

EXPLORING THE REGULATORY MECHANISMS BEHIND CHLOROPLAST
POPULATION MORPHOLOGY

By

Emily Jennings Tallerday

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Biology—Doctor of Philosophy
Molecular Plant Sciences—Dual Major

2022

PUBLIC ABSTRACT

EXPLORING THE REGULATORY MECHANISMS BEHIND CHLOROPLAST POPULATION MORPHOLOGY

By

Emily Jennings Tallerday

Plants are the primary food source for many organisms. Therefore, it is crucial to understand how plants obtain their food. Plants are green because their leaf cells contain an abundance of chloroplasts—small compartments that can capture light energy and convert it into food through a process known as photosynthesis. Most crops and other plants that live under constant direct sunlight have many very small chloroplasts, which can move around within the leaf and avoid damage that might occur after prolonged exposure to intense light. To generate these large populations within leaf cells, chloroplasts reproduce by dividing at the middle, much like bacteria. My work focuses on how chloroplast replication is controlled. Plants with fewer chloroplasts generally do not perform as well at capturing light energy and may be more susceptible to damage from intense amounts of light. I have laid the groundwork for establishing two *Peperomia* plant species as models for studying this regulation. *Peperomia*, also known as radiator plants, are popular houseplants due to their diverse ornamental foliage. This diversity extends within the leaf, as some *Peperomia* species have drastically different numbers and sizes of chloroplasts in their leaves. To understand how these physical differences might occur, I have studied two species—one with a few very large chloroplasts in its cells and the other with a much larger population of very small chloroplasts per cell.

ABSTRACT

EXPLORING THE REGULATORY MECHANISMS BEHIND CHLOROPLAST POPULATION MORPHOLOGY

By

Emily Jennings Tallerday

Eukaryotic cells tightly regulate their populations of endosymbiotically-derived organelles. Organelle populations can be described in terms of size, number, or coverage, the latter being the collective planar area taken up by the organellar population relative to that of the cell. As the photosynthetic organelle, chloroplasts are vital, and alterations to chloroplast population morphology can affect photosynthetic performance and biomass accumulation. However, how the cell perceives and regulates its chloroplast population remains a mystery. Division at the mid-plastid (binary fission) is the primary mechanism by which chloroplasts increase their population sizes. It has been well established that lower division rates result in a small population of enlarged chloroplasts, suggesting the existence of a compensatory mechanism ensuring that total chloroplast coverage within the cell is preserved through a tradeoff between chloroplast division and expansion.

Most model plants keep a relatively large number of chloroplasts in their leaf cells (>50 per cell). In expanding leaf cells, multiple rounds of chloroplast division typically increase the number of chloroplasts per cell. However, a number of natural adaptive alterations to chloroplast morphology have been observed in several tropical plant species, primarily those native to low-light environments. The tropical plant genus *Peperomia* (*Piperaceae*) offers a unique opportunity for understanding the regulation of chloroplast population morphology, as some *Peperomia* spp. contain two to six giant chloroplasts in their palisade mesophyll cells at maturity, while most others have higher numbers of small chloroplasts in their mesophyll cells.

I have characterized chloroplast population morphology in *Peperomia*, of which six species had not been studied previously, and shown that chloroplast division is inhibited in the palisade cells of *Peperomia pellucida*. Further, I have assembled and annotated the genome of *Peperomia dahlstedtii*, the first genome for this genus, and produced a novel transcriptome assembly for *P. pellucida*. Lastly, I have analyzed gene expression in these two species differing in palisade cell chloroplast population morphology and identified several candidate genes potentially underlying the differences in phenotype. For the first time, I also have described the expression of the chloroplast division genes in these two species. By characterizing variation in chloroplast population morphology in *Peperomia*, my work builds upon existing research on this trait over leaf development, provides the resources necessary for *Peperomia* to be used as a model, and identifies potential causes behind the large-chloroplast phenotype documented in several species.

This work is dedicated to each individual that committed their time to teach, mentor, and encourage me during my education. Your belief in me allowed me to believe in myself.

Christian A. Burr, I would not have made it here without you. You are an incredible mentor, and I am thrilled to call you my friend.

Joe Kieber and Gloria Muday, you both provided me with wonderful opportunities early on in my career. Your enthusiasm for science is inspiring, and I will never forget working in your labs.

Kathy, you took the time to speak with me and helped me navigate through an excellent program that I would not have known otherwise. Thank you for allowing me to take us both a little out of our comfort zones—this project has been lots of fun.

Thank you all so much for everything you have done and continue to do.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
KEY TO SYMBOLS AND ABBREVIATIONS	xi
Chapter 1. The Green Knight’s challenge: towards leveraging natural variation in chloroplast population morphology to understand this fundamental cell-biological trait	1
Introduction	2
Chloroplast replication and maintenance in the leaf	4
Chloroplast size and number are tightly controlled	5
Alterations to chloroplast size and number	9
Known genetic loci controlling chloroplast size and number.....	9
Variation in nature	11
Key points and moving forward.....	13
REFERENCES	16
Chapter 2. Chloroplast population morphology and division in the tropical plant genus <i>Peperomia</i>	24
Introduction	25
Results	27
Chloroplast population morphology in the mature leaves of nine <i>Peperomia</i> species.....	27
Quantifying chloroplast population morphology in nine <i>Peperomia</i> species.....	28
Chloroplast division during leaf development in <i>P. dahlstedtii</i> and <i>P. pellucida</i>	30
Discussion	35
Conclusions	39
Materials and methods.....	39
Plant materials, growth conditions, and propagation.....	39
Sampling of the leaf developmental series	40
Preparation and imaging of leaf tissue.....	41
Data manipulation, plots, and statistical analyses.....	42
Acknowledgments	42
REFERENCES	45
Chapter 3. Establishing novel <i>de novo</i> transcriptome assemblies for <i>Peperomia dahlstedtii</i> and <i>Peperomia pellucida</i>	49
Introduction	50
Results	51
Generation and assessment of novel <i>de novo</i> transcriptome assemblies for <i>P. pellucida</i> and <i>P. dahlstedtii</i>	51
Representation of the chloroplast division orthogroups	55

GIANT CHLOROPLAST 1 appears to be mutated to different degrees in <i>Peperomia</i> species	60
Identification of a potential gene of interest behind the ‘large chloroplast’ phenomenon in vascular plants	61
Discussion	68
Conclusions	70
Materials and methods.....	71
Sampling tissue for sequencing	71
RNA extraction and transcriptome sequencing	71
De novo transcriptome assembly	71
Assessing the quality and completeness of transcriptome assemblies.....	72
Assessment of transcriptome assembly quality and completeness	72
Orthogroup clustering, copy number, and expression	72
Multiple sequence alignments and protein domain identification	73
Data and statistical analyses.....	73
Data availability	73
Acknowledgments	74
REFERENCES	76

Chapter 4. Assembly and utilization of the first <i>Peperomia dahlstedtii</i> genome for differential expression analyses	80
Introduction	81
Results	82
Genome size, complexity, and the selection of representative species in <i>Peperomia</i>	82
Establishing the first genome assembly for the genus <i>Peperomia</i>	84
Exploring gene duplication in <i>P. dahlstedtii</i>	90
Generation of novel transcriptomic data over leaf development in <i>P. dahlstedtii</i> and <i>P. pellucida</i>	95
Revisiting BolA1 and GC1 as candidate regulators of chloroplast population morphology	96
Intra-species differential expression analyses: with a focus on the chloroplast division genes	103
Inter-species differential orthogroup expression	109
Discussion	114
Conclusions	117
Materials and methods.....	118
Plants and growth conditions	118
Genome size, ploidy, and heterozygosity estimates	118
DNA extraction, sequencing, and read QC.....	118
RNA extraction, sequencing, and read QC	120
Genome assembly and annotation	120
Genome-guided transcriptome assemblies	122
Orthogroup clustering, phylogenetic analyses, and multiple sequence alignments.....	123
Differential gene expression	124
Orthogroup expression.....	124
Data manipulation, plots, and statistical analyses.....	125

Data availability	125
Acknowledgments	125
REFERENCES	128
Chapter 5. Future directions	136
Potentially immediate-future experiments in <i>Peperomia</i> to characterize the inhibition of chloroplast division on the molecular level.....	137
FtsZ and ARC3 protein levels and immunolocalization.....	137
Regarding BolA1 and GC1 as candidate regulators of chloroplast population morphology in <i>P. pellucida</i>	139
Molecular characterization of the leaf developmental stages established for <i>P. dahlstedtii</i> and <i>P. pellucida</i>	139
On continuing with <i>Peperomia</i> as a model system	140
Moving forward: Single cell RNA sequencing and cytokinin signaling in rice	141
REFERENCES	143
APPENDICES	145
Appendix I. Establishing novel methods for the quantification of chloroplast coverage toward identifying regulators of this trait via natural variation.....	146
REFERENCES	155
Appendix II. Chloroplast and proplastid division in large-chloroplast-containing <i>Peperomia</i> species	158
REFERENCES	163

LIST OF TABLES

Table 3.1. Basic statistics for the <i>de novo</i> transcriptome assemblies.	53
Table 3.2. Alignment rates of paired end RNAseq reads to their respective <i>de novo</i> transcriptome assembly.....	53
Table 3.3. Spearman’s correlation tests between RNAseq replicates.....	56
Table 3.4. <i>BolAI</i> was not detected in species with a small population of large chloroplasts per cell.....	67
Table 4.1. Somatic cell genome size estimates derived from young leaf tissue by flow cytometry.	83
Table 4.2. Samples used for next-generation sequencing & read mapping statistics.	87
Table 4.3. Genome assembly statistics for <i>Peperomia dahlstedtii</i>	88
Table 4.4. Repeat content in the <i>Peperomia dahlstedtii</i> genome.	89
Table 4.5. Assessment of completeness in <i>Peperomia dahlstedtii</i> using BUSCOs.....	91
Table 4.6. Transcript annotation statistics for <i>Peperomia dahlstedtii</i>	91
Table 4.7. Genome-guided transcriptome assembly statistics.	97
Table 4.8. Assessment of completeness in <i>P. dahlstedtii</i> and <i>P. pellucida</i> genome-guided transcriptome assemblies using BUSCOs ¹	97
Table 4.9. Number of orthogroups significantly different in expression between species.	114

LIST OF FIGURES

Figure 1.1. Chloroplast division in <i>A. thaliana</i>	7
Figure 2.1. Chloroplast population morphology in <i>Peperomia</i>	29
Figure 2.2. Correlation between the number of chloroplasts per cell and planar cell area in <i>Peperomia</i>	31
Figure 2.3. Quantitative assessment of chloroplast population morphology in <i>P. pellucida</i> and <i>P. dahlstedtii</i>	32
Figure 2.4. Chloroplast division over leaf development in <i>P. pellucida</i> and <i>P. dahlstedtii</i>	36
Figure 3.1. Assessing the quality of the <i>de novo</i> transcriptome assemblies.....	54
Figure 3.2. Chloroplast division orthogroup expression and copy number.....	57
Figure 3.3. Possible MCD1 truncations within <i>Peperomia</i>	59
Figure 3.4. GC1 may be truncated, to varying extents, within <i>Peperomia</i>	62
Figure 3.5. <i>Peperomia BolA1</i> and the mutant phenotype of <i>AtbolA1</i>	65
Figure 4.1. Fixed metaphase chromosomes from four <i>Peperomia</i> species.....	83
Figure 4.2. WGS k-mer-coverage-based profile of <i>Peperomia dahlstedtii</i>	85
Figure 4.3. Gene duplication and heterozygosity in <i>P. dahlstedtii</i>	93
Figure 4.4. Orthogroup-based phylogeny and clustering of orthologous genes between <i>Peperomia dahlstedtii</i> , <i>Peperomia pellucida</i> , and representative vascular plants.....	98
Figure 4.5. <i>BolA1</i> is present in <i>P. pellucida</i>	99
Figure 4.6. Multiple protein sequence alignment for GC1.....	100
Figure 4.7. The ARC6/PARC6 clade.....	101
Figure 4.8. PARC6, if truly present in <i>Peperomia pellucida</i> , is severely truncated in critical domains.....	102
Figure 4.9. Normalization of expression across libraries and differential expression statistics.....	105
Figure 4.10. Differential expression of the chloroplast division genes in <i>P. dahlstedtii</i> and <i>P. pellucida</i> leaves over development.....	106

Figure 4.11. Distributions in orthogroup expression between species.	113
Figure II.1. Chloroplast numbers in guard cells from mature <i>P. dahlstedtii</i> and <i>P. pellucida</i> leaves.....	161
Figure II.2. Treatment with cytokinin increases the number of chloroplasts per cell in <i>P. metallica</i> leaves.....	161
Figure I.1. Natural variation in chloroplast coverage between <i>Arabidopsis thaliana</i> accessions.	148
Figure I.2. Concept and validation behind a new method for high-throughput measurement of chloroplast coverage.	150
Figure I.3. Chloroplast coverage is higher in palisade compared to spongy mesophyll cells of <i>P. dahlstedtii</i>	152
Figure I.4. Chloroplast coverage in mature palisade mesophyll cells from <i>P. dahlstedtii</i> and <i>P. pellucida</i>	153

KEY TO SYMBOLS AND ABBREVIATIONS

ARC	ACCUMULATION AND REPLICATION OF CHLOROPLASTS
%gc	Percent guanine-cytosine content
ALU	Alu element
FtsZ	Filamenting temperature-sensitive Z
GC1	GIANT CHLOROPLAST 1
IEM	Inner envelope membrane
IJM	ImageJ Macro Language
IMS	Inner membrane space
<i>ipt</i>	<i>Isopentenyl transferase</i>
LINE	Long Interspersed nuclear element
LTR	Long terminal repeat
MCD	MULTIPLE CHLOROPLAST DIVISION SITE
MIR	Mammalian-wide interspersed repeat
OEM	Outer envelope membrane
OG _{exp}	Orthogroup expression
PARC6	Paralog of ACCUMULATION AND REPLICATION OF CHLOROPLASTS 6
PDV	Plastid division
PE150	Paired-end 150 nucleotide long reads
REC	REDUCED CHLOROPLAST COVERAGE
RNAseq	RNA sequencing
Ruiz & Pav.	Ruiz y Pavón

SE50	Single-end 50 nucleotide long reads
SINE	Short interspersed nuclear element
WGS	Whole-genome sequencing
Z ring	Filamenting temperature-sensitive Z ring
$\Delta\text{OG}_{\text{exp}}$	The difference in orthogroup expression between species
$\Sigma\Delta\text{OG}_{\text{exp}}$	Cumulative difference in orthogroup expression between species

**Chapter 1. The Green Knight's challenge: towards leveraging natural variation in
chloroplast population morphology to understand this fundamental cell-
biological trait**

Introduction

Plant cells tightly regulate their plastid populations, and rightly so, as many essential functions are plastid-specific (e.g., photosynthesis, pigment production and storage, starch storage, gravity-sensing, etc.) (Pyke, 2009). However, how cells perceive and regulate the adjustment of their plastid populations remains a mystery (Honda et al., 1971; Ellis and Leech, 1985; Okie et al., 2016). Plastid population morphology can be described simply in terms of size, number, and coverage—the latter being the ratio of total plastid planar area to that of the cell area (Marshall, 2016; Okie et al., 2016). Alterations to chloroplast population morphology can affect photosynthetic performance, biomass accumulation (Pyke et al., 1994), and likely fitness (Külheim et al., 2002; Frenkel et al., 2006; Athanasiou et al., 2009). For example, remarkably large chloroplasts (discussed later) are known to experience higher rates of photodamage (Ghaffar et al., 2018) and are less photosynthetically efficient compared to those of ‘average’ size (Jotham and Webber, 2005; Weise et al., 2015; Dutta et al., 2017). Conversely, plants with populations of many small chloroplasts perform very well under high light (i.e., stressful) conditions (Jeong et al., 2002; Dutta et al., 2017; Xiong et al., 2017). While this phenomenon is somewhat related to photorelocation, as small chloroplasts may be better able to move around the cell (Jeong et al., 2002; Königer et al., 2008), there is more at play (Dutta et al., 2015; Dutta et al., 2017).

Division is the primary mechanism by which plastids replicate and increase their population sizes (Osteryoung and Pyke, 2014; Pyke, 2016). Chloroplasts divide by binary fission, yielding two daughter chloroplasts of equal size from a single parent, and all chloroplasts within a specific cell- or tissue-type are typically of similar size (Leech and Pyke, 1988; Osteryoung and Pyke, 2014). Reduced rates of chloroplast division have been shown to result in

an increase in individual chloroplast size (Butterfass, 1979; Pyke et al., 1994; Robertson et al., 1995; Osteryoung et al., 1998; Pyke, 1999). This well-documented phenomenon suggests the existence of a compensatory mechanism ensuring that total chloroplast coverage within the cell is maintained through a balance between chloroplast division and expansion (Butterfass, 1979; Osteryoung and Pyke, 2014).

Overall, small chloroplasts are better adapted for photo-stressful environments with fluctuation in/excess light. So why are there plants with large chloroplasts? Chloroplasts found in the leaves of shade-adapted and low-light-preferring plants are generally enlarged (and, thus, fewer in number), and their grana stacks tend to be thicker, wider, and more disorganized in appearance than those of high-light-adapted plants (Irmak, 1957; Bjorkman et al., 1971; Anderson et al., 1973; Tsel'niker, 1973; Boardman, 1977; Butterfass, 1979; Lichtenthaler et al., 1981; Machado et al., 1986; Nasrulhaq-Boyce and Duckett, 1991; Sheue et al., 2007; Liu et al., 2020). Recent work in *Selaginella* has shown that these features may be adaptive, optimizing light capture in environments with fluctuating indirect light (Liu et al., 2020). However, we do not know how this apparent tradeoff for larger chloroplast size over number is regulated in such plants.

In this thesis I have used species from the low light plant genus *Peperomia*, several of which have a small population of very large chloroplasts, to understand how the plant cell regulates individual chloroplast expansion over division. The work I have conducted and described in the following chapters builds upon our understanding of chloroplast population morphology in the primary photosynthetic tissues. I have characterized chloroplast population morphology in *Peperomia*, of which six species had not been studied previously, and shown that division is inhibited in the palisade cells of *Peperomia pellucida*. Further, I have assembled and

annotated the genome of *Peperomia dahlstedtii*, the first genome for this genus, and produced a novel transcriptome assembly for *P. pellucida*. Lastly, I have looked at gene expression between two species differing in palisade cell chloroplast population morphology and identified several candidate genes potentially underlying the differences in phenotype. For the first time, I also have described the expression of the chloroplast division genes in these two species. By characterizing variation in chloroplast population morphology in *Peperomia*, my work builds upon existing research on this trait during leaf development, provides the resources necessary for *Peperomia* to be used as a model, and identifies potential causes behind the large-chloroplast phenotype observed in several species. Overall, this work marks a path forward towards understanding the adjustment of chloroplast population morphology.

Here I begin with a review of chloroplast replication and maintenance over leaf development, as these are foundational concepts behind chloroplast population morphology. Variation in chloroplast population morphology is then discussed. Genetic loci known to affect chloroplast population morphology are reviewed, and their implications for natural variation are elaborated on.

Chloroplast replication and maintenance in the leaf

In plants, all plastids are derived from proplastids (plastid precursors) present in meristematic cells and gametes (Pyke, 2009). It is generally accepted that division is the primary mechanism by which both proplastids and chloroplasts replicate, though the exact proteins involved may differ slightly (Robertson et al., 1995; Pyke, 1997). The cell strictly controls the timing and frequency of plastid division (Honda et al., 1971; Cran and Possingham, 1972; Possingham, 1973; Butterfass, 1980; Boffey and Leech, 1982; Leech and Pyke, 1988; Pyke and Leech, 1992; Pyke, 1997). Proplastids differentiate into chloroplasts, amyloplasts, chromoplasts,

or other plastid types depending on cell fate (Pyke, 2009). Though plastid population size varies widely by cell type (described below), improper or insufficient proplastid division has been shown to produce differentiated cells apparently completely lacking plastids (i.e., aplastidic) (Butterfass, 1979; Chen et al., 2009; Pyke, 2009). This condition is generally thought to be lethal (Robertson et al., 1995; Pyke, 2009), as plastids perform several essential cellular functions beyond photosynthesis (i.e., starch storage and the synthesis of amino acids, lipids, hormones, etc.) (Pyke, 2009).

Chloroplast size and number are tightly controlled

Photosynthesis is primarily carried out in the leaf mesophyll cells, arranged in palisade and spongy mesophyll (Evert and Eichhorn, 2006). A mature spongy mesophyll cell contains 50-150 chloroplasts in most plants, though extreme cases have been reported (3-300 chloroplasts) (Butterfass, 1979; Evert and Eichhorn, 2006; Pyke, 2009). Epidermal, hypocotyl, and guard cells contain chloroplasts, though the individual chloroplast size and number per cell are much smaller (Butterfass, 1979; Pyke, 1997). This functionally-specific phenomenon is likely due to a combination of higher rates in chloroplast division and increased coverage in the primary photosynthetic cell layer(s) to maximize photosynthesis, as no other cell type contains as many plastids as mesophyll do chloroplasts (Pyke, 1997).

Leaf development can be loosely chronologically described in terms of cell differentiation, division, and expansion (Pyke et al., 1991; Evert and Eichhorn, 2006; Sakamoto et al., 2016). Upon illumination, the proplastids in leaf-destined cells develop into chloroplasts (Pyke, 2009). Expanding leaf mesophyll cells experience high rates of chloroplast division (Figure 1.1A-B; Honda et al., 1971; Possingham et al., 1988; Pyke, 2009). This results in a large

population (50-150) of relatively small chloroplasts (Butterfass, 1979; Evert and Eichhorn, 2006; Pyke, 2009).

The molecular process of chloroplast division has been well studied, primarily in *A. thaliana* (Figure 1.1C; Chen et al., 2018). Filamenting temperature-sensitive Z1 and Z2 (FtsZ1 and FtsZ2; tubulin-like GTPases) co-assemble inside the chloroplast stroma (light blue & green; Figure 1.1C; Osteryoung and Vierling, 1995; Osteryoung et al., 1998; Olson et al., 2010; Yoshida et al., 2016) and are herded to the mid-plastid (for binary fission) by ACCUMULATION AND REPLICATION OF CHLOROPLASTS (ARC; Pyke and Leech, 1992) 3 (ARC3; pink; Figure 1.1C), 11 (ARC11/MinD; dark blue; Figure 1.1C) and 12 (ARC12/MinE; dark grey; Figure 1.1C), in addition to membrane-bound MULTIPLE CHLOROPLAST DIVISION SITE 1 (MCD1; lavender; Figure 1.1C), (Colletti et al., 2000; Maple et al., 2007; Nakanishi et al., 2009; Miyagishima, 2011). ARC6 (yellow; Figure 1.1) interacts with FtsZ2, tethering the Z ring to the inner envelope membrane (IEM; Figure 1.1C; Maple and Møller, 2005). ARC6 and Paralog of ARC6 (PARC6; dark purple; Figure 1.1C) recruit PLASTID DIVISION 1 (PDV1; orange; Figure 1.1C) and 2 (PDV2; taupe; Figure 1.1C), respectively, to the outer envelope membrane (OEM) through interactions at the inner membrane space (IMS; light grey layer; Figure 1.1C) (Glynn et al., 2008; Wang et al., 2017). PDV1 and 2 help assemble the outer division ring by bringing ARC5 (also referred to as DRP5B; red; Figure 1.1C) to the OEM (Gao et al., 2003; Miyagishima, 2003). Lastly, PARC6 recruits ARC3 to the division site, facilitating Z-ring destabilization and remodeling, critical for chloroplast constriction (Zhang et al., 2016; Chen et al., 2019).

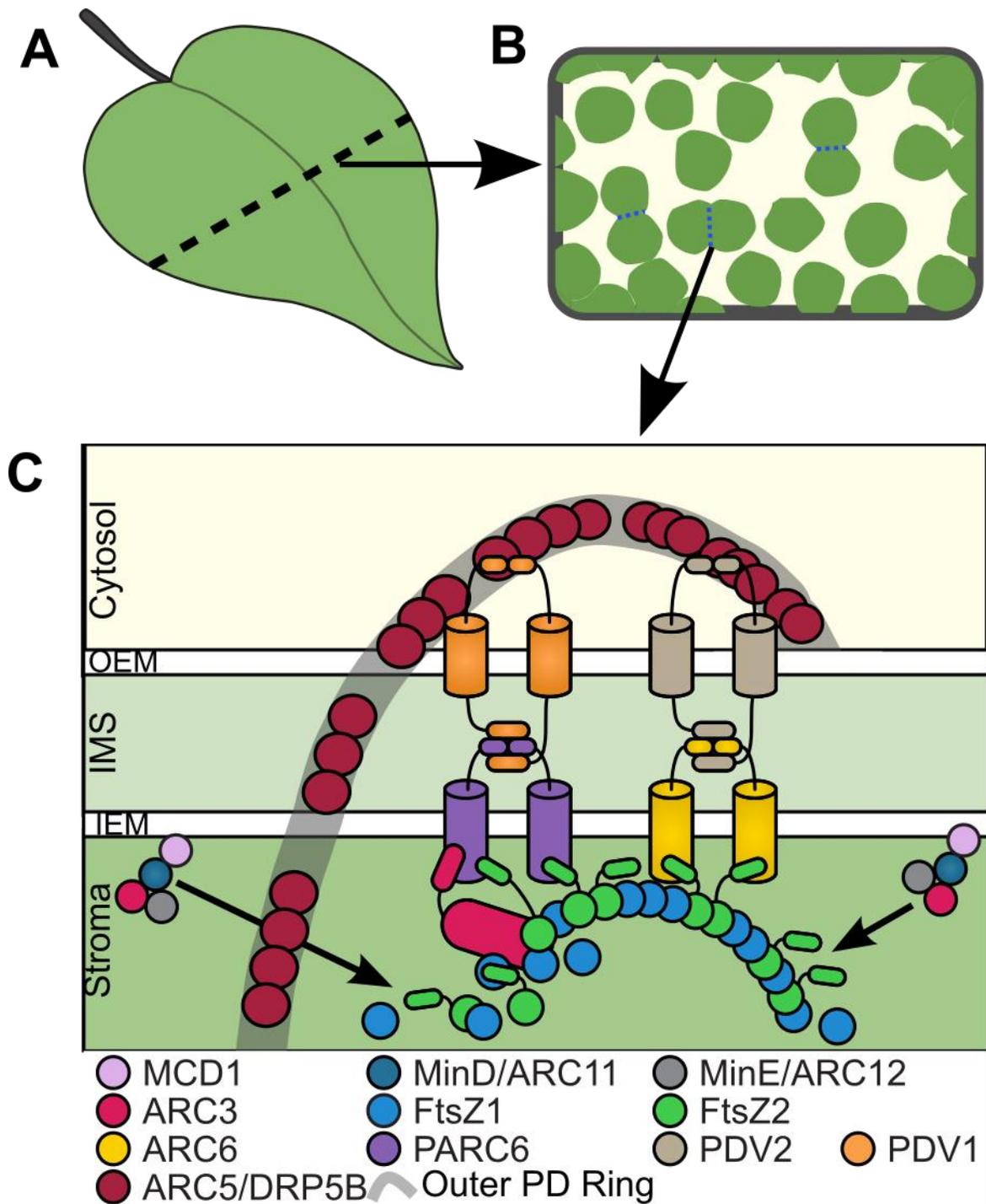


Figure 1.1. Chloroplast division in *A. thaliana*.

(A) The expanding leaf is the primary site of (B) chloroplast division and occurs at high rates in the mesophyll cells. Dashed blue lines indicate constriction sites. (C) The proteins involved in chloroplast division act in a coordinated system at the mid-plastid. Black arrows in C denote the targeting of Z-ring assembly to the mid-plastid by the Min system. OEM, outer envelope membrane; IMS, inner membrane space; IEM, inner envelope membrane. This figure was adapted from Chen et al. (2018).

Chloroplasts experience two distinct growth forms—elongation (i.e., lengthening) and equidistant expansion (Whatley, 1988). I consider these in two dimensions (area rather than volume) for simplicity. Just before division, the chloroplast elongates along one axis, changing from circular to ‘elongated peanut’ in shape (Leech and Pyke, 1988; Osteryoung and Pyke, 2014). The chloroplast assumes a ‘dumbbell’ shape during division due to constriction at the mid-plastid (Leech and Pyke, 1988; Gao et al., 2003). Immediately after fission, the daughter chloroplasts (similar in size) expand while maintaining a circular shape, ultimately doubling the area (Leech and Pyke, 1988). Chloroplast expansion can occur continuously (Whatley, 1988), though how this process is regulated is unknown.

Chloroplast coverage (i.e., ‘compartment size’ or ‘plastidome area’) is the ratio between the total area occupied by the chloroplast population to that of the cell area (Butterfass, 1979; Osteryoung and Pyke, 2014). While coverage can vary (Jellings et al., 1983; Larkin et al., 2016; Salesse et al., 2017), it is consistent by species and cell type—the leaf mesophyll cells of several plants maintain ~70% coverage (Pyke, 2009). In differentiated post-mitotic leaf mesophyll cells where cell expansion is ongoing (Evert and Eichhorn, 2006), chloroplast division rates increase to maintain coverage as the cellular area increases (Honda et al., 1971; Possingham et al., 1988; Pyke, 2009). As demonstrated in *A. thaliana* mutants for genes involved in chloroplast division, coverage is maintained by a tradeoff between chloroplast division and expansion (Pyke and Leech, 1992). Thus, if chloroplast division is inhibited, the plant will have a small population of enlarged chloroplasts. Despite its strict maintenance, we do not yet know how the cell senses or regulates the adjustment of this trait. Though the overall focus of my thesis is not on coverage, this trait is the subject of Appendix I and is further discussed there.

Alterations to chloroplast size and number

Most of what we know about the regulation of chloroplast population morphology is derived from studies on chloroplast division in *A. thaliana*. While we understand quite well how chloroplast replication occurs on the molecular level in *A. thaliana*, we know very little about how this mechanism might influence chloroplast population morphology in nature, or how the process of division is regulated itself. Below I begin with a summary of how the *A. thaliana* chloroplast division genes can influence chloroplast size and number. I then provide an overview of documented cases of natural variation in chloroplast population morphology and how they may relate to chloroplast division.

Known genetic loci controlling chloroplast size and number

As described above, chloroplast division is necessary in the leaves so that a large population of chloroplasts can be generated for the purpose of photosynthesis. The earliest mutants (*AtftsZ1* and *Atarc6*) for the chloroplast division genes identified, as described below, share very similar and distinct chloroplast morphology, where each cell contains a few very large chloroplasts. The FtsZs are required for constriction at the mid-plastid during division (Figure 1.1C); therefore, it is unsurprising that *AtftsZ* mutants contain a small population of enlarged chloroplasts. While the division is severely inhibited, chloroplast coverage is maintained by the continued expansion of the individual chloroplasts in these plants (Osteryoung et al., 1998; Schmitz et al., 2009). Interestingly, natural variation in chloroplast size between different *A. thaliana* accessions has been partially attributed to the truncation of *FtsZ2-2* (Kadirjan-Kalbach et al., 2019). It seems likely that reduced expression of any of the *FtsZs* could contribute to chloroplast expansion over division in natural plant systems.

A. thaliana arc6 exhibits the most extreme morphology of the division mutants, containing only 1-3 large chloroplasts per leaf mesophyll cell (Pyke et al., 1994; Robertson et al., 1995). It is thought that chloroplast division does not occur in *Atarc6*, as the number of chloroplasts per cell does not increase with cell expansion (Pyke et al., 1994; Robertson et al., 1995). Importantly, ARC6 is required to tether FtsZ2 to the IEM during Z-ring assembly (Figure 1.1C). Further, ARC6 may be necessary for proplastid replication, as ~30% of *Atarc6* mature guard cells were aplastidic—an extremely rare condition no matter the cell type (Robertson et al., 1995). While reduced expression of *ARC6* is a plausible mechanism for producing larger chloroplasts in natural plant populations, it may be less likely considering the potential role *ARC6* plays in proplastid division, as the cell aplastidic condition is extremely rare and thought to be lethal (Robertson et al., 1995).

Through direct interaction with the FtsZs, *ARC3* acts as a negative regulator of FtsZ assembly (Figure 1.1C). Independent overexpression of *ARC3* or *PARC6*, which interacts with *ARC3* at the constriction site to promote Z-ring disassembly (Figure 1.1C), produces large chloroplasts in *A. thaliana* (Maple et al., 2007; Zhang et al., 2013; Zhang et al., 2016; Chen et al., 2019). Interestingly, *Atparc6* plants overexpressing *ARC3* exhibit an extreme large chloroplast phenotype similar to that observed in *Atarc6* (Chen et al., 2019). Recent work in *A. thaliana* suggests that *PARC6* also interacts with MinD, which helps target Z-ring assembly to the mid-plastid, during chloroplast division in epidermal cells (Itoh et al., 2018). High expression of *ARC3* and or alterations to the expression of *PARC6* could therefore potentially explain the natural production of enlarged chloroplasts. Overall, it is likely that changes in the expression of more than one of the chloroplast division gene would contribute to naturally-derived large chloroplasts.

Moving away from the genes with clear roles in chloroplast division, an initial report observed that reduced expression of *A. thaliana* *GIANT CHLOROPLAST 1* (*GCI*) resulted in mesophyll cells containing one or two giant chloroplasts and that these plants experienced reduced CO₂ fixation (Maple et al., 2004). However, work by another group reported that overexpression of the same locus inhibited plastid division and was able to rescue lines overexpressing the *FtsZs* (Raynaud et al., 2004). A recent paper attempting to clarify the role of *GCI* found that multiple lines overexpressing or lacking *GCI* rarely exhibited the phenotypes described previously (Li et al., 2017). There is good agreement that *GCI* localizes to the chloroplast, where it resides at the stromal side of the inner envelope membrane (Maple et al., 2004; Li et al., 2017). I believe that *GCI* likely does play a role in chloroplast division, but that this role is more complex than those of the division genes described above, and that different levels of *GCI* expression, rather than knocking out the gene, might be required for observation of the phenotypes described previously. To this end, I have included *GCI* in my list of chloroplast division genes to assess for expression analyses in *Peperomia* (Chapters 3 & 4).

Variation in nature

While most work on chloroplast division and morphology has been conducted in *A. thaliana*, there is natural variation in chloroplast population morphology in many different plant species, and this variation could be attributable to the chloroplast division genes or yet unknown players. Land plants generally maintain a large population (>50) of chloroplasts in their primary photosynthetic cells (palisade and spongy mesophyll) (Evert and Eichhorn, 2006). However, shade-adapted plants and low-light species maintain a smaller population of larger chloroplasts in these cells (Butterfass, 1979; Lichtenthaler et al., 1981). Here I will provide some specific examples.

There is a lot of variation in chloroplast population morphology in hornworts. In some species, one giant chloroplast is found per gametophyte cell. These chloroplasts are unusually shaped (e.g., ‘dumbbell’ or ellipsoidal) (Vaughn et al., 1992). Other accounts describe ‘chains’ of small chloroplasts and cells containing 100 chloroplasts (Butterfass, 1979). These dumbbell and chain phenotypes are reminiscent of *Atarc5* mutants, where chloroplast division is arrested during constriction (Figure 1.1B, blue dotted lines; Gao et al., 2003; Chen et al., 2018).

Among the vascular plants, members of the lycophyte genus *Selaginella* consistently (70% of species observed) exhibit one of the most extreme cases of altered chloroplast population morphology, with a single gigantic chloroplast per cell (Butterfass, 1979; Sheue et al., 2007; Liu et al., 2020). Further, *Selaginella* meristematic cells contain one large proplastid (Butterfass, 1979). It is thought that the large size and shape of *Selaginella* chloroplasts may be adaptive to the low light environments in which most species are found (Sheue et al., 2007; Li et al., 2019; Liu et al., 2020). Measurement of photosynthesis and photodamage in *Selaginella* demonstrated that these plants experience high rates of photodamage under high-light conditions (Ghaffar et al., 2018). What could be the molecular cause of this phenotype? Reduced chloroplast division is one potential explanation given the phenotypes of *Atarc6*, *AtftsZ1*, *AtftsZ2*, and *Atpdv1* and *Atpdv2* (Pyke et al., 1994; Robertson et al., 1995; Stokes et al., 2000; Okazaki et al., 2009; Schmitz et al., 2009). Similarly, many *A. thaliana* chloroplast division mutants, particularly those with very large chloroplasts, are susceptible to high light and photodamage (Jotham and Webber, 2005; Weise et al., 2015; Dutta et al., 2017). Though not as well documented, some *Selaginella* species have chloroplasts that appear to be linked (i.e., are not completely separated). This has been observed in young guard cells and the stem and base of

the leaf (Butterfass, 1979) and is reminiscent of *Atarc5* (Gao et al., 2003; Chen et al., 2018) and *Atclmp1* (Yang et al., 2011).

Though not as extensively documented, *Isoetes* species (lycophyta) (Butterfass, 1979) and the fern *Teratophyllum rotundifoliatum* (Nasrulhaq-Boyce and Duckett, 1991) exhibit similar large chloroplast phenotypes to that of *Selaginella* (Butterfass, 1979).

Members of the angiosperm genus *Peperomia* can be found growing in the understory of tropical forests (Fosberg and Sachet, 1975; Kubitzki et al., 1993; Rasingam and Parthasarathy, 2009; Ashton-Butt et al., 2018) and possess an array of unique low-light-associated morphological features (e.g., palisade-localized calcium oxalate crystals, water-containing adaxial ‘window’ cells, etc.) (Gausman et al., 1975; McWilliams and McWilliams, 1978; Gibeaut and Thomson, 1989; Christensen-Dean and Moore, 1993; Ting et al., 1994; Kuo-Huang et al., 2007; Lee, 2007). For several species, one such feature includes extremely large chloroplasts in the leaf palisade and root cells. Though chloroplasts in *Peperomia* generally are less abundant (~20-30 chloroplasts/cell) than in other plants (>50), they are even more so (2-6) in the palisade mesophyll of *P. metallica* (Butterfass, 1979; Cherkashin et al., 1999; Ahmadabadi and Bock, 2012), *P. pellucida* (Machado et al., 1986), and *P. meridiana* (Chapter 2). As with *Selaginella*, one might expect differences in chloroplast division to be behind this phenomenon, and the work described in the following chapters addresses this thoroughly.

Key points and moving forward

There is a great need for a thorough understanding of endosymbiotic organelle population morphology, particularly for chloroplasts, as alterations to the size and number of chloroplasts in leaf cells can affect photosynthesis (Jotham and Webber, 2005; Weise et al., 2015; Dutta et al., 2017). While it is clear that chloroplast division affects population morphology, it remains a

mystery how the host cell senses and controls the adjustment of this trait. Several cases of altered chloroplast population morphology have been discussed. Due to the breadth of variation in *Peperomia* (Machado et al., 1986; Ahmadabadi and Bock, 2012), this genus is potentially a valuable model for studying variation in chloroplast population morphology. The primary barrier to this proposed path forward is the lack of molecular resources for *Peperomia*, as only two *de novo* transcriptome assemblies are published (Matasci et al., 2014; Batista et al., 2017), and no genome is available for the genus. My work fills these gaps by exploring chloroplast division and population morphology over leaf development in *Peperomia* (Chapter 2), establishing novel transcriptomic and genomic resources for the genus (Chapters 3 & 4), and identifying underlying candidate regulators for the enlarged chloroplast morphology in *P. pellucida* (Chapter 4). Overall, the work described in this thesis provides novel insight into leaf development and chloroplast division dynamics in the early-diverging eudicot genus *Peperomia* and characterizes the expression of the chloroplast division genes by RNAseq for the first time.

