Massive Exchange of mRNA between a Parasitic Plant and its Hosts

Gunjune Kim

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Plant Pathology, Physiology and Weed Science

James H. Westwood

Brenda S.J. Winkel

John M. McDowell

Boris A. Vinatzer

August 14, 2014

Blacksburg, Virginia

Keywords: plant parasitism, *Cuscuta pentagona*, lespedeza dodder, mRNA trafficking, bioinformatics, host-parasite interaction, *de novo* transcriptome, translatome, ribosome footprinting

Massive Exchange of mRNA between a Parasitic Plant and its Hosts

Gunjune Kim

Abstract

Cuscuta pentagona is an obligate parasitic plant that hinders production of crops throughout the world. Parasitic plants have unique morphological and physiological features, the most prominent being the haustorium, a specialized organ that functions to connect them with their host's vascular system. The *Cuscuta* haustorium is remarkable in that it enables mRNA movement to occur between hosts and parasite, but little is known about the mechanisms regulating cross-species mRNA transfer or its biological significance to the parasite. These questions were addressed with genomics approaches that used high throughput sequencing to assess the presence of host mRNAs in the parasite as well as parasite mRNAs in the host. For the main experiment *Cuscuta* was grown on stems of *Arabidopsis thaliana* and tomato (*Solanum lycopersicon*) hosts because the completely sequenced genomes of these plants facilitates identification of host and parasite transcripts in mixed mRNA samples. Tissues sequenced included the *Cuscuta* stem alone, the region of *Cuscuta*-host attachment, and the host stem adjacent to the attachment site. The sequences generated from each tissue were mapped to host reference genes to distinguish host sequences, and the remaining sequences were used in a *de novo* assembly of a *Cuscuta* transcriptome. This analysis revealed that thousands of different *Arabidopsis* transcripts, representing nearly half of the expressed transcriptome of *Arabidopsis*, were represented in the attached *Cuscuta*. RNA movement was also found to be bidirectional, with a substantial proportion of expressed *Cuscuta* transcripts found in host tissue. The

mechanism underlying the exchange remains unknown, as well as the function of mobile RNAs in either the parasite or host. An approach was developed to assay potential translation of host mRNAs by detecting them in the *Cuscuta* translatome as revealed by sequencing polysomal RNA and ribosome-protected RNA. This work highlights RNA trafficking as a potentially important new form of interaction between hosts and *Cuscuta*.

Acknowledgements

Over the past four years I have received fantasitic training, support, and encouragement from a great mentor and my advisor, Dr. James H. Westwood. He is a great role model and also very passionate supporter for students. I am very thankful to have this blessed meeting with Dr. Westwood in my life. He has been very patient with me and always encouraged me with many aspects of research and career. If I start listing many great things that Dr. Westwood provided me during my Ph.D years, the dissertation pages are not enough to fill out. I would like to thank my dream team committee of Drs. Brenda S.J. Winkel, John M. McDowell, and Boris A. Vinatzer for their support and thorough guidance for my project. They always had sharp eyes to correct me in research and trained me to become a better critical thinker in designing experiments and interpreting data. In addition, I would like to thank my graduate student collaborator, Dr. Megan Leblanc for helping me in many areas with our research. We had great teamwork throughout our Ph.D careers. I would like to thank Dr. Claude dePamphilis and Eric K. Wafula for their great collaboration with bioinformatics support. Also, I would like to thank Kristen Clermont, Hailey Larose, Rachel Warnock, Hope Gruszewski and Christopher Clarke for helping me in many areas of my research. The PPWS department, MPS community, and Latham community have provided me with tremendous support.

Attributions

Chapter 1: Authorship is shared with Dr. Megan LeBlanc and Dr. James Westwood. The article appears courtesy of *Frontiers in Plant Science* and is made available as an electronic reprint with the permission of the journal. The paper can be found at the following URL on the Frontiers website: http://journal.frontiersin.org/Journal/10.3389/fpls.2012.00203/abstract

Systematic or multiple reproduction or distribution to multiple locations via electronic or other means is prohibited and is subject to penalties under law.

Chapter 3: Authorship is shared with Dr. James Westwood, Dr. Megan LeBlanc, Eric Wafula, Dr. Claude dePamphilis. This article appears courtesy of *Science* and is made available as an electronic reprint with the permission of the journal. The paper can be found at the following URL on the Science website: http://www.sciencemag.org/content/345/6198/808.short

Systematic or multiple reproduction or distribution to multiple locations via electronic or other means is prohibited and is subject to penalties under law.

Table of Contents

Abstract ii

Acknowledgements iv

Attributions v

List of Tables ix

List of Figures xi

Chapter 1: RNA trafficking in parasitic plant systems 1

Abstract 2

Introduction 3

Parasitic plant connections: the perfect graft? 4

Host-parasite movement of macromolecules 7

RNA trafficking 13

A model for cell-to-cell RNA transport to parasites 13

Virus movement as a model for mRNA trafficking 15

Long distance RNA transport 16

Parasitic plants as tools for studying RNA trafficking 17

Potential significance of trans-species RNA trafficking 21

Conclusions and future directions 24

Ackowledgements 25

References 26

Chapter 2: *De novo* assembly and characterization of the parasitic plant *Cuscuta pentagona* stem and haustoria transcriptomes 40

Abstract 41

Introduction 42

Materials and Methods 45
Plant material 45
RNA isolation and sequencing 45
Sequence pre-processing and filtering 46
Filteration of parasite mRNA reads and *de novo* assembly of the*Cuscuta* transcriptome 47
Differential expression analysis for *Cuscuta* haustorium induced transcripts 48
GO enrichment analysis for up-regulated *Cuscuta* unigenes in haustorial tissue 48
Results 48
Cuscuta transcriptome *de novo* assembly, optimization test and evaluation of quality 48
Transcript expression abundance Gene ontology (GO) analysis across haustorium and parasite stem tissue 56
Discussion 59
References 64
Chapter 3: Genomic-Scale Exchange of mRNA between a Parasitic Plant and its Hosts ... 80
Abstract 81
Main text 82
Materials and methods 93
Plant material 93
RNA isolation and sequencing 93
Sequence pre-processing and filtering 95
Identification of mRNA movement between hosts and parasite 96
Estimating RNA mobility from read counts 97
RT-PCR validation 98
Analyses of mobile RNA properties 98

Comparison with phloem transcriptomes ... 99
References ... 148
Acknowledgments ... 150
Chapter 4: Summary ... 151
An overview of host-*Cuscuta* RNA trafficking ... 151
References ... 157
Appendix ... 159
Evaluating the fate of host mobile RNA in *Cuscuta* by transcriptome, translatome and ribosome profiling ... 159
Abstract ... 160
Introduction ... 161
Results and discussion ... 166
Experimental strategy to validating translatome and ribosome footprinting protocol ... 166
Materials and Methods ... 175
Plant material ... 175
Preparation of polysomal RNA and fractionation ... 176
Isolation of ribosome protected RNAs ... 177
Library construction of transcriptome, translatome and ribosome profiling. ... 177
References ... 179

