.

51. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, *et al*. GenBank. Nucleic Acids Res. 2013;41 Database issue:D36–42.

52. Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, *et al*. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181:1–20.

53. Liu G, Cao D, Li S, Su A, Geng J, Grover CE, *et al*. The complete mitochondrial genome of *Gossypium hirsutum* and evolutionary analysis of higher plant mitochondrial genomes. PLoS one. 2013;8:e69476.

54. Maier UG, Zauner S, Woehle C, Bolte K, Hempel F, Allen JF, *et al*. Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. Genome Biol Evol. 2013;5:2318–29.

55. Woloszynska M. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. J Exp Bot. 2009;3:657–671.

56. Shearman JR, Sonthirod C, Naktang C, Pootakham W, Yoocha T, Sangsrakru D, *et al*. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads. Sci Rep. 2016;6:31533.

57. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, *et al*. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 2006;313:1596–604.

58. Kuhn J, Tengler U, Binder S. Transcript lifetime is balanced between stabilizing stem-loop structures and degradation-promoting polyadenylation in plant mitochondria. Mol Cell Biol. 2001;21:731–42.

59. Hayes R, Kudla JO, Gruissem W. Degrading chloroplast mRNA: The role of polyadenylation. Trends Biochem Sci. 1999;24:199–202.

60. Valkov VT, Scotti N, Kahlau S, Maclean D, Grillo S, Gray JC, *et al*. Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: Transcriptional and posttranscriptional control. 2009;150 August:2030–44.

61. Kahlau S, Bock R. Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: chromoplast gene expression largely serves the production of a single protein. Plant Cell. 2008;20:856–74.

62. Woo HR, Koo HJ, Kim J, Jeong H, Yang JO, Lee IH, *et al*. Programming of plant leaf senescence with temporal and inter-organellar coordination of transcriptome in *Arabidopsis*. Plant Physiol. 2016;171:452–67.

63. Moison M, Roux F, Quadrado M, Duval R, Ekovich M, Lê D-H, *et al*. Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in *Arabidopsis thaliana*. Plant J. 2010;63:728–38.

64. Dobler R, Rogell B, Budar F, Dowling DK. A meta-analysis of the strength and nature of cytoplasmic genetic effects. J Evol Biol. 2014;27:2021–34.

65. Soltani A, Kumar A, Mergoum M, Pirseyedi SM, Hegstad JB, Mazaheri M, *et al*. Novel nuclear-cytoplasmic interaction in wheat (*Triticum aestivum*) induces vigorous plants. Funct Integr Genomics. 2016;16:171–82.

66. Roux F, Mary-Huard T, Barillot E, Wenes E, Botran L, Durand S, *et al*. Cytonuclear interactions affect adaptive traits of the annual plant *Arabidopsis thaliana* in the field. Proc Natl Acad Sci U S A. 2016;113:3687–92.

67. Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - A baiting and iterative mapping approach. Nucleic Acids Res. 2013;41:1–9.

68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al*. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

70. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014;47:11.12.1–34.

71. Beck N, Lang BF. MFannot, organelle genome annotation webserver. 2010. http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl.

72. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, *et al*. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.

73. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2016;:gkw955.

74. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004;20 17 :3252–5.

75. Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, *et al*. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genomics. 2012;13:715.

76. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics. 2009;:4–10.

77. Protsyuk IV, Grekhov GA, Tiunov AV, Fursov MY. Shared bioinformatics databases within the Unipro UGENE platform. J Integr Bioinform. 2015;12:257.

78. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 2013;41:W575–81.

79. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14:178–92.

80. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215:403–10.

81. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, *et al*. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45.

82. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al*. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

83. Anders S, Pyl PT, Huber W. HTSeq: Analysing high-throughput sequencing data with Python. 2010.

84. Love M, Anders S, Huber W. Differential analysis of count data–the DESeq2 package. Genome Biol. 2014. https://bioc.ism.ac.jp/packages/3.3/bioc/vignettes/DESeq2/inst/doc/DESeq2.pdf.

85. Team RS. RStudio: integrated development for R. RStudio, Inc, Boston, MA URL http://www rstudio com. 2015.

86. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag New York; 2009. http://had.co.nz/ggplot2/book.

87. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS one. 2010;5:e11147.

3.11. Tables

Table 3.1. *E. grandis* organellar genome and gene content characteristics

METRIC	MITOCHONDRIA	PLASTID
GENOME SIZE (BP)	478 813	160 109
GC CONTENT	44.80%	36.90%
WGS COVERAGE	~700x	~3 456x
WGS INSERT SIZE	475 bp	475 bp
WGS READ LENGTH	100 bp	100 bp
TOTAL LENGTH OF HOMOLOGOUS REGIONS IN THE NUCLEAR GENOME	1 256 558 bp	751 886 bp
% OF INTER-ORGANELLAR DNA TRANSFERS	6% (28 123 bp)	None
PROTEIN CODING GENES	39	84 (76 unique)
tRNA	20	38 (20 unique)
rRNA	4 (3 unique)	8 (4 unique)
CODING GENES WITH INTRONS	9	11
tRNA WITH INTRONS	0	8
# OF PREDICTED EDITING SITES (PREP/PREPACT)	470/505	49/53
EDITING SITES/GENE (PREP/PREPACT)	12/13	0.6/0.7
% OF GENOME REPEAT ELEMENTS	2.47%	3.05%

Table 3.2. *E. grandis* mitochondrial genome gene content

Complex I	nad1 (5*)	nad2 (5*)	nad3	nad4 (4)	nad4L
	nad5 (4*)	nad6	nad7 (5)	nad9	
Complex III	cob				
Complex IV	cox1	cox2 (2)	cox3		
Complex V	atp1	atp4	atp6	atp8	atp9
Cytochrome C biogenesis	ccmB	ccmC	ccmFC (2)	ccmFN	
Ribosomal large subunit	rpl2 (2)	rpl5	rpl10	rpl16[a]	
Ribosomal small subunit	rps3 (2)	rps4	rps7	rps10 (2)	rps11
	rps12	rps13	rps14		
Intron maturase	matR				
Protein translocase	mttB				
Other	psbM[pl]				
rRNA genes	26S rRNA	18S rRNA	5S rRNA (x2)		
tRNA genes	tRNA-Asn (x2)[pl]	tRNA-Asp[pl]	tRNA-Cys	tRNA-fMet (x2)	tRNA-Gln
	tRNA-Glu[pl]	tRNA-Gly	tRNA-His[pl]	tRNA-Ile	tRNA-Lys
	tRNA-Met (x2)[pl]	tRNA-Phe	tRNA-Pro	tRNA-Ser	tRNA-Trp[pl]
	tRNA-Tyr (x2)[pl]				

Genes with multiple exons are denoted with the number of exons shown in parenthesis, and trans-spliced genes are indicated with *.tRNAs underlying plastid transferred regions are indicated with [pl]. [a] Note that *rpl16* is annotated as a pseudogene due to an internal stop codon, but this gene has an in-frame GUG present downstream from the ATG start codon, which may be used as a start codon as in other plant species.

3.12. Figures

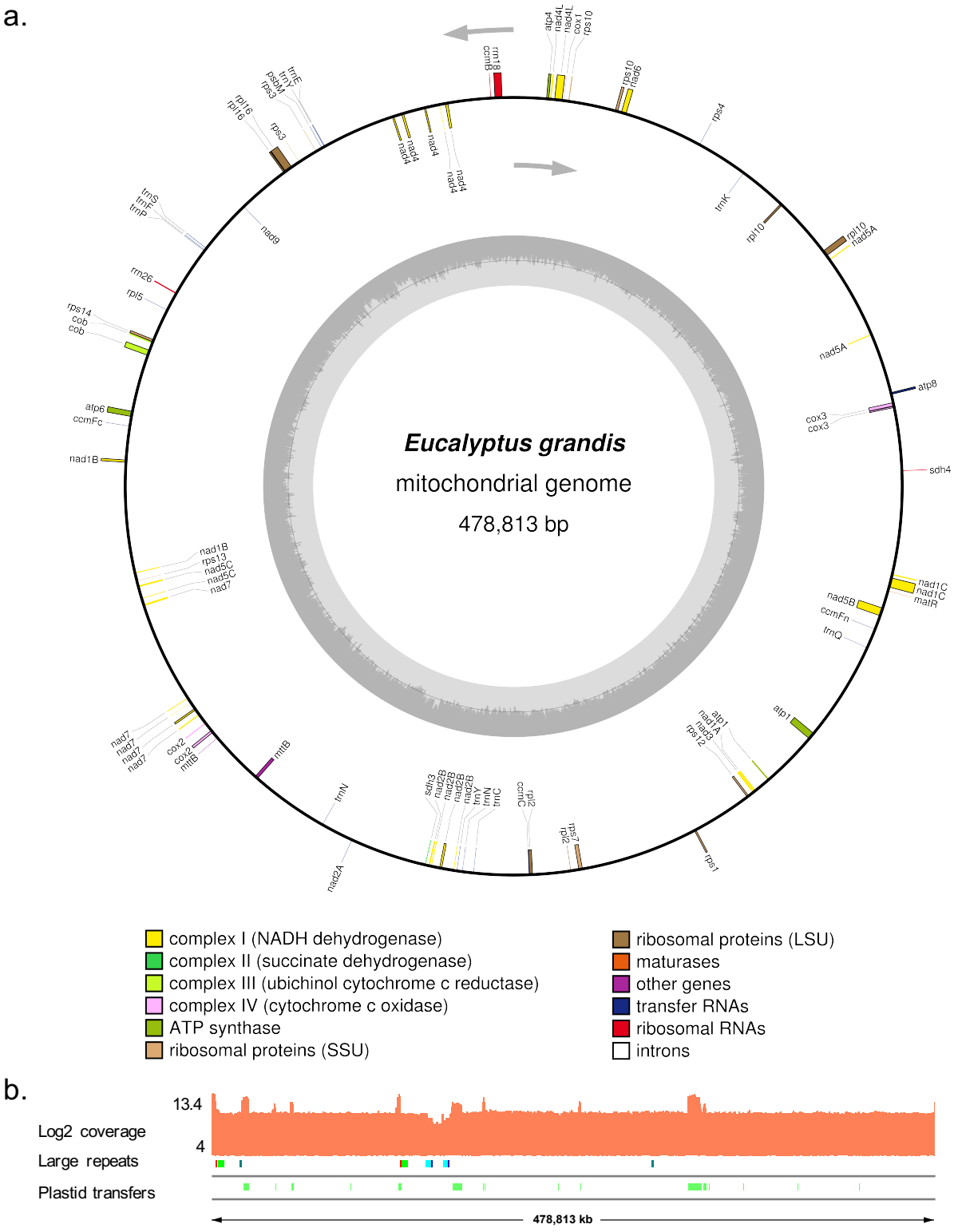


Figure 3.1. Mitochondrial genome of *E. grandis*.

a. Genomic features are shown facing outward (positive strand) and inward (negative strand) of the *E. grandis* mitochondrial genome represented as a circular molecule. The colour key shows the functional class of the mitochondrial genes, and introns are shown in white. The GC content is represented in the innermost circle. The figure was generated in OGDraw [76].

b. Genome coverage of *E. grandis* WGS reads in log2 scale (Log coverage) across the mitochondrial genome. WGS reads were mapped with Bowtie 2 [77] and visualized in IGV [78]. The second track (Large repeats) shows mitochondrial repeat regions > 1 000 bp in length, with pairs in matching colours, all repeat pairs are direct repeats, with the exception of the repeat pair shown in teal. The third track (Plastid transfers) shows plastid to mitochondrial DNA transfers longer than 100 bp, with e-value > $1x10^{-5}$ in green.

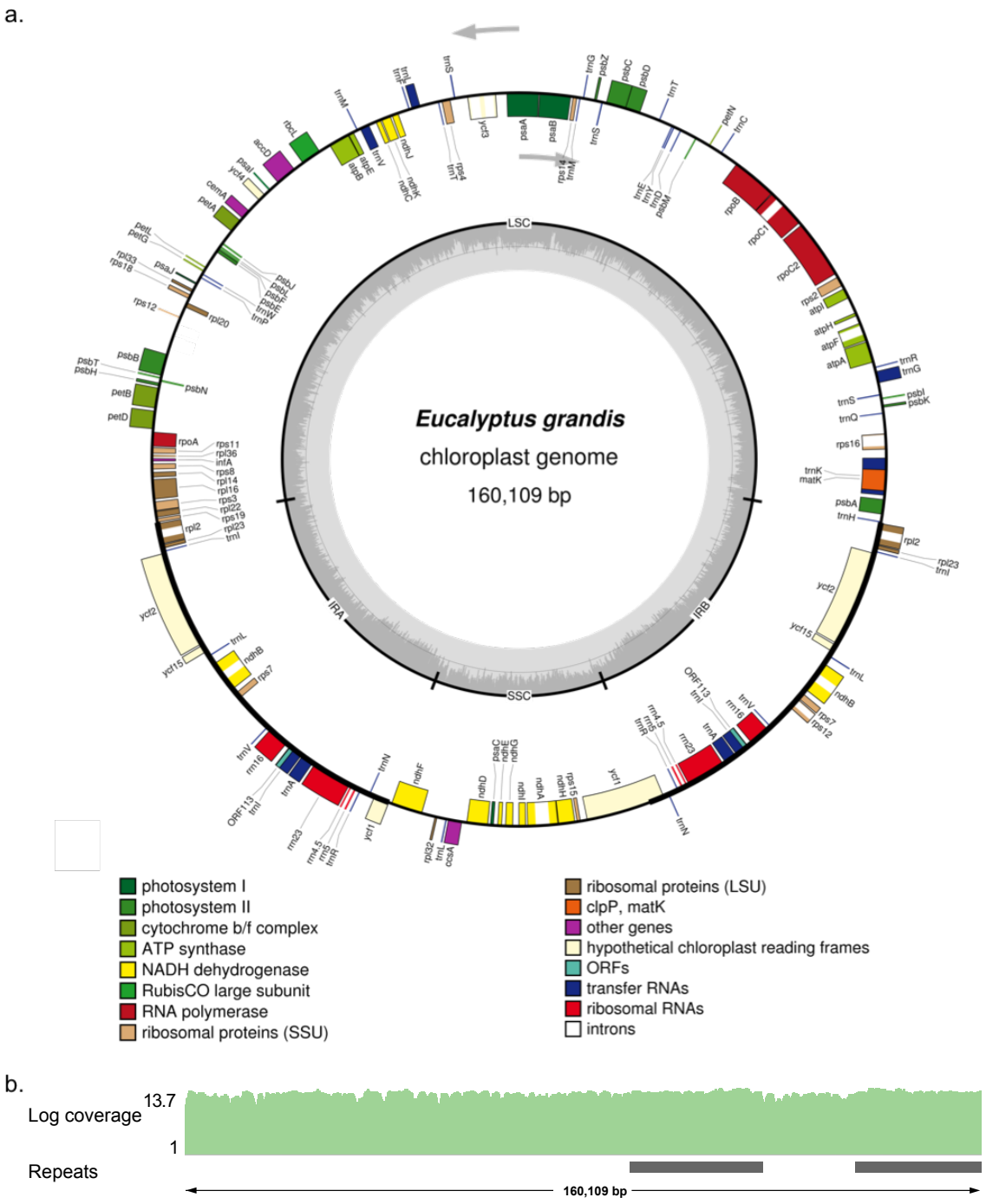


Figure 3.2. Plastid genome of *E. grandis*.

a. Genomic features are shown facing outward (positive strand) and inward (negative strand) of the circular *E. grandis* plastid genome. The colour key shows the functional class of the plastid genes, and introns are shown in white. The GC content is

represented in the innermost circle with the inverted repeat (IR) and single copy (SC) regions indicated. The figure was generated in OGDraw [76].

b. Genome coverage of *E. grandis* WGS reads in log2 scale (Log coverage) across the plastid genome. WGS reads were mapped with Bowtie 2 [77] and visualized in IGV [78]. The position of the plastid inverted repeat regions are shown below (Repeats) in grey.

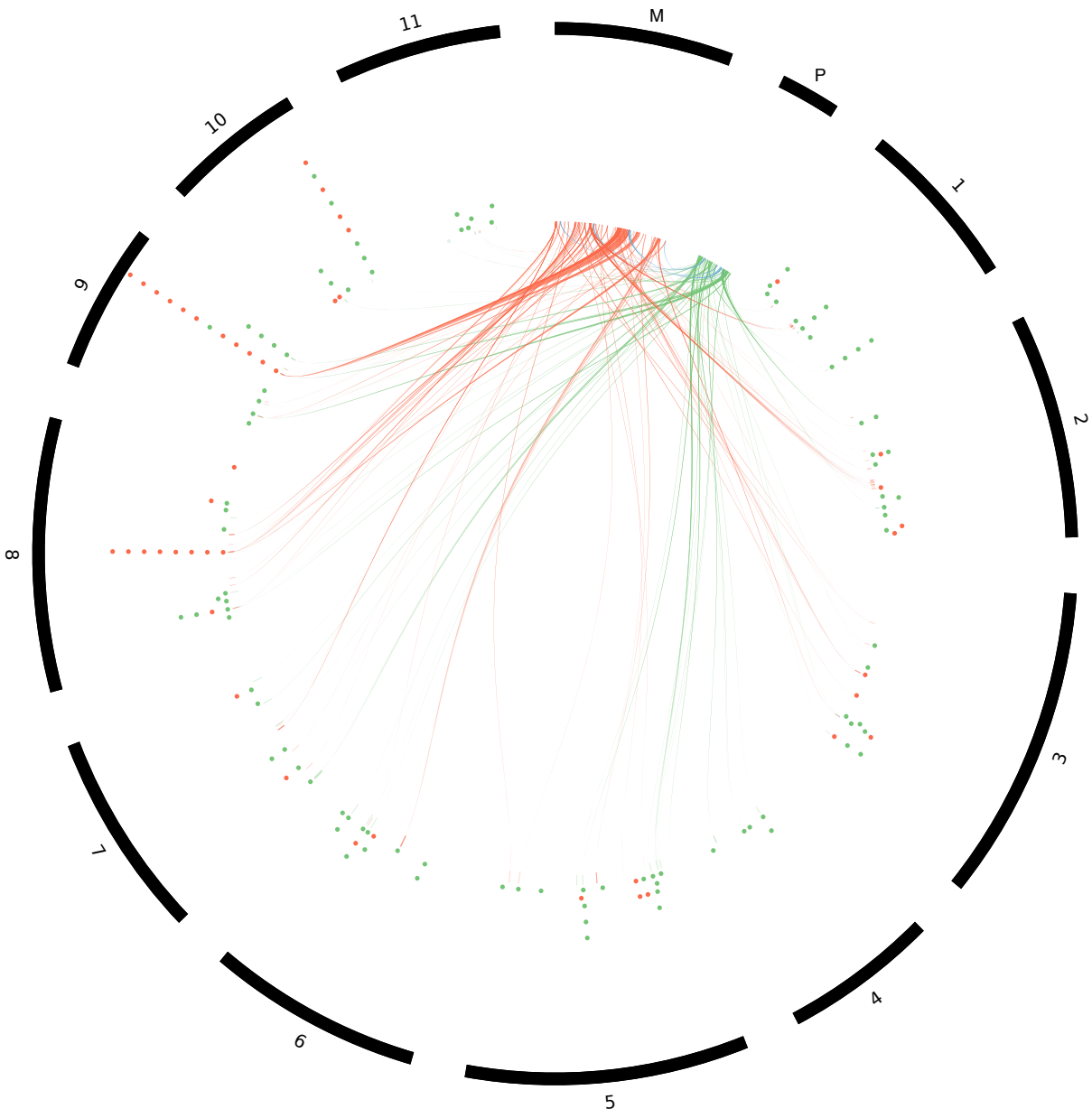
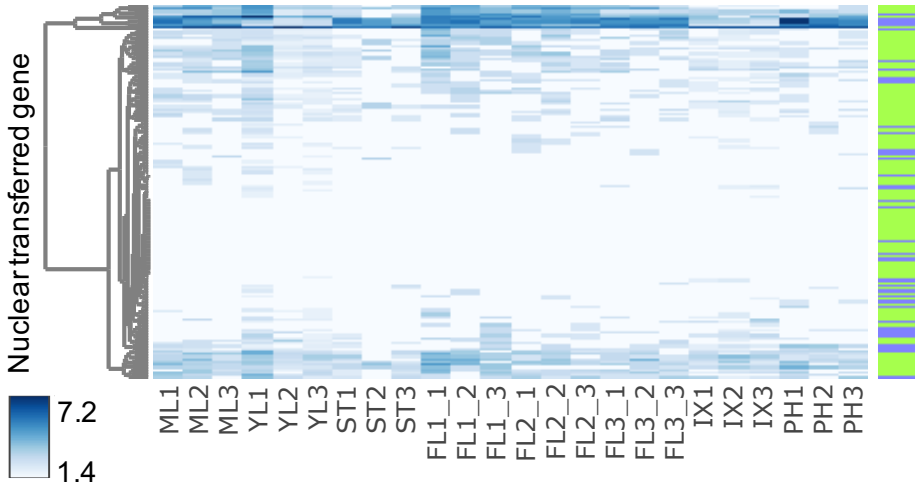


Figure 3.3. DNA and gene transfer between nuclear and organellar genomes in *E. grandis*.

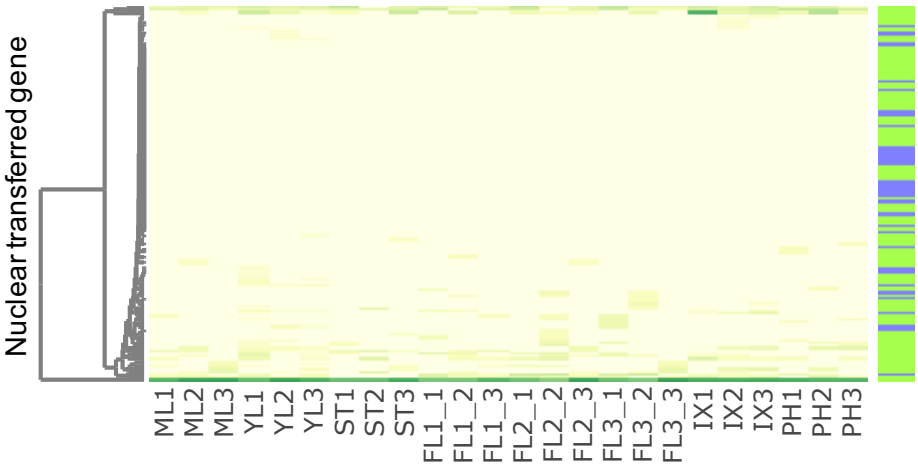
The outer track shows the relevant chromosomes of *E. grandis*, the inner track shows complete coding regions of NUMTs and NUPTs in red and green respectively. The red (mitochondria) and green (plastid) dots indicate full length gene transfers from the

organelles to the nuclear genome. The ribbons represent DNA transfers identified by BLAST analysis greater than 500 nt, with percentage identity greater than 75%. Red ribbons indicate mitochondrial to nuclear DNA transfer, green ribbons indicate plastid to nuclear DNA transfer, and blue ribbons represent plastid to mitochondrial DNA transfer. For clarity, the scale of the plastid and mitochondrial genome size has been increased by 100x.

a.


Reads mapped to nuclear genome only
Nuclear transferred gene
7.2
1.4
ML1
ML2
ML3
YL1
YL2
YL3
ST1
ST2
ST3
FL1_1
FL1_2
FL1_3
FL2_1
FL2_2
FL2_3
FL3_1
FL3_2
FL3_3
IX1
IX2
IX3
PH1
PH2
PH3

b.


Reads mapped to nuclear and organellar genomes
Nuclear transferred gene
ML1
ML2
ML3
YL1
YL2
YL3
ST1
ST2
ST3
FL1_1
FL1_2
FL1_3
FL2_1
FL2_2
FL2_3
FL3_1
FL3_2
FL3_3
IX1
IX2
IX3
PH1
PH2
PH3

c.

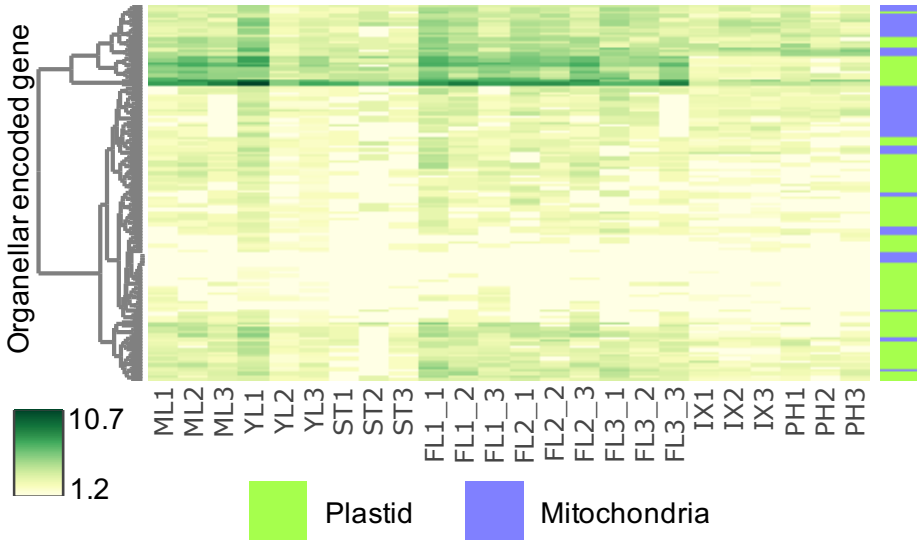

Organellar encoded gene
10.7
1.2
ML1
ML2
ML3
YL1
YL2
YL3
ST1
ST2
ST3
FL1_1
FL1_2
FL1_3
FL2_1
FL2_2
FL2_3
FL3_1
FL3_2
FL3_3
IX1
IX2
IX3
PH1
PH2
PH3
Plastid
Mitochondria

Figure 3.4. Poly-A selected organellar transcriptomes

Poly-A selected RNA read abundance of nuclear genes with homology or annotation suggesting organellar transfer (**a.** and **b.**), and organellar-encoded genes (**c.**) aligned to the nuclear genome only (blue) and the nuclear and organellar genomes of *E. grandis* (green). **a**. Variance stabilizing transformation (VST) counts of 141 organellar transferred genes in the nuclear genome of polyA selected RNA sequencing data aligned to the nuclear genome of *E. grandis* only. **b**. VST counts of full-length transferred genes in the nuclear genome of polyA selected RNA sequencing data aligned to the nuclear and organellar genomes of *E. grandis* simultaneously. **c**. VST counts of organellar-encoded genes of polyA selected RNA sequencing data aligned to the nuclear and organellar genomes of *E. grandis* simultaneously. Row dendrograms on the left-hand side of all three heat maps show clustering of genes based on expression variation between tissues. Tissue samples are shown at the bottom edge of each heatmap, three biological replicates per tissue. Tissues are abbreviated as follows: Mature leaf (ML), young leaf (YL), shoot tips (ST), flowers stage 1 (FL_1), flowers stage 2 (FL_2), flowers stage 3 (FL_3), immature xylem (IX), and phloem (PH). The range of VST count values per heatmap are represented from low (white) to high (blue) for the polyA selected RNA mapping to the nuclear genome only, and from low (yellow) to high (green) for the polyA selected RNA mapping to the nuclear and organellar genomes. The bar on the right of the heatmaps shows the organellar origin of each gene, either plastid (transferred or encoded- green) or mitochondrial (transferred or encoded- blue).