REFERENCES

REFERENCES

- Ahmadabadi M, Bock R** (2012) Plastid division and morphology in the genus *Peperomia*. *Biol Plant* **56**: 301–306
- Anderson JM, Goodchild DJ, Boardman NK** (1973) Composition of the photosystems and chloroplast structure in extreme shade plants. *Biochim Biophys Acta BBA - Bioenerg* **325**: 573–585
- Asano T, Yoshioka Y, Kurei S, Sakamoto W, Sodmergen3, Machida Y** (2004) A mutation of the CRUMPLED LEAF gene that encodes a protein localized in the outer envelope membrane of plastids affects the pattern of cell division, cell differentiation, and plastid division in *Arabidopsis*. *Plant J* **38**: 448–459
- Ashton-Butt A, Aryawan AAK, Hood ASC, Naim M, Purnomo D, Suhardi, Wahyuningsih R, Willcock S, Poppy GM, Caliman J-P, et al** (2018) Understory vegetation in oil palm plantations benefits soil biodiversity and decomposition rates. *Front For Glob Change* **1**: 10
- Athanasiou K, Dyson BC, Webster RE, Johnson GN** (2009) Dynamic Acclimation of Photosynthesis Increases Plant Fitness in Changing Environments. *PLANT Physiol* **152**: 366–373
- Batista ANL, Santos-Pinto JRA dos, Batista JM, Souza-Moreira TM, Santoni MM, Zanelli CF, Kato MJ, López SN, Palma MS, Furlan M** (2017) The combined use of proteomics and transcriptomics reveals a complex secondary metabolite network in *Peperomia obtusifolia*. *J Nat Prod* **80**: 1275–1286
- Bjorkman O, Boardman NK, Anderson JM, Thorne SW, Goodchild DJ, Pylotis NA** (1971) Effect of light intensity during growth of *Atriplex patula* on the capacity of photosynthetic reactions, chloroplast components and structure. *Annu Rep Dep Plant Biol Carnegie Inst*
- Boardman NK** (1977) Comparative Photosynthesis of Sun and Shade Plants. *Annu Rev Plant Physiol* **28**: 355–377
- Boffey SA, Leech RM** (1982) Chloroplast DNA Levels and the Control of Chloroplast Division in Light-Grown Wheat Leaves. *Plant Physiol* **69**: 1387–1391
- Butterfass T** (1979) Patterns of chloroplast reproduction: a developmental approach to protoplasmic plant anatomy. Springer-Verlag, Wien ; New York
- Butterfass Th** (1980) The Continuity of Plastids and the Differentiation of Plastid Populations. *In* J Reinert, ed, *Chloroplasts*. Springer, Berlin, Heidelberg, pp 29–44
- Chen C, Cao L, Yang Y, Porter KJ, Osteryoung KW** (2019) ARC3 Activation by PARC6 Promotes FtsZ-Ring Remodeling at the Chloroplast Division Site. *Plant Cell* **31**: 862–885

- Chen C, MacCready JS, Ducat DC, Osteryoung KW** (2018) The molecular machinery of chloroplast division. *Plant Physiol* **176**: 138–151
- Chen Y, Asano T, Fujiwara MT, Yoshida S, Machida Y, Yoshioka Y** (2009) Plant Cells Without Detectable Plastids are Generated in the crumpled leaf Mutant of *Arabidopsis thaliana*. *Plant Cell Physiol* **50**: 956–969
- Cherkashin AA, Bulychev AA, Vredenberg WJ** (1999) Outward photocurrent component in chloroplasts of *Peperomia metallica* and its assignment to the 'closed thylakoid' recording configuration. *Bioelectrochem Bioenerg* **48**: 141–148
- Christensen-Dean GA, Moore R** (1993) Development of chlorenchyma and window tissues in leaves of *Peperomia columella*. *Ann Bot* **71**: 141–146
- Colletti KS, Tattersall EA, Pyke KA, Froelich JE, Stokes KD, Osteryoung KW** (2000) A homologue of the bacterial cell division site-determining factor MinD mediates placement of the chloroplast division apparatus. *Curr Biol* **10**: 507–516
- Cran DG, Possingham JV** (1972) Two Forms of Division Profile in Spinach Chloroplasts. *Nature New Biol* **235**: 142–142
- Dutta S, Cruz JA, Imran SM, Chen J, Kramer DM, Osteryoung KW** (2017) Variations in chloroplast movement and chlorophyll fluorescence among chloroplast division mutants under light stress. *J Exp Bot* **68**: 3541–3555
- Dutta S, Cruz JA, Jiao Y, Chen J, Kramer DM, Osteryoung KW** (2015) Non-invasive, whole-plant imaging of chloroplast movement and chlorophyll fluorescence reveals photosynthetic phenotypes independent of chloroplast photorelocation defects in chloroplast division mutants. *Plant J* **84**: 428–442
- Ellis JR, Leech RM** (1985) Cell size and chloroplast size in relation to chloroplast replication in light-grown wheat leaves. *Planta* **165**: 120–125
- Evert RF, Eichhorn SE** (2006) *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development*, 3rd ed. John Wiley & Sons, Inc., Hoboken, New Jersey
- Fang J, Li B, Chen L-J, Dogra V, Luo S, Wu W, Wang P, Hwang I, Li H, Kim C** (2022) TIC236 gain-of-function mutations unveil the link between plastid division and plastid protein import. *Proc Natl Acad Sci* **119**: e2123353119
- Fosberg RF, Sachet M-H** (1975) *Flora of Micronesia, 2: Casuarinaceae, Piperaceae, and Myricaceae*. *Smithson Contrib Bot* 1–32
- Frenkel M, Bellafiore S, Rochaix J-D, Jansson S** (2006) Hierarchy amongst photosynthetic acclimation responses for plant fitness. *Physiol Plant* **129**: 455–459

Gao H, Kadirjan-Kalbach D, Froehlich JE, Osteryoung KW (2003) ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc Natl Acad Sci* **100**: 4328–4333

Gausman HW, Rodriguez RR, Escobar DE (1975) Ultraviolet radiation reflectance, transmittance, and absorptance by plant leaf epidermises. *Agron J* **67**: 720–724

Ghaffar R, Weidinger M, Mähner B, Schagerl M, Lichtscheidl I (2018) Adaptive responses of mature giant chloroplasts in the deep-shade lycopod *Selaginella erythropus* to prolonged light and dark periods. *Plant Cell Environ* **41**: 1791–1805

Gibeaut DM, Thomson WW (1989) Stereology of the internal structures of leaves in *Peperomia obtusifolia*, *P. campotricha*, and *P. scandens*. *Bot Gaz* **150**: 115–121

Glynn JM, Froehlich JE, Osteryoung KW (2008) Arabidopsis ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane space. *Plant Cell* **20**: 2460–2470

Gould KS, Lee DW (1996) Physical and ultrastructural basis of blue leaf iridescence in four Malaysian understory plants. *Am J Bot* **83**: 45–50

Honda SI, Hongladarom-Honda T, Kwanyuen P, Wildman SG (1971) Interpretations on chloroplast reproduction derived from correlations between cells and chloroplasts. *Planta* **97**: 1–15

Irmak L (1957) The chloroplasts of sun and shade leaves. *Rev Fac Sci Univ Istanbul Ser B* **22**: 191–95

Itoh RD, Ishikawa H, Nakajima KP, Moriyama S, Fujiwara MT (2018) Isolation and analysis of a stromule-overproducing Arabidopsis mutant suggest the role of PARC6 in plastid morphology maintenance in the leaf epidermis. *Physiol Plant* **162**: 479–494

Jacobs M, Lopez-Garcia M, Phrathep OP, Lawson T, Oulton R, Whitney HM (2016) Photonic multilayer structure of *Begonia chloroplasts* enhances photosynthetic efficiency. *Nat Plants* **2**: 16162

Jellings AJ, Usher MB, Leech RM (1983) Variation in the chloroplast to cell area index in *Deschampsia antarctica* along a 16°C latitudinal gradient. *Br Antarct Surv Bull* **61**: 13–20

Jeong WJ, Park Y-I, Suh K, Raven JA, Yoo OJ, Liu JR (2002) A Large Population of Small Chloroplasts in Tobacco Leaf Cells Allows More Effective Chloroplast Movement Than a Few Enlarged Chloroplasts. *Plant Physiol* **129**: 112–121

Jotham Austin II, Webber AN (2005) Photosynthesis in *Arabidopsis thaliana* mutants with reduced chloroplast number. *Photosynth Res* **85**: 373–384

Kadirjan-Kalbach DK, Turmo A, Wang J, Smith BC, Chen C, Porter KJ, Childs K, DellaPenna D, Osteryoung KW (2019) Allelic variation in the chloroplast division gene *FtsZ2-2* leads to natural variation in chloroplast size. *Plant Physiol* **181**: 1059–1074

Königer M, Delamaide JA, Marlow ED, Harris GC (2008) *Arabidopsis thaliana* leaves with altered chloroplast numbers and chloroplast movement exhibit impaired adjustments to both low and high light. *J Exp Bot* **59**: 2285–2297

Kubitzki K, Rohwer JG, Bittrich V, eds (1993) *Flowering Plants · Dicotyledons*.

Külheim C, Ågren J, Jansson S (2002) Rapid regulation of light harvesting and plant fitness in the field. *Science* **297**: 91–93

Kuo-Huang L-L, Ku MSB, Franceschi VR (2007) Correlations between calcium oxalate crystals and photosynthetic activities in palisade cells of shade-adapted *Peperomia glabella*. *Bot Stud* **48**: 155–164

Larkin RM, Stefano G, Ruckle ME, Stavoe AK, Sinkler CA, Brandizzi F, Malmstrom CM, Osteryoung KW (2016) REDUCED CHLOROPLAST COVERAGE genes from *Arabidopsis thaliana* help to establish the size of the chloroplast compartment. *Proc Natl Acad Sci U S A* **113**: E1116-25

Lee DW (2007) *Nature's palette: the science of plant color*. University of Chicago Press, Chicago

Leech RM, Pyke KA (1988) Chloroplast division in higher plants with particular reference to wheat. *Div. Segreg. Organelles*. Cambridge University Press, New York, pp 39–62

Li X, Liu S, Wang Q, Wu H, Wan Y (2019) The effects of environmental light on the reorganization of chloroplasts in the resurrection of *Selaginella tamariscina*. *Plant Signal Behav* **14**: 1–8

Li Y, Wang L, Wang G, Feng Y, Liu X (2017) AT2G21280 only has a minor role in chloroplast division. *Front Plant Sci* **8**: 2095

Lichtenthaler HK, Buschmann C, Döll M, Fietz HJ, Bach T, Kozel U, Meier D, Rahmsdorf U (1981) Photosynthetic activity, chloroplast ultrastructure, and leaf characteristics of high-light and low-light plants and of sun and shade leaves. *Photosynth Res* **2**: 115–141

Liu J-W, Li S-F, Wu C-T, Valdespino IA, Ho J-F, Wu Y-H, Chang H-M, Guu T-Y, Kao M-F, Chesson C, et al (2020) Gigantic chloroplasts, including bizonoplasts, are common in shade-adapted species of the ancient vascular plant family Selaginellaceae. *Am J Bot* **107**: 562–576

Machado RD, Sant'Anna WR, Benchimol M (1986) Internal dimorphism of chloroplasts in the leaves of *Peperomia pellucida* (L.) H.B.K. *Protoplasma* **134**: 73–77

- Maple J, Fujiwara MT, Kitahata N, Lawson T, Baker NR, Yoshida S, Møller SG** (2004) GIANT CHLOROPLAST 1 is essential for correct plastid division in Arabidopsis. *Curr Biol* **14**: 776–781
- Maple J, Møller SG** (2005) An emerging picture of plastid division in higher plants. *Planta* **223**: 1
- Maple J, Vojta L, Soll J, Møller SG** (2007) ARC3 is a stromal Z-ring accessory protein essential for plastid division. *EMBO Rep* **8**: 293–299
- Marshall WF** (2016) Cell Geometry: How Cells Count and Measure Size. *Annu Rev Biophys* **45**: 49–64
- Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, et al** (2014) Data access for the 1,000 Plants (1KP) project. *GigaScience* **3**: 17
- McWilliams EL, McWilliams AK** (1978) Leaf anatomy and ultrastructure of acclimatized and non-acclimatized *Peperomia obtusifolia*.
- Miyagishima S** (2011) Mechanism of plastid division: from a bacterium to an organelle. *PLANT Physiol.* doi: 10.1104/pp.110.170688
- Miyagishima S** (2003) An evolutionary puzzle: chloroplast and mitochondrial division rings. *Trends Plant Sci* **8**: 432–438
- Nakanishi H, Suzuki K, Kabeya Y, Miyagishima S** (2009) Plant-Specific Protein MCD1 Determines the Site of Chloroplast Division in Concert with Bacteria-Derived MinD. *Curr Biol* **19**: 151–156
- Nasrulhaq-Boyce A, Duckett JG** (1991) Dimorphic epidermal cell chloroplasts in the mesophyll-less leaves of an extreme-shade tropical fern, *Teratophyllum rotundifoliatum* (R. Bonap.) Holtt.: a light and electron microscope study. *New Phytol* **119**: 433–444
- Okazaki K, Kabeya Y, Suzuki K, Mori T, Ichikawa T, Matsui M, Nakanishi H, Miyagishima S** (2009) The PLASTID DIVISION1 and 2 Components of the Chloroplast Division Machinery Determine the Rate of Chloroplast Division in Land Plant Cell Differentiation. *Plant Cell* **21**: 1769–1780
- Okie JG, Smith VH, Martin-Cereceda M** (2016) Major evolutionary transitions of life, metabolic scaling and the number and size of mitochondria and chloroplasts. *Proc R Soc B Biol Sci* **283**: 20160611
- Olson BJSC, Wang Q, Osteryoung KW** (2010) GTP-dependent Heteropolymer Formation and Bundling of Chloroplast FtsZ1 and FtsZ2. *J Biol Chem* **285**: 20634–20643
- Osteryoung KW, Pyke KA** (2014) Division and Dynamic Morphology of Plastids. *Annu Rev Plant Biol* **65**: 443–472

- Osteryoung KW, Stokes KD, Rutherford SM, Percival AL, Lee WY** (1998) Chloroplast division in higher plants requires members of two functionally divergent gene families with homology to bacterial ftsZ. *Plant Cell* **10**: 1991–2004
- Osteryoung KW, Vierling E** (1995) Conserved cell and organelle division. *Nature* **376**: 473–474
- Pao S-H, Tsai P-Y, Peng C-I, Chen P-J, Tsai C-C, Yang E-C, Shih M-C, Chen J, Yang J-Y, Chesson P, et al** (2018) Lamelloplasts and minichloroplasts in *Begoniaceae*: iridescence and photosynthetic functioning. *J Plant Res* **131**: 655–670
- Possingham JV** (1973) Effect of Light Quality on Chloroplast Replication in Spinach. *J Exp Bot* **24**: 1247–1259
- Possingham JV, Hashimoto H, Oross J** (1988) Factors that influence plastid division in higher plants. *Div. Segreg. Organelles*. Cambridge University Press, New York, pp 1–20
- Pyke K** (1999) Plastid division and development. *Plant Cell* **11**: 549–556
- Pyke K** (2009) *Plastid biology*. Cambridge University Press, Cambridge, UK ; New York
- Pyke K A** (2016) *Plastid Division*. *Mol. Cell Biol. Growth Differentiation Plant Cells*. CRC Press, Boca Raton, FL, pp 37–50
- Pyke KA** (1997) The genetic control of plastid division in higher plants. *Am J Bot* **84**: 1017–1027
- Pyke KA, Leech RM** (1992) Chloroplast Division and Expansion Is Radically Altered by Nuclear Mutations in *Arabidopsis thaliana* 1. *Plant Physiol* **99**: 1005–1008
- Pyke KA, Marrison JL, Leech AM** (1991) Temporal and Spatial Development of the Cells of the Expanding First Leaf of *Arabidopsis thaliana* (L.) Heynh. *J Exp Bot* **42**: 1407–1416
- Pyke KA, Rutherford SM, Robertson EJ, Leech RM** (1994) arc6, a fertile *Arabidopsis* mutant with only two mesophyll cell chloroplasts. *Plant Physiol* **106**: 1169–1177
- Rasingam L, Parthasarathy N** (2009) Diversity of understory plants in undisturbed and disturbed tropical lowland forests of Little Andaman Island, India. *Biodivers Conserv* **18**: 1045–1065
- Raynaud C, Cassier-Chauvat C, Perennes C, Bergounioux C** (2004) An *Arabidopsis* homolog of the bacterial cell division inhibitor SulA is involved in plastid division. *Plant Cell* **16**: 1801–1811
- Robertson EJ, Pyke KA, Leech RM** (1995) Arc6, an extreme chloroplast division mutant of *Arabidopsis* also alters proplastid proliferation and morphology in shoot and root apices. *J Cell Sci* **108**: 2937–2944

- Sakamoto T, Sakamoto Y, Matsunaga S** (2016) Cell Division and Cell Growth. Mol. Cell Biol. Growth Differ. Plant Cells. CRC Press, Boca Raton, FL, pp 86–100
- Salesse C, Sharwood R, Sakamoto W, Stern D** (2017) The Rubisco chaperone BSD2 may regulate chloroplast coverage in maize bundle sheath cells. *Plant Physiol* **175**: 1624–1633
- Schmitz AJ, Glynn JM, Olson BJSC, Stokes KD, Osteryoung KW** (2009) Arabidopsis FtsZ2-1 and FtsZ2-2 are functionally redundant, but FtsZ-based plastid division is not essential for chloroplast partitioning or plant growth and development. *Mol Plant* **2**: 1211–1222
- Sheue CR, Sarafis V, Kiew R, Liu HY, Salino A, Kuo-Huang LL, Yang YP, Tsai CC, Lin CH, Yong JWH, et al** (2007) Bizonoplast, a unique chloroplast in the epidermal cells of microphylls in the shade plant *Selaginella erythropus* (*Selaginellaceae*). *Am J Bot* **94**: 1922–1929
- Stokes KD, McAndrew RS, Figueroa R, Vitha S, Osteryoung KW** (2000) Chloroplast division and morphology are differentially affected by overexpression of FtsZ1 and FtsZ2 genes in Arabidopsis. *PLANT Physiol* **124**: 1668–1677
- Ting IP, Patel A, Sipes DL, Reid PD, Walling LL** (1994) Differential expression of photosynthesis genes in leaf tissue layers of *Peperomia* as revealed by tissue printing. *Am J Bot* **81**: 414–422
- Tsel'niker YL** (1973) Rhythms of tissue and chloroplast growth and determination of characteristics of light and shade structure of the leaf in Norway maple. *Sov. Plant Physiol*.
- Vaughn KC, Ligrone R, Owen HA, Hasegawa J, Campbell EO, Renzaglia KS, Monge-Najera J** (1992) The anthocerot chloroplast: a review. *New Phytol* **120**: 169–190
- Wang W, Li J, Sun Q, Yu X, Zhang W, Jia N, An C, Li Y, Dong Y, Han F, et al** (2017) Structural insights into the coordination of plastid division by the ARC6-PDV2 complex. *Nat Plants* **3**: 1–9
- Weise SE, Carr DJ, Bourke AM, Hanson DT, Swarthout D, Sharkey TD** (2015) The arc mutants of Arabidopsis with fewer large chloroplasts have a lower mesophyll conductance. *Photosynth Res* **124**: 117–126
- Whatley JM** (1988) Mechanisms and morphology of plastid division. Div. Segreg. Organelles. Cambridge University Press, New York, pp 63–84
- Xiong D, Huang J, Peng S, Li Y** (2017) A few enlarged chloroplasts are less efficient in photosynthesis than a large population of small chloroplasts in Arabidopsis thaliana. *Sci Rep* **7**: 5782
- Yang Y, Sage TL, Liu Y, Ahmad TR, Marshall WF, Shiu S-H, Froehlich JE, Imre KM, Osteryoung KW** (2011) CLUMPED CHLOROPLASTS 1 is required for plastid separation in Arabidopsis. *Proc Natl Acad Sci* **108**: 18530–18535

Yoshida Y, Mogi Y, TerBush AD, Osteryoung KW (2016) Chloroplast FtsZ assembles into a contractible ring via tubulin-like heteropolymerization. *Nat Plants* **2**: 1–10

Zhang M, Chen C, Froehlich JE, TerBush AD, Osteryoung KW (2016) Roles of Arabidopsis PARC6 in Coordination of the Chloroplast Division Complex and Negative Regulation of FtsZ Assembly. *Plant Physiol* **170**: 250–262

Zhang M, Schmitz AJ, Kadirjan-Kalbach DK, TerBush AD, Osteryoung KW (2013) Chloroplast Division Protein ARC3 Regulates Chloroplast FtsZ-Ring Assembly and Positioning in Arabidopsis through Interaction with FtsZ2. *Plant Cell* **25**: 1787–1802

Chapter 2. Chloroplast population morphology and division in the tropical plant

genus Peperomia

Introduction

Peperomia Ruiz & Pav. is the eleventh largest angiosperm genus with just over 1,300 species (Simmonds et al., 2021). *Peperomia* spp. are early-diverging eudicots found in the *Piperaceae* family alongside *Piper nigrum* (black pepper). Further, *Peperomia* is a member of the Magnoliid clade, containing *Cinnamomum* spp. (cinnamon), and *Persea americana* (avocado) (Wanke et al., 2006; Frenzke et al., 2015; Soltis et al., 2018; Simmonds et al., 2021). *Peperomia* spp. are herbaceous perennial epiphytes (Frenzke et al., 2016), best known as commercially-available house plants, primarily due to their morphologically diverse and decorative foliage (Frenzke et al., 2015; Gutierrez et al., 2016). Several species of *Peperomia* have medicinal value, and the genus is well known for the production of biochemically active compounds (Mahiou et al., 1995; Mahiou et al., 1996; Gutierrez et al., 2016; Yu et al., 2016). Due to their phylogenetic position within the Magnolidae and the diversity within the genus, these species are important for understanding ecology, evolution, and diversification within the angiosperms (Egydio et al., 2014; Richards et al., 2015; Simmonds et al., 2021).

Peperomia is the most biogeographically diverse genus in the *Piperaceae* family, and these species can be found in the Neotropics, South Pacific, Asian tropics, and Africa (Frenzke et al., 2015; Frenzke et al., 2016; Simmonds et al., 2021). A number of *Peperomia* spp. reside in the shaded understory of tropical forests (Fosberg and Sachet, 1975; Kubitzki et al., 1993; Rasingam and Parthasarathy, 2009; Ashton-Butt et al., 2018) and exhibit leaf-morphological features thought to be beneficial to shade or low-light environments. Large water-containing ‘window’ cells just below the adaxial epidermis of the leaf are thought to help distribute light to the primary photosynthetic tissues below (Christensen-Dean and Moore, 1993; Egbert and Martin, 2000; Egbert and Martin, 2002). Calcium oxalate crystals are found in the palisade

mesophyll leaf cells and are also thought to distribute light to optimize capture (Kuo-Huang et al., 2007). Large grana-dense shade-like chloroplasts (Irmak, 1957; Bjorkman et al., 1971; Anderson et al., 1973; Tsel'niker, 1973; Boardman, 1977; Lichtenthaler et al., 1981; Sheue et al., 2007) have been reported in some species (Machado et al., 1986; Ahmadabadi and Bock, 2012) and utilized in early patch-clamp studies on photosynthesis (Bulychev et al., 1972; Bulychev et al., 1998; Cherkashin et al., 1999).

In general, leaves are comprised of adaxial epidermal, palisade, spongy mesophyll, and abaxial epidermal cells from the light-facing top to the underside. In plants that have been studied extensively, including *Arabidopsis thaliana*, spinach, and wheat, mature mesophyll cells serve as the primary photosynthetic tissue and contain a dense population of 50-150 chloroplasts that collectively occupy ~70% of the cell surface area (Butterfass, 1979; Evert and Eichhorn, 2006; Pyke, 2009). In many plants, the planar area of a single mature chloroplast is ~50 μm^2 (Pyke, 1997). Together, *Peperomia metallica* and *Peperomia pellucida* represent one of the most severe natural cases of altered chloroplast population morphology described amongst the angiosperms (Butterfass, 1979), as their palisade mesophyll cells specifically contain 2-8 enlarged chloroplasts, while those of *Peperomia serpens* and *Peperomia argyreia* had a larger population of much smaller chloroplasts (Machado et al., 1986; Ahmadabadi and Bock, 2012). Importantly, coverage appears to be similar in palisade cells between species (Ahmadabadi and Bock, 2012). Taken together, this suggests that chloroplast division may be reduced or inhibited completely in the palisade cells of *P. metallica* and *P. pellucida*, and that increased chloroplast expansion occurs to compensate for coverage. Here, I explore chloroplast morphology in nine *Peperomia* spp. and investigate to what extent chloroplast division occurs, if at all, in the species with large palisade-cell chloroplasts.

Results

Chloroplast population morphology in the mature leaves of nine Peperomia species

Despite there being over 1,000 species of *Peperomia* (Simmonds et al., 2021), chloroplast population morphology had only been described in four (Machado et al., 1986; Ahmadabadi and Bock, 2012). I obtained these species to confirm their phenotypes, but found that one of the large chloroplast species, previously called ‘*P. peduncularis*’ Ahmadabadi and Bock, 2012), was the same as *P. metallica*, reducing my sample size of large chloroplast species to two. This was further confirmed by an expert on *Peperomia* (Marie-Stéphanie Samain, Instituto de Ecología, A.C., Red de Diversidad Biológica del Occidente Mexicano). Further, I found it difficult to maintain large populations of all species except *P. pellucida*, primarily due to the requirement for clonal propagation and slow growth rate. To identify other *Peperomia* species that would be easier to work with, I obtained and characterized plants from various nurseries (see Methods). Plant morphology for each species was cross-referenced against herbaria and live-specimen images available through Tropicos and the Meise Botanic Gardens to confirm species identity. Nine species were identified with high confidence as accepted species and retained for assessment of chloroplast population morphology.

In order to observe chloroplast size and number in these nine species, mature leaf tissue was fixed and separated cells were imaged (Figure 2.1). There was great diversity in the size and number of chloroplasts per cell (Figure 2.1). Based on previous phylogenetic analyses, I anticipated the morphologies of *P. meridiana*, *P. metallica*, and *P. pellucida* would be similar, as they are closely related (Wanke et al., 2006; Frenzke et al., 2015). Indeed, these three species contained 3-6 very large chloroplasts ranging from ~150-400 μm^2 in the palisade mesophyll cells (Figure 2.1)—establishing *P. meridiana* as a ‘large chloroplast’ species for the first time.

Palisade chloroplast area in the other species observed was much smaller, while the number of chloroplasts per cell was higher, e.g., $\sim 12\text{-}50\ \mu\text{m}^2$ and ~ 27 chloroplasts/cell, respectively, for *P. dahlstedtii* (Figure 2.1).

Depending on the species, different types of chloroplasts appeared to be present in the mesophyll cells (palisade and spongy). Chloroplasts with blue iridescent-like coloring were occasionally observed in the spongy mesophyll cells of *P. dahlstedtii*, *P. incana*, *P. metallica*, *P. pellucida*, and *P. serpens* (Figure 2.1). Chloroplasts with similar coloring and morphology have been observed in other low-light plants such as *Begonia pavonina*, *Selaginella willdenowii*, and *Elaphoglossum herminieri*, the leaves or microphylls of which give off an iridescent blue color (Graham et al., 1993; Gould and Lee, 1996; Lee, 2007; Glover and Whitney, 2010; Thomas et al., 2010).

Clusters of chloroplasts were also frequently observed in spongy mesophyll cells of *P. metallica* and *P. pellucida* (Figure 2.1). This made me wonder whether the large-palisade chloroplasts of these species might be connected, as it is difficult to see whether they are separated within the tight confines of the palisade cell (Figure 2.1). However, several patch-clamp studies have described the isolation of individual chloroplasts from *P. metallica* leaves as part of their methods (Vredenberg and Tonk, 1975; De Grooth and Van Gorkom, 1981) indicating that they are distinct individuals.

Quantifying chloroplast population morphology in nine Peperomia species

It has been established in several plant species that there is a strong correlation between the number of chloroplasts per mesophyll cell and cell size, as chloroplast numbers typically increase by division during cell and leaf expansion to maintain coverage (Ellis and Leech, 1985; Pyke and Leech, 1987; Leech and Pyke, 1988; Pyke and Leech, 1992).

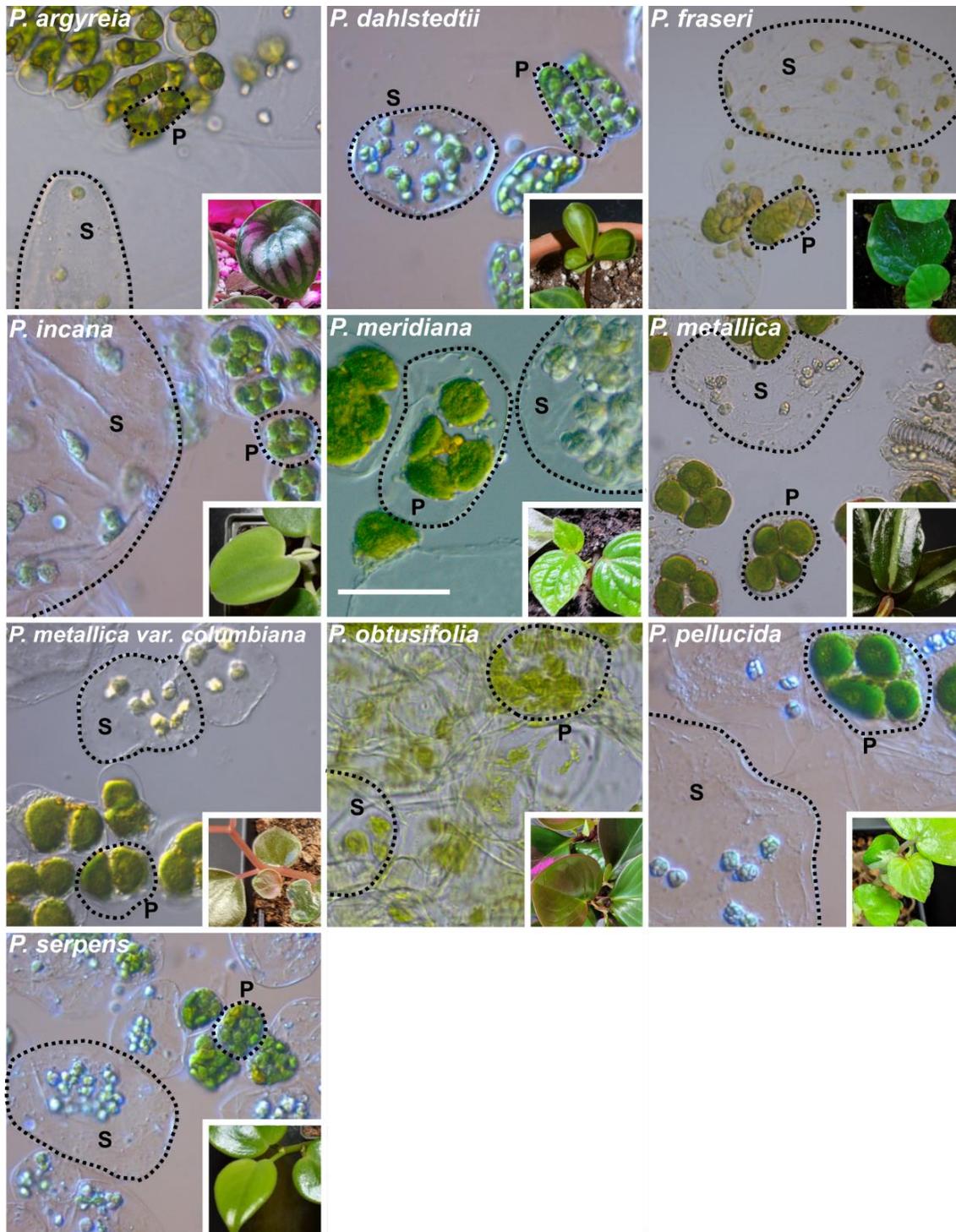


Figure 2.1. Chloroplast population morphology in *Peperomia*.

Palisade (P) and spongy (S) mesophyll cells (one of each kind outlined with dotted black lines/panel) are isolated from fixed mature leaves. Species names are indicated in the top left corner of each panel. The scale bar (centermost panel) represents 50 μm for all images. Insets (bottom right of each panel) show leaf morphology for each species. Note that *P. metallica* var. *columbiana* was erroneously referred to as *P. peduncularis* previously (Ahmadabadi and Bock, 2012).

The maintenance of chloroplast coverage has been demonstrated particularly well in the *A. thaliana* chloroplast division mutants, where chloroplast number does not increase with cell size—the result of perturbation to division (Pyke and Leech, 1992). The large chloroplast phenotype found in the palisade cells of *P. pellucida*, *P. metallica*, and *P. meridiana* suggested that the division of differentiated chloroplasts is reduced in these cells. To test this, I measured chloroplast number per cell and found that it did not increase with cell size in these species, as they maintained ~3-5 chloroplasts per cell regardless of cell area (Figure 2.2). In contrast, small chloroplast species such as *P. dahlstedtii*, *P. incana*, *P. fraseri*, and *P. serpens* accumulated more chloroplasts as cell size increased (Figure 2.2). Taken together, these results suggested that chloroplast division could occur in the palisade cells of small chloroplast species, but that it may not in the large chloroplast species.

In order to determine whether chloroplast division occurs in the palisade cells of the large-chloroplast *Peperomia* species, I selected two representatives to focus on—*P. dahlstedtii*, which has up to 30 chloroplasts (each averaging ~27 μm^2 in area) per cell, and *P. pellucida*, which maintains 3-6 chloroplasts (each averaging ~175 μm^2) per cell (Figure 2.3). These species were selected primarily for their clear differences in chloroplast number per cell and average chloroplast size, and for their ease of cultivation, similar growth habit, and the relatively small diploid genome size of *P. dahlstedtii*, which is discussed in Chapter 4.

Chloroplast division during leaf development in P. dahlstedtii and P. pellucida

As shown in Figures 2.2 and 2.3, the larger palisade cells of *P. pellucida* do not accumulate more chloroplasts, while those in *P. dahlstedtii* do.

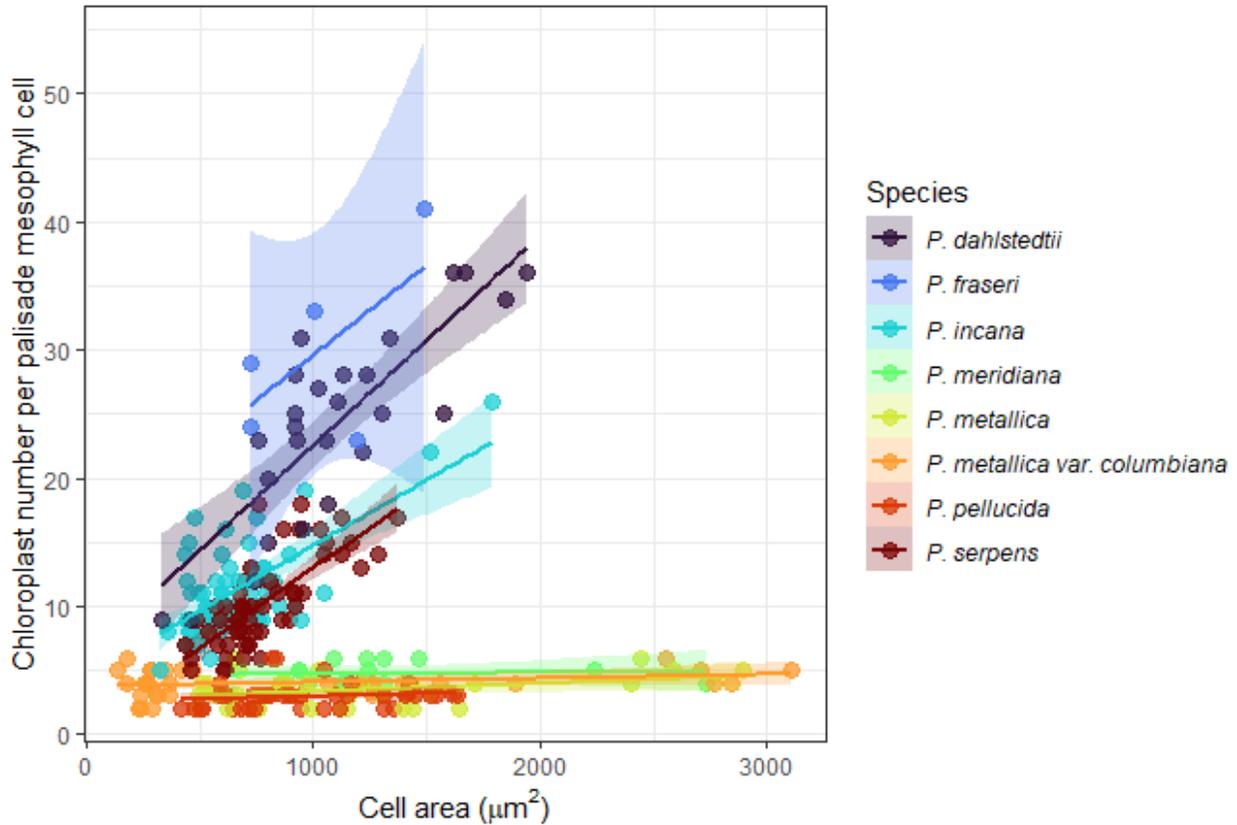


Figure 2.2. Correlation between the number of chloroplasts per cell and planar cell area in *Peperomia*.

Data were measured from fixed mature palisade mesophyll cells from the species listed in the legend. The 95% confidence intervals for each linear regression are indicated by semi-transparent coloring surrounding the line of best fit, for each species. *P. dahlstedtii*, *P. fraseri*, *P. incana*, and *P. serpens* had slopes significantly greater than zero ($p < 0.001$), while the remaining large-chloroplast species did not ($p > 0.05$).

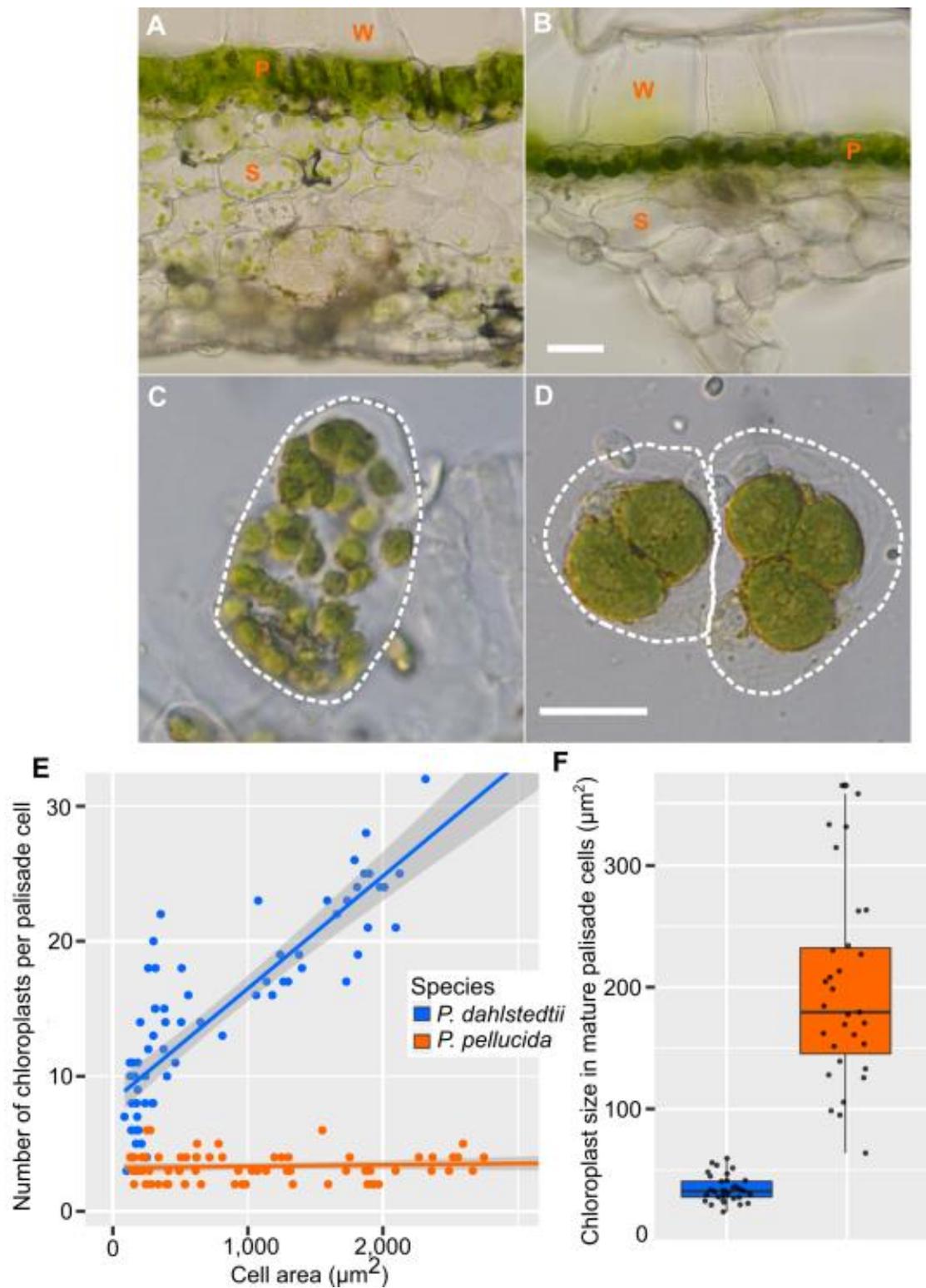


Figure 2.3. Quantitative assessment of chloroplast population morphology in *P. pellucida* and *P. dahlstedtii*.

(A, B) Live cross-sections were taken from the mature leaves of (A) *P. dahlstedtii* and (B) *P. pellucida*. W, water-containing window cell; P, palisade mesophyll; S, spongy mesophyll.

Figure 2.3 (cont'd)

The scale bar (bottom left, B) represents 100 μm for both images. (C) Fixed mature palisade mesophyll cells from *P. dahlstedtii* and *P. pellucida*. Individual cells are outlined with white dashed lines. The scale bar (bottom left, D) represents 20 μm . (E) Linear regressions for chloroplast number per cell as a function of cell area. 95% confidence intervals are represented with semi-transparent grey coloring around each line. (F) Quantitative comparison of chloroplast size (Wilcox test, $p < 0.001$) between the two species. The legend (center) serves for both plots.

