DUAL SUBCELLULAR LOCALIZATION AND NON-PHOTOSYNTHETIC FUNCTIONS OF PHYLLOQUINONE IN PARASITIC AND NON-PARASITIC PLANTS

by

XI GU

(Under the Direction of Chung-Jui Tsai)

ABSTRACT

Phylloquinone (PhQ) is a group of lipid-soluble naphthoquinone derivatives produced by photosynthetic organisms to support photosystem I electron transport. Involvement of PhQ in non-photosynthetic plasma membrane redox activities of plants has been reported but is not well characterized due to challenges in preventing chloroplast contamination. This research aimed to understand the non-canonical function(s) and subcellular localization of PhQ biosynthesis using a photosynthesis-free study system, and to leverage the gained knowledge to assist the investigation in photosynthetic species.

Non-photosynthetic holoparasites offer a photosynthesis-free system to explore the non-canonical function of PhQ. However, available transcriptome assemblies were not of sufficient quality to study the PhQ biosynthetic pathway. To overcome the limitation, a Parallelized Local Assembly of Sequences (PLAS) pipeline was developed that showed improved performance over other de novo assembly algorithms. PLAS successfully reconstructed full-length transcripts for the entire PhQ biosynthetic pathway genes for the holoparasite *Phelipanche aegyptiaca* and two of its photosynthetic relatives. Careful inspection of the sequences revealed that the terminal two enzymes of the PhQ pathway have been redirected to the plasma membrane in the holoparasite, but remain plastid-targeted in the photosynthetic parasites. Comparative gene coexpression network analyses reveal an association of PhQ with plasma membrane redox

activities in the holoparasite. Plasma membrane PhQ biosynthesis was also predicted to exist as a minor route in multiple photoautotrophic species, indicating that the association between PhQ and the plasma membrane is evolutionarily conserved. Despite the insight from the parasitic plant system, investigation in photoautotrophic plants remains challenging, even when using heterotrophic tissues. The results from gene expression analyses revealed a dominant role of PhQ in photosynthesis, regardless of tissue. However, multiple lines of evidence indicated a large degree of plasticity of the PhQ biosynthetic pathway through lineage-dependent gene duplication, retention, and functional divergence among higher plants.

This work was the first to investigate the plasma membrane biosynthesis of PhQ and its non-photosynthetic function in a photosynthesis-free system. Results from this work open new opportunities for future investigations to confirm the function of PhQ in parasitic plants and to characterize the PhQ pathway gene duplication in photoautotrophic plants.

INDEX WORDS: phylloquinone, vitamin K1, parasitic plants, photoautotrophic plants, plasma membrane, electron transport, alternative splicing, transcriptome *de novo* assembly.

DUAL SUBCELLULAR LOCALIZATION AND NON-PHOTOSYNTHETIC FUNCTIONS OF PHYLLOQUINONE IN PARASITIC AND NON-PARASITIC PLANTS

by

XI GU

BS, Beijing Normal University, P.R. China, 2011

MS, The University of Georgia, 2015

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

DUAL SUBCELLULAR LOCALIZATION AND NON-PHOTOSYNTHETIC FUNCTIONS OF PHYLLOQUINONE IN PARASITIC AND NON-PARASITIC PLANTS

by

XI GU

Major Professor:	Chung-Jui Tsai
Committee:	Jonathan Arnold
	Jessica Kissinger
	James Leebens-Mack
	Scott Jackson

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2017

DEDICATION

I dedicate this dissertation to my loving parents Tianping Gu and Qincheng Guan, without whom I would not be here. I also dedicate this dissertation to my dearest boyfriend Minglu Gao who have provided endless encouragement and support for me throughout the process.

ACKNOWLEDGEMENTS

I would first like to thank my dissertation advisor CJ Tsai. CJ is the most diligent and meticulous scientist I have ever known. She has shown me a wonderful example from whom I will always learn during my future career. I am very grateful for the training and guidance that I received from her, especially for many hours of training on my writing. Her wide knowledge in both biology and bioinformatics guided me to make breakthroughs along my research. She has been very generous and kind to allow me to obtain a secondary Master's degree in Statistics, and offered all her help to advance my career. To me, she is not just a research advisor, but also a life mentor.

My advisory committee also deserves my gratitude. I thank Jim Leebens-Mack and Scott Jackson for their guidance on the field of evolution and crop science. I also would like to thank Jessica Kissinger for her invaluable suggestions on my coursework and dissertation, and Jonathan Arnold for his kind offer to be the co-advisor for my Statistics degree. I also want to give my special thanks for Scott Jackson for his referral for a summer internship, without which I wouldn't be able to obtain such an invaluable experience.

I also want to thank all fellow members of CJ's lab, especially Scott Harding who has provided insightful suggestions and discussions on my research projects, Kavita Aulakh, Naomi Rodman, Batbayar Nyamdari (together with Scott Harding) who has performed experiments to confirm my findings, Liangjiao Xue who has offered Bioinforamtics guidance along my graduate career.

I am very grateful for the seminar given by Claude W. dePamphilis from Penn State University in 2014 which inspired a large part of my dissertation. I also truly appreciate the

collaboration with James H. Westwood at Virgia Tech University who made the experimental validation of my research findings feasible.

Above all, special thanks are extended to my parents for their endless love, and my boyfriend Minglu for his accompany, encouragement and patience through the good times and bad.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

CHAPTER

1 INTRODUCTION AND LITERATURE REVIEW 1

Photosynthetic Functions of Vitamin K 1

Evolution of the PhQ Biosynthetic Pathway 3

Compartmentalization of PhQ Biosynthetic Pathway 4

Photosynthesis in Developing Seeds 6

Potential Involvement of PhQ in Plasma Membrane Electron Transport 7

Other Non-photosynthetic Functions of PhQ 9

Non-photosynthetic Holoparasite as a Study System 10

Objectives and Overview of Dissertation Chapters 13

Significance of This Work 15

References 16

2 PLAS: PARALLELIZED LOCAL *DE NOVO* ASSEMBLY OF SEQUENCES 28

Abstract 29

Introduction 30

Materials and Methods 31

Results 35

Discussions 41

References 87

3 PLASMA MEMBRANE PHYLLOQUINONE BIOSYNTHESIS: CONSERVATION AND DIFFERENTIAL EVOLUTION IN GREEN PLANTS AND HOLOPARASITES 90

Abstract 91

Introduction 92

Materials and Methods 93

Results and Discussion 96

Conclusions 104

References 124

4 EXPLORING NON-PHOTOSYNTHETIC FUNCTION OF PHQ BIOSYNTHESIS IN ARABIDOPSIS, POPULUS AND GLYCINE: A COMPARATIVE APPROACH 130

Abstract 131

Introduction 132

Materials and Methods 135

Results 136

Discussion 143

References 174

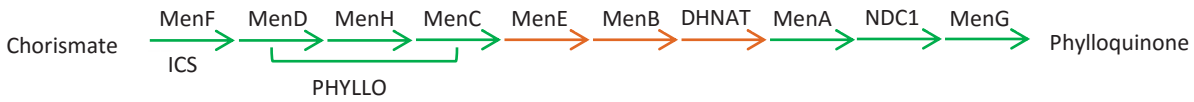
5 CONCLUDING REMARKS 181

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Photosynthetic Functions of Vitamin K

Phylloquinone (PhQ, 2-methyl-3-phytyl-1,4-naphthoquinone), also known as vitamin K1 (VK1), is a critical cofactor in the photosystem I (PSI) electron transport chain in plants (Itoh and Iwaki, 1989). Two PhQ molecules bind to the A_1 site of PSI (Brettel et al., 1986; Petersen et al., 1987), where each PhQ molecule transfers one electron from the chlorophyll *a* binding site (A_0) to the iron-sulphur center (F_x) (Sigfridsson et al., 1995; Boudreaux et al., 2001). This process involves a quinone/semi-quinone turnover.



PhQ is fat-soluble and contains a naphthoquinone ring and a phytyl moiety. More information about the phytyl moiety and its essential involvement in PhQ biosynthesis can be found in other recent papers (Lohr et al., 2011; Vranová et al., 2013; Wang et al., 2017). The scope of this review focuses on the biosynthesis of the naphthoquinone. The naphthoquinone ring originates from chorismate of the shikimate pathway, and is then modified via a series of enzymatic steps catalyzed by MenF (ICS), PHYLLO (MenD, MenH, MenC), MenE, MenB, DHNAT, MenA, NDC1 and MenG. The so-called Men proteins comprise the pathway as originally characterized in bacteria for biosynthesis of menaquinones (MKs or vitamin K2). The plant *Men* genes were identified and characterized in *Arabidopsis* in the late 2000s based on sequence similarity with *Men* genes of PhQ-synthesizing cyanobacterium *Synechocystis* PCC6803 and MK-synthesizing bacteria (reviewed in Van Oostende et al., 2011). *DHNAT*

encodes 1,4-dihydroxynaphthoyl-CoA thioesterase, which mediates a catalytic step that was thought until recently to be non-enzymatic (Widhalm et al., 2009). Subsequent genomic approaches and functional complementation experiments facilitated the identification of *DHNAT* in plants, but the deduced protein sequences reveal changes in the catalytic motif compared to cyanobacteria (Widhalm et al., 2012). *NDC1* encodes a type II NAD(P)H dehydrogenase, originally found to regulate the redox state of the plastoquinone pool of chloroplasts (Eugeni Piller et al., 2011) and to coordinate with Tocopherol cyclase (VTE1) in the redox cycle of tocopherol (Eugeni Piller, 2014). Recently, NDC1 was shown to catalyze the reduction of demethynaphthoquinone after MenA-mediated transfer of the phytyl moiety to the naphthoquinone (Eugeni Piller et al., 2012; Fatihi et al., 2015). This reduction step is necessary before methylation of demethynaphthoquinone by MenG.

Cyanobacteria *Synechocystis* sp. PCC 6803, mutants defective in PhQ biosynthesis grow normally in sub-optimal light because plastoquinone can bind to the A1 site of PSI and functionally compensate for the phylloquinone shortfall under those conditions (Semenov et al., 2000; Johnson et al., 2001; Lefebvre-Legendre et al., 2007). However, the mutants have compromised photosynthesis and growth under more intense light. The *ndc1* and *menG* mutants of Arabidopsis are also viable under low light conditions, because demethylphylloquinone (precursor of NDC1) can partially fulfill PhQ function in the PSI electron transport chain. However, under high light, the stability of PSI in *ndc1* and *menG* mutants is compromised as demethylphylloquinone cannot fully substitute the function of PhQ in PSI (Lohmann et al., 2006; Fatihi et al., 2015). Most *Arabidopsis* mutants deficient in PhQ are seedling-lethal due to severely impaired PSI assembly (Shimada et al., 2005; Gross et al., 2006; Garcion et al., 2008; Kim et al., 2008). *Arabidopsis ics1 ics2* (MenF) double mutants only survive on medium that contains sucrose and still show a severe reduction in growth compared to the wild type and single mutants (Garcion et al., 2008). The *phyllo* mutant (*pha*) is completely devoid of phylloquinone and has a 75%-95% decrease of PSI activity, along with a moderate,

25% reduction of PSII activity compared to the wild type (Gross et al., 2006). The *menA* mutant (*abc4*) cannot grow photoautotrophically, because it lacks functional PSI and exhibits large decreases in plastoquinone and PSII activity (Shimada et al., 2005). A common phenotype observed across those mutants is decreased PSI stability due to PhQ deficiency.

It has been reported that 60% of PhQ is detected in thylakoids but a smaller portion (30%) is associated with plastoglobules (Lohmann et al., 2006), lipoprotein bodies attached to thylakoid membranes in plastids (Austin et al., 2006). In mutants defective at the MenG step, 70% of the unmethylated PhQ can accumulate in plastoglobules (Lohmann et al., 2006). Biochemically, PhQ can either be reduced to semi-quinone via one electron reduction, or fully reduced to the quinol by two-electron reduction. In plants, PhQ is more generally reduced to the semi-quinone but not to the quinol (Oostende et al., 2008). Fully reduced quinol is mainly observed in animals as a cofactor for carboxylation of glutamate residues in blood clotting proteins (Furie et al., 1999). Nevertheless, the quinol form of PhQ has been detected in multiple dicots, monocots and cyanobacteria (Oostende et al., 2008; Widhalm et al., 2009). The quinol form constitutes 5-10% of the total PhQ pool in developing and mature leaves, but can increase to 25-35% in senescing leaves and dark-grown leaves (Oostende et al., 2008). The abundance of the quinol form under dark conditions suggests that phylloquinol is not generated during photosynthesis. The quinol is speculated to be involved in other redox activities distinct from photosynthetic electron transfer, though the exact roles remain unknown. At the same time it is possible that a small amount of the quinol form is formed normally as a product of PhQ biosynthesis, for example, via methylation of the NDC1 product demethylphylloquinol (Fatihi et al., 2015).

Evolution of the PhQ Biosynthetic Pathway

The Men genes of the PhQ biosynthetic pathway have a fascinating evolutionary history. Plant plastids originated as an outcome of endosymbiotic relationships between cyanobacteria and other bacteria that existed 1.5 billion years ago (Hedges et al., 2001; Yoon et al., 2004;

Reyes-Prieto et al., 2007). Expression control of an estimated 500-1500 cyanobacterial genes was transferred to the nucleus of the symbiotic host, while most of the protein products were targeted to the plastid (Sato et al., 2005; Reyes-Prieto et al., 2006; Moustafa and Bhattacharya, 2008). *Men* genes are nucleus-encoded in plants, green algae and diatom genomes, while in red algae Cyanidiales they reside in the plastid genome. Phylogenetic analyses revealed a complex evolutionary history of *Men* genes in photosynthetic eukaryotes. While *MenA*, *NDC1* and *MenG* genes in plants and green algae descended from cyanobacteria, phylogenetic reconstruction supported a Chlorobi and Gammaproteobacteria origin for *PHYLLO* (*MenD*, *MenC* and *MenH*) and *MenB* genes, respectively, and a Deltaproteobacteria ancestor for *MenE* genes (Michalecka et al., 2003; Gross et al., 2008). The *DHNAT* genes are derived from yet another bacteria: *Lactobacillales* (Widhalm et al., 2012). These observations indicate multiple rounds of horizontal gene transfer, involving different donor taxa, in the history of the progenitor plastid (Gross et al., 2008; Widhalm et al., 2012).

Men genes tend to cluster together as an operon in prokaryotic genomes for coordinated expression between different steps of the PhQ/MK biosynthetic pathway. Following endosymbiosis, *MenF*, *MenD*, *MenC* and *MenH* were fused into a composite gene named *PHYLLO* in diatoms, green algae, and plants (Gross et al., 2006). The encoded protein is composed of multiple domains which correspond to the ancestral gene products, possibly for efficient channeling of PhQ biosynthetic flux. In higher plants, the *MenF* module of *PHYLLO* is truncated and non-functional (Gross et al., 2006). A *MenF* duplication event earlier in higher plant evolution gave rise to an independent gene, known as isochorismate synthase (*ICS*) which also functions in salicylic acid (SA) biosynthesis (Wildermuth et al., 2001; Gross et al., 2006).

Compartmentalization of PhQ Biosynthetic Pathway

The *Men* genes of red algae are encoded in the plastid genomes and the proteins are localized in the plastid. During land plant evolution, while the early (ICS and PHYLLO) and late

(NDC1, MenA and MenG) steps remained in the plastid, the intermediate steps MenE, MenB and DHNAT relocated to the peroxisome after acquiring PTS targeting sequences. Consistent with this model, an intermediate, cytosolic phase of MenB is observed in *Chlamydomonas* and *Physcomitrella* (Babujee et al., 2010). With metabolic exchanges between plastids and peroxisomes, plants established compartmentalization of PhQ biosynthesis between the two organelles. The plastidic Men proteins contain an N-terminal transit peptide (Wildermuth et al., 2001; Shimada et al., 2005; Gross, 2006; Lohmann et al., 2006), which directs the premature proteins to the plastids. The transit peptide is cleaved upon arrival at the plastid and gives rise to mature proteins. This plastid localization has been validated by fusing a fluorescent protein to the C- terminus of the target protein (Wildermuth et al., 2001; Shimada et al., 2005; Gross et al., 2006; Lohmann et al., 2006). MenE may have dual targeting to both plastids and peroxisomes. When fused with a C-terminal GFP, MenE is observed in plastids (Kim et al., 2008). As the C-terminus of MenE harbors SSL>, a conserved Peroxisome Targeting Signal 1 (PTS1) (Reumann et al., 2007), peroxisomal targeting may have been abolished by the C-terminal tagging. Peroxisomal targeting of MenE is observed when it carries an N-terminal tag (Babujee et al., 2010). In fact, the peroxisome is likely to be the primary location for MenE since MenE has not been identified in any proteomic study of plastids (Babujee et al., 2010). MenB has a conserved PTS2 sequence at its N-terminus and is exclusively targeted to peroxisomes (Babujee et al., 2010). The subsequent step catalyzed by DHNAT1 also takes place in peroxisomes (Reumann et al., 2009), before the pathway is channeled back to plastids. The recently uncovered NDC1 is dually targeted to plastids and mitochondria (Xu et al., 2013).

Men proteins in plastids may form a multienzyme complex or metabolon to facilitate PhQ biosynthesis. The PHYLLO protein contains multiple domains, each corresponding to a distinct eubacterial Men protein and catalyzing a different reaction. In *pha* mutants, enzymatic conversion by MenA and MenG of fed naphthoate is strongly impaired, suggesting that without PHYLLO, the stability of the macromolecular enzyme complex is compromised (Gross et al.,

2006). This also suggests that MenA and/or MenG contribute to the metabolon as well. In support of a plastid metabolon, fluorescence signal localization of PHYLLO, MenE, NDC1 and MenG showed a punctate pattern inside the plastids (Gross et al., 2006; Lohmann et al., 2006; Kim et al., 2008; Eugeni Piller et al., 2011), consistent with the distributional of plastoglobules (Austin et al., 2006). However, it's unclear why the intermediate steps occur in the peroxisomes and how intermediates are shuttled.

Photosynthesis in Developing Seeds

Not surprisingly, green leafy vegetables are PhQ-rich (122－440 μg/100 g) (Booth and Suttie, 1998). Interestingly, PhQ levels can be fairly high in soybean seeds (47 μg/100 g) compared to most non-leafy foods (<10 μg/100 g). This is surprising because developing seeds/embryos are predominantly heterotrophic. Given the limited transmission of light through the protective tissues that surround developing seeds, the photosynthetic capacity of developing seeds is thought to be much lower than that of green tissues such as leaves or pods (Harvey et al., 1976; Atkins and Flinn, 1978; Saito et al., 1989; Wullschleger and Oosterhuis, 1990; Eastmond et al., 1996; Asokanthan et al., 1997). However, low levels of photosynthesis in chlorophyllous seeds and embryos of many angiosperms provide O_2 for respiration (Rolletschek et al., 2003), and generate ATP and NADHP to recycle respiratory CO_2 (Wullschleger and Oosterhuis, 1990) and support lipid biosynthesis (Asokanthan et al., 1997). Photosynthesis in developing seeds exhibits saturation at low light consistent with low chlorophyll *a/b* ratios (Eastmond et al., 1996).

Photosyntheic capacity peaks at different stages during seed development in species-specific fashion. In *Brassica napus*, developing seed photosynthesis activity is positively correlated with chlorophyll content and continues to increase during storage reserve accumulation until the onset of desiccation (Eastmond et al., 1996; Asokanthan et al., 1997). Such photosynthetic activity involves both PSI and PSII, and the PSI/PSII ratio decreases

during seed development (Asokanthan et al., 1997). In *Arabidopsis thaliana*, a PSII light-harvesting complex gene (LHCII) and a PSI gene exhibited their highest expression at the onset of reserve accumulation, with declines thereafter as seeds matured (Ruuska et al., 2002; Fait et al., 2006). In contrast, Broad bean and pea embryo photosynthesis activity is weak in the early stages, but increases during embryo differentiation, in concordance with chlorophyll content (Rolletschek et al., 2003).

In chloroplasts, the thylakoid membrane system can be divided into appressed grana and stroma lamellae interconnecting the grana. PSII is rich in the granal regions whereas PSI is primarily located in the intergranal stroma lamellae (Anderson, 1981; Anderson and Melis, 1983; Danielsson et al., 2004). Chloroplasts in developing embryos are characterized by increased grana stacking and poorly developed or lacking stroma lamellae (Fisc et al., 1988; Saito et al., 1989; Asokanthan et al., 1997), similar to the phenotype of mutant *menA* (Shimada et al., 2005). Plastoglobules, persist in mature soybean seeds as plastids senesce and lose their internal membrane structure, PSI, and chlorophyll (Saito et al., 1989). This might explain the unusually high PhQ content observed in soybean seeds (Booth and Suttie, 1998), and hint at a non-photosynthetic role of PhQ in soybean seeds.

Potential Involvement of PhQ in Plasma Membrane Electron Transport

Since a substantial fraction of PhQ is not associated with PSI, other PhQ functions have been speculated (Gross et al., 2006). Multiple lines of evidence suggest a close association between PhQ and the plasma membrane. For example, PhQ has been directly detected in the plasma membrane of maize (*Zea mays* L.) roots (Lüthje and Böttger, 1995). Electron transport across the plasma membrane has been repeatedly observed by applying membrane impermeable artificial electron acceptors, e.g. hexacyanoferrate III (HCF III), to intact plant roots (Döring et al., 1990; Döring et al., 1992; Lüthje et al., 1992). The electron acceptors were reduced concomitant with plasma membrane depolarization and medium acidification (Döring et

al., 1990). Destruction of quinones in cultured carrot cells by ultraviolet radiation blocked trans-plasma membrane electron transport, and significantly decreased reduction of external artificial electron acceptors (Barr et al., 1992). Subsequent addition of PhQ restored the transmembrane electron flux. Vitamin K antagonists, e.g. dicumarol and warfarin, inhibited the transmembrane redox flow and proton secretion, whereas Vitamin K_3 and PhQ applications rescued the inhibition and stimulated the reduction of the external electron acceptors (Döring et al., 1992; Lüthje et al., 1992; Lüthje et al., 1994; Lüthje and Böttger, 1995; Döring and Lüthje, 2001).

The involvement of PhQ in plasma membrane electron transport also gained support by its potential involvement as a cofactor in other plasma membrane redox activities. Naphthoquinone-dependent NADH dehydrogenase activities have been characterized in plasma membranes of onion roots (Serrano et al., 1994), maize roots (Lüthje et al., 1998) and soybean hypocotyls (Schopfer et al., 2008). In maize roots, the localization of the protein at the cytoplasmic surface of the lipid bilayer, together with its naphthoquinone-reducing activity (Lüthje et al., 1998), lent support to its role upstream of an electron transport chain. In soybean hypocotyls, the naphthoquinone-dependent NADH dehydrogenase has been demonstrated to generate superoxide radicals in the presence of menadione or 1,4-naphthoquinones through a single electron transfer (Schopfer et al., 2008). An NADH oxidase isolated from the plasma membrane of soybean hypocotyl possessed PhQ hydroquinone oxidase activity, which would enable it to function downstream of the electron flow (Bridge et al., 2000). This NADH oxidase can be stimulated by growth factors (Morré et al., 1986; Brightman et al., 1988) and was found to reside on the cell surface (DeHahn et al., 1997). The findings are consistent with the idea that the NADH oxidase functions at the cell surface as the terminal step of the plasma membrane electron flow from cytosolic donors to apoplastic acceptors (Bridge et al., 2000). Alternatively, a b-type cytochrome with a membrane-spanning structure has been suggested as the terminal step of the plasma-membrane-bound electron transfer in maize roots (Döring and Lüthje, 1996; Lüthje et al., 1998; Lochner et al., 2003; Lüthje et al., 2005). In conjunction with the NADH

oxidoreductases, PhQ can transfer electrons across the plasma membrane. A model of the putative plasma membrane redox system involving PhQ is proposed in which electrons are transferred from cytosolic donors (e.g. NADPH) to apoplastic acceptors (Lochner et al., 2003).

Other Non-photosynthetic Functions of PhQ

PhQ and ubiquinone occur in photosynthetic and respiratory electron transport of higher eukaryotes, respectively, whereas MK only occurs in bacteria. However, their functions can overlap or even replace each other depending on the species. Therefore, comparing the functions of PhQ and its counterpart MKs in different species may shed light on the functional evolution of PhQ in plants.

In the photosynthetic reaction center of a green sulfur bacteria *Chlorobium vibrioforme*, MK functions in a site which closely resembles the structure and function of the A_1 site in PSI (Kjær et al., 1998), supporting the functional analogy between MK and PhQ. MK is the sole isoprenoid quinone in certain photosynthetic bacteria (Frydman and Rapoport, 1963; Hale et al., 1983) and a special type of gram-positive bacteria *Heliobacterium chlorum* (Hiraishi, 1989), fulfilling the electron transfer role in both photosynthesis and respiration. Similarly, some purple bacteria synthesize ubiquinone exclusively, with photosynthesis-related functions (Hiraishi et al., 1984; Imhoff, 1984). The alternative functions of MK and ubiquinone may provide hints about non-canonical functions of PhQ in plants. Evidence for a direct role of menaquinone-8 (MK-8) and ubiquinone-8 (Q8) in bacterial signal transduction came from an investigation in the Arc Two-Component (ABC) system where oxidized MK-8 and Q8 inhibit the autophosphorylation of an ArcB transmembrane sensor kinase to regulate downstream gene expression (Georgellis et al., 2001). PhQ has been shown to modulate signal transduction in animal systems through a protein-tyrosine phosphorylation cascade (Saxena et al., 1997; Ni et al., 1998). Within the transmembrane electron chain, PhQ can play a signaling role by linking internal systems to

external redox states at the cell surface, which may be similar to aging and senescence mechanisms (Linnane et al., 1992; Takahashi et al., 1995; Lenaz et al., 1997).

Analogous to Coenzyme Q_{10} (CoQ_{10}) in animal plasma membranes (Perry and Harwood 1993), PhQ may protect plant plasma membranes during oxidative stress (Perry et al. 1999). PhQ or intermediates of the PhQ biosynthetic pathway were suggested to be involved in programmed cell death during plant defense against pathogens (Brodersen et al., 2005). The *menA* mutants grown on medium supplemented with sucrose are able to reach the flowering stage but fail to produce mature seeds, revealing the possibility that PhQ might have a role in seed development (Shimada et al., 2005). PhQ may also serve as a co-factor for the formation of protein disulfide bonds in the chloroplasts (Singh et al., 2008; Furt et al., 2010; Karamoko et al., 2011). It has also been established that PhQ promotes radish enlargement and pea stem elongation in the presence of auxin (Hemberg, 1953; Stowe and Obreiter, 1962). The growth of cultured carrot cells in the absence of PhQ (by UV-B treatment) was blocked and can be restored by supplementing external PhQ (Barr et al., 1992). Such growth inhibition and recovery was also observed in algae cells in the presence of naphthoquinone derivatives and by addition of PhQ, respectively (Gaffron, 1945). PhQ was proposed to perform dual functions as prooxidant and antioxidant in the non-photosynthetic redox system (Perry et al. 1999; Taylor et al. 2009).

Non-photosynthetic Holoparasite as a Study System

Investigations into PhQ functions not associated with photosynthesis have been challenging. First, PhQ defective mutants are seedling-lethal because of the impaired PSI activity (Shimada et al., 2005; Gross et al., 2006; Garcion et al., 2008; Kim et al., 2008), making it difficult to uncover phenotypes associated with the non-photosynthetic functions. Second, PhQ is predominantly detected in chloroplasts. Studies speculating on non-canonical functions of PhQ in plants were often based on indirect evidence, with potential contamination from

chloroplasts. Exploring alternative study systems is necessary in order to advance our knowledge on non-photosynthetic functions of PhQ.

Parasitic plants are a group of specialized plants with varying degrees of photosynthetic capacity that partially or completely depend on their hosts for nutrition. Species that attack agriculturally important crops can cause enormous damages and yield losses. For example, the economic impacts of *Striga*, a parasite widespread in sub-Saharan Africa, exceeds $3 billion annually (Parker, 2009). Seeds of parasitic plants can lay dormant underground for years before they sense the presence of the hosts and germinate. The parasites are therefore difficult to eliminate, and once a field is infected by parasitic plants, farmers will usually choose to abandon the whole field to prevent further infection. In these cases, actual losses are immeasurable (Abu Irmaileh et al., 2008).

Parasitism has evolved independently more than 12 times from photoautotrophic plants (Barkman et al., 2007; Westwood et al., 2010), giving rise to multiple parasitic plant families with diverse morphology (Yoshida et al., 2016). Based on their photosynthesis capacity, parasitic plants are grouped into hemiparasites which retain full or partial photosynthesis, and holoparasites which have completely lost photosynthetic ability. By the degree of host dependence, parasitic plants can be classified as facultative parasites and obligate parasites. Facultative parasites can complete their life cycles independently but will opportunistically parasitize the host when available. Obligate parasites require the presence of the host to germinate and develop into mature plants. Depending on the site where parasitic plants attack their hosts, parasitic plants can be classified as root parasites or stem parasites.

Among all parasite-containing families, only Orobanchaceae spans the full range of parasitic dependency and photosynthetic capability. Orobanchaceae is the second largest plant family and contains around 1800 species (Westwood et al., 2010). The Parasitic Plant Genome Project (PPGP) sequenced the transcriptomes of three representative species of different nutritional types from Orobanchaceae: *Triphysaria versicolor* is a facultative hemiparasite, *Striga*

hermonthica is an obligate hemiparasite, and *Phelipanche aegyptiaca* is an obligate holoparasite (Westwood et al., 2012). All three species are outcrossing with genome sizes ranging from 1.7 Gb (*S. hermonthica*) to 3.9 Gb (*P. aegyptiaca*) (Westwood et al., 2010). In accordance with photosynthetic capacity, nuclear-encoded photosynthetic genes, including PSI, PSII and Light Harvesting Complex (LHC), showed considerably lower expression in *S. hermonthica* compared with *T. versicolor*, and no corresponding transcripts were detected in *P. aegyptiaca* (Wickett et al., 2011). In addition, most of the plastid-encoded genes for photosynthesis have been completely lost or become pseudogenes in holoparasites (dePamphilis and Palmer, 1990; Wolfe et al., 1992; Wickett et al., 2008; Delannoy et al., 2011).

Despite multiple independent origins and distinct morphology and physiology, parasitic plants share a common organ, known as the haustorium, to invade their host for nutrient acquisition (Yoshida et al., 2016). Initiation of haustorial development requires host signals to be sensed by the parasitic plants. One such signaling molecule is 2,6 dimethyoxy-1,4-benzoquinone (DMBQ) which is derived from host cell wall phenolics (Kim et al., 1998). In addition to DMBQ, other haustorium inducing factors (HIFs) also exist in the host root exudates (Albrecht et al., 1999). It is hypothesized that parasitic genes involved in haustorial signaling, development, and penetration have been recruited from genes and biochemical pathways in root and floral tissues (Yang et al., 2015).

Several genes have been well established to participate in haustorial development. Two quinone oxidoreductases in *T. versicolor* (TvQR1 and TvQR2) are upregulated at root tips following exposure to HIFs (Matvienko et al., 2001a; Matvienko et al., 2001b). Both enzymes catalyze the reduction of quinones, including DMBQ, via one-electron (TvQR1) or two-electron (TvQR2) reactions (Sparla et al., 1996; Wrobel et al., 2002). Interestingly, transgenic plants with silenced expression of *TvQR1*, specifically, hosted a reduced number of haustoria (Bandaranayake et al., 2010). At the same time, only the orthologs of *TvQR2* are upregulated in response to HIFs in facultative parasite *Phtheirospermum japonicum* and obligate hemiparasite

Striga asiatica, indicating that the haustorial signaling pathway may vary across parasitic species (Ishida et al., 2016; Liang et al., 2016). Conversion of host cell wall phenolics to DMBQ requires parasite-generated peroxidases and H_2O_2. Two peroxidases from *Striga asiatica* (SaPOXA and SaPOXB) and two peroxidases from *Phelipanche ramosa* have been characterized for their involvement during haustorial development (Kim et al., 1998; González-Verdejo et al., 2006; Veronesi et al., 2007). An NADPH oxidase in *S. asiatica* (*SaNOX1*), belonging to a respiratory burst oxidase homolog (Rboh) family, was found to participate in ROS generation at root tips in response to DMBQ (Liang et al., 2016). Another HIF-induced gene is *TvPirin* which encodes a transcriptional factor that positively regulates haustorium-related genes (Bandaranayake et al., 2012).

Objectives and Overview of Dissertation Chapters

Despite the efforts in exploring alternative functions of PhQ in plants, investigations have been challenging due to the persistence of photosynthetic functions, even in heterotrophic tissues, of photoautotrophic plants. The possible existence of PhQ and its biosynthetic pathway in non-photosynthetic holoparasitic plants has not been explored. My dissertation research aimed to elucidate the non-photosynthetic functions of PhQ using the parasitic plants as a study system, and to understand the evolution of PhQ biosynthesis pathway in both parasitic and photoautotrophic plants. Although transcriptome resources of parasitic plants are available from PPGP, fragmented assembly prevented recovery of PhQ biosynthetic pathway gene transcripts for the research. Therefore, Chapter 2 was devoted to the development of an innovative pipeline for *de novo* transcriptome assembly in non-model species that lack a sequenced genome. RNA-Seq data sets of pollen from various flowering trees were used to assess the quality of assemblies compared to the results of other *de novo* assembly methods. This pipeline leverages the advantages of both reference-based and *de novo* assembly algorithms, by using proteome information from a closely related species as the reference for local *de novo* assembly.

This pipeline dramatically increases computing efficiency by organizing the input RNA-Seq reads into independent bins based on gene families for parallel assembly. Finally, this pipeline adopts iterative computing to improve the accuracy by using assembled sequences from the previous run as the reference to repeat the assembly. The results demonstrated improved performance compared to Trinity and the CLC assembly pipeline based on TransRate evaluation. The pipeline enabled reconstruction of full-length transcripts for the entire suite of PhQ biosynthesis genes from parasitic plants, critical for the research presented in Chapter 3.

In Chapter 3, the research goals were to investigate the occurrence, expression and function of the PhQ biosynthetic pathway in parasitic plants, and to understand the evolution of this pathway in angiosperms. Using the improved local assembly pipeline developed in Chapter 2, I successfully recovered full-length transcripts for all PhQ genes to support a functional PhQ biosynthesis pathway in the parasitic plants. The analysis revealed that the last two enzymatic steps of the PhQ biosynthesis pathway have been relocated from chloroplasts to plasma membranes in the holoparasite. The bioinformatics findings were validated, through collaboration, by subcellular localization experiments and by detection of PhQ in the holoparasite. Gene co-expression network analysis suggested a role of PhQ in the plasma membrane redox activities associated with the signaling of parasitic haustorium development. The plasma membrane localization of the terminal PhQ biosynthesis steps was found to be conserved in photoautotrophic species via alternative splicing, suggesting plasma membrane PhQ biosynthesis is evolutionarily conserved. This work provides the first molecular evidence for plasma membrane PhQ biosynthesis in plants.

In Chapter 4, the research goals were to explore the expression patterns of PhQ biosynthetic genes in *Arabidopsis thaliana*, *Glycine max* and *Populus tremula x alba* for non-photosynthetic functions of PhQ, and for evidence of functional diversification in photoautotrophic species. Discerning the non-photosynthetic function of PhQ in green plants proved to be difficult even with the use of heterotrophic tissues, as photosynthesis-related

activities remained as the dominant functions of PhQ. However, I found evidence of functional divergence among recently duplicated *ICS* and *DHNAT* genes in *Arabidopsis*. Similar divergence was also observed for *DHNAT* in *Glycine* and *Populus*.

Significance of This Work

This study advances our understanding of the dual function of PhQ in plants. PhQ has long been speculated to exhibit non-photosynthetic functions, but experimental support has been scarce due to the masking effect of its primary function in photosynthesis. This study is the first to use photosynthesis-free holoparasites as a study system to explore the evolution and the alternative functions of PhQ biosynthesis. The work established unequivocally that the plastidial PhQ biosynthesis in photoautotrophic species has been exploited by the holoparasites and redirected to the plasma membrane for redox regulation associated with haustorium development. Plasma membrane PhQ biosynthesis appeared to be conserved in photoautotrophic species, suggesting an ancient origin of dually localized PhQ biosynthesis in angiosperms. Given the conservation of plasma membrane PhQ biosynthesis, the results from parasitic plants shed lights on the non-photosynthetic roles of PhQ in photoautotrophic plants. Importantly, knowledge from this work on the signaling mechanisms of parasitic haustorial development may lead to potential targets for controlling parasitic plants that cause devastating losses to agriculture. The computational pipeline developed in this work improved upon existing methods for high-quality *de novo* transcriptome assembly. It should be valuable to the broad communities working on non-model species with limited genomics resources. Although investigation of non-photosynthetic function of PhQ in photoautotrophic species remained challenging, this study has already revealed some unexpected findings on the plasticity of the PhQ biosynthetic pathway. The unusual expression patterns of *ICS2* in response to osmotic stresses and of *DHNAT* in plant roots are examples that warrant future research.

References

Abu Irmaileh BE, FAO R (Italy). PP and PD, Eng, Labrada R, (ed.), FAO R (Italy). D de la PV et de la P des P, Fre (2008) Integrated Orobanche management. Prog. farmer Train. Parasit. weed Manag. Rome (Italy) FAO, pp 17–29

Albrecht H, Yoder JI, Phillips DA (1999) Flavonoids Promote Haustoria Formation in the Root Parasite *Triphysaria versicolor* 1. Plant Physiol **119**: 585–592

Anderson JM (1981) Consequences of spatial separation of photosystem I and II in thylakoid membranes of higher plant chloroplasts. FEBS Lett. 124:

Anderson JM, Melis A (1983) Localization of different photosystems in separate regions of chloroplast membranes. Proc Natl Acad Sci U S A **80**: 745–749

Asokanthan PS, Johnson RW, Griffith M, Krol M (1997) The photosynthetic potential of canola embryos. Physiol Plant **101**: 353–360

Atkins CA, Flinn AM (1978) Carbon dioxide fixation in the carbon economy of developing seeds of *Lupinus albus* (L.). **62**: 486–490

Austin JR, Frost E, Vidi P-A, Kessler F, Staehelin LA (2006) Plastoglobules Are Lipoprotein Subcompartments of the Chloroplast That Are Permanently Coupled to Thylakoid Membranes and Contain Biosynthetic Enzymes. Plant Cell Online **18**: 1693–1703

Babujee L, Wurtz V, Ma C, Lueder F, Soni P, Van Dorsselaer A, Reumann S (2010) The proteome map of spinach leaf peroxisomes indicates partial compartmentalization of phylloquinone (vitamin K_1) biosynthesis in plant peroxisomes. J Exp Bot **61**: 1441–1453

Bandaranayake PCG, Filappova T, Tomilov A, Tomilova NB, Jamison-McClung D, Ngo Q, Inoue K, Yoder JI (2010) A single-electron reducing quinone oxidoreductase is necessary to induce haustorium development in the root parasitic plant *Triphysaria*. Plant Cell **22**: 1404–19

Bandaranayake PCG, Tomilov A, Tomilova NB, Ngo QA, Wickett N, DePamphilis CW, Yoder JI (2012) The TvPirin Gene Is Necessary for Haustorium Development in the

Parasitic Plant *Triphysaria versicolor*. Plant Physiol **158**: 1046–1053

Barkman TJ, McNeal JR, Lim S-H, Coat G, Croom HB, Young ND, DePamphilis CW (2007) Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. BMC Evol Biol **7**: 248

Barr R, Pan RS, Crane FL, Brightman AO, Morré DJ (1992) Destruction of vitamin K_1 of cultured carrot cells by ultraviolet radiation and its effect on plasma membrane electron transport reactions. Biochem Int **27**: 449–56

Booth SL, Suttie JW (1998) Dietary intake and adequacy of vitamin K1. J Nutr **128**: 785–788

Boudreaux B, MacMillan F, Teutloff C, Agalarov R, Gu F, Grimaldi S, Bittl R, Brettel K, Redding K (2001) Mutations in Both Sides of the Photosystem I Reaction Center Identify the Phylloquinone Observed by Electron Paramagnetic Resonance Spectroscopy. J Biol Chem **276**: 37299–37306

Brettel K, Setif P, Mathis P (1986) Flash-induced absorption changes in photosystem I at low temperature: Evidence that the electron acceptor A_1 is vitamin K_1. FEBS Lett **203**: 220–224

Bridge A, Barr R, Morré DJ (2000) The plasma membrane NADH oxidase of soybean has vitamin K_1 hydroquinone oxidase activity. Biochim Biophys Acta - Biomembr **1463**: 448–458

Brightman AO, Barr R, Crane FL, Morré DJ (1988) Auxin-Stimulated NADH Oxidase Purified from Plasma Membrane of Soybean. Plant Physiol **86**: 1264–9

Brodersen P, Malinovsky FG, Hématy K, Newman M-A, Mundy J (2005) The role of salicylic acid in the induction of cell death in Arabidopsis *acd11*. Plant Physiol **138**: 1037–45

Danielsson R, Albertsson P-Å, Mamedov F, Styring S (2004) Quantification of photosystem I and II in different parts of the thylakoid membrane from spinach. Biochim Biophys Acta - Bioenerg **1608**: 53–61

DeHahn T, Barr R, Morré DJ (1997) NADH oxidase activity present on both the external and internal surfaces of soybean plasma membranes. Biochim Biophys Acta - Biomembr **1328**:

99–108

Delannoy E, Fujii S, Brundrett M, Small I (2011) Rampant Gene Loss in the Underground Orchid *Rhizanthella gardneri* Highlights Evolutionary Constraints on Plastid Genomes. Mol Biol Evol **28**: 2077–2086

dePamphilis CW, Palmer JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature **348**: 337–339

Döring O, Lüthje S (2001) Inhibition of trans-membrane hexacyanoferrate III reductase activity and proton secretion of maize (*Zea mays* L.) roots by thenoyltrifluoroacetone. Protoplasma **217**: 3–8

Döring O, Lüthje S (1996) Molecular components and biochemistry of electron transport in plant plasma membranes (Review). Mol Membr Biol **13**: 127–142

Döring O, Lüthje S, Böttger M (1992) Modification of the activity of the plasma membrane redox system of Zea mays L. roots by vitamin K_3 and dicumarol. J Exp Bot **43**: 175–181

Döring O, Lüthje S, Hilgendorf F, Böttger M (1990) Membrane Depolarization by Hexacyanoferrate (III), Hexabromoiridate (IV) and Hexachloroiridate (IV). J Exp Bot **41**: 1055–1061

Eastmond P, Koláčá L, Rawsthorne S (1996) Photosynthesis by developing embryos of oilseed rape (*Brassica napus* L.). J Exp Bot **47**: 1763–1769

Eugeni Piller L (2014) Role of plastoglobules in metabolite repair in the tocopherol redox cycle. Front Plant Sci **5**: 1–10

Eugeni Piller L, Abraham M, Dormann P, Kessler F, Besagni C (2012) Plastid lipid droplets at the crossroads of prenylquinone metabolism. J Exp Bot **63**: 1609–1618

Eugeni Piller L, Besagni C, Ksas B, Rumeau D, Bréhélin C, Glauser G, Kessler F, Havaux M (2011) Chloroplast lipid droplet type II NAD(P)H quinone oxidoreductase is essential for prenylquinone metabolism and vitamin K_1 accumulation. Proc Natl Acad Sci U S A **108**: 14354–9

Fait A, Angelovici R, Less H, Ohad I, Urbanczyk-Wochniak E, Fernie AR, Galili G (2006) Arabidopsis seed development and germination is associated with temporally distinct metabolic switches. Plant Physiol **142**: 839–54

Fatihi A, Latimer S, Schmollinger S, Block A, Dussault PH, Vermaas WFJ, Merchant SS, Basset GJ (2015) A dedicated type II NADPH dehydrogenase performs the penultimate step in the biosynthesis of vitamin K1 in *Synechocystis* and Arabidopsis. Plant Cell **27**: 1730–41

Fisc W, Bergf R, Pla C, Schäfer R, Schopfer P (1988) Accumulation of Storage Materials, Precocious Germination and Development of Desiccation Tolerance During Seed Maturation in Mustard (*Sinapis alba* L.). Bot Acta **101**: 344–354

Frydman B, Rapoport H (1963) Non-Chlorophyllous Pigments of *Chlorobium Thiosulfatophilum* Chlorobiumquinone. J Am Chem Soc **85**: 823–825

Furie B, Bouchard BA, Furie BC (1999) Vitamin K-Dependent Biosynthesis of γ-Carboxyglutamic Acid. Blood 93:

Furt F, Oostende C van, Widhalm JR, Dale MA, Wertz J, Basset GJC (2010) A bimodular oxidoreductase mediates the specific reduction of phylloquinone (vitamin K1) in chloroplasts. Plant J **64**: 38–46

Gaffron H (1945) Some effects of derivatives of vitamin K on the metabolism of unicellular algae. J Gen Physiol **28**: 259–268

Garcion C, Lohmann A, Lamodière E, Catinot J, Buchala A, Doermann P, Métraux J-P (2008) Characterization and biological function of the *ISOCHORISMATE SYNTHASE2* gene of Arabidopsis. Plant Physiol **147**: 1279–1287

Georgellis D, Kwon O, Lin EC, Parkinson JS, Kofoid EC, Iuchi S, Lin ECC, Kwon O, Georgellis D, Lynch AS, et al (2001) Quinones as the redox signal for the arc two-component system of bacteria. Science **292**: 2314–6

González-Verdejo CI, Barandiaran X, Moreno MT, Cubero JI, Di Pietro A (2006) A

peroxidase gene expressed during early developmental stages of the parasitic plant *Orobanche ramosa*. J Exp Bot **57**: 185–92

Gross J (2006) The biosynthesis of phylloquinone (vitamin K_1) in higher plants. 98

Gross J, Cho WK, Lezhneva L, Falk J, Krupinska K, Shinozaki K, Seki M, Herrmann RG, Meurer J (2006) A plant locus essential for phylloquinone (vitamin K_1) biosynthesis originated from a fusion of four eubacterial genes. J Biol Chem **281**: 17189–96

Gross J, Meurer J, Bhattacharya D (2008) Evidence of a chimeric genome in the cyanobacterial ancestor of plastids. BMC Evol Biol **8**: 117

Hale MB, Blankenship RE, Fuller RC (1983) Menaquinone is the sole quinone in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. BBA - Bioenerg **723**: 376–382

Harvey DM, Hedley CL, Keely R (1976) Photosynthetic and respiratory studies during pod and seed development in *Pisum sativum* L. Ann Bot **40**: 993–1001

Hedges SB, Chen H, Kumar S, Wang DY, Thompson AS, Watanabe H, Steinberger R, Eriksson A-S, Winkler H, Kurland C (2001) A genomic timescale for the origin of eukaryotes. BMC Evol Biol 2001 11 **24**: 1135–1135

Hemberg T (1953) The Effect of Vitamin K and Vitamin H' on the Root Formation in Cuttings of *Phaseolus vulgaris* L. Physiol Plant **6**: 17–20

Hiraishi A (1989) Occurrence of menaquinone as the sole isoprenoid quinone in the photosynthetic bacterium *Heliobacterium chlorum*. Arch Microbiol **151**: 378–379

Hiraishi A, Hoshino Y, Kitamura H (1984) Isoprenoid quinone composition in the classification of *Rhodospirillaceae*. J Gen Appl Microbiol **30**: 197–210

Imhoff JF (1984) Quinones of phototrophic purple bacteria. FEMS Microbiol Lett **25**: 85–89

Ishida JK, Wakatake T, Yoshida S, Takebayashi Y, Kasahara H, Wafula E, DePamphilis CW, Namba S, Shirasu K (2016) Local auxin biosynthesis mediated by a YUCCA flavin monooxygenase regulates haustorium development in the parasitic plant *Phtheirospermum*

japonicum. Plant Cell **28**: 1795–814

Itoh S, Iwaki M (1989) Vitamin K_1 (phylloquinone) restores the turnover of FeS centers in the ether-extracted spinach PS I particles. FEBS Lett **243**: 47–52

Johnson TW, Zybailov B, Jones AD, Bittl R, Zech S, Stehlik D, Golbeck JH, Chitnis PR (2001) Recruitment of a foreign quinone into the A1 site of photosystem I. *In vivo* replacement of plastoquinone-9 by media-supplemented naphthoquinones in phylloquinone biosynthetic pathway mutants of *Synechocystis* sp. PCC 6803. J Biol Chem **276**: 39512–21

Karamoko M, Cline S, Redding K, Ruiz N, Hamel PP (2011) Lumen Thiol Oxidoreductase1, a disulfide bond-forming catalyst, is required for the assembly of photosystem II in Arabidopsis. Plant Cell **23**: 4462–75

Kim D, Kocz R, Boone L, Keyes WJ, Lynn DG (1998) On becoming a parasite: evaluating the role of wall oxidases in parasitic plant development. Chem Biol **5**: 103–117

Kim HU, van Oostende C, Basset GJC, Browse J (2008) The *AAE14* gene encodes the Arabidopsis o-succinylbenzoyl-CoA ligase that is essential for phylloquinone synthesis and photosystem-I function. Plant J **54**: 272–83

Kjær B, Frigaard NU, Yang F, Zybailov B, Miller M, Golbeck JH, Scheller HV (1998) Menaquinone-7 in the reaction center complex of the green sulfur bacterium *Chlorobium vibrioforme* functions as the electron acceptor A1. Biochemistry **37**: 3237–3242

Lefebvre-Legendre L, Rappaport F, Finazzi G, Ceol M, Grivet C, Hopfgartner G, Rochaix J-D (2007) Loss of phylloquinone in *Chlamydomonas* affects plastoquinone pool size and photosystem II synthesis. J Biol Chem **282**: 13250–63

Lenaz G, Bovina C, Castelluccio C, Fato R, Formiggini G, Genova ML, Marchetti M, Pich MM, Pallotti F, Castelli GP, et al (1997) Mitochondrial Complex I defects in aging. Detect. Mitochondrial Dis. Springer US, Boston, MA, pp 329–333

Liang L, Liu Y, Jariwala J, Lynn DG, Palmer AG (2016) Detection and adaptation in parasitic

angiosperm host selection. Am J Plant Sci **7**: 1275–1290

Linnane AW, Zhang C, Baumer A, Nagley P (1992) Mitochondrial DNA mutation and the ageing process: bioenergy and pharmacological intervention. Mutat Res **275**: 195–208

Lochner K, Döring O, Böttger M (2003) Phylloquinone, what can we learn from plants? BioFactors **18**: 73–78

Lohmann A, Schottler MA, Brehelin C, Kessler F, Bock R, Cahoon EB, Dormann P (2006) Deficiency in phylloquinone (vitamin K1) methylation affects prenyl quinone distribution, photosystem I abundance, and anthocyanin accumulation in the Arabidopsis *AtmenG* mutant. J Biol Chem **281**: 40461–40472

Lohr M, Schwender J, Polle JEW (2011) Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. Plant Sci **185–186**: 9–22

Lüthje S, Böttger M (1995) On the function of a K-type vitamin in plasma membranes of maize (*Zea mays* L.) roots. Mitt Inst Allg Bot Hambg **25**: 5–13

Lüthje S, Böttger M, Döring O (2005) Proton channelling b-type cytochromes in plant plasma membranes? Prog. Bot. Springer-Verlag, Berlin/Heidelberg, pp 187–217

Lüthje S, Döring O, Böttger M (1992) The effects of vitamin K_3 and dicumarol on the plasma membrane redox system and H^+ pumping activity of *Zea mays* L roots measured over a long time scale. J Exp Bot **43**: 183–188

Lüthje S, Gestelen P, Córdoba-Pedregosa MC, González-Reyes J a., Asard H, Villalba JM, Böttger M (1998) Quinones in plant plasma membranes — a missing link? Protoplasma **205**: 43–51

Lüthje S, Gonzaléz-Reyes JA, Navas P, Döring O, Böttger M (1994) Inhibition of Maize (*Zea mays* L.) Root Plasma Membrane-Bound Redox Activities by Coumarins. Zeitschrift für Naturforsch C **49**: 447–452

Matvienko M, Torres MJ, Yoder JI (2001a) Transcriptional responses in the hemiparasitic

plant Triphysaria versicolor to host plant signals. Plant Physiol **127**: 272–282

Matvienko M, Wojtowicz A, Wrobel R, Jamison D, Goldwasser Y, Yoder JI (2001b) Quinone oxidoreductase message levels are differentially regulated in parasitic and non-parasitic plants exposed to allelopathic quinones. Plant J **25**: 375–387

Michalecka AM, Svensson AS, Johansson FI, Agius SC, Johanson U, Brennicke A, Binder S, Rasmusson AG (2003) Arabidopsis genes encoding mitochondrial type II NAD(P)H dehydrogenases have different evolutionary origin and show distinct responses to light. Plant Physiol **133**: 642–52

Morré DJ, Navas P, Penel C, Castillo FJ (1986) Auxin-stimulated NADH oxidase (semidehydroascorbate reductase) of soybean plasma membrane: Role in acidification of cytoplasm? Protoplasma **133**: 195–197

Moustafa A, Bhattacharya D (2008) PhyloSort: a user-friendly phylogenetic sorting tool and its application to estimating the cyanobacterial contribution to the nuclear genome of *Chlamydomonas*. BMC Evol Biol **8**: 6

Ni R, Nishikawa Y, Carr BI (1998) Cell growth inhibition by a novel vitamin K is associated with induction of protein tyrosine phosphorylation. J Biol Chem **273**: 9906–9911

Oostende C van, Widhalm JR, Basset GJC (2008) Detection and quantification of vitamin K_1 quinol in leaf tissues. Phytochemistry **69**: 2457–2462

Van Oostende C, Widhalm JR, Furt F, Ducluzeau A-L, Basset GJ (2011) Vitamin K_1 (Phylloquinone): function, enzymes and genes. Biosynth Vitam Plants Part B Vitam B6, B8, B9, C, E, K **59**: 229

Parker C (2009) Observations on the current status of *Orobanche* and *Striga* problems worldwide. Pest Manag Sci **65**: 453–459

Petersen J, Stehlik D, Gast P, Thurnauer M (1987) Comparison of the electron spin polarized spectrum found in plant photosystem I and in iron-depleted bacterial reaction centers with time-resolved K-band EPR; evidence that the photosystem I acceptor A_1 is a quinone.

Photosynth Res **14**: 15–30

Reumann S, Babujee L, Ma C, Wienkoop S, Siemsen T, Antonicelli GE, Rasche N, Lüder F, Weckwerth W, Jahn O (2007) Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. Plant Cell **19**: 3170–3193

Reumann S, Quan S, Aung K, Yang P, Manandhar-Shrestha K, Holbrook D, Linka N, Switzenberg R, Wilkerson CG, Weber APM, et al (2009) In-depth proteome analysis of Arabidopsis leaf peroxisomes combined with in vivo subcellular targeting verification indicates novel metabolic and regulatory functions of peroxisomes. Plant Physiol **150**: 125–43

Reyes-Prieto A, Hackett JD, Soares MB, Bonaldo MF, Bhattacharya D (2006) Report cyanobacterial contribution to algal nuclear genomes is primarily limited to plastid functions. Curr Biol **16**: 2320–2325

Reyes-Prieto A, Weber APM, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annu Rev Genet **41**: 147–168

Rolletschek H, Weber H, Borisjuk L (2003) Energy status and its control on embryogenesis of legumes. Embryo photosynthesis contributes to oxygen supply and is coupled to biosynthetic fluxes. Plant Physiol **132**: 1196–206

Ruuska SA, Girke T, Benning C, Ohlrogge JB (2002) Contrapuntal networks of gene expression during Arabidopsis seed filling. Plant Cell **14**: 1191–206

Saito GY, Chang YC, Walling LL, Thomson WW (1989) A correlation in plastid development and cytoplasmic ultrastructure with nuclear gene expression during seed ripening in soybean. New Phytol **113**: 459–469

Sato N, Fujiwara M, Ishikawa M, Sonoike K (2005) Mass identification of chloroplast proteins of endosymbiont origin by phylogenetic profiling based on organism-optimized homologous protein groups. Genome Informatics **16**: 56–68

Saxena SP, Fan T, Li M, Israels ED, Israels LG (1997) A novel role for vitamin K_1 in a tyrosine phosphorylation cascade during chick embryogenesis. J Clin Invest **99**: 602–7

Schopfer P, Heyno E, Drepper F, Krieger-Liszkay A (2008) Naphthoquinone-dependent generation of superoxide radicals by quinone reductase isolated from the plasma membrane of soybean. PLANT Physiol **147**: 864–878

Semenov AY, Vassiliev IR, van Der Est A, Mamedov MD, Zybailov B, Shen G, Stehlik D, Diner BA, Chitnis PR, Golbeck JH (2000) Recruitment of a foreign quinone into the A_1 site of photosystem I. Altered kinetics of electron transfer in phylloquinone biosynthetic pathway mutants studied by time-resolved optical, EPR, and electrometric techniques. J Biol Chem **275**: 23429–38

Serrano A, Córdoba F, Conzález-Reyes JA, Navas P, Villalba JM (1994) Purification and characterization of two distinct NAD(P)H dehydrogenases from onion (*Allium cepa*) Root Plasma Membrane. Plant Physiol **106**: 87–96

Shimada H, Ohno R, Shibata M, Ikegami I, Onai K, Ohto M, Takamiya K (2005) Inactivation and deficiency of core proteins of photosystems I and II caused by genetical phylloquinone and plastoquinone deficiency but retained lamellar structure in a T-DNA mutant of Arabidopsis. Plant J **41**: 627–637

Sigfridsson K, Hansson O, Brzezinski P (1995) Electrogenic light reactions in photosystem I: Resolution of electron-transfer rates between the iron-sulfur centers (electrogenic events/photosynthesis/photovoltage/spinach). Biophysics (Oxf) **92**: 3458–3462

Singh AK, Bhattacharyya-Pakrasi M, Pakrasi HB (2008) Identification of an atypical membrane protein involved in the formation of protein disulfide bonds in oxygenic photosynthetic organisms. J Biol Chem **283**: 15762–70

Sparla F, Tedeschi G, Trost P, Biologia D, Bologna U, S F, Fisiologia D, Biochimica V, Milano U, T G (1996) NAD(P)H:(quinone-acceptor) oxidoreductase of tobacco Leaves is a flavin mononucleotide-containing flavoenzyme. Plant Physiol **112**: 249–258

Stowe BB, Obreiter JB (1962) Growth promotion in pea stem sections. II. by natural oils & isoprenoid vitamins. Plant Physiol **37**: 158

Takahashi T, Yamaguchi T, Shitashige M, Okamoto T, Kishi T (1995) Reduction of ubiquinone in membrane lipids by rat liver cytosol and its involvement in the cellular defence system against lipid peroxidation. Biochem J **309 (Pt 3**: 883–90

Veronesi C, Bonnin E, Calvez S, Thalouarn P, Simier P (2007) Activity of secreted cell wall-modifying enzymes and expression of peroxidase-encoding gene following germination of *Orobanche ramosa*. Biol Plant **51**: 391–394

Vranová E, Coman D, Gruissem W (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol **64**: 665–700

Wang L, Li Q, Zhang A, Zhou W, Jiang R, Yang Z, Yang H, Qin X, Ding S, Lu Q, et al (2017) The phytol phosphorylation pathway is essential for the biosynthesis of phylloquinone, which is required for photosystem I stability in Arabidopsis. Mol Plant **10**: 183–196

Westwood JH, DePamphilis CW, Das M, Fernández-Aparicio M, Honaas L a., Timko MP, Wafula EK, Wickett NJ, Yoder JI (2012) The parasitic plant genome project: new tools for understanding the biology of *Orobanche* and *Striga*. Weed Sci **60**: 295–306

Westwood JH, Yoder JI, Timko MP, dePamphilis CW (2010) The evolution of parasitism in plants. Trends Plant Sci **15**: 227–235

Wickett NJ, Honaas LA, Wafula EK, Das M, Huang K, Wu B, Landherr L, Timko MP, Yoder J, Westwood JH, et al (2011) Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. Curr Biol **21**: 2098–104

Wickett NJ, Zhang Y, Kellon Hansen S, Roper JM, Kuehl J V, Plock SA, Wolf PG, Depamphilis CW, Boore JL, Goffinetà B (2008) Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. Mol Biol Evol **25**: 393–401

Widhalm JR, Ducluzeau AL, Buller NE, Elowsky CG, Olsen LJ, Basset GJC (2012)

Phylloquinone (vitamin K_1) biosynthesis in plants: Two peroxisomal thioesterases of lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-coa. Plant J **71**: 205–215

Widhalm JR, van Oostende C, Furt F, Basset GJC (2009) A dedicated thioesterase of the Hotdog-fold family is required for the biosynthesis of the naphthoquinone ring of vitamin K1. Proc Natl Acad Sci U S A **106**: 5599–603

Wildermuth MC, Dewdney J, Wu G, Ausubel FM (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature **414**: 562–565

Wolfe KH, Morden CW, Palmer JD (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc Natl Acad Sci U S A **89**: 10648–52

Wrobel RL, Matvienko M, Yoder JI (2002) Heterologous expression and biochemical characterization of an NAD(P)H:quinone oxidoreductase from the hemiparasitic plant *Triphysaria versicolor*. Plant Physiol Biochem **40**: 265–272

Wullschleger SD, Oosterhuis DM (1990) Photosynthetic and respiratory activity of fruiting forms within the cotton canopy. Plant Physiol **94**: 463–9

Xu L, Law SR, Murcha MW, Whelan J, Carrie C (2013) The dual targeting ability of type II NAD(P)H dehydrogenases arose early in land plant evolution. BMC Plant Biol **13**: 100

Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, Huang K, Bandaranayake PCG, Wu B, Der JP, et al (2015) Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol **32**: 767–90

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol **21**: 809–818

Yoshida S, Cui S, Ichihashi Y, Shirasu K (2016) The haustorium, a specialized invasive organ in parasitic pChopralants.