3.13. Additional files



Figure S3.1. Discordantly mapped read pairs flanking direct repeat 13 of the mitochondrial genome.

The insert size of the reads is ~118 000 bp, compared to the expected 475. These reads suggest a repeat mediated structural variation, supported by SVDetect analysis. Read pair insert is shown by the red lines and the direct repeat is shown in the blue track.

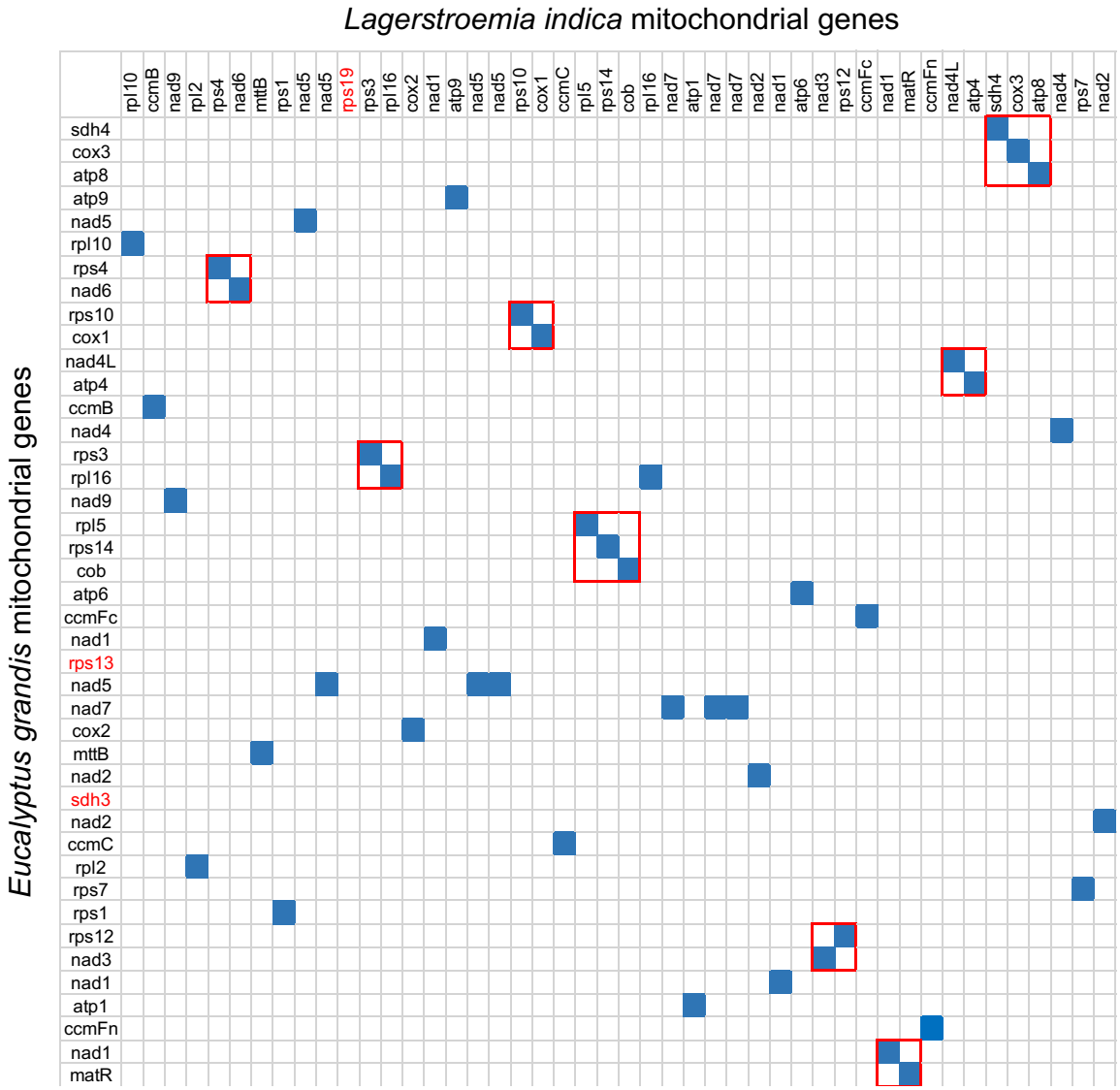


Figure S3.2. Mitochondrial gene order comparison between *Eucalyptus grandis* and *Lagerstroemia indica*.

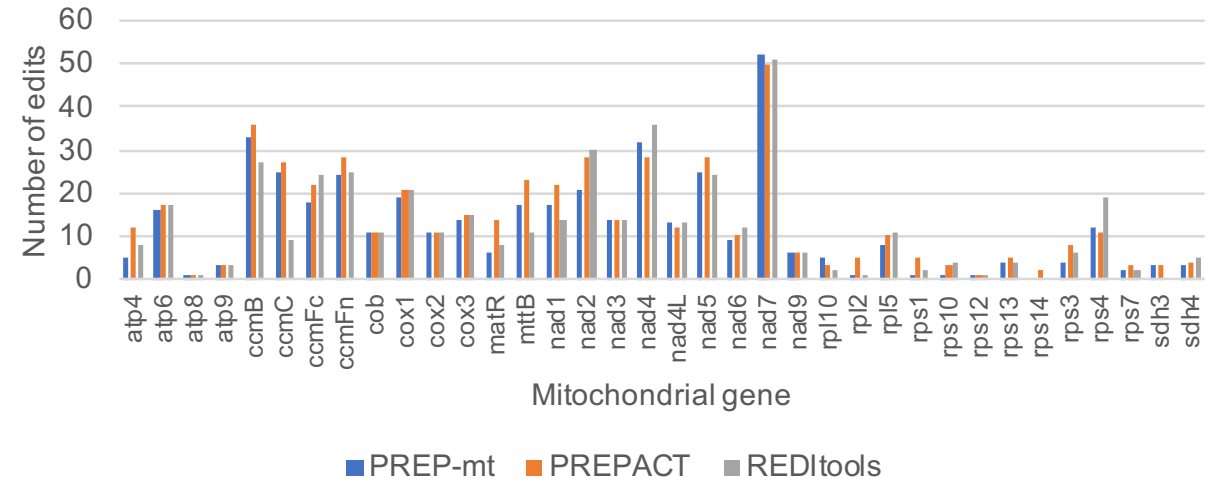
The gene order for the *E. grandis* mitochondrial genome is shown at the right of the figure, and that of *L. indica* on the top. Genes that are not found in each genome are indicated with red text. Collinear genes are indicated by red boxes.



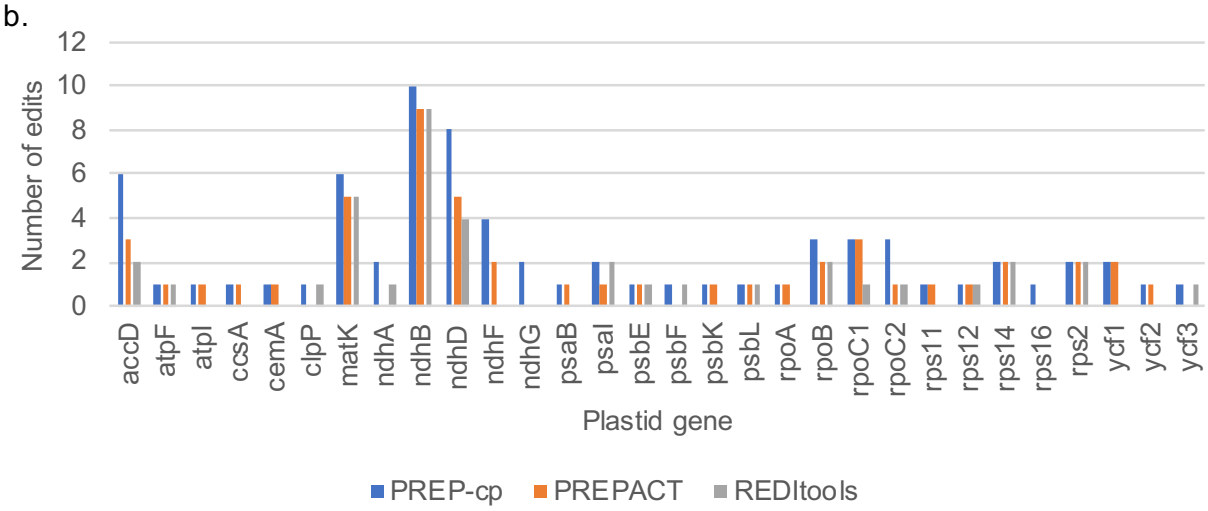
Figure S3.3. Multiple whole genome alignment of selected land plant mitochondrial genomes.

Alignment was performed using the progressiveMauve algorithm in Mauve multiple alignment tool [87], with the coloured blocks representing Locally Collinear Blocks of sequences between genomes. The red lines indicate the length of the mitochondrial genomes, and the name of the organism is shown at the bottom of each genome. This figure shows the widespread genome rearrangements present in plant mitochondrial genomes.

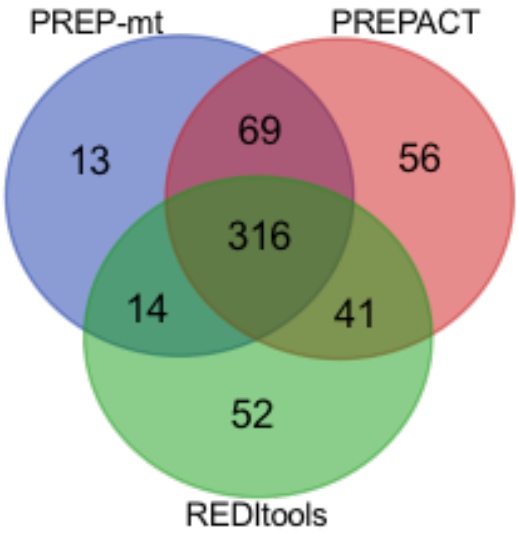
a.


Number of edits
60
50
40
30
20
10
0
atp4 atp6 atp8 atp9 ccmB ccmC ccmFc ccmFn cob cox1 cox2 cox3 matR mttB nad1 nad2 nad3 nad4 nad4L nad5 nad6 nad7 nad9 rpl10 rpl2 rpl5 rps1 rps10 rps12 rps13 rps14 rps3 rps4 rps7 sdh3 sdh4
Mitochondrial gene
PREP-mt
PREPACT
REDItools

b.


Number of edits
12
10
8
6
4
2
0
accD atpF atpI ccsA cemA clpP matK ndhA ndhB ndhD ndhF ndhG psaB psaI psbE psbF psbK psbL rpoA rpoB rpoC1 rpoC2 rps11 rps12 rps14 rps16 rps2 ycf1 ycf2 ycf3
Plastid gene
PREP-cp
PREPACT
REDItools

c.


PREP-mt
PREPACT
REDItools
13
69
56
316
14
41
52

d.

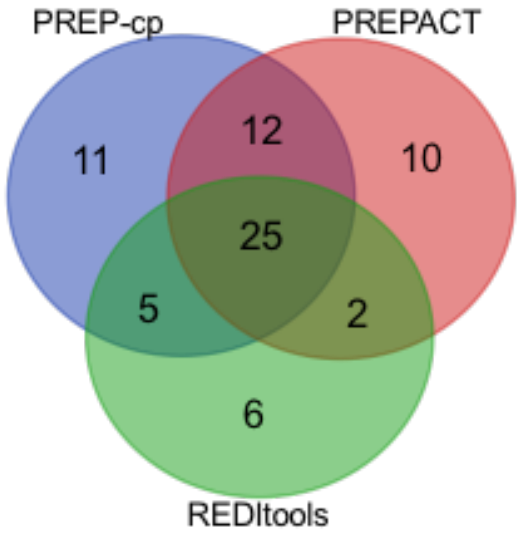

PREP-cp
PREPACT
REDItools
11
12
10
25
5
2
6

Figure S3.4. Number of predicted C to U editing sites in the mitochondrial and plastid genomes of *E. grandis*

a. Number of editing sites (y-axis) in *E. grandis* mitochondrial genes (x-axis) as predicted by PREP-Mt (blue), PREPACT (orange), and evidence from bulked polyA-selected reads from three samples each of eight tissues in *E. grandis* using REDItools (DNA-RNA algorithm: minimum read depth = 10, minimum amount of reads per editing event = 3) shown in grey.

b. Number of editing sites (y-axis) in *E. grandis* plastid genes (x-axis) as predicted by PREP-Cp (blue), PREPACT (orange), and evidence from polyA-selected reads using REDItools (grey). These figures show that bulked polyA selected reads are sufficient to detect the majority of predicted editing events in land plants, however the read depth lower than would be detected with total RNA sequencing.

c. Number of predicted editing sites in common between PREP-Mt, PREPACT, and REDItools in the *E. grandis* mitochondrial genome. d. Number of predicted editing sites in common between PREP-Cp, PREPACT, and REDItools in the *E. grandis* plastid genome.

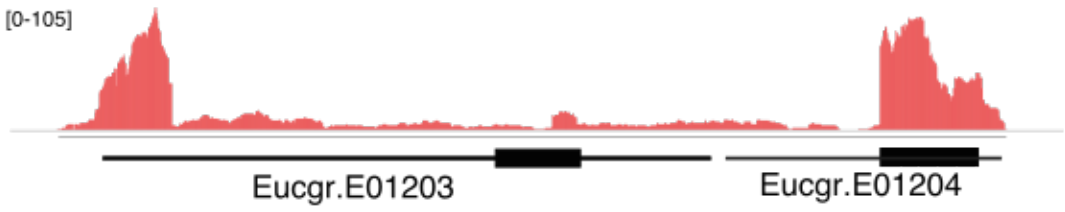


Figure S3.5. Sashimi plot of polyA-selected mRNA reads mapped to *Eucgr.E01203* in *E. grandis* mature leaf tissue.

The plot shows the count of reads (0 to 105) aligned to the annotated gene regions of *Eucgr.E01203.* Reads were aligned using GSNAP and visualized in the Integrated Genome Viewer. Black lines show the annotated gene regions, and thicker black bars show the annotated protein coding regions. The plot shows that the read coverage of *Eucgr.E01203* across the protein coding regions is lower than in the 5' UTR, indicating that the VST counts generated for this gene do not represent functional gene expression.

Table S3.1. *E. grandis* mitochondrial and plastid genome short repeat elements overview

Table S3.2. Amount of *E. grandis* organellar DNA transfer to nuclear chromosomes

Table S3.3. Inter-organellar DNA transfers in *E. grandis* show regions of high homology between *E. grandis* plastid and mitochondrial genomes.

Additional BLAST analysis with land plant organellar genomes show that the transferred regions are all transferred from the plastid to the mitochondria (see Additional file 4).

Table S3.4. Differentially expressed organellar-encoded genes in *E. grandis*.

Negative log2 fold change values indicate increased polyA selected RNA read abundance in mature leaf compared to immature xylem.

Additional file 3.1. Large mitochondrial repeats.

Excel spreadsheet of the results of UniPro UGENE analysis of large (length > 100 bp, identity > 95%) repeats in the *Eucalyptus grandis* mitochondrial genome.

Additional file 3.2. Mitochondrial genome structural variant analysis

Excel spreadsheet of the results of SVDetect analysis results showing breakpoints of structural variants in the *E. grandis* mitochondrial genome (Sheet 1), and breakpoints within 250 bp of mitochondrial genome large repeat regions (Sheet 2).

Additional file 3.3. Predicted and confirmed mitochondrial and plastid transcript editing sites

Excel spreadsheet of the results of predicted editing sites in the *E. grandis* mitochondrial (Sheet 1) and plastid (Sheet 2) genomes using PREPACT, PREP-suite, and REDItools, labelled by position of the edit in the coding sequence.

Additional file 3.4. Inter-organellar DNA transfer

Excel spreadsheet of the origin of inter-organellar DNA transfers, showing the organellar genomes of sequenced land plants, and the results of BLAST analysis of *E. grandis* mitochondrial genome (TAG0014_chr_M) regions that have significant

homology to the *E. grandis* plastid genome. Note that all mitochondrial genomes analysed have no significant matches to the transferred regions, suggesting that there is no transfer of DNA from the mitochondrial genome to the plastid genome.

Additional file 3.5. Transcription of nuclear genes of organellar origin

Excel spreadsheet of nuclear genes of organellar origin for mRNA-sequencing read mapping by homology using BLAST analysis (>80% full length of gene matches with >90% identity to organellar genome) or by annotation (annotation closest match is *Arabidopsis thaliana* organellar gene).

Chapter 4

Transcriptional coordination between nuclear and organellar genomes in carbon source, transport and sink tissues in *Eucalyptus grandis*

D. Pinard[1], D. Behrens[1], A. A. Myburg[1], E. Mizrachi[1]

[1] Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria, 0028, South Africa

E.M. was the lead investigator, conceived of the study, provided scientific input, and edited the chapter. D.B. generated the cosRNA data and edited the chapter. D.P. performed the sample collection, data analysis, generated the figures, and drafted the chapter. A.A.M. was a co- investigator for the study, and edited the chapter. The chapter has been formatted for publication in The New Phytologist.

4.1. Summary

- During wood formation, several essential carbon metabolic pathways take place in the plastids and mitochondria. These pathways are coordinated via spatiotemporal regulation - including subcellular macro and micro-compartmentation - to facilitate the large scale deposition of carbon during secondary xylem growth.
- In wood forming tissues, nuclear and organellar-encoded genes must be transcriptionally and post-transcriptionally regulated to coordinate organellar metabolism. We aimed to explore the role and regulation of plastid and mitochondrial DNA copy number and gene expression during wood formation using small RNA, total RNA, and polyA selected transcriptomic datasets from *Eucalyptus grandis* mature leaf, immature xylem, and phloem tissues.
- Analysis of organellar DNA copy number revealed that mitochondrial genome copy number does not correspond with transcript abundance between tissues, and that plastid genome copy number in immature xylem and phloem is much higher than previous analyses in non-green tissues, and may reflect the transcription of plastid genes by the nuclear-encoded RNA polymerase.
- Furthermore, in xylem and phloem plastids, genes transcribed by the nuclear-encoded RNA polymerase are transcriptionally upregulated to an extent not previously found in non-green plastids. We also found evidence of active upregulation of *ycf2* by as yet unidentified RNA binding proteins, which may be involved in the import of cytoplasmic proteins specific to heterotrophic plastid metabolism.

- Together, these results give novel insight into the regulation of organellar biology during xylogenesis, and suggest that plastid-encoded genes are uniquely regulated in wood forming tissues, adding evidence for a wood-specific plastid - the xyloplast.

4.2. Introduction

Plastid and mitochondrial localized metabolism plays a central role during xylogenesis in plants, and especially in large woody plants, through the supply of energy and phenolic biopolymer precursors. Previously, we have highlighted the widespread co-regulation of carbon metabolic organellar targeted genes with the major processes of xylogenesis, namely primary development, secondary cell wall biopolymer synthesis and deposition, and circadian-regulated metabolic processes (Pinard *et al.*, 2019a - Chapter 2). Mitochondria are known to vary in terms of their proteome, transcriptome, biogenesis, and morphology during development and in response to abiotic factors (Lee *et al.*, 2012; Jacoby *et al.*, 2012; Stone and Storchova, 2015; Welchen *et al.*, 2014). Plastid variation is more complex, as specialized plastid types with distinct functions are found in different tissues (Lopez-Juez & Pyke, 2005; Brillouet *et al.*, 2014). Chloroplasts in mature leaves predominantly sequester atmospheric carbon for conversion to organic carbon (sugars). Sugars are transported via the phloem (transport) cells to the sink tissues such as immature xylem (lignocellulosic sink), where a vast majority of the captured carbon is deposited in secondary cell walls (SCWs). Although plastids and mitochondria in photosynthetic tissues have been well characterized (Tseng *et al.*, 2013; Bramham & Pyke, 2017), the role of plastid and

mitochondrial biology in wood formation remains to be comprehensively explored (Pinard & Mizrachi, 2018).

Transcriptional regulation of plastid and mitochondrial encoded genes is regulated by nuclear-encoded genes, which in turn are regulated by circadian rhythm, biotic and abiotic stimuli, and tissue-type or developmental stage (Welchen *et al.*, 2014; Sakamoto *et al.*, 2008). Organellar size, number, genome copy number, and morphology vary between tissue and cell types according to cellular metabolic activities (Liere & Börner, 2013; Hanson & Sattarzadeh, 2014; Beltrán *et al.*, 2018). In *Arabidopsis* meristems, ~10 proplastids are present per cell, having ~10 nucleoids per proplastid, while leaf mesophyll cells contain ~100 chloroplasts with > 50 copies of plastid DNA per cell (Zoschke *et al.*, 2007). Similarly, the number of mitochondria and their genome copy number per cell are highly variable in plant cells (Welchen *et al.*, 2014). Not much is known about the determinants of tissue-specific differential regulation of mitochondrial encoded transcripts, as promoter usage is relaxed (Gualberto & Kühn, 2014), and mitochondrial gene expression is not accurately quantified in typical transcriptome profiling assays.

Plastid gene expression is differentially regulated between tissues according to their metabolic activities/type (e.g. chloroplasts vs. amyloplasts) via the interaction of nuclear and plastid-encoded transcription factors (Barkan, 2011; Zhang *et al.*, 2012; Tseng *et al.*, 2013). An overarching theme to plastid and mitochondrial dynamics (gene expression, type, number, protein/RNA/metabolite import), is the regulation by nuclear-encoded genes and inter-organellar coordination via retrograde signalling

(Hedtke *et al.*, 1999; Woodson & Chory, 2008; Welchen *et al.*, 2014; Singh *et al.*, 2015; Woo *et al.*, 2016). Nuclear-encoded RNA binding proteins (RNAbps) protect organellar transcripts against exo- and endonucleases, leaving behind small RNA footprints indicative of nuclear regulation, termed clustered organellar sRNAs (cosRNAs) (Ruwe *et al.*, 2016). cosRNAs may also regulate organellar transcript abundance by competition for RNAbps, or be generated in the antisense direction of genes derived from double-stranded RNA molecules (Ruwe *et al.*, 2016; Cavaiuolo *et al.*, 2017). Identification and quantification of organellar cosRNAs provide insight into the regulation of steady-state transcript levels by nuclear-encoded regulators.

A fundamental characteristic of plastid-encoded gene expression is the gene-specific action of two distinct polymerases, the plastid-encoded polymerase (PEP), and the nuclear-encoded polymerase (NEP) (Lysenko & Kuznetsov, 2005). Most plastid-encoded genes contain promoters for both types of polymerase, although in general, the so-called “housekeeping” plastid genes are predominantly transcribed by NEP, and photosystem genes are transcribed by PEP (Liebers *et al.*, 2017). In non-green plastids, it is assumed that gene-specific expression regulation is mainly passive, with the basal transcription of housekeeping genes, and down-regulation of photosystem genes due to lack of activating transcription factors that are normally expressed in response to, or in anticipation of, light in green tissues (Zhang *et al.*, 2012). A known exception is plastid-encoded *accD* in developing tomato fruits, which is significantly upregulated during the chloroplast to chromoplast transition (Kahlau & Bock, 2008). In the chloroplast to chromoplast transition, plastid housekeeping gene transcripts were less abundant in tomato fruits, although the repression was less pronounced than in

potato root tubers (Valkov *et al.*, 2009). These findings led to the hypothesis that increased expression of plastid-encoded *accD* may affect other plastid-encoded genes via the non-specific action of NEP (Kahlau & Bock, 2008), which has not been explored in other plastid types.

To date, no comprehensive study of organellar gene expression and cytonuclear coordination has been conducted in the secondary xylem forming tissues of vascular plants. This represents a major gap in our understanding of organellar biology and xylogenesis. Here, we aimed to characterize the transcriptional regulation of organelles in carbon source, sink, transport, and secondary xylem tissues of *Eucalyptus grandis*. Using ribosomal RNA depleted (total) mRNA-seq data we surveyed organellar-encoded genes in *E. grandis* mature leaf (source), phloem (transport), and immature xylem (sink) tissues, and found that plastid-encoded non-photosynthetic genes in xylem and phloem are specifically upregulated to levels not seen in other heterotrophic tissues. Small-RNA sequencing data identified signatures of wood specific regulation by as yet unknown nuclear-encoded RNA binding proteins. This system-wide analysis suggests that plastid-encoded genes may be actively regulated in secondary growth and wood formation through protein import, and that novel and important xylem-specific regulators of plastid biology remain to be discovered. Transcriptomic analysis of plastid-encoded and -targeted genes suggests that xylem plastid biology is distinctly regulated to facilitate carbon partitioning towards secondary cell wall biosynthesis, increasing the evidence for a wood-specific plastid, the xyloplast (Pinard & Mizrachi, 2018).

4.3. Methods

4.3.1. Tissue sample preparation and sequencing

Immature xylem, phloem, and mature leaf tissues were collected from three *E. grandis* TAG0014 clonal ramets (provided by Mondi South Africa) in KwaMbonambi, Kwazulu Natal, South Africa, in April 2014 between 08:00 and 14:00 as previously described in Mizrachi *et al*. (2010). The samples were immediately frozen in liquid nitrogen and kept at -80°C until sample preparation for RNA sequencing at the Beijing Genomics Institute (BGI Beijing). After RNA extraction, the samples were sequenced on an Illumina HiSeq2000 sequencer according to three different protocols : (a.) poly-A selection of mRNA transcripts, (b.) RiboZero RNA depletion for total RNA sequencing, and (c.) size selection for small RNA sequencing. For the poly-A selected and total RNA sequencing, Illumina TruSeq strand-specific library preparation and sequencing were performed. Standard quality control, and read trimming was performed by BGI.