List of Tables

Chapter 1

Table 1.1. Macromolecules demonstrated to transfer between hosts and parasitic plants 37

Chapter 2

Table 2.1. Summary statistics of transcriptome assemblies generated from *Cuscuta* haustoria (I) and stem (P) tissues 51

Table 2.2. Statistics of combined read sequence assembly using Trinity 54

Chapter 3

Table 3.1. Numbers of genes/unigenes with mobile transcripts from hosts into *Cuscuta* or *Cuscuta* into hosts 92

Table 3.S1. Summary of read sequence assembly for the Cuscuta transcriptome. 105

Table 3.S2. Summaries of Gene Ontology (GO) over-representation and under-representation of mobile and non-mobile transcripts from *Arabidopsis*, tomato and *Cuscuta* 105

Table 3.S2.1 Gene Ontology enrichment for overrepresentation of mobile transcripts that occur in at least two species 105

Table 3.S2.2 Gene Ontology enrichment for overrepresentation of non-mobile transcripts that occur in at least two species 107

Table 3.S2.3 Gene Ontology enrichment for 1,179 mobile *Arabidopsis* transcripts into *Cuscuta*. 109

Table 3.S2.4 Gene Ontology enrichment for 229 mobile *Solanum lycopersicum* transcripts into *Cuscuta*. 113

Table 3.S2.5 Gene Ontology enrichment for 998 mobile *Cuscuta* unigenes into *Arabidopsis* or *Solanum lycopersicum* hosts 114

Table 3.S2.6 Gene Ontology enrichment for 8,809 non-mobile *Arabidopsis* transcripts 116

Table 3.S2.7 Gene Ontology enrichment for 14,481 non-mobile *Solanum lycopersicum* transcripts. 140

Table 3.S2.8 Gene Ontology enrichment for 16,423 non-mobile *Cuscuta* unigenes from plants grown with *Arabidopsis* and *Solanum lycopersicum* hosts 141

Table 3.S3. Correlation of *Arabidopsis* transcript mobility into *Cuscuta* with reported phloem-associated transcripts from *Arabidopsis* and other species 143

Table 3.S4. Correlation of specific categories of *Arabidopsis*-to-*Cuscuta* transcript mobility with reported phloem-associated transcripts from *Arabidopsis* and other species 145

Table 3.S5. Primers used in amplifying transcripts from *Arabidopsis* and *Cuscuta*. 147

List of Figures

Chapter 1

Figure 1.1. Hypothesized model of RNA trafficking from host to *Cuscuta* via plasmodesmata 39

Chapter 2

Figure 2.1. N50 (A) and total nucleotide length (B) of three assemblies 53

Figure 2.2. Contig length distribution of *Cuscuta* transcriptome generated in Trinity 55

Figure 2.3. Volcano plot illustrating differential *Cuscuta* unigene expression between haustorial tissue (interface) compared to *Cuscuta* stem (parasite) 57

Figure 2.4. GO slim annotation for Biological Process (A) and Molecular function (B) of 522 haustorial up-regulated transcriptome using REVIGO 58

Figure 2.5. Number of *Cuscuta* haustorial associated unigenes orthologous to shoot- or root-specific Arabidopsis transcription factors that are *Cuscuta* transcripts .59

Chapter 3

Figure 3.1. Transcriptome compositions of host and parasite tissues at and near the region of haustorial attachment 89

Figure 3.2. Example of read assemblies of an *Arabidopsis* gene, *Translationally Controlled Tumor Protein* (*AtTCTP*), in host stem (HS), interface (I) and parasite stem (PS) tissues. 89

Figure 3.3. Properties of mobile and non-mobile RNAs 90

Figure 3.4. Scatterplots of transcript abundance (FPKM) in parasite stem vs. host-parasite interface 91

Figure 3.S1. Overview of parasite attached to host plants 100

Figure 3.S2. Pipeline of read filtering, mapping, and assembly to distinguish *Cuscuta* transcripts from host transcripts and other contaminants 101

Figure 3.S3. Confirmation of mobility of selected *Arabidopsis* and *Cuscuta* transcripts 102

Figure 3.S4. Summary of *Cuscuta* contig length distribution for the *Cuscuta* transcriptome. Contigs below 200bp was not included 104

Appendix

Figure A.1. Schematic diagram of the transcriptome, translatome and ribosome footprinting pipeline 168

Figure A.2. Identification of polysome, ribosome and ribosome subunits using EDTA control experiment to dissociate the polysomes from mRNA. 170

Figure A.3. Polysome extraction from three biological replicates of host and *Cuscuta* tissue by sucrose density gradient 171

Figure A.4. Ribosome protected mRNAs were treated with RNase I (nuclease digestion) and UV monitor shows content of polysomes and 80S monosomes 174

Figure A.5. Ribosome-protected mRNA library size selection quality check with bioanalyzer 174

This paper was published in *Frontiers in Plant Science* and is made available as an electronic reprint with the permission of the journal. The paper can be found at the following URL on the Frontiers website:

http://www.frontiersin.org/Plant_Physiology/10.3389/fpls.2012.00203/abstract. Systematic or multiple reproduction or distribution to multiple locations via electronic or other means is prohibited and is subject to penalties under law.

Chapter 1: RNA trafficking in parasitic plant systems

Megan LeBlanc, Gunjune Kim and James H. Westwood*

Department of Plant Pathology, Physiology and Weed Science, Virginia Tech, Blacksburg, VA, USA

Abstract

RNA trafficking in plants contributes to local and long-distance coordination of plant development and response to the environment. However, investigations of mobile RNA identity and function are hindered by the inherent difficulty of tracing a given molecule of RNA from its cell of origin to its destination. Several methods have been used to address this problem, but all are limited to some extent by constraints associated with accurately sampling phloem sap or detecting trafficked RNA. Certain parasitic plant species form symplastic connections to their hosts and thereby provide an additional system for studying RNA trafficking. The haustorial connections of *Cuscuta* and *Phelipanche* species are similar to graft junctions in that they are able to transmit mRNAs, viral RNAs, siRNAs and proteins from the host plants to the parasite. In contrast to other graft systems, these parasites form connections with host species that span a wide phylogenetic range, such that a high degree of nucleotide sequence divergence may exist between host and parasites and allow confident identification of most host RNAs in the parasite system. The ability to identify host RNAs in parasites, and vice versa, will facilitate genomics approaches to understanding RNA trafficking. This review discusses the nature of host parasite connections and the potential significance of host RNAs for the parasite. Additional research on host-parasite interactions is needed to interpret results of RNA trafficking studies, but parasitic plants may provide a fascinating new perspective on RNA trafficking.

Introduction

RNA trafficking in plants has received increasing attention in recent years. It is an important phenomenon in that it suggests that plants operate at the level of supracellular organisms, meaning that mRNA is not restricted to the cell in which it was synthesized, but rather moves from cell to cell and even over long distances through the phloem (Lucas & Lee, 2004a). Under this model mRNAs function in cells distant from their point of origin by carrying information throughout the plant and act as part of a system for coordinating plant development.