I measured chloroplast number per palisade cell in both species over the course of leaf expansion to determine whether chloroplast division occurs in *P. pellucida*, as chloroplast division is known to occur continuously throughout leaf development as the cells expand, increasing chloroplast numbers to maintain coverage (Pyke and Leech, 1992). To establish a developmental leaf series for each species, the smallest unfurled leaf at the shoot apex was identified, removed, measured for length and width, and placed into fixative (Figure 2.4B, *P. dahlstedtii* & E, *P. pellucida*). The oldest fully-expanded leaf closer to the base of the plant was taken next (Figure 2.4D, *P. dahlstedtii* & G, *P. pellucida*). In between these two extremes, both spatially along the stem and regarding leaf size, three more leaves varying in size and spaced as equidistantly as possible along the stem were identified and prepared in the same manner. This was done for three plants per species. Together, these leaf series span the period of leaf expansion through maturity. The fixed samples were stained and imaged for cell wall in addition to chlorophyll autofluorescence, allowing for more accurate chloroplast counts in the younger leaves, where chloroplast and cell size are generally smaller. Palisade cell chloroplast number, size, and cell size measurements were taken from these images, and the resulting data are shown in Figure 2.4. In *P. dahlstedtii*, chloroplast number per cell significantly increased between each stage, with approximately two doublings to chloroplast number overall (~5 to 22 chloroplasts per cell; Figure 2.4A). This is in line with previous observations on chloroplasts population increases in the expanding leaves of other plants, though perhaps a bit lower in net increase, as three to four rounds of division have been observed in *A. thaliana* and wheat (Leech and Pyke, 1988; Pyke, 2009). Similar to *P. dahlstedtii*, the youngest *P. pellucida* leaves had ~5 chloroplasts/cell (red; Figure 2.4A). However, these numbers did not increase significantly with leaf expansion (Figure 2.4A). In fact, the youngest *P. pellucida* leaves (stage A) had significantly more chloroplasts per

cell than to all other more mature leaves (stages B-E; *Figure 2.4A*). This may indicate that cell division was still occurring in the stage A leaves of *P. pellucida* while chloroplast division was not, meaning that the number of chloroplasts per cell would be ~2x lower in the daughter cells post-division. This is consistent with my results showing that the average number of chloroplasts per cell was ~6 in stage A leaves and ~3 in the successive stages (*Figure 2.4A*). Together, these data demonstrate that the reduced number of chloroplasts observed in mature *P. pellucida* palisade cells, and likely that of its sister species *P. metallica* and *P. meridiana* (*Figure 2.1*), can be attributed to a complete lack of chloroplast division. Further, I have established *P. dahlstedtii* as a positive control for palisade-cell chloroplast division, which is utilized for the work described in Chapters 3 and 4. While I have observed chloroplasts in the spongy mesophyll cells of *P. pellucida* and *P. dahlstedtii* (*Figure 2.3A-B*), it was particularly difficult to quantify their numbers and size in *P. pellucida*, so it is unclear whether chloroplast division is inhibited in the whole *P. pellucida* leaf or just in the palisade layer (*Figure 2.3B*).

Discussion

Chloroplast population morphology is an important trait that can vary depending on genetic or environmental conditions. However, we do not understand how the cell senses and regulates this trait. In order to study this trait, I have identified or confirmed small and large chloroplast phenotypes in the palisade mesophyll cells of five and four species, respectively, within the *Peperomia* genus (*Figure 2.1*; Machado et al., 1986; Ahmadabadi and Bock, 2012). In order to conduct fine-scale developmental studies on chloroplast population size and division, two representative species were selected as models for the small and large chloroplast phenotype in the palisade cells, *P. dahlstedtii* and *P. pellucida*, respectively (*Figure 2.3*).

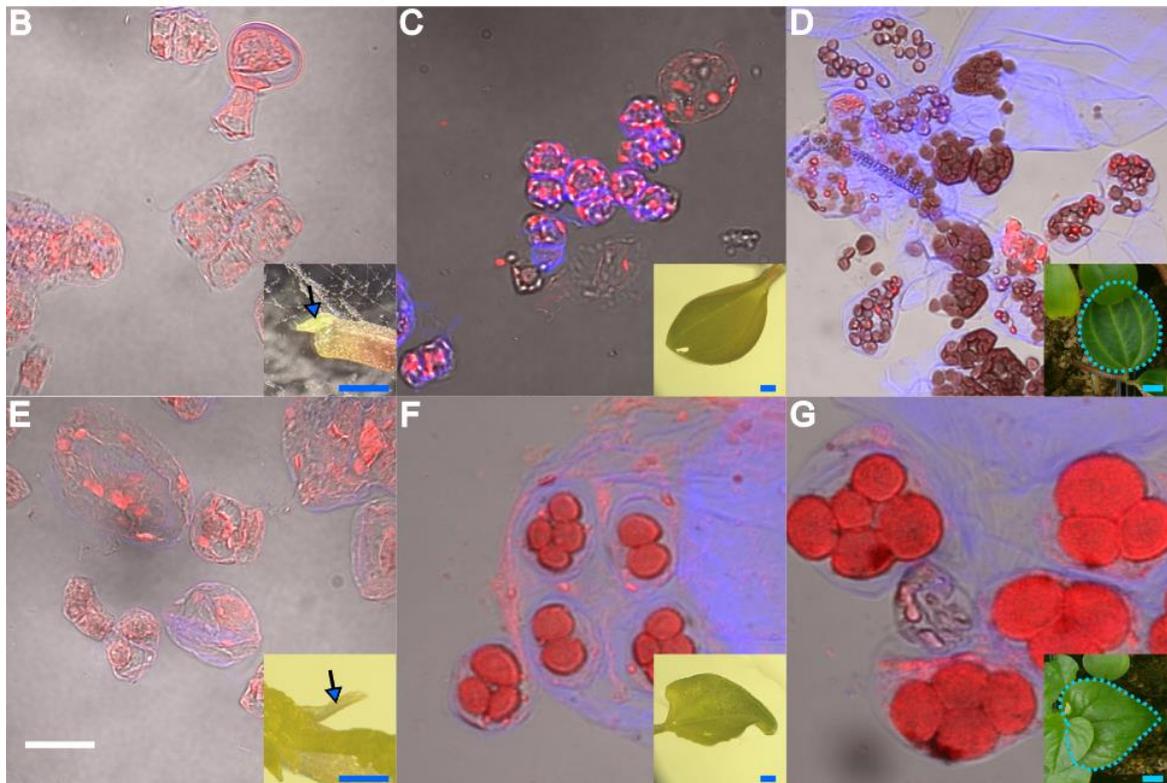
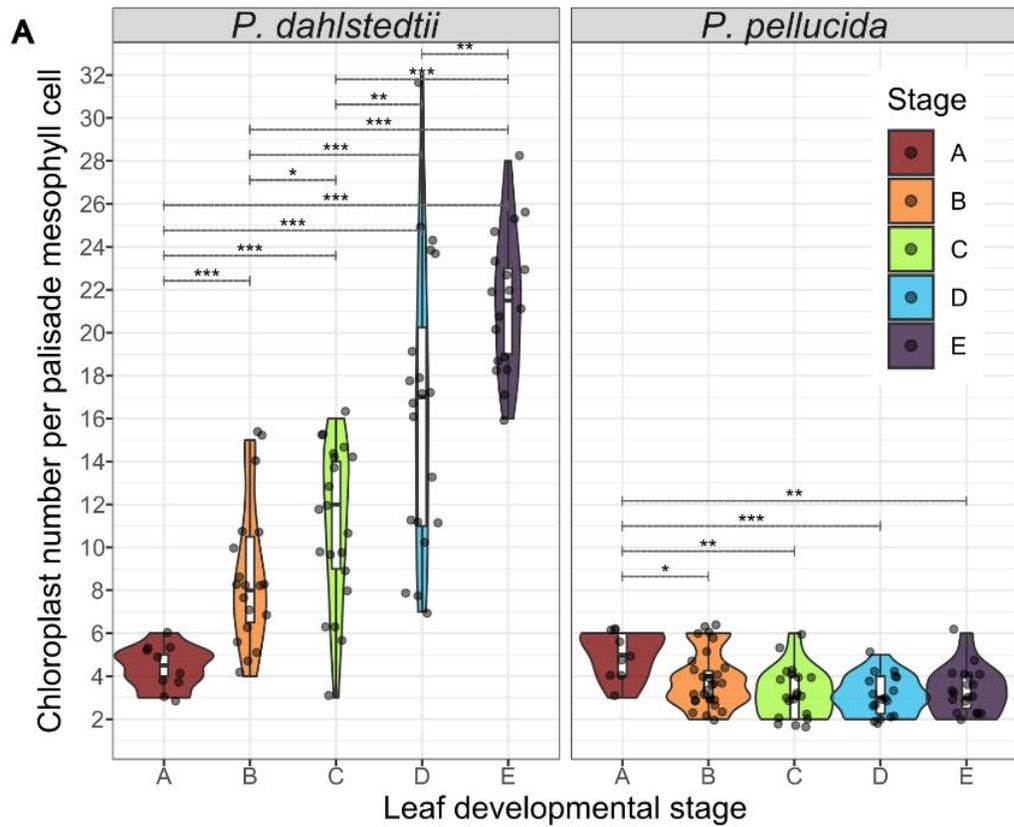


Figure 2.4. Chloroplast division over leaf development in *P. pellucida* and *P. dahlstedtii*.

Figure 2.4 (cont'd)

(A) The average number of chloroplasts per palisade cell in *P. dahlstedtii* and *P. pellucida* from young through mature leaves (stages A-E, respectively). Statistical significance was determined by Kruskal–Wallis tests with post-hoc Holm-correction (** <0.001 , ** <0.01 , and * <0.05). (B-D) *P. dahlstedtii* stage A, stage C, and stage E, merged bright field, calcofluor white fluorescence (cell wall), and chlorophyll autofluorescence (red), images of fixed and separated leaf tissue. (E-G) Equivalent stages for *P. pellucida* in the same order. The white scale bar (bottom leftmost panel) denotes 25 μm for all similar images. Image insets in the bottom right of each panel show leaf morphology for the corresponding stage. Blue arrows indicate the youngest-leaves isolated. The blue inset scale bars represent (from left to right) 1 mm, 1 mm, and 1 cm. Mature leaves are outlined with a blue dashed line.

After identifying five distinct leaf-developmental stages for both species, I was able to demonstrate that the large-chloroplast phenotype observed in *P. pellucida* palisade cells is caused by a complete lack of chloroplast division during leaf expansion (Figure 2.4). These data suggest that the chloroplast division genes could play a role in the large chloroplast phenotype observed in certain *Peperomia* species.

Interestingly, chloroplast size and number also vary between the palisade and spongy mesophyll cells within the three large chloroplast species—*P. meridiana*, *P. metallica*, and *P. pellucida* (Figure 2.1). Chloroplast division was only quantified in palisade mesophyll cells, so it is unclear whether chloroplast division occurs in the spongy mesophyll cells of these species. Further, it remains to be seen whether the spongy mesophyll chloroplasts in *P. pellucida* can be considered functionally equivalent to those in the spongy mesophyll of other species, as they do not appear to contain many chloroplasts, and those observed do not appear to have much chlorophyll (Figure 2.3B). It is possible that the large palisade chloroplasts block enough light to disrupt the development of those in the spongy mesophyll below (Figure 2.4B), especially given the well-developed dense grana documented in *P. pellucida* palisade chloroplasts (Machado et al., 1986). This would explain their coloring and smaller population sizes, as chloroplast division is regulated by the light environment (Butterfass, 1979). Overall, the cell-specific differences in chloroplast population morphology between the palisade and spongy mesophyll suggest that differential gene expression, rather than gene-level loss or mutation, may be responsible.

Most of what we know about alterations to chloroplast population morphology on the gene level comes from studies on *A. thaliana* mutants for the chloroplast division genes (Chen et al., 2018; Chapter 1). However, no studies have been done on the genetic regulation of this trait in a natural system. With the goal of identifying potentially novel chloroplast population

morphology-regulating genes and observing expression of the chloroplast division genes between the small and large chloroplast conditions, I established novel transcriptome data for the two model *Peperomia* species described here (Figure 2.3 & Figure 2.4). This work is described in Chapter 3 and expanded upon in Chapter 4.

Conclusions

In my efforts to characterize the regulation of chloroplast population morphology in plants I have demonstrated that the small population (~3 chloroplasts per cell) of enlarged chloroplasts in the low-light species *P. pellucida* is caused by a complete lack of division in the palisade cells. Further, I have shown that chloroplast division does occur in the small-chloroplast species *P. dahlstedtii*. Moving forward, both species can be used as models for their respective chloroplast size phenotypes. In Chapter 3, I describe the generation of *de novo* transcriptomes for *P. pellucida* and *P. dahlstedtii* as a first step towards identifying candidate regulators of chloroplast population morphology, which are likely expressed at different levels between these species.

Materials and methods

Plant materials, growth conditions, and propagation

P. fraseri, *P. dahlstedtii*, *P. argyreia*, *P. incana*, and *P. serpens* were purchased from Steve's Leaves (<https://stevesleaves.com/>). *P. meridiana* and occasionally *P. metallica* (when available) were purchased from Josh's Frogs (<https://www.joshsfrogs.com/>). *P. metallica* (var. *columbiana* or otherwise) was purchased from Walmart and *P. obtusifolia* from Meijer. *P. pellucida* was propagated from seeds initially purchased through Amazon (https://www.amazon.com/gp/product/B01M7T2E8I/ref=ppx_yo_dt_b_search_asin_title?ie=UTF8&psc=1). Seeds were sprinkled evenly on top of moist soil in square 4-inch pots and covered

with a clear plastic dome, which was removed ~2-3 weeks after germination (when the first two true leaves were visible).

For species identification, herbaria records (descriptions and images) available through Tropicos (<https://www.tropicos.org>) and the Meise Botanic Garden living plant collections database (LIVCOL; <http://www.br.fgov.be/research/COLLECTIONS/LIVING/LIVCOL>) were used. Further, only accepted species as determined using The Plant List (<http://www.theplantlist.org>) were included in this work.

Peperomia plants were grown in a growth chamber in a 2:1 mixture of ‘Redi Earth’ soil mix (SunGro Horticulture, TLRE3F) to perlite, under a light intensity of 61 μmol (12-h day), humidity of 55-68%, at temperatures of 21-25°C. Plants were watered when the soil was dry and fertilized (Miracle-Gro All Purpose Plant Food, 1000283), following the manufacturer’s instructions every two and five weeks during the spring/summer and fall/winter months. All species were propagated from leaf cuttings. Healthy leaves were removed and cut in half perpendicular to the direction of the veins using a sterile razor blade. Both halves were dipped (the basal end of each, closest to if not at the petiole) in FastRoot powder (MiracleGro, 1006451) and packed in moist soil. The leaf cuttings were covered with a plastic dome to increase humidity and incubated in the growth chamber two feet further from the lights than mature plants for 3-4 weeks. Successfully rooted cuttings were transferred to fresh soil and brought two feet closer to the chamber lights.

Sampling of the leaf developmental series

To start, the smallest unfurled leaf at the shoot apex was identified, removed, measured for length and width, and placed into 3.5% glutaraldehyde. Next, the oldest (fully-expanded) leaf closest to the base of the stem was processed as above. In between the locations of the first and

second samples, leaves of different sizes spaced as equidistantly along the stem were removed and prepared as above. This process was done in *P. dahlstedtii* and *P. pellucida* separately, with three biological replicates per species. A total of five distinct leaves were isolated in this manner, constituting five distinct developmental stages. All leaves were removed from the plant using a sterile double edge razor blade, measured for length and width, fixed in 3.5% glutaraldehyde, and prepared for imaging as described below.

Preparation and imaging of leaf tissue

Following techniques as described by Ruzin (1999), Live cross-sections were made using fresh mature leaves. Leaves were removed from the plant and promptly placed in a shallow dish of sterile distilled H₂O. The leaf was gently held down under a layer of parafilm, and a double edge razor blade cleaned with 70% ethanol was used to slice several cross-sections <1 mm in thickness.

For images such as those presented in Figure 2.1, leaves were fixed at room temperature in 3.5% glutaraldehyde for 2 h in the dark with shaking. The tissue was subsequently loosened in 0.2 M Na₂-EDTA pH 9.0 for 1 h at 55°C (Pyke and Leech, 1991; Khoshravesh and Sage, 2018) and could be stored at 4°C for three months in between imaging. Slides were prepared by placing a small amount of leaf tissue in a drop of H₂O or 0.2 M Na₂-EDTA pH 9.0 on a glass slide. After placing the coverslip, the sample was gently tapped down using a pencil eraser, effectively separating the cells. Differential Interference Contrast (DIC) images were taken on a Leica DMI3000 B microscope with a Leica FLEXACAM C1 camera (12730522) at 400x magnification.

For the developmental series, samples were prepared as described above. However, prior to mounting, samples were stained with 0.6% Calcofluor White (CFW) overnight in the dark at

4°C (Lee et al., 2021). The next day samples were mounted on slides as described above, but in sterile 75% glycerol in place of H₂O or EDTA. Further, slides were sealed (for long-term storage at 4°C as needed) by applying clear nail polish along the edges of the coverslip. Images of chlorophyll autofluorescence (Excitation 633nm, Emission 600-700nm) and CFW (Excitation 405nm, Emission 410-500nm) were acquired at 400x magnification using a FLUOVIEW (FV1000, Olympus) confocal microscope at the Michigan State University Center for Advanced Microscopy.

Data manipulation, plots, and statistical analyses

All cell and chloroplast area measurements and counts were performed using the freehand tracing and counter tools in Fiji (ImageJ Version 2.0.0-rc-69/1.52p; Schindelin et al., 2012) following previously established methods (Pyke and Leech, 1991). Unless otherwise noted, R version 3.6.3 (Team, 2020) and RStudio version 1.2.5033 (2019) were used for data manipulation. Plots were made with ggplot2 (Wickham, 2016) using color palettes from viridis (Garnier, 2018). The R code used for this project will be made available at <https://github.com/AFrolicOfFerns/peperomia-developmental-measurements>.

Acknowledgments

Dr. Marie-Stéphanie Samain (Instituto de Ecología, A.C., Red de Diversidad Biológica del Occidente Mexicano, Pátzcuaro 61600, Mexico) confirmed the identify of some *Peperomia* species. The University of Alberta Botanic Gardens and Members of the laboratory of Tammy Sage (The University of Toronto) provided and stored *Peperomia fraseri* specimens for us. Stefanie Sultmanis and Roxana Khoshravesh (former Sage lab members, University of Toronto) taught me the invaluable art of embedding and sectioning leaves. Thank you, as well, to Jameel Al-Haddad (Michigan State University, Dept. of Plant Biology) for providing advice and

materials for microscopy. I also want to thank ‘Mike from Michigan’ for his expertise and assistance regarding how to care for *Peperomia*.

REFERENCES

REFERENCES

- Ahmadabadi M, Bock R** (2012) Plastid division and morphology in the genus *Peperomia*. *Biol Plant* **56**: 301–306
- Anderson JM, Goodchild DJ, Boardman NK** (1973) Composition of the photosystems and chloroplast structure in extreme shade plants. *Biochim Biophys Acta BBA - Bioenerg* **325**: 573–585
- Ashton-Butt A, Aryawan AAK, Hood ASC, Naim M, Purnomo D, Suhardi, Wahyuningsih R, Willcock S, Poppy GM, Caliman J-P, et al** (2018) Understorey vegetation in oil palm plantations benefits soil biodiversity and decomposition rates. *Front For Glob Change* **1**: 10
- Bjorkman O, Boardman NK, Anderson JM, Thorne SW, Goodchild DJ, Pyliotis NA** (1971) Effect of light intensity during growth of *Atriplex patula* on the capacity of photosynthetic reactions, chloroplast components and structure. *Annu Rep Dep Plant Biol Carnegie Inst*
- Boardman NK** (1977) Comparative Photosynthesis of Sun and Shade Plants. *Annu Rev Plant Physiol* **28**: 355–377
- Bulychev AA, Andrianov VK, Kurella GA, Litvin FF** (1972) Micro-electrode Measurements of the Transmembrane Potential of Chloroplasts and its Photoinduced Changes. *Nature* **236**: 175–177
- Bulychev AA, Dassen JHA, Vredenberg WJ, Opanasenko VK, Semenova GA** (1998) Stimulation of photocurrent in chloroplasts related to light-induced swelling of thylakoid system. *Bioelectrochem Bioenerg* **46**: 71–78
- Butterfass T** (1979) Patterns of chloroplast reproduction: a developmental approach to protoplasmic plant anatomy. Springer-Verlag, Wien ; New York
- Cherkashin AA, Bulychev AA, Vredenberg WJ** (1999) Outward photocurrent component in chloroplasts of *Peperomia metallica* and its assignment to the 'closed thylakoid' recording configuration. *Bioelectrochem Bioenerg* **48**: 141–148
- Christensen-Dean GA, Moore R** (1993) Development of chlorenchyma and window tissues in leaves of *Peperomia columella*. *Ann Bot* **71**: 141–146
- De Grooth BG, Van Gorkom HJ** (1981) External electric field effects on prompt and delayed fluorescence in chloroplasts. *Biochim Biophys Acta BBA - Bioenerg* **635**: 445–456
- Egbert KJ, Martin CE** (2002) The influence of leaf windows on the utilization and absorption of radiant energy in seven desert succulents. *Photosynthetica* **40**: 35–39
- Egbert KJ, Martin CE** (2000) Light penetration via leaf windows does not increase photosynthesis in three species of desert succulents. *J Plant Physiol* **157**: 521–525

- Egydio APM, Yoshida NC, Gutiérrez LYV, Yamaguchi LLF, Motta LB, Tepe E, Salatino ML, Salatino A, Kato MJ** (2014) Phylogenetic studies of Piper and Peperomia: Reconstruction of secondary metabolites evolution. *Planta Med* **80**: P2O31
- Ellis JR, Leech RM** (1985) Cell size and chloroplast size in relation to chloroplast replication in light-grown wheat leaves. *Planta* **165**: 120–125
- Evert RF, Eichhorn SE** (2006) *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development*, 3rd ed. John Wiley & Sons, Inc., Hoboken, New Jersey
- Fosberg RF, Sachet M-H** (1975) *Flora of Micronesia, 2: Casuarinaceae, Piperaceae, and Myricaceae*. *Smithson Contrib Bot* 1–32
- Frenzke L, Goetghebeur P, Neinhuis C, Samain M-S, Wanke S** (2016) Evolution of Epiphytism and Fruit Traits Act Unevenly on the Diversification of the Species-Rich Genus *Peperomia* (Piperaceae). *Front Plant Sci* **7**: 13410
- Frenzke L, Scheiris E, Pino G, Symmank L, Goetghebeur P, Neinhuis C, Wanke S, Samain M-S** (2015) A revised infrageneric classification of the genus *Peperomia* (Piperaceae). *Taxon* **64**: 424–444
- Garnier S** (2018) viridis: Default Color Maps from “matplotlib.”
- Glover BJ, Whitney HM** (2010) Structural colour and iridescence in plants: the poorly studied relations of pigment colour. *Ann Bot* **105**: 505–511
- Gould KS, Lee DW** (1996) Physical and ultrastructural basis of blue leaf iridescence in four Malaysian understory plants. *Am J Bot* **83**: 45–50
- Graham RM, Lee DW, Norstog K** (1993) Physical and ultrastructural basis of blue leaf iridescence in two neotropical ferns. *Am J Bot* **80**: 198–203
- Gutierrez YV, Yamaguchi LF, Moraes MM de, Jeffrey CS, Kato MJ** (2016) Natural products from *Peperomia*: occurrence, biogenesis and bioactivity. *Phytochem Rev* **15**: 1009–1033
- Honda SI, Hongladarom-Honda T, Kwanyuen P, Wildman SG** (1971) Interpretations on chloroplast reproduction derived from correlations between cells and chloroplasts. *Planta* **97**: 1–15
- Irmak L** (1957) The chloroplasts of sun and shade leaves. *Rev Fac Sci Univ Istanb Ser B* **22**: 191–95
- Khoshravesh R, Sage TL** (2018) Creating Leaf Cell Suspensions for Characterization of Mesophyll and Bundle Sheath Cellular Features. *In* S Covshoff, ed, *Photosynthesis*. Springer New York, New York, NY, pp 253–261

- Kubitzki K, Rohwer JG, Bittrich V, eds** (1993) Flowering Plants · Dicotyledons. doi: 10.1007/978-3-662-02899-5
- Kuo-Huang L-L, Ku MSB, Franceschi VR** (2007) Correlations between calcium oxalate crystals and photosynthetic activities in palisade cells of shade-adapted *Peperomia glabella*. *Bot Stud* **48**: 155–164
- Lee D, Hua L, Khoshravesh R, Giuliani R, Kumar I, Cousins A, Sage TL, Hibberd JM, Brutnell TP** (2021) Engineering chloroplast development in rice through cell-specific control of endogenous genetic circuits. *Plant Biotechnol J* **19**: 2291–2303
- Lee DW** (2007) Nature's palette: the science of plant color. University of Chicago Press, Chicago
- Leech RM, Pyke KA** (1988) Chloroplast division in higher plants with particular reference to wheat. *Div. Segreg. Organelles*. Cambridge University Press, New York, pp 39–62
- Lichtenthaler HK, Buschmann C, Döll M, Fietz HJ, Bach T, Kozel U, Meier D, Rahmsdorf U** (1981) Photosynthetic activity, chloroplast ultrastructure, and leaf characteristics of high-light and low-light plants and of sun and shade leaves. *Photosynth Res* **2**: 115–141
- Machado RD, Sant'Anna WR, Benchimol M** (1986) Internal dimorphism of chloroplasts in the leaves of *Peperomia pellucida* (L.) H.B.K. *Protoplasma* **134**: 73–77
- Mahiou V, Roblot F, Hocquemiller R, Cavé A, Arias ARD, Inchausti A, Yaluff G, Fournet A** (1996) New Prenylated Quinones from *Peperomia galioides*. *J Nat Prod* **59**: 694–697
- Mahiou V, Roblot F, Hocquemiller R, Cavé A, Barrios AA, Fournet A, Ducrot P-H** (1995) Piperogalin, a new prenylated diphenol from *Peperomia galioides*. *J Nat Prod* **58**: 324–328
- Pyke K** (2009) *Plastid biology*. Cambridge University Press, Cambridge, UK ; New York
- Pyke KA** (1997) The genetic control of plastid division in higher plants. *Am J Bot* **84**: 1017–1027
- Pyke KA, Leech RM** (1987) The control of chloroplast number in wheat mesophyll cells. *Planta* **170**: 416–420
- Pyke KA, Leech RM** (1992) Chloroplast Division and Expansion Is Radically Altered by Nuclear Mutations in *Arabidopsis thaliana* 1. *Plant Physiol* **99**: 1005–1008
- Pyke KA, Leech RM** (1991) Rapid image analysis screening procedure for identifying chloroplast number mutants in mesophyll cells of *Arabidopsis thaliana* (L.) Heynh. *Plant Physiol* **96**: 1193–1195
- Rasingam L, Parthasarathy N** (2009) Diversity of understory plants in undisturbed and disturbed tropical lowland forests of Little Andaman Island, India. *Biodivers Conserv* **18**: 1045–1065

Richards LA, Dyer LA, Forister ML, Smilanich AM, Dodson CD, Leonard MD, Jeffrey CS (2015) Phytochemical diversity drives plant–insect community diversity. *Proc Natl Acad Sci* **112**: 10973–10978

Ruzin SE (1999) *Plant microtechnique and microscopy*. Oxford University Press

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676–682

Sheue CR, Sarafis V, Kiew R, Liu HY, Salino A, Kuo-Huang LL, Yang YP, Tsai CC, Lin CH, Yong JWH, et al (2007) Bizonoplast, a unique chloroplast in the epidermal cells of microphylls in the shade plant *Selaginella erythropus* (*Selaginellaceae*). *Am J Bot* **94**: 1922–1929

Simmonds SE, Smith JF, Davidson C, Buerki S (2021) Phylogenetics and comparative plastome genomics of two of the largest genera of angiosperms, *Piper* and *Peperomia* (*Piperaceae*). *Mol Phylogenet Evol* **163**: 107229

Soltis DE, Soltis PS, Endress PK, Chase MW, Manchester SR, Judd WS, Majure LC, Mavrodiev E (2018) *Phylogeny and evolution of the angiosperms*, Revised and updated edition. The University of Chicago Press, Chicago, IL

Team RC (2020) *R: A Language and Environment for Statistical Computing*. Vienna, Austria

Thomas KR, Kolle M, Whitney HM, Glover BJ, Steiner U (2010) Function of blue iridescence in tropical understory plants. *J R Soc Interface* **7**: 1699–1707

Tsel'niker YL (1973) Rhythms of tissue and chloroplast growth and determination of characteristics of light and shade structure of the leaf in Norway maple. *Sov. Plant Physiol.*

Vredenberg WJ, Tonk WJM (1975) On the steady-state electrical potential difference across the thylakoid membranes of chloroplasts in illuminated plant cells. *Biochim Biophys Acta BBA - Bioenerg* **387**: 580–587

Wanke S, Samain M -S., Vanderschaeve L, Mathieu G, Goetghebeur P, Neinhuis C (2006) Phylogeny of the genus *Peperomia* (*Piperaceae*) inferred from the *trnK/matK* region (cpDNA). *Plant Biol* **8**: 93–102

Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York

Yu D, Yang X, Lu X, Shi L, Feng B (2016) Ethyl acetate extract of *Peperomia tetraphylla* induces cytotoxicity, cell cycle arrest, and apoptosis in lymphoma U937 cells. *Biomed Pharmacother* **84**: 1802–1809

Chapter 3. Establishing novel *de novo* transcriptome assemblies for *Peperomia dahlstedtii* and *Peperomia pellucida*

Introduction

Chapter 1 reviewed literature establishing that variation in chloroplast population morphology can impact photosynthesis and likely plant fitness, and that there is variation in this trait both in *A. thaliana* mutants and in natural plant populations such as *Peperomia*. Though this trait is clearly important, we do not know how it is regulated. In Chapter 2, I showed that the large chloroplast phenotype found in the palisade cells of *P. pellucida* is the result of a complete lack of chloroplast division during leaf expansion (Figure 2.4). Though a lack of chloroplast division explains why we see a small population of enlarged chloroplasts in the palisade cells of *P. pellucida*, we still do not know how the cell senses chloroplast population morphology and sets the rate of chloroplast division. Could expression of one of the chloroplast division genes (described in Chapter 1) be regulated differently in the palisade cells of *P. pellucida*, or is some other mechanism involved? Presumably, there is some regulatory mechanism by which the cell perceives and adjusts its chloroplast population. This mechanism is related to the timing and frequency of chloroplast division. Understanding how this process plays out in the *Peperomia* spp. where division does not occur in the palisade cells will provide novel insight into this process.

In order to identify genes that might regulate chloroplast population morphology and the timing and rate of chloroplast division, I decided to investigate gene expression differences in the two model species I established in Chapter 2—*P. dahlstedtii* (small chloroplasts) and *P. pellucida* (large chloroplasts; Figure 2.3). Genomic resources for *Peperomia* are scarce; no genome has been sequenced, and only two *de novo* transcriptome assemblies are currently available—based on RNA sequencing from the mature leaves of *P. obtusifolia* (Batista et al., 2017) and various mature tissues of *P. fraseri* (Matasci et al., 2014), neither of which have large

chloroplasts (Figure 2.1). To understand the mechanism behind the inhibition of chloroplast division in *P. pellucida*, gene expression data were needed. Here I describe the generation of novel RNA sequencing data and *de novo* transcriptome assemblies for *P. pellucida* and *P. dahlstedtii*.

Results

Generation and assessment of novel de novo transcriptome assemblies for P. pellucida and P. dahlstedtii

As described in Chapter 2, *P. dahlstedtii* and *P. pellucida* were selected as models for the small and large chloroplast phenotypes, respectively (Figure 2.3). Very young expanding leaves, where chloroplast division is active in *P. dahlstedtii*, from both species were collected for RNA sequencing (stage A; Figure 2.4). While *P. fraseri* and *P. obtusifolia* do not exhibit chloroplast population morphology similar to *P. pellucida* (Figure 2.1), their previously reported *de novo* transcriptome assemblies, which are derived exclusively from mature tissues, were obtained (Matasci et al., 2014; Batista et al., 2017) and assessed alongside those described here for comparison of assembly quality and completeness. Similarly, a *de novo* transcriptome assembly was also generated from publicly available RNA sequencing data derived from young *A. thaliana* expanding leaf tissue (Klepikova et al., 2016) for comparison to the *Peperomia* assemblies.

De novo transcriptome assembly statistics are shown in Table 3.1. Overall, the novel *Peperomia* assemblies from *P. pellucida* and *P. dahlstedtii* were generated using more than twice the amount of sequencing than *P. obtusifolia* or *P. fraseri*. Thus, my assemblies have better coverage and are likely to be more complete. To assess the correct assembly of transcripts in my two species of interest, reads were mapped to their respective assemblies. For *P. pellucida*, ~88%

of reads were mapped, and for *P. dahlstedtii*, 92% (Table 3.2). These are highly satisfactory numbers, indicating good agreement between read alignment and orientation in the assembled transcripts.

To assess the completeness of predicted proteins in the assemblies, Embryophyta (land plant)-specific benchmarked universal single-copy orthologs (BUSCOs) were identified and quantified using BUSCO (Simão et al., 2015). Overall, *P. pellucida* and *P. dahlstedtii* had the lowest numbers of missing BUSCOs (Figure 3.1A; red). Following *A. thaliana*, *P. dahlstedtii* had the largest proportion of complete BUSCOs (Figure 3.1A).

To compare relative gene content and eventually expression between species, groups of orthologous genes were clustered into orthogroups (Emms and Kelly, 2015). Orthogroup clustering revealed that 6,390 orthogroups were shared between all species and assemblies represented (Figure 3.1B). The next largest cluster of orthogroups (2,263) was shared between the two *Peperomia* species, as expected since they are very closely related. The *A. thaliana* transcriptome assembly had no unique orthogroups (Figure 3.1B), meaning these *de novo*-assembled sequences were ‘correct’ enough to be grouped with others, mainly its corresponding genome.

For a more thorough investigation of *Peperomia* assembly BUSCO-based completeness, the percent alignment between individual *Peperomia* BUSCO sequences and the corresponding embryophyta reference sequence (Simão et al., 2015) was extracted and plotted (Figure 3.1C). In all assemblies, the majority of sequences aligned at 100%, indicating that these sequences are likely complete (Figure 3.1C).

Table 3.1. Basic statistics for the *de novo* transcriptome assemblies.

Species	% GC	Total assembled bases	Total genes	Total transcripts	Median contig length (bp)	Average contig length (bp)
<i>P. dahlstedtii</i> ¹	42.34	168,362,854	27,810	105,731	1,291	1,596.9
<i>P. pellucida</i> ²	42.38	213,829,199	36,502	143,653	1,158	1,482.1
<i>P. fraseri</i> ³	44.35	44,505,560	29,073	39,899	868	1,076.9
<i>P. obtusifolia</i> ⁴	42.56	94,998,144	25,757	63,639	1,182	1,494.4

¹ Very young expanding leaves

² Very young expanding leaves and whole seedlings

³ From the 1 KP project, Matasci et al. (2014); mature leaf, root, and shoots

⁴ From Batista et al. (2017); mature leaves

Table 3.2. Alignment rates of paired end RNAseq reads to their respective *de novo* transcriptome assembly.

Species	Overall alignment rate (%)	Concordant alignment rate exactly 1 time (%)	Concordant alignment rate >1 time (%)
<i>P. dahlstedtii</i>	92.2	16.21	64.47
<i>P. pellucida</i>	88.31	14.59	62.14

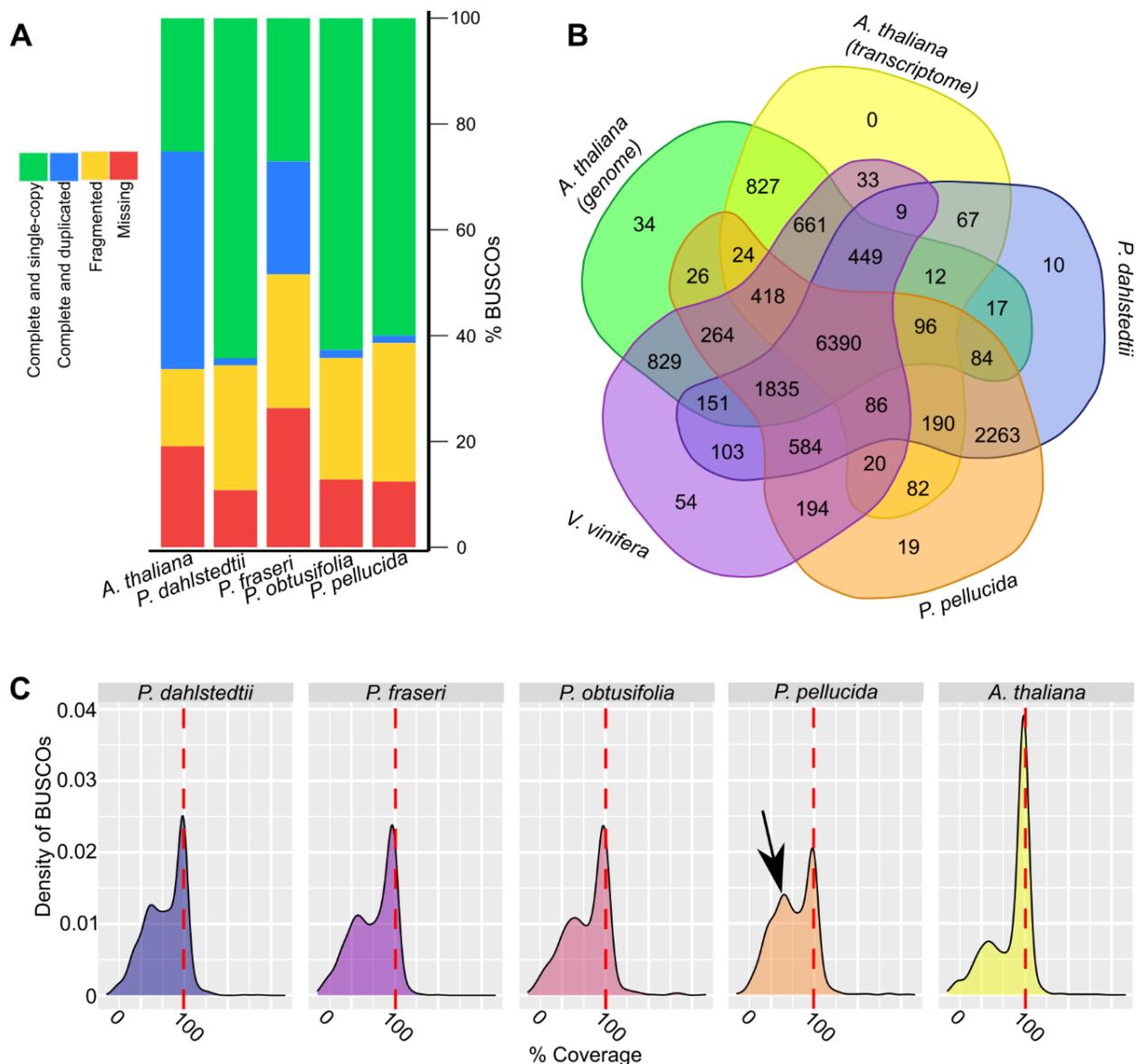


Figure 3.1. Assessing the quality of the *de novo* transcriptome assemblies.

(A) The percent of BUSCOs identified as complete and single-copy (green), complete and duplicated (blue), fragmented (yellow), or missing (red) for each *de novo* transcriptome assembly. (B) Orthogroup clustering between the *de novo* transcriptome assembly-predicted proteomes generated for *P. pellucida*, *P. dahlstedtii*, and *A. thaliana*, with reference genome-based proteomes from *A. thaliana* and *V. vinifera*. (C) Density plots for each of the *de novo* transcriptome assemblies generated or utilized, showing the distributions of the percent alignment (i.e., coverage) between the species-specific BUSCO hit and that of the reference sequence used by BUSCO, where 100% (vertical red dashed lines) or greater means the alignment was full length. The black arrow indicates one of the ‘humps’ referred to in the text.

Though the *A. thaliana* assembly had an overwhelming majority of alignments at 100%, all four *Peperomia* assemblies all had a sizeable proportion of alignments at ~50% (black arrow; Figure 3.1C)—indicating an abundance of shorter sequences in these transcriptomes, meaning they may be truncated or fragmented.

Representation of the chloroplast division orthogroups

Based on the similarity between the chloroplast population morphology in the palisade mesophyll cells of *P. pellucida* (Figure 2.3) and several *A. thaliana* mutants for the chloroplast division genes described in Chapter 1, I hypothesized that expression of one or more chloroplast division genes may be regulated differently in the palisade cells of *P. pellucida*. Though it would have been ideal to sequence palisade and spongy mesophyll cells separately, attempts at laser capture microdissection were not effective in these species, primarily due to the high water content of the leaves (window cells; Figure 2.3A). Thus, the expression analysis described here for the chloroplast division genes was based on whole-tissue mRNA sequencing in *P. pellucida* and *P. dahlstedtii*.

