The Pennsylvania State University
The Graduate School

The Huck Institutes of the Life Sciences

COMPARATIVE TRANSCRIPTOME AND PHYLOGENOMIC ANALYSES ON THE EVOLUTION OF PARASITIC OROBANCHACEAE

A Dissertation in
Plant Biology

by
Zhenzhen Yang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of
Doctor of Philosophy

May 2016

The dissertation of Zhenzhen Yang was reviewed and approved* by the following:

Claude W. dePamphilis
Professor of Biology
Dissertation Advisor
Chair of Committee

Charles T. Anderson
Assistant Professor of Biology

Naomi S. Altman
Professor of Statistics

Yinong Yang
Associate Professor of Plant Pathology and Environmental Microbiology

Teh-hui Kao
Distinguished Professor of Biochemistry and Molecular Biology
Chair of the Plant Biology Graduate Program

*Signatures are on file in the Graduate School

ABSTRACT

Parasitic plants are plants that form a parasitic association with their host using a connecting organ called the haustorium, a novel feeding structure through which parasites withdraw nutrients such as carbon and nitrogen from the conducting tissues of the host. Orobanchaceae is the only family containing species with the full spectrum of parasitic capabilities, including one nonparasitic autotrophic genus, *Lindenbergia*, and more than 90 genera (>2000 species) of parasites with varying degrees of photosynthesis and host dependence. To understand the genetic changes that led to a parasitic lifestyle, a transcriptome sequencing project was initialized to interrogate multiple stages of growth and development of three parasitic plants that span the range of parasitic dependence. Around 180 genes are upregulated during haustorial development following host attachment in at least two species, and these are enriched in proteases, cell wall modifying enzymes, and extracellular secretion proteins. The majority of parasitism genes were duplicated before the divergence of Orobanchaceae and *Mimulus*, a related nonparasitic plant. This suggests that gene duplication plays a role in the origin of parasitism. A comparative analysis of these genes' homologs in sequenced nonparasitic plant genomes revealed that parasitic plants make haustoria by co-opting genes from root and floral development. Gene duplication, often taking place in a nonparasitic ancestor of Orobanchaceae, followed by regulatory neofunctionalization, was an important process in the origin of parasitic haustoria. Horizontal gene transfer (HGT), acts as another evolutionary force contributing to parasite adaptation. At least 42 gene families accounting for 52 transfers – largely via genomic integration - result in functional transcripts that were primarily involved in translation, defense responses, transposable element (TE), and other diverse roles. Three lines of evidence indicate an adaptive role of HGT in parasite evolution - (i) A majority of HGT genes are upregulated in haustoria-related tissues in the most parasitic *Phelipanche aegyptiaca*; (ii) A higher frequency of HGTs are observed in parasites with a higher degree of parasitism; (iii) A portion of genes detected to have evolved some adaptive sites under positive selection. The study of strigolactone (SL) pathway in parasitic plants reveals that parasitic plants still retain genes in SL synthesis. Two SL biosynthesis genes and D14 (the receptor) are upregulated in haustorial structures, indicating a possibility of SL in haustoria development.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES x

ACKNOWLEDGEMENTS xi

Chapter 1 Introduction to parasitic plants and related research 1

1.1 Introduction to parasitic plants 2
1.1.1 Parasitic plants – classification and morphology 2
1.1.2 Parasitic Orobanchaceae 3
1.2 Biology of parasitic plants 3
1.2.1 Germination of parasitic plants 3
1.2.2 The haustorium 4
1.2.3 Haustorium initiation and early development 4
1.2.4 Haustorium penetration and development 5
1.2.5 Physiology of parasitic plants 6
1.2.6 Nutrient transfer 7
1.3 Parasite control 8
1.3.1 Problems and germination-based approaches 8
1.3.2 Known parasitism genes 9
1.3.3 Parasite and host defense – the arms race 10
1.4 Evolution of novel traits 12
1.4.1 Haustorium as a good model to study the evolution of novel traits 12
1.4.2 Mechanisms for the origin of novel traits 12
1.4.3 Gene duplication and regulatory networks (origin of flower) 13
1.4.4 Homeobox domain-mediated regulation of leaf development 14
1.4.5 Co-option of existing genes and gene family expansion 14
1.4.6 Mutation-driven positive selection 15
1.4.7 Haustoria origins – the exogenous model (HGT) 15
1.4.8 Haustoria origins – the endogenous model (gene recruitment) 17
1.5 Horizontal gene transfer 18
1.5.1 Mitochondrial HGT 18
1.5.2 HGT in parasitic plants 19
1.5.3 HGT of non-plant origin 20
1.5.4 Models for HGT 21
1.5.5 HGT – where do we go from here? 21
1.6 A transcriptomic and phylogenomic approaches to study parasitic plants 22
1.6.1 Driving questions for research on parasitic plants 22
1.6.2 Sequencing technologies allow the capture of transcriptomes and genomes 23
1.6.3 The power of phylogenomic approaches (polyploidy (gene duplication) and HGT) 24
1.6.4 Overview of PPGP 26

Chapter 2 Identification of parasitism genes and the origin of the haustorium 28

2.1 Introduction to novel traits 29
2.2 Results 31
2.2.1 Assembly statistics and coverage 31
2.2.2 Validation of genes with known roles in Orobanchaceae parasitism 34
2.2.3 Differential gene expression and clustering to identify haustorial genes 36
2.2.4 Shared haustorial genes are enriched for proteolysis and extracellular region localization 38
2.2.5 Two examples illustrating haustorial gene expression evolution 42
2.2.6 Identifying haustorial initiation genes and "parasitism genes" 44
2.2.7 A majority of parasitism genes evolved through gene duplication 47
2.2.8 Regulatory neofunctionalization and origin of the haustorium from root and flower 48
2.2.9 Parasitism genes show signatures of adaptive evolution or relaxed constraint in parasitic lineages 52
2.2.10 The majority of the parasite-specific sequences have unknown functions 54
2.3 Discussion 55
2.3.1 Summary of results 55
2.3.2 Cell wall degradation enzymes and the haustorium 56
2.3.3 Proteases, transporters, and the haustorium 57
2.3.4 Gene duplication and regulatory neofunctionalization – origin of parasitism 59
2.3.5 Origin of the haustorium involves co-option of root and/or flower genes 60
2.3.6 Parasite-specific genes – mobile elements 61
2.3.7 Conclusion 61
2.4 Materials and methods 62
2.4.1 Tissues, libraries, and sequence data 62
2.4.2 Assembly, cleaning, and annotation (including gene family classification) 63
2.4.3 Developing a component-orthogroup from each *de novo* assembly 64
2.4.4 Read mapping and expression normalization 64
2.4.5 PV-clustering of global transcriptional profile 65
2.4.6 Identification of differentially expressed genes for candidate parasite gene assignment 65
2.4.7 K-Means clustering to identify putative parasite feature within each species 65
2.4.8 Hierarchical clustering to identify putative parasite feature within each species 66
2.4.9 Self organizing maps (SOM clustering) to identify patterns of haustorial-specific expression 66
2.4.10 Enrichment analysis – parasite genes versus whole plant background 67
2.4.11 Phylogenies and Ka/Ks constraint analysis for parasite genes 67
2.4.12 Scoring gene duplications 68
2.4.13 Selective constraint analysis 69
2.4.14 Expression of haustorial orthologs in nonparasitic species 70

Chapter 3 HGT in parasitic Orobanchaceae 72

3.1 Introduction of HGT 73
3.1.1 HGT in bacteria 73
3.1.2 Review of current methods for HGT identification 73
3.1.3 HGT in plants 74
3.1.4 Host range and HGT detection 75
3.1.5 Objectives and overview of the analyses 76
3.2 Results 77
3.2.1 The phylogenomic pipeline and the analytical schema for HGT detection 77
3.2.2 52 high-confidence HGT events 78
3.2.3 Transfers from ancestral host lineages 80
3.2.4 Increased numbers of HGT with increased heterotrophic dependence 81
3.2.5 Integration of genomic fragments 87
3.2.6 Functional HGT 92
3.2.7 Adjacent HGTs in two *Striga* species 98
3.2.8 Absence of transfers from parasitic plants to their hosts 102
3.3 Discussion 104
3.3.1 A stringent and robust phylogenomic approach for HGT identification 104
3.3.2 Reasons for increased HGT with increase heterotrophic dependence 105
3.3.3 A proposed adaptive role of HGT in parasitic plants 105
3.3.4 Genomic integration, functional inference, the tip of an iceberg 106
3.3.5 HGT hotspot and possible transfer mechanisms 107
3.3.6 Conclusions 108
3.4 Materials and methods 110
3.4.1 Removal of contamination 110
3.4.2 Phylogenomic reconstruction of parasite gene trees 110
3.4.3 HGT screening on phylogenetic trees 111
3.4.4 HGT validation by increased taxon sampling 112
3.4.5 Intron analyses 113
3.4.6 Genome assembly of three parasites 114
3.4.7 Estimation of number of transfers 114

Chapter 4 Conclusions and future directions 115

4.1 Conclusions 116
4.2 Future directions 116
4.2.1 Studies based on experimental characterization 116
4.2.2 Reveal genetic mechanisms underlying physiological differences of three parasites with comparative transcriptome analyses (PPGP2) 118
4.2.3 Evolution of parasitic plants – phylogenetic inferences of species relationships and HGT 120

Appendix A HGT in *Striga asiatica* 122

A.1 Introduction 123
A.2 Phylogenomic-based approach 123
A.2.1 Constructing a species tree using 26 sequenced plant genomes 123
A.2.2 Phylogenomic screening 124
A.3 Validation of BLAST-predicted HGT 127

Appendix B Evolution of strigolactone pathway in parasitic Orobanchaceae 130

B.1 Introduction 131
B.1.1 Strigolactone-dependent germination of *Striga* and *Phelipanche* 131
B.1.2 Review of strigolactone pathway components and roles 131
B.1.3 Regulation of SL pathway 132
B.1.4 Working hypothesis for SL-mediated germination of parasites 132
B.2 Results and discussion 133
B.2.1 Conservation of strigolactone biosynthesis genes in parasitic plants 133
B.2.2 Review of SL receptor diversification in parasitic *Striga* 135
B.2.3 Relevance of SL receptor diversification in internal and external SL recognition 136
B.2.4 Relevance of *KAI2d (HTL)* in host recognition and specificity 138
B.2.5 Transcriptional dynamic of *KAI2* members among three parasites 139
B.2.6 Parasitic D14s show abundant expression in interface and haustoria 140
B.2.7 Haustorial expression of additional SL pathway genes implies a role of SL in haustorial development 140

References 146

LIST OF FIGURES

Figure **2-1**. An illustration of stages of each parasitic plant used in the Parasitic Plant Genome Project (Westwood et al. 2012) in this study. 33

Figure **2-2**. Gene expression profiles from RNA-seq data for two previously characterized parasitism genes in *Triphysaria* (*QR1*, left, and *Pirin*, right). 35

Figure **2-3**. Gene expression clustering (A) and heatmap (B) of upregulated genes in post attachment haustorial stages 3 & 4 ("haustorial genes") in parasitic Orobanchahceae. .. 36

Figure **2-4**. Gene family phylogeny and gene expression profile of two orthogroups showing co-option of haustorial genes from flower and root. 43

Figure **2-5**. Venn diagram illustrating the number of orthogroups with upregulated expression in stage 3 and/or stage 4 (by K-means and SOM) in *Triphysaria*, *Striga*, and *Phelipanche*. 45

Figure **2-6**. HIGs: Orthogroups containing genes upegulated in root or seedlings following haustorial initiation factor (HIF) exposure (stage 2) compared to germinating seedlings (stage 1) in *Triphysaria*, *Striga*, and *Phelipanche*. 46

Figure **2-7**. Overall similarity of transcriptional profiles of all stages in three parasites..... 50

Figure **2-8**. Haustorial genes in parasitic species were recruited from root, flower and other tissues. 51

Figure **2-9**. Haustorial genes show evidence of adaptive selection or relaxed selective constraint. 53

Figure **3-1**. Three models for phylogenomic identification of HGTs, and further examination of the preliminary-screened HGT candidates. 79

Figure **3-2**. HGT artifacts due to insufficient taxon sampling (A-B), contamination (C), and frame-shift errors (D-E)..... 83

Figure **3-3**. RAxML-based Maximum likelihood (ML) trees supporting HGT in two orthogroups, donor families and recipient taxa inferred from the 42 HGT set. 86

Figure **3-4**. Genomic horizontal transfer of a tRNAHis guanylyltransferase from an ancestor of *Fragaria* to *Phelipanche* parasites 90

Figure **3-5**. Intron phylogeny and intron-positions for several HGT orthogroups that encode tRNA synthetase/transferase. 91

Figure **3-6**. A heat map shows the expression of HGT transgenes in *Phelipanche aegyptiaca*. 93

Figure **3-7**. Phylogeny of orthogroup 11841 and predicted 3D structure. 94

Figure **3-8**. RAxML-based ML trees and comparisons of HGT genes between parasite and its donor supporting two HGTs being adjacent in the recipient genome (A1, A2, B) and donor genome (A3 and C). 101

Figure **3-9**. Falsely identified parasite-to-host HGTs due to frame-shift errors. 103

Figure **A-1**. RAxML-based maximum likelihood species tree for 26 selected genomic taxa. 124

Figure **A-2**. Six rosid-derived preliminary HGT trees from phylogenomic screening 126

Figure **A-3**. Two monocot-derived preliminary HGT trees from the phylogenomic screening. 127

Figure **A-4**. Phylogenetic tree of a horizontally-transferred alanine-tRNA synthetase. 128

Figure **A-5**. Maximum Likelihood tree of a transcribed gene encoding an amino-tRNA synthetase with the forced codon alignment using RAxML. 129

Figure **B-1**. RAxML-based maximum likelihood trees for four SL biosynthesis genes – D27, CCD7, CCD8, MAX1. 134

Figure **B-2**. Expression of D14 in all three parasitic plants. 142

Figure **B-3**. CCD8 expression in parasitic plants. 143

Figure **B-4**. MAX1 expression in parasitic plants 144

Figure **B-5**. Expression of KAI2s (KAI2c, KAI2i, KAI2d) mapping onto phylogeny in all three parasitic plants. 145

LIST OF TABLES

Table **2-1**.Transcriptome assembly statistics in the post-processed combined assembly for each study species. 34

Table **2-2**.Transcriptome gene capture statistics in three parasitic species..... 34

Table **2-3**. Enriched Pfam domains in the shared set of haustorial unigenes identified by either K-means or SOM clustering in *Triphysaria*, *Striga* and *Phelipanche*. 39

Table **2-4**. Enriched GO cellular component (GO-CC), biological process (GO-BP) and molecular function (GO-MF), KEGG pathway, and tissue expression terms among shared set of haustorial unigenes identified by either K-means or SOM clustering in *Triphysaria*, *Striga*, and *Phelipanche*. 41

Table **2-5**. Phylogenetic placement of gene duplications observed in gene families with shared parasitism genes. 47

Table **3-1**. A scoring scheme used to score each phylogenetic tree based on bootstrap support, depth of donor clades, and long branches. 80

Table **3-2**. Information of the 42 HGT orthogroups including the HGT recipient, donor, expression, dN/dS, functional category, and homology-based annotation. 84

Table **3-3**. SH test to evaluate number of transfers in HGT trees by constraining multiple HGT genes to one monophyletic clade. 87

Table **3-4**. PAML analyses with branch test on codon alignment of 42 HGT orthogroups testing presence of purifying selection, relaxed constraint, or positive selection. 95

Table **3-5**. PAML analyses with the branch-site model on codon alignment of 15 HGT orthogroups with greater dN/dS on HGT genes compared to the background, identifying the presence of sites under positive selection. 96

ACKNOWLEDGEMENTS

I would like to thank all the people for their assistance during my Ph.D. studies. First and foremost, thanks to my advisor, Claude dePamphilis, for his passion and expertise in research related to parasitic plants, evolution, bioinformatics, and plant taxonomy. His inspiring ideas, patience in advising, as well as nice personality have led to my completion of my Ph.D. It is of my great fortune to have met him and been supervised by him that results in an overall enjoyable Ph.D. journey. I would also like to thank my colleague, Eric Wafula. Being a computational programmer in the lab, he has stimulated my interest and accumulation of skills in programming, bioinformatics, and evolutionary approaches. Without his influence of persistence, kindness, and encouragement, time would have witnessed another several years for the completion of my Ph.D. My gratitude also goes to Huiting Zhang, a third-year graduate student, who was using functional characterization to validate some of my candidate genes in parasitic plants. Her research in this area has given me courage to keep me forward. My committee members, Naomi, has provided a lot of valuable suggestions and support in using a statistical point of view to address biological questions related to next-generation sequencing data and evolutionary analyses. The counseling and help she has offered was indispensable for my thesis projects. I thank her fully. Charlie's questions on my comprehensive exam have provided me with thoughts on strigolactone evolution in parasitic plants, which generated really interesting hypotheses on its possible role in haustorial development. His thoughts on pectate lyase have also been insightful in generating hypotheses that are currently being investigated by Huiting. Dr. Yang's knowledge on defense responses in haustorial genes provided us with thoughts on our haustorial gene list, which gave us a better understanding of the parasite-host defense and counter-defenses. His views have also led to our selection of a couple of candidate genes, for instance, leucine rich-repeat receptor like kinases for functional knockout. I am grateful to all of them, for giving me the motivation in the continuous pursuit of remarkably interesting questions related to fascinating parasitic plants. Last but not least, I show my utmost appreciation to my husband, Liye Zhang, who has given me endless encouragement, support, and help on my project and life. He also actively helped me with my practice talks for several important presentations. His involvement in my project and life, his broad interest in science but not limited to his field of cancer biology, make him a wonderful companion in my life.

Chapter 1

Introduction to parasitic plants and related research

1.1 Introduction to parasitic plants

1.1.1 Parasitic plants – classification and morphology

Parasitic plants are plants that have the ability to form specialized feeding structures called haustoria that allow them to extract water and nutrients from their host (Kuijt 1969). Almost 1% of all flowering plants are parasitic plants, including 4,500 species within 28 families (Thorne 2002). Parasitism has evolved independently at least 11 times during angiosperm evolution (Barkman, et al. 2007). Depending on the tissue to which the parasite attaches, they can be classified as either root or stem parasites. Root parasites contain pests from the family of Orobanchaceae, such as *Striga* and *Phelipanche*, whereas examples of stem parasites are represented by *Cuscuta* (dodder) and mistletoes (the family of Loranthaceae). Parasitic plants can also be classified as hemiparasites (if they retain some photosynthetic capability, and thus are at least partly autotrophic) or holoparasites (if they are entirely heterotrophic) (Kuijt 1969; dePamphilis and Palmer 1990; Heide-Jørgensen 2013b). A majority (90%) of parasitic plants are hemiparasites, and root parasites account for 60% of all parasitic plants (Heide-Jørgensen 2013b). Additionally, parasitic plants can be classified by their degree of host dependence. Facultative parasites must retain photosynthetic abilities and are opportunistic parasites that are able to complete their life cycle without attaching to a host (Kuijt 1969; Westwood, et al. 2010) whereas obligate parasites must form a host attachment in order to complete their life cycle. A holoparasite is often an obligate parasite (such as *Phelipanche*), whereas a hemiparasite can be either facultative (*Triphysaria*) or obligate (*Striga*).

Morphologies of parasitic plants also differ significantly from autotrophic plants. They often have characters such as shortened vegetative stem, reduced leaves, simplified inflorescences, and conversion from few, large seeds to numerous, small seeds (from facultative parasites to obligate parasites). For instance, in stem parasitic *Cassytha* and *Cuscuta*, the vegetative tissues are reduced to only stem and scale leaves. In *Hydnora*, a basal angiosperm, however, leaves are completely absent. In fact, the above ground vegetative tissues of this parasitic plant contain only flower tissues that are thick and succulent in texture with three openings.

1.1.2 Parasitic Orobanchaceae

Among all the parasitic lineages, Orobanchaceae is the only family containing species with a complete spectrum of parasitic capabilities from facultative to obligate parasitism and from hemiparasites to holoparasites (Westwood et al. 2010). Being the largest family of parasitic plants, it also includes a basal nonparasitic lineage (*Lindenbergia*), and thus this family is an ideal group for investigating the evolution of parasitism. Three other parasites in this family include *Triphysaria versicolor* (facultative hemiparasite), *Striga hermonthica* (obligate hemiparasite), and *Phelipanche aegyptiaca* (obligate holoparasite). *T. versicolor* has a wide host range including monocots and dicots (Estabrook and Yoder 1998; Jamison and Yoder 2001), and common hosts used in the laboratory include *Medicago* and maize. *S. hermonthica* and *S. asiatica* specializes on grasses including rice, maize, sorghum, and millet (De Groote, et al. 2008; Parker 2009), whereas *S. gesneroides* is a dicot feeder growing on cowpea (Timko, et al. 2007). *P. aegyptiaca* also has a wide host range that is nonetheless limited to dicots such as legumes, tomato, and *Arabidopsis* (Carlon, et al. 2005; Schneeweiss 2007; Parker 2009).

1.2 Biology of parasitic plants

1.2.1 Germination of parasitic plants

Parasitic plants have different requirements for germination. In facultative parasitic *Triphysaria*, they can germinate independent of a host and it is believed that they retain all the pathways for germination (Westwood et al. 2010). Obligate parasites such as *Striga* and *Phelipanche*, have to rely on a host for germination(Brown, et al. 1949; Brown, et al. 1951; Westwood et al. 2010). The germination stimulant later was discovered as Strigolactone (SL) (Butler 1995), a hormone synthesized and released by the roots of their host plants. It is thought that host-dependent germination is a novel adaptation and represents an evolutionary advantage for these parasitic plants, considering that germination without a host would still result in death in the absence of a host.

1.2.2 The haustorium

The presence of a haustorium is a unique feature and hallmark of parasitic plants, and Kuijt called the haustorium the "essence of parasitism" (Kuijt 1969). It was considered by Kuijt to be a modified root that forms a physiological and morphological link connecting the conducting tissues of the parasite and its host. The haustorium derives from root or radicles. In *Triphysaria*, for instance, the haustorium develops after seedlings have established autotrophically, and its haustorium as an extension of lateral root or adventitious root is termed a lateral haustorium. In *Striga* and *Phelipanche*, however, the haustorium that develops from the apex of the primary root is a terminal haustorium; in these two species there are both terminal and lateral haustoria. In general, terminal haustoria are the largest and can often support the parasite throughout its life cycle, whereas lateral haustoria often last for a few months (Heide-Jørgensen and Kuijt 1995). The haustorium formation involves two processes, initiation and penetration. Haustorial initiation is caused by inducing factors from host root exudates and morphologically speaking, is characterized by the presence of a bulge on the root tip and the formation of haustorial hairs. In facultative *Triphysaria*, haustorial initiation occurs without a host, whereas in *Striga* and *Phelipanche*, their haustorial initiation, which comes after germination induced by host signals, occurs with close contact to a host.

1.2.3 Haustorium initiation and early development

Haustorium initiation requires chemical and physiological stimuli from a compatible host in most Orobanchaceae (Baird and Riopel 1984). The process of initiation that happens on parasite root (hemiparasites) or radicles (holoparasites) is attributed to host root exudates or purified inducing factors from the host. In general, the common haustorial inducing factor (HIF) used in a laboratory is 2,6-dimethoxybenzoquinone (DMBQ) (Keyes, et al. 2000). Haustorium initiation involves the cessation of tip growth (Baird and Riopel 1984; Riopel and Baird 1987), interruption of cell replication (Torres, et al. 2005), swollen tip regions (Bandaranayake and Yoder 2013a), and the elongation of epidermal cells into long haustorial hairs that later allow for

grasping of the host root (Baird and Riopel 1985). In addition to the morphological changes, experiments with *Triphysaria* roots transformed with an auxin-responsive reporter showed that auxin was also involved in this process (Tomilov, et al. 2005), as well as cell wall loosening enzymes such as expansins that are transcriptionally regulated during this period (O'Malley and Lynn 2000; Wrobel and Yoder 2001; Torres et al. 2005).

The study of haustorium initiation in *T. versicolor* has allowed for the identification of genes in the redox signal transduction pathway – *TvQR1 and TvQR2* (Bandaranayake, et al. 2010). Both of these genes show upregulated expression in response to DMBQ, and silencing of *TvQR1* using RNAi results in reduced numbers of haustoria (Bandaranayake et al. 2010). Biochemical analyses showed that TvQR1 encodes an NADPH-dependent single electron quinone oxidoreductase that generates semiquinones that further form reactive oxygen species (ROS) in the presence of oxygen (Bandaranayake et al. 2010). The generated ROS are predicted to initiate the morphological changes associated with haustorial development, such as cortical cell expansion and haustorial hair elongation (Foreman, et al. 2003). As these semiquinones are toxic, the two-electron reduction enzyme TvQR2 acts as a detoxification enzyme to eliminate the semiquinones (Bandaranayake et al. 2010).

1.2.4 Haustorium penetration and development

The production of haustorial hairs allowing the parasite to anchor to the host root surface provides the first crucial step for the development of the attachment organ. The cells are then in preparation for penetration. For instance, the intrusive cells within haustoria of *Rhamphicarpa fistulosa* show dense cytoplasm, enlarged nucleoli, numerous mitochondria, and rough endoplasmic reticulum (Neumann, Vian, Weber and Sallé 1999). To penetrate the host vascular tissues, the parasite haustorium utilizes a combination of mechanical and enzymatic processes.

Evidence of host cells that are pushed aside by *Orobanche spp.* intrusive cells and the dissolution of the middle lamella between host cells supported the mechanical invasion of host cells (Joel and Losner-Goshen 1994b). In *P. aegyptiaca*, the endodermis of the host root vasculature is disrupted by the parasite in order to establish vascular connections, supported by dissolution of Casparian strips in the endodermis (Joel DM 1998). The penetration of host vascular tissues by *P. aegyptiaca* is contributed to by a list of pectolytic, cellulolytic and proteolytic enzymes (Shomer-Ilan 1992, 1993, 1999), as well as a series of cell wall-degrading enzymes including cellulases, polygalacturonases, xylanases and proteases (Joel DM 1998) in the tubercle.

1.2.5 Physiology of parasitic plants

Parasitic plants often have higher transpiration rate compared to their host (Ehleringer and Marshall 1995; Jiang, et al. 2003). Therefore, parasites such as *Striga* and mistletoes are often found in open, sunny habitats with unlimited access to sunlight where shading is avoided. In fact, most root-hemiparasitic plants inhabit temperate regions, in particular, Mediterranean climates or African countries. Orobanchaceae are most diverse in South Africa, Mediterranean, East Asia, and western America (Bennett and Mathews 2006; McNeal, et al. 2013b).

A high transpiration rate is found in *Striga*, the stomata of which remain open even under stress conditions or under high levels of ABA (Smith and Stewart 1990), to drive the transfer of nutrients from their host. In addition, parasitic plants upregulate their ABA in response to host attachment, and this can be 18-fold higher than the barley host upon attachment of *Rhinanthus minor* (Jiang et al. 2003). In addition, they also regulate the synthesis of ABAs of their hosts. For instance, *S. hermonthica* induced two-fold higher levels of ABA in its sorghum host leaf tissue and xylem sap (Taylor, et al. 1996). Although ABA is known to induce stomatal closure, parasitic plants seem to have evolved a reduced sensitivity to ABAs. This is thought to contribute to greater water flow into the parasite (Jiang, et al. 2004).

1.2.6 Nutrient transfer

Haustoria of parasitic plants transport high flows of xylem contents including water and minerals from their hosts (Ehleringer and Marshall 1995). Transfer of nutrients is driven by lowering their water potential relative to their hosts, which is contributed by accumulating high levels of osmotic compounds such as sugar alcohols (mannitol) (Ehleringer and Marshall 1995), or maintaining open stomata especially in many hemiparasites (Jiang et al. 2003). Notably, stomata of hemiparasites *Rhinanthus* and *S. hermonthica* keep open stomata when they are attached to hosts (Jiang et al. 2003), even if their host is under severe water stress (Smith and Stewart 1990). Interestingly, the free-living *Rhinanthus* has closed leaf stomata, but the stomata remain open when it attaches to a host (Jiang et al. 2003). Holoparasites, such as *Phelipanche*, remain underground for a long term before the emergence of above-ground vegetative structures and have reduced leaf; thus cannot achieve a high transpiration rate. But accumulation of high levels of mannitol may be the primary driving force for effective transfer of xylem contents in many parasites (Harloff and Wegmann 1993).

Although xylem connections of haustoria are mainly responsible for the transfer of inorganic compounds, haustoria also mobilize carbon compounds from their host via clear phloem connections. Hemiparasites that can fix carbon via their own photosynthesis parasitize their hosts mainly for the uptake of nitrogen and water. In holoparasites, however, the significantly reduced levels of photosynthesis require them to parasitize for both nitrogen and carbon. In general, the percentage of carbon uptake in hemiparasites is estimated to be up to 30% (10% in facultative hemiparasite *Triphysaria* and 60-70% in emerged obligate holoparasite *Striga hermonthica*) (Irving and Cameron 2009), whereas in *Phelipanche* which has a complete loss of photosynthesis, carbon uptake from the host is 100%. In underground *Striga* that has not produced green tissues for photosynthesis, their dependence on carbon from host is also 100%. Although mature emerged *S. hermonthica* obtained 60%-70% of its carbon from its host (Press, et al. 1987), this can vary a lot among different species of *Striga* and among different hosts. In *S. hermonthica*, the carbon uptake from millet can be up to 80%, whereas in *S. gesnerioides* on cowpea, the uptake can be as extreme as 99% (Press 1995). In holoparasites with a complete carbon dependence on their hosts, the sugar levels accumulated in haustorial tubercles ranges from 6- to 8-fold higher than in their hosts (Aber, et al. 1983).

The photosynthetic product that holoparasites obtain from their host can vary in different forms of sugar. For instance, in *P. ramosa*, sucrose is converted to other compounds such as

hexoses, mannitol and starch (Draie, et al. 2011), presumably acting to increase the osmotic potential of the parasite. Tubercles of *O. foetida* primarily accumulate and convert sugar to storage compounds like starch when attached to faba bean (Abbes, et al. 2009). Mannitol is a sugar alcohol that has the advantage of improving the osmotic potential for parasites to drive the flow of xylem contents. The level of mannitol can be accumulated from 34% in *O. hederae* stems (Abbes et al. 2009) up to 77% in *S. asiatica* leaves. Interestingly in *S. hermonthica*, its xylem sap contained 58% of mannitol, but none was detected in its sorghum host (Press and Graves 1991).

1.3 Parasite control

1.3.1 Problems and germination-based approaches

The ability of the haustorium to efficiently transfer host resources results in substantial yield loss to several economically important crop plants. For example, in sub-Saharan Africa, witchweed (*Striga spp*.) infests over 50 million hectares of arable farmland cultivated with corn and legumes, causing annual yield loss estimated to exceed $10 billion USD (Scholes and Press 2008). Two characters make *Striga* and *Phelipanche* notorious weeds and challenging for eradication. First, they are root parasites that attack crops underground, and often have already caused severe damage to host plants before farmers notice their emergence above the soil. Second, they produce up to a million tiny seeds that remain in the soil and are difficult to remove. Because of this, their infestations of staple crops can result in substantial or near complete yield loss, exacerbating problems of low food security.

The dependence of detrimental parasitic weeds, *Striga* and *Phelipanche*, on SLs from their hosts allows the development of control strategies that target their germination. The first strategy is suicidal germination, the induction of their germination in the absence of a host by application of synthetic analogs of SL in the soil (Kondo, et al. 2007; Mwakaboko and Zwanenburg 2011). The second approach includes trap crops, which is the introduction of non-host crops that can produce for instance, increased levels of germination stimulants (Chittapur, et al. 2000). The genes responsible for SL synthesis have been identified and include CCD7, CCD8,

MAX1, and D27 (Al-Babili and Bouwmeester 2015). Plants overexpressing these genes are expected to produce higher levels of SLs to act as trap crops. The third approach involves the use of herbicides that produce reduced levels of SLs resulting in decreased parasite germination. For instance, because SL biosynthesis includes some enzymes common in carotenoid biosynthesis (Matusova, et al. 2005; Lopez-Raez, et al. 2008), some inhibitors targeting carotenoid biosynthesis have been used on rice to cause decreased *Striga* germination and infection (Sergeant, et al. 2009; Ito, et al. 2010; Jamil, et al. 2010). Other approaches can be applied according to the characteristic of SL induction by low nutrient conditions such as low nitrogen and phosphate. Thus the application of fertilizers could be useful in resulting in reduced SL production (Yoneyama, et al. 2009; Jamil, et al. 2011).

1.3.2 Known parasitism genes

Identifying genes with key roles in parasitism may reveal novel strategies to control weedy agricultural pest species (Aly, et al. 2011; Alakonya, et al. 2012; Bandaranayake, et al. 2012; Westwood, et al. 2012; Bandaranayake and Yoder 2013b; Bandaranayake and Yoder 2013c; Ranjan, et al. 2014). Despite decades of research, only a few genes have been previously characterized with specific roles in the parasitic process in Orobanchaceae. One quinone oxidoreductase gene (*TvQR1*) is necessary for haustorium initiation through redox bioactivation of haustorial inducing factors (HIFs) in *Triphysaria* (Bandaranayake et al. 2010; Ngo, et al. 2013). Additionally, *TvPirin* is upregulated by HIFs (or by contact with host roots) and putatively functions as a positive regulator of other genes needed for haustorial development (Bandaranayake et al. 2012). Finally, *mannose 6-phosphate reductase* (M6PR) in *Phelipanche aegyptiaca* (a root parasite) was also shown to be involved in parasite metabolism. Silencing of the parasite M6PR gene by RNAi from the host resulted in decreased mannitol concentration in the haustorium tubercle and increased tubercle mortality, thus clarifying the role of mannitol in parasitism (Aly, et al. 2009).

Although not in the Orobanchaceae, experimental characterization also supports the role of two parasitism genes in *Cuscuta*: 1) a cysteine protease in the stem parasite, *Cuscuta reflexa* (Bleischwitz, et al. 2010), and 2) a *SHOOT MERISTEMLESS-Like* (STM) gene in *C. pentagona* (Alakonya et al. 2012). STM encodes a KNOTTED-like homeobox transcription factor (TF) with

a known role in promoting cytokinin biosynthesis in the shoot apical meristem. Silencing of the *STM* gene in *Cuscuta* by the production of small RNA by host plants resulted in reduced haustorial development and increased growth of infected host plants (Alakonya et al. 2012). Possible strategies by targeting these candidate genes with roles in haustoria development can be developed to control these parasitic weeds.

1.3.3 Parasite and host defense – the arms race

Plant often use two levels of innate immune response mechanism to resist pathogens (Jones and Dangl 2006). The first level of defense responses is called pathogen-triggered immunity (PTI) that recognize the pathogen-associated molecular patterns (PAMPs) or microbe-associated molecular patterns (MAMPs) (Boller and He 2009). PTI includes receptor-like kinases that activate host defense pathways (Boller and Felix 2009; Ronald and Beutler 2010). On the other hand, to evade host surveillance mechanisms, plant pathogens have evolved specific effectors that suppress host defense responses associated with PTI in order to invade host cells (Abramovitch, et al. 2006; Bent and Mackey 2007; Torto-Alalibo, et al. 2009). Correspondingly, host plants have evolved a second level of defense mechanism against these pathogens, called effector-trigger immunity (ETI) (Tameling and Joosten 2007). This includes a second-class of receptor proteins, typically containing a nucleotide-binding site (NBS), and a leucine-rich repeat (LRR) domain, also called R proteins (Takken, et al. 2006; Caplan, et al. 2008). It is now clear that the effect of R protein in certain hosts expressed in response to effectors from parasitic plants, such as in *S. gesneroides*-cowpea (Li and Timko 2009), and *O. crenata*-sunflower (Molinero-Ruiz, et al. 2006; Letousey, et al. 2007) interactions, acts as gene-for-gene resistance.

Parasitic plants are likely to be perceived by hosts as pathogens because their invasions of host vascular tissues are similar to many pathogens. So how can parasitic plants cope with host defense responses? Several hypotheses were suggested to provide potential mechanisms for parasites to overcome host defenses. The first possibility is that the host may fail to recognize the parasite as an alien as the intrusive growth of the haustoria mimics pollen tube growth (Lev-Yadun 2001). A similar hypothesis is that parasitic plants modulate host expression in a way that is similar to nodulation, based on the upregulation of nodulation-related genes in the host (Hiraoka, et al. 2009). A third possibility for failure of the host to recognize the parasite as alien

may be due to strikingly similar defense mechanisms that the parasite and host both use because both are plants. Lines of evidence of compatible hosts failing to respond against parasite attack come from the lack of ROS either from the parasite or host side (Mor, et al. 2008). The last hypothesis is that parasites may repress the host defense responses. Some evidence supports the idea of host recognizing parasitic plants as pathogens. In incompatible or resistant *Striga*-host and *Orobanche*-host interactions, localized cell death of host cells and an HR-like rapid necrosis at the attachment interface were observed (Lane, et al. 1993; Mohamed, et al. 2003; Gurney, et al. 2006). An HR response acted as a mechanism to block parasite invasion. This indicates that the host responds to parasitic plants in a way similar to its response to common plant pathogens. In *Arabidopsis*, expression of defense related genes encoding pathogenesis-related proteins, cell wall reconstruction proteins, and components of jasmonate, ethylene, phenylpropanoid biosynthesis pathways were induced by the attack of *O .(Phelipanche) ramosa* (Dos Santos, et al. 2003). Resistant hosts, compared to susceptible hosts, show increased lignification of cell walls with histological staining (Irving and Cameron 2009). In addition, differentially upregulated transcription levels of genes encoding peroxidase, an enzyme with known roles in cell wall cross-linking, occurred between resistant and susceptible pea against *O. crenata* infection (Perez-de-Luque, et al. 2006). Furthermore, a resistant sunflower differentially upregulated genes involved in ROS detoxification (a methionine synthase, a glutathione S-transferase and a quinone oxidoreductase) to respond to the observed oxidative burst during the incompatible interaction with *O. crenata* (Dos Santos et al. 2003). On the other hand, to establish successful connections, parasitic plants have developed several mechanisms to regulate host defenses against parasite attack. For instance, parasitic plants secrete peroxidases (Antonova and TerBorg 1996) or phenolic compounds that may repress host defense responses (Mayer 2006). The upregulated levels of ABA in hosts by parasitic plants, similar to that in many fungal and bacterial pathogens (Cao, et al. 2011), could contribute to the observed lack of salicylic acid (Hiraoka and Sugimoto 2008) associated with defenses in *S. hermonthica-Sorghum* and *Orobanche-Arabidopsis* interactions (Dos Santos et al. 2003; Griffitts, et al. 2004).

1.4 Evolution of novel traits

1.4.1 Haustorium as a good model to study the evolution of novel traits

The evolution of the haustorium represents a remarkable innovation for parasitic plants, and one that has occurred independently at least 11 times based on the phylogenetic distribution of parasitic plants across all flowering plants (Barkman et al. 2007). Compared to their autotrophic free-living ancestor, parasitic plants show a gain of a novel organ. How a plant transitions to a heterotrophic lifestyle by losing its photosynthetic capability and adapts to the environment has been a mystery. Because haustorium is the novel organ of parasitic plants that differentiates from autotrophic plants, we believe studying genes upregulated in haustorial tissues should allow us to understand how the novel haustorium structure has evolved.

1.4.2 Mechanisms for the origin of novel traits

Throughout evolutionary time, numerous complex adaptations have evolved in organisms. Beetles have evolved horns to combat male competitors (Moczek 2005), birds have evolved wings to fly (Ostrom 1979), and moths have evolved with decorated wings with eyespots to deter predators (Stevens 2005). Adaptation in optimization of shape or phenotype of the organism to best utilize the environment provides an advantage to the organism's fitness (Moczek 2005). The evolution of novel traits is believed to involve multiple ecological, developmental, and genetic mechanisms. At least five genetic mechanisms have been shown to contribute to the evolution of novel traits – regulatory networks, gene duplication, recruitment of pre-existing machinery, positive selection, and horizontal gene transfer.

1.4.3 Gene duplication and regulatory networks (origin of flower)

It has been suggested that novel traits may have evolved by recruiting not only single genes, but also pleiotropic *cis*-regulatory elements, and even network modules (Monteiro and Podlaha 2009). The origin of the flower and its subsequent diversification have contributions from gene duplication (Kramer, et al. 1998) and regulatory networks (Liu, et al. 2010) in the plant MADS-box gene family, which encode a large number of transcription factors that regulate downstream genes controlling for floral shape and floral organ identity (Becker and Theissen 2003). The *Amborella* genome (*Amborella* Genome Project 2013) and the analyses by Jiao et al (2011) revealed an ancient whole genome duplication that happened in the ancestor of all angiosperm lineages which gave rise to many genes that finally functioned in flower development. Genetic studies gave rise to the classic ABC model, identifying roles of classes A, B, and C genes of the MADS-box family in specifying the identities of four floral organs in different whorls of a flower – sepal, petal, stamen, and carpel (Coen and Meyerowitz 1991). Class A genes specify sepal (the first whorl), class A and B genes specify petal (the second whorl), class B and C specify stamen (the third whorl), class C genes specify carpel (the fourth whorl) (Coen and Meyerowitz 1991). Gene duplication in the MADS-box family has been the driving force in the diversification of floral shapes of diverse angiosperm lineages (Kramer et al. 1998). In addition, protein-protein interactions of many MADS-box proteins are often needed before they bind to the regulatory regions of their downstream genes (Liu et al. 2010). Several lines of evidence also show that MADS-box interaction are also required to determine the transition of vegetative organs to floral organs (Honma and Goto 2001; Li, Yu, et al. 2015). Furthermore, the hetero-dimer interaction of MADS-box proteins present in *Amborella* - the earliest surviving branch of angiosperms - but absent in the nonflowering gymnosperms, provides another

mechanism of regulatory network for the origin of flowers in diverse angiosperms (*Amborella* Genome Project 2013).

1.4.4 Homeobox domain-mediated regulation of leaf development

A similar example illustrating transcription factor-mediated regulation of complex plant traits has been demonstrated with another gene, the KNOTTED-LIKE HOMEODOMAIN (*KNOX*) gene. The homologous *Hox* genes in animals control many aspects of development of homologous appendages. Diversification of arthropods with numerous morphological innovations has been attributed to changes in *Hox* genes and their targets (Weatherbee, et al. 1999). In plants, *KNOX* gene expression plays a role in the maintenance of shoot apical meristem and its formation of homo- or hetero-dimers has been shown to determine leaf initiation (Hake, et al. 2004) and the determination of simple or compound leaf (Champagne and Sinha 2004; Piazza, et al. 2005). These lines of evidence collectively support the role of gene duplication and regulatory networks in diversification and evolution of novel complex traits.

1.4.5 Co-option of existing genes and gene family expansion

The development of horns in horned beetles has involved the co-option of the pre-existing appendage patterning genes (Moczek 2005), the mechanism of which also underpinned many other developmental traits, such as insect distal limbs (Panganiban, et al. 1994), the center of butterfly eyespots (Carroll, et al. 1994; Brakefield, et al. 1996). The genome sequences of diverse avian species shed light on the evolution of many traits in birds. The increased gene copy numbers of opsin genes in birds relative to mammals may be associated with enhanced avian vision (Zhang, Li, et al. 2014). Around two-fold increase in copy number in birds compared to

reptiles was observed for β-keratin gene family, a family involved in constructing structural proteins unique to the epidermal appendages of birds and reptiles (Zhang, Li, et al. 2014).

1.4.6 Mutation-driven positive selection

Positive selection has also been shown to play an important role in the evolution of novel traits. The earliest example of positive selection was seen in four non-synonymous substitutions in a hemoglobin gene where one amino acid difference was associated with enhanced affinity to oxygen, and allowed bar-headed geese to fly across the Himalayas (Petschow, et al. 1977). Similarly, recent genetic variations in humans show selective sweeps in two hypoxia-related genes associated with adaptation in the high altitude Himalayas region (Peng, et al. 2011). Another example is shown in honey bees where worker-biased proteins show signatures of positive selection (Harpur, et al. 2014), providing stronger evidence for the role of adaptive evolution in driving worker traits. It is worth mentioning that positive selection as the driving force of evolution was proposed by Darwin in his *Origin of Species* (Darwin 1859); however, Nei's mutation theory focused on mutation as a prime driving force for evolution (Nei 2013a). In his theory, Nei emphasized the importance of mutation, which is the first step that has to occur before selection acts on it (Nei 2013a). Without mutation, positive selection simply cannot exert its role. In this sense, "mutation-driven positive selection" should be a more accurate term than simply "positive selection".

1.4.7 Haustoria origins – the exogenous model (HGT)

The evolution of novel traits has also been impacted by horizontal gene transfer (HGT) (Jain, et al. 2003; Dagan, et al. 2008), especially in the evolution of many prokaryotic genomes in

which large proportions of their genomes are driven by HGT through processes such as transformation, conjugation, and transduction (Bapteste, et al. 2009). Many important traits involve the establishment or expansion of ecological niches such as the acquisition of antibiotic resistance (Davies and Davies 2010) and pesticide degradation (McGowan, et al. 1998) which were both impacted by HGT. The earlier hypothesis by Atsatt proposed a model of haustorial evolution based on horizontal gene transfer (Atsatt 1973). His rationale was that morphologies of haustoria resemble nodules and crown galls; he hypothesized that haustoria resulted from plant responses to endophytic microorganisms that have the ability to invade plant roots. A theory similar to this hypothesis is the endosymbiotic theory in which the eukaryotic organelles evolved from endosymbionts. Similarly, Kuijt also proposed that haustoria originated from mycorrhizal fungi that bridged roots of different plants (Kuijt 1969). Both of their hypotheses seem to propose that the ability of parasitic plants to attack their hosts was acquired by HGT from endosymbiotic bacteria or fungi. Interestingly, several expressed sequences from an endosymbiotic bacterium present in animal-parasitic nematodes and arthropods – *Wolbachia* – were found in a plant-parasitic nematode *R. similis* (Haegeman, et al. 2009), indicating that HGT from a bacterial endosymbioint may be one origin of parasitism in plant-nematodes.

There are many cases in eukaryotic organisms where parasitic or pathogenic capability was enabled by HGTs of bacteria genes. Parasitic nematodes secrete cell wall-degrading enzymes whose sequences are more similar to fungi and bacteria than to animals, indicating a potential HGT origin for these sequences in the nematode (Smant, et al. 1998). There are at least 46 proposed cases of HGTs that have contributed to the colonization of their plant hosts for pathogenic fungi and oomycetes (Soanes and Richards 2014). These HGT genes encode proteins associated with processes involved in invasion, degradation (for instance, beta-galactosidase, involved in the breakdown of hemicellulose and pectin (Zhuang, et al. 2006)), and manipulation of their host (antioxidative enzymes that scavenges ROS from the host (Klotz and Loewen 2003),

cytochrome P450 that break down phytoalexins, which are antifungal toxins (Maloney and VanEtten 1994). In addition, adaptive roles of specific HGT events have been demonstrated in several noteworthy examples. A bacterial hydrolase was horizontally transferred to an insect pest of coffee that allowed its digestion of coffee berry (Acuna, et al. 2012). Horizontal transfer of a chimeric photoreceptor (neochrome) from bryophytes to ferns enabled ferns to adapt to low-light conditions (Li, et al. 2014a).

1.4.8 Haustoria origins – the endogenous model (gene recruitment)

Haustoria, developmentally speaking, are more similar to roots than to any other plant structure. Terminal haustoria develop at the tip of the embryonic radicle, and lateral haustoria develop as extensions from lateral roots or adventitious roots. On the other hand, haustoria have an important role in obtaining nutrients from their host; their role in transferring carbon from the host is analogous to the action of the leaf vein that moves sugar from mesophyll cells into phloem cells of minor veins (Westwood 2013). In this sense, phloem loading activity particularly in obligate parasites are common in both leaf and haustoria.

Research efforts (Aly et al. 2009; Bandaranayake et al. 2010; Alakonya et al. 2012; Bandaranayake et al. 2012) on parasitic plants have identified a set of candidate genes that play a role in haustorium initiation and development, for instance *TvQR1* and *TvQR2* in haustorial initiation. Additionally, *TvPirin* is upregulated by HIFs (or by contact with host roots) and putatively functions as a positive regulator of other genes needed for haustorial development (Bandaranayake et al. 2012). Finally, mannose 6-phosphate reductase (M6PR) in *Phelipanche*, a root holoparasite, was also shown to be involved in parasite metabolism (Aly et al. 2009).

As these genes are also found in autotrophic plants, it is believed that these genes have evolved a role in parasitism by neofunctionalization of existing genes that play roles irrelevant to parasitism. Thus, it is believed that recruiting genes involved in other aspects of plant development could act as one origin of parasitism.

1.5 Horizontal gene transfer

Horizontal gene transfers from endosymbiotic bacteria or invasive microorganisms may provide one mechanism of haustorial origin, but horizontal transfers of genetic material from host plants may play an additional important role in parasite evolution. The intimate contact with host plants allowing the exchange of molecules including nucleic acids, acts as one mechanism resulting in horizontal transfers of mitochondrial genes (Davis, et al. 2005; Mower, et al. 2010; Xi, et al. 2013), plastid genes (Park, et al. 2007b), and nuclear genes (Yoshida, Maruyama, et al. 2010) in parasitic plants. In the following text, we review the currently published research progress on HGT in plants that has been published to date.

1.5.1 Mitochondrial HGT

Mitochondrial genes are particularly subject to horizontal transfer in plants, with the strongest example evidenced by the frequent transfers of mitochondrial DNA up to the whole mitochondrial genome in *Amborella* (Bergthorsson, et al. 2004; Rice, et al. 2013). Quite remarkably, repeated horizontal transfers have been demonstrated in *cox1* (cytochrome oxidase subunit 1) intron, a group I mobile element encoding a self-splicing endonuclease that can catalyze its movement from intron-containing to intron-less genes found primarily in organellar genomes and nuclear rRNA genes (Lambowitz and Belfort 1993). Since its probable first

horizontal landing into the basal angiosperm *Peperomia* from a distant fungal donor (Adams, et al. 1998), its explosive cross-species transfer has invaded increasingly many lineages of flowering plants more than 1,000 times during angiosperm evolution (Cho, et al. 1998). The widespread transfer of mitochondrial genes between distantly related flowering plants include ribosomal protein genes – *rps2*, *rps11*, and respiratory genes – *atp1* (Bergthorsson, et al. 2003). These mitochondrial HGTs result in the recapture of genes lost from functional transfer to a nucleus, the creation of a duplicated copy in the presence of a vertical copy, or a chimeric gene by recombination (Bergthorsson et al. 2003). Several lines of evidence suggest that gene conversion between the parasitic copy and the host copy for mitochondrial HGTs (Archibald and Richards 2010; Mower et al. 2010) may be one mechanism to produce chimeric mitochondrial genes (Hao and Palmer 2009; Hao, et al. 2010).