However, the study of RNA movement between cells of a single organism presents tremendous challenges. It is difficult to confidently determine the cell of origin of an RNA molecule when the genotypes of source and destination cells are identical. This inability to distinguish mRNA origins hampers research, and while it is currently thought that there are about 1,100 proteins and hundreds of mRNAs that undergo long distance movement (Atkins *et al.*, 2011) the actual scope of RNA trafficking over short and long distances is unknown.

Parasitic plants are relevant to the consideration of RNA trafficking because certain species have the ability to form symplastic unions with host species wherein the connections allow the transfer of RNA. In the case of lespedeza dodder (*Cuscuta pentagona*), the connections permit transfer of mRNA from host to parasite (Roney, J *et al.*, 2007; David-Schwartz R, 2008), and this raises exciting prospects for understanding RNA trafficking in plants. Because *Cuscuta* has a relatively wide host range and can effectively parasitize a number of species from a diverse range of plant families, this parasite can act as a sink for host mobile RNA from many different species. Furthermore, the evolutionary distance between *Cuscuta* and most of its hosts means that the majority of mRNAs synthesized in a host have sequences that are divergent from those of *Cuscuta*, thus simplifying the process of bioinformatically recognizing host mRNA that has

trafficked into the parasite. To the extent that *Cuscuta* connections to hosts approximate normal cell-to-cell connections within plants, *Cuscuta* can serve as an exceptionally wide heterograft to facilitate studies of mobile RNA. This review will examine the nature of host-parasite connections and consider the advantages and disadvantages of using parasites for studies of RNA trafficking in plants.

Parasitic plant connections: the perfect graft?

The connection between parasitic plants and their hosts has been compared to "the perfect graft" (Kuijt, 1983). The analogy of parasitic plant connections to graft unions is appropriate in that both involve fusing together separate plants to forge new cellular connections and vascular continuity. Both grafts and parasite connections establish symplastic connections (Although this is not true of all parasite species, it is accepted for *Cuscuta* and *Phelipanche* spp.), and have the ability to transmit RNA (Westwood *et al.*, 2009; Harada, 2010). However, whereas man-made grafts are the result of joining cut tissues, the parasitic connection involves a highly coordinated biological invasion (Joel, 1994; Lee, 2007). Although parasitism may elicit defense responses from the host (Borsics & Lados, 2002; Griffitts *et al.*, 2004; Swarbrick *et al.*, 2008), compatible reactions display little tissue necrosis and haustorial connections are characterized by close association of live cells from both species.

Another difference between graft unions and parasite connections is the greater breadth of compatibility between parasites and hosts compared to graft compatibilities. Parasites are able to form connections with plant species that are phylogenetically distant from themselves, which stands in contrast to grafting where success is greatest when stock and scion are from the same or closely related species (Mudge *et al.*, 2009). For example, a hetero-graft may consist of a pepper scion on a tomato stock, but both species are members of the Solanaceae family. Parasites in

contrast, commonly connect to host plants that are phylogenetically distant from themselves, with an excellent example being *Striga hermonthica*, a dicotyledonous members of the Orobanchaceae family that attacks grass (Poaceae) hosts. The host range of parasitic plants may also vary substantially, and some species are able to parasitize a wide range of host species while others are limited to just a single genera. Host-specificity is variable among parasite species and certain parasites have adapted to attack relatively broad ranges of crop plants and are economically important agricultural weeds (Parker & Riches, 1993b; Westwood *et al.*, 2010).

Parasitic plants live by tapping into the vascular system of a host plant and withdrawing the necessary water and nutrients needed to provide part or all of the parasite nutritional needs. Parasites connect to their hosts using a specialized structure, the haustorium, which penetrates host tissue and forms a bridge to the vascular system of the host. Haustoria vary substantially in their anatomy and function among different parasitic plant species, and are generally characterized by whether they form connections to the xylem only or to both xylem and phloem (Irving & Cameron, 2009). Xylem-feeding parasites primarily withdraw water and dissolved solutes from the host and are generally hemiparasitic in that they are able to photosynthesize to produce at least part of their carbon needs. Parasites that form both xylem and phloem connections are often holoparasitic, relying on their hosts for all nutritional needs. Among this latter group are the dodders (*Cuscuta* spp.) and broomrapes (*Orobanche* and *Phelipanche* spp.), two genera with relatively well-characterized haustoria. RNA trafficking to parasitic plants has been best characterized in these species, particularly *Cuscuta*, but even here the level of understanding of RNA movement is far from complete.

With respect to *Cuscuta*, the overwhelming physiological data indicates that the parasite absorbs phloem contents from the host. However, the exact mechanism is not clear as no direct phloem

connections have been demonstrated (Vaughn, 2006). Rather, cells of the parasite searching hyphae that encounter host sieve elements appear to grow around the phloem cells of the host (Felsenstein, 1978). These parasite cells differentiate in a manner consistent with development of sieve elements, although they also contain an elaborate network of smooth endoplasmic reticulum (ER) proximal the host cell, a feature of transfer cells (Christensen *et al.*, 2003). For this reason it remains a formal possibility that *Cuscuta* may acquire host resources by apoplastic transfer, although this seems to fall short of explaining the ability of *Cuscuta* to readily absorb macromolecules such as mRNA, proteins and viruses from their hosts. Physiological continuity of host and parasite phloem is sufficient to transfer the symplastic marker carboxyfluorescein within two hours of dye being applied to the host (Birschwilks, M. *et al.*, 2006). This dye, as well as GFP-tagged viral movement protein, moved readily through the phloem of established haustoria, yet was not observed extensively in host parenchyma cells outside the vascular bundle, suggesting that phloem comprises the major connection. The cell wall structure of *Cuscuta* phloic hyphae is extremely loose such that it could permit the passage of larger molecules via an apoplastic mechanism (Vaughn, 2006), but more research will be needed to definitively settle the question of phloem transfer.

In contrast to the scant anatomical evidence for direct phloem connections, *Cuscuta* has well documented plasmodesmata (PD) connections with host cells (Vaughn, 2003; Birschwilks, M. *et al.*, 2006). These occur along the cell walls of searching hyphae of the *Cuscuta* haustorium that traverse the host cortex to reach the host vascular tissue. The searching hyphae grow through host cells in a manner that does not puncture them, but rather forms what amounts to a tunnel through a host cell and results in the formation of new cell wall on either side, creating a chimeric structure composed of host and parasite cell walls (Vaughn, 2003). PD in these walls

may take simple or branched form and are spanned by desmotubules typical of PD (Vaughn, 2003; Birschwilks, M. *et al.*, 2006). It is possible that these PD contribute to the transfer of materials between host and parasite, but the question has never been addressed experimentally.

The symplastic connections between broomrapes and their hosts differ from those of *Cuscuta* in terms of anatomy, but appear to share many of the same physiological functions. In contrast to *Cuscuta*, direct connections between sieve elements of *Orobanche crenata* and those of its host *Vicia narbonensis* have been imaged using electron microscopy (Dorr & Kollmann, 1995). PD between these species have also been documented and are proposed to lead to formation of sieve pores between adjacent seive elements (Dorr & Kollmann, 1995). Anatomical differences between Convolvulaceae and Orobanchaceae parasites are not surprising given their different evolutionary origins of parasitism (Barkman *et al.*, 2007). Broomrapes also contrast with *Cuscutas* in the host organ targeted, as broomrapes attack roots while *Cuscutas* parasitize stems and leaves.