Orthogroup expression (Benoît Boachon et al., 2018; Liu et al., 2018) was used to assess differences in expression between species. To account for biological variation in gene expression between individual plants, three biological replicates were sequenced for each species. To ensure that the replicates within each species were similar in terms of gene expression patterns, relative expression levels for all genes were compared across the samples using a Spearman's correlation test, and only those with a correlation coefficient >0.87 were retained (Table 3.3). Using the same orthogroups shown earlier (Figure 3.1B), gene expression was calculated as transcripts per million (TPM; Wagner et al., 2012) for each orthogroup from each replicate. Finally, TPM was summed across each species, where the TPM values for genes in the same orthogroup were

Table 3.3. Spearman's correlation tests between RNAseq replicates.

Species			
<i>P. pellucida</i>			
	Rep 1	Rep 2	Rep 3
Rep 1	1	0.989603	0.887129
Rep 2	0.989603	1	0.885916
Rep 3	0.887129	0.885916	1
<i>P. dahlstedtii</i>			
	Rep 1	Rep 2	Rep 3
Rep 1	1	0.976703	0.945264
Rep 2	0.976703	1	0.932755
Rep 3	0.945264	0.932755	1
<i>A. thaliana</i>			
	Rep 1	Rep 2	
Rep 1	1	0.98747	
Rep 2	0.98747	1	

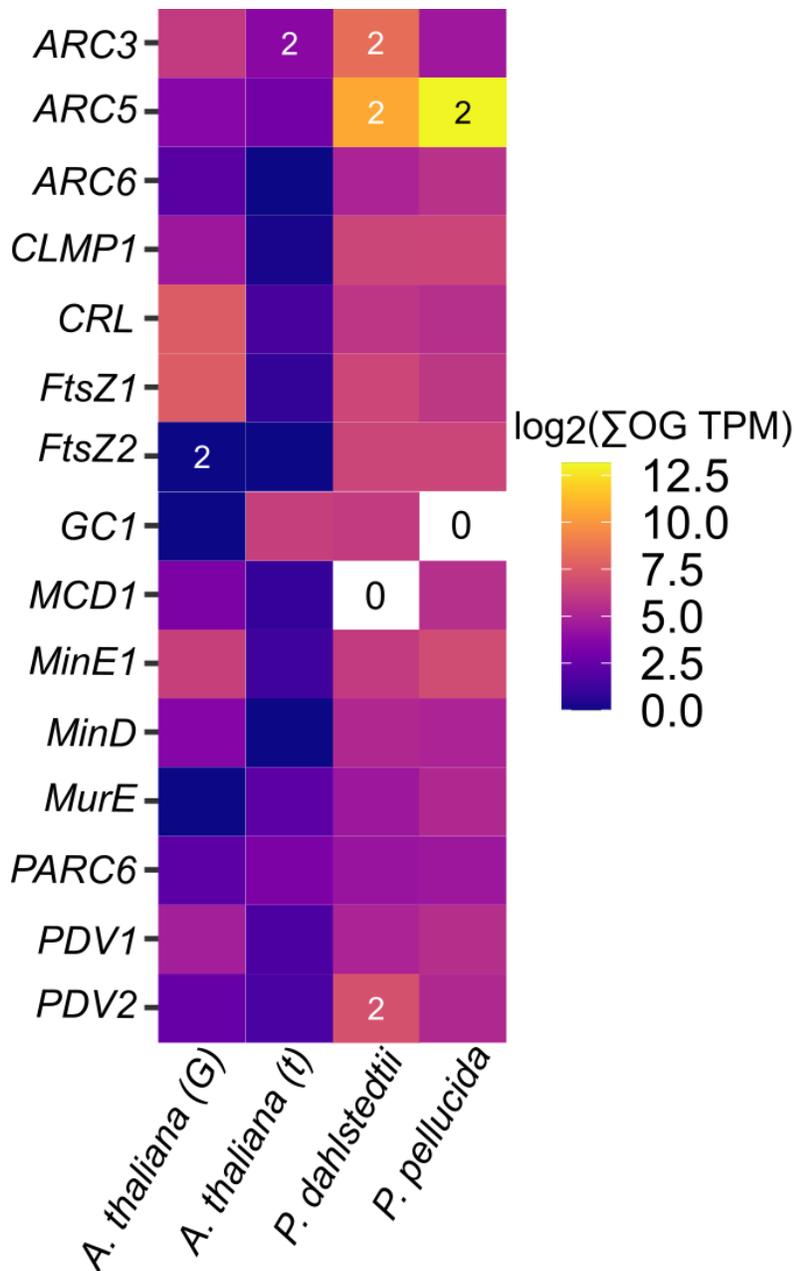


Figure 3.2. Chloroplast division orthogroup expression and copy number.

The heatmap shows the expression of the chloroplast division orthogroups indicated on the left in the four species shown at the bottom. The expression values represent the \log_2 value for the sum of the transcripts per million (TPM) values for all genes identified in each orthogroup. Copy number (the number of predicted genes present in a given orthogroup) is one unless indicated otherwise in the individual tiles. (G), genome; (t) transcriptome; 0's in the tiles (MCD1 and GC1) indicate that no transcripts were identified for that gene; OG, orthogroup.

added, yielding the expression value for each orthogroup. For visualization and ease of interpretation, these expression values are shown as the \log_2 of the orthogroup TPM (Figure 3.2). Overall ~70-80% of reads, from the young and expanding leaf datasets specifically, aligned to the final (longest isoform) transcriptome assemblies for both *P. pellucida* and *P. dahlstedtii*. Together, these data indicate that there was good alignment and consistency between replicate RNAseq samples.

I first compared expression of the chloroplast division orthogroups from *P. dahlstedtii* and *P. pellucida*, as I expected to find some differences in gene expression correlating with active chloroplast division in *P. dahlstedtii* and no division in *P. pellucida* (Figure 2.4). Orthologs for all chloroplast division genes (Chen et al., 2018) were identified in both species, except *GIANT CHLOROPLAST 1 (GCI)* (also known as Sula) (Maple et al., 2004; Raynaud et al., 2004) in *P. pellucida* (Pp) and *MULTIPLE CHLOROPLAST DIVISION SITE 1 (MCD1)* (Nakanishi et al., 2009) in *P. dahlstedtii* (Pd) (Figure 3.2). Truncated candidate *PpGCI* and *PdMCD1* transcripts were recovered from a search against all predicted transcripts (not just the longest open reading frames), though they contained many premature stop codons.

All *Peperomia* spp. except *P. fraseri* contained severely truncated MCD1 sequences lacking the N-terminal chloroplast transit peptide and conserved transmembrane domain (Figure 3.3). *PdMCD1* was also truncated at the C-terminus and lacked the conserved coiled-coil domain (Figure 3.3; Nakanishi et al., 2009). None of these *Peperomia* species exhibit the *Atmcd1* phenotype, where chloroplasts heterogeneous in size are found due to asymmetric division (Nakanishi et al., 2009). However, it is possible that loss of MCD1 could be associated with the larger-sized chloroplasts of *Peperomia* species in general compared to model plants such as *A. thaliana*.

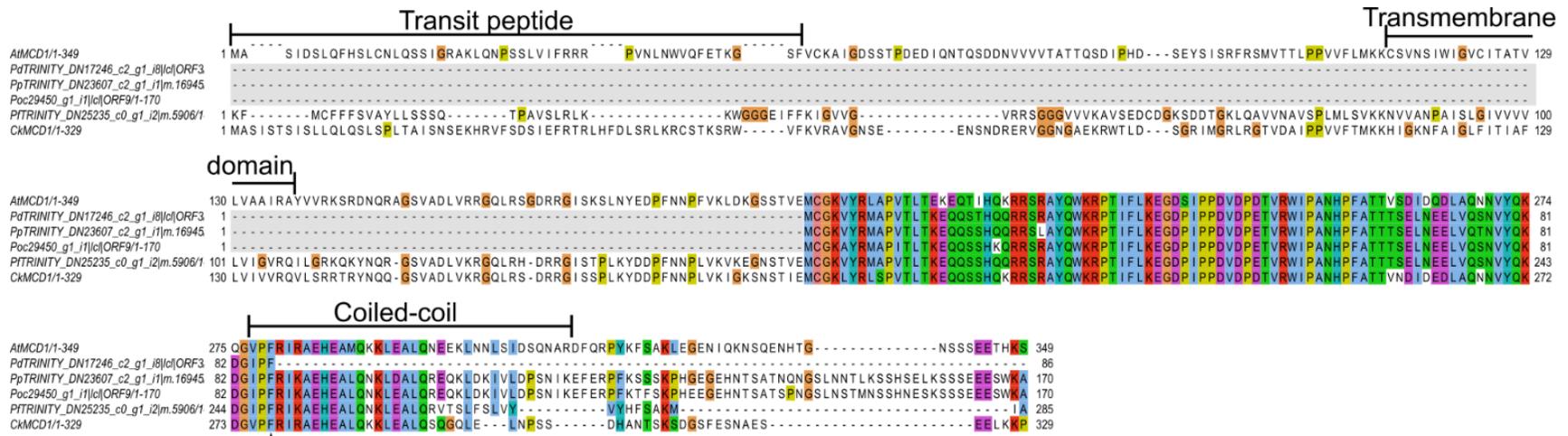


Figure 3.3. Possible MCD1 truncations within *Peperomia*.

MCD1 sequences from (top to bottom) *A. thaliana*, *P. dahlstedtii*, *P. pellucida*, *P. obtusifolia*, *P. fraseri*, and *Cinnamomum kanehirae* (stout camphor tree, the closest-related genome available at the time). Important domains identified in *A. thaliana* (Nakanishi et al., 2009) are annotated above with black lines. The severe early truncation in *P. dahlstedtii* is indicated by a black arrow below the alignment.

GIANT CHLOROPLAST 1 appears to be mutated to different degrees in Peperomia species

GC1 appears to be at least partially involved in chloroplast division in *A. thaliana*, making it a high priority gene of interest for the regulation of chloroplast size and number. Altering *GCI* expression levels using knock-down lines in *A. thaliana* produces mesophyll cells with a smaller population of enlarged chloroplasts than in wild type (Maple et al., 2004; Raynaud et al., 2004). Two assembled transcript sequences with >70% identity to *AtGCI* were identified in *P. pellucida*. However, the 6-frame translations for both transcripts were riddled with stop codons, yielding truncated predicted protein sequences, the longest and most complete of which is shown in Figure 3.4. GC1 is predicted to have an epimerase domain (Maple et al., 2004; Li et al., 2017) with active site residues at S161 and Y168, both of which are substituted in *P. pellucida* (Figure 3.4A). In *P. obtusifolia*, which does not have large chloroplasts (Figure 2.1), GC1 has a more severe truncation than in *P. pellucida* and is missing these key residues entirely (Figure 3.4A). Overall, all four *Peperomia* GC1 sequences contain multiple deletions or truncations varying in severity. Considering this evidence, *GCI* is a suitable candidate for the large-chloroplast phenotype observed in *P. pellucida*.

A. thaliana mutants for *GCI* have been observed for alterations to chloroplast size and number, but enlarged chloroplasts were not observed frequently or consistently (Li et al., 2017). However, all previously studied mutant alleles retained exon 7, where the predicted epimerase active site residues are located (Figure 3.4B). To address the possibility that retention of these active site residues in a truncated protein product might result in a milder phenotype, I obtained and phenotyped two T-DNA mutants—*Atgcl-1* and *Atgcl-2* (red arrows; Figure 3.4B). *Atgcl-1* is disrupted at the 5' end of exon 5, upstream of the active site residues. *Atgcl-2*, originally phenotyped by Li et al., (2017), is disrupted at the 3' end of exon 8, downstream of the active

site residues (Figure 3.4B). Neither mutant appeared to contain visibly larger chloroplasts compared to wild-type (Figure 3.4C-E). However, both mutants had reduced numbers of chloroplasts per cell with increasing cell size compared to wild-type, though they did not appear to differ from one another (Figure 3.4F). Overall, these data suggest that GC1 may play a role in chloroplast division, though it is likely not as direct as the division genes described at length in Chapter 1.

Identification of a potential gene of interest behind the ‘large chloroplast’ phenomenon in vascular plants

Enlarged-chloroplast phenotypes, similar to those I have observed in *P. pellucida*, have been reported in *Selaginella* species and in *Theobroma cacao* (cocoa), both of which are shade-preferring tropical species (Baker et al., 1975; Butterfass, 1979; Sheue et al., 2007; Liu et al., 2020). I wondered if the shared enlarged-chloroplast morphology observed in these three species might be caused by a common genetic factor. Using conservatory samples, I confirmed the phenotypes reported for *Selaginella* and cocoa in mature microphylls or leaves, respectively (Figure 3.5A-C). Both species have reference genome assemblies available, and orthogroup clustering between a set of small-chloroplast (*A. thaliana*, *V. vinifera*, *P. dahlstedtii*, and *P. obtusifolia*) and large-chloroplast (*T. cacao*, *S. moellendorffii*, and *P. pellucida*) species was done to identify orthogroups that might be uniquely present in or absent from one morphologically-distinct group over the other. Indeed, the *BolA1* orthogroup was missing from *T. cacao*, *S. moellendorffii* and *P. pellucida*, but was identified in all other species (Table 3.1). As for *PpGC1* and *PdMCD1* described above, I checked all predicted *P. pellucida* transcripts for *AtBolA1*-like sequences and found none, likely indicating that *BolA1* is either not present in the

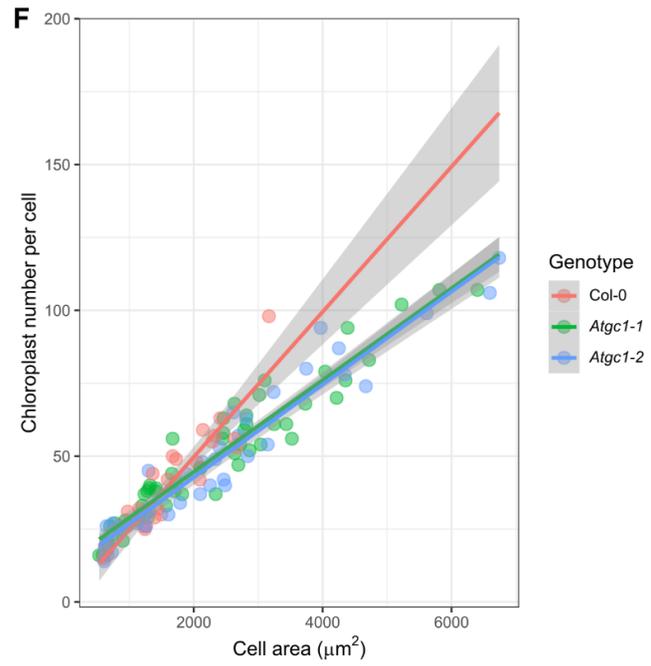
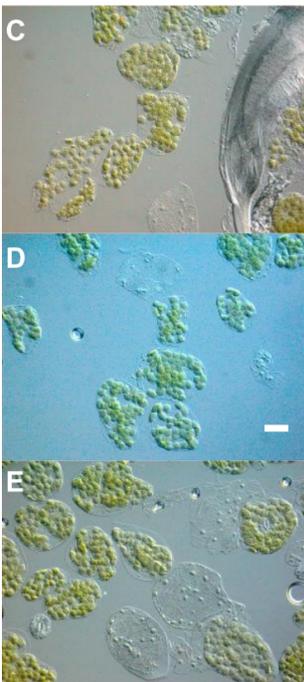
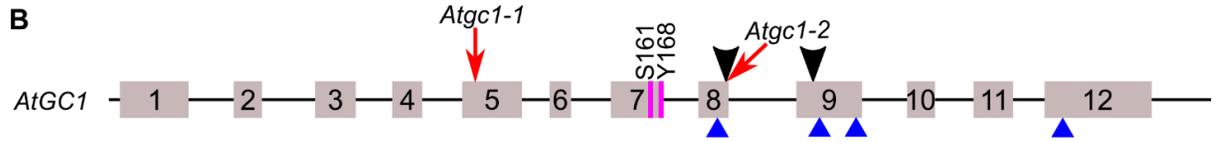
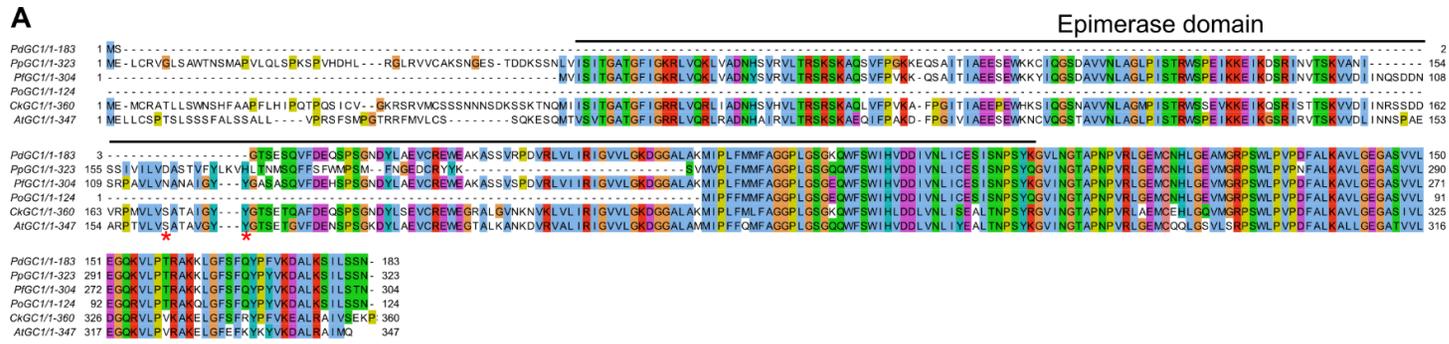


Figure 3.4. GC1 may be truncated, to varying extents, within *Peperomia*.

Figure 3.4 (cont'd)

(A) Protein sequence alignment of GC1 from the *Peperomia* species, *A. thaliana*, and *C. kanehirae*. The epimerase domain is indicated above and predicted active site residues by red asterisks below the *A. thaliana* sequence. (B) The exon structure of *AtGC1*, where exons are indicated by taupe boxes and relevant ones numbered, and introns and UTRs by a black line. The T-DNA insertion sites in the mutants used in this study are indicated by red arrows. T-DNA and CRISPR/Cas9 mutants used previously (Li et al., 2017) are indicated by black arrowheads above and blue triangles below. The locations of the predicted active site residues are shown in pink. (C-E) Fixed leaf mesophyll cells from (C) wild-type (Columbia-0), (D) *Atgc1-1* (CS65651), and (E) *Atgc1-2* (SALK_100683). The scale bar represents 20 μm for all images. (F) Correlation between chloroplast number and cells size. The slopes and R^2 values for each genotype were as follows: Col-0, 0.025 and 0.85; *Atgc1-1*, 0.016 and 0.92; *Atgc1-2*, 0.014 and 0.82. Transparent grey areas surrounding each line of best fit represents the 95% confidence intervals.

P. pellucida genome or was not expressed at high enough levels for reconstruction by *de novo* assembly.

To determine whether BolA1 might be worth pursuing as a potential regulator of chloroplast size, I looked into the literature on *AtBolA1* and found that it is a homolog of *Escherichia coli* BolA, which plays a role in bacterial cell division and is dependent upon FtsZ (Aldea et al., 1988; Aldea et al., 1989; Santos et al., 1999). Importantly, many of the chloroplast division genes discovered in *A. thaliana* were identified based on their homology to bacterial cell division genes (Colletti et al., 2000; Itoh et al., 2001; Vitha et al., 2003). Indeed, the *AtFtsZs* were first discovered for their homology to bacterial FtsZs (Osteryoung and Vierling, 1995; Osteryoung et al., 1998).

Given the possibility that plant BolAs might be involved in chloroplast division, perhaps via interaction with the FtsZs, I investigated the phylogenetics of this protein group. There are four *BolA* genes in *A. thaliana*—*AT1G55805* (*AtBolA1*), *AT4G26500* (*AtSufE1*), *AT5G17560* (*AtBolA4*), and *AT5G09830* (*AtBolA2*) (Couturier et al., 2014; Qin et al., 2015). BolA2 and BolA4 diverged prior to BolA1 and SufE1 and will not be considered further. Protein sequence alignments for BolA1s and SufE1s identified in nine plant species, including cocoa and *P. pellucida*, were made. BolA1 and SufE1 sequences share high sequence similarity, and both contain a BolA domain at their C-terminus (Figure 3.5D-E; Qin et al., 2015). BolA1 sequences are best-identified for their lack of a SufE domain at the N-terminus (Figure 3.5D-E). *P. pellucida* was the only *Peperomia* species lacking a BolA1 sequence, but did have a SufE1 (Figure 3.5D). Similarly, *T. cacao* contains one SufE1, but no BolA1 (Figure 3.5D; Table 3.4).

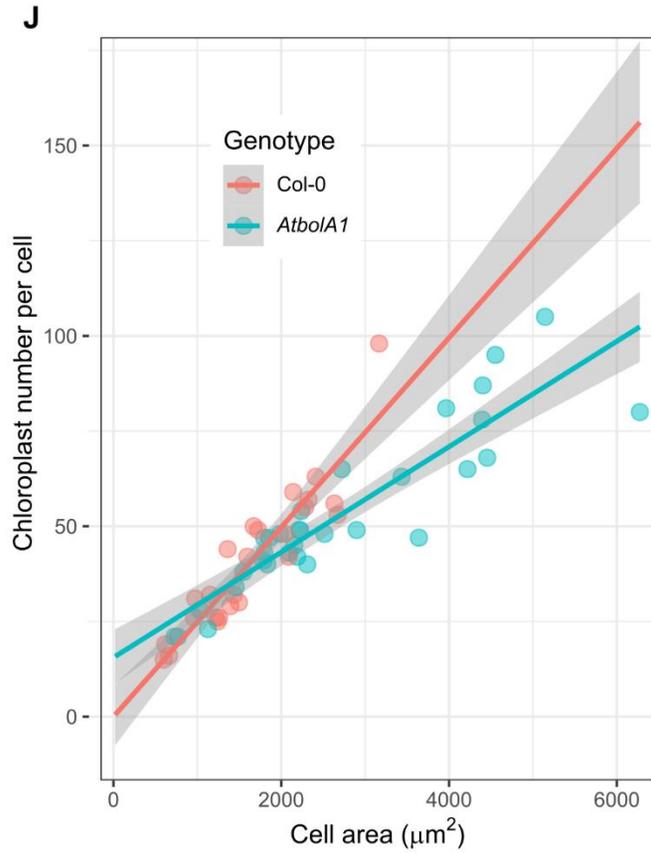
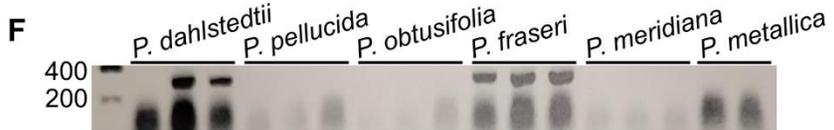
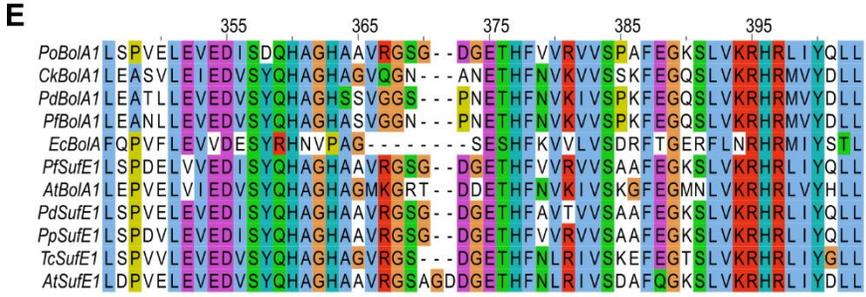
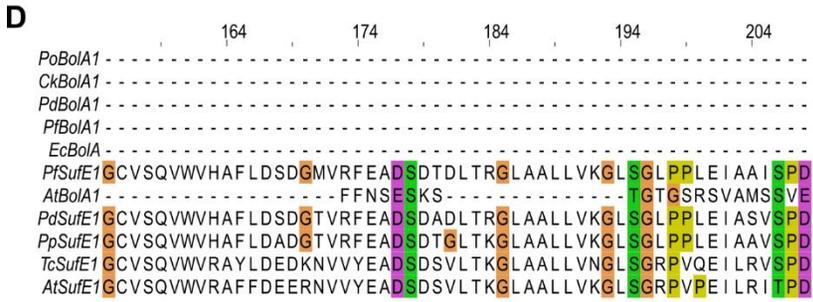
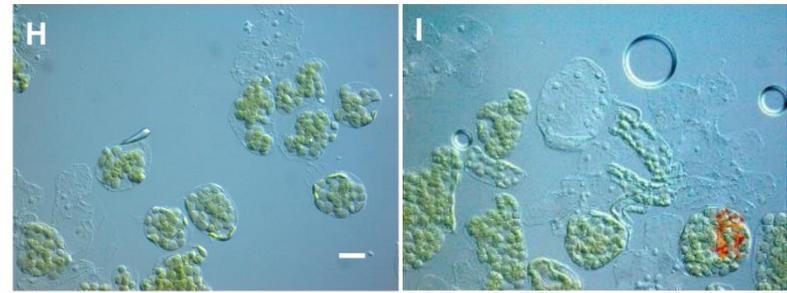
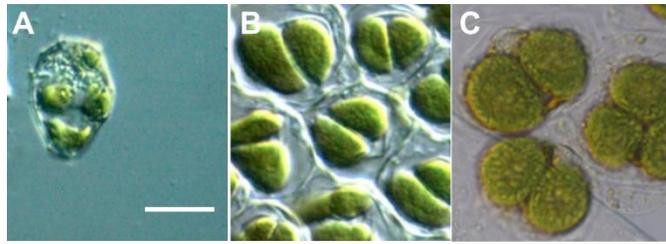


Figure 3.5. *Peperomia* *BolA1* and the mutant phenotype of *AtbolA1*.

Figure 3.5 (cont'd)

(**A**) Fixed mature leaf tissue from *T. cacao*, (**B**) *S. moellendorffii*, and (**C**) *P. pellucida*. The scale bar represents 20 μm for all three images. (**D**) Protein sequence alignments for the SufE and (**E**) BolA domains in candidate BolA1 and SufE1 proteins. (**F**) DNA agarose gels showing the PCR products amplified from *Peperomia* spp. with primers specific to *BolA1* and (**G**) targeting the BolA domain found in both *SufE1* and *BolA1*. Two or three biological replicates are shown for each species. (**H**) Fixed leaf tissue from wild type *A. thaliana* and (**I**) *AtbolA1* (SALK_022246; insertion in the 5' untranslated region). The scale bar represents 20 μm for both images. (**J**) Linear regressions for chloroplast number per cell relative to cell area. The slopes and R^2 values for each genotype were as follows: Col-0, 0.025 and 0.85; *AtbolA1*, 0.016 and 0.82. Transparent grey areas surrounding each regression line of best fit represent the 95% confidence intervals.

Table 3.4. *BolA1* was not detected in species with a small population of large chloroplasts per cell.

Species	Genes present in orthogroup OG0009268 (<i>BolA1</i>)
<i>A. thaliana</i>	AT1G55805.1 (<i>BolA1</i>)
<i>P. dahlstedtii</i>	TRINITY_DN6675_c0_g1_i1 m.24425
<i>P. obtusifolia</i>	c22121_g4_i1 m.19970
<i>P. pellucida</i>	-
<i>S. moellendorffii</i>	-
<i>T. cacao</i>	-
<i>V. vinifera</i>	VIT_217s0053g00420.1

I then investigated whether *BolA1* was indeed missing from the genome of *P. pellucida* by attempting to genotype for *BolA1* by PCR amplification with genomic DNA. Primers were designed from conserved regions identified by aligning coding sequences from the *Peperomia* spp., *Cinnamomum kanehirae* (stout camphor tree; magnoliid), and *A. thaliana*. Indeed, *BolA1*-specific primers yielded no product in *P. pellucida* (Figure 3.5F), while primers targeting the *BolA* domain found in *BolA1* and *SufE* did (Figure 3.5G). The same was observed for *P. metallica* (Figure 3.5G), which also has large chloroplasts (Fig. 2.1). Further, both sets of primers amplified fragments in *P. dahlstedtii* and *P. fraseri* (Figure 3.5F-G). The lack of PCR products from *P. obtusifolia* and *P. meridiana* was likely due to poor-quality DNA extractions, which were very low in yield (data not shown).

To observe what impact the potential loss of *BolA1* might have on chloroplast size, an *A. thaliana* T-DNA mutant was obtained and phenotyped for chloroplast morphology (Figure 3.5H-J). A slight reduction in the number of chloroplasts per unit cell area was observed in *AtbolA1* compared to wild-type (Figure 3.5J), indicating that chloroplast division may indeed be impaired with loss of *BolA1*.

Discussion

The overall goal of my work is to identify genes that regulate chloroplast population morphology using the large-chloroplast species *P. pellucida* as a model system. In *Peperomia*, the large-chloroplast phenotype is cell-specific, indicating that differences in gene expression are likely the underlying cause. Therefore, I sequenced mRNA from very young expanding leaves taken from *P. pellucida* (large chloroplasts) and *P. dahlstedtii* (small chloroplasts), and generated reference transcriptomes by *de novo* assembly for both species. I then utilized these assemblies to begin exploring expression of the chloroplast division orthogroups.

As demonstrated in Chapter 2, the large chloroplasts of *P. pellucida*, *P. metallica*, and *P. meridiana* are specifically found in the single layer of palisade cells in the leaf, not in the spongy mesophyll below. This cell-specific phenotype indicates that gene expression in the palisade and spongy mesophyll are likely distinct, as is the case in some C₄ plants with cell-specific chloroplast morphological phenotypes (Stata et al., 2014; Stata et al., 2016). I attempted laser capture microsection (LCMD) early on in the project to try and prepare cell-specific samples of palisade and spongy mesophyll from *P. pellucida* or *P. metallica* and *P. serpens* or *P. dahlstedtii*. However, due to the high water content of the leaves, sections were frequently destroyed. Further, intact sections were not able to be isolated by the laser—another water-related issue known to LCMD. Single-cell RNAseq (scRNAseq) is an increasingly popular method that has been successfully used to study cell-type-specific expression patterns in young developing tissues. I have had preliminary success in extracting protoplasts from mature leaves of *P. pellucida* and *P. metallica*. In the future, scRNAseq would be very useful for studying cell-specific expression patterns in the young leaves of these species.

Though tissue-specific gene expression would have been ideal for this project, I hypothesized that I would still be able to observe potentially significant differences in gene expression between samples of whole *P. pellucida* and *P. dahlstedtii* leaves. Additionally, I needed to generate reference transcriptomes for both species. Compared to existing *Peperomia* transcriptome assemblies, my *de novo* assemblies are more complete (Figure 3.1A). This is not surprising, as I opted for deeper coverage in my sequencing project. *P. dahlstedtii* and *P. pellucida* predicted orthogroups match the proteomes of *A. thaliana* and *V. vinifera*, another indication that the assemblies are complete and well representative (Figure 3.1B). It is important to note that orthogroups were identified after assembly filtering, which removed very short

transcripts and coding-sequence predictions. This led to the initial ‘loss’ of *GCI* and *MCD1* (Figure 3.3; Figure 3.4). While ultimately, this issue brought these proteins of interest to my attention, it also illustrates one of the pitfalls of using *de novo* transcriptome assemblies for non-model organisms.

The *de novo* transcriptome assemblies generated here were derived from very young expanding leaves because 1) chloroplast division (and presumably its regulation) occurs primarily in expanding leaves (Ellis and Leech, 1985), and 2) reference sequences of some kind were required to move forward with these species. Though I have a short list of candidate genes of interest, more work is needed to look into these genes. Further, it would be valuable to identify novel genes of interest that may contribute to the difference in phenotype. The simplest way to accomplish these goals is to do differential gene expression analysis between young expanding and mature leaves. I have demonstrated that chloroplast division occurs rapidly in the former (Chapter 2) but levels off as the leaf matures. Further, this work would be facilitated by incorporating a reference genome assembly for *Peperomia*, which would allow for a more thorough assessment of *GCI*, *BolA1*, and potentially other candidate genes of interest.

Conclusions

In this chapter, I have described the generation of novel transcriptomic datasets for *P. pellucida* and *P. dahlstedtii*, but many questions remain unanswered. The biggest issue is the fragmentation of the reconstructed sequences, which could explain the truncations and or mutations observed in *MCD1* and *GCI*. A reference genome assembly that enables comparative expression studies based on genome-guided transcriptome assembly and analysis is necessary to fully address these issues. This has been pursued in Chapter 4.

Materials and methods

Sampling tissue for sequencing

For RNA, young expanding leaves (2-3 mg for *P. dahlstedtii* and 1-6 mg for *P. pellucida*) were removed from plants using a sterile razor blade and forceps. Whole *P. pellucida* seedlings were gently removed from the soil and rinsed thoroughly with sterile nuclease-free water to remove debris. Three replicates for each tissue and species were taken. All samples were weighed, moved to a 2 mL plastic screw-cap tube (Sarstedt, 72.693) containing three sterilized glass beads, and immediately frozen in liquid nitrogen followed by storage at -80°C. Before RNA extraction, samples were ground to a powder using a FastPrep-24 benchtop tissue homogenizer (MP Biomedicals, SKU 116004500).

RNA extraction and transcriptome sequencing

Total RNA was extracted from powdered plant tissues using the RNeasy RNA Cell Miniprep System (Promega, Z6011). DNase digestion was performed on-column. Sample RNA concentrations were measured using a Qubit fluorometer (Invitrogen) and quality-checked with an Agilent 2100 Bioanalyzer by the Michigan State University Genomics Core. Samples passing quality checking (QC) were submitted to the Genomics Core for TruSeq stranded mRNA library (Illumina) preparation—where each replicate was prepared separately, after which they were multiplexed and sequenced on one PE150 lane of an Illumina HiSeq 4000.

De novo transcriptome assembly

Read quality was assessed before and after trimming using FastQC/0.11.7 (Andrews, 2010). Reads were trimmed using Trimmomatic/0.38 (Bolger et al., 2014), and only reads greater than 85 nt in length were kept for downstream analyses. De novo transcriptome assemblies (one

for each species) were built using Trinity/2.6.6 (Grabherr et al., 2011) with the --normalize_by_read_set flag and default settings.

Assessing the quality and completeness of transcriptome assemblies

Using a custom UNIX script, Trinity assemblies were filtered to remove all transcripts less than 500 nucleotides in length for all further analyses, and TransDecoder (TransDecoder/2.1.0) (Haas et al., 2013) was run to predict and translate likely protein-coding sequences.

Assessment of transcriptome assembly quality and completeness

As an assessment of assembly completeness, BLAST+ v. 2.8.1 (Camacho et al., 2009) BLASTX and BLASTP were run using the filtered Trinity assemblies as the query against a variety of reference plant proteomes (listed in Chapter 4 Methods), with the following settings: -evaluate 1e-5, -num_threads 8, -max_target_seqs 1, -outfmt 6. The Trinity script “analyze_blastPlus_topHit_coverage.pl” was run to estimate the number of full-length proteins present in the assembly compared to the reference proteomes.

In parallel, the completeness of single-copy orthologues was assessed using BUSCO version 10, and the % match (transcriptome protein: BUSCO) was plotted as a density plot for each species.

Orthogroup clustering, copy number, and expression

Orthogroups were identified from translated CDSs (predicted CDSs for the transcriptome assemblies) using OrthoFinder (Emms and Kelly, 2015). The *A. thaliana* reference proteome was included as a root for identifying and functional prediction of orthogroups. All other species used are from Phytozome v13 (<https://phytozome-next.jgi.doe.gov/>), except *P. fraseri* (Matasci et al., 2014) and *P. obtusifolia* (Batista et al., 2017).

SAM files were generated by mapping reads back to their respective transcriptome assemblies (longest isoform, with those > 500) using hisat2/2.1.0 (Kim et al., 2015). CDS-based count data were generated using htseq-count Version 0.11.0 (Anders et al., 2015) in Python/3.6.4. Count and feature data for all species was organized and cleaned in R using the following packages: tidyverse (Wickham et al., 2019), data.table (Dowle and Srinivasan, 2019), and magrittr (Bache and Wickham, 2014) (see code files “1.count_data_TPM.R” and “3.tidy_orthogroups.R” on GitHub). TPM was calculated (Wagner et al., 2012) using a custom function (see code file “1.count_data_TPM.R” on GitHub) for each replicate, which was then averaged for each species. Copy number was calculated as the number of unique genes in each orthogroup for each species.

Multiple sequence alignments and protein domain identification

Multiple sequence alignments were performed using MAFFT (version 7) with default settings (Kato et al., 2019). Alignments were assessed by hand and pruned as needed. Protein domains were identified using InterProScan (version 5.55-88.0) (Jones et al., 2014).

Data and statistical analyses

R version 3.6.3 (Team, 2020) and RStudio version 1.2.5033 (2019) were used for all data manipulation, plotting, and statistical analyses unless noted otherwise. Plots were made using ggplot2 (Wickham, 2016) using color palettes from viridis (Garnier, 2018).

Data availability

The code used for this project is available at <https://github.com/AFrolicOfFerns/peperomia-transcriptomes-2020>. The *P. dahlstedtii* and *P. pellucida* transcriptome assemblies are available at Zenodo (10.5281/zenodo.6423262). The novel mRNAseq datasets generated from *P. dahlstedtii* and *P. pellucida* were deposited with

NCBI (BioProject PRJNA824611). All code and data are set to be released publicly on July 8th, 2022.

Acknowledgments

Lisa Murphey (Michigan State University Plant Biology Teaching Conservatory) readily supplied *Selaginella* and *T. cacao* samples.