1.5.2 HGT in parasitic plants

In addition to the mitochondrial HGTs, increasing evidence of HGT have been reported in parasitic plants. This includes the massive transfer of 16 mitochondrial (mt) genes in the parasitic *Rafflesia* (Davis and Wurdack 2004; Nickrent, et al. 2004; Barkman et al. 2007; Xi, et al. 2012a; Xi et al. 2013), and several genes in various parasitic lineages including Apodanthaceae (atp1) (Nickrent et al. 2004; Barkman et al. 2007), *Cuscuta* (Convolvulaceae, 3 mt genes) (Mower et al. 2010), Mitrastemonaceae (combined phylogeny of three mt genes: *atp1*, *cox1*, *matR*) (Nickrent et al. 2004; Barkman et al. 2007). The plastid genome is more "immune" to HGT: despite the widespread transfers of mitochondrial genes in the large *Amborella* genome from various sources, there is no evidence supporting HGT in its plastid genome (Rice et al. 2013). Nevertheless, a significant number of examples have been identified in parasitic plants including 2 plastid genes in Orobanchaceae (Park et al. 2007b; Li, et al. 2013) and 29 plastid

genes in Rafflesiaceae (Xi et al. 2013). These lines of evidence collectively support an identified number of 20 mt genes and 31 plastid genes in lineages of parasitic plants. In addition to the frequent transfer from hosts to parasitic plants, transfers from parasite to host have been reported. These includes two cases of mt transfers from parasitic sandlewood (Santalales) to a fern (Davis et al. 2005) and two mt transfers involving *atp1* gene from parasitic plants Orobanchaceae and Convolvulaceae to the host *Plantago* (Mower, et al. 2004).

The instances of nuclear transfer in parasitic plants are relatively few; so far strong evidence include three nuclear genes in Orobanchaceae from their hosts – one unknown *Striga* gene from grasses (Yoshida et al. 2010), one legume-specific *albumin 1* in *P. aegyptiaca* (Zhang, et al. 2013a) and one Brassicaceae-specific *strictosidine synthase-like* gene in both *P. aegyptiaca* and *Cuscuta* (Zhang, Qi, et al. 2014b). Although Xi et al (2012a) claimed the identification of 47 nuclear HGTs in *Rafflesia*, they lack strong phylogenetic evidence for support.

1.5.3 HGT of non-plant origin

In addition to plant-plant transfers, horizontal transfers involving 57 nuclear genes have been identified in *Physcomitrella* from various donors including bacteria, fungi, and viruses (Yue, et al. 2012). These genes include auxin biosynthesis gene (YUCCA family monooxygenase), and stress-responsive genes (HAD-family hydrolase), and genes involved in metabolism (glutamine synthetase) and nutrient mobilization (subtilase) that are important to plant colonization of land (Yue et al. 2012). Although the phylogenetic trees clearly show the placement of *Physcomitrella* genes as sisters of distant lineages (bacteria for instance), caution needs to be taken especially since there are not many basal land plant genomes in support of sufficient taxon sampling.

1.5.4 Models for HGT

In light of these findings, Wang et al (2014) reviewed four models to explain the mechanisms of HGT in plants. The first model is the "intimate physical contact", as supported by frequent transfers in diverse lineages of parasitic plants. The second model is "mitochondrial fusion", inspired by the widespread transfers of genomic pieces in the giant *Amborella* mitochondria genome (Rice et al. 2013). The third model is the "weak-link model" proposed by Huang (2013) who suggested certain stages of the moss life-cycle such as zygotes, embryos, or spores, may represent a weak link that allows easier access of foreign genes to germline cells. The fourth model is "illegitimate pollination" assuming that pollen grains can germinate on the stigma of another species from which it is otherwise reproductively isolated, allowing the integration of the pollen DNA with egg cells during pollen tube elongation (Wang et al. 2014). This model could explain transfers between closely related species such as gene transfer between different genera of Poaceae (Diao, et al. 2006). The last model is through "transfer agents" such as aphids, bacteria, viruses, or fungi (Gao, et al. 2014), for instance transfers mediated by transformation potent - *Agrobacterium tumefaciens* in plants (Intrieri and Buiatti 2001), or invasive retrotransposons moved by a virus to bridge distantly related species (Roulin, et al. 2008; Roulin, et al. 2009; Gao et al. 2014).

1.5.5 HGT – where do we go from here?

HGT in several species seem to indicate a role in land plant evolution (Wang et al. 2014), with the strongest example from the horizontal acquisition of a neochrome in ferns (Li et al. 2014a) and 57 HGTs of genes with functions related to plant colonization to land in moss (Yue et al. 2012). Parasitic plants are known as an exemplary model for HGT discovery; however,

identification of nuclear HGT is still in its infancy. Considering that there are more than 10 independently evolved lineages of parasitic plants with varying extents of parasitic dependence and structural integration of parasite and host tissue (Barkman et al. 2007), it is worth identifying all nuclear HGTs in these parasitic lineages and exploring whether HGT plays a role in the evolution of parasitism. Important goals for researchers in the future are to examine if a set of homologous genes are shared among multiple parasitic lineages, which will shed light on whether HGT involving particular genes may contribute to a common mechanism of parasitism.

1.6 A transcriptomic and phylogenomic approaches to study parasitic plants

1.6.1 Driving questions for research on parasitic plants

One motivation that drives us to carry out research of parasitic plants is to provide a series of approaches for parasitic weed control (Westwood et al. 2012; Gressel and Joel 2013). In light of existing research advances on parasitic plants, we are also interested in genomic changes that led to the transition to a parasitic lifestyle. In particular, we are focused on two processes of parasite development: haustorial initiation and development. We hope to characterize genes that are important to these two stages, and by relating possible roles of their orthologs in existing studies, we hope to select a handful of candidate genes for functional characterization. In addition, we are interested in understanding what evolutionary forces have been involved in driving parasite evolution. As discussed in the previous context, are there gene duplications either from genome duplications or small-scale duplications in the history of parasitic plants? Increasing examples of HGT have been identified in several lineages of parasitic plants, how extensive has HGT been in the parasitic Orobanchaceae? How frequent and important is HGT to the

development of a parasitic lifestyle? In addition to haustoria related stages, an important feature of obligate parasites is the requirement of host signals (SLs) for germination. Studying the evolution of SLs in parasitic plants may also guide us to develop approaches for parasitic weed control.

1.6.2 Sequencing technologies allow the capture of transcriptomes and genomes

Thanks to the rapid advancement of sequencing technologies, we are able to target the genome and transcriptome of several parasitic plant taxa. Traditional sequencing technology was based on Sanger sequencing, which led to the sequence and map of the human genome in 2001 (Lander, et al. 2001). Later on to mitigate the high cost, laborious prep, and radioactive labeling, a series of affordable next-generation sequencing approaches were developed that included 454 in 2005 (which later was bought by Roche), and Solexa in 2006 which has now been largely abandoned in favor of Illumina (Liu, et al. 2012). The third-generation sequencing technologies are the single-molecule real-time (SMRT) technologies developed by Pacific Bioscience (PacBio) (Rhoads and Au 2015) and Bionano (Antón, et al. 2015). Illumina has unique features of producing high throughput and accurate reads with an average length of 250 bp (up to 500 bp today) (Turner 2014). PacBio is famous for its long fragment molecule up to 40 kb which can bridge repeats in a genome, and works well when combined with Bionaono that produces single molecule maps with an average sequence length up to hundreds of kilobases to megabases (Antón et al. 2015). Error rates vary widely among the different technologies with longer sequences generally having a much higher per base error rate. Efforts on developing algorithms to handle billions of reads from next-generation sequencing technologies have also generated various *de novo* transcriptome assembly softwares including Trinity (Haas, et al. 2013), SOAPdenovo (Xie, et al. 2014), and CLC Workbench software (CLC Bio-Qiagen, Aarhus, Denmark). By breaking

the transcripts into small fragments called reads, one can sequence these molecules and generate a whole catalog of all the expressed genes (including splice variants) in each developmental stage of a plant. By generating a reference assembly (containing all reads from different libraries) and read mapping, one can get each transcript's expression level across all stages. In terms of sequencing genomes, however, a combination of these technologies is needed to complement the shortcomings of each other (Pendleton, et al. 2015). Illumina's advantage of generating short accurate fragments can produce accurate contigs, which can be combined to scaffolds using PacBio to bridge repeats. BioNano's ability to generate even longer fragment can be used to construct a physical map (instead of providing direct sequence information). These efforts could finally lead to downstream analyses including differential expression analyses, gene annotation, and phylogenetic analyses, etc.

1.6.3 The power of phylogenomic approaches (polyploidy (gene duplication) and HGT)

The capture of sequence information for the protein-coding genes in a genome can lead to many phylogenomic analyses. Phylogenomics is a tool to study evolution of genes and gene families using phylogenetic trees. It is based on the fact that all species have a common ancestor; thus studying the phylogeny of each gene can provide inference on species phylogeny. Phylogenies can reveal evolutionary events in a gene family, such as gene duplication, gene loss, rapid rate of evolution, etc.

Polyploidy Plant genomes have a rich history of polyploid events, and most major lineages of flowering plants have undergone one or more whole genome duplication (WGD) events in their history. For instance, the poplar genome (*Populus trichocarpa*) revealed at least two WGDs (Tuskan, et al. 2006); *Vitis* has had a triplication event (1 WGD) (Jaillon, et al. 2007); one or two WGDs have been inferred in the rice genome (Paterson, et al. 2004); the *Striga asiatica* genome

also revealed one WGD after its divergence with its close relative *Mimulus guttatus* (Yoshida et al., submitted)[1]. WGD can also give rise to species diversification, which can provide raw material for the action of selection to facilitate innovations of important traits (Rensing, et al. 2007) and enhanced adaptation to diverse environment conditions (Fawcett, et al. 2009). For instance, tetraploid *Arabidopsis* plants, compared to the diploid ones, exhibited significantly higher levels of potassium uptake and salt tolerance (Chao, et al. 2013). Evidence shows that polyploidy can result in epigenetic reprogramming causing tissue-specific differential expression of duplicate pairs (Adams and Wendel 2005). Fawcett et al (2009) suggested that these changes may contribute to hybrid vigor and increase phenotypic variation allowing rapid adaptation to new ecological niches. By building large-scale phylogenetic trees for all the protein-coding genes in a species' genome, Jiao et al (2011) inferred two ancient WGDs – one in the common ancestor of all flowering plants, the other in the common ancestor of all seed plants (*Amborella* Genome Project 2013; Li, Baniaga, et al. 2015).

HGT Phylogenetic analyses can also reveal HGT events. The use of a phylogenomic approach allowed Xi et al (2013) to claim widespread HGTs in parasitic *Rafflesia* from its host *Tetrastigma*. Although he didn't show clear phylogentic evidence for each tree as strong support, the use of a phylogenomic approach in inferring HGT was clearly implementable. This was done by first selecting a number of sequenced plant genomes, which represents lineage from each clade and also provides enough resolution for HGT inference. Then a species tree can be constructed by using a concatenated matrix containing the single-copy genes across these taxa (Duarte, et al. 2010; Wickett, et al. 2014). Alternatively, one can use coalescent-based methods such as ASTRA to reduce bias from incomplete lineage sorting (Mirarab, et al. 2014). The selected sequenced genomes are used to construct an orthogroup classification so that each gene from HGT focal taxa

[1] Satoko Yoshida, Seuungill Kim, Eric K Wafula et al. 2015. Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. Nature plant (submitted).

can be classified into each orthogroup (gene family) by BLAST (McGinnis and Madden 2004) or Hidden Markov Model (HMM) (Eddy 2011a). Finally phylogenetic trees for each gene are constructed using either a maximum likelihood- or Bayesian-based approach. A vertical gene tree is shown as sisters of genes from its closely related taxa, whereas a gene from HGT focal taxa nested within a distantly related donor clade is inferred as resulting from HGT.

1.6.4 Overview of PPGP

The Parasitic Plant Genome Project (PPGP) was initiated to identify the key genes and evolutionary changes important for the establishment of a parasitic lifestyle (Westwood et al. 2010; Westwood et al. 2012). The PPGP used a transcriptome sequencing approach (RNA-seq) to interrogate multiple stages of the growth and developmental stages of parasitic plants. The study group is the family of Orobanchaceae, the only family that has a complete spectrum of parasitic capabilities – including one autotrophic free-living *Lindenbergia philippensis*, and three representative parasites *Triphysaria versicolor*, *Striga hermonthica*, and *Phelipanche aegyptiaca*. The stages include important stages of haustoria initiation and host attachment and invasion (stage 3 and stage 4). By applying a *de novo* transcriptome assembly approach (a combined assembly of reads from all stages of a parasitic plant), we constructed a complete catalog of all the transcribed sequences from each stage of each parasite. Followed by read mapping, differential expression, and clustering analyses, we identified genes that are specific to haustoria development. In light of a phylogenomic approach, gene trees containing all the transcribed sequences from all the four Orobanchaceae species were built to infer gene duplications and HGT. These findings were combined with expression data and gene annotations to infer the possible roles of gene duplication and HGT in parasite evolution. Candidate genes were selected for functional characterization using RNAi in the parasitic plant. As small-interfering RNAs

targeting parasite genes transformed into the host can move into the parasite through haustorium – leading to successful silencing of genes in parasitic *T. versicolor* (*GUS*) (Tomilov, et al. 2008), *P. aegyptiaca* (*M6PR* (Aly et al. 2009), *CCD7* and *CCD8* (Aly, et al. 2014)), and dodder (*STM*) (Alakonya et al. 2012), the approach using host-induced gene silencing (HIGS) can also be applied to engineer parasite-resistant crops.

Chapter 2

Identification of parasitism genes and the origin of the haustorium[2]

[2] This chapter has been published as:
Zhenzhen Yang, Eric K. Wafula, Loren A. Honaas, Huiting Zhang, Malay Das, Monica Fernández-Aparicio, Kan Huang, Pradeepa C.G. Bandaranayake, Biao Wu, Joshua P. Der, Christopher R. Clarke, Paula Ralph, Lena Landherr, Naomi S. Altman, Michael P. Timko, John I. Yoder, James H. Westwood, and Claude W. dePamphilis (2014) *Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evo*l. doi: 10.1093/molbev/msu343

2.1 Introduction to novel traits

Throughout evolutionary history, organisms have evolved a variety of sophisticated novel traits for survival and reproduction. For instance, insects have evolved wings for flying, and plants have evolved different patterns of floral shapes and colors to maximize the attraction of insects and other animals for pollination. The origin of such novel traits has been of longstanding interest to evolutionary biologists, and a wide range of approaches has been used to gain insights into the origin of specific traits. For example, examination of the sensory functions of cilia, the secretory structure of sponges, an early diverging group of multicellular animals, provided insights into the origin of the sensory system of metazoans (Ludeman, et al. 2014). Phylogenetic histories of genes known to be involved in eye development and phototransduction revealed that a greater variety of eye types as found in pancrustacean arthropods, appeared to be associated with a higher rate of gene duplication (Rivera, et al. 2010). Mutation and gene duplication have played an important role in generating complex pathways for refined eye development (Gehring 2011; Nei 2013b), which resulted in many different eye types including the camera eye, the compound eye and the mirror eye (Salvini-Plawen L 1961). The complete genome analysis of the basal angiosperm *Amborella trichopoda*, the sister species to all other extant flowering plants, revealed that a whole genome duplication led to the creation of many novel genes and functions associated with floral development and evolution, ultimately contributing to the diversification of flowering plants (*Amborella* Genome Project 2013).

As seen in the above case studies, gene duplication is frequently associated with the evolution of novel functions (Stephens 1951; Nei 1969; Ohno 1970; Kaessmann 2010; Liberles, et al. 2010). The most extensively documented proposal for the evolution of novel gene function is the classic gene duplication model proposed by Ohno (Ohno 1970) and extended by Force, Lynch, and many others (Force, et al. 1999; Lynch and Conery 2000; Tirosh and Barkai 2007; Liberles et al. 2010). Following gene duplication, one copy may retain its original function, while the other copy diverges, and can have a variety of different fates, including pseudogenization (Lynch and Conery 2000), hypofunctionalization (Duarte, et al. 2006), subfunctionalization, or neofunctionalization. Subfunctionalization is due to complementary loss of some of the functional attributes that are initially shared by the new paralogs following duplication, while neofunctionalization can occur when one of the paralogs evolves a new expression pattern or sequence attribute and acquires a new function (Force et al. 1999; Lynch and Conery 2000; Tirosh and Barkai 2007; Liberles et al. 2010). Subneofunctionalization was proposed to describe

processes that involved both (He and Zhang 2005). Polyploidy duplicates all of the genes in the genome at once (Otto and Whitton 2000), providing ample opportunities for the function of paralogous gene copies to diverge (Crow and Wagner 2006; Nei 2013b), especially in plants, where both angiosperms (Jiao et al. 2011; *Amborella* Genome Project 2013) and seed plants (Jiao et al. 2011) have been hypothesized to be ancestrally polyploid, and a large number of more recent polyploidy events have been detected (Schlueter, et al. 2004; Cui, et al. 2006; Soltis, et al. 2009; Jiao, et al. 2012; Vanneste, et al. 2013). Thus, novel gene creation through single gene duplications, large-scale genome duplication, and neofunctionalization may all play significant roles in the origin of a novel function.

For plants, the evolution of parasitism is one of the most extraordinary examples of evolution of novel traits, as parasitic plants have evolved the ability to form a connection that allows it to feed off plants of other species, allowing some parasites to completely abandon photosynthesis, one of the hallmarks of life for most plants. Parasitism is enabled by specialized feeding structures known as haustoria (Kuijt 1969; Heide-Jørgensen 2013b), which have evolved independently at least 11 times in angiosperm evolution (Barkman et al. 2007; Westwood et al. 2010). Most haustorial parasitic plants invade host roots, while some are able to form haustorial connections with stems, and rarely, leaves.

The origin of parasitism in plants has been proposed to follow two general mechanisms. The first considers the striking morphological similarity between some parasitic plant haustoria, root nodules, and crown galls; it was thus proposed that parasites may have evolved through endophytic association or horizontal gene transfer of genes from bacteria or other microorganisms that could confer parasitic ability (Atsatt 1973). The second mechanism, termed the endogenous model (Bandaranayake and Yoder 2013b), was that parasitic functions may have evolved through neofunctionalization from plant genes encoding nonparasitic functions. These mechanisms are not necessarily mutually exclusive, and both may have been important to the evolution of parasitism.

In this study, we have focused on seeking evidence relevant to the endogenous model for the origin of parasitism. We utilized differential expression analysis and expression clustering to identify upregulated genes associated with haustorium initiation, development, and physiology. Through identification of a core set of parasitism genes shared by multiple species of parasites, our results also shed light on the evolutionary mechanism(s) that led to the origin of the haustorium in the Orobanchaceae. As the haustorium is a novel structure at the core of the

parasitic process, comparative analysis of the genes and gene expression patterns of both parasitic and nonparasitic plants enabled us to propose its genetic origins.

2.2 Results

2.2.1 Assembly statistics and coverage

For each of the three parasitic plants in this study (*Triphysaria versicolor*, *Striga hermonthica*, and *Phelipanche aegyptiaca*), we generated 11 to 14 stage-specific libraries (Westwood et al. 2012), plus additional whole-plant normalized libraries using RNA from all developmental stages in each species (Figure 2-1). Additionally, a whole-plant normalized library of *Lindenbergia philippensis* was sequenced to represent the nonparasitic sister lineage of the parasites. A grand total of 2,995,494,710 Illumina reads and 3,153,353 Roche 454 GS-FLX reads were generated. Hybrid assemblies combining all sequencing data for each species resulted in unigene numbers ranging from 117,470 in *Striga* to 131,173 in *Triphysaria* (table 2-1). Average unigene length varied between 581 bp (*Triphysaria*) and 745 bp (*Striga*), while average N50 lengths ranged from 789 bp to 1183 bp with the N50 unigene counts ranging from 21,356 to 24,729. To evaluate the completeness of our transcriptome sequence datasets, we examined the frequency of capture of three known conserved sets of plant genes in the transcriptome assemblies, namely the universally conserved orthologs (UCOs) (Kozik, et al. 2008; Der, et al. 2011; Williams, et al. 2014), conserved single copy genes from COSII (Fulton, et al. 2002; Wu, et al. 2006; Williams et al. 2014) and the set of conserved single copy genes in PlantTribes2 (Wall, et al. 2008) (http://fgp.bio.psu.edu/tribedb/10_genomes/). The UCO list was obtained from the Compositae genome project (http://compgenomics.ucdavis.edu/compositae_reference.htp) and COSII gene list was obtained from SolGenomics (http://solgenomics.net/documents/markers/cosii.xls). The single copy gene list containing 970 single copy orthogroups from PlantTribes2.0 (http://fgp.bio.psu.edu/tribedb/10_genomes/) were identified as single copy in the seven angiosperm genomes included in the classification: *Arabidopsis thaliana* Columbia (version 7), *Carica papaya* (version 1), *Populus trichocarpa* (version 1), *Medicago truncatula* (version 1), *Oryza sativa* (version 5), *Sorghum bicolor* (version 1) and *Vitis vinifera* (version 1). The *Arabidopsis thaliana* proteins from each of

the three conserved single copy gene lists was used as the query in tblastn search of each transcriptome assembly. A gene was considered detected if it returned a hit with an E-value smaller than 1e-10 and at least 30 amino acids long. Results from this analysis shown in table 2-2 indicate that gene coverage ranged from at least 90% in *Phelipanche* (PlantTribes single copy analysis) to 100% (UCO analysis) in *Triphysaria* combined assemblies. These results suggested that our assemblies have excellent gene coverage and are very likely to capture the large majority of the expressed genes in a transcriptome. Additionally, to validate the accuracy of the *de novo* transcriptome assemblies, we used RT-PCR to amplify a total of 33 contigs spanning a range of assembly sizes in the three parasitic species. The estimated sizes from the amplified cDNAs agree well with the expected sizes from the *de novo* transcriptome assemblies (R^2 for the three species of *Triphysaria*, *Striga*, and *Phelipanche* range from 0.973 to 0.999), suggesting a high degree of accuracy in the *de novo* assemblies (supplementary fig. S1) (Yang, et al. 2015). Seven of the *Triphysaria* sequences were selected for validation sequencing and matched precisely the predicted length of the contig and differed from the reference assembly by at most a few SNPs, as would be expected for allelic variants in an outcrossing species. To conclude, we have produced high-quality large scale transcriptome assemblies that serve as a valuable resource for comparative studies of parasitic plant gene content and expression.

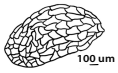
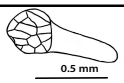
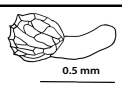
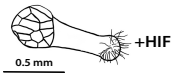
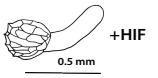
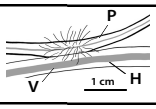
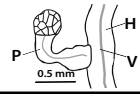
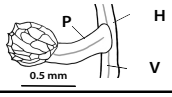
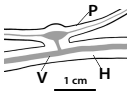
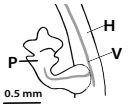
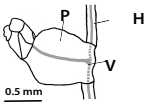
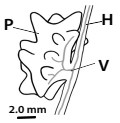
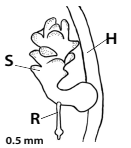
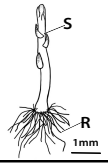
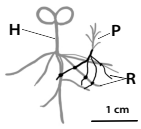
	Description	*Triphysaria versicolor*	*Striga hermonthica*	*Phelipanche aegyptiaca*
Stage 0	Imbibed seed			
Stage 1	Roots of germinated seedlings (*Triphysaria)* or germinated seedlings (after exposure to GR24 in *Striga* and *Phelipanche*)			
Stage 2	Roots of germinated seedlings (*Triphyaria*) or germinated seedlings (*Striga* and *Phelipanche*) after exposure to host roots (*Triphysaria* and *Phelipanche*) or DMBQ (*Striga*)			
Stage 3	Haustoria attached to host roots; penetration stages, pre-vascular connection (~ 48hrs)			
Stage 4.1	Haustoria attached to host roots; penetration stages after vascular connection (~72 hrs in *Striga* and *Phelipanche* and ~120 hrs in *Triphysaria)*			
Stage 4.2	Spider stage	N/A	N/A	
Stage 5	Late post-attachment stage from below ground plants 5.1 Pre-emerged shoots (S) from soil 5.2 Roots (R) on pre-emerged shoots	N/A		
Stage 6.1	Vegetative structures; leaves/stems			
Stage 6.2	Reproductive structures; floral buds (up through anthesis)			
Stage 6.3	Roots in late post-attachment		N/A	N/A

Figure **2-1**. An illustration of stages of each parasitic plant used in the Parasitic Plant Genome Project (Westwood et al. 2012) in this study. The sketches were contributed by Huiting Zhang. Drawings are based on original photographs, as shown in Nickrent et al. (1979), Musselman and Hepper (1986), Zhang (1988), and Rumsey and Jury (1991). Additional sequences from the parasite-host interface (Honaas, et al. 2013) were also used to study haustorial-specific gene

expression (Stage 4). Abbreviations: H (host); P (parasite); V (vasculature); R (root); S (shoot); HIF (haustorium inducing factor); N/A (not applicable).

Table **2-1**.Transcriptome assembly statistics in the post-processed combined assembly for each study species.

Species	Assembly ID	Number of contigs	Assembly size (Mbp)	Number of N50 contigs	N50 contig length (bp)	Mean contig length (bp)
Triphysaria versicolor	TrVeBC3	131,173	76.20	24,729	789	580.91
Striga hermonthica	StHeBC3	117,470	87.53	21,356	1,183	745.17
Phelipanche aegyptiaca	PhAeBC5	129,450	83.80	21,552	1,010	643.48

Table **2-2**.Transcriptome gene capture statistics in three parasitic species.

Gene set	Total	TrVeBC3	TrVeBC3 proportion (%)	StHeBC3	StHeBC3 proportion (%)	PhAeBC5	PhAeBC5 proportion (%)
COSII single copy	220	216	98.18	214	97.27	201	91.36
PlantTribes2.0 single copy	970	949	97.84	952	98.14	869	89.59
UCO	357	357	100.00	356	99.72	354	99.16

2.2.2 Validation of genes with known roles in Orobanchaceae parasitism

We validated the expression data from RNA-seq using expression profiles of two genes that are known to play a role in parasitism: *TvQR1* (Bandaranayake et al. 2010; Ngo et al. 2013) and *TvPirin* (Matvienko, et al. 2001; Bandaranayake et al. 2012; Ngo et al. 2013). Both genes are upregulated in *Triphysaria* roots exposed to the HIF (quinone 2,6 dimethoxy-1,4-benzoquinone; DMBQ) (stage 2) compared to roots without exposure (stage 1). All BLASTn alignments of the *TvQR1* gene with an E-value cutoff of e-10 or smaller were used to construct the putative full-

length transcript representing *TvQR1*. The expression level of *TvQR1* in each stage was calculated as the combined expression of all the unigene hits within TrVeBC3_12199 trinity component that contributed to the construction of the full-length reference (TrVeBC3_12199.1 to TrVeBC3_12199.9). For profiling gene expression we also used three libraries made from host-parasite interfaces of haustoria from the three parasitic plants (Honaas 2013). The interface tissues from the three parasitic plants were targeted by a laser-capture microdissection approach (Honaas 2013; Honaas et al. 2013). Reads from stage 4 interface libraries were mapped onto the combined assemblies to quantify the gene expression in the interface. The RNA-Seq data for the gene *TvQR1* showed high and specific expression for root tissue (stage 1 and stage 2) but low expression levels in other tissues (haustoria, seed, and above ground tissue). When root tissue was treated with DMBQ (stage 2), the expression of *TvQR1* increased relative to roots without any treatment (stage 1) (Figure 2-2), which is consistent with results obtained in previous studies (Bandaranayake et al. 2010). These results confirm the expected expression for *TvQR1* (Figure 2-2).

The only significant hit for gene *TvPirin* in the assembly was contig TrVeBC3_1063.1, which included the full-length CDS and 5' and 3' UTR regions. As expected (Matvienko et al. 2001; Bandaranayake et al. 2012), this gene showed the highest expression in stage 1 and stage 2 (root tissue), with the expression in stage 2 (root treated with DMBQ) higher than stage 1 (untreated root) (Figure 2-2). Post-attachment root tissue (6.3) also showed relatively high expression levels for *TvPirin*, suggesting that this gene is highly expressed in roots. Given the consistency in expression validation as well as assembly validation (supplementary Figure S1) (Yang et al. 2015) , our RNA-Seq assemblies should be able to provide good estimates of gene expression in the species within this study.

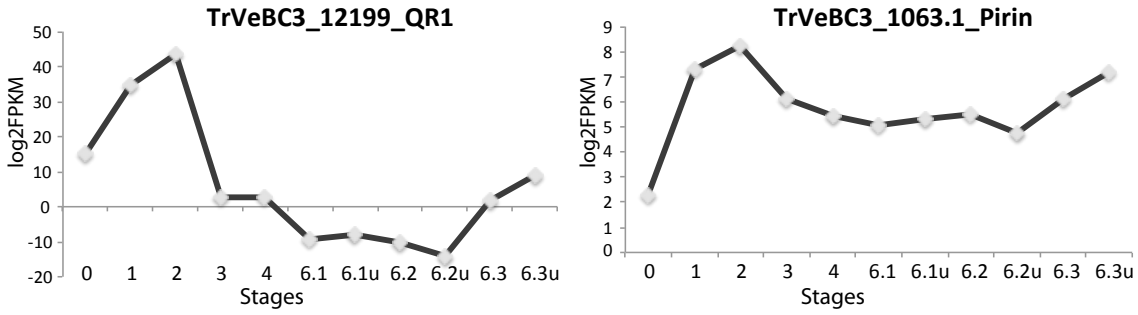


Figure **2-2**. Gene expression profiles from RNA-seq data for two previously characterized parasitism genes in *Triphysaria* (*QR1*, left, and *Pirin*, right). The two genes were shown to be upregulated in stage 2 relative to stage 1 by RT-PCR, which was confirmed by RNA-Seq data.

The organ/stage of the parasite sampled is shown along the x-axis. Labels on the X-axis refer to stage (see Figure 2-1) and 'u' means that this facultative parasite was growing 'unattached' to any host (for instance, 6.1u means unattached stage 6.1). The y-axis gives expression values as fragments per kb per million reads (FPKM) on a log2 scale.

2.2.3 Differential gene expression and clustering to identify haustorial genes

A differential expression (DE) analysis for the common stages present in all three parasitic plants (stages 0, 1, 2, 3, 4, 6.1 and 6.2) was performed to identify differential expression patterns for any pairwise comparison among the seven stages. Next, we conducted two clustering analyses using K-means clustering and self organizing maps (SOM clustering) to identify clusters of coexpressed genes with high expression in post-attachment haustorial stages (3 and/or 4) for each parasitic plant. Clusters of coexpressed genes that exhibited significantly higher gene expression in post-attachment haustorial stages were extracted from the K-means cluster analysis for each species. We refer to these upregulated genes in post-attachment haustorial stages as "haustorial genes". A boxplot and expression heat map for each cluster of haustorial genes in each species was used to visualize the specific expression patterns for each of the parasitic plants (Figure 2-3A & B). DE analyses and clustering approaches were performed for expression of both unigenes and a more inclusive putative transcript definition, the "component-orthogroup" (supplementary data 1, 2). The latter is defined as a representative sequence for a Trinity component containing all associated unigenes (e.g., splice forms, alleles, subassemblies), so long as they are assigned to the same orthogroup (i.e., clusters of homologous genes representing narrowly defined gene lineages) in the gene family classification used by the (*Amborella* Genome Project 2013). The unigenes and component-orthogroups identified by SOM clustering are shown in the supplementary section (supplementary data 3).

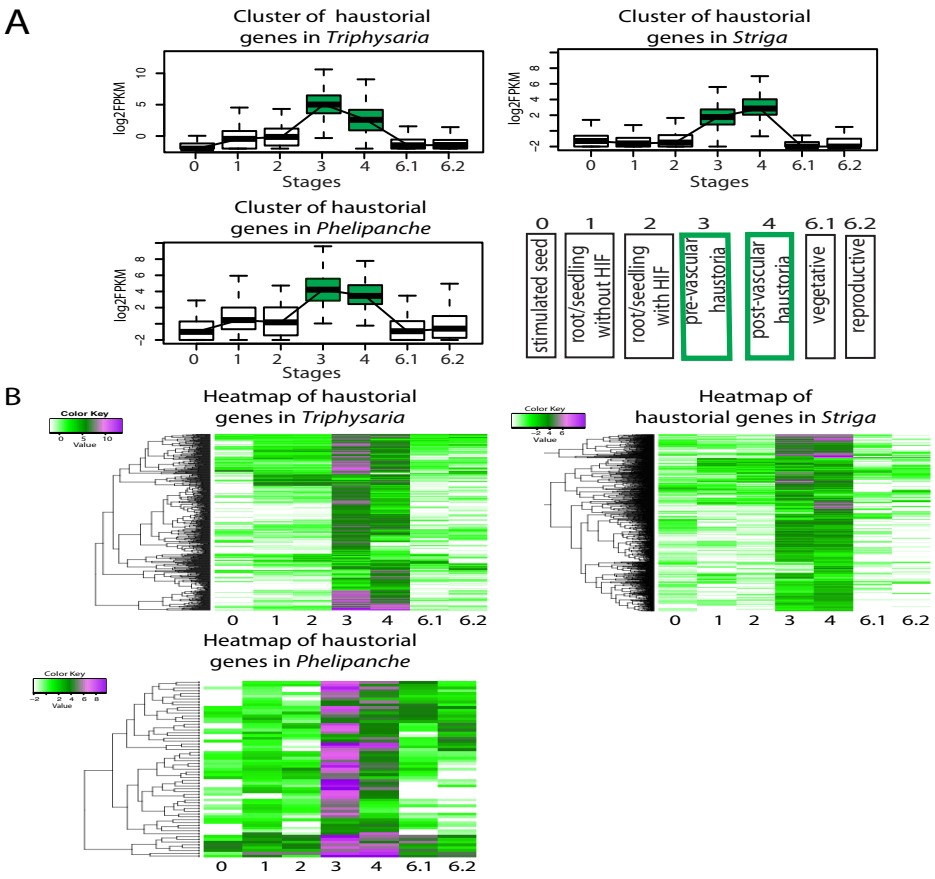


Figure **2-3**. Gene expression clustering (A) and heatmap (B) of upregulated genes in post attachment haustorial stages 3 & 4 ("haustorial genes") in parasitic Orobanchaceae. (A) One cluster of highest expression in post attachment haustorial stages with K-means clustering in each species (511 in *Triphysaria*, 958 in *Striga* and 126 in *Phelipanche*). Expression in each stage is represented by a boxplot. The upper whisker of the boxplot indicates the highest expression value for features within each cluster, the lower whisker, the lowest expression value, and the middle line, the median expression. The upper and lower edges of the box represent the 75th and 25th percentile, respectively. Expression of genes in the post attachment haustorial stages is highlighted in green. A description of the focal stages is shown on the lower right. (B) - Gene expression heat map of component-orthogroups with upregulated expression in post attachment haustorial (stage 3 and/or 4) stages identified by K-means and hierarchical clustering in *Triphysaria*, *Striga*, and *Phelipanche*. The color-intensity in the heat map represents expression value (log2FPKM).

Genes that are differentially upregulated in developmentally similar stages of haustorial development, and are evolutionarily conserved across species, are likely to play an important role in parasitism. We examined three species in Orobanchaceae that exhibit varying levels of host dependence and photosynthetic ability. These species serve as divergent biological replicates; shared gene sequences that show conserved upregulation in parasitic structures are likely to be important to the parasitic process. We used orthogroups to associate homologous (and putatively orthologous) genes across the Orobanchaceae species in this study (see Materials and Methods, supplementary data 4, 5). The final list of candidate haustorial genes was defined as the union of orthogroups represented by upregulated unigenes or component-orthogroups from each species. We also identified unique orthogroups and orthogroups present in only two of the three species. Both K-means clustering and SOM clustering were used to identify genes with high and specific expression after attachment to a host (supplementary data 3, 6). As a result, we identified 185 orthogroups that contained genes (874 unigenes and 488 component-orthogroups) highly expressed of at least two of the species (Figure 2-6). Forty orthogroups were identified that show their highest level of expression during haustorial development of all three parasitic plants. It is important to note that most of the two-way shared haustorial genes are likely to represent a true set of haustorial genes, because at least 70% of these two-way shared orthogroups also contain genes with increased expression in haustorial stages in the third species (supplementary data 7).

2.2.4 Shared haustorial genes are enriched for proteolysis and extracellular region localization

To determine whether the haustorial genes are enriched for specific molecular processes or biochemical pathways, we took the highly expressed unigenes belonging to an orthogroup shared by at least two species and aligned them with BLASTx to the TAIR database (Lamesch, et al. 2012). Best hits in *Arabidopsis* (E-value ≤ e-10) were used as the input for a DAVID enrichment analysis (Huang, et al. 2009) for enriched Pfam domains, GO molecular functions (MFs), biological processes (BPs), and cellular components (CCs) (supplementary data 8). All three parasites share enrichment for the Pfam term serine carboxypeptidase. In addition, the haustorial genes in *Triphysaria* and *Striga* are significantly enriched for eukaryotic aspartyl protease, peroxidase and leucine-rich repeat N-terminal domains (table 2-3, supplementary data 3 and 8). There was a similar high level of enrichment for eukaryotic aspartyl protease and leucine-

rich repeat N-terminal domains in *Phelipanche*, but this was not significant after correcting the P-value for multiple testing. The smaller number of haustorial genes in *Phelipanche* limited the power to detect significantly enriched terms. Finally, pectate lyase was significantly enriched among shared genes in *Triphysaria*.

Table **2-3**. Enriched Pfam domains in the shared set of haustorial unigenes identified by either K-means or SOM clustering in *Triphysaria, Striga and Phelipanche*. Significance levels for category enrichment relative to background are given as Bonferroni-adjusted P-values. NA means enrichment information for the particular term is not identified by the test, NS means non-significant.

Pfam Term	***Triphysaria***		***Striga***		***Phelipanche***	
	Fold change	Bonferroni-adjusted P-value	Fold change	Bonferroni-adjusted P-value	Fold change	Bonferroni-adjusted P-value
Serine carboxypeptidase	24.0	5.08E-07*	24.0	4.26E-08*	39.1	3.58E-05*
Eukaryotic aspartyl protease	39.6	2.64E-06*	0.7	9.94E-08*	41.4	NS
Peroxidase	22.4	7.56E-11*	14.0	3.80E-05*	NA	NA
Leucine rich repeat N-terminal domain_2	6.1	1.67E-02*	7.4	1.07E-04*	6.7	NS
FAD_binding_4	16.1	3.85E-02*	23.1	6.99E-06*	NA	NA
Pectate lyase	41.2	2.02E-06*	5.9	NS	NA	NA

The largest GO BP category in terms of the number of genes (supplementary data 8) is "proteolysis", represented by genes encoding aspartyl protease or serine-type peptidase and subtilase, followed by oxidation-reduction processes, such as peroxidases, and protein phosphorylation, such as kinases. There are also two genes involved in transport activity; one is an oligopeptide transporter and the other is a glutamate-receptor protein. Moreover, six genes

involved in cell wall modification were identified, including three genes encoding pectate lyase or pectate lyase-like proteins, one encoding pectin methylesterase inhibitor, one encoding cellulase, and one encoding carbohydrate-binding X8 protein. GO MF terms such as “serine-type peptidase activity” and “aspartic-type endopeptidase activity” were significantly enriched (table 2-4). The KEGG pathway terms “phenylalanine metabolism” and “methane metabolism” and “phenylpropanoid biosynthesis” were also significantly enriched in the haustorial upregulated gene set. In addition, other enriched GO terms, such as “cellular response to hydrogen peroxide” and “cellular response to reactive oxygen species (ROS)”, support the suggestion by Torres et al (2006) that ROS may be an important signaling intermediate during the parasitic plant-host plant interaction.

Table **2-4**. Enriched GO cellular component (GO-CC), biological process (GO-BP) and molecular function (GO-MF), KEGG pathway, and tissue expression terms among shared set of haustorial unigenes identified by either K-means or SOM clustering in *Triphysaria, Striga,* and *Phelipanche*. Significance levels for category enrichment relative to background are given as Bonferroni-adjusted P-values. NA means enrichment information for the particular term was not identified by the test, NS means non-significant. P-value less than 0.05 is marked with an asterisk.

Term	***Triphysaria***		***Striga***		***Phelipanche***	
	Fold change	Bonferroni-adjusted P-value	Fold change	Bonferroni-adjusted P-value	Fold change	Bonferroni-adjusted P-value
GO-CC: external encapsulating structure	5.3	3.65E-12*	5.1	1.65E-12*	3.6	NS
GO-CC: cell wall	5.2	1.69E-11*	5.0	7.09E-12*	3.7	NS
GO-CC: extracellular region	3.5	4.53E-11*	3.3	3.97E-10*	2.8	NS
GO-BP: proteolysis	3.4	1.92E-06*	4.4	7.51E-14*	4.9	3.88E-05*
GO-BP: cellular response to hydrogen peroxide	18.0	1.87E-09*	10.2	4.18E-03*	NA	NA
GO-BP: cellular response to reactive oxygen species	16.1	7.26E-09*	9.1	8.65E-03*	NA	NA
GO-MF: serine-type peptidase activity	13.3	4.35E-13*	15.8	9.14E-21*	21.9	3.01E-10*
GO-MF: aspartic-type endopeptidase activity	15.0	2.59E-06*	18.3	7.31E-11*	15.7	NS
GO-MF: electron carrier activity	3.6	6.71E-04*	2.8	4.98E-02*	NA	NA
KEGG: Phenylalanine metabolism	14.2	8.68E-10*	10.6	7.95E-06*	NA	NA
KEGG: Methane metabolism	14.0	9.99E-10*	10.5	8.77E-06*	NA	NA
KEGG: Phenylpropanoid biosynthesis	10.9	1.62E-08*	8.2	6.07E-05*	NA	NA
Tissue_specificity: Specifically expressed in root cap cells	49.7	7.06E-01 (NS)	47.2	7.68E-01 (NS)	NA	NA
Tissue_specificity: Expressed in flowers, but not in leaves	33.2	8.40E-01 (NS)	NA	NA	NA	NA

Significant enrichment of genes with the GO CC category "external encapsulating structure", "cell wall", and "extracellular regions" were found in both *Triphysaria* and *Striga* (table 2-4). A concordant pattern of enrichment was observed for these categories in *Phelipanche*, though, like the domain analysis, the conservative test, corrected for multiple comparisons, did not find these enrichments significant in this species. Future research with experimental evidence is needed to determine if the extracellular predictions for these candidate haustorial genes hold true or not, and whether the proteins they encode affect the parasite-host interactions.

2.2.5 Two examples illustrating haustorial gene expression evolution

To further understand the evolutionary origin of some of these haustorial-related genes, we examined patterns of gene expression in orthologs in three nonparasitic model plant or crop species: thale cress (*Arabidopsis thaliana*), barrel medic (*Medicago truncatula*), and tomato (*Solanum lycopersicum*). To illustrate this approach, we present a detailed analysis of the pectate lyase and peroxidase gene families, which were identified by the presence of enriched Pfam domains in *Triphysaria* (pectate lyase) or in both *Triphysaria* and *Striga* (peroxidase). Following an early, possibly angiosperm-wide duplication in the pectate lyase gene family (fig 2-4A), a subsequent gene duplication gave rise to two paralogous gene lineages shared by *Mimulus* and members of Orobanchaceae. One paralogous gene in parasitic Orobanchaceae and their orthologs in *Arabidopsis* and tomato show principal expression levels in floral tissues (supplementary data 21; (Goda, et al. 2004; Sun and van Nocker 2010)); the other paralog took on abundant expression in the haustorium of parasitic Orobanchaceae (Figure 2-4A). The more recent gene duplication in a common ancestor of *Mimulus* and Orobanchaceae gave rise to two *Striga* genes, one having peak expression in haustorial tissue and the other in flower (Figure 2-4A). Conservation of principal gene expression in floral tissues of non-parasites and the maintenance of a floral-expressed ortholog in *Striga* strongly suggests that the haustorial expression of this gene was co-opted from an ancestral gene acting in flowers, and was recruited to haustoria following gene duplication through regulatory neofunctionalization. Alternatively, because the ancestral gene may have been expressed in both floral tissue and root, a two-step process in which the ancestral gene first subfunctionalized and then shifted to haustoria would be an example of subneofunctionalization (He and Zhang 2005).

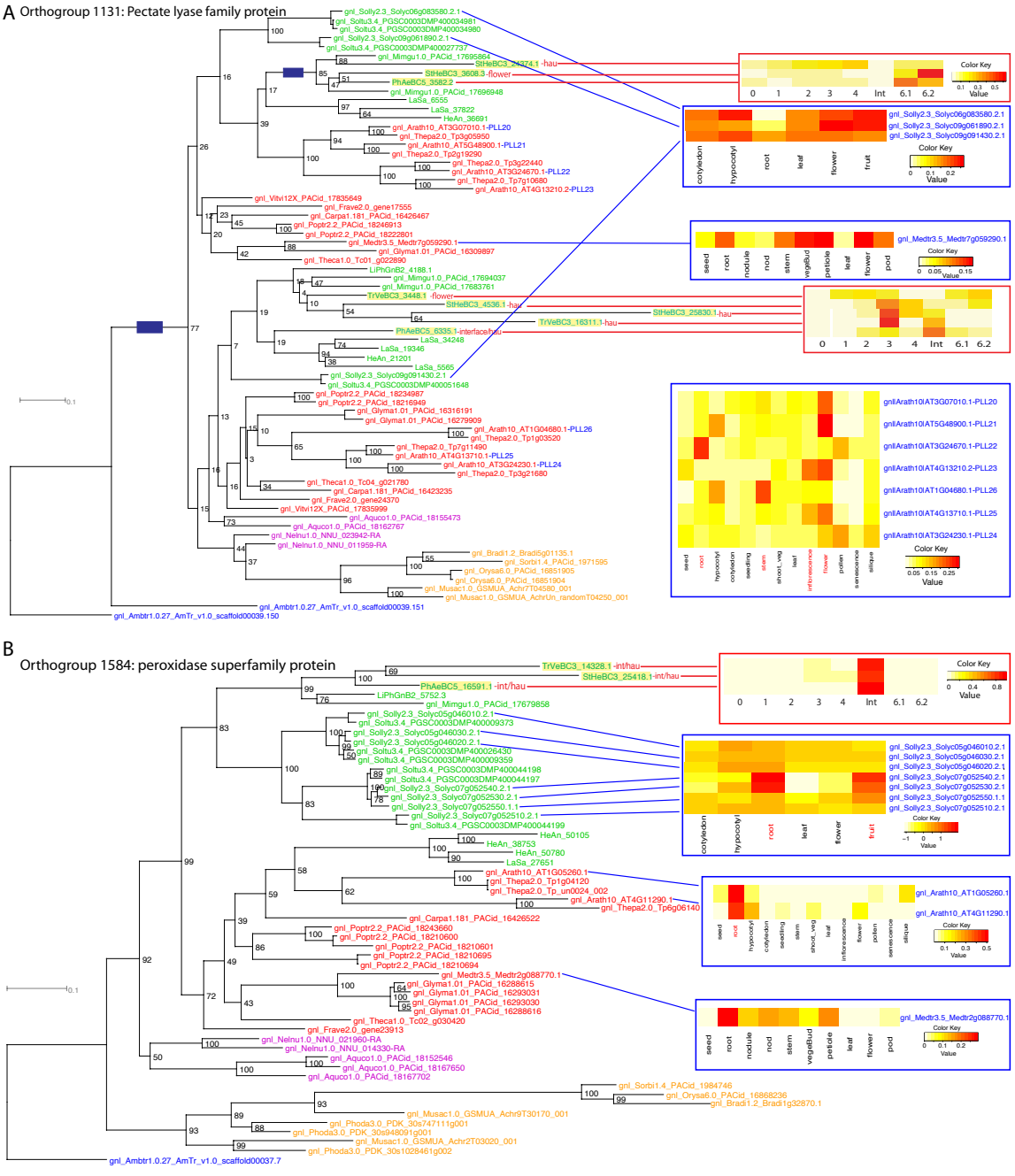


Figure **2-4**. Gene family phylogeny and gene expression profile of two orthogroups showing co-option of haustorial genes from flower and root. A) shift of gene expression from flower to haustorium following gene duplication, and B) shift of gene expression from root to haustorium without gene duplication. Gene duplication events relevant to the origin of parasite genes are shown on the tree with blue rectangular bars. Sequence names are color-coded to represent different lineages: basal angiosperms (blue), monocots (yellow), basal eudicots (purple), rosids

(red) and asterids (green). Parasite genes are highlighted with yellow background and green foreground. The expression of parasite genes and nonparasite genes in *Arabidopsis*, *Medicago*, and Tomato are shown using heat maps. Green and red lines are connecting genes from the phylogeny to the heatmap for nonparasitic genes (except in *Arabidopsis* orthogroup 1131 where genes are labeled as PLLs) and parasitic genes. The tissue with the highest expression was labeled in red. The color intensity in heatmaps refers to expression measurements with an RNA-Seq approach in parasites and tomato, and with microarrays in *Arabidopsis* and *Medicago*. Int or int means "interface" tissue of haustoria (~ stage 4) (Honaas 2013), and hau means "haustoria".

In contrast to the pectate lyase gene, a peroxidase gene family shows a shift of gene expression from roots in all related nonparasitic model species to haustorial tissue of parasitic plants without gene duplication (Figure 2-5B). The peroxidase gene highly expressed in roots of *Arabidopsis* was characterized to be involved in the production of ROS (Kim, et al. 2010), stress response (Llorente, et al. 2002) and pathogenic responses (Ascencio-Ibanez, et al. 2008).