CHAPTER 2

PLAS: PARALLELIZED LOCAL *DE NOVO* ASSEMBLY OF SEQUENCES[1]

[1] Gu, X., Queen S.J., and C.J. Tsai. To be submitted to *Molecular Plant*

Abstract

Rapid accumulation of sequenced transcriptome data (RNA-Seq) enables powerful and cost-efficient studies in comparative analyses, but it also poses great challenges for transcriptome assembly, particularly in non-model species where a reference genome is absent. Here we present a Parallelized Local *de novo* Assembly of Sequences (PLAS) pipeline that combines reference-based mapping and *de novo* assembly to improve both computing efficiency and assembly quality. PLAS uses quality-filtered RNA-Seq reads and a reference proteome from a closely-related species as input. The reference proteome is first clustered by gene family before read mapping, which effectively groups the input RNA-Seq data into bins for local *de novo* assembly. Because read assembly is performed independently for each bin, PLAS employs parallel computing to improve processing efficiency and memory usage. This group-and-assemble process is repeated and each iteration uses assembled sequence from the previous iteration as the new reference to re-group reads. The iterative process allows the assembled contigs to extend, thereby improving the assembly quality. To capture sequences that may be divergent from the reference proteome, input reads that do not map to the reference-guided assembly are subject to *de novo* assembly. The combined assemblies are quality-checked for redundancy to produce the final transcriptome assembly. The performance of PLAS was compared against Trinity on multiple datasets from species with or without a reference genome. PLAS showed robust improvement on both sensitivity (reconstructing more full-length transcripts) and specificity (achieving higher accuracy when comparing the assembly against the reference). PLAS is currently being implemented into CyVerse for broader accessibility.

Introduction

Generating high-quality transcriptomes from RNA-Seq data is important for gene expression assessments and comparative analyses. Several transcriptome assembly strategies have been developed to tackle the computational challenges posed by a large amount of short read sequence data. Depending on the availability of a reference genome, transcriptome analysis approaches can be classified as reference-based assembly or *de novo* assembly. Popular algorithms include Cufflinks (Trapnell et al., 2010) and Scripture (Guttman et al., 2010) for reference-based assembly, and Trinity (Grabherr et al., 2011), Oases (Schulz et al., 2012), Trans-AbySS (Robertson et al., 2010), and SOAPDenovo-Trans (Xie et al., 2014) for *de novo* assembly.

Reference-based assembly aligns short reads to a reference genome and reconstructs transcripts from the aligned reads. As the information of the reference genome is used to guide assembly, reference-based assemblers are able to reconstruct transcripts of low abundance and fill small gaps caused by low read coverage (Denoeud et al., 2008). Therefore, reference-based assemblers are more sensitive than *de novo* assemblers (Grabherr et al., 2011; Vijay et al., 2013; Marchant et al., 2016). In addition, reference-based assembly is more computationally efficient by distributing millions of reads into independent loci that can be assembled in parallel. As each locus usually contains less than thousands of reads, the analysis can be performed with a relatively low requirement on computing resources compared to *de novo* assembly. However, the performance of reference-based assembly depends heavily on the quality of the reference genome. This strategy is not suitable when the reference genome is unavailable or is of low quality.

De novo assembly does not require a reference genome. Instead, it leverages the depth of information contained in the reads to reconstruct transcripts. In *de novo* assembly, De Brujin graphs are built based on overlapping reads, then traversed to reconstruct isoforms. Even when the reference sequence is available, *de novo* assembly is useful for uncovering novel transcripts

and isoforms that are not annotated in the genome. However, d*e novo* assembly is computationally demanding, and time-consuming. Among the *de novo* assembly algorithms, Trinity is widely used to generate full-length transcripts and spliced isoforms efficiently (Grabherr et al., 2011; Zhao et al., 2011). However, in a simulation study, Trinity showed poor performance when a complex transcriptome with considerable number of paralogs was used (Vijay et al., 2013). In addition, Trinity is known to produce erroneous transcript isoforms that are not present in the transcriptome.

An assembly strategy that combines high sensitivity and computational efficiency of reference-based assembly with novel transcript detection capabilities of *de novo* assembly will be a powerful research tool. A similar strategy integrating reference-based and *de novo* assembly approaches has been described (Martin and Wang, 2011), but an associated software pipeline has not yet been made available. Here, we describe a Parallelized Local *de novo* Assembly of Sequences (PLAS) pipeline that was built upon the local assembly idea of aTRAM (automated target restricted assembly method) (Allen et al., 2015). aTRAM can assemble a small number of target genes across distantly related taxa using BLAST and a reference-guided, iterative process. However, the design of aTRAM is not applicable to genome-scale applications, and is limited to only one sequence library. PLAS was designed to extend the idea of aTRAM for whole transcriptome assembly from multiple libraries by employing parallel computing.

Materials and Methods

Overview

The PLAS pipeline, illustrated in Figure 2.1 and described in detail below, requires two inputs: quality-controlled RNA-Seq read data in fastq format; and a reference proteome or transcriptome from a closely-related species in fasta format. The pipeline consists of two major components. The first component assembles conserved (mappable) sequences to full length. The second component assembles diverged (unmapped) sequences, species-specific

sequences, and partial sequences. The final output will be a *de novo* assembled transcriptome from the input data. Note that a transcriptome reference can be an alternative when the reference species shares a highly-similar genome with the target species.

Reference organization

The reference transcriptome or proteome is first organized by orthologous gene families defined by the Markov Cluster Algorithm (MCL; Fraley et al., 2012; http://micans.org/mcl/). An all-against-all sequence similarity matrix is computed for the reference transcriptome or proteome. This similarity measurement is log-transformed E-value (-log10) from the results of WU-BLAST 2.2.6 (http://blast.wustl.edu/). The MCL classification results are confirmed by manually examining a known gene family (sucrose transporter, SUT). A cutoff E-value is chosen as 1e-5. The inflation parameter of MCL is set to 1.5 as suggested by OrthoMCL (Li et al., 2003).

A “hybrid” sequence can be generated when two similar genes (genes A1 and A2) are split into two groups and cross-attract each other’s reads during mapping, due to mismatch tolerance in Bowtie. As a result, some part of the hybrid sequence agrees with gene A1 and the other part aligns with gene A2. Such a hybrid sequence is an assembly artefact and does not exist in the real transcriptome. To avoid the cross-assembly, highly-similar genes need to be classified into the same group.

The OrthoMCL-sorted gene families are further combined into meta-groups in a manner that results in a roughly equal number of genes. *De novo* assembly is then performed on each read set (by bin) aligned to the meta-group independently for parallel computing. The number of groups is user tunable and should fit the amount of available resources, such as available computing nodes.

Assembly of conserved sequences

Input reads are first mapped to the reference proteome meta-groups by DIAMOND and organized into bins based on the presence or absence of significant hits against each meta-group. Significance is defined as a cutoff of E value = 1e-3 and is user-configurable. The read

bins are independent from one another and are *de novo* assembled by Trinity version r20140717 (Grabherr et al., 2011) in parallel to speed up the computing. The resulting contigs were mapped to the reference sequences in each bin using BLAST and only contigs with significant hits (Evalue = 1e-5) are retained for further extension.

Assembled contigs are then used as a starting point for the next round of assembly. From this point on, Bowtie2 (Langmead and Salzberg, 2012) is used for read mapping, and re-binned reads are used for the second iteration of *de novo* assembly. As the new reference is derived from the target transcriptome, it will recruit more closely-related reads to generate longer contigs of higher quality. The process is repeated until a user-defined number of iterations has taken place. The resulting contigs, residing in parallel bins (Intermediate Assembly I in Figure 2.1) are BLASTN-mapped against each other to remove redundancy. Sequences are considered redundant when they either share an overall identity above 95%, or when 90% of the shorter sequence(s) align with more than 99% identity across the alignment. The longer sequence is retained unless the shorter sequence is more similar to the reference sequence. The product is Intermediate Assembly II, which consists of conserved transcripts recovered to full length. A transcript is defined as full length when the aligned portion is either ≥98% of the reference coding sequence when the reference is from the same species, or ≥90% of the protein reference when the reference comes from a different species.

Assembly of species-specific and partial sequences

Using the reference proteome or transcriptome from a closely-related species can only recover conserved sequences. To assemble divergent, species-specific and partial sequences, the input reads are mapped to the Intermediate Assembly II using Bowtie2 and only unmapped reads are retained. Those unmapped reads are assembled *de novo* using Trinity to generate Intermediate Assembly III. The Intermediate Assembly III is BLASTN-searched against the Intermediate Assembly II to remove redundancy using the criteria described above. The combined non-redundant set is the final PLAS-assembled transcriptome.

Assembly Evaluation

The assemblies are evaluated in several aspects: 1) assembly score and associated components as defined by TransRate v1.0.1 (Smith-Unna et al., 2016); 2) the number of fully recovered transcripts and the corresponding gene models; 3) the alignment quality between the assemblies and the reference; 4) the reconstruction of highly-similar genes.

Datasets

Pollen datasets used in this study were from *Populus trichocarpa*, *Cornus florida* 'Appalachian Spring' (dogwood), *Lagerstroemia* sp. (crepe myrtle), *Quercus robur* (oak), *Prunus persica* (peach), *Prunus mume* (plum) and *Salix purpurea* L. (willow). These datasets include species with or without a sequenced genome, allowing the assessment of PLAS and Trinity performance on model and non-model species. Three biological replicates were sequenced for *Populus trichocarpa* with ~21M paired-end 75bp (PE75) reads after quality control, four biological replicates for dogwood with ~72M PE75 reads, two replicates for peach with ~33M PE75 reads, two replicates for oak with ~19M PE75 reads, two replicates for plum with ~13M PE75 reads, three replicates for crepe myrtle with ~94M PE75 reads, and two replicates for willow with ~35M PE75 reads. Raw reads were pre-processed by Cutadapt 1.9.dev1 (Martin, 2011), Trimmomatic 0.32 (Bolger et al., 2014) and custom scripts to remove adapter, non-coding RNA, organellar sequences, and low-quality reads. After quality control, PLAS and Trinity were used to assemble transcriptomes for the seven datasets. *Populus trichocarpa* proteome was used as the reference for all seven datasets. For the *Populus trichocarpa* pollen dataset, the *Mimulus guttatus* proteome was also used as the reference to assess the effects of the evolutionary distance between the reference and the target species on the accuracy of the resulting assembly. Transcript abundance was estimated by eXpress 1.5.1 (Roberts and Pachter, 2013). Gene expression of *P. trichocarpa* pollen was also estimated by aligning the reads to *P. trichocarpa* reference genome with Tophat 2.0.13 (Kim et al., 2013) followed by read count with HTseq 0.6.1p1 (Anders et al., 2015) and expression estimation with DEseq2 (Love et

al., 2014). This reference-based gene expression analysis was used to assess the effects of gene properties, such as sequence depth, expression level and copy number, on assembly quality.

Results

Populus trichocarpa pollen transcriptome assembly

General Statistics

As PLAS adopts Trinity as the tool for *de novo* assembly, the output format of contig IDs from PLAS is the same as that from Trinity. The contig IDs consist of three parts joined together with an underscore. The first part is read cluster ID, the second is (inferred) gene ID and the third is (inferred) isoform ID. The number of unique gene IDs and isoform IDs obtained by Trinity and PLAS assemblies are summarized in Table 2.1. PLAS generated more isoforms and putative genes than Trinity (Table 2.1). To determine the proportions of transcripts recovered to full length or near-full length, the assemblies were aligned to *P. trichocarpa* transcriptome (cds). PLAS recovered 26% more full-length transcripts and 34% more full-length genes than Trinity (Table 2.1). Although PLAS recovered more isoforms and putative genes than Trinity, the average number of isoforms per gene (45141/39913=1.13, or full-length 8604/7184=1.20) is slightly smaller than that (41896/35561=1.18, or full-length 6828/5360=1.27) of Trinity. To examine the read representation of the assembled transcriptomes, the input reads were aligned back to the two assemblies. PLAS exhibited higher mapping rates than Trinity.

Two versions of PLAS assemblies were generated, with *Populus trichocarpa* (the same species) or *Mimulus guttatus* (a distantly related species) proteome as the reference. The *Populus*-derived assembly showed better performance than the *Mimulus*-derived assembly in the number of full-length transcripts and genes recovered, suggesting that the assembly quality is correlated with phylogenetic distance between the reference and target species. It is worth noting that the *Mimulus*-derived assembly still showed superior performance compared to Trinity,

although with a smaller margin of improvements. This indicates that even when using a less related species as the reference, PLAS can generate better assemblies than Trinity.

TransRate Evaluation

TransRate (Smith-Unna et al., 2016) was used to quantitatively assess the quality of the assemblies, based on the comparison of input reads to assemblies, and of reference transcriptome to assemblies. The assembly score is a quantitative measure of the accuracy and completeness of the assembly, with a higher score indicative of a more biologically accurate assembly. PLAS achieved an assembly score of 0.36 and 0.34 in TransRate, superior to Trinity (Table 2.1) and the majority (>78%) of the 155 published *de novo* assemblies available in the NCBI Transcriptome Shotgun Assembly (TSA) database complied by (Smith-Unna et al., 2016).

When compared to the *P. trichocarpa* reference transcriptome, PLAS showed better coverage than Trinity (Figure 2.2A). Given that the assembly score is a summarized measure of individual contig score, we further examined the density distribution of contig scores of the two assemblies. The distribution of PLAS contig scores was shifted to the right compared to that of Trinity, revealing more contigs with higher scores by PLAS (Figure 2.2B). The two components of the contig score, “p_good” and “p_seq_true”, were extracted from the TransRate output and their density distributions were displayed. The “p_good score” is a measure of structural correctness, with low scores indicating incompleteness, spurious insertions, or improper assembly. “P_seq_true” measures how well a contig is supported by reads, with low scores indicating gene collapse within a gene family. PLAS showed superior performance in both components (Figure 2.2C, D). Taken together, TransRate assessments indicate that the PLAS assembly is of higher quality than the assembly produced by Trinity.

Effects of gene properties

To understand how gene properties affect assembly, we classified the *Populus* genes based on (1) their expression levels estimated by reference-based read mapping to the *P. trichocarpa* genome (Figure 2.3A, see Material and Methods) and (2) the size of the corresponding gene families according to MCL (Figure 2.3B, see Material and Methods). Highly expressed genes are better assembled to full-length due to greater coverage depth. Except for the first class (FPKM <1) where genes were barely detected, PLAS showed consistently higher recovery rates than Trinity across the full range of expression levels. When the expression reached FPKM >10, the recovery rate of both methods began to plateau, with PLAS (~80%) outperforming Trinity (~60%). The leveling off in recovery score of expressed genes has also been reported previously (Zhao et al., 2011). Surprisingly, PLAS did not fully recover highly-expressed genes within the data set. To further understand the cause of this limitation, the bin with the most highly expressed gene (>1000 FPKM) was manually examined. Among the 129 genes in this bin, 20 genes were missed by PLAS (84% recovery). Ten genes were not retrieved due to the presence of paralogs with high nucleotide and amino acid identities (97% and 100%, respectively). Five genes were not recovered due to gene model mis-annotation or alternative splicing events. In both cases, PLAS-assembled transcripts deviated from the reference gene models, but were supported by read mapping. This attested to the power of PLAS to retrieve true transcripts. Three genes failed to be assembled because the corresponding contigs were hybrids derived from two genes sharing a short stretch of common sequences. Partial transcripts were recovered for the two remaining genes, missing the 5' end. All the manually examined cases were documented in Supplemental File S2.1.

The genome of *Populus trichocarpa* has experienced multiple whole-genome, segmental and tandem duplication events, which pose a significant challenge to sequence assembly. PLAS showed a slightly higher recovery rate than Trinity for single-copy genes. However, the improvements were substantially higher for duplicated genes (Figure 2.3B). This suggested that

PLAS is better able to distinguish highly-similar genes and recover more full-length transcripts of multi-copy genes. Tubulin gene family was used as an example to examine the sensitivities of PLAS and Trinity to highly-similar genes. In the case of tubulin alpha (TUA), *TUA2* and *TUA4* are highly-similar duplicates. PLAS was able to correctly reconstruct *TUA2*. However, Trinity generated a hybrid sequence with the N-terminus matched to TUA2 and the C-terminus to TUA4 (Figure 2.3C). This example provided evidence for higher accuracy of PLAS over Trinity when assembling highly-similar genes.

Quality of fully assembled transcripts

We extracted the portion of fully assembled transcripts (Table 2.1) from both assemblies for comparison. The majority of both assemblies overlapped with each other, but PLAS recovered more full-length transcripts than Trinity (Figure 2.4A). When aligned to the *P. trichocarpa* reference, PLAS assembled transcripts exhibited fewer mismatches than those assembled by Trinity (Figure 2.4B and 2.4C). A similar trend was observed for BLAST bit scores (Figure 2.4D and E). Together, the results showed that PLAS produces a higher-quality assembly than Trinity for transcripts that were recovered by both methods.

Pollen transcriptome assembly of non-model species

Both PLAS and Trinity were applied to pollen datasets of dogwood for transcriptome assembly. Four samples were pre-processed and around 70 millions of PE-75 reads were retained after quality filtering (70 M dataset). To examine the effects of sequencing depth (read number) on the assembly quality, a smaller dataset containing just two samples with a total of 16 million reads (16 M dataset) was independently assembled. For the 16M dataset, both *P. trichocarpa* and *Mimulus guttatus* proteomes were used as the reference.

The results showed again that PLAS reconstructed more unique isoforms and putative genes than Trinity in all test datasets (Table 2.2). PLAS also recovered more full-length transcripts and genes than Trinity, with higher mapping rates in all conditions (Table 2.2). Although we were not able to generate a Transrate score for PLAS assembly of the 70M

dataset, PLAS achieved higher scores than Trinity in other scenarios (Table 2.2). For the 70M dataset, PLAS showed larger improvement over Trinity compared to that for the 16M dataset, indicating PLAS can better employ the information contained in large datasets. The use of different proteome references had little effect on the assembly quality (Table 2.2). We noted, however, slightly higher recovery rates with an Asterids reference (*Mimulus*) than with a Rosids reference (*Populus*), consistent with the classification of dogwood in Asterids.

The performance improvements of PLAS over Trinity for the dogwood pollen data were not as pronounced as the *P. trichocarpa* pollen data using a heterologous proteome (*Mimulus*) reference, especially from the smaller (16M) datasets. Overall, fewer full-length transcripts and genes were recovered from the dogwood pollen dataset than the *Populus* pollen dataset, despite similar number of contigs. The results may suggest differences in pollen transcriptomes between the two species. Because PLAS was more effective than Trinity in reconstructing paralogs to full length, the smaller gains observed in the dogwood dataset may reflect different extents of duplication in the two genomes. To test this idea, ks (synonymous substitution) distributions were calculated for the pollen transcriptomes of the two species (Figure 2.5). *P. trichocarpa* pollen transcriptome showed considerably more paralogs with ks values in 0.0-0.4 than dogwood pollen transcriptome. Paralogs with small ks values are likely derived from recent duplication events and share higher levels of sequence similarity than paralogs with large ks values. Given the strength of PLAS in handling paralogs, the ks analysis may explain the greater improvement of PLAS over Trinity for poplar *P. trichocarpa* (with more recent duplicates) than dogwood transcriptomes.

To test the robustness of PLAS performance, we applied both PLAS and Trinity to the pollen data sets of crepe myrtle, oak, peach, plum, and willow. The results supported the superior performance of PLAS over Trinity across a wide range of data sets, with greater rates of full-length transcript/gene reconstruction, and read mapping, as well as better TransRate assembly scores (Table S2.1).

Parasitic plant transcriptome assembly

The PLAS pipeline has also been used to assemble parasitic plant transcriptomes to recover transcripts that were missed by Trinity and CLC Assembly Cell (CLC) (Yang et al., 2015), as part of the Chapter 3 investigation. Parasitic plants cause huge economic losses by feeding off agriculturally important crops. There have been significant interests in understanding the molecular mechanisms of parasitism, including transcriptomics resources generated by the Parasitic Plant Genome Project (PPGP) for three representative species from Orobanchaceae (Westwood et al., 2012). As part of our investigation into non-photosynthetic function of phylloquinone, we searched the PPGP Trinity and CLC assemblies using as queries phylloquinone biosynthetic genes from *Mimulus* which is a closely-related species of Orobanchaceae. Putative orthologs identified by BLASTx were largely fragmented (Table S2.2-2.4), in part because individual samples were assembled independently due to computational constraints (Westwood et al., 2012). PLAS was able to assemble all samples from the same species, including both 454 and Illumina data, into a single transcriptome. Nearly all *Men* genes of the three species were fully reconstructed by PLAS (Figure 2.6), enabling identification of non-canonical phylloquinone biosynthesis in the plasma membrane (see Chapter 3). Full-length sequence of a parasitism gene *TvQR1* was also obtained (Supplemental File S2.2). TvQR1 transcript was highly fragmented in the PPGP database (visited May 25, 2017), and multiple BLASTn alignments of *TvQR1* was used to construct the full-length sequence in a previous study (Yang et al., 2015). Taking together, application of PLAS to parasitic plant transcriptome data further supports its robustness in handling various RNA-Seq data.

Discussions

Here we report a new pipeline, PLAS, for reference-guided *de novo* assembly that shows improved performance compared to the Trinity and CLC assembly pipelines. Several key features of PLAS include (1) the use of protein sequences as reference, (2) the organization of input reads into independent bins for simultaneous processing, and (3) iterative assembly. Protein sequences from a closely-related species are used to organize data from a non-model organism into independent bins by gene families. This pre-organization effectively reduces data complexity, as only reads matching the reference are used for the assembly. This alleviates the limitation of Trinity in handling large dataset, while facilitating parallel computing. The iterative assembly extends the assembled sequence length and improves assembly quality. Together, these features allow PLAS to achieve higher coverage, accuracy, and computational efficiency. A previous method aTRAM incorporated features #1 and #3, but was designed for a limited target sequences, not scalable for transcriptome-wide applications etc.

Improved de novo assembly performance of PLAS

As the volume of sequence data grows, the computing requirements by Trinity and other *de novo* assemblers increase as well because *de novo* assembly is treated as an indivisible problem that can only be finished in a shared-memory environment. If the memory requirement of the assembly job exceeds the available memory, a common solution is to perform the assembly on smaller datasets separately (e.g., on a sample-by-sample basis) before the results are combined. PLAS assigns reads into bins organized by gene families, thus dramatically reducing the actual data volume used for assembly, and consequently, the memory requirement for each assembly task. By reducing the problem to a set of smaller tasks, PLAS can avoid the shared memory requirement and complete large *de novo* assembly jobs on a computing cluster without extensive memory demand. PLAS can use as many nodes as there are available to further reduce computing time. For example, it is impossible to generate a Trinity assembly from a parasitic plant RNA-Seq data set (14 samples and 281 Gb total for *Triphysaria versicolor*) at

once in a server with 48-core, 1 TB RAM, and AMD Opteron processors. As a result, individual samples were assembled separately before being combined into a complete assembly in a previous study (Westwood et al., 2012). In contrast, PLAS can assemble all samples at once (multiple nodes with 8-core, 48 GB RAM, and Intel Xeon processors. The number of nodes can be configured by users) without requesting resources from a large memory queue.

PLAS outperformed Trinity in both sensitivity and specificity. Specifically, PLAS reconstructed a greater number of full-length transcripts than Trinity across a wide range of expression levels (FPKM10-1000). PLAS also demonstrated higher specificity in that the alignments between PLAS assembly and the reference showed fewer mismatches and higher bit scores. These results suggest that with the same amount of sequence reads, PLAS can better utilize the information for assembly by limiting the process to only relevant sequences through data pre-organization.

Complex transcriptomes with the presence of paralogs are more challenging for *de novo* assemblers (Vijay et al., 2013). PLAS showed higher sensitivity and better resolution when reconstructing paralogous transcripts. However, for highly-similar genes with identity above 97%, no short-read assembly methods can properly distinguish the paralogs. When the distance between two SNPs of a set of paralogs is larger than the length of the reads, it is beyond the power of the data to join the two SNPs together correctly.

It is worth noting that for genes with relatively high expression levels, PLAS was not able to reconstruct their transcripts to full length after accounting for highly-similar paralogs, annotation errors or alternative splicing. Similar observations have been documented elsewhere that highly expressed genes are often assembled into incomplete and sometimes hybrid transcripts (Grabherr et al., 2011; Zhao et al., 2011). The underlying reasons remain elusive, but may be related to the exaggerated sequence errors that occur when many reads are generated. Regardless, PLAS correctly reconstructed more highly expressed genes than Trinity.

Comparison to the genome-guided assembly method of Trinity

In addition to *de novo* assembly, Trinity also offers a genome-guided assembly function where reads are first aligned to a reference genome and grouped by genomic locus, followed by *de novo* assembly at each locus. This differs from reference-based assembly because the genome is only used to organize the reads before assembly, instead of being used for map-based assembly. Trinity's genome-guided assembly function has been used in constructing the pine (*Pinus patula*) juvenile shoot transcriptome (Visser et al., 2015).

Trinity's genome-guided assembly function requires a high-quality reference genome, which poses considerable limitations. Because of the relatively large variation in coding sequences when compared to protein sequences, genome-guided assembly must employ a reference from the same species, which limits its utility to model species with a sequenced genome. Additionally, when the reference genome is highly fragmented, as is typical in draft versions, ambiguities in the genome such as misassembled or gapped regions will likely introduce errors to the assembled transcripts. Genome-guided assembly also highly depends on the quality of gene structure annotation and the performance of alignment methods to handle gaps (introns). Aligning RNA-Seq reads to a genome is always more challenging than to transcripts or proteins due to intervening intronic sequences and alternative splicing. Most importantly, as described above, *de novo* assembly of genes with similar sequences (e.g. paralogs) tends to generate artificial hybrid sequences. Because genes with similar sequences are usually located at different positions of the genome (with the exception of tandem duplicates), hybrid artifacts of genome-guided assembly will likely to be severe, especially for plant species that have undergone several rounds of genome duplications.

In contrast, the PLAS pipeline is designed to bypass the limitations described above by employing a "reference proteome" from a non-self species. Protein sequences are relatively well-conserved even across evolutionary distance, thus enabling the use of a high-quality proteome from a closely-related species to guide transcriptome assembly in non-model species.

Without intervening intronic sequences, reads can be aligned continuously to the reference, which eases the alignment challenge. Proteins with similar sequences are clustered before being used to organize reads, which significantly alleviates the issue of artificial hybrids. Iterative computing that uses sequences assembled in a previous run extends the assembly length in the next run. Thus, compared to the Trinity genome-guided assembly function, PLAS provides more flexibility and power for transcriptome assembly in non-model species.

Adaptability to other de novo assembly approaches and future improvements

Transcriptome assembly is a complex problem with considerable variations in input data. The quality of the assembly is affected by a wide range of factors, such as transcript abundance, RNA heterogeneity (pre-mRNA, mature RNA and degraded RNA), alternative splicing, and sequence polymorphism associated with outcrossing species. Different *de novo* assembly algorithms have different strengths and weaknesses. Compared to other *de novo* assemblers like SOAPdenovo and Trans-ABySS, Trinity reconstructed more full-length transcripts in *S. pombe* and mouse (Grabherr et al., 2011). Trinity was found to generate assemblies with better contiguity and longer contigs (Vijay et al., 2013; Chopra et al., 2014) when compared to other *de novo* assemblers. The performance of Trinity was also more robust to a broad range of parameter configuration and input data features (Grabherr et al., 2011; Li et al., 2014). However, Trinity tends to erroneously infer artificial transcript isoforms that do not reflect the real transcriptome (Vijay et al., 2013). The number of artificial isoforms increases dramatically when the transcriptome becomes more complex in terms of size and paralogs. In these cases, SOAPdenovo-trans showed better performance than Trinity (Vijay et al., 2013). We also found that Trinity can generate hybrid assemblies derived from different similar genes (**Figure 2.3C**).

As PLAS employs Trinity to perform reference-guided *de novo* assembly, PLAS cannot fully resolve the inherent limitations of Trinity. Therefore, future improvements of PLAS may consider alternative *de novo* assembly algorithms like SOAPdenovo-trans. To make PLAS

accessible to a broader scientific community, future work includes implementation into CyVerse, an open-source platform hosting numerous bioinformatics tools.

Table 2.1. Assembly statistics of *P. trichocarpa* pollen transcriptome by Trinity and PLAS

	Trinity	PLAS	
Reference proteome	-	*P. trichocarpa*	*M. guttatus*
Contig (Transcript) Number	41896	45141	45129
Putative Genes	35561	39913	39315
Full Length Transcripts	6828	8604	7921
% increased (vs. Trinity)		26.01%	16.01%
Full Length Genes	5360	7184	6506
% increased (vs. Trinity)		34.03%	21.38%
Read Mapping Rate	87.91%	88.76%	92.48%
Assembly Score	0.2833	0.3582	0.3398

Table 2.2. Statistics summary for comparison of Trinity and PLAS in assembling dogwood pollen transcriptome

	70 M		16 M		
Dataset	Trinity	PLAS	Trinity	PLAS	
Reference proteome	-	*Ptr*	-	*Ptr*	*Mgu*
Contig Number	69,044	70,366	34,972	35,957	35,959
Putative Genes	58382	60182	31376	32697	32744
Full Length Transcripts	6,779	8,053	4,259	4,902	5,063
% increased (vs. Trinity)		18.79%		15.09%	18.88%
Full Length Genes	4,881	5,426	3,353	3,673	3,700
% increased (vs. Trinity)		11.17%		9.54%	10.34%
Read Mapping Rate	85.13%	93.30%	86.81%	88.30%	87.99%
Assembly Score	0.3059	NA*	0.3510	0.3678	0.3746

* unable to be obtained from TransRate. *Ptr: P. trichocarpa. Mgu: M. guttatus.*

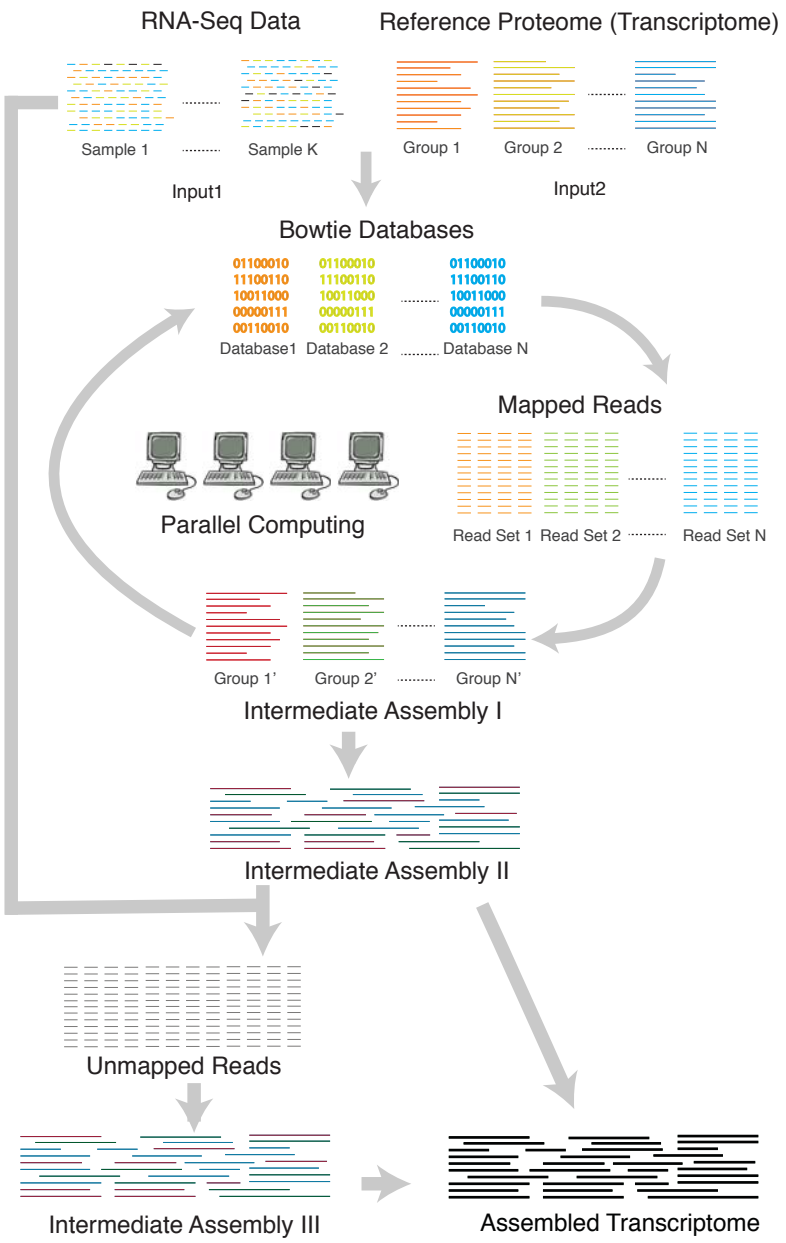


Figure 2.1. Schematic of the PLAS pipeline to perform *de novo* local assembly using parallel computing. Long lines in each group represent transcript or protein sequences. Different colors indicate different groups defined in the Method. Short line fragments in each sample represent short RNA-Seq reads. Different colors indicate the reads come from genes which have orthologs in the corresponding colored reference group. For example, orange reads are generated from genes with orthologs in the orange group of the reference.

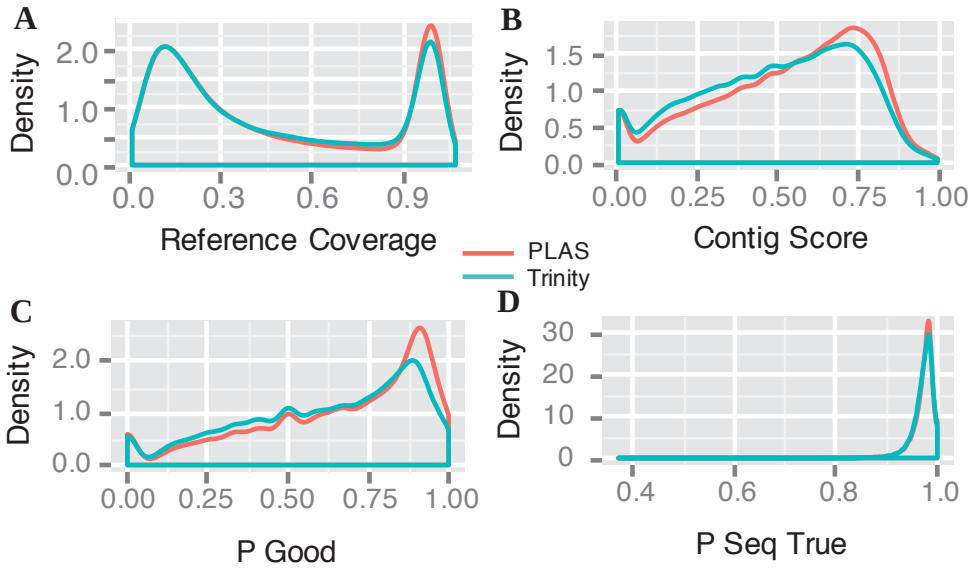


Figure 2.2. Density distribution of quality scores measured by comparing assemblies vs. reference and assemblies vs. input reads. (**A**) Reference coverage refers to the proportion of reference transcriptome recovered by the corresponding assembly. (**B**) TransRate contig score is calculated by summarizing various aspects of supportive evidence from the input reads for each assembly. (**C**) and (**D**) are two components of a contig score. “P Good” is a measure of structural correctness, with low scores indicating incompleteness, spurious insertions, or misassembly. “P Seq True” measures how well a contig is supported by reads, with low scores indicating gene collapse within a gene family.

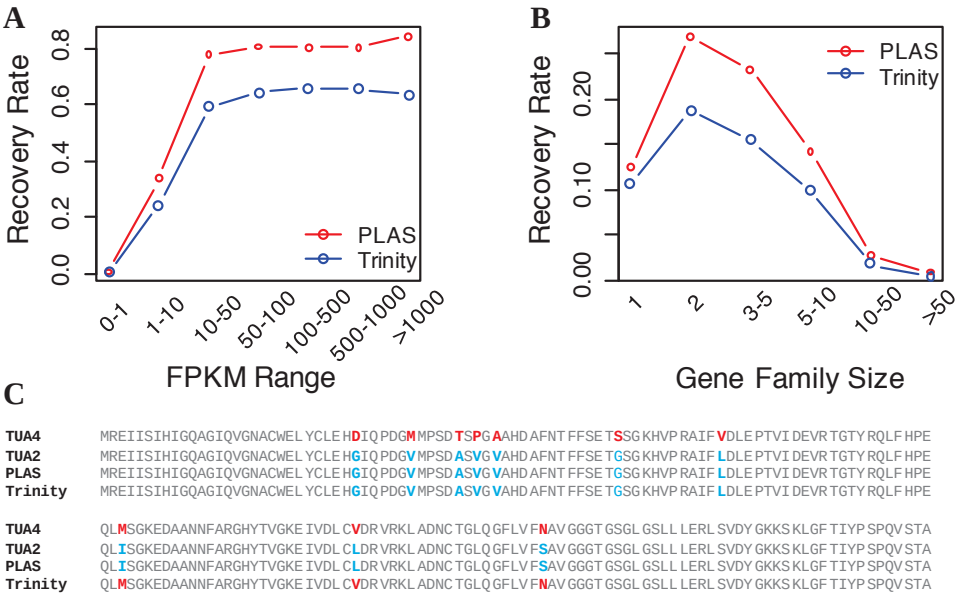


Figure 2.3. Proportions of transcriptome fully recovered by PLAS or Trinity. (**A**) Genes were binned based on the expression levels. In the first two bins (FPKM <10) where genes were lowly expressed, both methods showed poor performance with little difference. When the expression increased to FPKM >10, both method started to show stable performance (~80% for PLAS and 60% for Trinity). (**B**) Genes were binned based on the duplication copy number. The data supported that PLAS is better able to distinguish highly-similar genes and recover more full-length transcripts of multi-copy genes. (**C**) Alignment of two tubulin alpha duplicates.

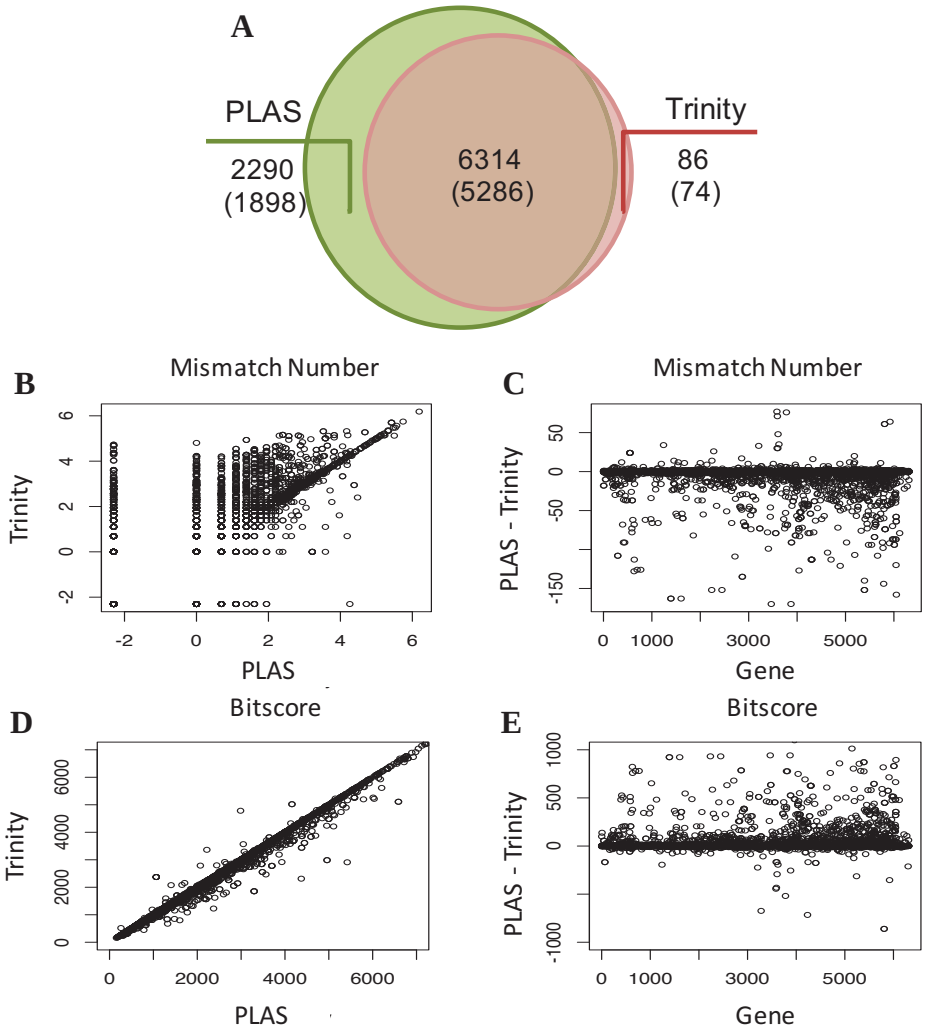


Figure 2.4. Comparisons of fully assembled transcripts by PLAS and Trinity. (A) Overlap of fully assembled transcripts (genes) between the two methods. The majority of Trinity transcripts (98.66%) were also recovered by PLAS, whereas PLAS recovered 2,290 (1898) extra transcripts (genes). (**B**-**E**). Sequence accuracy of the overlapped set in (A). Transcripts were BLAST against reference cds. (**B**-**C**) Total mismatch number of each alignment. (**D**-**E**) Bitscore of each alignment. F. Assemblies of highly-similar genes.

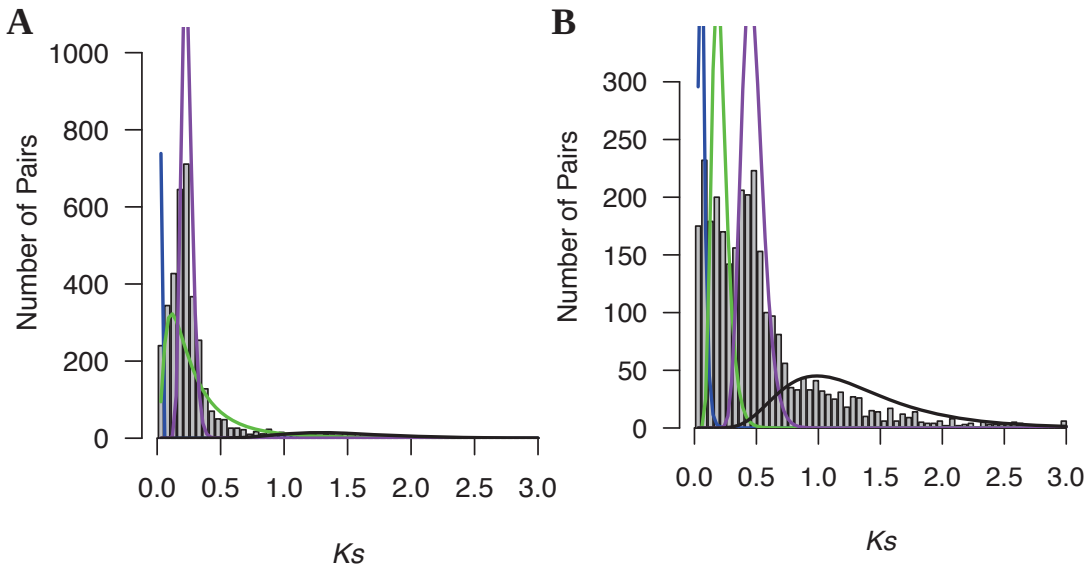


Figure 2.5. Ks distributions of *P. trichocarpa* (A) and dogwood (B) pollen transcriptomes assembled by PLAS.

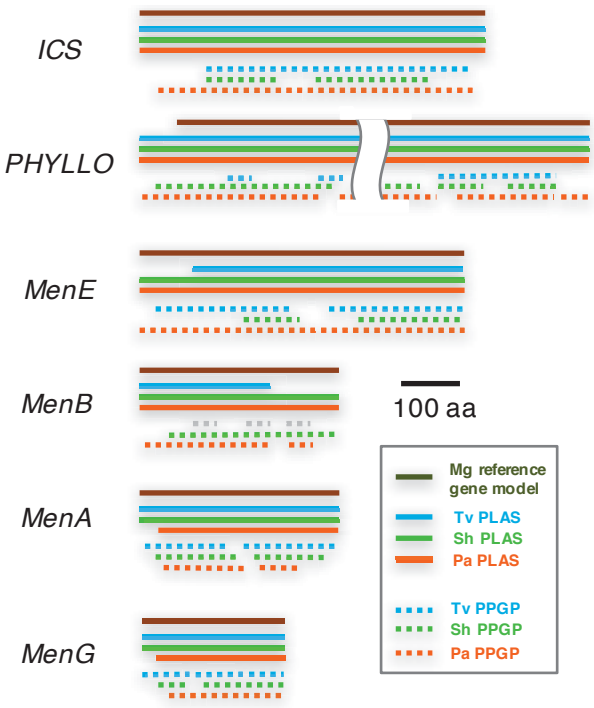


Figure 2.6. Assembled transcripts for PhQ genes in three parasitic plants (visited October 08,2014)**.** The length of the lines in the diagram is scaled to the protein length except for *PHYLLO*. Tv, *Triphysaria versicolor*; Sh, *Striga hermonthica*; Pa, *Phelipanche aegyptiaca*; Mg, *Mimulus guttatus*; PPGP, assembly provided by the PPGP database.

Table S2.1. Assembly statistics of multipe pollen transcriptomes by Trinity and PLAS

	Peach		Oak		Plum		crepe myrtle		willow	
	Trinity	PLAS	Trinity	PLAS	Trinity	PLAS	Trinity	PLAS	Trinity	PLAS
Reference proteome	-	*Pt* †	-	*Pt*	-	*Pt*	-	*Pt*	-	*Pt*
Total Number of Reads	33,261,317		18,886,945		12,983,418		93,856,451		34,706,057	
Contig Number	43,056	45,517	48,817	51,450	23,605	24,824	67,719	70,479	67,042	71,286
Putative Genes	31,424	34,100	36,067	39,129	20,708	22,377	55,290	57,469	48,258	52,808
Full Length Transcripts	3,656	4,134	4,140	4,790	3,963	4,285	10,825	14,029	5,375	6,755
% increased (vs. Trinity)		13.07%		15.70%		8.13%		29.60%		25.67%
Full Length Genes	2,642	2,985	3,017	3,467	3,093	3,387	7,150	8,033	3,653	4,595
% increased (vs. Trinity)		12.98%		14.92%		8.68%		12.35%		25.79%
Read Mapping Rate	78.74%	79.90%	75.88%	80.40%	91.39%	92.58%	89.02%	95.28%	83.02%	85.68%
Assembly Score	0.0001	17.38%	0.0001	13.96%	0.3956	40.49%	0.3224	N/A §	0.0009	18.58%

† *Pt* is short for *Populus trichocarpa*

§ unable to obtain from TransRate

Table S2.2. TBLASTN results of PhQ biosynthetic genes against PPGP database for *P. aegyptiaca* (visited December 16, 2015)

query ID	hit ID	identity	align len	mis-match	query start	query end	hit start	hit end	E-value	bit score
Migut.I00130.1	OrAe61GB1_8917*	78.87	497	96	89	579	14	1495	0	798
(MgICS1) †	OrAeGnB1_40972	78.39	509	101	69	571	9	1526	0	788
	OrAeBC5_348.1	76.81	526	107	28	547	1553	3	0	781
	OrAe41GB1_35235	78.77	457	88	91	541	5	1366	0	736
	OrAe0GB1_32434	79.64	447	84	134	575	1464	130	0	736
Migut.I00129.1	OrAeBC5_348.1	78.24	533	106	19	546	1586	3	0	815
(MgICS2)	OrAe61GB1_8917	80.82	490	90	89	574	14	1483	0	812
	OrAeGnB1_40972	80.04	506	97	69	570	9	1526	0	809
	OrAe41GB1_35235	80.22	455	86	90	540	2	1366	0	752
	OrAe0GB1_32434	80.94	446	82	133	575	1464	127	0	748
Migut.L01140.1	OrAeBC5_5680.1	72.85	1444	329	249	1634	4637	321	0	2108
(MgPHYLLO)	OrAeBC5_5680.1	71.09	256	65	1	251	5382	4627	2.00E-98	355
	OrAeBC5_5680.2	75.36	1376	324	269	1634	4433	321	0	2096
	OrAeGnB1_138378	69.35	757	162	9	703	2	2248	0	1010
	OrAe42GB1_75264	72.89	653	176	982	1634	3	1958	0	989
	OrAe3GB1_55516	76.97	521	115	612	1131	3	1553	0	775
Migut.H01327.1	OrAeBC5_6326.1 §	78.01	564	115	1	557	67	1752	0	915
(MgMenE)	OrAeGnB1_19007	77.66	564	117	1	557	1753	68	0	908
	OrAe41G2B1_66532	77.32	560	118	1	553	1676	3	0	904
	OrAe41GB1_49325	77.72	552	115	1	545	1654	2	0	898
	OrAe3GB1_77840	77.66	555	115	10	557	1813	155	0	894
Migut.E00173.1	OrAe61GB1_13875	86.39	338	46	4	341	1092	79	0	621
(MgMenB)	OrAe2FB1_524	86.09	338	47	4	341	68	1081	0	618
	OrAe1FB1_1140	85.21	338	50	4	341	60	1073	0	614
	OrAe1FB1_1283	80	335	64	4	336	71	1072	0	554
	OrAe2FB1_1775	90.7	215	20	127	341	438	1082	4.00E-165	415
	OrAe2FB1_1775	72.64	106	29	29	134	143	460	4.00E-165	169
	OrAe2FB1_1775	65.38	26	9	4	29	67	144	4.00E-165	39.7
Migut.B00584.1	OrAe42GB1_55606	70.36	307	91	75	381	85	1005	1.00E-134	396
(MgMenA1)	OrAeBC5_3495.2	66.67	321	104	75	392	1214	252	4.00E-129	384
	OrAe2FB1_110	66.98	321	103	75	392	1228	266	4.00E-129	385
	OrAeBC5_3495.1	66.67	321	104	75	392	1214	252	4.00E-128	384
	OrAe3GB1_45644	68.66	268	83	96	362	805	2	1.00E-115	344

Table S2.2 (continued). TBLASTN results of PhQ biosynthetic genes against PPGP database for *P. aegyptiaca* (visited December 16, 2015)

query ID	hit ID	identity	align len	mis-match	query start	query end	hit start	hit end	E-value	bit score
Migut.B01155.1	OrAe2FB1_110	67.43	304	98	86	388	1255	344	3.00E-150	440
(MgMenA2)	OrAeBC5_3495.2	67.67	300	96	90	388	1229	330	4.00E-150	439
	OrAeBC5_3495.1	61.14	350	129	44	388	1373	330	4.00E-149	439
	OrAe42GB1_55606	67.85	311	94	84	388	34	966	2.00E-140	411
	OrAe3GB1_45644	68.66	268	83	116	382	805	2	1.00E-135	396
Migut.B01157.1	OrAe2FB1_110	66.2	284	95	86	368	1255	404	3.00E-135	401
(MgMenA3)	OrAeBC5_3495.2	66.43	280	93	90	368	1229	390	9.00E-135	399
	OrAeBC5_3495.1	59.7	330	126	44	368	1373	390	3.00E-134	400
	OrAe42GB1_55606	66.67	291	91	84	368	34	906	1.00E-125	373
	OrAe3GB1_45644	68.11	254	80	116	368	805	44	2.00E-125	370
Migut.E00183.1	OrAe2FB1_4148	67.57	259	74	4	260	9	761	4.00E-117	345
(MgMenG)	OrAe1FB1_4532	72.53	233	62	30	260	79	777	5.00E-117	345
	OrAe2FB1_1928	71.67	233	64	30	260	55	753	7.00E-114	336
	OrAeBC5_992.1	71.65	194	53	69	260	821	240	5.00E-91	276
	OrAe2FB1_32440	73.83	149	37	30	176	772	326	1.00E-81	229
	OrAe2FB1_32440	68.12	69	22	192	260	278	72	1.00E-81	94.4

† protein length is 583 aa for MgICS1, 582 aa for MgICS2, 1637 aa for MgPHYLLO, 557 aa for MgMenE, 341 aa for MgMenB, 394 aa for MgMenA1, 414 aa for MgMenA2, 397 aa for MenA3, and 260 aa for MgMenG

§ records highlighted in red represent contigs reconstructed to their full length
* Hit IDs are from PPGP assemblies. OrAe is short for *Orobanche aegyptiaca* which is the old species name for *Phelipanche aegyptiaca.*

Table S2.3. TBLASTN results of PhQ biosynthetic genes against PPGP database for *S. hermonthica* (visited December 16, 2015)

query ID	hit ID	identity	align len	mis-match	query start	query end	hit start	hit end	E-value	bit score
Migut.I00130.1	StHeBC3_8527.1 §	75.9	556	127	17	571	195	1844	0	815
(MgICS1) †	StHeBC3_8527.2	75.39	516	120	17	531	195	1724	0	750
	StHe1GB1_47414	77.78	405	86	179	583	1209	7	0	645
	StHeBC3_8527.3	86.23	247	34	325	571	84	824	4.00E-150	445
	StHe61GB1_35983	81.2	234	44	248	481	703	2	1.00E-132	394
Migut.I00129.1	StHeBC3_8527.1	77.66	555	120	16	569	186	1841	0	852
(MgICS2)	StHeBC3_8527.2	76.94	516	115	16	530	186	1724	0	781
	StHe1GB1_47414	81.4	387	72	196	582	1167	7	0	657
	StHe1GB1_47414	63.73	193	66	16	207	1702	1133	7.00E-57	206
	StHeBC3_8527.3	88.66	247	28	324	570	84	824	1.00E-155	459
	StHe61GB1_35983	82.48	234	41	247	480	703	2	3.00E-136	403
Migut.L01140.1	StHe2GB1_49012	71.23	1477	347	219	1635	1	4377	0	2079
(MgPHYLLO)	StHe1GB1_6717	70.11	833	178	93	867	2461	2	0	1147
	StHe1GB1_6476	72.37	778	206	861	1636	2324	12	0	1129
	StHeBC3_4281.3	76.92	702	157	583	1283	1	2094	0	1075
	StHeBC3_4281.7	77.81	658	141	583	1239	1	1962	0	1016
Migut.H01327.1	StHe3FB1_9446	60.26	234	78	191	416	289	969	6.00E-102	286
(MgMenE)	StHe3FB1_9446	58.33	48	20	127	174	93	236	6.00E-102	65.9
	StHe3FB1_9446	84.62	26	4	93	118	8	85	6.00E-102	49.3
	StHe3FB1_9446	93.75	16	1	176	191	242	289	6.00E-102	33.9
	StHe4FB1_14386	75.17	149	37	180	328	447	1	3.00E-97	233
	StHe4FB1_14386	85.11	47	7	67	113	770	630	3.00E-97	82.8
	StHe4FB1_14386	66.07	56	19	126	181	610	443	3.00E-97	82.8
	StHeBC3_5002.3	73.82	191	50	115	305	3	575	9.00E-97	299
	StHeBC3_5002.2	71.67	180	51	126	305	142	681	1.00E-90	270
	StHeBC3_5002.2	80	50	10	74	123	2	151	1.00E-90	85.9
	StHe1G2B1_76249	66.35	211	61	356	557	677	48	2.00E-90	284
Migut.E00173.1	StHe1G2B1_57706	86.14	339	44	3	341	1050	43	0	611
(MgMenB)	StHe51GB1_9027	86.14	339	44	3	341	51	1058	0	611
	StHe61GB1_40383	86.14	339	44	3	341	100	1107	0	611
	StHe3G2B1_68039	86.14	339	44	3	341	93	1100	0	611
	StHeBC3_1263.1	86.14	339	44	3	341	196	1203	0	611

Table S2.3 (continued). TBLASTN results of PhQ biosynthetic genes against PPGP database for *S. hermonthica* (visited December 16, 2015)

query ID	hit ID	identity	align len	mis-match	query start	query end	hit start	hit end	E-value	bit score
Migut.B00584.1	StHeBC3_8351.1	79.15	307	64	88	394	4	924	5.00E-157	453
(MgMenA1)	StHe4GB1_96328	80.15	262	52	133	394	796	11	9.00E-131	383
	StHe61GB1_7666	80.15	262	52	133	394	2	787	5.00E-130	383
	StHe3G2B1_30523	79.77	262	53	133	394	2	787	1.00E-129	382
	StHe1GB1_74001	79.39	262	54	133	394	906	121	4.00E-129	380
Migut.B01155.1	StHeBC3_8351.1	69.51	305	93	108	412	4	918	7.00E-144	421
(MgMenA2)	StHe4GB1_96328	70.38	260	77	153	412	796	17	8.00E-122	361
	StHe61GB1_7666	70.38	260	77	153	412	2	781	3.00E-121	361
	StHe3G2B1_30523	70	260	78	153	412	2	781	5.00E-121	360
	StHe1GB1_74001	69.62	260	79	153	412	906	127	2.00E-120	359
Migut.B01157.1	StHeBC3_8351.1	64.54	282	98	108	387	4	849	7.00E-124	369
(MgMenA3)	StHeBC3_28148.1	65.14	218	75	89	305	36	689	5.00E-102	308
	StHe4GB1_96328	64.56	237	82	153	387	796	86	5.00E-102	310
	StHeBC3_28148.2	68.45	206	65	100	305	30	647	1.00E-101	307
	StHe61GB1_7666	64.56	237	82	153	387	2	712	2.00E-101	310
Migut.E00183.1	StHe61GB1_10274	81.01	258	49	1	258	47	820	6.00E-153	436
(MgMenG)	StHe62GB1_10982	81.25	256	48	3	258	915	148	5.00E-152	433
	StHe1GB1_13566	81.01	258	49	1	258	939	166	1.00E-151	434
	StHeBC3_13089.1	81.4	258	48	1	258	61	834	2.00E-150	437
	StHe51GB1_7807	82.79	244	42	15	258	16	747	1.00E-147	422

† protein length is 583 aa for MgICS1, 582 aa for MgICS2, 1637 aa for MgPHYLLO, 557 aa for MgMenE, 341 aa for MgMenB, 394 aa for MgMenA1, 414 aa for MgMenA2, 397 aa for MenA3, and 260 aa for MgMenG

§ records highlighted in red represent contigs reconstructed to their full length

Table S2.4. TBLASTN results of PhQ biosynthetic genes against PPGP database for *T. versicolor* (visited December 16, 2015)

query ID	hit ID	identity	align len	mis-match	query start	query end	hit start	hit end	E-value	bit score
Migut.I00130.1	TrVe62GB1_40684	80.64	470	80	113	573	1552	149	0	775
(MgICS1) †	TrVeBC3_9373.1	85.86	396	56	178	573	1192	5	0	710
	TrVeGnuB1_74381	85.44	364	53	210	573	1456	365	0	647
	TrVe1GB1_29254	85.04	361	54	213	573	1087	5	0	639
	TrVe2GB1_33075	84.38	333	52	199	531	1000	2	0	590
Migut.I00129.1	TrVe62GB1_40684	80.72	472	84	112	576	1552	137	0	781
(MgICS2)	TrVeBC3_9373.1	86.11	396	55	177	572	1192	5	0	712
	TrVeGnuB1_74381	85.05	368	55	209	576	1456	353	0	653
	TrVe1GB1_29254	85.32	361	53	212	572	1087	5	0	642
	TrVe2GB1_33075	84.68	333	51	198	530	1000	2	0	593
Migut.L01140.1	TrVeBC3_4440.4 §	70.93	1703	392	1	1636	72	5072	0	2365
(MgPHYLLO)	TrVeBC3_4440.3	70.93	1703	392	1	1636	72	5072	0	2365
	TrVeBC3_4440.2	70.93	1703	392	1	1636	72	5072	0	2365
	TrVeBC3_4440.1	70.93	1703	392	1	1636	72	5072	0	2365
	TrVe2GB1_42461	72.51	291	75	1048	1337	862	2	4.00E-137	432
Migut.H01327.1	TrVeBC3_12294.1	76.01	567	119	1	557	222	1901	0	890
(MgMenE)	TrVe2GB1_54642	75.84	567	120	1	557	63	1742	0	887
	TrVe61GB1_51874	76.33	507	104	60	557	1632	133	0	800
	TrVe1GB1_51969	74.63	469	102	99	557	1693	308	0	726
	TrVe1GB1_51968	73.83	428	95	99	516	1760	498	0	653
Migut.E00173.1	TrVe2GB1_15810	83.08	331	54	1	331	988	2	0	574
(MgMenB)	TrVe41GB1_22856	72.11	337	80	7	341	1144	170	4.00E-178	506
	TrVe63GB1_24197	80.14	282	54	1	282	123	962	5.00E-167	475
	TrVe63GB1_24198	82.05	273	47	10	282	35	847	5.00E-165	469
	TrVe0GB1_19681	82.03	256	44	14	269	762	1	6.00E-155	442
Migut.B00584.1	TrVe0GB1_15679	68.67	399	108	10	394	1350	163	1.00E-176	508
(MgMenA1)	TrVe62GB1_22684	76.09	230	55	165	394	695	1384	2.00E-138	328
	TrVe62GB1_22684	59.88	172	48	10	165	148	648	2.00E-138	185
	TrVeBC3_10305.2	75.65	230	56	165	394	907	218	1.00E-136	327
	TrVeBC3_10305.2	59.2	174	48	10	165	1460	954	1.00E-136	180
	TrVe61GB1_16513	75.65	230	56	165	394	750	1439	1.00E-136	327
	TrVe61GB1_16513	59.2	174	48	10	165	197	703	1.00E-136	180
	TrVeBC3_10305.1	67.97	256	56	165	394	985	218	2.00E-132	313
	TrVeBC3_10305.1	59.2	174	48	10	165	1538	1032	2.00E-132	180

Table S2.4 (continued). TBLASTN results of PhQ biosynthetic genes against PPGP database for *T. versicolor* (visited December 16, 2015)

query ID	hit ID	identity	align len	mis-match	query start	query end	hit start	hit end	E-value	bit score
Migut.B01155.1	TrVe0GB1_15679	56.73	416	155	3	412	1359	169	4.00E-144	426
(MgMenA2)	TrVe63GB1_52709	69.6	273	83	100	372	109	927	8.00E-136	399
	TrVe62GB1_22684	68.42	228	72	185	412	695	1378	4.00E-113	285
	TrVe62GB1_22684	43.39	189	82	3	185	139	648	4.00E-113	144
	TrVeBC3_10305.2	67.98	228	73	185	412	907	224	7.00E-113	284
	TrVeBC3_10305.2	43.68	190	82	2	185	1466	954	7.00E-113	144
	TrVe61GB1_16513	67.98	228	73	185	412	750	1433	7.00E-113	284
	TrVe61GB1_16513	43.68	190	82	2	185	191	703	7.00E-113	144
Migut.B01157.1	TrVe63GB1_52709	69.52	269	82	100	368	109	915	2.00E-132	389
(MgMenA3)	TrVe0GB1_15679	52.93	393	158	3	387	1359	238	1.00E-126	380
	TrVeBC3_10305.2	66.85	184	61	185	368	907	356	3.00E-99	239
	TrVeBC3_10305.2	43.68	190	82	2	185	1466	954	3.00E-99	144
	TrVe61GB1_16513	66.85	184	61	185	368	750	1301	3.00E-99	239
	TrVe61GB1_16513	43.68	190	82	2	185	191	703	3.00E-99	144
	TrVe62GB1_22684	66.85	184	61	185	368	695	1246	3.00E-99	239
	TrVe62GB1_22684	43.39	189	82	3	185	139	648	3.00E-99	144
Migut.E00183.1	TrVeRBC1_91	74.73	186	45	30	213	73	630	6.00E-98	292
(MgMenG)	TrVe62GB1_39955	88.97	136	15	47	182	147	554	3.00E-81	248
	TrVe63GB1_10227	77.08	144	31	30	171	1	432	1.00E-75	233
	TrVe2GB1_17652	77.08	144	31	30	171	434	3	3.00E-75	233
	TrVe3GB1_45744	77.24	145	31	115	257	438	4	3.00E-73	227

† protein length is 583 aa for MgICS1, 582 aa for MgICS2, 1637 aa for MgPHYLLO, 557 aa for MgMenE, 341 aa for MgMenB, 394 aa for MgMenA1, 414 aa for MgMenA2, 397 aa for MenA3, and 260 aa for MgMenG

§ records highlighted in red represent contigs reconstructed to their full length

Supplemental File S2.1

Case 1. Out of the 20 genes (>1000 FPKM) not fully assembled by PLAS, 10 genes were not retrieved due to the presence of paralogs with high nucleotide (nt) and amino acid (aa) identities.