Host-parasite movement of macromolecules

Parasitic plants acquire a wide range of macromolecules from their hosts that are relevant to the current discussion (Table 1). The capacity of *Cuscuta* to acquire mRNAs from hosts was first demonstrated for mRNAs of specific genes known to be mobile in pumpkin phloem (Roney, J *et al.*, 2007). Total RNA was isolated from *Cuscuta* stems near the point of host attachment and used in reverse transcriptase (RT)-PCR and the resulting amplified products were then sequenced to confirm that the mRNA detected in the parasite was identical to that of the host gene. Additional experiments used tomato hosts and hybridized RNA extracted from *Cuscuta* onto a tomato microarray, leading to the identification of 474 putatively mobile mRNAs. Confirmation tests on a subset of these using RT-PCR, along with an additional study (David-Schwartz R,

2008) bring the current total of confirmed trafficked mRNAs into *Cuscuta* to 27 (Westwood, James H *et al.*, 2009). Mobility of mRNA into *Cuscuta* from three different hosts (alfalfa in addition to tomato and pumpkin) provides evidence that the phenomenon is not specific to just one host-parasite interaction.

Host mRNAs trafficked into the parasite cover a range of biological functions that are typical of those reported in other mobile transcriptomes. For example, the list of host mobile mRNAs in *Cuscuta* include transcription factors such as the *CmNACP* and *CmWRKYP* (Ruiz-Medrano, R *et al.*, 1999) and *GIBBERELLIC ACID INSENSITIVE* (GAI) (Haywood, V. *et al.*, 2005; Huang & Yu, 2009a). In addition, *Cuscuta* contains host RNAs associated with protein synthesis such as translation initiation factors and ribosomal proteins, the presence of which in phloem of other species has fueled speculation over the possibility of protein translation in sieve elements (Kragler, 2010a). Finding mRNAs for genes associated with defense responses such as a cathepsin D proteinase inhibitor (Ryan, 1990) are not surprising given that the system is based on a pathogenic attack of one plant on another. One potential paradox is the finding of both large and small subunits of host RuBisCO mRNA in *Cuscuta* (David-Schwartz R, 2008), which is surprising because the small subunit is generally thought to be immobile and has been used as an indicator of non-phloem contamination in studies of phloem-mobile RNAs (Ruiz-Medrano, R. *et al.*, 1999).

The finding of host *rbcS* mRNA in *Cuscuta* supports the hypothesis that *Cuscuta* is accessing contents of parenchyma cells as well as phloem. In an *in situ* RT-PCR experiment, mRNA of the tomato phosphofructokinase gene was shown to be concentrated in host parenchyma as well as adjacent *C. pentagona* parenchyma cells closest to the host (David-Schwartz R, 2008). This, along with detection of transcripts localized to sieve element and companion cell regions in the

parasite, suggests cell-to-cell translocation of the transcript mediated by parenchyma connections. It is possible that transcripts cross multiple layers of parenchyma cells to ultimately reach the parasite phloem, but it is also possible that the *Cuscuta* searching hyphae cells themselves differentiate into phloem or take on the role of assisting phloem cells, which would simplify the pathway to long-distance movement. Host mRNAs do move long distances as they have been found distant from the point of haustorial attachment. RT-PCR with specificity to the host transcript showed that tomato phosphofructokinase mRNA was detected in *Cuscuta* shoots up to 30 cm away from the point of contact with the host, but no target mRNA was detected beyond that point (David-Schwartz R, 2008). This would seem to indicate a limit to mRNA translocation, although it is premature to draw any conclusions based on findings from a single gene. Additional studies are needed that characterize mobility of multiple transcripts, ideally using quantitative techniques that can reveal the dynamics of host to parasite transfer as well as long distance movement and fate of host mRNA in the parasite.

The mobility of host mRNAs into *Cuscuta* suggests that a similar transfer would occur into broomrapes from their hosts. However this has not been demonstrated, and studies with *P. aegyptiaca* growing on pumpkin did not detect mRNA transfer even though conditions were identical to those used to demonstrate mRNA transfer into *Cuscuta* (Roney, J *et al.*, 2007). More recently, data from an EST sequencing project (The Parasitic Plant Genome Project, http://ppgp.huck.psu.edu/) revealed no host mRNAs in the shoots of attached *P. aegyptiaca*, suggesting that host mRNA mobility into this parasite may be limited at the point of haustorial transfer into the parasite or between the tubercle and shoot of the parasite. Further experiments are needed to specifically answer this question.

Other examples of host-parasite exchange of RNA include viruses, viroids and siRNA signals (Table 1). Cuscuta has long been recognized for the ability to accept and transmit viruses with its hosts and in fact has often been used as a vector for transmitting plant viruses between different plants (Bennet, 1944). Fifty-six viruses have been reported to thus move through Cuscuta (Hosford, 1967). More recent work quantified potato virus Y isolate N (PVYn) transmitting through a Cuscuta bridge between two Nicotiana tabacum plants, yet showed little virus accumulation in the Cuscuta bridge itself compared to recipient host plant and suggesting that virus movement occurs in Cuscuta without multiplication in the parasite (Birschwilks, M. et al., 2006). In P. aegyptiaca the uptake of three positive ssRNA viruses Cucumber Mosaic Virus (CMV), Tomato Mosaic Virus (ToMV), Potato Virus Y (PVY) and the ssDNA virus Tomato Yellow Leaf Curl Virus (TYLCV) were shown to move into the parasite from infected hosts and, in the case of CMV, is able to replicate in the parasite (Gal-On et al., 2009).

Viroid transport is useful to understanding long distance RNA trafficking in general (Wang & Ding, 2010), and is interesting to consider with respect to movement into parasitic plants. Potato Spindle Tuber Viroid (PSTVd) was shown to be taken up from a tomato host by *Phelipanche ramosa* and translocated through the tubercle and into stems of floral shoots (Vachev *et al.*, 2010). The viroid in this study appeared to replicate in the parasite but was not mobile in the reverse direction back into the host, suggesting a dominant sinkward flow.

RNA-based silencing signals also move from host into parasite where they are able to influence gene expression. The facultative parasite *Tryphysaria versicolor* was transformed to express the *β*-glucuronidase (GUS) reporter gene and then allowed to parasitize lettuce plants that expressed a fragment of the GUS gene in a hairpin orientation (Tomilov *et al.*, 2008). Parasites attached to siRNA-expressing host plants showed a decrease in GUS staining in the root tissue near the point

of host attachment, indicating transmission of the silencing signal. However, the silencing signal did not move into tissues distant from the site of attachment, suggesting that movement was somehow restricted. The signal was able to move through a *T. versicolor* bridge and into a second host plant expressing GUS. As with *Cuscuta* bridges that transmit viruses from one plant to another, it seems that parasites are able to transfer silencing signals. A similar approach was taken to silence mannose 6-phosphate reductase (M6PR) in *P. aegyptiaca* using constructs expressed in a tomato host (Huang & Yu, 2009a). This caused a decrease in the expression of the parasite gene and raises the prospect of trans-specific gene silencing as a potential strategy for controlling parasitic weeds through engineered hosts.