2.2.6 Identifying haustorial initiation genes and "parasitism genes"

Genes that are upregulated in response to the haustorial initiation factor (HIF) DMBQ might also play an important role in parasitism. For example, *TvQR1*, which is upregulated following DMBQ exposure, encodes a quinone reductase that acts early in the HIF signaling pathway (Bandaranayake et al. 2010). Stage 1 in the tiny-seeded *Striga* and *Phelipanche* (Westwood, 2012) consists of whole germinated seedlings, while in the much larger-seeded species (*Triphysaria*) stage 1 is comprised of excised radicles of germinated seedlings. Upon treatment of stage 1 seedlings or roots with DMBQ or a host root, the plants progress to stage 2 (haustorial initiation). DESeq was used to identify "haustorial initiation genes" (HIGs), defined here as unigenes and component-orthogroups that have significantly higher gene expression in stage 2 compared to stage 1 (supplementary data 9).

Triphysaria HIGs were enriched for GO MF terms such as "magnesium ion binding", "calcium ion binding", "calcium-transporting ATPase activity", "calcium ion transmembrane transporter activity", and "cation-transporting ATPase activity" (supplementary data 10). The co-

occurrence of putative functions of calcium ion transport and ion binding with ATPase activity suggest a possible involvement of a Ca^{2+}ATPase (Brini and Carafoli 2011) in the regulation of the haustorium initiation pathway. In *Striga*, however, a distinct set of enriched GO MF terms was detected, including "nucleotide binding", "ATPase activity", and "ATP-dependent helicase activity". This may suggest a different picture associated with HIF exposure between facultative and obligate parasitic plants.

By combining the list of shared upregulated genes in haustoria (Figure 2-5, supplementary data 5) with HIGs (Figure 2-6, supplementary data 11), we define a joint set of genes in these parasites that we call "*parasitism genes*". Parasitism genes are defined by having enhanced expression in parasitic structures, and likely playing a role in parasite biology, as opposed to parasite-specific sequences which are defined only by their joint presence in parasitic and absence from nonparasitic plants (see below). In total, we identify 1809 parasitism genes in these three parasitic species that are assigned to 298 orthogroups that were shared by at least two species. As most of the parasitism genes shared by two parasitic plants also show upregulated pattern in the third species (supplementary data 7), this set of genes upregulated in parasitic process constitutes a shared set of parasitism genes in Orobanchaceae. There are almost 300 gene families shared by at least two species that have genes upregulated in the parasitic processes.

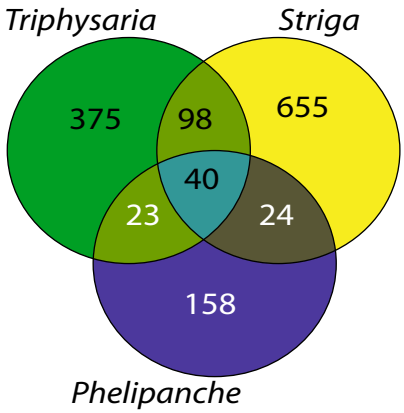


Figure **2-5**. Venn diagram illustrating the number of orthogroups with upregulated expression in stage 3 and/or stage 4 (by K-means and SOM) in *Triphysaria, Striga*, and *Phelipanche*.

The number of orthogroups containing HIGs varied widely in the three parasites (2285 in *Striga*, 17 in *Phelipanche*, and 249 in *Triphysaria*). While the low number of HIGs in

Phelipanche could be a result of lower power to detect differential expression (fewer replicated libraries and a smaller total volume of data; supplementary data 12), we also observed that a large number of genes are highly upregulated in *Phelipanche* stage 1, suggesting that *Phelipanche* may automatically begin haustorial initiation without HIF exposure, which was reported by a previous study where haustorial initiation, as an exception, doesn't require the application of HIF (Joel and Losner-Goshen 1994a). To examine this possibility, we expanded the list of HIGs by including genes highly expressed in stage 1 (a cluster of upregulated expression in stage 1 relative to any other stages 0, 2, 3, 4, 6.1, and 6.2) of *Phelipanche* and performed another Venn diagram analysis. In this expanded set, there are eight additional orthogroups including HIGs that were shared by all three parasitic plants as well as an additional 65 orthogroups shared by two species (Figure 2-6 – number in parenthesis). An examination of the eight orthogroups revealed genes coding for the following functions: cytochrome P450, heat shock protein 70, ribosomal protein, peptidase C48, oleosin, ATPase, pyruvate kinase and integrase. This is consistent with the possibility that *Phelipanche* starts haustorial initiation at an earlier stage (stage 1) than other two hemiparasites. Alternatively, because haustorium development in response to HIFs in *Phelipanche* is not as evident as in *Striga* or *Triphysaria* (Joel and Losner-Goshen 1994a), there may actually be fewer HIGs in *Phelipanche*.

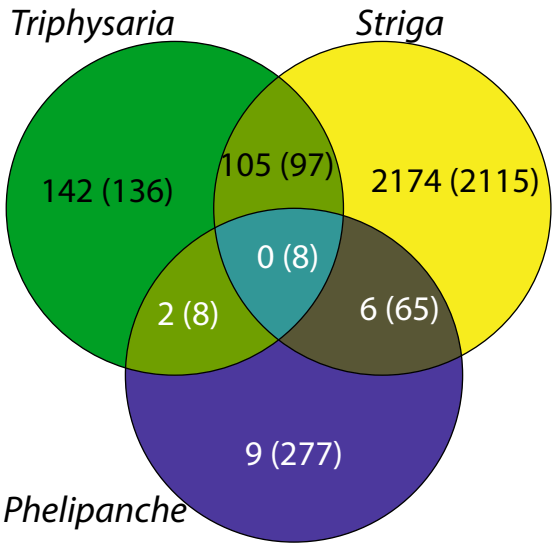


Figure **2-6**. HIGs: Orthogroups containing genes upregulated in root or seedlings following haustorial initiation factor (HIF) exposure (stage 2) compared to germinating seedlings (stage 1) in *Triphysaria*, *Striga* and *Phelipanche.* The numbers in parenthesis are the corresponding set of

orthogroups when including genes that are highly expressed in stage 1 (relative to any other stages of stage 0, 2, 3, 4, 6.1, and 6.2 – by K-means clustering) of *Phelipanche*.

2.2.7 A majority of parasitism genes evolved through gene duplication

We next explored the possibility that these putative parasitism genes evolved via gene duplication. We utilized an automated tree-building pipeline (Jiao et al. 2011; *Amborella* Genome Project 2013) to construct gene families for 114 orthogroups containing parasitism genes with a manageable number of genes to score for gene duplications (orthogroup number ranging from 1000 to 9999). Each orthogroup contained homologs from 22 other plant species used in the construction of the classification (*Amborella* Genome Project 2013), plus the genes from Orobanchaceae identified here. Manual inspection of the gene tree phylogenies was performed to find parasite genes that may have been missing in one or more "whole plant" combination assemblies. We also manually examined each alignment (and resulting tree) for frame shift and translation errors that could result in extremely long (> 10x others) branches. These errors were corrected when possible, or the sequence was eliminated from the matrix to avoid spurious topologies. Together, these gene family phylogenies give us a broad view of how parasitism genes evolved.

Of the 114 orthogroups (supplementary data 13) containing parasitism genes, gene duplications were detected in 58 trees at ≥ 50% bootstrap support and 38 trees at ≥ 80% bootstrap value support. By mapping the duplication events observed in parasitic plants onto phylogenetic species trees, and examining bootstrap support values for key supporting nodes, we determined when the putative parasite paralogs were duplicated (supplementary data 14). A detailed scheme illustrating various duplication events for when parasitism genes were duplicated is shown in supplementary data 15. The greatest proportion of duplicated gene families supported a gene duplication event that occurred in a common ancestor of *Mimulus* and Orobanchaceae (but not seen in Solanaceae, other asterids, or rosids) (table 2-5).

Table **2-5**. Phylogenetic placement of gene duplications observed in gene families with shared parasitism genes.

Duplicated lineages	Orthogroups with duplication	
	BS≥ 80%	BS≥ 50%
Parasite-wide	9 (23.68%)	13 (22.41%)
Orobanchaceae-wide	4 (10.53%)	9 (15.52%)
Orobanchaceae+Mimulus	21 (55.26%)	28 (48.28%)
EuAsterid1-wide	1 (2.63%)	1 (1.72%)
Asterid-wide	1 (2.63%)	1 (1.72%)
Core-eudicot-wide	5 (13.165)	8 (13.79%)
Eudicot-wide	3 (7.89%)	8 (13.79%)
Total	38 (100%)	58 (100%)

As with the parasite genes in general, most of the duplicated parasitism genes detected in this analysis were annotated with terms related to peptidase activity (such as aspartyl protease, serine carboxypeptidase) and cell wall modification processes (pectate lyase, pectin methylesterase inhibitor, carbohydrate-binding X8 protein and glycosyl hydrolase). In addition, three transcription factors (homeodomain-like transcription factor, ethylene responsive transcription factor and LOB domain-containing protein), genes with transporter activity (cationic amino acid transporter, major facilitator family protein, NOD26-like intrinsic protein and an oligopeptide transporter), a peroxidase, and a leucine-rich repeat-containing protein were also derived from scorable gene duplications.

2.2.8 Regulatory neofunctionalization and origin of the haustorium from root and flower

Tissue expression clustering - To obtain a global view of transcriptional profiles throughout parasite growth and development in each species, expression values for each unigene were clustered by tissue and stage expression levels using complete linkage and correlation distances using the pvclust routine in R (Racine 2012) (supplementary data 16). The expression clustering in each of the three parasitic plants (Figure 2-7) shows that overall, gene expression from vegetative and reproductive above-ground tissues is quite different from the below-ground structures. In both *Striga* and *Triphysaria*, above-ground stage 6.1 (vegetative structures) clustered with above-ground stage 6.2 (reproductive structures; floral buds), and were separated from the remaining tissues with 100% bootstrap support. In contrast, pre-emergent shoots

occurring underground (stage 5.1) in *Phelipanche* clustered with above-ground shoots (stage 6.1), and the two shoot transcriptomes clustered with floral buds (stage 6.2). It is notable that *Phelipanche* and *Striga* both produce pre-emergent shoots (stage 5.2), but the overall expression patterns of this stage in the two species is somewhat different. The fact that *Striga* pre-emergent shoots (stage 5.1) do not cluster with emergent shoots (stage 6.1), but are more similar to other below-ground stages (haustorium - stage 4 and pre-attachment roots - stage 5.2), may be due to the fact that photosynthetic activity in *Striga* shoots only becomes active after emergence, while *Phelipanche* is not capable of photosynthetic activity at all.

Cluster analysis (Figure 2-7) also shows that in all three species haustorial gene expression is overall most similar to root expression. In *Triphysaria* and *Phelipanche*, expression patterns from both stage 3 and stage 4 haustorial tissues cluster with roots. [Roots from *Triphysaria* were taken from germinated seedlings (stages 1 and 2), while roots for *Phelipanche* were from the late post attachment stage, but prior to shoot emergence from the soil (stage 5.2).] In *Striga*, a late stage 4 haustorial tissue is most similar to roots prior to the above-ground emergence of shoots, a scenario similar to *Phelipanche*.

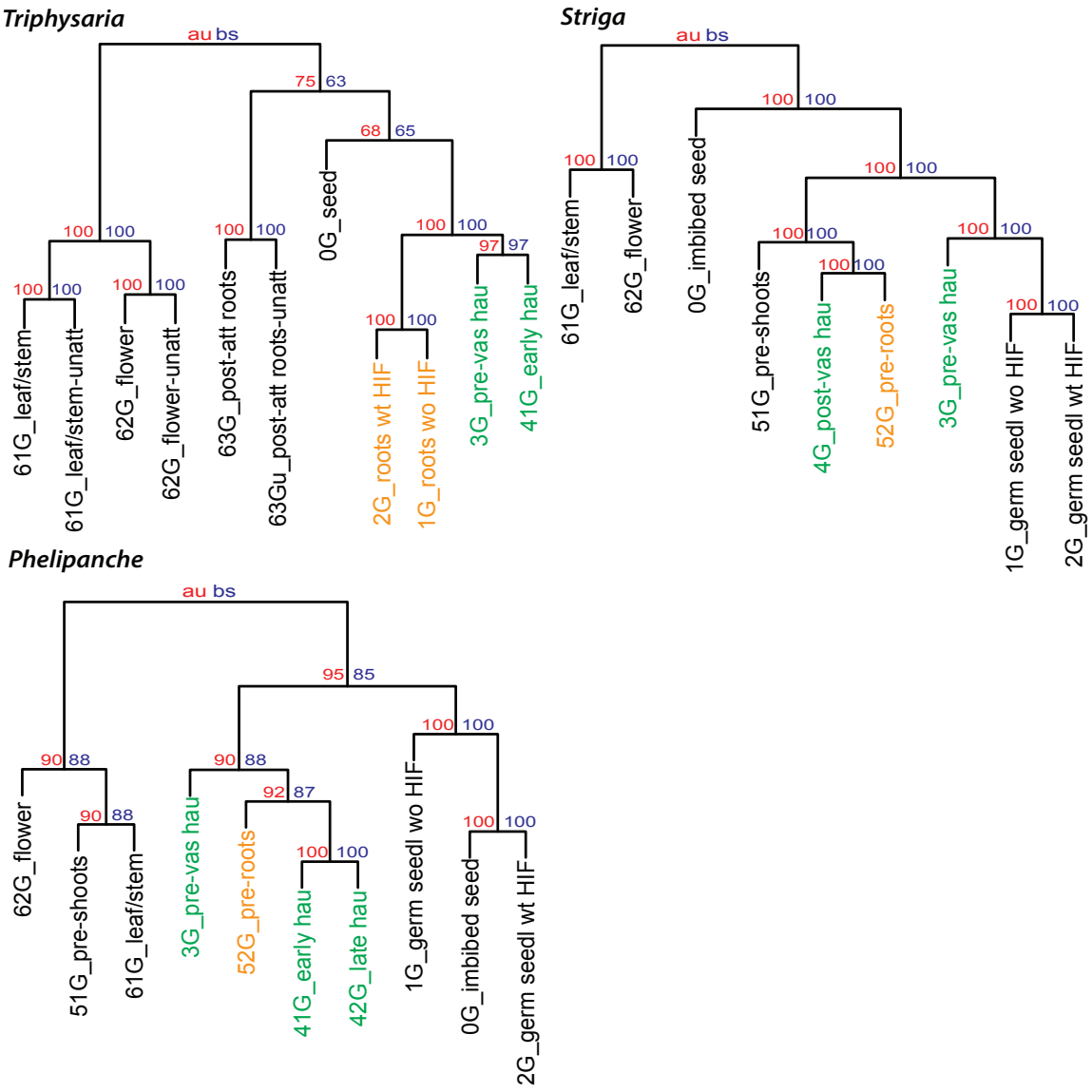


Figure **2-7**. Overall similarity of transcriptional profiles of all stages in three parasites. Numerical values represent supports as estimated by the approximately unbiased (on left, in red) and bootstrap (on right, in blue) as described in the methods. Clustering was performed with complete linkage and correlation distance. Haustoria tissues (from stages 3 and 4) are labeled in green, while root tissues (from stages 1, 2, and 5.2) are labeled in orange.

Expression of orthologs of parasite genes in nonparasitic models - The pattern we see of haustorial expression being most similar to root is consistent with the longstanding hypothesis that the haustorium was derived from a modified root (Kuijt 1969; Musselman and Dickison

1975; Joel 2013). However, it is also possible that individual genes functioning in the haustorium have been recruited from genes normally expressed in other plant organs. To investigate this scenario, we compared gene expression of candidate parasitism genes with extensive gene expression data from multiple tissues and organs of *Arabidopsis thaliana*, *Medicago truncatula,* and tomato to examine the evolution of expression patterns across large evolutionary distances. When we trace the haustorial gene expression back to orthologous genes in nonparasitic plants, we found significantly higher organ-specific expression in root and floral tissue, than in leaf, seed, or hypocotyl (Figure 2-8 and supplementary data 22).

In all three nonparasitic model plant species, the orthologs of haustorial genes are expressed most highly in root and floral (or fruit) tissues, suggesting that these were the major sources of genes recruited to the haustorium. For both *Medicago* and tomato, root is the most frequent source, which is consistent with the similarity between root and haustorial expression in the parasites (Figure 2-7), In *Arabidopsis*, orthologs of haustorial genes are also commonly upregulated in roots, but even more are upregulated in pollen. It is possible that the slightly different picture obtained from the three nonparasitic plants might be due to differences in tissue sampling in the nonparasitic model species. For example, floral tissue sampling is less extensive in tomato and *Medicago* than in *Arabidopsis*.

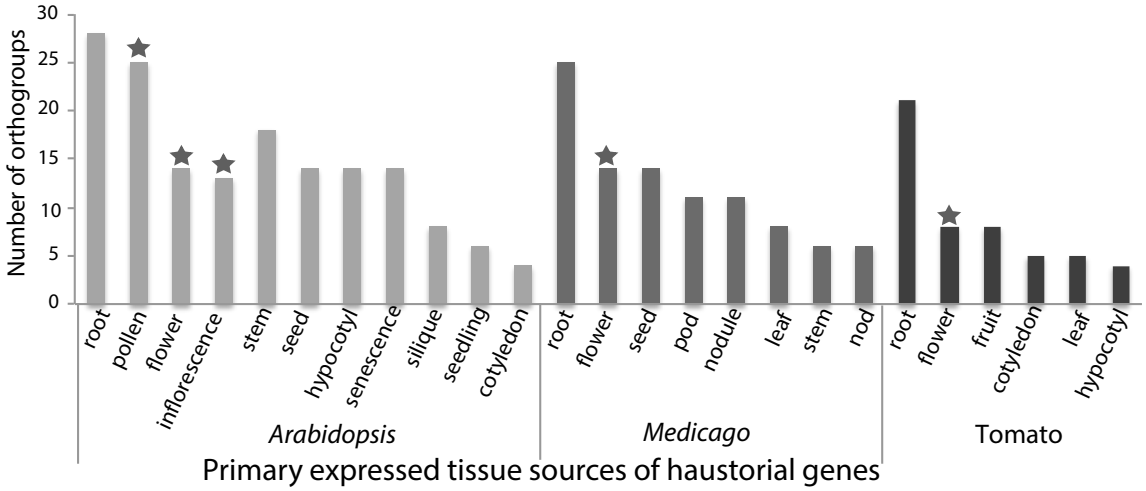


Figure **2-8**. Haustorial genes in parasitic species were recruited from root, flower and other tissues. Values on the Y-axis show the number of orthogroups containing haustorial genes as identified from expression analysis of nonparasitic model species *Arabidopsis*, *Medicago* and tomato. Tissues on the X-axis represent the principally expressed tissue for orthologs of upregulated haustorial genes in *Arabidopsis* (light black), *Medicago* (grey), and tomato (black)

are shown in, grey, and black, respectively. Floral tissues are highlighted with stars on top of the bars.

2.2.9 Parasitism genes show signatures of adaptive evolution or relaxed constraint in parasitic lineages

To examine whether altered selection patterns play a role in the evolution of parasitism, we utilized the branch model in PAML for hypothesis testing. We labeled the parasite genes as the foreground and the remaining genes as the background, and then identified genes that show accelerated evolution (dN/dS ratio) in parasitic lineages compared to the background. We focused on parasitism genes identified in our analyses. The branch model implemented in PAML was used to identify orthogroups that show a significantly higher dN/dS ratio in parasitic lineages compared to nonparasitic lineages. Twenty-seven orthogroups were found to have greater dN/dS in parasitic lineages compared to nonparasitic lineages, whose GO biological processes include proteolysis, cell wall modification, oxidation-reduction process, transport, protein glycosylation, cytokinin metabolic process and ubiquitin-dependent protein catabolic process (supplementary table S1). To examine if there are sites that have evolved under positive selection, the branch-site model was performed on orthogroups that show an elevated dN/dS ratio in foreground parasite lineages relative to background nonparasitic lineages. Nine orthogroups were found to contain sites under positive selection. They include two orthogroups encoding aspartyl protease and one orthogroup each encoding serine carboxypeptidase, expansin, glycosyl transferase, pectin methylesterase inhibitor, PAR1 protein, and C2H2 and C2H2 zinc finger family protein (supplementary data 17), respectively.

We then compared the dN/dS ratio of haustoria-specific genes with nonhaustorial specific genes. The dN/dS ratio was calculated by selecting orthologous pairs across three different parasitic species by best blast hit based on an E-value cutoff of e-10 (PaSh: between *P. aegyptiaca* and *S. hermonthica*, PaTv: between *P. aegyptiaca* and *T. versicolor* TvSh: between *T. versicolor* and *S. hermonthica*). dN and dS were calculated separately by codeml in PAML (supplementary data 18). The distribution for genome-wide dN/dS values was represented with a symmetric violin plot from each pairwise species comparison. This overall distribution was made by calculating the dN/dS ratio for the orthologous pairs from all unigenes from each species pair.

The distributions of all haustorial gene pairs (in red) and a randomly chosen equally-sized set of nonhaustorial gene pairs (in blue) were represented by a dotplot and a density plot. In all three cases, the peaks of the density plots for the haustorial genes are above the peak for nonhaustorial genes. The haustorial genes exhibit significantly greater dN/dS ratio compared to the nonhaustorial genes for all three pairwise species comparisons (Wilcoxon rank, P-value < 0.01) (Figure 2-9).

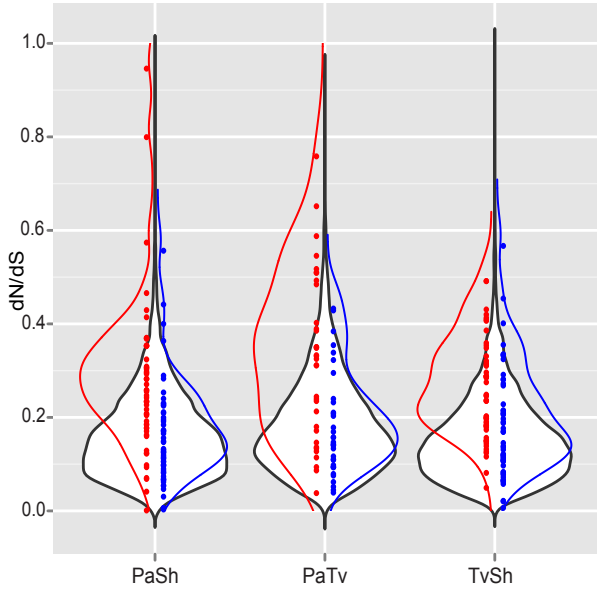


Figure **2-9**. Haustorial genes show evidence of adaptive selection or relaxed selective constraint. Symmetric violin plots show the genome-wide distributions of dN/dS values for comparisons of *P. aegyptiaca* (Pa), *S. hermonthica* (Sh) and *T. versicolor* (Tv). Red dots represent haustorial genes shared by all three species, while blue dots represent randomly selected non-haustorial genes from the relevant species. The density plots colored in red and blue represent the frequency distributions of the individual dN/dS values as seen in the dot distributions of haustorial and nonhaustorial genes.

2.2.10 The majority of the parasite-specific sequences have unknown functions

The transcriptome assemblies allowed us to identify parasite-specific sequences (Yoshida, Ishida, et al. 2010), some of which may be associated with the role of parasitism. This was done by building a secondary OrthoMCL orthogroup classification of all of the genes and transcripts that were not assigned to the initial orthogroup classification of 22 plant genomes. In addition to the singleton genes from the 22 genomes, we used in the secondary classification all of the unassigned transcripts from the three parasitic species and from *Lindenbergia, Lactuca and Helianthus*. This strategy, which includes multiple nonparasitic lineages closely related to the parasites, allows a highly sensitive means of distinguishing sequences found only in the parasitic Orobanchaceae. We identified 84 novel orthogroups that contain sequences from all three parasitic Orobanchaceae species, but lack sequences from any nonparasitic plant (supplementary table S2). 178, 180, and 139 unigenes were found in these orthogroups from *Triphysaria*, *Striga,* and *Phelipanche*, respectively. The large majority of these sequences had no significant BLAST alignments to any of the nonparasitic species, while a few of the predicted peptide sequences (6, 18, and 13 from *Triphysaria*, *Striga*, and *Phelipanche*, respectively) had hits to genes of unknown function in the annotation databases. Most of the significant hits in the annotation databases corresponded to sequences that were transposon and retrotransposon related, such as GAG-pre-integrase domain protein and retrotransposon gag protein (supplementary table S2). A sequence with homology to a homing intron endonuclease with two LAGLIDADG motifs (Belfort and Roberts 1997) was also among the sequences with annotations. One orthogroup contained sequences with distant homology to genes annotated as mitochondrial aconitase with a putative role in mitochondrial oxidative electron transport (Yan, et al. 1997). We also obtained the stage-specific expression profiles for each of parasite specific genes. Six and four of the parasite specific unigenes in *Striga* and *Triphysaria* were also among the list of significantly upregulated haustorial genes (supplementary table S2). In addition, three parasite specific orthogroups contained genes from all three parasite species that exhibited a pattern of increased expression in the haustorial post attachment penetration stages (supplementary table S2).

2.3 Discussion

2.3.1 Summary of results

The Parasitic Plant Genome Project has used large-scale transcriptome sequencing to interrogate multiple stages of parasite growth and development of three related parasitic plants spanning a wide range of parasite ability, enabling an integrated analysis of genes upregulated in parasitic processes in Orobanchaceae. The large comparative framework has made it possible to identify, for the first time, a set of genes we believe are essential or core to parasitism. These candidate parasitism genes will be a valuable resource for future functional studies as we strive to understand the genetic changes that led to the parasitic lifestyle, as well as those that resulted from the transition to heterotrophy. Among the core parasitism genes are specific members of gene families encoding cell wall modifying enzymes (cellulase, pectate lyases, glycosyl hydrolases, and pectin methylesterase), and peroxidase enzymes, proteins that are known to be involved in the parasite invasion process (Singh and Singh 1993b; Antonova and TerBorg 1996; Losner-Goshen, et al. 1998; Pe´rez-de-Luque 2013). We also identified a variety of genes (encoding proteases, transporters, regulatory proteins [transcription factors and receptor protein kinases], and others including many genes of unknown function) that are co-expressed in parasitic stages and may be important in haustorial development and function. Because homologs of these haustorially-expressed genes encode proteins also functioning in nonparasitic plants, this supports the endogenous mechanism for the origin of parasitism (Bandaranayake and Yoder 2013b). For this collection of genes, the evolution of regulatory sequences resulting in novel expression in the haustorium were likely essential to the evolution of parasitic functions. By expanding our analyses to orthologous genes in non-parasitic model plants, we have gained insights into the evolution of parasitism and the source of genes that shifted expression to parasite tissues, and presumably function there. Although some genes have gained haustorial expression in the absence of detectable gene duplication, a majority of the parasitism genes originated following a gene duplication event.

2.3.2 Cell wall degradation enzymes and the haustorium

Enzymatic degradation of the host cell walls has been suggested to be important in the penetration of the parasite across the host root cortex as it attempts to reach and connect with host vascular tissues (Kuijt 1977). Baird and Riopel (1984) observed that the intrusive cells at the tip of the penetration peg of the haustorium contained a densely staining cytoplasm, indicative of high levels of cell wall hydrolytic activity. Additionally, early histological studies observed that during penetration of the host endodermis by the haustoria of *P. aegyptiaca*, there was the dissolution of the middle lamella between host cell walls (Joel and Losner-Goshen 1994b) and degradation of the cutin of the Casparian strips (Joel, et al. 1998). Cell-wall degrading enzymes such as cellulase, polygalacturonase, and xylanase were found in the tubercles of *P. aegyptiaca* suggesting that they play a role in the penetration process necessary for establishing haustorial connections with the host vasculature (Pe´rez-de-Luque 2013). In the enriched set of upregulated prehaustorial and haustorial genes in dodder (*Cuscuta pentagona*), many genes encoding cell wall modifying enzymes were found including pectate lyases, pectin methylesterase, cellulases, and expansins (Ranjan et al. 2014).

In our study, four glycosyl hydrolase and five pectate lyase (PL) genes were upregulated in haustorial tissues in at least two species of Orobanchaceae. GO enrichment analysis of the cellular component terms identified cell wall and extracellular localization annotation terms as being significantly enriched among the upregulated haustorial genes, suggesting that proteins encoded by these genes tend to be secreted where they could impact cell wall integrity of the parasite or the host. Consistent with this idea was the evidence of disintegration of the middle lamella in *Striga gesnerioides* attacking cowpea (Reiss and Bailey 1998). Glycosyl hydrolases were shown to have a role in hydrolysis and degradation of structural or storage polysaccharides, including cellulose and hemicellulose (Henrissat, et al. 1995). Pectic enzymes have long been recognized as important proteins in cell wall loosening or disassembly. Almost all cell-wall penetration processes, including pollen tube growth and bacterial or fungal pathogenic invasion, involve the modification of pectins that are integral for cell wall stability. For instance, studies reported the role of pectic enzymes in degrading the plant cell wall during the invasion process by bacterial or fungal pathogens (Delorenzo, et al. 1991; Volpi, et al. 2011). Similarly, a recent study identified a PL to be required for root infection by rhizobia during nodulation (Xie, et al. 2012). Additional evidence also supports a direct role of PLs in loosening of the cell wall in fruit

ripening (Marin-Rodriguez, et al. 2002). Thus, it is likely that PLs play a role in loosening or separating the host cell wall for invasion by parasitic plants.

Immunological detection of pectin methyl esterase (PME) at the penetration site and de-methylated pectins at the cell walls adjacent to the intrusive cells of *Orobanche* (Losner-Goshen et al. 1998) implies a role for pectin degradation in the haustorial penetration process. Three orthogroups (orthogroup 2875, 6176 and 19181) encoding putative pectin methylesterase inhibitors (PMEI) were identified to contain genes that were specific to haustorial tissue in at least two species of Orobanchaceae. It would be interesting to determine whether PMEs and PMEIs have distinct roles in parasite-host interactions. A previous analysis revealed another gene involved in cell wall modification, a beta-expansin, which was highly expressed in the haustorial interface between *Triphysaria versicolor* and its grass family host (Honaas et al. 2013).

2.3.3 Proteases, transporters, and the haustorium

Genes involved in proteolysis, largely proteases and proteinases, account for a great proportion of transcripts upregulated in the haustorium and that are shared by at least two of the parasite species we investigated. The upregulated haustorial genes identified in this study include four genes encoding subtilisin-like serine protease similar to those required for virulence in bacterial pathogens (Kennan, et al. 2010). In addition, a subtilisin-like protein from soybean was reported to activate defense-related genes (Pearce, et al. 2010). In nonparasitic plants, serine proteases often play a role in various processes including protein degradation/processing, hypersensitive response, and signal transduction (Antao and Malcata 2005), but what roles they take on in Orobanchaceae parasites is not yet clear.

In addition to serine protease, the eukaryotic aspartyl proteases are also enriched among upregulated haustorial genes. Aspartyl proteases (APs), a large gene family with members present in all living organisms, play central roles in protein degradation, processing, and maturation (Chen, et al. 2009). Plant APs are expressed in various organs including seed, root, grain, leaf, and flower (Chen et al. 2009). Also they play a role in seed germination, where they degrade seed-storage proteins to provide amino acids to growing plants (Higgins 1984). Other studies identified APs of blood-feeding malaria parasites to play a role in degrading hemoglobin proteins to amino acids for nutrition (Brinkworth, et al. 2001). In addition, a marked expansion within the AP gene family was found in the xylem-feeding hemiparasites, *Triphysaria* and *Striga*

hermonthica (Dorr 1997; Neumann, Vian, Weber and Salle 1999), but not in the phloem-feeding obligate parasite, *Phelipanche* (Aly et al. 2011), which has a lower rate of nutrient uptake from the xylem stream, suggesting that these proteins may play a pivotal role in nutrient mobilization only in the hemiparasites.

Our analyses indicate that four APs show peak expression in haustorial tissue, while their orthologs in *Arabidopsis* and tomato (or paralogs in the parasitic plants) have peak expression in root and flower (supplementary fig. S2) (Yang et al. 2015). This provides another line of evidence for gene recruitment to haustorial function, and suggests that this has occurred through regulatory neofunctionalization. A recent study reported that a rice aspartyl protease plays an indispensible role in pollen tube germination and growth, and that loss of function results in reduced male fertility (Huang, et al. 2013). In addition, a secretome analysis (Kall, et al. 2007) of predicted peptides for upregulated haustorial genes revealed that haustorial genes are more likely than nonhaustorial genes to be extracellularly localized or contain a signal peptide structure (supplementary data 19). The fact that upregulated haustorial genes are enriched for proteases with signal peptides suggests that the evolution of parasitism may be associated with an expansion of the suite of secreted proteases to aid parasite attack and/or feeding

Analysis of selective constraints showed that proteases with expression specific to haustoria in the parasites show a greater dN/dS as compared to their orthologous genes in nonparasitic plants. The greater dN/dS ratio for these upregulated haustorial proteases suggests either a relaxation of purifying selection or adaptive evolution of these protease-encoding genes associated with the evolution of parasitism. The fact that particular sites were indicated as evolving adaptively, especially in the functional domains, provides support for the latter hypothesis (supplementary data 17).

We also found five genes encoding transporters upregulated in the haustorium that are shared by at least two species: one ABC transporter, two oligopeptide transporters, one zinc transporter, and one glutamate transporter. Upregulated genes encoding transporters including sugar transporter, amino acid transporter, and ammonium transporter were also identified to be enriched in haustorial tissue of the parasite Cuscuta (Ranjan et al. 2014). Interestingly, *Striga hermonthica* infection has been shown to increase amino acid levels in xylem sap of its *Sorghum* host, with glutamate being the predominant form of translocated nitrogen (Pageau, et al. 2003). The assimilation of host ^{15}N-labeled nitrate into the parasite (Pageau et al. 2003), provided evidence for the potential role of a glutamate transporter in nitrogen translocation between the host and parasite.

2.3.4 Gene duplication and regulatory neofunctionalization – origin of parasitism

The identification of haustorium genes allowed us to gain new insights into the evolution of parasitism. We have used phylogenetic analysis to show that a majority of the genes with a putative role in parasite functions arose by gene duplication. Most of the duplications occurred in a nonparasitic common ancestor of the parasitic Orobanchaceae species and *Mimulus* (a nonparasitic plant in the related nonparasitic family Phrymaceae) (Schaferhoff, et al. 2010; Refulio-Rodriguez and Olmstead 2014). This suggests that either multiple independent gene duplications, or a whole-genome duplication event occurring before the divergence of *Mimulus* and Orobanchaceae (Wickett, et al. 2011), may have resulted in the diversification of genes important to haustorial development, and ultimately contributed to the rise of parasitic plants. In contrast, relatively few of the parasitism genes arose through duplications occurring in a more recent common ancestor of just the parasites or Orobanchaceae. The fact that the gene duplications that produced parasitism genes occurred in a nonparasitic ancestor, well before the origin of the parasitic Orobanchaceae about 32 my ago (Naumann, et al. 2013), is consistent with the idea that gene duplication does not immediately give rise to novel functions, as described recently by the WGD-Radiation Lag-Time Model (Schranz, et al. 2012).

We mapped expression data from multiple stages of parasite development onto parasite genes and found that under most circumstances, the two copies derived from a gene duplication event show different expression profiles. In addition, we interrogated the expression profiles of orthologous genes from tomato, *Medicago* and *Arabidopsis*. By comparing the parasite duplicate's expression with orthologous gene expression in these related nonparasitic plants, we gained insight into how parasitism genes evolved, both in gene sequence and gene expression. The fact that these parasitism genes shift their expression from root or flower in related nonparasitic plants to haustoria of parasitic plants, through gene duplication or otherwise (Figure 2-4 and supplementary fig. S2), supports inferences of neofunctionalization in the evolution of parasitism (Conant and Wolfe 2008; Innan and Kondrashov 2010). While investigating the evolution of parasitism genes, we also found that a number of them play a role in symbiotic nodulation in non-parasites such as genes encoding ERF transcription factors (TFs) (Vernie, et al. 2008), oligopeptide transporter (Nogales, et al. 2009), peroxidase, pectinesterase inhibitor (Young, et al. 2011; Zouari, et al. 2014), suggesting possible parallels between the evolution of parasitism and that of mutualism, both of which involve invasion of host tissues. Similar evidence

that *Phelipanche aegyptiaca* induced upregulation of genes involved in nodulation in *Lotus japonicus* supports this idea (Hiraoka et al. 2009).

2.3.5 Origin of the haustorium involves co-option of root and/or flower genes

The results of this study shed light on the origin of haustoria in Orobanchaceae, where molecular phylogenetic studies have identified the nonparasitic *Lindenbergia* as sister to the parasitic Orobanchaceae and are thus consistent with a single origin of haustorial parasitism in Orobanchaceae (dePamphilis, et al. 1997; Olmstead, et al. 2001; Schneeweiss, et al. 2004; Bennett and Mathews 2006; Angiosperm Phylogeny Group 2009; McNeal, et al. 2013a). Two lines of evidence - global gene expression data in the parasites, and expression specificity in related nonparasitic plants - suggest that gene expression patterns in haustorial tissues are most similar to those of root. The second largest number of haustorial genes shows floral specific expression patterns in nonparasitic models. These observations suggest a possible mechanism of parasitism through neofunctionalization, where genes with a role in root and floral biology in non-parasite species were co-opted to haustorial function in parasite species.

The root is a likely source for processes useful to subterranean haustorial structures. Both haustoria and roots operate underground, are physically adjacent, and haustoria are derived from apex of the primary root, sometimes from lateral root extensions (Heide-Jørgensen 2013a). Additionally, both haustoria and root are highly specialized organs for nutrient uptake and transfer. The recruitment of many haustorial genes from those normally expressed in floral tissue such as pollen is more surprising, but the idea that haustorial growth was similar to the intrusive growth of pollen tubes was explored recently (Thorogood and Hiscock 2010; Pe´rez-de-Luque 2013). The authors propose that neighboring host cells recognize the parasite as alien without reacting against the “invasion”, similar to the way that plants recognizes intrusive growth of pollen tubes (Thorogood and Hiscock 2010; Pe´rez-de-Luque 2013). Specifically we found that genes, like pectate lyases that are used in polarized pollen tube growth to rapidly invade stylar tissue (Krichevsky, et al. 2007), are expressed during the invasion of host tissue by the growing parasitic haustorium. One possible explanation is that the penetration peg of the haustorium, which grows rapidly into the host tissue, may have co-opted genes from the polarized, invasive growth found in the pollen tubes of flowers (Sampedro and Cosgrove 2005; Krichevsky et al. 2007; Honaas et al. 2013). It is also likely that some pectic enzymes involved in loosening the

pollen tube cell walls so that they can elongate into the female reproductive tissues (Taniguchi, et al. 1995) are recruited in the penetration and growth of the haustoria towards the host vascular tissue.

2.3.6 Parasite-specific genes – mobile elements

Of the 84 parasite-specific orthogroups detected in our analysis, most contained sequences with no known function. However, almost all of the remaining sequences with an annotated Pfam domain have significant BLAST alignments with genes encoding proteins involved in the transfer of mobile elements, including retrontransposon gag protein, GAG-pre-integrase domain and a LAGLIDADG homing endonuclease (supplementary table S2). Retrontransposon gag proteins are associated with the transposition of retrotransposons to telomere-associated structures in *Drosophila* (Rashkova, et al. 2002), while GAG-pre-integrase domain proteins are associated with chromosomal rearrangements by retrovirus insertion activity (Houzet, et al. 2003). Interestingly, some LAGLIDADG endonucleases (Belfort and Roberts 1997) encoded by self-splicing group I introns are implicated in the highly mobile transfer and insertion of copies of the intron to specific target sequences. Intron homing by the LAGLIDADG endonuclease activity is implicated in the widespread horizontal gene transfer of the self-splicing group I intron in plant mitochondrial *cox*1 genes (Vaughn, et al. 1995; Cho et al. 1998; Barkman et al. 2007; Sanchez-Puerta, et al. 2008; Sanchez-Puerta, et al. 2011). Thus, both transposable elements and homing introns by the retrontransposon gag proteins, GAG-pre-integrase domain proteins, and LAGLIDADG homing endonuclease are involved in horizontal gene transfer (Daniels, et al. 1990; Rodelsperger and Sommer 2011), supporting the possibility of a mechanistic link between parasitism in plants and at least some horizontal gene transfers (Barkman et al. 2007; Xi et al. 2013; Zhang et al. 2013a).

2.3.7 Conclusion

In this paper we have shown that parasitic plants have evolutionarily recruited many genes for haustorial development and host penetration from genes that were involved in other processes in related nonparasitic plants, primarily root or flower development. These candidate

parasitism genes are being functionally characterized to determine if they are essential to parasite function and survival. The observation that genes with similar GO classifications (cell wall modification process and transporters) are also upregulated in the haustoria of *Cuscuta* (Ranjan et al. 2014), increases the likelihood that these genes do play important roles in haustorial function. In Orobanchaceae, genes recruited from root or pollen tube development show evidence of potentially adaptive changes in elevated dN/dS ratios and sites with excess non-synonymous changes in parasitic lineages. The study of parasitic haustoria in Orobanchaceae indicates that two modes of regulatory neo-functionalization – either following gene duplication or in unduplicated orthogroup lineages – have provided the mechanism through which a novel structure has evolved.

2.4 Materials and methods

2.4.1 Tissues, libraries, and sequence data

Multiple stages of parasite development from the species *Triphysaria versicolor*, *Striga hermonthica* and *Phelipanche aegyptiaca* within Orobanchaceae were interrogated by transcriptome sequencing. Detailed descriptions of the stages ranging from seed and seedling, though haustorial development to above ground tissues such as leaf, stem, and flowering, are illustrated by Westwood et al (2012) and in figure 1. Tissues from each stage were subjected to RNA extraction and library preparation, followed by subsequent Illumina paired-end sequencing (Honaas et al. 2013). Methods for Illumina and 454 paired-end mRNA-Seq library construction and sequencing are as described in Wickett et al (2011). Additional Illumina transcriptome sequences were also obtained from a single normalized library for each species using pooled RNAs of all stages. Finally, a normalized whole plant library was also constructed and Illumina sequenced, as above, for *Lindenbergia philippensis*, representing the nonparasitic sister group to the parasitic Orobanchaceae (dePamphilis et al. 1997; Young, et al. 1999; Olmstead et al. 2001; Schneeweiss et al. 2004; Angiosperm Phylogeny Group 2009; McNeal et al. 2013a).

2.4.2 Assembly, cleaning, and annotation (including gene family classification)

Duplicate reads in the Illumina sequence data were removed with CLC Assembly Cell version 3.2.0 (http://www.clcbio.com/products/clc-assembly-cell/). Adapters and bases with a quality score lower than Q20 were trimmed from the ends of the reads, and these reads were retained only if at least half of the sequence had quality ≥Q20. Raw Roche 454 sequence files in Standard Flowgram Format (SFF) were converted to FASTA and associated quality files along with clipping of sequence adapters and low-quality bases using sff_extract version 0.2.10 (http://bioinf.comav.upv.es/sff_extract/).

De novo assemblies of Illumina reads from each species were performed using Trinity release 2011-10-29 (Grabherr, et al. 2011) and *de novo* hybrid assemblies of combined 454 and Illumina reads were performed using CLC Assembly Cell version 3.2.0 with default parameters. Assembled transcripts from both assemblies were combined by assigning hybrid CLC transcripts to Trinity components that yielded the best bitscore with BLASTN (E-value = 1e-10). The resulting combined assemblies were filtered by removing contigs without coding regions (Iseli, et al. 1999) as well as redundant transcripts (Edgar 2010). The assemblies for parasite species (*Triphysaria*, *Striga*, and *Phelipanche*) were then cleaned to remove contaminant sequences using a three-step process: 1) the transcripts were screened against the NCBI non-redundant protein database using BLASTX (E-value = 1e-5) to remove non-plant transcripts, 2) the transcripts were then screened with BLASTN (E-value = 1e-10) against a collection of publicly available genomes and ESTs data sets from the experimental host plants (to identify and remove host transcripts), and 3) after performing this screen on each of the parasite species, BLASTN (E-value = 1e-10) of host candidate sequences were screened against the Orobanchaceae species (*Lindenbergia*, *Triphysaria*, *Striga*, and *Phelipanche*) databases (not including the parasite species being cleaned) to retain the transcripts that were much better matches to other Orobanchaceae family members than to the host plant. ORFs and protein sequences were predicted from the reconstructed transcript assemblies with ESTScan version 2.0 (Iseli et al. 1999). We experimented with different reference sequences to guide the ESTScan predictions, and other protein prediction programs such as GeneWise (Birney, et al. 2004). Finally we chose ESTscan with an *Arabidopsis* reference to obtain the best balance of length and protein number in the resulting protein set.

The predicted protein sequences were used for BLASTP (E-value = 1e-5) searches against Swissprot, TAIR10 and trEMBL databases to assign putative functional annotations in the form of human readable descriptions using the automated assignment of human readable

descriptions (AHRD) pipeline (https://github.com/groupschoof/AHRD). AHRD uses similarity searches and lexical analysis for automatic assignment of human readable descriptions to protein sequences. These translated transcripts were also annotated with Pfam domains using InterProScan version 4.8 (McDowall and Hunter 2011), and identified domains were directly translated into Gene Ontology terms.

2.4.3 Developing a component-orthogroup from each *de novo* assembly

Very large transcriptome sequence datasets, including those produced by this project, result in complex *de novo* assemblies including many splice variants and distinct alleles (Grabherr et al. 2011). Due to the assembly complexity, we combined expression information for unigenes that were assigned to the same Trinity component and mapped to the same orthogroup (Wickett et al. 2011). We call these unigenes from the same component and orthogroup a "component-orthogroup".

2.4.4 Read mapping and expression normalization

High-quality non-redundant Illumina reads from individual stage-specific samples were independently mapped on each parasite's post-processed transcripts using CLC Genomic Workbench version 6.0.4 (parameters: mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity = 0.8, min insert size = 100, and max insert size = 300). Transcript abundance was then estimated using the CLC Genomic Workbench RNA-Seq program with unique reads counted for their matching transcripts, and non-specifically mapped reads allocated on a proportional basis relative to the number of uniquely mapped reads. The numbers of reads mapped per library were normalized by the fragments per kilobase per million mapped reads (FPKM) (Mortazavi, et al. 2008) method that corrects for biases in the total transcript size, and normalizes for the total number of read sequences obtained in each sample library. The read counts and FPKM values of transcripts for each Trinity component classified as an orthogroup were summed up to obtain each component-orthogroup's expression in each library.

2.4.5 PV-clustering of global transcriptional profile

To get an overall picture of the global gene expression profile, stages were clustered by the pvclust (Suzuki and Shimodaira 2006) command in R using the expression of all unigenes in each stage. Pvclust not only clusters stages, but also infers confidence support with approximately unbiased multi-scale resampling (AU) and bootstrap resampling (BP) values, obtained by resampling genes from the total population of unigenes.

2.4.6 Identification of differentially expressed genes for candidate parasite gene assignment

We first used a differential expression analysis to limit the number of genes for candidate parasitism gene identification. DE analysis was performed within each species using DESeq package in R (Anders and Huber 2010), which utilizes the negative binomial distribution to model the read count data for variance estimation. Only the stages that were shared among the three parasite species were used in this analysis (stage 0, 1, 2, 3, 4, 6.1, and 6.2). DE analysis using pairwise comparison was undertaken to identify genes that were differentially expressed in at least two stages.

2.4.7 K-Means clustering to identify putative parasite feature within each species

DE analysis resulted in a list of genes with varying expression patterns across the seven shared stages. K-means clustering was used to identify clusters of co-expressed features with high expression in haustoria tissue. The expression of the DE unigenes measured by log2FPKM in stage 0, 1, 2, 3, 4, 6.1 and 6.2 constituted an expression matrix as the input for clustering analysis. To determine the optimum number of clusters needed to identify a single cluster with high expression in haustorium tissue, an R script was developed to generate a series of pdf files to represent each cluster's expression profile for a total number of specified clusters ranging from 2 to 30. Each cluster contained a set of co-expressed genes whose expression in each stage was reflected in a boxplot with the median expression of the co-expressed genes connected by a line. The criteria used to find the appropriate number of clusters was the smallest number of clusters

that enabled the visualization of a haustorial-specific (stage 3 and or stage 4) cluster. This means that for a smaller cluster number, we cannot identify the haustoria-specific pattern; a large cluster number may split the haustoria-specific cluster into several clusters, but is unnecessary in terms of gene identification.

2.4.8 Hierarchical clustering to identify putative parasite feature within each species

After the appropriate number of clusters was chosen, hierarchical clustering was used to identify genes with expression patterns of interest; that cluster was extracted with the cutree function in R. Each cluster's expression profile was determined by plotting a heat map using the heatmap.2 function from the gplots package in R.

2.4.9 Self organizing maps (SOM clustering) to identify patterns of haustorial-specific expression

SOM clustering was used to maximize the identification of genes that show a high and specific expression pattern. The analysis was performed on the web server with the unsupervised learning SOM clustering of GenePattern developed by the Broad Institute (Reich, et al. 2006). SOM clustering clearly reveals the overall pattern from the genome-wide gene expression data by reducing dimensionality of the original data. SOM clustering of GenePattern involved three steps: data preprocessing, SOM clustering and SOMClusterViewer. The expression matrix (log2FPKM) of the differentially expressed features was used as the input file for preprocessing, in which a row normalization and gene filtering were performed. The threshold and filter were performed with default parameters (floor: -3; ceiling: 18; min fold change: 1.5; min delta: 5). SOM clustering was performed by manually selecting the cluster-range. The default was chosen at the beginning and changed until the pattern with haustorial high and specific expression was revealed. The SOMClusterViewer displayed the expression profile of each cluster identified by SOM clustering. Finally for all three datasets from the three species, the cluster range was chosen at 6-8, which identified one or more clusters with high and specific expression in haustorial tissue.

2.4.10 Enrichment analysis – parasite genes versus whole plant background

The identification of a list of parasite genes allowed us to ask what biological functions are enriched compared to the whole plant background within each species. For each parasite gene, we obtained its BLASTx best hit in *Arabidopsis* and used these TAIR hits to identify Gene Ontology (GO) assignments (Ashburner, et al. 2000) and perform enrichment analysis. *Arabidopsis* was selected for this analysis because of its relatively complete GO-term annotation. Enrichment analysis was performed with DAVID using Bonferoni adjusted P-values for multiple tests (Huang et al. 2009) by comparing GO assignments for foreground (putative orthologs of parasite genes in *Arabidopsis*) vs. background (all genes from the *Arabidopsis* genome). Enriched components with statistical significance, with annotations including Pfam domains, Gene Ontology (GO) Molecular Function (MF), GO Biological Process (BP), GO Cellular Component (CC), and KEGG pathway, were identified for a set of shared parasite genes from each species.