4.3.2. Organellar genome copy number estimation

Using a QuantStudio OpenArray (ThermoFisher Scientific, Waltham, USA) assay developed for the organellar genomes of *E. grandis*, we estimated the genome copy number of organellar genomes from total DNA extracted from the three tissues analyzed. This was done by qPCR and delta-CT value analysis of organellar genes compared to nuclear gene copy number (Schmittgen & Livak, 2008; Supplementary Method 1)

4.3.3. Total, polyA, and small RNA transcriptome analysis

The total RNA (RiboZero) sequencing reads were aligned to the nuclear (v2.0) and organellar genomes (Myburg *et al.*, 2014; Bartholomé *et al.*, 2015; Pinard *et al.*, 2019b - Chapter 3) using the GSNAP alignment tool (gmap version 2016-09-23) (Wu *et al.*, 2016). GSNAP alignment was done using a file specifying predicted editing sites in both organellar genomes as SNPs to prevent preferential mapping to the unedited nuclear copies of organellar genes (Pinard *et al.*, 2019b - Chapter 3), and a splice sites file generated per the gmap user guide. Additional GSNAP parameters used were as follows: quality-protocol=illumina and all other parameters were kept at default. For the plastid genome, inverted repeat region A was hard masked using bedtools maskFastaFromBed command before read alignment to correct for duplicate read alignment to these regions (bedtools version 2.27.0) (Quinlan, 2014). This was done as IRA contains a truncated copy of *ycf1* (Pinard *et al.*, 2019b - Chapter 3), and it is not yet known if transcripts are derived from each repeat equally. As the aim of the study was to compare transcript abundance *between* tissues, and all reads were aligned with the same parameters between samples, the higher read counts for these genes will be treated equally across tissues. The number of reads per chromosome was assessed with samtools flagstat and idxstat function (v1.8) (Li *et al.*, 2009). Reads mapped to annotated exons were then counted and normalized using HTSeq count (HTSeq version 0.8.0, python version 2.7), with mode = union strandedness = reverse (Anders *et al.*, 2015). Differentially expressed genes between tissues were detected using DESeq2 in R (R version 3.4.4, DESeq2 version 1.18.1) (Love *et al.*, 2014). The polyA RNA sequencing reads were aligned to the organellar and nuclear genomes and quantified as above for the total mRNA sequencing reads. For the nuclear-

encoded genes, polyA RNA sequencing values were used, and the annotation and localization of genes were taken from Pinard *et al.* (2019a - Chapter 2).

For read coverage analysis across the organellar genomes, bam files of mapped paired-end reads were separated by strand using deepTools bamCoverage (Ramírez *et al.*, 2014). Coverage of poly-A and total RNA sequencing reads from the strand-separated bam files was then visualized using IGV (Thorvaldsdóttir *et al.*, 2013). The strand-separated bam files were converted to bed format, and bedtools intersect was used with gff files of the organellar genomes (one edited to reverse strand) to count the average read depth per feature in the sense and antisense directions using a custom shell script (Quinlan, 2014).

Small RNA reads were mapped to the organellar genomes using bowtie v1.2, with mismatch set to 1, and all highest scoring hits reported (-a --best --strata -v 2) (Langmead & Salzberg, 2012). The resulting samfiles were converted to bam format, sorted, indexed, and merged using samtools v1.8 (Li *et al.*, 2009). Read groups were added to the resulting bam file for sRNA miner analysis. The sRNA miner cosRNA detection pipeline was used to identify cosRNAs in the plastid and mitochondrial genomes, with all parameters set to default (read count 40 and end sharpness 0.75, 5' or 3' ends) (Ruwe *et al.*, 2016). Bed files of cosRNA were exported and edited for count analysis using ShortStack v3.8.5 (Axtell, 2013). ShortStack default parameters were changed to account for the size difference between cosRNA and dicer processed nuclear microRNAs, with --dicermin 18 --dicermax 50 and hairpin test option deactivated with --nohp. BLAST+ analysis (Altschul *et al.*, 1997) was used to identify

cosRNAs which may originate from the nuclear genome or match significantly to previously identified nuclear microRNAs (Behrens *et al.*, personal communication). To identify which organelle genomic features overlap with the identified cosRNAs, bedtools closest was used (Quinlan, 2014). Note that bedtools closest assigns any overlap of more than a nucleotide as a complete overlap (distance = 0). Similar to the coverage analysis, two gff files, one with the strand information reversed, were used to assign cosRNA to the sense or antisense strand of a feature. Differential abundance of cosRNAs between tissues was determined using DESeq2 in R from the count data reported by ShortStack (R version 3.4.4, DESeq2 version 1.18.1) (Axtell, 2013, Love *et al.*, 2014; Ruwe *et al.*, 2016).

4.3.4. Organellar transcript editing analysis

To detect the editing of organellar transcripts, the REDO package was implemented using the aligned total RNA seq bam files per tissue for each organelle separately (Wu *et al.*, 2018). First, samtools mpileup (v1.8) and bcftools (v1.9) (Li *et al.*, 2009; Danecek *et al.*, 2011) were used to call alignment variants from the total RNA seq bam files and previously aligned bam files from whole genome sequencing reads (Pinard *et al.*, 2019b - Chapter 3). The output RNA seq vcf files were then used as input for the REDO pipeline with the organellar genomes in fasta format and annotations in the tbl format provided. The parameters used to call edits were minimum read depth = 100, and the minimum proportion of edited reads = 0.1. Potential DNA variants were filtered from the output if the site was also detected in the vcf file from the DNA

sequence alignment using custom shell scripts. REDO outputs sites per feature, the proportion of reads edited, and read depth, which was used to visualize the results.

4.4. Results

4.4.1. Plastid and mitochondrial genome copy numbers vary between carbon source, sink, and transport tissues, but are not directly related to organellar transcription

To determine the relative copy numbers of organellar genomes in *E. grandis*, TAQman qPCR assays of plastid and mitochondrial protein-coding genes were used (Supplementary Method **4.1**). These assays were specifically designed as a panel to cover all possible organellar encoded genes and avoid regions of homology to the nuclear genome and include technical replicates (Supplementary Method **4.1**). Single-copy nuclear control genes were used to estimate the relative organellar genome copy number using the delta CT method (Schmittgen & Livak, 2008), as the diploid nuclear genome of *E. grandis* does not undergo somatic endopolyploidization (Grattapaglia & Bradshaw, 1994). The results showed that mitochondrial genome copy number was significantly higher in immature xylem (~260 copies/cell) and phloem (~280 copies/cell) compared to mature leaf (~51 copies/cell) (Figure **4.1a**). Due to inter-assay variation, and the fact that whole tissues were sampled, not single cell types, structural variants in the mitochondrial genome could not be identified (Supplementary Method 1) (Preuten *et al.*, 2010). Although plastid genome copy numbers were significantly higher in mature leaf (~407 copies/cell) compared immature xylem (~332

copies/cell) and phloem (~345 copies/cell) (Figure **4.1a**), the difference was lower than expected from literature for photosynthetic vs non-photosynthetic tissues (Valkov *et al.*, 2009).

Additional data from one sample each of *Eucalyptus grandis* TAG0014 young leaf and developing flower tissues were also analyzed to determine if the differences in genome copy number between heterotrophic and phototrophic tissues were due to developmental stages. Young leaf had an average mitochondrial genome copy number of ~65 copy numbers per cell relative to the diploid nuclear genome, and developing flowers had ~70 mitochondrial genome copies per cell (Figure **S4.1**). Plastid genome copy number in young leaf was similar to that in immature xylem and phloem at ~358 copies per cell, while in developing flowers, plastid genome copy number was low compared to the other tissues analyzed at ~144 copies per cell (Figure **S4.1**). These results suggest that tissue specific biology, rather than developmental stage, is mainly responsible for the relatively high organellar genome copy numbers in woody tissues.

Post-transcriptional editing of organellar transcripts and organellar DNA transfer to the nuclear genome poses unique challenges for read alignment in plants, as edited reads create mismatches between the transcript and organellar DNA sequences. Therefore, the GSNAP SNP-aware alignment method (Wu *et al.*, 2016) was used to accurately account for organellar transcript editing-related read alignment bias and inter-genomic DNA transfer. Ribosomal RNA depleted total-RNA and polyA-selected reads from mature leaf, immature xylem, and phloem were aligned to the nuclear and organellar

genomes of *E. grandis* simultaneously using a "SNP" file of predicted organellar transcript editing sites (Table **S4.1**) (Pinard *et al.*, 2019b - Chapter 3).

The aligned reads show that both strands of the plastid and mitochondrial genomes were transcribed in all three tissues (Figure **4.1c**), and average read coverage was highest in sense regions of protein-coding genes and introns (Figure **S4.2**), which are standard features for organellar transcriptomes (Castandet *et al.*, 2016; Sanitá Lima & Smith, 2017). Between 5-7% of the total mapped reads aligned to the mitochondrial genome in phloem and developing secondary xylem tissue, compared to 3% in mature leaf (Figure **4.1b**). The transcription of the mitochondrial genome was not directly proportional to genome copy number in immature xylem and phloem, with ~5x genome copy number *vs* ~2x reads mapped compared to mature leaf. In mature leaf, an average of 38% of total RNA-seq reads aligned to the plastid genome, compared to 1.5% in immature xylem, and 2% in phloem, although it should be noted that transcripts from a single gene (*psbA*) accounted for ~35% of plastid reads in mature leaf (Figure **4.1b**, File **S4.1**). The relatively high plastid genome copy number in immature xylem and phloem was not proportional to transcription, due to the massive upregulation of photosystem genes that is characteristic of chloroplasts (Kahlau & Bock, 2008; Valkov *et al.*, 2009; Zhelyazkova *et al.*, 2012). Together, the organellar genome copy number analysis and transcriptome mapping show that in *E. grandis* immature xylem and phloem, genome copy number is not directly related to whole-genome transcription level.

Previously, we have shown that a small number of polyadenylated reads align to the organellar genomes of *E. grandis*, and that photosystem genes were differentially polyadenylated in green tissues compared to immature xylem and phloem (Pinard *et al.*, 2019b - Chapter 3). To confirm if polyadenylation is a mechanism of tissue-specific degradation of organellar-encoded transcripts (Lisitsky & Schuster, 1999; Levy & Schuster, 2016), the proportion of polyA captured reads to ribosomal rRNA depleted reads aligned to the organellar genomes were compared. In general, the polyadenylation percentage of plastid and mitochondrial genes was very low, for an average of 0.29% in immature xylem, 0.35% in phloem, and 0.31% in mature leaf (Figure **S4.3**, File **S4.5**). The highest percentage of polyadenylated reads was found for *rps12* in phloem, at 4.9% of total reads polyadenylated. This suggests that polyadenylation is not a widespread mechanism for tissue-specific organellar gene regulation in *E. grandis*, and may rather be mitigating the effects of imprecise transcription termination (Lisitsky & Schuster, 1999), as polyadenylation is positively correlated to transcript abundance, where highly expressed genes were more polyadenylated than lowly expressed genes (Figure **S4.4**). Alternatively, polyadenylation may act rapidly to degrade organellar transcripts, in which case polyA-selection transcriptome sequencing would not accurately detect polyadenylation as a mechanism of post-transcriptional regulation in organelles.

4.4.2. Mitochondrial genes are ubiquitously expressed, while plastid genes are differentially regulated across tissues

In the few examples of tissue specific expression of mitochondrial genes in plants, it has been found that transcript levels of individual genes differ within a tissue, and reproductive tissues tend to have the highest expression levels (Fujii *et al.*, 2011; Fang *et al.*, 2012; Hameed *et al.*, 2017). In *E.* grandis, mitochondrial gene expression varied per gene, where VST counts ranged from 11 to 17 (File **S4.2b**). Between tissues, mitochondrial gene expression was relatively consistent (Figure **4.2b**), with *rps7* the only gene with L2FC > 2 in immature xylem compared to mature leaf (File **S4.3**). Intergenic regions of plant mitochondria have previously been shown to have tissue specific transcription (Fujii *et al.*, 2011), which may account for the ubiquitous expression of individual mitochondrial genes despite the higher overall proportion of reads mapped in immature xylem and phloem compared to mature leaf.

As expected for plastids, photosystem genes were highly expressed in mature leaf, and 13 genes were significantly upregulated (log2 fold change > 2 in mature leaf) (File **S4.1**). Of all plastid-encoded genes, *psbA* was the highest expressed, with VST count of 22 in mature leaf, while *petN* was the lowest expressed gene, with VST of 2.7 in some samples (File **S4.1**). Plastid gene expression in xylem and phloem were very similar, with no differentially expressed genes between them (Figure **4.2a**, File **S4.1**). In *E. grandis* immature xylem and phloem plastid housekeeping genes were up-regulated compared to mature leaf (Figure **4.2a**, File **S4.1**). Specifically, 13 out of 75 single copy plastid genes (genes in Inverted Repeat Region A hard masked before alignment) had a statistically significant log2 fold change > 2 in immature xylem

compared to mature leaf (File **S3**). Surprisingly, *ycf2* was the highest up-regulated gene in immature xylem and phloem, with log2 fold change > 4 compared to mature leaf (File **S4.1**). *ycf2* encodes an essential protein recently associated with the plastid inner membrane protein transporter complex (TIC) (Drescher *et al.*, 2000; Kikuchi *et al.*, 2018). Additionally, *rpo* operon genes, *rpl* subunit genes, and another component of the TIC complex, *ycf1* were also highly upregulated in immature xylem and phloem (File **S4.1**). The upregulation of housekeeping and *ycf* genes in immature xylem and phloem show that NEP is the dominant polymerase used in woody tissues.

To date, no previous research has quantified plastid gene expression in secondary xylem and phloem or examined phototrophic versus heterotrophic plastid gene expression in woody plants. However, compared to available array based RT-qPCR data for tomato fruit chromoplasts and potato tuber amyloplasts *vs* leaf chloroplasts show that *E. grandis* housekeeping and "other" genes (*ycf1, ycf2*, and *accD*) were upregulated in immature xylem to an extent not seen in chromoplasts or amyloplasts (Figure **4.2c**) (Kahlau & Bock, 2008; Valkov *et al.*, 2009). These results indicate that secondary xylem and phloem have a unique regulatory program for plastid gene expression, where the upregulation of *ycf1* and *ycf2* may play a specific role in non-photosynthetic metabolism not related to lipid metabolism (chromoplasts) or starch metabolism (amyloplasts).

4.4.3. Post-transcriptional editing in *E. grandis* plastid and mitochondrial genomes

In land plant organelles, post-transcriptional C to U editing is a mechanism of transcriptional regulation that may act in a tissue-specific manner to regulate organellar metabolism (Picardi *et al.*, 2010; Tseng *et al.*, 2013). To determine the degree of post-transcriptional editing in *E. grandis* organelles, the aligned total RNA-seq reads were used to quantify editing in the plastid and mitochondrial genomes of mature leaf, immature xylem, and phloem. After variant calling with samtools mpileup and vcftools, edits were analyzed and annotated using the REDO package (Li *et al.*, 2009; Danecek *et al.*, 2011; Wu *et al.*, 2018). In total, 57 and 471 editing sites were identified in plastid and mitochondrial protein coding regions, respectively (File **S4.4**, Figure **4.3**, Figure **S4.5** and **S4.6**). Editing sites were retained for downstream analysis if they were edited in all three samples in at least one tissue and resulted in non-synonymous amino acid changes in protein-coding genes (Figure **4.3**). In Figure **4.3** and Figure **S4.6**, the average proportion of edited reads per site in three samples per tissue are shown, with serine to leucine the most commonly identified amino acid change (17 in the plastid and 114 in the mitochondria). The editing of serine to leucine is a feature of organellar coding region transcripts, acting to increase the hydrophobicity of the protein interaction interface (Wodak & Janin, 2002; Yura & Go, 2008; Kaila *et al.*, 2019). Mitochondrial editing percentage was variable between samples in immature xylem and phloem (Figure **S4.5c**, Figure **S4.6b**), especially in samples IX2 and PH3, and therefore an average of editing percentage across all three tissues was used for further analysis.

Immature xylem and phloem editing sites were edited at roughly the same average frequency in both plastids and mitochondria (within 10%), while many mature leaf editing sites differ from the woody tissues (Figure **4.3**, File **S4.4**). The creation of canonical start codons by C to U editing is critical to regulating the number of functional proteins produced per tissue. In the plastid genome, the start codon of *psbL* was completely unedited in xylem and phloem, and 75% edited in mature leaf, as previously seen in spinach (File **S4.4**) (Bock *et al.*, 1993). Further, in the mitochondrial genome, the start codons of *rps10* (~50% vs ~80%), *nad4L* (~50% vs ~90%), and *rps13* (~45% vs ~70%) were alternatively edited in woody tissues compared to mature leaf (File **S4.2**). Therefore, start codon editing is a regulatory mechanism that tailors both plastid and mitochondrial gene expression to tissue-specific requirements in non-photosynthetic tissues.

In the *E. grandis* mitochondrial genome, editing was more variable between tissues, with 58% of the 71 editing sites preferentially edited in mature leaf compared to xylem and phloem (>50% difference in editing percentage), while three editing sites in the cytochrome c biosynthetic genes *ccmC*, *ccmFc*, and *ccmFn* were preferentially edited in xylem and phloem (>20% difference in editing percentage) (File **S4.4**, Figure **S4.6a**). Given the ubiquitous expression of mitochondrial genes between tissues, it was surprising to identify such a large number of sites that were preferentially edited in the mitochondrial genome (Figure **S4.6a**). Alignment visualization across the region of *matR*, which was almost completely unedited in immature xylem and phloem, shows that the transcripts were edited below the threshold specified in the REDO analysis (> 10% of reads edited) (Figure **S4.6b**). Genes that were preferentially edited in mature

leaf compared to xylem and phloem include cytochrome oxidase complex maturation and biosynthetic genes *ccmC*, *ccmFn*, *ccmB*, *ccmFc*, and NADH dehydrogenase subunit *nad3*. These mitochondrial transcripts have a positive relationship between editing percentage and translation compared to others (Planchard *et al*, 2018). Therefore, the increased editing of mitochondrial transcripts in leaf compared to xylem and phloem suggests that mitochondrial transcript editing may be functionally coordinated with plastids in leaf for photosynthesis.

In other plastid genes, the most significant differences in editing were seen in the photosystem genes *psbF* and *psbE*, ribosomal protein subunit *rps2*, PEP subunits *rpoC1*, and *rpoB*, and NADH dehydrogenase subunit genes *ndhA*, *ndhB*, and *ndhD*, which were unedited in immature xylem and phloem (Figure **4.3**, File **S4.4**). The first exon of *ndhB* was edited to a similar extent in all three tissues, while a single site in the second exon was edited to a much lower degree in xylem and phloem (Figure **4.3**). Previous research has shown that differential splicing and editing of *ndhB* in non-photosynthetic tissues that experience relatively low levels of photo-oxidation may allow for non-functional NDH complex proteins (Tseng *et al.*, 2013). Significantly, the reduced editing of PEP subunits *rpoC1* and *rpoB* may act to functionally decrease the protein complex level of these NEP expressed genes, which are upregulated in immature xylem and phloem by the non-specific action of NEP, to reduce photosystem gene expression.

4.4.4. Tissue specific small RNA footprints highlight nuclear-encoded RNA binding protein regulation of organellar transcriptomes

The release of sRNA-miner, a cosRNA identification pipeline allowed for the identification of small RNA footprints in the organellar genomes of *E. grandis* using small RNA sequencing data (Ruwe *et al.*, 2016). In total, 444 plastid cosRNAs, and 419 mitochondrial cosRNAs were identified across all nine samples analyzed (Figure **4.4a**, Figure **S4.7**, File **S4.6**), almost double the amount previously reported in *Arabidopsis thaliana* leaf tissue (229 plastid and 222 mitochondrial) (Ruwe *et al.*, 2016). Small RNAs may be targeted to/originate from organellar DNA that has been transferred to the nuclear genome (Ruwe *et al.*, 2016), and potential nuclear small RNAs were identified by BLAST analysis against the *E. grandis* nuclear genome and nuclear small RNAs (Behrens *et al.*, personal communication). In total, 43 unique plastid cosRNAs and 20 mitochondrial cosRNAs had significant homology to the nuclear genome. Of these, 21 unique plastid cosRNAs and eight mitochondrial cosRNAs had significant homology to nuclear small RNAs, while 42 and 20 unique cosRNA were homologous to 169 and 140 regions of the *E. grandis* genome respectively (Behrens *et al.*, personal communication, File **S4.6**). In contrast, ~45% of *Arabidopsis* cosRNAs mapped equally well to the nuclear genome, as *Arabidopsis* chromosome 2 has a large, recent insertion of the mitochondrial genome (Michalovova *et al.*, 2013; Ruwe *et al.*, 2016). In *E. grandis*, mitochondrial DNA insertions are smaller and more interspersed and therefore may not need nuclear siRNAs to maintain a heterochromatic state (Voinnet, 2009; Pinard *et al.*, 2019b - Chapter 3). In plastids, 99 cosRNAs are found to overlap genes in the sense direction, and 94 were found in intergenic regions (Figure **S4.7a**). In mitochondria, almost half of the identified

cosRNAs were found in intergenic regions (193), with 29 found in the sense orientation of coding regions (Figure **S4.7a**). Given the wide range of functions of RNAbps in organelles, the identified genic cosRNA may indicate 5' end maturation, 3' end protection, or sites of post-transcriptional editing.

Several organelle-specific cosRNAs are significantly differentially expressed between tissues, the majority being in higher abundance in mature leaf compared to immature xylem and phloem (Figure **4.4a** and Figure **S4.7b**, File **S4.3**). As with plastid and mitochondrial gene expression and editing patterns, immature xylem and phloem have similar cosRNA profiles (Figure **4.4a** and Figure **S4.7b**, File **S4.3**, File **S4.6**). In the plastid genome, a total of 74 cosRNAs are differentially up-regulated in immature xylem and phloem compared to mature leaf (log2 fold change > 2) (File **S4.3**, File **S4.6**). To determine the relationship between plastid cosRNA abundance and plastid gene expression, the correlation between the log2 fold change of differentially expressed genes and differentially abundant cosRNAs, either overlapping or within 200 nt of the coding region, was analyzed (Figure **4.4b**). There were 34 plastid differentially abundant cosRNAs within 200 bps of 18 differentially expressed plastid genes, whose log2 fold change between immature xylem and mature leaf had a Pearson correlation of 0.69 (Figure **4.4b**), indicating that these cosRNAs may be the footprints of tissue-specific nuclear-encoded organellar RNA-binding proteins (Ruwe *et al.*, 2016).

Given that some antisense cosRNAs may play an inhibitory role by binding to sense mRNA (Zghidi-Abouzid *et al.*, 2011; Cavaiuolo *et al.*, 2017), the function of

differentially abundant cosRNAs between tissues is difficult to determine given the complex polycistronic transcription and transcript degradation in *E. grandis* organelles (Figure **4.1c**). Nevertheless, two examples that may illustrate the functional significance of cosRNA abundance between tissues are *psbH* and *ycf2* (Figure **4.4c**). *psbH* encodes a subunit of photosystem II (Komenda *et al.*, 2003), and has previously been shown to be downregulated in heterotrophic plastids (Kahlau & Bock, 2008; Valkov *et al.*, 2009; Zhelyazkova *et al.*, 2012). In *E. grandis*, the PEP expressed *psbH* transcript is highly abundant in mature leaf and has a cosRNA footprint (plastid cosRNA 82) 16 nt upstream of the start site, which is also highly abundant (Figure **4.4c**). HCF107, the RNA binding protein responsible for this cosRNA has been characterized (Hammani *et al.*, 2012). The nuclear-encoded HCF107 protein binds upstream of the start site of *psbH*, stabilizing its transcript abundance, and was highly and significantly upregulated in mature leaf compared to immature xylem (Figure **4.4c**). The *ycf2* transcript has not been previously associated with any known RNA binding proteins, but was significantly upregulated in immature xylem and phloem. A xylem and phloem preferentially abundant cosRNA footprint (plastid cosRNA 438) overlapping the 3' end may be due to an RNA binding protein protecting the transcripts from degradation, increasing the steady state level of *ycf2* (Figure **4.4c**).

4.4.5. Organellar targeted gene regulation in *E. grandis* reflects tissue-specific metabolic processes

Nuclear-encoded genes are largely responsible for the regulation of morphology, transcription, and metabolism within organelles in plant cells (Woodson & Chory,

2008; Weihe *et al.*, 2012; Ng *et al.*, 2014). To identify nuclear-encoded genes that regulate organellar metabolism during wood formation in *E. grandis*, the expression of nuclear-encoded, organellar-targeted genes across tissues was analyzed. Plastid-targeted genes were highly expressed in mature leaf compared to immature xylem and phloem, with ~40 % (1 130 of 2 778) plastid-targeted genes significantly up-regulated (log2 fold change > 2, q-value > 0.05) in mature leaf (Figure **4.5a**, File **S4.7**). These genes were predominantly photosystem and Calvin cycle genes, and the requisite photosynthetic regulatory and signalling factors (File **S4.7**). In mature leaf, plastid RNAbps were generally higher expressed compared to immature xylem and phloem (File **S4.7**). Plastid *GUN* genes have been identified by mutant screens that result in altered retrograde signalling from the plastid to the nuclear genome (Susek *et al.*, 1993). Mature leaf upregulated *Gun* genes are those involved in the retrograde signalling pathways that promote nuclear photosynthetic gene expression such as *GUN4* and G*UN5 (Eucgr.H00371* and *Eucgr.A02348*) (File **S4.7**) (Terry & Smith, 2013; Brzezowski *et al.*, 2014). In total, 45 out of 92 *known* plastid-targeted RNAbps (Castandet *et al.*, 2016; Reichel *et al.*, 2016) were significantly upregulated in mature leaf, reflecting the increased expression and editing of photosystem genes in leaf (Figure **4.2a**, File **S4.7**). Following this, these 45 RNAbps would not be the first candidates for functional studies of xyloplast regulation in *Eucalyptus*.

In the plastid immature xylem and phloem upregulated genes, there were no known regulators of plastid gene expression such as PPR proteins or other plastid targeted RNA binding proteins (File **S4.7**). Despite this, 140 plastid targeted genes were significantly upregulated in immature xylem compared to mature leaf (log2 fold change

> 2), of which 73 were above the 3rd quartile of expression values in immature xylem. In immature xylem and phloem, plastid-targeted gene expression showed that carbon metabolism between these two tissues varied in key pathways (Figure **4.5**, File **S4.7**). Of these, the pentose phosphate pathway and the shikimate pathway showed xylem-specific expression of several homologs (Figure **4.5c** and **d**). In developing xylem, the pentose phosphate pathway produces erythrose-4-phosphate (E-4-P) from the reactions of transketolase (TK) and/or transaldolase (TAL) (Henkes *et al.*, 2001; Caillau & Quick, 2005). The TK protein plays a role in the Calvin Cycle and AAA synthesis in sink tissues (Henkes *et al.*, 2001), but how it plays this role is unclear as of yet. Two TK homologs were highly and specifically expressed in either mature leaf (*Eucgr.D02467*) or immature xylem (*Eucgr.E01909*), respectively, suggesting that in *E. grandis* plastids, TK genes have sub-functionalized. Additionally, a TAL homolog (*Eucgr.D02030*) was highly and specifically expressed in immature xylem, which may act with the xylem specific TK to jointly regulate carbon allocation between primary metabolism and aromatic amino acid production (Caillau & Quick, 2005).