In addition to RNAs, proteins have also been shown to traffic from hosts to parasitic plants. The characterization of protein mobility employed transgenic host plants expressing green fluorescent protein (GFP) controlled by the phloem targeting *Arabidopsis* SUC2 promoter (AtSUC2-GFP) (Imlau *et al.*, 1999). This construct was expressed in tobacco and the GFP signal translocated into *Cuscuta reflexa*, confirming the symplastic connection between host and this parasite as well with evidence of unloading of the fluorescent protein in sink tissues (Haupt *et al.*, 2001a). The same construct was used to study host protein uptake by *P. aegyptiaca* parasitizing transgenic tomato. Although an endoplasmic reticulum-targeted version of the protein did not move, a soluble version accumulated in tubercles of the parasite (Aly *et al.*, 2011).

Parasitic tissue is not the final destination of every host molecule taken up. The leafhopper-borne yellows disease, originally believed to originate from a virus, is caused by a mycoplasma-like bacteria that flows through plant phloem (Doi *et al.*, 1967). In two studies these bacteria, called phytoplasmas, have been shown to traverse *Cuscuta odorata* bridges, and infect healthy secondary *Catharanthus roseus* (periwinkle) hosts. *Cuscuta* bridges connecting Alder Yellows

(ALY) infected alder to non-infected periwinkle were able to transfer the phytoplasma, and symptoms were slowly manifested four months after the *Cuscuta* bridge was established (Marcone *et al.*, 1997). In another study phytoplasmas were transmitted in 50% of the cases from *Lilium* (hybrid Casablanca) to periwinkle through the *Cuscuta* bridge, with the recipient plant developing symptoms of stunting and flower bud deficiency 2-3 weeks following connection by *Cuscuta* (Kamińska & Korbin, 1999)

The final example of host-parasite macromolecule exchange is Horizontal Gene Transfer (HGT). Although HGT and RNA trafficking both involve movement of nucleic acids, it is not certain whether they use similar mechanisms. HGT has been reported between several non-parasitic plants, but seems to occur at a higher frequency in parasitic associations (Richardson & Palmer, 2007). The high frequency of transfer events to or from parasites is probably attributable to the greater opportunity for nucleic acid exchange provided by the close physical association of the parasitic interaction. HGT events are generally discovered during phylogenetic studies in which genes that are otherwise reliable indicators of species phylogenies instead show high homology to versions in distant families. All HGT events described to date are the result of transfers that occurred thousands of years ago, and little is known about the frequency with which HGT events occur, or the frequency at which the transferred genes are introgressed into the recipient genome.

The precise mechanism of HGT remains unresolved, but two possibilities exist. One mechanism that relates directly to the topic of mRNA trafficking is suggested by an HGT event in which a gene of unknown function moved from a grass host into *Striga* spp. (Yoshida *et al.*, 2010). The introduced gene in *Striga* has high homology to the version in grass, but lacks introns and seems to have the remnant of a poly-A sequence, suggesting that the transfer occurred via an mRNA intermediate. In contrast, most other reported HGT events involve mitochondrial genes (e.g.,

atp1, *atp6* and *matK*), and these have been proposed to transfer between plants as large sections of mitochondrial DNA (Mower *et al.*, 2010). The transfer of DNA across graft junctions has been shown to occur in both directions, although the movement of DNA is limited to the region of the graft so is likely a cell-to-cell movement that does not involve phloem (Stegemann & Bock, 2009). This may be similar to what happens in parasitic plant and other natural grafts, but limited mobility of the transgenes would require development of a shoot from the graft junction to incorporate the foreign gene into the germline of the recipient plant. The study of HGT will benefit from increased sequencing of all plants, including parasitic species.

RNA trafficking

A model for cell-to-cell RNA transport to parasites

As a starting point, we base our understanding of host-parasite symplastic connections on what is known from cell-to-cell connections in autotrophic plants. In land plants intracellular connections need to be small enough to allow structural integrity of cell walls, while allowing for macromolecular movement throughout the plant. Cell-to-cell movement of RNA occurs through PD which are complex structures embedded in cell walls between cells and are composed of plasma membrane lined with microtubules and a continuous span of ER (Wolfe & dePamphilis, 1998; Hyun *et al.*, 2011). Water and solutes are able to passively travel via concentration gradients but movement is limited to molecules below 1 kDa. Facilitated movement of larger molecules is selective and usually requires localization to, and interaction with, PD (Wolfe & dePamphilis, 1998; Hyun *et al.*, 2011).

In order for mRNA to move in a non-cell-autonomous manner, recognition of signal sequences, or motifs, such as 3'untranslated regions (UTRs), interact with specific RNA binding proteins much like the movement proteins (MP) of RNA viruses (discussed below). The MPs bind RNAs,

repress translation, and enable movement of the ribonuclearprotein (RNP) complex to a PD (Baillaud, 1953; Hyun *et al.*, 2011). Mutational studies have determined that cell-to-cell movement of RNA is also motif-dependent in order for RNPs to associate with PD trafficking proteins. NON-CELL-AUTONOMOUS PATHWAY PROTEIN1 (NCAPP1) acts as a receptor for pumpkin (*Cucurbita maxima*) Phloem Protein 16 (CmPP16) in order to enter the phloem stream (Xoconostle-Cazares *et al.*, 1999), and requires phosphorylation and glycosylation (of Tyr63 and Ser66, respectively) for entrance to PD (Lee *et al.*, 2003). SHORTROOT (SHR) requires multiple motifs for mobility, but a singular specific sequence that confers mobility has not been identified (Gallagher & Benfey, 2009).

One of the best characterized mobile mRNAs is KNOTTED1 (KN1). The KNOTTED1-like family of transcription factors are ubiquitous plant cell constituents responsible for regulating pattern formation of the apical meristem. The KN1 homeodomain protein complexed to its own mRNA was detected in phloem and was associated with determining cell fate (Lucas *et al.*, 1995a). Fluorescently-labeled *E. coli* KNOTTED1 mRNA transcript was microinjected into maize mesophyll adjacent to vascular tissue, and was seen to migrate from cell to cell, presumably via PD connections. Mutated KN1 proteins were microinjected to show that the homeodomain protein was necessary for translocation.