2.4.11 Phylogenies and Ka/Ks constraint analysis for parasite genes

Transcripts from the Orobanchaceae were assigned into orthogroups defined by 586,228 protein-coding genes of 22 representatives of sequenced land plant genomes (*Amborella* Genome Project 2013) using OrthoMCL. The selected taxa includes nine rosids (*Arabidopsis thaliana*, *Thellungiella parvula*, *Carica papaya*, *Theobroma cacao*, *Populus trichocarpa*, *Fragaria vesca*, *Glycine max*, *Medicago truncatula*, *Vitis vinifera*), three asterids (*Solanum lycopersicum*, *Solanum tuberosum*, *Mimulus guttatus*), two basal eudicots (*Nelumbo nucifera*, *Aquilegia coerulea*), five monocots (*Oryza sativa*, *Brachypodium distachyon*, *Sorghum bicolor*, *Musa acuminata*, *Phoenix dactylifera*), one basal angiosperm (*Amborella trichopoda*), one lycophyte (*Selaginella moellendorffii*), and one moss (*Physcomitrella patens*). Of the plants with sequenced genomes, *Mimulus*, an emerging asterid model plant of family Phyrmaceae, is the most closely related to Orobanchanchaceae (Schaferhoff et al. 2010; Refulio-Rodriguez and Olmstead 2014). Candidate orthogroups for unigenes from transcriptome assemblies of *Lindenbergia*, *Triphysaria*, *Striga*, *Phelipanche*, and two Asteraceae species, *Lactuca sativa* and *Helianthus annuus,* were identified by retaining BLASTP (McGinnis and Madden 2004) hits with E-value <=1e-5 for predicted peptide searches against the orthogroup-classified proteomes from those 22 sequenced

plant genomes. HMM (Eddy 2011a) searches of the translated transcripts were then performed on constructed candidate HMM orthogroup classification profiles, and orthogroups yielding the best bitscore were assigned to the transcripts. Once unigenes were found that had high and differential expression in haustoria, phylogenies were estimated for their corresponding orthogroups. Amino acid alignments of sequences within these orthogroups (including any translated transcripts that were assigned to the orthogroup as described above) were generated with MAFFT (Katoh and Standley 2013a) and the corresponding DNA sequences were forced onto the amino acid alignments using a custom perl script. DNA alignments were then trimmed with trimAL (Capella-Gutierrez, et al. 2009) to remove sites with less than 10% of the taxa. Orthogroup alignments were required to contain transcripts with alignment coverage of at least 50%. Otherwise, the failing transcripts were removed from the orthogroup amino acids and DNA alignments, and the alignments were re-generated. This process was iterated until all of the sequences covered at least 50% of the alignment. Finally, maximum likelihood (ML) phylogenetic trees of DNA alignments for orthogroups containing parasite sequence(s) were generated using RAxML (Stamatakis 2006) with the GTRGAMMA model. To evaluate the reliability of the branches on the tree, 100 pseudosamples for the alignment were generated to estimate branch support using the bootstrap method (Felsenstein 1985).

2.4.12 Scoring gene duplications

Gene duplication events were scored by referring to each rooted gene tree. Genes from *Physcomitrella* and/or *Selaginella* were used as outgroups to root each tree; when these outgroups were not present in the orthogroup, *Amborella* and/or or monocots were used. Because our interest in this paper was focused on when *parasitism genes* evolved, we limited our analysis to gene trees that contain *parasitism genes*. Possible topologies showing gene duplications giving rise to *parasitism genes* are illustrated in supplementary data 15. Gene duplications were scored if a parasitism gene from one or more of the three parasitic species (*Triphysaria*, *Striga* and *Phelipanche*) were present in each duplicated clade, and if bootstrap values for key nodes met defined criteria. In addition, a sequence from at least one taxon had to be present in each duplicated clade. To illustrate, a *Mimulus*+*Orobanchaceae*-wide duplication (including nonparasitic *Lindenbergia* and three parasites- *Triphysaria*, *Striga* and *Phelipanche*), and shown as (((M1O1)bootstrap1, (M2O2)bootstrap2)bootstrap3), is required to meet the following criteria:

1) each clade defined by the nodes M1O1 and M2O2 contains at least one gene from the parasite taxa; 2) at least one taxon of *Mimulus* or Orobanchaceae has to be present in both clades defined by M1O1 or M2O2; 3) bootstrap2 and at least one of either bootstrap1 or bootstrap2 must be greater than or equal to the bootstrap stringency cutoffs of 50% or 80%. An example of a scored gene tree, with a supported gene duplication is given in supplementary figure S3.

2.4.13 Selective constraint analysis

To perform the constraint analysis to infer adaptive or purifying selection, PAML (Yang 1997, 2007) software based on maximum likelihood was utilized for hypothesis testing. The branch model in PAML was used to test if the foreground branch of interest has significantly different dN/dS ratio (omega, ω) compared to the background ω. The codeml tool in PAML was used to perform such analyses. To estimate significance of one particular hypothesis, a likelihood ratio test was used. The one–ratio model and branch model in codeml of PAML were used to test if the branch model fits the model significantly better than the one-ratio model. When the one-ratio model is correct, the distribution of the likelihood ratio test statistic follows a chi-square distribution with the degrees of freedom being equal to the number of additional parameters in the branch model test. The test statistics was calculated by taking twice the difference between log likelihood-values from the two tests. These models were fitted to examine which model was more appropriate for the data. The branch-site model was further used when the branch-model identified significant differences between the foreground and background lineages. The branch-site model was also used to identify sites under positive selection for the indicated foreground lineages. To perform the branch-site model, the codeml file was set to model = 2, NSsites=2. The null model was set to fix ω at 1, while the alternative model was set to estimate ω. Sites identified by PAML with a probability greater than 0.95 by Bayes Empirical Bayes (BEB) analysis were examined further through looking at the peptide alignment.

2.4.14 Expression of haustorial orthologs in nonparasitic species

We utilized the existing expression profile data for growth and developmental stages in *Arabidopsis thaliana*, *Medicago truncatula* and *Solanum lycopersicum*, which we refer to as *Arabidopsis*, *Medicago* and tomato in this analysis. The expression profiles for the *Arabidopsis* and *Medicago* genes were extracted from the PLEXdb database (Dash, et al. 2012) that contains microarray data [*Arabidopsis*, AT40: Expression Atlas of *Arabidopsis* Development (AtGenExpress); *Medicago*, ME1: The *Medicago truncatula* Gene Expression Atlas], while the data for tomato was from the digital expression (RNA-Seq) experiment (D007: Transcriptase analysis of various tissues in wild species *S. pimpinellifolium*, LA1589) in the Tomato Functional Genomics Database (Fei, et al. 2011). First, parasite genes with upregulated haustorial expression were identified, and their putative orthologs were obtained as the best blast hits within the same orthogroup in *Arabidopsis*, *Medicago* and tomato among the sequences. To find the expression information of the genes in *Arabidopsis* and *Medicago* used in the gene family analysis, we used the microarray expression information of probes for these genes by BLASTn. For tomato, expression for each gene was retrieved directly from the RNA-Seq database. To make the expression easily comparable across species, the expression values for similar tissues were averaged (supplementary data 20). In the *Arabidopsis* gene expression atlas, all vegetative_leaf and rosette_leaf were combined as "leaf", the sepal, petal, stamen, and carpel were combined as "flower", different root tissues (root7, root_MS1, root_GM) were combined as "root". In the *Medicago* gene expression atlas, expression data for tissue responding to nod factors (Nod 4d, Nod 10d, Nod 14d) were combined as "nod", root and root-0d as "root", and seed 10d, seed 12d, seed 16d, seed 20d, seed 36d as "seed". In the tomato RNA-Seq data, newly developed leaves and mature green leaflets were combined and labeled as "leaf", flower buds 10 days before anthesis or younger and flowers at anthesis (0DPA) as "flower", and fruit at10 DPA, 20 DPA, 33 DPA as "fruit". We scored the expression for upregulated haustorial orthogroups based on the principally expressed tissue. To do this, we divided the expression of a gene in a given tissue by its summed expression across all tissues to obtain a normalized expression in each tissue. Then we identified genes as "principally expressed" in a tissue if its normalized expression was ≥ 2 fold higher in that tissue than in any other. If similarly high expression was found in two tissues, both tissues were scored. Genes with broad expression in more than three tissues were not scored. In addition to this binary scoring of the principally expressed tissue for haustorial genes, we also scored each tissue quantitatively using the average of each tissue's expression across all haustorial gene

orthologs. We excluded genes whose highest expressions across all tissues was in the lower 25th percentile. We then averaged expression of each tissue across all genes and ranked each tissue based on the expression. All orthogroups were subject to this step and finally each tissue type that supported a possible haustorial origin was scored by the number of upregulated haustorial orthogroups and the average expression across all orthogroups.

Chapter 3

HGT in parasitic Orobanchaceae[3]

[3] A modified version (more brief) of this chapter has been prepared for submission as:

Zhenzhen Yang*, Yeting Zhang*, Eric Wafula, Loren A. Honaas, Paula E. Ralph, Sam Jones, Huiting Zhang, Naomi S. Altman, Michael P. Timko, John I. Yoder, James H. Westwood, Claude W. dePamphilis (2016) You are what you eat: Horizontal gene transfer is more frequent with increased heterotrophy and may contribute to parasite adaptation. Proc. Natl. Acad. Sci. U.S.A.

3.1 Introduction of HGT

3.1.1 HGT in bacteria

Horizontal Gene Transfer (HGT), as opposed to vertical transmission where organisms inherit their genetic material from their parental generation, is any process in which an organism acquire genes from another organism without being that organism's offspring (Richardson and Palmer 2007; Acuna et al. 2012). The phenomenon was first reported in bacteria in the 1950s, when multidrug resistance emerged on a worldwide scale because antibiotic resistance traits were transferred across taxa instead of generated *de novo* within each taxon (Davies and Davies 2010). Over time, numerous examples of HGT have been identified in bacteria, and HGT is now known to be common in bacterial evolution. A substantial amount of HGT is associated with plasmid-, phage- or transposon-related sequences (Ochman, et al. 2000), whereas fewer well-documented cases of HGT among eukaryotes (Keeling and Palmer 2008) are known. Many of these cases appear to result in short-lived, nonfunctional sequences (Feschotte and Pritham 2007; Keeling and Palmer 2008; Schaack, et al. 2010). Consequently, the long-term evolutionary impact of HGT in multicellular eukaryotes remains largely unknown.

3.1.2 Review of current methods for HGT identification

To understand the long-term impact of horizontal gene transfer, we examine if many examples of HGTs exist to indicate a pattern in favor of certain functional groups. Three methods have been commonly used to detect and explore putative HGT events: an approach based on codon bias, the BLAST-based approach, and a phylogenomic approach. The approach relying on codon bias has mainly been used in prokaryotes where codon biases can be strong and differ markedly among organisms (Sueoka 1962; Tuller 2011). This approach was used to identify recent transfer events in prokaryotes based on the fact that a newly introduced gene into the recipient genome often has a different codon bias (because they often reflect the codon usage and nucleotide composition of the host genomes) compared to the native genes. The BLAST-based approach (Zhaxybayeva 2009) predicts HGT candidates when the top blast hits of the HGT focal

taxa derive from distantly related species instead of its close relatives. Past studies have relied on the best BLAST hit for HGT identification; however, as best BLAST hit does not guarantee the closest neighbor on the phylogenetic tree (Koski and Golding 2001), it has been prone to errors. Phylogenomic analyses involve the use of large-scale phylogenetic trees to identify supported incongruence between a well-resolved species tree and a gene tree (Galtier and Daubin 2008). This method is more straightforward as it provides direct phylogenetic evidences for HGT inference. The primary drawback is being computationally expensive and only a few studies have applied this approach on parasitic plants (Xi et al. 2012a; Xi et al. 2013).

3.1.3 HGT in plants

HGT events have been well documented in several autotrophic plants, most often involving the acquisition of mitochondrial sequences. These mitochondrial sequences include *atp*1, *rps*2 and *rps*11 genes transferred among different angiosperm species (Bergthorsson et al. 2003). Transfers also occurred via repeated invasions of a group 1 homing intron of the *cox1* gene in a wide range of angiosperm genera and species (Cho et al. 1998; Sanchez-Puerta et al. 2011), or involve the transfer of a group II intron in *nad*1 gene and its adjacent exons (exon b and exon c) from an asterid to the gymnosperm *Gnetum* (Won and Renner 2003). A notable example showing massive horizontal acquisition of mitochondrial sequences was recently reported in *Amborella* (Cho et al. 1998; Bergthorsson et al. 2004; Rice et al. 2013), the first emerging flowering plant. *Amborella*'s large mitochondrial genome not only contains numerous genes obtained from a wide range of other angiosperms and non-angiosperms, but entire mitochondrial genomes from distantly related moss and algal species. HGT events reported in other autotrophic plants include transfer of MULE transposons among grasses (Diao et al. 2006), and the evolution of C4 photosynthetic pathways via HGT among Panicoid species (Christin, et al. 2012). Most recently, researchers have discovered that ferns horizontally acquired a neochrome gene from hornworts (Li, et al. 2014b).

The other major set of HGT events identified in plants involves parasitic plants. Parasitic plants invade their host plants' tissues through either the shoots or roots with a haustorium (Kuijt 1969), a novel organ for heterotrophic feeding, which allows parasitic plants to acquire nutrients and water from their host plants. The channel allowing the exchange of genetic and other biochemical compounds between parasites and their host plants increases the possibility of HGT

events (Richardson and Palmer 2007). Mitochondrial atp1 sequences have transferred from *Tetrastigma* to *Rafflesiacae* (Davis and Wurdack 2004). In another report, nad1 and matR were found to have transferred from Santalales to a fern (Davis et al. 2005). Researchers also reported that four endoparasitic lineages acquired a mitochondrial atp1 from their host lineages, and repeated horizontal transfers of group I cox1 intron were discovered in parasitic plants (Barkman et al. 2007). In addition to mitochondrial genes, a plastid gene was transferred from the parasitic *Orobanche* to *Phelipanche* (Park, et al. 2007a). Several transfers of nuclear genes were also discovered: a nuclear gene of unknown function from *Sorghum bicolor* to *Striga hermonthica* (Yoshida et al. 2010), a legume-specific albumin 1 from legumes to both genera of *Phelipanche* and *Orobanche* (Zhang, et al. 2013b), a Brassicaceae-specific strictosidine synthase-like gene from Brassicaceae to both *Phelipanche aegyptiaca* and *Cuscuta australis* (Zhang, Qi, et al. 2014a).

With an increase in available genomic and transcriptomic data for plants, it is now possible to detect HGT events *en masse* using a comprehensive phylogenomic approach. To identify HGTs in *Rafflesia cantleyi*, an endophytic holoparasite which lacks leaves and stems and only attaches to members of the grapevine family (Vitaceae), Xi et al (Xi, et al. 2012b) generated both transcriptomic data for both the parasite (*Rafflesia*) and its obligate host (*Tetrastigma*). A phylogenomic approach of constructing phylogenies from nine sequenced plant genomes plus the parasitic *Rafflesia* allowed Xi et al (Xi et al. 2012a) to conclude that *Rafflesia* horizontally acquired several dozen of its actively transcribed nuclear genes from *Tetrastigma*. Later, the same group reported massive mitochondrial gene transfer in the same species using a similar approach (Xi et al. 2013): 11 of 38 examined mitochondrial genes showed phylogenetic patterns suggestive of HGT. While Rafflesiaceae has a fairly narrow host plant range, other parasite lineages have much broader host preferences (Naumann et al. 2013). Thus, we are interested in discovering if other parasitic lineages that are widely distributed and interact with a wide range of host plants undergo HGT frequently and result in exchanges of genes across a much broader set of plants.

3.1.4 Host range and HGT detection

The three parasitic plant species in Orobanchaceae have widely varied host plant preferences. *Triphysaria versicolor*'s host plants range from monocots to dicots (Estabrook and

Yoder 1998; Jamison and Yoder 2001). *Striga hermonthica*, and most other *Striga* species exclusively parasitize member of Poaceae (grasses) (De Groote et al. 2008; Parker 2009), while *Phelipanche aegyptiaca* parasitizes a range of dicot hosts from rosid and asterid angiosperm lineages, including the Solanaceae tomato, potato, eggplant, tobacco, crops in Fabaceae (beans and other legumes), Apiaceae and Asteraceae (Carlon et al. 2005; Schneeweiss 2007; Parker 2009). Because *Striga hermonthica* has a relatively narrow range of host plant species, a stringent BLAST-based screening for HGT candidates is feasible (Yoshida et al. 2010). Zhang et al. (2013b) reported a unique gene, *albumin1*, identified in *Phelipanche aegyptiaca* and other related broomrape species with BLAST (Zhang et al. 2013b). However, for parasitic plants with a wide range of host plant species, such as *Triphysaria versicolor*, the previous BLAST-based HGT screening approach is less effective. Furthermore, ancient HGT events may have taken place in an ancestor of the parasite in question that fed on host plants different from the current species. Therefore, a phylogenomic approach (Jiao et al. 2011; Xi et al. 2012b), which leverages genome scale data to analyze every parasite gene in a phylogenetic context, is an efficient and powerful approach to detect evidence of HGT.

3.1.5 Objectives and overview of the analyses

In this study, with a goal of detecting functional transfers from the hosts to the parasites, we implemented a phylogenomic approach to detect cases of HGT in transcribed sequences in members of Orobanchaceae that together represent a wide range of parasitic ability, including the parasites *Triphysaria versicolor*, *Striga hermonthica*, and *Phelipanche aegyptiaca*, and the closely related nonparasitic *Lindenbergia philippensis*. Stringent screening plus careful verification were used to identify a number of high-confidence HGT events. The goals of this study will address the following questions: 1) How frequently has HGT resulted in expressed transgenes in the Orobanchaceae? 2) Do the HGT events detected involve only known host plants or other plant lineages? 3) What is the mechanism involved in the HGT event, e.g., has HGT occurred through an RNA intermediate or direct transfer of the genomic fragment? 4) Are any HGT events shared among the three parasitic species, suggesting transfers involving an ancestral parasite taxon? 5) Are there any HGT genes that appear to have a function related to parasitism? Through a comprehensive study of all possible HGT genes, we hope to shed light on whether HGT plays an adaptive role in parasite evolution.

3.2 Results

3.2.1 The phylogenomic pipeline and the analytical schema for HGT detection

Most of the currently reported HGTs have been identified by a BLAST-based approach, which is mostly due to short running time of a single BLAST search. However, because the best BLAST hit does not guarantee the closest neighbor on the tree (Koski and Golding 2001), and the BLAST results have to be routinely verified by a phylogenetic tree, a phylogenomic approach is the main focus of this analysis. At first, large-scale phylogenetic trees were reconstructed for every unigene in four focal species of Orobanchaceae, including the nonparasite *Lindenbergia philippensis,* and three parasites -*Triphysaria versicolor, Striga hermonthica,* and *Phelipanche aegyptiaca.* A total of 13254 orthogroup trees that involves taxa from 22 sequenced plant genomes and two large EST datasets from Asteraceae (*Lactuca sativa* and *Helianthus annuus*), as well as the four Orobanchaceae taxa were used as the basis for HGT identification.

Customized python scripts were applied to the 13,245 orthogroup phylogenies to automatically identify gene trees where the parasitic genes were unexpectedly placed within a putative donor clade. The phylogenetic position of Orobanchaceae is well established within the Lamiales (Young and dePamphilis; Olmstead et al; Soltis et al 17 gene) containing Mimulus guttatus (Phyrmaceae) (Olmstead et al. 2001; Soltis, et al. 2011), the closely related species with a sequenced genome (Hellsten, et al. 2013). Due to complexities of gene tree evolution such as gene loss, incomplete lineage sorting, as well as insufficient taxon sampling within this group, we searched for HGTs only from distantly related taxa in this study - rosids and monocots.

Three models illustrate topologies indicative of HGTs we sought to detect (Fig. 1). With a goal of identifying unambiguous HGTs, we focused our search on rosid and monocot donors because of the relatively large number of finished genomes in these groups, and the greater genetic distance from our Orobanchaceae focal group. In all three models, we identified "ancestral" nodes, defined here as the node containing exclusively parasite genes and genes in the donor clade. The first model describes a scenario where genes from the parasitic plant or nonparasitic relative are strongly supported as nested within a donor clade (Figure 3-1A). The second model (Figure 3-1B) describes a case where the parasite's genes are placed outside of the donor clade. In both cases, two nodes (with bootstrap >50) supporting the grouping of parasitic

genes with donor clades were required. The third model (Figure 3-1C) requires only one node supporting the placement of the parasite's gene(s) as sisters of donor clades.

3.2.2 52 high-confidence HGT events

Custom scripts searching for topologies consistent with these three models resulted in the identification of a set of 192 gene trees with preliminary evidence for HGT (143 orthogroups from rosids and 49 orthogroups from monocots) from the three focal genera. Only one orthogroup (3861) was identified as including a potential HGT from the nonparasitic "control" species *Lindenbergia*, all others involved the parasitic species only. We then applied a scoring scheme to assign these 192 candidate HGT trees to low-, medium-, and high-confidence groups based on tree characteristics including the bootstrap support for key nodes surrounding the inferred HGT event, sampling of the donor clades grouped with the HGT genes, and branch length heterogeneity (Figure 3-3-1D, see Table 3-1 for detailed criteria). The authenticity of each HGT tree in the medium- and high-confidence groups was then validated with follow-up analyses, including manual examination of branch lengths and sequence alignments, correcting any translation or alignment errors, and examining the phylogenetic stability with increased taxon sampling. We examined low coverage transcriptome sequences from eight more parasitic Orobanchaceae species, and also from related non parasitic species from the 1kp transcriptome project (Matasci, et al. 2014) four sequenced asterid genomes (Phytozome) (Goodstein, et al. 2012), and the *Striga asiatica* genome (Yoshida and Shirasu, pers. comm.) were also included (see Materials and methods "HGT validation by increased taxon sampling"). This resulted in a final set of 42 HGT orthogroups (Table 3-2), with the remaining 158 putative HGTs determined to be artifactual or merely low confidence (Figure 3-1E and Figure 3-2). The primary artifact (106 out of 158) arose from insufficient taxon sampling (Figure 3-2A, B), especially of the order (Lamiales) that contains Orobanchaceae (Soltis et al. 2009). Other artifacts came from frame-shift errors (Figure 3-2D, E) and contamination, either from the experimental host (nine orthogroups) (Figure 3-2C) or from fungal contamination (one orthogroup). The topology and bootstrap values (bs) for the 42 HGT orthogroup trees strongly support the placement of parasitic genes within the donor clade, indicating a clear HGT origin (Figure 3-3A, 3-4A). 11 out of 42 trees suggested a polyphyletic origin of HGT genes (one or more transfers in a gene family), thus we used the

Shimodaira-Hasegawa (SH) test to evaluate whether more than one transfer was most likely based on the available data (Table 3-3). A single transfer could not be rejected for orthogroup 3861 (Table 3-3). Interestingly, all the remaining trees did support more than one transfer (Table 3-3), suggesting a propensity for certain gene families to include fixed horizontal transfers. A minimum of 52 horizontal transfer events were thus inferred from these 42 gene families. That a majority, fully 78%, of the candidates (67% of initial medium and high confidence) were 1) excluded due to low support, 2) identified as artifacts (via increased taxon sampling), or 3) identified as contamination from host or other organisms, illustrates the challenge of HGT discovery.

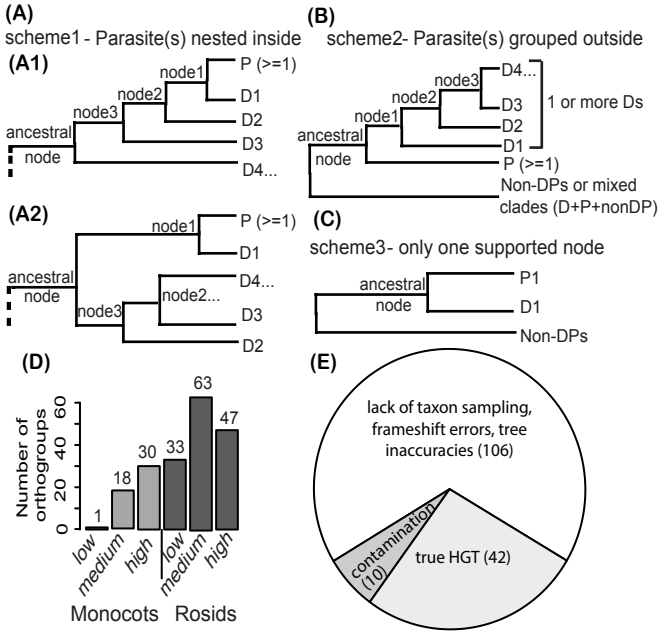


Figure **3-1**. Three models for phylogenomic identification of HGTs, and further examination of the preliminary-screened HGT candidates. Scheme 1, parasite genes (P) are nested inside donor clades (D); scheme 2, parasitic genes group outside of the donor clade; scheme 3, only one node of donor sequence is sister to parasitic genes. In this study, donor refers to distantly related monocot and rosid sequences. Ancestral node is defined to be composed of exclusively parasitic and donor sequences. In A1, at least two nodes within the ancestral node (including the ancestral) are required to have bootstrap support (BS) >=50, in A2, both the ancestral node and node 1 are required to have BS >=50. In scheme 2 (B), the ancestral node and at least one node within the ancestral node are required to have BS >=50. Scheme 3 (C), only the

node that supports the grouping of the parasitic gene and donor sequence is required to have BS>=50. "Non-DPs" refers to non-parasitic, non-donor sequences. (D) A number of 192 HGT orthogroup trees from the initial screening were classified into low-, medium-, and high-confidence categories based on a scoring scheme (table S2). Grey colors represent the HGT orthogroups identified in the monocots, darker grey represents the rosids. (E) The number of HGT-candidate orthogroups manually curated as true HGTs (light grey), artifacts resulting from insufficient taxon sampling, frame shift errors or tree inaccuracies (white), or fungal or host contamination (dark grey).

Table **3-1.** A scoring scheme used to score each phylogenetic tree based on bootstrap support, depth of donor clades, and long branches.

Criterion1		**Criterion2**		**Criterion3**		**Scoring**	
Bootstrap support	score	Sampling depth	score	Branch length	score	Summed score	confidence
x=100	10	rich sampling in one inner node and strong support	5	branch length not long	3	x≤9	low
90≤x<100	8						
80≤x<90	6	rich sampling in one inner node and low support	3	branch length < 2 times of the average	2	10≤x≤14	medium
70≤x<80	4						
60≤x<70	3						
50≤x<60	2	mixed samples in inner nodes	1	branch length >=2 times	1	15≤x≤18	high
x<50	1						

3.2.3 Transfers from ancestral host lineages

A majority of these HGTs could be assigned to ancestral donors from known host lineages. All the HGTs from grass donors (Poaceae) were discovered in *Striga* (Table 3-1 and

Figure 3-3C), which (except for *S. gesnerioides*) are specialized parasites of Poaceae (Musselman 1980). In *Phelipanche*, however, inferred donors reflected a wide range of dicot families with the majority from Rosaceae and Fabaceae, also consistent with feeding preferences for this plant and its congeners (Westwood et al. 2010) (Figure 3-3C). In 38 orthogroups, the transfer was inferred to be unique to one genus (15 are unique in *Phelipanche*, 8 are unique in *Striga*), or in two closely related genera (15 occurred both in *Phelipanche* and *Orobanche*).

Any HGT events leading to the origin of parasitism would have occurred in a common ancestor of the parasites. Previously reported cases of HGT in microbial parasites or pathogens primarily encode cell wall-degrading enzymes of plants (Keeling 2009) and thus are implicated in host invasion. Surprisingly, although cell wall modifying enzymes are well represented in haustorial tissues (Yang et al. 2015), no such proteins were identified in our HGT search. Instead, numerous proteins involved in cell wall modification processes in the haustorium were attributed to gene duplications that occurred in an ancestor of all parasitic lineages of Orobanchaceae (Yang et al. 2015). Our HGT phylogenies, however, supported predominantly recent occurrences that were unique to individual genera. In only one case (orthogroup 218), the transfer was detected in almost all the parasitic taxa (*Phelipanche*, *Striga*, *Triphysaria*, *Alectra*) (Figure 3-3D), but the SH test indicated that this likely involved at least two (more recent) transfers instead of a single ancestral HGT event (Table 3-3). Therefore, while gene duplications often preceded and underpin the origin of parasitism in Orobanchaceae (Yang et al. 2015). HGT events are more recent, and may have contributed to subsequent parasite adaptation (see below).

3.2.4 Increased numbers of HGT with increased heterotrophic dependence

The number of HGT events appears to increase in parasites with greater host dependence. In the free-living sister lineage to all parasitic Orobanchaceae, *Lindenbergia*, we detected only

one HGT. In *T. versicolor*, the facultative hemiparasite, two HGT events were found (Table 3-2). In *Striga*, the obligate hemiparasite, ten orthogroup trees support HGTs and seven were from grasses in the Poaceae family. A majority (34 orthogroups) of the HGTs were detected in *Phelipanche*, the obligate holoparasite with the strongest host dependence (Fig3-3C, D).

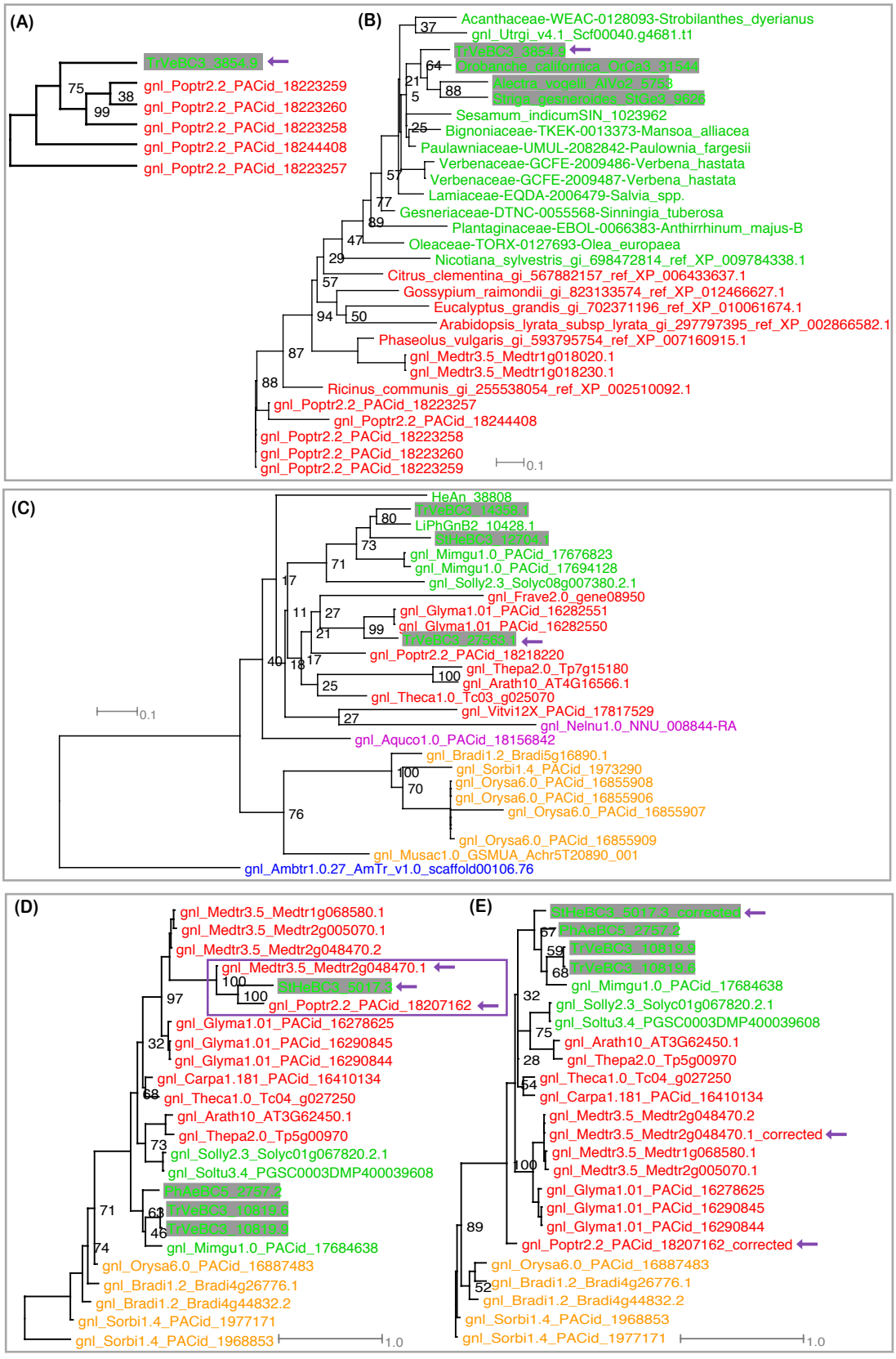

(A)
TrVeBC3_3854.9
gnl_Poptr2.2_PACid_18223259
gnl_Poptr2.2_PACid_18223260
gnl_Poptr2.2_PACid_18223258
gnl_Poptr2.2_PACid_18244408
gnl_Poptr2.2_PACid_18223257
(B)
Acanthaceae-WEAC-0128093-Strobilanthes_dyerianus
gnl_Utrgi_v4.1_Scf00040.g4681.t1
TrVeBC3_3854.9
Orobanche_californica_OrCa3_31544
Alectra_vogelii_AlVo2_5758
Striga_gesneroides_StGe3_9626
Sesamum_indicumSIN_1023962
Bignoniaceae-TKEK-0013373-Mansoa_alliacea
Paulawniaceae-UMUL-2082842-Paulownia_fargesii
Verbenaceae-GCFE-2009486-Verbena_hastata
Verbenaceae-GCFE-2009487-Verbena_hastata
Lamiaceae-EQDA-2006479-Salvia_spp.
Gesneriaceae-DTNC-0055568-Sinningia_tuberosa
Plantaginaceae-EBOL-0066383-Anthirrhinum_majus-B
Oleaceae-TORX-0127693-Olea_europaea
Nicotiana_sylvestris_gi_698472814_ref_XP_009784338.1
Citrus_clementina_gi_567882157_ref_XP_006433637.1
Gossypium_raimondii_gi_823133574_ref_XP_012466627.1
Eucalyptus_grandis_gi_702371196_ref_XP_010061674.1
Arabidopsis_lyrata_subsp_lyrata_gi_297797395_ref_XP_002866582.1
Phaseolus_vulgaris_gi_593795754_ref_XP_007160915.1
gnl_Medtr3.5_Medtr1g018020.1
gnl_Medtr3.5_Medtr1g018230.1
Ricinus_communis_gi_255538054_ref_XP_002510092.1
gnl_Poptr2.2_PACid_18223257
gnl_Poptr2.2_PACid_18244408
gnl_Poptr2.2_PACid_18223258
gnl_Poptr2.2_PACid_18223260
gnl_Poptr2.2_PACid_18223259
0.1
(C)
HeAn_38808
TrVeBC3_14358.1
LiPhGnB2_10428.1
StHeBC3_12704.1
gnl_Mimgu1.0_PACid_17676823
gnl_Mimgu1.0_PACid_17694128
gnl_Solly2.3_Solyc08g007380.2.1
gnl_Frave2.0_gene08950
gnl_Glyma1.01_PACid_16282551
gnl_Glyma1.01_PACid_16282550
TrVeBC3_27563.1
gnl_Poptr2.2_PACid_18218220
gnl_Thepa2.0_Tp7g15180
gnl_Arath10_AT4G16566.1
gnl_Theca1.0_Tc03_g025070
gnl_Vitvi12X_PACid_17817529
gnl_Nelnu1.0_NNU_008844-RA
gnl_Aquco1.0_PACid_18156842
gnl_Bradi1.2_Bradi5g16890.1
gnl_Sorbi1.4_PACid_1973290
gnl_Orysa6.0_PACid_16855908
gnl_Orysa6.0_PACid_16855906
gnl_Orysa6.0_PACid_16855907
gnl_Orysa6.0_PACid_16855909
gnl_Musac1.0_GSMUA_Achr5T20890_001
gnl_Ambtr1.0.27_AmTr_v1.0_scaffold00106.76
0.1
(D)
gnl_Medtr3.5_Medtr1g068580.1
gnl_Medtr3.5_Medtr2g005070.1
gnl_Medtr3.5_Medtr2g048470.2
gnl_Medtr3.5_Medtr2g048470.1
StHeBC3_5017.3
gnl_Poptr2.2_PACid_18207162
gnl_Glyma1.01_PACid_16278625
gnl_Glyma1.01_PACid_16290845
gnl_Glyma1.01_PACid_16290844
gnl_Carpa1.181_PACid_16410134
gnl_Theca1.0_Tc04_g027250
gnl_Arath10_AT3G62450.1
gnl_Thepa2.0_Tp5g00970
gnl_Solly2.3_Solyc01g067820.2.1
gnl_Soltu3.4_PGSC0003DMP400039608
PhAeBC5_2757.2
TrVeBC3_10819.6
TrVeBC3_10819.9
gnl_Mimgu1.0_PACid_17684638
gnl_Orysa6.0_PACid_16887483
gnl_Bradi1.2_Bradi4g26776.1
gnl_Bradi1.2_Bradi4g44832.2
gnl_Sorbi1.4_PACid_1977171
gnl_Sorbi1.4_PACid_1968853
1.0
(E)
StHeBC3_5017.3_corrected
PhAeBC5_2757.2
TrVeBC3_10819.9
TrVeBC3_10819.6
gnl_Mimgu1.0_PACid_17684638
gnl_Solly2.3_Solyc01g067820.2.1
gnl_Soltu3.4_PGSC0003DMP400039608
gnl_Arath10_AT3G62450.1
gnl_Thepa2.0_Tp5g00970
gnl_Theca1.0_Tc04_g027250
gnl_Carpa1.181_PACid_16410134
gnl_Medtr3.5_Medtr2g048470.2
gnl_Medtr3.5_Medtr2g048470.1_corrected
gnl_Medtr3.5_Medtr1g068580.1
gnl_Medtr3.5_Medtr2g005070.1
gnl_Glyma1.01_PACid_16278625
gnl_Glyma1.01_PACid_16290845
gnl_Glyma1.01_PACid_16290844
gnl_Poptr2.2_PACid_18207162_corrected
gnl_Orysa6.0_PACid_16887483
gnl_Bradi1.2_Bradi4g26776.1
gnl_Bradi1.2_Bradi4g44832.2
gnl_Sorbi1.4_PACid_1968853
gnl_Sorbi1.4_PACid_1977171
1.0

Figure **3-2**. HGT artifacts due to insufficient taxon sampling (A-B), contamination (C), and frame-shift errors (D-E). RAxML-based maxiumum likelihood tree for orthogroup 20348 from the initial automated pipeline (A) shows the placement of a parasitic TrVeBC3_3854.9 gene (purple arrow) as sister of rosid clades composed of many Poplar sequences (Poptr), indicative of HGT. Increase of taxon sampling in the Lamiales order (green non-shading) and the rosid groups (red) converts the *Triphysaria* gene (TrVeBC3_3854.9) to be a vertically inherited gene which groups with its closely related parasitic taxa (green foreground, grey shading) in the Lamiales order. Contamination rather than HGT (orthogroup 8224) explains the placement of the *Triphysaria* gene (TrVeBC3_27583.1) with Glyma sequences (C), which is a close relative of its experimental donor – *Medicago truncatula*. (D-E) show an HGT artifact due to frame-shift errors for orthogroup (20348). (D) – Purple branch shows a clade with long branch composed of three sequences – StHeBC3_5017.3, gnl_Medtr3.5_Medtr2g048470.1, and gnl_Poptr2.2_PACid_18207162. Striga sequence was screened as an HGT sequence because of strong placement with the rosid donor. Careful examination of the alignment revealed frame-shift errors of all these three sequences, and repair of this error resolved the *Striga* gene as a vertically inherited gene (E). The two rosid sequences from *Poplar* and *Medicago* also went to its expected position within the rosid clades without long branch (E).

Table **3-2**. Information of the 42 HGT orthogroups including the HGT recipient, donor, expression, dN/dS, functional category, and homology-based annotation.

ortho group	**recipient**	**donor**	**intron**	**expression**	**dN/dS**	**functional category**	**annotation based on homology**
226	P	Poptr	Y/Y	1, 42	P	defense	cytochrome P450
1685	P	Gyma+Medtr	Y/Y	>2	P	defense	cysteine-rich receptor-like kinase
2376	P	Poptr+Theca	Y/Y	>2	RP	defense	Proteasome subunit alpha type
14624	S	Sorbi+Zea	N/N - 5'UI	NA	P	defense (disease resistance)	BTB/POZ
23343	P	Theca	-	52	P	defense (disease	disease resistance protein
11841	P	Frave	Y/Y	51	SP	defense	hyoscyamine 6-dioxygenase-like
1886	P	Frave	-	62	RP	defense	Ankyrin repeat family protein
11437	P	Frave	-	>2	RP	defense and nodule development	Kelch modif related to galactose oxidase
8888	P	Frave	-	2, 41	RP	transcription	Poly A polymerase
18709	P	Arath	Y/Y	int	POS	transcription	nucleolin 2-like
806	P	Theca	Y/Y	int	RP	translation	valyl-tRNA synthetase
2270	P, S	Theca	Y/Y	>2	RP	translation	methionyl-tRNA synthetase
4067	P	Frave	Y/Y	41	RP	translation	$tRNA^{His}$ guanylyltransferase
10050	P	Frave	-	42	RP	translation	histidine-tRNA ligase
13892	P	Medtr	Y/Y	int	P	translation	Ribosomal protein S13

17	P	Betvu	Y/Y	42, 51, 62	P	nutrient transport	ABC transporter C family member 3
9613	P	Glyma	Y/Y	0	RP	nodule	cytosolic purine 5-nucleotidase
15246	P	Medtr	-	62	P	defense-related (insect toxin)	Albumin I (Zhang et al. 2013a)
1226	P	Poptr	Y/Y	3, int	RP	diverse	alpha/beta-Hydrolases
10143	P	Frave	-	42	RP	diverse	Tubulin-specific chaperone D
3861	P	Glyma	Y/Y	>2	POS	diverse	Poly(ADP-ribose)
4598	P	Medtr	-	int	RP	diverse	nuclear pore complex protein
19696	P	Poptr	Y/Y	41, 62	SP	diverse	ubiquitin-like-specific protease
16703	S	Orysa	Y/Y	52	P	diverse	Zinc finger, GRF-type
4572	P	Frave	-	3, 41	P	diverse	FBD-associated F-box protein
5896	S	Glyma	Y/Y	52, 61	P	plastid-to-nucleus signaling	uroporphyrinogen-III synthase
218	P, S,	Prunus	Y/Y	>2	P	TE	hAT transposon
1021	P, T	Frave+Malus	Y/Y	>2	P	TE	hAT transposon
5002	P	Prunus	Y/Y	int	P	TE	hAT transposon
12835	S	Sorbi	N/N	0	P	TE	Putative harbinger transposase-
14230	P	Prunus	Y/Y	42	P	TE	MULE transposase
15149	P	Frave	Y/Y	int	P	TE	hAT transposon
13512	P	Frave	-	51	POS	unknown	unknown
13656	S	Sorbi+Orysa	-	int	SP	unknown	hypothetical protein
14233	S	Sorbi+Orysa	-	61	SP	unknown	unknown
14675	P	Frave	-	62	P	unknown	unknown
18354	P	Frave	-	int	P	unknown	unknown
18774	S	Orysa	Y/Y	0	P	unknown	unknown
19297	S	Bradi	N/N-3'UI	NA	P	unknown	unknown (Yoshida et al. 2010)
20188	P	Frave	Y/Y	int	P	unknown	unknown
20190	P	Frave	-	int, 2	P	unknown	unknown
23480	P	Frave	-	int	P	unknown	unknown

recipient column: P – *Phelipanche*, S – *Striga*, T – *Triphysaria*; Intron column: "-" means not determined, Y/Y represents presence of intron in both donor and recipient gene, N/N represents absence of introns in both donor and recipient; 3'UI and 5'UI mean 3'/5'-UTR introns. Expression: int – interface, - >2 means highly expressed in more than two stages; dN/dS: P – purifying selection, RP – relaxed purifying selection, SP – stronger purifying selection; POS – positive selection; functional category: TE – transposable element.

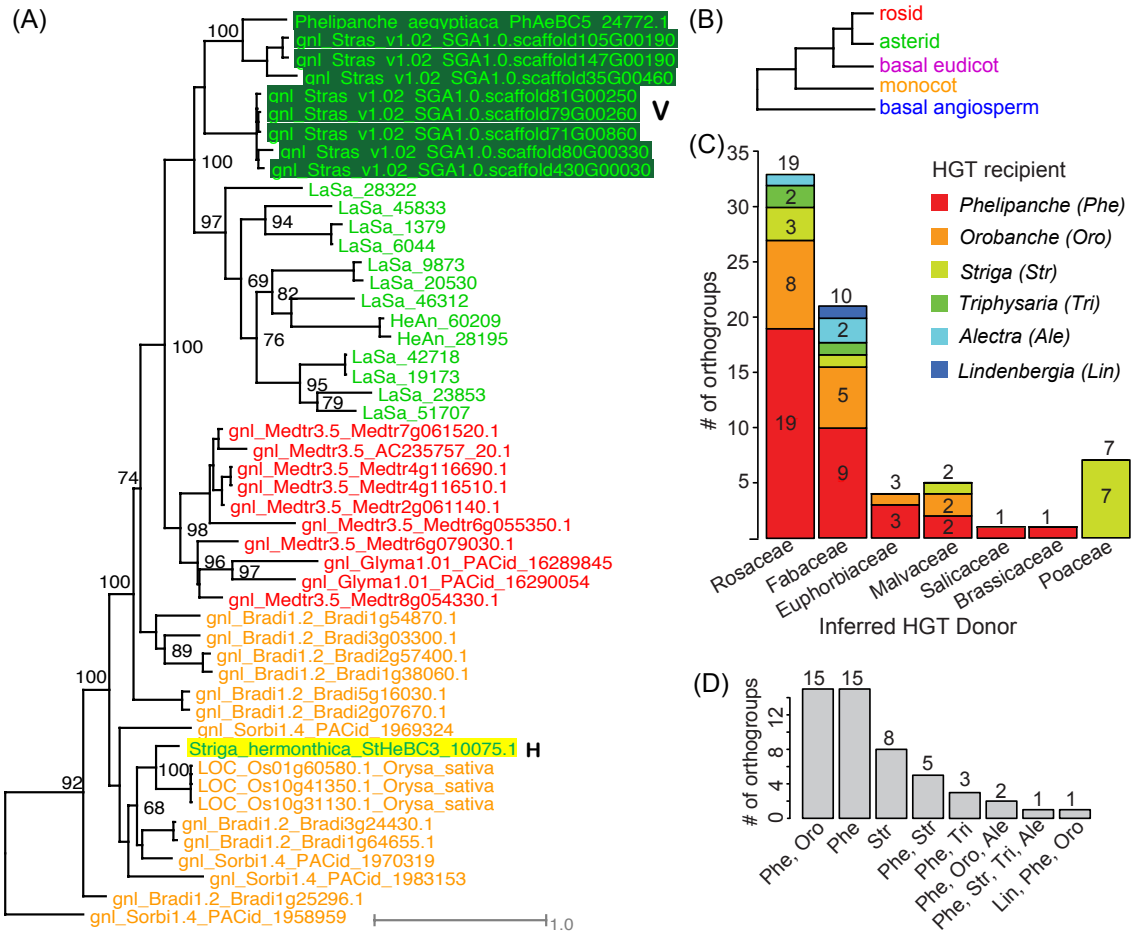


Figure **3-3**. RAxML-based Maximum likelihood (ML) trees supporting HGT in two orthogroups, donor families and recipient taxa inferred from the 42 HGT set. Orthogroup trees support one *Phelipanche* HGT from a eudicot donor (*Frave–Fragaria vesca*) (A), and a *Striga* HGT from a grass donor (*Brachypodium*) (B). A hypothetical tree illustrates the color-coding system for each angiosperm lineage represented in (A) and (B) is shown in (A) outlined with a box. (C) Mapping of parasitic recipient taxa onto inferred donor family (X-axis). Each genus in HGT recipient is followed with a three-letter code used in (D). Total number of HGT orthogroups inferred from each donor family are placed on top of each bar. Numbers within each bar represent number of orthogroups, the number of singletons is not shown due to space limitations. (D) Number of HGT orthogroups support transfers from shared and unique parasitic genera.

3.2.5 Integration of genomic fragments

Signatures of the donor molecule should persist in the genome, giving clues to the mechanism of transfer. For instance, a nuclear HGT reported in *Striga* supports a possible mRNA mediated transfer since the HGT lacked introns and seemed to contain a remnant poly-A tail, whereas the donor *Sorghum* gene lacked a poly-A tail (Yoshida et al. 2010). Documented translocation of

Table **3-3.** SH test to evaluate number of transfers in HGT trees by constraining multiple HGT genes to one monophyletic clade.

orthogroup	logL (original)	logL (constrained)	Significant (0.01)?	# of transfers
218	-106049.78	-106490.58	Y	≥ 2
1021	-40178.73	-40281.56	Y	≥ 2
3861	-42295.88	-42296.95	N	1
5002	-95429.05	-95694.82	Y	≥ 2
8888	-35555.79	-35813.86	Y	≥ 2
11437	-121967.75	-122293.12	Y	≥ 2
13512	113878.47	-114045.44	Y	≥ 2
14233	-88970.83	-89052.98	Y	≥ 2
18354	-11889.45	-11941.91	Y	≥ 2
18774	-48200.19	-48661.65	Y	≥ 2
19297	-48200.19	-48661.65	Y	≥ 2

host RNA into *Triphysaria* (Tomilov et al. 2008) and *Phelipanche* (Aly et al. 2009), as well as the massive movement of host RNA into *Cuscuta* would support an RNA-based mechanism for HGT in parasitic plants (Kim, et al. 2014). In contrast, a horizontally-acquired *albumin 1* gene in *Phelipanche* and related taxa (Zhang et al. 2013a) and horizontally acquired Brassicaceae-specific strictosidine synthase-like (SSL) genes contained introns in genomic sequences of both donor and

the parasites (*Phelipanche* and *Cuscuta*), all consistent with direct genomic transfers without an RNA intermediate (Zhang, Qi, et al. 2014b). To test the hypothesis of mRNA mediated transfer we examined the coding sequence structure (exon-intron boundaries) in the 42 HGT orthogroups. We had sufficient genomic data to examine 26 of the HGT orthogroups (Table 3-1), though three orthogroups contained HGT genes that lacked CDS introns in both donor and recipient. While these gene lacked CDS introns, 2 of these had introns in the UTRs. A gene in Orthogroup 14624 (BTB/POZ domain containing protein) was transferred from *Sorghum bicolor* into *Striga hermonthica*, and the 5'-UTR intron shows 87% sequence identity between the donor and recipient gene (CDS: 91%, 5'-UTR: 87%, 3'-UTR: 68%). In the other case, a gene in orthogroup 19297, a conserved 3'-UTR intron (3'-UTR intron: 78%, CDS: 85%, 5'-UTR: 54%, 3'-UTR: 82%) is present in both the donor and recipient. It is noteworthy that this HGT event was previously identified by Yoshida et al. (2010) who speculated, in part, based upon a remnant poly-A tail in the cDNA, that this HGT event may have been mediated by an mature mRNA rather than a genomic fragment. Our analyses identified the presence of a high identity intron in the 3' UTR, suggesting that this event (like the majority of cases reported here) was mediated by a genomic fragment rather than an mRNA. Only one orthogroup (12835 – a *Pong*-like transposable element), lacked intron in both the donor and recipient gene, and the non-conserved flanking region failed to provide evidence whether genomic or mRNA-mediated transfer was supported. The remaining 23 HGT orthogroups contained genes whose donor and recipient contained CDS introns. We further reduced the list to 13 orthogroups with full-length genes allowing us to examine similarities and differences in intron positions and sequences between donor and recipient.