The plastid localized shikimate pathway is the only route to the production of aromatic amino acid phenylalanine in angiosperm plant cells and represents an irreversible shunt between sugars and phenolic biopolymers (Maeda & Dudareva, 2012). In immature xylem, shikimate pathway genes were highly and specifically expressed compared to phloem and mature leaf (Figure **4.5d**). The shikimate pathway uses erythrose-4-phosphate (E-4-P) and phosphoenolpyruvate (P-E-P) as substrates, which can also be utilized by other biosynthetic pathways in the plastid, such as fatty acid biosynthesis (Nishida, 2004). Phosphoenolpyruvate levels within plastids are

regulated by tissue specific functional differences in the P-E-P/phosphate translocator 1 (PPT/CUE1), which imports phosphoenolpyruvate into photosynthetic plastids. In non-photosynthetic root plastids, CUE1 acts as a net exporter of phosphoenolpyruvate, where a plastid specific enolase (ENO1) produces phosphoenolpyruvate in *Arabidopsis* (Prabhakar *et al.*, 2009; Lundquist *et al.*, 2014). In *E. grandis,* there is no homolog of *Eno1*, only the cytosolic *EnoC*, and phosphoenolpyruvate must be transported into plastids by CUE1 for the shikimate pathway and *de novo* fatty acid biosynthesis (Neuhaus & Emes, 2000). In support of the lack of plastidial phosphoenolpyruvate synthesis in *E. grandis*, a *Cue1* homolog was highly expressed in immature xylem compared to mature leaf and phloem (*Eucgr.B00299*), and another was expressed at a lower level almost exclusively in mature leaf (*Eucgr.K02835*) (Figure **4.5d**). Mesophyll specific *PPT2* (*Eucgr.J02222*) is specifically expressed in leaf tissue, at a lower absolute level than *PPT1* (*Eucgr.B00299*) in xylem (Figure **4.5d**, File **S4.7**), in agreement with its association with terpene biosynthesis in *Eucalyptus* leaves (Kainer *et al.*, 2019). Plastidial phosphoenolpyruvate concentrations in *E. grandis* may, therefore, be regulated by the immature xylem specific *Cue1* gene to ensure that phosphoenolpyruvate is available for the shikimate pathway during secondary growth.

Unlike mature leaf and immature xylem, phloem upregulated genes do not indicate that specific plastid metabolic processes define phloem metabolism, despite the fact that phloem is involved in the regulation of carbon partitioning between source and sink tissues through the exchange and remobilization of non-structural carbohydrates (Van Bel, 2003; Furze *et al.*, 2018). Sugar signalling plays a major role in whole plant

carbon allocation, especially in regulating the transient partitioning of non-storage carbohydrates over the diurnal cycle and in response to environmental and biotic stimulus (Griffiths *et al.*, 2016; Seki *et al.*, 2017). A regulator of carbon metabolism in land plants is SNF1-related kinases (SnRK1s - aka AKIN isoforms), which modulate carbon metabolism in a tissue-specific manner (Wurzinger *et al.*, 2018). In phloem, the *E. grandis* AKIN10 homolog *Eucgr.B00544* and its plastid-targeted interacting partner FCS-like zinc finger 10 (*FLZ10* - *Eucgr.C02876*) were highly and specifically expressed (Fragoso *et al.*, 2009; Bayer *et al.*, 2012; Nietzsche *et al.*, 2014; Jamsheer *et al.*, 2018) (File **S4.7**). A function of AKINs in plants is the down-regulation of energy intensive biosynthetic processes via enzyme phosphorylation (Fragoso *et al.*, 2009). The plastid-targeted metabolic pathways of lipid and starch biosynthesis are repressed by AKIN10/11, and several plastid-specific proteins are targets of AKIN10 (Nukarinen *et al.*, 2016; Broeckx *et al.*, 2016). Although the specific action of protein kinases cannot be determined by gene expression, the phloem-specific upregulation of *AKIN10* and *FLZ10* in phloem indicate a mechanism of preventing plastids from storing transported sugars as starch, and may explain the lack of phloem upregulated plastid carbon metabolic genes in an excess sugar environment (Robinson & Mansfield, 2011).

Therefore, although plastid-encoded transcriptomes between immature xylem and phloem are very similar, the predominant metabolic function of xylem plastids is the large-scale production of aromatic amino acids for monolignol synthesis, while phloem plastids may be regulated to facilitate effective signalling, transport, and unloading of sugars into sink tissues. These differences in the expression of nuclear-encoded

proteins increase the likelihood that plastids in xylem are uniquely adapted to produce phenylalanine for secondary cell wall monolignol biosynthesis, and gives more support for the xyloplast as a unique organelle.

Mitochondrial-targeted genes showed less variation across tissues compared to plastid-targeted genes, although some significant differences were observed (Figure **4.5b**). In mature leaf, the mitochondrial targeted components of the photorespiratory pathway are amongst the highest expressed mitochondrial targeted genes (File **S4.7**), with *serine transhydroxymethyltransferase I* (*SHMT1 - Eucgr.I00580*) and *glycine decarboxylase P-protein I* (*GLDP1 - Eucgr.C03047*) almost exclusively expressed in mature leaf (99% and 100% of total FPKM between tissues) (Figure **S4.8**). Mitochondrial genes in the tricarboxylic acid cycle are however not differentially expressed between tissues, as is malate dehydrogenase (File **S4.7**). Leaf mitochondria therefore specifically invest in serine biosynthesis to maintain photorespiration in chloroplasts as compared to immature xylem and phloem. The few differentially expressed genes in immature xylem and phloem mitochondria are generally lowly expressed, and involved in general stress response and do not indicate any wood specific mitochondrial metabolic functions (File **S4.7**). The transcript abundance of mitochondrial targeted proteins related to the TCA cycle may not accurately reflect biological variation between tissue types, as TCA cycle enzymes are known to be regulated at the post-translational level by oxidation and allosteric activation by pyruvate (Mackenzie & McIntosh, 1999). Taken together, nuclear-encoded mitochondrial gene expression shows that photorespiration is the primary

driver of metabolic variation between *E. grandis* leaf and woody tissues, which is reflected in the predominant editing seen in leaf mitochondrial transcripts.

4.5. Discussion

This study provides a novel, high-resolution profile of organellar transcriptome regulation in wood forming tissues by integrating small, total, and polyA-selected RNA sequencing data in leaf, xylem, and phloem. The main conclusions drawn from this study are that: (i.) Immature xylem and phloem plastids are regulated by similar mechanisms, indicated by comparable transcript abundance, editing, cosRNA footprints, and nuclear-encoded RNA binding protein expression; (ii.) Carbon metabolism in mature leaf plastids and mitochondria is optimized for photosynthesis (carbon sequestration and photorespiration), while the xylem plastid is distinctly dedicated to the synthesis of the aromatic amino acid phenylalanine for lignin deposition in secondary cell walls. This manifests in xylem through increased transcript abundance of plastid-encoded transcripts relating to heterotrophic function, with the potential action of stabilizing RNA-binding proteins indicating *active* regulation by the nuclear genome.

In *E. grandis* mature leaf, immature xylem, and phloem, the percentage of total reads derived from the mitochondrial genome is relatively similar, despite large differences in genome copy number (Figure **4.1a** and **4.b**). This agrees with findings from *Arabidopsis* that showed little correlation between mitochondrial gene copy number and transcript level or respiration rate between tissues and developmental stages

(Preuten *et al*., 2010). The vast majority of mitochondrial editing sites are preferentially edited in *E. grandis* mature leaf, similar to *Vitis vinifera* where 71% of sites were preferentially edited in leaf, compared to 0.4%, 21%, and 7.6% in stem, root, and callus (Picardi *et al.*, 2010) (Figure **S4.5b**, Figure **S4.6**). The decreased editing in immature xylem and phloem mitochondria may be due to dual targeted proteins that coordinate organellar editing in photosynthetic tissues, and have large effects, such as *RIP1* which affects 266 editing sites in mitochondria (Bentolila *et al.*, 2012; Andrés-Colás *et al.*, 2017). Mitochondrial genome copy number thus plays a limited role in transcript abundance, and the editing status and nuclear-encoded metabolic genes of *E. grandis* mitochondria may largely be regulated by the inter-organellar coordination determined by photosynthesis in leaves.

Plastid genome copy number and the maintenance of plastid genomes during leaf development have been a somewhat controversial subject in the past (Golczyk *et al.*, 2014; Oldenburg *et al.*, 2014). Previously, it was found that during *Arabidopsis* leaf development, plastid DNA copy number increases dramatically during the proplastid to chloroplast transition, and then declines as leaves mature, reaching stability during the later stages of maturation up to senescence (Sakamoto & Takami, 2018). A more recent comprehensive study has found that experimental deficits are responsible for the apparent decline in plastid genome copy number in plants over development (Greiner *et al*. 2019). In this study, we have designed a robust panel for profiling organellar DNA copy numbers in *Eucalyptus*, given the number of individual assays providing replicated evidence measurements (Supplementary Method **4.1**). In *E. grandis*, plastid DNA copy number is significantly higher in mature leaf at ~ 400 per

haploid nuclear genome copy number compared to immature xylem and phloem at ~ 320 and ~ 350 (Figure **4.1a**). For leaf tissues, this is well within the reported range of plastid genome copy numbers (Zoschke *et al.*, 2007; Ma & Li, 2015). The relatively small difference between mature leaf and immature woody tissues (immature xylem and phloem) is somewhat unexpected, as the ratio between *E. grandis* xylem and leaf copy number is ~ 0.8, compared to ~0.4 between potato tubers and leaf, and 0.2 in maize green vs non-green tissues (Valkov *et al.*, 2009; Kumar *et al.*, 2014). This could be because plastid transcripts in leaf tissues do not need to be dynamically transcribed, as plastid transcripts can have half-lives as long as 40 hours in barley (Kim *et al.*, 1993). Alternatively, a relationship between plastid genome copy number and transcript abundance has been suggested in the analysis of plastid DNA replication impaired *w2* mutants in maize leaf, where transcription of NEP expressed genes are related to genome copy number, although this may be due to a threshold effect in the mutants (Udy *et al.*, 2012). Immature xylem and phloem genome copy numbers may, therefore, be regulated in a tissue-specific manner to ensure the efficient transcription of NEP expressed genes (Figure **4.2c**).

In *E. grandis* immature xylem and phloem, *ycf2* is the most significantly upregulated gene compared to mature leaf (log2 fold change > 4) and is the most highly upregulated NEP gene identified to date in non-photosynthetic plastids (Figure **4.2a**) (Kahlau & Bock, 2008, Valkov *et al.*, 2009). Previously, it was suggested that NEP expressed genes in tomato fruit chromoplasts are expressed to support the expression *accD*, which plays an important role in fatty acid biosynthesis in fruit (Kahlau & Bock, 2008). Similarly, the upregulation of *ycf2* (and *ycf1*) in immature xylem and phloem

may be driven by increased NEP activity, and the non-specific action of NEP leads to the concomitant upregulation of housekeeping genes, including the PEP subunit *rpo* genes for photosystem gene expression (Legen *et al.*, 2002; Cho *et al.*, 2009). The PEP subunit *rpo* genes are lowly edited across all sites in immature xylem and phloem, which implies that they may not form fully functional protein complexes, essentially "down-regulating" photosystem genes in immature xylem and phloem (Figure **4.3**). In immature xylem and phloem plastids, tissue specific editing may act to inhibit the translation of photosynthetic genes that are expressed by NEP, which is highly active to express *ycf2*.

The *ycf2* gene has recently been characterized as a component of the translocon of the inner chloroplast membrane (TIC) complex (Kikuchi *et al.*, 2018). Nuclear-encoded proteins are transported into plastids by the translocon of the outer chloroplast membrane (TOC), and into the stroma by the TIC complex. Components of the TOC complex have long been known to preferentially import photosynthetic proteins (TOC33 and TOC159), or non-photosynthetic proteins in heterotrophic tissues such as roots (TOC34, TOC132, and TOC120) (reviewed in Demarsy *et al.*, 2014). Similarly, *ycf1* (TIC214) and TIC20-IV have also been associated with non-photosynthetic protein import (Kikuchi *et al.*, 2013; Nakai, 2015). Evidence for a predominantly non-photosynthetic import function of *ycf2* is mounting, given the interaction of *ycf2* and plastidial NAD-dependent malate dehydrogenase (pdNAD-MDH), which is associated with heterotrophic plastidial metabolism, and moonlights as a structural protein in TIC complex assembly in *Arabidopsis* embryos (Beeler *et al.*, 2014; Schreier *et al.*, 2018; Kikuchi *et al.*, 2018). The distinct up-regulation of *ycf2* in

E. grandis immature xylem and phloem highlights the importance of specific protein import in heterotrophic metabolism of plastids. The similarity of immature xylem and phloem plastid-encoded gene expression, editing profiles, cosRNA abundance, and nuclear-encoded RNA binding protein expression profiles suggests that the transcriptomes of woody plastids are regulated by the same mechanisms in these functionally different tissues.

The analysis of non-photosynthetic organellar transcriptome dynamics presented here is a novel resource for the systems biology analysis of secondary growth wood formation in ecologically and economically important tree species. The specific upregulation of the shikimate pathway in xylem adds further evidence for the metabolic specialization of the xyloplast compared to the phloem plastid. The transcriptomic and cosRNA abundance profiles of immature xylem and phloem plastids suggest a common regulatory mechanism of cytonuclear coordination in woody tissues. It would be interesting to determine if the upregulation of *ycf2* in immature xylem and phloem means that the regulation of heterotrophic metabolism protein import is uniquely adapted to secondary growth, and how this impacts xylem and phloem plastidial metabolism. Compared to previous models of “maintenance” levels of plastid regulation in non-photosynthetic tissues, we hypothesize that woody plastid transcriptomes are actively regulated by increased activity of NEP and the regulation of *ycf2* transcript levels by a nuclear-encoded RNA binding protein. In conclusion, the unique organellar regulation in *E. grandis* xylem and phloem shows that there has been a concerted co-evolution of organellar biology and xylogenesis in vascular plants.

4.6. Data availability

All raw sequencing files will be submitted to the NCBI Sequence Read Archive as the chapter is submitted for publication. Any data is available for the purpose of examination upon request.

4.7. Acknowledgements & funding

The authors would like to acknowledge the following people for their kind assistance:

- Stephan Henning for loading and running the organellar gene panels on the QuantStudio12K Flex platform at the University of Pretoria.
- Pius Brzoska, Joanne Mizrachi, and Phil Smith from Thermo Fischer Scientific for their assistance in designing the organellar gene panels.
- Petrus Gostmann for DNA extractions and PCR confirmation for the quantification of organellar genome copy number.

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). DP was supported by the National Research Foundation of South Africa Scarce Skills grant.

The authors have no conflicts of interest to declare.

4.8. Author contributions

E.M. was the lead investigator, conceived of the study, provided scientific input, and edited the manuscript. D.B. generated the cosRNA data and edited the manuscript. D.P. performed the data analysis, generated the figures, and drafted the manuscript. A.A.M. was a co- investigator for the study, and edited the manuscript. All authors have read and commented on the manuscript.

4.9. References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. **1997**. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389–3402.

Anders S, Pyl PT, Huber W. **2015**. HTSeq - A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31:** 166–169

Andrés-Colás N, Zhu Q, Takenaka M, De Rybel B, Weijers D, Van Der Straeten D. **2017**. Multiple PPR protein interactions are involved in the RNA editing system in *Arabidopsis* mitochondria and plastids. *Proceedings of the National Academy of Sciences* **114**: 8883–8888.

Axtell MJ. 2013. ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* **19:** 740–751.

Barkan A. **2011**. Expression of plastid genes: Organelle-specific elaborations on a prokaryotic scaffold. **155**: 1520–1532.

Bartholomé J, Mandrou E, Mabiala A, Jenkins J. 2015. High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *New Phytologist* **206**: 1283-1296.

Bauwe H, Hagemann M, Fernie AR. 2010. Photorespiration: Players, partners and origin. *Trends in Plant Science* **15:** 330–336.

Bayer RG, Stael S, Rocha AG, Mair A, Vothknecht UC, Teige M. **2012**. Chloroplast-localized protein kinases: a step forward towards a complete inventory. *Journal of Experimental Botany* **63**: 1713–1723.

Beeler S, Liu H-C, Stadler M, Schreier T, Eicke S, Lue W-L, Truernit E, Zeeman SC, Chen J, Kötting O. **2014**. Plastidial NAD-dependent malate dehydrogenase is critical for embryo development and heterotrophic metabolism in *Arabidopsis*. *Plant Physiology* **164**: 1175–1190.

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

Bentolila S, Heller WP, Sun T, Babina AM, Friso G, van Wijk KJ, Hanson MR. **2012**. RIP1, a member of an *Arabidopsis* protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proceedings of the National Academy of Sciences* **109**: E1453–61.

Bock R, Hagemann R, Kössel H, Kudla J. 1993. Tissue- and stage-specific modulation of RNA editing of the *psbF* and *psbL* transcript from spinach plastids- A new regulatory mechanism? *Molecular Genetics and Genomics* **240**: 238–44.

Bramham L, Pyke K. **2017**. Changing plastid dynamics within early root and shoot apical meristem-derived tissue of *A. thaliana*. *Bioscience Horizons* **10**. doi.org/10.1093/biohorizons/hzx001.

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.

Broeckx T, Hulsmans S, Rolland F. **2016**. The plant energy sensor: Evolutionary conservation and divergence of SnRK1 structure, regulation, and function. *Journal Of Experimental Botany* **67**: 6215–6252.

Brzezowski P, Schlicke H, Richter A, Dent RM, Niyogi KK, Grimm B. **2014**. The GUN4 protein plays a regulatory role in tetrapyrrole biosynthesis and chloroplast-to-nucleus signalling in *Chlamydomonas reinhardtii. The Plant Journal* **79**: 285–298.

Bykova NV, Møller IM, Gardeström P, Igamberdiev AU. 2014. The function of glycine decarboxylase complex is optimized to maintain high photorespiratory flux via buffering of its reaction products. *Mitochondrion* **19:** 357–364.

Caillau M, Quick PW. **2005**. New insights into plant transaldolase. *The Plant Journal* **43**: 1–16.

Castandet B, Hotto AM, Strickler SR, Stern DB. **2016**. ChloroSeq, an optimized chloroplast RNA-Seq bioinformatic pipeline, reveals remodelling of the organellar transcriptome under heat stress. *G3 Genes, Genomes, Genetics* **9**: 1–32.

Cavaiuolo M, Kuras R, Wollman FA. **2017**. Small RNA profiling in *Chlamydomonas*: insights into chloroplast RNA metabolism. *Nucleic Acids Research* **45:** 10783–10799.

Cho WK, Geimer S, Meurer J. **2009**. Cluster analysis and comparison of various chloroplast transcriptomes and genes in *Arabidopsis thaliana*. *DNA Research* **16**: 31–44.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Demarsy E, Lakshmanan AM, Kessler F. **2014**. Border control: Selectivity of chloroplast protein import and regulation at the TOC-complex. *Frontiers in Plant Science* **5**: 483–486.

Drescher A, Ruf S, Calsa T, Carrer H, Bock R. **2000**. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *The Plant Journal* **22**: 97–104.

Fang Y, Wu H, Zhang T, Yang M, Yin Y, Pan L, Yu X, Zhang X, Hu S, Al-Mssallem IS, *et al.* 2012. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS one* **7**: e37164.

Fragoso S, Espíndola L, Páez-Valencia J, Gamboa A, Camacho Y, Martínez-Barajas E, Coello P. **2009**. SnRK1 isoforms AKIN10 and AKIN11 are differentially regulated in Arabidopsis plants under phosphate starvation. *Plant Physiology* **149**: 1906–1916.

Fujii S, Toda T, Kikuchi S, Suzuki R, Yokoyama K, Tsuchida H, Yano K, Toriyama K. **2011**. Transcriptome map of plant mitochondria reveals islands of unexpected transcribed regions. *BMC Genomics* **12**: 279.

Furze ME, Trumbore S, Hartmann H. **2018**. Detours on the phloem sugar highway: stem carbon storage and remobilization. *Current Opinion in Plant Biology* **43**: 89–95.

Golczyk H, Greiner S, Wanner G, Weihe A, Bock R. **2014**. Chloroplast DNA in mature and senescing leaves: a reappraisal. *The Plant Cell* **26**: 847–854.

Grattapaglia D, Bradshaw HD Jr. **1994**. Nuclear DNA content of commercially important *Eucalyptus* species and hybrids. *Canadian Journal of Forest Research* **24**: 1074–1078.

Greiner S, Golczyk H, Malinova I, Pellizzer T, Bock R. **2019**. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. *bioRxiv* DOI: 10.1101/632240

Griffiths CA, Paul MJ, Foyer CH. **2016**. Metabolite transport and associated sugar signalling systems underpinning source/sink interactions. *Biochimica et Biophysica Acta* **1857**: 1715–1725.

Gualberto JM, Kühn K. **2014**. Mitochondrion DNA-binding proteins in plant mitochondria: Implications for transcription. *Mitochondrion* **10**: 1–6.

Hameed MW, Juszczak I, Bock R, van Dongen JT. **2017**. Comparison of mitochondrial gene expression and polysome loading in different tobacco tissues. *Plant Methods* **13**: 112–119.

Hammani K, Cook WB, Barkan A. **2012**. RNA binding and RNA remodelling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. *Proceedings of the National Academy of Sciences* **109**: 5651–5656.

Hanson MR, Sattarzadeh A. **2014**. Fluorescent labeling and confocal microscopic imaging of chloroplasts and non-green plastids. *Methods in Molecular Biology* **1132**: 125–143.

Hedtke B, Wagner I, Börner T, Hess WR. **1999**. Inter-organellar crosstalk in higher plants: Impaired chloroplast development affects mitochondrial gene and transcript levels. *The Plant Journal* **19**: 635–643.

Henkes S, Sonnewald U, Badur R, Flachmann R, Stitt M. **2001**. A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *The Plant Cell* **13**: 535–551.

Jamsheer K M, Sharma M, Singh D, Mannully CT, Jindal S, Shukla BN, Laxmi A. **2018**. FCS-like zinc finger 6 and 10 repress SnRK1 signalling in *Arabidopsis*. *The Plant Journal* **94**: 232–245.

Kahlau S, Bock R. **2008**. Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: Chromoplast gene expression largely serves the production of a single protein. *The Plant Cell* **20**: 856–874.

Kaila T, Saxena S, Ramakrishna G, Tyagi A, Tribhuvan KU, Srivastava H, Sandhya, Chaudhury A, Singh NK, Gaikwad K. **2019**. Comparative RNA editing profile of mitochondrial transcripts in cytoplasmic male sterile and fertile pigeonpea reveal significant changes at the protein level. *Molecular Biology Reports* **42:** 2067–2084.

Kainer D, Padovan A, Degenhardt J, Krause S, Mondal P, Foley WJ, Külheim C. 2019. High marker density GWAS provides novel insights into the genomic architecture of terpene oil yield in *Eucalyptus*. *New Phytologist* **223**: 1489–1504.

Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bédard J, Hirabayashi-Ishioka Y, *et al.* 2018. A Ycf2-FtsHi heteromeric

AAA-ATPase complex is required for chloroplast protein import. *The Plant Cell* **30**: 2677–2703.

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. **2013**. Uncovering the protein translocon at the chloroplast inner envelope membrane. *Science* **339**: 571–574.

Kim M, Christopher DA, Mullet JE. **1993**. Direct evidence for selective modulation of *psbA*, *rpoA*, *rbcL* and *16S* RNA stability during barley chloroplast development. *Plant Molecular Biology* **22**: 447–463.

Komenda J, Štys D, Lupínková L. **2003**. The PsbH protein of photosystem 2. *Photosynthetica* **41**: 1–7.

Kumar RA, Oldenburg DJ, Bendich AJ. **2014**. Changes in DNA damage, molecular integrity, and copy number for plastid DNA and mitochondrial DNA during maize development. *Journal of Experimental Botany* **65**: 6425–6439.

Langmead B, Salzberg SL. **2012**. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Legen J, Kemp S, Krause K, Profanter B, Herrmann RG, Maier RM. **2002**. Comparative analysis of plastid transcription profiles of entire plastid chromosomes from tobacco attributed to wild-type and PEP-deficient transcription machineries. *The Plant Journal* **31**: 171–188.

Levy S, Schuster G. **2016**. Polyadenylation and degradation of RNA in the mitochondria. *Biochemical Society Transactions* **44**: 1475–1482.

Liebers M, Grübler B, Chevalier F, Lerbs-Mache S, Merendino L, Blanvillain R, Pfannschmidt T. **2017**. Regulatory shifts in plastid transcription play a key role in

morphological conversions of plastids during plant development. *Frontiers in Plant Science* **8**: 23-31.

Liere K, Börner T. **2013**. Development-dependent changes in the amount and structural organization of plastid DNA. In: Biswal B, Krupinska K, Biswal UC, eds. *Advances in Photosynthesis and Respiration. Plastid Development in Leaves during Growth and Senescence*. Springer Netherlands, 215–237.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. **2009**. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lisitsky I, Schuster G. **1999**. Polyadenylation and degradation of mRNA in the chloroplast. In: *The Chloroplast: From Molecular Biology to Biotechnology*. Springer, Dordrecht, 85–90.

Lopez-Juez E, Pyke K a. 2005. Plastids unleashed: their development and their integration in plant development. *The International Journal of Developmental Biology* **49**: 557–577.

Love M, Anders S, Huber W. **2014**. Differential analysis of count data – the DESeq2 package. *Genome Biology* **15**:1-54.

Lundquist PK, Rosar C, Bräutigam A, Weber APM. **2014**. Plastid signals and the bundle sheath: Mesophyll development in reticulate mutants. *Molecular Plant* **7**: 14–29.

Lysenko EA, Kuznetsov VV. **2005**. Plastid RNA polymerases. *Molecular Biology* **39**: 661–674.

Maeda H, Dudareva N. **2012**. The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annual Review of Plant Biology* **63**: 73–105.

Ma J, Li XQ. **2015**. Organellar genome copy number variation and integrity during moderate maturation of roots and leaves of maize seedlings. *Current Genetics* **61**: 591–600.

Mackenzie S, McIntosh L. 1999. Higher plant mitochondria. *The Plant Cell* **4:** 571-585.

Michalovova M, Vyskot B, Kejnovsky E. **2013**. Analysis of plastid and mitochondrial DNA insertions in the nucleus (NUPTs and NUMTs) of six plant species: size, relative age and chromosomal localization. *Heredity* **111**: 314–320.

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**: 681–693.

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.

Nakai M. **2015**. YCF1: A green TIC?: Response to the de Vries *et al*. Commentary. *The Plant Cell* **27**: 1834–1838.

Neuhaus HE, Emes MJ. **2000**. Nonphotosynthetic metabolism in plastids. *Annual Review of Plant Physiology and Plant Molecular Biology* **51**: 111–140.

Ng S, De Clercq I, Van Aken O, Law SR, Ivanova A, Willems P, Giraud E, Van Breusegem F, Whelan J. **2014**. Anterograde and retrograde regulation of nuclear genes encoding mitochondrial proteins during growth, development, and stress. *Molecular Plant* **7**: 1075–1093.

Nietzsche M, Schießl I, Börnke F. **2014**. The complex becomes more complex: protein-protein interactions of SnRK1 with DUF581 family proteins provide a

framework for cell- and stimulus type-specific SnRK1 signalling in plants. *Frontiers in Plant Science* **5**: 54.

Nishida I. **2004**. Plastid metabolic pathways for fatty acid metabolism. In: Daniell H, Chase C, eds. *Molecular Biology and Biotechnology of Plant Organelles*. Springer Netherlands, 543–564.