Given this understanding of cellular RNA transport, we propose a model for physical interaction and RNA trafficking between host and parasite cells (Figure 1). We focus on host-parasite connections via a shared PD because in the case of *Cuscuta* there is both physical and experimental evidence for trafficking through PD (discussed above) and for broomrapes the PD may initiate the formation of sieve pores between species (Dorr & Kollmann, 1995). The model is based on the assumption that chimeric PD are congruent to self PD in terms of structure and

function. Therefore, host RNP complexes destined for for transport to another cell would associate with appropriate chaperone and PD-associated proteins and be directed through the PD. Once in the parasite cell, the fate of the RNP complex is unknown, but presumably could be processed in a manner similar to other translocated RNAs. Of course one of the interesting aspects of host parasite interactions is that the interaction is not necessarily equal and while the parasite requires an open PD for nutrient acquisition, the host would benefit by shutting down the connection. One or both plants must maintain the shared PD, with its associated plasma membrane, ER and embedded proteins, but nothing is known about this aspect of the interaction.

Virus movement as a model for mRNA trafficking

The mechanism of intercellular virus delivery, facilitated by MPs in cell-to-cell movement, informs our understanding of RNA transmission into and through parasitic plant attachments. Most viruses make MPs and coat proteins which assemble onto viral RNA and form viral nucleoprotein complexes (vNPCs) to facilitate passage through PD (Lough & Lucas, 2006; Lucas *et al.*, 2009; Hyun *et al.*, 2011). Two models demonstrate the mechanism of MP-mediated intercellular transmission of viral RNA through PD. First, the MP of a well-characterized phloem mobile pathogen, Tobacco Mosaic Virus (TMV), may associate with ER, F-actin and microtubules (MTs) for delivery of vNPCs to PD. The MP may send a PD dilation signal to adjacent cells to allow vNPCs transport (Waigmann *et al.*, 1994b; Waigmann & Zambryski, 1995; Kawakami *et al.*, 2004), and has been shown to bind to calreticulin in the ER to aid in cytoskeletal trafficking (Chen *et al.*, 2005). A second mechanism for virus transmission based on cell-to-cell trafficking of potato virus X (PVX) presents some of the complexities of movement of vNPCs to PD. PVX encodes three viral proteins, called a triple-gene block (TGB) that is associated with cell-to-cell movement of vNPCS to PD. TGBp2 and TGBp3 are integral

endomembrane ER proteins and TGBp1 forms a TGBp1-viral RNA complex to increase SEL of PD. Anchoring TGBp1-viral RNA complex to TGBp2 and TGBp3 mediates the delivery of the complex to the PD. In both TMV and PVX, movement proteins encounter SEL-binding motifs on PD proteins, which results in dilation of the channel and movement of the virus-protein complex through the PD (Haupt *et al.*, 2005).

Long distance RNA transport

Generally, long distance transport of RNA occurs when mRNAs from companion cells are loaded into sieve elements and are thus able to move systemically in the phloem stream. The RNP complex is stabilized as the sieve elements do not contain ribosomes for translation and lack RNAse activity that would lead to degradation, making an ideal conduit for transport of mRNAs (Kragler, 2010a). mRNAs may be specifically unloaded at their destination (Haywood, V. *et al.*, 2005) based on a hypothesized targeting signal (or "zip code") incorporated into the RNA or accompanying protein. Mounting evidence indicates that RNA motifs enhance long-distance trafficking, specifically the UTRs of mRNA and chaperone proteins that direct movement through the symplast. UTRs from *StBEL5* fused to the coding sequence of a less mobile BEL-homolog increased mobility (Hannapel, 2010a) and the 3'UTR of *GAI* in *Arabidopsis* has been shown to be required for transcript movement (Huang & Yu, 2009a)

Another important example is *FLOWERING LOCUS T* (*FT*) RNA, which travels by means of a *cis*-acting element on the transcript, independent of the essential FT protein, from leaf to shoot apical meristem to induce flowering and suggests that mRNA has a role in systemic signaling of major developmental transitions (Li *et al.*, 2009; Li *et al.*, 2011). If parasites form compatible cellular connections with their hosts it is possible to infer these mechanisms of RNA mobility

function in much the same way in a heterologous systems, and raising questions about the potential for direct information exchange across species.

Parasitic plants as tools for studying RNA trafficking

Studying RNA mobility over short and long distances in plants is technically challenging because of the need to pinpoint source and destination cells of the mRNA. Techniques range from the elegant application of transgenics, phloem-sap collection (either directly or using aphid stylets), grafting, or combinations of these (Atkins *et al.*, 2011). Each approach has advantages, but they are also subject to limitations in scope and potential for artifacts. We propose that parasitic plants can be added to this suite of techniques.

Studies of short-distance movement of mRNA have relied on transgenics and the use of genes with high specificity of expression (Kim *et al.*, 2005), microinjection (Xoconostle-Cazares *et al.*, 1999) or particle bombardment (Bennet, 1944). These approaches are excellent for demonstrating cell-to-cell movement and characterizing the features controlling trafficking of specific genes. However, these experiments are laborious and target individual genes, so are less useful for genomic-scale studies.

Investigations of long distance trafficking are somewhat simpler because RNAs identified from phloem sap are assumed to be in the process of moving from source to destination. Direct collection of phloem sap exuded from stem incisions is the ideal, but relatively few herbaceous species are copious exuders like cucurbits (Liere & Maliga, 2001) and legumes (Sharkey & Pate, 1976). The phloem stream of most plant species is rapidly shut down by callose plugs resulting in insufficient quantities of sap. To circumvent this limitation, phloem sap can be gathered by placing the cut stem in a solution of EDTA, a chelating agent that delays formation of the

calcium cation callose plug that clogs the sieve plates (Turgeon & Wolf, 2009). However, in all cases of phloem sap collection, care must be taken to avoid contamination with cellular contents from the cut cells at the incision site. This is generally done by blotting the wound surface immediately after the incision and discarding the first several µl of sap exudate, but this minimizes contamination rather than completely eliminating it.

Another ingenious approach for sampling phloem is aphid stylectomy, which uses the ability of aphids to precisely insert their stylets into plant sieve elements. Subsequent severing of the aphid from its stylet leaves a tube that exudes small droplets of nearly pure phloem sap (Fisher & Frame, 1984; Doering-Saad *et al.*, 2002). Disadvantages of this method are that it is technically challenging and yields low volumes of sap, generally in the nano- or microliter range (Atkins *et al.*, 2011). A relatively high throughput method for collecting barley phloem sap has been used that attached microcapillary tubes to the embedded stylets of 600 aphids on 30 plants, and over the course of 6 hours captured 10 µl of phloem sap for protein and mRNA analysis (Gaupels *et al.*, 2008).

Grafting is an excellent method for detecting phloem-mobile molecules where the combination of stock and scion enable differentiation of the mobile signal. For example, wide grafts (or heterografts) between related species such as pumpkin and cucumber have been used in several key studies of RNA trafficking (Lucas *et al.*, 1995a; Lee *et al.*, 2003; Ham *et al.*, 2009). Solanceous species are conducive to grafting and have been used in experiments that showed that tuber formation in potato is regulated by a mobile transcription factor, *StBEL5* (Banerjee *et al.*, 2006), and that tuber formation is increased in grafts with transgenic overexpressing scions. Where one member of the graft carries a mutation or transgene, it is possible to demonstrate

trafficking of a specific mRNA and coincident transmission of a phenotype as was done for *StBEL5* in potato (Rosin *et al.*, 2003) and the leaf shape phenotype Mouse ears (ME) in tomato (Kim *et al.*, 2001). A procedure for grafting together different *Arabidopsis* plants has been used for detecting flowering signals, but the success rate is relatively low compared to other graft systems at approximately 11% (Yunker, 1932). A disadvantage of grafting is the relatively narrow range of plants that can be joined together. Even wide grafts are restricted to members of the same plant family (Mudge *et al.*, 2009), which leaves ambiguity in distinguishing mobile RNAs between stock and scion.