Surprisingly, all 13 showed congruence of CDS structure between donor and recipient, suggesting a transfer of a genomic fragment containing the gene, rather than an mRNA intermediate. Intron positions are overall quite conserved (Table S5, S6[4], Fig. S3) (although with occasional intron loss, Fig. 3B), suggesting maintenance of intron structure for functional transcription. To infer the intron origin, we constructed phylogenies using the intron sequences only and compared them to phylogenies constructed with exons only. Only three orthogroup phylogenies were well-resolved due to a high level of intron sequence divergence (orthogroup 4067, 806, and 2270 -supplementary table S1). Reassuringly, the intron phylogenies were congruent with the CDS phylogenies, indicating the same donor lineage and providing strong support of a genomic fragment-mediated HGT (Figure 3-4 and Figure 3-5).

The strongest example, orthogroup 4067 (tRNA(His) guanylyltransferase - required for translation (Heinemann, et al. 2012)), not only exhibits strong CDS similarity with its inferred *Fragaria* donor (~74%) (Figure 3-3A), but the intron sequences maintain ~51% similarity (Figure 3-3B, 3C), even higher than that between the vertically inherited parasite gene and its close relative in *Mimulus* (~21%) (Figure 3-3C). These results show that all of the resolvable HGT events were likely mediated by genomic fragments containing most or all of donor genes rather than by RT-mediated transfer.

[4] All the supplementary data are not shown in this dissertation as there are too many, please refer to the submitted manuscript.

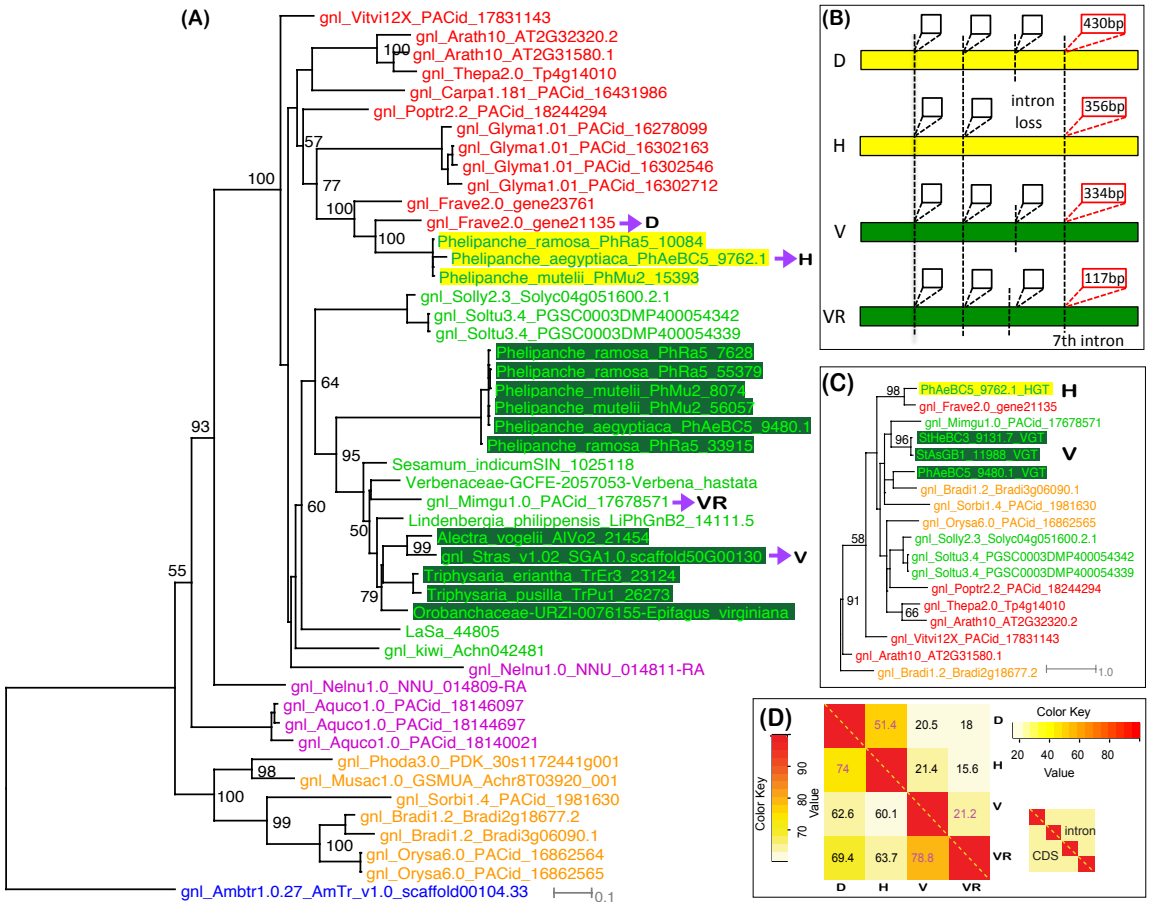
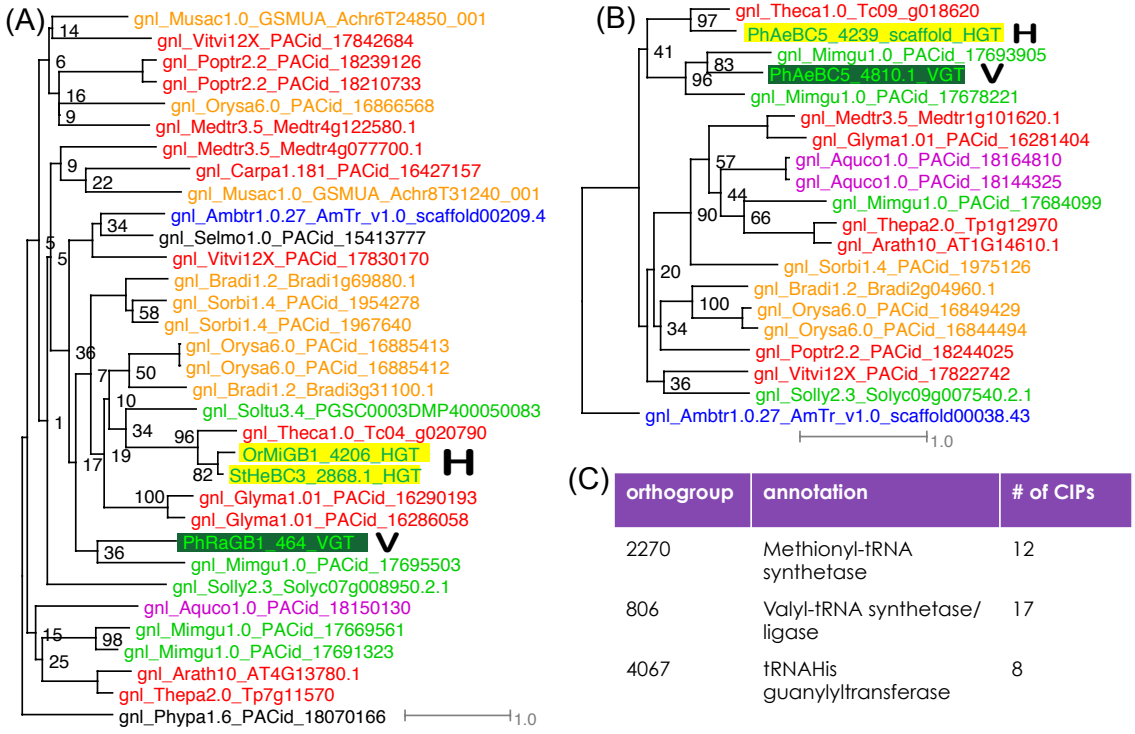


Figure 3-4. Genomic horizontal transfer of a tRNAHis guanylyltransferase from ancestor of *Fragaria* to *Phelipanche* parasites. (A) A coding-sequence (CDS) tree by RAxML from represented species across angiosperm lineages, H: the parasitic HGT gene, D: inferred donor (in *Fragaria*), V: vertical parasitic gene, VR: related sequence of the vertical parasitic gene (in *Mimulus*). (B) Gene structure with four selected introns for the four sequences (H, D, V, VR). Yellow and green bars represent coding sequence, the vertical dashed lines represent the intron positions; the boxes represent introns. At least four conserved intron positions were shown on the gene structure; the third intron was lost in the HGT gene, the fourth intron on the graph (which is the seventh intron of the *Mimulus* gene) showed strong sequence similarity between the HGT gene and its donor (marked by red intron boxes with length within). (C) The seventh intron (marked red in B) sequence phylogeny of genes on the CDS tree: the HGT gene groups with its donor supported by 98% bootstrap support (BS), whereas the vertically inherited genes group with its close relative (*Mimulus* sequence). (D) A heat map shows the pairwise sequence similarities for CDS (below diagonal) and the last

intron region (above diagonal) among the four genes (H, D, V, VR). CDS similarity between the HGT gene (H) and its donor sequence (D) is 74%, between the vertically inherited parasitic gene (V) and its relative (VR) is 78.8%; intron similarity for the former pair is 51.4%, for the latter is 21.2%.



(C)

orthogroup	annotation	# of CIPs
2270	Methionyl-tRNA synthetase	12
806	Valyl-tRNA synthetase/ ligase	17
4067	tRNAHis guanylyltransferase	8

Figure **3-5.** Intron phylogeny and intron-positions for several HGT orthogroups that encode tRNA synthetase/transferase. (A) and (B) RAxML-based ML trees of intron sequences in the corresponding cds tree of orthogroup 2270 (A) and orthogroup 806 (B). "H" symbol and yellow highlighting represent parasitic HGT genes, while "V" and green highlighting represent the parasitic vertically inherited genes. (C) Annotation of three HGT orthogroups, all encoding tRNA synthetase/tranferases involved in the attachment of the codon and tRNA anticodon. Also, the number of conserved intron positions (CIPs) in the forced cds alignment were also shown, 4067 is the orthogroup in main Figure 3-4 that also encodes tRNA transferase.

3.2.6 Functional HGT

Tissue specific HGT expression 37 out of 49 HGT transgenes show an expression with maximum FPKM in any stage greater than 5, indicating that most are actively transcribed. In addition, 36 out of the 42 HGT orthogroups contain HGT genes from more than one parasitic taxon (Table 3-1) suggesting a conserved role in the parasite. The species with the most HGTs is *P. aegyptiaca,* and all candidates show tissue specific expression (Figure 3-6). Moreover, the expression profile of *P. aegyptiaca* HGT genes (Table 3-1) revealed a distinctive cluster of interface-specific expression (Figure 3-6), and an equal number with abundant expression in haustorial tissues. A subset of these genes encodes functions related to transcription and protein synthesis (Table 3-1), consistent with the role of metabolically active haustoria in loading host nutrients (characterized by large nuclei, organelle-rich cytoplasm and abundance of rough endoplasmic reticulum (Visser, et al. 1984; Pielach, et al. 2014)).

HGTs are evolving under constraint For each of the HGT orthogroup phylogenies, we performed a branch test to estimate the level of constraint in protein sequences. This compares the foreground HGT genes and the background orthogroup members. The same or even stronger levels of purifying selection in parasitic HGT genes observed in 27 orthogroups show that HGT-encoded proteins are under strong constraint (Table 3-1, 3-4), indicating a likely functional role in parasitic plants. Additional evidence comes from conservation of predicted 3-D structure for HGT proteins in comparison to their nonparasitic orthologs in *Arabidopsis thaliana* (Figure 3-7).

In summary, four lines of evidence support a functional role for these horizontally acquired sequences in parasitic Orobanchaceae: (i) HGT sequences are detected and commonly

conserved across species boundaries, (ii) the sequences are actively and differentially transcribed, frequently with a bias toward haustorial expression, (iii) the genes are evolving under purifying selection consistent with the conservation of functional protein structures, and (iv) surviving HGTs are obtained from ancestral host lineages suggesting that HGTs play a role in the parasite-host interaction.

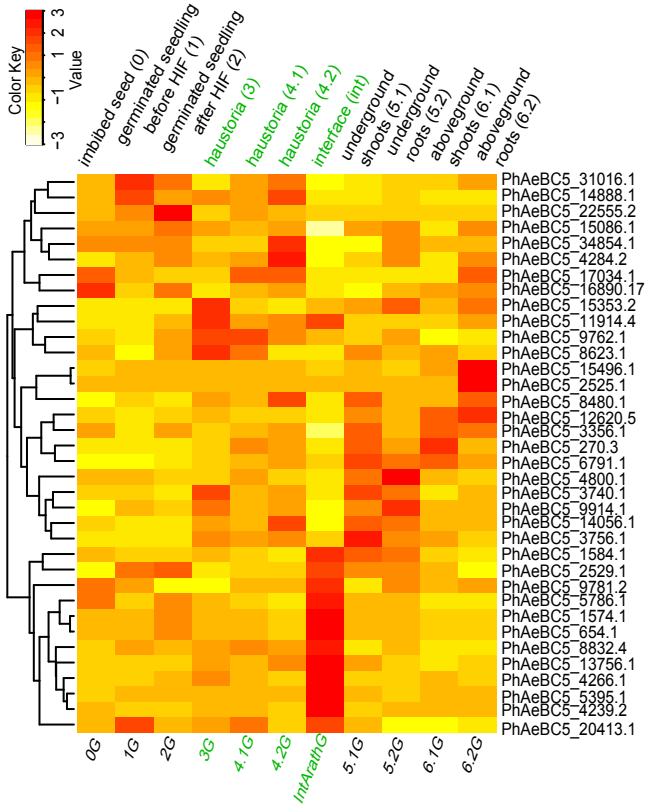


Figure **3-6**. A heat map shows the expression of HGT transgenes in *Phelipanche aegyptiaca*. Expression is shown with FPKM-transformed z-scores to ensure even signal intensity across stages. Rows represent HGT genes, columns represent stages (below) or tissues (above). Haustorial and interface tissues are colored in green. Genes were clustered on the left to show similarity.

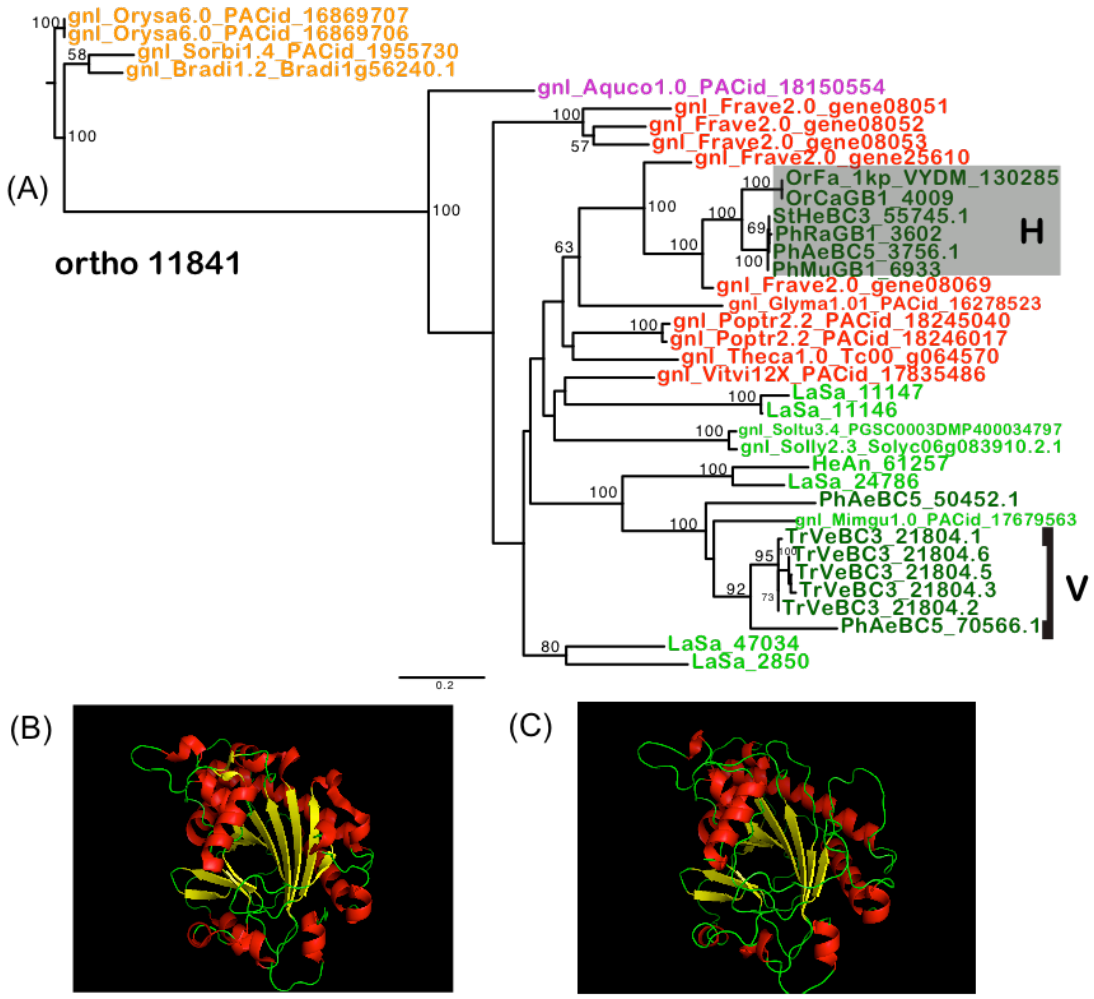


Figure **3-7.** Phylogeny of orthogroup 11841 and predicted 3D structure. The HGT genes identified are PhAeBC5_3756.1 in *Phelipanche aegyptiaca* and StHeBC3_55745.1 in *Striga hermonthica*. For this particular orthogroup, the HGT gene was also identified in *Orobanche fasciculate* (OrFa_1kp_VYDM_130285) from the 1KP database. Additional Orobanchaceae orthologs were also found through our private dataset, including *Orobanche californica* (OrCaGB1_4009), *Phelipanche ramosa* (PhRaGB1_3602) and *Phelipanche mutelii* (PhMuGB1_6933). HGT and vertical clade was labeled with "H" and "V" respectively on the phylogeny. (A). phylogeny of orthogroup 11841 supporting HGT; (B). Protein 3D structure of Anthocyanidin synthase from *Arabidopsis thaliana* (PDB ID: 1GP4); (C). Predicted protein 3D structure of PhAeBC5_3756.1. Color scheme: alpha helixes shown in red; beta sheets shown in yellow; loops shown in green.

Table **3-4.** PAML analyses with branch test on codon alignment of 42 HGT orthogroups testing presence of purifying selection, relaxed constraint, or positive selection.

Ortho ID	dN	dS	Foreground ω	background ω	P-value	Significant at 0.05	Selection
806	0.09	0.33	0.27	0.14	< 0.00001	S	RP
1886	0.48	0.82	0.58	0.16	< 0.00001	S	RP
2270	0.09	0.28	0.31	0.14	< 0.00001	S	RP
2376	0.03	0.15	0.18	0.04	0.000108	S	RP
4067	0.10	0.24	0.43	0.20	0.000187	S	RP
8888	0.12	0.28	0.43	0.23	0.001553	S	RP
9613	0.51	1.03	0.50	0.14	< 0.00001	S	RP
10050	0.13	0.26	0.51	0.20	< 0.00001	S	RP
10143	0.10	0.21	0.46	0.20	< 0.00001	S	RP
1226	0.22	0.41	0.55	0.27	0.002847	S	RP
4598	0.22	0.56	0.39	0.27	0.016327	S	RP
11437	0.40	0.53	0.74	0.36	< 0.00001	S	RP
13512	0.29	0.17	1.65	0.34	< 0.00001	S	POS
3861	0.14	0.10	1.48	0.28	< 0.00001	S	POS
18709	0.19	0.17	1.16	0.22	0.000495	S	POS
19696	0.56	73.08	0.01	0.29	0.022212	S	SP
11841	0.03	0.20	0.14	0.26	0.03422	S	SP
13656	0.01	3.08	0.00	0.30	< 0.00001	S	SP
14233	0.32	1.16	0.27	0.48	0.005556	S	SP
5896	0.06	0.21	0.29	0.30	0.943628	NS	P
1685	0.06	0.21	0.29	0.27	0.811464	NS	P
17	0.21	0.78	0.27	0.23	0.054126	NS	P
218	0.07	0.54	0.12	0.20	0.431345	NS	P
226	0.12	0.51	0.24	0.27	0.449545	NS	P
1021	0.13	0.72	0.18	0.19	0.943886	NS	P
4572	0.13	0.42	0.31	0.47	0.054707	NS	P
5002	0.17	1.00	0.17	0.23	0.172164	NS	P
12835	0.11	0.68	0.16	0.18	0.731762	NS	P
13892	0.02	0.09	0.26	0.24	0.917998	NS	P
14230	0.28	2.22	0.13	0.16	0.331354	NS	P
14624	0.22	1.02	0.22	0.26	0.420936	NS	P
14675	0.49	1.22	0.40	0.47	0.705675	NS	P
15149	0.31	1.24	0.25	0.28	0.734189	NS	P
15246	0.06	0.10	0.65	0.24	0.139996	NS	P
16703	0.40	70.39	0.01	0.33	0.09305	NS	P
18354	0.54	1.01	0.54	0.62	0.908563	NS	P
18774	0.20	0.32	0.61	0.51	0.26167	NS	P
19297	0.08	0.14	0.61	0.51	0.26167	NS	P
20188	0.55	0.66	0.83	0.82	0.971238	NS	P
20190	0.19	0.22	0.88	0.91	0.946516	NS	P
23343	0.08	0.32	0.26	0.34	0.05642	NS	P
23480	0.22	0.26	0.56	0.84	0.218423	NS	P

Significance at 0.05 column: S – significant, NS – non-significant. Selection type column: P – purifying, SP – stronger purifying, RP – relaxed purifying, POS – positive selection

Evidence of adaptive evolution of HGTs Our observation of abundant haustorial expression in a majority of the HGT genes suggests a likely contribution of HGT to parasitic adaptation (evolution of haustoria). To corroborate this idea, we examined the possibility of adaptive signatures on protein sequences of these HGTs. Evidence of relaxed purifying selection and positive selection were observed for more than 15 HGT orthogroups (Table 3-4), 13 of which contain adaptive sites present in HGT parasite genes but absent in the closely related nonparasitic genome (Table 3-5) (*Mimulus* for instance). Of these, six orthogroups have genes encoding functions related to transcription and translation (orthogroup 8888, 18709, 806, 4067, 10050, 13512), and four orthogroups contain genes with abundant haustorial expression (orthogroup 1226, 8888, 18709, 806) (Table 3-1). The signatures of adaptive sites and their retention as haustorial genes in the genome suggest that these changes in HGT proteins are largely under positive selection and may have provided novel functions contributing to increased parasite fitness.

Table **3-5**. PAML analyses with the branch-site model on codon alignment of 15 HGT orthogroups with greater dN/dS on HGT genes compared to the background, identifying the presence of sites under positive selection.

HGT transgene IDs	**Orthogroup ID**	**Sites identified having positive selection (P<=0.05)**
PhAeBC5_4239.1 OrMi2_4015 PhRa5_26371 PhMu2_16115	806*	556 D;
PhAeBC5_11914.4 PhRa5_49596 PhRa5_27280 PhMu2_10869 PhMu2_12766	1226*	135 E; 250 V;

PhAeBC5_15496.1 PhRa5_103413 PhRa5_58721	1886*	1 M; 39 E; 83 F; 84 S; 86 E; 92 K; 146 L; 153 S; 170 T; 173 D; 176 C; 210 T; 214 L; 230 T; 236 T; 247 R; 249 Q; 251 C; 254 D; 256 V; 317 G; 318 K; 322 E; 330 P; 347 S; 363 A; 403 G; 404 F ; 406 S; 416 C; 421 G; 426 F; 428 G; 429 P; 438 C; 444 W; 472 A; 485 A; 488 V; 493 A; 500 G; 530 S; 533 E; 535 V; 581 D; 582 L;
StHeBC3_2868.1 PhAeBC5_3356.1 PhRa5_6157 PhMu2_12619 PhMu2_44156 PhMu2_20937 PhMu2_5516 PhMu2_14031 OrMi2_4206 StGe3_7730 StGe3_20099	2270	NA
PhAeBC5_270.3 StHeBC3_48088.1 PhRa5_4383 PhMu2_6326 PhRa5_4382	2376*	20 V;
PhAeBC5_9914.1OrMi2_2808 PhRa5_12205 PhMu2_12820 PhMu2_11052	3861*	23 S; 45 E; 68 A; 71 D; 93 L; 96 D; 97 D; 254 D; 289 K; 301 V; 486 C; 543 W; 568 P;
PhAeBC5_9762.1 PhRa5_10084 PhMu2_15393	4067*	218 K;
PhAeBC5_13756.1 PhMu2_15647 PhRa5_33568	4598	31 R; 32 P; 156 S; 159 G; 160 L; 169 K; 187 L; 188 S; 233 Q; 235 K; 240 K; 244 T; 247 E; 248 A; 249 M; 251 L; 253 P; 282 L; 284 S; 379 K; 481 N; 495 E; 500 G; 535 A; 536 T; 539 V; 540 A; 543 C; 545 P; 547 N;
Phelipanche_ramosa_PhRa5_6603 Phelipanche_mutelii_PhMu2_37370 Phelipanche_ramosa_PhRa5_6605 PhAeBC5_15086.1 Phelipanche_ramosa_PhRa5_6604	8888*	209 Q; 538 D; 559 P;

PhAeBC5_16890.17 PhRa5_14643 PhRa5_14643	9613*	3 P; 4 S; 22 S; 26 R; 33 F; 42 K; 50 K; 53 T; 55 N; 73 L; 74 P; 77 D; 78 A; 81 I; 82 G; 85 L; 86 Q; 87 I; 90 E; 95 V; 96 E; 99 F; 100 V; 101 H; 102 L; 104 F; 106 C; 107 E; 109 K; 110 P; 112 H; 114 V; 116 S; 121 S; 122 K; 123 P; 126 K; 127 F; 162 T;
PhAeBC5_4284.2 PhMu2_5340 PhRa5_11783 PhMu2_31112	10050*	472 S; 666 S; 709 G; 710 S; 763 F;
PhAeBC5_14056.1 OrMi2_39134 OrMi2_14052	10143*	521 H; 599 G;
PhAeBC5_1584.1 PhAeBC5_3740.1	11437*	1201 Y; 1242 E; 1287 T; 1417 A; 1637 G;
PhAeBC5_1584.1 OrCa3_3600 PhRa5_1522 PhMu2_40952	13512*	367 R; 479 C; 480 R; 497 S; 520 S; 533 F; 536 E; 547 L; 768 G; 831 N; 844 N; 886 E; 1005 P; 1107 Y; 1129 L; 1174 K; 1391 S;
PhAeBC5_654.1 PhRa5_6642	18709*	664 C; 676 Y; 710 K; 715 S; 724 R;

NA indicates no adaptive sites were identified. "*" after each orthogroup indicates presence of adaptive sites were identified only in parasitic lineages, not in a nonparasitic genome, such as *Mimulus*.

3.2.7 Adjacent HGTs in two *Striga* species[5]

While we were seeking genomic evidence for HGT transgenes (StHeBC3_16619.1, orthogroup 14233), we identified another contig (StHeGnB1_80049, orthogroup 14624) as a high confidence HGT transgene. The latter is located 5' upstream of the former on the same genomic contig (Figure 3-8). Both of these two HGT events showed a transfer from *Sorghum* to two parasitic *Striga* species, *S. hermonthica* and *S. gesnerioidie* (Figure 3-8A1 and Figure 3-8A2). As the monophelytic HGT clade doesn't include *S. asiatica*, it suggests that these two events

[5] This section is not present in the submitted HGT manuscript.

occurred in the ancestor of *S. hermonthica* and *S. gesneroidie*. Interestingly, each of these two orthogroups represents a monocot-specific gene family that lacked genes from any eudicot taxon. A further BLASTX analyses against NCBI non-redundant (nr) databases revealed that StHeGnB1_80049 has 90% similarity at the amino acid level and 91% at the nucleotide level with a *Sorghum bicolor* gene (SORBIDRAFT_10g026740/ Sb10g026740, top hit from NCBI Blast results). BTB superfamily domain was identified with the predicted peptide sequence and functional annotations of its homologs in monocots are speckle-type POZ protein, the ortholog of which in *Arabidopsis* encodes a disease resistance protein. In contrast, the best hit of StHeBC3_16619.1 is a hypothetical protein of unknown function in *Setaria italica* (sequence ID: XP_004978140.1) with 43% identity at the amino acid level.

Our sequence comparison analyses revealed a likelihood of genomic integration. For StHeGnB1_80049, both the donor (*Sorghum bicolor* - SORBIDRAFT_10g026740) and the recipient lacked introns; however, a TATA Box was identified upstream of HGT transgene StHeGnB1_80049. Interestingly, high DNA sequence similarity up to 88.54% between the *Striga* gene and the host genomic sequence was also extended to a 300 bp-long intergenic region of these two *Striga* transgenes (Figure 3-8C – light and dark yellow region between the recipient and donor). This showed the transfer likely happened on the genomic level. As low coverage DNA-Seq failed to cover the coding sequences for StHeBC3_16619.1, we were unable to carry out the same analysis to infer the transfer mechanism. In conclusion, this example revealed two adjacent recipient genes from two different donor positions in the host genome.

Another example shows HGTs were derived from two adjacent genes in the donor genome. In the same orthogroup 14233, the *Striga asiatica* ortholog of StHeBC3_16619.1, gnl_Stras_v1.02_SGA1.0.scaffold994G00050, was placed within grass clade, supporting an HGT origin. The best NR hit of the *S. asiatica* gene in orthogroup 14233 is from *Setaria italica*, the same donor as the *S. hermonthica* gene StHeBC3_16619.1. To examine if all the HGT genes in

this orthogroup were the result of one transfer, we constrained the five sequences to form a monophyletic clade. The constrained tree with SH test proved to be significantly worse than the original tree (Figure 3-8A2) with P-value less than 0.01, suggesting two independent transfers, instead of one single transfer that happened in the ancestor of all three *Striga* species. Interestingly, another HGT gene in *S. asiatica* encoding an unknown protein from Satoko et al 2016 had a best NR hit also from *Setaria italica*. BLASTN analyses of the *Striga asiatica* HGT genes into the *Setaria italica* genome revealed their donor sequences being adjacent to each other in chromosome 3 of the *Setaria italica* genome. The analyses showed that the unknown *Striga asiatica* HGT gene of orthogroup 14233 had homology with four genes in the donor genome - two genes in the donor genome are next to each other in chromosome 3, the other two are next to each other in chromosome 2 (Figure 3-8 A2 and C). Amino acid and DNA identity between the donor and recipient sequences are as high as 80.4% and 91.7%, respectively (Figure 3-8 A2 and C). The unknown gene phylogeny of Figure 3-8 A3 revealed four HGT genes in *Striga asiatica*, one in *Striga hermonthica*, and one in *Striga gesneroides*, suggesting the transfer happened in the ancestor of *Striga* genera from *Setaria italica*. These four genes are primarily homologous to *Setaria italica* gene Si021037m.g of chromosome 3, downstream of the donor genes of the *Striga asiatica* gene in orthogroup 14233. Indeed, the detailed BLASTN analyses showed that two HGT genes in Figure 6-8 A3 were chimeric HGT sequences showing homologies with two genes in chromosome 3 and chromosome 2 of the donor *Setaria italica* genome. The chimeric origin of the two *Striga asiatica* genes – both in red and yellow color, indicated a possible gene conversion event involved in the horizontal gene transfer events. The amino acid and DNA identity between the unknown gene in Figure 3-8 A3 and the donor genes are 61.8% and 91%. In both Figure 3-8 A2 and A3 cases, the donor and recipient genes show presence of introns, indicating a likelihood of genomic transfers (orthogroup 14233, *Si029853m.g* and *gnl_Stras_v1.02_SGA1.0.scaffold994G00050* both have intron sequence, the unknown gene in

Figure 3-8 A3 gene in both *Striga asiatica* and *Setaria italica* gene have multiple introns). In conclusion, this piece of analyses revealed another scenario of two adjacent donor genes in the HGT history, whereas the recipient sequences are not adjacent.

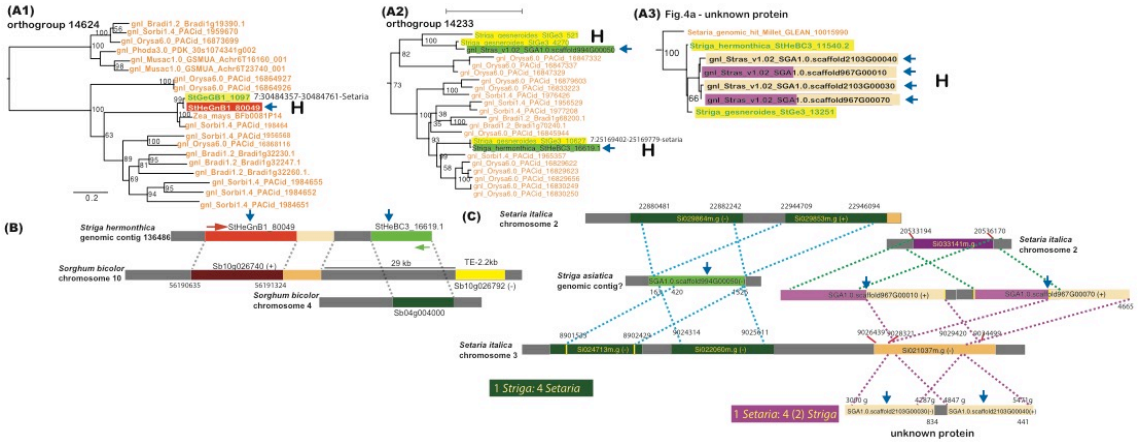


Figure **3-8**. RAxML-based ML trees and comparisons of HGT genes between parasite and its donor supporting two HGTs being adjacent in the recipient genome (A1, A2, B) and donor genome (A3 and C). ML trees of two orthogroups (14624 and 14233) support two HGT events likely to have occurred from grasses to an ancestor of *Striga hermonthica* and *Striga gesneroides*. HGT clades are highlighted in yellow and with "H". (C), Two *Striga hermonthica* genes are located adjacent to each other on the same genomic contig (136486) in *Striga* genome. The comparison with BLASTN shows that these two genes (StHeGn_8049 and StHeBC3_16619.1) have homologies with two genes located on two different chromosomes in *Sorghum bicolor* (Sb10g026740 on chromosome 10 and Sb04g004000 on chromosome 4). StHeGn_80049 not only has homology with its donor sequence (Sb10g026740) in the genic region (lighter and dark red), but also in the right flanking region (light orange). 29 kb downstream of the Sb10g026740 is gene Sb10g026792 of 2.2 kb encoding a MULE-transposase (yellow). The downstream gene StHeBC3_16619.1 is an ortholog of gnl_Stras_v1.02_SGA1.0.scaffold994G00050 in the *S. asiatica* genome (A2), which together with the four genes in (A3) represent HGTs of adjacent genes in the donor genome. The *S. asiatica* gene in A2 is syntenic to four donor genes in *Setaria*

genome. Another gene family containing four *S. asiatica* genes is syntenic to two genes in *Setaria* genome. Two *Setaria* donor genes for the unknown gene in A2 are upstream of one donor gene encoding Alanine tRNA-synthetase of A3. Two of the *S. asiatica* genes in the recipient genome are chimeras, with contributions from two donor genes located in chromosome 2 and 3 of the *Setaria* genome. The HGT genes in the donor-recipient genome comparisons are marked with a blue arrow in phylogenies and syntenic comparisons.

3.2.8 Absence of transfers from parasitic plants to their hosts[6]

As intimate contact between the parasite and host is one mechanism for HGT, we also expect to observe transfers in the opposite direction – from parasite to host. Using a similar screening approach, we screened on phylogenetic trees to look for orthogroups in which the host sequences (rosid and monocot clades) strongly group within the parasitic clades. The initial screening yielded a total number of 35 orthogroup trees (17 in rosids and 18 in monocots), none of which made it through secondary careful validation. In fact, most of them were artifacts from lack of taxon sampling, and frame-shift errors (Figure 3-9), a feature of which is presence of a long-branch on the phylogenetic tree. Orthogroup 3542 shows a Vitis sequence (gnl_Vitvi12X_PACid_17832271) groups with Orobanchaceae clade with a long branch. Manual examination of the peptide alignment revealed frame-shift errors in two sequences (LiPhGnB2_1399.1 and OrAeBC5_1982.1), the frame-shift correction of which resulted in a vertical placement of the Vitis gene (Figure 3-9 (A) and (B)). In orthogroup 8060, a Carica papaya sequence (gnl_Carpa1.181_PACid_16424322) was nested within a parasitic clade with 68 BS value, suggesting a parasite-to-host HGT (Figure 3-9 (C)). However, the C. papaya gene exhibited a long branch, which later proved to be caused by a frame-shift error (Figure 3-9 (D)).

[6] This section is also excluded from the HGT manuscript due to limited space.

In other cases, increase in taxon sampling by adding more taxa in the donor clades all refuted the remaining parasite-to-host HGTs, either belonging to the case host-to-parasite transfers, or became vertical transmitted genes. In a nutshell, we have not identified a single transfer from parasitic Orobanchaceae to hosts.

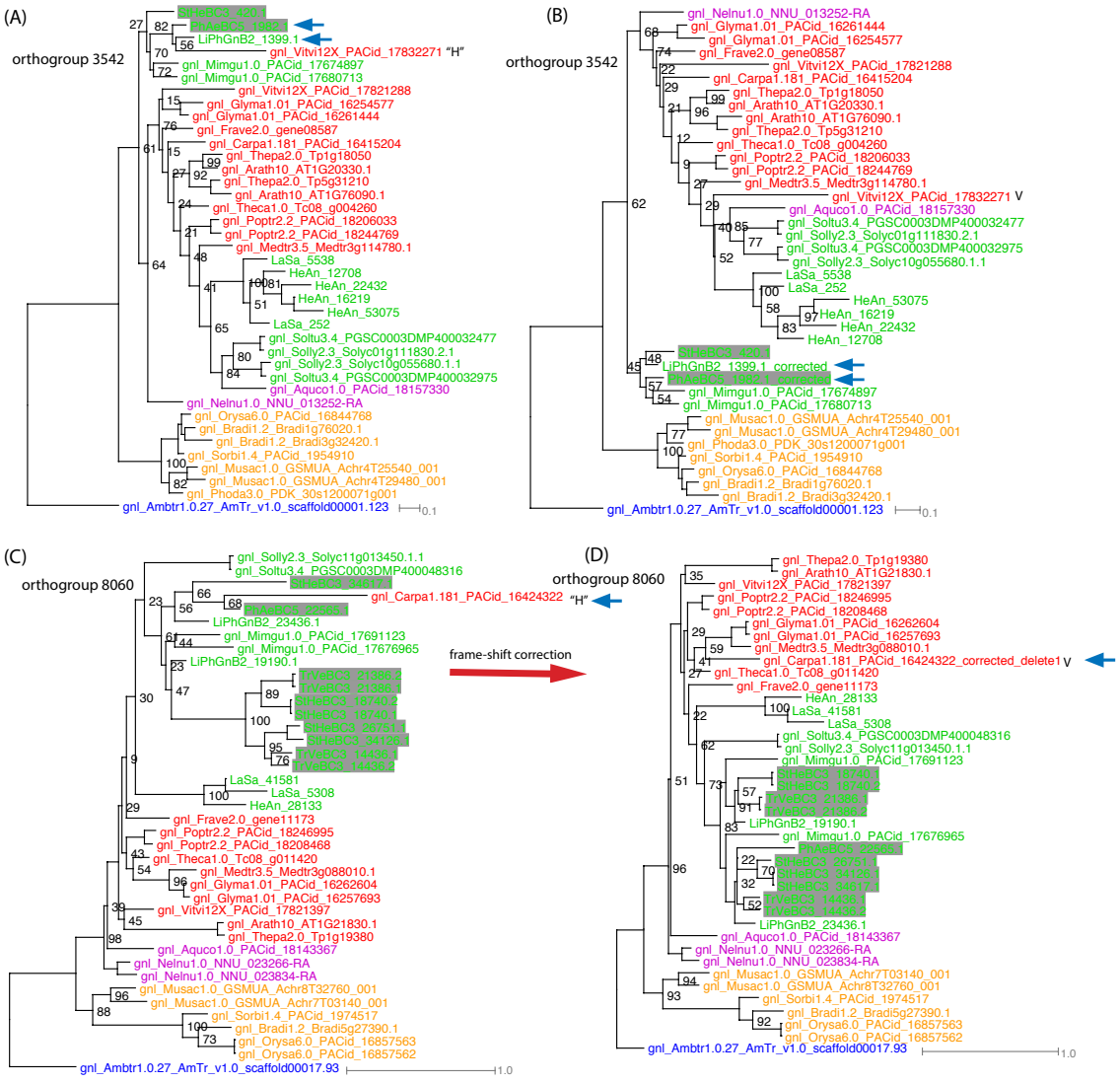


Figure **3-9**. Falsely identified parasite-to-host HGTs due to frame-shift errors. (A) and (C), RAxML-based ML trees of two orthogroups, orthogroup 3542 (A) and orthogroup 8060 (C), indicating possible HGTs from parasite to host as one rosid sequences (in red) of each tree is

nested within the parasitic clade. (B) and (D), RAxML trees of the same gene set in (A) and (C) respectively, except the aberrant sequences were fixed by correcting frame-shift errors at the region from which bad alignment started to occur. "H" and blue arrow pointed to the sequence identified to have undergone HGT from host to parasite. Blue arrows also point to sequences that had frame-shift errors. After frame-shift error correction in either the parasite sequences (A and B) or host sequence (C and D), the host "HGT" sequence became vertical (V).

3.3 Discussion

3.3.1 A stringent and robust phylogenomic approach for HGT identification

Our analyses with explicit phylogenetic schema and stringent evaluation by the use of increased taxon sampling represent a novel and robust approach for HGT identification as it includes two of the published high-confidence HGT cases (Yoshida et al. 2010; Zhang et al. 2013a). In addition, these 42 orthogroups include five orthogroups encoding transposable elements (TEs), which is consistent their invasive nature and their representation of HGT events in the recently sequenced genome of *Striga asiatica*. Three of them containing genes with abundant expression in haustorial tissues (orthogroup 5002, 14230, 15149) suggests a likely active role in parasites. This approach represents the first study of the use of a stringent and robust assessment for HGT identification at the genomic level in the plant kingdom. We believe that this approach has the potential for application in a large number of studies of other species.

3.3.2 Reasons for increased HGT with increase heterotrophic dependence

We observed a clear pattern of a gradually increased number of HGT events from the hemiparasitic facultative *T. versicolor*, to hemiparasitic obligate *S. hermonthica*, to the holoparasitic obligate *P. aegyptiaca*. Several factors could account for the increasing number of HGTs in parasites with increased host dependence. First, the lifestyle of the obligate parasites results in a shorter distance between host tissues at the parasite-host interface and the germline cells, increasing the chance that genomic fragments will be integrated into the genome and passed to offspring. The seedlings of the obligate parasites, which require host plant induced germination stimulation, are in contact with host plants from a very young developmental stage, thus increasing the chances that undifferentiated cells will experience HGT events (Huang 2013).There is also clear evidence for phloem connections between host and *P. aegyptiaca* (Aly et al. 2011), allowing for more HGTs along with the genetic exchange of nucleic acids via phloem (Kim et al. 2014). Finally, haustoria have a high metabolic rate associated with loading host nutrients (Visser et al. 1984; Pielach et al. 2014) and HGTs related to transcription and translation are highly expressed in this novel plant organ (Table 3-2). This increased host dependence and high metabolic flux may create a positive feedback loop where mobile nucleic acids are transferred to the parasite at a greater rate as a consequence of increased acquisition of host resources.

3.3.3 A proposed adaptive role of HGT in parasitic plants

Adaptive roles for HGT have been indicated in many lineages. Notably in bacteria, HGT has significant impact in adaptation to ecological niches due to the frequent cases of HGT from transformation, conjugation, etc. In eukaryotes, only a few examples have been reported to

suggest HGT with an adaptive role. In plants, the efforts were previously focused on the identification of HGT events. With the increasing numbers being discovered, we were able to determine if there is a pattern for the potential role of HGT in parasitic plants.

Our analyses of an interface-specific expression forming a clear cluster in the parasitic *P. aegyptiaca*, suggests that HGT may play an adaptive role for parasitic plants. Previously reported cases of HGT in parasites or pathogens primarily encoded cell wall-degrading enzymes of plants (Keeling 2009) and thus are implicated in host invasion. Surprisingly, no such proteins were identified in our HGT search. Instead, a number of proteins involved in cell wall modification processes were attributed to gene duplications that happened in the ancestor of all parasitic lineages of Orobanchaceae (Yang et al. 2015). Events leading to the origin of parasitism should occur in an ancestor of parasites. Our HGT phylogenies, however, supported predominantly recent occurrences that were unique to individual genera. Therefore, gene duplications may precede and underpin the origin of parasitism, while HGTs contributed to further transition to an increased heterotrophic dependence.

3.3.4 Genomic integration, functional inference, the tip of an iceberg

We found strong evidence supporting genomic integration for vast majority of the transfers in transcriptomes of three parasites. Only one event was indicative of an RNA-mediated event as the donor has two introns, whereas the recipient has no introns. Notably, our evidence from intron phylogenies provides stronger support for the mechanism of genome integration. Compared to mRNA-mediated transfers, genomic integration is more likely to result in genes with promoters that drive their active transcription. We proposed that a predominant genomic integration may account for the transcribed HGT sequences and it is likely that these 42 HGT

families just represented the tip of the iceberg. The sequencing of the genome in the near future should reveal more HGTs, many of which degrade with time and evolve as pseudogenes.

3.3.5 HGT hotspot and possible transfer mechanisms

Comparisons of the HGT CDS with the genomic sequences reveal two adjacent HGT genes in *S. hermonthica* genome. The nonadjacent positions of the donor genes in the *Sorghum bicolor* genome and the much lower sequence similarity between the downstream HGT recipient and donor gene indicate two separate transfers of *Striga hermonthica* genes from the donor *Sorghum bicolor* genome. Another example involves a *Striga asiatica* ortholog of the downstream *Striga hermonthica* gene, the donor sequence of which is adjacent to the donor sequence of a second *S. asiatica* gene (encoding an unknown protein), supporting two transfers from two adjacent genes in the donor *Setaria italica* genome. In *Escherichia coli*, evidence indicates that operons are more likely to be fixed than individual genes following HGT due to coregulation (Koonin and Wolf 2008). Future evidence of more expression data for the StHeGnB1_80049 may show whether they show similar expression indicating co-expression, which might be indicative of the mechanisms facilitating their survival in the genome. In the second case, the HGT donor genes are adjacent to each other in the genome but the different locations in the recipient genome and the different levels of similarity with donor sequences are more likely to infer two separate transfers as well.

The two transfers involving adjacent donor or recipient genes indicate a likelihood of a HGT hotspot. The recurrent independent horizontal transfers in two related *Striga* species – two orthologous HGT genes in *S. asiatica* and *S. hermonthica* in orthogroup 14233, may indicate a potentially functional role for this gene. Despite relatively little evidence supporting function, the annotation of this orthogroup based on sequence similarity supports an unknown protein, and the

expression of StHeBC3_16619.1 shows a shoot specific expression pattern. Interestingly, the *S. hermonthica* alanine tRNA-synthetase gene StHeBC3_11540.2 shows highest expression in interface tissues, a similar pattern to tRNA synthetase genes with abundant haustorial and interface expression in *P. aegyptiaca*. The fact that particular regions show higher frequency of transfer than other regions in the genome may indicate a potential role for HGT as an adaptive strategy for parasitic plants.

TE could act as one mechanism of HGT -- especially in bacteria such mechanism has supported the transfer of adjacent DNA upstream of transposase genes (Toleman, et al. 2006). A previous bacteria-derived hydrolase in an insect pest of coffee revealed the HGT gene being flanked by two transposase genes (Acuna et al. 2012). To examine if a similar mechanism was also involved in plant horizontal gene transfer, we looked for evidence of transposases in the surrounding regions of the HGT genes in the donor genome. Only in the case of StHeGnB1_80049 (Figure 3-8 B), we identified a transposase located 29.4 kb downstream of the HGT donor gene in *Sorhgum bicolor* genome (other cases are much farther apart). With such a long distance between the HGT gene and TE in the donor genome, TE is unlikely to mediate gene transfer; we thus propose TEs may reside in the recipient genomes or a different mechanism may be involved in transferring eukaryotic genes. Future efforts that leverage whole genome sequencing of *S. hermonthica* and *P. aegyptiaca* could elucidate whether the neighboring regions of the adjacent HGT sequences contain sequences such as TEs or recombination hotspots that facilitated the HGT events.