(1) Potri.002G194900 was absent in both PLAS and Trinity assemblies.

Potri.002G194900 and Potri.002G194800 share 98% nt identity and 97% aa identity. The expression value of Potri.002G194900 (FPKM = 920.208) is lower than that of Potri.002G194800 (FPKM = 7381.31).

(2) Potri.002G202200 was absent in both PLAS and Trinity assemblies.

Potri.002G202200 and Potri.002G202100 share 99% nt identity and 98% aa identity. The expression value of Potri.002G202200 (FPKM = 1860.14) is lower than that of Potri.002G202100 (FPKM = 4389.03).

(3) Potri.004G045000 was absent in both PLAS and Trinity assemblies.

Potri.004G045000 and Potri.004G044900 share 99% nt identity and 100% aa identity. The expression value of Potri.004G045000 (FPKM = 3718.78) is lower than that of Potri.004G044900 (FPKM = 7556.81).

(4) Potri.007G073700 was absent in both PLAS and Trinity assemblies.

Potri.007G073700 and Potri.007G073800 share 99.92% nt identity and 100% aa identity. The expression value of Potri.007G073700 (FPKM = 1250.3) is lower than that of Potri.007G073800 (FPKM = 1811.88).

(5) Potri.008G150800 was absent in both PLAS and Trinity assemblies.

Potri.008G150800 and Potri.008G150700 share 98% nt identity and 97% aa identity. The expression value of Potri.008G150800 (FPKM = 4305.26) is higher than that of Potri.008G150700 (FPKM = 917.19).

(6) Potri.012G010900 was absent in PLAS assembly, but present in Trinity assembly.

Potri.012G010900 and Potri.012G005300 share 100% nt identity and 100% aa identity. They are identical sequences.

(7) Potri.012G112800 was absent in both PLAS and Trinity assemblies.

Potri.012G112800 and Potri.012G114900 share 99% nt identity and 99% aa identity. The expression value of Potri.012G112800 (FPKM = 1730.01) is lower than that of Potri.012G114900 (FPKM = 5149.88).

(8) Potri.013G030000 was absent in both PLAS and Trinity assemblies.

Potri.013G030000 and Potri.013G030200 share 99% nt identity and 100% aa identity. The expression value of Potri.013G030000 (FPKM = 3270.49) is lower than that of Potri.013G030200 (FPKM = 4159.58).

(9) Potri.019G067100 was absent in both PLAS and Trinity assemblies.

Potri.019G067100 and Potri.019G067200 share 99% nt identity and 99% aa identity. The expression value of Potri.019G067100 (FPKM = 2677.64) is lower than that of Potri.019G067200 (FPKM = 8928.91).

(10) Potri.T130300 was absent in both PLAS and Trinity assemblies.

Potri.T130300 and Potri.018G005100 share 99% nt identity and 99% aa identity. The expression value of Potri.T130300 (FPKM = 1525.48) is lower than that of Potri.018G005100 (FPKM = 5568.38).

Case 2. Out of the 20 genes (>1000 FPKM) not fully assembled by PLAS, 5 genes not recovered due to gene model mis-annotation or alternative splicing events.

(1) Potri.001G469000. One region of the gene was skipped by the PLAS assembled transcript.

```
Potri.001G469000      ------------------------------------------------------------
c11646_g1_i1          TGTCCACACCAAAAATATAGAAGTAAGATTTGACTAAAAATAAATGGAATAGGCAAAAAA

Potri.001G469000      ------------------------------------------------------------
c11646_g1_i1          AACAGAGTTTAAAAACAGGGTAAAACAGAGCATACTGCAATTCTACTGTGAAAGGAAATT

Potri.001G469000      -------------------------ATGTTAGCTTTGTT--CCTAGAATCATTCTTATC
c11646_g1_i1          TCGCAATTGAGAGAGAAACAATGAGGCTCTTAAGGTTGTCTCCCTTAGCTCTATC-TGTC
                                              * ***   ****   ***    **  ** * **
```

```
Potri.001G469000      CATCAAAATTGAGGATTCCATGGCTGTCCACACCAAAAATATAGAATCTTACATAGTTTA
c11646_g1_i1          CTCCTCCTTTATCTCTTTCAGACTTGTGCTG-CCAA------AAAGTCTTACATAGTTTA
                      *  *    **     ** **    *** *   ****      * * **************

Potri.001G469000      CATGGGAGAGTCTTCGTTTTCCCCATTATCATCAACTGGAGAGAGTTC---TTCAGAATT
c11646_g1_i1          CATGGGAGAGTCTTCGTTTTCCCCATTATCATCAACTGGAGAGAGTTCTTCTTCAGAATT
                      ********************************************   *********

Potri.001G469000      AGATGTGCAGCACATGACCAAATCACACTTTGATCTTCTTGGATCATGCTTGGAGAGCAA
c11646_g1_i1          AGATGTGCAGCACATGACCAAATCACACTTTGATCTTCTTGGATCATGCTTGGAGAGCAA
                      ************************************************************

Potri.001G469000      AGAGAATGTCCAGGATGTGATGATTTACTCTTACACTAAATGTATCAATGGTTTTGCTGC
c11646_g1_i1          AGAGAATGTCCAGGATGTGATGATTTACTCTTACACTAAATGTATCAATGGTTTTGCTGC
                      ************************************************************

Potri.001G469000      AAACCTCAATGAAGCTCAAGTAGCTGCCATGAAAGGTAATCCGGGAGTGATATCGGTTTT
c11646_g1_i1          AAACCTCAATGAAGCTCAAGTAGCTGCCATGAAAGGTAATCCGGGAGTGATATCGGTTTT
                      ************************************************************

Potri.001G469000      CGAGAACAAAGAAAGAATGTTGCACACAACACATTCATGGGAATTTATGGGATTTGAAGC
c11646_g1_i1          CGAGAACAAAGAAAGAATGTTGCACACAACACATTCATGGGAATTTATGGGATTTGAAGC
                      ************************************************************

Potri.001G469000      AAATGGAGCACCTACCCTCAGCTCGCTTCAGAAGAAGGCAAACTTTGGTGAAGGTGTAAT
c11646_g1_i1          AAATGGAGCACCTACCCTCAGCTCGCTTCAGAAGAAGGCAAACTTTGGTGAAGGTGTAAT
                      ************************************************************

Potri.001G469000      CATTGCAAATCTTGATACCGGTAAGGTTCTGTCCTTGAAGCTCCAAGGGAAAAATCTTAA
c11646_g1_i1          CATTGCAAATCTTGATACCG----------------------------------------
                      ********************

Potri.001G469000      CAGTGTTCATATTGGTTCTTTGCCAATTGTAATACTCTCTTACATTTTTTGGCTGCGTAC
c11646_g1_i1          ------------------------------------------------------------

Potri.001G469000      TATCACTATAGGTGTATGGCCAGAATCCAAGAGCTTTAATGATGAGGGAATGGGCCCTGT
c11646_g1_i1          ----------GTGTATGGCCAGAATCCAAGAGCTTTAATGATGAGGGAATGGGCCCTGT
                             ************************************************

Potri.001G469000      GCCATCAAGATGGAAGGGAACTTGTCAGGCTGGAGGTGGATTTAAGTGCAACAAAAAGCT
c11646_g1_i1          GCCATCAAGATGGAAGGGAACTTGTCAGGCTGGAGGTGGATTTAAGTGCAACAAAAAGCT
                      ************************************************************

Potri.001G469000      TATTGGTGCAAGGTACTTCAACAAAGGTTTTGCTTCTGCAAGTCCTACTCCAATCCCTAC
c11646_g1_i1          TATTGGTGCAAGGTACTTCAACAAAGGTTTTGCTTCTGCAAGTCCTACTCCAATCCCTAC
                      ************************************************************

Potri.001G469000      AGAGTGGAATACTGCCCGTGACACCGAGGGCCACGGTTCCCACACCTTATCTACAGCTGG
c11646_g1_i1          AGAGTGGAATACTGCCCGTGACACCGAGGGCCACGGTTCCCACACCTTATCTACAGCTGG
                      ************************************************************

Potri.001G469000      TGGTAGCTTTGTTCCTGGAGCGAGTATTTTTGGTTATGGGAATGGAACTGCTAAAGGTGG
c11646_g1_i1          TGGTAGCTTTGTTCCTGGAGCGAGTATTTTTGGTTATGGGAATGGAACTGCTAAAGGTGG
                      ************************************************************

Potri.001G469000      ATCGCCAAAAGCTCACGTGGCTGCTTACAAAGTATGCTGGCCTTCAGACAATGGTGGATG
c11646_g1_i1          ATCGCCAAAAGCTCACGTGGCTGCTTACAAAGTATGCTGGCCTTCAGACAATGGTGGATG
                      ************************************************************

Potri.001G469000      CTTTGATGCAGATATCTTGGCAGCTTTTGATGCTGCCATAGGTGATGGAGTTGATGTGAT
c11646_g1_i1          CTTTGATGCAGATATCTTGGCAGCTTTTGATGCTGCCATAGGTGATGGAGTTGATGTGAT
                      ************************************************************

Potri.001G469000      CTCAATGTCGTTGGGTCCACATCAAGCTGTAGAATTTTTGCAAGATGGAATGGCCATTGG
c11646_g1_i1          CTCAATGTCGTTGGGTCCACATCAAGCTGTAGAATTTTTGCAAGATGGAATGGCCATTGG
                      ************************************************************

Potri.001G469000      CTCCTTCAATGCCATCAAGAAAGGCATTCCCGTCGTTGCTTCAGCCGGCAATAGCGGACC
c11646_g1_i1          CTCCTTCAATGCCATCAAGAAAGGCATTCCCGTCGTTGCTTCAGCCGGCAATAGCGGACC
                      ************************************************************
```

```
Potri.001G469000      AGTTGCTGGGTCCGTGGCTCATGGGGCTCCCTGGTTGTTTACGATTGGTGCTAGTACACT
c11646_g1_i1          AGTTGCTGGGTCCGTGGCTCATGGGGCTCCCTGGTTGTTTACGATTGGTGCTAGTACACT
                      ************************************************************

Potri.001G469000      GGATCGTGAGTTTTCAGCCACTGTTACTCTTGGCAACAAGAAGTTTTTCAAGGGATCAAG
c11646_g1_i1          GGATCGTGAGTTTTCAGCCACAGTTACTCTTGGCAACAAGAAGTTTTTCAAGGGATCAAG
                      ******************** ***************************************

Potri.001G469000      TGTTGCAAGTAAAGGCTTACCAGCTGGGAAATTCTATCCATTGATCAATGCCGCAGAAGC
c11646_g1_i1          TGTTGCAAGTAAAGGCTTACCAGCTGGGAAATTCTATCCATTGATCAATGCCGCAGAAGC
                      ************************************************************

Potri.001G469000      AAGGCTTCCTACAGCACCAGCTGCAGATGCTCAGCTATGCCAAAATGGAACACTTGATCC
c11646_g1_i1          AAGGCTTCCTACAGCACCAGCTGCAGATGCTCAGCTATGCCAAAATGGAACACTTGATCC
                      ************************************************************

Potri.001G469000      CAAGAAGGTTGCAGGGAAAATTATAGTATGCCTTCGAGGAATAAACAGTAGAGTAGTAAA
c11646_g1_i1          CAAGAAGGTTGCAGGGAAAATTATAGTATGCCTTCGAGGAATAAACAGTAGAGTAGTAAA
                      ************************************************************

Potri.001G469000      AGGACATGAGGCTGAGCTTGCTGGTGCCGTTGGGATGATATTGGCAAATGATGAAGAAAG
c11646_g1_i1          AGGACATGAGGCTGAGCTTGCTGGTGCCGTTGGGATGATATTGGCAAATGATGAAGAAAG
                      ************************************************************

Potri.001G469000      TGGAAGTGAAATTTTGTCCGATCCTCATATGCTCCCTGCTGCCCACCTCACGTTCACTGA
c11646_g1_i1          TGGAAGTGAAATTTTGTCCGATCCTCATATGCTCCCTGCTGCCCACCTCACGTTCACTGA
                      ************************************************************

Potri.001G469000      TGGTCAAGCTGTAATGAACTACATCAAGTCGACCAAAAATCCTACAGCATCAATTAGTCC
c11646_g1_i1          TGGTCAAGCTGTAATGAACTACATCAAGTCGACCAAAAATCCTACAGCATCAATTAGTCC
                      ************************************************************

Potri.001G469000      AGTACATACAGATTTAGGAGTCGTGCCGAATCCTGTGATGGCTGCATTCTCATCAAGGGG
c11646_g1_i1          AGTACATACAGATTTAGGAGTCGTGCCGAATCCTGTGATGGCTGCATTCTCATCAAGGGG
                      ************************************************************

Potri.001G469000      ACCTAGTTTAATTGAGCCAGCAATACTCAAGCCTGATGTCACTGCACCTGGGGTTGATGT
c11646_g1_i1          ACCTAGTTTAATTGAGCCAGCAATACTCAAGCCTGATGTCACTGCACCTGGGGTTGATGT
                      ************************************************************

Potri.001G469000      AATCGCTGCTTACACTGAAGCTCTAGGGCCATCTGAACTACCTTTTGACAAGCGTCGGAC
c11646_g1_i1          AATCGCTGCTTACACTGAAGCTCTAGGGCCATCTGAACTACCTTTTGACAAGCGTCGGAC
                      ************************************************************

Potri.001G469000      ACCTTACATCACCATGTCTGGCACTTCAATGTCATGCCCTCATGTTTCCGGCATTGTTGG
c11646_g1_i1          ACCTTACATCACCATGTCTGGCACTTCAATGTCATGCCCTCATGTTTCCGGCATTGTTGG
                      ************************************************************

Potri.001G469000      CCTCCTTAGAGCTATCCATCCAGATTGGAGTCCAGCTGCTCTTAAATCTGCAATCATGAC
c11646_g1_i1          CCTCCTTAGAGCTATCCATCCAGATTGGAGTCCAGCTGCTCTTAAATCTGCAATCATGAC
                      ************************************************************

Potri.001G469000      AACAGCAAAAACAATATCTAACTCCAAGAAGAGAATACTCGATGCTGATGGCCAACCTGC
c11646_g1_i1          AACAGCAAAAACAATATCTAACTCCAAGAAGAGAATACTCGATGCTGATGGCCAACCTGC
                      ************************************************************

Potri.001G469000      GACACCATTTGCATATGGTGCAGGACATGTGAATCCAAATCGTGCAGCAGATCCTGGCCT
c11646_g1_i1          GACACCATTTGCATATGGTGCAGGACATGTGAATCCAAATCGTGCAGCAGATCCTGGCCT
                      ************************************************************

Potri.001G469000      AGTTTATGACACGAACGAGATTGATTACCTTAACTTCTTATGTGCCCATGGCTATAACAG
c11646_g1_i1          AGTTTATGACACGAACGAGATTGATTACCTTAACTTCTTATGTGCCCATGGCTATAACAG
                      ************************************************************

Potri.001G469000      TACCTTCATAATAGAATTCTCAGGCGTGCCTTATAAATGTCCTGAGAATGCTAGCTTGGC
c11646_g1_i1          TACCTTCATTATAGAATTCTCAGGCGTGCCTTATAAATGTCCTGAGAATGCTAGCTTGGC
                      ********* ****************************************************
```

```
Potri.001G469000      TGAATTCAACTATCCTTCAATCACAGTACCTGATCTCAATGGCCCAGTGACTGTTACTCG
c11646_g1_i1          TGAATTCAACTATCCTTCAATCACAGTACCTGATCTCAATGGCCCAGTGACTGTTACTCG
                      ************************************************************

Potri.001G469000      CCGAGTGAAGAACGTAGGGGCTCCGGGCACATACACAGTCAAAGCTAAGGCACCACCTGA
c11646_g1_i1          CCGAGTGAAGAACGTAGGGGCTCCGGGCACATACACAGTCAAAGCTAAGGCACCACCTGA
                      ************************************************************

Potri.001G469000      GGTTTCAGTGGTTGTTGAACCTTCAAGCTTGGAATTCAAGAAAGCCGGTGAAGAGAAGAT
c11646_g1_i1          GGTTTCAGTGGTTGTTGAACCTTCAAGCTTGGAATTCAAGAAAGCCGGTGAAGAGAAGAT
                      ************************************************************

Potri.001G469000      TTTCAAGGTTACTTTTAAGCCTGTAGTGAATGGAATGCCGAAAGACTACACATTTGGGCA
c11646_g1_i1          TTTCAAGGTTACTTTTAAGCCTGTAGTGAATGGAATGCCGAAAGACTACACATTTGGGCA
                      ************************************************************

Potri.001G469000      CCTTACGTGGTCAGATAGCAACGGCCATCATGTCAAGAGTCCTCTTGTGGTGAAGCATGC
c11646_g1_i1          CCTTACGTGGTCAGATAGCAACGGCCATCATGTCAAGAGTCCTCTTGTGGTGAAGCATGC
                      ************************************************************

Potri.001G469000      GTAG------------------------------------------------------
c11646_g1_i1          GTAGATGTTCATTGTAGATGACAATTTCAGTACACACAACTGTACTTTCATCTATAATCT
                      ****

Potri.001G469000      ------------------------------------------------------------
c11646_g1_i1          TCCCACTGATTCAATTCAATTCAATTTATTTATTTTTTCATGTTAATTTTCCCATACCAT


Potri.001G469000      ------------------------------------------------------------
c11646_g1_i1          GAATTAACAACATTCTGAAGGAGTGGGGGAACTATTGTTCACCCCCGAACTATAAACACA


Potri.001G469000      ------------------------------------------------------
c11646_g1_i1          CACTCATGCGCACAATAGATTTACTGTGCCCATGAGTTTTTTTTTTTAAAAAAAAA
```

(2) Potri.006G128300. An mis-annotation is likely to exist in this gene model.

```
Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          CAAAAACCCTCTTGGTAAAAAGATTCTCCTTTTCAGCTCTAAGATTTGTCTCTCTTCCTC


Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          CGTTGATTGGTAAAGAGGGACAAATATTGCCAGCAATTTTGCCTGAATACAAACGACTAC


Potri.006G128300      ----------------------ATGGCTAAGGAGGTTAGCGGTTCACGTTCTTGGATTGA
c32677_g1_i1          TGCTACTTTGCTTCAAAATTCAATGGCTAAGGAGGTTAGCGGTTCACGTTCTTGGATTGA
                                            ************************************

Potri.006G128300      GGTGGCTCCAGCTCCAATCATTTATCCCCGGAAGCCTTCAAATGCTCCCCGTTTGGAGCC
c32677_g1_i1          GGTGGCTCCAGCTCCAATCATTTATCCCCGGAAGCCTTCAAATGCTCCCCGTTTGGAGCC
                      ************************************************************

Potri.006G128300      GATAGCCGAAGAGGGCCACGAGGAACATGATGAAGATTCACAAGCCTTCCAGTCGTCT--
c32677_g1_i1          GATAGCCGAAGAGGGCCACGAGGAACATGATGAAGATTCACAAGCCTTCCAGTAATCCTC
                      ************************************************  **

Potri.006G128300      ---ATGTAG------------------------------------------------
c32677_g1_i1          CCCATTTGGATCCTTATTTCTTTTTTGGTTCTTTGGGGATTCAACTGATCACTACCTCTG
                         ** * *

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          CATTGCCATGTTGAAGTTCCATTGCAGCACTACTCTGACTTTAGCTGAAAGGCCATGTTA


Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          ATTAATCCAGCTGCCCAATGAGTTCATAAATGCACCAACGCCTAGATATGCAGCCAACAT
```

```
Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          GGTGTAAATCTTGCAGATTAGTCTTTTCTATTTAATTTAGATTATTTAATTATATATTTT

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          TCTTATTTAATATCTATACGTTTATATTTTATTTGGTGTCTTGTTGGTGGATTAAAATTA

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          ACTCATTTTTGCCGCGCATCTGGATCTAGTTAATTAAAACAAAAAAAGGAACCAGGGAGA

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          AGAAGAGAGGCAGCAGAAATAAGTTACAAAAGAGTGGCGAGGATATAAAGCTACATTCTG

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          CAATGTGTTATTTAATTTATATTGTAGACGATAGGTCGTCTATGTAGCCAATTAGAAGAA

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          CACCAACCTCGGTGCCCGAGCTTTCCTTCAAGGCCACACTATTTTAATAAAACAAAAAAC

Potri.006G128300      -------------
c32677_g1_i1          AAAATGACGCGCC

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          CAAAAACCCTCTTGGTAAAAAGATTCTCCTTTTCAGCTCTAAGATTTGTCTCTCTTCCTC

Potri.006G128300      ------------------------------------------------------------
c32677_g1_i1          CGTTGATTGGTAAAGAGGGACAAATATTGCCAGCAATTTTGCCTGAATACAAACGACTAC

Potri.006G128300      ----------------------ATGGCTAAGGAGGTTAGCGGTTCACGTTCTTGGATTGA
c32677_g1_i1          TGCTACTTTGCTTCAAAATTCAATGGCTAAGGAGGTTAGCGGTTCACGTTCTTGGATTGA
                                            ************************************

Potri.006G128300      GGTGGCTCCAGCTCCAATCATTTATCCCCGGAAGCCTTCAAATGCTCCCCGTTTGGAGCC
c32677_g1_i1          GGTGGCTCCAGCTCCAATCATTTATCCCCGGAAGCCTTCAAATGCTCCCCGTTTGGAGCC
                      ************************************************************

Potri.006G128300      GATAGCCGAAGAGGGCCACGAGGAACATGATGAAGATTCACAAGCCTTCCAGTAATCCTC
c32677_g1_i1          GATAGCCGAAGAGGGCCACGAGGAACATGATGAAGATTCACAAGCCTTCCAGTAATCCTC
                      ************************************************************

Potri.006G128300      CCCATTTGGATCCTTATTTCTTTTTTGGTTCTTTGGGGATTCAACTGATCACTACCTCTG
c32677_g1_i1          CCCATTTGGATCCTTATTTCTTTTTTGGTTCTTTGGGGATTCAACTGATCACTACCTCTG
                      ************************************************************

Potri.006G128300      CATTGCCATGTTGAAGTTCCATTGCAGCACTACTCTGACTTTAGCTGAAAGGCCATGTTA
c32677_g1_i1          CATTGCCATGTTGAAGTTCCATTGCAGCACTACTCTGACTTTAGCTGAAAGGCCATGTTA
                      ************************************************************

Potri.006G128300      ATTAATCCAGCTGCCCAATGAGTTCATAAATGCACCAACGCCTAGATATGCAGCCAACAT
c32677_g1_i1          ATTAATCCAGCTGCCCAATGAGTTCATAAATGCACCAACGCCTAGATATGCAGCCAACAT
                      ************************************************************

Potri.006G128300      GGTGTAAATCTTGCAGATTAGTCTTTTCTATTTAATTTAGATTATTTAATTATATATTTT
c32677_g1_i1          GGTGTAAATCTTGCAGATTAGTCTTTTCTATTTAATTTAGATTATTTAATTATATATTTT
                      ************************************************************

Potri.006G128300      TCTTATTTAATATCTATACGTTTATATTTTATTTGGTGTCTTGTTGGTGGATTAAAATTA
c32677_g1_i1          TCTTATTTAATATCTATACGTTTATATTTTATTTGGTGTCTTGTTGGTGGATTAAAATTA
                      ************************************************************
```

```
Potri.006G128300      ACTCATTTTTGCCGCGCATCTGGATCTAGTTAATTAAAACAAA-AAAGGAACCAGGGAGA
c32677_g1_i1          ACTCATTTTTGCCGCGCATCTGGATCTAGTTAATTAAAACAAAAAAAGGAACCAGGGAGA
                      ****************************************** ***************

Potri.006G128300      AGAAGAGAGGCAGCAGAAATAAGTTACAAAAGAGTGGCGAGGATCTAAAACTACGTTCTG
c32677_g1_i1          AGAAGAGAGGCAGCAGAAATAAGTTACAAAAGAGTGGCGAGGATATAAAGCTACATTCTG
                      ****************************************** **** **** *****

Potri.006G128300      CAATGTGTTATTTAATTTATATTGTAGACGATAGGTCGTCTATGTAGCCAATTAGAAGAA
c32677_g1_i1          CAATGTGTTATTTAATTTATATTGTAGACGATAGGTCGTCTATGTAGCCAATTAGAAGAA
                      ************************************************************

Potri.006G128300      CACCAACCTCGGTGCCCGAGCTTTCCTTCAAGGCCACACTATTTTAATAAAACAAAAAAC
c32677_g1_i1          CACCAACCTCGGTGCCCGAGCTTTCCTTCAAGGCCACACTATTTTAATAAAACAAAAAAC
                      ************************************************************

Potri.006G128300      AAAATGACGCTCAAA
c32677_g1_i1          AAAATGACGCGCC--
                      ********** *
```

(3) Potri.006G219700. The start position of this gene model is likely to be wrong.

```
Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       TTCCAACCAACCATTATATGTATATATATATAACTAGCAACTGAATTCTTTCCCACACAT

Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       CACATTTATAGCTAGCTCAAATAATTAACCATCTTCTAAGAGATCCAAGAGCCTAACCGC

Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       CTCTTTCCTTCGCTTTATATATAAACCCCGCAGCATTTCTAGCAAAAACACATCCAATTC

Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       TCTCATTCTCTATAAATCATTCCCTTTAATATTTCTCTACTTGTTCTTGGATTCTCTAAT

Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       AGGCTCTTGGATTCTGAGTTTGAATTTGTTTTGTTTTTGCCTCATGGGTACCTCGATGAG

Potri.006G219700      ----------------------------------ATGGCAATCAAGAAATGGCGCGTGC
c335_g1_i1.2858       ATCATGGATGGCCATTTTGATCATGGTTTGGCTTACTCTCGTTCAAGGAATTT-TTGTGG
                                                       *  *  ***** ***     ***

Potri.006G219700      ATGTGAAAGATTGTTCTACCTTCTACGATGCCCTTACTAAGTCTATTATTTTCCTAGAAG
c335_g1_i1.2858       CTGTTGATGCTACTTTTAACTACAAGGATGCCCTTACTAAGTCTATTATTTTCCTAGAAG
                       ***  * * *  ** ** ** * * **********************************

Potri.006G219700      CACAAAGATCAGGAAAACTTCCTCCAAACCACAGGCCACAATGGAGAGGAGATTCTGGCC
c335_g1_i1.2858       CACAAAGATCAGGAAAACTTCCTCCAAACCATAGGCCACAATGGAGAGGAGATTCTGGCC
                      ****************************** **************************

Potri.006G219700      TCGACGATGGTAAACTTGCAAATGTGGACCTTGTTGGGGGGATATTATGATGCAGGAGACA
c335_g1_i1.2858       TCGACGATGGTAAACTTGCAAATGTGGACCTTGTTGGGGGGATATTATGATGCAGGAGACA
                      ************************************************************

Potri.006G219700      ATGTGAAATATGGACTGCCAATGGCTTTTACTGTTACCACTCTGGCTTGGGGTGCTCTCG
c335_g1_i1.2858       ATGTGAAATATGGACTGCCAATGGCTTTTACTGTTACCACTCTGGCTTGGAGTGCTCTCG
                      ************************************************* *********

Potri.006G219700      CTTATCACAAAGAGCTCCATGCCACAGGCGAGCTGCCCCATGTACGTTCTGCCATTAAAT
c335_g1_i1.2858       CTTATCACAAAGAGCTCCATGCCACAGGCGAGCTGCCCCATGTACGTTCTGCCATTAAAT
                      ************************************************************
```

```
Potri.006G219700      GGGGCACAGATTATTTTCTTAAAGCCAGTTCCAGGAAGAACCGTTTGTACGTGCAGGTGG
c335_g1_i1.2858       GGGGCACAGATTATTTTCTTAAAGCCAGTTCCAGGAAGAACCGTTTGTACGTGCAGGTGG
                      ************************************************************

Potri.006G219700      GAGACCCAGTGCTGGATCATCAATGTTGGGTTAGACCAGAAAATATGAGGACACCAAGAA
c335_g1_i1.2858       GAGACCCAGTGCTGGATCATCAATGTTGGGTTAGACCAGAAAATATGAGGACACCAAGAA
                      ************************************************************

Potri.006G219700      CTGTGTTGAGGATTGATGAGAATAACCCGGGAACAGAGATTGCAGCTGAAACTTCAGCTG
c335_g1_i1.2858       CTGTGTTGAGGATTGATGAGAATAACCCGGGAACAGAGATTGCAGCTGAAACTTCAGCTG
                      ************************************************************

Potri.006G219700      CAATGGCTGCTGCTTCCATTGTTTTTCGACACACTAATCGTACCTATTCCCGTAGACTCC
c335_g1_i1.2858       CAATGGCTGCTGCTTCCATTGTTTTTCGACACACTAATCGTACCTATTCCCGTAGACTCC
                      ************************************************************

Potri.006G219700      TCAACAAAGCCAAGTTGCTGTTTGAATTTGCTAAAACACACAAGAAAACCTTTGATGGAG
c335_g1_i1.2858       TCAACAAAGCCAAGTTGCTGTTTGAATTTGCTAAAACACACAAGAAAACCTTTGATGGAG
                      ************************************************************

Potri.006G219700      AATGCCCATTTTATTGCTCTTTCTCAGGCTACAATGATGAGCTGTTGTGGTCAGCAACAT
c335_g1_i1.2858       AATGCCCATTTTATTGCTCTTTCTCAGGCTACAATGATGAGCTGTTGTGGTCAGCAACAT
                      ************************************************************

Potri.006G219700      GGTTGTACAAGGCCACCACTAAGCCTATGTACTTAAAGTACATCAAAGAAGAAGCCACTA
c335_g1_i1.2858       GGTTGTACAAGGCCACCACTAAGCCTATGTACTTAAAGTACATCAAAGAAGAAGCCACTA
                      ************************************************************

Potri.006G219700      GTGCTGCTGTGGCTGAGTTTAGCTGGGACCTTAAATACGCTGGAGCCCAAGTCCTCCTCT
c335_g1_i1.2858       GTGCTGCTGTGGCTGAGTTTAGCTGGGACCTTAAATACGCTGGAGCCCAAGTCCTCCTCT
                      ************************************************************

Potri.006G219700      CTAAGCTGTATTTTGAGGGAGTGAAGGATTTGGAATCCTATAAGAAAGACGCTGACAGTT
c335_g1_i1.2858       CTAAGCTGTATTTTGAGGGAGTGAAGGATTTGGAATCCTATAAGAAAGACGCTGACAGTT
                      ************************************************************

Potri.006G219700      TTATATGCTCAGTGCTGCCTGGTAGCCCCTTCCATCAAGTATATATCTCTCCTGGTGGTA
c335_g1_i1.2858       TTATATGCTCAGTGCTGCCTGGTAGCCCCTTCCATCAAGTATATATCTCTCCTGGTGGTA
                      ************************************************************

Potri.006G219700      TGATTAACTTGAGAGATGGGGCCAACACTCAATATGTTACCAGCACAGCTTTCTTGTTTA
c335_g1_i1.2858       TGATTAACTTGAGAGATGGGGCCAACACTCAATATGTTACCAGCACAGCTTTCTTGTTTA
                      ************************************************************

Potri.006G219700      GCGTCTACAGTGATATCCTTGCCGAACACAATCAAAAAGTACAGTGTGGAAACCAAGCAT
c335_g1_i1.2858       GCGTCTACAGTGATATCCTTGCCGAACACAATCAAAAAGTACAGTGTGGAAACCAAGCAT
                      ************************************************************

Potri.006G219700      TTGACTCTACCCGCGTCATGGCATTCGCCAAGCAACAGATAGATTACTTGCTAGGGAGCA
c335_g1_i1.2858       TTGACTCTACCCGCGTCATGGCATTCGCCAAGCAACAGATAGATTACTTGCTAGGGAGCA
                      ************************************************************

Potri.006G219700      ACCCTGAAAAAAGATCATATATGGTAGGGTTTGGACACAATCCACCAGTGCAAGCACACC
c335_g1_i1.2858       ACCCTGAAAAAAGATCATATATGGTAGGGTTTGGACACAATCCACCAGTGCAAGCACACC
                      ************************************************************

Potri.006G219700      ATAGAGGCGCTTCTGTTCCAGTGATGTCTACTAATACAATAGTGAACTGTGGCACGAGCT
c335_g1_i1.2858       ATAGAGGCGCTTCTGTTCCAGTGATGTCTACTAATACAATAGTGAACTGTGGCACGAGCT
                      ************************************************************

Potri.006G219700      TTGCTAACTGGTTCAACAAGGATGCACCAAACCCTCATGAACTAACTGGTGCCTTTGTGG
c335_g1_i1.2858       TTGCTAACTGGTTCAACAAGGATGCACCAAACCCTCATGAACTAACTGGTGCCTTTGTGG
                      ************************************************************

Potri.006G219700      GTGGACCTGACCGGTTCGACAACTTTGTTGATAAGCGTTGGGATTCATCTAAAACCGAGC
c335_g1_i1.2858       GTGGACCTGACCGGTTCGACAACTTTGTTGATAAGCGTTGGGATTCATCTAAAACCGAGC
                      ************************************************************

Potri.006G219700      CTTGCACGTACGTTAACTCTATTTCAGTTGGTGTTTTGGCAAAGCTTGCAACAGATGGCC
c335_g1_i1.2858       CTTGCACGTACGTTAACTCTATTTCAGTTGGTGTTTTGGCAAAGCTTGCAACAGATGGCC
                      ************************************************************
```

```
Potri.006G219700      GTGTCTAG----------------------------------------------------
c335_g1_i1.2858       GTGTCTAGTAATCAGTCACTAATCCATTCCATTATCTGTTGTTTAGTGATTGATCATGAG
                      ********

Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       CACTCAAAGTACGTGTGTGTGTGTGTGTGTGTCTATATATATATATAGTCCAAAGACTGA


Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       ATTTTGCATGCCTCAACACTTCTTCTTTTTCCATCAATTCCATTGACACCAGAGAATAAT


Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       ATGATATGTAATATGATTTTCAGTGGTGTTTAGTGTAGGGAAATAAAGCTAGTGTAGTTT


Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       TGCCAGCATTCTATCTGGGTATGTATGTTTTTCCACACAAAGTGGATTAATTTGCAACTA


Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       TGAACAGATACACTGGTTGCCCTAATTGTGAACTATAATTGAATCTTGGCTATGATCATT


Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       TGCGAGGATATGGCACTTCAATTTGTATTTGTAGAACGATGGGTTATATTGTTGAGATAC


Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       ATTACCGGTTTTTATACTTCTTTTCTTTCTTTATCAAACCAAATATTTTACAAGAAAGGG


Potri.006G219700      ------------------------------------------------------------
c335_g1_i1.2858       TTTGGTTTGATAAATCATGTTTGTTAAATCCTAGTCTTTTAGATTCTAAGATGGCATGGA


Potri.006G219700      -----------------------------------
c335_g1_i1.2858       TTATTTCATATGTTTATTAATTTATTTAAAAATTAC
```

(4) Potri.010G110900. The last exon is likely to be mis-annotated.

```
Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       AACCACCAACGGGAAGGGCGTGCATAGCGGCGGGCCACAAAGGATTGTCTAATACGAGGC


Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       TTTTACTGTTTATAGTTTTTCTTGGCCAACATATATTTTGTCTAATACCTGGATTGTTGC


Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       CTCGTTTCACCCGATCAATCTACATGTTGCTTGGATTTTTCAAGCAACATTTATATATTC


Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       CTTGTTCTAGGCCATAAAATCCAGTATTCCGTATCTGTTTGACCTGAACAATAGGCAGTA


Potri.010G110900      --------------ATGGACAACCTTCTTGGCCTTCTCAGAATCCGGGTGAAACGAGGCA
c365_g1_i1.1339       GGAGCTTTACAGACATGGACAACCTTCTTGGCCTTCTCAGAATCCGGGTGAAACGAGGCA
                                    *******************************************

Potri.010G110900      ACAATCTTGCCGTTCGCGATCTTGGTACCAGTGATCCTTATGCTGTCATCACCATGGGAA
c365_g1_i1.1339       ACAATCTTGCCGTTCGCGATCTTGGTACCAGTGATCCTTATGCTGTCATCACCATGGGAA
                      ************************************************************
```

```
Potri.010G110900      AACAGAAATTGAAAACTCGAGTGGTGAAAAAAAACTGCAATCCAGAGTGGAACGAGGAGC
c365_g1_i1.1339       AACAGAAATTGAAAACTCGAGTGGTGAAAAAAAACTGCAATCCAGAGTGGAACGAGGAGC
                      ************************************************************

Potri.010G110900      TTACTCTTTCAATCACAGATCTCAATGTTCCAATCAATTTAACTGTTTTTGACAAAGACA
c365_g1_i1.1339       TTACTCTTTCAATCACAGATCTCAATGTTCCAATCAATTTAACTGTTTTTGACAAAGACA
                      ************************************************************

Potri.010G110900      GATTTACCGTGGATGATAAAATGGGTGAAGCAGAAATAGACATCAAAGCATATATCGCGA
c365_g1_i1.1339       GATTTACCGTGGATGATAAAATGGGTGAAGCAGAAATAGACATCAAAGCATATATCGCGA
                      ************************************************************

Potri.010G110900      GTCTAAAGATGGGATTGCAAAATCTCCCAAACGGTTGTGTGGTCTCAAGAATTAAGCCAA
c365_g1_i1.1339       GTCTAAAGATGGGATTGCAAAATCTCCCAAACGGTTGTGTGGTCTCAAGAATTAAGCCAA
                      ************************************************************

Potri.010G110900      GCCGGAACAACTGCCTTGCTGACGAGAGCTGCGTTGTTTGGGATAACGGCAAAATCCTGC
c365_g1_i1.1339       GCCGAAACAACTGCCTTGCTGACGAGAGCTGCGTTGTTTGGGATAACGGCAAAATCCTGC
                      **** *********************************************************

Potri.010G110900      AAGACATGATTCTCAGATTAAGAAATGTAGAGTCCGGTGAAGTGATGATTCAAATCGAGT
c365_g1_i1.1339       AAGACATGATTCTCAGATTAAGAAATGTAGAGTCCGGTGAAGTGATGATTCAAATCGAGT
                      ************************************************************

Potri.010G110900      GGATGAATGTTCCAGGTTGTCGGGGATTGGAAATTGGAGGTACGAGATAA----------
c365_g1_i1.1339       GGATGAATGTTCCAGGTTGTCGGGGATTGGAAATTGGAGACAGGGAAGCACCATGGAACA
                      ************************************  * *  *  *

Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       GGTCGAAGAGACTCGACTGATCGAATGATTTTCCTAATTTTACCATCCACAGCGGTATCC

Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       TCTATGGGTGAGATGCGACTGCTCATTTATAGCGTCCTAACCGTGAAGATGATGACTGTT

Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       TTCTCTTGAATCAAGAATTACTTGGATATCTACTCCATTAATTTGGTGTCTTCTTTCTTT

Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       CTTGTTTTGGTCGTATTTTAAATAATTTTTTTCGATTTTTATGAATCACATTTTGATTTT

Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       TTTTTCCCAAGGATAATGGCATTCTCGAACTTGGCTTGCATTGTATTATATAAAAACCAA

Potri.010G110900      ------------------------------------------------------------
c365_g1_i1.1339       TAAATTCTTATATAAAATCATATATCGTGAAACAATTATTCACCTGATTTTCTCAAAAAA

Potri.010G110900      ------
c365_g1_i1.1339       AAAAAA
```

(5) Potri.010G140800. The gene model is likely to be mis-annotated.

```
c253_g1_i1.3797             CGCACAACAAAACACAAATACATCAAAATAGAACTCTGAATTATCCCATTATTTTTCTCT
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     ------------------------------------------------------------

c253_g1_i1.3797             AACAGTCTAATTATTCATTGCCGCAGTATATTCTCTAGCTGTACCTTGTTACGTAATCCA
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     ------------------------------------------------------------
```

```
c253_g1_i1.3797             GCTAGTAATGAACAAACCTCTTCCTCTCCCTCGTATATATAAAGAATACCAATAATTTGG
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     ------------------------------------------------------------

c253_g1_i1.3797             CACCAGAACCTTCACAGCCTGAGGCGTAACATTCCGATGTTTCTCTTAATTTCTCTCCTT
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     ------------ACAGCCTGAGGCGGAACATTCCGATGTTTCTCT----TTCTCTCCTT

c253_g1_i1.3797             GCCTTCTAGCTAGCTAGTACTCCTGGTGGAAAAGTGTAAGAGAAATGGAAAATCACTTTC
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     GCCTTCTAGCTAGCTAGTACTCCTGGTGGAAAAGTGTAAGAGAAATGGAAAATCACTTTC

c253_g1_i1.3797             AGGCATCTAATGTTAACAATCAAATATATGGCAACGTTGGTGATTCGAAGTTAGAGAAAA
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AGGCATCTAATGTTAACAATCAAATATATGGCAACGTTGGTGATTCGAAGTTAGAGAAAA

c253_g1_i1.3797             GGTGTAGACACCACATACCTCAATCTTGCATGCTATCATCCTCACCAATGCCACCAGTAC
Potri.010G140800            ----------------------------ATGCTATCATCCTCACCAATGCCACCAGTAC
Potri.010G140800_genome     GGTGTAGACAC-ACATACCTCAATCTTGCATGCTATCATCCTCACCAATGCCACCAGTAC
                                                         ******************************

c253_g1_i1.3797             TATCTTCATGGAATGTTCATAGCGATCATAGCATATATAAATCAAAACGGAGTTTCAACG
Potri.010G140800            TATCTTCATGGAATGTTCATAGCGATCATAGCATATATAAATCAAAACGGAGTTTCAACG
Potri.010G140800_genome     TATCTTCATGGAATGTTCATAGCGATCATAGCATATATAAATCAAAACGGAGTTTCAACG
                            ************************************************************

c253_g1_i1.3797             ACTCGGCTGAAGCCAAGAGACAAAAGAGAGTTATGAAGTATAAGGCCTATGCTGTTGAAG
Potri.010G140800            ACTCGGCTGAAGCCAAGAGACAAAAGAGAGTTATGAAGTATAAGGCCTATGCTGTTGAAG
Potri.010G140800_genome     ACTCGGCTGAAGCCAAGAGACAAAAGAGAGTTATGAAGTATAAGGCCTATGCTGTTGAAG
                            ************************************************************

c253_g1_i1.3797             GGAAAATGAAGACCTCTTTCAGGAATGGGATACGTTGGGTCAAGGACAAGTATTGTTCAC
Potri.010G140800            GGAAAATGAAGACCTCTTTCAGGAATGGGATACGTTGGGTCAAGGACAAGAGGTCATGGC
Potri.010G140800_genome     GGAAAATGAAGACCTCTTTCAGGAATGGGATACGTTGGGTCAAGGACAAGTATTGTTCAC
                            ********************************************   *  *  *

c253_g1_i1.3797             ----TTGTGCATAGATATTGATTGACTATGTGAAAACATGAATTTATC------TGTGTT
Potri.010G140800            TCAAAATTGCAA--------AACGCCTTTTTCGAAGGAGGAAATTCTTGAGCTTTGAGCT
Potri.010G140800_genome     ----TTGTGCATAGATATTGATTGACTATGTGAAAACATGAATTTATC------TGTGTT
                                  ****         *  * ** * *  **  * *** ** *       ** * *

c253_g1_i1.3797             TCTTGGATATATATAGAATTTTCTTCCCTTATGAACAATATTTAAGGTTTTTGGTTGTTC
Potri.010G140800            TCCAGGA----------AT------------------------------TTGA------
Potri.010G140800_genome     TCTTGGATATATATAGAATTTTCTTCCCTTATGAACAATATTTAAGGTTTTTGGTTGGTC
                            **  ***          **                                 ***

c253_g1_i1.3797             TGATGATGATGATTTAAATATTGTGGTCACTCAATATGTATGCCTATAATTCTTCGATGC
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     TGATG---ATGATTTAAATATTGTGGTCACTCAATATGTATGCTTATAATTCTTCGATGC

c253_g1_i1.3797             TTGGTTTGTCCACGCAAAAACCTTGGAGAGAATGATTAAAAGAGCTTTGTCAAGGAAATA
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     TTGGTTTGTCCACGCAAAAACCTTGGAGAGAATGATTAAAAGAGCTTTGTCACGGAAATA

c253_g1_i1.3797             TATGGGAACCAATTCTCTTTTCAAAAACGTTACCTGCTCGTGAAATCTCAGATCGACTAG
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     TATGGGAACTAATTCTCTTTTCATAAACGCTGCCTGCTCGTGAAATCTCAGATCGACTAG

c253_g1_i1.3797             AACATCAACGACTCCTCTAGTTCAATTATATAGCTCTGACCTATGGTGATGTTATTTATT
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AACATCAACCACTCCTCTAGTTCAAATATATAGCTTTGACCTATGGTGATGTTATTTATT
```

```
c253_g1_i1.3797             TTTTTGGCAGGAGGTCATGGCTCAAAATTGCAAAACGCCGTACGCATACAAATATCCAAT
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     TTTTTGGCAGGAGGTCATGGCTCAAAATTGCAAAACGCCGTACGCATACAAATATCCAAT

c253_g1_i1.3797             TATTAAATCTTCATTAATTGAGAC-----------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     TATTAAATCTTCATTAATTGAGACACCTTTTCTTTTCTTTCCCTGGTAGTTTTTCGAAGG

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AGGAAATTCTTGAGCTTTGAGCTTCCAGGAATTTGAGCAATTAATTACCTCGTATTAATT

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AAAAAAAAAACAATAAACTTAACGTTAGCTGGGTCAGAGAGCAAGAAGGTGTAGAATAGA

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     TGTGTGTGCATTAGTCCTGTGAAGATTTTATTTCTCTAGGAGGTTTATTTGATCTTGAAT

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AGAGATAACTAGCCATCCACTTTTCTAGATTCATATCCTTTCTAATGTGAAAAAAATCTA

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AACATTAAATTGCCTGTTTTTATTTTTACATTTTAAAAATGTTTTTGAAAAAATTTGATT

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     TTTTTTTTGTTTTTTTTACTTCAAATTAATATTTTATTAGTGATTTCTAATGTTAAAATA

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AATTTTTTTAAAATAAAAAATATTATTTTAATATATTTTTAAATAAAATTTTTTTAAAAA

c253_g1_i1.3797             ------------------------------------------------------------
Potri.010G140800            ------------------------------------------------------------
Potri.010G140800_genome     AATAACCAAACATCCCAAATACATACTTTAACACTCGAAACCTCACCGGGACAACAAACT

c253_g1_i1.3797             ------------------------------------------------------
Potri.010G140800            ------------------------------------------------------
Potri.010G140800_genome     AGGCTAGCCCATCACCTAAAAACAAAAGGTGGTCTGTCCTCATTCGTTTCATATAA
```

Case 3. Out of the 20 genes (>1000 FPKM) not fully assembled by PLAS, 3 genes were reconstructed as hybrids or chimera of two genes sharing a short stretch of common sequences.

(1) Potri.004G146400 was missed by PLAS assembly. The contig is a hybrid of Potri.004G146400 and Potri.009G108100.

```
Potri.004G146400      ------------------------------------------------------------
c466_g1_i1.2043       GGAAAATTTCGGGCGCCTATAACTGGAGATGTCTTTCCTCAACCTCTCCATGGCTGTAAT

Potri.004G146400      ------------------------------------------------------------
c466_g1_i1.2043       TCTTGCCTTCCTTTTCCCTTGATATTACGTCCTTCTCATCTTCATTCCCTCCTATAAGCA

Potri.004G146400      --------------------------ATGGCGAACCCTCTCAGAGCCTTCGGCTGCTTT
c466_g1_i1.2043       GCCTTGGGCCTTCCAAGGTGGTCATTAATGGCGAACCCTCTCAGAGCCTTCGGCTGCTTT
                                                 *********************************

Potri.004G146400      TTCTTTTTCTCCCTGCTCTTCTCTTCTTTCTCAACCTTCTCTCTTGCCTTAACTGATGCT
c466_g1_i1.2043       TTCTTTTTCTCCCTGCTCTTCTCTTCTTTCTCAACCTTCTCTCTTGCCTTAACTGATGCT
                      ************************************************************

Potri.004G146400      GAAGCATCTTATATTGCTCGTCGCCAGCTCTTGACGTTAAATGAAAATAGTGAGCTTCCT
c466_g1_i1.2043       GAAGCATCTTATATTGCTCGTCGCCAGCTCTTGACGTTAAATGAAAATAGTGAGCTTCCT
                      ************************************************************

Potri.004G146400      CATGAGTTTGAGTATGAGGTCGATGTGAAAATAACCTTCGCAAACCAAAGGCTGAGGAGA
c466_g1_i1.2043       CATGAGTTTGAGTATGAGGTCGATGTGAAAATAACCTTCGCAAACCAAAGGCTGAGGAGA
                      ************************************************************

Potri.004G146400      GCGTATATTGGTCTTCAGGCCTGGAAAAAGGCAATATACTCCGACCCATTTAACACTACT
c466_g1_i1.2043       GCGTATATTGGTCTTCAGGCCTGGAAAAAGGCAATATACTCCGACCCATTTAACACTACT
                      ************************************************************

Potri.004G146400      GGCAATTGGGTTGGCGCCAATGTGTGTGCCTATAATGGTGTGTTTTGTGCACCAGCTCTA
c466_g1_i1.2043       GGCAATTGGGTTGGCGCCAATGTGTGTGCCTATAATGGTGTGTTTTGTGCACCAGCTCTA
                      ************************************************************

Potri.004G146400      GACGACCCCAGTCTGAGCGTTGTGGCAGGTGTTGATCTTAACGGTGCTGACATTGCTGGG
c466_g1_i1.2043       GACGACCCCAGTCTGAGCGTTGTGGCAGGTGTTGATCTTAACGGTGCTGACATTGCTGGG
                      ************************************************************

Potri.004G146400      CACCTTCCAGCTGAATTAGGGCTTATGACAGATGTTGCATTATTCCACATTAACTCTAAC
c466_g1_i1.2043       CACCTTCCAGCTGAATTAGGGCTTATGACAGATGTTGCATTATTCCACATTAACTCTAAC
                      ************************************************************

Potri.004G146400      AGGTTTTGTGGTATCATTCCCGAGAGCTTTTCCAAGCTCACACTCATGTATGAGTTTGAT
c466_g1_i1.2043       AGGTTTTGTGGTATCATTCCCGAGAGCTTTTCCAAGCTCACACTCATGTACGAGTTTGAT
                      ************************************************* *********

Potri.004G146400      GTCAGCAACAACCGCTTTGTTGGTGATTTCCCTTCTGTTGTTCTATCCTGGCCAAGCCTC
c466_g1_i1.2043       GTCAGCAACAACCGCTTTGTTGGTGATTTCCCTTCTGTTGTTCTATCCTGGCCAAGCCTC
                      ************************************************************

Potri.004G146400      AAGTATCTTGACGTCAGATTCAACGATTTCGAAGGTAGTTTGCCTCCAGAACTCTTCAAC
c466_g1_i1.2043       AAGTATCTTGACGTCAGATTCAACGATTTCGAAGGTAGTTTGCCTCCAGAACTCTTCAAC
                      ************************************************************

Potri.004G146400      AAGGAACTCGATGCTTTGTTCTTGAATGACAACCGATTCACATCCACCATCCCGGAGACA
c466_g1_i1.2043       AAGGACCTCGATGCTTTGTTCTTGAATGACAACCGATTCACATCCACCATCCCGGAGACA
                      ***** ********************************************************

Potri.004G146400      ATAGGCAACTCCGCAGTTTCTGTTGTCACATTTGCTAACAACAAATTCACCGGCTGCATT
c466_g1_i1.2043       ATAGGCAACTCCGCAGTTTCTGTTGTCACATTTGCTAACAACAAATTCACCGGCTGCATT
                      ************************************************************

Potri.004G146400      CCACACAGCGTCGGCAAGATGGCCAACTTGAACGAGGTCATCTTTATGGGCAATGACCTT
c466_g1_i1.2043       CCACACAGCGTCGGCAAGATGGCCAACTTGAACGAGGTCATCTTTATGGGCAATGATCTT
                      ********************************************************** ***

Potri.004G146400      GGTGGTTGCTTCCCAGCAGAAATTGGGCTGCTTCGTAATGTGACTGTCTTTGATGCCAGC
c466_g1_i1.2043       GGTGGTTGCTTCCCAGCAGAAATTGGGCTGCTTGGTAATGTGACTGTGTTTGATGCCAGC
                      ********************************** ************* ************

Potri.004G146400      CACAATGGGTTCACAGGAATCTTGCCGCCCAGCTTTGCAGGCCTAAAGAAGGTTGAACTC
c466_g1_i1.2043       CACAATGGGTTCACAGGAATCTTGCCGTCCAGCTTTGCAGGGCTAAAGAAGGTTGAACTC
                      ************************** ************* ******************
```

```
Potri.004G146400      TTGGATCTTGCCGACAACAAGCTGACAGGATTTGTGCctgagaacatttgcaagttgcca
c466_g1_i1.2043       TTGGATCTTGCAGACAACAAGCTGACAGGATTTGTGCCTGAGAACATTTGTAGGTTGTCA
                      ********** ************************************ * **** **

Potri.004G146400      agcttgacaaacttcacattctcgtatAACTACTTCAAGGGCGAGGCTCAAGCTTGCGTG
c466_g1_i1.2043       AGCTTGACGAACTTCACATTCTCGTATAACTACTTCAAGGGGGAGGCTCAAGCTTGCGTG
                      ******** ****************************** ******************

Potri.004G146400      CCTCCATCAAGGAAAGACATTGTGTTGGATGATACCAGCAATTGCCTGTCTGACAGGCCA
c466_g1_i1.2043       CCTCCATCAAGGAAGGACACCGTGTTGGATGATACCAGCAATTGCCTGTCTGACAGGCCA
                      ************** ****  ************************************

Potri.004G146400      AAGCAGAAGTCAGCCAGGACATGTTATCCAGTGGTGAGCCGACCTGTGGATTGCAGCAAG
c466_g1_i1.2043       AAGCAGAAGTCAGCCAGGACATGTTACCCAGTGGTGAGCCGACCTGTGGATTGCAGCAAG
                      ************************* ********************************

Potri.004G146400      GACAAGTTTCACT------------CTccaccaccacctgtcca--gtcaccaccaccac
c466_g1_i1.2043       GACAAGTGTTCTGGAGGAGGAGGTTCTTCAA--ACCCCCATCCAAAACCACAACCCACAC
                      ******* *                ** **    * **  ****    *** ***  ***

Potri.004G146400      ---------cagttcact------ctcccccgccaccc--------gtcca----gtcac
c466_g1_i1.2043       CACCTACTCCAGAACATAAACAAACCCCATCTCCACCTAAATCTACTTCTACTCCAACAC
                             ***  **        * **  * *****          ** *      ***

Potri.004G146400      ctccaccaccagttcactctccaccaccacccgtacactcaccgccaccacctgtccagt
c466_g1_i1.2043       CATCATCACCAATCCCTGCCCCTCGAACACCAG---AATTACCAAAACCAGAACCTAAGT
                      *  ** ***** * *   * ** * * **** *   * * ***   ****       ***

Potri.004G146400      cacctccaccaccagttcac---------tctccaccaccacccgtacactcaccgccac
c466_g1_i1.2043       TACCGCTGGCTCCAGTTGAACCAATTAGTCCATCAACACCAGAGGTATCCTTACCACCAT
                       *** *   * ****** *           *  ** *****   ***  ** *** ***

Potri.004G146400      cac---cagttc---actcaccacc----accacctgtccactctccgccaccaccagtc
c466_g1_i1.2043       CTTTATCAATTAGTCCTTCAACTCCGGAGATATCCTCACCACCATCTTCATCAATTAGTC
                      *     ** **      *** * **    *   ***  ****  **  *  * *  ****

Potri.004G146400      cagtcacccccctccacctgtc------------------------------cactctcca
c466_g1_i1.2043       CATCTACCCCATC-ATCTGACCCATACAATCCAGGACCTGGTGGGCATGACGAGACACCG
                      **   ***** ** * *** *                                 *  * **

Potri.004G146400      cctc------cacccGTACACTCACCTC------------------------CTCCTGTC
c466_g1_i1.2043       CCATCACCAAAATCTGCACCGTCACCTGATTCATTTGATAATTCATCAATTAGGCATGAC
                      **         * * * **  ******                              * ** *

Potri.004G146400      CAGTCACCCCCTCCACCTGTTCACTCTccaccaccaccagtacactcacccccctcc-tcc
c466_g1_i1.2043       AAGACACCACCACCATCATCTGAGATTGCTATACCACCATCATCCTCATTTAGTCCATCA
                       ** **** ** *** *   * *   * *   *******  *  ****     *** **

Potri.004G146400      agttcactca--ccc---tctccacccgtacaatcccccccaccaccacc-agtacac--
c466_g1_i1.2043       ACTTCAGAGAAACCCATGCCTCCATCACCTAATTCTA---CACCATCACCTAATTCATAT
                      * ****   *  ***    ***** *     * **     ***** **** * * **

Potri.004G146400      ------------------------ttacctcctccaccagtacactctccaccaccacat
c466_g1_i1.2043       AACCTAGAACTTGGTGGGCATGATGGACCACCACCATCACTAATGCCTTCAGTAAAGCCC
                                             *** ** *** ** **    ** **  *   *

Potri.004G146400      gttaaatcaccaccaccaccacgaccagtc-----------aaatcatc----tcc---a
c466_g1_i1.2043       GAGACACCAGAACCATCACCACAACCAGAAATACCAATAATAAATCATCATTCTCCTTTA
                      *  * * **  **** ****** *****             ********    ***   *

Potri.004G146400      cttccaattttct---ctccacc---------accaccaACTGTATTTCCTCATCCTCCT
c466_g1_i1.2043       CCTTTGGTTCACTCACCTCCACCACAACCAGAAACACCAATTGTAAATCATCACTCTCCT
                      * *    **  **   *******         * ****** ****  ** ***  *****

Potri.004G146400      CGTGC---T------TTTCCTCCACCACCCCCAAA-TGA-----AGATATAGTCCTTCCA
c466_g1_i1.2043       CAACTTCCAGTTCACTCACCTCCACCATCATCACAATCCCTACCCCCTCTAGTCCATTCA
                      *             *  ********* *  ** * *          * ****** * **
```

```
Potri.004G146400      CCAAACCTCGGATTCCAATA--CGCATCGCCACCTCCACCAGT------GTTCCCAG---
c466_g1_i1.2043       C-----------TTCCACCACCAGCACATTCACCCCCACCATCAATTCATTTCCCACCAC
                      *           *****  *   ***    **** ******         ******

Potri.004G146400      ------GCTACTAG
c466_g1_i1.2043       CACCTATCTATT--
                             *** *
```

```
CLUSTAL O(1.2.4) multiple sequence alignment

Potri.004G146400.1      ATGGCGAACCCTCTCAGAGCCTTCGGCTGCTTTTTCTTTTTCTCCCTGCTCTTCTCTTCT
Potri.009G108100.1      ATGGCTGAACCTCTCAGAGTCTTGGGCTGCTTTTTCTTTTTTTCCTTTCTCTTATCTTCT
                        *****  * ********** *** ***************** *** * ***** ******

Potri.004G146400.1      TTCTCAACCTTCTCTCTTGCCTTAACTGATGCTGAAGCATCTTATATTGCTCGTCGCCAG
Potri.009G108100.1      TTCTCAAACTTCTCTCTTGCCTTAACTGATGCTGAAGCATCTTCTATTGCTCGTCGCCAG
                        ******* ************************************ ****************

Potri.004G146400.1      CTCTTGACGTTAAATGAAAATAGTGAGCTTCCTCATGAGTTTGAGTATGAGGTCGATGTG
Potri.009G108100.1      CTATTGACATTACATGAAAATGGTGAACTTCCCGATGATTTTGAGTATGAGGTGGATGTG
                        ** ***** *** ******** **** *****  **** ************** ******

Potri.004G146400.1      AAAATAACCTTCGCAAACCAAAGGCTGAGGAGAGCGTATATTGGTCTTCAGGCCTGGAAA
Potri.009G108100.1      AAAGAAACCTTTGCAAACCAAAGGCTCAGGAGGGCATATATTGGTCTCCAGGCCTGGAAA
                        ***  ****** ************** ***** ** ************ ************

Potri.004G146400.1      AAGGCAATATACTCCGACCCATTTAACACTACTGGCAATTGGGTTGGCGCCAATGTGTGT
Potri.009G108100.1      AAGGCAATGTACTCCGACCCGTTTAATACAACTGGCAATTGGGTTGGCGCCGATGTGTGT
                        ******** ************ ***** ** ********************** *********

Potri.004G146400.1      GCCTATAATGGTGTGTTTTGTGCACCAGCTCTAGACGACCCCAGTCTGAGCGTTGTGGCA
Potri.009G108100.1      GCTTATAATGGTGTGTTTTGTGCACCGGCTCTTGACGACTCTGGTCTAAGCGTTATGGCA
                        ** ************************ ***** ****** *  **** ****** *****

Potri.004G146400.1      GGTGTTGATCTTAACGGTGCTGACATTGCTGGGCACCTTCCAGCTGAATTAGGGCTTATG
Potri.009G108100.1      GGTGTTGATCTTAACGGTGCTGATATTGCTGGGTACCTTCCAGCTGAATTGGGGCTTTTG
                        ********************** ********* *************** ****** **

Potri.004G146400.1      ACAGATGTTGCATTATTCCACATTAACTCTAACAGGTTTTGTGGTATCATTCCCGAGAGC
Potri.009G108100.1      ACAGATGTTGCATTGTTCCACATTAACTCTAACAGGTTTTGTGGAATCATCCCCAAGAGC
                        ************** ****************************** ***** *** *****

Potri.004G146400.1      TTTTCCAAGCTCACACTCATGTATGAGTTTGATGTCAGCAACAACCGCTTTGTTGGTGAT
Potri.009G108100.1      TTTTCCAAGCTCACACTCATGTACGAGTTTGATGTCAGCAACAACCGCTTTGTTGGTGAC
                        ********************** ************************************

Potri.004G146400.1      TTCCCTTCTGTTGTTCTATCCTGGCCAAGCCTCAAGTATCTTGACGTCAGATTCAACGAT
Potri.009G108100.1      TTCCCTTCTGTTGTTTTAACCTTGCCAAGCCTCAAGTATCTTGACATCAGATTCAATGAT
                        *************** ** *** ********************** ********** ***

Potri.004G146400.1      TTCGAAGGTAGTTTGCCTCCAGAACTCTTCAACAAGGAACTCGATGCTTTGTTCTTGAAT
Potri.009G108100.1      TTCGAAGGTAGTTTGCCTCCAGAACTCTTCAACAAGGACCTCGATGCTTTGTTCTTGAAT
                        ************************************* ********************

Potri.004G146400.1      GACAACCGATTCACATCCACCATCCCGGAGACAATAGGCAACTCCGCAGTTTCTGTTGTC
Potri.009G108100.1      GACAACCGGTTCACATCCACCATTCCGGAGACAATAGGCAACTCCCCAGTTTCTGTAGTC
                        ******** ************** ********************** ********** ***

Potri.004G146400.1      ACATTTGCTAACAACAAATTCACCGGCTGCATTCCACACAGCGTCGGCAAGATGGCCAAC
Potri.009G108100.1      ACATTTGCGAACAACAAATTCACTGGCTGCATTCCACACAGCATCGGCAAGATGACAAAC
                        ******** ************** ****************** *********** * ***
```

```
Potri.004G146400.1      TTGAACGAGGTCATCTTTATGGGCAATGACCTTGGTGGTTGCTTCCCAGCAGAAATTGGG
Potri.009G108100.1      TTGAACGAGGTCATCTTTATGGGCAATGATCTTGGTGGTTGCTTCCCAGCAGAAATTGGG
                        ****************************** ******************************

Potri.004G146400.1      CTGCTTCGTAATGTGACTGTCTTTGATGCCAGCCACAATGGGTTCACAGGAATCTTGCCG
Potri.009G108100.1      CTGCTTGGTAATGTGACTGTGTTTGATGCCAGCCACAATGGGTTCACAGGAATCTTGCCG
                        ****** ************* ************************************

Potri.004G146400.1      CCCAGCTTTGCAGGCCTAAAGAAGGTTGAACTCTTGGATCTTGCCGACAACAAGCTGACA
Potri.009G108100.1      TCCAGCTTTGCAGGGCTAAAGAAGGTTGAACTCTTGGATCTTGCAGACAACAAGCTGACA
                         ************* ****************************** ***************

Potri.004G146400.1      GGATTTGTGCCTGAGAACATTTGCAAGTTGCCAAGCTTGACAAACTTCACATTCTCGTAT
Potri.009G108100.1      GGATTTGTGCCTGAGAACATTTGTAGGTTGTCAAGCTTGACGAACTTCACATTCTCGTAT
                        ********************** * **** ********** ******************

Potri.004G146400.1      AACTACTTCAAGGGCGAGGCTCAAGCTTGCGTGCCTCCATCAAGGAAAGACATTGTGTTG
Potri.009G108100.1      AACTACTTCAAGGGGGAGGCTCAAGCTTGCGTGCCTCCATCAAGGAAGGACACCGT----
                        ************** ******************************** ****  **

Potri.004G146400.1      GATGATACCAGCAATTGCCTGTCTGACAGGCCAAAGCAGAAGTCAGCCAGGACATGTTAT
Potri.009G108100.1      ----------------------------GCCAAAGCAGAAGTCAGCCAGGACATGTTAC
                                                    ******************************

Potri.004G146400.1      CCAGTGGTGAGCCGACCTGTGGATTGCAGCAAGGACAAGTTTCACT------------C-
Potri.009G108100.1      CCAGTGGTGAGCCGACCTGTGGATTGCAGCAAGGACAAGTGTTCTGGAGGAGGAGGTTCT
                        ************************************** *               *

Potri.004G146400.1      -----------TCC---------------------------------------------
Potri.009G108100.1      TCAAACCCCCATCCAAAACCACAACCCACACCACCTACTCCAGAACATAAACAAACCCCA
                              ***

Potri.004G146400.1      --------------------------ACCA-------------------------------
Potri.009G108100.1      TCTCCACCTAAATCTACTTCTACTCCAACACCATCATCACCAATCCCTGCCCCTCGAACA
                                          * **

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      CCAGAATTACCAAAACCAGAACCTAAGTTACCGCTGGCTCCAGTTGAACCAATTAGTCCA

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      TCAACACCAGAGGTATCCTTACCACCATCTTTATCAATTAGTCCTTCAACTCCGGAGATA

Potri.004G146400.1      ---------CCA---------------------------CCTG-----------------
Potri.009G108100.1      TCCTCACCACCATCTTCATCAAGTCCATCTACCCCATCATCTGACCCATACAATCCAGGA
                              ***                               ***

Potri.004G146400.1      -------------TCCAGT------------------------------------------
Potri.009G108100.1      CCTGGTGGGCATGACGAGACACCGCCATCACCAAAATCTGCACCGTCACCTAATCCATTT
                              * **

Potri.004G146400.1      ----------------------------CACCACCACC----------------------
Potri.009G108100.1      AATAATTCACCAGTTGGGCACAATGAGACACCACCATCACCGGAGTCTGCACCGTCACCC
                                           ******** *

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      GATCCATTCAATAATTCACCAGATGGGCATAACGAGACACCGCTATCACCGGAGTCTGCA

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      CCGTCACCCAATCCATTCAATAATTCACCAGATGGGCACGACGAGACACCGCTATCACCG
```

```
Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      GAGTCTGCACCGTCACAATCACCGGAGTCTGCACCGTCACCCGATCCATTTAATAATTCA

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      CCAGATGGGCACGATGAGACACCACAATCATCGGAGTCTGCACTGTCACCCGATCCATTC

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      AATAATTCACCAGATGGGCACGACGAGATACCACCATCACCAGAGCCGTCACCGGATCCA

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      TTTAATAATTCACCGAATGGGCATGATGAGACACCAACATCACCAGAGTCCGCACAATCA

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      CCTGATCCATTTAATAATTCACCAATTGGGCACGACAAGACACCACCACCATCATCTGAG

Potri.004G146400.1      ------------------------------------------------------------
Potri.009G108100.1      ATATCTATACCACCATCACCCTTAATTAGTCCACCAACATCGGAGAAACATATACCACCA

Potri.004G146400.1      ------------------------------------------ACCAGTTCACTCTCC----
Potri.009G108100.1      TCATCAGAGTTTGCTCCATCACCTGATTCATATAATTTACGACCTGTTCACTCACCTCCA
                                                                *** ******** **

Potri.004G146400.1      -------------------------------------CCCGCCACCCGTCCAGTCACCT
Potri.009G108100.1      CCATCATCACAATCCCTACCCCCTCTAGTCTATTCACTTCCACCACCAGCACATTCACCC
                                                             ** ***** *  ** *****

Potri.004G146400.1      CCACCACCAGTTCACTCTCCACCACCACCCGTACACTCA---------------------
Potri.009G108100.1      CCACCATCAATTCATTTCCCACCACCACCTGTACACTCTCCCCCACCACCCCCTGTTTAC
                        ****** ** **** *  *********** ********

Potri.004G146400.1      ------------------------CCGCCACCACCT------------------------
Potri.009G108100.1      TCTCCCCCTCTGCCAGTACATTCACCGCCACCATCAGTGCACTCTCCCCCACCACCAATG
                                             ********* *

Potri.004G146400.1      ------------------------------------------GTCCAGTCACCTCCACCA
Potri.009G108100.1      CACTCTCCCCCACCACCACCAGTTTACTCTCCCCCACCGCCAGTACAATCATTCCCACCA
                                                           ** ** ***   ******

Potri.004G146400.1      CCAGTTCACTCTCCACCACCACCCGTACACTCACCGC---CACCACCAGTTCA-------
Potri.009G108100.1      CCAGTGCACTCTCCCCCACCACCTGTACACTCACCCCCTCCACCACCAGTTTACTCTCCC
                        ***** ******** ******** *********** *   *********** *

Potri.004G146400.1      --------------CTCACCACCACCACCTGTCCACTCTCCGCCACCACCAGTCCAGTCA
Potri.009G108100.1      CCTCCGCCAGTACATTCACCGCCACCACCAGTGTACTCTCCCCCACCGCTGGTACAATCA
                                    ***** ******** **  ******* ***** *  ** ** ***

Potri.004G146400.1      CCCCCTCCACCTGTCCACTCTCCACCTCCACCCGTACACTCACCTCCT------------
Potri.009G108100.1      CCCCCACCACCAGTGCACTCTCCCCCACCACCTTTACACTCACCCCCTCCACCACCAGTT
                        ***** ***** ** ******** ** *****  ********** ***

Potri.004G146400.1      ---------------CCTGTCCAGTCACCCCCTCCACCTGTTCACTCTCCACCACCACCA
Potri.009G108100.1      TACTCTCCCCCTCCGCCAGTACATTCACCGCCACCACCAGTGCACTCTCCCCCACCGCCG
                                    ** ** ** ***** ** ***** ** ******** ***** **
```

```
Potri.004G146400.1      GTACACTCACCCCCTCCTCCAGTT-----------------------CACTCACCCTCT
Potri.009G108100.1      ATACAATCACCCCCACCACCAGTGCACTCACCCCCTCCACCACCTATACACTCACCCCCG
                         **** ******** ** *****                          ********* *

Potri.004G146400.1      CCACCCGTACAATCCCCCCCACCACCACCAGTACACTTACCTCCTCCACCAGTACACTCT
Potri.009G108100.1      CCACCCGTGCAATCTCTCCCTCCACCACCTGTAAACTCACCCCTGCCACCCGTGCACTCC
                        ******** ***** * *** ******** *** *** *** *  ***** ** *****

Potri.004G146400.1      CCACCACCACATGTTAAATCACCACC---------------------------------
Potri.009G108100.1      CCACCACCACCGGTTCATTCTCCTACATCACCCATACACTCCCATCCACCACCTGTAAAC
                        **********  *** * ** **  *

Potri.004G146400.1      --ACCACCACGACCAGTCAAATCATCTCCACTTCCAATTTTCTCTCCACCACCACCAACT
Potri.009G108100.1      TCACCCCCGCCACCCGTGCAATCACCTCCACCTCCAGTTTTCTCTCCACCACCAGTAATT
                          *** ** * *** **  ***** ****** **** ****************  ** *

Potri.004G146400.1      GTATTTCCTCATCCTCCTCGTGCTTTTCCTCCACCACCCCCAAATGAAGATATAGTCCTT
Potri.009G108100.1      GTATCTCCTCC------------------TCCACCTCCCCCGGAAGAAGACTTCATCCTT
                        **** *****                  ****** *****  * *****  *  *****

Potri.004G146400.1      CCACCAAACCTCGGATTCCAATACGCATCGCCACCTCCACCAGTGTTCCCAGGCTACTAG
Potri.009G108100.1      CCACCAAACCTCGGATTCCAATATGCATCACCACCTCCACCAACGTTCCCAGGCTACTAA
                        ********************** ***** ************  ***************
```