Nukarinen E, Nägele T, Pedrotti L, Wurzinger B, Mair A, Landgraf R, Börnke F, Hanson J, Teige M, Baena-Gonzalez E, *et al.* 2016. Quantitative phosphoproteomics reveals the role of the AMPK plant ortholog SnRK1 as a metabolic master regulator under energy deprivation. *Scientific Reports* **6**: 31697.

Oldenburg DJ, Rowan BA, Kumar RA, Bendich AJ. **2014**. On the fate of plastid DNA molecules during leaf development: Response to the Golczyk *et al*. Commentary. *The Plant Cell* **26**: 855–861.

Picardi E, Horner DS, Chiara M, Schiavon R. **2010**. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. *Nucleic Acids Research* **38:** 4755–4767.

Pinard D, Fierro AC, Marchal K, Myburg AA. **2019a**. Organellar carbon metabolism is co-ordinated with distinct developmental phases of secondary xylem. *New Phytologist* **222**: 1832-1845.

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

Pinard D, Myburg AA, Mizrachi E. **2019b**. The plastid and mitochondrial genomes of *Eucalyptus grandis*. *BMC Genomics* **20**: 132.

Planchard N, Bertin P, Quadrado M, Dargel-Graffin C, Hatin I, Namy O, Mireau H. 2018. The translational landscape of *Arabidopsis* mitochondria. *Nucleic Acids Research* **46:** 6218–6228.

Prabhakar V, Löttgert T, Gigolashvili T, Bell K. **2009**. Molecular and functional characterization of the plastid-localized phosphoenolpyruvate enolase (ENO1) from *Arabidopsis thaliana*. *FEBS Letters* **583:** 983-991.

Preuten T, Cincu E, Fuchs J, Zoschke R, Liere K, Börner T. **2010**. Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *The Plant Journal* **64**: 948–959.

Quinlan AR. **2014**. BEDTools: The Swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics* **47**: 11.12.1–34.

Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. **2014**. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research* **42**: W187–191.

Reichel M, Liao Y, Rettel M, Ragan C, Evers M, Alleaume A-M, Horos R, Hentze MW, Preiss T, Millar AA. **2016**. *In planta* determination of the mRNA-Binding proteome of *Arabidopsis* etiolated seedlings. *The Plant Cell* **28**: 2435–2452.

Robinson AR, Mansfield SD. Metabolomics in poplar. In: Joshi CP, Difazio S, Chittaranjan K, eds. *Genetics, genomics and breeding of poplar*. Enfield, NH, USA: CRC Press, Science Publishers, Inc. (Imprint of Edenbridge Ltd.), 166–192.

Ruwe H, Wang G, Gusewski S, Schmitz-linneweber C. **2016**. Systematic analysis of plant mitochondrial and chloroplast small RNAs suggests organelle-specific mRNA stabilization mechanisms. *Nucleic Acids Research* **44**: 1–12.

Sakamoto W, Takami T. **2018**. Chloroplast DNA dynamics: Copy number, quality control and degradation. *Plant & Cell Physiology* **59**: 1120–1127.

Sanitá Lima M, Smith DR. **2017**. Pervasive transcription of mitochondrial, plastid, and nucleomorph genomes across diverse plastid-bearing species. *Genome Biology and Evolution* **9:** 2650–2657.

Schmittgen TD, Livak KJ. **2008**. Analyzing real-time PCR data by the comparative C(T) method. *Nature Protocols* **3**: 1101–1108.

Schreier TB, Cléry A, Schläfli M, Galbier F, Stadler M, Demarsy E, Albertini D, Maier BA, Kessler F, Hörtensteiner S, *et al.* 2018. Plastidial NAD-dependent malate dehydrogenase: A moonlighting protein involved in early chloroplast development through its interaction with an FtsH12-FtsHi protease complex. *The Plant Cell* **30**: 1745–1769.

Seki M, Ohara T, Hearn TJ, Frank A, da Silva VCH, Caldana C, Webb AAR, Satake A. **2017**. Adjustment of the *Arabidopsis* circadian oscillator by sugar signalling dictates the regulation of starch metabolism. *Scientific Reports* **7**: 8305-8312.

Singh R, Singh S, Parihar P, Singh VP, Prasad SM. **2015**. Retrograde signalling between plastid and nucleus: A review. *Journal of Plant Physiology* **181**: 55–66.

Small ID, Rackham O, Filipovska A. **2013**. Organelle transcriptomes: Products of a deconstructed genome. *Current Opinion in Microbiology* **16**: 652–658.

Susek RE, Ausubel FM, Chory J. **1993**. Signal transduction mutants of *Arabidopsis* uncouple nuclear *CAB* and *RBCS* gene expression from chloroplast development. *Cell* **74**: 787–799.

Terry MJ, Smith AG. **2013**. A model for tetrapyrrole synthesis as the primary mechanism for plastid-to-nucleus signalling during chloroplast biogenesis. *Frontiers in Plant Science* **4**: 14.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. **2013**. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14**: 178–192.

Tseng C-C, Lee C-J, Chung Y-T, Sung T-Y, Hsieh M-H. **2013**. Differential regulation of *Arabidopsis* plastid gene expression and RNA editing in non-photosynthetic tissues. *Plant Molecular Biology* **82**: 375–392.

Udy DB, Belcher S, Williams-Carrier R, Gualberto JM, Barkan A. **2012**. Effects of reduced chloroplast gene copy number on chloroplast gene expression in maize. *Plant Physiology* **160**: 1420–1431.

Valkov VT, Scotti N, Kahlau S, Maclean D, Grillo S, Gray JC, Bock R, Cardi T, Nazionale C, Vegetale G, *et al.* 2009. Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: Transcriptional and posttranscriptional control. *Plant Physiology* **150**: 2030–2044.

Van Bel AJE. **2003**. Transport phloem: Low profile, high impact. *Plant Physiology* **131**: 1509–1510.

Voinnet O. **2009**. Origin, biogenesis, and activity of plant microRNAs. *Cell* **136**: 669–687.

Weihe A, Liere K, Börner T. **2012**. Transcription and transcription regulation in chloroplasts and mitochondria of higher plants. In: Bullerwell CE, ed. *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression*. Berlin, Heidelberg: Springer Berlin Heidelberg, 297–325.

Welchen E, García L, Mansilla N, Gonzalez DH. **2014**. Coordination of plant mitochondrial biogenesis: Keeping pace with cellular requirements. *Frontiers in Plant Science* **4**: 551-558.

Wodak SJ, Janin J. **2002**. Structural basis of macromolecular recognition. *Advances in Protein Chemistry* **61**: 9–73.

Woodson JD, Chory J. **2008**. Coordination of gene expression between organellar and nuclear genomes. *Nature Reviews Genetics* **9**: 383–395.

Woodson JD, Perez-Ruiz JM, Chory J. **2011**. Heme synthesis by plastid ferrochelatase i regulates nuclear gene expression in plants. *Current Biology* **21**: 897–903.

Woo HR, Koo HJ, Kim J, Jeong H, Yang JO, Lee IH, Jun JH, Choi SH, Park SJ, Kang B, *et al.* 2016. Programming of plant leaf senescence with temporal and inter-organellar coordination of transcriptome in *Arabidopsis*. *Plant Physiology* **171**: 452–467.

Wu S, Liu W, Aljohi HA, Alromaih SA, Alanazi IO, Lin Q, Yu J, Hu S. **2018**. REDO: RNA editing detection in plant organelles based on variant calling results. *Journal of Computational Biology* **25**: 509–516.

Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. **2016**. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. *Methods in Molecular Biology* **1418**: 283–334.

Wurzinger B, Nukarinen E, Nägele T, Weckwerth W, Teige M. **2018**. The SnRK1 kinase as central mediator of energy signalling between different organelles. *Plant Physiology* **176**: 1085–1094.

Yura K, Go M. **2008**. Correlation between amino acid residues converted by RNA editing and functional residues in protein three-dimensional structures in plant organelles. *BMC Plant Biology* **8**: 79.

Zghidi-Abouzid O, Merendino L, Buhr F, Malik Ghulam M, Lerbs-Mache S. **2011**. Characterization of plastid *psbT* sense and antisense RNAs. *Nucleic Acids Research* **39**: 5379–5387.

Zhang J, Ruf S, Hasse C, Childs L, Scharff LB, Bock R. **2012**. Identification of cis-elements conferring high levels of gene expression in non-green plastids. *The Plant Journal* **72**: 115–128.

Zhelyazkova P, Sharma CM, Forstner KU, Liere K, Vogel J, Borner T. **2012**. The primary transcriptome of barley chloroplasts: Numerous noncoding RNAs and the dominating role of the plastid-encoded RNA polymerase. *The Plant Cell* **24**: 123–136.

Zoschke R, Liere K, Börner T. **2007**. From seedling to mature plant: *Arabidopsis* plastidial genome copy number, RNA accumulation and transcription are differentially regulated during leaf development. *The Plant Journal* **50**: 710–722.

4.10. Figures

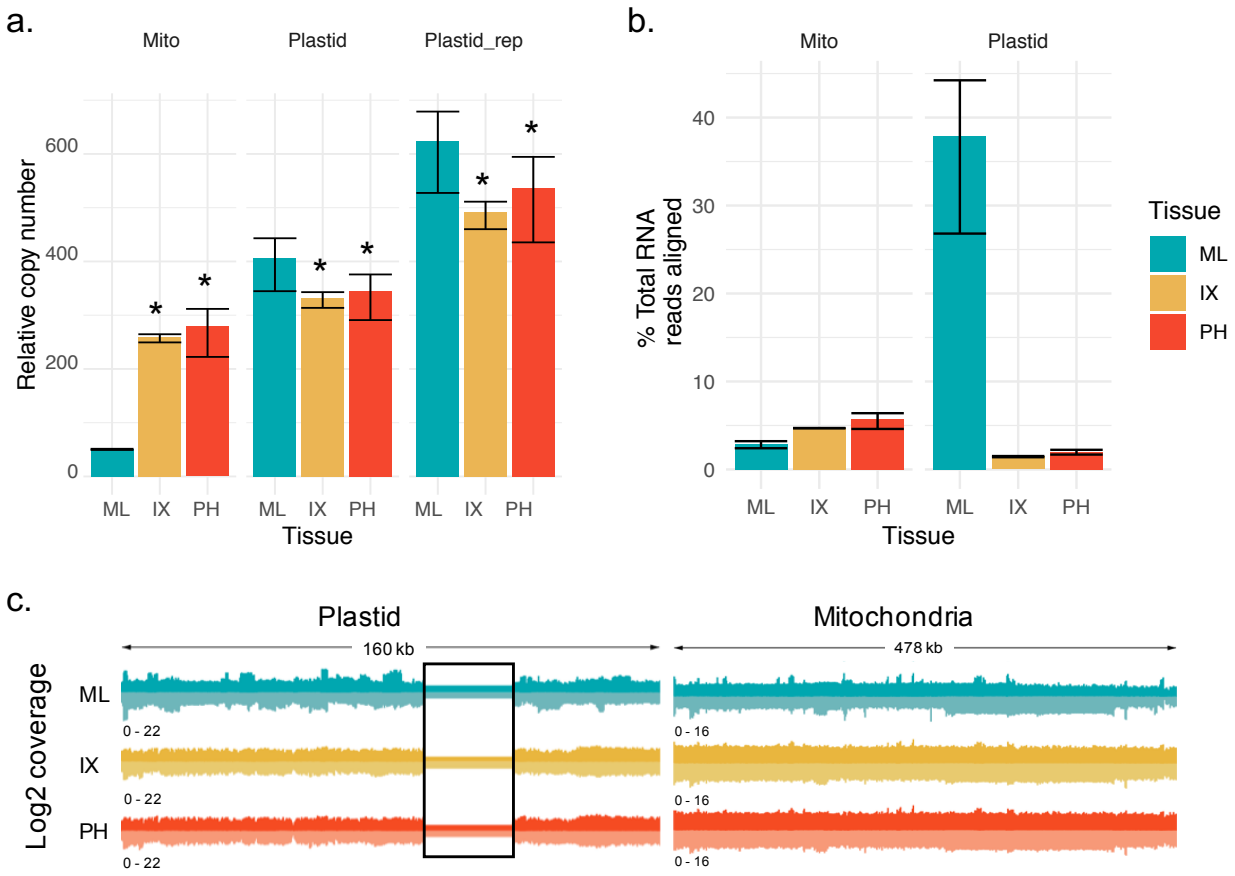


Figure 4.1. Relative DNA copy number and total RNA derived reads aligned to the mitochondrial and plastid genomes of *E. grandis*.

a. Relative DNA copy number of mitochondria, plastids, and plastid repeat regions in mature leaf (ML – Blue), immature xylem (IX), and developing phloem (PH) determined relative to nuclear genome copy number using qPCR. The x-axis shows the tissue and organelle analysed, and the y-axis gives the average organellar copy number (21 mitochondrial and 84 plastid assays) in units relative to the average diploid copy number of six nuclear control genes, determined using the Δ - CT method. The bars show standard error of means between three

biological replicates. Significant differences between the mean copy number between tissues was determined using ANOVA and Tukeys HSD. The asterisks denote statistical significant differences to the mean of ML (p-adj < 0.05).

b. Percentage of total RNA seq reads mapped to the mitochondrial and plastid genomes in ML, IX, and PH. The x-axis represents the tissue and organelle analysed, and the y-axis is the percentage of all sequenced reads aligned to the nuclear, mitochondrial, and plastid genomes. The bars indicate standard error across the biological repeats per tissue.

c. Total RNA seq read coverage across the plastid and mitochondrial genomes in ML, IX, and PH. The average read coverage across three biological replicates for ML (blue), IX (yellow), and PH (orange) are shown in log2 scale on the y-axis. The forward strand is indicated on top, and the reverse below in the lighter shade. The plastid genome is shown on the left, and the mitochondrial genome is on the right. Plastid inverted repeat region B was hard-masked prior to read alignment, and is indicated by the black box.

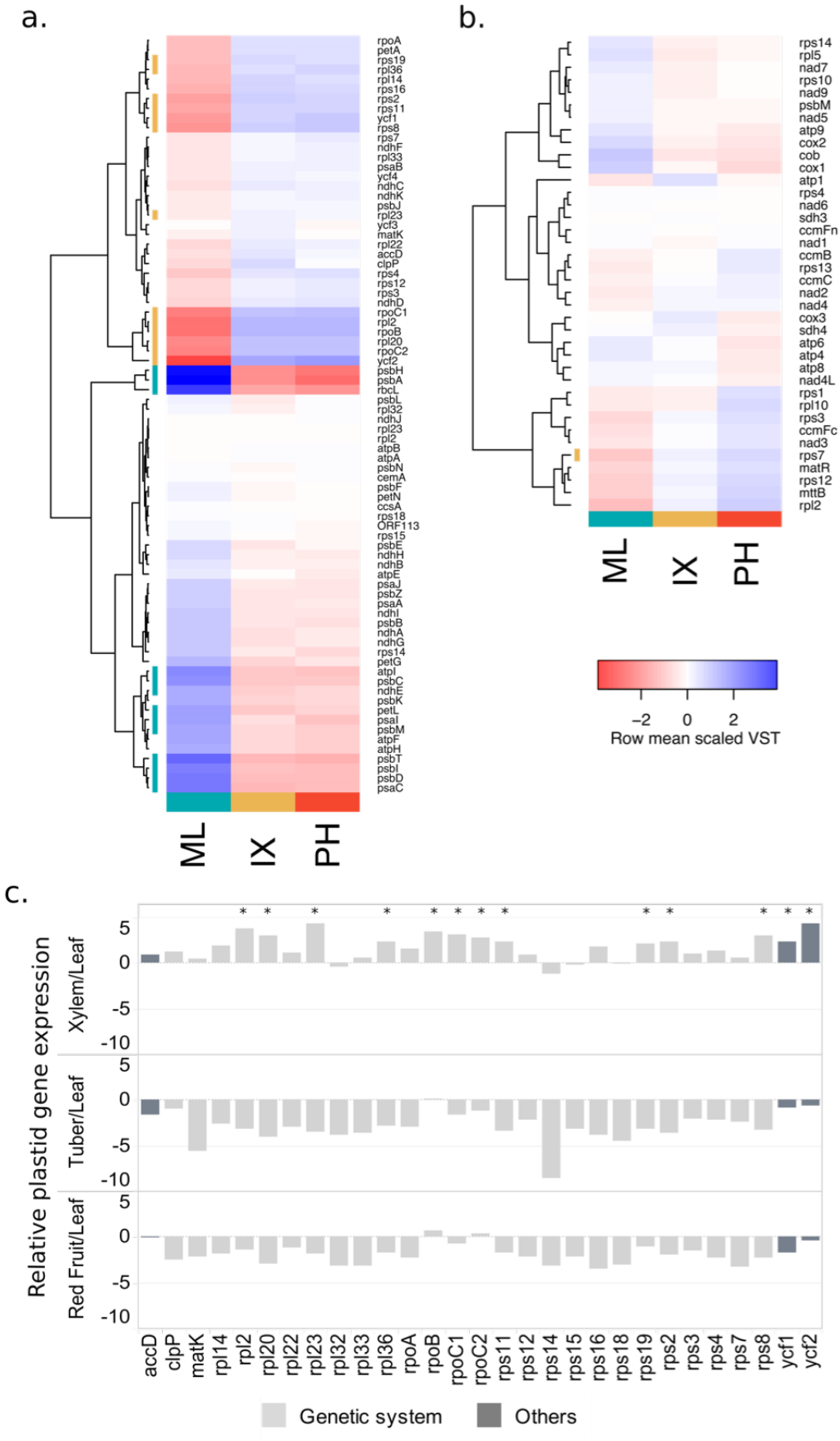

a.
b.
c.
ML
IX
PH
Row mean scaled VST
−2
0
2
Relative plastid gene expression
Xylem/Leaf
Tuber/Leaf
Red Fruit/Leaf
Genetic system
Others

Figure 4.2. Average tissue specific expression of plastid and mitochondrial encoded genes.

a. Heatmap representing row mean scaled VST of plastid-encoded genes in *E. grandis* mature leaf (ML - blue), immature xylem (IX - yellow), and developing phloem (PH) derived from total RNA seq reads. The bars next to the row dendrogram represent genes significantly differentially expressed in mature leaf (blue) or immature xylem (yellow).
b. Heatmap representing row mean scaled VST of mitochondrial encoded genes in *E. grandis*, with the colours matching those in (a). The scale bar is shown below the heatmap, and is scaled to the minimum and maximum values across the plastid and mitochondrial values for comparison.
c. Relative expression of differentially regulated housekeeping plastid genes in phototrophic and heterotrophic tissues in *E. grandis* (top – immature xylem vs leaf), potato (middle – tubers vs leaf), and tomato (bottom – red fruit vs leaf). Relative expression values are shown on the y-axis, with values determined using oligonucleotide microarrays for potato and tomato from Valkov *et al.*, (2009) and Kahlau and Bock (2008). Genetic system genes are coloured light grey, and genes that are classified as “others” in previous literature are coloured dark grey in the figure. All axes were standardized across graphs, and significantly differentially expressed *E. grandis* plastid genes are indicated with *.

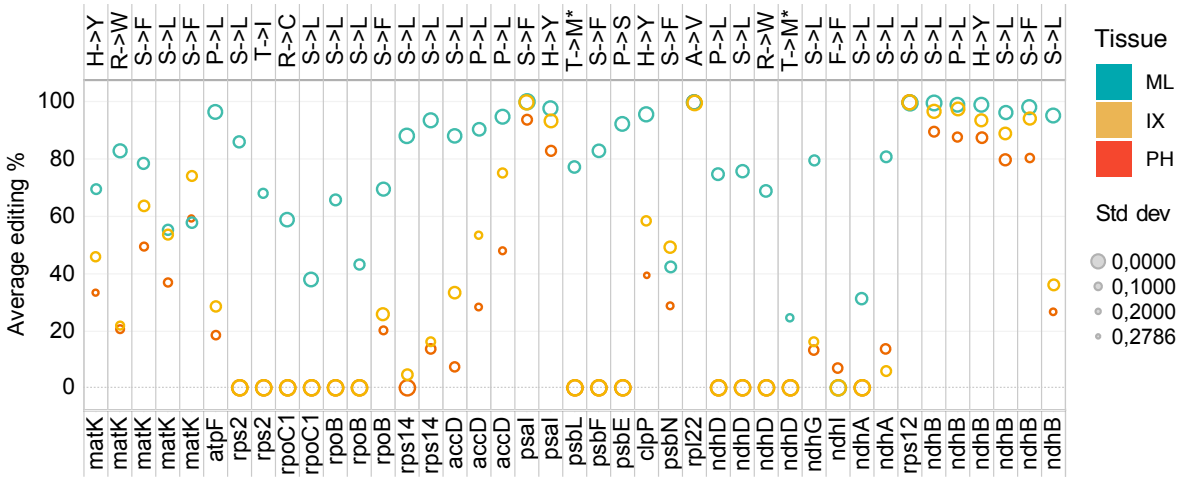


Figure 4.3. Plastid gene editing sites in *E. grandis* carbon source, sink and transport tissues

Editing sites per plastid gene showing the average editing percentage in mature leaf (blue), immature xylem (yellow), and phloem (orange). Sites were retained if at least one tissue showed editing frequency > 0 in all three replicates. The size of the circles represent the standard deviation of three biological replicates per tissue, ranging from 0 (large circle) to 0.2786 (small circles).

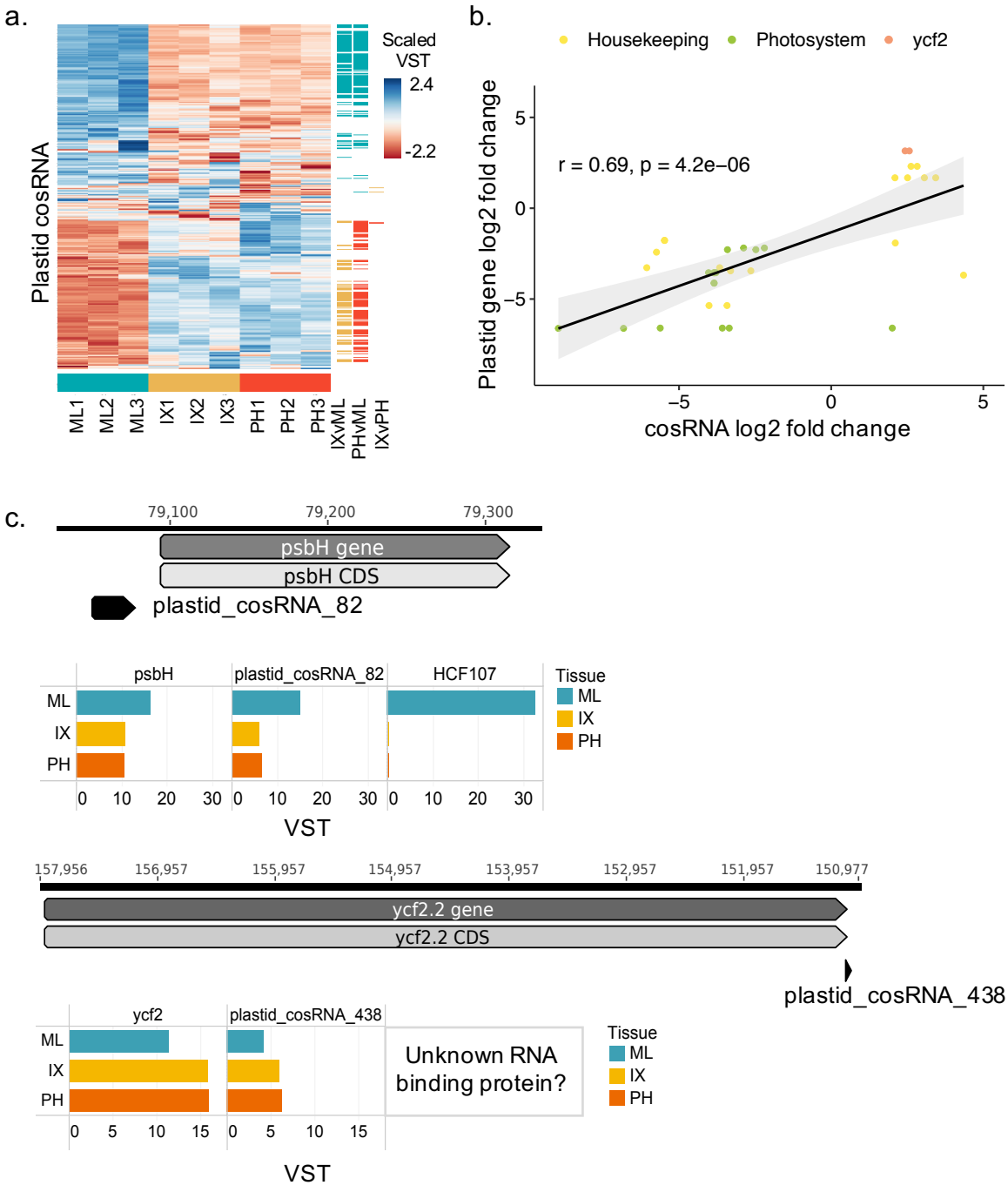

a.
Plastid cosRNA
Scaled VST
2.4
-2.2
ML1
ML2
ML3
IX1
IX2
IX3
PH1
PH2
PH3
IXvML
PHvML
IXvPH
b.
Housekeeping
Photosystem
ycf2
r = 0.69, p = 4.2e−06
Plastid gene log2 fold change
cosRNA log2 fold change
c.
79,100
79,200
79,300
psbH gene
psbH CDS
plastid_cosRNA_82
psbH
plastid_cosRNA_82
HCF107
Tissue
ML
IX
PH
VST
157,956
156,957
155,957
154,957
153,957
152,957
151,957
150,977
ycf2.2 gene
ycf2.2 CDS
plastid_cosRNA_438
ycf2
plastid_cosRNA_438
Unknown RNA binding protein?
VST

Figure 4.4. Tissue-specific abundance of cosRNAs in the *Eucalyptus grandis* plastid genome, and association of differentially abundant plastid cosRNAs and genes.

a. Heatmap of plastid cosRNA abundance in *Eucalyptus grandis* mature leaf (blue), phloem (orange), and immature xylem (yellow). cosRNA abundance is represented as row mean centred VST values from low (red) to high (blue). Significantly differentially abundant cosRNA between tissues are indicated in the bars to the right of the heatmap, where blue is significantly upregulated in mature leaf, yellow in immature xylem, and orange in phloem.
b. Correlation between differentially expressed (VST log2 fold change >2 or <-2) of plastid cosRNAs and genes in immature xylem and mature leaf. The x-axis is the log2 fold change of plastid genes, and the y-axis is the log2 fold change of plastid cosRNAs. The points are coloured according to plastid gene type, photosystem genes are shown in green, housekeeping genes in yellow, and *ycf2* (other) is in peach. Positive log2 fold change values indicate preferential xylem expression, and negative values show preferential mature leaf expression. The cosRNAs were filtered to either overlap with the genes, or to be within 200 bp of the gene. The Pearson correlation and p-value is shown on the graph.
c. Examples of relevant cosRNAs and their abundance between tissues compared to their associated plastid transcripts. Plastid cosRNA 148 overlaps the start site of *ndhB*, and the transcript and cosRNA are more abundant in mature leaf. The known RNA binding protein causing this cosRNA footprint *hcf107* expression is similarly highly expressed in mature leaf. Plastid cosRNA

438 overlaps with the 3' end of *ycf2*, and both are significantly upregulated in immature xylem and phloem, There is no known RNA binding protein which may cause this footprint. Tissues are coloured as above, and average VST is shown for cosRNA and plastid transcripts, and *hcf107* abundance is average FPKM.