Parasites form connections to their hosts that have many of the features of grafts as well as specific advantages. Chief among these is the ability of parasites to form interspecific grafts with a wide range of species. Given the ability of parasites to form symplastic connections with many of the species commonly used for RNA trafficking studies such as *Arabidopsis* (Westwood & Foy, 1998; Birschwilks *et al.*, 2007), tomato, and cucurbits (Roney, J *et al.*, 2007), it is possible to conduct comparative studies on different plants using the parasite as a common “scion”. More importantly, the phylogenetic differences between parasites and these hosts facilitates the use of genomics approaches to understand RNA trafficking. *Cuscuta* that was grown on host species with sequenced genomes yields a mixture of host and self RNA that can be distinguish using microarrays (Roney, J *et al.*, 2007) and it presents an excellent application for next-generation sequencing. Even without an extensive *Cuscuta* genome, the host sequences can be determined for most genes based on exact matches to known host sequences. An extensive database for expressed genes of Orobanchaceae parasites exists (Westwood *et al.*, 2012), which will improve confidence in distinguishing host and parasite RNAs. The broad host range of *Cuscuta* will also

facilitate studies on species that do not graft well or from which phloem exudates are difficult to obtain.

A concern over analyses of phloem sap obtained from incisions is that the sudden release in pressure would create artifacts such as dislodged macromolecules that would otherwise not be mobile in sieve elements (Atkins et al., 2011). Parasitic plants would avoid this concern because the process of parasitism unfolds gradually, so presumably is accomplished without the artifacts arising from sudden pressure changes. Of course the parasite creates a strong sink that draws material from the host, but with negligible leaf surface area (for *Cuscuta* and *Phelipanche*), the process of withdrawing host vascular contents is likely within the range of normal plant translocation.

As with other methods of studying RNA trafficking, host-*Cuscuta* connections have certain disadvantages. Among these are the technical issues of establishing connections on specific host locations and at specific times. Parasites are unpredictable, and generating uniform tissues from synchronized attachment points can be challenging. In *Cuscuta*, the haustoria also follow a developmental progression through stages of host penetration, vascular connection, and eventually occlusion (Vaughn, 2003), and although the period of open transfer of macromolecules likely spans many days or weeks, there is little information on how to distinguish actively translocating haustoria from those that are too young or old to function well. Thus, the most important limitation may well be the dearth of information on the precise functioning of haustorial connections. Finally, the interaction of host and parasite is ultimately one of pathogen and host, and it is reasonable to expect that control of the PD and macromolecule exchange is a point of contention between the two species. One consequence of this is that the mRNAs trafficked from host to parasite will be enriched in pathogen/defense

response functions, but this is also true to some degree for all other methods of sampling phloem contents.

Potential significance of trans-species RNA trafficking

The most intriguing questions regarding cross species movement of RNA are whether and how RNA from one species functions in another species. Answering this question is difficult because mechanisms of action of native trafficked RNAs are not fully understood. If mRNAs function through translation into protein as indicated for *KN1* (Kim *et al.*, 2005), then parasites should be able to process host transcripts as well as they process their own mobile mRNA (Figure 1). On the other hand, if trafficked RNA acts in a sequence specific manner, then function in a parasite would depend on the presence of sufficiently homologous genes in the parasite. Many trafficked mRNA may not have a metabolic or regulatory function and may simply serve as a nutrient source for the parasite. Of course all three of these possibilities may occur, depending on the transcript in question.

Mobile RNAs and proteins have been shown to influence leaf shape (Kim *et al.*, 2001; Haywood, V. *et al.*, 2005), tuberization (Banerjee *et al.*, 2006), and flowering (Li *et al.*, 2011). There is little evidence for transmission of a host phenotype in parasites, as the general morphology of the parasite does not depend on which host is used. Of course, parasites may be deficient in key regulatory pathways, for example those that lack expanded leaves would not be expected to perceive and respond to altered signals for leaf shape. A particularly intriguing area is flowering, and some authors have suggested that *Cuscuta* flowering time depends on timing of host flowering, thus invoking the possibility that *FT* RNA and protein from the host are capable of inducing flowering in the parasite (Fratianne, 1965; Mortazavi *et al.*, 2008). If the trafficked mRNA is a transcription factor, it is possible that the delivery of just a few molecules to the right

cells could effectively throw the switch to turn on a new developmental program (Kragler, 2010a).

From an evolutionary standpoint, it is reasonable to think that parasites would be under selective pressure to decode and use information from host mRNAs and proteins that reveal the host physiological status, thereby enabling the parasite to respond to any changes in the host system. For example, by recognizing when the host is preparing to flower or senesce, the parasite can complete its own reproductive cycle in time. Considering that the host forms a dominant feature of the parasite's environment, it would not be surprising if the parasites were found to have mechanisms to monitor the health and developmental status of their hosts.

One mechanism for communication that has empirical support is post transcriptional gene silencing. The ability to silence GUS in transgenic *Triphysaria* and M6PR in *Phelipanche* indicates that the process works across species as long as the RNAi construct matches the target gene in the parasite. We have no information to date on whether parasites contain homologs of silencing targets known from other plant species, but we hypothesize that parasites may encounter miRNA signaling molecules that are generated in stressed plants (Stefanovic *et al.*, 2002; Lu & Huang, 2008). Of the many identified miRNAs, miRNA399, a phosphate starvation response signal, has been shown to move into grafted tissues (Stefanovic *et al.*, 2002), so likely passes into parasites. It will be interesting to learn whether parasites have sufficiently conserved homologs to be affected by this type of regulation. For perspecitive we asked this question of four *C. pentagona* genes with sequences available in the 1KP Project (http://www.onekp.com). Genes for a pentatricopeptide repeat-containing protein (Solyc01g081290.2.1), DNA mismatch repair protein (Solyc02g082660.2.1), clathrin heavy chain (Solyc06g051310.2.1) and ATPase subunit 1 (Solyc11g039980.1.1) are 80% (464/583), 83% (166/199), 85% (720/852) and 96%

(251/261) identical, respectively, at the nucleotide level between tomato and *C. pentagona* (Bombarely *et al.*, 2011). This suggests that for certain genes it is likely that a highly homologous 21 nt silencing signal, such as miRNA, could act between host and parasite.