REFERENCES

REFERENCES

- Aldea M, Garrido T, Hernández-Chico C, Vicente M, Kushner SR** (1989) Induction of a growth-phase-dependent promoter triggers transcription of *bolA*, an *Escherichia coli* morphogene. *EMBO J* **8**: 3923–3931
- Aldea M, Hernández-Chico C, Campa AG de la, Kushner SR, Vicente M** (1988) Identification, cloning, and expression of *bolA*, an *ftsZ*-dependent morphogene of *Escherichia coli*. *J Bacteriol* **170**: 5169–5176
- Baker NR, Hardwick K, Jones P** (1975) Biochemical and physiological aspects of leaf development in cocoa (*Theobroma cacao*) II. Development of chloroplast ultrastructure and carotenoids. *New Phytol* **75**: 513–518
- Batista ANL, Santos-Pinto JRA dos, Batista JM, Souza-Moreira TM, Santoni MM, Zanelli CF, Kato MJ, López SN, Palma MS, Furlan M** (2017) The combined use of proteomics and transcriptomics reveals a complex secondary metabolite network in *Peperomia obtusifolia*. *J Nat Prod* **80**: 1275–1286
- Benoît Boachon, Buell CR, Crisovan E, Dudareva N, Garcia N, Godden G, Henry L, Kamileen MO, Kates HR, Kilgore MB, et al** (2018) Phylogenomic mining of the mints reveals multiple mechanisms contributing to the evolution of chemical diversity in Lamiaceae. *Mol Plant* **11**: 1084–1096
- Butterfass T** (1979) Patterns of chloroplast reproduction: a developmental approach to protoplasmic plant anatomy. Springer-Verlag, Wien ; New York
- Chen C, MacCready JS, Ducat DC, Osteryoung KW** (2018) The molecular machinery of chloroplast division. *Plant Physiol* **176**: 138–151
- Colletti KS, Tattersall EA, Pyke KA, Froelich JE, Stokes KD, Osteryoung KW** (2000) A homologue of the bacterial cell division site-determining factor MinD mediates placement of the chloroplast division apparatus. *Curr Biol* **10**: 507–516
- Couturier J, Wu H-C, Dhalleine T, Pégeot H, Sudre D, Gualberto JM, Jacquot J-P, Gaymard F, Vignols F, Rouhier N** (2014) Monothiol Glutaredoxin–*BolA* Interactions: Redox Control of *Arabidopsis thaliana* *BolA2* and *SufE1*. *Mol Plant* **7**: 187–205
- Ellis JR, Leech RM** (1985) Cell size and chloroplast size in relation to chloroplast replication in light-grown wheat leaves. *Planta* **165**: 120–125
- Emms DM, Kelly S** (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157

Gao H, Kadirjan-Kalbach D, Froehlich JE, Osteryoung KW (2003) ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc Natl Acad Sci* **100**: 4328–4333

Garcia M, Myouga F, Takechi K, Sato H, Nabeshima K, Nagata N, Takio S, Shinozaki K, Takano H (2008) An Arabidopsis homolog of the bacterial peptidoglycan synthesis enzyme MurE has an essential role in chloroplast development. *Plant J* **53**: 924–934

Glynn JM, Froehlich JE, Osteryoung KW (2008) Arabidopsis ARC6 coordinates the division machineries of the inner and outer chloroplast membranes through interaction with PDV2 in the intermembrane space. *Plant Cell* **20**: 2460–2470

Glynn JM, Yang Y, Vitha S, Schmitz AJ, Hemmes M, Miyagishima S, Osteryoung KW (2009) PARC6, a novel chloroplast division factor, influences FtsZ assembly and is required for recruitment of PDV1 during chloroplast division in Arabidopsis. *Plant J* **59**: 700–711

Itoh R, Fujiwara M, Nagata N, Yoshida S (2001) A Chloroplast Protein Homologous to the Eubacterial Topological Specificity Factor MinE Plays a Role in Chloroplast Division. *Plant Physiol* **127**: 1644–1655

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–1240

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**: 1160–1166

Klepikova AV, Kasianov AS, Gerasimov ES, Logacheva MD, Penin AA (2016) A High Resolution Map of the Arabidopsis Thaliana Developmental Transcriptome Based on RNA-seq Profiling. *The Plant Journal* **80**: 1058-1070

Li Y, Wang L, Wang G, Feng Y, Liu X (2017) AT2G21280 only has a minor role in chloroplast division. *Front Plant Sci* **8**: 2095

Liu J-W, Li S-F, Wu C-T, Valdespino IA, Ho J-F, Wu Y-H, Chang H-M, Guu T-Y, Kao M-F, Chesson C, et al (2020) Gigantic chloroplasts, including bizonoplasts, are common in shade-adapted species of the ancient vascular plant family Selaginellaceae. *Am J Bot* **107**: 562–576

Liu M-J, Sugimoto K, Uygun S, Panchy N, Campbell MS, Yandell M, Howe GA, Shiu S-H (2018) Regulatory Divergence in Wound-Responsive Gene Expression between Domesticated and Wild Tomato. *Plant Cell* **30**: 1445–1460

Maple J, Fujiwara MT, Kitahata N, Lawson T, Baker NR, Yoshida S, Møller SG (2004) GIANT CHLOROPLAST 1 is essential for correct plastid division in Arabidopsis. *Curr Biol* **14**: 776–781

- Marrison JL, Rutherford SM, Robertson EJ, Lister C, Leech RM** (1999) The distinctive roles of five different ARC genes in the chloroplast division process in Arabidopsis. *Plant J* **18**: 651–662
- Matasci N, Hung L-H, Yan Z, Carpenter EJ, Wickett NJ, Mirarab S, Nguyen N, Warnow T, Ayyampalayam S, Barker M, et al** (2014) Data access for the 1,000 Plants (1KP) project. *GigaScience* **3**: 17
- Miyagishima S, Froehlich JE, Osteryoung KW** (2006) PDV1 and PDV2 mediate recruitment of the dynamin-related protein ARC5 to the plastid division site. *Plant Cell* **18**: 2517–2530
- Nakanishi H, Suzuki K, Kabeya Y, Miyagishima S** (2009) Plant-Specific Protein MCD1 Determines the Site of Chloroplast Division in Concert with Bacteria-Derived MinD. *Curr Biol* **19**: 151–156
- Osteryoung KW, Stokes KD, Rutherford SM, Percival AL, Lee WY** (1998) Chloroplast division in higher plants requires members of two functionally divergent gene families with homology to bacterial ftsZ. *Plant Cell* **10**: 1991–2004
- Osteryoung KW, Vierling E** (1995) Conserved cell and organelle division. *Nature* **376**: 473–474
- Pyke KA, Rutherford SM, Robertson EJ, Leech RM** (1994) arc6, a fertile Arabidopsis mutant with only two mesophyll cell chloroplasts. *Plant Physiol* **106**: 1169–1177
- Qin L, Wang M, Zuo J, Feng X, Liang X, Wu Z, Ye H** (2015) Cytosolic BolA Plays a Repressive Role in the Tolerance against Excess Iron and MV-Induced Oxidative Stress in Plants. *PLOS ONE* **10**: e0124887
- Raynaud C, Cassier-Chauvat C, Perennes C, Bergounioux C** (2004) An Arabidopsis homolog of the bacterial cell division inhibitor SulA is involved in plastid division. *Plant Cell* **16**: 1801–1811
- Santos JM, Freire P, Vicente M, Arraiano CM** (1999) The stationary-phase morphogene bolA from Escherichia coli is induced by stress during early stages of growth. *Mol Microbiol* **32**: 789–798
- Sheue CR, Sarafis V, Kiew R, Liu HY, Salino A, Kuo-Huang LL, Yang YP, Tsai CC, Lin CH, Yong JWH, et al** (2007) Bizonoplast, a unique chloroplast in the epidermal cells of microphylls in the shade plant *Selaginella erythropus* (*Selaginellaceae*). *Am J Bot* **94**: 1922–1929
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM** (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212

Stata M, Sage TL, Rennie TD, Khoshravesh R, Sultmanis S, Khaikin Y, Ludwig M, Sage RF (2014) Mesophyll cells of C4 plants have fewer chloroplasts than those of closely related C3 plants. *Plant, Cell and Environment* **37**: 2587–2600

Stata M, Sage TL, Hoffmann N, Covshoff S, Ka-Shu Wong G, Sage RF (2016) Mesophyll Chloroplast Investment in C3, C4 and C2 Species of the Genus *Flaveria*. *Plant and Cell Physiology* **57**: 904–918

Vitha S, Froehlich JE, Koksharova O, Pyke KA, Erp H van, Osteryoung KW (2003) ARC6 is a J-domain plastid division protein and an evolutionary descendant of the cyanobacterial cell division protein Ftn2. *Plant Cell* **15**: 1918–1933

Wagner GP, Kin K, Lynch VJ (2012) Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* **131**: 281–285

Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, et al (2012) Evolution of GOLDEN2-LIKE gene function in C3 and C4 plants. *Planta* **237**: 481–495

Weise SE, Carr DJ, Bourke AM, Hanson DT, Swarthout D, Sharkey TD (2015) The arc mutants of *Arabidopsis* with fewer large chloroplasts have a lower mesophyll conductance. *Photosynth Res* **124**: 117–126

Yang Y, Sage TL, Liu Y, Ahmad TR, Marshall WF, Shiu S-H, Froehlich JE, Imre KM, Osteryoung KW (2011) CLUMPED CHLOROPLASTS 1 is required for plastid separation in *Arabidopsis*. *Proc Natl Acad Sci* **108**: 18530–18535

Chapter 4. Assembly and utilization of the first *Peperomia dahlstedtii* genome for differential expression analyses

Introduction

Although chloroplast population morphology is important for photosynthesis and likely plant fitness, we do not know how this trait is sensed and controlled by the cell (Chapter 1). In Chapter 2, I identified several species from the low-light tropical plant genus *Peperomia* with small and large chloroplast phenotypes, and demonstrated that a complete lack of chloroplast division is responsible for the palisade-specific large-chloroplast phenotype. Utilizing two representative species of *Peperomia*, I sequenced mRNA from young expanding leaves in order to create reference transcriptomes and identify candidate regulators of chloroplast population morphology (Chapter 3). However, considering the fragmentation observed in the *de novo* assemblies, it has become clear that higher-quality genomic resources would benefit this project immensely. Here, in the final data-focused chapter of my thesis, I describe the sequencing, assembly, and annotation of the *P. dahlstedtii* genome—the very first genome sequence for *Peperomia*—one of the largest and most diverse angiosperm genera (Frenzke et al., 2015; Frenzke et al., 2016; Simmonds et al., 2021). Further, alongside *Piper nigrum* (black pepper), the *P. dahlstedtii* genome described here one of only two available for the *Piperaceae*, the largest family in the early-diverging angiosperm magnoliid clade (Hu et al., 2019).

With a genome in hand, I can conduct better quality expression analyses. Thus, I have built upon the work described in Chapter 3 using both the RNAseq described there in addition to two more developmental stages for both species following the developmental series I established in Chapter 2. Further, I was able to assemble a much higher-quality transcriptome for *P. pellucida* by assembling it with the *P. dahlstedtii* genome as a guide. These analyses are therefore more thorough than those conducted using the *de novo* transcriptome assemblies described in Chapter 3. Below I describe likely biologically relevant trends in the expression of

some of the chloroplast division genes in both species. Further, I utilize orthogroup expression over leaf development between species to identify lists of candidate regulators of chloroplast population morphology.

Results

Genome size, complexity, and the selection of representative species in Peperomia

Use of a reference genome assembly is the best way to validate potentially missing or truncated genes of interest identified in Chapter 3. To date there is no genome sequence available for *Peperomia*. Therefore, I set out to generate the first genome assembly for the genus. To identify the best candidate reference species for sequencing, I investigated genome diversity in *Peperomia*.

Sequencing depth is an important factor to consider when generating a reference genome, as deeper coverage allows for a more complete genome assembly. Larger genomes require more sequencing in order to obtain sufficient coverage. Genome sizes were estimated from the young leaves of six *Peperomia* species using flow cytometry and ranged from 1.03 to 6.86 Gb (Table 4.1). Interestingly, the two large-chloroplast species observed, *P. metallica* and *P. pellucida*, had the largest genomes (~5.5 and 7 Gb, respectively; Table 4.1), making them less desirable targets for assembly as they would require more sequencing.

Ploidy is another important factor to consider for genome assembly, as polyploid genomes can be more challenging to assemble and work with in downstream analysis. Ploidy was estimated by counting chromosomes in young root tips taken from *P. dahlstedtii*, *P. fraseri*, *P. obtusifolia*, and *P. pellucida*. All except for *P. pellucida* were diploid with $2n=22$ chromosomes, while *P. pellucida* was tetraploid with $4n=44$ chromosomes (Figure 4.1).

Table 4.1. Somatic cell genome size estimates derived from young leaf tissue by flow cytometry.

Species	pg/2C	Gb
<i>P. argyreia</i>	1.05	1.03
<i>P. dahlstedtii</i>	2.59	2.53
<i>P. incana</i>	2.63	2.57
<i>P. metallica</i>	5.63	5.51
<i>P. metallica</i> (var. <i>Columbiana</i>)	5.26	5.14
<i>P. pellucida</i>	7.01	6.86
<i>P. serpens</i>	4.05	3.96

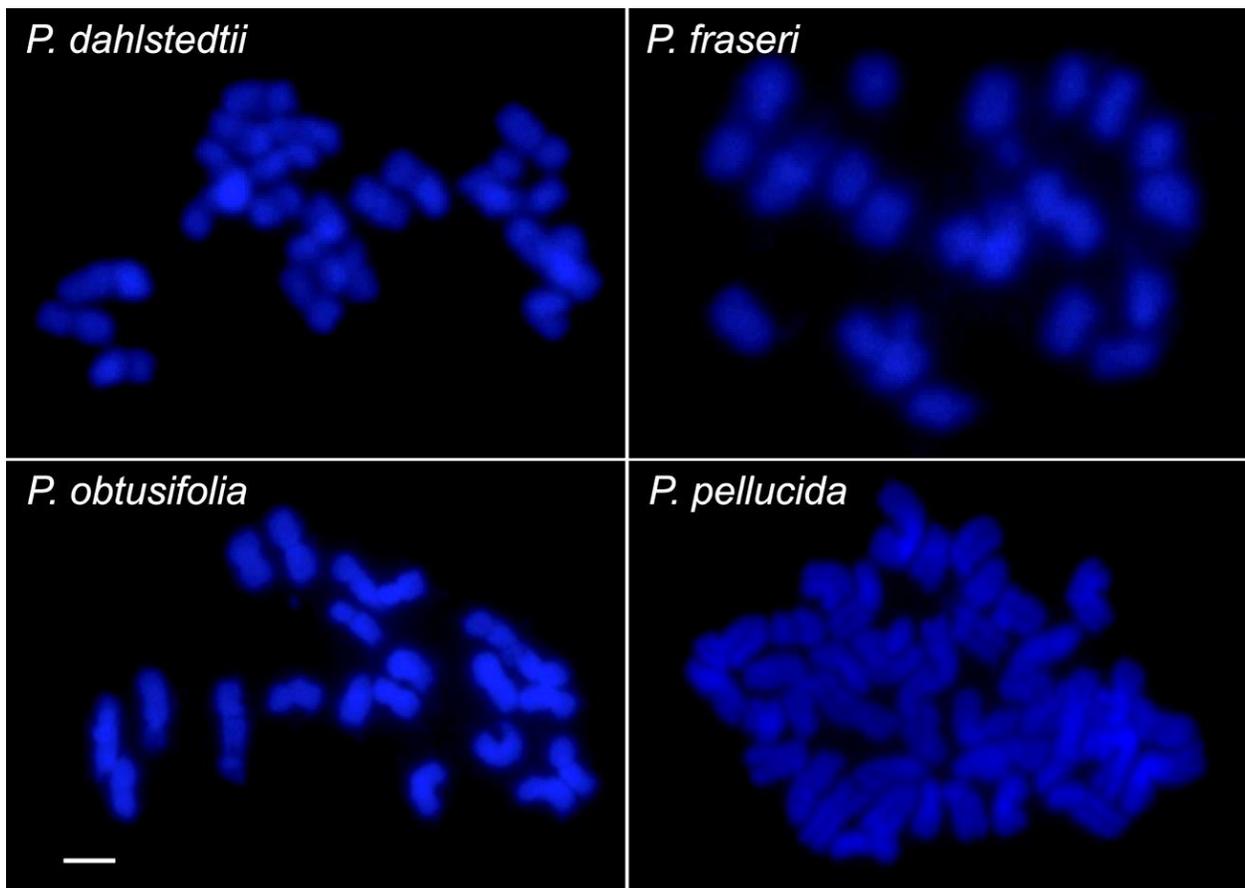


Figure 4.1. Fixed metaphase chromosomes from four *Peperomia* species.

Fixed DAPI (4',6-diamidino-2-phenylindole)-stained metaphase chromosomes from young root tip meristematic cells extracted from the indicated species. *P. dahlstedtii*, *P. fraseri*, and *P. obtusifolia* are each $2n=22$ chromosomes; *P. pellucida* is $4n=44$ chromosomes. The scale bar represents 5 μm .

Considering the smaller genome size and ploidy of *P. dahlstedtii* (2.53 Gb and $2n=22$; Table 4.1 & Figure 4.1) and the extensive work I did on this species in Chapters 2 and 3 on chloroplast morphology and expression in very young leaves, *P. dahlstedtii* was selected for genome sequencing.

Establishing the first genome assembly for the genus Peperomia

Prior to assembly, it is helpful to know how heterozygous the genome is, as this may affect assembly and annotation (Ranallo-Benavidez et al., 2020). High molecular weight genomic DNA (gDNA) was extracted from *P. dahlstedtii* leaves, and k-mer-based analysis of Illumina whole-genome sequencing (WGS) revealed that *P. dahlstedtii* is 3.4% heterozygous (Figure 4.2). For perspective, an *Arabidopsis thaliana* F1 plant is ~1.04% heterozygous (Vurture et al., 2017). The high heterozygosity of the *P. dahlstedtii* genome is not surprising, as *Peperomia* are almost exclusively clonally propagated, allowing for the accumulation of somatic mutations and repetitive elements (Ahmadabadi and Bock, 2010; Gutierrez et al., 2016; Ramu et al., 2017; Chen et al., 2019c).

Based on the high levels of heterozygosity (3.44%; Figure 4.2) and relatively large genome size (2.53 Gb; Table 4.1) of *P. dahlstedtii* compared to other sequenced plant genomes (e.g., 1.33% and ~0.76 Gb for black pepper; Hu et al., 2019) long-read sequencing was selected as the best strategy for assembly (Dumschott et al., 2020; Michael and VanBuren, 2020). To this end, high molecular weight genomic DNA (gDNA) was extracted from *P. dahlstedtii* leaves and sequenced using the Oxford Nanopore Technologies platform (ONT) to ~55x coverage of the diploid genome size (Table 4.1).

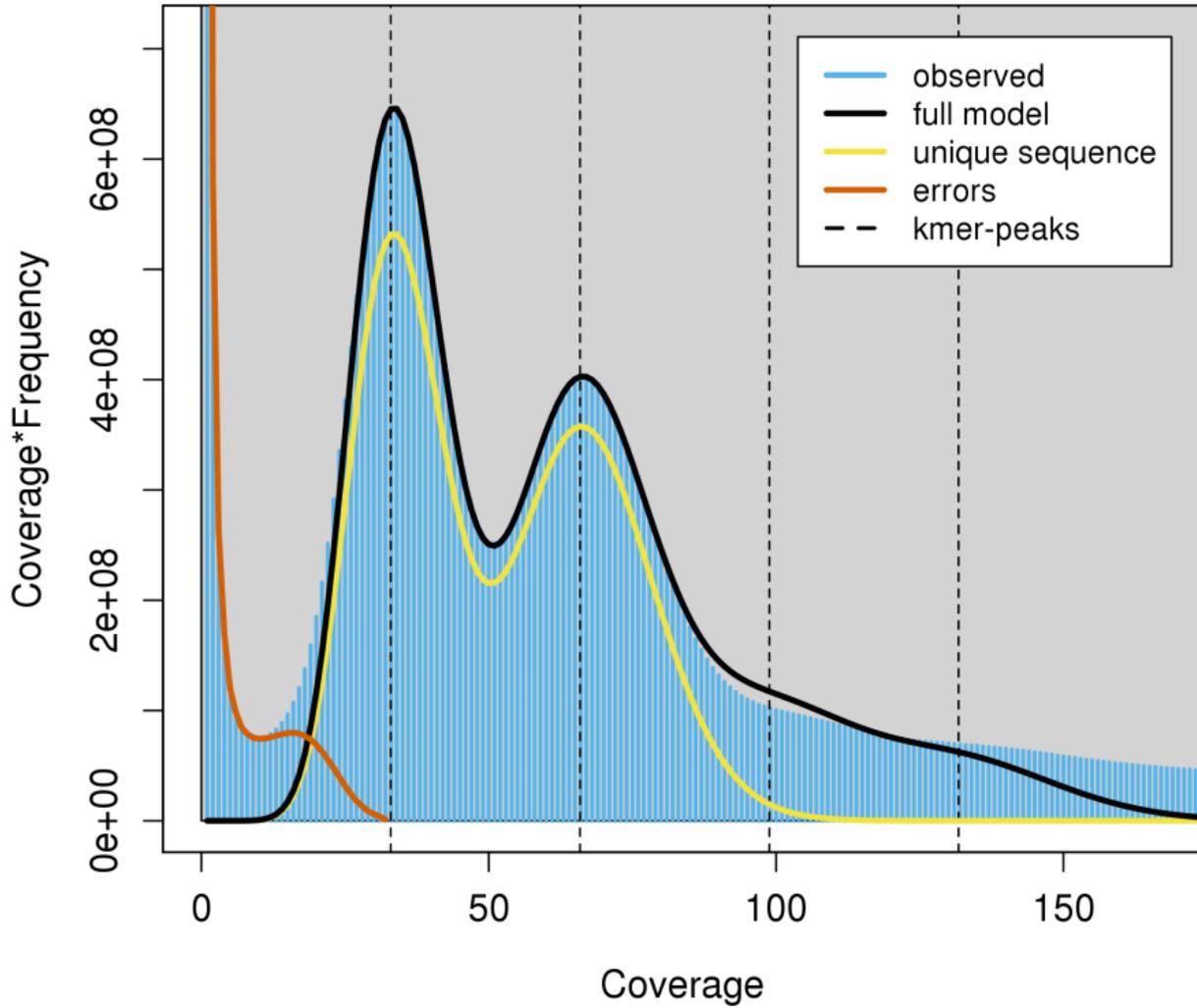


Figure 4.2. WGS k-mer-coverage-based profile of *Peperomia dahlstedtii*.

GenomeScope (Vurture et al., 2017) was used to generate a k-mer coverage profile from *P. dahlstedtii* Illumina PE150 WGS data. From this profile, heterozygosity was estimated to be 3.44%.

A draft genome assembly was generated using the *P. dahlstedtii* ONT gDNA reads with Canu (Table 4.3; Koren et al., 2017). As ONT reads are known to be noisy and error-prone, the resulting assembly was polished and corrected using the Illumina WGS reads generated for heterozygosity estimation (Table 4.2). Following this step, ~89.5% of the assembly was contained in contigs >50 kb and 92.4% in contigs >35 kb (Table 4.3). To reduce computation time moving forward, contigs <35 kb were purged from the assembly. The final assembly contained 21,477 contigs totaling 2.42 Gb (Table 4.3), approximately equal to the estimated diploid genome size of *P. dahlstedtii* (2.53 Gb; Table 4.1), suggesting that heterozygosity prevented the collapse of the two haplotypes. The largest contig assembled was 5.46 Mb (Table 4.3).

To assess assembly quality, all available gDNA- and mRNA-seq reads derived from *P. dahlstedtii* tissue were mapped to the genome. Overall, regardless of the sequencing technology used, gDNA- and mRNA-derived reads mapped to the *P. dahlstedtii* genome at >80%, indicating good agreement in content and orientation between the reads and assembled sequence.

The assembly was scanned for repetitive elements using RepeatModeler (Flynn et al., 2020) and masked with RepeatMasker (Smit et al., 2016). The genome of black pepper (*Piper nigrum*, 2n=52 chromosomes), the closest related available, contains ~55% repeat regions (Hu et al., 2019). The *P. dahlstedtii* genome has a much higher repeat content of ~79.6%, with the majority (>50%) being long terminal repeat retrotransposons (LTRs) (Table 4.4). This is not entirely surprising, as the genomes of asexually-propagated plants are known to have higher amounts of repetitive elements (Ramu et al., 2017; Chen et al., 2019c).

Table 4.2. Samples used for next-generation sequencing & read mapping statistics.

Strategy ¹	Species	Library	Described ²	Use	Stage ³	% reads mapped ⁴	% properly paired
Illumina WGS⁵	<i>P. dahlstedtii</i>	Stranded, PE150 ⁶	Ch. 4	Genome polishing	-	99.25	88.71
ONT⁷ WGS	<i>P. dahlstedtii</i>	-	Ch. 4	Genome assembly	-	100	-
ONT cDNA	<i>P. dahlstedtii</i>	-	Ch. 4	Gene annotation	C, E	99.06	-
Illumina mRNA	<i>P. dahlstedtii</i>	Stranded, PE150	Ch. 3	Expression & gene annotation	A	97.56	90
Illumina mRNA	<i>P. dahlstedtii</i>	SE50 ⁸	Ch. 4	Expression & gene annotation	C, E	82.22-89.14	-
Illumina mRNA	<i>P. pellucida</i>	Stranded, PE150	Ch. 3	Expression	A	67.09-69.51	100
Illumina mRNA	<i>P. pellucida</i>	SE50	Ch. 4	Expression	C, E	74.82-83.77	-

¹ For each sample listed here three biological replicates were prepared and sequenced as separate libraries

² See the indicated Chapter's methods for details on sequencing and tissue used

³ For details on the leaf developmental stages used see Chapter 2

⁴ For *P. dahlstedtii* mapping statistics are based on alignment to the genome. For *P. pellucida* statistics are based on alignment to the genome-guided transcriptome assembly

⁵ Whole genome sequencing

⁶ Paired end reads, 150 nt in length

⁷ Oxford Nanopore Technologies

⁸ Single end reads, 50 nt in length

Table 4.3. Genome assembly statistics for *Peperomia dahlstedtii*.

Statistic	Value
Pre-purging of small contigs	-
# contigs >= 35,000 bp	21,682
Total length >= 35,000 bp	2,422,927,263 bp
# contigs >= 50,000 bp	15,444
Total length >= 50,000 bp	2,163,057,567 bp
Post-purging of contigs <35 kb	-
# contigs	21,477
Largest contig	5,460,211 bp
Total length	2,415,797,839 bp
GC (%)	37.33
N50	150,373
N75	77,668
L50	4,131
L75	9,785
# N's per 100 kbp	0

Table 4.4. Repeat content in the *Peperomia dahlstedtii* genome.

Element type	Subtype	Number of elements	% of sequence
SINEs¹	-	5,328	0.02
	ALUs ²	0	0
	MIRs ³	0	0
LINEs⁴	-	48,784	2.4
	LINE1	34,950	2.09
	LINE2	0	0
	L3/CR1	377	0.01
LTR⁵	-	588,562	51.41
	ERVL	0	0
	ERVL-MaLRs	0	0
	ERV_classI	0	0
	ERV_classII	0	0
DNA	-	89,543	3.39
	hAT-Charlie	0	0
	TcMar-Trigger	0	0
Unclassified	-	844,666	21.51
Small RNA	-	6,916	0.24
Satellites	-	430	0
Simple Repeats	-	298,684	0.71
Low complexity	-	48,730	0.10
% total sequence masked	-	-	79.60

¹ Short interspersed nuclear elements

² Alu elements

³ Mammalian-wide interspersed repeats

⁴ Long interspersed nuclear elements

⁵ Long terminal repeats

To assess gene content and completeness, 1,614 benchmarked universal single-copy orthologs (BUSCOs) (Simão et al., 2015) were identified in *P. dahlstedtii*. Overall, 94.5% of *P. dahlstedtii* BUSCOs were complete, 2.4% fragmented, and 3.1% missing (Table 4.5). A large proportion of BUSCOs (58.2%) was duplicated, likely due to the uncollapsed haplotypes resulting from the high heterozygosity (Table 4.5). Satisfied with the completeness of the genome, I performed gene annotation next.

A total of 45,295 genes were annotated using MAKER (Bowman et al., 2017). Transposable element-related genes were identified and removed, reducing the gene set to 44,140. deFusion was used to identify potentially fused tandem-duplicate loci, yielding a final gene count of 44,156 (Table 4.6; Wang et al., 2021). On average, predicted transcripts were 2,588 nt and coding sequences 1,092 nt, with ~5 exons per gene (Table 4.6).

Exploring gene duplication in P. dahlstedtii

I expected predicted gene duplication rates to be high in *P. dahlstedtii* (58.2% of BUSCOs duplicated; Table 4.5) due to elevated genomic heterozygosity and repeat-content, as both conditions could lead to the inability for haplotypes to collapse during assembly (Figure 4.2; Table 4.4). However, I was initially surprised when attempts to artificially collapse haplotypes were unsuccessful—pseudohaploid (Chen et al., 2019c) only reduced BUSCO duplication rates by 0.8%. I decided to look at intragenomic collinearity to ensure that the high rates of duplication were due to uncollapsed haplotypes rather than something else, such as polyploidy. I used gene annotations and BLASTP output from the *P. dahlstedtii* and *P. nigrum* genomes to identify regions of collinearity and synteny, respectively, using MCSCanX (Wang et al., 2012). First, I looked at collinearity within the *P. dahlstedtii* genome, which should show regions of the genome that were not collapsed as separate contigs sharing gene content and order.

Table 4.5. Assessment of completeness in *Peperomia dahlstedtii* using BUSCOs¹.

% BUSCOs	<i>P. dahlstedtii</i>
Complete (C)	94.5
C single-copy	36.3
C duplicated	58.2
Fragmented	2.4
Missing	3.1

¹ Benchmarked universal single-copy orthologs (Simão et al., 2015)

Table 4.6. Transcript annotation statistics for *Peperomia dahlstedtii*.

Metric	Value
Count	44,156
Average Length	2,588 nt
Median Length	1,919 nt
Total Length	114,275,177 nt
Average Coding Length	1,092 nt
Median Coding Length	879 nt
Total Coding Length	48,198,317 nt
Ave Exons Per	5
Median Exons Per	4
Total Exons	234,046

Only 4.78% of all *P. dahlstedtii* genes were identified as collinear with another region in the *P. dahlstedtii* genome. I expected at least half of genes to be identified as collinear with another region in the genome, as duplication rates were just over 50% (Table 4.5). The low proportion of collinear genes identified in *P. dahlstedtii* (<5%) concerned me, and I elected to investigate this further.

Polyploidy could explain both high apparent gene duplication rates and the inability to resolve haplotypes in *P. dahlstedtii*. To determine whether *P. dahlstedtii* might be polyploid, orthologous predicted protein sequences—one per gene—were clustered between *P. dahlstedtii* and three model diploids with chromosome-scale genome assemblies—*P. nigrum*, *Amborella trichopoda*, and *A. thaliana*. Only orthogroups found in all four species were retained for analysis. *P. nigrum* was included because it shares some similarities in genome complexity with *P. dahlstedtii* (i.e., heterozygosity), but is known to be diploid (Hu et al., 2019). *A. thaliana* and *A. trichopoda* are both high-quality reference genomes, neither of which is known for high rates of heterozygosity (Platt et al., 2010; Poncet et al., 2012). For two diploid species, the number of genes in a given orthogroup is expected to be equal. For example, if *A. trichopoda* has five genes in an orthogroup, then *A. thaliana* would likely have five genes in that orthogroup as well—meaning the ratio of genes per orthogroup between two diploid species would be one-to-one for most orthogroups. I observed the ratio of genes per orthogroup in *A. trichopoda*, *P. dahlstedtii*, and *P. nigrum*, relative to *A. thaliana* (Figure 4.3A). As I predicted, the majority of orthogroups were one-to-one between *A. trichopoda* and *A. thaliana* (Figure 4.3A; pink). Interestingly, *P. nigrum*, though diploid, had twice the number of genes per orthogroup relative to *A. thaliana* in ~50% orthogroups (Figure 4.3A; blue).

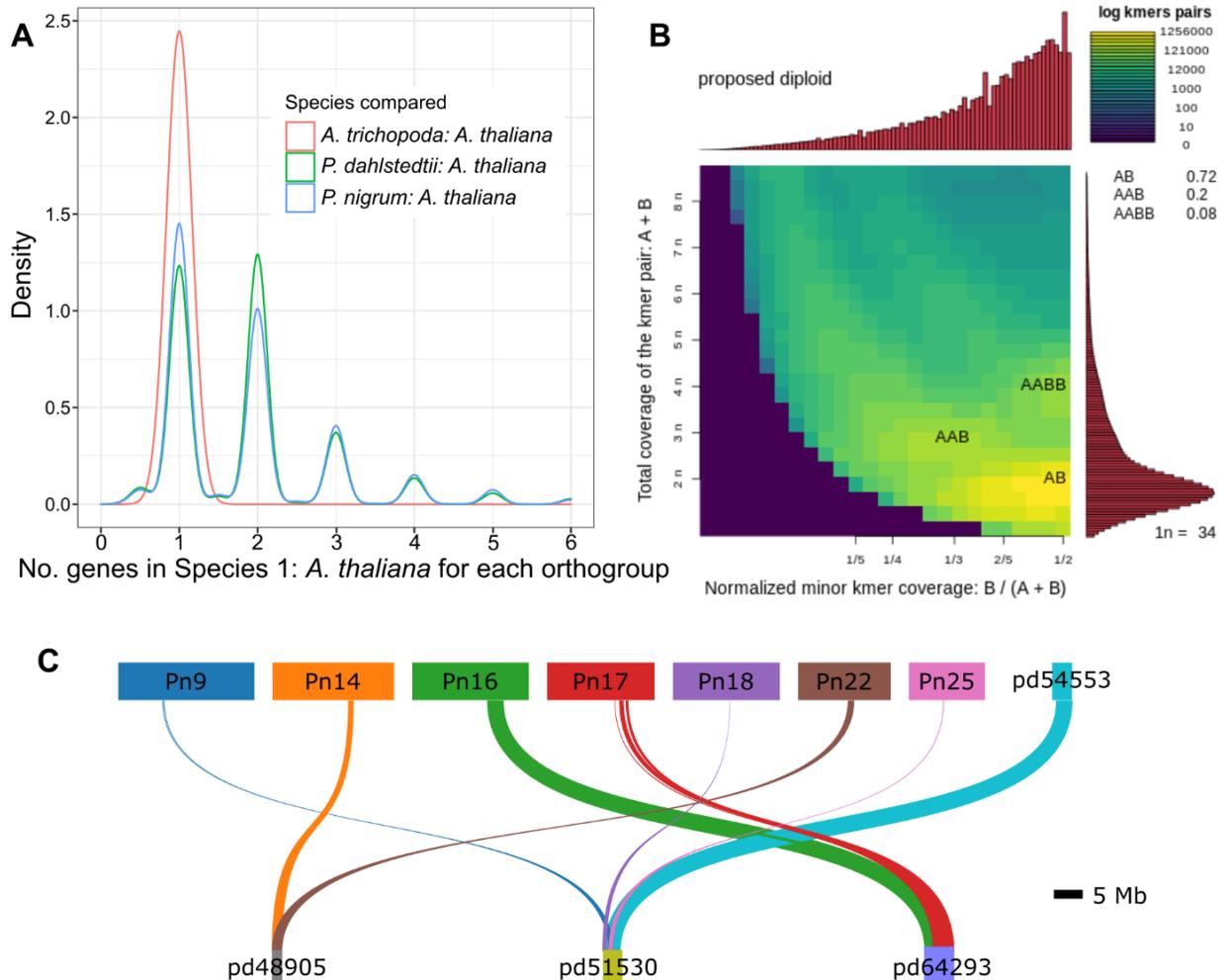


Figure 4.3. Gene duplication and heterozygosity in *P. dahlstedtii*.

(A) The density (i.e., frequency) of orthogroups with different ratios of genes identified in Species 1:Species 2 comparisons (see legend). (B) Log₁₀-scale smudgeplot (Ranallo-Benavidez et al., 2020) showing a high concentration of AB-identified k-mer pairs (yellow ‘smudge’ in the lower right corner), indicating that *P. dahlstedtii* is diploid. (C) For the three largest *P. dahlstedtii* contigs (‘pd,’ bottom; numbers indicate contig number), synteny with the chromosome-scale genome assembly for *Piper nigrum* (‘Pn,’ top left; numbers indicate whole chromosomes) and the whole *P. dahlstedtii* genome (‘pd’ top right) is shown. The scale bar represents 5 Mb.

P. dahlstedtii behaved similarly to *P. nigrum*, though with a slight majority of 2:1 orthogroups relative to *A. thaliana* (Figure 4.3A; green). The large proportion of 2:1 orthogroups in *P. nigrum* and *P. dahlstedtii* is likely reflective of the higher rates of heterozygosity in both genomes (1.33 and 3.44%, respectively). Taken together, these data support *P. dahlstedtii* being diploid with high apparent duplication rates due to heterozygous uncollapsed haplotypes. Further, reassessment of ploidy and heterozygosity using smudgeplot (Ranallo-Benavidez et al., 2020) further supported *P. dahlstedtii* being a heterozygous diploid (Figure 4.3B).

Lastly, I revisited the collinearity data described above from a different perspective. While only 4.78% of genes were identified as colinear, this is likely due to a combination of cutoff criteria set by MCSCanX (Wang et al., 2012) and fragmentation within the *P. dahlstedtii* assembly. MCSCanX requires a five-gene minimum when calling collinear regions (Wang et al., 2012). The *P. dahlstedtii* genome N50 was ~150 kb, which is relatively low—likely indicating the presence of many small contigs containing fewer genes (Table 4.3). Indeed, following the assembly and gene metrics, one gene is expected every 54.7 kb—or just under three genes every ~150 kb (2,415,797.8 kb genome size divided by 44,156 genes; Table 4.3), thus, leading to lower rates of collinearity than expected. To confirm this, I evaluated collinearity of the three largest *P. dahlstedtii* contigs against the whole genome (Figure 4.3C). Indeed, one of these three contigs was collinear with a different contig in the assembly, representing a concrete example where these alleles were not collapsed (Figure 4.3C; pd51530). I also briefly assessed synteny between these three largest *P. dahlstedtii* contigs and the *P. nigrum* chromosomes and found that each of the three *P. dahlstedtii* contigs shared syntenic regions with more than one *P. nigrum* chromosome (Figure 4.3C). This could be interpreted as evidence of fragmentation of the *P. dahlstedtii* assembly, which is not problematic but nonetheless good to be aware of. Similarly,

the presence of uncollapsed haplotypes in the *P. dahlstedtii* genome is not necessarily problematic, but should be kept in mind for downstream gene expression analyses. Therefore, having assessed the *P. dahlstedtii* genome and found it to be satisfactory in terms of quality (e.g., 94% complete), I was able to move forward towards preparing additional data for gene expression analyses.

Generation of novel transcriptomic data over leaf development in P. dahlstedtii and P. pellucida

In order to identify genes regulating chloroplast population morphology and division, I built upon the very young leaf expression data I generated in Chapter 3 by sequencing mature and middle-stage expanding leaves from *P. dahlstedtii* and *P. pellucida* (stages C and E, respectively; Table 4.2). These stages were chosen for their measurable differences in chloroplast division in *P. dahlstedtii* (Chapter 2)—as potential regulators of this condition are likely expressed in concert.

Prior to performing expression analysis, I decided to generate a genome-guided transcriptome for *P. pellucida* using the *P. dahlstedtii* genome and additional sequencing data described above, as genome-guided assembly should help to improve completeness and reduce fragmentation, both of which were issues with the *de novo* transcriptomes described in Chapter 3. *P. pellucida* mRNAseq reads were aligned to the *P. dahlstedtii* genome and the resulting alignment files were used for transcriptome assembly (Haas et al., 2013). A genome-guided transcriptome assembly was also made for *P. dahlstedtii*—both as a control for *P. pellucida* and for annotation of the genome. The primary assembly statistics for both transcriptomes were similar (Table 4.7). Both assemblies were assessed for completeness using BUSCO and were found to be very high quality, at >94% complete (Table 4.8).

As a final assessment of completeness and phylogenetic standing, orthologs were clustered between *P. dahlstedtii*, *P. pellucida*, and a number of other representative plant species (Figure 4.4). A phylogenetic tree derived from these data supports the placement of *Peperomia* and the magnoliid clade (including cinnamon, black pepper, avocado, and many others) as sister to the eudicots, having diverged after the monocots (Figure 4.4A). This is in line with some other recent studies on other magnoliid species (Soltis and Soltis, 2019). Orthogroup clustering revealed that ~3,500 orthogroups were shared between all species included (Figure 4.4B). With a high-quality reference genome for *P. dahlstedtii* and transcriptome for *P. pellucida* prepared, gene expression analysis was the next step.

Revisiting BolA1 and GC1 as candidate regulators of chloroplast population morphology

In Chapter 3, I concluded that loss of *BolA1* and or severe truncation of *GC1* might cause the large chloroplast phenotype observed in *P. pellucida*. I reinvestigated the presence, completeness, and expression of these genes using the new reference assemblies described above.

In Chapter 3, I found that *BolA1* appeared to be missing from three species with similar large chloroplast phenotypes—*Theobroma cacao*, *Selaginella moellendorffii*, and *P. pellucida*. I searched for *BolA1* in the *P. dahlstedtii* genome and *P. pellucida* genome-guided transcriptome assembly and was able to identify sequences that appear to be full length in both species (Figure 4.5). The only potential interesting *P. pellucida*-specific differences include a substitution to the *Peperomia*-specific insertion at V62G and another substitution at S112N (black arrows; Figure 4.5). The apparent presence of *PpBolA1* conflicts with my hypothesis that loss of this gene may contribute to the large-chloroplast phenotype, making *BolA1* a lower-priority candidate of interest.

Table 4.7. Genome-guided transcriptome assembly statistics.

Statistic	Version	<i>P. pellucida</i>	<i>P. dahlstedtii</i>
Total trinity genes		42,902	56,488
Total trinity transcripts		84,822	104,010
%gc¹		44.38	42.38
	Stats based on all transcript contigs	-	-
Contig n10		3,632	3,907
Contig n20		2,817	3,159
Contig n30		2,335	2,660
Contig n40		1,989	2,281
Contig n50		1,717	1,975
Median contig length (nt)		1,185	1,366
Average contig length (nt)		1,435.22	1,611.08
Total assembled bases		121,737,947	167,568,353
	Stats based on only the longest isoform per gene	-	-
Contig n10		3,500	3,752
Contig n20		2,720	2,973
Contig n30		2,276	2,477
Contig n40		1,937	2,110
Contig n50		1,674	1,808
Median contig length (nt)		1,108	1,149.5
Average contig length (nt)		1,369.62	1,441.25
Total assembled bases		58,759,500	81,413,523

¹ Percent guanine-cytosine content

Table 4.8. Assessment of completeness in *P. dahlstedtii* and *P. pellucida* genome-guided transcriptome assemblies using BUSCOs¹.