(2) Potri.006G192000 was missed by PLAS assembly. The contig is a chimera of Potri.006G192000 and Potri.005G201800 because of ATTATTATTATTATTATTATTAT repetitive sequence.

```
Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           CAAATGGAAGAGCAAAGGTAGGAGGAAGGGAGCGGTCGCGATATTGCTTAAAATACTGGC


Potri.006G192000      -----CCCCTTGTGCCTTTGCTCAAGGGCCGGCTACGCCTATTATTGCACGCCCGCAAGC
c8982_g1_i1           CTCTCCCCCTTGTGCCTTTGCTCAAGGGCCGGCTACGCCTATTATTGCACGCCCGCAAGC
                           **************************************************


Potri.006G192000      ACATAGAAATAGAGAGGATTTTGGGAAAAGGAGAAGCGAGGCTTCAAGAGAAGGAGAGTC
c8982_g1_i1           ACATAGAAATAGAGAGGATTTTGGGAAAAGGAGAAGCGAGGCTTCAAGAGAAGGAGAGTC
                      ************************************************************


Potri.006G192000      ACTCCCTTTCTCCCTTATCAGGTAATAATACATTTCCTCTCCTCTTTAGCTCTCTGCTCT
c8982_g1_i1           ACTCCCTTTCTCCCTTATCA---------------------------------------
                      ********************


Potri.006G192000      GTTTCAATCACCAACCTGCATTTTTGTGTTTGATTCCTTTGCGGGTTCTGCGCGCGAATG
c8982_g1_i1           ------------------------------------------------------------


Potri.006G192000      CTTGTTTTCTTCTCTCCTGGGTCTTAGATATTTCAGATCCAGGTCTTGAAACAATGGGTG
c8982_g1_i1           -----------------------------------------GGTCTTGAAACAATGGGTG
                                                                ******************


Potri.006G192000      CTGGCGGTAGAATGTCTGTTCCTCCTTCCTCTGCTAAGGTGGAATCTGATGTTTTCAAGC
c8982_g1_i1           CTGGCGGTAGAATGTCTGTTCCTCCTTCCTCTGCTAAGGTGGAATCTGATGTTTTCAAGC
                      ************************************************************


Potri.006G192000      GAGCTCCTGACTCGAAGCCGCCATTTACACTCGGCCAGATCAAGAAAGCCATCCCACCTC
c8982_g1_i1           GAGCTCCTGACTCGAAGCCGCCATTTACACTCGGCCAGATCAAGAAAGCCATCCCACCTC
                      ************************************************************
```

```
Potri.006G192000      ATTGTTTCCAGCGTTCTGTTCTTCGCTCATTCTCTTACGTTGCTCATGACCTGATCATTG
c8982_g1_i1           ATTGTTTCCAGCGTTCTGTTCTTCGCTCATTCTCTTACGTTGCTCATGACCTGATCATTG
                      ************************************************************

Potri.006G192000      CCTCTATCTTCTATTATGTTGCGACCAATTACTTCCACCTCCTTCCTCACCCTCTCTCCT
c8982_g1_i1           CCTCTATCTTCTATTATGTTGCGACCAATTACTTCCACCTCCTTCCTCACCCTCTCTCCT
                      ************************************************************

Potri.006G192000      ATGTGGCCTGGCCGATTTATTGGGCTGTCCAGGGATGTGTCCTCACCGGCGTTTGGGTTA
c8982_g1_i1           ATGTGGCCTGGCCGATTTATTGGGCTGTCCAGGGATGTGTCCTCACCGGCGTTTGGGTTA
                      ************************************************************

Potri.006G192000      TAGCTCATGAGTGTGGTCATCATGCCTTTAGCGACTATCAATTGCTTGATGACATCGTTG
c8982_g1_i1           TAGCTCATGAGTGTGGTCATCATGCCTTTAGCGACTATCAATTGCTTGATGACATCGTTG
                      ************************************************************

Potri.006G192000      GCCTTGTCCTCCATTCTTGTCTCCTCGTCCCTTATTTTTCATGGAAACATAGCCATCGTC
c8982_g1_i1           GCCTTGTCCTCCATTCTTGTCTCCTCGTCCCTTATTTTTCATGGAAACATAGCCATCGTC
                      ************************************************************

Potri.006G192000      GCCATCATTCCAACACAGGCTCTCTGGATAGGGATGAAGTGTTTGTACCGAAGAAGAAAT
c8982_g1_i1           GCCATCATTCCAACACAGGCTCTCTGGATAGGGATGAAGTGTTTGTACCGAAGAAGAAAT
                      ************************************************************

Potri.006G192000      CTGGTATCCGTTGGTACTCCAAATACCTTAACAACCCGCTAGGTCGTTTCCTCACCATTA
c8982_g1_i1           CTGGTATCCGTTGGTACTCCAAATACCTTAACAACCCGCTAGGTCGTTTCCTCACCATTA
                      ************************************************************

Potri.006G192000      CCATCACCCTTACTCTTGGCTGGCCTCTTTACCTTGCATTCAATGTTTCAGGCAGACCTT
c8982_g1_i1           CCATCACCCTTACTCTTGGCTGGCCTCTTTACCTTGCATTCAATGTTTCAGGCAGACCTT
                      ************************************************************

Potri.006G192000      ATGATAGGTTTGCTTGCCACTACGATCCATATGGCCCTATCTACAATGATCGTGAGCGTG
c8982_g1_i1           ATGATAGGTTTGCTTGCCACTACGATCCATATGGCCCTATCTACAATGATCGTGAGCGTG
                      ************************************************************

Potri.006G192000      TGGAGATATTTATATCTGATGCTGGTATTCTTGCTGTCACTTACGGGCTCTACCGCCTTG
c8982_g1_i1           TGGAGATATTTATATCTGATGCTGGTATTCTTGCTGTCACTTACGGGCTCTACCGCCTTG
                      ************************************************************

Potri.006G192000      CAGTCGCAAAGGGACTTGGTTGGGTTCTTTGTGTTTATGGAGGGCCATTACTTGTGGTGA
c8982_g1_i1           CAGTCGCAAAGGGACTTGGTTGGGTTCTTTGTGTTTATGGAGGGCCATTACTTGTGGTGA
                      ************************************************************

Potri.006G192000      ATGCATTCCTTGTTCTGATCACATATCTGCAGCATACCCATCCTTCATTGCCGCATTACG
c8982_g1_i1           ATGCATTCCTTGTTCTGATCACATATCTGCAGCATACCCATCCTTCATTGCCGCATTACG
                      ************************************************************

Potri.006G192000      ATTCATCTGAGTGGGACTGGTTAAAAGGGGCTCTAGCAACCGTCGATAGAGATTATGGAA
c8982_g1_i1           ATTCATCTGAGTGGGACTGGTTAAAAGGGGCTCTAGCAACCGTCGATAGAGATTATGGAA
                      ************************************************************

Potri.006G192000      TCTTGAACAAGGTCTTCCATAACATAACAGACACTCATGTAGCTCACCATTTGTTCTCAA
c8982_g1_i1           TCTTGAACAAGGTCTTCCATAACATAACAGACACTCATGTAGCTCACCATTTGTTCTCAA
                      ************************************************************

Potri.006G192000      TGATGCCACACTACCATGCTATGGAGGCAACGAAGGCAATCAAACCAATTTTGGGAGATT
c8982_g1_i1           TGATGCCACACTACCATGCTATGGAGGCAACGAAGGCAATCAAACCAATTTTGGGAGATT
                      ************************************************************

Potri.006G192000      ACTACCAACATGACGGAACTCCAGTCTATAAGGCAACGTGGAGAGAGGCCAAGGAATGCA
c8982_g1_i1           ACTACCAACATGACGGAACTCCAGTCTATAAGGCAACGTGGAGAGAGGCCAAGGAATGCA
                      ************************************************************

Potri.006G192000      TTTATGTACATCCAGACGACGACGACGACGACAAACAGAAGAACAAAGGCGTCTTTTGGT
c8982_g1_i1           TTTATGTACATCCAGACGACGACGACGACGACAAACAGAAGAACAAAGGCGTCTTTTGGT
                      ************************************************************

Potri.006G192000      ACAGAAATAAATTGGATTGAAGATGTCATCATGAATGTATCGGGGAGTGGAGGTTTCTGT
c8982_g1_i1           ACAGAAATAAATTGGATTGAAGATGTCATCATGAATGTATCGGGGAGTGGAGGTTTCTGT
                      ************************************************************
```

```
Potri.006G192000      TTGTTGCTAGGGATTATAGCCTCCCTGTCTTGTTGGCTTGGAGATCGTTTCAGTTGTTTT
c8982_g1_i1           TTGTTGCTAGGGATTATAGCCTCCCTGTCTTGTTGGCTTGGAGATCGTTTCAGTTGTTTT
                      ************************************************************

Potri.006G192000      TGTCGAACTTTAAACTTAGTTGTGCTCCTTTTTTGAATAACCCTAAGCATCAAGGTCCAG
c8982_g1_i1           TGTCGAACTTTAAACTTAGTTGTGCTCCTTTTTTGAATAACCCTAAGCATCAAGGTCCAG
                      ************************************************************

Potri.006G192000      CATGCATTGGCATGGGACATTCAAGAGGATGCTCCTTTGGCAAACAATTATCAATATTTC
c8982_g1_i1           CATGCATTGGCATGGGACATTCAAGAGGATGCTCCTTTGGCAAACAATTATCAATATTTC
                      ************************************************************

Potri.006G192000      AAAGGCTTTAGCATTGCCACTCATGattattattattattattattattattaCATT
c8982_g1_i1           AAAGGCTTTAGCATTGCCACTCATGATTATTATTATTATTATTATCATTATTATTTT
                      ********************************************* *******  **

Potri.006G192000      CAAATTGGGTATCCTTAATTTATAAAACA-ATTGGCTGAAT-------------------
c8982_g1_i1           ------GTGTAGGGTTAGTT-AGAGGACAGATTGATGGAATAAGGAATGCAAGAAGAGGT
                            * ***   *** ** * *  *** ****   ****

Potri.006G192000      --TGGGCTCGTTCC-----------------------------------------------
c8982_g1_i1           ATTCTCATCGTTGTTGTTGTCAAAAAGAGACGGTGGCTGTTGATTTATGATAGTTAT
                        *    *****

Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           TGATATTATATTGGAGGGAGGGAATTTGGAGCTTGGGATTGGGGGAGGAAGAAAATTGGA

Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           GGAGTGCATGAAGGGATTGGTCTTGAATGCATGCTCGGTTAAGGATTCGAGGAGGATAAG

Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           GTGGGATTCTGGTCTTGTAATAAGGAGCAGAGGTAGATGGGGCAAATGCGGTTGATCAAC

Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           TTTGATTGATCAGGATTTGTAACCTGTTTTATATATAATTCAAACGAGTTTCAGATTTTA

Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           GAGGTCGGAAATTTTTTTTAAGTCGAAATTTTGGGATTGGAGAACAAAAAATATGCAGCA

Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           GCAAGATCATAGAAAAAAGAATTCAACTGAAATGGACTTCTTCTCAGAATATGGTGATGC

Potri.006G192000      ------------------------------------------------------------
c8982_g1_i1           CAATAGGTACAAAATTCAGGAAGTTATCGGGAAAGGCAGTTATGGTGTTGTTTGCTCTGC

Potri.006G192000      -------------------------------------
c8982_g1_i1           AATTGACACTCACACCGGTGAGAAAGTGGCAATAAAGA
```

(3) Potri.009G108100 was missed by PLAS assembly. The contig is a hybrid of Potri.009G108100 and Potri.004G146400.

```
Potri.009G108100      ------------------------------------------------------------
c10302_g1_i1          CGGGCGCCTATAACTGGAGATGTCTTTCCTCAACCTCTCCATGGCTGTAATTCTTGCCTT
Potri.004G146400      ------------------------------------------------------------

Potri.009G108100      ------------------------------------------------------------
c10302_g1_i1          CCTTTTCCCTTGATATTACGTCCTTCTCATCTTCATTCCCTCCTATAAGCAGCCTTGGGA
Potri.004G146400      ------------------------------------------------------------
```

```
Potri.009G108100      ------------------ATGGCTGAACCTCTCAGAGTCTTGGGCTGCTTTTTCTTTTTT
c10302_g1_i1          CTTCCAAGGTAGTCAGTAATGGCTGAACCTCTCAGAGTCTTGGGCTGCTTTTTCTTTTTT
Potri.004G146400      ------------------ATGGCGAACCCTCTCAGAGCCTTCGGCTGCTTTTTCTTTTTC
                                    *****  * ********** *** *****************

Potri.009G108100      TCCTTTCTCTTATCTTCTTTCTCAAACTTCTCTCTTGCCTTAACTGATGCTGAAGCATCT
c10302_g1_i1          TCCTTTCTCTTATCTTCTTTCTCAAACTTCTCTCTTGCCTTAACTGATGCTGAAGCATCT
Potri.004G146400      TCCCTGCTCTTCTCTTCTTTCTCAACCTTCTCTCTTGCCTTAACTGATGCTGAAGCATCT
                      *** * ***** ************* **********************************

Potri.009G108100      TCTATTGCTCGTCGCCAGCTATTGACATTACATGAAAATGGTGAACTTCCCGATGATTTT
c10302_g1_i1          TCTATTGCTCGTCGCCAGCTATTGACATTACATGAAAATGGTGAACTTCCCGATGATTTT
Potri.004G146400      TATATTGCTCGTCGCCAGCTCTTGACGTTAAATGAAAATAGTGAGCTTCCTCATGAGTTT
                      * ******************* ***** *** ******** **** *****  **** ***

Potri.009G108100      GAGTATGAGGTGGATGTGAAAGAAACCTTTGCAAACCAAAGGCTCAGGAGGGCATATATT
c10302_g1_i1          GAGTATGAGGTGGATGTGAAAGAAACCTTTGCAAACCAAAGGCTCAGGAGGGCATATATT
Potri.004G146400      GAGTATGAGGTCGATGTGAAAATAACCTTCGCAAACCAAAGGCTGAGGAGAGCGTATATT
                      *********** *********  ****** ************** ***** ** ******

Potri.009G108100      GGTCTCCAGGCCTGGAAAAAGGCAATGTACTCCGACCCGTTTAATACAACTGGCAATTGG
c10302_g1_i1          GGTCTCCAGGCCTGGAAAAAGGCAATGTACTCCGACCCGTTTAATACAACTGGCAATTGG
Potri.004G146400      GGTCTTCAGGCCTGGAAAAAGGCAATATACTCCGACCCATTTAACACTACTGGCAATTGG
                      ***** ******************** ************ ***** ** ************

Potri.009G108100      GTTGGCGCCGATGTGTGTGCTTATAATGGTGTGTTTTGTGCACCGGCTCTTGACGACTCT
c10302_g1_i1          GTTGGCGCCGATGTGTGTGCTTATAATGGTGTGTTTTGTGCACCGGCTCTTGACGACTCT
Potri.004G146400      GTTGGCGCCAATGTGTGTGCCTATAATGGTGTGTTTTGTGCACCAGCTCTAGACGACCCC
                      ********* ********** ************************ ***** ****** *

Potri.009G108100      GGTCTAAGCGTTATGGCAGGTGTTGATCTTAACGGTGCTGATATTGCTGGGTACCTTCCA
c10302_g1_i1          GGTCTAAGCGTTATGGCAGGTGTTGATCTTAACGGTGCTGATATTGCTGGGTACCTTCCA
Potri.004G146400      AGTCTGAGCGTTGTGGCAGGTGTTGATCTTAACGGTGCTGACATTGCTGGGCACCTTCCA
                       **** ****** ***************************** ********* ********

Potri.009G108100      GCTGAATTGGGGCTTTTGACAGATGTTGCATTGTTCCACATTAACTCTAACAGGTTTTGT
c10302_g1_i1          GCTGAATTGGGGCTTTTGACAGATGTTGCATTGTTCCACATTAACTCTAACAGGTTTTGT
Potri.004G146400      GCTGAATTAGGGCTTATGACAGATGTTGCATTATTCCACATTAACTCTAACAGGTTTTGT
                      ******** ****** ***************** ***************************

Potri.009G108100      GGAATCATCCCCAAGAGCTTTTCCAAGCTCACACTCATGTACGAGTTTGATGTCAGCAAC
c10302_g1_i1          GGAATCATCCCCAAGAGCTTTTCCAAGCTCACACTCATGTACGAGTTTGATGTCAGCAAC
Potri.004G146400      GGTATCATTCCCGAGAGCTTTTCCAAGCTCACACTCATGTATGAGTTTGATGTCAGCAAC
                      ** ***** *** ***************************** ******************

Potri.009G108100      AACCGCTTTGTTGGTGACTTCCCTTCTGTTGTTTTAACCTTGCCAAGCCTCAAGTATCTT
c10302_g1_i1          AACCGCTTTGTTGGTGACTTCCCTTCTGTTGTTTTAACCTTGCCAAGCCTCAAGTATCTT
Potri.004G146400      AACCGCTTTGTTGGTGATTTCCCTTCTGTTGTTCTATCCTGGCCAAGCCTCAAGTATCTT
                      ***************** *************** ** *** ******************

Potri.009G108100      GACATCAGATTCAATGATTTCGAAGGTAGTTTGCCTCCAGAACTCTTCAACAAGGACCTC
c10302_g1_i1          GACATCAGATTCAATGATTTCGAAGGTAGTTTGCCTCCAGAACTCTTCAACAAGGACCTC
Potri.004G146400      GACGTCAGATTCAACGATTTCGAAGGTAGTTTGCCTCCAGAACTCTTCAACAAGGAACTC
                      *** ********** ****************************************** ***

Potri.009G108100      GATGCTTTGTTCTTGAATGACAACCGGTTCACATCCACCATTCCGGAGACAATAGGCAAC
c10302_g1_i1          GATGCTTTGTTCTTGAATGACAACCGGTTCACATCCACCATTCCGGAGACAATAGGCAAC
Potri.004G146400      GATGCTTTGTTCTTGAATGACAACCGATTCACATCCACCATCCCGGAGACAATAGGCAAC
                      ************************** ************** ******************

Potri.009G108100      TCCCCAGTTTCTGTAGTCACATTTGCGAACAACAAATTCACTGGCTGCATTCCACACAGC
c10302_g1_i1          TCCCCAGTTTCTGTAGTCACATTTGCGAACAACAAATTCACTGGCTGCATTCCACACAGC
Potri.004G146400      TCCGCAGTTTCTGTTGTCACATTTGCTAACAACAAATTCACCGGCTGCATTCCACACAGC
                      *** ********** *********** *************** ******************

Potri.009G108100      ATCGGCAAGATGACAAACTTGAACGAGGTCATCTTTATGGGCAATGATCTTGGTGGTTGC
c10302_g1_i1          ATCGGCAAGATGACAAACTTGAACGAGGTCATCTTTATGGGCAATGACCTTGGTGGTTGC
Potri.004G146400      GTCGGCAAGATGGCCAACTTGAACGAGGTCATCTTTATGGGCAATGACCTTGGTGGTTGC
                       *********** * ********************************* ************
```

```
Potri.009G108100      TTCCCAGCAGAAATTGGGCTGCTTGGTAATGTGACTGTGTTTGATGCCAGCCACAATGGG
c10302_g1_i1          TTCCCAGCAGAAATTGGGCTGCTTCGTAATGTGACTGTCTTTGATGCCAGCCACAATGGG
Potri.004G146400      TTCCCAGCAGAAATTGGGCTGCTTCGTAATGTGACTGTCTTTGATGCCAGCCACAATGGG
                      ************************ ************* **********************

Potri.009G108100      TTCACAGGAATCTTGCCGTCCAGCTTTGCAGGGCTAAAGAAGGTTGAACTCTTGGATCTT
c10302_g1_i1          TTCACAGGAATCTTGCCGCCCAGCTTTGCAGGCCTAAAGAAGGTTGAACTCTTGGATCTT
Potri.004G146400      TTCACAGGAATCTTGCCGCCCAGCTTTGCAGGCCTAAAGAAGGTTGAACTCTTGGATCTT
                      ****************** ************* **************************

Potri.009G108100      GCAGACAACAAGCTGACAGGATTTGTGCCTGAGAACATTTGTAGGTTGTCAAGCTTGACG
c10302_g1_i1          GCCGACAACAAGCTGACAGGATTTGTGCCTGAGAACATTTGCAAGTTGCCAAGCTTGACA
Potri.004G146400      GCCGACAACAAGCTGACAGGATTTGTGCCTGAGAACATTTGCAAGTTGCCAAGCTTGACA
                      ** ************************************ * **** **********

Potri.009G108100      AACTTCACATTCTCGTATAACTACTTCAAGGGGGAGGCTCAAGCTTGCGTGCCTCCATCA
c10302_g1_i1          AACTTCACATTCTCGTATAACTACTTCAAGGGCGAGGCTCAAGCTTGCGTGCCTCCATCA
Potri.004G146400      AACTTCACATTCTCGTATAACTACTTCAAGGGCGAGGCTCAAGCTTGCGTGCCTCCATCA
                      ******************************** **************************

Potri.009G108100      AGGAAGGACACCGT---------------------------------GCCAAAGCAGAAG
c10302_g1_i1          AGGAAAGACATTGTGTTGGATGATACCAGCAATTGCCTGTCTGACAGGCCAAAGCAGAAG
Potri.004G146400      AGGAAAGACATTGTGTTGGATGATACCAGCAATTGCCTGTCTGACAGGCCAAAGCAGAAG
                      ***** ****  **                                 *************

Potri.009G108100      TCAGCCAGGACATGTTACCCAGTGGTGAGCCGACCTGTGGATTGCAGCAAGGACAAGTGT
c10302_g1_i1          TCAGCCAGGACATGTTATCCAGTGGTGAGCCGACCTGTGGATTGCAGCAAGGACAAGTGT
Potri.004G146400      TCAGCCAGGACATGTTATCCAGTGGTGAGCCGACCTGTGGATTGCAGCAAGGACAAGTGT
                      ***************** ****************************************

Potri.009G108100      TCTGGAGGAGGAGGTTCTTCAAACCCCCATCCAAAACCACAACCCACACCACCTACTCCA
c10302_g1_i1          GCTGGAGGAGGAGGTTCTTCAAACCCTCATCCAAAGCCCCAACCCACACCACCTACTTCA
Potri.004G146400      GCTGGAGGAGGAGGTTCTTCAAACCCTCATCCAAAGCCCCAACCCACACCACCTACTTCA
                       ************************* ******** ** ****************** **

Potri.009G108100      GAACATAAACAAACCCCATCTCCACCTAAATCTACTTCTACTCCAACA---------CCA
c10302_g1_i1          AAACATGAACCAACTCCATCTCCTCCCAAATCTATTTCTATTTCTACACCAACGCCACCA
Potri.004G146400      AAACATGAACCAACTCCATCTCCTCCCAAATCTATTTCTATTTCTACACCAACGCCACCA
                       ***** *** *** ******** ** ******* ***** * * ***         ***

Potri.009G108100      TCATCACCAATCCCTGCCCCTCGAACACCAGAATTACCAAAACCAGAACCTAAGTTACCG
c10302_g1_i1          TCAGCACGGGTCCCCACCCCTCAAACAGCAGAATCACCAAAACCAGAACATGAGTTGCCA
Potri.004G146400      TCAGCACGGGTCCCCACCCCTCAAACAGCAGAATCACCAAAACCAGAACATGAGTTGCCA
                      *** ***   ****  ****** **** ****** ************** * **** **

Potri.009G108100      CTGGCTCCAGTTGAACCAATTAGTCCATCAACACCAGAGGTATCCTTACCACCATCTTTA
c10302_g1_i1          CAAACTCCGGTTGAACCTATTAGGCCATCGACTCCAAAGATACCCTTACCGTCATCTCCA
Potri.004G146400      CAAACTCCGGTTGAACCTATTAGGCCATCGACTCCAAAGATACCCTTACCGTCATCTCCA
                      *   **** ******** ***** ***** ** *** ** ** *******  *****  *

Potri.009G108100      TCAATTAGTCCTTCAACTCCGGAGATATCCTCACCACCATCTTCATCAAGTCCATCTACC
c10302_g1_i1          TCAATCAA----------------------------------------TTCATCTGCC
Potri.004G146400      TCAATCAA----------------------------------------TTCATCTGCC
                      ***** *                                         * ***** **

Potri.009G108100      CCATCATCTGACCCATACAATCCAGGACCTGGTGGGCATGACGAGACACCGCCATCACCA
c10302_g1_i1          CCATCATTTGATCCATATAATCCAGGATCTGGTGGTCATGGCGAGACACCATTATCACCG
Potri.004G146400      CCATCATTTGATCCATATAATCCAGGATCTGGTGGTCATGGCGAGACACCATTATCACCA
                      ******* *** ***** ********* ******* **** *********   ******

Potri.009G108100      AAATCTGCACCGTCACCTAATCCATTTAATAATTCACCAGTTGGGCACAATGAGACACCA
c10302_g1_i1          AATTATGCACCGTCACCTGATTCATTTGGTAACTCACCTATTGGCCACCACGATACACCG
Potri.004G146400      AATTATGCACCGTCACCTGATTCATTTGGTAACTCACCTATTGGCCACCACGATACACCG
                      ** * ************* ** *****  *** *****  **** *** * ** *****

Potri.009G108100      CCATCACCGGAGTCTGCACCGTCACCCGATCCATTCAATAATTCACCAGATGGGCATAAC
c10302_g1_i1          CCATCACTCTCT--------------------AT----TAGTCCATC-------------
Potri.004G146400      CCATCACTCTCT--------------------AT----TAGTCCATC-------------
                      *******                          **    ** * ** *
```

```
Potri.009G108100      GAGACACCGCTATCACCGGAGTCTGCACCGTCACCCAATCCATTCAATAATTCACCAGAT
c10302_g1_i1          ----------------------------ATCA--------------------TC----
Potri.004G146400      ----------------------------ATCA--------------------TC----
                                                   ***                    *

Potri.009G108100      GGGCACGACGAGACACCGCTATCACCGGAGTCTGCACCGTCACAATCACCGGAGTCTGCA
c10302_g1_i1          --------AAAGATACCTGT-----------------ATCACCATCACCAAAGTCTGCA
Potri.004G146400      --------AAAGATACCTGT-----------------ATCACCATCACCAAAGTCTGCA
                              *** ***  *                  **** ******  ********

Potri.009G108100      CCGTCACCCGATCCATTTAATAATTCACCAGATGGGCACGATGAGACACCACAATCATCG
c10302_g1_i1          CCATCACCTGATGATGAATACAATCCA--------------------------------
Potri.004G146400      CCATCACCTGATGATGAATACAATCCA--------------------------------
                      ** ***** ***       * *** **

Potri.009G108100      GAGTCTGCACTGTCACCCGATCCATTCAATAATTCACCAGATGGGCACGACGAGATACCA
c10302_g1_i1          --------------------------------GGAGCTGGTGGACATGGCGAGACACCA
Potri.004G146400      --------------------------------GGAGCTGGTGGACATGGCGAGACACCA
                                                       * * * *** ** * ***** ****

Potri.009G108100      CCATCACCAGAGCCGTCACCGGATCCATTTAATAATTCACCGAATGGGCATGATGAGACA
c10302_g1_i1          TCATCACCATCACC---------------------------------------------
Potri.004G146400      TCATCACCATCACC---------------------------------------------
                       ********   **

Potri.009G108100      CCAACATCACCAGAGTCCGCACAATCACCTGATCCATTTAATAATTCACCAATTGGGCAC
c10302_g1_i1          -----------AACCTCCTCACTAAAACCCGA----------------------GGCAC
Potri.004G146400      -----------AACCTCCTCACTAAAACCCGA----------------------GGCAC
                               *   *** *** *  *** **                       *****

Potri.009G108100      GACAAGACACCACCACCATCATCTGAGATATCTATACCACCATCACCCTTAATTAGTCCA
c10302_g1_i1          CAA-AAACATCACCACAACCA---AAA--------------------------------
Potri.004G146400      CAA-AAACATCACCACAACCA---AAA--------------------------------
                       *  * *** ****** * **    *

Potri.009G108100      CCAACATCGGAGAAACATATACCACCATCATCAGAGTTTGCTCCATCACCTGATTCATAT
c10302_g1_i1          ------------------ATACCA-----------G------------------------
Potri.004G146400      ------------------ATACCA-----------G------------------------
                                  ******           *

Potri.009G108100      AATTTACGACCTGTTCACTCACCTCCACCATCATCACAATCCCTACCCCCTCTAGTCTAT
c10302_g1_i1          ---TTATAAAC-CCTCACT---TTCCATCATC------ATCCCCACCACTTGTGCTC---
Potri.004G146400      ---TTATAAAC-CCTCACT---CTCCATCATC------ATCCCCACCACTTGTGCTC---
                         ***  * *   *****    **** ****      ***** *** * * *  **

Potri.009G108100      TCACTTCCACCACCAGCACATTCACCCCCACCATCAATTCATTTCCCACCACCACCTGTA
c10302_g1_i1          TCAACCCCTTCA--------------------CCAGTTCACTCTCCACCACCACCTGTC
Potri.004G146400      TCAACCCCTTCACTGGTCCACTCTCCCCCACCACCAGTTCACTCTCCACCACCACCTGTC
                      ***   **  **                    ** **** *  **************

Potri.009G108100      CACTCTCCCCCACCACCCCCTGTTTACTCTCCCCCTCTGCCAGTACATTCACCGCCACCA
c10302_g1_i1          CAGTCACCCCCACC---ACCAGTTCACTCTCCCCCGCCACCTGTCCAATCACCAShowfl
Potri.004G146400      CAGTCACCACCACC---ACCAGTTCACTCTCCCCCGCCACCCGTCCAGTCACCTCCACCA
                      ** ** ** *****    ** *** ********** *  ** ** ** *****

Potri.009G108100      TCAGTGCACTCTCCCCCACCACCAATGCACTCTCCCCCACCACCACCAGTTTACTCTCCC
c10302_g1_i1          ankingsequenceups--------treamdow--------------------------
Potri.004G146400      CCAGTTCACTCTCCACCACCACCCGTACACTCACC-----------------------G
                                *           *  *

Potri.009G108100      CCACCGCCAGTACAATCATTCCCACCACCAGTGCACTCTCCCCCACCACCTGTACACTCA
c10302_g1_i1          --------nstreamSubmit---------------------------------------
Potri.004G146400      CCACCACCTGTCCAGTCACCTCCACCACCAGTTCACTCTCCACCACCACCCGTACACTCA
                                *  *

Potri.009G108100      CCCCCTCCACCACCAGTTTACTCTCCCCCTCCGCCAGTACATTCACCGCCACCACCAGTG
c10302_g1_i1          ---------------------------------------------------------
Potri.004G146400      CCGC---CACCACCAGTTCAC--------------------TCACCACCACCACCTGTC
```

```
Potri.009G108100      TACTCTCCCCCACCGCTGGTACAATCACCCCCACCACCAGTGCACTCTCCCCCACCACCT
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      CACTCTCCGCCACCACCAGTCCAGTCACCCCCTCCACCTGTCCACTCTCCACCTCCACCC


Potri.009G108100      TTACACTCACCCCCTCCACCACCAGTTTACTCTCCCCCTCCGCCAGTACATTCACCGCCA
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      GTACACTCACCTCC--------------------------TCCTGTCCAGTCACCCCCT


Potri.009G108100      CCACCAGTGCACTCTCCCCCACCGCCGATACAATCACCCCCACCACCAGTGCACTCACCC
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      CCACCTGTTCACTCTCCACCACCA--------------------CCAGTACACTCACCC


Potri.009G108100      CCTCCACCACCTATACACTCACCCCCGCCACCCGTGCAATCTCTCCCTCCACCACCTGTA
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      CCTCCTCCA---GTTCACTCACCCTCTCCACCCGTACAATCCCCCCCACCACCACCAGTA


Potri.009G108100      AACTCACCCCTGCCACCCGTGCACTCCCCACCACCACCGGTTCATTCTCCTACATCACCC
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      CACTTACCTCCTCCACCAGTACACTCTCCACCACCACATGTTAAATCACCACC-------


Potri.009G108100      ATACACTCCCATCCACCACCTGTAAACTCACCCCCGCCACCCGTGCAATCACCTCCACCT
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      ----------------------------ACCACCACGACCAGTCAAATCATCTCCACTT


Potri.009G108100      CCAGTTTTCTCTCCACCACCAGTAATTGTATCTCCTCC------------------TCCA
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      CCAATTTTCTCTCCACCACCACCAACTGTATTTCCTCATCCTCCTCGTGCTTTTCCTCCA


Potri.009G108100      CCTCCCCCGGAAGAAGACTTCATCCTTCCACCAAACCTCGGATTCCAATATGCATCACCA
c10302_g1_i1          ------------------------------------------------------------
Potri.004G146400      CCACCCCCAAATGAAGATATAGTCCTTCCACCAAACCTCGGATTCCAATACGCATCGCCA


Potri.009G108100      CCTCCACCAACGTTCCCAGGCTACTAA
c10302_g1_i1          ---------------------------
Potri.004G146400      CCTCCACCAGTGTTCCCAGGCTACTAG
```

Case 4. Out of the 20 genes (>1000 FPKM) not fully assembled by PLAS, 2 genes were missing 5’ end.

(1) Potri.006G276200. The contig is missing 5’ end.

```
Potri.006G276200      ATGATCCACCAATATTGTCATTGCTATTGCTGGCTTAAGACATTGTCAAACTACAGAAAG
c8950_g1_i1           ------------------------------------------------------------


Potri.006G276200      TTATCCACTCTTTTGTTCATCCCCTACATAACAAGACCTCTCATCCTGATCTGTAGATTG
c8950_g1_i1           ------------------------------------------------------------


Potri.006G276200      AGAAATTCATTCTCGACCATGAATAGCCCAGCCACTCATTTCTGGTCAACATCTCTACAT
c8950_g1_i1           ------------------------------------------------------------T
                                                                                   *


Potri.006G276200      GCATATAAAGGCCTCTGCATTATCAGCTTTGAATTCAAACACAAGGGAAGTGATAATAAT
c8950_g1_i1           GCATATAAAGGCCTCTGCATTATCAGCTTTGAATTCAAACACAAGGGAAGTGATAATAAT
                      ************************************************************
```

```
Potri.006G276200      AAGGAAGGCTTTGGCTTCATCTTTTTGGCAAAGATGGATTCCAAGGCTTCTTCTCTCCTC
c8950_g1_i1           AAGGAAGGCTTTGGCTTCATCTTTTTGGCAAAGATGGATTCCAAGGCTTCTTCTCTCCTC
                      ************************************************************

Potri.006G276200      TTCATTGCATTCTTGTGCTTAATCTCAACTTCCACAGCCTTCAACAGCACCAAGATCCTT
c8950_g1_i1           TTCATTGCATTCTTGTGCTTAATCTCAACTTCCACAGCCTTCAACATCACCAAGATCCTT
                      ******************************************** **************

Potri.006G276200      GCACAGTACCCTGAGTTTGCTAACTTTAATGATCTCCTCAGCCAGAGCGGGCTCGCCCAG
c8950_g1_i1           GCACAGTACCCTGAGTTTGCTAACTTTAATGATCTCCTCAGCCAGAGCGGGCTCGCCCAG
                      ************************************************************

Potri.006G276200      GAAATGAACAGCCGCCAAACCATCACTGTCCTTGTGCTTGATAACGGATCAATCGATGGA
c8950_g1_i1           GAAATGAACAGCCGCCAAACCATCACTGTCCTTGTGCTTGATAACGGATCAATCGATGGA
                      ************************************************************

Potri.006G276200      CTCTCTGGCAGACCCTTAGACATTGCAAAGAGGATCTTGAGTGCACATGTAATCCTTGAT
c8950_g1_i1           CTCTCTGGCAGACCCTTAGACATTGCAAAGAGGATCTTGAGTGCACATGTAATCCTTGAT
                      ************************************************************

Potri.006G276200      TACTATGATCAAATAAAGCTTTCGAAACTTCAAAAGGCCAGCACTATCGTTACCACCTTG
c8950_g1_i1           TACTATGATCAAATAAAGCTTTCGAAACTTCAAAAGGCCAGCACTATCGTTACCACCTTG
                      ************************************************************

Potri.006G276200      TACCAAGCTAGTGGTGTTGCAGATAATCGACAAGGTTTCCTGAATATTAGCAGAACTGCT
c8950_g1_i1           TACCAAGCTAGTGGTGTTGCAGATAATCGACAAGGTTTCCTGAATATTAGCAGAACTGCT
                      ************************************************************

Potri.006G276200      GAGGGAATCAAATTCGGTTCAGCAATGAAAGGTGCTCCTCTCGTTGCATCACTTGTGAAA
c8950_g1_i1           GAGGGAATCAAATTCGGTTCAGCAATGAAAGGTGCTCCTCTCGTTGCATCACTTGTGAAA
                      ************************************************************

Potri.006G276200      TCCATCTACTCGCAGCCTTACAACATCTCGGTGCTACAAGTCAGCGAACCTATTGAGACT
c8950_g1_i1           TCCATCTACTCGCAGCCTTACAACATCTCGGTGCTACAAGTCAGCGAACCTATTGAGACT
                      ************************************************************

Potri.006G276200      CCAGGGATTGAGAACATGGCTCCACCACCACCACCTGGTACTGCTGCCGTTCCCAAGAAG
c8950_g1_i1           CCAGGGATTGAGAACATGGCTCCACCACCACCACCTGGTACTGCTGCCGTTCCCAAGAAG
                      ************************************************************

Potri.006G276200      GCACCTGCTCCAGCTCCAAGCACTAAAACGCCACCAGCTGCACCTCCAACTGCCAAGACT
c8950_g1_i1           GCACCTGCTCCAGCTCCAAGCACTAAAACGCCACCAGCTGCACCTCCAACTGCCAAGACT
                      ************************************************************

Potri.006G276200      CCAGCCAAATCCCCTGCCAAATCTCCTTCCAAGGCTCCTGCACCATCCAAGGAGGGACCA
c8950_g1_i1           CCAGCCAAATCCCCTGCCAAATCTCCTTCCAAGGCTCCTGCACCATCCAAGGAGGGACCA
                      ************************************************************

Potri.006G276200      TCTACACCAACTAAAGCACCAGCCGAGGGGCCAGTGGCTGCTGATGGGCCAGTGGCTGCT
c8950_g1_i1           TCTACACCAACTGAAGCACCAGCCGAGGGGCCAGTGGCTGCTGATGGGCCAGTGGTTGCT
                      ************ ****************************************** ****

Potri.006G276200      GGTGGCCCAGTAGCTGATGTGCCCGCAGAGTCCCCAGAGGCTGATACAGAAGTGGCTGAG
c8950_g1_i1           GATGGGCCAGTGGCTGATGTGCCCGCAGACTCCCCAGAGGCTGATACAGAAGTGGCTGAG
                      * *** ***** ***************** ****************************

Potri.006G276200      GAAGCACCAGCTGTAGCACCTGCAAAAGCTGCTTCTTCACGTATGCATGTTGCTGGTGCA
c8950_g1_i1           GAAGCACCAGCTGTAGCACCTGCAAAAGCTGCTTCTTCACGTATGCATGTTGCTGGTGCA
                      ************************************************************

Potri.006G276200      ACCGTGGTTATCGGATTGTTTGCCTGCATAATGGGTTTTTAA------------------
c8950_g1_i1           ACCGTGGTTATCGGATTGTTTGCCTGCATAATGGGTTTTTAAAAGGCAAGCAATAGAACA
                      ****************************************

Potri.006G276200      ------------------------------------------------------------
c8950_g1_i1           AGGCAGAACAAAAACCGGGTAAAAAAATGGAGTGAATGGGAACTGGATAGTACGAGGCTA


Potri.006G276200      ------------------------------------------------------------
c8950_g1_i1           GCACAAGAATCAATTTATTATTTTTTTTGTATTTCCTTTGATTGTATCCACCCCTGATCA
```

```
Potri.006G276200      ------------------------------------------------------------
c8950_g1_i1           CTGTTCGAAATGAAGCTTGAACCCTAATCAGGGGCAAACATAATTTGATTAATCTGATAT

Potri.006G276200      ------------------------------------------------------------
c8950_g1_i1           TTACTGATAGATTGCAAATTGAAATTTTTCCACAGTATATTTTATATGATTTATCAACAT

Potri.006G276200      ------------------------------
c8950_g1_i1           CCCTTGATTTTTCATCAAAAAAAAAAAAAA
```

(2) Potri.010G055300. The contig is missing 5’ end.

```
Potri.010G055300      ATGAAGAACCATTTTATTTGGCGTCGCCCGGACTCGAACCGGAGACCTTCAGGCTTGAAA
c32651_g1_i1          ------------------------------------------------------------

Potri.010G055300      CGGCCGCCAATAAAGGAAAGGAAAAGGAAACCGACTAATTTTATGGTTGGTGGGGTTTTG
c32651_g1_i1          --------AATACAGGAAAGGAAAAGGAAACCGACTAATTTTATGGTTGGTGGGGTTTTG
                              **** **********************************************

Potri.010G055300      CTGATCATTCTTGACCTATCCTTTTTCCTTTGGGCACCCACTCTTTTCCTGAATTTTAGG
c32651_g1_i1          CTGATCATTGTTGACCAATCCTTTTTCCTTTGGGCACCCACTGTTTTCCTGAATTTTAGG
                      ********* ****** ************************ *****************

Potri.010G055300      CGTTTCTATCCACTACAAATTCTCTCAATATCAATCACTTCCCTTTCATACTCTCCTTTC
c32651_g1_i1          CGTTTCTATCCACTACAAATTCTCTCAATATCAATCACTTCCCTTTCATACTCTCCTTTC
                      ************************************************************

Potri.010G055300      CTTCATTTCTCTCTAAACTTGATTTTTCTGGAGAAGGGGATTTCTCTGTTTCTCCTTCTC
c32651_g1_i1          CTTCATTTCTCTCTAAACTTGATTTTTCTGGAGAAGGGGATTTCTCTGTTTCTCCTTCTC
                      ************************************************************

Potri.010G055300      TCTTGCAATTCTAGCATGGCTCCCTCACGATGGATAAGGCCTGAGGTGTTTCCACTCTTT
c32651_g1_i1          TCTTGCAATTCTAGCATGGCTCCCTCACGATGGATAAGGCCTGAGGTGTTTCCACTCTTT
                      ************************************************************

Potri.010G055300      GCATCTGTTGGTGTAGCTGTTGGCATTTGTGGCATGCAACTTCTTAGGAATATAACCACC
c32651_g1_i1          GCATCTGTTGGTGTAGCTGTTGGCATTTGTGGCATGCAACTTCTTAGGAATATAACCACC
                      ************************************************************

Potri.010G055300      AACCCTGAAGTAAGGGTGACGAAAGAGAACAGGGCAGCAGGAGTGCTTGACAACTTTAAA
c32651_g1_i1          AACCCTGAAGTAAGGGTGACGAAAGAGAACAGGGCAGCAGGAGTGCTTGACAACTTTAAA
                      ************************************************************

Potri.010G055300      GAGGGCGAGAAATATGCAGAACATGGTCTTAGGAAGTATGTCCGAAAGAGAACTCCTCAG
c32651_g1_i1          GAGGGCGAGAAATATGCAGAACATGGTCTTAGGAAGTATGTCCGAAAGAGAACTCCTCAG
                      ************************************************************

Potri.010G055300      ATCATGCCATCCATCAACGGTTTCTTCTCAGACCCAGATCTTCCAACTAACTAA------
c32651_g1_i1          ATCATGCCATCCATCAACGGTTTCTTCTCAGACCCAGATCTTCCAACTAACTAAAACCCG
                      ********************************************************

Potri.010G055300      ------------------------------------------------------------
c32651_g1_i1          CTGACCACTCTATTTAGGATTGCATCTTTTCAAGTTTGAGATCTGGAGGAAGAAAGCAAT

Potri.010G055300      ------------------------------------------------------------
c32651_g1_i1          GACAACGTTATTCATTTATCATTTGCATAGGATAAGGAAGAGTATGATACATTGGTTGTT

Potri.010G055300      ---------------------------------------------
c32651_g1_i1          CCATCTTTTTGTTACTCTATCTCAATACAATTTAATAATAACAGAG
```

Supplemental File S2.2

Protein and nucleotide sequences for *TvQR1*

Protein sequence

>TrVe.c107009_g2_i1.16950_QR1

MAGKLMRAVQYDGYSGGAAGLKHDEVPIPSPGKGEVLIKLEAISLNQLDWKLQNGMVRPFLPRKFPFIPA
TDVAGEVVRIGPDVKNFKPGDKVVAMLGSFGGGGLAEYGVASEKLTVHRPPEVSAAESSGLPIAGLTAHM
ALTQHIGLNLDKSGPHKNILITAASGGVGQYAVQLAKLGNTHVTATCGSRNFDLVKSLGADEVIDYKTPE
GAALKSPSGKKYDAVIHCASPLPWSVFKPNLSKHGKVIDITPGPRVMLTSAMTKLTCSKKRLVTLLVVIK
GEHLSYLVELMREGKLKTVIDSKFPLSKAEEAWAKSIDGHATGKIVVEP*

Nucleotide sequence

>TrVe.c107009_g2_i1.16950_QR1

AAAAAATCGCTGAAATTTAATTTAATTATGGCCGGAAAGCTTATGCGTGCGGTTCAGTACGACGGTTATA
GCGGTGGAGCTGCTGGTTTGAAGCATGATGAAGTTCCAATACCTAGTCCTGGCAAGGGCGAGGTCCTTAT
AAAGCTTGAAGCCATAAGCTTAAATCAACTTGATTGGAAGCTTCAGAATGGCATGGTTCGTCCTTTTCTT
CCTCGGAAATTCCCTTTTATACCTGCTACCGACGTGGCTGGGGAGGTGGTCCGGATCGGACCGGATGTCA
AAAACTTTAAACCCGGTGACAAAGTTGTTGCTATGCTTGGCAGTTTTGGAGGAGGTGGCTTAGCCGAATA
CGGCGTAGCAAGTGAAAAGCTAACAGTCCATAGGCCGCCCGAGGTATCAGCTGCCGAGAGCTCAGGCCTT
CCCATTGCCGGCCTTACAGCCCACATGGCCCTAACCCAACACATTGGCCTAAACCTCGACAAAAGTGGTC
CCCACAAAAACATCCTCATCACAGCCGCCTCCGGTGGTGTTGGCCAATACGCCGTTCAGCTCGCAAAGCT
AGGAAACACACATGTAACCGCCACATGTGGGTCCCGAAACTTTGACTTGGTCAAAAGCCTCGGAGCCGAC
GAGGTTATTGACTATAAAACCCCCGAAGGGGCAGCCCTTAAGAGCCCGTCGGGCAAAAAGTATGATGCGG
TTATTCATTGTGCATCGCCTTTGCCATGGTCCGTTTTTAAACCGAACTTGAGCAAACATGGGAAAGTGAT
CGATATAACTCCCGGTCCGAGGGTTATGTTGACTTCGGCTATGACAAAACTTACGTGCTCGAAGAAACGA
TTGGTGACGTTACTTGTTGTGATCAAGGGCGAGCATTTGAGTTATCTTGTTGAGTTAATGAGAGAAGGGA
AACTTAAGACGGTTATCGACTCTAAGTTTCCGTTAAGTAAGGCTGAGGAGGCTTGGGCTAAGAGCATCGA
CGGCCATGCTACCGGGAAGATCGTTGTCGAGCCATAAGTTAGTAAGATTTTGTTTTGTTTTATGATATTG
TAATGTGGAATTTGGCTTATGACTTGTTTTGGTGATCTTTATGTTTTGATATGTACTCTTTTGTTAACCT
ACTTGTGGTTAGAGTGGCAATTTGTGTCACCATGGTTGTGTTTGTTCGTGTCCTTAAGTCCTATAATGTA
ATTTTCATATTTTATACTTTATTTAGTC

References

Allen JM, Huang DI, Cronk QC, Johnson KP (2015) aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. BMC Bioinformatics **16**: 98

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

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120

Chopra R, Burow G, Farmer A, Mudge J, Simpson CE, Burow MD, Seijo G, Lavia G, Fernández A, Krapovickas A, et al (2014) Comparisons of *de novo* transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-Seq data. PLoS One **9**: e115055

Denoeud F, Aury J-M, Da Silva C, Noel B, Rogier O, Delledonne M, Morgante M, Valle G, Wincker P, Scarpelli C, et al (2008) Annotating genomes with massive-scale RNA sequencing. Genome Biol **9**: R175

Fraley C, Raftery AE, Murphy TB (2012) mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29**: 644–52

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, et al (2010) *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol **28**: 503–10

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.

Genome Biol **14**: R36

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

Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, Dewey CN (2014) Evaluation of *de novo* transcriptome assemblies from RNA-Seq data. Genome Biol. doi: 10.1186/s13059-014-0553-5

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13**: 2178–89

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**: 550

Marchant A, Mougel F, Mendonça V, Quartier M, Jacquin-Joly E, da Rosa JA, Petit E, Harry M (2016) Comparing *de novo* and reference-based transcriptome assembly strategies by applying them to the blood-sucking bug Rhodnius prolixus. Insect Biochem Mol Biol **69**: 25–33

Martin JA, Wang Z (2011) Next-generation transcriptome assembly. Nat Rev Genet **12**: 671–682

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. [Miyashita mitsunori]

Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods **10**: 71–3

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al (2010) *De novo* assembly and analysis of RNA-seq data. Nat Methods **7**: 909–12

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics **28**: 1086–92

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S (2016) TransRate: reference-free

quality assessment of *de novo* transcriptome assemblies. Genome Res **26**: 1134–44

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28**: 511–515

Vijay N, Poelstra JW, Künstner A, Wolf JBW (2013) Challenges and strategies in transcriptome assembly and differential gene expression quantification. A comprehensive *in silico* assessment of RNA-seq experiments. Mol Ecol **22**: 620–634

Visser EA, Wegrzyn JL, Steenkmap ET, Myburg AA, Naidoo S (2015) Combined *de novo* and genome guided assembly and annotation of the *Pinus patula* juvenile shoot transcriptome. BMC Genomics **16**: 1057

Westwood JH, DePamphilis CW, Das M, Fernández-Aparicio M, Honaas L a., Timko MP, Wafula EK, Wickett NJ, Yoder JI (2012) The parasitic plant genome project: new tools for understanding the biology of *Orobanche* and *Striga*. Weed Sci **60**: 295–306

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, et al (2014) SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics **30**: 1660–6

Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, Huang K, Bandaranayake PCG, Wu B, Der JP, et al (2015) Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol **32**: 767–90

Zhao Q-Y, Wang Y, Kong Y-M, Luo D, Li X, Hao P, Graveley B, Brooks A, Carlson J, Duff M, et al (2011) Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics **12**: S2

CHAPTER 3

PLASMA MEMBRANE PHYLLOQUINONE BIOSYNTHESIS: CONSERVATION AND DIFFERENTIAL EVOLUTION IN GREEN PLANTS AND HOLOPARASITES[1]

[1] Gu, X., S.A. Harding, B. Nyamdari, K.B. Aulakh, J.H. Westwood, and C.J. Tsai, to be submitted to *Science*

Abstract

Phylloquinone (PhQ, or vitamin K1) is an essential electron carrier for photosynthesis found in plastids. Accordingly, PhQ is most abundant in photosynthetic tissues of green pants. Here we report unexpected biosynthesis and subsequent accumulation of PhQ in the plasma membrane of a non-photosynthetic holoparasite *Phelipanche aegyptiaca.* The entire suite of PhQ genes were found to be transcribed in this holoparasite, with high expression levels at early developmental stages associated with parasitism. Enzymes catalyzing the last two steps of PhQ biosynthesis were characterized to be redirected from plastids to the plasma membrane, indicating a plasma membrane destination of PhQ. Co-expression analysis in the holoparasite revealed a reduced association of PhQ genes with photosynthesis compared to photosynthetically-competent parasites, consistent with the loss of its photosynthesis capacity. The representation trend was reversed for transcripts associated with oxidation-reduction and defense, suggesting their association with plasma-membrane-destined PhQ. Some candidate genes highly connected with PhQ genes in the non-photosynthetic parasite, including peroxidases and quinone reductases, have been experimentally validated to participate in haustorial development. Our results support a model where PhQ functions as an electron carrier in plasma membrane redox systems to mediate parasitism. Plasma membrane localization of the last two enzymatic steps was also predicted for photosynthetic species via alternatively splicing. Together with the holoparasite results, this observation suggested that non-plastidial PhQ is evolutionarily conserved.