3.3.6 Conclusions

In this study, we developed a phylogenomic-based pipeline that parsed large-scale phylogenetic trees for preliminary HGT identification, followed by careful validation with further

analyses and increased taxon sampling. Our criteria for HGT identification requires the focal gene to be nested within donor clades and supported by two strong nodes, instead of just being a sister of the donor sequence. This proves to be stringent, but also robust to the challenges of genome scale HGT discovery - we identified 52 high confidence HGT events in three parasites in Orobanchaceae. Our analyses of intron sequences and structure support genomic fragment integration of HGTs, with only one transposon family supporting an RNA-mediated retroprocessing event. Although unexpected, considering the well documented mRNA transfer that occurs between the parasite and host, we hypothesize that transfers of genomic fragments will more often result in functional transfers than mRNA because genomic regions can contain intact promoters that may be recognized by the recipient plant species. Cross-species promoter recognition is common in experimental transformation studies (Atkinson and Halfon 2014; Oo, et al. 2014), even among very distantly related plant species (Xu, et al. 2014). Furthermore, because promoters from other eudicots may be more recognizable than promoters of Poaceae (if it involves enhancers), it could also help to explain why fewer functional transfers are observed in grass-feeding *Striga* species. These hypotheses could be tested experimentally by comparing the capacity of Orobanchaceae to recognize and transcribe sequences with foreign promoters (from other eudicots, from monocots) versus the likelihood of substantial transcription of a randomly inserted cDNA.

Functional roles conferred by these HGT genes, for the first time, identified HGT as a mechanism contributing to the adaptive evolution of parasitic plants. Our methods likely have underestimated the number of HGT gene because (i) the phylogenomic-approach in this study relies on an farily complete and accurate construction of gene family phylogenies, (ii) large and complex gene families are not amenable to this approach, and (iii) we restricted our search of donor lineages to distantly related monocot and rosid groups for enhanced signal to noise ratio.

With the increasing availability of genome sequences and other genomic scale data, along with increasingly rigorous standards for discovery, many more examples of true HGT are likely to be revealed.

3.4 Materials and methods

3.4.1 Removal of contamination

Sequences were cleaned by removing non-plant transcripts and host plant transcripts (*Medicago* or *Zea* for *Triphysaria*, *Sorghum* for *Striga*, and *Arabidopsis* for *Phelipanche*; (Westwood et al. 2010)) with BLASTN (E-value of 1e-10).

3.4.2 Phylogenomic reconstruction of parasite gene trees

Open reading frames (ORF) and protein sequences encoded by assembled transcripts were predicted with ESTScan version 2.0 (Iseli et al. 1999). 586,228 protein coding gene of 22 representatives of sequenced land plant genomes were classified into 53,136 orthogroups using OrthoMCL (Li, et al. 2003). The selected taxa includes nine rosids (*Arabidopsis thaliana*, *Thellungiella parvula*, *Carica papaya*, *Theobroma cacao*, *Populus trichocarpa*, *Fragaria vesca*, *Glycine max*, *Medicago truncatula*, *Vitis vinifera*), three asterids (*Solanum lycopersicum*, *Solanum tuberosum*, *Mimulus guttatus*), two basal eudicots (*Nelumbo nucifera*, *Aquilegia coerulea*), five monocots (*Oryza sativa*, *Brachypodium distachyon*, *Sorghum bicolor*, *Musa acuminate*, *Phoenix dactylifera*), one basal angiosperm (*Amborella trichopoda*) (Project 2013),

one lycophyte (*Selaginella moellendorffii*), and one moss (*Physcomitrella patens*). Unigenes from *Lindenbergia*, *Triphysaria*, *Striga*, *Phelipanche*, and two Asteraceae species, *Lactuca sativa* and *Helianthus annuus*, were assigned into the 22-genome orthogroup classifications by BLASTP (Altschul, et al. 1997) with E-value ≤1e-5 and HMM (Eddy 2011c). This resulted in 13,125 orthogroup phylogenetic trees containing at least one parasitic species in the phylogeny. Procedure used to generate orthogroup phylogenies, annotation, and expression quantification followed the same approach as Yang et al. (Yang et al. 2015).

3.4.3 HGT screening on phylogenetic trees

Customized Python scripts were developed to screen incongruent phylogenies. The python script utilized the tree-parsing functions available in the ete2 libraries (Huerta-Cepas, et al. 2010) to traverse one node at a time and extracted members of each node. To decrease the false positive rate for HGT discovery, the script searched for donors in distantly related rosid and monocot groups. Ancestral node was determined when traversing to a node whose left and right branches were exclusively composed of parasite and donor sequences. The script then examined all the inner nodes within the ancestral node for bootstrap support (BS) values that support the grouping of parasite and donor sequences. Three models of topology (Figure 3-1) represent HGTs with decreasing degrees of confidence. The script reported orthogroups that match any of them. After the automated screening, the HGT candidate orthogroups were further classified into three categories: low-confidence, medium-confidence, and high-confidence trees. The classifying criterion was based on a scoring scheme considering whether the donor clade contained at least two donor sequences, bootstrap values supporting the grouping of the parasite gene and donor sequences, and presence of long-branch clades. Each of these three factors was assigned a score, and the summed score indicates the confidence level of the HGT events and dictated the grouping

of each category (Table 3-1). The medium and high-confidence orthogroup trees were then examined carefully for possible sources of errors including contamination, long-branch attractions, and insufficient taxon sampling. Frame-shift errors were fixed by manually introducing 1-2 bp into sequences that caused long branches.

3.4.4 HGT validation by increased taxon sampling

For HGT validation, we added more taxa from related species, including five sequenced asterid genomes and 10 transcriptomes from 1kp in the Lamiales order (Matasci et al. 2014). The genomes include: *Beta vulgaris* (beet), *Actinidia chinensis* (kiwifruit), *Utricularia gibba*, *Sesamum indicum*, and *Striga asiatica* (parasite in Orobanchaceae). The transcriptomes include: *Strobilanthes dyeriana* (Acanthaceae), *Mansoa alliacea* (Bignoniaceae), *Sinningia tuberosa* (Gesneriaceae), *Salvia spp.* (Lamiaceae), *Olea europaea* (Oleaceae), *Epifagus virginiana* (Orobanchaceae), *Paulownia fargesii* (Paulowniaceae), *Antirrhinum majus* (Plantaginaceae), *Rehmannia glutinosa* (Rehmannia), and *Verbena hastata* (Verbenaceae). Also, we added genes from eight additional parasite transcriptomes in the family of Orobanchaceae: *Alectra vogelii*, *Orobanche californica*, *Orobanche minor*, *Phelipanche mutelii*, *Phelipanche ramosa*, *Striga gesneroides*, *Triphysaria eriantha*, and *Triphysaria pusilla*. To make sure that all the HGTs were captured from these added taxa, we used HMM approach (Eddy 2011b) (hmmsearch with 1e-5). For lineage-specific HGT orthogroups, a superorthogroup tree was constructed to ensure the inclusion of all homologous sequences.

3.4.5 Intron analyses

Intron positions were extracted to examine if they were conserved in multiple sequence alignment (MSA). For each orthogroup, the peptide sequences were aligned using MAFFT version 7 (Katoh and Standley 2013b), which were then forced onto coding sequences (CDS) to generate the CDS alignment. A customized Perl script was used to extract the intron positions in each coding sequence, and the corresponding positions were mapped onto the CDS alignment. For transcripts in the Orobanchaceae species, we used the genomic sequences from *Triphysaria versicolor*, *Striga hermonthica*, and *Phelipanche aegyptiaca*. Coding sequence was predicted for each transcript using ESTScan (Iseli et al. 1999), and was then aligned to genomic sequences using BLASTN with E-value cutoff of 1e-05. Manual curation was performed for each CDS-genomic DNA alignment to make sure introns start with GT, and end with AG. To extract intron sequences for each gene from sequenced plant genomes, we used the gff file for the intron regions of each gene. Intron sequences of genes in sequenced genomes were extracted from genomic sequence in Phytozome 10 (Goodstein et al. 2012) using samtools (Li, et al. 2009) and betools (index command "samtools faidx", fastaFromBed in bedtools (Quinlan and Hall 2010) was used by indicating the reference genome fasta using "-fi" and the gff file using "-bed", which generated an output file using "-fo"). Intron sequences for parasite genes were obtained by blasting the coding sequence onto genomic sequences. For intron phylogenies, introns were concatenated to increase the number of informative sites for tree reconstruction with the same approach as building the tree of CDS.

3.4.6 Genome assembly of three parasites

Genomic DNA sequencing data (Illumina data for *T. versicolor*, *S. hermonthica* and *P. aegyptiaca*) was *de novo* assembled using CLC Assembly Cell v 4.1 (http://www.clcbio.com/products/clc-assembly-cell/):

"novo_assemble -o contigs.fasta -p fb ss 180 250 -q -i reads1.fq reads2.fq".

3.4.7 Estimation of number of transfers

We utilized a Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999) to estimate the number of transfer events from 42 HGT orthogroup trees. Trees in which HGT genes didn't form a monophyletic clade were constrained to represent one event, and a RAxML tree with constrained HGT clade (indicated by parameter "-g" in RAxML) was produced. An SH test was performed in RAxML version 7.2.7 (Stamatakis 2006) to test if the constrained tree was significantly worse than the original tree. The SH test was executed with parameters "- t" followed by original tree and "-z" followed by the constrained tree. The ouput of the test returned the likelihood values and significance level at 0.05, 0.02, and 0.01. Significance result from the test supports more than one transfer, whereas insignificance indicates one transfer event.

Chapter 4[7]

Conclusions and future directions

[7] Make sure you read the appendix chapter 2 before reading this chapter.

4.1 Conclusions

Using the comparative transcriptome analyses of three related parasitic plants within Orobanchaceae, we have identified a list of candidate parasitism genes that are important to haustorial initiation and development. Haustorium development involves a group of cell wall modification enzymes such as pectate lyase likely to establish vascular connections with the host, and a group of proteolytic enzymes likely to be involved in nutrient mobilization. The specific functions of these genes are under characterization, and may reveal how these genes contribute to the invasion and attachment processes. Gene duplication and neofunctionalization constitute two major evolutionary processes resulting in genes that are key to a parasitic transition. Haustorium development also involves the co-option of genes known to function in root and flower development. Horizontal gene transfer (HGT) represents another "signature" process of parasitic plants. More than 52 horizontal transfer events were detected, often from known host plant lineages, and with an increasing number of HGT events in species with the greatest parasitic dependence. Analyses of intron sequences in putative donor and recipient lineages provide evidence for integration of genomic fragments, which may carry along regulatory elements that increase the likelihood of functional transfers. HGT acquired genes are preferentially expressed in the haustorium - the novel organ of parasitic plants - indicating products of horizontally acquired genes are also contributing to the unique adaptive feeding structure of parasitic plants.

4.2 Future directions

4.2.1 Studies based on experimental characterization

One of the primary goals of research on parasitic plants is to ultimately reduce their harm to crop plants. Using a transcriptome approach, our study identified a list of candidate genes, including cell-wall degradation enzymes, defense-related genes, proteases, etc (Yang et al. 2015). Establishing a feasible system to characterize their role in parasitic plants is

challenging and could benefit from opinions of experts from each field. There are a few current research directions that have been initiated but involve challenges and thus may need collaborative work. (i) The upregulated genes in haustorial tissues include cell wall modifying enzymes, such as expansins, pectate lyases, pectin methylesterase inhibitors (Yang et al. 2015). Studies revealed presence of cell-wall degrading enzymes including cellulose, polygalacturonase, xylanase and protease in haustoria of *P. aegyptiaca* (Singh and Singh 1993a). Immunocytochemical studies with specific antibodies show direct involvement of pectin methylesterase at the penetration site (Losner-Goshen et al. 1998), implicating a role in establishing a vascular connection with the host conducting tissues. The haustorial interface has one side that faces inside towards the parasite invading peg, and the other side that faces the host cells. How the parasites can degrade the cell walls of the host side without degrading the cell walls of its own side is worth further investigation. The use of *in-situ* studies and some microscopy technologies should allow a better understanding of this question. Current RNAi knock-out studies are underway; however, the study of possible phenotype in this process remains challenging. (ii) Transcriptome studies of three parasites revealed NBS-LRR disease resistance genes upregulated in haustorial stages (Yang et al. 2015). Considering the similarity of plant defense mechanisms between parasitic plants and their hosts, how they regulate host resistance is unclear. It is unknown if this gene acts to mimic host defense to evade the host immune attack or it acts as a counter-defense response. Future experiments may overexpress host R genes in parasites and see if that would result in enhanced parasitism on host. The knowledge of plant-fungal pathogenic interactions may also be applied in understanding the disease resistance mechanisms between parasitic plants and their hosts. In particular, the isolation of susceptible and resistant host lines in the field may be a good entry point. (iii) Current technology used in gene characterization used in the lab is primarily RNAi on hairy roots (composite roots) (Bandaranayake et al. 2010). CRISPR-Cas9 has advantages over RNAi in genome editing (Sander and Joung 2014) instead of post-transcriptional regulation. Both however, rely on an established transformation system. The development of transformation systems may need continuous efforts. Alternatively, host-induced gene silencing (HIGS) represents a unique strategy that can be applied on parasite-host interactions considering the movement of small RNA (Alakonya et al. 2012), mRNA (Kim et al. 2014), and big molecules up to 70kDa (Aly et al. 2011) from host to parasite. Currently, HIGS with RNAi has achieved success in *Triphysaria versicolor* (Tomilov et al. 2008), *Phelipanche aegyptiaca* (Aly et al. 2009), *Cuscuta pentagona* (Alakonya et al. 2012), however, the use of HIGS with CRISPR-Cas9

has not been attempted. CRISPR-Cas9 also has its advantage in targeting multiple genes (Xie, et al. 2015), so it may have potential to be applied in parasitic plants for gene characterization. (iv) Current research progress on parasitic plants has allowed a better understanding of parasitism, which mainly has contributions from three processes: 1) host-induced seed germination, 2) haustorium initiation, and 3) haustorial development. All three processes involve the interaction between parasites and hosts. Investigation of the haustorium initiation process has achieved much success mainly because hairy root mutants with defects in haustorial hairs are easy to assay. In terms of seed germination, facultative parasites (*Triphysaria*) can germinate without stimulation from a host (Westwood et al. 2010), whereas obligate parasite *Striga* and *Phelipanche* have to rely on host strigolactone (SL) signal for seed germination (Hauck, et al. 1992; Bouwmeester, et al. 2003; Yoneyama, et al. 2010). This has been shown to attribute to external SL receptors, represented by three publications in *Science* of last year (2015) (Conn, et al. 2015; Toh, et al. 2015; Tsuchiya, et al. 2015). The research revealed the diversification of KAI2 (a karrikin receptor) in parasitic *Striga*, followed by divergent evolution, finally resulting in neofunctionalization as a SL receptor. Despite this progress, many more questions have not been resolved in this process. For instance, plant seed germination is regulated by several hormones such as gibberellin acid (GA) (Peng and Harberd 2002), but how SL signaling interacts with GA signaling in parasitic plants has been unclear. Also, it is unknown whether this process is related to host preference. *S. hermonthica* and *S. asiatica* are known as grass specialists (Musselman 1980); whether specific host recognition is regulated at the stage of interaction between host-derived SL and *Striga* SL receptors requires additional work. The future work on parasitic plants should be able to differentiate which process of the three (germination, haustorium initiation, haustorial development) is involved in parasite-host resistance.

4.2.2 Reveal genetic mechanisms underlying physiological differences of three parasites with comparative transcriptome analyses (PPGP2)

A second future research direction is inspired by the differences in physiologies among the three parasitic plants. It is worthwhile to investigate how comparative transcriptome studies allow us to identify the underlying mechanisms that control the differences. This should provide insights on how parasites progressively evolve. These insights can benefit parasitic weed

control. Some questions that may gain insights from comparative transcriptome analyses include: (i) It is known that hemiparasitic *Striga* keep their stomata open when attached to host (Jiang et al. 2003), even the host is under severe water stress (Smith and Stewart 1990). This ensures that it reduces water potential in the parasite so as to drive the xylem flow from host to parasite. It is possible that the ABA receptor is less sensitive - could one clone the gene and examine if this is the causal gene? Is it possible to reduce nutrient transfer by increasing water potential of the parasitic *Striga*? In a holoparasite such as *Phelipanche* (which doesn't have much leaf area to keep open stomata for efficient transpiration), however, xylem flow is thought to rely on the accumulation of osmotic compounds such as sugar alcohols (mannitol for instance) that decreases the water potential of parasites (Ehleringer and Marshall 1995). Previously mannose-6-phosphate receptor (M6PR) (Aly et al. 2009) has been shown to play a role in accumulation of mannitol in *Phelipanche aegyptiaca*. It is possible that hemiparasitic *Striga* and *Triphysaria* have different *M6PR* transcription profiles, or have fewer copies compared to *Phelipanche*. The PPGP2 transcriptomes with replicated libraries in each stage of the parasites may help us answer this question. This poses another aspect in weed control where the nutrient transfer mechanism could also be targeted to shut down nutrient transfer from host to parasite. (ii) Obligate holoparasitic *Phelipanche* (Aly et al. 2011) and *Cuscuta* (Kim et al. 2014)(Convolvulaceae of *Solanum*) can form clear phloem connections with their host, whereas hemiparasitic plants often don't. The focus on phloem formation in *Phelipanche* and *Cuscuta* may reveal genetic differences between holoparasites and hemiparasites in phloem formation. For instance, which transcription factors regulate phloem formation? APL is a phloem identity gene, and mutation causes the formation of xylem where phloem is supposed to form (Bonke, et al. 2003). It is possible that ectopic expression of APL in haustorial stages cause phloem connections with host in *Phelipanche*, whereas no APL expression is found in haustoria of *Triphysaria* and *Striga*. SUC2 is a phloem specific plasma membrane sucrose transporter (Gottwald, et al. 2000); another sucrose transporter, SUT1 (Slewinski, et al. 2010), is localized to the plasma membrane of sieve elements. Both of them are involved in efficient phloem loading. Considering much more host-derived carbon uptake of *Phelipanche* by phloem connections that are absent in *Striga*, and *Triphysaria*, haustoria upregulated *SUC2* and *SUT1* may be absent in *Triphysaria* and *Striga* yet retained in *Phelipanche*. (iii) Both obligate parasitic *Striga* and *Phelipanche* have to rely on a host for germination. It has been shown that this is mediated by a number of SL receptors (more than 10 KAI2d copies) in *Striga hermonthica* (Conn et al. 2015). Two groups both characterize at least 10 SL receptors (KAI2d1, KAI2d2…)

in *Striga hermonthica*, and found consistently that these show a wide range of SL sensitivity, implicated with a role in sensing host SLs of different concentrations (Toh et al. 2015; Tsuchiya et al. 2015). *Phelipanche aegyptiaca*, also requires a host for seed germination, however, my analyses show it has only three KAI2d gene family members (my analyses): what could be the underlying reason? (iv) Parasitic plants need SLs for shoot branching (together with auxin), and the canonical SL receptor in non-parasitic *Arabidopsis* is D14, a paralog of KAI2 (karrikin receptor) (Nelson, et al. 2009). D14 is likely to be an internal SL receptor in parasitic plants, as phylogenetic history showed that it remains non-duplicated and has quite conserved sequence evolution in parasites (Conn et al. 2015). The duplication of KAI2 gave rise to three classes of KAI2, conserved KAI2 named as KAI2c, the intermediate KAI2 named as KAI2i, and the divergent KAI2d, which are essentially the external SL receptors responding to host SLs for seed germination (Conn et al. 2015; Toh et al. 2015; Tsuchiya et al. 2015). How could one characterize the role of KAI2c, D14, and KAI2d in parasitic plants in terms of the differentiation of internal and external SL signaling? Expression analyses reveal upregulated D14 expression in haustoria and interface tissues of all three parasitic plants; does this indicate additional role of SL in haustorium development? (v) Developmental stages of *Triphysaria* are quite similar to an autotrophic plant: the development of haustoria comes after the development of root (Westwood et al. 2010). In *Striga* and *Phelipanche*, the haustoria occurred much early in development, and root (5.1) develops after haustoria; the occurrence of underground shoots (5.2) are also quite unique (Westwood et al. 2010). How the development of haustoria affect the development of underground roots and shoots is unknown and should attract researchers' attention.

4.2.3 Evolution of parasitic plants – phylogenetic inferences of species relationships and HGT

A third aspect of future directions is regarding the evolution of parasitic plants, including (i) The relationship of *Triphysaria*, *Striga*, and *Phelipanche* is still unclear. The phylogenetic relationship is important to infer which are the ancestral traits and which are the derived traits. Transcriptome sequencing of many species in each genus could help resolve the relationship of Orobanchaceae species. (ii) The genome sequence of *Striga asiatica* revealed two whole genome duplication events in addition to the eudicot-wide triplication (gamma) (Jiao et al.

2012): one in the common ancestor of *Mimulus* and Orobanchaceae, the other more recent[8]. It is unknown if the recent duplication occurred in the ancestor of all Orobanchaceae, or is unique to *Striga asiatica*. The genome sequence of future parasitic taxa could reveal this. In particular, the genome sequence of *Phelipanche* should be important to reveal losses of many genes as a fully evolved holoparasite, but also may reveal additional genes that it has acquired to make it advanced in heterotrophic feeding. (iii) Secondly, dozens of horizontal gene transfer events have been revealed in three parasitic plants of Orobanchaceae[9]. Identification of HGT in other parasitic lineages are also needed to indicate if HGT of same genes is repeatedly happening across independent parasitic lineages, which may shed light on additional roles of HGT in parasite evolution. (iv) Functional characterization of HGT genes is also needed to infer their specific roles in parasitic lifestyle. (v) The inference of HGT is currently focused on the transcribed genes, and all of the resolvable HGTs point to a transfer mechanism mediated by genomic DNA (manuscript in preparation). This suggests that more non-transcribed HGTs such as transposable elements (TEs) await discovery with a better genome annotation. (vi) Factors that facilitate HGT, and that drive HGT gene expression need further investigation. HGT mechanisms could involve TEs (Acuna et al. 2012), or recombination machinery (Lawrence and Retchless 2009). The fate of HGT genes can also be revealed with an understanding of all HGTs that occurred in parasite history.

[8] Satoko Yoshida, Seuungill Kim, Eric K Wafula et al. 2015. Genome sequence of Striga asiatica provides insight into the evolution of plant parasitism. Nature plant (submitted & revised version under review)

[9] Zhenzhen Yang*, Yeting Zhang*, Eric Wafula, Loren A. Honaas, Paula E. Ralph, Sam Jones, Huiting Zhang, Naomi S. Altman, Michael P. Timko, John I. Yoder, James H. Westwood, Claude W. dePamphilis (2016) You are what you eat: Horizontal gene transfer is more frequent with increased heterotrophy and may contribute to parasite adaptation. Proc. Natl. Acad. Sci. U.S.A. (in preparation)

Appendix A

HGT in *Striga asiatica*[10]

[10] I contributed this material for a publication by Yoshida et al. 2016 (Satoko Yoshida, Seuungill Kim, Eric K Wafula et al. 2015. Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. Nature plant (submitted & revised version under review)).

A.1 Introduction

A recently sequenced parasitic plant, *Striga asiatica*, provides us with opportunities to mine for HGT. In contrast to *S. hermonthica* with estimated genome size of 1.6 G, *S. asiatica* has a fairly small genome of only 400 Mb. *Striga asiatica*, similar to *Striga hermonthica*, is a grass specialist attached to specifically grasses including maize, rice, and sorghum.

Two approaches were used to identify HGTs in *S. asiatica*. First is the phylogenomic approach developed for HGT identification in Orobanchaceae, the second is a blast-based analyses performed by Yoshida et al 2016[11]. The phylogenomic approach looks for conflicts between a well-resolved species tree and a gene tree, and the detailed procedure was the same as described in Chapter 3 of this dissertation. A species tree including *S. asiatica* needs to be constructed using a number of single copy genes from a selected group of genomes.

A.2 Phylogenomic-based approach

A.2.1 Constructing a species tree using 26 sequenced plant genomes[12]

A phylogenetic tree using 613 single copy genes from 26 selected plant genomes places *S. asiatica* in the expected group, a sister of *Mimulus guttatus*, the most closely related sequenced genome in the family of Phymaceae in the Lamiales order (Figure A-1). The 26 genome include one moss - *Physcomitreella patens*, one lycophyte - *Selaginella moellendorffii*, one gymnosperm - *Pinus teeda*, one basal angiosperm – *Amborella trichopoda*, five grasses – *Spirodella polyrhiza*, *Orysa sativa*, *Sorghum bicolor*, *Elaeis guineensis*, and *Musa acuminate*, two basal eudicots – *Aquilegia coerulea* and *Nulumbo nucifera*, 10 rosid genomes – *Vitis vinifera*, *Eucalyptus grandis*, *Medicago truncatula*, *Phaseolus vulgaris*, *Prunus persica*, *Carica papaya*, *Arabidopsis thaliana*, *Theobroma cacao*, *Populus trichocarpa*, and *Manihot esculenta*,

[11] Satoko Yoshida, Seuungill Kim, Eric K Wafula et al. 2015. Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Nature plant* (submitted).

[12] This section is taken from the submitted *Striga asiatica* genome paper, analysis of which was performed by Eric Wafula.

five asterid genomes – *Beta vulgaris*, *Utricularia gibba*, *Solanum lycopersicum*, *Mimulus guttatus*, and *Striga asiatica* (Figure A-1).

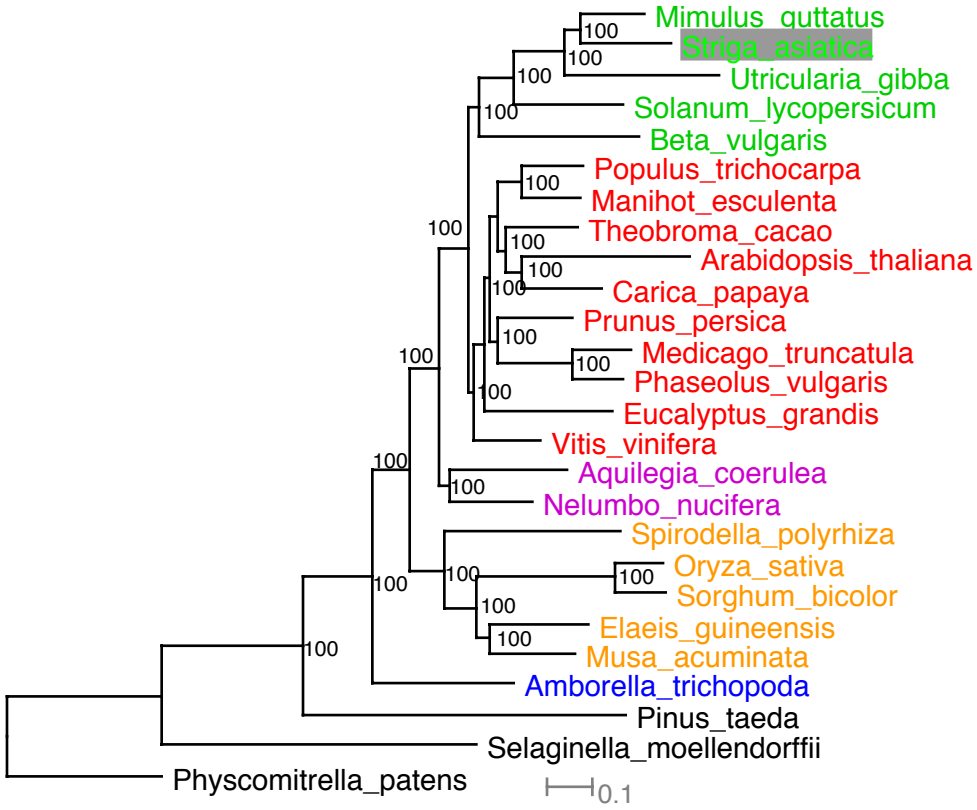


Figure **A-1**. RAxML-based maximum likelihood species tree for 26 selected genomic taxa. Parasitic plant Striga asiatica is highlighted with a grey background. Taxa in other lineages were color-coded: black – *Physcomitrella* (moss), *Selaginella* (lycophyte), *Pinus* (gymnosperm), green – basal angiosperms, yellow – monocots, purple – basal eudicots, red – rosids, green – asterids.

A.2.2 Phylogenomic screening

Phylogenomic-based screening of HGT in *S. asiatica* followed the same schema as described in Chapter 3 of this dissertation. At first, 17052 DNA sequence-based phylogenetic trees for all the orthogroups (17052 orthogroups) containing protein-coding genes in the annotated *S. asiatica* genome were constructed with the forced CDS alignment using RAxML

version 7.2.7 (Stamatakis 2006)[13]. *S. asiatica* genes that fit three schema in Figure 3-1 monocot or rosid clades were screened as preliminary HGT candidates, which were manually evaluated by increased taxon sampling to identify true HGT genes. The script returned six candidate rosid-derived HGTs (Figure A-2) and two monocot-derived HGTs (Figure A-3). Five of the six rosid-derived HGTs were found to be artifacts from insufficient taxon sampling as NR-blast analyses resulted in the best BLAST from closely related taxa (*Nicotiana* and *Mimulus* in orthogroup 103, *Sesamum* in orthogroup 1365 and 2030, *Mimulus* in 2309, *Mimulus* and *Sesamum* in orthogroup 13763), indicating the phylogenetic pipeline exhibited certain weaknesses from genes missing from the genome scaffold that were present in the NR database. Monocot-derived orthogroup 13948 is also a false positive as top blast hits are all from *Mimulus*. Orthogroup 205 is likely to be a real HGT as top hits are from rosid family, which is consistent with it being placed within rosid clades. However, this tree needs extra validation as the *S. asiatica* sequence encodes only a 74-AA protein, while homologs in other species are all over 300 AA. A similar case was observed in the monocot-derived HGT orthogroup 9369, where blast validation appears to agree with tree inference but the encoded protein product is only 58 AA long. Further examination is needed to confirm if this is due to contamination or a chimeric sequence assembly.

[13] The phylogentic trees were built by Eric Wafula.

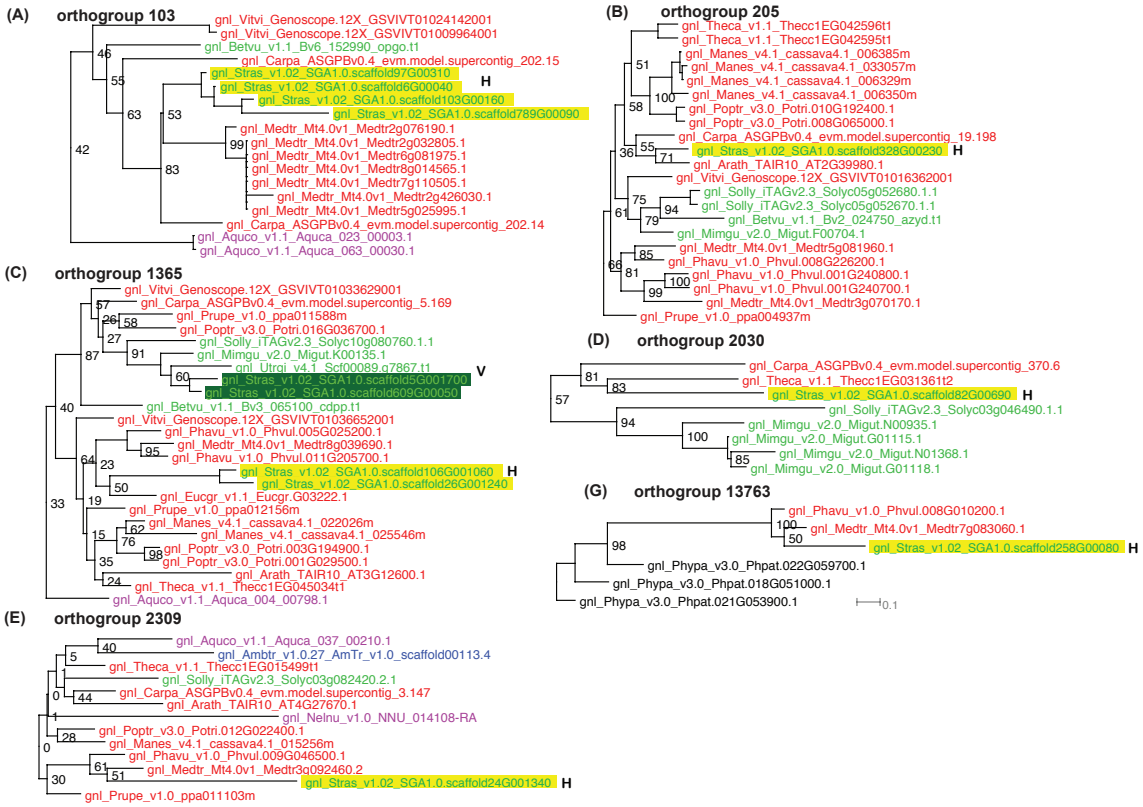


Figure **A-2**. Six rosid-derived preliminary HGT trees from phylogenomic screening. HGT clades are labeled with "H" and yellow highlighting, vertical clades are labeled with "V" and green highlighting.

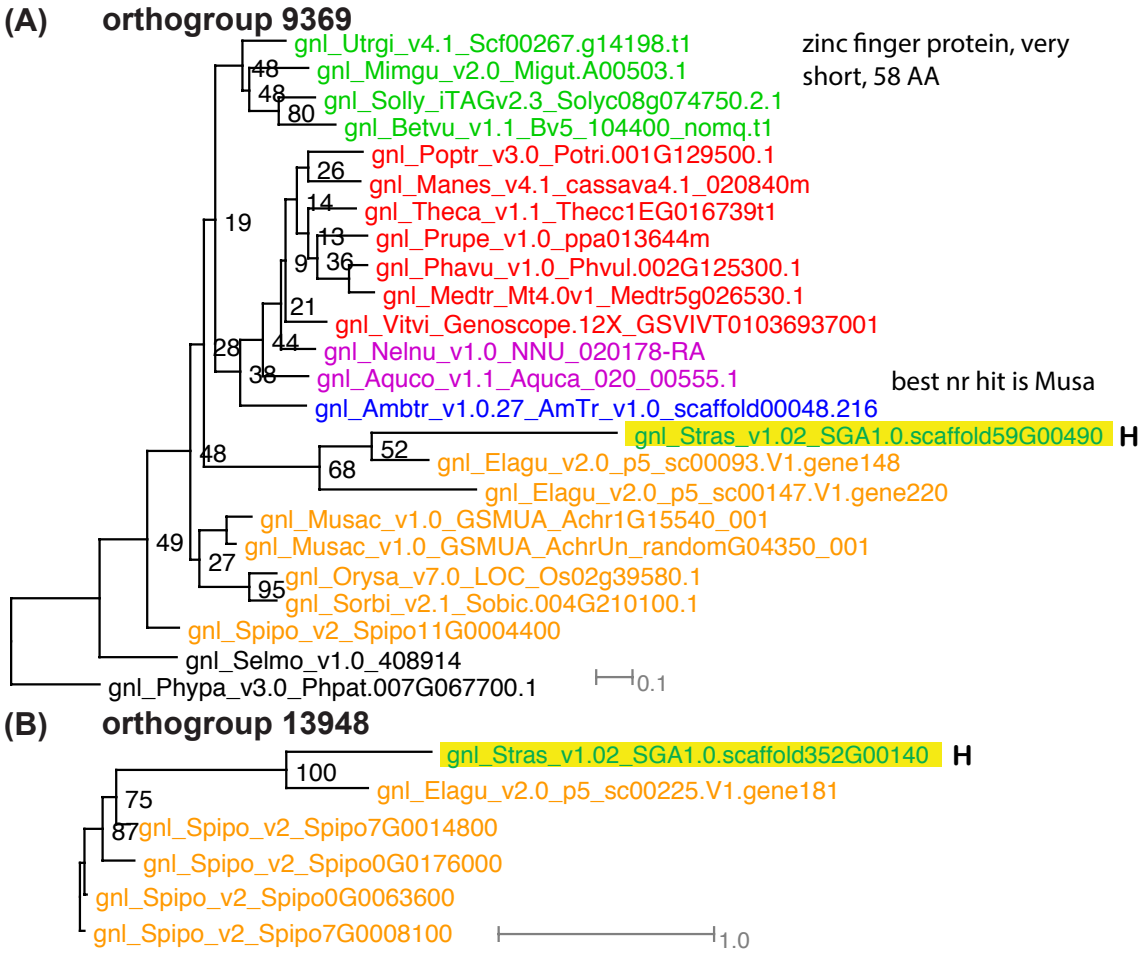


Figure **A-3**. Two monocot-derived preliminary HGT trees from the phylogenomic screening.

A.3 Validation of BLAST-predicted HGT

As a good HGT tree requires correct sampling, careful tree construction, and proper rooting, a preliminary HGT tree needs to have several rounds of follow-up analysis to determine if there is convincing evidence in support for HGT. Here we present an example of an HGT with a preliminary tree and a good HGT tree to reflect such processes. The preliminary tree was an original tree without proper rooting and only one strong node (bootstrap support value of 100) supports the grouping of two *S. asiatica* genes as sisters of a *Setaria* gene (Figure A-4). Because our criteria for strong evidences of HGT require the gene from focal taxa to be nested within its donor clade, this tree is viewed with some skepticism. To improve the tree, we first extracted the sequences and corresponding orthogroup assignment in the 26-genome orthogroup

classification. Our orthogroup doesn't reveal this as an HGT event because it doesn't contain this putative *Setaria* donor sequence. By adding more homologous sequences from NR into orthogroup, we improved the tree significantly. Interestingly, the clade with the *Striga* HGT gene and its donor sequence shows a long branch, indicating its significant divergence from genes in other clades. Moreover, we also found this transfer was shared by *S. hermonthica* and *S. gesneroides*, indicating a transfer likely occurred in an ancestor of the two *Striga* species (Figure A-5).

Conclusion The above different set of genes by BLAST and phylogenomic approach suggest that BLAST and a phylogenomic approach can be complementary to each other. In their BLAST-based approach, the donor search was restricted to grasses, for which they had a rich sampling including many grass species. In addition, our phylogenomic-approach was focused on protein-coding genes only, whereas their BLAST-based approach involve the whole genome, with potential pseudogenes. Many HGTs from host to parasite could finally degrade as pseudogenes, especially if they failed to be properly transcribed after transferring into the recipient genome such as particular types of transposons. On the other hand, due to the computational cost of building large-scale phylogenetic trees, a database composed of a predefined number species has to be constructed, which could miss sequences from additional donor taxa. All of these suggest the phylogenomic-based approach could identity a different set from a BLAST-based approach. In the future HGT identification, a combination of both approaches is desirable.

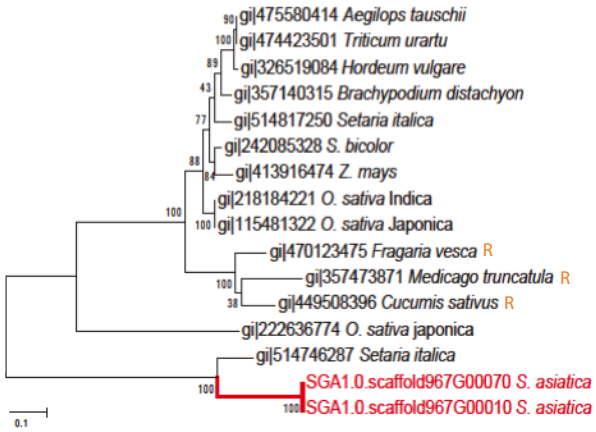


Figure **A-4.** Phylogenetic tree of a horizontally-transferred alanine-tRNA synthetase. The tree was drawn by the maximum likelihood method based on JTT-based model. Bootstrap values (%) were determined using 100 replicates and are shown for branches with more than 50% bootstrap

support. Red branches and red sequences represent *S. asiatica* HGT sequences, orange "R" indicates the rosid genes from strawberry, *Medicago*, and cucumber, respectively (up to down).

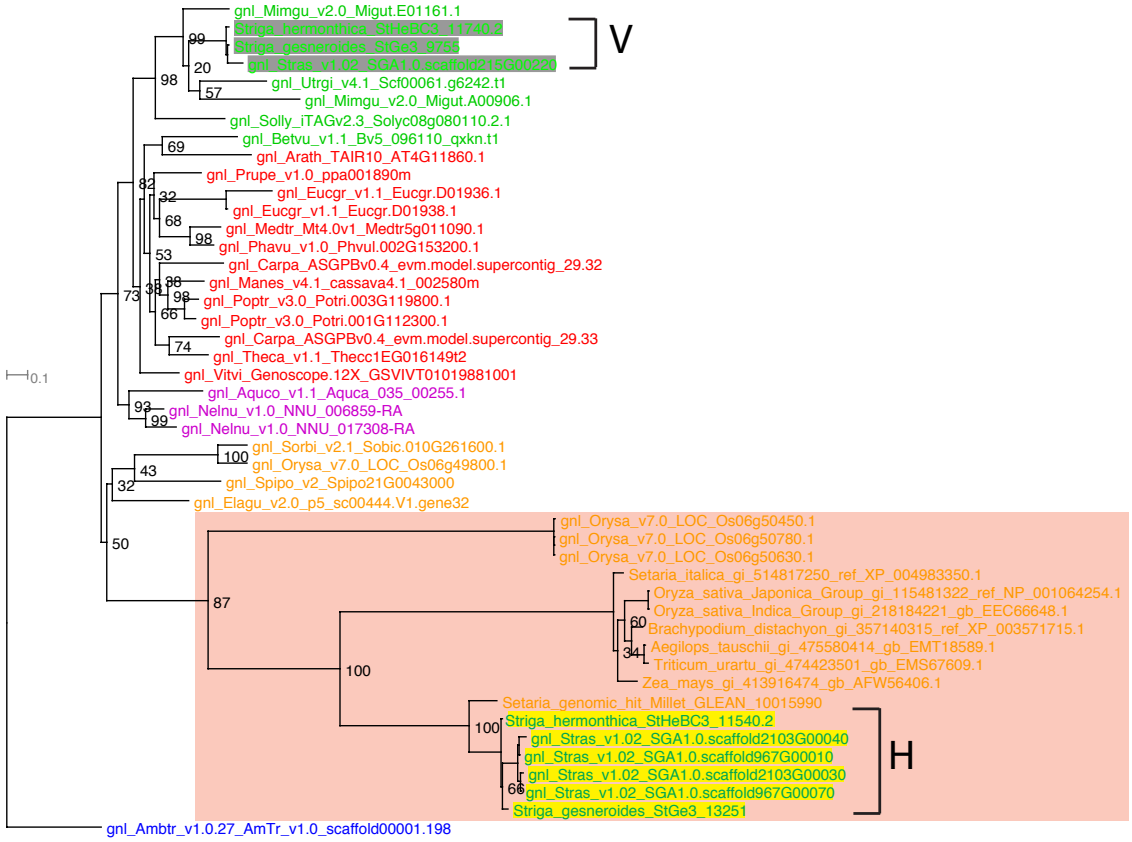


Figure **A-5**. Maximum Likelihood tree of a transcribed gene encoding an unknown protein with the forced codon alignment using RAxML. The HGT clade (labeled "H") identifies a transfer from an ancestor of *Setaria* to a shared ancestor of *Striga asiatica*, *Striga gesnerioides*, and *Striga hermonthica.* A vertically-transmitted lineage ("V") is also identified. The divergent clade that involves the horizontal transfer is highlighted with a pink background, and contains sequences from diverse grasses annotated as an unknown protein, whereas the remaining genes encode an amino-tRNA synthetase.

Appendix B

Evolution of strigolactone pathway in parasitic Orobanchaceae[14]

[14] This is an expanded work based on phylogenetic analyses I performed for the Das et al 2015 paper, and three additional papers.

B.1 Introduction

B.1.1 Strigolactone-dependent germination of *Striga* and *Phelipanche*

Obligate parasitic plants such as *Striga* and *Phelipanche* rely on secondary metabolites produced by the host to stimulate seed germination. These metabolites have been shown to be the plant hormone strigolactones (SL), carotenoid-derived germination stimulant that also serve a vital function in regulation of several other growth processes in plants. In addition, germination of *Striga* and *Phelipanche* (or *Orobanche*) spp. seeds require a (pre) conditioning step, which involves exposure to suitable moisture and temperature for several days before they become responsive to SLs (Bouwmeester et al. 2003).

B.1.2 Review of strigolactone pathway components and roles

Strigolactone (SL) was first discovered as root-derived hormone that can induce seed germination of parasitic plants (Hauck et al. 1992). Later it was also shown to have other primary roles in stimulating hyphae branching of arbuscular mycorrhizal fungi (Akiyama, et al. 2005) and inhibition of shoot branching (Gomez-Roldan, et al. 2008) . Studies of shoot branching mutants whose phenotype was complemented by the application of SL or GR24 (SL analog) revealed the identification of a complete catalog of genes in the SL pathway, for instance, MAX1 (more axillary growth) in *Arabidopsis*, which encodes a cytochrome P450 (Booker, et al. 2005). Orthologs of this gene were also identified in other species – ramosus (rms) in pea (Morris, et al. 2001), drawf (d) /high tillering dwarf (htd) in rice (Zhang, van Dijk, et al. 2014), and decreased apical dominance (dad) in petunia (Drummond, et al. 2011). Other genes involved in SL biosynthesis include *AtD27* (Carotene isomerase) (Waters, et al. 2012), *CCD7* and *CCD8* (carotenoid cleavage dioxygenase 7, 8) (Brewer, et al. 2009) and SL perception/signaling (*AtD14* - α/β-Hydrolase (Chevalier, et al. 2014), MAX2 – an F-box protein

(Nelson, et al. 2011a)). Starting with the role of isomerization of β-carotene by D27, followed by CCD7 and CCD8 converting carotenoids to carlactone, SL synthesis was completed by MAX1 which converts β-carlactone to the parent SLs of SL-like compounds. These four enzymes were shown sufficient to produce SLs in *Nicotiana benthamiana* (Beveridge and Kyozuka 2010; Al-Babili and Bouwmeester 2015). The binding of SL with SL receptor (D14) changes the confirmation of D14, which causes its interaction with the F-box protein MAX2, further targeting D53 for degradation. Degradation of the negative regulator – D53, instead activates the downstream SL signaling (Al-Babili and Bouwmeester 2015). The elucidation of SL pathway components also provides further insights on the functional roles of SL in plant growth and development. In addition to shoot branching, SLs also regulate primary root length and lateral root density, stimulate root hair elongation (Snowden, et al. 2005), inhibit adventitious root (Sun, et al. 2014), increase stem thickness by interaction with auxin, and accelerate leaf senescence (Al-Babili and Bouwmeester 2015).

B.1.3 Regulation of SL pathway

The regulation of the SL pathway is a fine-tuned process involving the induction by nutrient status such as nitrogen and phosphorus levels (Bonneau, et al. 2013). In particular, phosphorus starvation will stimulate SL production, which presumably acts to stimulate plants to establish symbiosis with mycorrhizae by stimulating hyphae branching of arbuscular fungi. In addition, the production of SLs is also affected by other plant hormones (Akiyama and Hayashi 2006). Interaction of SLs with auxin is believed to play a role in regulating shoot branching (Al-Babili and Bouwmeester 2015). In addition, SLs stimulate parasitic seed germination by upregulating genes encoding enzymes involved in degrading ABA, which is known to be involved in seed dormancy (Liu, et al. 2013).

B.1.4 Working hypothesis for SL-mediated germination of parasites

As obligate parasitic plants, *Striga* and *Phelipanche* depend on host signals for germination; this suggests that they either have evolved a mechanism to differentiate internal

SLs from external SLs, or they have completely lost their SL pathway (Das, et al. 2015). To test the latter hypothesis Das et al (2015) investigated the SL pathway genes in parasitic plants, in which I contributed to the phylogenetic analyses of each gene. The result showed that these parasitic plants have retained all the pathway components, suggesting a still functional SL biosynthesis pathway. This also suggested the differentiation between internal and external SL recognition, the result of which was supported by at least three groups by investigating the SL receptors.

B.2 Results and discussion

B.2.1 Conservation of strigolactone biosynthesis genes in parasitic plants

To test the first hypothesis, we examined each gene in the SL biosynthesis pathway. The four genes that are sufficient for the biosynthesis of SLs are D27, CCD7, CCD8, and MAX1. By identifying the gene in *Arabidopsis*, the use of 22 genome orthogroup classifications (Yang et al. 2015) allowed us to extract orthogroups to which the *Arabidopsis* gene is assigned. We then optimized the orthogroup tree for each gene (Figure B-1). To our surprise, all the parasitic plants including the SL-dependent species – *Striga* and *Phelipanche* retain these four genes involved in SL biosynthesis in full length. In addition, they are expressed in life stages of parasitic plants (Das et al. 2015). We then performed selective constraint analyses on all these four genes, and two additional genes involved in SL signaling – *D14* (SL receptor) and *MAX2* (an F-box protein). They showed strong purifying selection, and one of the genes - CCD7 exhibited even stronger purifying selection than orthologs in the non-parasitic ancestral lineage. These results indicate that SL biosynthesis is still retained and functional in parasitic plants, and still evolving under strong purifying selection. SLs are primarily produced in root, and consistently, SL biosynthesis genes CCD7, CCD8, and D27 primarily express in underground roots (Das et al. 2015). SL receptor, KAI2, is predominantly expressed in above ground shoots and inflorescences (Das et al. 2015), implicating a canonical role of SLs in shoot branching. This clearly suggests that SL biosynthesis and perception are essential in the parasite and are likely to be involved in plant growth and development such as shoot branching, as they are in nonparasitic plants. Additionally,

it suggests that losing the SL pathway is not the mechanism that parasites rely on to be able to distinguish host SL signals for germination. This implies an alternative hypothesis – parasitic plants must have evolved mechanisms to differentiate internal and external SL signals.