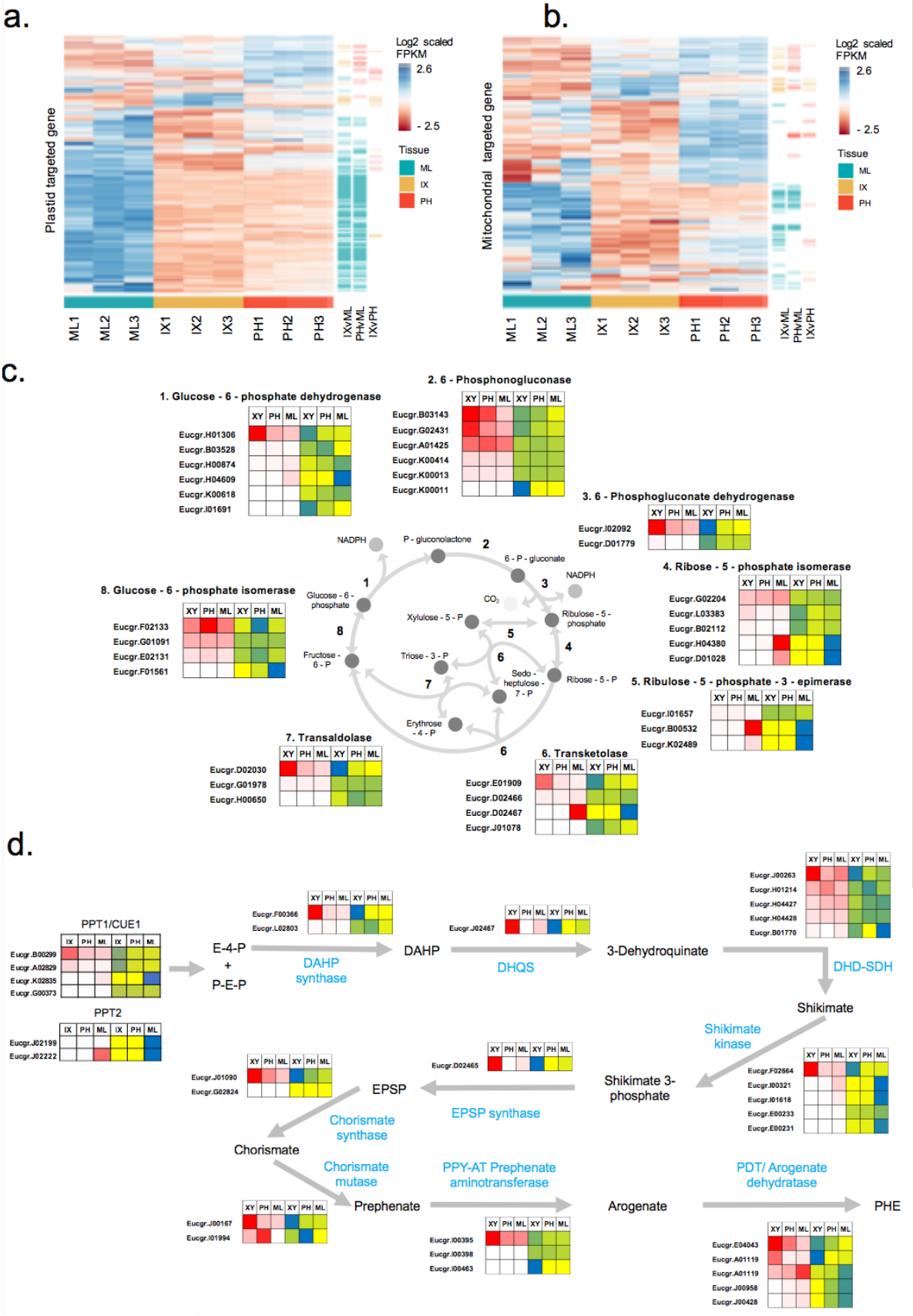

a.
Plastid targeted gene
Log2 scaled FPKM
2.6
- 2.5
Tissue
ML
IX
PH
ML1
ML2
ML3
IX1
IX2
IX3
PH1
PH2
PH3
IXvML
PHvML
IXvPH
b.
Mitochondrial targeted gene
c.
1. Glucose - 6 - phosphate dehydrogenase
Eucgr.H01306
Eucgr.B03528
Eucgr.H00874
Eucgr.H04609
Eucgr.K00618
Eucgr.I01691
2. 6 - Phosphonogluconase
Eucgr.B03143
Eucgr.G02431
Eucgr.A01425
Eucgr.K00414
Eucgr.K00013
Eucgr.K00011
3. 6 - Phosphogluconate dehydrogenase
Eucgr.I02092
Eucgr.D01779
4. Ribose - 5 - phosphate isomerase
Eucgr.G02204
Eucgr.L03383
Eucgr.B02112
Eucgr.H04380
Eucgr.D01028
5. Ribulose - 5 - phosphate - 3 - epimerase
Eucgr.I01657
Eucgr.B00532
Eucgr.K02489
6. Transketolase
Eucgr.E01909
Eucgr.D02466
Eucgr.D02467
Eucgr.J01078
7. Transaldolase
Eucgr.D02030
Eucgr.G01978
Eucgr.H00650
8. Glucose - 6 - phosphate isomerase
Eucgr.F02133
Eucgr.G01091
Eucgr.E02131
Eucgr.F01561
NADPH
P - gluconolactone
6 - P - gluconate
NADPH
CO2
Ribulose - 5 - phosphate
Ribose - 5 - P
Sedo - heptulose - 7 - P
Erythrose - 4 - P
Triose - 3 - P
Fructose - 6 - P
Xylulose - 5 - P
Glucose - 6 - phosphate
d.
PPT1/CUE1
Eucgr.B00299
Eucgr.A02829
Eucgr.K02835
Eucgr.G00373
PPT2
Eucgr.J02199
Eucgr.J02222
E-4-P
+
P-E-P
Eucgr.F00366
Eucgr.L02803
DAHP synthase
DAHP
Eucgr.J02467
DHQS
3-Dehydroquinate
Eucgr.J00263
Eucgr.H01214
Eucgr.H04427
Eucgr.H04428
Eucgr.B01770
DHD-SDH
Shikimate
Shikimate kinase
Eucgr.F02664
Eucgr.I00321
Eucgr.I01618
Eucgr.E00233
Eucgr.E00231
Shikimate 3-phosphate
Eucgr.D02465
EPSP synthase
EPSP
Eucgr.J01090
Eucgr.G02824
Chorismate synthase
Chorismate
Chorismate mutase
Eucgr.J00167
Eucgr.I01994
Prephenate
PPY-AT Prephenate aminotransferase
Eucgr.I00395
Eucgr.I00398
Eucgr.I00463
Arogenate
PDT/ Arogenate dehydratase
PHE
Eucgr.E04043
Eucgr.A01119
Eucgr.A01119
Eucgr.J00958
Eucgr.J00428

Figure 4.5. Expression of plastid and mitochondrial targeted genes in *E. grandis* mature leaf, immature xylem, and phloem.

a and b: Heatmaps of plastid and mitochondrial targeted nuclear gene expression as mean scaled and centred log2FPKM values in *E. grandis* mature leaf (ML - blue), immature xylem (IX - yellow), and phloem (PH). Differentially expressed genes between tissues are shown on the right of the heatmap: immature xylem compared to mature leaf (IXvML), phloem compared to mature leaf (PHvML), and immature xylem compared to phloem (IXvPH). c. Average absolute and relative expression of pentose phosphate pathway (PPP) genes in *E. grandis* immature xylem, mature leaf and phloem. Average expression of PPP homologs in *E. grandis* are shown as absolute FPKM from high (red) to low (white), and as a percentage of total expression between tissues from 100% (blue) to 0% (yellow) in order to identify tissue specific homologs. Enzymes in the pathway are labelled 1 to 8, and are shown in bold on the pathway. d. Average absolute and relative expression of shikimate pathway genes in *E. grandis* are shown as in c. Enzymes in the pathway are indicated in blue in the figure, and PPT/CUE1 P-E-P transporter is also shown in black text.

4.11. Supplementary Tables

Table S4.1:. RNA-sequencing mapping statistics

Total RNA-seq						
Sample	# of Reads	% of Reads Mapped	# Plastid Mapped	% Plastid Mapped	# Mito Mapped	% Mito Mapped
IX1	2.68E+08	94.63%	4.20E+06	2%	1.25E+07	5%
IX2	2.17E+08	94.68%	3.12E+06	1%	1.02E+07	5%
IX3	2.04E+08	94.71%	2.88E+06	1%	9.62E+06	5%
PH1	2.03E+08	94.91%	4.93E+06	2%	1.43E+07	7%
PH2	1.64E+08	94.23%	3.00E+06	2%	8.26E+06	5%
PH3	2.00E+08	94.34%	3.67E+06	2%	1.02E+07	5%
ML1	2.11E+08	96.50%	5.60E+07	27%	4.95E+06	2%
ML2	1.65E+08	98.44%	8.04E+07	49%	5.51E+06	3%
ML3	1.79E+08	97.50%	6.84E+07	38%	5.49E+06	3%
PolyA RNA-seq						
Sample	# of Reads	% of Reads Mapped	# Plastid Mapped	% Plastid Mapped	# Mito Mapped	% Mito Mapped
IX1	1.04E+08	97.18%	1.07E+04	0.01%	2.86E+04	0.03%
IX2	9.39E+07	97.26%	7.14E+03	0.01%	2.50E+04	0.03%
IX3	1.08E+08	97.22%	8.35E+03	0.01%	2.90E+04	0.03%
PH1	8.73E+07	96.58%	9.98E+03	0.01%	2.97E+04	0.03%
PH2	1.06E+08	96.92%	3.41E+04	0.03%	1.23E+05	0.12%
PH3	1.03E+08	96.54%	1.25E+04	0.01%	3.56E+04	0.03%
ML1	9.84E+07	97.50%	1.08E+05	0.11%	8.86E+04	0.09%
ML2	9.61E+07	97.46%	1.13E+05	0.12%	9.16E+04	0.10%
ML3	5.43E+07	97.06%	6.23E+04	0.11%	5.28E+04	0.10%
Small RNA-seq						
Sample	# of Reads	% of Reads Mapped [1]	# Plastid Mapped	% Plastid Mapped	# Mito Mapped	% Mito Mapped
IX1	2.38E+07	-	3.70E+05	1.56%	3.51E+05	1.47%
IX2	2.32E+07	-	3.51E+05	1.52%	2.87E+05	1.24%
IX3	2.42E+07	-	5.69E+05	2.35%	4.12E+05	1.71%
PH1	2.29E+07	-	6.54E+05	2.86%	4.73E+05	2.07%
PH2	2.43E+07	-	5.06E+05	2.08%	3.66E+05	1.51%
PH3	2.30E+07	-	4.78E+05	2.08%	3.32E+05	1.45%
ML1	2.25E+07	-	1.35E+06	6.01%	7.29E+05	3.25%
ML2	2.23E+07	-	1.82E+06	8.15%	6.44E+05	2.88%
ML3	2.43E+07	-	2.80E+06	11.54%	7.82E+05	3.22%

[1] Note that small RNA-reads were mapped to the organellar genomes only, and cosRNA were cross-referenced with the nuclear genome after identification as in Ruwe *et al.*, 2016

4.12. Supplementary Figures

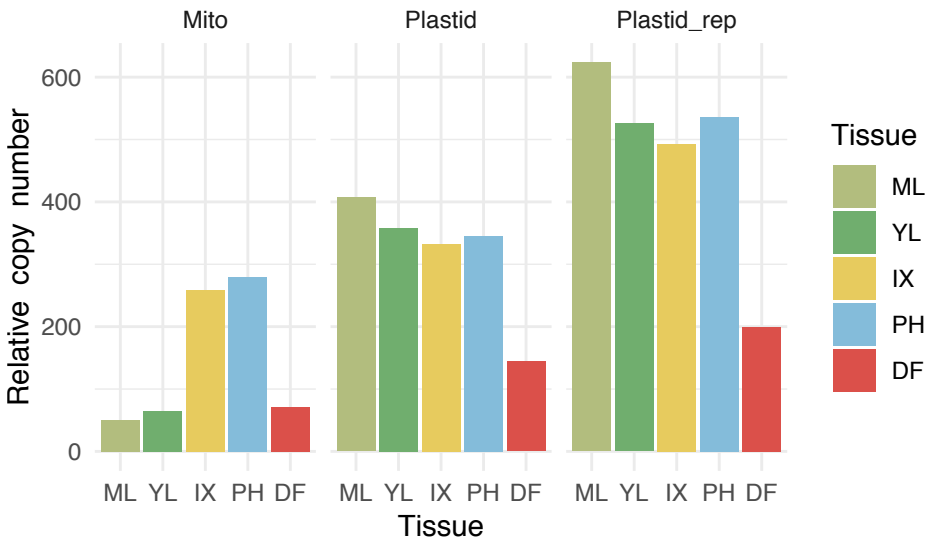


Figure S4.1. Plastid and mitochondrial genome copy number relative to diploid nuclear genome copy number in five *E. grandis* tissues at different developmental stages.

Young leaf (YL) and developing flower (DF) genome copy numbers are estimated using the average delta CT values of all probes in one tissue sample. Mature leaf (ML), immature xylem (IX), and phloem (PH) were calculated as for Figure 1 from three samples per tissue.

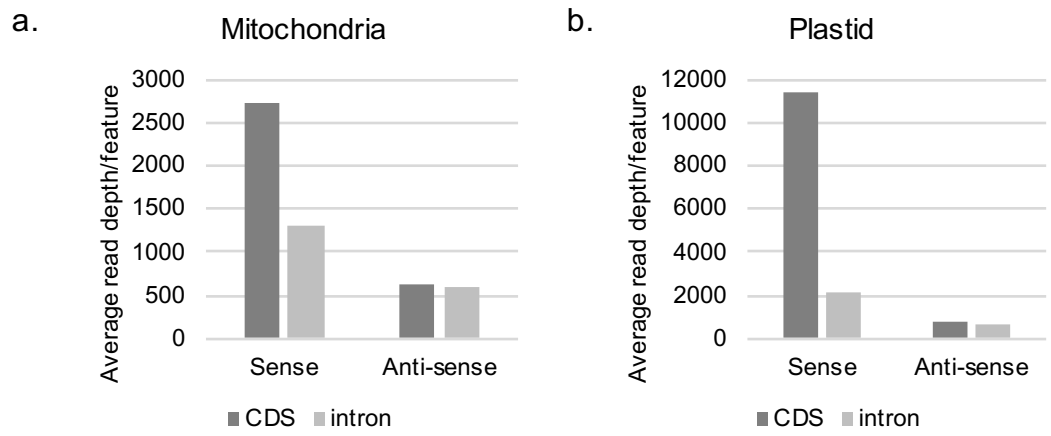


Figure S4.2. Average read depth per feature in sense and anti-sense orientation in (a.) mitochondria and (b.) plastids.

Average read depth across all nine samples (3 tissues with 3 biological replicates each) was calculated with bedtools intersect for each stranded total RNA seq library, and sense (dark grey) and antisense (light grey) strands were separated per feature (CDS and intron).

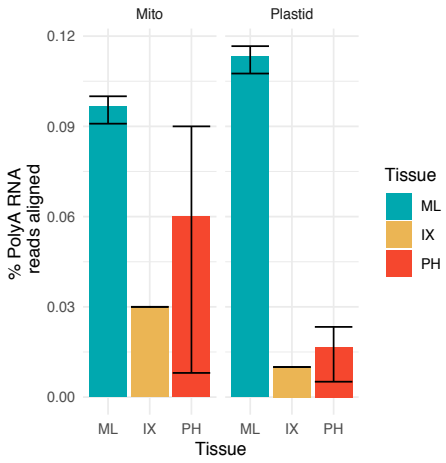


Figure S4.3. Percentage of polyA selected RNA seq reads mapped to the mitochondrial and plastid genomes in ML, IX, and PH.

The x-axis represents the tissue and organelle analysed, and the y-axis is the percentage of all sequenced reads (aligned and unaligned) to the nuclear, mitochondrial, and plastid genomes. The bars indicate standard error across the biological repeats per tissue.

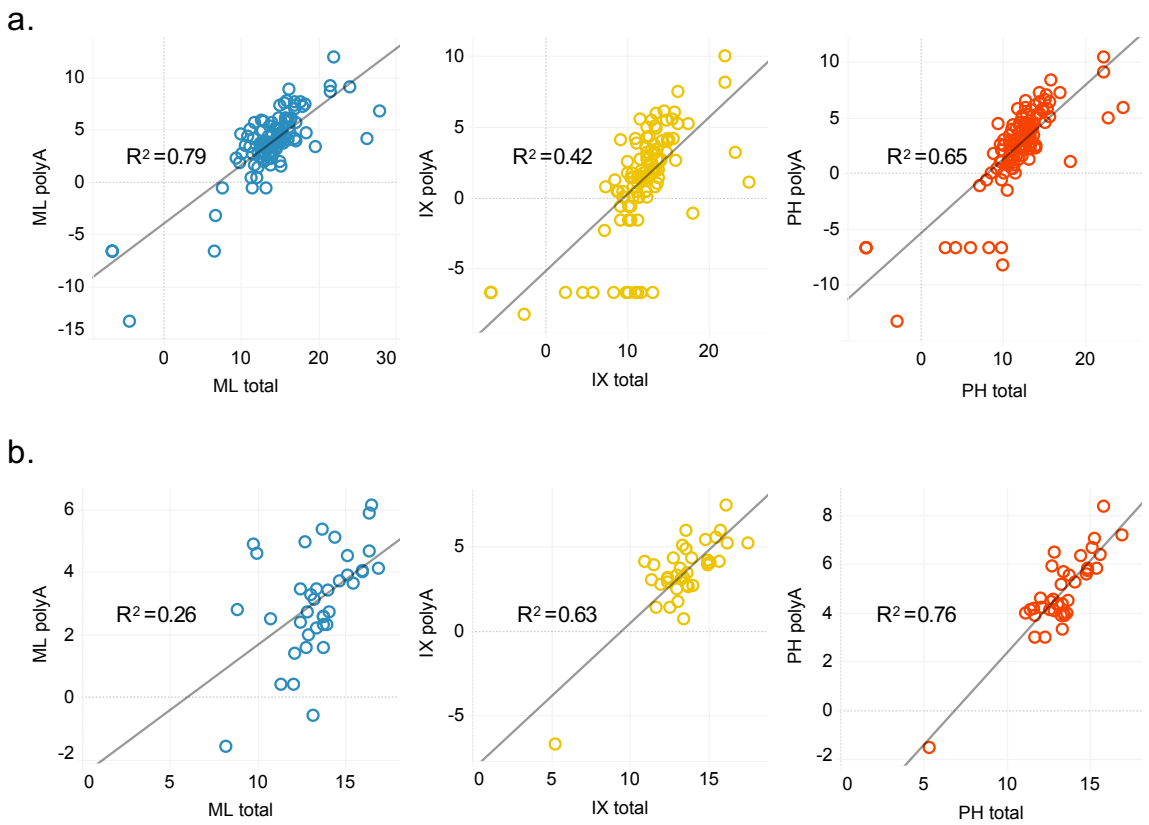
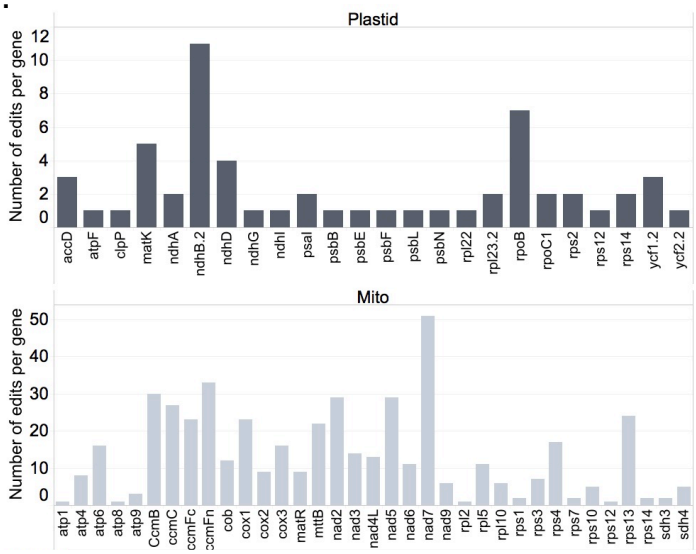


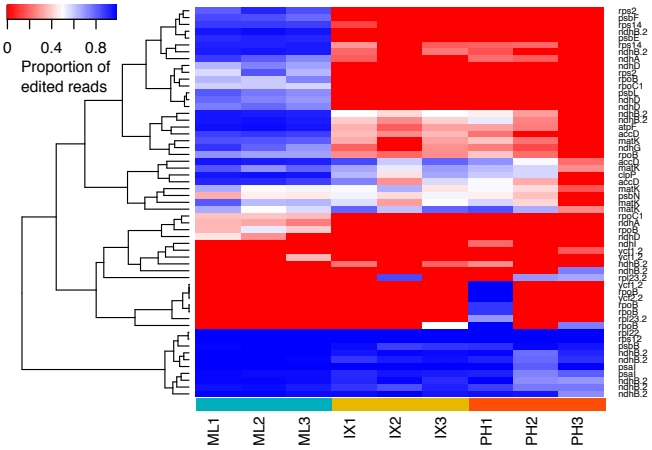
Figure S4.4. Correlation between total and polyA selected RNA seq read counts for *Eucalyptus grandis* plastid and mitochondrial genes in mature leaf, immature xylem, and phloem.

a. Log2 read count of plastid-encoded gene RNA seq of total (x-axis) and polyA selected RNA seq reads in mature leaf (ML - blue), immature xylem (IX - yellow) and phloem (PH - orange). b. Log2 read count of mitochondrial encoded gene RNA seq of total (x-axis) and polyA selected RNA seq reads in mature leaf (ML - blue), immature xylem (IX - yellow) and phloem (PH - orange). For all correlations, the Pearson R^2 value is shown on the graph next to the linear trendline.

a.


Plastid
Number of edits per gene
0
2
4
6
8
10
12
accD
atpF
clpP
matK
ndhA
ndhB.2
ndhD
ndhG
ndhI
psaI
psbB
psbE
psbF
psbL
psbN
rpl22
rpl23.2
rpoB
rpoC1
rps2
rps12
rps14
ycf1.2
ycf2.2
Mito
Number of edits per gene
0
10
20
30
40
50
atp1
atp4
atp6
atp8
atp9
CcmB
ccmC
ccmFc
ccmFn
cob
cox1
cox2
cox3
matR
mttB
nad2
nad3
nad4L
nad5
nad6
nad7
nad9
rpl2
rpl5
rpl10
rps1
rps3
rps4
rps7
rps10
rps12
rps13
rps14
sdh3
sdh4

b.


0
0.4
0.8
Proportion of
edited reads
ML1
ML2
ML3
IX1
IX2
IX3
PH1
PH2
PH3

c.

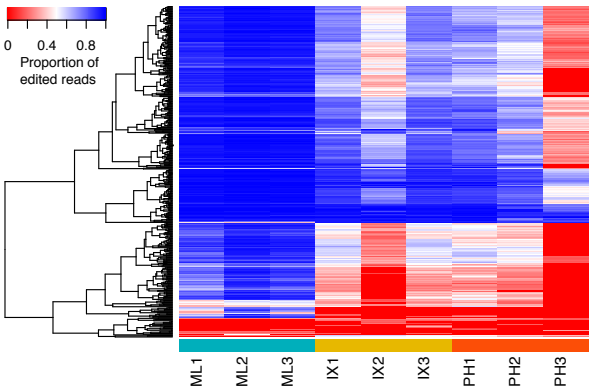

0
0.4
0.8
Proportion of
edited reads
ML1
ML2
ML3
IX1
IX2
IX3
PH1
PH2
PH3

Figure S4.5. Summary of identified editing sites in the plastid and mitochondrial genomes of *E. grandis* in mature leaf, immature xylem, and phloem.

a. Number of confirmed editing sited per gene in the plastid (top – dark grey) and mitochondria (bottom – light grey).
b. Heat map of plastid transcript editing sites. Individual plastid editing sites are coloured according to proportion of reads that are edited, ranging from 1 (blue – completely edited) to 0 (red – completely unedited) in three biological replicates each of mature leaf (ML - blue), immature xylem (IX - yellow), and phloem (PH - orange). Gene names are shown at the right-hand side of the heatmap.
c. Heat map of mitochondrial transcript editing sites Individual mitochondrial editing sites are coloured according to proportion of reads that are edited, ranging from 1 (blue – completely edited) to 0 (red – completely unedited) in three biological replicates each of mature leaf (ML - blue), immature xylem (IX - yellow), and phloem (PH - orange). Due to the number of editing sites, the gene names are omitted, and can be found in Table S5.