% BUSCOs	<i>P. dahlstedtii</i>	<i>P. pellucida</i>
Complete (C)	94.7	94.2
C single-copy	13.6	28.8
C duplicated	81.1	65.4
Fragmented	2.2	1.9
Missing	3.1	3.9

¹ Benchmarked universal single-copy orthologs (Simão et al., 2015)

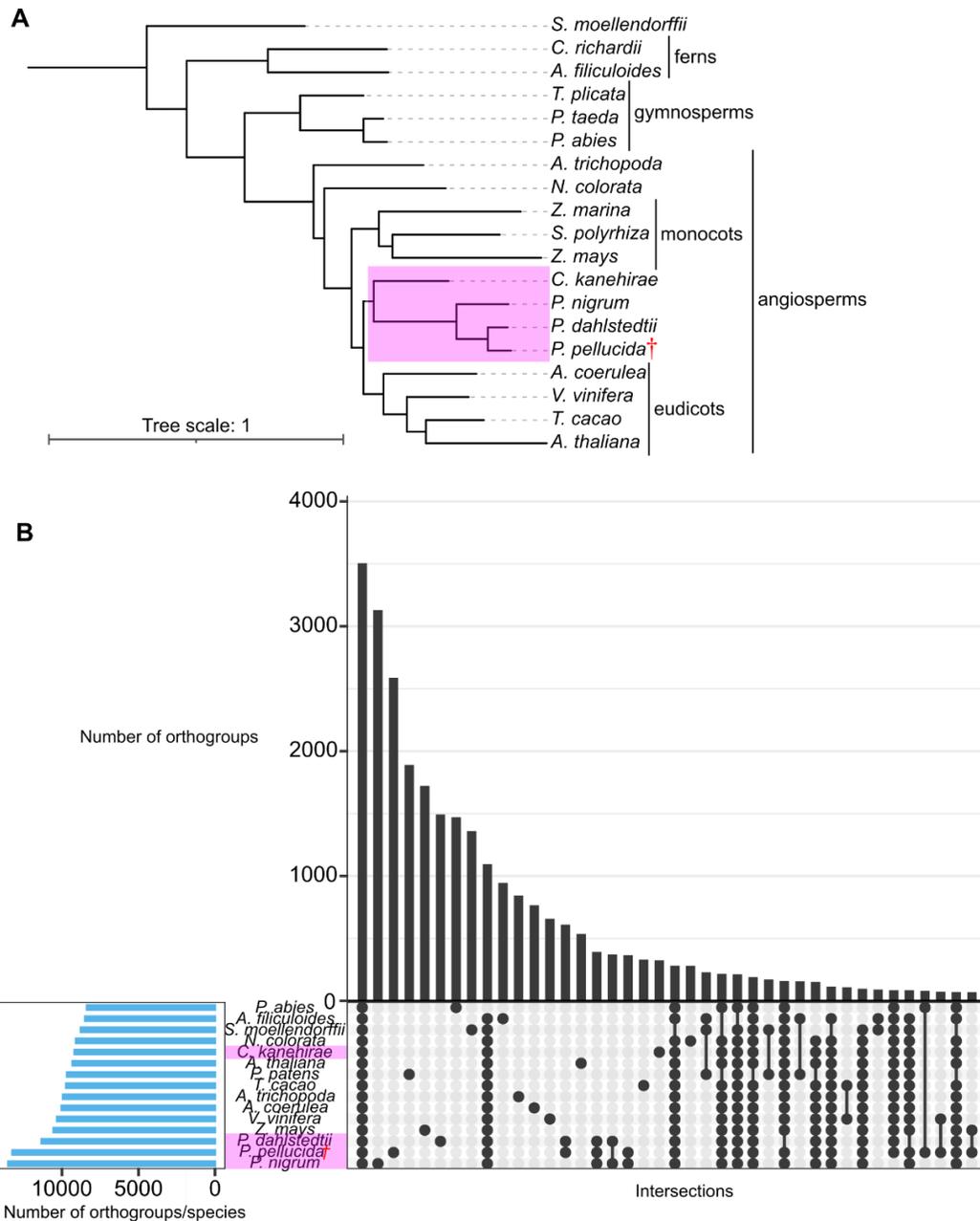


Figure 4.4. Orthogroup-based phylogeny and clustering of orthologous genes between *Peperomia dahlstedtii*, *Peperomia pellucida*, and representative vascular plants. (A) OrthoFinder-based species tree (using the orthogroups with genes present from all species, ~3,500) showing the placement of the magnoliids (*Peperomia*, *Piper*, and *Cinnamomum*) as sister to the eudicots. Relevant clades are labeled to the right. The scale bar represents a distance of 1. (B) UpSet (Conway et al., 2017) plot showing the largest-populated orthogroups between species. *Selaginella moellendorffii*, *Ceratopteris richardii*, *Azolla filiculoides*, *Tiquilia plicata*, *Pinus taeda*, *Picea abies*, *Amborella trichopoda*, *Nymphaea colorata*, *Zostera marina*, *Spirodela polyrhiza*, *Zea mays* (corn), *Cinnamomum kanehirae* (stout camphor tree), *Piper nigrum* (black pepper), *Peperomia dahlstedtii*, *Aquilegia coerulea* (columbine), *Vitis vinifera* (grape), *Theobroma cacao* (cocoa), and *Arabidopsis thaliana* sequences are all derived from genome assemblies. *Peperomia pellucida*[†] is the only species represented by a transcriptome assembly..

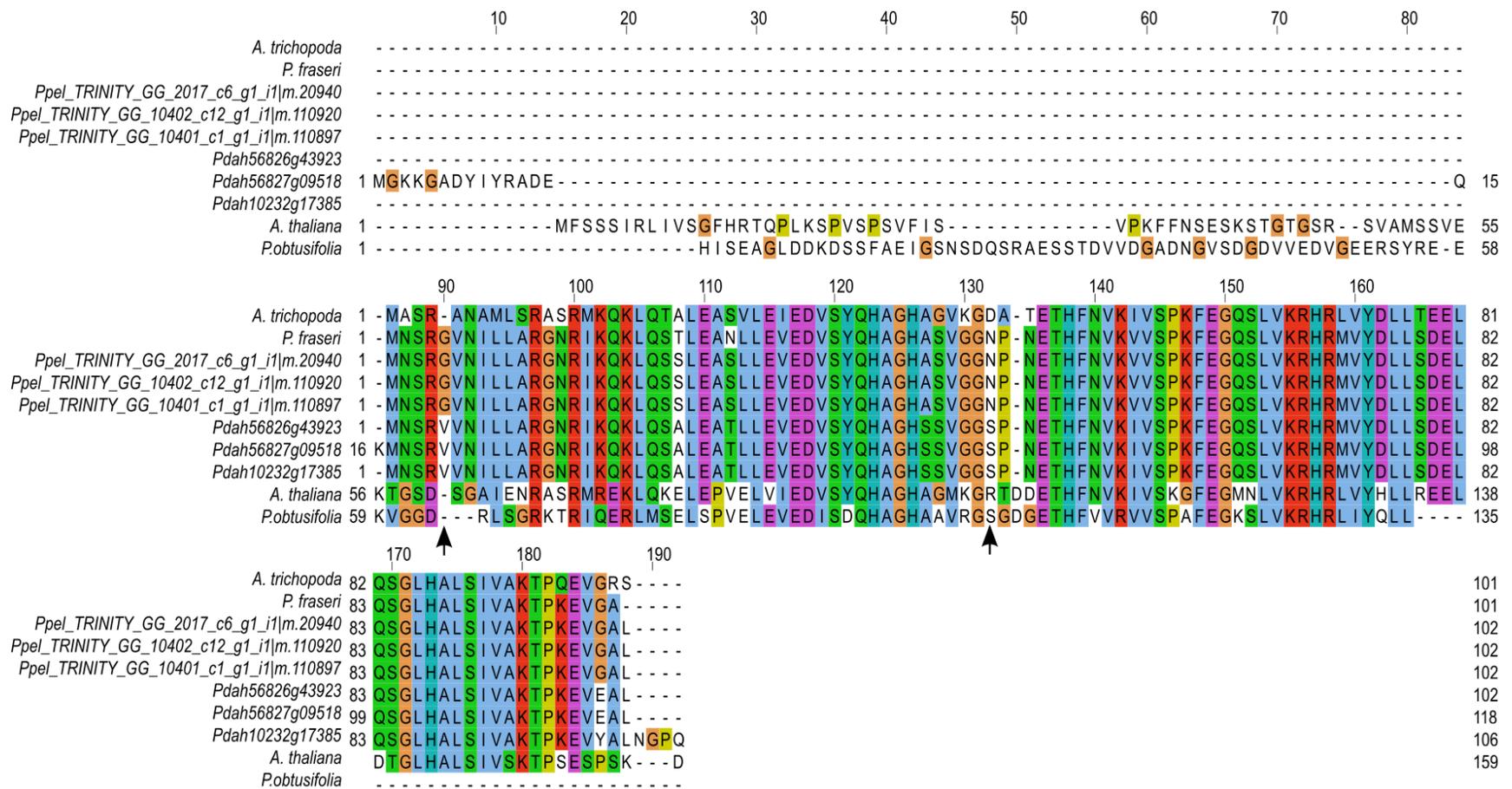


Figure 4.5. BolA1 is present in *P. pellucida*. Protein sequence alignment for BolA1 proteins identified in *P. dahlstedtii* (genome), *P. pellucida* (genome-guided transcriptome), *P. fraseri* (*de novo* transcriptome), and *P. obtusifolia* (*de novo* transcriptome), *Amborella trichopoda*, and *A. thaliana*. Black arrows (bottom) denote a possible insertion of interest at position 90 and a substitution at R132N (top scale).

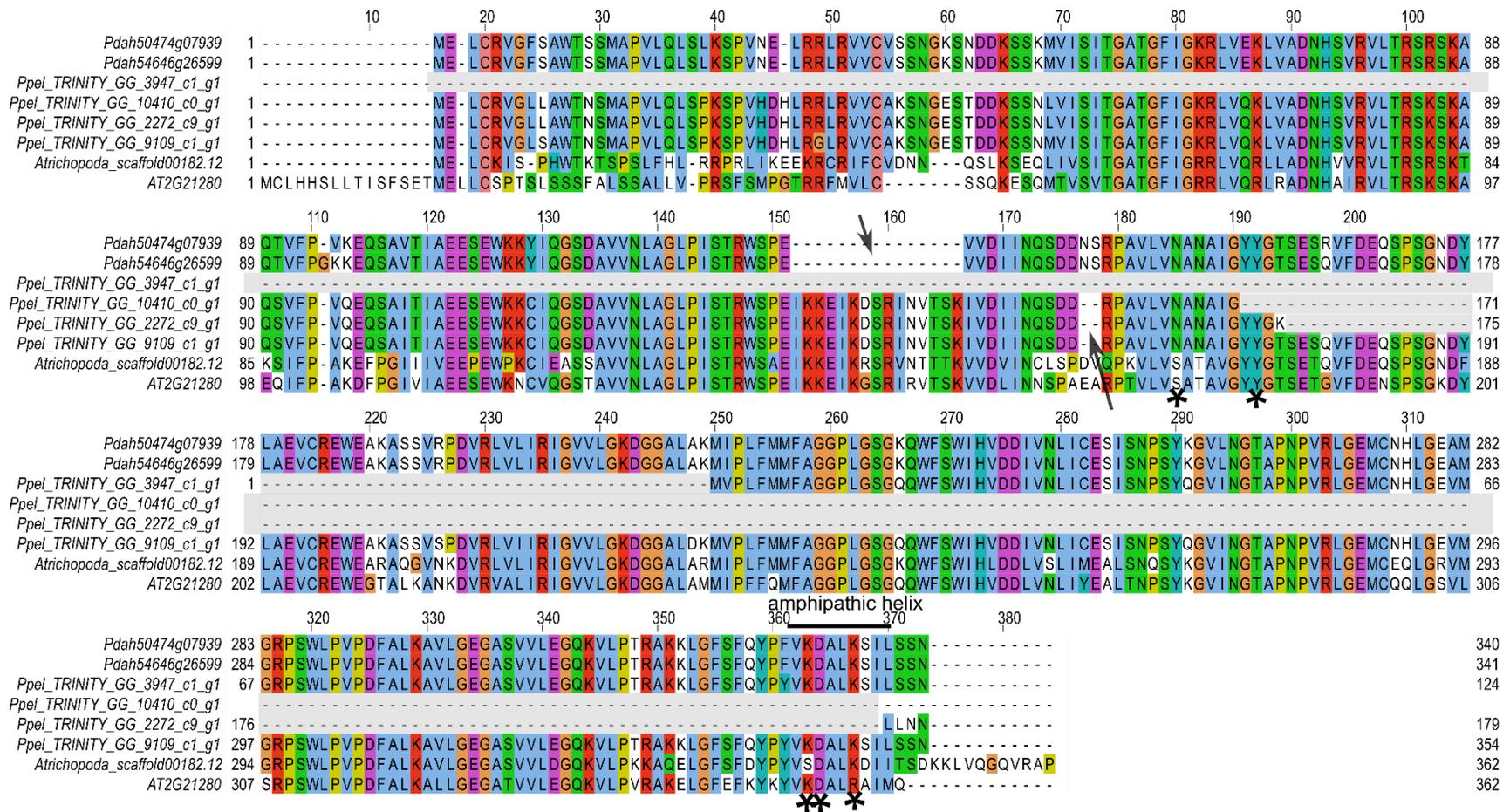
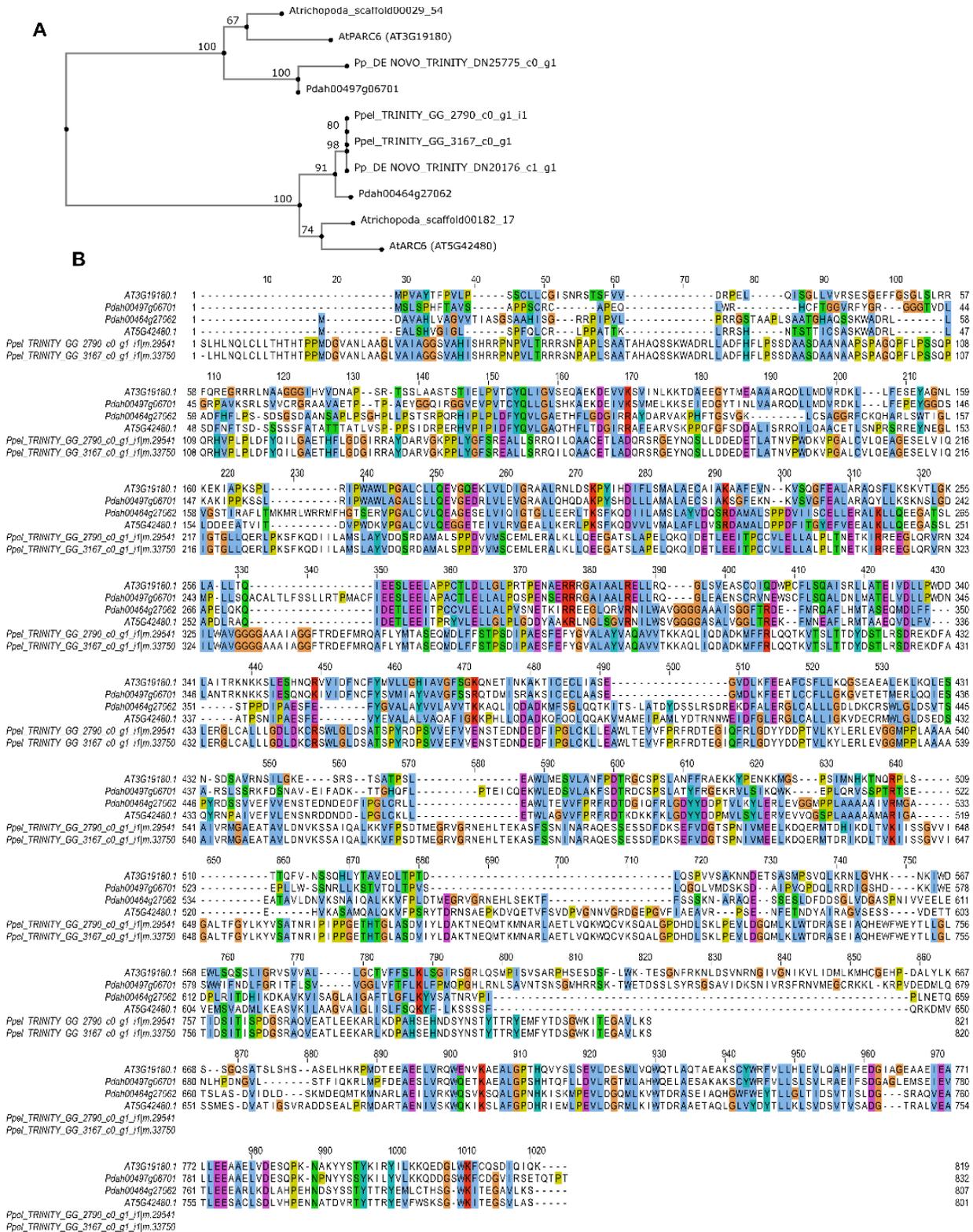


Figure 4.6. Multiple protein sequence alignment for GC1.

Protein sequence alignment of GC1s identified from *P. dahlstedtii* (Pdah; genome), *P. pellucida* (Ppel; genome-guided transcriptome), and the reference proteomes of *A. trichopoda* (Atrichopoda) and *A. thaliana* (AT). All references to positions refer to the scale at the top of the alignment. Grey highlights point out the successive series of severe truncations found in three of the four *P. pellucida* GC1s. Black arrows indicate *P. dahlstedtii*- (positions 152-166) and *P. pellucida*-specific (positions 177-178) deletions common amongst all their respective proteins. Black asterisks indicate amino acids of potential functional significance as identified in *A. thaliana*—S185 and Y192 are predicted epimerase active site residues; K363, D364, and R367 are part of the amphipathic helix (black bar, above) required for GC1 to tether itself to the inner envelope membrane (Maple et al., 2004).



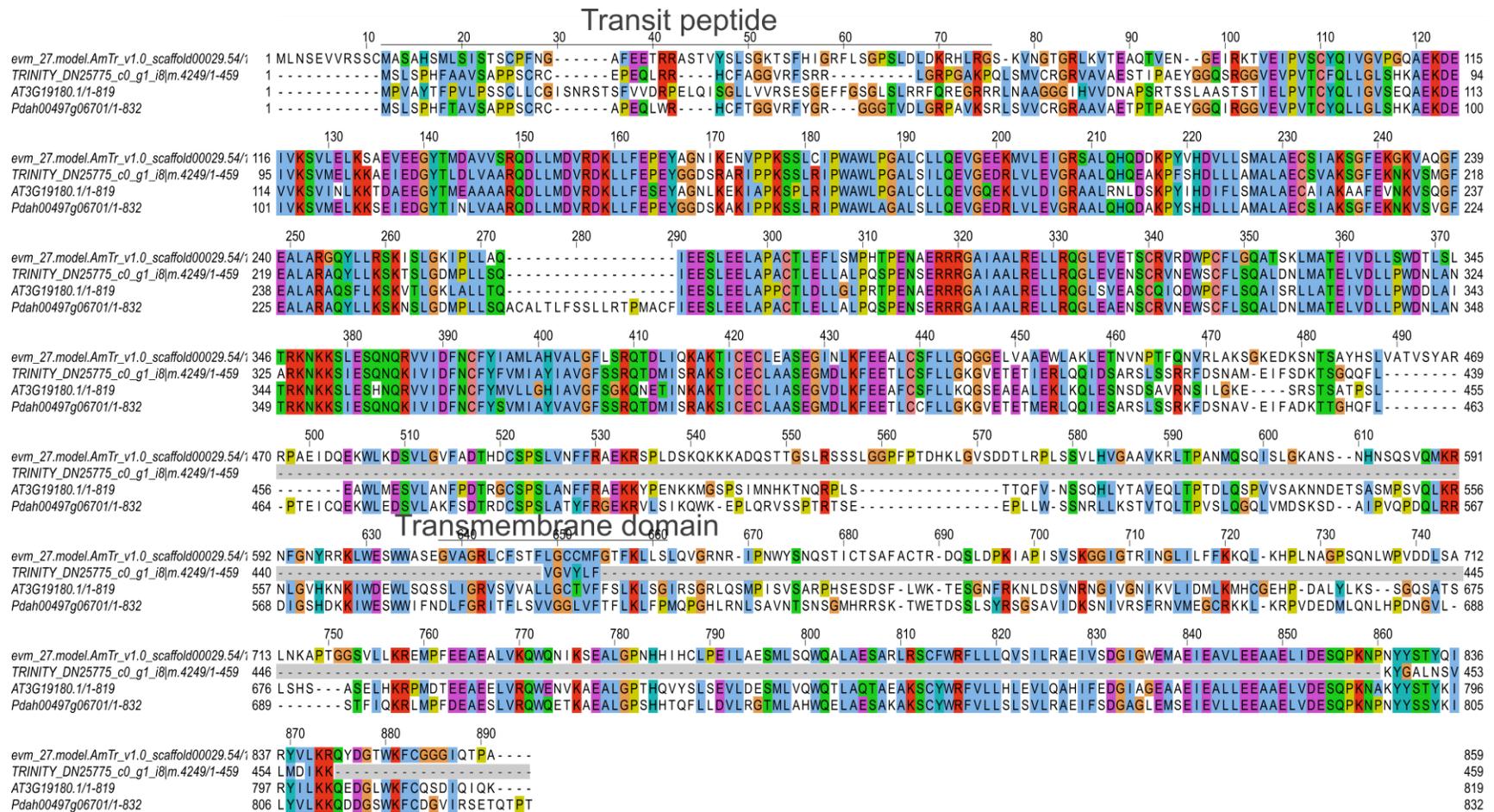


Figure 4.8. PARC6, if truly present in *Peperomia pellucida*, is severely truncated in critical domains.

Multiple sequence alignment for PARC6 identified in *Amborella* (evm_27_modelAmTr), the *de novo* transcriptome assembly of *P. pellucida* (“TRINITY...”; Chapter 3), *A. thaliana*, and *P. dahlstedtii*. The PARC6 transit peptide (TP) and transmembrane domain (TMD) regions are labeled above. The stroma region spans from the end of the transit peptide to the beginning of the TMD. The inner membrane space regions span from the end of the transmembrane domain to the end of the sequence. Large deletions present in *PpPARC6* are highlighted in gray.

GCI was initially identified as a candidate of interest behind the large chloroplast phenotype in *P. pellucida* due to severe truncation (Chapter 3). Two and four predicted *GCI* genes were identified in the reference assemblies established here for *P. dahlstedtii* and *P. pellucida*, respectively (Figure 4.6). Notably, all four *PpGC1*s exhibit successive large deletions or truncations (grey regions; Figure 4.6). Further, all four *PpGC1*s share a deletion at positions 177-178 (Figure 4.6). Unexpectedly, both *PdGC1*s contain a large deletion starting at position 152 (Figure 4.6). Overall, these data support my theory that truncation to *GCI* may contribute to the large chloroplast phenotype, though it is likely not alone in the regulation of this phenotype

Intra-species differential expression analyses: with a focus on the chloroplast division genes

P. pellucida palisade cell chloroplasts do not undergo division, while those in *P. dahlstedtii* do (Chapter 2). Therefore, I expected *P. dahlstedtii* chloroplast division gene expression to follow established protein-level patterns in *A. thaliana* during leaf expansion. Further, I anticipated that the expression of these genes in *P. pellucida* may not be upregulated in younger leaves compared to mature.

To compare gene expression within species between each of the three leaf developmental stages sequenced, mRNAseq reads were mapped to their respective species-specific transcripts and relative transcript abundance was quantified using Kallisto (Bray et al., 2016). To reduce variation between the sequencing libraries, the data were normalized so that the average relative log expression value, derived from all genes, was shrunk closer to zero for each library (Figure 4.9A-B & D-E) (Risso et al., 2014). Differential gene expression for all genes between each pairwise comparison of the three leaf developmental stages sampled (i.e., three different comparisons; see rows with corresponding leaf diagrams in Figure 4.10) using edgeR after normalization (Robinson et al., 2009). The numbers of significantly differentially expressed

genes for each comparison are shown in Figure 4.9C and F for *P. dahlstedtii* and *P. pellucida*, respectively. As one might expect, the largest proportion of differentially expressed genes was between the youngest expanding (stage A) and mature (stage E) leaf in both species (Figure 4.9). This is likely due to the ongoing leaf expansion and development in the youngest (stage A) and middle-expanding leaves (stage C). Next, I looked for significant developmental changes in expression of the chloroplast division genes for each species individually.

As demonstrated in Chapter 2, chloroplast division is active in *P. dahlstedtii* expanding leaves (stages A and C), where chloroplast number per cell increased with each successive stage (Figure 2.4). Therefore, I anticipated upregulation of some of the chloroplast division genes in the younger leaves sampled (stages A and C). In *P. dahlstedtii*, expression of two (of the five) *PdFtsZ1* genes identified was significantly upregulated in the youngest leaves compared to mature (Figure 4.10A; middle row), while *PdFtsZ2* levels were not significantly differentially regulated (Figure 4.10A). In *A. thaliana*, *AtFtsZ2* protein levels have been shown to decrease as the plant ages, accompanied by decreasing promoter activity (McAndrew et al., 2008; Schmitz et al., 2009). However, one study has shown that *AtFtsZ2* protein levels remain unchanged in expanding leaf tissue, which is consistent with my observations of *PdFtsZ2* expression (Okazaki et al., 2009; Figure 4.10A). One of the (three) *PdPDV2* genes was significantly upregulated in mid-stage expanding compared to mature leaves (Figure 4.10A; top row). In line with this finding, *AtPDV2* promoter activity is known to be elevated in very young expanding leaves, and protein levels decrease sharply as the leaf matures (Okazaki et al., 2009). Overall, reduced expression of the chloroplast division genes appears to be uncommon in developing *P. dahlstedtii* leaves, which is in line with the documented increase rates of chloroplast division in this species (Figure 2.4).

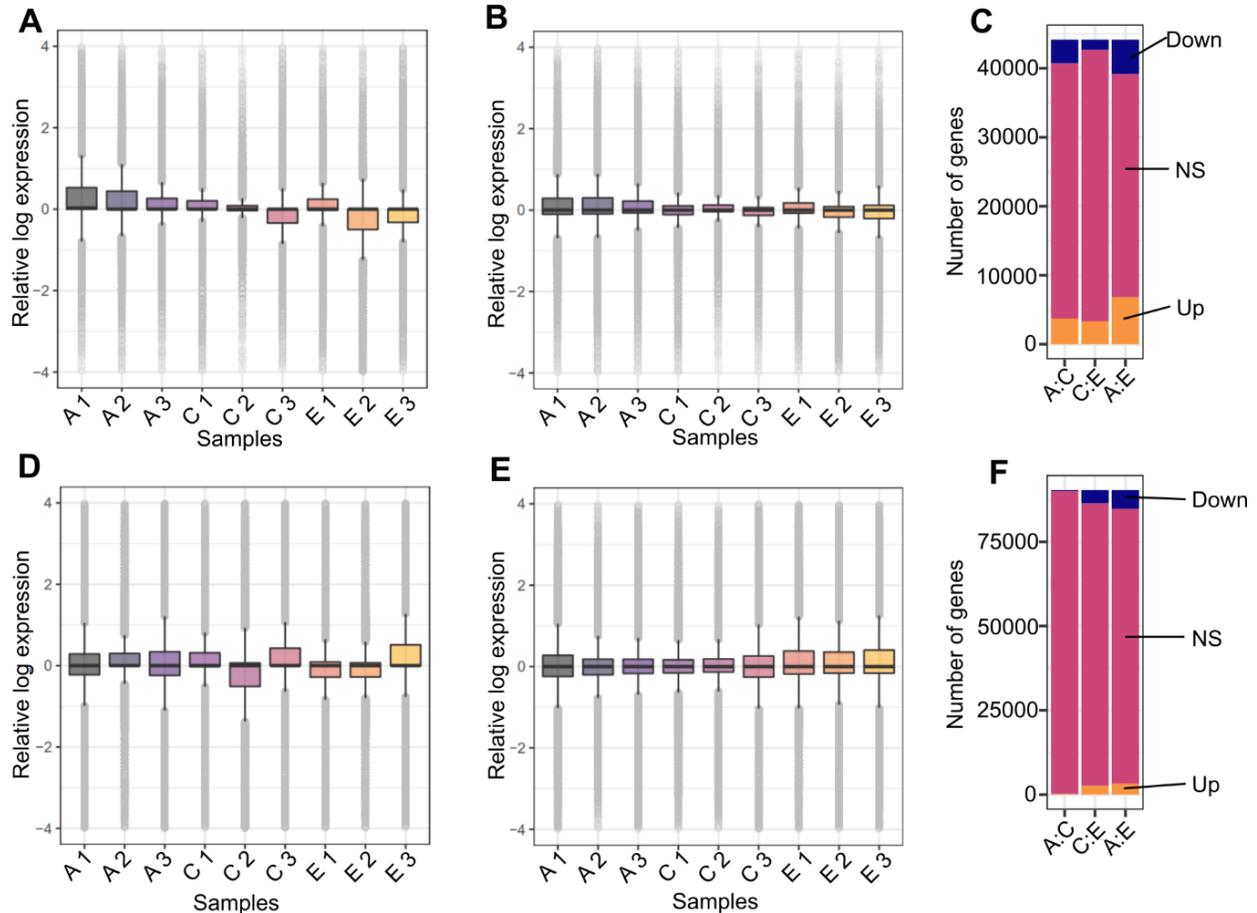


Figure 4.9. Normalization of expression across libraries and differential expression statistics.

(A) *P. dahlstedtii* pre-normalized and (B) post-normalized gene relative log expression for all leaf stage and biological replicate libraries. (C) The number of significantly up-regulated, down-regulated, and not-significant differentially regulated genes in *P. dahlstedtii*, derived from the normalized data in (C). (D) *P. pellucida* pre- and (E) post-normalized gene relative log expression for all leaf stage and biological replicate libraries. (F) The number of significantly up-, down-regulated, and not-significant differentially regulated genes in *P. pellucida* at each developmental comparison, derived from the normalized data in (E).

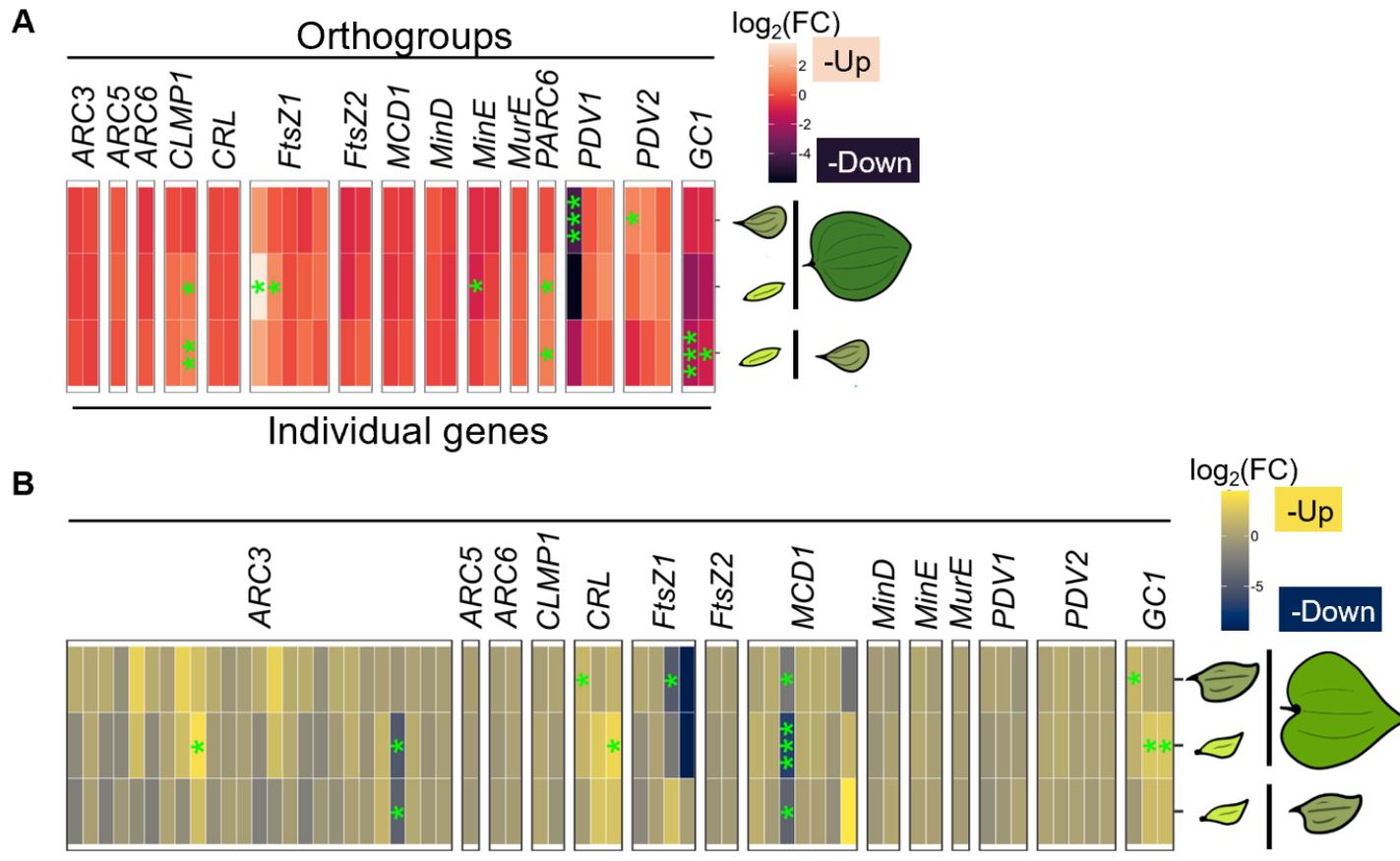


Figure 4.10. Differential expression of the chloroplast division genes in *P. dahlstedtii* and *P. pellucida* leaves over development. (A, B) Log₂FC for each of the chloroplast division genes (one genes per column of tiles, with orthogroups grouped within black outlined boxes) identified in (A) *P. dahlstedtii* and (B) *P. pellucida* for the leaf developmental-stage comparisons indicated to the right of each heatmap. Lighter-colored tiles (see heatmap legends) indicate upregulation in the younger leaf (left; see row-specific diagrams to the right of each map), and *vice versa*. Green asterisks denote significance (***, $p < 0.001$; **, $p < 0.01$, and *, $p < 0.05$) as determined using edgeR (Robinson et al., 2009).

In *P. pellucida*, I anticipated fewer significant changes in expression of the division genes, as division does not occur in the palisade cells of this species (Figure 2.4). This does appear to be the general case, as the majority of the heat map contains tiles darker in color, likely indicating unchanged expression of these genes in younger compared to mature leaf tissue—for example, see the *PpARC5*, *PpARC6*, *PpFtsZ2*, *PpMinD*, and *PpPDV1* groups (Figure 4.10B).

Further, several genes were significantly downregulated in young leaf tissue. One *PpFtsZ* was significantly downregulated in mid-expanding compared to mature leaves (Figure 4.10B; top row))—while *PdFtsZ1* expression was highest in young *P. dahlstedtii* leaf tissue (Figure 4.10A; middle row). Reduced expression of the *AtFtsZs* results in a reduced number of enlarged chloroplasts in the mature leaf (Osteryoung et al., 1998; Schmitz et al., 2009). *PpMCD1* was downregulated in all young versus mature leaf comparisons (Figure 4.10B; all rows). Loss of *AtMCD1* is known to produce a heterogeneous population of enlarged chloroplasts (Nakanishi et al., 2009). Overall, it appears that there is indeed reduced expression of some of the chloroplast division genes in *P. pellucida*, which is in line with the inhibition of chloroplast division in the palisade cells of this species (Figure 2.4).

Importantly, *PdPARC6* expression was consistently significantly upregulated in the youngest-expanding leaf (Figure 4.10A; bottom two rows), while *PARC6* was not identified in *P. pellucida* (Figure 4.8). A TBLASTN search (Camacho et al., 2009) with *AtPARC6* and *PdPARC6* against the *P. pellucida* genome-guided transcriptome identified five predicted transcripts, none of which clade with *PARC6*, indicating they are more like *ARC6* (Figure 4.7A). Further, both sequences clade exclusively with *ARC6* and share less sequence homology with *PARC6* (Figure 4.7). A *PARC6* sequence was identified in the *de novo* transcriptome of *P. pellucida* (Figure 4.7A), but closer inspection revealed numerous severe deletions (Figure 4.8).

Atparc6 mutants have enlarged chloroplasts (Glynn et al., 2009; Zhang et al., 2016; Itoh et al., 2018); thus, it is possible that disruption of PARC6 function via the observed deletions could contribute to the large chloroplast phenotype observed in *P. pellucida*.

Overall, two observations were perplexing to me. First, loss of *AtGCI* is thought to produce enlarged chloroplasts (Maple et al., 2007; Li et al., 2017), and potential loss or truncation of *GCI* in *P. pellucida* was one of my main findings in Chapter 3 (Figure 3.4). Conversely, I found that *PpGCI* was significantly upregulated in both younger leaf stages compared to mature (Figure 4.10B; top two rows). Further, *PdGCI* was significantly downregulated in the youngest compared to mid-stage leaves (Figure 4.10A; bottom row). While it would make sense for *GCI* to be regulated differently between these two species, as one undergoes chloroplast division in the palisade cells and the other does not, it remains unclear to me why this gene would be upregulated in *P. pellucida*.

The second confounding observation was primarily due to the expansion in the number of *ARC3* orthologs identified in *P. pellucida* (25 columns within the *ARC3* orthogroup; Figure 4.10B) in addition to the opposite direction in which those that were significantly different were regulated (Figure 4.10B; middle row). Differences in gene copy number between species was anticipated, due to the heterozygosity of *P. dahlstedtii* and tetraploid condition of *P. pellucida* (Figure 4.1; Figure 4.3). However, I anticipated that gene copy number would not differ as much as it did for *ARC3*, where *P. dahlstedtii* has 2 and *P. pellucida* 25 (Figure 4.10). Overall, I felt that the visualization and interpretation of these data would be improved using orthogroup rather than gene-specific expression (as described in Chapter 3), as this would effectively remove gene copy number differences and allow for comparison between species. In the next section I

describe the two methods taken to compare expression differences and correlation over development for all orthogroups common between *P. dahlstedtii* and *P. pellucida*.

Inter-species differential orthogroup expression

P. pellucida palisade cell chloroplasts do not undergo division, while those in *P. dahlstedtii* do (Chapter 2). Identifying distinct patterns in the correlation or differences in expression of the chloroplast division orthogroups between species should provide a set of parameters to be used to identify novel regulators of chloroplast population morphology. To compare expression between species, orthogroup expression—the collective expression of all individual genes or isoforms belonging to a given orthogroup (see Chapter 3)—was calculated within each species for all orthogroups shared between *P. dahlstedtii* and *P. pellucida* following

Equation 4.1.

Based on the chloroplast morphology of *P. dahlstedtii* and *P. pellucida*, in addition to the intraspecies differential expression analyses described above, I anticipated chloroplast division orthogroup expression would correlate poorly if not strongly negatively, meaning expression patterns are very different between species or follow opposite developmental trends. To test this, rank-based correlation coefficients (Kendall's Tau) were calculated for all orthogroups shared between *P. dahlstedtii* and *P. pellucida* using orthogroup expression at each developmental stage, and their distribution was observed (Figure 4.11A; green). As equivalent developmental stages were included for both species, I expected the majority of orthogroups to have a positive correlation, meaning they follow similar developmental patterns in expression between species. However, the distribution of correlation coefficients was negatively skewed (Figure 4.11A; green). Interestingly, expression patterns of a subset of the chloroplast division orthogroups was significantly negatively correlated—*PDVI*, *MurE*, *MCDI*, *FtsZ1*, and *GCI* (Figure 4.11A; orange dots). In an attempt to isolate a 'control' set of orthogroups for comparison to the chloroplast division orthogroups, distributions were also plotted for a subset of orthogroups from specific GO-term categories, the chloroplast division orthogroups, and single-copy orthologs for statistical comparison (Figure 4.11A). Kolmogorov-Smirnov tests revealed that the chloroplast division orthogroup subset did not differ significantly from any other (Figure 4.11A). Overall, the correlation of orthogroup expression between species did not seem amenable to isolating a subset of orthogroups potentially involved in chloroplast population morphology; therefore, another metric was considered.

I expected that the difference in expression between some of the chloroplast division orthogroups and potentially novel orthogroups regulating chloroplast population morphology

might be significant between species. To assess this, I calculated the developmental-stage-specific difference in expression between species for all orthogroups following Equation 4.2. Further, the cumulative difference in orthogroup expression between species over development was calculated using Equation 4.3. Distributions for these data were plotted, and the chloroplast division orthogroups identified (Figure 4.11B-E; orange points). Overall, *ARC3* expression was consistently higher in *P. pellucida* (Figure 4.11B-E), in line with what was observed for some of the individual *PpARC3* genes (Figure 4.10B). In the youngest leaf, *FtsZ1* expression was higher in *P. dahlstedtii* (Figure 4.11B), also consistent with the earlier differential expression observations (Figure 4.10A). Lastly, in mature leaves, *GCI* expression was higher in *P. dahlstedtii* (Figure 4.11D). Taken together, these results do indeed show that there are some significant differences in the expression of some of the chloroplast division orthogroups between these two morphologically distinct species, supporting the use of these distributions in identifying novel regulators in chloroplast population morphology.