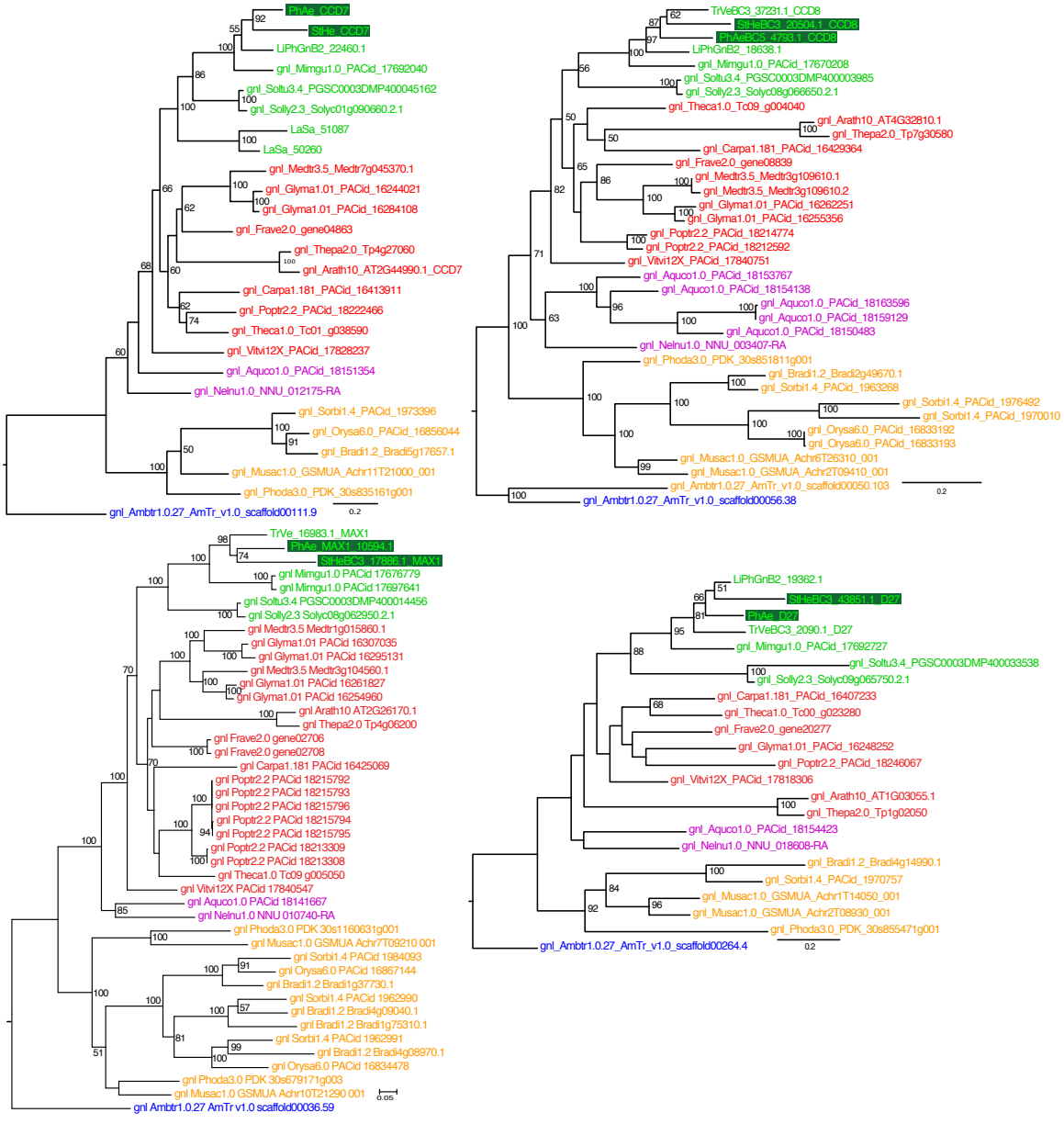


Figure **B-1**. RAxML-based maximum likelihood trees for four SL biosynthesis genes – CCD7, CCD8, D27, and MAX1 (Das et al. 2015). Genes in SL-dependent species – *Striga* and

Phelipanche are highlighted in blue, in *Triphysaria*, *Lindenbergia*, and *Mimulus* are highlighted in green.

B.2.2 Review of SL receptor diversification in parasitic *Striga*

In *Arabidopsis*, *D14* encodes an SL receptor that controls shoot branching and seed germination (by regulating the biosynthesis of several hormones including ABA, GA (Toh, et al. 2012), and ethylene (Sugimoto, et al. 2003)) whereas its paralog KAI2 encodes a karrikin receptor that stimulates seed germination by karrikin (Nelson et al. 2009), a butenolide present in smoke that stimulates seed germination for many plant species (Flematti, et al. 2004). However, karrikin-induced germination was inactive in parasitic weeds (Chiwocha, et al. 2009). Studies show the downstream signaling pathway induced by SL and karrikin shares some components such as an F-box protein MAX2 (Liu, et al. 2014). David Nelson's group reported two rounds of gene duplication in KAI2 of *Striga hermonthica* - one happened in the lineage of asteridae clade, the other was a lineage-specific duplication unique to parasitic plants (Conn et al. 2015). The latter duplication gave rise to many *S. hermonthica* genes named as *KAI2d* (and up to 11 homologs. These KAI2d sequences were highly divergent and as a result appeare on long branches (Conn et al. 2015). Functional divergence of these *karrikin* receptors resulted in neofunctionalization that switches them from a karrikin receptor to SL receptor for germination (Conn et al. 2015). D14 in parasitic plants, however, remains single-copy as in its non-parasitic ancestors (Conn et al. 2015). Because KAI2ds act a similar role as KAI2's paralog – D14 in both inducing SL-mediated seed germination, it's considered as convergent evolution (Conn et al. 2015). KAI2ds in parasitic plants, showed relaxed selective constraint compared to their non-parasitic ancestral orthologs, and a subset of the codons exhibited positive selection (Conn et al. 2015). Relaxed constraint of KAI2d has likely resulted in a switch of ligand binding from karrikin to SLs, as a germination-associated defect of *Atkai2* mutant was rescued by parasite KAI2ds only in the presence of GR24, not karrikin (Conn et al. 2015). Host-dependent germination of parasitic plants appeared to rely on MAX2, an F-box protein that interacts with D14 in parasites (Al-Babili and Bouwmeester 2015). In fact, MAX2 is able to mediate both karrikin and SL dependent signaling (Nelson, et al. 2011b) by directing the negative regulators of SL signaling for ubiquitin-

dependent proteasomal degradation (Zhou, et al. 2013) and its role is quite conserved across plants. *MAX2* from *Striga* is able to complement the *Arabidopsis max2* mutant phenotype of shoot branching (Liu et al. 2014).

Later additional studies by Toh et al (2015) showed that these KAI2ds are responsible for the sensitivity of parasites to SLs (Toh et al. 2015). Tsuchiya et al (2015) showed that SL receptor is essentially a hydrolase that catalyses SLs into ring-like compounds and he designed a probe Yoshimulactone Green (YLG) which activates SL signaling and produces fluorescence, and can work both *in vitro* and *in vivo* to probe the activity of SL receptors (Tsuchiya et al. 2015). Moreover, they showed *Striga ShHTL7* could complement the *Arabidopsis Athtl-3* mutant phenotype. The mutant was defective in SL (GR24)-stimulated seed germination at high temperature (a phenomenon called thermoinhibition, which could be alleviated by GR24 in the presence of *AtHTL3* (Toh et al. 2015)), proving its activity as SL receptor (Tsuchiya et al. 2015). His assay showed that multiple *ShHTLs* could recognize a structurally diverse array of SLs, whereas Toh et al (2015) showed they also diversified to show varying degrees of sensitivity to SLs (Toh et al. 2015). In particular, StHTL7 shows the highest sensitivity to SLs as it stimulated germination at picomolar concentrations for naturally occurring strigolactones (Toh et al. 2015). The diversification of SLs with varying degrees of SL sensitivity allows parasitic *Striga* to sense SLs of a wide range of concentration, a trait that can enhance parasite germination.

B.2.3 Relevance of SL receptor diversification in internal and external SL recognition

These studies repeatedly point to one conclusion that diversification of KAI2d in parasitic lineages (at least in *Striga*), the orthologs of *Arabidopsis AtKAI2* plays a role in perceiving SLs produced by their hosts for seed germination. In this sense, these genes act as external SL regulators for seed germination. However, it is unknown which genes are involved in internal SL signaling involved in other processes such as shoot branching or root growth. The possible candidates are KAI2c or D14; which of the two genes is the internal SL receptor could likely be determined by careful analysis of their expression or through a functional complementation assay. Based on our expression analyses, the KAI2c in all three parasitic taxa show upregulated

expression in above-ground tissues (Figure B-5). D14, however, showed elevated expression in haustorial tissues of all three parasites, but high expression was found in only *S. hermonthica* and *P. aegyptiaca* (Figure B-2). This suggests that *KAI2c* is likely the canonical internal SL receptor.

D14 in nonparasitic plants also acts as SL areceptor, and D14 shows strong purifying selection (Conn et al. 2015), indicating D14 in parasitic plants may still act as a SL receptor. KAI2ds, however, act as SL receptors to external SL signals. It thus has become clear that there is a differentiation of SL receptor responding to internal and external SLs in parasitic plants. Our study, however, shows that parasitic plants still retain all the pathway components for SL biosynthesis and perception (Das et al. 2015), suggesting that SLs may regulate other processes in addition to the canonical role of shoot branching and root growth. The pathway involving SL biosynthesis seems rather conserved between nonparasitic plants and parasitic plants (for instance, *Striga* SL biosynthesis genes (*ShMAX2* for instance) could rescue *Arabidopsis max2* mutant (Liu et al. 2014)), whereas the downstream signaling is expected to differ significantly. This is suggested because there has been a diversification of SL receptors, and there are differences between D14 (the internal SL receptor) and external receptor ShKAI2d (or ShHTLs) responding to host signals. In *Arabidopsis*, D14 cannot rescue the *Atkai2* mutant phenotype, indicating that they interact with different downstream components (Conn et al. 2015). It is known that D14 controls shoot branching phenotype, a role similar to auxin. It is possible that interaction of D14 and MAX2 activates genes that regulate auxin biosynthesis. As KAI2ds are involved in SL-mediated seed germination, it makes it possible to speculate that KAI2d's interaction with other components may activate ABA-related genes that control seed dormancy.

In short, we proposed in parasitic plants, D14 acts as an internal SL receptor regulating downstream signaling in shoot branching, perhaps in cooperation with the auxin signaling pathway. KAI2d, acts as a SL receptor to respond to exogenous SLs, for instance, those produced by their host. To test this hypothesis, one proposed experiment is functional complementation. It is expected that *Striga* D14 will complement the *Atd14* mutant phenotype in shoot branching, whereas *Striga* KAI2d will not. On the other hand, the *Atkai2* mutant phenotype of reduced germination rate at high temperature in response to GR24 is expected to be complemented by the *Striga KAI2ds* by application of SLs GR24, whereas it cannot be rescued by *Striga D14*. To further explore what downstream genes are involved in internal SL-regulated shoot branching,

and external SL-mediated seed germination, one may generate *D14* and *KAI2d* mutants in parasites, and examine what downstream pathway genes could be affected. We predict the silencing of D14 may affect genes at least involved in auxin-signaling, while silencing of *KAI2d* may affect genes at least encoding ABA-degrading enzymes.

B.2.4 Relevance of *KAI2d (HTL)* in host recognition and specificity

These papers also suggest that multiple KAI2ds (or HTLs) could be involved in recognition of different hosts or contribute to host specificity. A recent study showed all *Striga* ShHTLs have high affinity to a type of SL called 5DS, which is produced by many grass hosts of *Striga*. This may explain why *Striga* is a grass specialist (Tsuchiya et al. 2015). In addition, differing levels of sensitivity to SLs of varying concentration may also be a strategy used by *Striga* to establish successful connection with hosts. Certain resistant *Sorghum* cultivars produce a reduced amount of SLs that failed to stimulate germination of parasites (Yoneyama et al. 2010).

On the other hand, our analyses (Figure B-5) as well as a list of other studies reported around a dozen HTLs or KAI2ds in parasitic *Striga* (Conn et al. 2015; Toh et al. 2015; Tsuchiya et al. 2015). However, less than a dozen was present in *Phelipanche* (Figure B-5), which also exhibits host-dependent germination like *Striga*. *Striga* appears to recognize strigols produced by grass hosts, whereas *Phelipanche* or *Orobanche* recognizes orobanchol from a host for germination (Yoneyama et al. 2010). A synthetic SL GR24 is able to initiate germination of both *Striga* and *Phelipanche* (Matusova et al. 2005). It is unknown if different types of naturally occurring SL could induce specific parasites for germination. Future efforts could include performing a germination assay in which SLs produced from grasses (*Striga*'s host) are isolated and used to treat *Phelipanche* to measure the germination rate, and vice versa. An alternative hypothesis could be that multiple *ShHTLs* are required to initiate downstream germination signaling, whereas *Pheliepanche PaHTLs* have high sensitivity to SLs so that a small number of HTLs are enough to initiate germination. The study by Toh et al (2015) showed a comparison of key amino acids involved in ligand binding of multiple *ShHTLs*. He found that the more identical key amino acids determining the substrate binding activity of the enzyme (HTLs) shared between

ShHTL peptide sequence and AtHTL peptide sequence, the lower the sensitivity to SL. There appeared to be a correlation between key amino acid divergence (relative to *Arabidopsis* AtHTL (KAI2)) and SL sensitivity. He showed that the predicted peptide sequences of ShHTL7 are more divergent than the peptide sequences of *Arabidopsis thaliana* HTL (AtKAI2) in terms of key amino acids that determine the enzyme activity, and exhibit the highest sensitivity to SL based on a complementation-based germination assay. So one can compare all members of PaHTLs (PaKAI2ds) with *Arabidopsis* AtHTL and examine if there are more copies that are as divergent as ShHTL7 relative to AtHTL. An additional potentially useful experiment would be to overexpress all members of *PaHTL* (PaKAI2d) into *Arabidopsis Athtl1* mutant followed by a GR24 induced germination assay, to examine if they show high levels of sensitivity to GR24.

B.2.5 Transcriptional dynamic of *KAI2* members among three parasites

We also examined the expression of KAI2s in parasitic plants. Consistent with the new findings, the expression of KAI2ds indicates a role in germination and early seedling growth. In all three parasites (*Triphysaria*, *Striga*, and *Phelipanche*), KAI2ds show upregulated expression in early seedlings (stage 1 and stage 2) (Figure B-5). In addition, in both *Striga* and *Phelipanche*, its expression is also observed in stage 0 (Figure B-5), indicating its role in controlling seed germination of *Striga* and *Phelipanche* may differ from *Triphysaria*. Both *Striga* and *Phelipanche* depend on host SLs for germination, whereas *Triphysaria* could germinate independent of host germination stimulant, indicating fine spatial and temporal regulation of KAI2d could be important for SL-induced germination.

Patterns of KAI2c and KAI2d expression also differ among the three parasitic plants. KAI2is, the intermediate KAI2s, are implicated to respond to both karrikins and SLs. Based on the predicted pocket structure of the encoded enzyme for KAI2i (Conn et al. 2015), as well as their upregulated expression in both stages 0 (germination) and 5.1 (shoots) (Figure B-5), KAI2is are predicted to play roles in both shoot branching and germination. It is unclear why *Phelipanche* doesn't have the intermediate KAI2is. Studies by Toh et al (2015) showed that some copies of ShHTL show binding affinity to karrikin yet fail to induce germination, suggesting

binding to stimulant is not sufficient for a predicted biological role. We speculate that as an advanced holoparasite, it may have shed the intermediate genes. The upregulated expression of KAI2c in shoots and floral stages of *Phelipanche* and *Striga* (Figure B-5) indicate that it is not playing the canonical role in germination, but more similar to roles related to shoot branching, a role similar to D14 (the internal SL receptor). However, an additional expression in stage 1 (Figure B-5) indicates facultative *Triphysaria* may still involve a role of KAI2c in regulating early seed germination.

B.2.6 Parasitic D14s show abundant expression in interface and haustoria

In addition to the reported roles of *ShHTLs* in germination, we also examined the expression profiles of SL-pathway components in parasitic plants by looking at their expression in parasitic developmental stages with the PPGP2 data (http://ppgp.huck.psu.edu/). In terms of D14, expression in underground shoots (5.1) and above-ground shoots (6.1) in *Phelipanche* and expression in above-ground shoots (6.1) in *Striga* indicate a role of this gene involved in shoot branching (Figure B-2). Interestingly, this gene also shows high expression in interface tissues of both *Triphysaria* and *Phelipanche*, as well as abundant expression in haustoria (stage 3) of *Striga* (Figure B-2). This indicates a likely role of the SL receptor involved in haustoria development. D14 is an SL receptor of internal SLs, indicating a likelihood of internal SLs transported to haustoria tissue. Future efforts could examine the expression of SL transporter (PDR12) in parasites to support whether SLs either from the parasite or host side can regulate haustorial growth.

B.2.7 Haustorial expression of additional SL pathway genes implies a role of SL in haustorial development

The comparison of gene expression for D14 and KAI2s revealed that D14, instead of KAI2s, show upregulated haustorial expression in parasites. As D14 is the internal SL receptor involved in shoot branching in nonparasitic species, its upregulation in haustorial tissues of

parasites suggests that D14 may have neofunctionalized to function in haustoria development. Thus, SLs that are known as a germination stimulant for parasitic seeds may have additional roles in haustoria development in parasitic plants. This idea is supported by additional evidence of upregulated haustoria or interface expression for CCD8 (Figure B-3) and MAX1 (Figure B-4), two of which are among the four enzymes required for SL production. Significantly, Aly et al (2014) showed knock down of *P. egyptiaca CCD7* and *CCD8* resulted in reduced tubercle growth by using host-induced gene silencing (HIGS) system in parasite-tobacco interactions. Their paper suggested a novel role of SLs involved in tubercle growth and development, however, they concluded a role of SLs in seed germination. On the other hand, haustoria growth towards the host may mimic the hyphae growth of arbuscular fungi towards their plant host, making a scenario for diverse process of SL regulating similar biological processes. Last but not least, the formation of shoots in underground soil seems a rather unexpected occurrence if we believe shoots are expected to occur above ground. We would also like to propose a possible link between the development of underground shoots and the production of SLs at the preceding tubercle stages, especially considering that SLs have a canonical role in regulating shoot branching (Gomez-Roldan et al. 2008). Future experiments are needed to examine which partner interacts with D14, which may lead to the identification of downstream signaling components involved in haustoria development.

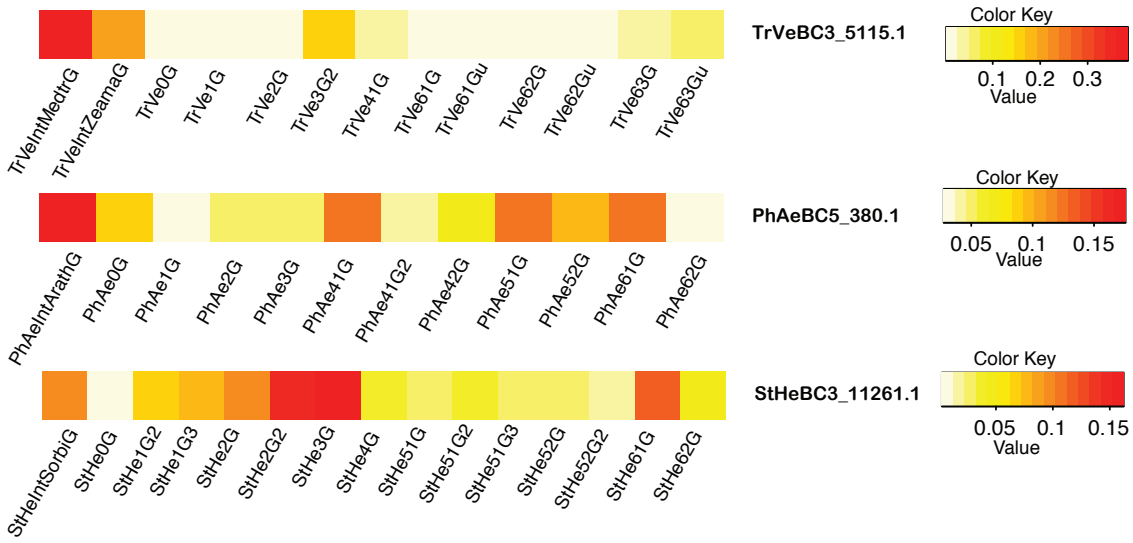


Figure **B-2**. Expression of D14 in all three parasitic plants. Expression is shown with a heat map scaled for across-gene comparison, the intensity represents normalized z-scores from FPKM.

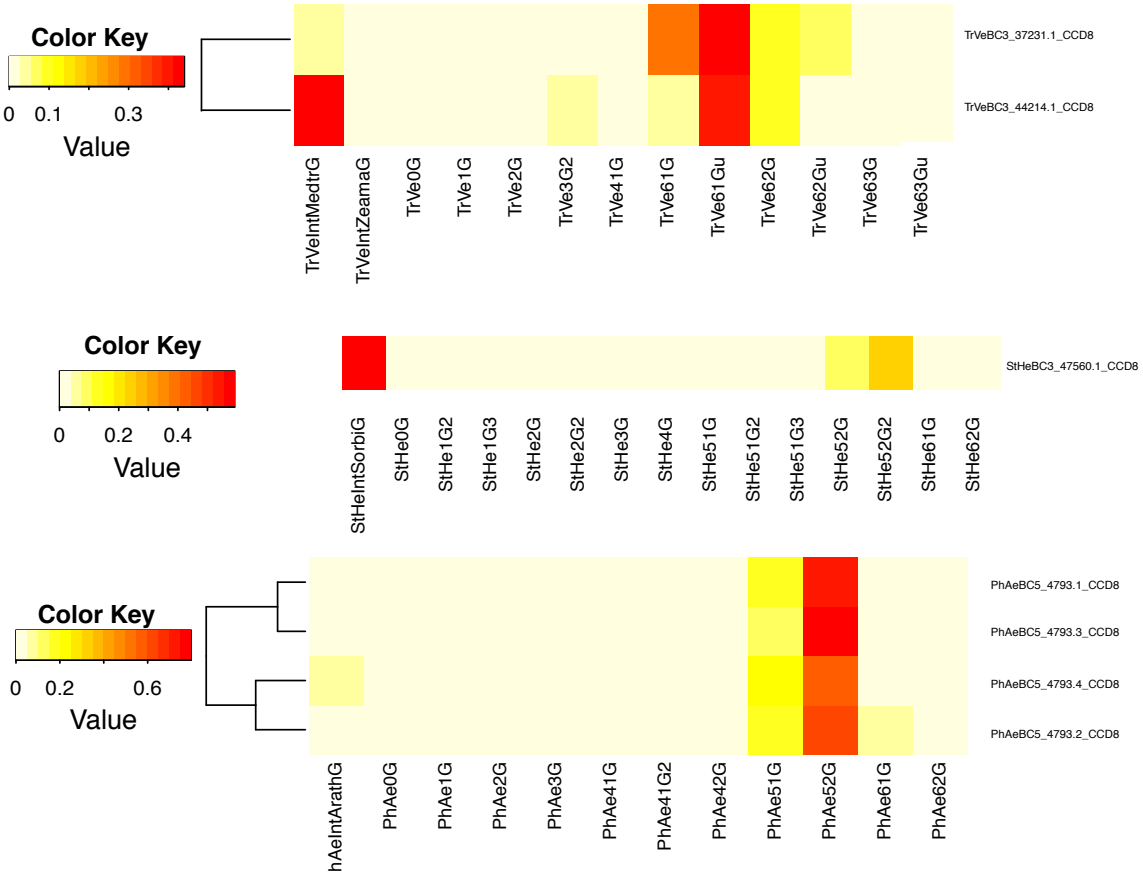


Figure **B-3**. CCD8 expression in parasitic plants.

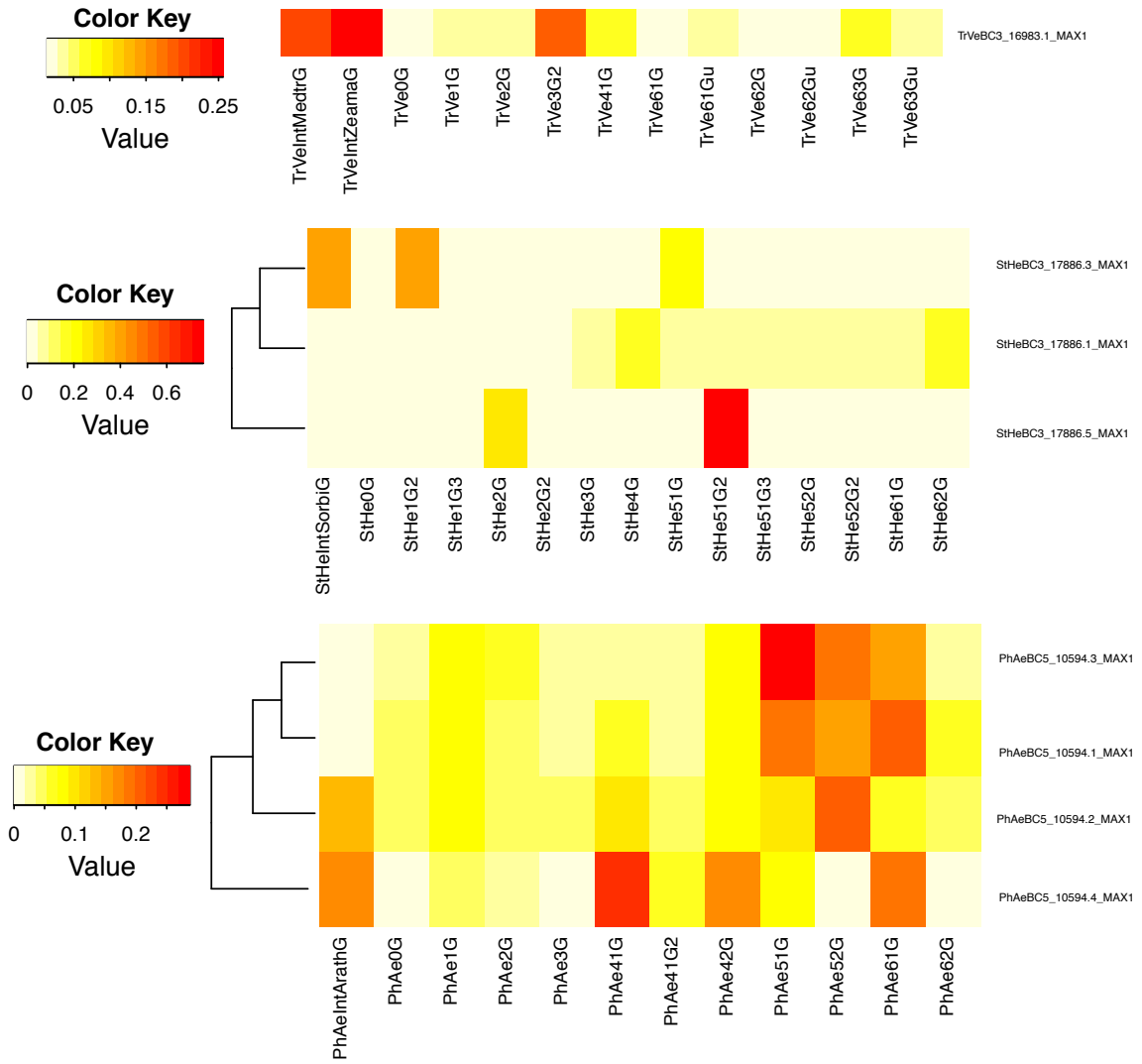


Figure **B-4**. MAX1 expression in parasitic plants.

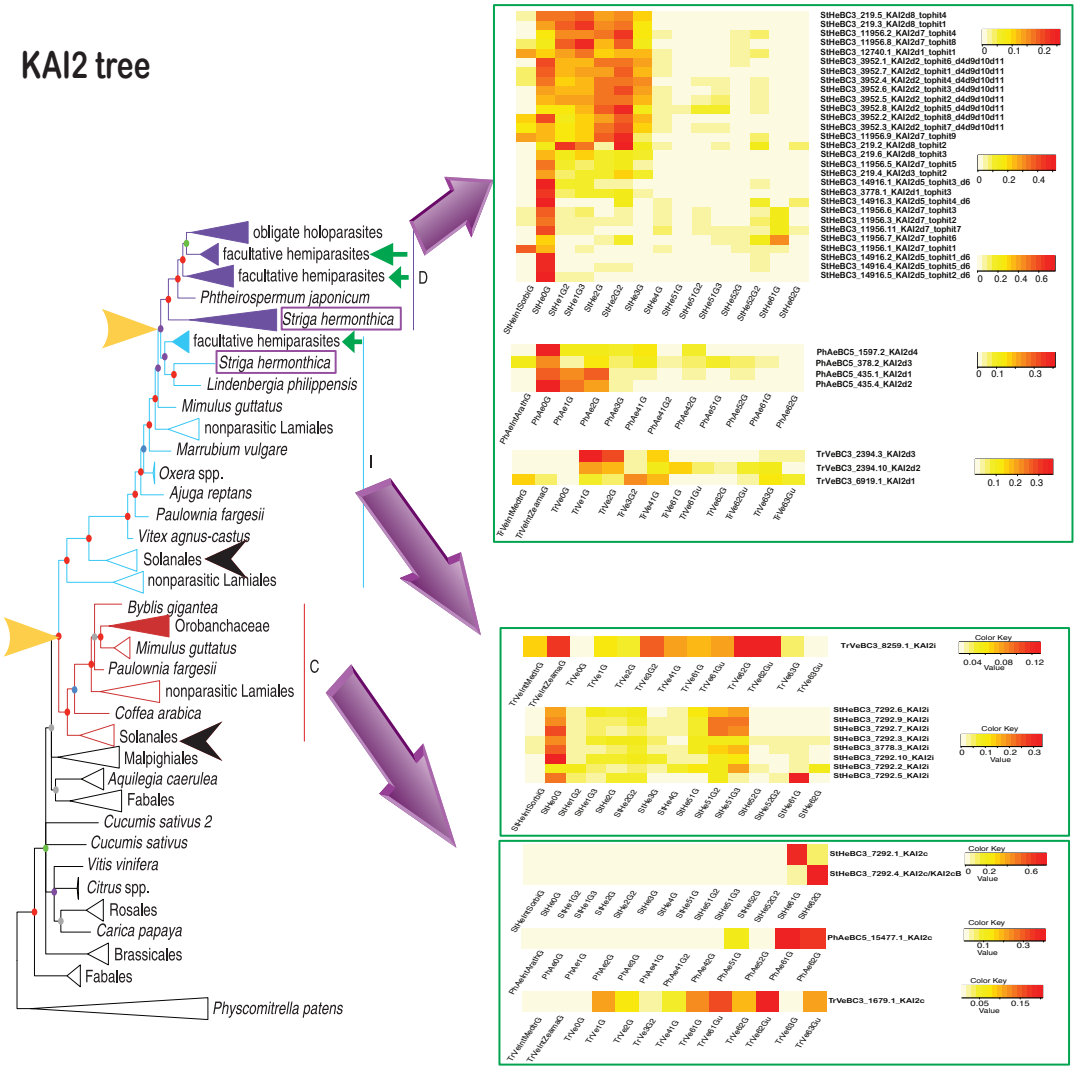


Figure **B-5**. Expression of KAI2s (KAI2c, KAI2i, KAI2d) mapping onto phylogeny in all three parasitic plants. Left part represents the KAI2 phylogeny from David Nelson's paper, right are expression heatmaps for all unigenes encoding KAI2c (C), KAI2i (I) and KAI2ds (D) in all three parasites. Arrows link the expression with phylogeny. Expression is shown with a heat map, the intensity represents normalized z-scores from FPKM.

References

Abbes Z, Kharrat M, Delavault P, Chaïbi W, Simier P. 2009. Nitrogen and carbon relationships between the parasitic weed *Orobanche foetida* and susceptible and tolerant faba bean lines. Plant Physiol Biochem 47:153-159.

Aber M, Fer A, Salle´ G. 1983. Etude du transfert des substances organiques de l'hoˆte (*Vicia faba*) vers le parasite (*Orobanche crenata* Forsk.). Z Pflanzenphysiol 112:297-308.

Abramovitch RB, Anderson JC, Martin GB. 2006. Bacterial elicitation and evasion of plant innate immunity. Nat Rev Mol Cell Biol 7:601-611.

Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee SJ, Yeats TH, Egan AN, Doyle JJ et al. 2012. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proc Natl Acad Sci U S A 109:4197-4202.

Adams KL, Clements MJ, Vaughn JC. 1998. The Peperomia mitochondrial coxI group I intron: timing of horizontal transfer and subsequent evolution of the intron. J Mol Evol 46:689-696.

Adams KL, Wendel JF. 2005. Novel patterns of gene expression in polyploid plants. Trends Genet 21:539-543.

Akiyama K, Hayashi H. 2006. Strigolactones: chemical signals for fungal symbionts and parasitic weeds in plant roots. Ann Bot 97:925-931.

Akiyama K, Matsuzaki K, Hayashi H. 2005. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. Nature 435:824-827.

Al-Babili S, Bouwmeester HJ. 2015. Strigolactones, a novel carotenoid-derived plant hormone. Annu Rev Plant Biol 66:161-186.

Alakonya A, Kumar R, Koenig D, Kimura S, Townsley B, Runo S, Garces HM, Kang J, Yanez A, David-Schwartz R et al. 2012. Interspecific RNA interference of *SHOOT MERISTEMLESS-Like* disrupts *Cuscuta pentagona* plant parasitism. Plant Cell 24:3153-3166.

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

Aly R, Cholakh H, Joel DM, Leibman D, Steinitz B, Zelcer A, Naglis A, Yarden O, Gal-On A. 2009. Gene silencing of mannose 6-phosphate reductase in the parasitic weed Orobanche aegyptiaca through the production of homologous dsRNA sequences in the host plant. Plant Biotechnol J 7:487-498.

Aly R, Dubey N, Yahyaa M, Abu-Nassar J, Ibdah M. 2014. Gene silencing of CCD7 and CCD8 in Phelipanche aegyptiaca by tobacco rattle virus system retarded the parasite development on the host. Plant Signal Behav 9.

Aly R, Hamamouch N, Abu-Nassar J, Wolf S, Joel DM, Eizenberg H, Kaisler E, Cramer C, Gal-On A, Westwood JH. 2011. Movement of protein and macromolecules between host plants and the parasitic weed *Phelipanche aegyptiaca* Pers. Plant Cell Rep 30:2233-2241.

Amborella Genome Project. 2013. The *Amborella* genome and the evolution of flowering plants. Science 342:1241089.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol 11:R106.

Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot J Linn Soc 161:105-121.

Antao CM, Malcata FX. 2005. Plant serine proteases: biochemical, physiological and molecular features. Plant physiology and biochemistry 43:637-650.
Antón PS, Silberglitt R, Schneider J. 2015. The Global Technology Revolution: Bio/Nano/Materials Trends and Their Synergies with Information Technology by 2015 (1st ed.): RAND Corporation.
Antonova TS, TerBorg SJ. 1996. The role of peroxidase in the resistance of sunflower against *Orobanche cumana* in Russia. Weed Res 36:113-121.
Archibald JM, Richards TA. 2010. Gene transfer: anything goes in plant mitochondria. BMC Biol 8:147.
Ascencio-Ibanez JT, Sozzani R, Lee TJ, Chu TM, Wolfinger RD, Cella R, Hanley-Bowdoin L. 2008. Global analysis of *Arabidopsis* gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. Plant physiology 148:436-454.
Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics 25:25-29.
Atkinson TJ, Halfon MS. 2014. Regulation of gene expression in the genomic context. Comput Struct Biotechnol J 9:e201401001.
Atsatt PR. 1973. Parasitic flowering plants: how did they evolve? The American Naturalist 107:502-510.
Baird WV, Riopel JL. 1984. Experimental studies of haustorium initiation and early development in *Agalinis purpurea* (L) Raf (Scrophulariaceae). American Journal of Botany 71:803-814.
Baird WV, Riopel JL. 1985. Surface characteristics of root haustorial hairs of parasitic Scrophulariaceae. Bot Gaz 146:63-69.
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 22:1404-1419.
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 Physiology 158:1046-1053.
Bandaranayake PCG, Yoder J. 2013a. Early haustorium development. In: Joel DM, Gressel J, Musselman LJ, editors. Parasitic Orobanchaceae - parasitic mechanisms and control strategies. Springer Heidelberg New York Dordrecht London: Springer. p. 62.
Bandaranayake PCG, Yoder J. 2013b. Evolutionary origins. In: Joel DM, Gressel J, Musselman LJ, editors. Parasitic Orobanchaceae - parasitic mechanisms and control strategies. Springer Heidelberg New York Dordrecht London: Springer. p. 69-70.
Bandaranayake PCG, Yoder JI. 2013c. Trans-specific gene silencing of acetyl-CoA carboxylase in a root-parasitic plant. Mol Plant Microbe In 26:575-584.
Bapteste E, O'Malley MA, Beiko RG, Ereshefsky M, Gogarten JP, Franklin-Hall L, Lapointe FJ, Dupre J, Dagan T, Boucher Y et al. 2009. Prokaryotic evolution and the tree of life are two different things. Biol Direct 4:34.
Barkman TJ, McNeal JR, Lim SH, 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.
Becker A, Theissen G. 2003. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol 29:464-489.
Belfort M, Roberts RJ. 1997. Homing endonucleases: keeping the house in order. Nucleic Acids Res 25:3379-3388.

Bennett JR, Mathews S. 2006. Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. American Journal of Botany 93:1039-1051.
Bent AF, Mackey D. 2007. Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. Annu Rev Phytopathol 45:399-436.
Bergthorsson U, Adams KL, Thomason B, Palmer JD. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424:197-201.
Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm Amborella. Proc Natl Acad Sci U S A 101:17747-17752.
Beveridge CA, Kyozuka J. 2010. New genes in the strigolactone-related shoot branching pathway. Curr Opin Plant Biol 13:34-39.
Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. Genome Research 14:988-995.
Bleischwitz M, Albert M, Fuchsbauer HL, Kaldenhoff R. 2010. Significance of Cuscutain, a cysteine protease from *Cuscuta reflexa*, in host-parasite interactions. BMC Plant Biology 10:227.
Boller T, Felix G. 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Annu Rev Plant Biol 60:379-406.
Boller T, He SY. 2009. Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. Science 324:742-744.
Bonke M, Thitamadee S, Mahonen AP, Hauser MT, Helariutta Y. 2003. APL regulates vascular tissue identity in Arabidopsis. Nature 426:181-186.
Bonneau L, Huguet S, Wipf D, Pauly N, Truong HN. 2013. Combined phosphate and nitrogen limitation generates a nutrient stress transcriptome favorable for arbuscular mycorrhizal symbiosis in Medicago truncatula. New Phytol 199:188-202.
Booker J, Sieberer T, Wright W, Williamson L, Willett B, Stirnberg P, Turnbull C, Srinivasan M, Goddard P, Leyser O. 2005. MAX1 encodes a cytochrome P450 family member that acts downstream of MAX3/4 to produce a carotenoid-derived branch-inhibiting hormone. Dev Cell 8:443-449.
Bouwmeester HJ, Matusova R, Zhongkui S, Beale MH. 2003. Secondary metabolite signalling in host-parasitic plant interactions. Curr Opin Plant Biol 6:358-364.
Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, Monteiro A, French V, Carroll SB. 1996. Development, plasticity and evolution of butterfly eyespot patterns. Nature 384:236-242.
Brewer PB, Dun EA, Ferguson BJ, Rameau C, Beveridge CA. 2009. Strigolactone acts downstream of auxin to regulate bud outgrowth in pea and Arabidopsis. Plant Physiol 150:482-493.
Brini M, Carafoli E. 2011. The plasma membrane Ca^{2+} ATPase and the plasma membrane sodium calcium exchanger cooperate in the regulation of cell calcium. CSH Perspect Biol 3:a004168.
Brinkworth RI, Prociv P, Loukas A, Brindley PJ. 2001. Hemoglobin-degrading, aspartic proteases of blood-feeding parasites - substrate specificity revealed by homology models. J Biol Chem 276:38844-38851.
Brown R, Greenwood AD, Johnson AW, Long AG. 1951. The stimulant involved in the germination of Orobanche minor Sm. I. Assay technique and bulk preparation of the stimulant. Biochem J 48:559-564.
Brown R, Johnson AW, et al. 1949. The stimulant involved in the germination of Striga hermonthica. Proc R Soc Lond B Biol Sci 136:1-12.
Butler LG. 1995. Chemical communication between the parasitic weed Striga and its crop host. A new dimension in allelochemistry. In: Inderjit KM, Dakshini M, Enhelling FA, editors. Allelopathy, organisms, processes and applications. Wahington, DC. p. 158-166.

Cao FY, Yoshioka K, Desveaux D. 2011. The roles of ABA in plant-pathogen interactions. J Plant Res 124:489-499.
Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-1973.
Caplan J, Padmanabhan M, Dinesh-Kumar SP. 2008. Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. Cell Host Microbe 3:126-135.
Index of Orobanchaceae [Internet]. 2005. Available from: http://www.farmalierganes.com/Otrospdf/publica/Orobanchaceae Index.htm
Carroll SB, Gates J, Keys DN, Paddock SW, Panganiban GE, Selegue JE, Williams JA. 1994. Pattern formation and eyespot determination in butterfly wings. Science 265:109-114.
Champagne C, Sinha N. 2004. Compound leaves: equal to the sum of their parts? Development 131:4401-4412.
Chao DY, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE. 2013. Polyploids exhibit higher potassium uptake and salinity tolerance in Arabidopsis. Science 341:658-659.
Chen J, Ouyang Y, Wang L, Xie W, Zhang Q. 2009. Aspartic proteases gene family in rice: Gene structure and expression, predicted protein features and phylogenetic relation. Gene 442:108-118.
Chevalier F, Nieminen K, Sanchez-Ferrero JC, Rodriguez ML, Chagoyen M, Hardtke CS, Cubas P. 2014. Strigolactone promotes degradation of DWARF14, an alpha/beta hydrolase essential for strigolactone signaling in Arabidopsis. Plant Cell 26:1134-1150.
Chittapur BM, Hunshal CS, Shenoy H. 2000. Allelopathy in parasitic weeds management: Role of catch and trap crops. Allelopathy J 8:147-160.
Chiwocha SDS, Dixon KW, Flematti GR, Ghisalberti EL, Merritt DJ, Nelson DC, Riseborough JAM, Smith SM, Stevens JC. 2009. Karrikins: A new family of plant growth regulators in smoke. Plant Science 177:252-256.
Cho Y, Qiu YL, Kuhlman P, Palmer JD. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc Natl Acad Sci U S A 95:14244-14249.
Christin PA, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP. 2012. Adaptive evolution of C(4) photosynthesis through recurrent lateral gene transfer. Curr Biol 22:445-449.
Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. Nature 353:31-37.
Conant GC, Wolfe KH. 2008. Turning a hobby into a job: How duplicated genes find new functions. Nature Reviews Genetics 9:938-950.
Conn CE, Bythell-Douglas R, Neumann D, Yoshida S, Whittington B, Westwood JH, Shirasu K, Bond CS, Dyer KA, Nelson DC. 2015. PLANT EVOLUTION. Convergent evolution of strigolactone perception enabled host detection in parasitic plants. Science 349:540-543.
Crow KD, Wagner GP. 2006. What is the role of genome duplication in the evolution of complexity and diversity? Molecular Biology and Evolution 23:887-892.
Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A et al. 2006. Widespread genome duplications throughout the history of flowering plants. Genome Res 16:738-749.
Dagan T, Artzy-Randrup Y, Martin W. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. Proc Natl Acad Sci U S A 105:10039-10044.
Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. 1990. Evidence for Horizontal Transmission of the P-Transposable Element between Drosophila Species. Genetics 124:339-355.
Darwin C. 1859. On the origin of species by means of natural selection or the preservation of favoured races in the struggle of life. London.

Das M, Fernandez-Aparicio M, Yang Z, Huang K, Wickett N, Alford S, Wafula E, Depamphilis C, Bouwmeester H, Timko MP et al. 2015. Parasitic Plants *Striga* and *Phelipanche* dependent upon exogenous strigolactones for germination have retained genes for strigolactone biosynthesis. American Journal of Plant Sciences 6.
Dash S, Van Hemert J, Hong L, Wise RP, Dickerson JA. 2012. PLEXdb: gene expression resources for plants and plant pathogens. Nucleic acids research 40:D1194-D1201.
Davies J, Davies D. 2010. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 74:417-433.
Davis CC, Anderson WR, Wurdack KJ. 2005. Gene transfer from a parasitic flowering plant to a fern. Proc Biol Sci 272:2237-2242.
Davis CC, Wurdack KJ. 2004. Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. Science 305:676-678.
De Groote H, Wangare L, Kanampiu F, Odendo M, Diallo A, Karaya H, Friesen D. 2008. The potential of a herbicide resistant maize technology for Striga control in Africa. Agricultural Systems 97:83-94.
Delorenzo G, Cervone F, Hahn MG, Darvill A, Albersheim P. 1991. Bacterial endopectate lyase - evidence that plant-cell wall pH prevents tissue maceration and increases the half-life of elicitor-active oligogalacturonides. Physiol Mol Plant P 39:335-344.
dePamphilis CW, Palmer JD. 1990. Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. Nature 348:337-339.
dePamphilis CW, Young ND, Wolfe AD. 1997. Evolution of plastid gene *rps2* in a lineage of hemiparasitic and holoparasitic plants: Many losses of photosynthesis and complex patterns of rate variation. P Natl Acad Sci USA 94:7367-7372.
Der JP, Barker MS, Wickett NJ, dePamphilis CW, Wolf PG. 2011. *De novo* characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. BMC genomics 12:99.
Diao X, Freeling M, Lisch D. 2006. Horizontal transfer of a plant transposon. PLoS Biol 4:e5.
Dorr I. 1997. How *Striga* parasitizes its host: A TEM and SEM study. Ann Bot-London 79:463-472.
Dos Santos CV, Letousey P, Delavault P, Thalouarn P. 2003. Defense Gene Expression Analysis of Arabidopsis thaliana Parasitized by Orobanche ramosa. Phytopathology 93:451-457.
Draie R, Peron T, Pouvreau JB, Veronesi C, Jegou S, Delavault P, Thoiron S, Simier P. 2011. Invertases involved in the development of the parasitic plant Phelipanche ramosa: characterization of the dominant soluble acid isoform, PrSAI1. Mol Plant Pathol 12:638-652.
Drummond RS, Sheehan H, Simons JL, Martinez-Sanchez NM, Turner RM, Putterill J, Snowden KC. 2011. The Expression of Petunia Strigolactone Pathway Genes is Altered as Part of the Endogenous Developmental Program. Front Plant Sci 2:115.
Duarte JM, Cui L, Wall PK, Zhang Q, Zhang X, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. Mol Biol Evol 23:469-478.
Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW. 2010. Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. BMC Evol Biol 10:61.
Eddy SR. 2011a. Accelerated profile HMM searches. Plos Computational Biology 7:e1002195.
Eddy SR. 2011b. Accelerated Profile HMM Searches. PLoS Comput Biol 7:e1002195.
Eddy SR. 2011c. Accelerated Profile HMM Searches. Plos Computational Biology 7.
Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.
Ehleringer J, Marshall J. 1995. Water relations. In: Press MC, Graves JD, editors. Parasitic plants. London: Chapman & Hall. p. 125-140.

Ellers J, Kiers ET, Currie CR, McDonald BR, Visser B. 2012. Ecological interactions drive evolutionary loss of traits. Ecol Lett 15:1071-1082.
Estabrook EM, Yoder JI. 1998. Plant-plant communications: Rhizosphere signaling between parasitic angiosperms and their hosts. Plant physiology 116:1-7.
Fawcett JA, Maere S, Van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proc Natl Acad Sci U S A 106:5737-5742.
Fei ZJ, Joung JG, Tang XM, Zheng Y, Huang MY, Lee JM, McQuinn R, Tieman DM, Alba R, Klee HJ et al. 2011. Tomato Functional Genomics Database: a comprehensive resource and analysis package for tomato functional genomics. Nucleic acids research 39:D1156-D1163.
Felsenstein J. 1985. Confidence-limits on phylogenies - an approach using the bootstrap. Evolution 39:783-791.
Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331-368.
Flematti GR, Ghisalberti EL, Dixon KW, Trengove RD. 2004. A compound from smoke that promotes seed germination. Science 305:977.
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.
Foreman J, Demidchik V, Bothwell JH, Mylona P, Miedema H, Torres MA, Linstead P, Costa S, Brownlee C, Jones JD et al. 2003. Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. Nature 422:442-446.
Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD. 2002. Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457-1467.
Galtier N, Daubin V. 2008. Dealing with incongruence in phylogenomic analyses. Philos Trans R Soc Lond B Biol Sci 363:4023-4029.
Gao C, Ren X, Mason AS, Liu H, Xiao M, Li J, Fu D. 2014. Horizontal gene transfer in plants. Funct Integr Genomics 14:23-29.
Gehring WJ. 2011. Chance and necessity in eye evolution. Genome biology and evolution 3:1053-1066.
Goda H, Sawa S, Asami T, Fujioka S, Shimada Y, Yoshida S. 2004. Comprehensive comparison of auxin-regulated and brassinosteroid-regulated genes in *Arabidopsis*. Plant physiology 134:1555-1573.
Gomez-Roldan V, Fermas S, Brewer PB, Puech-Pages V, Dun EA, Pillot JP, Letisse F, Matusova R, Danoun S, Portais JC et al. 2008. Strigolactone inhibition of shoot branching. Nature 455:189-194.
Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178-1186.
Gottwald JR, Krysan PJ, Young JC, Evert RF, Sussman MR. 2000. Genetic evidence for the in planta role of phloem-specific plasma membrane sucrose transporters. Proc Natl Acad Sci U S A 97:13979-13984.
Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644-U130.
Gressel J, Joel DM. 2013. Weedy Orobanchaceae: The Problem. In. Parasitic Orobanchaceae - Parasitic Mechanisms and Control Strategies. Springer Heidelberg New York Dordrecht London: Springer. p. 309-312.