Introduction

Phylloquinone (PhQ), also known as vitamin K1, is a membrane-bound, lipid-soluble naphthoquinone derivative essential to plants. PhQ functions as an electron transfer cofactor in photosystem I (PSI) during photosynthesis (Brettel et al., 1986). *Arabidopsis* mutants deficient in PhQ biosynthesis are often seedling-lethal or growth-compromised due to impaired PSI assembly (Gross et al., 2006). The eubacterial counterpart menaquinone (MK, vitamin K2) plays an important role in respiratory electron transport and redox regulation across the plasma membrane (Frydman and Rapoport, 1963; Hale et al., 1983; Frigaard et al., 1997). A similar function has been speculated for PhQ in plant plasma membrane (Lochner et al., 2003; Gross et al., 2006). However, PhQ is found predominantly in thylakoids and plastoglobules (Lohmann et al., 2006). Experimental support for a plasma membrane localization and/or function(s) of PhQ remains scarce.

PhQ is synthesized by a series of "Men" proteins named after their eubacterial homologs: MenF (isochorismate synthase, ICS), MenD (2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase), MenH (2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase), MenC (*o*-succinylbenzoate synthase), MenE (*o*-succinylbenzoyl-CoA synthase), MenB (1,4-dihydroxynaphthoyl-CoA synthase), DHNAT(1,4-dihydroxynaphthoyl-CoA thioesterase), MenA (1,4-dihydroxynaphthoate phytyltransferase), and MenG (demethylphylloquinone methyltransferase) (reviewed in (Van Oostende et al., 2011)). In plants, *MenD*, *MenH* and *MenC,* along with a truncated *MenF*, are fused into a composite gene named *PHYLLO* (Gross et al., 2006). The early (ICS and PHYLLO) and late (MenA and MenG) steps of PhQ biosynthesis occur in the plastid (Shimada et al., 2005; Gross et al., 2006; Lohmann et al., 2006; Garcion et al., 2008; Kim et al., 2008), whereas enzymes catalyzing the intermediate steps are either dually targeted to plastids and peroxisomes (MenE) (Kim et al., 2008; Babujee et al., 2010), or are exclusively peroxisomal (MenB and DHNAT)(Reumann et al., 2007; Babujee et al., 2010; Widhalm et al., 2012). This suggests that there is intracellular trafficking of PhQ pathway intermediates (Figure 1). The subcellular localization patterns of PhQ pathway enzymes are corroborated by the presence of corresponding signal peptides for

plastidic and/or peroxisomal targeting (Shimada et al., 2005; Reumann et al., 2007; Babujee et al., 2010; Widhalm et al., 2012).

We have observed in multiple photosynthetic taxa that PhQ pathway genes have measurable expression in heterotrophic tissues where photosynthetic genes were weakly or negligibly expressed (see Chapter 4). This raises the possibility that PhQ may have function(s) outside the PSI. However, decoupling the presumed non-photosynthetic function of PhQ from its dominant role in photosynthesis has been challenging. We therefore exploited parasitic plants for investigating the possibility, reasoning that retention of the PhQ pathway in an obligate, non-photosynthetic parasite would provide evidence in support of a non-photosynthetic role of PhQ. Parasitic plants are classified into three major groups based on the degree of host dependency and photosynthetic capacity: 1) facultative parasites that are fully photosynthetic, capable of completing their lifecycle independently, but will take advantage of the hosts when available; 2) obligate hemiparasites that are partially photosynthetic and require the presence of hosts for their development; and 3) obligate holoparasites that are non-photosynthetic and obtain all of their carbon from the hosts (Irving and Cameron, 2009). Here, we report biosynthesis and subsequent accumulation of PhQ in the holoparasite *Phelipanche aegyptiaca*. Alternative targeting to the plasma membrane supports a role of PhQ in plasma membrane electron transport chain. The data suggest a previously unidentified link between PhQ and cellular oxidation-reduction processes associated with parasitic haustorial development.

Materials and Methods

Transcriptome assembly of parasitic plants

RNAseq data of *Triphysaria versicolor*, *Striga hermonthica* and *Phelipanche aegyptiaca* were downloaded from the Parasitic Plant Genome Project (PPGP, http://ppgp.huck.psu.edu/) database. Raw data were pre-processed using an in-house pipeline to remove adapters, non-coding RNAs and low-quality reads. Reads with low complexity were removed by dustMasker of NCBI BLAST+ 2.2.29 to reduce assembly

complexity. Cleaned reads were then assembled using Parallelized Local Assembly of Sequences (PLAS). Transcriptomes of host plant, protozoa, invertebrate, bacteria, fungi and human were downloaded from NCBI to remove potential contaminations in the assembly. Redundant contigs sharing at least 95% sequence identity were also removed from the transcriptome. The transcriptome was annotated after BLASTing the sequences against the *Arabidopsis* proteome, *Mumulus* proteome and Uni-prot database. Transcript abundance was estimated using eXpress 1.5.1 (Roberts and Pachter, 2013). Additional MenA and MenG sequences were obtained from the 1000 Plants database (Matasci et al., 2014) by BlastN (https://db.cngb.org/blast4onekp/) against the 'Core Eudicots/Asterids' clade using P. aegyptiaca sequences as query.

Subcellular and transmembrane domain prediction and gene structure

PhQ gene sequences of photosynthetic species were downloaded from Phytozome v11 (https://phytozome.jgi.doe.gov). Subcelluar localization was predicted by Predotar 1.04 (Small et al., 2004), TargetP 1.1 (Emanuelsson et al., 2007), Protein Prowler 1.2 (Boden and Hawkins, 2005), and WolF PSORT (Horton et al., 2007). Transmembrane domain was predicted by TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). The predicted scores were exported and plotted in R. Gene annotation files were downloaded from Phytozome v11 (https://phytozome.jgi.doe.gov/pz/portal.html). Gene structures were drawn by Gene Structure Display Server (GSDS) 2.0 (http://gsds.cbi.pku.edu.cn/). Sequence alignment was performed with Clustal Omega 1.2.1 (Sievers et al., 2011) and visualized using Color Align Conservation (http://www.bioinformatics.org/sms2/color_align_cons.html).

Co-expression analysis

The transcriptomes were first filtered to remove components with low or invariant expression profiles. Only transcripts with FPKM values ≥ 2 in at least two samples and with a coefficient covariance no less than 0.35 were retained. Pair-wise Gini Correlation Coefficient was calculated using an in-house python script. PhQ-co-expressed transcripts were defined as the 500 most highly correlated transcripts or those with a GCC ≥0.8 for each PhQ pathway gene. The union sets were used for Gene Ontology enrichment analysis using topGO R

package 2.26.0 (Alexa and Rahnenfuhrer, 2010). To facilitate comparative analysis between the three species, ortholog groups were detected by OrthoFinder 1.0.8 (Emms and Kelly, 2015). Network visualization was performed in Cytoscape 3.4.0 (Shannon et al., 2003) using edge-weighted spring embedded layout, with a GCC cutoff of 0.6.

RNA-seq data processing of photoautotrophic species

RNA-seq data of *Arabidopsis thaliana*, *Glycine max* and *Populus tremula x alba* were downloaded from the NCBI Short Read Archive and processed by Cutadapt 1.9.dev1 (Martin, 2011), Trimmomatic 0.32 (Bolger et al., 2014) and custom scripts to remove adapter, low-quality reads, rRNA and organellar sequences. Reads were mapped by Tophat 2.0.13 (Kim et al., 2013), alignment sorted by Samtools 1.2 (Li et al., 2009), and read count and expression estimation obtained by HTseq 0.6.1p1 (Anders et al., 2015) and DESeq2 (Love et al., 2014). *Arabidopsis thaliana* datasets used for GCC computation were SRA236885, SRA091517, SRA269936, SRA219425, SRA308579, SRA050132, SRA067724, SRA291734, SRA269101, SRA098075, SRA100242, SRA122395, SRA163488, SRA064368, SRA246225, SRA248861, SRA202878, SRA201550, SRA303151, SRA221137, SRA272654, and SRA221060 (stressed samples were excluded). *Glycine max* datasets included SRA187830, SRA047293, SRA036577, SRA116533, SRA091756, SRA187830, SRA036538, SRA036577, and SRA129337. *Populus tremula x alba* datasets were SRA274261 and SRA097208.

Phylogenetic Tree Construction

The protein sequences of *Phelipanche ramose* OrPRX1 (AAY89058) and OrPOX1 (AAU04440), *Striga asiatica* SaPOXA (AAB97853) and SaPOXB (AF043235) were searched against the transcriptomes of *Triphysaria versicolor*, *Striga hermonthica* and *Phelipanche aegyptiaca* to identify orthologs. Their protein sequences were aligned by MUSCLE 3.8.31 (Edgar, 2004) and the alignments were cleaned by Gblocks. Bayesian phylogenetic tree was constructed by MrBayes 3.2.5 (Huelsenbeck and Ronquist, 2001; Ronquist et al., 2012).

Results and Discussion

Detection of PhQ biosynthetic genes in parasitic plants.

Among angiosperm parasite families, only the Orobanchaceae contains species that span the full spectrum of photosynthetic capacities, and for which rich transcriptomic resources (Westwood et al., 2012) for exploring the evolution and function of PhQ biosynthesis are available. To this end, PhQ protein sequences of *Mimulus* (family Phrymaceae), a photosynthetic relative of Orobanchaceae, were searched against the transcript assemblies available on the Parasitic Plant Genome Project (PPGP) website using TBLASTX. Full-length coding sequences were identified for *ICS* and *MenE* genes in *P. aegyptiaca*, along with partial assemblies of other PhQ pathway genes. This supports the possibility that some PhQ genes are transcriptionally active in the parasitic plants. However, fragmented or incomplete assembly of other PhQ transcripts prevented confirmation that a complete and functional PhQ pathway exists in parasitic plants and further assessment of expression and functions.

To address the *de novo* assembly challenge, a target-restricted assembly approach (Allen et al., 2015) was adopted to develop a “Parallelized Local Assembly of Sequences” (PLAS) pipeline for transcriptome-wide applications with parallel computing (Chapter 2). When applied to the parasitic RNA-Seq datasets from PPGP (Yang et al., 2015), I successfully recovered full-length or near full-length transcripts for all PhQ genes with intact open reading frames from all three parasitic species. These transcripts showed moderate to high abundances in most tissues examined (Figure 3.1). Two multifunctional genes recently implicated in PhQ biosynthesis were either poorly expressed (NAD(P)H DEHYDROGENASE C1, *NDC1*, FPKM<10) or not recovered (PHYTYL-PHOSPHATE KINASE, *VTE6*) in the holoparasite. These genes are also involved in the biosynthesis (*VTE6*) and redox cycle (*NDC1*) of plastid α-tocopherol (Fatihi et al., 2015; Wang et al., 2017). The interconnection between PhQ biosynthesis and other plastid-derived metabolites (Basset, 2016) is therefore absent or reduced in the holoparasite. While *PHYLLO*, *MenE* and *DHNAT* transcripts were detected at similar levels between species, *ICS*, *MenB*, *MenA* and *MenG* transcript levels were much more abundant in the non-photosynthetic *P. aegyptiaca* than the photosynthetically

competent *S. hermonthica* and *T. versicolor*, especially during seed germination and haustorial development (Figure 3.1). The data strongly support a role for PhQ beyond photosynthesis, and hint at divergent regulation of PhQ biosynthesis associated with photosynthetic and non-photosynthetic functions. HPLC analysis performed by Batbayar Nyamdari and Scott Harding confirmed the presence of PhQ in germinated *P. aegyptiaca* seeds prior to haustorial initiation (Figure 3.2). Retention and active transcription of PhQ genes and accumulation of PhQ in the holoparasite argues that PhQ has non-photosynthetic functions.

Subcellular prediction of PhQ proteins in the holoparasite

The predicted polypeptides of MenA and MenG involved in the last two steps of PhQ biosynthesis were substantially shorter in *P. aegyptiaca* than in autotrophic species. Sequence alignment with the *Arabidopsis thaliana* orthologs revealed a high level of conservation for both proteins, except for their N-termini (Figure S3.4-3.5). Since the N-terminus of AtMenA and AtMenG harbors a transit peptide for plastid-localization (Shimada et al., 2005; Lohmann et al., 2006), N-truncation of PaMenA and PaMenG may impact their signal peptides and affect subcellular destination of PhQ. To test this possibility, the deduced PhQ protein sequences from parasitic and autotrophic species were subjected to plastid localization prediction analysis. Because not all subcellular localization predicting tools achieved accurate predictions of the experimentally verified plastidic ICS, PHYLLO, MenA and MenG from *Arabidopsis* (Shimada et al., 2005; Gross et al., 2006; Lohmann et al., 2006; Garcion et al., 2008), four different programs were used and proteins with a high prediction score from at least one program were deemed potentially plastid-localized (Figure 3.3A, Table S3.1). Orthologs of ICS, PHYLLO, MenA and MenG from the other autotrophic species were all predicted to be plastid-localized (Figure 3.3A, Table S3.1), consistent with the involvement of PhQ in PSI electron transport. In the holoparasite, however, only PaICS and PaPHYLLO that catalyze the early steps of PhQ biosynthesis were predicted to be plastidic (Table S3.1). By contrast, the N-truncated PaMenA and PaMenG scored poorly for plastid targeting with all four prediction programs (Figure 3.3A). The penultimate step catalyzed by MenA occurs at

the thylakoid membrane of photosynthetic species (Schultz et al., 1981; Kaiping et al., 1984), consistent with the prediction of AtMenA as an integral membrane protein (Figure 3.4). The N-truncated PaMenA was also predicted to contain eight transmembrane domains (Figure 3.4). The absence of an N-terminal plastidic targeting peptide in PaMenA thus suggests that it is likely localized to other cell membranes.

It has recently been shown that the intermediate steps of PhQ biosynthesis catalyzed by MenE, MenB and DHNAT occur in peroxisomes of green plants (Figure 3.1) (Reumann et al., 2007; Babujee et al., 2010; Widhalm et al., 2012). The experimentally verified peroxisomal AtMenE and AtDHNAT harbor the peroxisome targeting signal PTS1 at their C-termini (SSL> and AKL>, respectively), while AtMenB contains the peroxisome targeting signal PTS2 (RLX_5HL) at its N-terminus (Babujee et al. 2010; Widhalm et al. 2012; Reumann et al. 2007). The parasitic orthologs were also predicted to be peroxisomal proteins, harboring the conserved PTS1 in the cases of MenE and DHNAT, or PTS2 in the case of MenB (Figure S3.1-3.3). Together, the data support compartmentalization of PhQ biosynthesis between plastids and peroxisomes before delivery to thylakoid membranes in photosynthetic species. In the holoparasite, however, the post-peroxisome route likely involves other cellular membranes. Plasma membranes are an attractive target, based on the reported occurrence of PhQ there (Lüthje and Böttger, 1995; Lüthje et al., 1998).

The PM localization of PaMenA was demonstrated by expressing *35S:PaMenA-GFP* in stably transformed *Nicotiana benthamiana* plants (Figure 3.5A). The GFP experiment was performed by Kavita Aulakh and Naomi Rodman. The N-truncated *PaMenG-GFP* fusion was observed to be localized to PM in transgenic *N. benthamiana* leaves (Figure 3.5B). The data indicated a redirection of PhQ biosynthesis to PM in the non-photosynthetic *P. aegyptiaca*. To bolster this finding, we mined the 1000 Plants database (Matasci et al., 2014) for additional parasitic orthologs. We found that *MenA* and *MenG* transcripts from several other holoparasites, including *Phelipanche fasciculata* and *Conopholis americana* of Orobanchaceae, and *Cuscuta pentagona* of Convolvulacea, also lack plastid-targeting sequences (Figure S3.4-3.5). By contrast, the predicted N-termini of MenA and MenG from

closely-related photosynthetic taxa resemble those of *S. hermonthica* and *T. versicolor* (Figure S3.4-3.5). The results suggested convergent evolution of N-truncated MenA and MenG, and hence PM-PhQ biosynthesis, in unrelated holoparasites.

Dual subcellular localization of PhQ proteins in photosynthetic species

Interestingly, the *Arabidopsis AtMenA* and *AtMenG* are annotated with alternative transcripts (TAIR10 genome release) predicted to encode N-terminal truncated isoforms (Figure 3.3B), similar to what we observed for *PaMenA* and *PaMenG*. The alternative *AtMenA* (At1g60600.1) and *AtMenG* (At1g23360.2) transcripts were supported by EST (AV832198 and AV829761, respectively). To investigate whether 5' alternative splicing also occurs in other taxa, we surveyed all sequenced genomes available at Phytozome (v11) for alternative transcripts of *MenA* and *MenG*. Indeed, multiple dicot and monocot species showed alternative splicing at both loci, giving rise to N-truncated isoforms (Figure 3.3B). When subjected to subcellular localization prediction, the truncated MenA and MenG isoforms scored poorly for plastidic localization with all four prediction programs (Figure 3.3A). The findings support alternative, non-plastidic localization of MenA and MenG, and hence PhQ, in photosynthetic species. In agreement with the prediction, stable expression of *35S:AtMenA.1-GFP* in transgenic *N. benthamiana* leaves showed localization to PMs (Figure 3.5C) instead of plastids as reported for AtmenA.2 (Shimada et al., 2005). Together, the results suggest that non-plastidic targeting of MenA and MenG is evolutionarily conserved in angiosperms. Dual subcellular localization in photosynthetic species is afforded via alternative splicing. In holoparasites that are devoid of photosynthesis, relaxed selection in plastid-targeting sequences might have led to accumulation of mutations, resulting in degeneration of the plastidic signal peptide and exclusive plasma membrane localization, as shown for PaMenA and PaMenG. The fact that both PaMenA and PaMenG share a high level of sequence similarity (~76%) with the predicted mature protein of their *Arabidopsis* orthologs is consistent with a strong selective pressure to retain a functioning PhQ pathway, presumably to fulfill non-photosynthetic functions in holoparasites.

Co-expression patterns of PhQ genes differed between parasitic plants

To shed light on potential non-photosynthetic functions of PhQ in parasitic plants, we computed pairwise Gini Correlation Coefficient (GCC) among QC-filtered transcripts in each species. The expression of PhQ genes was highly coordinated in photosynthetic species like *Arabidopsis*, Soybean and *Populus* (Figure S3.6), presumably for synthesis of the plastid PhQ pool dominant in the sampled tissues. High levels of co-expression were also observed in the holoparasite *P. aegyptiaca* (Figure 3.6A), suggesting that biosynthesis of plasma membrane PhQ for non-photosynthetic functions is also tightly co-regulated. However, such coordination was not observed in photosynthetically competent parasites, particularly between early- and late-pathway genes (Figure 3.6B-C). Given the attenuated photosynthesis in these species (Wickett et al., 2011), weakened co-expression among PhQ genes is consistent with a heterogeneous PhQ pool from both the plastid and plasma membrane routes in *S. hermonthica* and *T. versicolor*.

We extracted the top 500 most highly correlated transcripts for each PhQ pathway gene, and the union set contained 2447, 3677 and 3930 unique transcripts for *P. aegyptiaca*, *S. hermonthica* and *T. versicolor*, respectively (hereafter referred to as PhQ-coexpressed transcripts or PhQ-CET). The smaller PhQ-CET set of the holoparasite is consistent with stronger coexpression of PhQ genes when compared to *S. hermonthica* and *T. versicolor* as described above (Figure 3.6A-C). Subsets of PhQ-CETs with GO (Biological Process) annotation (645, 1199 and 1173 for *P. aegyptiaca*, *S. hermonthica* and *T. versicolor*, respectively) were subject to functional enrichment analysis. GO terms that were enriched in at least two species were retained for comparison. Transcripts associated with photosynthesis-related processes comprised 3-4% of the GO-annotated PhQ-CETs in *S. hermonthica* and *T. versicolor*, but were negligible in *P. aegyptiaca* (Figure 3.6D). In contrast, the proportions of transcripts associated with oxidation-reduction process, protein phosphorylation, and defense response were higher in *P. aegyptiaca* than *S. hermonthica* or *T. versicolor* (Figure 3.6D). Similar patterns were observed when we used a different criterion (GCC≥0.8) to define the PhQ-CETs (Figure S3.7). With the gradual decline and eventual loss

of photosynthesis capacity from *T. versicolor* to *P. aegyptiaca*, non-photosynthetic functions of PhQ are expected to become more enriched in *P. aegyptiaca*, due to lessened masking by photosynthesis-related functions. On this basis, plasma membrane-destined PhQ are likely involved in oxidation-reduction and defense-related functions.

To further explore the functional evolution of PhQ genes, we focused on PhQ-CETs assigned to oxidation-reduction, defense response, response to biotic stimulus, as well as photosynthesis GO terms for gene coexpression network analysis. To facilitate comparative analysis, we included orthologs from all three parasitic species based on orthogroups that were constructed using OrthoFinder 0.2.5 (Emms and Kelly, 2015). This resulted in 359, 544 and 560 non-redundant transcripts from *P. aegyptiaca*, *S. hermonthica* and *T. versicolor*, respectively. Network visualization of their co-expression patterns revealed striking differences between photosynthetically competent and incompetent parasites. Two dense modules were detected for *S. hermonthica* and *T. versicolor*, one of them encompassing most of the PhQ-coexpressed photosynthesis genes (Figure 3.7, green nodes). However, the network topology was distinctly different for *P. aegyptiaca* that is devoid of phostosynthesis (Figure 3.7). The PhQ genes were highly interconnected in the *P. aegyptiaca* network (seven orange-colored nodes), but were scattered over the *S. hermonthica* (nine nodes) and *T. versicolor* (10 nodes) networks (Figure 3.7**)**, consistent with the correlation patterns of PhQ genes shown in Figure 3.6A-C. We ranked genes by the number of edges they shared with PhQ genes (referred to as EG_{PhQ}) in each network, and observed a striking enrichment of PhQ-interconnected genes in the smaller *P. aegyptiaca* network. More than 23% of *P. aegyptiaca* nodes had an EG_{PhQ} =4-6 (*i.e.*, connected with a majority of the PhQ genes). However, less than 3% of the *S. hermonthica* and *T. versicolor* nodes met the same criterion (EG_{PhQ} ≥5 of 9-10 PhQ genes), and only 10 and 15% of their respective nodes had an EG_{PhQ} ≥4 (Figure 3.7, vertical bars). Close examination of the *P. aegyptiaca* PhQ-subnetwork genes identified several candidates potentially involved in haustorium signaling and plasma membrane electron transport associated with parasitism, as discussed below.

PhQ Association with Parasitism

The Class III secretory peroxidases are of particular interest because of their potential involvement in oxidation of cell wall-derived phenols and generation of reactive oxygen species (ROS, such as H_2O_2) during early development of parasitic plants (González-Verdejo et al., 2006; Keyes et al., 2007; Lynn and Chang, 1990). Specifically, two peroxidase genes from *S. asiatica* (*SaPOXA* and *SaPOXB*) were previously shown to be highly induced by haustorium-inducing factors (HIFs), and their encoded proteins were capable of oxidizing a range of host cell wall-derived phenolics into benzoquinones necessary for haustorial induction (Kim et al., 1998). The *P. ramose* orthologs *PrPOX1* and *PrPRX1* were also specifically expressed during early development that coincided with active secretion of peroxidase enzymes (González-Verdejo et al. 2006; Veronesi et al. 2007). In the present study, seven corresponding orthologs of the peroxidases were identified from the three parasitic species investigated here (Figure 3.8A, blue clade). Transcript levels of the *P. aegyptiaca* orthologs were one to two orders of magnitude higher than those of *S. hermonthica* and *T. versicolor*, especially during seed germination and haustorial initiation (Figure 3.8B), reminiscent of the patterns observed for several PhQ genes (Figure 3.1). Network connectivity with PhQ genes was highest for the *P. aegyptiaca* orthologs (EG_{PhQ} = 5 and 3), followed by *S. hermonthica* orthologs (EG_{PhQ} = 4 and 3). However, PhQ-coexpression was not observed for the *T. versicolor* orthologs, or orthologs in the neighboring clade of the phylogenetic tree (Figure 3.8A, C). Thus, both the transcript levels and PhQ-coexpression of the secretory peroxidase were both positively correlated with parasitism. Another parasitism gene *QR1* encodes an NAD(P)H-dependent quinone reductase (Bandaranayake et al., 2010), which reduces host-derived quinones into highly reactive semiquinones necessary for haustorium induction via a plasma membrane-localized electron transport chain (Keyes et al., 2000). *QR1* transcripts were identified in all three parasites, but were not coexpressed with PhQ genes. In *P. aegyptiaca* specifically, *QR1* transcript levels were higher in imbibed than germinated seeds, whereas PhQ gene expression peaked at/after seed germination (Figure 3.1 vs. 3.8D). These

findings placed PhQ and PhQ-coexpressed genes downstream of *QR1* in haustorium signaling.

PhQ Involvement in Plasma Membrane Electron Transport

We next explored the *P. aegyptiaca* PhQ subnetwork for redox proteins involved in transmembrane electron flow. Membrane-associated NAD(P)H-oxidoreductases (QRs/NQRs) are an integral component of electron transport (reviewed in Keyes et al., 2000; Moller and Lin, 1986). Two groups of flavin-containing QRs/NQRs are potential candidates, one represented by *QR2* (Wrobel et al., 2002) and its *Arabidopsis* orthologs/genome duplicates At5g54500 and At4g27270, and the other by *NQR1*/At3g27890 (Heyno et al., 2013). The parasitic *NQR1* orthologs were not captured in the PhQ networks. However, *QR2* orthologs exhibited strong coexpression with PhQ genes in obligate parasites *P. aegyptiaca* (EG_{PhQ} =4) and *S. hermonthica* (EG_{PhQ} =3) (Figure 3.8F), in line with *QR2* responsiveness to HIFs during haustorium formation of *S. asiatica* and *Phtheirospermum japonicum* (Ishida et al., 2016; Liang et al., 2016). In soybean, the NQR1 ortholog was found in plasma membranes, whereas QR2-like immunosignals were detected in the cytosolic protein fraction (Schopfer et al., 2008). Paradoxically, the *Arabidopsis* QR2 orthologs, but not NQR1, are detected in the plasma membrane (Marmagne et al., 2004; and see Table S1 in Marmagne et al., 2007). These and our finding of strong *QR2* co-expression with PhQ genes in obligate parasites raised the possibility that QR2 participate in the plasma membrane redox system.

Membrane-bound NAD(P)H oxidase (NOX) is another key component of the electron transport chain, analogous to mammalian systems (Bridge et al., 2000; Keyes et al., 2000; Lochner et al., 2003). Partially purified NADP(H) oxidase from soybean plasma membranes was shown to catalyze oxidation of reduced PhQ (Bridge et al., 2000). Despite this early finding and despite extensive studies of the plant 'respiratory burst oxidase homologs' (Rboh)(Kaur et al., 2014), identity of the exact oxidases involved in the plasma membrane electron transfer chain remains elusive. We found only one *NOX* ortholog in the *S. hermonthica* network (k_{PhQ} = 3), and its counterpart in *S. asiatica* (*SaNOX1*) was recently shown to be root-specific and HIF-responsive (Liang et al., 2016). The *P. aegyptiaca*

transcriptome lacked *NOX1*, but a ferric-chelate reductase (*PaFRO1*) exhibited strong PhQ-coexpression (k_{PhQ} = 4) (Figure 3.8G). Interestingly, *FRO1* transcript was not recovered in the *NOX1*-harboring *S. hermonthica*. FROs are plasma membrane-localized and like Rbohs, belong to the flavocytochrome superfamily involved in electron transport (Sagi and Fluhr, 2006). The PhQ-coexpressed *PaFRO* is orthologous to the *Arabidopsis FRO4/FRO5* tandem duplicates that encode root surface copper-chelate reductase necessary for copper acquisition from the soil (Bernal et al., 2012). Some of these associations are consistent with previous reports that the plasma membrane NAD(P)H oxidase is tightly coupled to auxin-stimulated growth and resides on the cell surface (Brightman et al., 1988; DeHahn et al., 1997). This suggests that the transmembrane redox system for copper uptake might have been co-opted for parasitic signaling and haustorium development in *P. aegyptiaca* following its divergence from *Striga*. The NOX/FRO may also facilitate redox exchange for disulfide bond formation in oxidative protein folding (Bridge et al., 2000), a process that has been shown to involve PhQ as a cofactor in both plants and cyanobacteria (Furt et al., 2010; Li et al., 2010).

Conclusions

This study presents multiple lines of evidence at the transcriptional, protein subcellular localization and metabolite levels to support active biosynthesis of PhQ in the non-photosynthetic holoparasite *P. aegyptiaca*. Plasma membrane-destined PhQ appears to be evolutionarily conserved in angiosperms. In autotrophic dicots and monocots, alternative splicing of *MenA* and *MenG* results in truncated isoforms that are targeted to plasma membrane. In the heterotrophic *P. aegyptiaca*, non-plastid targeting of MenA and MenG is the default due to loss of the N-terminal transient peptides. Retention of the alternative-targeting pathway in the holoparasite would presumably confer upon PhQ a beneficial non-photosynthetic function, most likely, we argue, in plasma membrane electron transport.

The long-proposed role of PhQ in plasma membrane electron transport of higher plants (Lüthje et al., 1998; Bridge et al., 2000; Lochner et al., 2003) gains molecular support in this study. The plasma membrane-localized redox machinery has been a missing link in

understanding haustorium signaling and parasitism (Boone et al., 1995; Keyes et al., 2000). The PhQ-centered data mining presented in this work now offers a rich source of candidate genes for hypothesis-driven research to ascertain their roles in transmembrane redox regulation. Our comparative analyses also revealed previously unrecognized transcriptional dose responses across parasitic species with different levels of photosynthetic capability. Expression of many PhQ network genes was found to be positively correlated with parasitism (highest in *P. aegyptiaca* followed by *S. hermonthica* and then *T. versicolor*). Because host attachment within days after germination is most critical for obligate parasites, these species have evolved advanced host recognition and haustorium signaling systems that are tightly coupled to sophisticated seed germination requirements to ensure survival (Keyes et al., 2001; Yoder, 2001). Data presented here suggest that these requirements involve complex redox regulation with plasma membrane PhQ as a key player. Given our finding of the conservation of plasma membrane PhQ in angiosperm evolution, the PhQ role in parasitic signaling can shed light on its non-photosynthetic function in autotrophic species.

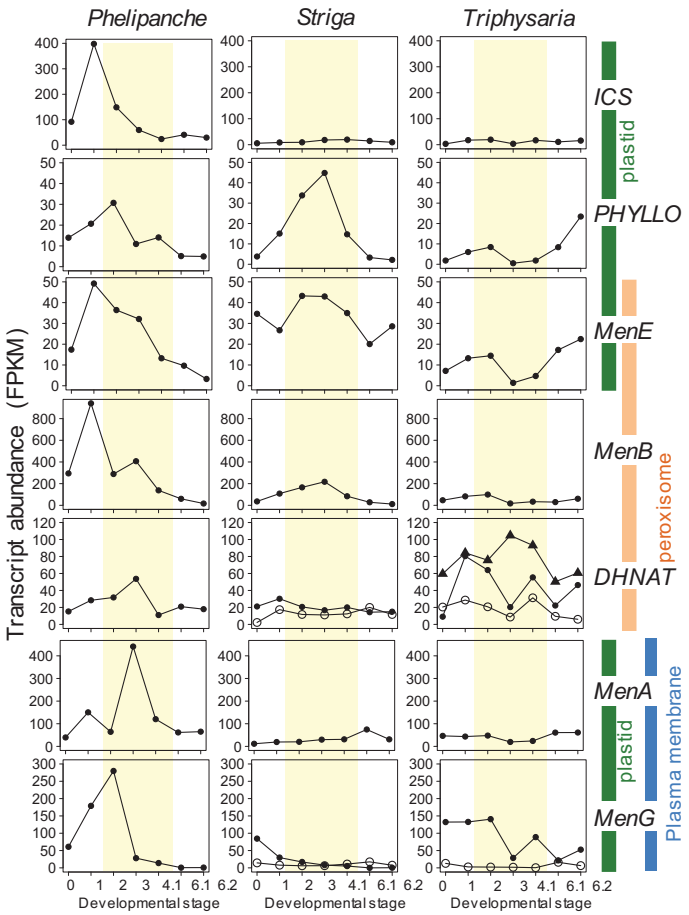


Figure 3.1. **Expression profiles of phylloquinone biosynthetic genes in parasitic plants *Phelipanche aegyptiaca* (holoparasite), *Striga hermonthica* (hemiparasite) and *Triphysaria versicolo* (facultative parasite)**. 0, imbibed seeds; 1, germinated seedlings after exposure to GR24 for *Striga hermonthica* and *Phelipanche aegyptiaca*, or roots of germinated *Triphysaria versicolor* seedlings; 2, seedlings after exposure to haustorial inducing factors for *Striga hermonthica* and *Phelipanche aegyptiaca* or germinated roots of *Triphysaria versicolor*; 3, haustoria attached to host root prior to vascular connection; 4.1 haustoria attached to host root after vascular connection; 6.1 leaves/stems; 6.2, floral buds. The predicted or experimentally verified subcellular localizations are color-coded on the right.

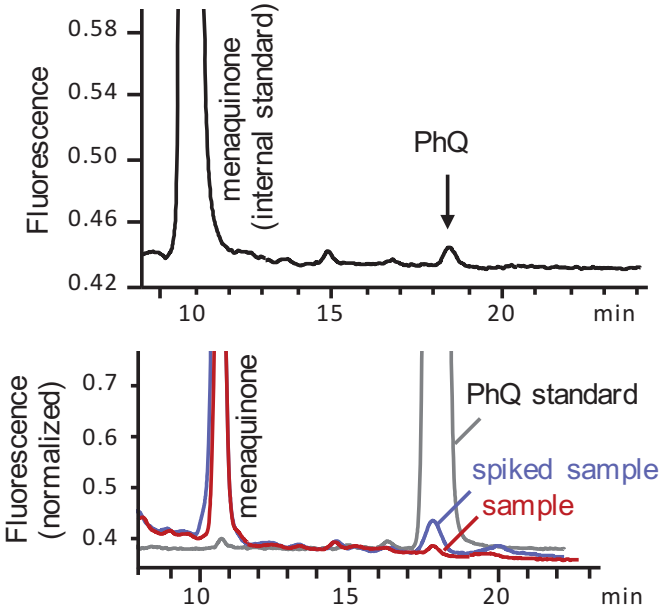


Figure 3.2. **HPLC detection of PhQ in germinated *Phelipanche aegyptiaca* seeds**. The lower panel overlays the authentic standard and sample with or without PhQ spike.

A

Gene model	Predotar	TargetP	Protein Prowler	WoIF PSORT	Prediction
MenA					
At1g60600.2*	0.9	0.5	0.0	0	plastid
At1g60600.1	0.0	0.0	0.0	0	-
Glyma.08G070600	0.2	0.4	0.5	9	plastid
Glyma.13G042800	0.3	0.6	0.6	1	plastid
Migut.B00584.1	1.0	0.7	0.2	0	plastid
Migut.B00584.3	0.2	0.0	0.0	0	-
Migut.B01155	0.6	0.4	0.2	10.5	plastid
Migut.B01157	0.6	0.4	0.2	0	plastid
Ciclev10028615m	0.2	0.5	0.7	13	plastid
Ciclev10028855m	0.1	0.1	0.0	2	-
Os03g09060	0.0	0.9	0.6	14	plastid
Bradi1g71860.1	0.0	0.8	0.5	12	plastid
Bradi1g71860.2	0.0	0.0	0.1	4	-
TvMenA	0.9	0.8	0.8	7	plastid
ShMenA	0.7	0.2	0.0	3	plastid
PaMenA	0.0	0.1	0.1	4	-
MenG					
At1g23360.1*	0.7	0.9	0.8	10	plastid
At1g23360.2	0.0	0.1	0.0	6	-
At1g23360.3	0.0	0.1	0.0	6	-
Glyma.01G028200	1.0	0.4	0.8	10.5	plastid
Glyma.02G037000.1	0.8	0.7	0.8	12.5	plastid
Glyma.02G037000.2	0.2	0.1	0.0	5	-
Migut.E00183	0.5	0.5	0.3	13	plastid
Ciclev10032509m	1.0	0.7	0.8	12	plastid
Ciclev10032796m	0.0	0.1	0.0	4	-
Os04g42870.2	1.0	0.9	0.9	13	plastid
Os04g42870.1	0.0	0.1	0.0	7	-
Os04g42870.4	0.0	0.1	0.0	3	-
Bradi5g15167.3	0.7	1.0	1.0	14	plastid
Bradi5g15167.4	0.0	0.0	0.0	6	-
TvMenG1	0.1	0.6	0.1	12	plastid
TvMenG2	0.0	0.1	0.0	5	-
ShMenG1	0.4	0.5	0.7	14	plastid
ShMenG2	0.0	0.1	0.0	7	-
PaMenG	0.0	0.1	0.0	7	-

B

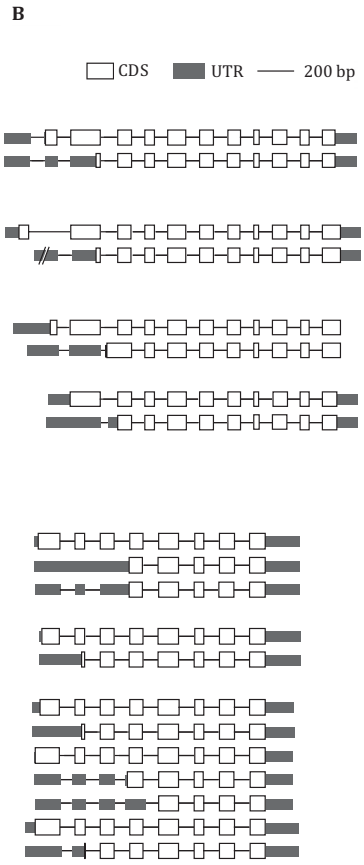


Figure 3.3. **Subcellular localization prediction of PhQ biosynthetic genes in parasitic plants and photoautotrophic plants.** (**A**) Plastid-targeting prediction of MenA and MenG polypeptides from various species. Heatmaps show prediction scores above the 50th percentile of each method, and asterisks denote experimentally verified plastidic proteins. (**B**) Exon-intron structures of representative angiosperm *MenA* and *MenG* genes with alternative splicing that affects the plastid transient peptides. Introns are not drawn to scale.

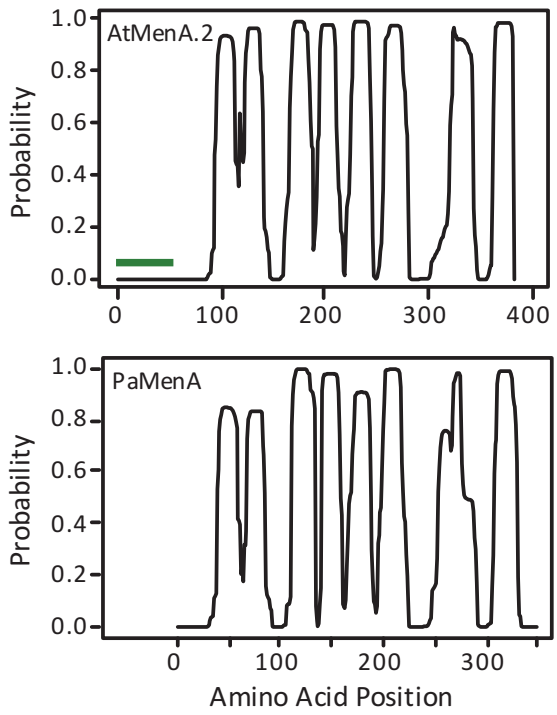


Figure 3.4. **Transmembrane domain prediction of AtMenA.2 and PaMenA.** The green line in the top panel corresponds to the plastid transient peptide, and the arrowhead denotes the start of alternative isoform AtMenA.1. The x-axis in the bottom panel is shifted for alignment with AtMenA.2.

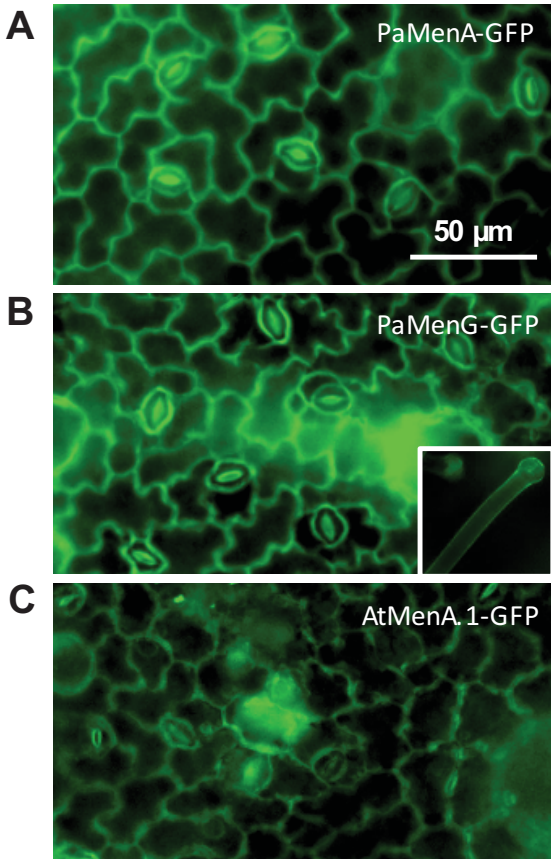


Figure 3.5. **Plasma membrane localization of GFP fusion. (A)** PaMenA, (**B**) PaMenG, and (**C**) AtMenA.1 (I). Inset in **(B)** shows trichome signal.

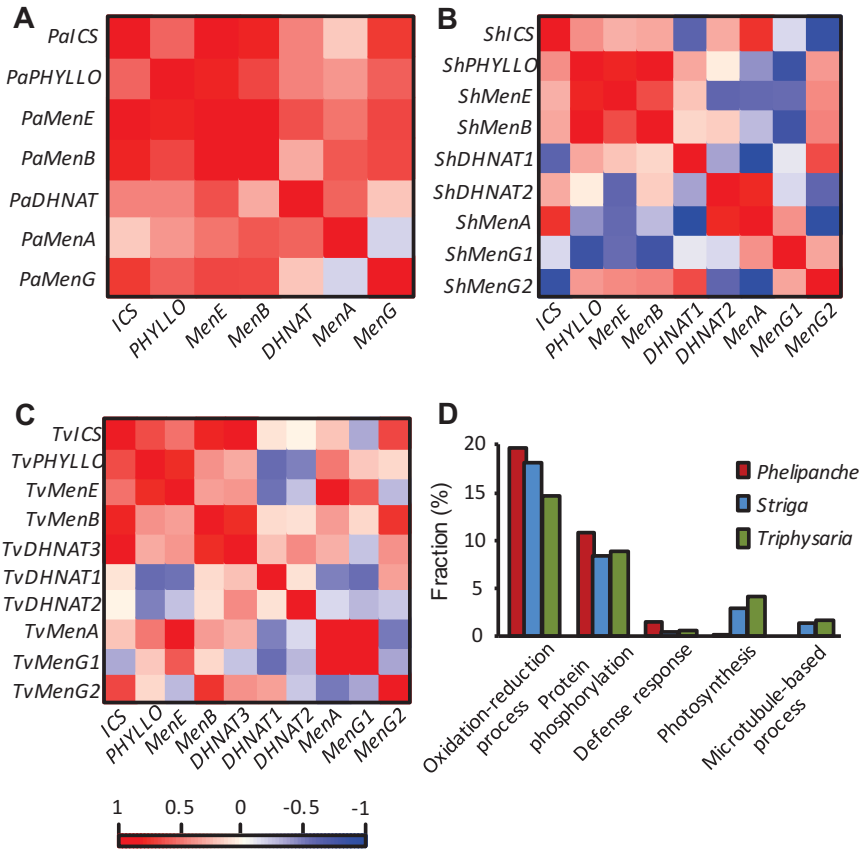


Figure 3.6. **Coexpression of PhQ genes.** (**A-C**) Coexpression patterns among PhQ biosynthesis genes based on Gini correlation coefficient (GCC). (**D**) GO enrichments of PhQ-coexpressed genes (union of top 500). Only GO terms with differential enrichment in *Phelipanche* relative to photosynthetically competent *Striga* and *Triphysaria* are shown. The distribution of top ten GO terms, and analysis using a gene coexpression cutoff of GCC≥ 0.8 are shown in figure S3.7.

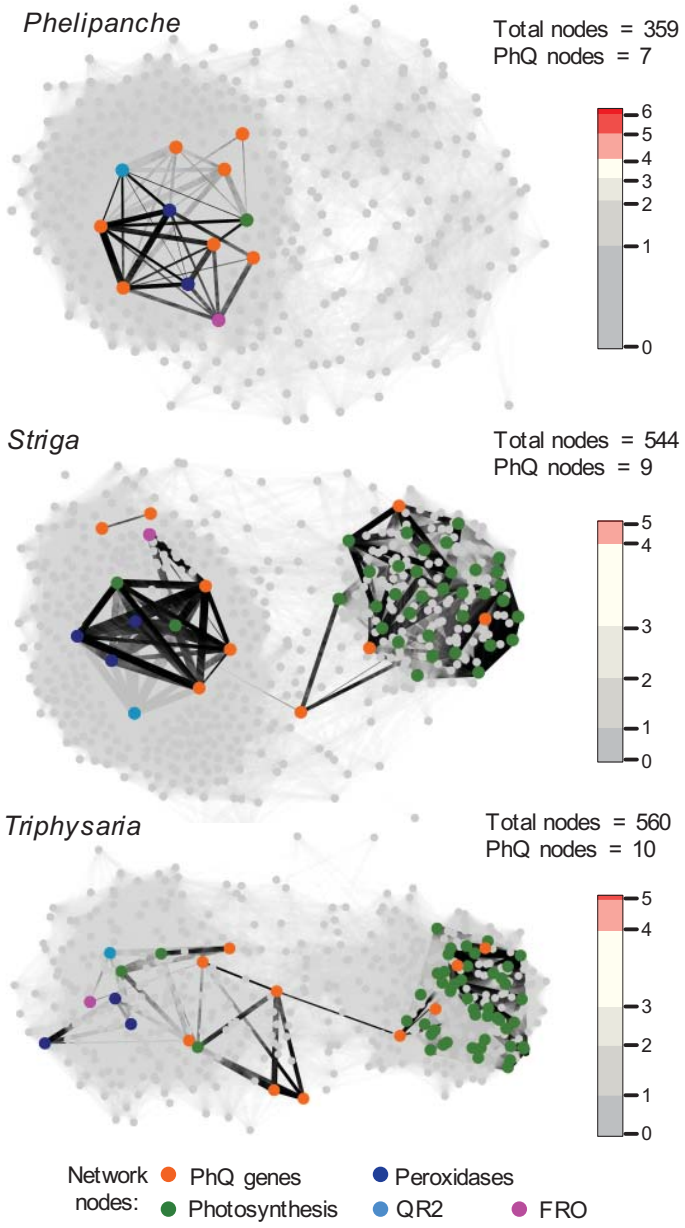


Figure 3.7. **Network visualization of PhQ-subnetworks of the parasitic plants**. PhQ genes are shown in orange, photosynthesis genes in green, peroxidases in blue, *QR2* in cyan and *FRO* in magenta. Edge thickness reflects the coexpression strength. Vertical bars depict the distribution of network genes according to their connectivity with PhQ genes (k_{PhQ}).

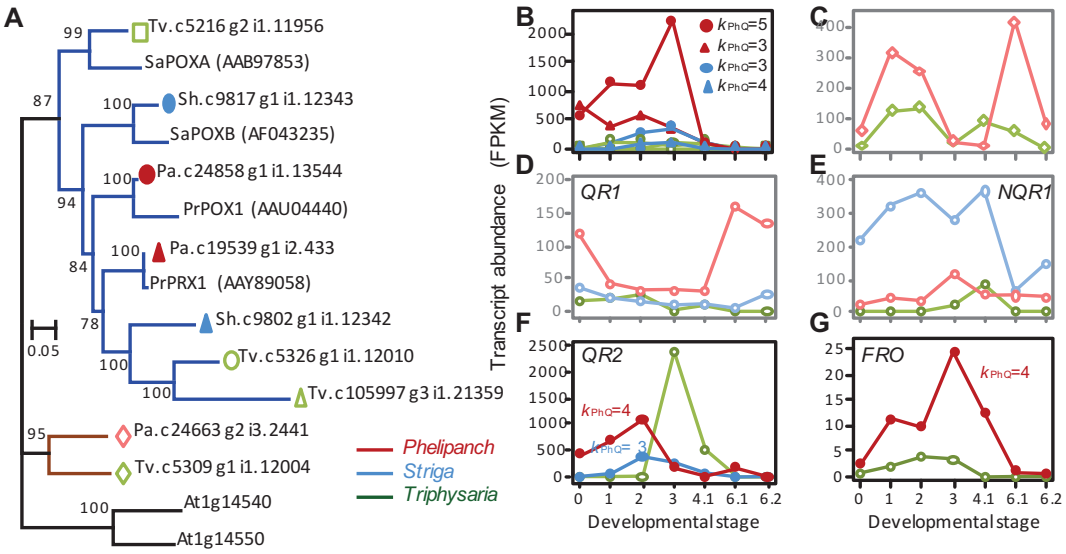


Figure 3.8. Bayesian phylogeny and PhQ-coexpressed genes. **(A)** Phylognetic tree of peroxidases in the three parasitic plants. Orthologs of experimentally characterized peroxidases cluster in the blue clade and color-coded by species. Their expression profiles are shown in **(B).** **(C)** Expression of unrelated peroxidases (brown clade in A). (**D-G**) Expression profiles of *QR1* (D), *NQR1* (E), *QR2* (F), and *FRO* (G) orthologs. Solid symbols and dark colors denote PhQ-coexpressed genes (k_{PhQ} ≥3). Genes not coexpressed with PhQ genes are shown in open symbols/light colors. Developmental stages are the same as in Figure 3.1.

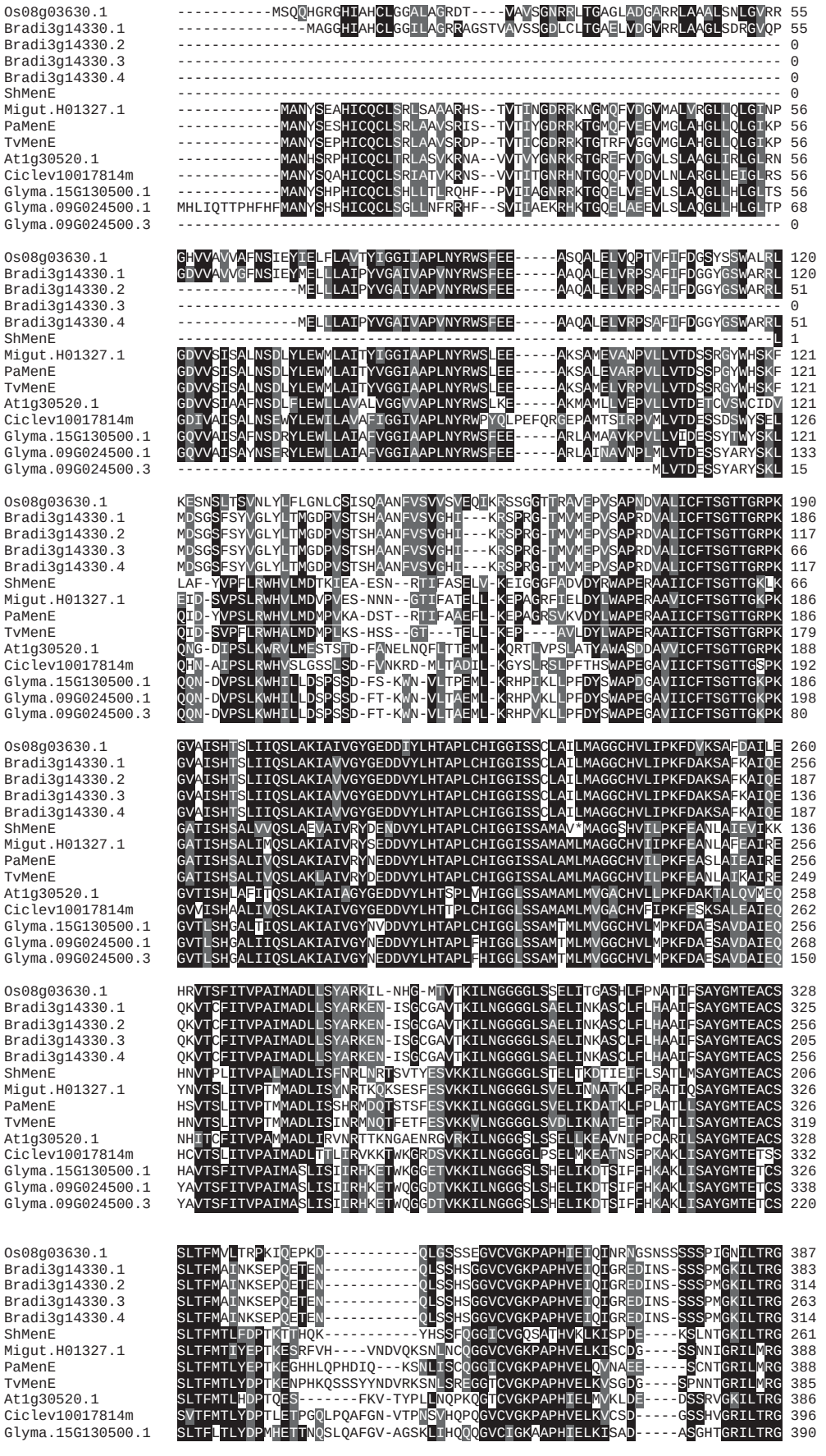
Table S3.1. Plastid-targeting Prediction for *ICS* and *PHYLLO*

	Predotar	TargetP	Protein Prowler	WoIF PSORT	Prediction
ICS					
AT1G18870.1*	0.79	0.95	0.97	13	plastid
AT1G18870.2	0.00	0.31	0.04	1	-
AT1G74710	0.82	0.77	0.99	11	plastid
Glyma.01G104100	0.19	0.20	0.00	12	plastid
Glyma.03G070600.1	0.16	0.19	0.01	14	plastid
Glyma.03G070600.3	0.02	0.09	0.01	1	-
Migut.I00130	0.70	0.16	0.08	10	plastid
Migut.I00129	0.00	0.03	0.00	13	plastid
Ciclev10024236m	0.29	0.75	0.87	3	plastid
Os09g19734	0.90	0.93	0.98	12	plastid
Bradi4g28670	0.77	0.57	0.30	12.5	plastid
TvICS	0	0.18	0	4	-
ShICS	0.01	0.3	0.07	6	plastid
PaICS	0.18	0.22	0.03	8	plastid
Phyllo					
AT1G68890*	0.37	0.94	0.89	12	plastid
Glyma.15G276800	0.01	0.16	0.04	9	plastid
Migut.L01140.1	0.20	0.64	0.31	13.5	plastid
Ciclev10030492m	0.00	0.60	0.04	7	plastid
Os02g37090 (DAA34846)[1]	0.01	0.60	0.54	12	plastid
Bradi3g47067	0.02	0.72	0.67	13	plastid
TvPHYLLO	0.37	0.45	0.89	11	plastid
ShPHYLLO	0.57	0.74	0.98	12	plastid
PaPHYLLO	0.63	0.401	0.28	13	plastid

Heatmaps show prediction scores above the 50th percentile of each method

*Experimentally verified for plastid-targeting.

[1] GenBank sequence DAA34846 was used due to erroneous gene model annotation in the reference genome


Os08g03630.1
Bradi3g14330.1
Bradi3g14330.2
Bradi3g14330.3
Bradi3g14330.4
ShMenE
Migut.H01327.1
PaMenE
TvMenE
At1g30520.1
Ciclev10017814m
Glyma.15G130500.1
Glyma.09G024500.1
Glyma.09G024500.3

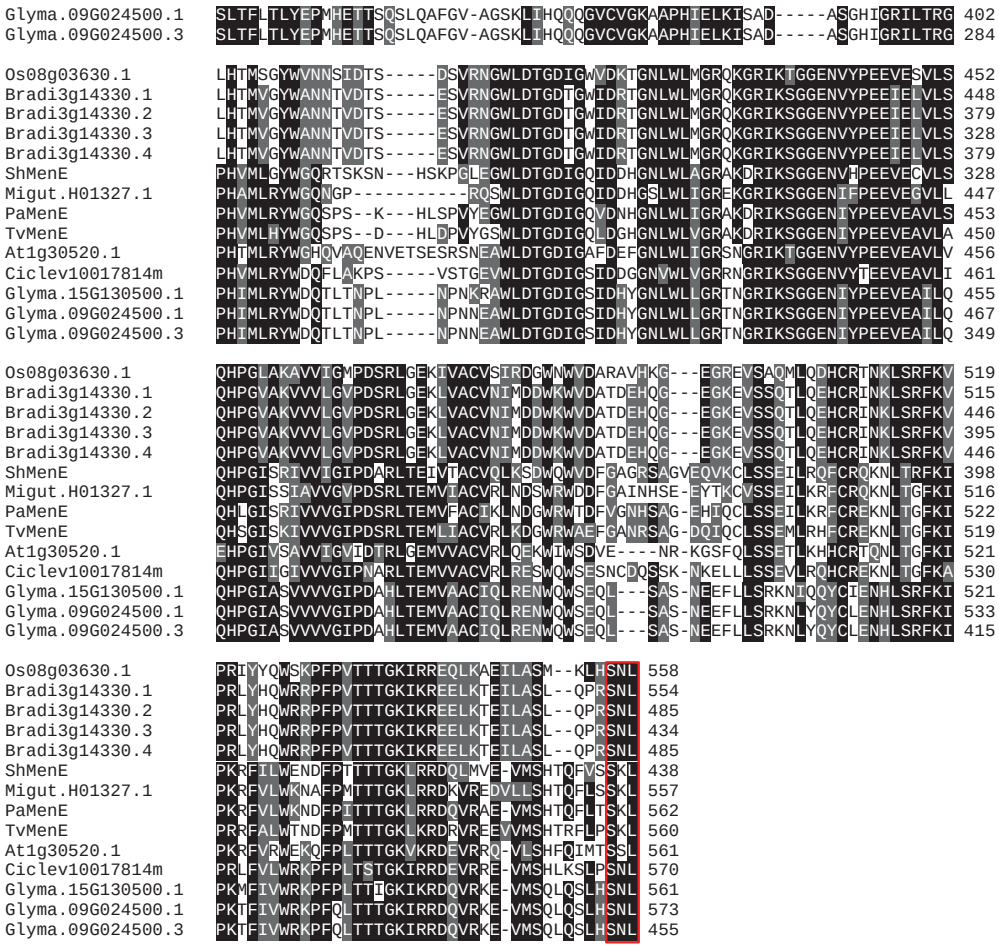


Figure S3.1. MenE sequence alignment with C-terminal peroxisome targeting signal PTS1. MenE sequences from the three parasitic species and representative non-parasitic plants from Phytozome v11 were aligned by Clustal Omega 1.2.1 (http://www.ebi.ac.uk/Tools/msa/clustalo/) and visualized using Color Align Conservation (http://www.bioinformatics.org/sms2/color_align_cons.html). The C-terminal PTS1 is boxed in red.