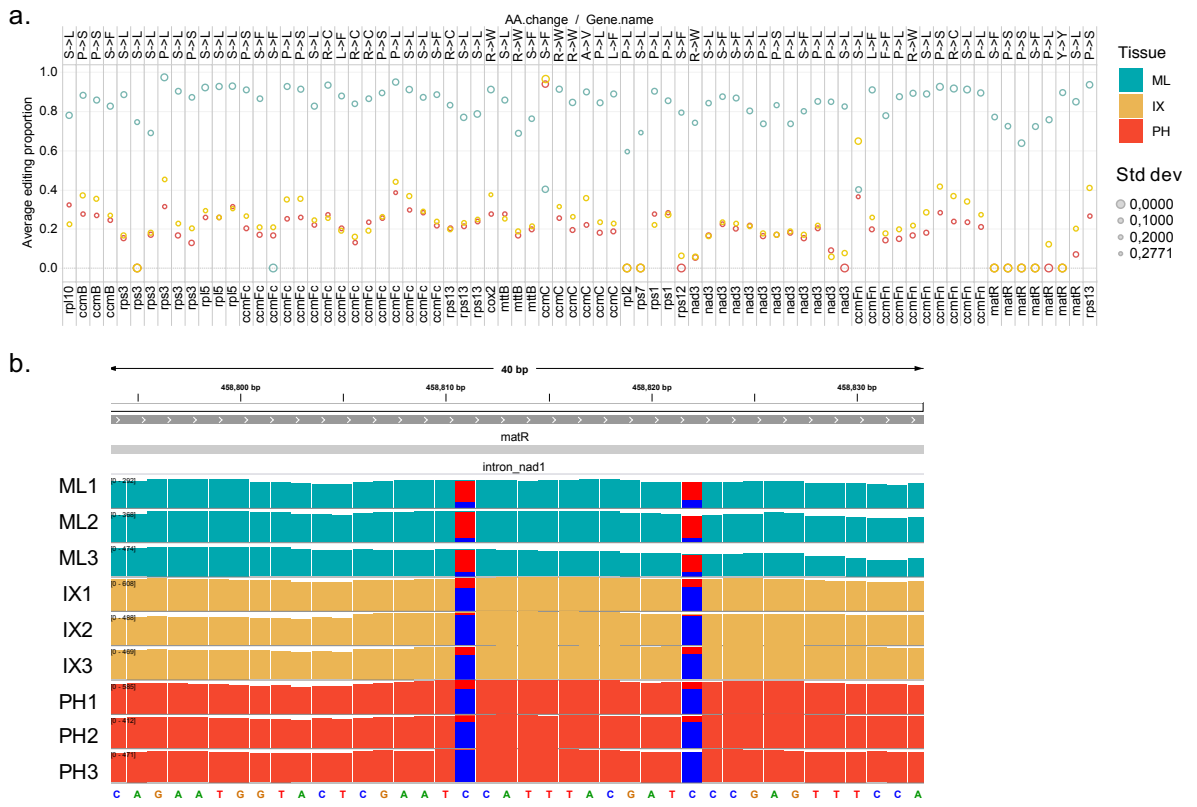


Figure S4.6. Mitochondrial editing and read alignment

a. Editing sites per mitochondrial gene showing the percentage of sites edited in mature leaf (blue), immature xylem (yellow), and phloem (orange). Sites were filtered to contain at least one tissue where all samples had editing frequency > 0, and the values shown are the average per tissue. Mitochondrial editing sites are further filtered if they had a difference of greater than 50% in mature leaf compared to immature xylem, and if they had a difference of greater than 20% in immature xylem compared to mature leaf. The size of the circles represent the standard deviation of three biological replicates per tissue, ranging from 0 (large circle) to 0.2771 in mitochondria (small circles).

b. Read alignment across all nine samples of forward read pairs visualized in IGV for a region of mitochondrial *matR* showing decreased editing in immature xylem and phloem, where red indicates edited reads, and blue unedited reads. Note the variation between IX2 and PH3 compared to other samples in those tissues, which were filtered out as stated above.

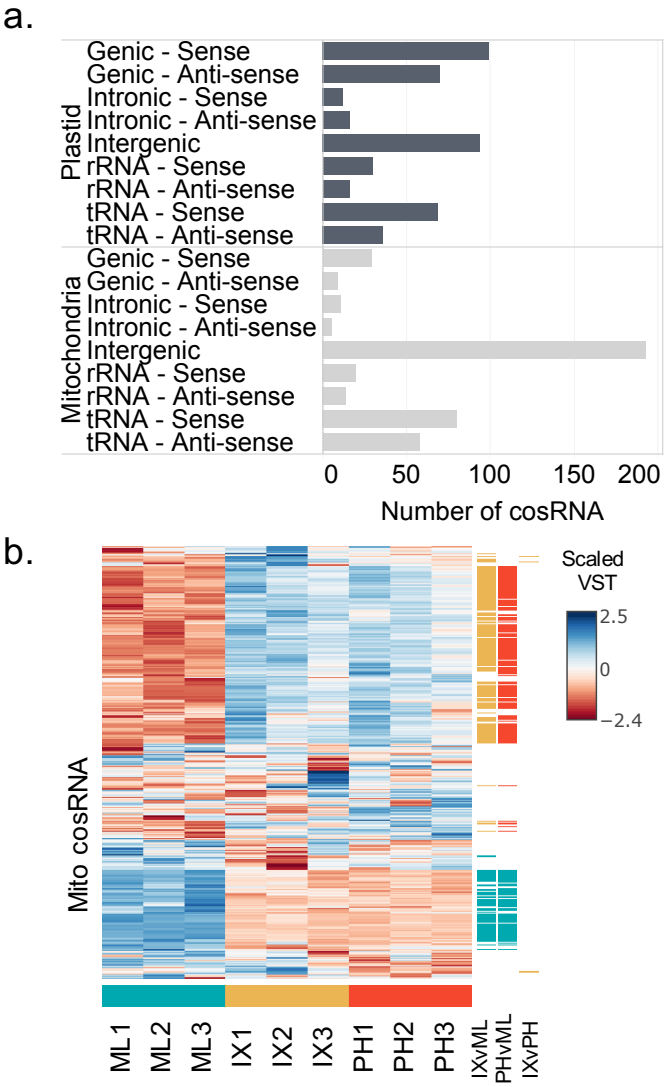


Figure S4.7. Plastid and mitochondrial cosRNA genomic feature overlap and abundance in *E. grandis* mitochondria

a. Number of *E. grandis* plastid and mitochondrial cosRNA per genomic feature. cosRNAs overlapping features were identified with bedtools intersect using the annotated plastid and mitochondrial genomes of *E. grandis*.

b. Heatmap of mitochondrial cosRNA abundance and significant differential abundance *Eucalyptus grandis* mature leaf (blue), phloem (orange), and immature xylem (yellow). cosRNA abundance is represented as row mean

centred VST values from low (red) to high (blue). Significantly differentially abundant cosRNA between tissues are indicated in the bars to the right of the heatmap.

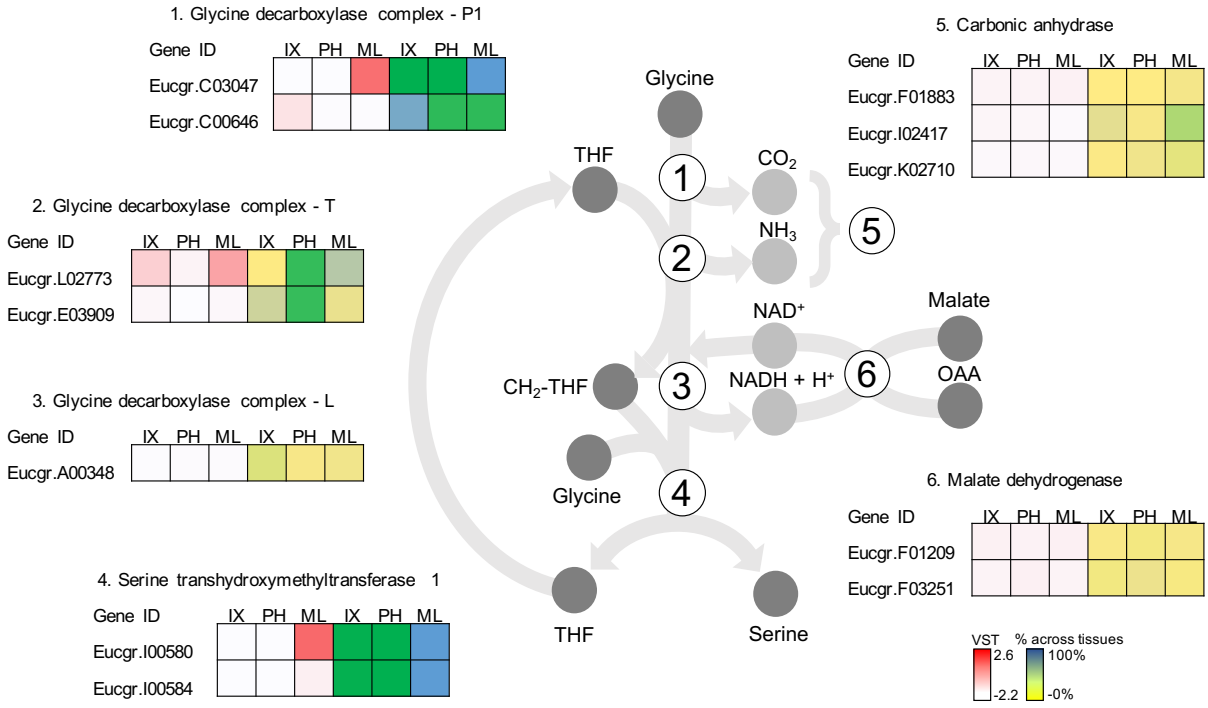


Figure S4.8. Biochemical pathway and tissue specific expression of *E. grandis* mitochondrial targeted photorespiration genes.

Average FPKM values for immature xylem (IX), phloem (PH), and mature leaf (ML) are shown from high (red) to low (white), and as a percentage of total expression between tissues from 100% (blue) to 0% (yellow) in order to identify tissue specific homologs. Reactions are numbered from 1 to 6, indicated in white circles, and numbered as such on the tables. In the mitochondrial localized reactions of photorespiration, two molecules of glycine are used to form serine and recycle tetrahydrofolate by the combined action of the glycine dehydrogenase complex (GDC - enzymes 1 to 3) and serine transhydroxymethyltransferase (SHM1 - enzyme 4). These reactions are buffered by carbonic anhydrase, and malate dehydrogenase. All genes shown are the mitochondrial targeted orthologs determined using PLAZA 3.0. Figure adapted from Bykova *et al*., 2014 and Bauwe *et al*., 2010.

4.13. Supplementary data files

Supplementary file 4.1 - *E. grandis* plastid gene expression quantification

Excel spreadsheet of plastid-encoded gene expression in *E. grandis* mature leaf (ML), immature xylem (IX), and phloem (PH) quantified as raw reads mapped, TPM, and VST.

Supplementary file 4.2 - *E. grandis* mitochondrial gene expression quantification

Excel spreadsheet of mitochondrial encoded gene expression in *E. grandis* mature leaf (ML), immature xylem (IX), and phloem (PH) quantified as raw reads mapped, TPM, and VST.

Supplementary file 4.3 - *E. grandis* plastid and mitochondrial differentially expressed genes

Excel spreadsheet of significantly differentially expressed plastid and mitochondrial genes between *E. grandis* tissues mature leaf (ML), immature xylem (IX), and phloem (PH), where significantly differentially expressed genes have log2 fold change > 2 and adjusted p-value < 0.05.

Supplementary file 4.4 - Plastid and mitochondrial encoded transcript editing sites across tissues in *E. grandis*

Excel spreadsheet of editing sites in the *E. grandis* plastid and mitochondrial encoded genes showing genomic and gene position, amino acid change, and proportion of edited sites between mature leaf (ML), immature xylem (IX), and phloem (PH) as a percentage of total reads mapped at that position. Sites were classified as edited by the REDO algorithm if read depth > 100, and % of edited reads > 10%.

Supplementary file 4.5 - Polyadenylated reads as a % of total read count in *Eucalyptus grandis* plastid and mitochondrial encoded genes

Excel spreadsheet of polyA-captured RNA-seq read count as a percentage of total RNA-seq read counts of *E. grandis* plastid and mitochondrial encoded genes between tissues.

Supplementary file 4.6 - *E. grandis* plastid and mitochondrial cosRNA position, orientation to closest genomic feature, and abundance across tissues

Excel spreadsheet showing *E. grandis* organellar cosRNA genomic position, orientation relative to genes, strandedness compared to closest genomic feature, and abundance across tissues, with differentially abundant transcripts between immature xylem (IX) and mature leaf (ML) indicated as in Supplementary file 3.

Supplementary file 4.7 - *E. grandis* plastid and mitochondrial targeted gene expression across tissues, and gene expression of known and potential plastid targeted RNA binding proteins

Excel spreadsheet of plastid and mitochondrial targeted gene expression, where differentially expressed genes are shown as in Supplementary file 3. RNA binding proteins (known and predicted) targeted to the plastid are also shown, taken from ChloroSeq (Castandet *et al*, 2016) and Reichel *et al*., 2016.

4.14. Supplementary Methods

qPCR of plastid and mitochondrial genes in *E. grandis* mature leaf, xylem, and phloem

Background

Due to the cost and technical difficulty associated with total RNA sequencing, a QuantStudio 12K flex panel of 168 assays was designed to facilitate future quantification of organellar gene expression. In this study, this panel was used to quantify organellar genome copy number using qPCR (See Udy *et al.*, 2012 for a study using the same primers for qPCR and RT-qPCR).

Methods and Materials

1. Primer design - Control genes

Nuclear control genes - To design the nuclear control genes, available RNAseq experiment were used to determine the stability of nuclear genes using co-efficient of variation (cv). The cv was calculated using the formula: cv = (Standard Deviation / Mean) * 100. The experiments included 156 *E. urophylla* xylem tissue samples (EUR), 126 *E. grandis* xylem tissue samples (EGR), and three replicates per tissue of six different *E. grandis* tissue samples (TISSUE) (Table 1). Additionally, control genes were chosen based on their absolute expression level relative to all other genes (>75th percentile), and if they were present as single copy genes in the genome (Myburg *et al.*, 2014).

Table 1:Final nuclear control genes with mean and cv shown per experiment.

gene_id	mean_Egr	cv_Egr	mean_Eur	cv_Eur	mean_Tissue	cv_Tissue
EUCGR.A00860	2.56E+06	13.74	2.13E+06	12.94	1.56E+06	13.29
EUCGR.A02027	3.65E+06	13.89	3.36E+06	12.52	1.37E+06	14.40
EUCGR.B01797	1.35E+07	14.79	1.55E+07	11.73	8.16E+06	9.19
EUCGR.B03019	7.39E+06	9.61	7.78E+06	12.28	4.71E+06	13.87
EUCGR.C01102	1.77E+06	11.27	1.54E+06	12.41	1.25E+06	15.33
EUCGR.F04124	1.22E+06	11.73	1.20E+06	12.28	7.86E+05	15.65
EUCGR.G03085	3.61E+06	14.20	4.07E+06	12.98	2.22E+06	13.75
EUCGR.K01846	1.66E+06	14.53	1.49E+06	12.45	9.17E+05	15.38
EUCGR.K02078	4.40E+06	13.95	4.65E+06	12.72	2.57E+06	15.35
EUCGR.K02619	1.08E+06	14.61	1.11E+06	15.46	5.43E+05	15.82
EUCGR.L00637	5.08E+06	11.75	5.42E+06	12.77	3.00E+06	13.75

The control genes were chosen to assay gene expression, not gene structure, and therefore 4 out of the 10 genes chosen did not amplify from DNA (Black in Table 1). This is probably due to intron structure which was not taken into account in the primer design, and cannot be evaluated here as the primer sequences are proprietary to Thermo Fischer Scientific.

2. Primer design - organellar-encoded genes

For the design of the organellar-encoded genes, a number of criteria were used to maximize the accuracy of the designed probes, and to ensure that the panel could be used for future experiments in the Forest Molecular Genomics programme. First, the probes had to avoid SNPs between the *E. grandis* (Pinard *et al.*, 2019b - Chapter 3) and *E. urophylla* organellar genomes (unpublished). Secondly, the probes were designed to be within exons, not flanking introns. Thirdly, the probes needed to avoid predicted editing sites in the organellar genomes (Pinard *et al.*, 2019b - Chapter 3).

Lastly, the probes had to avoid overlapping with transferred regions to the nuclear genome (Pinard *et al*., 2019b - Chapter 3), which was especially difficult for the mitochondrial genome.

Given these constraints, the final number of organellar probes designed were as follows:

- 75 unique plastid-encoded genes
- 6 technical replicates of *ycf2*, *psbL*, *clpP*, *accD*, *matK* and *ycf1* - same gene, different probe
- 3 Probes for the additional exons of *rps12*, *ndhB*, and *clpP*
- 26 mitochondrial genes

To complete the 168-assay panel, 48 nuclear genes of interest were chosen for their differential expression in xylem, targeting to plastid, and inclusion in the plastid targeted network constructed previously (Pinard *et al*., 2019a - Chapter 2).

DNA extraction and PCR confirmation of organellar DNA extraction

DNA extraction from frozen and finely ground tissues was done using the NucleoSpin II Plant Kit (Machery-Nagel) according to the manufacturer's instruction. gDNA presence was confirmed using gel electrophoresis, and was assessed for quality and concentration using the nanodrop instrument (Table 2). Next, PCR analysis using primers spanning genes in the plastid and mitochondria, respectively, was done to confirm that organellar DNA was extracted along with nuclear DNA.

Table 2: Nanodrop results for DNA extraction from three biological replicates of *E. grandis* mature leaf (ML), phloem (PH), and immature xylem (IX).

Sample	ng/μL	260/280	260/230
ML 1	422.72	1.83	2
ML 2	243.45	1.89	2.22
ML 3	253.9	1.89	2.26
PH 1	107.55	1.89	2.07
PH 2	128.15	1.89	1.98
PH 3	108.74	1.95	2.1
IX 1	86.62	1.93	1.91
IX 2	77.08	1.93	2.09
IX 3	99.53	1.94	2.07

The gel electrophoresis results are shown below in Figure 1.

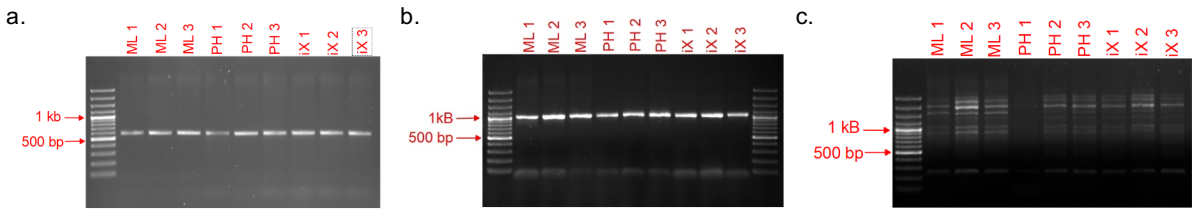


Figure 1. Gel electrophoresis confirmation of plastid and mitochondrial DNA extraction in *E. grandis* mature leaf (ML), immature xylem (IX), and phloem (PH). a. Positive control for nuclear gene *pipB* across tissues. b. Plastid DNA confirmation using primers flanking the *accD* gene. c. Mitochondrial DNA confirmation using primers flanking mitochondrial *cox1* and *rps10*. Ladder in all three gels is the GeneRuler 100bp ladder. The banding pattern in the mitochondrial primer set is due to the multiple mitochondrial DNA insertions into the nuclear genome of *E. grandis*. Future experiments will include primers designed to avoid these regions.

Table 3: Primer sequences used to confirm plastid and mitochondrial DNA extraction:

	pipB	*cox1 - rps10*	*accD*
Forward	5' gcgctcgtgactctcatgtg '3	5' tccagtgccgatcgacacaaga '3	5' tttcaggagtggcttgcttt '3
Reverse	5' gtatgctgctgctgcgagtg '3	5' agcagccgcggtaatacaga '3	5' accatctctgaacaaggttct '3

qPCR using the QuantStudio12K Flex platform

The qPCR experiment was performed using the QuantStudio (QS) 12K Platform according to the manufacturer's instruction. Each sample was analysed with three technical replicates, and an interplate control (ML1) and a negative control (distilled water instead of sample DNA) was included for each assay panel.

In brief, the data was analyzed as follows: Raw data were downloaded (.ado files) from the computer associated with the QS instrument. These files were uploaded to the Thermo Fischer cloud dashboard (https://apps.thermofisher.com/apps/dashboard/#/), and the preliminary data analysis and amplification curve quality control was performed. A table of cq values for all samples was downloaded for analysis in R (version 3.4.4). In R, assays with no amplification in target, and amplification in negative control were removed. The remaining control genes were used to get a reference cq value for each sample (average of controls). The reference gene average cq was subtracted from the value of each target gene per sample, giving the delta cq value. Relative organellar gene copy number was calculated using the formula: delta cq = 2^(-delta cq). For each class of gene (plastid, plastid repeat, and mitochondrial) all values per sample were averaged. Finally, statistical analysis to assess significant

variance from mature leaf copy number was performed following the procedure laid out in http://www.sthda.com/english/wiki/one-way-anova-test-in-r.

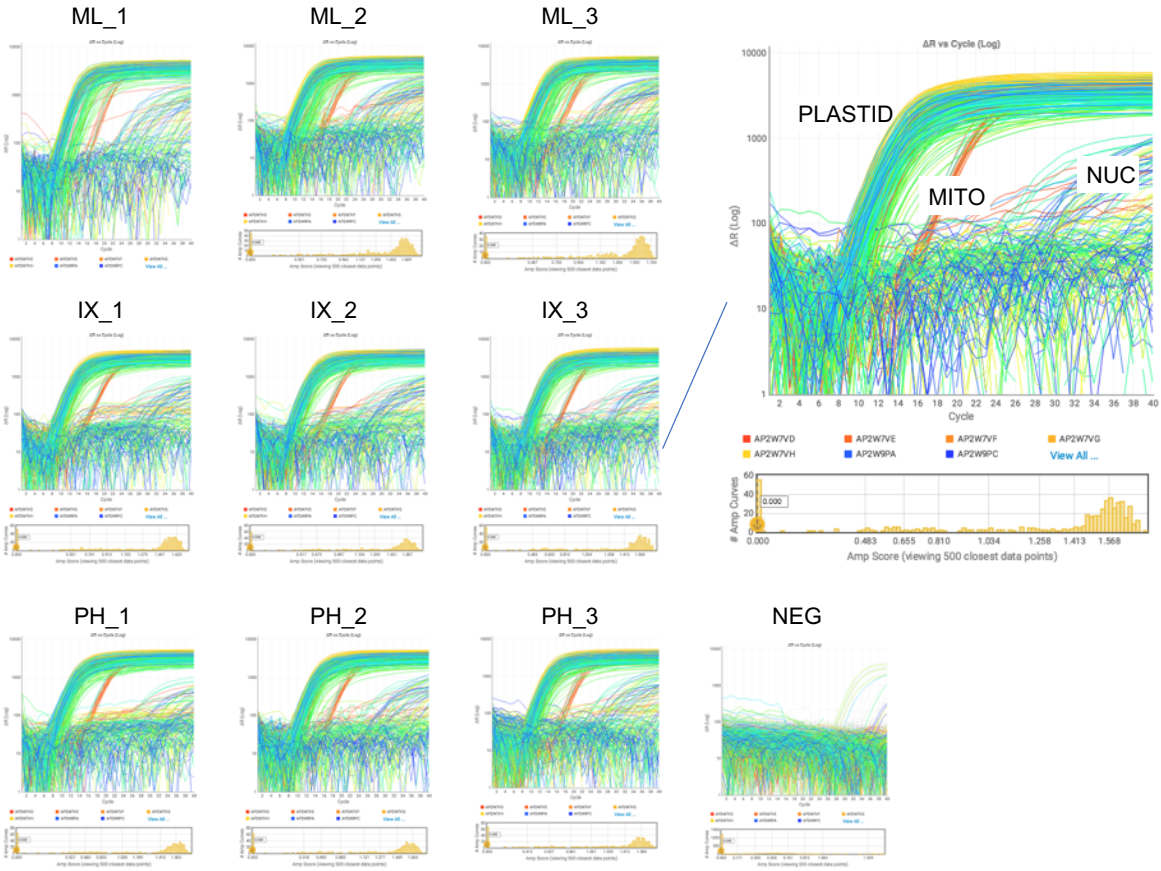


Figure 2. Figure showing qPCR amplification curves from the Thermo Fischer cloud dashboard for each sample and the negative control. Insert shows the amplification curves for sample IX3, with the plastid, mitochondrial and nuclear probes indicated. Probes with amplification in the negative control are removed during data processing.

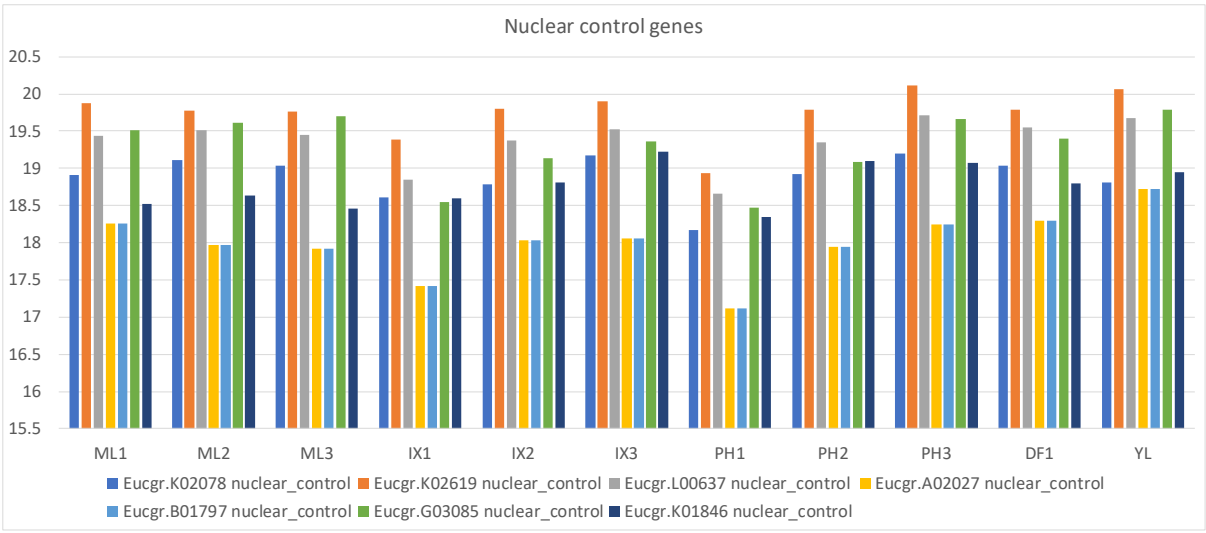


Figure 3. *E. grandis* nuclear control gene cq value across samples, including *Eucgr.B01797*, which was removed due to amplification in the negative control.

To determine if the means for all assays vary significantly between tissues, an ANOVA and Tukeys Honest Significant Differences test implemented in R version3.4.4 for mitochondria, plastid, and plastid repeats was performed (Table 4).

Table 4: Results of Tukeys HSD between tissues.

	Tukey's multiple comparisons of means 95% family-wise confidence level				
	Comparison	diff	lwr	upr	p adj
Mito	ML-IX	-114.18	-137.88	-90.49	**0**
	PH-IX	-10.01	-33.7	13.68	0.5789
	PH-ML	104.17	80.48	127.87	**0**
Plastid	ML-IX	31.18	11.17	51.19	**0.0008**
	PH-IX	-6.28	-26.29	13.73	0.7414
	PH-ML	-37.46	-57.47	-17.45	**0**
Plastid_rep	ML-IX	43.54	1.78	85.3	**0.03893**
	PH-IX	-21.94	-63.69	19.82	0.42664
	PH-ML	-65.48	-107.24	-23.72	**0.00093**

Chapter 5

5.1. Concluding remarks

Forests play a major role in the global carbon cycle, sequestering ~2 billion tonnes of atmospheric carbon a year (http://www.fao.org/state-of-forests/en/). Sustainable forestry has the potential to offset the effect of industrial carbon emissions, and provide renewable forms of energy and products towards the Sustainable Development Goals of the United Nations (Timko *et al.*, 2018). In addition to classical breeding, biotechnology strategies to improve biomass production in forest tree crops are in development to enhance their value to the bio-economy (Mottiar *et al.*, 2016; Van de Wouwer & Boerjan, 2018; Chang *et al.*, 2018; Grattapaglia *et al.*, 2018; Tuskan *et al.*, 2018). Forest trees partition a large amount of their sequestered carbon as cellulosic and phenolic biopolymers in the secondary xylem during xylogenesis, representing a major carbon sink in the environment (Plomion *et al.*, 2001; Zhong *et al.*, 2019). Ongoing research aims to model xylogenesis as a biological system to rationally design breeding strategies and biotechnology solutions in forest tree species for the bio-economy (Vanholme *et al.*, 2012; Mizrachi & Myburg, 2016).