Griffitts AA, Cramer CL, Westwood JH. 2004. Host gene expression in response to Egyptian broomrape (*Orobanche aegyptiaca*). Weed Sci 52:697-703.
Gurney AL, Slate J, Press MC, Scholes JD. 2006. A novel form of resistance in rice to the angiosperm parasite Striga hermonthica. New Phytol 169:199-208.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494-1512.
Haegeman A, Vanholme B, Jacob J, Vandekerckhove TT, Claeys M, Borgonie G, Gheysen G. 2009. An endosymbiotic bacterium in a plant-parasitic nematode: member of a new Wolbachia supergroup. Int J Parasitol 39:1045-1054.
Hake S, Smith HM, Holtan H, Magnani E, Mele G, Ramirez J. 2004. The role of knox genes in plant development. Annu Rev Cell Dev Biol 20:125-151.
Hao W, Palmer JD. 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. Proc Natl Acad Sci U S A 106:16728-16733.
Hao W, Richardson AO, Zheng Y, Palmer JD. 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proc Natl Acad Sci U S A 107:21576-21581.
Harloff HJ, Wegmann D. 1993. Evidence for a mannitol cycle in *Orobanche ramosa* and *Orobanche crenata*. J Plant Physiol 141:513-520.
Harpur BA, Kent CF, Molodtsova D, Lebon JM, Alqarni AS, Owayss AA, Zayed A. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. Proc Natl Acad Sci U S A 111:2614-2619.
Hauck C, Muller S, Schildknecht H. 1992. A germination stimulant for parasitic flowering plants from Sorghum bicolor, a genuine host plant. J Plant Physiol 139:474-478.
He XL, Zhang JZ. 2005. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169:1157-1164.
Heide-Jørgensen HS. 2013a. The haustorium. In: Joel DM, Gressel J, Musselman LJ, editors. Parasitic Orobanchaceae - Parasitic Mechanisms and Control Strategies. Springer Heidelberg New York Dordrecht London: Springer. p. 4.
Heide-Jørgensen HS. 2013b. Introduction: The parasitic syndrome in higher plants. In: Joel DM, Gressel J, Musselman LJ, editors. Parasitic Orobanchaceae - Parasitic Mechanisms and Control Strategies. Springer Heidelberg New York Dordrecht London: Springer. p. 1.
Heide-Jørgensen HS, Kuijt J. 1995. The haustorium of the root parasite *Triphysaria* (Scrophulariaceae), with special reference to xylem bridge ultrastructure. American Journal of Botany 82:782-797.
Heinemann IU, Nakamura A, O'Donoghue P, Eiler D, Soll D. 2012. tRNAHis-guanylyltransferase establishes tRNAHis identity. Nucleic Acids Res 40:333-344.
Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, Wessler SR, Schmutz J, Willis JH, Rokhsar DS. 2013. Fine-scale variation in meiotic recombination in Mimulus inferred from population shotgun sequencing. Proceedings of the National Academy of Sciences of the United States of America.
Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP, Davies G. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proceedings of the National Academy of Sciences of the United States of America 92:7090-7094.
Higgins TJV. 1984. Synthesis and Regulation of Major Proteins in Seeds. Annu Rev Plant Phys 35:191-221.
Hiraoka Y, Sugimoto Y. 2008. Molecular Responses of Sorghum to Purple Witchweed (*Striga hermonthica*) Parasitism. Weed Sci 56:356-363.

Hiraoka Y, Ueda H, Sugimoto Y. 2009. Molecular responses of Lotus japonicus to parasitism by the compatible species Orobanche aegyptiaca and the incompatible species Striga hermonthica. J Exp Bot 60:641-650.
Honaas LA. 2013. Tissue specific *de novo* transcriptomics in the parasitic Orobanchaceae. [Dissertation]: The Pennsylvania State University.
Honaas LA, Wafula EK, Yang Z, Der JP, Wickett NJ, Altman NS, Taylor CG, Yoder JI, Timko MP, Westwood JH et al. 2013. Functional genomics of a generalist parasitic plant: laser microdissection of host-parasite interface reveals host-specific patterns of parasite gene expression. BMC Plant Biol 13:9.
Honma T, Goto K. 2001. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature 409:525-529.
Houzet L, Battini JL, Bernard E, Thibert V, Mougel M. 2003. A new retroelement constituted by a natural alternatively spliced RNA of murine replication-competent retroviruses. The EMBO journal 22:4866-4875.
Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44-57.
Huang J. 2013. Horizontal gene transfer in eukaryotes: the weak-link model. Bioessays 35:868-875.
Huang JY, Zhao XB, Cheng K, Jiang YH, Ouyang YD, Xu CG, Li XH, Xiao JH, Zhang QF. 2013. OsAP65, a rice aspartic protease, is essential for male fertility and plays a role in pollen germination and pollen tube growth. J Exp Bot 64:3351-3360.
Huerta-Cepas J, Dopazo J, Gabaldon T. 2010. ETE: a python Environment for Tree Exploration. BMC Bioinformatics 11:24.
Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11:97-108.
Intrieri MC, Buiatti M. 2001. The horizontal transfer of Agrobacterium rhizogenes genes and the evolution of the genus Nicotiana. Mol Phylogenet Evol 20:100-110.
Irving LJ, Cameron DD. 2009. You are what you eat: interactions between root parasitic plants and their hosts. Adv Bot Res 50:87-138.
Iseli C, Jongeneel CV, Bucher P. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol:138-148.
Ito S, Kitahata N, Umehara M, Hanada A, Kato A, Ueno K, Mashiguchi K, Kyozuka J, Yoneyama K, Yamaguchi S et al. 2010. A new lead chemical for strigolactone biosynthesis inhibitors. Plant Cell Physiol 51:1143-1150.
Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463-467.
Jain R, Rivera MC, Moore JE, Lake JA. 2003. Horizontal gene transfer accelerates genome innovation and evolution. Mol Biol Evol 20:1598-1602.
Jamil M, Charnikhova T, Cardoso C, Jamil T, Ueno K, Verstappen F, Asami T, Bouwmeester HJ. 2011. Quantification of the relationship between strigolactones and *Striga hermonthica* infection in rice under varying levels of nitrogen and phosphorus. Weed Res 51:373-385.
Jamil M, Charnikhova T, Verstappen F, Bouwmeester H. 2010. Carotenoid inhibitors reduce strigolactone production and Striga hermonthica infection in rice. Arch Biochem Biophys 504:123-131.
Jamison DS, Yoder JI. 2001. Heritable variation in quinone-induced haustorium development in the parasitic plant Triphysaria. Plant physiology 125:1870-1879.

Jiang F, Jeschke WD, Hartung W. 2004. Abscisic acid (ABA) flows from Hordeum vulgare to the hemiparasite Rhinanthus minor and the influence of infection on host and parasite abscisic acid relations. J Exp Bot 55:2323-2329.
Jiang F, Jeschke WD, Hartung W. 2003. Water flows in the parasitic association Rhinanthus minor/Hordeum vulgare. J Exp Bot 54:1985-1993.
Jiao Y, Leebens-Mack J, Ayyampalayam S, Bowers JE, McKain MR, McNeal J, Rolf M, Ruzicka DR, Wafula E, Wickett NJ et al. 2012. A genome triplication associated with early diversification of the core eudicots. Genome biology 13:R3.
Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al. 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473:97-100.
Joel DM. 2013. Functional structure of the mature haustorium. In: Joel DM, Gressel J, Musselman LJ, editors. Parasitic Orobanchaceae - parasitic mechanisms and control strategies. Springer Heidelberg New York Dordrecht London: Springer. p. 54-55.
Joel DM L-GD, Goldman-Guez T, Portnoy VH editor. Proceedings of the 4th international workshop on *Orobanche*. Current problems in *Orobanche* research; 1998 Albena.
Joel DM, Losner-Goshen D. 1994a. The attachment organ of the parasitic angiosperms *Orobanche cumana* and *O. aegyptiaca* and its development. Can J Bot 72:564-574.
Joel DM, Losner-Goshen D editors. Proceedings of the third international workshop on Orobanche and related Striga research. Biology and management of *Orobanche*; 1994b Amsterdam.
Joel DM, Losner-Goshen D, Goldman-Guez T, Portnoy VH editors. Proceedings of the 4th international workshop on *Orobanche*. Current problems in Orobanche research; 1998 Institute for Wheat and Sunflower Dobroudja, Albena.
Jones JD, Dangl JL. 2006. The plant immune system. Nature 444:323-329.
Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. Genome Res 20:1313-1326.
Kall L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic acids research 35:W429-432.
Katoh K, Standley DM. 2013a. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30:772-780.
Katoh K, Standley DM. 2013b. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-780.
Keeling PJ. 2009. Functional and ecological impacts of horizontal gene transfer in eukaryotes. Curr Opin Genet Dev 19:613-619.
Keeling PJ, Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9:605-618.
Kennan RM, Wong W, Dhungyel OP, Han XY, Wong D, Parker D, Rosado CJ, Law RHP, McGowan S, Reeve SB et al. 2010. The subtilisin-like protease AprV2 is required for virulence and uses a novel disulphide-tethered exosite to bind substrates. Plos Pathog 6:e1001210.
Keyes WJ, O'Malley RC, Kim D, Lynn DG. 2000. Signaling Organogenesis in Parasitic Angiosperms: Xenognosin Generation, Perception, and Response. J Plant Growth Regul 19:217-231.
Kim G, LeBlanc ML, Wafula E, dePamphilis CW, Westwood JH. 2014. Genomic-scale exchange of mRNA between a parasitic plant and its hosts. Science 345:808-811.
Kim MJ, Ciani S, Schachtman DP. 2010. A peroxidase contributes to ROS production during *Arabidopsis* root response to potassium deficiency. Molecular plant 3:420-427.

Klotz MG, Loewen PC. 2003. The molecular evolution of catalatic hydroperoxidases: evidence for multiple lateral transfer of genes between prokaryota and from bacteria into eukaryota. Mol Biol Evol 20:1098-1112.
Kondo Y, Tadokoro E, Matsuura M, Iwasaki K, Sugimoto Y, Miyake H, Takikawa H, Sasaki M. 2007. Synthesis and seed germination stimulating activity of some imino analogs of strigolactones. Biosci Biotechnol Biochem 71:2781-2786.
Koonin EV, Wolf YI. 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res 36:6688-6719.
Koski B, Golding B. 2001. The closest BLAST hit is often not the nearest neighbor. J Mol Evol 52:540-542.
Kozik A, Matvienko M, Kozik I, Vvan Leeuwen H, Van DDeynze A, Michelmore RM editors. Plant and Animal Genome Conference. 2008 San Diego, CA, USA.
Kramer EM, Dorit RL, Irish VF. 1998. Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages. Genetics 149:765-783.
Krichevsky A, Kozlovsky SV, Tian GW, Chen MH, Zaltsman A, Citovsky V. 2007. How pollen tubes grow. Dev Biol 303:405-420.
Kuijt J. 1969. The biology of parasitic flowering plants. Berkeley, CA: University of California Press.
Kuijt J. 1977. Haustoria of phanerogamic parasites. Annu Rev Phytopathol 15:91-118.
Lambowitz AM, Belfort M. 1993. Introns as mobile genetic elements. Annu Rev Biochem 62:587-622.
Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M et al. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40:D1202-1210.
Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.
Lane JA, Bailey JA, Butler RC, Terry PJ. 1993. (1993) Resistance of cowpea [*Vigna unguiculata* (L.) Walp.] to *Striga gesnerioides* (Willd.) Vatke, a parasitic angiosperm. New Phytol 125:405-412.
Lawrence JG, Retchless AC. 2009. The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. Methods Mol Biol 532:29-53.
Letousey P, De Zélicourt A, Vieira Dos Santos C, Thoiron S, Monteau F, Simier P, Thalouarn P, Delavault P. 2007. Molecular analysis of resistance mechanisms to *Orobanche cumana* in sunflower. Plant Pathol:536-546.
Lev-Yadun S. 2001. Intrusive growth – the plant analog of dendrite and axon growth in animals. New Phytol 150:508-512.
Li FW, Villarreal JC, Kelly S, Rothfels CJ, Melkonian M, Frangedakis E, Ruhsam M, Sigel EM, Der JP, Pittermann J et al. 2014a. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. Proc Natl Acad Sci U S A 111:6672-6677.
Li FW, Villarreal JC, Kelly S, Rothfels CJ, Melkonian M, Frangedakis E, Ruhsam M, Sigel EM, Der JP, Pittermann J et al. 2014b. Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. Proceedings of the National Academy of Sciences of the United States of America.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Li J, Timko MP. 2009. Gene-for-gene resistance in Striga-cowpea associations. Science 325:1094.
Li L, Stoeckert CJ, Jr., Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178-2189.
Li L, Yu XX, Guo CC, Duan XS, Shan HY, Zhang R, Xu GX, Kong HZ. 2015. Interactions among proteins of floral MADS-box genes in *Nuphar pumila* (Nymphaeaceae) and the most recent common ancestor of extant angiosperms help understand the underlying mechanisms of the origin of the flower. Journal of Systematics and Evolution 53:285-296.
Li X, Zhang TC, Qiao Q, Ren Z, Zhao J, Yonezawa T, Hasegawa M, Crabbe MJ, Li J, Zhong Y. 2013. Complete chloroplast genome sequence of holoparasite Cistanche deserticola (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host Haloxylon ammodendron (Chenopodiaceae). PLoS One 8:e58747.
Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, Barker MS. 2015. Early genome duplications in conifers and other seed plants. Sci Adv 1:e1501084.
Liberles DA, Kolesov G, Dittmar K. 2010. Understanding gene duplication through biochemistry and population genetics. In: Dittmar K, Liberles D, editors. Evolution after gene duplication. Hoboken, New Jersey, USA: John Wiley & Sons. p. 10-11.
Liu C, Zhang J, Zhang N, Shan H, Su K, Zhang J, Meng Z, Kong H, Chen Z. 2010. Interactions among proteins of floral MADS-box genes in basal eudicots: implications for evolution of the regulatory network for flower development. Mol Biol Evol 27:1598-1611.
Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. 2012. Comparison of next-generation sequencing systems. J Biomed Biotechnol 2012:251364.
Liu Q, Zhang Y, Matusova R, Charnikhova T, Amini M, Jamil M, Fernandez-Aparicio M, Huang K, Timko MP, Westwood JH et al. 2014. Striga hermonthica MAX2 restores branching but not the Very Low Fluence Response in the Arabidopsis thaliana max2 mutant. New Phytol 202:531-541.
Liu X, Zhang H, Zhao Y, Feng Z, Li Q, Yang HQ, Luan S, Li J, He ZH. 2013. Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. Proc Natl Acad Sci U S A 110:15485-15490.
Llorente F, Lopez-Cobollo RM, Catala R, Martinez-Zapater JM, Salinas J. 2002. A novel cold-inducible gene from *Arabidopsis*, RCI3, encodes a peroxidase that constitutes a component for stress tolerance. The Plant journal : for cell and molecular biology 32:13-24.
Lopez-Raez JA, Charnikhova T, Gomez-Roldan V, Matusova R, Kohlen W, De Vos R, Verstappen F, Puech-Pages V, Becard G, Mulder P et al. 2008. Tomato strigolactones are derived from carotenoids and their biosynthesis is promoted by phosphate starvation. New Phytol 178:863-874.
Losner-Goshen D, Portnoy VH, Mayer AM, Joel DM. 1998. Pectolytic activity by the haustorium of the parasitic plant *Orobanche* L. (Orobanchaceae) in host roots. Ann Bot-London 81:319-326.
Ludeman DA, Farrar N, Riesgo A, Paps J, Leys SP. 2014. Evolutionary origins of sensation in metazoans: functional evidence for a new sensory organ in sponges. BMC Evol Biol 14:3.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155.
Maloney AP, VanEtten HD. 1994. A gene from the fungal plant pathogen Nectria haematococca that encodes the phytoalexin-detoxifying enzyme pisatin demethylase defines a new cytochrome P450 family. Mol Gen Genet 243:506-514.
Marin-Rodriguez MC, Orchard J, Seymour GB. 2002. Pectate lyases, cell wall degradation and fruit softening. J Exp Bot 53:2115-2119.

Matasci N, Hung LH, 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.
Matusova R, Rani K, Verstappen FW, Franssen MC, Beale MH, Bouwmeester HJ. 2005. The strigolactone germination stimulants of the plant-parasitic Striga and Orobanche spp. are derived from the carotenoid pathway. Plant Physiol 139:920-934.
Matvienko M, Torres MJ, Yoder JI. 2001. Transcriptional responses in the hemiparasitic plant *Triphysaria versicolor* to host plant signals. Plant Physiol 127:272-282.
Mayer AM. 2006. Pathogenesis by fungi and by parasitic plants: similarities and differences. Phytoparasitica 34:3-16.
McDowall J, Hunter S. 2011. InterPro protein classification. Bioinformatics for Comparative Proteomics 694:37-47.
McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic acids research 32:W20-W25.
McGowan C, Fulthorpe R, Wright A, Tiedje JM. 1998. Evidence for interspecies gene transfer in the evolution of 2,4-dichlorophenoxyacetic acid degraders. Appl Environ Microbiol 64:4089-4092.
McNeal JR, Bennett JR, Wolfe AD, Mathews S. 2013a. Phylogeny and origins of holoparasitism in Orobanchaceae. American Journal of Botany 100:971-983.
McNeal JR, Bennett JR, Wolfe AD, Mathews S. 2013b. Phylogeny and origins of holoparasitism in Orobanchaceae. Am J Bot 100:971-983.
Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541-548.
Moczek AP. 2005. The evolution and development of novel traits, or how beetles got their horns. BioSci 55:937-951.
Mohamed A, Ellicott A, Housley TL, Ejeta G. 2003. Hypersensitive Response to *Striga* Infection in *Sorghum*. Crop Sci 43:1320-1324.
Molinero-Ruiz ML, Melero-Vera JM, Garcia-Ruiz R, Dominguez J. 2006. Pathogenic diversity within field populations of Orobanche cumana and different reactions on sunflower genotypes. Weed Res 46:462-469.
Monteiro A, Podlaha O. 2009. Wings, horns, and butterfly eyespots: how do complex traits evolve? PLoS Biol 7:e37.
Mor A, Mayer AM, Levine A. 2008. Possible peroxidase functions in the interaction between the parasitic plant, *Orobanche aegyptiaca*, and its host, *Arabidopsis thaliana*. Weed Biol Manag 8:1-10.
Morris SE, Turnbull CG, Murfet IC, Beveridge CA. 2001. Mutational analysis of branching in pea. Evidence that Rms1 and Rms5 regulate the same novel signal. Plant Physiol 126:1205-1213.
Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621-628.
Mower JP, Stefanovic S, Hao W, Gummow JS, Jain K, Ahmed D, Palmer JD. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biol 8:150.
Mower JP, Stefanovic S, Young GJ, Palmer JD. 2004. Plant genetics: gene transfer from parasitic to host plants. Nature 432:165-166.
Musselman LJ. 1980. The Biology of Striga, Orobanche, and other Root-Parasitic Weeds. Annual Review of Phytopathology 18:463-489.
Musselman LJ, Dickison WC. 1975. The structure and development of the haustorium in parasitic Scrophulariaceae. Botanical Journal of the Linnean Society 70:183-212.

Musselman LJ, Hepper FN. 1986. The Witchweeds (*Striga*, *Scrophulariaceae*) of the Sudan Republic. Kew Bulletin 41:205-221.
Mwakaboko AS, Zwanenburg B. 2011. Strigolactone analogs derived from ketones using a working model for germination stimulants as a blueprint. Plant Cell Physiol 52:699-715.
Naumann J, Salomo K, Der JP, Wafula EK, Bolin JF, Maass E, Frenzke L, Samain MS, Neinhuis C, dePamphilis CW et al. 2013. Single-copy nuclear genes place haustorial Hydnoraceae within piperales and reveal a cretaceous origin of multiple parasitic angiosperm lineages. PLoS One 8:e79204.
Nei M. 2013a. Darwin's theory of evolution. In: Nei M, editor. Mutation-driven evolution. United Kingdom: Oxford University Press.
Nei M. 2013b. Evolution of eyes and photoreceptors. In: Nei M, editor. Mutation-Driven Evolution. United Kingdom: Oxford University Press. p. 156-157.
Nei M. 1969. Gene duplication and nucleotide substitution in evolution. Nature 221:40-42.
Nelson DC, Riseborough JA, Flematti GR, Stevens J, Ghisalberti EL, Dixon KW, Smith SM. 2009. Karrikins discovered in smoke trigger Arabidopsis seed germination by a mechanism requiring gibberellic acid synthesis and light. Plant Physiol 149:863-873.
Nelson DC, Scaffidi A, Dun EA, Waters MT, Flematti GR, Dixon KW, Beveridge CA, Ghisalberti EL, Smith SM. 2011a. F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in *Arabidopsis thaliana*. P Natl Acad Sci USA 108:8897-8902.
Nelson DC, Scaffidi A, Dun EA, Waters MT, Flematti GR, Dixon KW, Beveridge CA, Ghisalberti EL, Smith SM. 2011b. F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in Arabidopsis thaliana. Proceedings of the National Academy of Sciences 108:8897-8902.
Neumann U, Vian B, Weber HC, Salle G. 1999. Interface between haustoria of parasitic members of the Scrophulariaceae and their hosts: a histochemical and immunocytochemical approach. Protoplasma 207:84-97.
Neumann U, Vian B, Weber HC, Sallé G. 1999. Interface between haustoria of parasitic members of the Scrophulariaceae and their hosts: a histochemical and immunocytochemical approach. Protoplasma 207:84-97.
Ngo QA, Albrecht H, Tsuchimatsu T, Grossniklaus U. 2013. The differentially regulated genes TvQR1 and TvPirin of the parasitic plant Triphysaria exhibit distinctive natural allelic diversity. BMC Plant Biol 13:28.
Nickrent DL, Blarer A, Qiu YL, Vidal-Russell R, Anderson FE. 2004. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. BMC Evol Biol 4:40.
Nickrent DL, Musselman LJ, Riopel JL, Eplee RE. 1979. Haustorial initiation and non-host penetration in Witchweed *(Striga asiatica*). Ann Bot-London 43:233-236.
Nogales J, Munoz S, Olivares J, Sanjuan J. 2009. Genetic characterization of oligopeptide uptake systems in Sinorhizobium meliloti. Fems Microbiol Lett 293:177-187.
O'Malley RC, Lynn DG. 2000. Expansin message regulation in parasitic angiosperms: marking time in development. Plant Cell 12:1455-1465.
Ochman H, Lawrence JG, Groisman EA. 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405:299-304.
Ohno S. 1970. Evolution by Gene Duplication. New York: Springer-Verlag.
Olmstead RG, dePamphilis CW, Wolfe AD, Young ND, Elisons WJ, Reeves PA. 2001. Disintegration of the Scrophulariaceae. American Journal of Botany 88:348-361.
Oo MM, Bae HK, Nguyen TD, Moon S, Oh SA, Kim JH, Soh MS, Song JT, Jung KH, Park SK. 2014. Evaluation of rice promoters conferring pollen-specific expression in a heterologous system, Arabidopsis. Plant Reprod 27:47-58.

Ostrom JH. 1979. Bird flight: how did it begin? Am Sci 67:46-56.
Otto SP, Whitton J. 2000. Polyploid incidence and evolution. Annu Rev Genet 34:401-437.
Pageau K, Simier P, Le Bizec B, Robins RJ, Fer A. 2003. Characterization of nitrogen relationships between *Sorghum bicolor* and the root-hemiparasitic angiosperm *Striga hermonthica* (Del.) Benth. using (KNO3)-N-15 as isotopic tracer. Journal of Experimental Botany 54:789-799.
Panganiban G, Nagy L, Carroll SB. 1994. The role of the Distal-less gene in the development and evolution of insect limbs. Curr Biol 4:671-675.
Park J-M, Manen J-F, Schneeweiss GM. 2007a. Horizontal gene transfer of a plastid gene in the non-photosynthetic flowering plants Orobanche and Phelipanche(Orobanchaceae). Mol Phylogenet Evol 43:974-985.
Park JM, Manen JF, Schneeweiss GM. 2007b. Horizontal gene transfer of a plastid gene in the non-photosynthetic flowering plants Orobanche and Phelipanche (Orobanchaceae). Mol Phylogenet Evol 43:974-985.
Parker C. 2009. Observations on the current status of Orobanche and Striga problems worldwide. Pest Manag Sci 65:453-459.
Paterson AH, Bowers JE, Chapman BA. 2004. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci U S A 101:9903-9908.
Pe´rez-de-Luque A. 2013. Haustorium invasion into host tissues. In: Joel DM, Gressel J, Musselman LJ, editors. Parasitic Orobanchaceae - Parasitic Mechanisms and Control Strategies. Springer Heidelberg New York Dordrecht London: Springer p. 82-83.
Pearce G, Yamaguchi Y, Barona G, Ryan CA. 2010. A subtilisin-like protein from soybean contains an embedded, cryptic signal that activates defense-related genes. Proceedings of the National Academy of Sciences of the United States of America 107:14921-14925.
Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, Stutz AM, Stedman W, Anantharaman T, Hastie A et al. 2015. Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods 12:780-786.
Peng J, Harberd NP. 2002. The role of GA-mediated signalling in the control of seed germination. Curr Opin Plant Biol 5:376-381.
Peng Y, Yang Z, Zhang H, Cui C, Qi X, Luo X, Tao X, Wu T, Ouzhuluobu, Basang et al. 2011. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol 28:1075-1081.
Perez-de-Luque A, Gonzalez-Verdejo CI, Lozano MD, Dita MA, Cubero JI, Gonzalez-Melendi P, Risueno MC, Rubiales D. 2006. Protein cross-linking, peroxidase and beta-1,3-endoglucanase involved in resistance of pea against Orobanche crenata. J Exp Bot 57:1461-1469.
Petschow D, Wurdinger I, Baumann R, Duhm J, Braunitzer G, Bauer C. 1977. Causes of high blood O2 affinity of animals living at high altitude. J Appl Physiol Respir Environ Exerc Physiol 42:139-143.
Piazza P, Jasinski S, Tsiantis M. 2005. Evolution of leaf developmental mechanisms. New Phytol 167:693-710.
Pielach A, Leroux O, Domozych DS, Knox JP, Popper ZA. 2014. Arabinogalactan protein-rich cell walls, paramural deposits and ergastic globules define the hyaline bodies of rhinanthoid Orobanchaceae haustoria. Ann Bot 114:1359-1373.
Press MC. 1995. Carbon and nitrogen relations. In: Press MC, Graves JD, editors. Parasitic plants. London: Chapman & Hall. p. 103-124.
Press MC, Graves JD editors. Progress in Orobanche research. 1991 Tübingen
Press MC, Shah N, Tuohy JM, Stewart GR. 1987. Carbon isotope ratios demonstrate carbon flux from c(4) host to c(3) parasite. Plant Physiol 85:1143-1145.

Project AG. 2013. The Amborella genome and the evolution of flowering plants. Science 342:1241089.
Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.
Racine JS. 2012. RStudio: A Platform-Independent IDE for R and Sweave. J Appl Economet 27:167-172.
Ranjan A, Ichihashi Y, Farhi M, Zumstein K, Townsley B, David-Schwartz R, Sinha NR. 2014. *De novo* assembly and characterization of the transcriptome of the parasitic weed *Cuscuta pentagona* identifies genes associated with plant parasitism. Plant Physiol.
Rashkova S, Karam SE, Kellum R, Pardue ML. 2002. Gag proteins of the two Drosophila telomeric retrotransposons are targeted to chromosome ends. The Journal of cell biology 159:397-402.
Refulio-Rodriguez NF, Olmstead RG. 2014. Phylogeny of Lamiidae. Am J Bot 101:287-299.
Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. 2006. GenePattern 2.0. Nature genetics 38:500-501.
Reiss GC, Bailey JA. 1998. *Striga gesnerioides* parasitising cowpea: Development of infection structures and mechanisms of penetration. Ann Bot-London 81:431-440.
Rensing SA, Ick J, Fawcett JA, Lang D, Zimmer A, Van de Peer Y, Reski R. 2007. An ancient genome duplication contributed to the abundance of metabolic genes in the moss Physcomitrella patens. BMC Evol Biol 7:130.
Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics 13:278-289.
Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science 342:1468-1473.
Richardson AO, Palmer JD. 2007. Horizontal gene transfer in plants. J Exp Bot 58:1-9.
Riopel JL, Baird WV. 1987. Morphogenesis of the early development of primary haustoria in Striga asiatica. In: Musselman LJ, editor. Parasitic weeds in agriculture. Boca Raton, FL: CRC press. p. 107-125.
Rivera AS, Pankey MS, Plachetzki DC, Villacorta C, Syme AE, Serb JM, Omilian AR, Oakley TH. 2010. Gene duplication and the origins of morphological complexity in pancrustacean eyes, a genomic approach. BMC evolutionary biology 10:123.
Rodelsperger C, Sommer RJ. 2011. Computational archaeology of the Pristionchus pacificus genome reveals evidence of horizontal gene transfers from insects. BMC Evol Biol 11:239.
Ronald PC, Beutler B. 2010. Plant and animal sensors of conserved microbial signatures. Science 330:1061-1064.
Roulin A, Piegu B, Fortune PM, Sabot F, D'Hont A, Manicacci D, Panaud O. 2009. Whole genome surveys of rice, maize and sorghum reveal multiple horizontal transfers of the LTR-retrotransposon Route66 in Poaceae. BMC Evol Biol 9:58.
Roulin A, Piegu B, Wing RA, Panaud O. 2008. Evidence of multiple horizontal transfers of the long terminal repeat retrotransposon RIRE1 within the genus Oryza. Plant J 53:950-959.
Rumsey FJ, Jury SL. 1991. An account of *Orobanche* L. in Britain and Ireland. Watsonia 18:257-295.
Salvini-Plawen L ME. 1961. On the evolution of photoreceptors and eyes. Evolutionary Biology 10:63-207.
Sampedro J, Cosgrove DJ. 2005. The expansin superfamily. Genome biology 6:242.
Sanchez-Puerta MV, Abbona CC, Zhuo S, Tepe EJ, Bohs L, Olmstead RG, Palmer JD. 2011. Multiple recent horizontal transfers of the cox1 intron in Solanaceae and extended co-conversion of flanking exons. BMC Evol Biol 11:277.

Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. 2008. Frequent, phylogenetically local horizontal transfer of the cox1 group I Intron in flowering plant mitochondria. Mol Biol Evol 25:1762-1777.
Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32:347-355.
Schaack S, Gilbert C, Feschotte C. 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. Trends Ecol Evol 25:537-546.
Schaferhoff B, Fleischmann A, Fischer E, Albach DC, Borsch T, Heubl G, Muller KF. 2010. Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. BMC Evol Biol 10:352.
Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC. 2004. Mining EST databases to resolve evolutionary events in major crop species. Genome / National Research Council Canada = Genome / Conseil national de recherches Canada 47:868-876.
Schneeweiss GM. 2007. Correlated evolution of life history and host range in the nonphotosynthetic parasitic flowering plants Orobanche and Phelipanche (Orobanchaceae). J Evol Biol 20:471-478.
Schneeweiss GM, Colwell A, Park JM, Jang CG, Stuessy TF. 2004. Phylogeny of holoparasitic *Orobanche* (Orobanchaceae) inferred from nuclear ITS sequences. Mol Phylogenet Evol 30:465-478.
Scholes JD, Press MC. 2008. *Striga* infestation of cereal crops - an unsolved problem in resource limited agriculture. Current Opinion in Plant Biology 11:180-186.
Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Current opinion in plant biology 15:147-153.
Sergeant MJ, Li JJ, Fox C, Brookbank N, Rea D, Bugg TD, Thompson AJ. 2009. Selective inhibition of carotenoid cleavage dioxygenases: phenotypic effects on shoot branching. J Biol Chem 284:5257-5264.
Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114-1116.
Shomer-Ilan A. 1992. Enzymes with pectinolytic and cellulolytic activity are excreted by the haustorium of *Orobanche aegyptiaca*. Phytoparasitica 20:343.
Shomer-Ilan A. 1993. Germinating seeds of the root parasite *Orobanche aegyptiaca* Pers. excrete enzymes with carbohydrase activity. Symbiosis 15:61-70.
Shomer-Ilan A. 1999. Proteolytic activity of germinating *Orobanche aegyptiaca* seeds controls the degrading level of its own excreted pectinase and cellulase. Phytoparasitica 27:111.
Singh A, Singh M. 1993a. Cell-wall degrading enzymes in *Orobanche aegyptiaca* and its host *Brassica campestris*. Physiol Plant 89:177-181.
Singh A, Singh M. 1993b. Cell-wall degrading enzymes in *Orobanche aegyptiaca* and its host *Brassica campestris*. Physiol Plantarum 89:177-181.
Slewinski TL, Garg A, Johal GS, Braun DM. 2010. Maize SUT1 functions in phloem loading. Plant Signal Behav 5:687-690.
Smant G, Stokkermans JP, Yan Y, de Boer JM, Baum TJ, Wang X, Hussey RS, Gommers FJ, Hernrissa B, Davis EL et al. 1998. Endogenous cellulases in animals: isolation of beta-1, 4-endoglucanase genes from two species of plant-parasitic cyst nematodes. Proc Natl Acad Sci U S A 95:4906-4911.
Smith S, Stewart GR. 1990. Effect of potassium levels on the stomatal behavior of the hemi-parasite *Striga hermonthica*. Plant Physiol 94:1472-1476.
Snowden KC, Simkin AJ, Janssen BJ, Templeton KR, Loucas HM, Simons JL, Karunairetnam S, Gleave AP, Clark DG, Klee HJ. 2005. The Decreased apical dominance1/Petunia hybrida

CAROTENOID CLEAVAGE DIOXYGENASE8 gene affects branch production and plays a role in leaf senescence, root growth, and flower development. Plant Cell 17:746-759.
Soanes D, Richards TA. 2014. Horizontal gene transfer in eukaryotic plant pathogens. Annu Rev Phytopathol 52:583-614.
Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. Am J Bot 96:336-348.
Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlsward BS et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am J Bot 98:704-730.
Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.
Stephens SG. 1951. Possible significances of duplication in evolution. Advances in genetics 4:247-265.
Stevens M. 2005. The role of eyespots as anti-predator mechanisms, principally demonstrated in the Lepidoptera. Biol Rev Camb Philos Soc 80:573-588.
Sueoka N. 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci U S A 48:582-592.
Sugimoto Y, Ali AM, Yabuta S, Kinoshita H, Inanaga S, Itai A. 2003. Germination strategy of *Striga hermonthica* involves regulation of ethylene biosynthesis. Physiol Plantarum 119:137-145.
Sun H, Tao J, Liu S, Huang S, Chen S, Xie X, Yoneyama K, Zhang Y, Xu G. 2014. Strigolactones are involved in phosphate- and nitrate-deficiency-induced root development and auxin transport in rice. J Exp Bot 65:6735-6746.
Sun L, van Nocker S. 2010. Analysis of promoter activity of members of the PECTATE LYASE-LIKE (PLL) gene family in cell separation in *Arabidopsis*. BMC Plant Biol 10:152.
Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22:1540-1542.
Takken FLW, Albrecht M, Tameling WIL. 2006. Resistance proteins: molecular switches of plant defence. Curr Opin Plant Biol 9:383-390.
Tameling WIL, Joosten MHAJ. 2007. The diverse roles of NB-LRR proteins in plants. Physiol Mol Plant Pathol 71:126-134.
Taniguchi Y, Ono A, Sawatani M, Nanba M, Kohno K, Usui H, Kurimoto M, Matuhasi T. 1995. Cry j I, a major allergen of Japanese cedar pollen, has pectate lyase enzyme activity. Allergy 50:90-93.
Taylor A, Martin J, Seel WE. 1996. Physiology of the parasitic association between maize and witchweed (*Striga hermonthica*): is ABA involved? J Exp Bot 47:1057-1065.
Thorne RF. 2002. How many species of seed plants are there? Taxon 51:511-512.
Thorogood CJ, Hiscock SJ. 2010. Compatibility interactions at the cellular level provide the basis for host specificity in the parasitic plant *Orobanche*. New Phytol 186:572-575.
Timko MP, Gowda BS, Quedrogo J, Ousmane B. 2007. Integrating New Technologies for Striga Control: Towards Ending the Witch-Hunt. Singapore: World Scientific.
Tirosh I, Barkai N. 2007. Comparative analysis indicates regulatory neofunctionalization of yeast duplicates. Genome biology 8:R50.
Toh S, Holbrook-Smith D, Stogios PJ, Onopriyenko O, Lumba S, Tsuchiya Y, Savchenko A, McCourt P. 2015. Structure-function analysis identifies highly sensitive strigolactone receptors in Striga. Science 350:203-207.
Toh S, Kamiya Y, Kawakami N, Nambara E, McCourt P, Tsuchiya Y. 2012. Thermoinhibition uncovers a role for strigolactones in Arabidopsis seed germination. Plant Cell Physiol 53:107-117.

Toleman MA, Bennett PM, Walsh TR. 2006. ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev 70:296-316.
Tomilov AA, Tomilova NB, Abdallah I, Yoder JI. 2005. Localized hormone fluxes and early haustorium development in the hemiparasitic plant Triphysaria versicolor. Plant Physiol 138:1469-1480.
Tomilov AA, Tomilova NB, Wroblewski T, Michelmore R, Yoder JI. 2008. Trans-specific gene silencing between host and parasitic plants. Plant J 56:389-397.
Torres MA, Jones JD, Dangl JL. 2006. Reactive oxygen species signaling in response to pathogens. Plant physiology 141:373-378.
Torres MJ, Tomilov AA, Tomilova N, Reagan RL, Yoder JI. 2005. Pscroph, a parasitic plant EST database enriched for parasite associated transcripts. BMC Plant Biol 5:24.
Torto-Alalibo T, Collmer CW, Lindeberg M, Bird D, Collmer A, Tyler BM. 2009. Common and contrasting themes in host cell-targeted effectors from bacterial, fungal, oomycete and nematode plant symbionts described using the Gene Ontology. BMC Microbiol 9 Suppl 1:S3.
Tsuchiya Y, Yoshimura M, Sato Y, Kuwata K, Toh S, Holbrook-Smith D, Zhang H, McCourt P, Itami K, Kinoshita T et al. 2015. PARASITIC PLANTS. Probing strigolactone receptors in Striga hermonthica with fluorescence. Science 349:864-868.
Tuller T. 2011. Codon bias, tRNA pools and horizontal gene transfer. Mob Genet Elements 1:75-77.
Turner FS. 2014. Assessment of insert sizes and adapter content in fastq data from NexteraXT libraries. Front Genet 5:5.
Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A et al. 2006. The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:1596-1604.
Vanneste K, Van de Peer Y, Maere S. 2013. Inference of genome duplications from age distributions revisited. Mol Biol Evol 30:177-190.
Vaughn JC, Mason MT, Sper-Whitis GL, Kuhlman P, Palmer JD. 1995. Fungal origin by horizontal transfer of a plant mitochondrial group I intron in the chimeric CoxI gene of Peperomia. J Mol Evol 41:563-572.
Vernie T, Moreau S, de Billy F, Plet J, Combier JP, Rogers C, Oldroyd G, Frugier F, Niebel A, Gamas P. 2008. EFD Is an ERF Transcription Factor Involved in the Control of Nodule Number and Differentiation in Medicago truncatula. Plant Cell 20:2696-2713.
Visser JH, Dörr I, Kollmann R. 1984. The “hyaline body” of the root parasite *Alectra* orobanchoides Benth. (Scrophulariaceae), its anatomy, ultrastructure and histochemistry. Protoplasma 121:146-156.
Volpi C, Janni M, Lionetti V, Bellincampi D, Favaron F, D'Ovidio R. 2011. The ectopic expression of a pectin methyl esterase inhibitor increases pectin methyl esterification and limits fungal diseases in wheat. Mol Plant Microbe Interact 24:1012-1019.
Wall PK, Leebens-Mack J, Muller KF, Field D, Altman NS, dePamphilis CW. 2008. PlantTribes: a gene and gene family resource for comparative genomics in plants. Nucleic Acids Res 36:D970-976.
Wang Q, Sun H, Huang JL. 2014. The evolution of land plants: a perspective from horizontal gene transfer. Acta Soc Bot Pol 83:363-368.
Waters MT, Brewer PB, Bussell JD, Smith SM, Beveridge CA. 2012. The Arabidopsis ortholog of rice DWARF27 acts upstream of MAX1 in the control of plant development by strigolactones. Plant Physiol 159:1073-1085.
Weatherbee SD, Nijhout HF, Grunert LW, Halder G, Galant R, Selegue J, Carroll S. 1999. Ultrabithorax function in butterfly wings and the evolution of insect wing patterns. Curr Biol 9:109-115.

Westwood JH. 2013. The physiology of the established parasitte-host association. In: Joel DM, Gressel J, Musselman LJ, editors. Parasitic Orobanchaceae - parasitic mechanisms and control strategies. Springer Heidelberg New York Dordrecht London: Springer. p. 89-90.
Westwood JH, dePamphilis CW, Das M, Fernandez-Aparicio M, Honaas LA, 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 BA, 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 21:2098-2104.
Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A 111:E4859-4868.
Williams JS, Der JP, dePamphilis CW, Kao TH. 2014. Transcriptome analysis reveals the same 17 S-Locus F-box genes in two haplotypes of the self-incompatibility locus of *Petunia inflata*. Plant Cell.
Won H, Renner SS. 2003. Horizontal gene transfer from flowering plants to Gnetum. Proceedings of the National Academy of Sciences of the United States of America 100:10824-10829.
Wrobel RL, Yoder JI. 2001. Differential RNA expression of alpha-expansin gene family members in the parasitic angiosperm Triphysaria versicolor (Scrophulariaceae). Gene 266:85-93.
Wu F, Mueller LA, Crouzillat D, Petiard V, Tanksley SD. 2006. Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (COSII) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. Genetics 174:1407-1420.
Xi Z, Bradley RK, Wurdack KJ, Wong K, Sugumaran M, Bomblies K, Rest JS, Davis CC. 2012a. Horizontal transfer of expressed genes in a parasitic flowering plant. BMC Genomics 13:227.
Xi Z, Bradley RK, Wurdack KJ, Wong KM, Sugumaran M, Bomblies K, Rest JS, Davis CC. 2012b. Horizontal transfer of expressed genes in a parasitic flowering plant. BMC genomics 13:227.
Xi Z, Wang Y, Bradley RK, Sugumaran M, Marx CJ, Rest JS, Davis CC. 2013. Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genet 9:e1003265.
Xie F, Murray JD, Kim J, Heckmann AB, Edwards A, Oldroyd GED, Downie A. 2012. Legume pectate lyase required for root infection by rhizobia. Proceedings of the National Academy of Sciences of the United States of America 109:633-638.
Xie K, Minkenberg B, Yang Y. 2015. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. Proc Natl Acad Sci U S A 112:3570-3575.
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 30:1660-1666.
Xu B, Ohtani M, Yamaguchi M, Toyooka K, Wakazaki M, Sato M, Kubo M, Nakano Y, Sano R, Hiwatashi Y et al. 2014. Contribution of NAC transcription factors to plant adaptation to land. Science 343:1505-1508.
Yan LJ, Levine RL, Sohal RS. 1997. Oxidative damage during aging targets mitochondrial aconitase. Proc Natl Acad Sci U S A 94:11168-11172.
Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, Huang K, Bandaranayake PC, 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 32:767-790.

Yang ZH. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular biology and evolution 24:1586-1591.
Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555-556.
Yoneyama K, Awad AA, Xie X, Yoneyama K, Takeuchi Y. 2010. Strigolactones as germination stimulants for root parasitic plants. Plant Cell Physiol 51:1095-1103.
Yoneyama K, Xie X, Yoneyama K, Takeuchi Y. 2009. Strigolactones: structures and biological activities. Pest Manag Sci 65:467-470.
Yoshida S, Ishida JK, Kamal NM, Ali AM, Namba S, Shirasu K. 2010. A full-length enriched cDNA library and expressed sequence tag analysis of the parasitic weed, *Striga hermonthica*. BMC Plant Biol 10:55.
Yoshida S, Maruyama S, Nozaki H, Shirasu K. 2010. Horizontal gene transfer by the parasitic plant Striga hermonthica. Science 328:1128.
Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H et al. 2011. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480:520-524.
Young ND, Steiner KE, dePamphilis CW. 1999. The evolution of parasitism in Scrophulariaceae/Orobanchaceae: Plastid gene sequences refute an evolutionary transition series. Annals of the Missouri Botanical Garden 86:876-893.
Yue J, Hu X, Sun H, Yang Y, Huang J. 2012. Widespread impact of horizontal gene transfer on plant colonization of land. Nat Commun 3:1152.
Zhang D, Qi J, Yue J, Huang J, Sun T, Li S, Wen JF, Hettenhausen C, Wu J, Wang L et al. 2014a. Root parasitic plant Orobanche aegyptiaca and shoot parasitic plant Cuscuta australis obtained Brassicaceae-specific strictosidine synthase-like genes by horizontal gene transfer. BMC Plant Biol 14:19.
Zhang DL, Qi JF, Yue JP, Huang JL, Sun T, Li SP, Wen JF, Hettenhausen C, Wu JS, Wang L et al. 2014b. Root parasitic plant *Orobanche aegyptiaca* and shoot parasitic plant *Cuscuta australis* obtained Brassicaceae-specific strictosidine synthase-like genes by horizontal gene transfer. Bmc Plant Biology 14.
Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311-1320.
Zhang Y, Fernandez-Aparicio M, Wafula EK, Das M, Jiao Y, Wickett NJ, Honaas LA, Ralph PE, Wojciechowski MF, Timko MP et al. 2013a. Evolution of a horizontally acquired legume gene, albumin 1, in the parasitic plant *Phelipanche aegyptiaca* and related species. BMC Evol Biol 13:48.
Zhang Y, van Dijk AD, Scaffidi A, Flematti GR, Hofmann M, Charnikhova T, Verstappen F, Hepworth J, van der Krol S, Leyser O et al. 2014. Rice cytochrome P450 MAX1 homologs catalyze distinct steps in strigolactone biosynthesis. Nat Chem Biol 10:1028-1033.
Zhang YT, Fernandez-Aparicio M, Wafula EK, Das M, Jiao YN, Wickett NJ, Honaas LA, Ralph PE, Wojciechowski MF, Timko MP et al. 2013b. Evolution of a horizontally acquired legume gene, albumin 1, in the parasitic plant Phelipanche aegyptiaca and related species. BMC evolutionary biology 13.
Zhang ZY. 1988. Taxonomy of the Chinese *Orobanche* and its relationships with related genera. Acta Phytotaxonomica Sinica 26:394-403.
Zhaxybayeva O. 2009. Detection and quantitative assessment of horizontal gene transfer. Methods Mol Biol 532:195-213.

Zhou F, Lin Q, Zhu L, Ren Y, Zhou K, Shabek N, Wu F, Mao H, Dong W, Gan L et al. 2013. D14-SCF(D3)-dependent degradation of D53 regulates strigolactone signalling. Nature 504:406-410.
Zhuang JP, Su J, Li XP, Chen WX. 2006. Cloning and expression analysis of beta-galactosidase gene related to softening of banana (Musa sp.) fruit. Zhi Wu Sheng Li Yu Fen Zi Sheng Wu Xue Xue Bao 32:411-419.
Zouari I, Salvioli A, Chialva M, Novero M, Miozzi L, Tenore GC, Bagnaresi P, Bonfante P. 2014. From root to fruit: RNA-Seq analysis shows that arbuscular mycorrhizal symbiosis may affect tomato fruit metabolism. BMC genomics 15.

VITA

Zhenzhen Yang

403 Life Science Building, State College, PA, 16802 • 814-753-0214 • zzy5028@psu.edu

EDUCATION

- Sixth year Ph.D graduate student, Plant Biology — GPA: 3.85
  Ph.D. Advisor: Dr. Claude dePamphilis
  The Pennsylvania State University, PA — 2010 August - now
- Bachelor of Science, Applied Biotechnology — GPA: 3.67
  Huazhong Agricultural University, China — 2006-2010

SUMMARY OF SKILLS

- **Programming Skill**: skilled in Perl and R, familiar with Python
- **Bioinformatics tools**: BLAST, DESeq, bowtie, hierarchical clustering, K-means clustering, PCA, SOM, PAML, MEGA, RAxML, MUSCLE, Mafft, Geneious, DAVID, agriGO, bowtie, BWA, Trinity, pvclust, FeatureCounts, DupliPHY, ete2 module in python, adobe illustrator

PUBLICATIONS

- **Zhenzhen Yang**, Eric K. Wafula, Loren A. Honaas, Huiting Zhang, Malay Das, Monica Fernández-Aparicio, Kan Huang, Pradeepa C.G. Bandaranayake, Biao Wu, Joshua P. Der, Christopher R. Clarke, Paula Ralph, Lena Landherr, Naomi S. Altman, Michael P. Timko, John I. Yoder, James H. Westwood, and Claude W. dePamphilis (2014) *Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol* doi: 10.1093/molbev/msu343
- **Zhenzhen Yang***, Yeting Zhang*, Eric Wafula, Loren A. Honaas, Paula E. Ralph, Sam Jones, Huiting Zhang, Naomi S. Altman, Michael P. Timko, John I. Yoder, James H. Westwood, Claude W. dePamphilis (2016) You are what you eat: Horizontal gene transfer is more frequent with increased heterotrophy and may contribute to parasite adaptation. *Proc. Natl. Acad. Sci. U.S.A.* (in preparation)
- Satoko Yoshida, Seuungill Kim, Eric K Wafula et al. **Zhenzhen Yang**, et al. 2015. Genome sequence of *Striga asiatica* provides insight into the evolution of plant parasitism. *Nature plant* (revised manuscript under review).
- Loren A. Honaas, Eric K. Wafula, **Zhenzhen Yang**, Joshua P. Der, Norman J. Wickett, Naomi S. Altman, Christopher G. Taylor, John I. Yoder, Michael P. Timko, James H. Westwood, Claude W. dePamphilis (2013*) Functional genomics of a generalist parasitic plant: Laser microdissection of host-parasite interface reveals host-specific patterns of parasite gene expression. BMC Plant Bio.* 2013 Jan 9;13:9. doi: 10.1186/1471-2229-13-9
- Malay Das, Monica Fernández-Aparicio, **Zhenzhen Yang**, Kan Huang, Norman J. Wickett, Shannon Alford, Eric K. Wafula, Claude dePamphilis, Harro Bouwmeester, Michael P. Timko, John I. Yoder, James H. Westwood (2015). *Parasitic plants Striga and Phelipanche dependent upon exogenous strigolactones for germination have retained genes for strigolactone biosynthesis*. *American Journal of Plant Sciences* 6(8): 1151-1166

TEACHING EXPERIENCE

- TA of Bio 110—Basic Concepts and Biodiversity (Fall of 2011, 2012, 2013, and 2014)

CONFERENCES

- PLANT & ANIMAL GENOME XXIV Conference (2016) (oral and poster presentation)
- The 13th World Congress on Parasitic Plants (2015) (oral presentation)
- Plant & Animal Genome XXII Conference (2014) (poster)
- The SMBE 2013 – Society for Molecular Biology and Evolution (2013) (poster)
- The bioinformatics and Genomic Retreat at Penn State (2012 & 2013 & 2014) (poster)
- Maize Genetic Conferences (2011) (poster)

AWARDS AND HONORS

- Shannon scholarship of Plant Biology Program at Penn State (2015)
- Best student & postdoctoral oral presentation award at the 13th World Congress on Parasitic Plants (2015)
- Student Registration Subsidy for Attendance of WCPP13 (2015) & IMEG travel fund at Penn State (2013 & 2014 & 2015)
- Huck Student Travel Fund (2015) & Department of Biology travel fund at Penn State (2013 & 2014 & 2015)
- The Braddock Scholarship at Penn State (2010) & The FEGR Award at Penn State (2010)
- Monsanto Scholarship (2009) & Qifa Scholarship (2008) & The National Encouragement Scholarship (2007-2009)