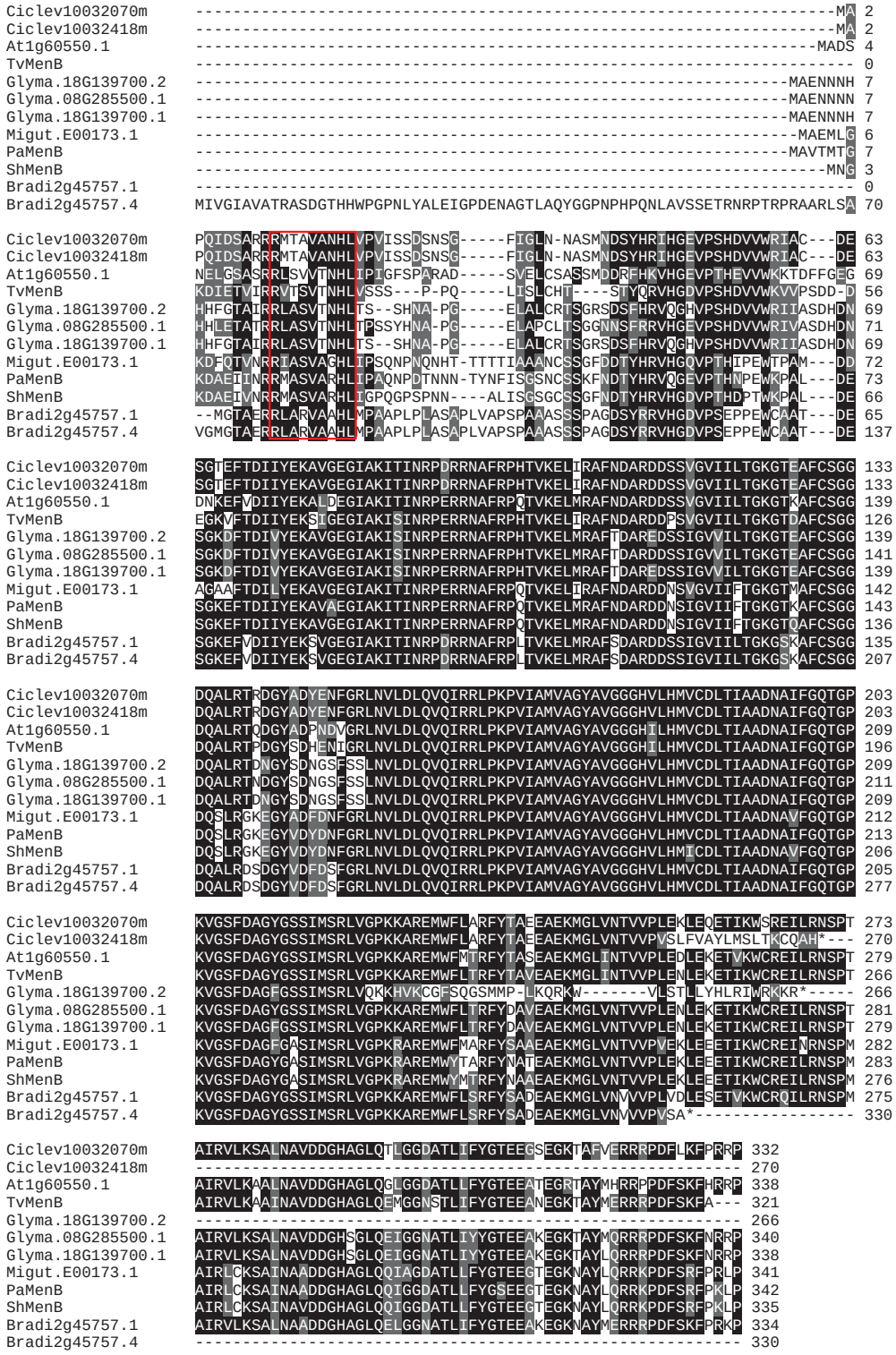


Figure S3.2. MenB sequence alignment with N-terminal peroxisome targeting signal PTS2. The N-terminal PTS2 is red-boxed. Sequence analysis was performed as in figure S3.1.

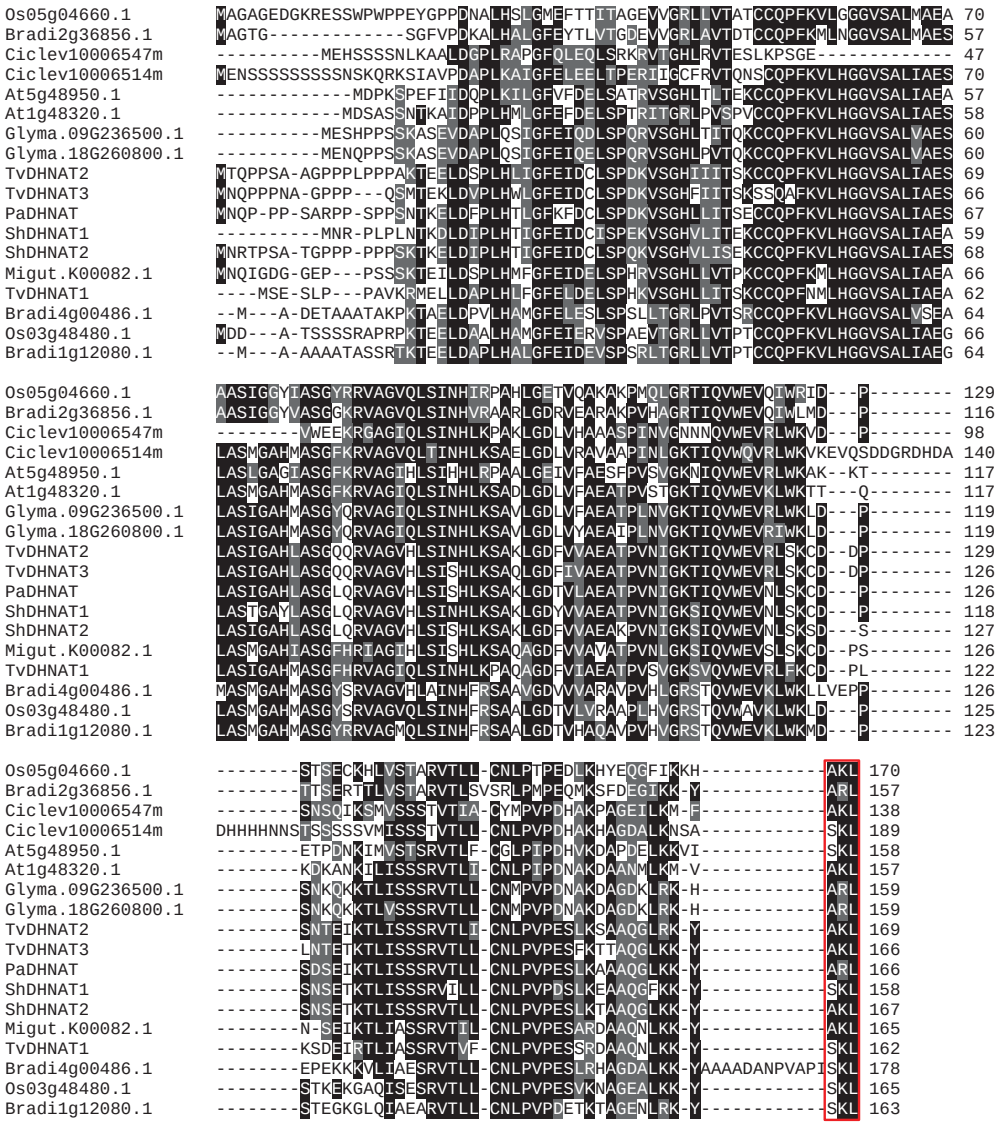


Figure S3.3. DHNAT sequence alignment with C-terminal peroxisome targeting signal PTS1. The C-terminal PTS1 is boxed in red. Sequence analysis was the same as in figure S3.1.

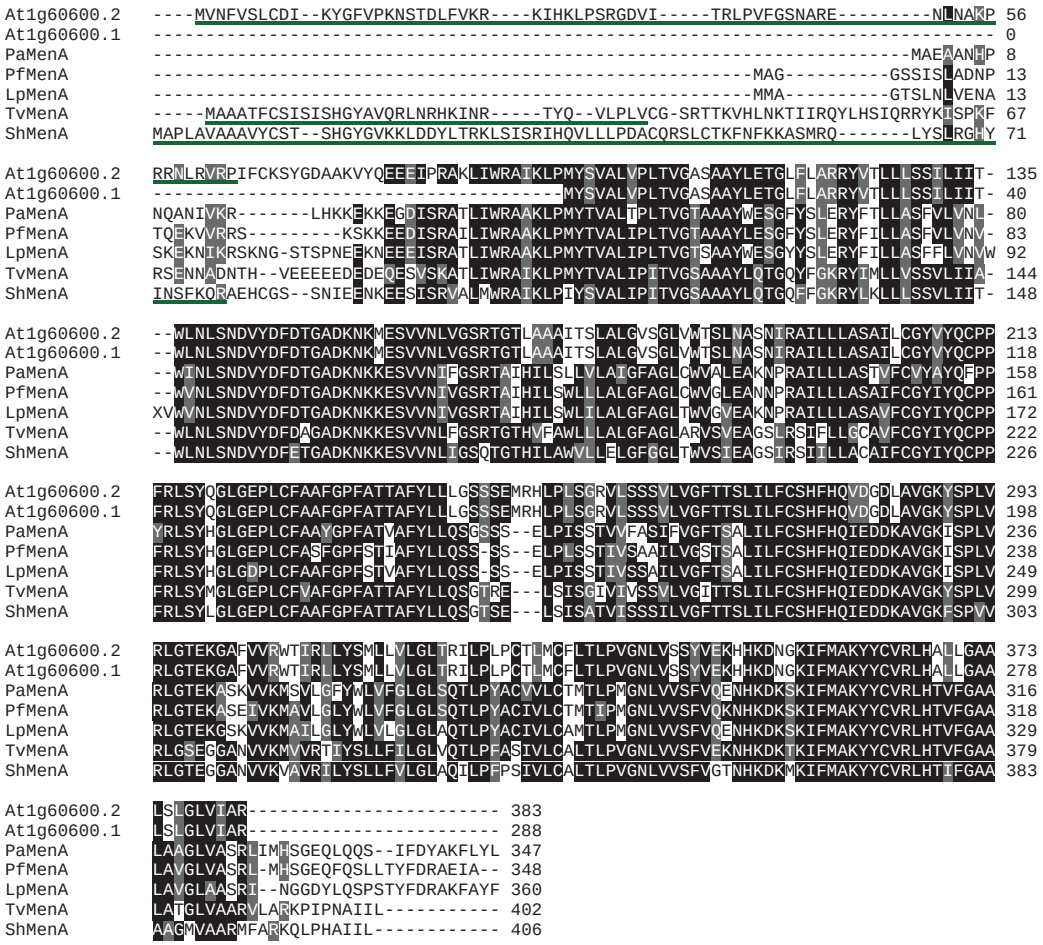


Figure S3.4. Alignment of parasitic MenA sequences with *Arabidopsis* isoforms. Additional MenA sequences were identified from holoparasites *Phelipanche fasciculata* and *Lindenbergia philippensis* available from the 1000 Plants (1KP) database (https://db.cngb.org/blast4onekp/) (Matasci et al., 2014). Sequence analysis was the same as in figure S3.1. The predicted plastid transient peptides in AtMenA, ShMenA and TvMenA are underlined in green. The IKP identifiers for PfMenA and LpMenA are: scaffold-VYDM-2034145-VYDM-Orobanche_fasciculata-2_samples_combined and scaffold-EJCM-2018390-EJCM-Lindenbergia_philippensis-2_samples_combined, respectively

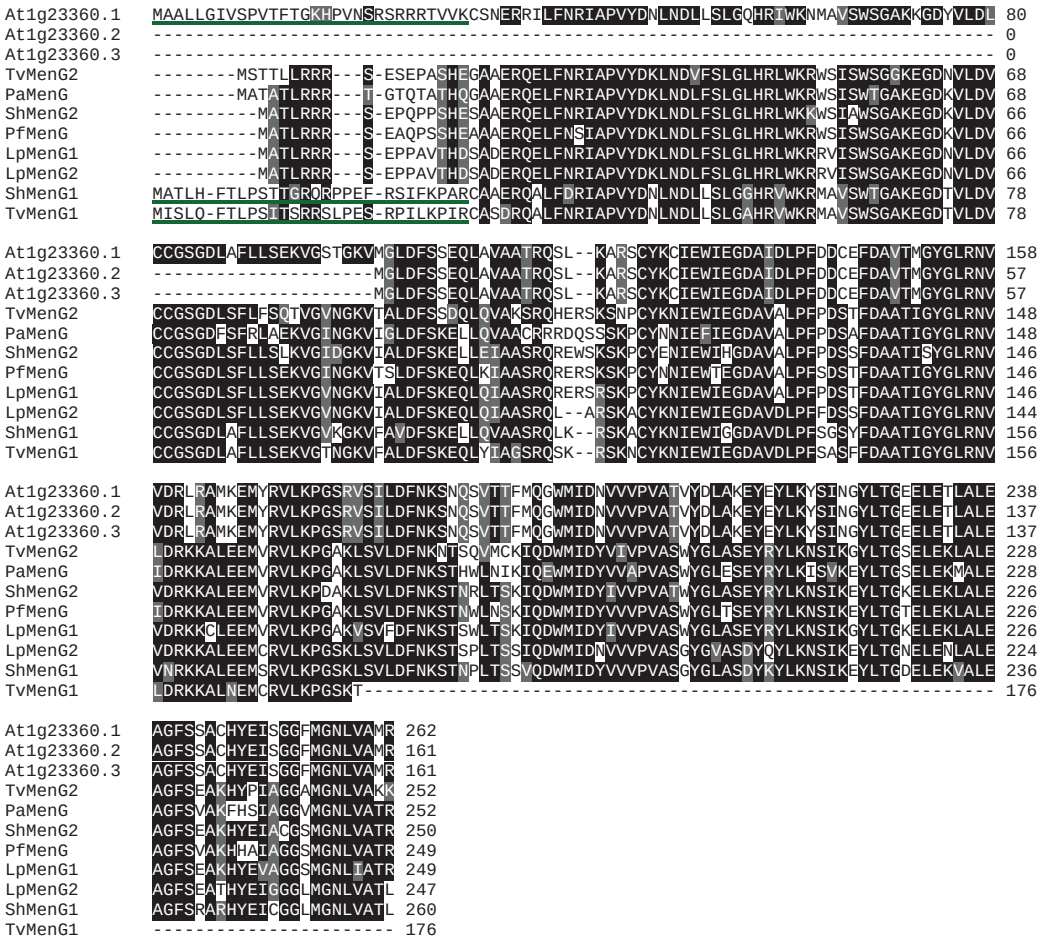


Figure S3.5. Alignment of parasitic MenG sequences with *Arabidopsis* isoforms. Additional MenG sequences of holoparasites *Phelipanche fasciculata* and *Lindenbergia philippensis* were identified from the 1KP database as in figure S3.4. The predicted plastid transient peptides in AtMenG.1, ShMenG1 and TvMenG1 are underlined in green. The IKP identifiers for PfMenG, LpMenG1 and LpMenG2 are: scaffold-VYDM-2129491-VYDM-Orobanche_fasciculata-2_samples_combined, scaffold-EJCM-2011303-EJCM-Lindenbergia_philippensis-2_samples_combined, and scaffold-EJCM-2011302-EJCM-Lindenbergia_philippensis-2_samples_combined, respectively.

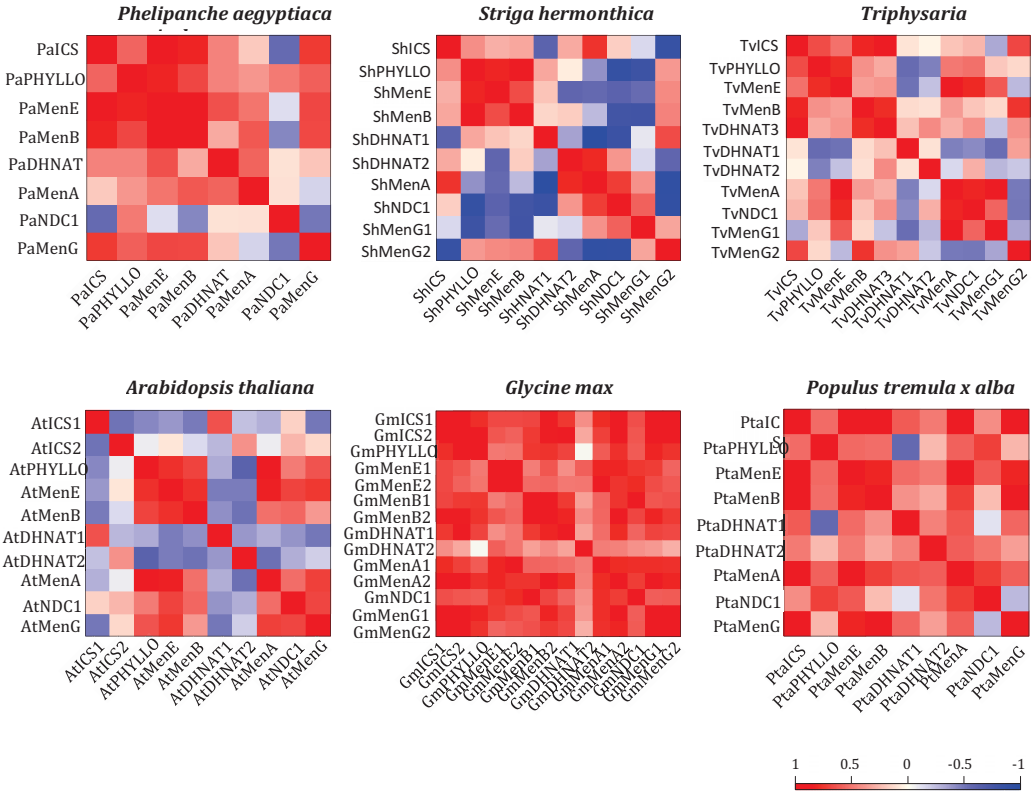


Figure S3.6. Coexpression of PhQ genes in parasitic and non-parasitic plants. Coexpression patterns of PhQ genes based on Gini correlation coefficient (GCC). The top three panels are the same as in Figure 3.5, except with the addition of the multifunctional *NDC1*. The bottom three panels show strong coexpression among PhQ genes in photoautototrophic species. The exceptions are *A. thaliana AtICS1* that is involved in salicylic acid biosynthesis for defense (Wildermuth et al., 2001), and *AtDHNATs* that are involved in peroxisomal β-oxidation (Cassin-Ross and Hu, 2014), besides PhQ biosynthesis. NDC1 exhibited strong coexpression with PhQ genes in photosynthetic species than in the holoparasite. Data used to compute GCC were downloaded from the Sequence Read Archive (SRA). *Arabidopsis thalinana* data sets include SRA236885, SRA091517, SRA269936, SRA219425, SRA308579, SRA050132, SRA067724, SRA291734, SRA269101, SRA098075, SRA100242, SRA122395, SRA163488, SRA064368, SRA246225, SRA248861, SRA202878, SRA201550, SRA303151, SRA221137, SRA272654 and SRA221060. Biotic and abiotic stress treated samples were excluded from

the data sets. *Glycine max* data sets include SRA187830, SRA047293, SRA036577, SRA116533, SRA091756, SRA187830, SRA036538, SRA036577 and SRA129337. *Populus tremula x alba* data sets include SRA274261 and SRA097208.

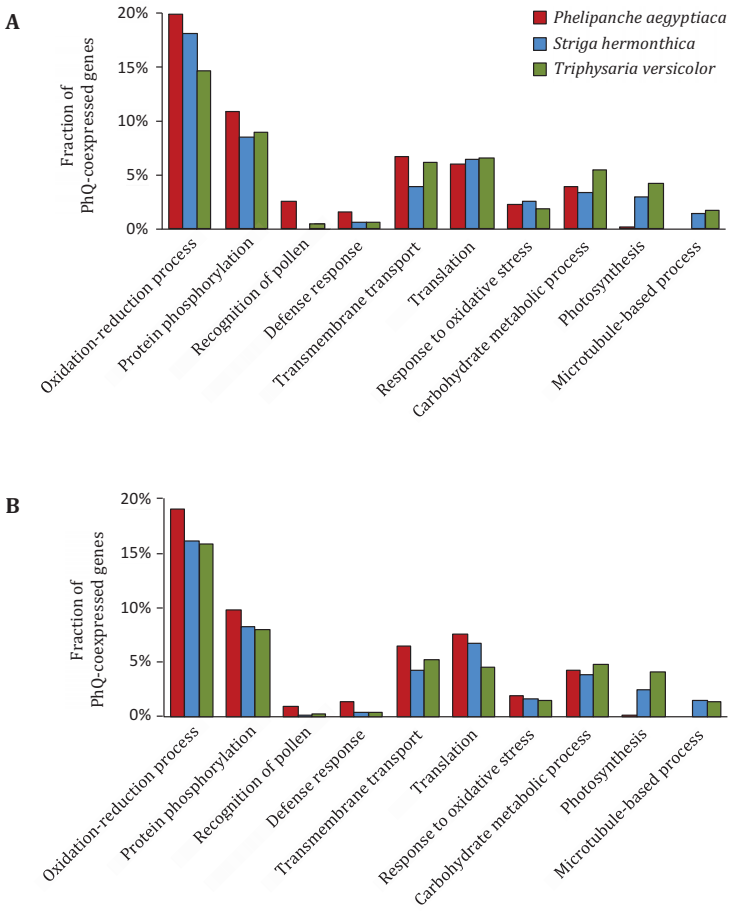


Figure S3.7. GO enrichment of PhQ-coexpressed genes. The top ten GO terms of PhQ-coexpressed genes, defined as the union set of the top 500 most highly-correlated transcripts for each PhQ gene (A), or as the union set of transcripts with a Gini correlation coefficient ≥0.8 for each PhQ gene (B). GO terms were considered differentially enriched if the fraction differed by at least 1% between the holoparasite *Phelipanche aegyptiaca* and the photosynthetically competent *Striga hermonthica* and *Triphysaria versicolor* in both analyses. Five GO terms that satisfied this criterion (oxidation-reduction process, protein phosphorylation, defense response, photosynthesis and microtubule-based process) from A are shown in Figure 3.6.

References

Alexa A, Rahnenfuhrer J (2010) topGO: enrichment analysis for gene ontology. R Packag. version 2:

Allen JM, Huang DI, Cronk QC, Johnson KP (2015) aTRAM - automated target restricted assembly method: a fast method for assembling loci across divergent taxa from next-generation sequencing data. BMC Bioinformatics **16**: 98

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

Babujee L, Wurtz V, Ma C, Lueder F, Soni P, Van Dorsselaer A, Reumann S (2010) The proteome map of spinach leaf peroxisomes indicates partial compartmentalization of phylloquinone (vitamin K_1) biosynthesis in plant peroxisomes. J Exp Bot **61**: 1441–1453

Basset GJ (2016) Phylloquinone (vitamin K1): occurrence, biosynthesis and functions. Mini-Reviews Med Chem **16**: 1–11

Boden M, Hawkins J (2005) Prediction of subcellular localization using sequence-biased recurrent networks. Bioinformatics **21**: 2279–2286

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120

Brettel K, Setif P, Mathis P (1986) Flash-induced absorption changes in photosystem I at low temperature: Evidence that the electron acceptor A_1 is vitamin K_1. FEBS Lett **203**: 220–224

Bridge A, Barr R, Morré DJ (2000) The plasma membrane NADH oxidase of soybean has vitamin K_1 hydroquinone oxidase activity. Biochim Biophys Acta - Biomembr **1463**: 448–458

Cassin-Ross G, Hu J (2014) Systematic phenotypic screen of Arabidopsis peroxisomal mutants identifies proteins involved in β-oxidation. Plant Physiol **166**: 1546–59

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32**: 1792–1797

Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell

using TargetP, SignalP and related tools. Nat Protoc **2**: 953–971

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

Fatihi A, Latimer S, Schmollinger S, Block A, Dussault PH, Vermaas WFJ, Merchant SS, Basset GJ (2015) A dedicated type II NADPH dehydrogenase performs the penultimate step in the biosynthesis of vitamin K1 in *Synechocystis* and Arabidopsis. Plant Cell **27**: 1730–41

Frigaard N-U, Takaichi S, Hirota M, Shimada K, Matsuura K (1997) Quinones in chlorosomes of green sulfur bacteria and their role in the redox-dependent fluorescence studied in chlorosome-like bacteriochlorophyll c aggregates. Arch Microbiol **167**: 343–349

Frydman B, Rapoport H (1963) Non-Chlorophyllous Pigments of *Chlorobium Thiosulfatophilum* Chlorobiumquinone. J Am Chem Soc **85**: 823–825

Furt F, Oostende C van, Widhalm JR, Dale MA, Wertz J, Basset GJC (2010) A bimodular oxidoreductase mediates the specific reduction of phylloquinone (vitamin K1) in chloroplasts. Plant J **64**: 38–46

Garcion C, Lohmann A, Lamodière E, Catinot J, Buchala A, Doermann P, Métraux J-P (2008) Characterization and biological function of the *ISOCHORISMATE SYNTHASE2* gene of Arabidopsis. Plant Physiol **147**: 1279–1287

González-Verdejo CI, Barandiaran X, Moreno MT, Cubero JI, Di Pietro A (2006) A peroxidase gene expressed during early developmental stages of the parasitic plant *Orobanche ramosa*. J Exp Bot **57**: 185–92

Gross J, Cho WK, Lezhneva L, Falk J, Krupinska K, Shinozaki K, Seki M, Herrmann RG, Meurer J (2006) A plant locus essential for phylloquinone (vitamin K_1) biosynthesis originated from a fusion of four eubacterial genes. J Biol Chem **281**: 17189–96

Hale MB, Blankenship RE, Fuller RC (1983) Menaquinone is the sole quinone in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. BBA -

Bioenerg **723**: 376–382

Heyno E, Alkan N, Fluhr R (2013) A dual role for plant quinone reductases in host-fungus interaction. Physiol Plant **149**: n/a-n/a

Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res **35**: W585-7

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinforma Appl NOTE **17**: 754–755

Ishida JK, Wakatake T, Yoshida S, Takebayashi Y, Kasahara H, Wafula E, DePamphilis CW, Namba S, Shirasu K (2016) Local auxin biosynthesis mediated by a YUCCA flavin monooxygenase regulates haustorium development in the parasitic plant *Phtheirospermum japonicum*. Plant Cell **28**: 1795–814

Kaiping S, Soll J, Schultz G (1984) Site of methylation of 2-phytyl-1,4-naphthoquinol in phylloquinone (vitamin K_1) synthesis in spinach chloroplasts. Phytochemistry **23**: 89–91

Kim D, Kocz R, Boone L, Keyes WJ, Lynn DG (1998) On becoming a parasite: evaluating the role of wall oxidases in parasitic plant development. Chem Biol **5**: 103–117

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**: R36

Kim HU, van Oostende C, Basset GJC, Browse J (2008) The *AAE14* gene encodes the Arabidopsis o-succinylbenzoyl-CoA ligase that is essential for phylloquinone synthesis and photosystem-I function. Plant J **54**: 272–83

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

Li W, Schulman S, Dutton RJ, Boyd D, Beckwith J, Rapoport TA (2010) Structure of a bacterial homologue of vitamin K epoxide reductase. Nature **463**: 507–512

Liang L, Liu Y, Jariwala J, Lynn DG, Palmer AG (2016) Detection and adaptation in parasitic angiosperm host selection. Am J Plant Sci **7**: 1275–1290

Lochner K, Döring O, Böttger M (2003) Phylloquinone, what can we learn from plants? BioFactors **18**: 73–78

Lohmann A, Schottler MA, Brehelin C, Kessler F, Bock R, Cahoon EB, Dormann P (2006) Deficiency in phylloquinone (vitamin K1) methylation affects prenyl quinone distribution, photosystem I abundance, and anthocyanin accumulation in the Arabidopsis *AtmenG* mutant. J Biol Chem **281**: 40461–40472

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**: 550

Lüthje S, Böttger M (1995) On the function of a K-type vitamin in plasma membranes of maize (*Zea mays* L.) roots. Mitt Inst Allg Bot Hambg **25**: 5–13

Lüthje S, Gestelen P, Córdoba-Pedregosa MC, González-Reyes J a., Asard H, Villalba JM, Böttger M (1998) Quinones in plant plasma membranes — a missing link? Protoplasma **205**: 43–51

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. [Miyashita mitsunori]

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

Van Oostende C, Widhalm JR, Furt F, Ducluzeau A-L, Basset GJ (2011) Vitamin K_1 (Phylloquinone): function, enzymes and genes. Biosynth Vitam Plants Part B Vitam B6, B8, B9, C, E, K **59**: 229

Reumann S, Babujee L, Ma C, Wienkoop S, Siemsen T, Antonicelli GE, Rasche N, Lüder F, Weckwerth W, Jahn O (2007) Proteome analysis of Arabidopsis leaf peroxisomes reveals novel targeting peptides, metabolic pathways, and defense mechanisms. Plant Cell **19**: 3170–3193

Roberts A, Pachter L (2013) Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods **10**: 71–3

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol **61**: 539–542

Schultz G, Soll J, Ellerbrock BH (1981) Site of prenylation reaction in synthesis of phylloquinone (vitamin K_1) by spinach chloroplasts. Eur J Biochem **117**: 329–332

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res **13**: 2498–504

Shimada H, Ohno R, Shibata M, Ikegami I, Onai K, Ohto M, Takamiya K (2005) Inactivation and deficiency of core proteins of photosystems I and II caused by genetical phylloquinone and plastoquinone deficiency but retained lamellar structure in a T-DNA mutant of Arabidopsis. Plant J **41**: 627–637

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7**: 539

Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: A tool for rapidly screening proteomes forN-terminal targeting sequences. Proteomics **4**: 1581–1590

Veronesi C, Bonnin E, Calvez S, Thalouarn P, Simier P (2007) Activity of secreted cell wall-modifying enzymes and expression of peroxidase-encoding gene following germination of *Orobanche ramosa*. Biol Plant **51**: 391–394

Wang L, Li Q, Zhang A, Zhou W, Jiang R, Yang Z, Yang H, Qin X, Ding S, Lu Q, et al (2017) The phytol phosphorylation pathway is essential for the biosynthesis of phylloquinone, which is required for photosystem I stability in Arabidopsis. Mol Plant **10**: 183–196

Wickett NJ, Honaas LA, Wafula EK, Das M, Huang K, Wu B, Landherr L, Timko MP, Yoder J, Westwood JH, et al (2011) Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. Curr Biol **21**: 2098–104

Widhalm JR, Ducluzeau AL, Buller NE, Elowsky CG, Olsen LJ, Basset GJC (2012) Phylloquinone (vitamin K_1) biosynthesis in plants: Two peroxisomal thioesterases of lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-coa. Plant J **71**: 205–215

Wildermuth MC, Dewdney J, Wu G, Ausubel FM (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature **414**: 562–565

Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, Huang K, Bandaranayake PCG, Wu B, Der JP, et al (2015) Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol **32**: 767–90

CHAPTER 4

EXPLORING NON-PHOTOSYNTHETIC FUNCTION OF PHQ BIOSYNTHESIS IN *ARABIDOPSIS*, *POPULUS* AND *GLYCINE*: A COMPARATIVE APPROACH[1]

[1] Gu X. and C.J. Tsai, to be submitted to *PLoS ONE*

Abstract

PhQ that is biosynthesized in chloroplasts is essential for photosynthetic electron transport. PhQ is also essential for certain transmembrane electron transport activities in non-photosynthetic parasites, and there is evidence that it may have similar functions in the plasma membrane of photoautotrophic plant cells. What remains uncharacterized is the identity of the cellular function(s) that depend on plasma-membrane PhQ in plants. Here we leveraged an RNA-Seq Atlas of *Arabidopsis*, *Populus* and *Glycine* to explore the expression patterns of the PhQ biosynthetic genes in both photosynthetic and heterotrophic tissues. This approach was expected to yield co-expression data that would be informative to discern PhQ functions between photosynthetic sink and source tissues. Strong to moderate transcript abundance was observed for some PhQ biosynthetic genes in sink tissues, but further functional analyses of their non-photosynthetic function remain inconclusive. Multiple episodes of PhQ pathway gene duplication and expression divergence were observed. Whole pathway duplication and retention was only observed in soybean. *ICS* divergence was specific to *Arabidopsis*, and the expression profile differences between *DHNAT* duplicates differed across species. This species-dependent divergence provided evidence for substantial plasticity in the PhQ biosynthetic pathway that cross-talks with various plastidial, peroxisomal and plasma membrane-associated processes. Further investigation will be required to fully resolve the non-photosynthetic function of PhQ and the associated evolutionary mechanisms.

Introduction

Vitamin K comprises a group of membrane-bound, lipid-soluble naphthoquinone derivatives essential to plants, animals and bacteria. The most abundant form of vitamin K in nature, vitamin K1 (phylloquinone or PhQ), is synthesized in plants as an electron transfer cofactor in photosystem I (PSI) (Brettel et al., 1986). *Arabidopsis* mutants deficient in PhQ biosynthesis are generally seedling-lethal (Shimada et al., 2005; Gross et al., 2006; Garcion et al., 2008; Kim et al., 2008). Bacteria synthesize vitamin K2 (menaquinone) that functions in respiratory, and in some cases, photosynthetic electron transport chains (Hale et al., 1983). Menaquinones also protect cells from oxidative stresses (Frigaard et al., 1997) and participate in signaling processes (Georgellis et al., 2001). Animals depend on dietary intake and intestinal bacteria for conversion and/or synthesis of vitamin K, which plays important roles in blood coagulation (Hirsh et al., 2001), vascular calcification (Price et al., 1998), and bone metabolism (Price and Williamson, 1981).

Despite the large body of evidence for functional multiplicity of vitamin K in both animals and prokaryotes, PhQ function has largely been regarded in plants as being tied to photosynthesis. PhQ is predominantly found in chloroplasts (Lohmann et al., 2006) where it binds to the A1 site of photosystem I and transfers an electron from chlorophyll a to an iron-sulfur cluster (Brettel et al., 1986; Petersen et al., 1987). However, small pools of PhQ have also been detected in the plasma membrane of non-photosynthetic organs like maize (*Zea mays*) roots (Lüthje and Böttger, 1995). PhQ participation in plasma membrane electron transport activity has also been demonstrated in heterotrophic carrot cell cultures by the use of UV treatments and PhQ feeding (Barr et al., 1992). A putative plasma membrane redox system involving PhQ was proposed in which electrons are transferred from cytosolic donors (e.g. NADPH) to apoplastic acceptors in plants (Lochner et al., 2003). Interestingly, naphthoquinone-dependent NADH dehydrogenase activities have been characterized in plasma membranes of onion roots (Serrano et al., 1994), maize roots (Lüthje et al., 1998) and soybean hypocotyls (Schopfer et al., 2008). A

naphthoquinone-dependent NADH oxidase (NOX) was also isolated from the plasma membrane of soybean hypocotyls (Bridge et al., 2000). Activity of such oxidoreductases in PhQ-containing plasma membranes would be consistent with PhQ function in non-photosynthetic electron transport.

Our previous work has shed light on an alternative biosynthetic route and function of PhQ in a photosynthesis-free system (Chapter 3). We showed that through evolutionary changes in the subcellular localization of the last two enzymes of the PhQ biosynthetic pathway, the biosynthesis of PhQ has been redirected from the plastid to the plasma membrane. In addition, co-expression network analysis revealed that PhQ genes were strongly co-expressed with genes encoding peroxidases, NAD(P)H-oxidoreductases and NAD(P)H oxidases. These proteins are involved in the development of haustorium, a specialized structure for nutrient absorbance and host invasion (Ishida et al., 2016; Liang et al., 2016), or known to participate in plasma membrane electron transport. The findings support a role for PhQ in transmembrane redox activities associated with parasitism.

Interestingly, the biosynthesis of PhQ in photoautotrophic plants occurs both in plastids and plasma membranes, via alternative splicing of the last two steps (Chapter 3). How the PhQ biosynthesis pathway evolved its dual subcellular localization and functioning in photoautotrophic plants remains elusive. The knowledge gained from the parasitic study described above may provide direction for elucidating the non-photosynthetic function of PhQ in photoautotrophic species. One common mechanism for genes to evolve novel functions in plants is through gene duplication followed by neofunctionalization or subfunctionalization, as first proposed by Ohno (Susumu, 1970) and extended by many others (Force et al., 1999; Lynch and Conery, 2000; Tirosh et al., 2007; Liberles et al., 2011). An additional way to gain functional diversity is through alternative splicing that produces functional distinct isoforms (Palusa et al., 2007; Zhang and Mount, 2009).

Genes of the PhQ biosynthesis pathway have a complex evolutionary history with multiple rounds of gene duplication and retention. *ICS*, which encodes isochorismate synthase for conversion of chorismate to isochorismate, is duplicated in Brassicales-Malvales (Macaulay et al., 2017). *AtICS1*, one of the duplicate in *Arabidopsis thaliana*, is important for the biosynthesis of both PhQ and salicylic acid (SA) (Wildermuth et al., 2001; Garcion et al., 2008). *AtICS2* is also involved in PhQ and SA biosynthesis but the role is minor (Garcion et al., 2008). Two *ICS* protein isoforms were characterized from the cell cultures of *Catharanthus roseus* and *Rubia tinctorum* (van Tegelen et al., 1999; Van Tegelen et al., 1999). Interestingly, *ICS* is also present in two copies in the genome of *Escherichia coli* and *Bacillus subtilis*, with one (*MenF*) involved in menaquinone biosynthesis, a counterpart of PhQ in the bacterial respiratory chain, and the other (*entC*) in synthesis of salicylic acid-derived siderophores which are involved in iron chelation (Daruwala et al., 1996; Müller et al., 1996; Rowland and Taber, 1996; Daruwala et al., 1997; Dahm et al., 1998). Another PhQ pathway gene with multiple duplication events is *DHNAT*, encoding a DHNA-CoA thioesterase that catalyzes the hydrolysis of DHNA-CoA in peroxisomes (Widhalm et al., 2012). *DHNAT* is present in duplicate in *Arabidopsis thaliana*, *Glycine max*, and *Populus tremula x alba*. Interestingly, *Glycine max* has retained all PhQ pathway genes as duplicates except for *PHYLLO* and *NDC1*. Except for the *Arabidopsis ICS* paralogs, functional conservation or divergence has not been explored for any of the PhQ gene duplicates in plants.

To understand the non-photosynthetic function of PhQ in photoautotrophic species and the associated evolutionary mechanism, heterotrophic tissues, like roots and xylem with little photosynthesis, were targeted for the investigation for comparison with photosynthetic tissues. We mined public available transcriptome datasets of *Arabidopsis thaliana*, *Populus tremula x alba* (poplar), and *Glycine max* (soybean) that comprise both photoautotrophic and heterotrophic tissues. Gene co-expression and functional enrichment analyses were performed on the two tissue types to discern the non-photosynthetic function of PhQ.

Materials and Methods

RNA-Seq Data Collection and Processing

RNA-Seq data were downloaded from the Sequence Read Archive (SRA). *A. thalinana* data sets included SRA236885, SRA091517, SRA269936, SRA219425, SRA308579, SRA050132, SRA067724, SRA291734, SRA269101, SRA098075, SRA100242, SRA122395, SRA163488, SRA064368, SRA246225, SRA248861, SRA202878, SRA201550, SRA303151, SRA221137, SRA272654 and SRA221060. *G. max* data sets included SRA187830, SRA047293, SRA036577, SRA116533, SRA091756, SRA187830, SRA036538, SRA036577 and SRA129337. *P. tremula x alba* data sets included SRA274261 and SRA097208. Raw reads were pre-processed by Cutadapt 1.9.dev1 (Martin, 2011), Trimmomatic 0.32 (Bolger et al., 2014) and custom scripts to remove adapter, non-coding RNA, organellar sequences, and low-quality reads. After quality control, reads were aligned to the corresponding reference genome with Tophat 2.0.13 (Kim et al., 2013) followed by read count with HTseq 0.6.1p1 (Anders et al., 2015) and expression estimation with DEseq2 (Love et al., 2014). Genome used for alignment were downloaded from Phytozome v11 for *Arabidopsis thaliana* (TAIR10) and *Glycine max* (Wm82.a2.v1). A variant-substitute genome was used for *Populus tremula x alba* (Xue et al., 2015). Expression values were normalized by Z-score transformation and visualized in heatmaps using *pheatmap* package in R.

Co-expression Network Construction

Biological replicates from the same experiment were averaged for each gene. Genes with poor expression (FPKM < 2) in at least 80% samples or with little expression variance across samples (Coefficient Variance < 0.4) were removed. Hierarchical clustering of samples was performed using the Euclidean distance matrix based on PSI/PSII gene expression. Gini correlation coefficient (GCC) was computed for source tissues and sink tissues separately. The resulting GCC matrix was used to extract the top500-source set and top500 sink set. For each

PhQ gene, the GCCs with other genes were ranked and the top 500 genes were selected for further analysis.

GO Enrichment Analysis

Gene annotations for *Arabidopsis thaliana* (TAIR10), *Populus trichocarpa* (v3.0), and *Glycine max* (Wm82.a2.v1) were downloaded from Phytozome v11. Annotation for *Arabidopsis* with better quality was downloaded from TAIR10 and combine with Phytozome annotation. To improve the annotation quality of poplar and soybean, orthology was constructed between the three species using OrthoFinder with default settings (Emms and Kelly, 2015). *Arabidopsis* annotation was assigned to and combined with poplar and soybean annotation based on the orthology. GO enrichment was performed on the top500-source set and top500-sink set using *topGO* R package.

AS Analysis

Transcript isoforms of *AtMenA* and *AtMenG* were obtained from Phytozome v11. Pre-processed reads were aligned to the isoforms using Bowtie2 (Langmead and Salzberg, 2012). Reads spanning isoform-specific junctions with a minimal length of four nucleotides (two on each side of the junction) were extracted and counted with a custom Perl script.

Results

Expression profiles of PhQ genes

RNA-Seq data of *Arabidopsis thaliana* (166 samples), *Populus tremula x alba* (68 samples) and *Glycine max* (64 samples), were analyzed. The samples span a wide range of tissue types at different developmental stages and under various environmental conditions. To separate heterotrophic tissues with weak to no photosynthesis from photoautotrophic tissues, all samples were clustered based on expression of photosystem I (PSI) and photosystem II (PSII) genes. As shown in Figure S4.1-4.3, two major clades were found in all three species, with one clade predominantly enriched in chlorophyllous tissues like leaves, shoots and seedlings, and the

other clade composed primarily of roots, xylem, phloem, flowers and seeds. However, some chlorophyllous tissues were grouped into the second clade, for example, green cotyledons, seed pods, young seedlings and leaves. Although photosynthetically active, those tissues are not self-sufficient and need to import photoassimilates from other source tissues. Therefore, we named the two clades as photosynthetic-source (source) clade and photosynthetic-sink (sink) clade.

Given the important role of PhQ in photosynthesis, it is expected that the biosynthesis of PhQ is active in source tissues. Conversely, exploring PhQ gene expression patterns in sink tissues where photosynthesis is weak or absent is expected to shed light on the non-photosynthetic function of PhQ. Expression of *PsaO* and *PsaD* which encode two subunits of the PSI protein complexes was used to gauge photosynthetic activity in each sample. As expected, *PsaO* and *PsaD* showed strong expression in the source tissues and very weak expression in sink tissues across all three species (Figure 4.1A-C). Consistent with its role in photosynthesis, PhQ genes also showed high expression in source tissues. Interestingly, unlike the weak expression of *PsaO* and *PsaD*, PhQ genes were moderately expressed in sink tissues, including roots, in all three species (Figure 4.1A-C). This indicated that the PhQ biosynthetic pathway is active in heterotrophic tissues, potentially with a non-photosynthetic function.

Gene enrichment analysis of PhQ-coexpressed genes in source and sink tissues

With a few exceptions (*ICS* and *DHNAT,* see below), PhQ genes exhibited highly similar expression profiles with each other, suggesting they are tightly co-regulated. Gini-correlation coefficient (GCC) was computed for QC-filtered genes among source and sink tissues separately. The top 500 most highly correlated genes were extracted for each PhQ gene for Gene Ontology Biological Process (GOBP) functional enrichment analysis. The gene sets from source and sink tissues were named as “top500-source set” and “top500-sink set”, respectively.

For *AtPHYLLO, AtMenE, AtMenB, AtMenA, AtNDC1*, and *AtMenG*, the enrichment patterns of the top500-source set were largely distinct from those of the top500-sink set (Figure 4.4A green and blue labels for source and sink sets, respectively). Both conditions were

significantly enriched in photosynthesis-related processes (Figure 4.4A). Similar patterns were observed for poplar and soybean. However, the top500-sink set was more significantly enriched in photosynthesis than the top500-sink set in both polar and soybean (Figure 4.4B-C, green and blue labels). In *Arabidopsis*, compared to the top500-source set, the top500-sink set was more enriched in shoot system morphogenesis, defense response to bacterium, embryo development ending in seed dormancy, cell differentiation, response to cold, and detection of biotic stimulus (Figure 4.4A). However, such patterns were not observed in poplar and soybean data (Figure 4.4B-C).

Taken together, the functional enrichment results did not reveal clear differences between sink and source tissues that were informative for understanding the non-photosynthetic role of PhQ in heterotrophic tissues. This limitation was likely because some photosynthetically competent tissues were included in the sink dataset as explained above.

High resolution microarray data from Arabidopsis roots

To get a less ambiguous view about PhQ gene expression in true heterotrophic tissues, we investigated the expression of PhQ genes in roots with cellular resolution using published *Arabidopsis* microarray data (Brady et al., 2007; Cartwright et al., 2009) available on the eFP browser (Figure 4.2, Table S4.1, http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). We observed localized expression in several cases. *AtICS1* was detected in cortex, phloem companion cells, and phloem pole pericycle, and *AtICS2* in cortex. *AtNDC1* and *AtMenG* was found in phloem companion cells, phloem pole pericycle and xylem pole pericycle. *AtPHYLLO*, *AtMenE*, *AtDHNAT1/2*, and *AtMenA* exhibited high expression in procambium. *AtDHNAT1/2* also showed high expression levels in several other root cell types. The photosynthesis marker genes *AtPsaD2* and *AtPsaO* were not expressed in roots, but *AtPsaD1* showed procambium expression. Thus, while clear evidence of heterotrophic tissue expression was obtained for PhQ genes in roots, it remains difficult to exclude the possibility that their expression in roots was associated with photosynthesis.

Abundance estimation of alternatively spliced isoforms

As characterized previously in Chapter 3, the plasma membrane PhQ biosynthesis also occurs in photoautotrophic plants via alternative splicing of the last two genes (*AtMenA* and *AtMenG*) in the PhQ biosynthetic pathway. The primary isoform encodes a longer protein with plastid-targeting signal peptide, whereas the secondary isoform encodes a shorter protein missing the transit peptide. The abundance of each isoform in various tissues under multiple conditions can reveal the relative importance of the two biosynthesis routes, and hence provide clues for PhQ functions in the plasma membrane.

The two alternatively spliced isoforms of *AtMenA* and *AtMenG* differ in exon-intron junctions near their 5' end (Figure 3.3). Thus, relative abundance of the isoforms can be estimated by counting reads mapped to the isoform-specific junctions. As shown in Table S4.2, the plastid isoforms of both *AtMenA* and *AtMenG* are more abundant than the plasma membrane isoforms in nearly all samples examined. In samples where the plastid and plasma membrane isoforms exhibited comparable read counts, the numbers were too low to be reliable due to the poor expression of both isoforms. While these results were not informative due to the limited sensitivity of the data, there were clear expression evidence for the plasma membrane isoform of both *AtMenA* and *AtMenG* in several tissues (read count >50) in support of a non-canonical role of PhQ in photoautotrophic species. Further experiments silencing the secondary isoforms that target the plasma membrane may help us understand the non-photosynthetic function of PhQ.

Expression patterns of duplicated PhQ biosynthetic genes in soybean

All three species analyzed in this study have experienced multiple rounds of whole genome duplication (WGD), however, only soybean has retained duplicates of essentially the entire PhQ pathway. As the most common fate for duplicated PhQ biosynthetic genes is gene loss, this pathway-level retention is of particular interest. One possible explanation is that the duplicates are not yet lost due to the relatively young age of the recent WGD in soybean (Schmutz et al., 2010). Alternatively, this pathway might have been under selection to retain the duplicates.

Although the exact mechanism is unclear, it should be noted that soybean seeds are known to accumulate high levels of PhQ (Booth and Suttie, 1998).

Using available RNA-Seq data from soybean, we examined whether the PhQ gene paralogs have started to diverge or remained redundant in expression. Figure 4.1 showed that PhQ biosynthetic gene paralogs shared very similar expression profiles overall. However, as the color scheme of the heatmap was scaled to reflect relative transcript abundance across tissues, high expression of PhQ biosynthetic genes in source tissues might have masked subtle differences between relatively less expressed paralogs in sink tissues. For this reason, a separate analysis was performed for sink tissues only. Although the expression profiles of PhQ gene paralogs remained similar, some differences were observed (Figure 4.3). For example, *GmMenE1* showed higher expression than *GmMenE2* in multiple stages of developing seeds. *GmMenB2* tended to have higher expression levels than *GmMenB1* in seedlings and seed coats, but lower levels in developing seeds (Figure 4.3). Taken together, soybean paralogs of the PhQ biosynthetic genes showed large redundancy in their expression, however, some expression divergence might have occurred in the sink tissues.

Expression and functional Divergence of ICS and DHNAT paralogs

As described above, distinct expression profiles were observed for *ICSs* in *Arabidopsis* and for *DHNATs* in all three species (Figure 4.1A-C). This suggested a potential functional divergence of (part of) the pathway, which was further investigated. *AtICS1* was dramatically up-regulated in photoautotrophic tissues by oxidative stress and upon pathogen infection (Figure 4.1A**,** Figure 4.3), supporting its essential role in biosynthesis of salicylic acid (SA) for defense (Wildermuth et al., 2001). Interestingly, we observed a strong up-regulation of *AtICS2* under dehydration and osmotic stresses like mannitol and salt treatments (Figure 4.1A, Figure 4.2, Figure 4.3), suggesting a potential role of *AtICS2* in mediating plant responses to these stimuli. These observations indicated that the two *ICS* copies were regulated independently and induced under different conditions.

Expression divergence was also observed for *DHNAT* duplicates in all three species (Figure 4.1A-C). *AtDHNAT1* was up-regulated by abiotic stresses including oxidative stress, dehydration, low Mg and salt stress, and biotic stresses including virulent and avirulent pathogen infections. *AtDHNAT1* was strongly induced only in photoautotrophic tissues and remained low in photoheterotrophic tissues (Figure 4.1A). In contrast, *AtDHNAT2* did not exhibit any stress response and was poorly expressed in photoautotrophic tissue. However, it showed high expression in photoheterotrophic tissues like seeds, germinated cotyledons, and nectary tissues (Figure 4.1A). No stress response was observed for poplar and soybean *DHNATs.* However, they exhibited specific expression patterns in some heterotrophic tissues. For example, *PtDHNAT1* was poorly expressed in photoautotrophic tissues but showed high expression in sink tissues like tension wood xylem and sepals of young flowers. *PtDHNAT2* displayed moderate expression in leaves and strong expression in callus and sepals of young flowers (Figure 4.1B). *GmDHNAT1* had relatively high expression in both source and sink tissues, whereas *GmDHNAT2* showed poor expression in leaves but high expression in roots and seed coats (Figure 4.1C). It appeared that in all three cases, one *DHNAT* gene was preferentially expressed in sink tissues.

The divergent expression patterns of *ICS* (in *Arabidopsis*) and *DHNAT* duplicates from the rest of the PhQ genes suggested an alternative role for these genes distinct from PhQ biosynthesis. Strong induction of *AtICS1*, *AtICS2*, and *AtDHNAT1* under stress conditions supported their involvement in plant stress responses. In addition, the preferential expression of one *DHNAT* duplicate in sink tissues of all three species hinted at DHNAT involvement in other non-photosynthetic processes. As no other PhQ genes shared this pattern, the non-photosynthetic functions of sink-tissue-expressed *DHNAT* likely involve dihydroxynaphthoate rather than PhQ *per se*.

Functional associations of the duplicated genes were inferred from GOBP enrichment analysis of their co-expressed genes. The enrichment patterns were quite similar for *AtICS1* and *AtDHNAT1* in both source and sink tissues (Figure 4.4A). Both genes were associated with

multiple abiotic and biotic stress responses, including hypersensitive response, response to cold, response to bacterium/fungus, response to water deprivation, salicylic acid mediated signaling pathway, and aging. Such enrichment patterns are consistent with the upregulation of *AtICS1* and *AtDHNAT1* by various stress treatments (Figure 4.1), and with the essential role of *AtICS1* in salicylic acid biosynthesis for defense (Wildermuth et al., 2001; Strawn et al., 2007).

AtICS2 and *AtDHNAT2* were clustered together in the GOBP enrichment analysis of their co-expressed genes, and the patterns differed from those for *AtICS1*/*AtDHNAT1* and the other PhQ genes. In source tissues, *AtICS2*-coexpressed genes were strongly associated with response to abscisic acid, response to water, response to osmotic stress cuticle development (Figure 4.4A), consistent with the strong induction of *AtICS2* to dehydration, salt and osmotic stresses (Figure 4.1, Figure 4.2). The co-expressed gene set of *AtDHNAT2* in source tissues was significantly enriched in post-embryonic root development, brassinosteroid metabolic process, cytokinin biosynthetic process, and response to nitrate (Figure 4.4A). The enrichment patterns of their co-expressed gene sets in sink tissues were weaker, likely due to the overall lower expression of *AtICS2* and *AtDHNAT2* in sink tissues. *AtICS2* was significantly associated with response to hormone and nitrate, reminiscent of *AtDHNAT2* associations in source tissues. *AtDHNAT2* in sink tissues was co-expressed with genes involved in fatty acid biosynthetic process and oxidation-reduction process. Together, *ICS* paralogs in *Arabidopsis* and *DHNAT* paralogs in all three species have exhibited distinct patterns in expression and functional association, providing evidence for their functional divergence. However, divergence of genes for the intermediate steps in the PhQ biosynthetic pathway does not necessarily reflect functional divergence of PhQ.

Discussion

Non-photosynthetic functions of PhQ

Here, we explored the non-photosynthetic functions of PhQ in photosynthetic taxa by mining the expression profiles of PhQ pathway genes across a wide range of tissue types and conditions, and by examining the co-expression gene sets of PhQ biosynthetic genes between photosynthetic source and sink tissues. PhQ biosynthetic genes exhibited strong expression in source tissues, consistent with the essential role of PhQ in photosynthesis. Moderate expression was also observed for some PhQ biosynthetic genes in the sink tissues, supporting a potential non-photosynthetic role there. Comparison of PhQ-coexpressed genes in sink versus source tissues in *Arabidopsis* revealed an over-representation of biological processes like defense response to bacterium, response to cold, detection of biotic stimulus, shoot system morphogenesis and pigment biosynthesis, suggesting a potential link with non-photosynthetic function of PhQ. However, this pattern was not observed in *Glycine max* or *Populus tremula x alba*.

The PhQ-coexpressed genes in source tissues were enriched in photosynthesis-related processes in all three species, as expected. However, this GO category was also significantly over-represented in the sink-tissue dataset, suggesting that photosynthesis-related processes cannot be fully excluded. In fact, the functional enrichment of photosynthesis in the sink dataset was likely due to the difficulty of sample classification and the low tissue-resolution (mixture of photosynthetic and non-photosynthetic cells) of publicly available data. We employed two PSI genes as a marker to gauge photosynthesis activity, but as expression varied along a continuum, no clear cutoff can be determined. Detection of low levels of PSI gene expression in sink tissues raised the possibility that PhQ gene expression there may still be associated with photosynthesis.

We also examined the expression patterns of the PhQ biosynthetic genes using microarray dataset from *Arabidopsis* roots with cellular resolution. High levels of expression were detected for some PhQ biosynthetic genes in some cell types, such as phloem companion cells,

cortex, and phloem pole pericycle, where PSI genes were poorly expressed. This localized expression pattern supported a role for PhQ in heterotrophic tissues non-related to PSI. However, datasets with such high resolution are limited, thereby hindering our ability to elucidate non-photosynthetic functions of PhQ at the present time.

Evolution of PhQ biosynthetic pathway genes

PhQ biosynthetic genes have undergone multiple rounds of gene duplication. In soybean, nearly the entire pathway has been retained as gene duplicates from the recent genome duplication 13 million years ago (Schmutz et al., 2010). Similar expression patterns were observed between paralogs within the pathway (except for *GmDHNATs*). It remains unclear whether such retention is a consequence of selection or whether divergence will require more time. Soybean seeds are well-known for the high PhQ content. It is possible that expression redundancy of PhQ gene paralogs influence the PhQ levels in soybean seeds via a dosage-dependent manner. Future work is needed to investigate the relationships between PhQ biosynthetic pathway duplication, high PhQ content in soybean seeds, and PhQ non-photosynthetic functions in the seeds.

In addition to the pathway-level duplication in soybean, single gene duplications have also been observed in *Arabidopsis* and poplar. Two copies of *ICS* were experimentally characterized in *Arabidopsis* previously (Widhalm et al., 2012). *AtICS1* is well known to participate in both PhQ and SA biosynthesis (Wildermuth et al., 2001; Yuan et al., 2009). Strong up-regulation of *AtICS2* expression in response to osmotic stresses observed in this work revealed a potential role of *AtICS2* in SA biosynthesis under these specific conditions. This induction was not observed for *AtICS1,* supporting the functional divergence between the *AtICS* paralogs. Interestingly, safflower *ICS* (*CtICS*) was shown to be involved in salt stress response (Sadeghi et al., 2013). Our observation suggested that *AtICS1* and *AtICS2* might have subfunctionalized since their divergence from *CtICS*. It is worth noting that SA promotes seed germination under salt stress (Lee et al., 2010), raising the possibility that *AtICS2* might contribute to SA biosynthesis for salt

tolerance during seed germination. Further investigations about the phenotype of *ics2* mutants under osmotic stresses, e.g. measuring seed germination rate, would provide clues to uncover the unique role of *AtICS2* in plant defense responses.

DHNAT was duplicated in *Arabidopsis* (Widhalm et al., 2012), as well as soybean and poplar as reported here. Differing expression profiles and functional enrichments were observed for the *DHNAT* paralogs in each of the three species, suggesting functional divergence after the duplication. In addition, *DHNAT*s exhibited similar expression patterns and functional associations with *ICS* in *Arabidopsis*, indicating that *DHNAT* may coop with *ICS* in certain defense responses in this species. A recent study showed that *DHNAT* was associated with peroxisomal β-oxidation during jasmonic acid metabolism, seed germination and early seedling growth (Cassin-Ross and Hu, 2014). Further experiments are needed to examine the phenotypes of *Arabidopsis dhnat* null mutants under various conditions, particularly in response to pathogen attack, Mg deficiency and salt stress when strong transcription induction was observed.

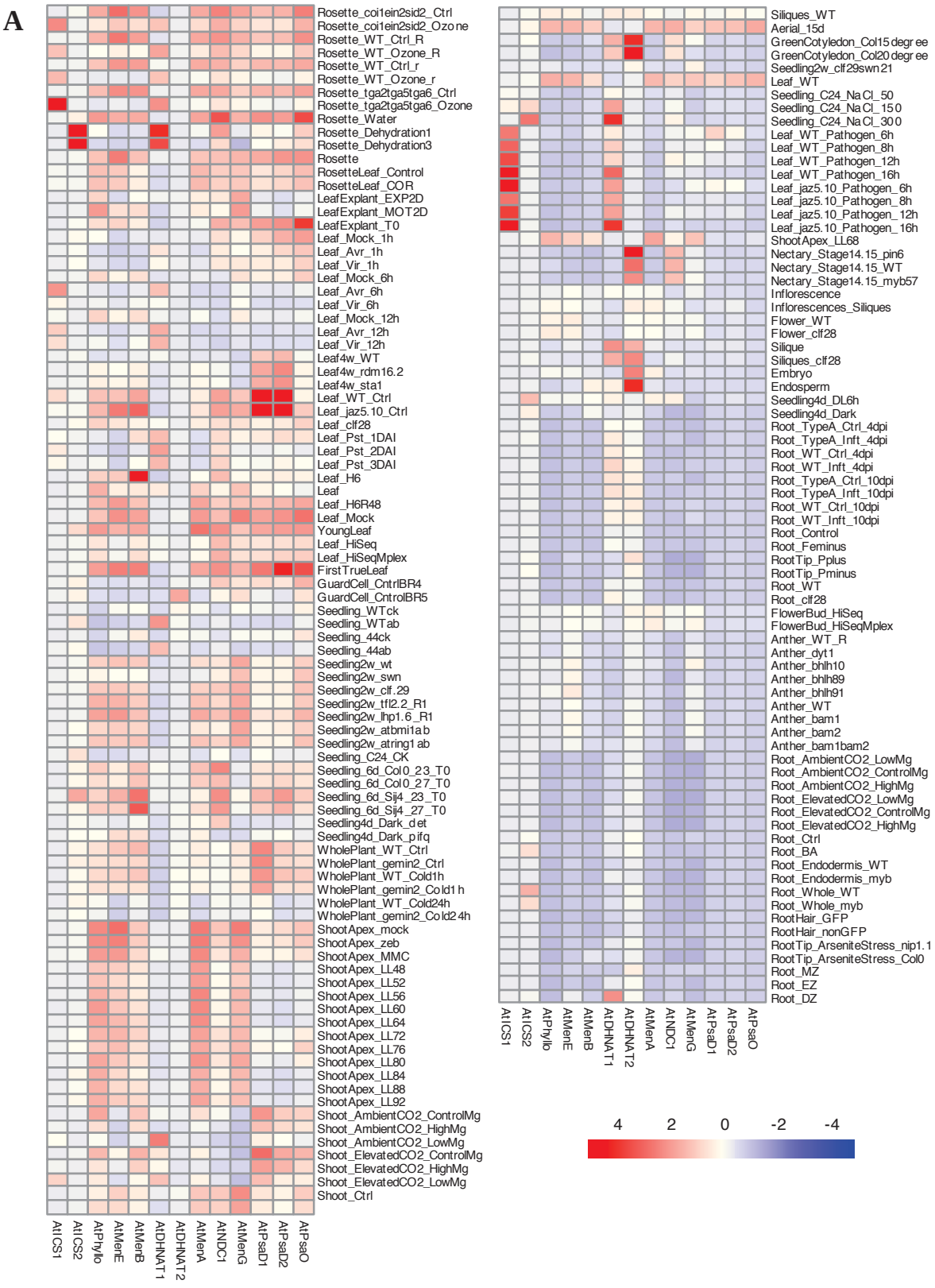
The gene duplication and expression divergence observed here were species-specific. The pathway-level retention was only observed in soybean with unusually high level of PhQ accumulation in seed. *ICS* divergence was specific to *Arabidopsis*, and is associated with defense-related SA biosynthesis (Wildermuth et al., 2001; Yuan et al., 2009). The duplication of *DHNAT* and the preferential expression of one of the paralogs in sink tissues were conserved in all three species examined here. This sink preference of *DHNAT* might be related to its involvement in peroximal β-oxidation, a multi-functional pathway associated with seed germination, embryo and flower development, as well as phytohormone biogenesis (Poirier et al., 2006). Previously, we also found that *MenG,* involved in the terminal step of PhQ biosynthesis, was present in duplicates in the two photosynthetic parasites, but not the non-photosynthetic holoparasite. One copy encodes a plastidic protein, and the other a plasma membrane protein due to loss of the plastid-targeting signal peptide. Together, these taxon-dependent duplication

events reveal a large degree of plasticity of the PhQ biosynthetic pathway that gives rise to functional divergence via gene duplication, retention and sub-/neo-functionalization.