Equation 4.1. Orthogroup expression (OG_{exp}) within-species.

$$OG_{\text{exp}} = \sum \text{orthogroup } \log_2(\text{CPM})$$

Equation 4.2. The difference in orthogroup expression between species (ΔOG_{exp}).

$$\Delta OG_{\text{exp}} = P. \text{dahlstedtii } OG_{\text{exp}} - P. \text{pellucida } OG_{\text{exp}}$$

Equation 4.3. Cumulative developmental difference in orthogroup expression between species ($\Sigma \Delta OG_{\text{exp}}$).

$$\sum \Delta OG_{\text{exp}} = \sum |\Delta OG_{\text{exp}}|$$

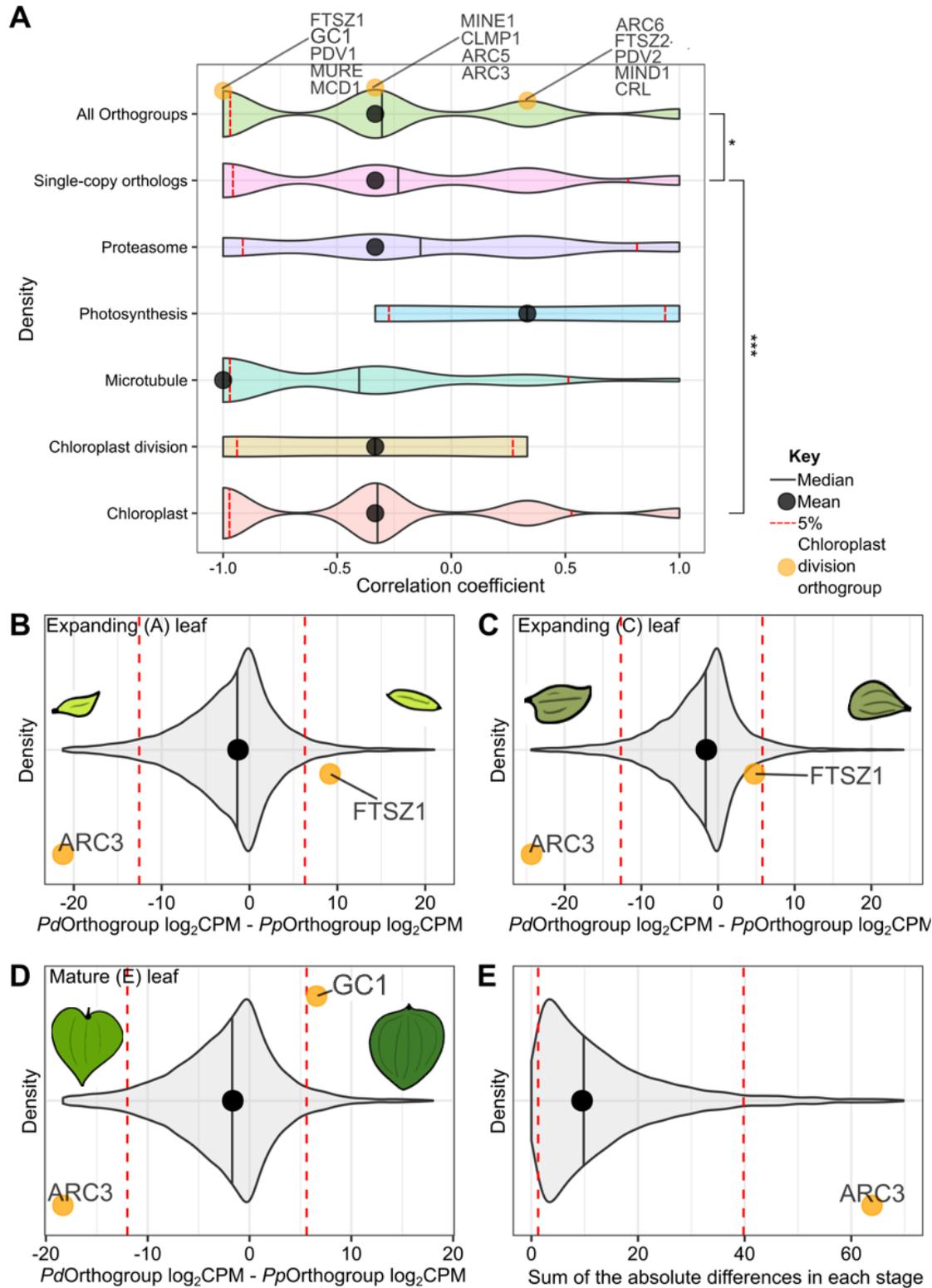


Figure 4.11. Distributions in orthogroup expression between species.

(A) Correlation coefficients (Kendall's Tau) were calculated for each orthogroup using the expression data from both species across all three developmental stages.

Figure 4.11 (cont'd)

All distributions were compared to one another (Kolmogorov-Smirnov test with Benjamini-Hochberg correction; *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$; non-significant differences are not shown). **(B-D)** The difference in orthogroup expression between species in **(B)** expanding stage A, **(C)** expanding stage C, and **(D)** mature stage E leaves (Chapter 2). Orthogroups positioned further left/negative had higher expression in *P. pellucida* than *P. dahlstedtii* and *vice versa*. **(E)** The absolute value of the difference in orthogroup expression from each leaf stage is summed. Only significant orthogroups are indicated in B-E. For all plots, medians are denoted by black and the chloroplast division orthogroups by orange dots, respectively. Means are denoted by a black line inside each violin plot. The 5 and 95% quantiles are indicated by dashed red lines for each distribution.

Table 4.9. Number of orthogroups significantly different in expression between species.

Leaf developmental stage examined ¹	<0.05 percentile	<Sig. chloroplast division orthogroup ²	>0.95 percentile	>Sig. chloroplast division orthogroup ³
A	519	217	535	342
C	676	149	457	-
E	926	374	572	441

¹ See Chapter 2 details on leaf stages

² Cutoff for significance made more stringent following the chloroplast division orthogroup identified in the tail, if any. For example, *ARC3* was used as the cutoff in all three stages.

³ Cutoff for significance made more stringent following the chloroplast division orthogroup identified in the tail, if any. For example, *GCI* was used as the cutoff in mature leaves.

In order to identify potentially novel regulators of chloroplast population morphology I utilized the distributions shown in Figure 4.11B-E to isolate orthogroups expressed significantly higher in *P. pellucida* or *P. dahlstedtii* at each developmental stage. Overall, 3.8-7.8% of all orthogroups were found in upper and lower 5th percentiles (Table 4.9). These subsets were further condensed to 3.1-1.3% of all orthogroups by using the significant chloroplast division orthogroup (where present) as more stringent cut-offs rather than the 5th percentile (Table 4.9). A brief look at these lists revealed that *GROWTH REGULATING FACTOR 5 (GRF5)*, which is thought to be involved in regulating chloroplast development and possibly division (Vercruyssen et al., 2015), was more highly expressed in *P. dahlstedtii* young leaves (stage A) compared to *P. pellucida*. Overexpression of *GRF5* in *A. thaliana* increases the number of chloroplasts per cell area (Vercruyssen et al., 2015)—in line with higher *GRF5* expression and chloroplast number in *P. dahlstedtii*. This is a positive indicator that these lists contain relevant potential regulators of chloroplast population morphology.

Discussion

The goal of this work was to characterize expression of the chloroplast division genes over leaf development and identify novel potential regulators of chloroplast population morphology. To this end, I established the first genome assembly for *Peperomia* and novel expression datasets for three different leaf developmental stages for two species of *Peperomia* that differ in rates of chloroplast division and number in their palisade mesophyll cells.

Peperomia is a member of the magnoliid clade of early-diverging angiosperms, which also houses black pepper, *Cinnamomum* spp. (cinnamon), *Persea americana* (avocado) and many other agriculturally and biochemically interesting species (Wanke et al., 2006; Frenzke et al., 2015; Soltis et al., 2018; Simmonds et al., 2021). Though the magnoliid clade is not well

represented with genomic or transcriptomic data, a handful of genome assemblies have been published over the last three years, including that of black pepper (Hu et al., 2019), *C. kanehirae* (stout camphor tree) (Chaw et al., 2019) and *Liriodendron tulipifera* (tulip tree) (Chen et al., 2019b). However, the exact placement of the magnoliids relative to the monocots and eudicots has remained somewhat unclear, as recent data are conflicting (Soltis and Soltis, 2019).

The inclusion of the *P. dahlstedtii* genome in phylogenetic analysis is in agreement with several studies in that the magnoliids are sister to the eudicots (Figure 4.4).

Peperomia is the most biogeographically diverse genus in the *Piperaceae* family, and these species can be found in the Neotropics, South Pacific, Asian tropics, and Africa (Frenzke et al., 2015; Frenzke et al., 2016; Simmonds et al., 2021). A number of *Peperomia* spp. reside in the shaded understory of tropical forests (Fosberg and Sachet, 1975; Kubitzki et al., 1993; Rasingam and Parthasarathy, 2009; Ashton-Butt et al., 2018) and exhibit leaf-morphological features thought to be beneficial to shade or low-light environments, including the large chloroplasts observed in three species, which I have demonstrated is due to a lack of chloroplast division (Chapter 2). Differences in the developmental expression patterns of some of the chloroplast division genes nicely complemented my findings in Chapter 2 (Figure 4.10 & Figure 4.11). Further, there is actually little published data on the expression of the division genes. Rather, the majority of work has reported on protein levels—meaning that my work is one of the first comprehensive studies on the expression of the division genes and the only one to my knowledge describing their differential expression.

Though the literature remains unclear on whether *GCI* is involved in chloroplast division, it does influence chloroplast population morphology on some level. I found a consistent reduction in chloroplast number relative to cell size in two different *Atgc1* lines compared to

wild type (Chapter 3), supporting a possible role for GC1 in chloroplast division. *AtGC1* was originally identified as a possible homolog of *Escherichia coli* SulA, which represses bacterial cell division by inhibiting the polymerization of FtsZ (Raynaud et al., 2004). In this same study, overexpression of *AtGC1* alone produced enlarged chloroplasts but was able to rescue the large-chloroplast phenotype caused by FtsZ overexpression, indicating that these proteins operate in the same pathway. However, they reported that reduced *AtGC1* expression causes the large chloroplast phenotype and that this reduction is dose-dependent (Maple et al., 2004). Overall, GC1 remains a good candidate for contribution to the large-chloroplast phenotype observed in *P. pellucida*, though the mechanism requires some unraveling. It is possible that fine-tuned changes to GC1 expression result in the most impactful alterations to chloroplast morphology. *GCI* orthogroup expression was higher in *P. dahlstedtii* compared to *P. pellucida* in mature leaves (Figure 4.11D) Overall, the consensus from my data is that *GCI* developmental expression patterns are very different between *P. dahlstedtii* and *P. pellucida*. For example, transgenic expression of the *P. dahlstedtii* and *P. pellucida* *GCI*s in *A. thaliana* might provide insight to how the mutations or truncations observed in these species might affect chloroplast morphology. Separately, reevaluation of *AtGC1* knock-down and overexpression lines would be helpful to determine whether GC1 under or overexpression (or both) inhibits chloroplast division.

Overall, *ARC3* orthogroup expression was consistently higher in *P. pellucida* than in *P. dahlstedtii* (Figure 4.11B-E). *ARC3* is a negative regulator of chloroplast division, as it acts as part of the Min system in sequestering FtsZ assembly to the mid-plastid (Maple et al., 2007). Further, while *Atarc3* plants have heterogeneously enlarged chloroplasts, overexpression lines have a smaller population of larger chloroplasts (Zhang et al., 2013). Considering that *ARC3* is a negative regulator of chloroplast division, the high expression levels of *ARC3* in *P. pellucida*

could very well be responsible (Figure 4.11B-E). Additionally, the possibility of *P. pellucida* lacking *PARC6* (functional or otherwise; Figure 4.8) is interesting, as *Atparc6* mutants exhibit enlarged chloroplasts (Glynn et al., 2009). Overexpression of *ARC3* in *Atparc6* plants, which may be equivalent to what I have found in *P. pellucida*, completely inhibits Z-ring assembly, producing huge chloroplasts (Chen et al., 2019a). It would be interesting to look at FtsZ immunolocalization in the chloroplasts of *P. pellucida*, as the morphological patterns could indicate whether overexpression of *ARC3* and or reduction of *PARC6* are causing the inhibition of chloroplast division directly in the palisade cells of *P. pellucida*. Further, regarding the cell-specificity of the large-chloroplast phenotype in *P. pellucida*, it is possible that the scenarios outlined above act in an additive manner. For example, if *P. pellucida* has lost *PARC6*, perhaps palisade-specific overexpression of *ARC3* determines the tissue-specificity observed.

Conclusions

Together, the *P. dahlstedtii* genome and *P. pellucida* transcriptome contribute to the representation of the early-diverging angiosperms and serve as high-quality references for my work on identifying regulators of chloroplast population morphology. Truncation or reduced expression of *PARC6* coupled with overexpression of *ARC3* may explain the large chloroplast phenotype observed in *P. pellucida*. Similarly, differences in the expression of *GCI* in *P. pellucida* compared to *P. dahlstedtii* could also contribute to this phenotype. Additionally, I have a new list of potential orthogroups of interest expressed differently between the two species that could be explored in the future, as this cell-specific phenotype is likely regulated on multiple levels.

Materials and methods

Plants and growth conditions

See Chapter 2 Methods.

Genome size, ploidy, and heterozygosity estimates

Genome sizes were estimated from healthy young leaf tissue using flow cytometry carried out by Aru K. Arumuganathan (Flow Cytometry Core, Benaroya Research Institute at Virginia Mason, Seattle, WA) as described previously (Arumuganathan and Earle, 1991). Three technical replicates were included for each species.

Ploidy was estimated by counting DAPI (4',6-diamidino-2-phenylindole)-stained metaphase chromosomes derived from root tip meristematic cells by Livia do Vale Martins (Jiming Jiang lab, Michigan State University Dept. of Plant Biology) as described previously (Brose et al., 2021).

Genome heterozygosity was estimated using GenomeScope 2.0 (Ranallo-Benavidez et al., 2020) and Jellyfish/2.2.10 (Marçais and Kingsford, 2011) with the PE150 genomic DNA (gDNA) Illumina data from *P. dahlstedtii* (described below, after trimming). Ploidy and heterozygosity were further confirmed using smudgeplot v0.2.4 (Ranallo-Benavidez et al., 2020) with KMC/3.1.2rc1-Python-3.9.5 (Kokot et al., 2017).

DNA extraction, sequencing, and read QC

As a general note, all library prep and sequencing described in this thesis was done by the Michigan State University (MSU) Research Technology Support Facility (RTSF) Genomics Core. For extraction of genomic DNA, *P. dahlstedtii* plants were dark-adapted for 48 hours, after which 5-10 g of leaf tissue (20-25 leaves) were harvested from the healthiest largest individual, frozen in liquid nitrogen, and stored at -80°C. High molecular weight (HMW) gDNA was

extracted by Krystle Wiegert-Rininger (MSU RTSF) using methods adapted from the Qiagen Genomic-tip Protocols and Qiagen Genomic DNA Handbook. Tissue was ground and 1 g added to Qiagen lysis Buffer G2 (10 mL) supplemented with lysing enzyme (0.5 mg/mL; Sigma Cat#L1412-5G), Pectinase (0.5mg/mL; Sigma Cat# P2401), and Viscozyme L (5%; Millipore Sigma Cat# V2010-50) and incubated for 1 hour at 37°C with gentle agitation. DNase-free RNase A (200 µL at 10 mg/mL; Thermo Fisher Cat# EN0531) was added, followed by incubation for 1 hr at 37°C. Next, 400 µL of Proteinase K (20 mg/mL; Qiagen Cat# 19133) was added, and the lysate was incubated for 2 hr at 50°C, after which the lysate was filtered using Qiagen Genomic-tips (20/G; Qiagen Cat# 10223) following the manufacturer's instructions, followed by isopropanol precipitation and elution. The resulting gDNA was stored at -80°C and subsequently used for all genome-based sequencing.

For estimating *P. dahlstedtii* heterozygosity and assembly polishing, one Illumina TruSeq Nano DNA stranded library (PE150) was prepared by the RTSF and sequenced on a single Illumina HiSeq 4000 lane (~ Gb). Illumina reads, quality was assessed before and after trimming using FastQC/0.11.7 (Andrews, 2010). Reads were trimmed using Trimmomatic/0.38 (Bolger et al., 2014).

The HMW gDNA was prepared using the Oxford Nanopore SQK-LSK109 Ligation Sequencing Kit and sequenced on two Oxford Nanopore PromethION FLO-MIN111 flow cells following the manufacturer's recommendations (<https://nanoporetech.com/>). Approximately 20 hours after the start of the run, it was paused, the flow cell was flushed with nuclease following Nanopore's recommended procedure, and a second aliquot of the library was loaded for the run to continue. The total flow cell run time was 72 hours. Raw Nanopore (ONT) reads were base-called using guppy v4.0.11 (<https://nanoporetech.com/>). Reads passing QC (13,099,351; 122.82

Gb) were filtered for >10 kb. ONT read trimming was done using Porechop/0.2.4-Python-3.6.6 (Wick et al., 2018).

RNA extraction, sequencing, and read QC

mRNAseq from the earliest leaf stage (A; Chapter 2) for both species and whole *P. pellucida* seedlings is described in Chapter 3. Mid- and late-stage leaves (stages C and E, respectively; Chapter 2) from *P. dahlstedtii* and *P. pellucida* were identified in Chapter 2, and RNA was extracted as described in Chapter 3.

Illumina TruSeq stranded mRNA libraries (SE50) were prepared and sequenced on two lanes of an Illumina HiSeq 4000. For the PE150 libraries described in Chapter 3, only reads greater than 85 bp were kept for downstream analyses. Read quality was assessed before and after trimming using FastQC/0.11.7 (Andrews, 2010). Reads were trimmed using Trimmomatic/0.38 (Bolger et al., 2014).

For *P. dahlstedtii*, two replicates each of cDNA-PCR sequencing libraries (PCS109) from mid- and late-stage leaves (stages C and E, respectively; Chapter 2) were sequenced together on one Oxford Nanopore GridION flow cell (FLO-MIN106). Bases were called using guppy v4.0.11 (<https://nanoporetech.com/>) and reads trimmed using Porechop/0.2.4-Python-3.6.6.

Genome assembly and annotation

canu/2.0-Java-11 (Koren et al., 2017) was used to assemble Nanopore genomic DNA reads derived from *P. dahlstedtii* leaf tissue with the following commands—`canu genomeSize=1.27g -nanopore *.fastq.gz gridOptions="--time=120:00:00" gridOptionscns="--mem-per-cpu=64g" corOverlapper=minimap`. The resulting assembly was successively polished using Nanopore gDNA reads (described above) with four iterations of Racon/1.4.0 (Vaser et al., 2017) followed by medaka/0.10.0.Py3 (<https://github.com/nanoporetech/medaka>). A final

polishing step was performed using Pilon/1.23 (Walker et al., 2014) with clean Illumina PE150 gDNA reads (described above). Nanopore cDNA reads were mapped to the unmasked assembly using minimap2/2.17 (Li, 2018), and a random subset of small contigs (<35 kb) were manually assessed for cDNA alignments (evidence of gene presence) using SAMTools/1.11 (Li et al., 2009) and IGV (Thorvaldsdóttir et al., 2013). Contigs <35 kb were purged from the assembly using a custom UNIX script, as the small contigs observed did not contain cDNA alignments, indicating a lack of gene presence. Assembly completeness was assessed at each step described above against BUSCO's embryophyta_odb10 database (Creation date: 2020-09-10, number of species: 50, number of BUSCOs: 1614) using BUSCO/5.0.0-Python-3.7.4 (Simão et al., 2015). Assembly statistics were calculated using QUAST/5.0.2.Py3 (Gurevich et al., 2013).

For annotation, the assembly was first scanned for repetitive elements using RepeatModeler/2.0.1 (Flynn et al., 2020), and a soft-masked version of the assembly was generated using RepeatMasker/4.0.5 (Smit et al., 2016). *P. dahlstedtii* cDNA Nanopore reads and genome-guided assembled transcripts (described below) were mapped to the genome using minimap2/2.17 (Li, 2018). The resulting BAM files were sorted using SAMtools/1.11 (Li et al., 2009), and GTF files were converted to GFF3 using a custom script (reformat_exonerate_transcript_output_gff.pl). Reference protein sequences from *P. nigrum* (Hu et al., 2019) and *A. thaliana* (Araport11_pep_20210622; Cheng et al., 2017) were mapped to the genome using Exonerate/2.2.0 (Slater and Birney, 2005). A handful of alignments were assessed manually in IGV (Version 2.10.3; Thorvaldsdóttir et al., 2013).

For gene annotation, MAKER/2.31.9 (Campbell et al., 2014; Bowman et al., 2017), SNAP (FATHOM version 2006-07-28) (Korf, 2004), and augustus/3.3.3.Py3 (Stanke et al., 2008) were run sequentially, followed by a second round of MAKER that incorporated all gene

predictions. A random selection of annotations was manually assessed in IGV (Version 2.10.3; Thorvaldsdóttir et al., 2013) alongside reference protein and native RNAseq aligned evidence to assess gene-prediction accuracy. Protein domain prediction was conducted using hmmscan (HMMER/3.2.1; Eddy, 2011) using the Pfam-A.hmm database (15-Nov-2021; Finn et al., 2016). Genes lacking evidence of any kind (RNAseq, reference protein, or a Pfam domain) were purged, generating the ‘MAKER standard gene set.’ Two approaches identified transposable element-related genes. First, the genome was searched via BLAST+/2.9.0 (Camacho et al., 2009) against the LTRretriever/2.7 Tspaces020812DNA and Tspaces020812LINE databases (Ou and Jiang, 2018) with an E-value cutoff of 1e-10. Second, HMMER/3.2.1 (Eddy, 2011) was used to search the genome against the gypsy_db_3.1b2.hmm database with flags --domE 1e-5 -E 1e-5. All TE genes identified were purged from the genome annotation files. Next, deFusion (Wang et al., 2021) was used to identify potentially-fused tandem duplicated genes, which were observed by hand in IGV (Version 2.10.3; Thorvaldsdóttir et al., 2013) and re-annotated as needed using MAKER/2.31.9 (Campbell et al., 2014; Bowman et al., 2017). Putative functional annotations for the *P. dahlstedtii* predicted protein sequences were identified by a BLAST+/2.9.0 (Camacho et al., 2009) search against the *A. thaliana* proteome (Araport11_pep_20210622; Cheng et al., 2017) with -evalue 1e-6 -max_hsps 1 -max_target_seqs 5. Gene statistics were calculated using eval-2.2.8 (Keibler and Brent, 2003), and annotation edit distance (AED) scores were extracted from the MAKER output.

Genome-guided transcriptome assemblies

P. dahlstedtii and *P. pellucida* Illumina SE and PE reads were mapped to the *P. dahlstedtii* genome assembly using hisat2/2.1.0 (Kim et al., 2015) with flags --rna-strandness RF --mp 4,1 --score-min L,-0.1,-0.8. The resulting SAM files (one for each tissue type and species)

were converted to BAM format, sorted, and merged by species using SAMtools/1.9 (Li et al., 2009). Genome-guided transcript assemblies were built using Trinity/2.8.5 (Haas et al., 2013) with flags `--genome_guided_max_intron 10000 --max_memory 512G --CPU 32 --jaccard_clip --min_contig_length 500`.

Transcript assembly completeness was assessed using BUSCO/5.0.0-Python-3.7.4 (Simão et al., 2015) in ‘transcriptome’ mode against the *embryophyta_odb10* database. TransDecoder/2.1.0-Perl-5.24.1 (Haas et al., 2013) was run to predict and translate protein-coding sequences from the transcriptome assemblies, and functional annotations for the predicted protein sequences were identified as described above for the genome.

Orthogroup clustering, phylogenetic analyses, and multiple sequence alignments

For orthogroup clustering and phylogenetic analyses, the most recent reference protein sequence databases for *Selaginella moellendorffii*, *Ceratopteris richardii*, *Thuja plicata*, *Amborella trichopoda*, *Nymphaea colorata*, *Zostera marina*, *Spirodela polyrhiza*, *Zea mays*, *Cinammomum kanehirae*, *Aquilegia coerulea*, *Vitis vinifera*, and *Theobroma cacao* were obtained from Phytozome v13 (<https://phytozome-next.jgi.doe.gov/>; Goodstein et al., 2012). In addition, *Arabidopsis thaliana* (Araport11, 2022-01-03; Berardini et al., 2015), *Pinus taeda* v1.0 and *Picea abies* v1.0 (congenie; Sundell et al., 2015), *Azolla filiculoides* v1.1 (FernBase; Li et al., 2018), and *Piper nigrum* (Hu et al., 2019) were obtained from their respective sources. Orthologous protein sequences were clustered using OrthoFinder/2.5.4-Python-3.7.4 and DIAMOND/2.0.1 with the default settings (Emms and Kelly, 2015). The R package UpSet (Conway et al., 2017) was used to view species intersections between orthogroups. All phylogenetic trees were visualized using iTOL (Letunic and Bork, 2021). Initial multiple sequence alignments were generated with OrthoFinder/2.5.4-Python-3.7.4—the addition of other

sequences, editing, or refinement was done using MAFFT/7.453 (Katoh et al., 2019). Alignments were viewed in Jalview (version 2) (Waterhouse et al., 2009).

Differential gene expression

The PE150 Illumina mRNA reads (described in Chapter 2) were trimmed to a length of 50 nt using Trimmomatic/0.38 (Bolger et al., 2014) and treated as SE reads for expression analyses alongside the SE50 libraries described above. Reads were pseudomapped to the transcripts belonging to their species of origin, and counts were quantified with kallisto/0.46.1 (Bray et al., 2016). Gene expression levels were normalized between replicate libraries using RUVr from the RUVseq package (Risso et al., 2014). Within-species differential gene expression analyses were conducted using edgeR using the GLM approach (version 3.36.0) (Robinson et al., 2009). Contrasts were made between each pairwise comparison of leaf developmental stages available (A: C, A: E, and C: E).

Orthogroup expression

Orthogroups were identified using OrthoFinder (Emms and Kelly, 2015), as described above. All other species included in the original run were removed, leaving only sequences from *P. dahlstedtii* and *P. pellucida*. Species-specific orthogroups were removed, meaning each orthogroup retained contained at least one gene from both species. Orthogroup expression (OG_{exp}) was calculated within species at each leaf developmental stage—edgeR-derived \log_2 CPM values (Robinson et al., 2009) were summed for the genes present in each orthogroup (

Equation 4.1). The difference in OG_{exp} between species (ΔOG_{exp}) was calculated separately for each stage, where *P. pellucida* OG_{exp} was subtracted from *P. dahlstedtii* (Equation 4.2). To calculate the cumulative ΔOG_{exp} over development ($\Sigma \Delta OG_{exp}$), the absolute value of ΔOG_{exp} was taken for each developmental stage. These values were then summed (Equation 4.3). The intra-species correlation in OG_{exp} over leaf development ($r_{OG_{exp}}$) was determined using the rank-based Kendall's tau statistic with VGAM in R (Yee, 2020). Violin plots showed the distribution of these data, and $\alpha=0.05$ was set to identify significantly different orthogroups. The *Peperomia* GO-term category orthogroups (proteasome, actin, chloroplast, and photosynthesis) included in the $r_{OG_{exp}}$ plot were identified using the GO-term list available for *A. thaliana* at UniProt (UP000006548; The UniProt Consortium, 2021).

Data manipulation, plots, and statistical analyses

Unless otherwise noted, R version 4.1.2 (2021-11-01) -- "Bird Hippie" and RStudio 2021.09.0+351 "Ghost Orchid" Release (077589bcad3467ae79f318afe8641a1899a51606, 2021-09-20) for Windows Mozilla/5.0 (Windows NT 10.0; Win64; x64) were used for data and statistical analyses (RStudio Team, 2020; R Core Team, 2021). Plots were made using ggplot2 (Wickham, 2016). Viridis was used for all non-ggplot2 color palettes (Garnier, 2018).

Data availability

The code used for this project can be found at https://github.com/AFrolicOfFerns/peperomia_genome_expression_analyses_2022. All raw sequencing files have been deposited at NCBI and will be released upon publication. The *P. dahlstedtii* genome has been deposited with CoGe (#62812, unmasked genome with annotations; #62859, masked assembly). The genome-guided transcriptome assembly for *P. pellucida* has

been deposited with Zenodo under Pep_pel_assembly_MSU_V2_2021

10.5281/zenodo.5974753. All data and code are set for public release upon publication.

Acknowledgments

This work was supported by the James E. Rodman Botany Endowment Fund in 2019 and 2020 and by the Illumina Grant Program in 2020. Kevin Childs (Michigan State University, Dept. of Plant Biology) dedicated a lot of time walking me through the genome annotation process in addition to providing insight on sequencing strategy and other genomics issues. Robin Buell (University of Georgia, Crop and Soil Sciences) provided essential guidance and input on genome assembly and assessment. Shin-Han Shiu (Michigan State University, Dept. of Plant Biology) was immensely helpful in the orthogroup expression work. Aru K. Arumuganathan (Flow Cytometry Core, Benaroya Research Institute at Virginia Mason, Seattle, WA) made Flow cytometry genome size estimates. Chromosome counts were made by Livia do Vale Martins (Jiming Jiang lab, Michigan State University Dept. of Plant Biology). *P. dahlstedtii* genomic DNA was isolated and prepared for ONT sequencing by Krystle Wiegert-Rininger (Michigan State University RTSF).

REFERENCES

REFERENCES

- Ahmadabadi M, Bock R** (2010) Development of a highly responsive leaf-based regeneration system for *Peperomia* species. **34**: 329–334
- Andrews S** (2010) FastQC: a quality control tool for high throughput sequence data.
- Arumuganathan K, Earle ED** (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Report* **9**: 208–218
- Ashton-Butt A, Aryawan AAK, Hood ASC, Naim M, Purnomo D, Suhardi, Wahyuningsih R, Willcock S, Poppy GM, Caliman J-P, et al** (2018) Understorey vegetation in oil palm plantations benefits soil biodiversity and decomposition rates. *Front For Glob Change* **1**: 10
- Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, Huala E** (2015) The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *genesis* **53**: 474–485
- Bolger AM, Lohse M, Usadel B** (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120
- Bowman MJ, Pulman JA, Liu TL, Childs KL** (2017) A modified GC-specific MAKER gene annotation method reveals improved and novel gene predictions of high and low GC content in *Oryza sativa*. *BMC Bioinformatics* **18**: 522
- Bray NL, Pimentel H, Melsted P, Pachter L** (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–527
- Brose J, Lau KH, Dang TTT, Hamilton JP, Martins L do V, Hamberger B, Hamberger B, Jiang J, O’Connor SE, Buell CR** (2021) The *Mitragyna speciosa* (Kratom) Genome: a resource for data-mining potent pharmaceuticals that impact human health. *G3 GenesGenomesGenetics* **11**: jkab058
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL** (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421–429
- Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al** (2014) MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. *Plant Physiol* **164**: 513–524
- Chaw S-M, Liu Y-C, Wu Y-W, Wang H-Y, Lin C-YI, Wu C-S, Ke H-M, Chang L-Y, Hsu C-Y, Yang H-T, et al** (2019) Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. *Nat Plants* **5**: 63–73
- Chen C, Cao L, Yang Y, Porter KJ, Osteryoung KW** (2019a) ARC3 Activation by PARC6 Promotes FtsZ-Ring Remodeling at the Chloroplast Division Site. *Plant Cell* **31**: 862–885

Chen J, Hao Z, Guang X, Zhao C, Wang P, Xue L, Zhu Q, Yang L, Sheng Y, Zhou Y, et al (2019b) *Liriodendron* genome sheds light on angiosperm phylogeny and species–pair differentiation. *Nat Plants* **5**: 18–25

Chen L-Y, VanBuren R, Paris M, Zhou H, Zhang X, Wai CM, Yan H, Chen S, Alonge M, Ramakrishnan S, et al (2019c) The bracteatus pineapple genome and domestication of clonally propagated crops. *Nat Genet* **51**: 1549–1558

Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J* **89**: 789–804

Colletti KS, Tattersall EA, Pyke KA, Froelich JE, Stokes KD, Osteryoung KW (2000) A homologue of the bacterial cell division site-determining factor MinD mediates placement of the chloroplast division apparatus. *Curr Biol* **10**: 507–516

Conway JR, Lex A, Gehlenborg N (2017) UpSetR: An R Package for the Visualization of Intersecting Sets and their Properties. *bioRxiv* 120600

Dumschott K, Schmidt MH-W, Chawla HS, Snowdon R, Usadel B (2020) Oxford Nanopore Sequencing: New opportunities for plant genomics? *J Exp Bot* **71**: eraa263-

Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**: e1002195-16

Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**: 157

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279–D285

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF (2020) RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci* **117**: 9451–9457

Fosberg RF, Sachet M-H (1975) *Flora of Micronesia, 2: Casuarinaceae, Piperaceae, and Myricaceae*. *Smithson Contrib Bot* 1–32

Frenzke L, Goetghebeur P, Neinhuis C, Samain M-S, Wanke S (2016) Evolution of Epiphytism and Fruit Traits Act Unevenly on the Diversification of the Species-Rich Genus *Peperomia* (Piperaceae). *Front Plant Sci* **7**: 13410

Frenzke L, Scheiris E, Pino G, Symmank L, Goetghebeur P, Neinhuis C, Wanke S, Samain M-S (2015) A revised infrageneric classification of the genus *Peperomia* (Piperaceae). *Taxon* **64**: 424–444

Gao H, Kadirjan-Kalbach D, Froehlich JE, Osteryoung KW (2003) ARC5, a cytosolic dynamin-like protein from plants, is part of the chloroplast division machinery. *Proc Natl Acad Sci* **100**: 4328–4333

Garnier S (2018) viridis: Default Color Maps from “matplotlib.”

Glynn JM, Yang Y, Vitha S, Schmitz AJ, Hemmes M, Miyagishima S, Osteryoung KW (2009) PARC6, a novel chloroplast division factor, influences FtsZ assembly and is required for recruitment of PDV1 during chloroplast division in Arabidopsis. *Plant J* **59**: 700–711

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* **40**: D1178–D1186

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075

Gutierrez YV, Yamaguchi LF, Moraes MM de, Jeffrey CS, Kato MJ (2016) Natural products from *Peperomia*: occurrence, biogenesis and bioactivity. *Phytochem Rev* **15**: 1009–1033

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**: 1494–1512

Hu L, Xu Z, Wang M, Fan R, Yuan D, Wu B, Wu H, Qin X, Yan L, Tan L, et al (2019) The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. *Nat Commun* **10**: 4702

Itoh R, Fujiwara M, Nagata N, Yoshida S (2001) A Chloroplast Protein Homologous to the Eubacterial Topological Specificity Factor MinE Plays a Role in Chloroplast Division. *Plant Physiol* **127**: 1644–1655

Itoh RD, Ishikawa H, Nakajima KP, Moriyama S, Fujiwara MT (2018) Isolation and analysis of a stromule-overproducing Arabidopsis mutant suggest the role of PARC6 in plastid morphology maintenance in the leaf epidermis. *Physiol Plant* **162**: 479–494

Kadirjan-Kalbach DK, Turmo A, Wang J, Smith BC, Chen C, Porter KJ, Childs K, DellaPenna D, Osteryoung KW (2019) Allelic variation in the chloroplast division gene FtsZ2-2 leads to natural variation in chloroplast size. *Plant Physiol* **181**: 1059–1074

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* **20**: 1160–1166

Keibler E, Brent MR (2003) Eval: A software package for analysis of genome annotations. *BMC Bioinformatics* **4**: 50

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357–360

Kokot M, Długosz M, Deorowicz S (2017) KMC 3: counting and manipulating k-mer statistics. *Bioinformatics* **33**: 2759–2761

- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM** (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**: 722–736
- Korf I** (2004) Gene finding in novel genomes. *BMC Bioinformatics* **5**: 59
- Kubitzki K, Rohwer JG, Bittrich V, eds** (1993) Flowering Plants · Dicotyledons. doi: 10.1007/978-3-662-02899-5
- Letunic I, Bork P** (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**: W293–W296
- Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, de Vries J, Delaux P-M, Eily A, Koppers N, Kuo L-Y, Li Z, et al** (2018) Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat Plants* **4**: 460–472
- Li H** (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100
- 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
- Li Y, Wang L, Wang G, Feng Y, Liu X** (2017) AT2G21280 only has a minor role in chloroplast division. *Front Plant Sci* **8**: 2095
- Maple J, Fujiwara MT, Kitahata N, Lawson T, Baker NR, Yoshida S, Møller SG** (2004) GIANT CHLOROPLAST 1 is essential for correct plastid division in Arabidopsis. *Curr Biol* **14**: 776–781
- Maple J, Vojta L, Soll J, Møller SG** (2007) ARC3 is a stromal Z-ring accessory protein essential for plastid division. *EMBO Rep* **8**: 293–299
- Marçais G, Kingsford C** (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764–770
- Mcandrew RS, Olson BJSC, Kadirjan-Kalbach DK, Chi-Ham CL, Vitha S, Froehlich JE, Osteryoung KW** (2008) In vivo quantitative relationship between plastid division proteins FtsZ1 and FtsZ2 and identification of ARC6 and ARC3 in a native FtsZ complex. *Biochem J* **412**: 367–378
- Michael TP, VanBuren R** (2020) Building near-complete plant genomes. *Curr Opin Plant Biol* **54**: 26–33
- Miyagishima S, Froehlich JE, Osteryoung KW** (2006) PDV1 and PDV2 mediate recruitment of the dynamin-related protein ARC5 to the plastid division site. *Plant Cell* **18**: 2517–2530

Nakanishi H, Suzuki K, Kabeya Y, Miyagishima S (2009) Plant-Specific Protein MCD1 Determines the Site of Chloroplast Division in Concert with Bacteria-Derived MinD. *Curr Biol* **19**: 151–156

Okazaki K, Kabeya Y, Suzuki K, Mori T, Ichikawa T, Matsui M, Nakanishi H, Miyagishima S (2009) The PLASTID DIVISION1 and 2 Components of the Chloroplast Division Machinery Determine the Rate of Chloroplast Division in Land Plant Cell Differentiation. *Plant Cell* **21**: 1769–1780

Osteryoung KW, Stokes KD, Rutherford SM, Percival AL, Lee WY (1998) Chloroplast division in higher plants requires members of two functionally divergent gene families with homology to bacterial ftsZ. *Plant Cell* **10**: 1991–2004

Ou S, Jiang N (2018) LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons1[OPEN]. *Plant Physiol* **176**: 1410–1422

Platt A, Horton M, Huang YS, Li Y, Anastasio AE, Mulyati NW, Ågren J, Bossdorf O, Byers D, Donohue K, et al (2010) The Scale of Population Structure in *Arabidopsis thaliana*. *PLoS Genet* **6**: e1000843

Poncet V, Couderc M, Tranchant-Dubreuil C, Gomez C, Hamon P, Hamon S, Pillon Y, Munzinger J, de Kochko A (2012) Microsatellite markers for *Amborella* (*Amborellaceae*), a monotypic genus endemic to New Caledonia. *Am J Bot* **99**: e411–e414

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria

Ramu P, Esuma W, Kawuki R, Rabbi IY, Egesi C, Bredeson JV, Bart RS, Verma J, Buckler ES, Lu F (2017) Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. *Nat Publ Group* **49**: 959–963

Ranallo-Benavidez TR, Jaron KS, Schatz MC (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**: 1432

Rasingam L, Parthasarathy N (2009) Diversity of understory plants in undisturbed and disturbed tropical lowland forests of Little Andaman Island, India. *Biodivers Conserv* **18**: 1045–1065

Raynaud C, Cassier-Chauvat C, Perennes C, Bergounioux C (2004) An *Arabidopsis* homolog of the bacterial cell division inhibitor Sula is involved in plastid division. *Plant Cell* **16**: 1801–1811

Risso D, Ngai J, Speed TP, Dudoit S (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32**: 896–902

Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma Oxf Engl* **26**: 139–140

RStudio Team (2020) RStudio: Integrated Development Environment for R. RStudio, PBC., Boston, MA

Schmitz AJ, Glynn JM, Olson BJSC, Stokes KD, Osteryoung KW (2009) Arabidopsis FtsZ2-1 and FtsZ2-2 are functionally redundant, but FtsZ-based plastid division is not essential for chloroplast partitioning or plant growth and development. *Mol Plant* **2**: 1211–1222

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212

Simmonds SE, Smith JF, Davidson C, Buerki S (2021) Phylogenetics and comparative plastome genomics of two of the largest genera of angiosperms, Piper and Peperomia (Piperaceae). *Mol Phylogenet Evol* **163**: 107229

Slater GSC, Birney E (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**: 31

Smit A, Hubley R, Green P (2016) RepeatMasker at <http://repeatmasker.org>. *Nucleic Acids Res*

Soltis DE, Soltis PS (2019) Nuclear genomes of two magnoliids. *Nat Plants* **5**: 6–7

Soltis DE, Soltis PS, Endress PK, Chase MW, Manchester SR, Judd WS, Majure LC, Mavrodiev E (2018) Phylogeny and evolution of the angiosperms, Revised and updated edition. The University of Chicago Press, Chicago, IL

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**: 637–644

Sundell D, Mannapperuma C, Netotea S, Delhomme N, Lin Y-C, Sjödin A, Van de Peer Y, Jansson S, Hvidsten TR, Street NR (2015) The Plant Genome Integrative Explorer Resource: PlantGenIE.org. *New Phytol* **208**: 1149–1156

The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**: D480–D489

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

Vaser R, Sović I, Nagarajan N, Šikić M (2017) Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* **27**: 737–746

- Vercruyssen L, Tognetti VB, Gonzalez N, Dingenen JV, Milde LD, Bielach A, Rycke RD, Breusegem FV, Inzé D** (2015) GROWTH REGULATING FACTOR5 stimulates Arabidopsis chloroplast division, photosynthesis, and leaf longevity. *Plant Physiol* **167**: 817–832
- Vitha S, Froehlich JE, Koksharova O, Pyke KA, Erp H van, Osteryoung KW** (2003) ARC6 is a J-domain plastid division protein and an evolutionary descendant of the cyanobacterial cell division protein Ftn2. *Plant Cell* **15**: 1918–1933
- Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC** (2017) GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**: 2202–2204
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al** (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**: e112963
- Wang J, Zhao D, Childs KL** (2021) deFusion: a tool to improve predictions of tandemly duplicated genes created by the MAKER annotation pipeline.
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, et al** (2012) MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**: e49
- Wanke S, Samain M -S., Vanderschaeve L, Mathieu G, Goetghebeur P, Neinhuis C** (2006) Phylogeny of the genus *Peperomia* (Piperaceae) inferred from the trnK/matK region (cpDNA). *Plant Biol* **8**: 93–102
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ** (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191
- Weise SE, Carr DJ, Bourke AM, Hanson DT, Swarthout D, Sharkey TD** (2015) The arc mutants of *Arabidopsis* with fewer large chloroplasts have a lower mesophyll conductance. *Photosynth Res* **124**: 117–126
- Wick R, Volkening J, Loman N** (2018) Porechop.
- Wickham H** (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York
- Yee TW** (2020) The VGAM package for negative binomial regression. *Aust N Z J Stat* **62**: 116–131
- Zhang M, Chen C, Froehlich JE, TerBush AD, Osteryoung KW** (2016) Roles of *Arabidopsis* PARC6 in Coordination of the Chloroplast Division Complex and Negative Regulation of FtsZ Assembly. *Plant Physiol* **170**: 250–262
- Zhang M, Schmitz AJ, Kadirjan-Kalbach DK, TerBush AD, Osteryoung KW** (2013) Chloroplast Division Protein ARC3 Regulates Chloroplast FtsZ-Ring Assembly and Positioning in *Arabidopsis* through Interaction with FtsZ2. *Plant Cell* **25**: 1787–1802

Oxford Nanopore Technologies. <https://nanoporetech.com/>

Chapter 5. Future directions

Potentially immediate-future experiments in *Peperomia* to characterize the inhibition of chloroplast division on the molecular level

In this thesis I have covered what is known about the variation and regulation of chloroplast population morphology in the literature, both in terms of variation in nature and genetic mechanisms (Chapter 1), characterized such variation for the first time in several species of *Peperomia*, and established that chloroplast division is replaced entirely by expansion to maintain coverage in the palisade mesophyll cells of *P. pellucida* (Figure 2.4 & Figure I.4). In order to identify genes that regulate chloroplast population morphology in terms of division versus expansion, I produced novel RNAseq datasets for *Peperomia* spanning leaf development, primarily during leaf and cell expansion (Chapters 3 & 4). I have identified several of the chloroplast division genes as potential regulators of the large-chloroplast phenotype observed in *P. pellucida*, in addition to lists of orthologs expressed at different levels in the developing leaves of *P. dahlstedtii* and *P. pellucida*—novel candidates of interest in the regulation of this phenotype (Chapter 4). Lastly, I produced the first assembled genome for the *Peperomia* genus and leveraged these data to extract lists of orthologs that are differentially expressed in *P. pellucida* or *P. dahlstedtii* at different developmental stages—candidate genes of interest in the regulation of chloroplast morphology (Chapter 4; Table 4.9). There are several potential experiments are being considered to strengthen the impact and context of the work. These experiments are described below.