Organellar biology is a critical aspect of plant development that has barely been explored in the study of wood formation, even though monolignol precursors are known to be synthesized by the plastidial shikimate pathway, and the energy requirements for xylogenesis are met by the mitochondria (Henkes *et al.*, 2001; Millar *et al.*, 2008). Additionally, plastids and mitochondria are intricately linked to multiple aspects of plant cell biology that extend beyond cellular respiration, photosynthesis,

and phenylalanine biosynthesis - many of which are crucial to xylogenesis. For example, the circadian rhythm of plants is regulated by signals originating from the plastid, mediated by nuclear-encoded receptors, which have been shown to play a role in xylogenesis, and the entrainment of tissue-specific circadian clocks in plant species (Thain, 2002; Solomon *et al.*, 2010; Jones, 2018). Additionally, plastids and mitochondria integrate environmental signals, and relay these signals to the nuclear genome, activating cellular responses to environmental changes and stress (Kmiecik *et al.*, 2016). Plastid localized *de novo* fatty acid biosynthesis, which in dicots includes a plastid-encoded gene, is a frequently overlooked carbon sink in developing xylem that is essential for cell division and elongation (Sasaki & Nagano, 2004). Mitochondrial localized tricarboxylic acid cycle supplies carbon skeletons and reducing agents necessary for xylogenesis, and mitochondrial reactive oxygen species play an important role in the developmentally programmed cell death responsible for xylem vessel and fibre cell maturation (Yu *et al.*, 2002; Millar *et al.*, 2008).

Land plant plastids and mitochondria are generally inherited maternally, with a few well-known exceptions; such pine species that exhibit paternal inheritance of the plastid (Neale & Sederoff, 1989; Nagata, 2010; Wicke *et al.*, 2011). This uniparental (mono-allelic) inheritance presents novel considerations in the study of systems biology in terms of the interactions between the organellar genomes and the nuclear copies of alleles inherited from the seed parent. Allele-specific interactions, epistasis, and trans-eQTLs are of particular interest when examining hybrid plants, where the organellar genome is strictly uniparental, yet varying proportions of the nuclear genome are derived from parents that may be different species. Surprisingly, given

the size and coding potential of organellar genomes, organellar genome variation can have large effects on the metabolism and fitness of a plant (Budar & Roux, 2011; Joseph *et al.*, 2013, 2015; Roux *et al.*, 2016). Organellar genome variation in hybrid populations of *Arabidopsis* can account for a large proportion of metabolite variation, often having a larger effect on metabolite variation than individual nuclear loci (Joseph *et al.*, 2013). Organellar genome inheritance and variation have also been shown to impact adaptive, fitness, and growth traits in several plant species (Wright *et al.*, 2009; Joseph *et al.*, 2013; Bock *et al.*, 2014; Budar & Roux, 2016). A recent study has found that a biomass increase of 23% from the optimal combination of nuclear and organellar genomes is possible in *Arabidopsis* (Flood *et al.*, 2018). Organellar genomes and their coordinated evolution with the nuclear genome is thus a key aspect to be understood in the context of vascular plant development, as well as in the generation of hybrid genetically diverse and/or clonal forest trees for industrial applications (Weng *et al.*, 2016).

The vast majority of the proteins localized to plastids and mitochondria are encoded by the nuclear genome, and for the most part, the nuclear genome regulates organellar biology (Weihe *et al.*, 2012). Interorganellar signalling coordinates synthesis of protein complexes that have subunits encoded by the organellar and nuclear genomes by organellar retrograde signalling, and nuclear-encoded proteins that regulate organellar transcription and translation (Woodson & Chory, 2008). In light of the complex interactions between the nuclear and organellar genomes that govern many aspects of xylem metabolism, understanding organellar biology during wood formation will

improve our understanding of the system as a whole (Kersten *et al.*, 2016; Pinard & Mizrachi, 2018).

Some fundamental questions that need to be addressed to begin integrating organellar biology into the systems biology model of xylogenesis are: To what extent, and how, are the carbon metabolic processes targeted to plastids and mitochondria co-regulated with the biological process of xylogenesis (Chapter 2)? Are recent lateral genome transfers of organellar genes to the nuclear genome functional, and do they exhibit any tissue-specific expression patterns (Chapter 3)? How are organellar-encoded genes regulated during xylogenesis, and how do nuclear-encoded genes coordinate organellar biology within tissues (Chapter 4)? This thesis aimed to address these questions and provide resources for the future study of xylogenic organelles, including a qPCR panel designed to rapidly and accurately assess organellar gene expression and organellar DNA copy number in two commonly used *Eucalyptus* commercial plantation species globally, *E. grandis* and *E. urophylla.*

The co-expression networks generated in Chapter 2 describe the emergent co-expression structure of plastid and mitochondrial carbon metabolism. The network structure is driven primarily by genetic variation in regulatory factors, as well as temporal variation with diurnal rhythms that are specifically associated with the plastid. This is distinct from co-expression networks derived from experimental datasets over multiple conditions and tissues, or a developmental series in a single tissue type, and reveals new architecture that may be missed using those methodologies (Serin *et al.*, 2016). For example, we provide xylem functional context for 197 plastid-targeted

genes with no discernible tissue-specific differential expression, highlighting both temporal and genetically linked genes, many of which have unknown functions (see Chapter 3, Notes S3). Together with the *Populus tremula* xylem developmental series of Sundell *et al* (2017), we find that primary metabolic pathways such as protein and nucleotide synthesis are associated with morning-specific diurnal clock genes, perhaps in preparation for secondary cell wall formation (Sundell *et al.*, 2017). Novel insight into plastid-targeted carbon metabolism was gained by examining the temporal regulation of carbon flux between *de novo* fatty acid biosynthesis and the shikimate pathway. The plastid-targeted network showed that the recently discovered role of plastids in epigenetic regulation may also be important during xylogenesis (Beltrán *et al.*, 2018).

In Chapter 3, we assembled and annotated the mitochondrial genome of *E. grandis*, and provided a high-quality annotation for the plastid genome of *E. grandis*. The annotated genomes were then used to characterize the genome dynamics of *E. grandis* by quantifying DNA and gene transfer between nuclear and organellar genomes. This work has highlighted the transfer of organellar DNA and intact coding regions of organellar genes into the *E. grandis* nuclear genome, and the minimal role of polyadenylation in organellar-encoded gene regulation. Further, we quantified the expression of intact organellar gene copies in the nuclear genome and concluded that nuclear-encoded transcripts of organellar genes are not functionally expressed. Thus, extra-organellar transcription or significant variation in transcript steady-state levels due to polyadenylation could be ruled out as factors impacting organellar transcription in xylem, leaf, and phloem.

Finally, in Chapter 4 the expression and post-transcriptional regulation of organellar-encoded genes were analyzed across the source, sink, and transport tissues of *E. grandis*; and nuclear regulators of xylem organellar biology were identified. We have shown that organellar genes in *E. grandis* tissues are differentially regulated in phototrophic *vs*. heterotrophic tissues at the transcriptional and post-transcriptional level, and have identified signatures of active organellar gene upregulation in wood forming tissues. A major finding from this work is the xylem and phloem-specific upregulation of genes transcribed by the Nuclear Encoded Polymerase (NEP), including the highly expressed *ycf2* gene, which has recently been associated with protein transport across the inner plastidial membrane (Kikuchi *et al.*, 2018). Together with the identification of RNA-binding protein footprints, we show that NEP expressed plastid-encoded genes in xylem and phloem are *actively* upregulated compared to mature leaf, suggesting a distinct function for non-photosynthetic plastids in secondary growth wood formation.

In this thesis, a number of experimental techniques were utilized and developed to ensure the accuracy of the conclusions drawn, as there are very few related studies of tissue-specific organellar biology in plants for comparison (aside from potato amyloplasts and tomato chromoplasts) (Kahlau & Bock, 2008; Valkov *et al.*, 2009). Whenever possible, future research should aim to include: (i.) Genomic DNA sequences to identify DNA variants from organellar editing analyses; (ii.) SNP aware alignment of RNA-sequencing paired with organellar editing predictions to decrease the permissible number of mismatches allowed when performing read alignment; and

(iii.) Simultaneous alignment of any generated RNA sequencing data to the nuclear and organellar genomes to reduce spurious read alignment, similar to the best practice for dual RNA sequencing of host-pathogen interactions.

In Pinard & Mizrachi (2018), we proposed that the secondary growth xylem plastid may be functionally and morphologically distinct from other types of leucoplasts, and suggest that this plastid be termed the “xyloplast”. A limitation to this thesis is the fact that could not directly assess the role of plastids in xylem fibre development, as xylem fibre cell lignification may be due to cell non-autonomous lignification by “good neighbour” parenchyma cells (Smith *et al.*, 2013, 2017; De Meester *et al.*, 2018). This may be studied in the future using inducible cell-type-specific knock-downs of plastid-targeted and encoded genes to determine the relative contribution of parenchyma plastids to xylem fibre cell lignification in trees such as *Populus* and *Eucalyptus* (Smith *et al.*, 2017). Further, the mix of cell types inherently present in the samples we define as “mature leaf”, “immature xylem” and “phloem”, complicates the analysis of organellar function as specific organelle populations are found in different cell types (Beltrán *et al.*, 2018). An example of this is the sensory plastid, which is found in the epidermal cells of *Arabidopsis* leaves but has a distinct proteome and function compared to mesophyll cell chloroplasts (Beltrán *et al.*, 2018). This may also impact the genome copy number estimates from the qPCR analysis, where specific cell types may have dramatically different genome copy numbers (Greiner *et al.*, 2019), and we have only identified the average per tissue. Additionally, the qPCR analysis performed in Chapter 4 was not suitable for quantifying sub-genomic mitochondrial DNA molecules and will be repeated using a dilution series to account for qPCR probe

inhibition according to the MIQE guidelines (Huggett *et al.*, 2019). Another limitation to this study is the lack of direct comparisons of experimental procedures for evaluating genome copy number - organellar DNA isolation vs whole-cell DNA isolation; and gene expression between plants - qPCR based vs RNA sequencing (Nakamura *et al.*, 2003; Castandet *et al.*, 2013; Golczyk *et al.*, 2014; Ma & Li, 2015). Future work will include using the qPCR panel designed in this work to assess organellar transcript abundance for comparison to the RNA sequencing values (Chapter 4, Supplementary Method 1), and quantitative proteomics to determine the effects of post-transcriptional regulation on organellar protein abundance.

From the available literature, there is very little consensus of what defines separate plastid types beyond morphological differences. For secondary growth xylem and phloem plastids in wood, updated microscopy and biochemical staining will be critical in validating the xyloplast and its function (Cronshaw & Wardrop, 1964). In the plant organellar research community, there may be a need to better evaluate the biological differences between plastid types beyond morphology, and how they relate to transcriptional and post-transcriptional regulation between tissues. This may be especially important since more evidence is emerging that plastid-targeted and encoded genes can function differently between photosynthetic and heterotrophic tissues. Specific examples include CUE1, which functions as a net importer or exporter of phosphoenolpyruvate, depending on tissue type (Staehr *et al.*, 2014), and pdNAD-MDH which utilizes different protein domains for specific functions in photosynthetic and heterotrophic metabolism (Schreier *et al.*, 2018).

A limitation to Chapter 2 and Chapter 4 is the lack of *Eucalyptus* - and tree - specific functional characterization of genes. In the case of large gene families, functional differences may occur between species, especially in the case of duplicated genes in *Eucalyptus*. The assumption of functional conservation between species is an inherent concern in any large-scale study in non-model organisms, which rely on homology and automatically generated functional annotations. These annotations are in some cases incomplete, outdated, or outright incorrect (Reviewed in Valencia, 2005). In Chapter 2, this was addressed in part by the comparative analyses with the *P. tremula* xylem developmental series (Sundell *et al.*, 2017) and the *Eucalyptus* eQTL:phenotype associations (Mizrachi *et al.,* 2017). These analyses showed that the broader functional conclusions drawn from the networks are supported, even though the fine-scale resolution of gene functions may be different in *Eucalyptus* compared to the species in which the gene was functionally characterised.

An interesting follow up to this research will be to further investigate the function of *ycf2* in non-green plastids, including comparative transcriptomic analysis in the secondary growth xylem of other tree species. As *ycf2* mutants are lethal in dicots, a strategy for studying YCF2 function in wood formation could be to identify the RNA-binding protein of the *ycf2* cosRNA footprint identified in Chapter 4 by RNA-protein pull-down assay (Iioka *et al.*, 2011). If the RNA-binding protein can be identified, tissue-specific knock-out or overexpression of the hypothetical protein may lead to a range of YCF2 protein levels and an observable phenotype (Kim *et al.*, 2017). This strategy is similar to one used by Yu *et al.* (2019) to increase transgene expression in non-photosynthetic plastids by engineering a PRR10 protein that activates plastid-

encoded gene expression (Yu *et al.*, 2019). Additionally, researchers have identified cis-elements for increased expression of transgenes specifically in non-photosynthetic plastids to avoid photosynthetic yield penalty (Mellor *et al.*, 2018), and the upstream region of *ycf2* may be a candidate promoter for transgene expression to specifically modify the shikimate pathway in wood forming tissues.

In addition to the function of *ycf2* and other highly expressed plastid-encoded genes in secondary growth wood tissues, many interesting questions have arisen from the work presented in this thesis. From the association of circadian clock-related genes in the plastid-targeted gene network specifically, to what extent are plastid-encoded and targeted genes diurnally regulated in wood formation? How is diurnal gene expression variation related to the temporal regulation of fatty-acid biosynthesis and the shikimate pathway, given that the fatty-acid biosynthetic gene *accD* is encoded by the plastid genome? Additionally, although mitochondrial-targeted metabolism was not specifically associated with the circadian clock, does the expression of mitochondrial glycine metabolic genes reflect the diurnal variation of lignin biosynthesis in light of the finding that the serine-glycine interconversion pathway supports lignification (Villalobos *et al.*, 2012)? At the tissue level, how is phloem carbon metabolism in plastids regulated, considering that sieve elements maintain their organelles after enucleation (Knoblauch *et al.*, 2018) and that the phloem transport stream is a relatively high sugar environment (Furze *et al.*, 2018)? Given our scant understanding, even today, of ray and axial parenchyma cells in angiosperm secondary growth, do parenchyma cells have distinct organellar populations compared to fibre and vessel cells – i.e. how is organellar metabolism regulated at the level of individual cell type?

The recent success in generating transplastomic *Arabidopsis* lines should stimulate research into the effects of manipulating plastid gene expression in the plastids of non-green tissues (Ruf *et al.*, 2019). As with most proposed biotechnological interventions in living organisms, care must be taken in interpreting the results of any perturbation of organellar functions. For example, transgene expression to modify terpene synthesis showed that the local metabolite pools in plastids may play a large role in the success of engineering plastid-targeted pathways, as transgene expression from the plastid, nucleus, or both, had different metabolic effects (Pasoreck *et al.*, 2016).

Since the first application of next-generation RNA sequencing in *Eucalyptus* was performed (Mizrachi *et al.*, 2010), and the release of the *E. grandis* genome (Myburg *et al.*, 2014), *Eucalyptus* has rapidly become a model genus of tree species for systems biology research in woody biomass tissue (Ribeiro *et al.*, 2016; Mizrachi *et al.*, 2017; Ployet *et al.*, 2019; Wierzbicki *et al.*, 2019). With this work, *Eucalyptus* becomes the only woody tree genus to date to have protein-coding, small RNA, and genome copy number data available for the organellar genomes in multiple tissues, giving novel insight into the role of organelles during xylogenesis, and the specific adaptation of plastids during the evolution of angiosperm wood formation.

5.2. References

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

Budar F, Roux F. **2011**. The role of organelle genomes in plant adaptation: time to get to work! *Plant Signalling & Behavior* **6**: 635–639.

Castandet B, Hotto AM, Fei Z, Stern DB. **2013**. Strand-specific RNA sequencing uncovers chloroplast ribonuclease functions. *FEBS Letters* **587**: 3096–3101.

Chang S, Mahon EL, MacKay HA, Rottmann WH, Strauss SH, Pijut PM, Powell WA, Coffey V, Lu H, Mansfield SD, *et al.* 2018. Genetic engineering of trees: progress and new horizons. *In Vitro Cellular & Developmental Biology - Plant* **54**: 341–376.

Cronshaw J, Wardrop AB. **1964**. The organization of cytoplasm in differentiating xylem. *Australian Journal of Botany* **12**: 15–23.

De Meester B, de Vries L, Özparpucu M, Gierlinger N, Corneillie S, Pallidis A, Goeminne G, Morreel K, De Bruyne M, De Rycke R, *et al.* 2018. Vessel-specific reintroduction of CINNAMOYL-COA REDUCTASE1 (CCR1) in dwarfed *ccr1* mutants restores vessel and xylary fibre integrity and increases biomass. *Plant Physiology* **176**: 611–633.

Flood PJ, Tom P J, Schneeberger K, Keizer P, Kruijer W, Severing E, Kouklas E, Hageman JA, Becker FFM, Schnabel SK, *et al.* 2018. Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *bioRxiv*: https://doi.org/10.1101/477687.

Furze ME, Trumbore S, Hartmann H. **2018**. Detours on the phloem sugar highway: stem carbon storage and remobilization. *Current Opinion in Plant Biology* **43**: 89–95.

Golczyk H, Greiner S, Wanner G, Weihe A, Bock R. **2014**. Chloroplast DNA in mature and senescing leaves: a reappraisal. *The Plant Cell* **26:** 847-54.

Grattapaglia D, Silva-Junior OB, Resende RT, Cappa EP, Müller BSF, Tan B, Isik F, Ratcliffe B, El-Kassaby YA. **2018**. Quantitative genetics and genomics converge to accelerate forest tree breeding. *Frontiers in Plant Science* **9**: 1693-1699.

Greiner S, Golczyk H, Malinova I, Pellizzer T, Bock R. **2019**. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. *bioRxiv*.

Henkes S, Sonnewald U, Badur R, Flachmann R, Stitt M. **2001**. A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *The Plant Cell* **13**: 535-541.

Huggett J, Nolan T, Bustin SA. **2019**. MIQE: Guidelines for the design and publication of a reliable real-time PCR Assay. *Polymerase Chain Reaction: Theory and Technology*.

Iioka H, Loiselle D, Haystead TA, Macara IG. **2011**. Efficient detection of RNA–protein interactions using tethered RNAs. *Nucleic Acids Research* **39**: e53–e53.

Jones MA. **2018**. Retrograde signalling as an informant of circadian timing. *New Phytologist* **221:** 1749-1753.

Joseph B, Corwin JA, Li B, Atwell S, Kliebenstein DJ. **2013**. Cytoplasmic genetic variation and extensive cytonuclear interactions influence natural variation in the metabolome. *eLife* **2**: e00776.

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.

Kahlau S, Bock R. **2008**. Plastid transcriptomics and translatomics of tomato fruit development and chloroplast-to-chromoplast differentiation: chromoplast gene expression largely serves the production of a single protein. *The Plant Cell* **20**: 856–874.

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.

Kikuchi S, Asakura Y, Imai M, Nakahira Y, Kotani Y, Hashiguchi Y, Nakai Y, Takafuji K, Bédard J, Hirabayashi-Ishioka Y, *et al.* 2018. A Ycf2-FtsHi heteromeric AAA-ATPase complex is required for chloroplast protein import. *The Plant Cell* **30**: 2677–2703.

Kim J, Le Nguyen T-N, Kang H. **2017**. Artificial targeting of a nucleus-encoded RNA-binding protein AtRZ1a to chloroplasts affects flowering and ABA response of *Arabidopsis thaliana. Journal of Plant Biology* **60**: 278–284.

Kmiecik P, Leonardelli M, Teige M. **2016**. Novel connections in plant organellar signalling link different stress responses and signalling pathways. *Journal of Experimental Botany*: erw136.

Knoblauch M, Peters WS, Bell K, Ross-Elliott TJ, Oparka KJ. **2018**. Sieve-element differentiation and phloem sap contamination. *Current Opinion in Plant Biology* **43**: 43–49.

Ma J, Li XQ. **2015**. Organellar genome copy number variation and integrity during moderate maturation of roots and leaves of maize seedlings. *Current Genetics* **61**: 591–600.

Mellor SB, Behrendorff JBYH, Nielsen AZ, Jensen PE, Pribil M. **2018**. Non-photosynthetic plastids as hosts for metabolic engineering. *Essays in Biochemistry* **62**: 41–50.

Millar AH, Small ID, Day DA, Whelan J. **2008**. Mitochondrial biogenesis and function in *Arabidopsis*. *The Arabidopsis Book : American Society of Plant Biologists* **6**: e0111.

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**: 681–693.

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.

Mottiar Y, Vanholme R, Boerjan W, Ralph J, Mansfield SD. **2016**. Designer lignins: harnessing the plasticity of lignification. *Current Opinion in Biotechnology* **37**: 190–200.

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.

Nagata N. **2010**. Mechanisms for independent cytoplasmic inheritance of mitochondria and plastids in angiosperms. *Journal of Plant Research* **123**: 193–199.

Nakamura T, Furuhashi Y, Hasegawa K, Hashimoto H, Watanabe K, Obokata J, Sugita M, Sugiura M. **2003**. Array-based analysis on tobacco plastid transcripts: preparation of a genomic microarray containing all genes and all intergenic regions. *Plant & Cell Physiology* **44**: 861–867.

Neale DB, Sederoff RR. **1989**. Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theoretical and Applied Genetics* **77**: 212–216.

Pasoreck EK, Su J, Silverman IM. **2016**. Terpene metabolic engineering via nuclear or chloroplast genomes profoundly and globally impacts off-target pathways through metabolite signalling. *Plant Biotechnology* **14** 1862-1875.

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

Plomion C, Leprovost G, Stokes A. **2001**. Wood formation in trees. *Plant Physiology* **127**: 1513–1523.

Ployet R, Veneziano Labate MT, Regiani Cataldi T, Christina M, Morel M, San Clemente H, Denis M, Favreau B, Tomazello Filho M, Laclau J-P, *et al.* 2019. A systems biology view of wood formation in *Eucalyptus grandis* trees submitted to different potassium and water regimes. *The New Phytologist*. doi.org/10.1111/nph.15802.

Ribeiro T, Barrela RM, Bergès H, Marques C, Loureiro J, Morais-Cecílio L, Paiva JAP. **2016**. Advancing *Eucalyptus* genomics: Cytogenomics reveals conservation of *Eucalyptus* genomes. *Frontiers in Plant Science* **7**: 510.

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.

Ruf S, Forner J, Hasse C, Kroop X, Seeger S, Schollbach L, Schadach A, Bock R. **2019**. High-efficiency generation of fertile transplastomic *Arabidopsis* plants. *Nature Plants* **5**: 282–289.

Sasaki Y, Nagano Y. **2004**. Plant Acetyl-CoA Carboxylase: Structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Bioscience, biotechnology, and biochemistry* **68**: 1175–1184.

Schreier TB, Cléry A, Schläfli M, Galbier F, Stadler M, Demarsy E, Albertini D, Maier BA, Kessler F, Hörtensteiner S, *et al.* 2018. Plastidial NAD-dependent malate dehydrogenase: a moonlighting protein involved in early chloroplast development through its interaction with an FtsH12-FtsHi protease complex. *The Plant Cell* **30**: 1745–1769.

Serin EAR, Nijveen H, Hilhorst HWM, Ligterink W. **2016**. Learning from co-expression networks: Possibilities and challenges. *Frontiers in Plant Science* **7**: 444-453.

Smith RA, Schuetz M, Karlen SD, Bird DA, Tokunaga N, Sato Y, Mansfield SD, Ralph J, Samuels LA. **2017**. Defining the diverse cell populations contributing to lignification in *Arabidopsis thaliana* stems. *Plant Physiology* **174**: 1028-1036.

Smith RA., Schuetz M, Roach M, Mansfield SD, Ellis B, Samuels L. **2013**. Neighboring parenchyma cells contribute to *Arabidopsis* xylem lignification, while lignification of interfascicular fibres is cell autonomous. *The Plant Cell* **25**: 3988–3999.

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.

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**: 126.

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.

Thain SC. **2002**. The circadian clock that controls gene expression in *Arabidopsis* Is tissue specific. *Plant Physiology* **130**: 102–110.

Timko J, Le Billon P, Zerriffi H, Honey-Rosés J, de la Roche I, Gaston C, Sunderland TCH, Kozak RA. **2018**. A policy nexus approach to forests and the SDGs: tradeoffs and synergies. *Current Opinion in Environmental Sustainability* **34**: 7–12.

Tuskan GA, Groover AT, Schmutz J, DiFazio SP, Myburg AA, Grattapaglia D, Smart LB, Yin T, Aury J-M, Kremer A, *et al.* 2018. Hardwood tree genomics: Unlocking woody plant biology. *Frontiers in Plant Science* **9**: 1799.

Valencia A. 2005. Automatic annotation of protein function. *Current Opinion in Structural Biology* 15: 267–274.

Valkov VT, Scotti N, Kahlau S, Maclean D, Grillo S, Gray JC, Bock R, Cardi T. **2009**. Genome-wide analysis of plastid gene expression in potato leaf chloroplasts and tuber amyloplasts: transcriptional and posttranscriptional control. *Plant Physiology* **150**: 2030–2044.

Van de Wouwer D, Boerjan W. **2018**. Plant cell wall sugars: Sweeteners for a bio-based economy. *Physiologia Plantarum* **164**: 27-44.

Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W. **2012**. A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *The Plant Cell* **24**: 3506–3529.

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.

Weihe A, Liere K, Börner T. **2012**. Transcription and transcription regulation in chloroplasts and mitochondria of higher plants. In: Bullerwell CE, ed. *Organelle Genetics: Evolution of Organelle Genomes and Gene Expression*. Berlin, Heidelberg: Springer Berlin Heidelberg, 297–325.

Weng M-L, Ruhlman TA, Jansen RK. **2016**. Plastid-nuclear interaction and accelerated coevolution in plastid ribosomal genes in Geraniaceae. *Genome Biology and Evolution*: evw115.

Wicke S, Schneeweiss GM, dePamphilis CW, Müller KF, Quandt D. **2011**. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76**: 273–297.

Wierzbicki MP, Christie N, Pinard D, Mansfield SD, Mizrachi E, Myburg AA. **2019**. A systems genetics analysis in *Eucalyptus* reveals coordination of metabolic pathways associated with xylan modification in wood forming tissues. *The New Phytologist*. doi.org/10.1111/nph.15972.

Woodson JD, Chory J. **2008**. Coordination of gene expression between organellar and nuclear genomes. *Nature Reviews Genetics* **9**: 383–395.

Yu Q, Barkan A, Maliga P. **2019**. Engineered RNA-binding protein for transgene activation in non-green plastids. *Nature Plants* **5**: 486–490.

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.

Zhong R, Cui D, Ye Z-H. **2019**. Secondary cell wall biosynthesis. *The New Phytologist* **221**: 1703–1723.