In addition to SA biosynthesis and β-oxidation, the PhQ biosynthetic pathway was also found to share the same phytol biogenesis with tocopherol and chlorophyll biosynthesis (Van Oostende et al., 2011). Furthermore, NDC1, the enzyme recently shown to catalyze reduction of the naphthoquinone ring prior to the terminal methylation step in PhQ biosynthesis, was also involved in the redox regulation of the plastoquinone pool in chloroplasts (Eugeni Piller et al., 2011) and the redox cycle of tocopherol (Eugeni Piller, 2014). These examples begin to suggest a complex picture of cross talk between the PhQ biosynthetic pathway and other plastidial or peroxisomal biological processes (Basset, 2016). The work described in Chapter 3 suggested that this cross talk also involves plasma membrane-associated activities, which opens new opportunities for future investigations.

In closing, despite the progress made using the parasitic plant study system in Chapter 3, it remains a challenge to investigate the non-photosynthetic function of PhQ in photoautotrophic species. Currently available data from photoautotrophic species were predominantly derived from photosynthetic tissues and lack the tissue-level or cellular resolution to discern true heterotrophic tissues/cells. Further experiments expanding on the high cellular-resolution data from heterotrophic tissues are needed to establish a robust system to aid investigation of non-photosynthetic PhQ function.

A


Rosette_coi1ein2sid2_Ctrl
Rosette_coi1ein2sid2_Ozone
Rosette_WT_Ctrl_R
Rosette_WT_Ozone_R
Rosette_WT_Ctrl_r
Rosette_WT_Ozone_r
Rosette_tga2tga5tga6_Ctrl
Rosette_tga2tga5tga6_Ozone
Rosette_Water
Rosette_Dehydration1
Rosette_Dehydration3
Rosette
RosetteLeaf_Control
RosetteLeaf_COR
LeafExplant_EXP2D
LeafExplant_MOT2D
LeafExplant_T0
Leaf_Mock_1h
Leaf_Avr_1h
Leaf_Vir_1h
Leaf_Mock_6h
Leaf_Avr_6h
Leaf_Vir_6h
Leaf_Mock_12h
Leaf_Avr_12h
Leaf_Vir_12h
Leaf4w_WT
Leaf4w_rdm16.2
Leaf4w_sta1
Leaf_WT_Ctrl
Leaf_jaz5.10_Ctrl
Leaf_clf28
Leaf_Pst_1DAI
Leaf_Pst_2DAI
Leaf_Pst_3DAI
Leaf_H6
Leaf
Leaf_H6R48
Leaf_Mock
YoungLeaf
Leaf_HiSeq
Leaf_HiSeqMplex
FirstTrueLeaf
GuardCell_CntrlBR4
GuardCell_CntrolBR5
Seedling_WTck
Seedling_WTab
Seedling_44ck
Seedling_44ab
Seedling2w_wt
Seedling2w_swn
Seedling2w_clf.29
Seedling2w_tfl2.2_R1
Seedling2w_lhp1.6_R1
Seedling2w_atbmi1ab
Seedling2w_atring1ab
Seedling_C24_CK
Seedling_6d_Col0_23_T0
Seedling_6d_Col0_27_T0
Seedling_6d_Sij4_23_T0
Seedling_6d_Sij4_27_T0
Seedling4d_Dark_det
Seedling4d_Dark_pifq
WholePlant_WT_Ctrl
WholePlant_gemin2_Ctrl
WholePlant_WT_Cold1h
WholePlant_gemin2_Cold1h
WholePlant_WT_Cold24h
WholePlant_gemin2_Cold24h
ShootApex_mock
ShootApex_zeb
ShootApex_MMC
ShootApex_LL48
ShootApex_LL52
ShootApex_LL56
ShootApex_LL60
ShootApex_LL64
ShootApex_LL72
ShootApex_LL76
ShootApex_LL80
ShootApex_LL84
ShootApex_LL88
ShootApex_LL92
Shoot_AmbientCO2_ControlMg
Shoot_AmbientCO2_HighMg
Shoot_AmbientCO2_LowMg
Shoot_ElevatedCO2_ControlMg
Shoot_ElevatedCO2_HighMg
Shoot_ElevatedCO2_LowMg
Shoot_Ctrl
AtICS1
AtICS2
AtPhyllo
AtMenE
AtMenB
AtDHNAT1
AtDHNAT2
AtMenA
AtNDC1
AtMenG
AtPsaD1
AtPsaD2
AtPsaO
Siliques_WT
Aerial_15d
GreenCotyledon_Col15degree
GreenCotyledon_Col20degree
Seedling2w_clf29swn21
Leaf_WT
Seedling_C24_NaCl_50
Seedling_C24_NaCl_150
Seedling_C24_NaCl_300
Leaf_WT_Pathogen_6h
Leaf_WT_Pathogen_8h
Leaf_WT_Pathogen_12h
Leaf_WT_Pathogen_16h
Leaf_jaz5.10_Pathogen_6h
Leaf_jaz5.10_Pathogen_8h
Leaf_jaz5.10_Pathogen_12h
Leaf_jaz5.10_Pathogen_16h
ShootApex_LL68
Nectary_Stage14.15_pin6
Nectary_Stage14.15_WT
Nectary_Stage14.15_myb57
Inflorescence
Inflorescences_Siliques
Flower_WT
Flower_clf28
Silique
Siliques_clf28
Embryo
Endosperm
Seedling4d_DL6h
Seedling4d_Dark
Root_TypeA_Ctrl_4dpi
Root_TypeA_Inft_4dpi
Root_WT_Ctrl_4dpi
Root_WT_Inft_4dpi
Root_TypeA_Ctrl_10dpi
Root_TypeA_Inft_10dpi
Root_WT_Ctrl_10dpi
Root_WT_Inft_10dpi
Root_Control
Root_Feminus
RootTip_Pplus
RootTip_Pminus
Root_WT
Root_clf28
FlowerBud_HiSeq
FlowerBud_HiSeqMplex
Anther_WT_R
Anther_dyt1
Anther_bhlh10
Anther_bhlh89
Anther_bhlh91
Anther_WT
Anther_bam1
Anther_bam2
Anther_bam1bam2
Root_AmbientCO2_LowMg
Root_AmbientCO2_ControlMg
Root_AmbientCO2_HighMg
Root_ElevatedCO2_LowMg
Root_ElevatedCO2_ControlMg
Root_ElevatedCO2_HighMg
Root_Ctrl
Root_BA
Root_Endodermis_WT
Root_Endodermis_myb
Root_Whole_WT
Root_Whole_myb
RootHair_GFP
RootHair_nonGFP
RootTip_ArseniteStress_nip1.1
RootTip_ArseniteStress_Col0
Root_MZ
Root_EZ
Root_DZ
AtICS1
AtICS2
AtPhyllo
AtMenE
AtMenB
AtDHNAT1
AtDHNAT2
AtMenA
AtNDC1
AtMenG
AtPsaD1
AtPsaD2
AtPsaO
4
2
0
-2
-4

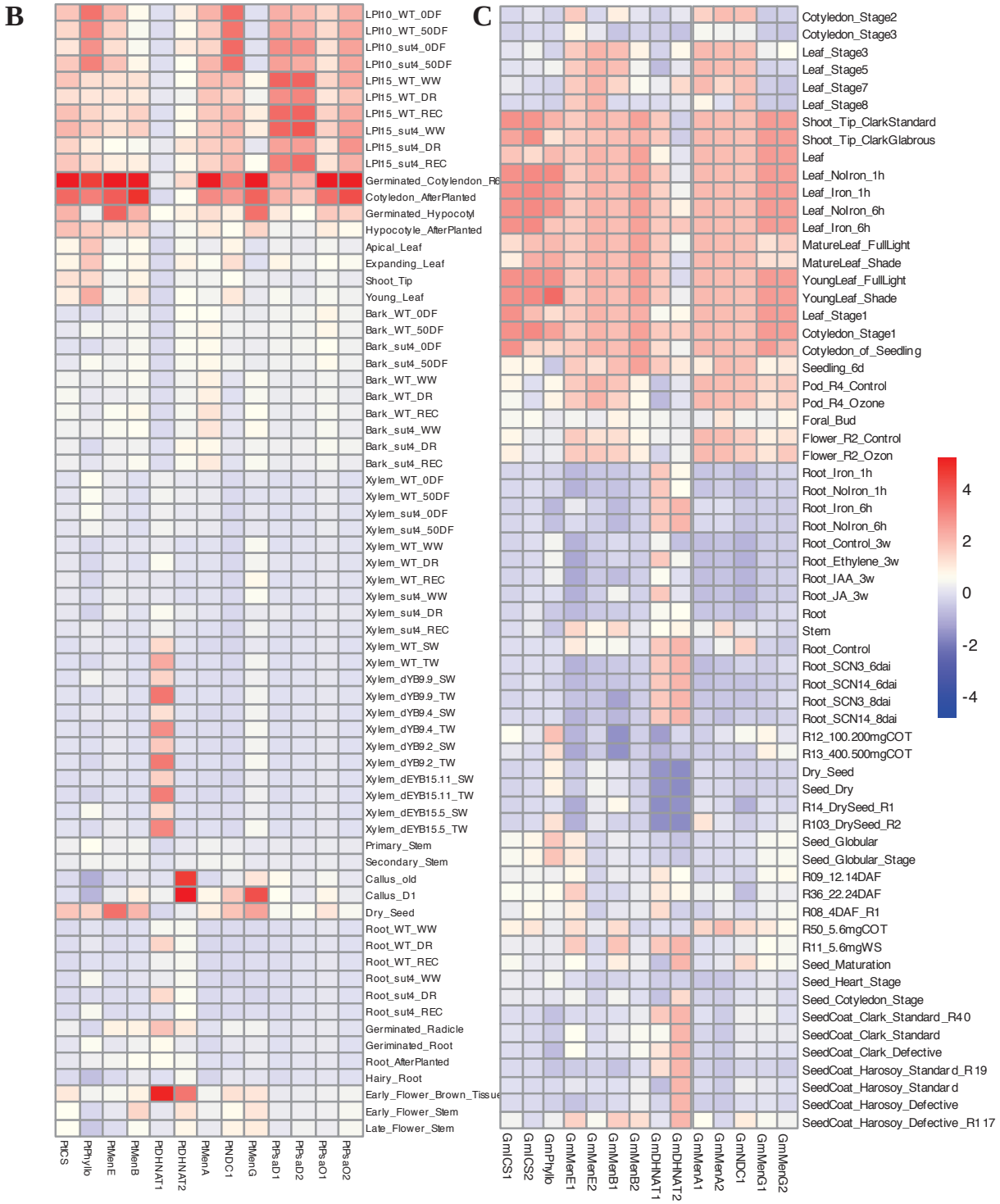

B
LPI10_WT_0DF
LPI10_WT_50DF
LPI10_sut4_0DF
LPI10_sut4_50DF
LPI15_WT_WW
LPI15_WT_DR
LPI15_WT_REC
LPI15_sut4_WW
LPI15_sut4_DR
LPI15_sut4_REC
Germinated_Cotylendon_R6
Cotyledon_AfterPlanted
Germinated_Hypocotyl
Hypocotyle_AfterPlanted
Apical_Leaf
Expanding_Leaf
Shoot_Tip
Young_Leaf
Bark_WT_0DF
Bark_WT_50DF
Bark_sut4_0DF
Bark_sut4_50DF
Bark_WT_WW
Bark_WT_DR
Bark_WT_REC
Bark_sut4_WW
Bark_sut4_DR
Bark_sut4_REC
Xylem_WT_0DF
Xylem_WT_50DF
Xylem_sut4_0DF
Xylem_sut4_50DF
Xylem_WT_WW
Xylem_WT_DR
Xylem_WT_REC
Xylem_sut4_WW
Xylem_sut4_DR
Xylem_sut4_REC
Xylem_WT_SW
Xylem_WT_TW
Xylem_dYB9.9_SW
Xylem_dYB9.9_TW
Xylem_dYB9.4_SW
Xylem_dYB9.4_TW
Xylem_dYB9.2_SW
Xylem_dYB9.2_TW
Xylem_dEYB15.11_SW
Xylem_dEYB15.11_TW
Xylem_dEYB15.5_SW
Xylem_dEYB15.5_TW
Primary_Stem
Secondary_Stem
Callus_old
Callus_D1
Dry_Seed
Root_WT_WW
Root_WT_DR
Root_WT_REC
Root_sut4_WW
Root_sut4_DR
Root_sut4_REC
Germinated_Radicle
Geriminated_Root
Root_AfterPlanted
Hairy_Root
Early_Flower_Brown_Tissue
Early_Flower_Stem
Late_Flower_Stem
PtCS
PtPhyllo
PtMenE
PtMenB
PtDHNAT1
PtDHNAT2
PtMenA
PtNDC1
PtMenG
PtPsaD1
PtPsaD2
PtPsaO1
PtPsaO2
C
Cotyledon_Stage2
Cotyledon_Stage3
Leaf_Stage3
Leaf_Stage5
Leaf_Stage7
Leaf_Stage8
Shoot_Tip_ClarkStandard
Shoot_Tip_ClarkGlabrous
Leaf
Leaf_NoIron_1h
Leaf_Iron_1h
Leaf_NoIron_6h
Leaf_Iron_6h
MatureLeaf_FullLight
MatureLeaf_Shade
YoungLeaf_FullLight
YoungLeaf_Shade
Leaf_Stage1
Cotyledon_Stage1
Cotyledon_of_Seedling
Seedling_6d
Pod_R4_Control
Pod_R4_Ozone
Foral_Bud
Flower_R2_Control
Flower_R2_Ozon
Root_Iron_1h
Root_NoIron_1h
Root_Iron_6h
Root_NoIron_6h
Root_Control_3w
Root_Ethylene_3w
Root_IAA_3w
Root_JA_3w
Root
Stem
Root_Control
Root_SCN3_6dai
Root_SCN14_6dai
Root_SCN3_8dai
Root_SCN14_8dai
R12_100.200mgCOT
R13_400.500mgCOT
Dry_Seed
Seed_Dry
R14_DrySeed_R1
R103_DrySeed_R2
Seed_Globular
Seed_Globular_Stage
R09_12.14DAF
R36_22.24DAF
R08_4DAF_R1
R50_5.6mgCOT
R11_5.6mgWS
Seed_Maturation
Seed_Heart_Stage
Seed_Cotyledon_Stage
SeedCoat_Clark_Standard_R40
SeedCoat_Clark_Standard
SeedCoat_Clark_Defective
SeedCoat_Harosoy_Standard_R19
SeedCoat_Harosoy_Standard
SeedCoat_Harosoy_Defective
SeedCoat_Harosoy_Defective_R117
GmCS1
GmCS2
GmPhyllo
GmMenE1
GmMenE2
GmMenB1
GmMenB2
GmDHNAT1
GmDHNAT2
GmMenA1
GmMenA2
GmNDC1
GmMenG1
GmMenG2
4
2
0
-2
-4

Figure 4.1. Expression profiles of PhQ genes and PSI genes in RNA-Seq Atlas of *Arabidopsis thaliana* (**A**), *Populus tremula x alba* (**B**), and *Glycine max* (**C**). PhQ genes were organized by the order in the PhQ biosynthetic pathway. Tissues were organized manually organized according to the clustering in Figure S4.1-4.3.

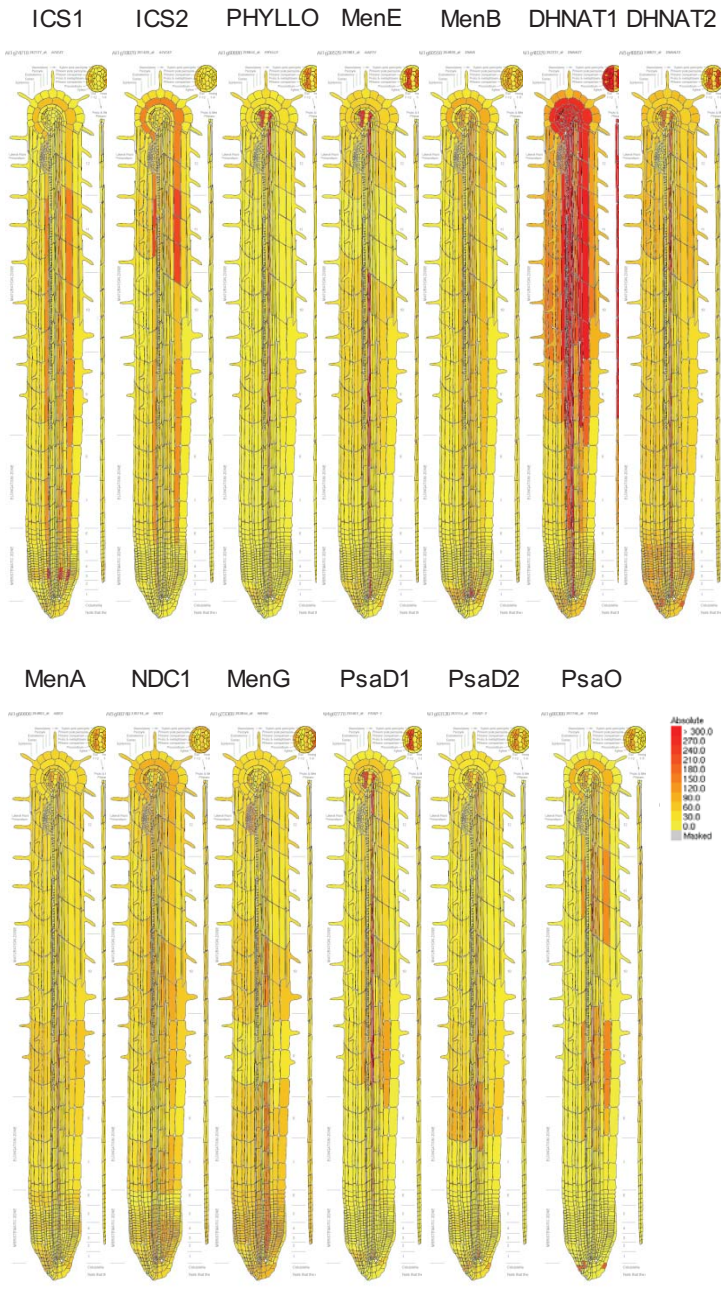


Figure 4.2. Expression patterns of PhQ and PSI genes in *Arabidopsis* root microarray data with cellular resolution. PhQ genes were organized by the order in the PhQ biosynthetic pathway.

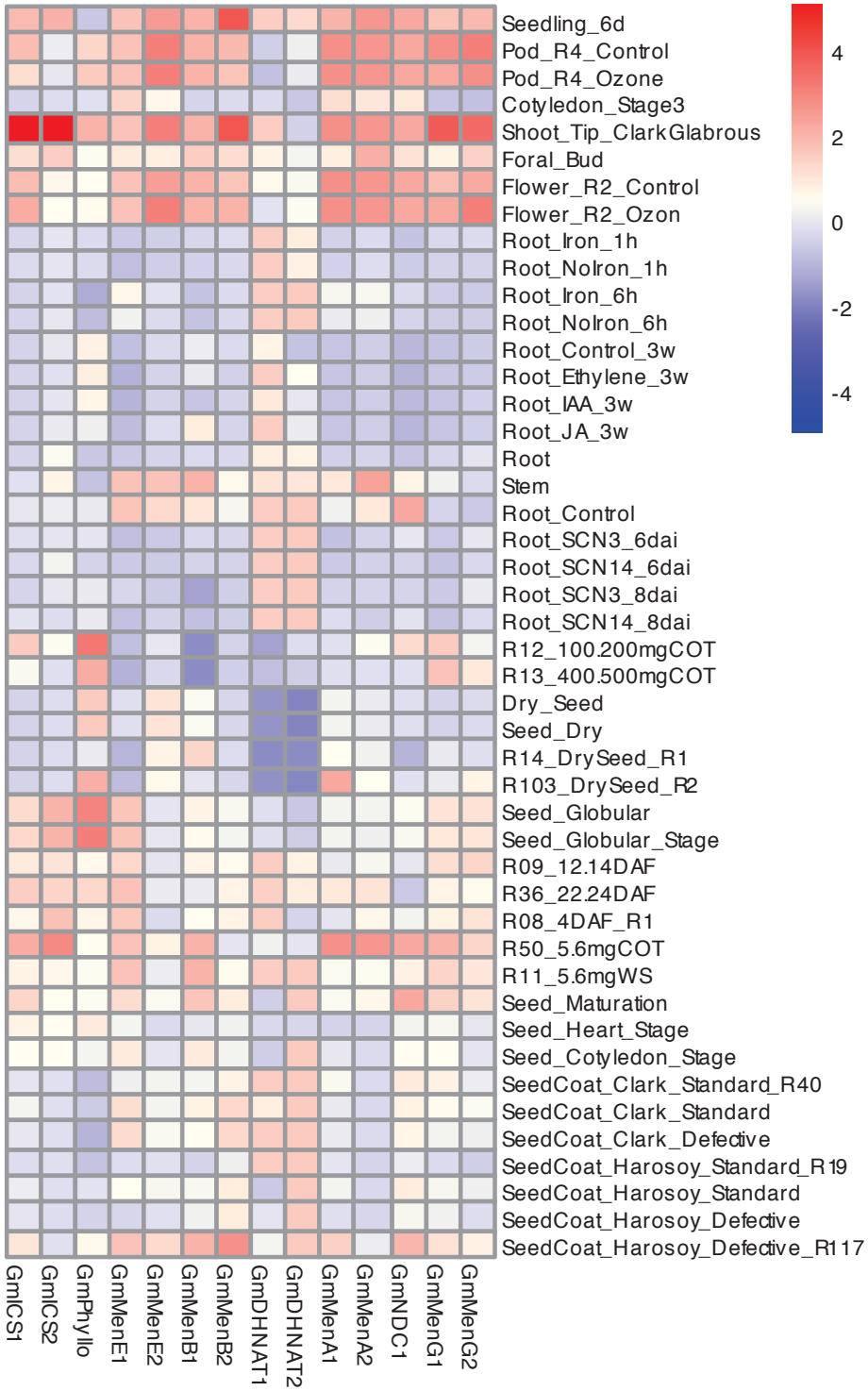


Figure 4.3. Expression conservations and divergence of PhQ paralogs in heterotrophic tissues of *Glycine max*. PhQ genes were organized by the order in the PhQ biosynthetic pathway.


A
Color Key
0
4
8
Value
regulation of cell cycle
G2/M transition of mitotic cell cycle
gene silencing
cytokinesis by cell plate formation
microtubule-based process
stomatal complex development
cell proliferation
regulation of cell death
response to nitrate
response to hormone
defense response by callose deposition
respiratory burst involved in defense re...
toxin metabolic process
vesicle-mediated transport
microtubule nucleation
pigment metabolic process
asymmetric cell division
cuticle development
fatty acid biosynthetic process
response to osmotic stress
response to water
response to abscisic acid
carotenoid biosynthetic process
photosynthesis, light reaction
aromatic amino acid family metabolic pro...
glyceraldehyde-3-phosphate metabolic pro...
oxidation-reduction process
cytokinin biosynthetic process
brassinosteroid metabolic process
post-embryonic root development
chloroplast relocation
embryo development
stomatal complex morphogenesis
leaf morphogenesis
tetrapyrrole biosynthetic process
salicylic acid mediated signaling pathwa...
response to endoplasmic reticulum stress
response to ethylene
response to bacterium
response to water deprivation
response to fungus
regulation of immune response
defense response
response to hypoxia
signal transduction
plant-type hypersensitive response
aging
detection of biotic stimulus
regulation of multi-organism process
response to cold
thylakoid membrane organization
photosynthesis
positive regulation of catalytic activit...
isopentenyl diphosphate biosynthetic pro...
pentose-phosphate shunt
ovule development
pigment biosynthetic process
cell differentiation
embryo development ending in seed dorman...
translation
lipoate metabolic process
defense response to bacterium
glucosinolate metabolic process
shoot system morphogenesis
AtDHNAT1
AtICS1
AtICS1
AtDHNAT1
AtICS2
AtDHNAT2
AtDHNAT2
AtICS2
AtMenA
AtMenG
AtPhyllo
AtMenE
AtNDC1
AtMenE
AtMenB
AtNDC1
AtMenG
AtMenB
AtPhyllo
AtMenA


B
Color Key
0
4
8
Value
positive regulation of catalytic activit...
glucosinolate metabolic process
cell differentiation
detection of biotic stimulus
regulation of multi-organism process
carbon fixation
systemic acquired resistance, salicylic ...
lipoate metabolic process
vitamin metabolic process
cellular response to water deprivation
response to karrikin
de-etiolation
embryo development ending in seed dorman...
translation
carotenoid biosynthetic process
chlorophyll biosynthetic process
defense response to bacterium
chloroplast relocation
cuticle development
photosynthesis
response to cold
thylakoid membrane organization
shoot system morphogenesis
glyceraldehyde-3-phosphate metabolic pro...
glycerol ether metabolic process
aromatic amino acid family metabolic pro...
radial pattern formation
maintenance of protein localization in o...
water transport
pigment biosynthetic process
response to misfolded protein
fatty acid oxidation
chlorophyll catabolic process
microgametogenesis
autophagy
vesicle-mediated transport
microtubule-based movement
small GTPase mediated signal transductio...
response to gibberellin
endosomal transport
response to molecule of fungal origin
transmembrane transport
D-ribose metabolic process
Golgi vesicle transport
embryo development
gravitropism
lipid storage
sister chromatid cohesion
cell-cell signaling
ovule development
tetrapyrrole biosynthetic process
auxin biosynthetic process
pigment metabolic process
photosynthesis, light reaction
response to wounding
response to endogenous stimulus
toxin catabolic process
programmed cell death
hydrogen peroxide biosynthetic process
heat acclimation
drug transport
response to fungus
lipid transport
defense response, incompatible interacti...
response to nitrate
somatic embryogenesis
cell wall macromolecule catabolic proces...
lipid localization
cell wall modification involved in multi...
defense response to fungus, incompatible...
pentose-phosphate shunt
isopentenyl diphosphate biosynthetic pro...
response to herbivore
response to salt stress
response to desiccation
PtMenA
PtNDC1
PtMenE
PtICS
PtMenB
PtDHNAT2
PtMenG
PtPhyllo
PtNDC1
PtDHNAT1
PtDHNAT1
PtDHNAT2
PtPhyllo
PtMenB
PtMenG
PtMenE
PtMenA
PtICS

Figure 4.4. GOBP enrichment for top 500 genes most highly correlated with PhQ genes in both autotrophic and heterotrophic tissues. (A). *Arabidopsis thaliana*, (**B**). *Glycine max*, (**C**). *Populus trichocarpa*. Green and blue labels represent top500-source and top500-sink sets, respectively.

Table S4.1. Gene Expression of PhQ and PSI in *Arabidopsis* Microarray in Root

Tissue	G1*	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
longitudinal zone 1, 140 mM NaCl	18	9	11	37	21	20	28	19	24	72	9	24	2
longitudinal zone 1, standard conditions	17	6	9	41	29	25	39	22	17	57	5	2	3
longitudinal zone 2, standard conditions	26	6	6	17	18	43	31	14	19	45	8	2	1
longitudinal zone 2, 140 mM NaCl	22	4	5	12	11	68	38	11	25	37	6	22	1
longitudinal zone 3, standard conditions	35	11	3	9	11	184	13	17	22	27	40	6	6
longitudinal zone 3, 140 mM NaCl	31	9	9	14	15	258	23	19	19	21	53	14	23
longitudinal zone 4, 140 mM NaCl	43	69	10	11	18	438	32	16	20	29	22	7	39
longitudinal zone 4, standard conditions	41	64	7	23	17	371	31	12	22	33	36	4	25
epidermis and lateral root cap, standard conditions	15	5	9	14	20	20	23	21	35	28	6	4	1
epidermis and lateral root cap, 140 mM NaCl	14	10	5	23	13	17	16	15	34	32	9	22	3
epidermis and lateral root cap, -Fe	14	9	3	29	13	19	14	13	42	30	8	5	4
columella root cap, standard conditions	19	12	15	9	19	69	47	30	44	45	11	24	5
columella root cap, 140 mM NaCl	14	16	5	6	7	90	26	15	32	60	7	30	3
columella root cap, -Fe	13	4	4	12	22	47	20	12	25	27	6	14	2
cortex, standard conditions	111	11 1	11	19	43	142	19	19	59	31	33	17	49
cortex, 140 mM NaCl	86	12 1	17	18	46	342	52	29	42	24	60	43	165
cortex, -Fe	71	83	6	26	71	209	26	14	30	20	54	17	174
endodermis and quiescent center, 140 mM NaCl	19	19	9	96	14	160	22	16	32	33	24	20	10
endodermis and quiescent center, -Fe	18	5	10	81	13	237	28	21	25	25	44	5	15
endodermis and quiescent center, standard conditions	16	16	10	167	17	175	17	17	36	44	49	15	13
stele, 140 mM NaCl	22	13	9	9	13	108	18	10	31	39	13	25	20
stele, -Fe	20	15	5	23	9	198	14	8	35	33	15	6	21
stele, standard conditions	18	25	6	24	20	58	23	26	53	65	17	9	10
protophloem, 140 mM NaCl	32	16	6	11	7	57	19	14	43	39	8	19	6
protophloem, standard conditions	23	12	13	21	21	159	19	22	34	42	13	9	16
Whole root, standard conditions (control)	22	19	4	17	25	154	20	7	40	35	18	10	20
Whole root, 32 hours of 140 mM NaCl exposure	22	13	4	9	25	172	27	10	24	40	60	19	119
Whole root, 30 minutes of 140 mM NaCl exposure	21	19	4	13	12	196	23	5	25	32	11	13	18
Whole root, 16 hours of 140 mM NaCl exposure	20	24	4	14	22	199	30	9	28	31	27	11	47
Whole root, 4 hours of 140 mM NaCl exposure	18	41	4	13	22	201	21	6	22	31	13	5	17
Whole root, 1 hour of 140 mM NaCl exposure	14	20	5	14	15	216	18	7	26	34	9	18	18
longitudinal zone 1, standard conditions	22	4	1	53	42	20	26	19	65	75	79	58	421
longitudinal zone 1, -Fe conditions	20	5	3	63	44	16	32	15	60	86	46	19	238
longitudinal zone 2, standard conditions	33	4	3	25	20	28	24	15	60	41	19	12	31
longitudinal zone 2, -Fe conditions	23	5	1	16	10	19	19	14	43	31	18	2	15

Table S4.1 (continued). Gene Expression of PhQ and PSI in *Arabidopsis* Microarray in Root

Tissue	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
longitudinal zone 3, standard conditions	49	18	8	53	42	157	20	15	42	23	99	24	97
longitudinal zone 3, -Fe conditions	37	6	5	32	28	170	8	10	39	19	47	14	58
longitudinal zone 4, standard conditions	41	42	4	34	57	241	18	22	71	47	77	17	162
longitudinal zone 4, -Fe conditions	32	165	5	24	46	284	17	8	42	24	43	14	124
Whole root, -Fe, 48 hour	43	32	1	20	10	315	17	8	36	31	5	4	5
Whole root, -Fe, 72 hour	36	19	2	19	10	361	21	8	36	29	6	8	9
Whole root, -Fe, 6 hour	30	53	1	25	29	181	5	16	52	31	49	15	85
Whole root, -Fe, 3 hour	24	39	7	23	29	179	32	16	43	17	65	12	51
Whole root, standard conditions, control	23	34	2	17	22	175	17	16	49	20	74	10	55
Whole root, -Fe, 24 hour	21	48	4	14	35	221	19	17	42	27	49	13	84
Whole root, -Fe, 12 hour	20	48	4	16	30	219	29	16	45	22	62	15	82
Lateral Root Cap root cells 2hr continuous KNO3 treated	328	159	270	85	187	302	257	284	387	354	304	139	345
Lateral Root Cap root cells 2hr KCl control treated	295	427	24	359	57	154	126	92	219	309	254	229	268
Epidermis and Cortex root cells 2hr KCl control treated	273	155	69	210	207	996	241	211	333	270	544	421	1398
Epidermis and Cortex root cells 2hr continuous KNO3 treated	163	193	192	80	35	731	250	148	426	260	89	87	103
Endodermis and Pericycle root cells 2hr KCl control treated	390	130	112	166	65	1048	331	296	407	300	174	96	37
Endodermis and Pericycle root cells 2hr continuous KNO3 treated	292	198	66	251	84	1270	261	220	494	227	185	223	152
Pericycle root cells 2hr KCl control treated	324	88	53	132	134	895	168	173	337	332	255	255	372
Pericycle root cells 2hr continuous KNO3 treated	300	108	110	70	116	1373	155	107	501	258	315	312	281
Stele root cells 2hr continuous KNO3 treated	161	69	77	71	108	456	267	149	449	338	9	93	11
Stele root cells 2hr KCl control treated	139	106	65	94	91	839	219	214	514	305	42	93	6
cortex 3	295	63	9	27	67	34	76	14	119	80	12	6	8
phloem companion cell 3	247	21	3	11	59	19	54	9	116	187	4	13	5
cortex 11	149	225	10	8	62	321	60	24	59	25	45	12	131
phloem pole pericycle 3	149	32	3	5	62	48	48	13	130	184	8	7	7
cortex 9	140	116	24	46	72	183	36	45	75	21	102	30	138
cortex 7	132	129	5	19	47	72	31	22	85	53	26	13	17
cortex 10	130	71	13	47	76	374	35	25	96	58	67	18	35
phloem companion cell 11	125	75	4	3	55	181	42	15	57	58	16	26	86
phloem companion cell 9	117	39	9	20	64	103	26	29	72	48	36	63	90
cortex 6	115	102	7	14	45	38	44	37	12	64	31	8	3
Phloem companion cells, young (average of levels in PC cells in sections 2-6)	114	23	3	11	44	20	41	17	74	170	5	28	8
phloem companion cell 7	110	43	2	8	41	41	22	14	83	124	9	27	11
phloem companion cell 10	109	24	5	20	67	210	25	16	93	135	23	38	23

Table S4.1 (continued). Gene Expression of PhQ and PSI in *Arabidopsis* Microarray in Root

Tissue	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
cortex 4	101	60	8	25	15	39	54	37	105	80	10	25	29
Phloem companion cells, old (average of levels in PC cells in sections 7-12)	99	45	6	14	61	120	30	20	71	100	19	55	52
phloem companion cell 6	97	34	3	6	40	21	31	24	11	150	11	16	2
cortex 1	95	54	14	32	230	44	49	28	120	70	27	36	5
cortex 2	94	83	5	43	83	25	45	22	82	75	3	14	16
cortex 8	87	127	11	37	63	66	44	34	34	61	24	61	79
phloem companion cell 4	85	20	3	11	13	22	38	24	102	188	3	53	19
phloem companion cell 2	79	28	2	18	74	14	32	14	79	177	1	28	10
cortex 5	77	30	10	20	39	39	73	19	60	62	16	14	8
phloem pole pericycle 11	75	113	4	2	57	451	38	22	64	57	30	13	114
phloem companion cell 8	73	42	4	16	56	37	31	22	33	143	8	128	51
phloem pole pericycle 9	71	58	8	9	66	258	23	42	82	47	69	32	120
cortex 12	70	144	28	45	92	264	47	33	89	38	59	24	81
non root hair cell 3	66	14	5	26	20	14	112	18	45	63	3	9	0
phloem pole pericycle 7	66	65	2	4	43	101	20	20	93	122	18	14	15
phloem pole pericycle 10	65	35	4	9	69	525	22	23	105	133	45	19	30
Phloem Pole Pericycle, young (average of levels in PPP cells in sections 1-6)	65	33	3	5	73	51	36	24	91	166	11	18	10
phloem companion cell 5	64	10	4	8	34	22	52	12	59	147	5	28	5
lateral root cap	62	12	5	76	31	101	198	77	15	82	7	101	197
Phloem Pole Pericycle, old (average of levels in PPP cells in sections 7-12)	59	68	5	7	63	300	27	28	80	98	36	28	70
phloem companion cell 12	58	48	11	19	81	148	33	21	87	89	21	50	53
phloem pole pericycle 6	58	51	2	3	41	53	28	35	13	148	21	8	3
xylem pole pericycle 3	57	11	9	7	15	34	46	11	116	61	3	3	0
meta protophloem 3	56	9	9	24	26	56	46	13	65	72	7	4	4
hair 3	52	8	11	14	13	8	71	18	71	89	2	6	0
phloem pole pericycle 4	51	30	3	5	14	55	34	34	115	184	7	27	26
endodermis 3	50	9	7	35	12	60	109	19	80	50	11	5	2
phloem pole pericycle 1	48	27	5	7	212	61	31	26	131	161	18	38	4
phloem pole pericycle 2	47	42	2	9	77	36	29	20	89	174	2	15	14
lateral root primordium 11	47	57	4	8	40	175	13	29	76	123	28	26	31
phloem pole pericycle 8	44	64	4	7	58	92	28	32	37	141	16	66	69
phloem pole pericycle 5	39	15	3	4	36	55	46	18	66	144	11	14	7
columella	36	13	3	7	11	58	56	26	38	76	3	29	2
phloem pole pericycle 12	35	72	10	9	84	370	30	30	98	88	40	26	71

Table S4.1 (continued). Gene Expression of PhQ and PSI in *Arabidopsis* Microarray in Root

Tissue	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
non root hair cell 11	34	50	6	8	19	132	88	32	22	19	11	17	8
lateral root cap 3	33	8	29	27	24	51	155	26	42	35	23	5	2
xylem 3	33	20	12	18	21	99	66	11	100	97	9	7	1
non root hair cell 9	31	26	14	44	21	75	54	60	28	16	26	42	8
non root hair cell 7	30	28	3	18	14	30	46	29	32	42	7	18	1
non root hair cell 10	29	16	8	45	23	153	52	33	36	45	17	25	2
xylem pole pericycle 11	29	38	10	2	14	321	36	19	57	19	10	6	2
meta protophloem 11	28	33	11	7	24	533	36	23	32	22	25	7	57
xylem pole pericycle 9	27	19	23	12	16	183	22	35	73	16	23	15	3
meta protophloem 9	26	17	24	41	28	305	22	43	41	19	55	17	60
hair 11	26	28	13	4	13	74	56	32	35	28	7	12	5
non root hair cell 6	26	22	4	14	13	16	65	49	4	50	8	11	0
Protophloem and Metaphloem, young (average of phloem levels in sections 2-6)	26	10	8	23	19	58	36	24	42	65	8	8	6
xylem pole pericycle 7	25	22	5	5	10	72	19	17	83	40	6	7	0
endodermis 11	25	32	8	11	11	563	86	33	39	15	40	11	27
xylem pole pericycle 10	25	12	13	13	17	373	21	20	93	44	15	9	1
Xylem Pole Pericycle, young (average of levels in XPP cells in sections 1-6)	25	11	9	7	18	37	34	20	81	55	4	9	0
meta protophloem 7	25	19	5	17	18	120	19	21	47	48	14	7	7
meta protophloem 10	25	10	13	42	29	620	21	24	53	52	36	10	15
hair 9	24	14	29	24	14	42	34	60	44	23	15	29	6
endodermis 9	24	17	18	60	12	322	52	63	50	13	91	26	28
hair 7	23	16	6	10	9	17	29	29	51	59	4	12	1
non root hair cell 4	23	13	5	25	4	16	80	49	40	63	2	36	2
hair 10	23	9	16	24	15	86	33	33	57	65	10	17	1
Xylem Pole Pericycle, old (average of levels in XPP cells in sections 7-12)	23	23	15	9	15	213	26	24	71	33	12	13	1
Protophloem and Metaphloem, old (average of phloem levels in sections 7-12)	22	20	15	30	26	354	26	29	40	38	29	15	35
endodermis 7	22	18	4	24	8	127	45	30	57	33	23	11	3
xylem pole pericycle 6	22	17	7	4	10	38	27	29	11	49	7	4	0
endodermis 10	22	10	10	61	13	655	51	35	64	36	59	15	7
meta protophloem 6	22	15	7	13	17	63	27	35	6	58	17	4	1
non root hair cell 1	21	12	8	31	69	18	72	37	45	55	7	50	0
non root hair cell 2	21	18	3	42	25	10	66	28	31	59	1	19	1
quiescent center 1	20	5	12	127	70	15	46	52	120	139	6	29	0
hair 6	20	12	8	7	9	9	41	49	7	72	5	7	0

Table S4.1 (continued). Gene Expression of PhQ and PSI in *Arabidopsis* Microarray in Root

Tissue	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
non root hair cell 8	20	28	6	36	19	27	64	45	13	48	6	87	5
endodermis 6	19	15	5	19	8	67	63	52	8	40	27	7	1
xylem pole pericycle 4	19	10	8	7	3	39	33	29	102	61	2	13	1
meta protophloem 4	19	9	8	23	6	65	33	35	58	72	5	14	13
xylem pole pericycle 1	18	9	14	9	51	44	30	22	117	54	6	18	0
xylem pole pericycle 2	18	14	5	12	18	25	27	17	80	58	1	7	0
meta protophloem 2	18	12	5	39	32	42	28	20	45	68	2	8	7
hair 4	18	7	10	13	3	9	50	49	62	89	1	24	1
non root hair cell 5	17	7	6	19	12	16	108	25	23	49	4	19	0
endodermis 4	17	9	7	33	3	69	78	51	70	50	9	22	6
xylem 11	17	71	14	6	19	932	52	19	49	30	32	14	14
xylem pole pericycle 8	17	21	11	10	14	66	27	27	33	47	5	31	1
hair 1	17	7	17	17	46	10	46	38	71	78	4	34	0
hair 2	16	10	6	22	17	6	42	28	49	84	0	13	1
meta protophloem 8	16	19	11	33	24	109	27	32	19	55	13	35	34
endodermis 1	16	8	11	42	40	77	71	39	80	44	24	30	1
endodermis 2	16	12	4	57	14	44	65	30	55	47	3	12	3
xylem 9	16	37	32	31	22	532	32	37	63	25	72	33	15
non root hair cell 12	16	32	17	43	27	108	70	43	34	30	15	34	5
hair 8	15	16	13	19	13	15	41	45	20	68	4	59	3
xylem 7	15	41	7	13	14	210	27	18	72	64	18	14	2
xylem pole pericycle 5	15	5	9	5	9	39	44	15	59	48	4	7	0
endodermis 8	15	18	8	48	11	115	63	47	23	38	21	52	16
xylem 10	15	22	17	32	23	1084	31	20	80	70	47	20	4
Xylem, young (average of xylem levels in sections 1-6)	15	21	12	18	25	106	49	21	70	88	12	19	1
meta protophloem 5	15	4	10	18	15	65	45	18	33	56	8	8	4
hair 5	13	4	12	10	8	9	68	25	36	70	2	13	0
xylem pole pericycle 12	13	24	28	12	20	263	29	25	87	29	13	12	1
Xylem, old (average of xylem levels in sections 7-12)	13	43	20	23	21	619	37	25	61	52	38	29	9
meta protophloem 12	13	21	29	40	35	438	29	31	49	34	32	14	35
xylem 6	13	32	9	10	14	110	38	30	10	78	22	9	0
endodermis 5	13	4	7	26	7	69	106	26	40	39	14	12	2
hair 12	12	18	35	23	19	60	44	43	53	43	9	23	3
endodermis 12	12	21	22	59	16	462	68	45	60	24	53	21	17
lateral root cap 4	11	8	28	26	5	59	110	71	37	35	18	19	7
xylem 4	11	19	11	17	5	114	47	30	88	98	7	28	3
lateral root cap 1	11	7	46	33	82	66	100	54	42	31	50	27	1

Table S4.1 (continued). Gene Expression of PhQ and PSI in *Arabidopsis* Microarray in Root

Tissue	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
xylem 1	11	17	18	22	71	127	43	23	101	85	19	40	1
lateral root cap 2	11	11	17	44	30	38	92	41	29	33	5	10	3
xylem 2	11	26	7	29	26	73	39	17	68	92	2	15	2
xylem 8	10	40	14	25	19	191	38	28	28	75	17	68	9
lateral root cap 5	9	4	32	20	14	59	149	36	21	28	29	10	2
xylem 5	9	10	13	14	12	114	64	15	51	76	11	15	1
xylem 12	8	46	38	30	28	765	41	26	75	46	42	27	9
columella 1	6	8	28	3	29	38	28	18	108	28	19	8	0
procambium 9	0	8	216	290	89	2189	140	142	0	0	412	127	167
procambium 10	0	5	119	294	94	4457	136	79	0	0	269	76	42
procambium 11	0	16	96	51	77	3831	230	75	0	0	183	52	158
procambium 12	0	10	260	282	114	3146	182	102	0	0	239	102	99
procambium 1	0	4	126	204	285	521	189	89	0	0	108	151	6
procambium 2	0	6	48	273	103	302	174	67	0	0	11	58	19
procambium 3	0	5	81	169	83	405	292	43	0	0	50	27	10
procambium 4	0	4	77	160	18	469	208	116	0	0	40	107	36
procambium 5	0	2	87	126	48	467	282	60	0	0	63	57	10
procambium 6	0	7	60	90	56	454	169	117	0	0	123	33	4
procambium 7	0	9	47	118	58	862	121	69	0	0	105	55	20
procambium 8	0	9	97	233	78	784	169	107	0	0	96	260	95
Procambium, young (average of procambium levels in sections 1-6)	0	5	80	171	99	436	219	82	0	0	66	72	14
Procambium, old (average of procambium levels in sections 7-12)	0	10	139	211	85	2545	163	96	0	0	217	112	97

* Genes are ordered in the same way as in Figure 4.1A

Table S4.2. Junction Sepcific Read Counts for Alternative Spliced Isoforms

	MenA			MenG								
	Plastid		PM	Plastid			PM					
Sample_Name	J1	J2	J1	J1	J2	J3	J1	J2	J3	J4	J5	J6
Aerial_15d_R1	233	225	13	166	163	120	2	2	1	3	19	40
Aerial_15d_R2	213	201	29	130	119	96	2	2	0	4	20	33
FirstTrueLeaf_R1	508	490	58	258	267	252	5	8	9	6	54	108
FirstTrueLeaf_R2	293	282	55	273	275	251	2	1	4	6	54	88
FirstTrueLeaf_R3	709	677	91	613	646	538	6	11	10	18	100	160
Leaf_Avr_12h_R1	12	12	2	32	56	16	2	1	0	1	5	14
Leaf_Avr_12h_R2	9	10	3	24	22	35	1	3	1	1	21	19
Leaf_Avr_1h_R1	5	4	0	16	53	26	1	0	0	1	6	17
Leaf_Avr_1h_R2	19	19	2	22	31	96	2	0	0	2	13	12
Leaf_Avr_6h_R1	11	10	0	36	49	20	1	1	1	1	8	15
Leaf_Avr_6h_R2	6	7	1	36	70	30	2	0	1	0	13	22
Leaf_clf28	400	369	62	361	354	287	4	7	4	9	83	100
Leaf_jaz5-10_Ctrl	55	102	9	60	61	80	1	1	1	3	9	18
Leaf_jaz5-10_Pathogen_12h	8	16	1	26	36	49	1	1	0	0	5	5
Leaf_jaz5-10_Pathogen_16h	3	4	1	12	14	22	0	0	0	0	2	3
Leaf_jaz5-10_Pathogen_6h	7	13	1	42	27	43	0	0	1	2	6	18
Leaf_jaz5-10_Pathogen_8h	5	13	0	20	27	47	2	0	0	1	5	12
Leaf_Mock_12h_R1	15	11	2	42	42	18	0	0	0	1	1	13
Leaf_Mock_12h_R2	25	25	11	53	121	47	1	2	3	3	11	42
Leaf_Mock_1h_R1	15	15	1	8	23	15	0	1	1	1	2	12
Leaf_Mock_1h_R2	26	26	2	29	35	77	0	1	2	3	28	36
Leaf_Mock_6h_R1	10	9	0	61	52	49	1	0	0	2	11	29
Leaf_Mock_6h_R2	19	22	2	39	98	66	3	0	1	1	13	28
Leaf_Vir_12h_R1	11	12	1	17	22	9	0	0	0	0	2	2
Leaf_Vir_12h_R2	4	4	3	22	33	23	1	0	0	1	8	8
Leaf_Vir_1h_R1	9	10	4	27	48	31	1	0	0	0	14	8
Leaf_Vir_1h_R2	47	42	1	34	56	76	3	3	0	1	24	33
Leaf_Vir_6h_R1	13	15	3	78	97	59	1	2	2	0	9	27
Leaf_Vir_6h_R2	33	34	6	37	102	61	4	2	0	0	16	30
Leaf_WT	632	584	75	411	419	365	5	2	6	8	100	132
Leaf_WT_Ctrl	83	142	13	68	58	101	0	0	1	2	13	21
Leaf_WT_Pathogen_12h	10	13	2	61	68	83	0	0	1	1	5	18
Leaf_WT_Pathogen_16h	5	15	0	22	23	51	1	0	0	1	4	11
Leaf_WT_Pathogen_6h	24	37	6	63	37	70	0	0	0	0	5	16
Leaf_WT_Pathogen_8h	8	18	0	51	33	43	0	0	1	2	5	9
Leaf1a_HiSeq	545	526	105	536	558	432	0	9	5	20	188	216
Leaf1a_HiSeqMplex	99	99	10	104	115	88	0	0	0	4	46	42
Leaf1b_HiSeqMplex	79	78	21	122	117	63	1	3	1	6	32	41
Leaf2_HiSeq	497	493	59	603	598	526	13	15	13	10	137	194
Leaf2_HiSeqMplex	92	88	18	135	129	114	1	5	5	10	31	31
Leaf3_GAII	56	56	5	53	30	43	0	0	0	2	15	14
Leaf4_GAII	93	90	10	87	58	65	0	0	0	1	20	30

Table S4.2 (continued). Junction Sepcific Read Counts for Alternative Spliced Isoforms

	MenA			MenG								
	Plastid		PM	Plastid			PM					
Sample_Name	J1	J2	J1	J1	J2	J3	J1	J2	J3	J4	J5	J6
Leaf4w_rdm16-2	88	87	9	38	15	5	2	1	1	3	3	4
Leaf4w_sta1	82	77	10	34	18	9	4	2	0	0	3	2
Leaf4w_WT	49	48	4	37	25	7	0	0	0	0	2	4
LeafExplant_EXP2D	209	209	23	411	439	318	2	6	6	10	73	157
LeafExplant_MOT2D	185	177	71	334	363	254	0	3	7	20	113	248
LeafExplant_T0	119	113	18	221	262	173	5	6	4	3	72	170
Seedling_6d_Col0_23_T0_R1	106	105	13	38	41	33	0	1	2	4	11	9
Seedling_6d_Col0_23_T0_R2	79	79	13	26	27	25	0	0	0	2	8	13
Seedling_6d_Col0_23_T0_R3	59	55	3	44	22	4	0	0	0	0	5	8
Seedling_6d_Col0_27_T0_R1	76	72	18	30	39	42	1	0	0	1	10	15
Seedling_6d_Col0_27_T0_R2	67	69	12	25	28	31	0	0	1	0	7	9
Seedling_6d_Col0_27_T0_R3	55	47	4	21	10	4	1	1	0	0	2	3
Seedling_6d_Sij4_23_T0_R1	6	5	12	37	28	15	2	2	0	0	0	0
Seedling_6d_Sij4_23_T0_R2	15	14	42	79	78	75	2	5	1	2	3	2
Seedling_6d_Sij4_27_T0_R1	10	6	33	50	35	20	2	7	2	3	0	0
Seedling_6d_Sij4_27_T0_R2	10	11	58	64	65	49	3	6	4	10	5	1
Seedling_7d_ga1max1_120mMock_R1	34	42	4	17	35	14	0	0	0	2	9	7
Seedling_C24_CK	23	22	8	32	39	38	1	1	1	0	0	0
Seedling_C24_NaCl_150	6	8	0	17	23	10	0	0	0	0	0	1
Seedling_C24_NaCl_300	13	11	2	16	26	35	1	1	0	0	0	2
Seedling_C24_NaCl_50	12	12	2	28	28	42	0	2	2	2	0	0
Seedling2w_atbmi1ab_R1	87	83	21	237	250	139	2	6	6	10	57	110
Seedling2w_atbmi1ab_R2	233	226	49	245	239	173	2	1	2	6	60	148
Seedling2w_atring1ab_R1	225	230	28	281	242	174	1	1	3	14	47	104
Seedling2w_atring1ab_R2	297	282	31	194	196	167	2	4	6	7	50	139
Seedling2w_clf-29_R1	142	154	15	240	197	154	6	4	4	15	27	78
Seedling2w_clf-29_R2	277	260	27	242	268	216	2	3	4	9	54	99
Seedling2w_clf29swn21_R1	60	58	2	92	115	92	2	3	5	11	41	67
Seedling2w_clf29swn21_R2	74	75	13	139	100	68	1	2	1	5	11	53
Seedling2w_lhp1-6_R1	315	296	50	243	221	232	4	2	1	5	43	150
Seedling2w_swn_R1	66	69	17	251	188	147	4	1	2	7	38	73
Seedling2w_swn_R2	51	49	21	187	165	117	4	2	2	6	41	73
Seedling2w_tfl2-2_R1	273	261	38	243	227	219	7	3	4	8	69	156
Seedling2w_wt_R1	105	111	17	253	203	157	2	1	1	9	31	76
Seedling2w_wt_R2	290	278	24	267	305	265	4	2	2	4	58	160
Seedling4d_4hLight	435	434	51	91	74	59	1	1	1	2	13	18
Seedling4d_4hLight_Translatome	60	65	8	22	35	30	0	0	0	0	1	7
Seedling4d_Dark	43	43	8	23	34	26	4	0	0	3	17	13
Seedling4d_Dark_Translatome	7	9	7	3	7	22	0	0	0	0	5	3
Shoot_AmbientCO2_ControlMg_R1	29	32	6	5	9	2	0	0	0	0	0	1
Shoot_AmbientCO2_ControlMg_R2	49	55	6	6	6	4	0	0	0	0	0	1
Shoot_AmbientCO2_HighMg_R1	17	17	0	4	11	2	0	0	0	0	0	0

Table S4.2 (continued). Junction Sepcific Read Counts for Alternative Spliced Isoforms

	MenA			MenG								
	Plastid		PM	Plastid			PM					
Sample_Name	J1	J2	J1	J1	J2	J3	J1	J2	J3	J4	J5	J6
Shoot_AmbientCO2_HighMg_R2	20	21	0	4	10	1	0	0	0	0	0	3
Shoot_AmbientCO2_LowMg_R1	29	33	1	12	12	1	0	0	0	0	0	5
Shoot_AmbientCO2_LowMg_R2	25	25	1	4	3	7	0	0	0	2	2	1
Shoot_BA_R1	64	62	8	51	53	56	1	1	0	0	6	15
Shoot_BA_R2	76	78	9	70	83	68	0	0	0	1	9	18
Shoot_BA_R3	33	31	3	28	30	23	1	2	2	1	11	17
Shoot_Ctrl_R1	63	66	1	54	58	63	2	1	2	2	9	22
Shoot_Ctrl_R2	99	93	8	120	108	73	2	0	1	1	9	16
Shoot_Ctrl_R3	11	11	0	16	28	19	0	0	0	1	7	6
Shoot_ElevatedCO2_ControlMg_R1	37	39	4	7	6	2	0	0	0	0	0	1
Shoot_ElevatedCO2_ControlMg_R2	23	25	2	3	2	1	0	0	0	0	0	0
Shoot_ElevatedCO2_HighMg_R1	27	27	3	7	6	8	2	1	1	1	0	1
Shoot_ElevatedCO2_HighMg_R2	28	31	5	6	1	5	2	2	1	0	0	1
Shoot_ElevatedCO2_LowMg_R1	50	50	3	1	3	4	0	0	0	0	0	0
Shoot_ElevatedCO2_LowMg_R2	41	41	2	5	7	1	0	0	0	0	1	3
ShootApex_LL48	145	134	21	51	74	77	0	1	1	10	14	25
ShootApex_LL52	299	280	31	151	148	115	0	0	3	11	31	35
ShootApex_LL56	275	281	22	110	130	119	4	2	1	4	29	17
ShootApex_LL60	350	346	31	153	169	170	1	3	2	10	34	34
ShootApex_LL64	272	255	26	127	159	127	2	1	1	3	32	30
ShootApex_LL68	198	190	21	111	113	97	0	4	4	2	25	15
ShootApex_LL72	158	144	16	104	109	104	1	1	1	4	27	29
ShootApex_LL76	279	271	24	155	164	122	1	1	0	4	23	39
ShootApex_LL80	197	179	8	104	126	103	2	2	0	4	26	36
ShootApex_LL84	207	202	18	120	138	104	0	1	0	5	38	36
ShootApex_LL88	176	169	22	109	120	86	5	2	1	5	21	27
ShootApex_LL92	204	197	11	124	105	83	3	4	2	2	35	28
Siliques_clf28	154	145	20	188	204	150	4	1	3	8	40	46
Siliques_WT	340	319	41	301	299	250	4	5	4	10	94	76
WholePlant_gemin2_Cold1h_R1	47	49	4	63	47	99	2	3	0	3	19	12
WholePlant_gemin2_Cold1h_R2	56	52	4	55	55	60	0	0	0	13	17	17
WholePlant_gemin2_Cold1h_R3	45	46	2	67	50	91	0	0	0	0	14	11
WholePlant_gemin2_Cold24h_R1	10	11	0	18	21	25	0	0	0	4	3	7
WholePlant_gemin2_Cold24h_R2	7	6	1	36	27	41	0	1	1	2	4	18
WholePlant_gemin2_Cold24h_R3	10	8	0	27	28	49	1	1	0	0	8	11
WholePlant_gemin2_Ctrl_R1	40	41	3	36	26	70	0	0	0	2	11	7
WholePlant_gemin2_Ctrl_R2	55	51	1	50	47	78	1	1	1	1	13	13
WholePlant_gemin2_Ctrl_R3	39	36	2	44	42	63	0	2	0	2	5	6
WholePlant_WT_Cold1h_R1	24	22	2	31	23	34	0	0	0	0	13	4
WholePlant_WT_Cold1h_R2	31	32	3	33	31	46	0	0	0	1	7	19
WholePlant_WT_Cold1h_R3	41	41	8	50	42	72	0	0	0	1	11	6
WholePlant_WT_Cold24h_R1	9	7	0	13	13	21	0	0	0	0	4	6

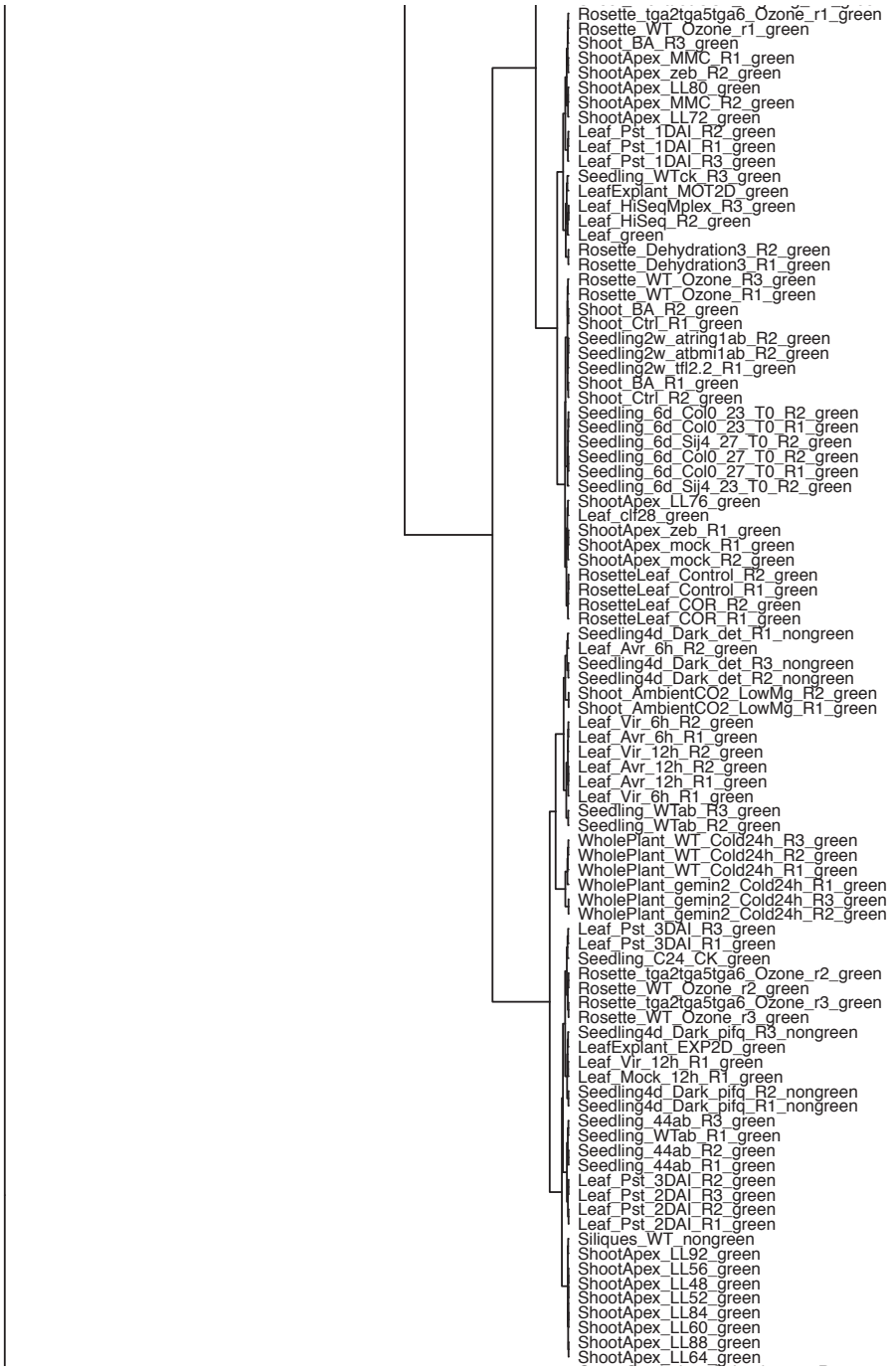
Table S4.2 (continued). Junction Sepcific Read Counts for Alternative Spliced Isoforms