FtsZ and ARC3 protein levels and immunolocalization

In Chapter 4 I found that *ARC3* was expressed much more in *P. pellucida* compared to *P. dahlstedtii* leaves, and that *PARC6* was not detectable in *P. pellucida*. Further, I found that *FtsZ1* was upregulated in the young leaves of *P. dahlstedtii*, but not in *P. pellucida*. I hypothesized that

overexpression of *ARC3* and lower expression levels of *FtsZ1* are contributing to the large chloroplast phenotype observed in *P. pellucida*. It would be particularly interesting to observe FtsZ protein levels and localization in the chloroplasts of *P. pellucida*, as these data may provide support to the gene expression patterns I described in Chapter 4, where *P. dahlstedtii* has higher expression levels of *FtsZ1* compared to *P. pellucida* in younger leaves (Figure 4.11).

Western blots with protein extracts from *P. dahlstedtii* and *P. pellucida* incubated with FtsZ antibodies will be performed, time and funding permitting. These blots will indicate whether the FtsZ antibodies can bind to *Peperomia* FtsZs and, if so, provide us with relative protein levels in both species, which may help validate the elevated expression of *FtsZ1* I observed in *P. dahlstedtii* compared to *P. pellucida* (Figure 4.11).

As described in the literature and discussed more extensively in Chapter 4, overexpression of *ARC3* in *A. thaliana* coupled with the loss of *PARC6* results in a population of only two extremely large chloroplasts per cell—a more dramatic phenotype compared to either condition on its own, which can be attributed to the degree of disruption of FtsZ-ring formation (Zhang et al., 2013; Chen et al., 2019). Provided that *Peperomia* FtsZs are detected in the western blots described above, then we may consider imaging *P. dahlstedtii* and *P. pellucida* chloroplasts for FtsZ immunolocalization following the protocol established previously (Stokes et al., 2000; McAndrew et al., 2001; Vitha et al., 2001). I would expect to see punctate aggregates of FtsZ in *P. pellucida*, and normal Z-ring formation in *P. dahlstedtii*, following the observations made in *Atparc6* plants overexpressing *ARC3* (Chen et al., 2019).

Further, for any of the above proposed experiments, I may consider including two other large-chloroplast species, *P. metallica* and *P. meridiana*, and another small-chloroplast species such as *P. fraseri*, as consistent results between species with similar phenotypes would suggest

that the molecular mechanisms are the same within the genus (Figure 2.1). I would also employ *A. thaliana*, wild-type and mutants for the proteins being examined, as controls. However, should none of the *A. thaliana* antibodies proposed for use above cross-react with *Peperomia* proteins, *Peperomia*-specific antibodies for the FtsZs and ARC3 could be generated to address FtsZ expression and immunolocalization.

Regarding BolA1 and GCI as candidate regulators of chloroplast population morphology in P. pellucida

In Chapter 3 I identified the loss or truncation of *BolA1* and *GCI* as candidate regulators of chloroplast size in *P. pellucida*. In Chapter 4 I looked at chloroplast population morphology in *A. thaliana* T-DNA mutants for both genes and found a slight reduction in the number of chloroplasts per cell as a function of cell size, indicating that both genes may be moderate regulators of chloroplast population morphology or division. To better assess the reduced chloroplast number per cell size in these T-DNA mutants, I will take more measurements from cells varying more widely in size from the images that I already have in-hand—as my current dataset does not include a wider variation of cell sizes. I will then compare these data statistically with wild-type measurements. In *Peperomia*, I will also look at differential expression, both within and between species, of *BolA1*, as reduced expression, rather than complete loss of the gene as I initially thought was the case in Chapter 3, in *P. pellucida* could redeem *BolA1* as a potential regulator of chloroplast population morphology.

Molecular characterization of the leaf developmental stages established for P. dahlstedtii and P. pellucida

To better characterize the leaf developmental stages established between my two *Peperomia* species as described in Chapter 2, better quantification of the rate of cell expansion

between each developmental stage using existing cell size measurements would be informative. This would provide a quantitative measure by which I can compare these stages between species, ensuring that I am indeed comparing equivalent stages in gene expression analyses. Further, I am exploring the literature in an attempt to identify several key genes that are known to be expressed at defined developmental time points, such as for chloroplast expansion and development, as well as leaf maturity. Expression values for these genes could then be observed using the data described in Chapter 4. Together, these data might show trends in the correlation of certain developmental-stage-specific genes with the novel orthologs I identified in Chapter 4, which may allow for targeted selection of genes to pursue further as regulators of chloroplast population morphology during leaf development.

On continuing with *Peperomia* as a model system

I have had the unique opportunity during my PhD, during which I have learned how to fund my own work in addition to picking up many technical skills from microscopy to bioinformatics. I will carry all of these skills forward with me as a continue in academia. Further, I independently established a novel model system for the study of chloroplast population morphology. Considering the cell-specific large-chloroplast phenotype in *P. pellucida* and the issues I encountered with laser capture microdissection, novel pipelines such as single cell RNA sequencing (Efroni and Birnbaum, 2016) would be perfect for *Peperomia*, and I will consider proposing such an experiment to address studying the cell-specific differences in chloroplast population morphology observed in *Peperomia* (Figure 2.1; Figure I.3) in post-doc fellowships or early-career grant applications.

Moving forward: Single cell RNA sequencing and cytokinin signaling in rice

I may incorporate some of the data described in this thesis in post-doc and early career grant applications, as I am interested in continuing to study chloroplast population morphology and have established a preliminary link between cytokinin and the control of chloroplast division in *P. pellucida* (Figure II.2). This summer, I am returning to North Carolina to start a post-doc position with Joe Kieber, my former undergraduate advisor at the University of North Carolina, Chapel Hill. While I am returning to rice, my shift in focus will not be too distant, as I will be characterizing gene expression in the monocot shoot apical meristem. Excitingly, I will learn how to perform single-cell RNA sequencing—a technique that I plan to master and utilize in my career moving forward. While this project is funded, I am applying for one or two post-doc fellowships (NSF-PGRP and USDA, likely) to establish my independence as an investigator in preparation for applying for faculty positions in the future. I think this will be a very nice continuation of the work, especially considering the effort I have gone to establish *Peperomia* as a model system.

REFERENCES

REFERENCES

Chen C, Cao L, Yang Y, Porter KJ, Osteryoung KW (2019) ARC3 Activation by PARC6 Promotes FtsZ-Ring Remodeling at the Chloroplast Division Site. *Plant Cell* **31**: 862–885

Efroni I, Birnbaum KD (2016) The potential of single-cell profiling in plants. *Genome Biol* **17**: 65

McAndrew RS, Froehlich JE, Vitha S, Stokes KD, Osteryoung KW (2001) Colocalization of Plastid Division Proteins in the Chloroplast Stromal Compartment Establishes a New Functional Relationship between FtsZ1 and FtsZ2 in Higher Plants. *Plant Physiol* **127**: 1656–1666

Stokes KD, McAndrew RS, Figueroa R, Vitha S, Osteryoung KW (2000) Chloroplast division and morphology are differentially affected by overexpression of FtsZ1 and FtsZ2 genes in arabidopsis. *PLANT Physiol* **124**: 1668–1677

Vitha S, McAndrew RS, Osteryoung KW (2001) Ftsz Ring Formation at the Chloroplast Division Site in Plants. *J Cell Biol* **153**: 111–120

Zhang M, Schmitz AJ, Kadirjan-Kalbach DK, TerBush AD, Osteryoung KW (2013) Chloroplast Division Protein ARC3 Regulates Chloroplast FtsZ-Ring Assembly and Positioning in Arabidopsis through Interaction with FtsZ2. *Plant Cell* **25**: 1787–1802

APPENDICES

Appendix I

**Establishing novel methods for the quantification of chloroplast coverage toward
identifying regulators of this trait via natural variation**

Summary

Chloroplast coverage varies between species, though is generally high in the primary photosynthetic tissue(s) and does not differ drastically between closely related species. However, this has not been studied extensively.

The *Arabidopsis* leaf bundle sheath cells (surrounding the vasculature) have lower chloroplast coverage than spongy mesophyll (Kinsman and Pyke, 1998). Individual chloroplast area and coverage are higher in C₄ bundle sheath cells compared to C₃ species (Sage, 2004; Stata et al., 2014; Reeves et al., 2017). Similarly, the spongy mesophyll cells of C₄ species have smaller chloroplasts and lower coverage than C₃ (Stata et al., 2014). Presumably, these alterations are associated with different modes of photosynthesis and the subsequent functions of these cell types (Kinsman and Pyke, 1998; Stata et al., 2014).

A set of *A. thaliana* mutants had reduced chloroplast coverage (Larkin et al., 2016). These *REDUCED CHLOROPLAST COVERAGE (REC)* genes are not well understood, though they may be involved in retrograde signaling; thus, these proteins may play a role in the sensing mechanism by which coverage is maintained by the cell (Larkin et al., 2016).

Coverage is difficult to quantify, meaning it is not well suited to genetic or high-throughput screen experiments. However, previous work in the lab identified several accessions of *Arabidopsis* with a maximum difference in chloroplast coverage of nearly 30% (Figure I.1). I have confirmed this difference in coverage (Figure I.1) and established a modern protocol for measuring coverage that can be applied to future high-throughput experiments.

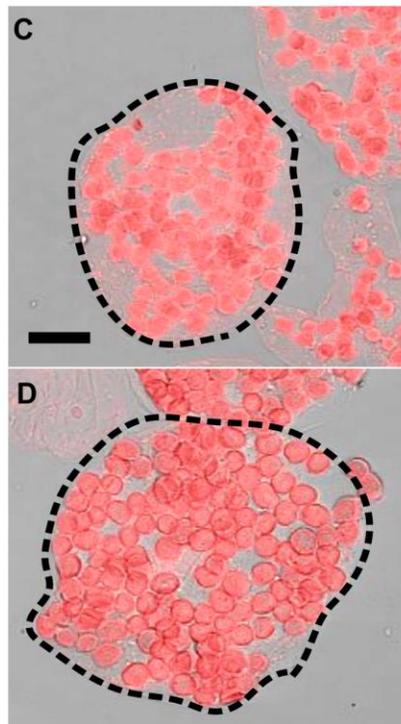
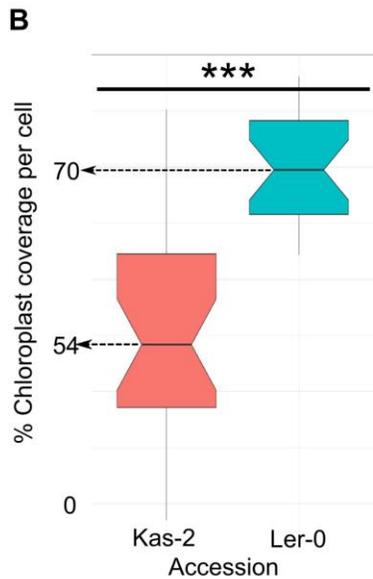
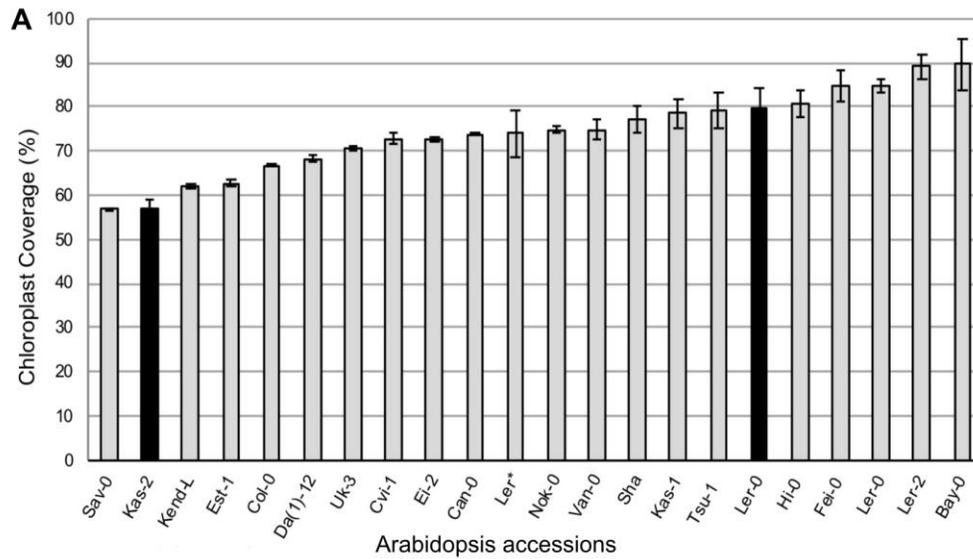


Figure I.1. Natural variation in chloroplast coverage between *Arabidopsis thaliana* accessions.

(A) % chloroplast coverage for each of the accessions shown. Error bars represent the standard error of the mean. Accessions of interest, Kas-2 and Ler-0 are colored with black bars. These data were collected and analyzed by Deena Kadirjan-Kalbach (Michigan State University). (B) Confirmation of the difference in % chloroplast coverage between Kas-2 and Ler-0 in my own hands using the original method without chlorophyll autofluorescence (Wilcox test; ***, $p < 0.001$). (C) Merged bright field and chlorophyll autofluorescence images from fixed mature leaf mesophyll cells taken from Kas-2 and (D) Ler-0. The cell wall is outlined with black dashed lines. The scale bar represents 20 μm for both images.

Establishing a new method for high-throughput screening of chloroplast coverage

To expedite the phenotyping necessary for large-scale projects, I adapted the traditional method leveraging modern tools (Figure I.2). Leaf tissue was fixed and treated as described previously (Pyke and Leech, 1991). Samples (small portions of fixed tissue, ~2x2 mm) were mounted in 75% sterile glycerol on glass slides and sealed using clear nail polish at the margins of the coverslip. In this manner, samples can be stored for months at 4°C, ready for imaging at any time. Mounted samples were imaged for chlorophyll autofluorescence using a confocal microscope. The resulting images—bright field and chlorophyll autofluorescence—were used to measure 2-D cell area and chloroplast coverage, respectively (Figure I.2). I further explored whether these measurements could be fully automated using machine learning with ilastik (Berg et al., 2019) and cellpose (Stringer et al., 2021). Ultimately, cell area was measured by hand, as it proved difficult for the boundaries of the cell to be accurately predicted. Chloroplast coverage was fully automated using FIJI custom IJM and macro batch processing scripts (Schindelin et al., 2012). Autofluorescence images were prepared and saved using another custom IJM script. Merged (bright-field and autofluorescence) images were opened in FIJI, and cell boundaries were traced by hand using the ‘freehand’ tool. These images were saved for later use. A custom macro batch processing script was used to process the remainder of the measurements. All pairs of cell-traced and modified autofluorescence images from a specified directory were opened and processed in batch. For each pair of images, FIJI converts and thresholds the autofluorescence image and measures the total area of fluorescence (chloroplast) within the traced cell boundary, yielding the percent coverage. A table of measurements was output by FIJI and saved by hand.

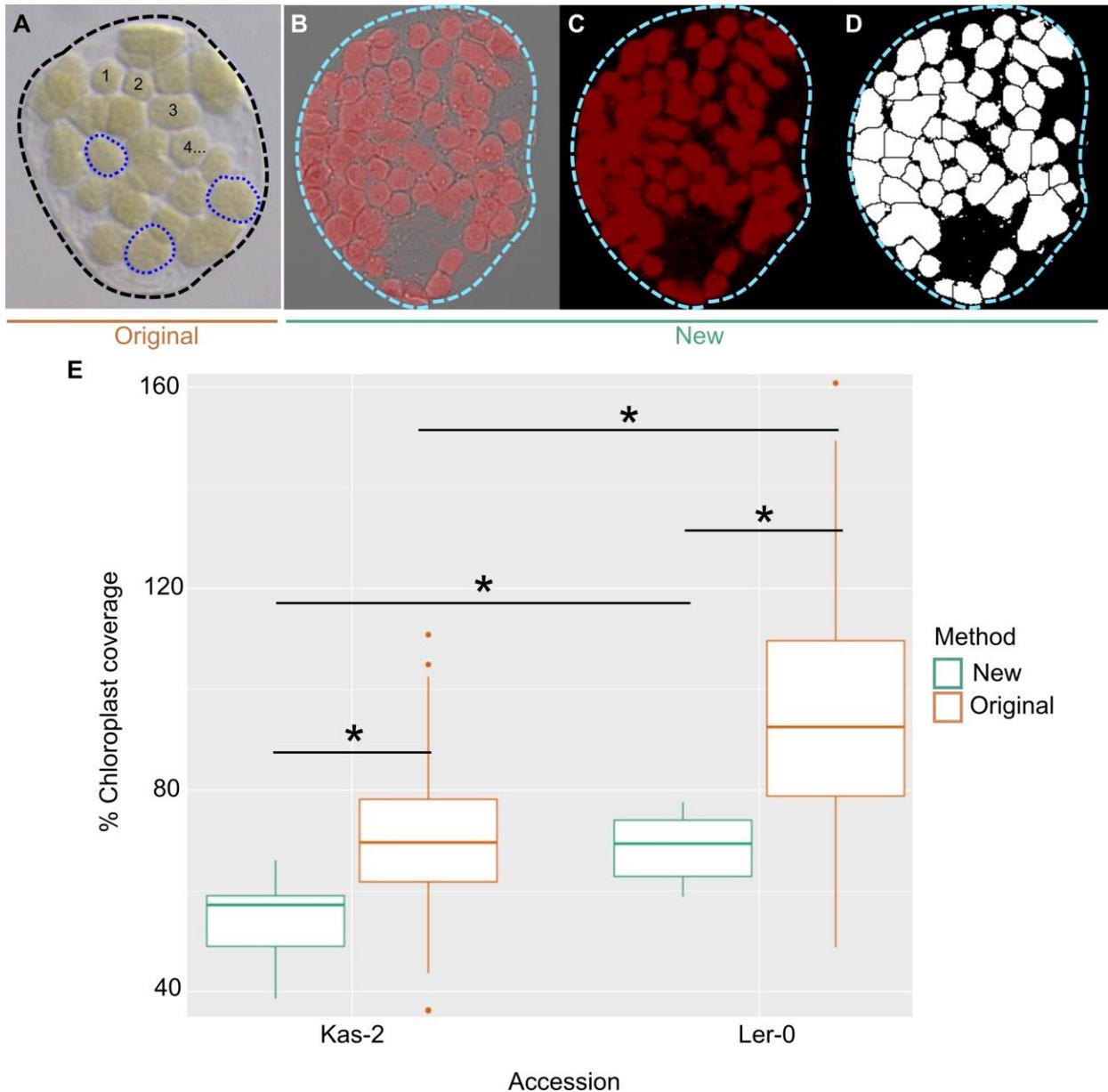


Figure I.2. Concept and validation behind a new method for high-throughput measurement of chloroplast coverage.

(A) Visualizations for the current method by which chloroplast coverage is determined and (B-D) the steps implemented in the new method. (B) Merged bright field and chlorophyll autofluorescence image, (C) autofluorescence alone, and (D) a mask representing the total chloroplast area relative to that of the cell generated using FIJI (Schindelin et al., 2012). (E) Comparison of the original and new methods for chloroplast coverage measurement. (Kruskal-Wallis test with Bonferroni correction; $\alpha=0.05$).

Natural variation in coverage in *Peperomia* species

Chloroplast coverage is also known to vary in plants naturally. In all *Peperomia* spp. observed, it seems that coverage differs greatly between palisade and spongy mesophyll cells (Figure 2.1; Ahmadabadi and Bock, 2012). Indeed, in *P. dahlstedtii*, coverage was significantly lower in spongy (<25%) versus palisade (~60%) mesophyll cells (Figure I.3). This phenomenon demonstrates a clear difference in the regulation of chloroplast coverage between palisade and spongy mesophyll cells in *Peperomia*.

Importantly, palisade cell chloroplast coverage appears to be maintained at similar levels between different *Peperomia* species. (Figure 2.1; Ahmadabadi and Bock, 2012). I confirmed that *P. dahlstedtii* and *P. pellucida* palisade cells maintain an average chloroplast coverage of 52-55% (Figure I.4). These phenomena are fascinating and should be studied further, as they may provide direct insight into how and why chloroplast coverage is regulated.

Materials and methods

Plant materials used

Arabidopsis lines were purchased from the Arabidopsis Biological Resource Center (ABRC) at Ohio State University (Kadirjan-Kalbach et al., 2019).

Plant growth conditions

Arabidopsis seeds were germinated after 3 nights of cold treatment for vernalization either in moist soil (Sure Mix on top of 1-inch vermiculite) or on sterile LS-plates and grown under 100 μ E m⁻²s⁻¹ (16-hour days) in a chamber. Seeds were sterilized with chlorine gas for 2-3 hours.

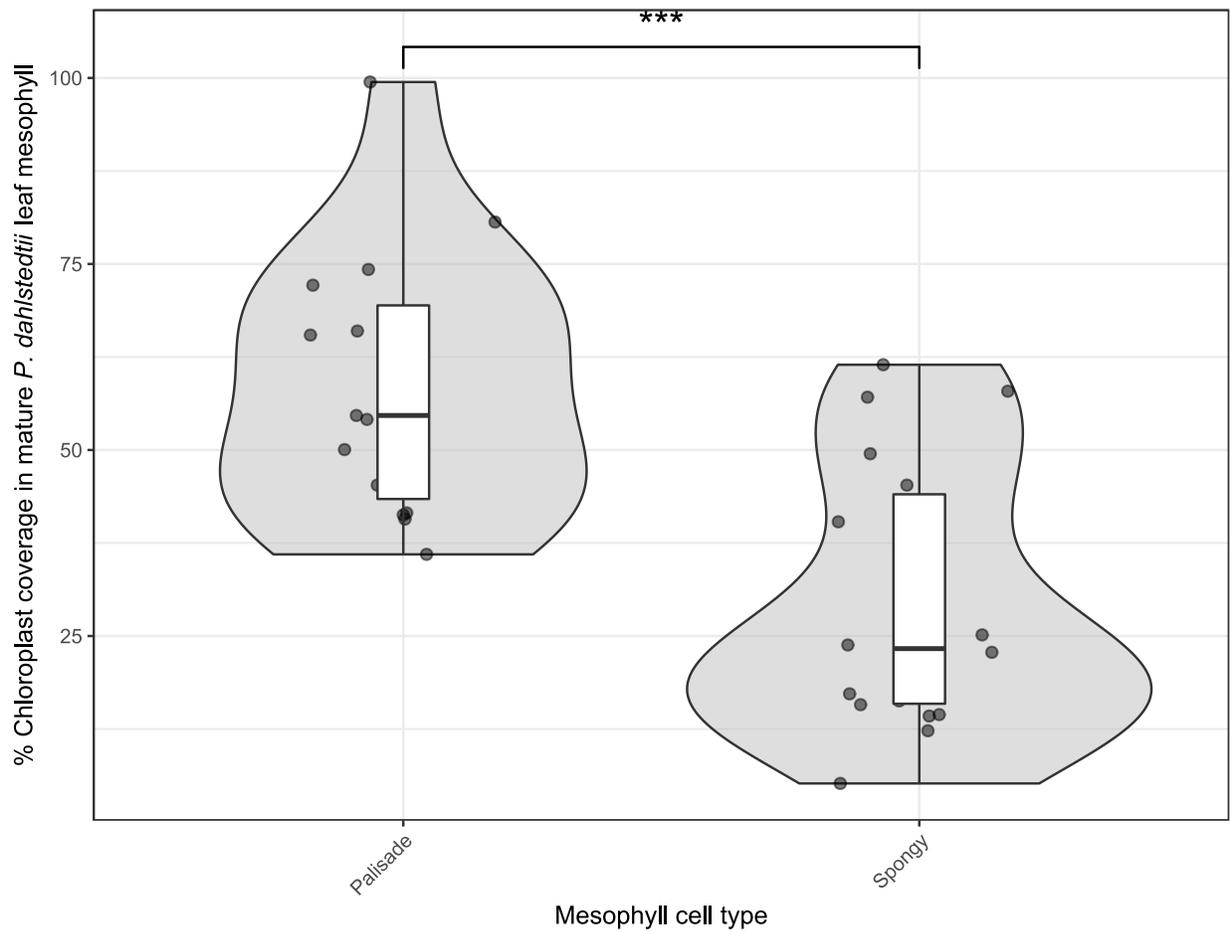


Figure I.3. Chloroplast coverage is higher in palisade compared to spongy mesophyll cells of *P. dahlstedtii*.

Chloroplast coverage (proportion of planar area occupied by the collective chloroplast population to that of the cell, shown as %) is significantly higher in palisade (P) compared to spongy mesophyll (SM) cells. Measurements were taken from mature leaves. (Wilcox test, $\alpha=0.05$).



Figure I.4. Chloroplast coverage in mature palisade mesophyll cells from *P. dahlstedtii* and *P. pellucida*.

Chloroplast coverage was measured using the new chlorophyll autofluorescence-based method established and described above in palisade cells extracted from mature leaves. A Wilcoxon test revealed no significant difference in coverage between species ($\alpha=0.05$).

Tissue fixation and microscopy

For chloroplast number and size observations, leaves were fixed in 3.5% glutaraldehyde for 1-2 hours with shaking in the dark. The tissue was subsequently loosened in 0.2 M Na₂-EDTA pH 9.0 for 30 minutes to one hour at 55°C (Pyke and Leech, 1994; Khoshnavesh and Sage, 2018). Slides were prepared by placing a small amount of leaf tissue in a drop of 75% glycerol (for long-term storage). After placing the coverslip, the sample was tapped down using a pencil eraser, effectively ‘squishing’ it and further separating the cells. Slides prepared with glycerol were sealed using nail polish and stored at 4°C. Differential Interference Contrast (DIC) images were taken on a Leica DMI3000 B microscope at 40x magnification. Fluorescent images of chlorophyll autofluorescence for chloroplasts (Excitation 633 nm, Emission 600-700 nm) were taken on a Nikon C2 confocal microscope at the MSU Center for Advanced Microscopy (CAM). Cell and chloroplast area measurements and counts were performed using the freehand tracing and counter tools in Fiji (Schindelin et al., 2012) (ImageJ Version 2.0.0-rc-69/1.52p).

Funding sources

The U.S. Department of Energy funded this project, award No. DE-FG02-06ER15808.

Acknowledgments

Deena Kadirjan-Kalbach (Department of Plant Biology, Michigan State University) originally identified the differences in coverage between *A. thaliana* accessions. Jonathan Sakkos (Plant Research Laboratory, Michigan State University) provided vital insight and advice on the automation-based portion of the methodology.

REFERENCES

REFERENCES

- Ahmadabadi M, Bock R** (2012) Plastid division and morphology in the genus *Peperomia*. *Biol Plant* **56**: 301–306
- Arends D, Prins P, Jansen RC, Broman KW** (2010) R/qtl: high-throughput multiple QTL mapping: Fig. 1. *Bioinformatics* **26**: 2990–2992
- Berg S, Kutra D, Kroeger T, Strahle CN, Kausler BX, Haubold C, Schiegg M, Ales J, Beier T, Rudy M, et al** (2019) ilastik: interactive machine learning for (bio)image analysis. *Nat Methods* **16**: 1226–1232
- El-Lithy ME, Bentsink L, Hanhart CJ, Ruys GJ, Rovito D, Broekhof JLM, Poel HJA van der, Eijk MJT van, Vreugdenhil D, Koornneef M** (2006) New Arabidopsis Recombinant Inbred Line Populations Genotyped Using SNPWave and Their Use for Mapping Flowering-Time Quantitative Trait Loci. *Genetics* **172**: 1867–1876
- Kadirjan-Kalbach DK, Turmo A, Wang J, Smith BC, Chen C, Porter KJ, Childs K, DellaPenna D, Osteryoung KW** (2019) Allelic variation in the chloroplast division gene *FtsZ2-2* leads to natural variation in chloroplast size. *Plant Physiol* **181**: 1059–1074
- Kinsman EA, Pyke KA** (1998) Bundle sheath cells and cell-specific plastid development in Arabidopsis leaves. *Development* **125**: 1815–1822
- Larkin RM, Stefano G, Ruckle ME, Stavoe AK, Sinkler CA, Brandizzi F, Malmstrom CM, Osteryoung KW** (2016) REDUCED CHLOROPLAST COVERAGE genes from Arabidopsis thaliana help to establish the size of the chloroplast compartment. *Proc Natl Acad Sci U S A* **113**: E1116-25
- Pyke KA, Leech RM** (1991) Rapid image analysis screening procedure for identifying chloroplast number mutants in mesophyll cells of Arabidopsis thaliana (L.) Heynh. *Plant Physiol* **96**: 1193–1195
- Reeves G, Grangé-Guermente MJ, Hibberd JM** (2017) Regulatory gateways for cell-specific gene expression in C4 leaves with Kranz anatomy. *J Exp Bot* **68**: 107–116
- Sage RF** (2004) The evolution of C4 photosynthesis. *New Phytol* **161**: 341–370
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, et al** (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**: 676–682
- Stata M, Sage TL, Rennie TD, Khoshravesh R, Sultmanis S, Khaikin Y, Ludwig M, Sage RF** (2014) Mesophyll cells of C4 plants have fewer chloroplasts than those of closely related C3 plants. *Plant Cell Environ* **37**: 2587–2600

Stringer C, Wang T, Michaelos M, Pachitariu M (2021) Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**: 100–106

Appendix II

Chloroplast and proplastid division in large-chloroplast-containing *Peperomia* species

***Peperomia dahlstedtii* and *Peperomia pellucida* likely retain similar numbers of proplastids in their meristematic cells**

Reduced rates of proplastid division in meristematic cells are known to affect chloroplast population morphology in *A. thaliana* (Pyke, 2009). Therefore, I wanted to see if *P. pellucida* might experience reduced proplastid division compared to *P. dahlstedtii*. Though proplastids are colorless and difficult to image (Butterfass, 1988; Pyke, 2009), previous observations suggest that the number of chloroplasts present in the mature guard cells of a given plant is reflective of its meristematic-cell proplastid numbers (Butterfass, 1988). Therefore, if *P. pellucida* experiences reduced rates of proplastid division, it is likely that the number of guard cell chloroplasts would be smaller than that of *P. dahlstedtii*. Interestingly, the average guard cell from *P. dahlstedtii* had 12 chloroplasts and *P. pellucida* 11.5, with no significant differences (Figure II.1). This suggests that both species keep similar numbers of proplastids in their meristematic cells, and that proplastid division is likely not disrupted in *P. pellucida*, further supporting my findings in Chapter 2 that a complete lack of chloroplast division in the palisade cells of *P. pellucida*, compensated for by expansion in individual chloroplast size, is responsible for the large-chloroplast phenotype observed in this species (Figure 2.4).

Chloroplast division is likely not wholly abolished in *Peperomia* species with large chloroplasts

Though chloroplast division does not occur in the palisade cells of *P. pellucida*, I wondered whether it is abolished or inhibited. While the mechanism is not yet understood, the phytohormone cytokinin is involved in chloroplast division (Cortleven and Schmölling, 2015; Vercruyssen et al., 2015). A cytokinin-sensitive *Physcomitrella patens* mutant with giant chloroplasts could be rescued by applying exogenous cytokinin or by transgenic expression of

Agrobacterium tumefaciens isopentenyl transferase (ipt)—the product of which is a protein required for cytokinin biosynthesis (Abel et al., 1989; Reski et al., 1991). Further, *A. thaliana* leaves treated with exogenous cytokinin have significantly increased numbers of chloroplasts per cell (Okazaki et al., 2009). In line with this work, I found that mature *P. metallica* leaves treated with cytokinin had significantly higher numbers of chloroplasts in their palisade mesophyll cells compared to the control group (Figure II.2). This indicates that chloroplast division is possible in the palisade mesophyll of *P. metallica* and its sister species (*P. pellucida* and *P. meridiana*) and that it likely is down-regulated or suppressed rather than abolished.

Materials and methods

Plant materials, growth conditions, and propagation

As described in Chapter 2.

Live cross-sections and leaf peels for imaging of guard cells

The following techniques were adapted from (Ruzin, 1999). Live cross-sections were made using fresh mature leaves. Leaves were removed from the plant and promptly placed in a shallow dish of sterile distilled H₂O. The leaf was gently held down under a layer of parafilm, and a double edge razor blade cleaned with 70% ethanol was used to slice several cross-sections <1 mm in thickness. Abaxial epidermal peels were taken directly from still-attached leaves by applying clear tape to the underside of the leaf and gently peeling away the epidermis or using a set of fine-tipped forceps to grasp the epidermis at the leaf edge and pull it away for removal. All live tissues were immediately mounted on a glass slide in distilled H₂O with a coverslip and imaged on a Leica DMI3000 B microscope.

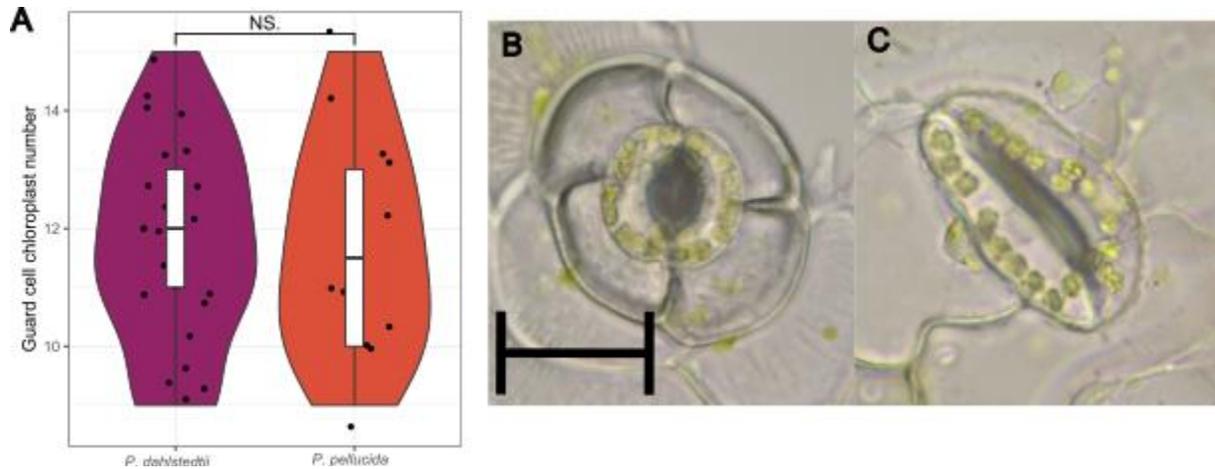


Figure II.1. Chloroplast numbers in guard cells from mature *P. dahlstedtii* and *P. pellucida* leaves.

(A) The number of chloroplasts was counted in guard cells from the mature leaves of *P. dahlstedtii* and *P. pellucida*. NS, no significant difference (Wilcox test, $\alpha=0.05$). (B-C) Representative images of guard cells used for quantification from (B) *P. dahlstedtii* and (C) *P. pellucida*. The scale bar represents 50 μm for both images.

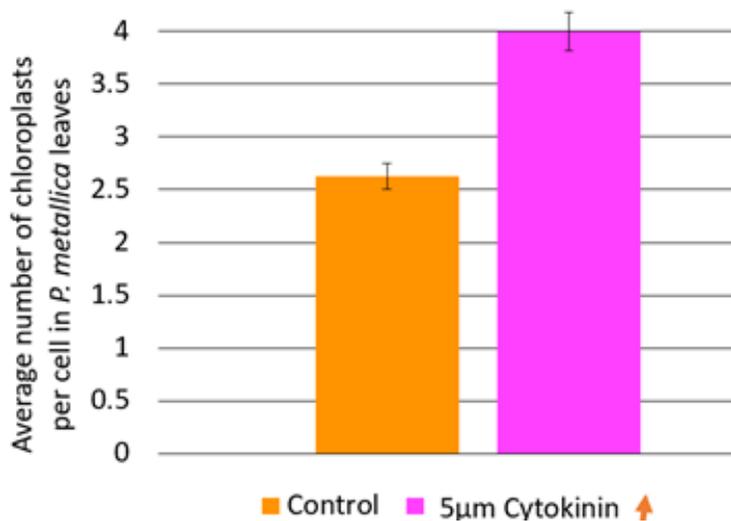


Figure II.2. Treatment with cytokinin increases the number of chloroplasts per cell in *P. metallica* leaves.

Mature *P. metallica* leaves were treated with solvent (control, orange) or 5 μM of cytokinin (pink) for two days. A student's t-test indicated statistical significance with $p < 0.05$ (not shown on the plot).

Cytokinin treatment assay

Healthy mature *P. metallica* leaves were removed from the plant and cut in half length-wise using a sterile razor blade, after which they were gently pressed (cut side down) into Murashige and Skoog (MS; Sigma, M5519) plates supplemented with 5 μ M N6-Benzyladenine solubilized in 1 M NaOH (BA, an artificial cytokinin; Sigma, B3408) or 1 M NaOH (as the negative control). The plates were sealed using mesh tape and moved to the growth chamber (same conditions described in Chapter 2) for three days. The leaves were then fixed, imaged, and quantified for chloroplast number, area, and cell area as described in Chapter 2.

Statistical analyses

As described in Chapter 2.

REFERENCES

REFERENCES

Abel WO, Knebel W, Koop H-U, Marienfeld JR, Quader H, Reski R, Schnepf E, Spörlein B (1989) A cytokinin-sensitive mutant of the moss, *Physcomitrella patens*, defective in chloroplast division. *Protoplasma* **152**: 1–13

Butterfass Th (1988) Nuclear control to plastid division. *Div. Segreg. Organelles*. Cambridge University Press, New York, pp 21–38

Cortleven A, Schmülling T (2015) Regulation of chloroplast development and function by cytokinin. *J Exp Bot* **66**: 4999–5013

Okazaki K, Kabeya Y, Suzuki K, Mori T, Ichikawa T, Matsui M, Nakanishi H, Miyagishima S (2009) The PLASTID DIVISION1 and 2 Components of the Chloroplast Division Machinery Determine the Rate of Chloroplast Division in Land Plant Cell Differentiation. *Plant Cell* **21**: 1769–1780

Pyke K (2009) *Plastid biology*. Cambridge University Press, Cambridge, UK ; New York

Reski R, Wehe M, Hadel B, Marienfeld JR, Abel WO (1991) Cytokinin and Light Quality Interact at the Molecular Level in the Chloroplast-Mutant PC22 of the Moss *Physcomitrella*. *J Plant Physiol* **138**: 236–243

Ruzin SE (1999) *Plant microtechnique and microscopy*. Oxford University Press

Vercruyssen L, Tognetti VB, Gonzalez N, Dingenen JV, Milde LD, Bielach A, Rycke RD, Breusegem FV, Inzé D (2015) GROWTH REGULATING FACTOR5 stimulates Arabidopsis chloroplast division, photosynthesis, and leaf longevity. *Plant Physiol* **167**: 817–832