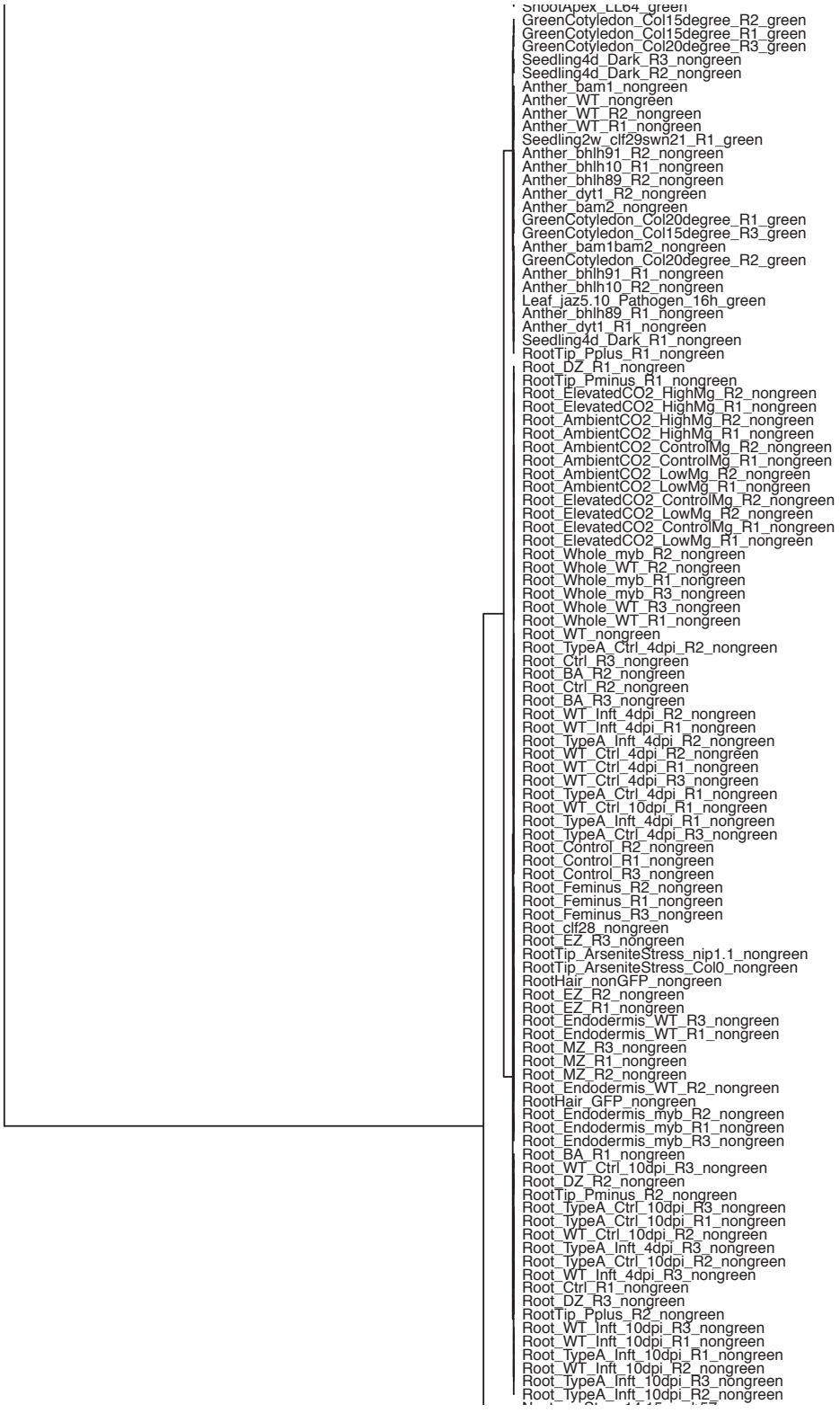
	MenA			MenG								
	Plastid		PM	Plastid			PM					
Sample_Name	J1	J2	J1	J1	J2	J3	J1	J2	J3	J4	J5	J6
WholePlant_WT_Cold24h_R2	16	17	0	55	47	54	0	0	0	0	17	16
WholePlant_WT_Cold24h_R3	12	14	0	35	18	29	0	0	0	0	11	14
WholePlant_WT_Ctrl_R1	19	18	1	20	17	24	0	0	0	0	0	8
WholePlant_WT_Ctrl_R2	23	19	2	35	31	35	0	0	0	0	0	3
WholePlant_WT_Ctrl_R3	24	23	4	42	45	36	0	0	0	0	4	10
Anther_bam1	14	12	3	51	43	26	3	4	2	5	11	20
Anther_bam1bam2	29	28	8	63	42	52	3	6	6	9	22	35
Anther_bam2	21	19	5	69	56	38	2	3	5	8	16	28
Anther_bhlh10_R1	29	29	7	69	67	51	3	4	6	15	27	26
Anther_bhlh10_R2	7	6	2	34	20	18	0	1	1	1	15	13
Anther_bhlh89_R1	3	2	0	0	1	0	0	0	0	0	0	0
Anther_bhlh89_R2	15	14	3	30	26	29	0	1	2	2	8	16
Anther_bhlh91_R1	0	0	0	1	1	2	0	0	0	1	1	1
Anther_bhlh91_R2	17	17	3	27	31	18	1	2	1	2	6	10
Anther_dyt1_R1	8	8	1	33	18	8	2	1	0	5	11	13
Anther_dyt1_R2	10	9	0	15	19	18	1	1	4	5	6	13
Anther_WT	14	13	5	64	53	26	0	0	1	4	13	27
Anther_WT_R1	15	14	3	28	21	12	1	0	2	2	5	12
Anther_WT_R2	20	19	0	26	22	14	2	2	0	1	10	16
Flower_clf28	371	354	58	286	311	225	3	11	16	13	94	118
Flower_WT	409	390	73	249	281	232	5	3	2	9	84	88
FlowerBud1_HiSeq	942	931	82	805	868	657	15	15	6	19	231	319
FlowerBud1_HiSeqMplex	221	221	25	170	184	139	3	1	0	8	41	53
FlowerBud2_HiSeq	771	740	137	565	595	429	0	2	3	11	109	141
FlowerBud2_HiSeqMplex	169	165	25	111	114	85	1	2	1	6	34	37
GreenCotyledon_Col15degree_R1	5	5	3	71	70	57	2	2	1	6	20	32
GreenCotyledon_Col15degree_R2	9	10	2	79	85	89	3	7	1	1	24	51
GreenCotyledon_Col15degree_R3	7	7	6	66	56	56	0	0	0	5	17	37
GreenCotyledon_Col20degree_R1	5	4	1	28	21	19	0	0	0	0	8	23
GreenCotyledon_Col20degree_R2	4	4	2	25	16	16	2	2	0	2	6	9
GreenCotyledon_Col20degree_R3	5	6	0	23	22	15	1	2	3	4	10	6
Root_AmbientCO2_ControlMg_R1	1	1	0	0	1	3	0	0	0	0	1	5
Root_AmbientCO2_ControlMg_R2	1	1	0	0	0	0	0	0	0	0	1	4
Root_AmbientCO2_HighMg_R1	0	0	0	1	0	0	0	0	0	0	1	3
Root_AmbientCO2_HighMg_R2	0	0	0	3	3	0	0	0	0	1	3	3
Root_AmbientCO2_LowMg_R1	0	0	0	1	2	0	0	0	0	0	0	0
Root_AmbientCO2_LowMg_R2	0	0	0	0	0	0	0	0	0	0	0	1
Root_BA_R1	5	4	0	8	6	7	0	0	0	0	3	1
Root_BA_R2	4	4	0	11	10	8	0	1	0	0	1	7
Root_BA_R3	5	5	0	7	6	8	0	0	0	0	8	6
Root_clf28	10	7	7	107	126	41	1	3	0	3	54	58
Root_Ctrl_R1	3	2	1	17	19	7	0	1	0	0	5	3

Table S4.2 (continued). Junction Sepcific Read Counts for Alternative Spliced Isoforms

	MenA			MenG								
	Plastid		PM	Plastid			PM					
Sample_Name	J1	J2	J1	J1	J2	J3	J1	J2	J3	J4	J5	J6
Root_Ctrl_R2	1	1	2	12	10	4	0	0	1	0	2	5
Root_Ctrl_R3	0	0	1	7	8	4	0	0	0	0	1	3
Root_ElevatedCO2_ControlMg_R1	0	0	0	2	1	1	0	0	0	0	0	1
Root_ElevatedCO2_ControlMg_R2	0	0	0	1	2	2	0	0	0	0	0	4
Root_ElevatedCO2_HighMg_R1	0	0	0	1	1	0	0	0	0	0	0	0
Root_ElevatedCO2_HighMg_R2	1	1	0	5	2	2	0	0	0	0	0	1
Root_ElevatedCO2_LowMg_R1	0	0	0	3	6	1	0	0	0	0	1	2
Root_ElevatedCO2_LowMg_R2	1	1	0	3	0	0	0	0	0	0	0	0
Root_WT	19	16	3	111	107	72	2	3	4	0	49	48
RootHair_GFP	0	0	0	13	11	11	0	0	3	3	4	4
RootHair_nonGFP	14	13	1	41	31	31	4	10	7	4	15	18
RootTip_ArseniteStress_Col0	2	0	0	2	4	17	0	0	0	0	0	1
RootTip_ArseniteStress_nip1.1	0	0	0	8	20	12	1	0	0	0	4	6


WholePlant_WT_Ctrl_R3_green
WholePlant_WT_Ctrl_R2_green
WholePlant_gemin2_Ctrl_R3_green
WholePlant_gemin2_Ctrl_R1_green
WholePlant_gemin2_Cold1h_R3_green
WholePlant_gemin2_Cold1h_R2_green
WholePlant_gemin2_Ctrl_R2_green
WholePlant_WT_Cold1h_R2_green
WholePlant_WT_Cold1h_R1_green
WholePlant_WT_Ctrl_R1_green
WholePlant_gemin2_Cold1h_R1_green
WholePlant_WT_Cold1h_R3_green
Shoot_AmbientCO2_ControlMg_R2_green
Shoot_AmbientCO2_ControlMg_R1_green
Shoot_ElevatedCO2_ControlMg_R2_green
Shoot_ElevatedCO2_ControlMg_R1_green
Shoot_AmbientCO2_HighMg_R2_green
Shoot_AmbientCO2_HighMg_R1_green
Leaf_Avr_1h_R1_green
Leaf_Mock_1h_R1_green
Aerial_15d_R2_green
Aerial_15d_R1_green
Leaf_HiSeqMplex_R1_green
Leaf_HiSeq_R1_green
Leaf_HiSeqMplex_R2_green
GuardCell_CntrlBR4_green
Leaf_Vir_1h_R2_green
Shoot_Ctrl_R3_green
Leaf_WT_green
Leaf_H6R48_R2_green
Leaf_H6R48_R1_green
Rosette_R2_green
Rosette_WT_Ctrl_r2_green
Rosette_WT_Ozone_R2_green
Rosette_coi1ein2sid2_Ozone_R3_green
Rosette_coi1ein2sid2_Ozone_R1_green
Seedling2w_lhp1.6_R1_green
Seedling2w_clf.29_R2_green
Seedling2w_wt_R2_green
Seedling2w_atring1ab_R1_green
Leaf_jaz5.10_Ctrl_green
Leaf_WT_Ctrl_green
Rosette_WT_Ctrl_R3_green
Rosette_WT_Ctrl_R2_green
Rosette_coi1ein2sid2_Ozone_R2_green
Rosette_coi1ein2sid2_Ctrl_R2_green
Rosette_tga2tga5tga6_Ctrl_r1_green
Rosette_WT_Ctrl_r1_green
Rosette_tga2tga5tga6_Ctrl_r2_green
Rosette_WT_Ctrl_r3_green
Rosette_tga2tga5tga6_Ctrl_r3_green
YoungLeaf_green
Rosette_R1_green
Rosette_Water_R2_green
Rosette_Water_R1_green
FirstTrueLeaf_R3_green
FirstTrueLeaf_R2_green
LeafExplant_T0_green
FirstTrueLeaf_R1_green
Leaf_Mock_R2_green
Rosette_coi1ein2sid2_Ctrl_R3_green
Rosette_coi1ein2sid2_Ctrl_R1_green
Rosette_WT_Ctrl_R1_green
Leaf_Mock_R1_green
Seedling_6d_Sij4_27_T0_R1_green
Seedling_6d_Col0_27_T0_R3_green
Seedling_6d_Col0_23_T0_R3_green
Leaf4w_WT_green
Seedling_6d_Sij4_23_T0_R1_green
Leaf4w_sta1_green
Leaf4w_rdm16.2_green
Rosette_Dehydration1_R2_green
Rosette_Dehydration1_R1_green
GuardCell_CntrolBR5_green
Seedling2w_atbmi1ab_R1_green
Seedling2w_wt_R1_green
Seedling2w_swn_R2_green
Seedling2w_swn_R1_green
Seedling2w_clf.29_R1_green
Seedling_44ck_R3_green
Seedling_WTck_R2_green
Seedling_44ck_R2_green
Seedling_44ck_R1_green
Seedling_WTck_R1_green
Leaf_Mock_6h_R2_green
Leaf_Vir_1h_R1_green
Leaf_Mock_12h_R2_green
Leaf_Mock_6h_R1_green
Leaf_Avr_1h_R2_green
Leaf_Mock_1h_R2_green
Leaf_H6_R2_green
Leaf_H6_R1_green
Shoot_ElevatedCO2_LowMg_R2_green
Shoot_ElevatedCO2_LowMg_R1_green
Shoot_ElevatedCO2_HighMg_R2_green
Shoot_ElevatedCO2_HighMg_R1_green


Rosette_tga2tga5tga6_Ozone_r1_green
Rosette_WT_Ozone_r1_green
Shoot_BA_R3_green
ShootApex_MMC_R1_green
ShootApex_zeb_R2_green
ShootApex_LL80_green
ShootApex_MMC_R2_green
ShootApex_LL72_green
Leaf_Pst_1DAI_R2_green
Leaf_Pst_1DAI_R1_green
Leaf_Pst_1DAI_R3_green
Seedling_WTck_R3_green
LeafExplant_MOT2D_green
Leaf_HiSeqMplex_R3_green
Leaf_HiSeq_R2_green
Leaf_green
Rosette_Dehydration3_R2_green
Rosette_Dehydration3_R1_green
Rosette_WT_Ozone_R3_green
Rosette_WT_Ozone_R1_green
Shoot_BA_R2_green
Shoot_Ctrl_R1_green
Seedling2w_atring1ab_R2_green
Seedling2w_atbmi1ab_R2_green
Seedling2w_tfl2.2_R1_green
Shoot_BA_R1_green
Shoot_Ctrl_R2_green
Seedling_6d_Col0_23_T0_R2_green
Seedling_6d_Col0_23_T0_R1_green
Seedling_6d_Sij4_27_T0_R2_green
Seedling_6d_Col0_27_T0_R2_green
Seedling_6d_Col0_27_T0_R1_green
Seedling_6d_Sij4_23_T0_R2_green
ShootApex_LL76_green
Leaf_clf28_green
ShootApex_zeb_R1_green
ShootApex_mock_R1_green
ShootApex_mock_R2_green
RosetteLeaf_Control_R2_green
RosetteLeaf_Control_R1_green
RosetteLeaf_COR_R2_green
RosetteLeaf_COR_R1_green
Seedling4d_Dark_det_R1_nongreen
Leaf_Avr_6h_R2_green
Seedling4d_Dark_det_R3_nongreen
Seedling4d_Dark_det_R2_nongreen
Shoot_AmbientCO2_LowMg_R2_green
Shoot_AmbientCO2_LowMg_R1_green
Leaf_Vir_6h_R2_green
Leaf_Avr_6h_R1_green
Leaf_Vir_12h_R2_green
Leaf_Avr_12h_R2_green
Leaf_Avr_12h_R1_green
Leaf_Vir_6h_R1_green
Seedling_WTab_R3_green
Seedling_WTab_R2_green
WholePlant_WT_Cold24h_R3_green
WholePlant_WT_Cold24h_R2_green
WholePlant_WT_Cold24h_R1_green
WholePlant_gemin2_Cold24h_R1_green
WholePlant_gemin2_Cold24h_R3_green
WholePlant_gemin2_Cold24h_R2_green
Leaf_Pst_3DAI_R3_green
Leaf_Pst_3DAI_R1_green
Seedling_C24_CK_green
Rosette_tga2tga5tga6_Ozone_r2_green
Rosette_WT_Ozone_r2_green
Rosette_tga2tga5tga6_Ozone_r3_green
Rosette_WT_Ozone_r3_green
Seedling4d_Dark_pifq_R3_nongreen
LeafExplant_EXP2D_green
Leaf_Vir_12h_R1_green
Leaf_Mock_12h_R1_green
Seedling4d_Dark_pifq_R2_nongreen
Seedling4d_Dark_pifq_R1_nongreen
Seedling_44ab_R3_green
Seedling_WTab_R1_green
Seedling_44ab_R2_green
Seedling_44ab_R1_green
Leaf_Pst_3DAI_R2_green
Leaf_Pst_2DAI_R3_green
Leaf_Pst_2DAI_R2_green
Leaf_Pst_2DAI_R1_green
Siliques_WT_nongreen
ShootApex_LL92_green
ShootApex_LL56_green
ShootApex_LL48_green
ShootApex_LL52_green
ShootApex_LL84_green
ShootApex_LL60_green
ShootApex_LL88_green
ShootApex_LL64_green


ShootApex_LL64_green
GreenCotyledon_Col15degree_R2_green
GreenCotyledon_Col15degree_R1_green
GreenCotyledon_Col20degree_R3_green
Seedling4d_Dark_R3_nongreen
Seedling4d_Dark_R2_nongreen
Anther_bam1_nongreen
Anther_WT_nongreen
Anther_WT_R2_nongreen
Anther_WT_R1_nongreen
Seedling2w_clf29swn21_R1_green
Anther_bhlh91_R2_nongreen
Anther_bhlh10_R1_nongreen
Anther_bhlh89_R2_nongreen
Anther_dyt1_R2_nongreen
Anther_bam2_nongreen
GreenCotyledon_Col20degree_R1_green
GreenCotyledon_Col15degree_R3_green
Anther_bam1bam2_nongreen
GreenCotyledon_Col20degree_R2_green
Anther_bhlh91_R1_nongreen
Anther_bhlh10_R2_nongreen
Leaf_jaz5.10_Pathogen_16h_green
Anther_bhlh89_R1_nongreen
Anther_dyt1_R1_nongreen
Seedling4d_Dark_R1_nongreen
RootTip_Pplus_R1_nongreen
Root_DZ_R1_nongreen
RootTip_Pminus_R1_nongreen
Root_ElevatedCO2_HighMg_R2_nongreen
Root_ElevatedCO2_HighMg_R1_nongreen
Root_AmbientCO2_HighMg_R2_nongreen
Root_AmbientCO2_HighMg_R1_nongreen
Root_AmbientCO2_ControlMg_R2_nongreen
Root_AmbientCO2_ControlMg_R1_nongreen
Root_AmbientCO2_LowMg_R2_nongreen
Root_AmbientCO2_LowMg_R1_nongreen
Root_ElevatedCO2_ControlMg_R2_nongreen
Root_ElevatedCO2_LowMg_R2_nongreen
Root_ElevatedCO2_ControlMg_R1_nongreen
Root_ElevatedCO2_LowMg_R1_nongreen
Root_Whole_myb_R2_nongreen
Root_Whole_WT_R2_nongreen
Root_Whole_myb_R1_nongreen
Root_Whole_myb_R3_nongreen
Root_Whole_WT_R3_nongreen
Root_Whole_WT_R1_nongreen
Root_WT_nongreen
Root_TypeA_Ctrl_4dpi_R2_nongreen
Root_Ctrl_R3_nongreen
Root_BA_R2_nongreen
Root_Ctrl_R2_nongreen
Root_BA_R3_nongreen
Root_WT_Inft_4dpi_R2_nongreen
Root_WT_Inft_4dpi_R1_nongreen
Root_TypeA_Inft_4dpi_R2_nongreen
Root_WT_Ctrl_4dpi_R2_nongreen
Root_WT_Ctrl_4dpi_R1_nongreen
Root_WT_Ctrl_4dpi_R3_nongreen
Root_TypeA_Ctrl_4dpi_R1_nongreen
Root_WT_Ctrl_10dpi_R1_nongreen
Root_TypeA_Inft_4dpi_R1_nongreen
Root_TypeA_Ctrl_4dpi_R3_nongreen
Root_Control_R2_nongreen
Root_Control_R1_nongreen
Root_Control_R3_nongreen
Root_Feminus_R2_nongreen
Root_Feminus_R1_nongreen
Root_Feminus_R3_nongreen
Root_clf28_nongreen
Root_EZ_R3_nongreen
RootTip_ArseniteStress_nip1.1_nongreen
RootTip_ArseniteStress_Col0_nongreen
RootHair_nonGFP_nongreen
Root_EZ_R2_nongreen
Root_EZ_R1_nongreen
Root_Endodermis_WT_R3_nongreen
Root_Endodermis_WT_R1_nongreen
Root_MZ_R3_nongreen
Root_MZ_R1_nongreen
Root_MZ_R2_nongreen
Root_Endodermis_WT_R2_nongreen
RootHair_GFP_nongreen
Root_Endodermis_myb_R2_nongreen
Root_Endodermis_myb_R1_nongreen
Root_Endodermis_myb_R3_nongreen
Root_BA_R1_nongreen
Root_WT_Ctrl_10dpi_R3_nongreen
Root_DZ_R2_nongreen
RootTip_Pminus_R2_nongreen
Root_TypeA_Ctrl_10dpi_R3_nongreen
Root_TypeA_Ctrl_10dpi_R1_nongreen
Root_WT_Ctrl_10dpi_R2_nongreen
Root_TypeA_Inft_4dpi_R3_nongreen
Root_TypeA_Ctrl_10dpi_R2_nongreen
Root_WT_Inft_4dpi_R3_nongreen
Root_Ctrl_R1_nongreen
Root_DZ_R3_nongreen
RootTip_Pplus_R2_nongreen
Root_WT_Inft_10dpi_R3_nongreen
Root_WT_Inft_10dpi_R1_nongreen
Root_TypeA_Inft_10dpi_R1_nongreen
Root_WT_Inft_10dpi_R2_nongreen
Root_TypeA_Inft_10dpi_R3_nongreen
Root_TypeA_Inft_10dpi_R2_nongreen

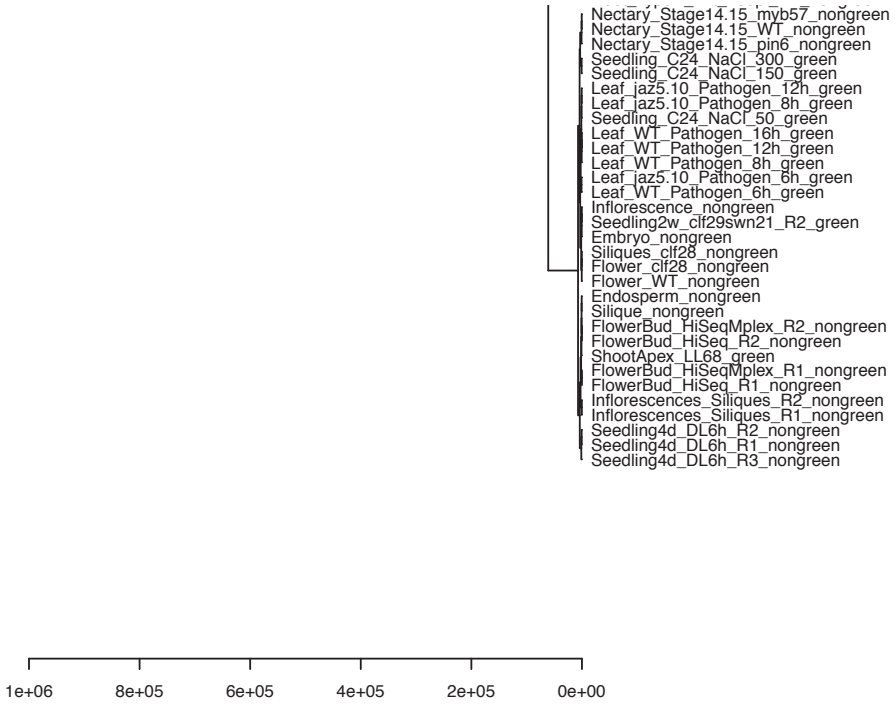
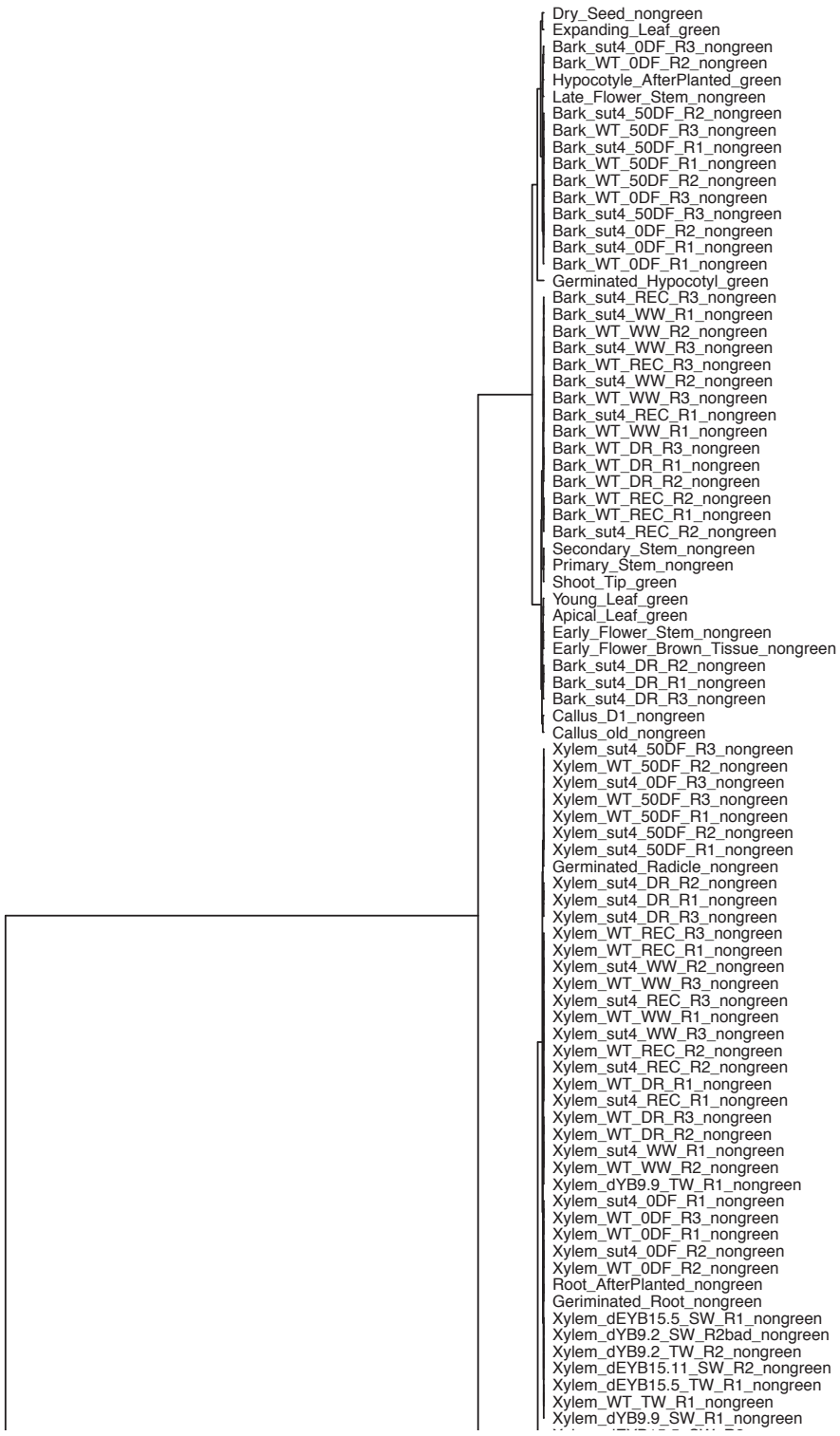


Figure S4.1. Tissue clustering in *Arabidopsis* based on the expression of PSI and PSII genes.


Dry_Seed_nongreen
Expanding_Leaf_green
Bark_sut4_0DF_R3_nongreen
Bark_WT_0DF_R2_nongreen
Hypocotyle_AfterPlanted_green
Late_Flower_Stem_nongreen
Bark_sut4_50DF_R2_nongreen
Bark_WT_50DF_R3_nongreen
Bark_sut4_50DF_R1_nongreen
Bark_WT_50DF_R1_nongreen
Bark_WT_50DF_R2_nongreen
Bark_WT_0DF_R3_nongreen
Bark_sut4_50DF_R3_nongreen
Bark_sut4_0DF_R2_nongreen
Bark_sut4_0DF_R1_nongreen
Bark_WT_0DF_R1_nongreen
Germinated_Hypocotyl_green
Bark_sut4_REC_R3_nongreen
Bark_sut4_WW_R1_nongreen
Bark_WT_WW_R2_nongreen
Bark_sut4_WW_R3_nongreen
Bark_WT_REC_R3_nongreen
Bark_sut4_WW_R2_nongreen
Bark_WT_WW_R3_nongreen
Bark_sut4_REC_R1_nongreen
Bark_WT_WW_R1_nongreen
Bark_WT_DR_R3_nongreen
Bark_WT_DR_R1_nongreen
Bark_WT_DR_R2_nongreen
Bark_WT_REC_R2_nongreen
Bark_WT_REC_R1_nongreen
Bark_sut4_REC_R2_nongreen
Secondary_Stem_nongreen
Primary_Stem_nongreen
Shoot_Tip_green
Young_Leaf_green
Apical_Leaf_green
Early_Flower_Stem_nongreen
Early_Flower_Brown_Tissue_nongreen
Bark_sut4_DR_R2_nongreen
Bark_sut4_DR_R1_nongreen
Bark_sut4_DR_R3_nongreen
Callus_D1_nongreen
Callus_old_nongreen
Xylem_sut4_50DF_R3_nongreen
Xylem_WT_50DF_R2_nongreen
Xylem_sut4_0DF_R3_nongreen
Xylem_WT_50DF_R3_nongreen
Xylem_WT_50DF_R1_nongreen
Xylem_sut4_50DF_R2_nongreen
Xylem_sut4_50DF_R1_nongreen
Germinated_Radicle_nongreen
Xylem_sut4_DR_R2_nongreen
Xylem_sut4_DR_R1_nongreen
Xylem_sut4_DR_R3_nongreen
Xylem_WT_REC_R3_nongreen
Xylem_WT_REC_R1_nongreen
Xylem_sut4_WW_R2_nongreen
Xylem_WT_WW_R3_nongreen
Xylem_sut4_REC_R3_nongreen
Xylem_WT_WW_R1_nongreen
Xylem_sut4_WW_R3_nongreen
Xylem_WT_REC_R2_nongreen
Xylem_sut4_REC_R2_nongreen
Xylem_WT_DR_R1_nongreen
Xylem_sut4_REC_R1_nongreen
Xylem_WT_DR_R3_nongreen
Xylem_WT_DR_R2_nongreen
Xylem_sut4_WW_R1_nongreen
Xylem_WT_WW_R2_nongreen
Xylem_dYB9.9_TW_R1_nongreen
Xylem_sut4_0DF_R1_nongreen
Xylem_WT_0DF_R3_nongreen
Xylem_WT_0DF_R1_nongreen
Xylem_sut4_0DF_R2_nongreen
Xylem_WT_0DF_R2_nongreen
Root_AfterPlanted_nongreen
Geriminated_Root_nongreen
Xylem_dEYB15.5_SW_R1_nongreen
Xylem_dYB9.2_SW_R2bad_nongreen
Xylem_dYB9.2_TW_R2_nongreen
Xylem_dEYB15.11_SW_R2_nongreen
Xylem_dEYB15.5_TW_R1_nongreen
Xylem_WT_TW_R1_nongreen
Xylem_dYB9.9_SW_R1_nongreen

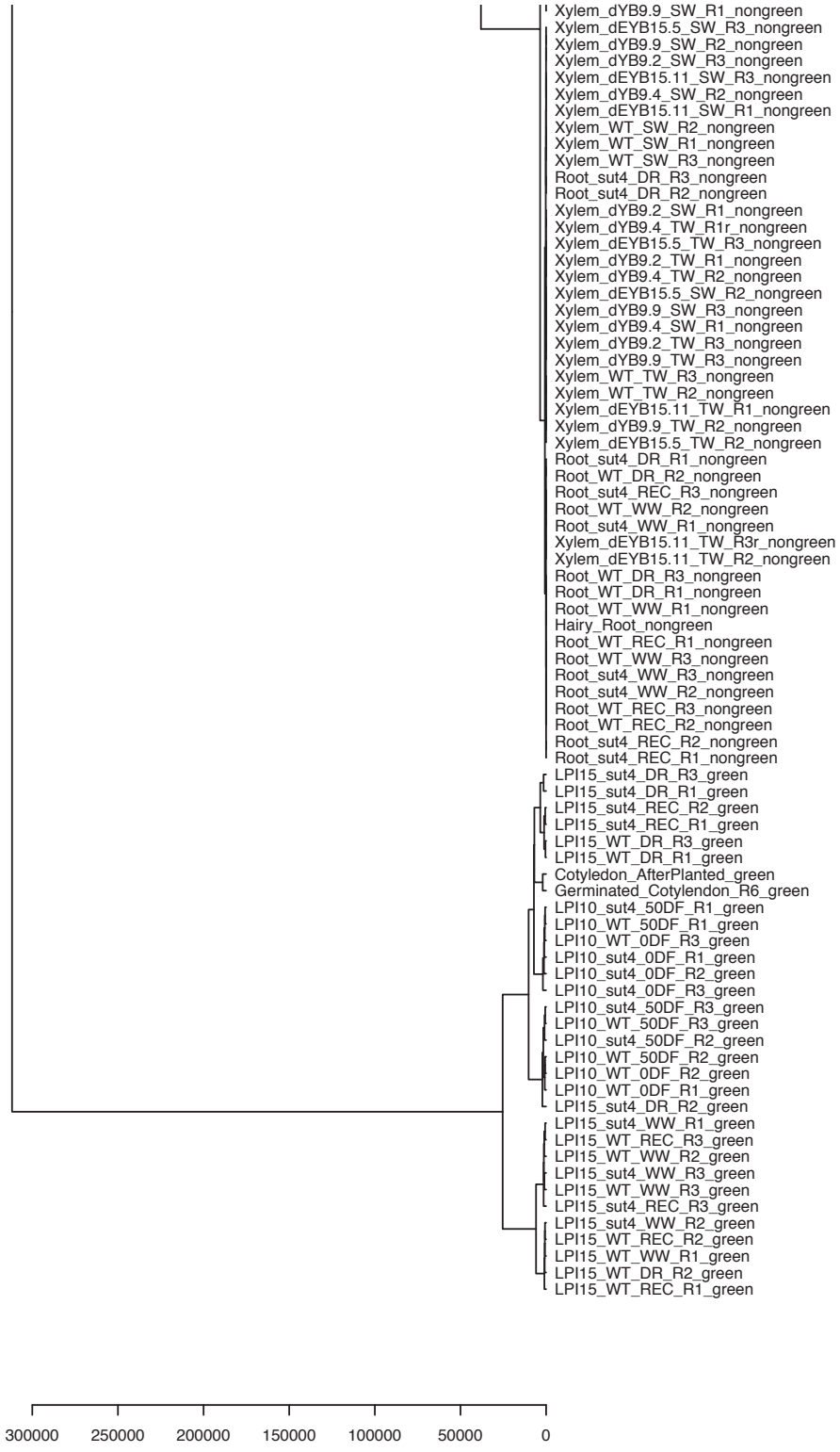
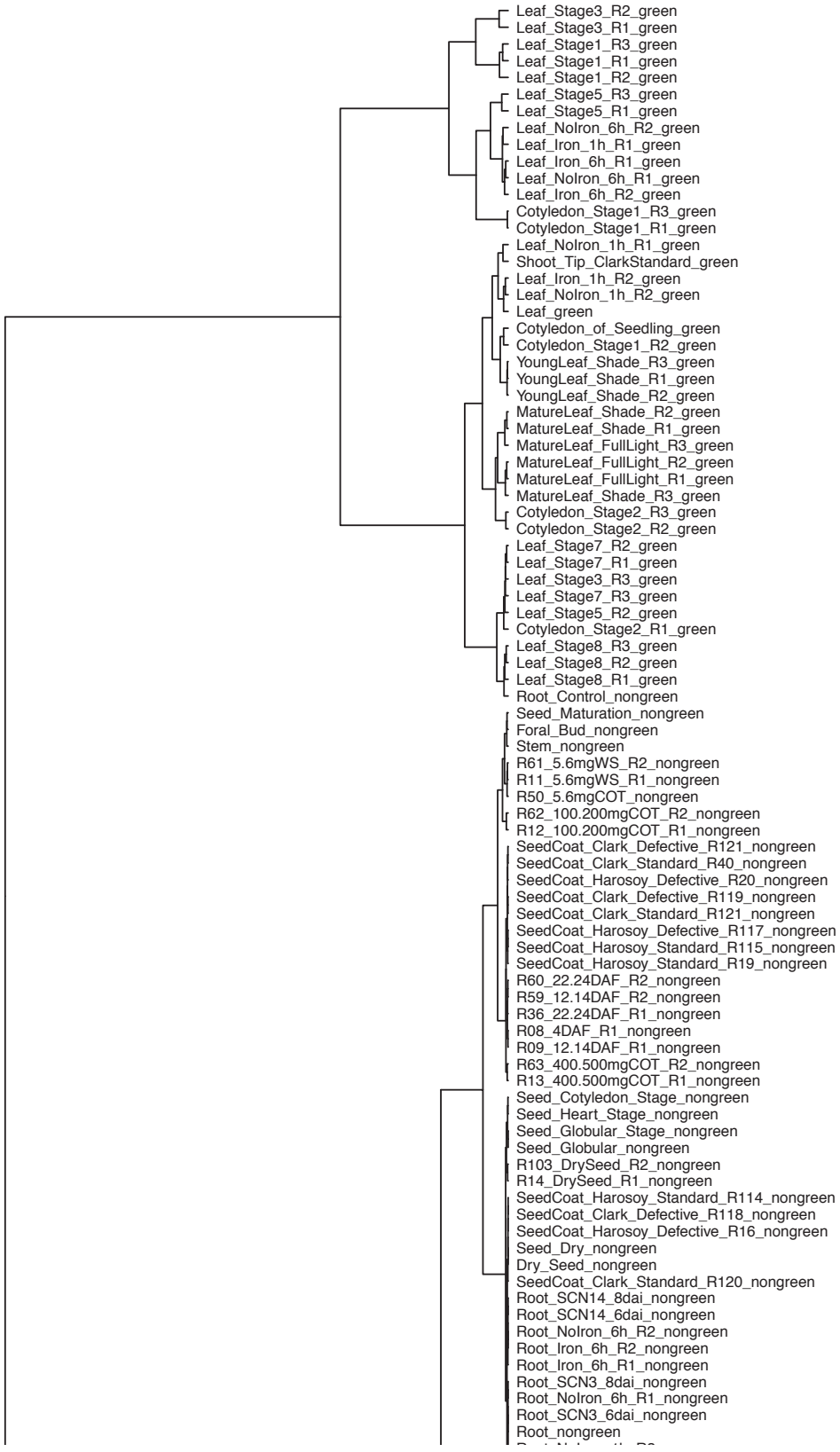


Figure S4.2. Tissue clustering in *Populus tremula x alba* based on the expression of PSI and PSII genes.


Leaf_Stage3_R2_green
Leaf_Stage3_R1_green
Leaf_Stage1_R3_green
Leaf_Stage1_R1_green
Leaf_Stage1_R2_green
Leaf_Stage5_R3_green
Leaf_Stage5_R1_green
Leaf_NoIron_6h_R2_green
Leaf_Iron_1h_R1_green
Leaf_Iron_6h_R1_green
Leaf_NoIron_6h_R1_green
Leaf_Iron_6h_R2_green
Cotyledon_Stage1_R3_green
Cotyledon_Stage1_R1_green
Leaf_NoIron_1h_R1_green
Shoot_Tip_ClarkStandard_green
Leaf_Iron_1h_R2_green
Leaf_NoIron_1h_R2_green
Leaf_green
Cotyledon_of_Seedling_green
Cotyledon_Stage1_R2_green
YoungLeaf_Shade_R3_green
YoungLeaf_Shade_R1_green
YoungLeaf_Shade_R2_green
MatureLeaf_Shade_R2_green
MatureLeaf_Shade_R1_green
MatureLeaf_FullLight_R3_green
MatureLeaf_FullLight_R2_green
MatureLeaf_FullLight_R1_green
MatureLeaf_Shade_R3_green
Cotyledon_Stage2_R3_green
Cotyledon_Stage2_R2_green
Leaf_Stage7_R2_green
Leaf_Stage7_R1_green
Leaf_Stage3_R3_green
Leaf_Stage7_R3_green
Leaf_Stage5_R2_green
Cotyledon_Stage2_R1_green
Leaf_Stage8_R3_green
Leaf_Stage8_R2_green
Leaf_Stage8_R1_green
Root_Control_nongreen
Seed_Maturation_nongreen
Foral_Bud_nongreen
Stem_nongreen
R61_5.6mgWS_R2_nongreen
R11_5.6mgWS_R1_nongreen
R50_5.6mgCOT_nongreen
R62_100.200mgCOT_R2_nongreen
R12_100.200mgCOT_R1_nongreen
SeedCoat_Clark_Defective_R121_nongreen
SeedCoat_Clark_Standard_R40_nongreen
SeedCoat_Harosoy_Defective_R20_nongreen
SeedCoat_Clark_Defective_R119_nongreen
SeedCoat_Clark_Standard_R121_nongreen
SeedCoat_Harosoy_Defective_R117_nongreen
SeedCoat_Harosoy_Standard_R115_nongreen
SeedCoat_Harosoy_Standard_R19_nongreen
R60_22.24DAF_R2_nongreen
R59_12.14DAF_R2_nongreen
R36_22.24DAF_R1_nongreen
R08_4DAF_R1_nongreen
R09_12.14DAF_R1_nongreen
R63_400.500mgCOT_R2_nongreen
R13_400.500mgCOT_R1_nongreen
Seed_Cotyledon_Stage_nongreen
Seed_Heart_Stage_nongreen
Seed_Globular_Stage_nongreen
Seed_Globular_nongreen
R103_DrySeed_R2_nongreen
R14_DrySeed_R1_nongreen
SeedCoat_Harosoy_Standard_R114_nongreen
SeedCoat_Clark_Defective_R118_nongreen
SeedCoat_Harosoy_Defective_R16_nongreen
Seed_Dry_nongreen
Dry_Seed_nongreen
SeedCoat_Clark_Standard_R120_nongreen
Root_SCN14_8dai_nongreen
Root_SCN14_6dai_nongreen
Root_NoIron_6h_R2_nongreen
Root_Iron_6h_R2_nongreen
Root_Iron_6h_R1_nongreen
Root_SCN3_8dai_nongreen
Root_NoIron_6h_R1_nongreen
Root_SCN3_6dai_nongreen
Root_nongreen

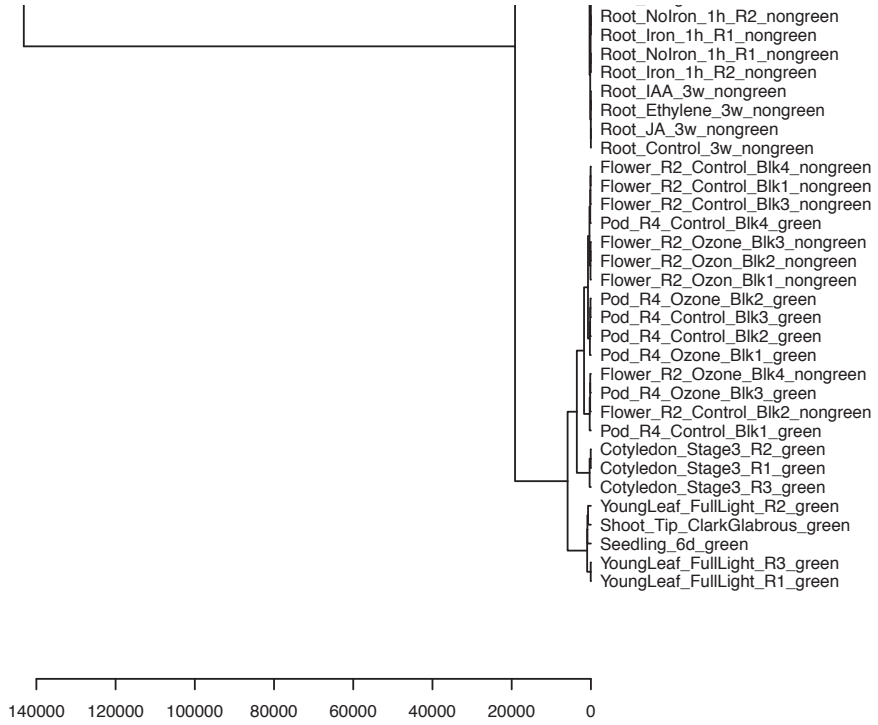


Figure S4.3. Tissue clustering in *Glycine max* based on the expression of PSI and PSII genes.

References

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

Barr R, Pan RS, Crane FL, Brightman AO, Morré DJ (1992) Destruction of vitamin K_1 of cultured carrot cells by ultraviolet radiation and its effect on plasma membrane electron transport reactions. Biochem Int **27**: 449–56

Basset GJ (2016) Phylloquinone (vitamin K1): occurrence, biosynthesis and functions. Mini-Reviews Med Chem **16**: 1–11

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**: 2114–2120

Booth SL, Suttie JW (1998) Dietary intake and adequacy of vitamin K1. J Nutr **128**: 785–788

Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN (2007) A high-resolution root spatiotemporal map reveals dominant expression patterns. Science (80-) **318**: 801–806

Brettel K, Setif P, Mathis P (1986) Flash-induced absorption changes in photosystem I at low temperature: Evidence that the electron acceptor A_1 is vitamin K_1. FEBS Lett **203**: 220–224

Bridge A, Barr R, Morré DJ (2000) The plasma membrane NADH oxidase of soybean has vitamin K_1 hydroquinone oxidase activity. Biochim Biophys Acta - Biomembr **1463**: 448–458

Cartwright DA, Brady SM, Orlando DA, Sturmfels B, Benfey PN (2009) Reconstructing spatiotemporal gene expression data from partial observations. Bioinformatics **25**: 2581–2587

Cassin-Ross G, Hu J (2014) Systematic phenotypic screen of Arabidopsis peroxisomal mutants identifies proteins involved in β-oxidation. Plant Physiol **166**: 1546–59

Dahm C, Müller R, Schulte G, Schmidt K, Leistner E (1998) The role of isochorismate hydroxymutase genes *entC* and *menF* in enterobactin and menaquinone biosynthesis in

Escherichia coli. Biochim Biophys Acta - Gen Subj **1425**: 377–386

Daruwala R, Bhattacharyya DK, Kwon O, Meganathan R (1997) Menaquinone (vitamin K_2) biosynthesis: overexpression, purification, and characterization of a new isochorismate synthase from *Escherichia coli*. J Bacteriol **179**: 3133–8

Daruwala R, Kwon O, Meganathan R, Hudspeth MES (1996) A new isochorismate synthase specifically involved in menaquinone (vitamin K2) biosynthesis encoded by the *menF* gene. FEMS Microbiol Lett **140**: 159–163

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

Eugeni Piller L (2014) Role of plastoglobules in metabolite repair in the tocopherol redox cycle. Front Plant Sci **5**: 1–10

Eugeni Piller L, Besagni C, Ksas B, Rumeau D, Bréhélin C, Glauser G, Kessler F, Havaux M (2011) Chloroplast lipid droplet type II NAD(P)H quinone oxidoreductase is essential for prenylquinone metabolism and vitamin K_1 accumulation. Proc Natl Acad Sci U S A **108**: 14354–9

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**: 1531–1545

Frigaard N-U, Takaichi S, Hirota M, Shimada K, Matsuura K (1997) Quinones in chlorosomes of green sulfur bacteria and their role in the redox-dependent fluorescence studied in chlorosome-like bacteriochlorophyll c aggregates. Arch Microbiol **167**: 343–349

Garcion C, Lohmann A, Lamodière E, Catinot J, Buchala A, Doermann P, Métraux J-P (2008) Characterization and biological function of the *ISOCHORISMATE SYNTHASE2* gene of Arabidopsis. Plant Physiol **147**: 1279–1287

Georgellis D, Kwon O, Lin EC, Parkinson JS, Kofoid EC, Iuchi S, Lin ECC, Kwon O, Georgellis D, Lynch AS, et al (2001) Quinones as the redox signal for the arc two-component system of bacteria. Science **292**: 2314–6

Gross J, Cho WK, Lezhneva L, Falk J, Krupinska K, Shinozaki K, Seki M, Herrmann RG, Meurer J (2006) A plant locus essential for phylloquinone (vitamin K_1) biosynthesis originated from a fusion of four eubacterial genes. J Biol Chem **281**: 17189–96

Hale MB, Blankenship RE, Fuller RC (1983) Menaquinone is the sole quinone in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. BBA - Bioenerg **723**: 376–382

Hirsh J, Dalen JE, Anderson DR, Poller L, Bussey H, Ansell J, Deykin D (2001) Oral anticoagulants: mechanism of action, clinical effectiveness, and optimal therapeutic range. Chest **119**: 8S–21S

Ishida JK, Wakatake T, Yoshida S, Takebayashi Y, Kasahara H, Wafula E, DePamphilis CW, Namba S, Shirasu K (2016) Local auxin biosynthesis mediated by a YUCCA flavin monooxygenase regulates haustorium development in the parasitic plant *Phtheirospermum japonicum*. Plant Cell **28**: 1795–814

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol **14**: R36

Kim HU, van Oostende C, Basset GJC, Browse J (2008) The *AAE14* gene encodes the Arabidopsis o-succinylbenzoyl-CoA ligase that is essential for phylloquinone synthesis and photosystem-I function. Plant J **54**: 272–83

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

Lee S, Kim S-G, Park C-M (2010) Salicylic acid promotes seed germination under high salinity by modulating antioxidant activity in Arabidopsis. New Phytol **188**: 626–637

Liang L, Liu Y, Jariwala J, Lynn DG, Palmer AG (2016) Detection and adaptation in parasitic angiosperm host selection. Am J Plant Sci **7**: 1275–1290

Liberles DA, Kolesov G, Dittmar K (2011) Understanding gene duplication through

biochemistry and population genetics. Evol after Gene Duplic 1–21

Lochner K, Döring O, Böttger M (2003) Phylloquinone, what can we learn from plants? BioFactors **18**: 73–78

Lohmann A, Schottler MA, Brehelin C, Kessler F, Bock R, Cahoon EB, Dormann P (2006) Deficiency in phylloquinone (vitamin K1) methylation affects prenyl quinone distribution, photosystem I abundance, and anthocyanin accumulation in the Arabidopsis *AtmenG* mutant. J Biol Chem **281**: 40461–40472

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **15**: 550

Lüthje S, Böttger M (1995) On the function of a K-type vitamin in plasma membranes of maize (*Zea mays* L.) roots. Mitt Inst Allg Bot Hambg **25**: 5–13

Lüthje S, Gestelen P, Córdoba-Pedregosa MC, González-Reyes J a., Asard H, Villalba JM, Böttger M (1998) Quinones in plant plasma membranes – a missing link? Protoplasma **205**: 43–51

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science (80-) **290**: 1151–1155

Macaulay KM, Heath GA, Ciulli A, Murphy AM, Abell C, Carr JP, Smith AG (2017) The biochemical properties of the two *Arabidopsis thaliana* isochorismate synthases. Biochem. J. 474:

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. [Miyashita mitsunori]

Müller R, Dahm C, Schulte G, Leistner E (1996) An isochorismate hydroxymutase isogene in *Escherichia coli*. FEBS Lett **378**: 131–134

Van Oostende C, Widhalm JR, Furt F, Ducluzeau A-L, Basset GJ (2011) Vitamin K_1 (Phylloquinone): function, enzymes and genes. Biosynth Vitam Plants Part B Vitam B6, B8,

B9, C, E, K **59**: 229

Palusa SG, Ali GS, Reddy ASN (2007) Alternative splicing of pre-mRNAs of Arabidopsis serine/arginine-rich proteins: regulation by hormones and stresses. Plant J **49**: 1091–1107

Petersen J, Stehlik D, Gast P, Thurnauer M (1987) Comparison of the electron spin polarized spectrum found in plant photosystem I and in iron-depleted bacterial reaction centers with time-resolved K-band EPR; evidence that the photosystem I acceptor A_1 is a quinone. Photosynth Res **14**: 15–30

Poirier Y, Antonenkov VD, Glumoff T, Hiltunen JK (2006) Peroxisomal β-oxidation-A metabolic pathway with multiple functions. Biochim Biophys Acta - Mol Cell Res **1763**: 1413–1426

Price PA, Faus SA, Williamson MK (1998) Warfarin causes rapid calcification of the elastic lamellae in rat arteries and heart valves. Arterioscler Thromb Vasc Biol **18**: 1400–1407

Price PA, Williamson MK (1981) Effects of warfarin on bone. Studies on the vitamin K-dependent protein of rat bone. J Biol Chem **256**: 12754–12759

Rowland BM, Taber HW (1996) Duplicate isochorismate synthase genes of *Bacillus subtilis*: regulation and involvement in the biosyntheses of menaquinone and 2,3-dihydroxybenzoate. J Bacteriol **178**: 854–61

Sadeghi M, Dehghan S, Fischer R, Wenzel U, Vilcinskas A, Kavousi HR, Rahnamaeian M (2013) Isolation and characterization of isochorismate synthase and cinnamate 4-hydroxylase during salinity stress, wounding, and salicylic acid treatment in *Carthamus tinctorius*. Plant Signal Behav **8**: e27335

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, et al (2010) Genome sequence of the palaeopolyploid soybean. Nature **463**: 178–183

Schopfer P, Heyno E, Drepper F, Krieger-Liszkay A (2008) Naphthoquinone-dependent generation of superoxide radicals by quinone reductase isolated from the plasma

membrane of soybean. PLANT Physiol **147**: 864–878

Serrano A, Córdoba F, Conzález-Reyes JA, Navas P, Villalba JM (1994) Purification and characterization of two distinct NAD(P)H dehydrogenases from onion (*Allium cepa*) Root Plasma Membrane. Plant Physiol **106**: 87–96

Shimada H, Ohno R, Shibata M, Ikegami I, Onai K, Ohto M, Takamiya K (2005) Inactivation and deficiency of core proteins of photosystems I and II caused by genetical phylloquinone and plastoquinone deficiency but retained lamellar structure in a T-DNA mutant of Arabidopsis. Plant J **41**: 627–637

Strawn MA, Marr SK, Inoue K, Inada N, Zubieta C, Wildermuth MC (2007) Arabidopsis isochorismate synthase functional in pathogen-induced salicylate biosynthesis exhibits properties consistent with a role in diverse stress responses. J Biol Chem **282**: 5919–33

Susumu O (1970) Evolution by gene duplication. Springer-Verlag ISBN 0-04-575015-7

van Tegelen LJ, Moreno PR, Croes AF, Verpoorte R, Wullems GJ (1999) Purification and cDNA cloning of isochorismate synthase from elicited cell cultures of *Catharanthus roseus*. Plant Physiol **119**: 705–12

Van Tegelen LJP, Bongaerts RJM, Croes AF, Verpoorte R, Wullems GJ (1999) Isochorismate synthase isoforms from elicited cell cultures of *Rubia tinctorum*. Phytochemistry **51**: 263–269

Tirosh I, Barkai N, Franke J, Hartmann E, Wiedmann M, Prehn S, Wiedmann B, Heinisch J, Anderson K, Andre B (2007) Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. Genome Biol 2007 84 **180**: 5682–5688

Widhalm JR, Ducluzeau AL, Buller NE, Elowsky CG, Olsen LJ, Basset GJC (2012) Phylloquinone (vitamin K_1) biosynthesis in plants: Two peroxisomal thioesterases of lactobacillales origin hydrolyze 1,4-dihydroxy-2-naphthoyl-coa. Plant J **71**: 205–215

Wildermuth MC, Dewdney J, Wu G, Ausubel FM (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. Nature **414**: 562–565

Xue L-J, Alabady MS, Mohebbi M, Tsai C-J (2015) Exploiting genome variation to improve next-generation sequencing data analysis and genome editing efficiency in *Populus tremula* × *alba* 717-1B4. Tree Genet Genomes **11**: 82

Yuan Y, Chung J-D, Fu X, Johnson VE, Ranjan P, Booth SL, Harding S a, Tsai C-J (2009) Alternative splicing and gene duplication differentially shaped the regulation of isochorismate synthase in *Populus* and *Arabidopsis*. Proc Natl Acad Sci U S A **106**: 22020–22025

Zhang X-N, Mount SM (2009) Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. Plant Physiol **150**: 1450–1458

CHAPTER 5

CONCLUDING REMARKS

The work presented here advances our understanding of the dual function of PhQ in flowering plants using a multi-disciplinary approach. The photosynthetic role of PhQ in PSI electron transport chain has been well established by previous studies, and gained validation in this work. However, the non-photosynthetic role of PhQ remained largely unexplored and thus was the focus of this work. Given the challenge of dissecting the non-photosynthetic function of PhQ from its photosynthetic function in photoautotrophic species, the work leveraged a photosynthesis-free system, a non-photosynthetic holoparasite, to study the non-canonical role of PhQ. However, as the holoparasite is a non-model species with no genome sequence available, molecular analysis at a genome scale presented a challenge. Although RNA-Seq data is available, published assemblies were highly fragmented. To surmount the difficulties, a novel pipeline was developed for improved assembly of the holoparasite transcriptome. This resulted in a solid foundation for the molecular analysis pertinent to the study focus. The findings were subsequently leveraged for exploration of the non-photosynthetic function of PhQ in *Arabidopsis thaliana*, *Glycine max*, and *Populus tremula x alba*.

RNA-Seq data accumulate substantially faster than genome data and provide a valuable molecular resource for investigations in non-model species. However, large outputs of short reads pose a challenge for current *de novo* assembly algorithms. Due to memory restrictions, existing *de novo* assembly methods often cannot load all the data, which lead to computational complexity and loss of information. Assembly algorithms that are reference-based are more sensitive and accurate than *de novo* assembly, but require a high-quality reference genome. A hybrid approach, known as ‘reference-guided’ assembly, was devised to improve assembly quality (Martin and

Wang, 2011). However, its implementation has been challenging because of difficulty in choosing an appropriate reference. Genome sequences are poorly suited for this purpose, even for closely related species, because of the high rates of natural variations, genome rearrangements and intronic alignment gaps. The PLAS pipeline developed in this study represents an innovative solution by employing protein sequences to guide assembly. RNA-Seq reads were organized by gene families into independent bins followed by parallel *de novo* assembly on each bin, and this process was repeated to increase the length and quality of assembled contigs. The original complex problem of *de novo* assembly was divided into multiple, less complex subtasks which were completed independently and in parallel. In this way, the memory requirement of *de novo* assembly was reduced with PLAS. More importantly, the pre-organization before assembly proved to be an efficient way to reconstruct more full-length transcripts with higher accuracy.

PLAS facilitated high-quality transcriptome reconstruction of the holoparasite, *Phelipanche aegyptiaca*. All PhQ biosynthetic genes were fully recovered, supporting the possibility that a functional PhQ biosynthesis pathway exists in this non-photosynthetic species. HPLC detection of PhQ in the imbibed seeds of *P. aegyptiaca* validated this possibility. Studies in photoautotrophic plants have established a compartmentalization of PhQ biosynthesis between plastids and peroxisomes. Enzymes catalyzing the early and late steps of PhQ biosynthetic pathway are targeted to plastids and intermediate steps are completed within peroxisomes. In this study, a detailed analysis of the protein sequences of PhQ biosynthetic genes in the holoparasite revealed that enzymes for the last two steps have lost their plastid-targeting signal peptide. GFP experiments revealed targeting of those proteins to the plasma membrane, and a likelihood that PhQ is therefore synthesized in plasma membranes of the holoparasite. The findings provide molecular support for previous reports that PhQ is involved in plasma membrane redox activities.

Transcriptomes of two other parasitic plants, *Triphysaria versicolor* and *Striga hermonthica*, were also reconstructed. They are close relatives of *P. aegyptiaca* but capable of photosynthesis and therefore valuable for comparative analysis. Subcellular localization

predictions showed that PhQ biosynthetic enzymes were targeted to plastids as in photoautotrophic species. Co-expression network analysis revealed strong connections between PhQ biosynthetic genes and those implicated in parasitism, such as peroxidases and quinone reductases. Those connections are weakened with increasing photosynthetic competence in *S. hermonthica* and *T. versicolor*. As a component of plasma-membrane electron transport, PhQ might be associated with haustorial development and parasite establishment.

This work also explored non-canonical PhQ function and biosynthesis pathway evolution in three photoautotrophic model species that offer rich genomic resources. RNA-Seq data of both photosynthetic-source and heterotrophic-sink tissues were used to mine the expression patterns of PhQ biosynthetic genes. PhQ biosynthetic genes were strongly expressed in photosynthetic-source tissues, in concordance with a predominant role in photosynthetic electron transport. However, some PhQ biosynthetic genes also exhibited moderate expression in heterotrophic-sink tissues. GO enrichment analyses of PhQ-coexpressed genes in both sink and source tissues showed similar patterns, potentially revealing the difficulty of disassociating non-photosynthetic aspects from photosynthetic function. Alternatively, the results may reflect the common functionality of the electron transport chains in both photosynthetic and non-photosynthetic contexts. The challenge is also partially due to data limitations and complications of separating heterotrophic tissues from photosynthetic tissues. Future experiments with high-resolution tissue sampling are needed to facilitate investigations on this topic.

This work has uncovered a potential role of *ICS2* from *Arabidopsis* and *DHNAT* from all three photoautotrophic species in certain plant defense responses. *AtICS2* was specifically induced in leaves by osmotic stresses like dehydration, salt and mannitol treatments, whereas *DHNAT* genes were variably expressed in heterotrophic-sink tissues depending on the species and abiotic stress. Their associations with stress responses gained support from GO enrichment analyses. Our observations of *AtICS2* and *DHNAT* have not been described in previous studies, and will provide direction for further investigations of their specific roles.

In conclusion, this study proved the plasma-membrane localization of PhQ biosynthesis conserved in both parasitic and photoautotrophic plants, and uncovered an association of PhQ with parasitism. The work also provided a bioinformatics tool for transcriptome assembly in non-model species with improved performance over existing *de novo* assembly methods. Although the investigation into the non-photosynthetic role of PhQ in photoautotrophic species remained inconclusive, unexpected findings about a link between duplicated PhQ biosynthetic genes and plant defense responses opened a door for future research regarding the functional plasticity of the PhQ biosynthetic pathway.