Because parasites are foremost feeding on their hosts, it is possible that the RNA is taken up as a nutrient and carried with the bulk movement of solutes from the host plant with no informational significance for the parasite. The host trafficked mRNA in parasites may undergo catabolism to provide inorganic phosphate and sugars for conservation of cellular and organismal homeostatsis of the parasite or to maintain haustorial function. However, for this to be true, the parasite must have a mechanism to distinguish host trafficked RNA from its own. This might be possible given sufficient differences between host and parasite RNA and protein sequences if the limitation on function was at the level of uptake into parasite cells. By this hypothesis, parasite PD would discriminate and only allow uptake of self RNPs. However, if such precise recognition is possible, it would suggest that the parasite ability to differentiate RNPs would decrease when parasitizing closely related species, or for certain highly conserved genes. No evidence exists that parasite host ranges are specifically aimed at targeting the most distantly related species.

There may be an evolutionary benefit to trafficking of nucleic acids into the parasite, such as for horizontal gene transfer (Yoshida *et al.*, 2010). Many parasites lineage including Rafflesiaceae, Orobanchaceae, *Cuscuta,* Mitrastemonaceae, Santalales, and *Pilostyles* (Apodanthaceae) show evidence of HGT events in phylogenetic studies (Table 1). However, the functional significance of most HGT events involving parasites is unclear and a benefit to the parasite has not yet been demonstrated for any gene acquired in this way.

Finally, it is interesting to speculate whether RNAs from the parasite could be used as pathogenic factors in establishing and maintaining host connections. Although an assay for parasite-to-host movement of a *C. pentagona PYROPHOSPHATE-DEPENDENT PHOSPHOFRUCTOKINASE* β subunit mRNA was negative (David-Schwartz R, 2008), the movement of viruses and RNAi signals through parasite bridges from one host into another suggests that bidirectional movement of mRNA is also possible (Birschwilks, M. *et al.*, 2006; Tomilov *et al.*, 2008). If RNA trafficking is important in cell-to-cell communication, it would be reasonable to expect that parasites have evolved a way to use this system to their advantage.

Conclusions and future directions

Parasitic plants that form symplastic connections with their hosts provide a new perspective on cell-to-cell and long-distance trafficking of RNAs. The haustorium creates a union with the host that resembles a graft in many ways, but is established through invasive growth of the parasite and is coordinated with host cells and therefore presents a unique type of junction. Macromolecules including RNAs, proteins and DNAs, are able to traffic between the plants and the system presents new opportunities for studying RNA movement. Parasitic plants are ultimately pathogens, and the interactions may not be precisely equivalent to the connections between two cells of the same plant, and it is almost certain that parasitic plants have adapted the normal structure and function of PD and RNA trafficking machinery to meet their needs. Unraveling the intricacies of the parasite-host interface is likely to contribute to understanding these phenomena in the same way that studies of viruses high-jacking the RNA trafficking system played a major role in elucidating components of PD function (Carrington *et al.*, 1996). The use of parasitic plants for understanding RNA trafficking will benefit from deeper understanding of haustorial function. The interface between two plants involves coordination at

many levels, from structural (shared PD) to signaling. Further research on this interaction will provide new insights into cell-cell interactions

An important feature of the haustorial connection is the ability to connect to diverse host species and this leads to a unique system in which the mobile transcriptome of one species is mixed into that of another. The situation raises intriguing questions about the functions of RNA trafficking in plants. Which RNAs are trafficked, are they targeted specifically, and how do they function at their destination? Are they translated into protein, do they modify gene expression, or are they recycled into raw material for nutrition of the recipient cells? Many of these questions can be addressed using parasitic plants. The ability to readily distinguish host and parasite sequences facilitates identification of mobile transcripts and their movement and fate in the parasite can be tracked using quantitative methods. Next generation sequencing provides the power to reconstruct mobile transcriptomes of hosts from parasite-derived RNA populations.

Ackowledgements

This work was supported by NSF award IOS-0843372 to JHW, with additional support from NSF Plant Genome DBI-0701748 and U.S. Department of Agriculture Hatch Project no. 135798.

References

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

Atkins, C., Smith, P., and Rodriguez-Medina, C. (2011). Macromolecules in phloem exudates—a review. *Protoplasma* 248**,** 165-172.

Ayre, B.G., and Turgeon, R. (2004). Graft transmission of a floral stimulant derived from CONSTANS. *Plant Physiol* 135**,** 2271-2278.

Baillaud, L. (1953). Effect of temperature on the period of nutation of the volubile stems of *Cuscuta*. *CR Hebd Acad Sci* 236**,** 1986-1988.

Banerjee, A., Chatterjee, M., Yu, Y., Suh, S.-G., Miller, W., and Hannapel, D. (2006). Dynamics of a mobile RNA of potato involved in a long-distance signaling pathway. *Plant Cell* 18**,** 3443-3457.

Barkman, T.J., Mcneal, J.R., Lim, S.H., Coat, G., Croom, H.B., Young, N.D., and Depamphilis, C.W. (2007). Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evol Biol* 7.

Bennet, C.W. (1944). Studies of dodder tranmission of plant viruses. *Phytopathology* 34**,** 905-932.

Birschwilks, M., Haupt, S., Hofius, D., and Neumann, S. (2006). Transfer of phloem-mobile substances from the host plants to the holoparasite *Cuscuta* sp. *J Exp Bot* 57**,** 911-921.

Birschwilks, M., Sauer, N., Scheel, D., and Neumann, S. (2007). Arabidopsis thaliana is a susceptible host plant for the holoparasite *Cuscuta* spec. *Planta* 226**,** 1231-1241.

Bombarely, A., Menda, N., Tecle, I.Y., Buels, R.M., Strickler, S., Fischer-York, T., Pujar, A., Leto, J., Gosselin, J., and Mueller, L.A. (2011). The sol genomics network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res* 39**,** D1149-D1155.

Borsics, T., and Lados, M. (2002). Dodder infection induces the expression of a pathogenesis-related gene of the family PR-10 in alfalfa. *J Exp Bot* 53**,** 1831-1832.

Buhtz, A., Pieritz, J., Springer, F., and Kehr, J. (2010). Phloem small RNAs, nutrient stress responses, and systemic mobility. *BMC Plant Biol* 10**,** 64.

Burch-Smith, T.M., and Zambryski, P.C. (2012). Plasmodesmata paradigm shift: regulation from without versus within. *Annu Rev Plant Biol* 63**,** 239-260.

Carrington, J.C., Kasschau, K.D., Mahajan, S.K., and Schaad, M.C. (1996). Cell-to-cell and long-distance transport of viruses in plants. *Plant Cell* 8**,** 1669-1681.

Chen, M.-H., Tian, G.-W., Gafni, Y., and Citovsky, V. (2005). Effects of calreticulin on viral cell-to-cell movement. *Plant Physiol* 138**,** 1866-1876.

Christensen, N.M., Dörr, I., Hansen, M., Van Der Kooij, T.a.W., and Schulz, A. (2003). Development of *Cuscuta* species on a partially incompatible host: induction of xylem transfer cells. *Protoplasma* 220**,** 131-142.

Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., Giakountis, A., Farrona, S., Gissot, L., Turnbull, C., and Coupland, G. (2007). FT protein movement contributes to long-distance signaling in floral induction of arabidopsis. *Science* 316**,** 1030-1033.

Crafts, A.S. (1932). Phloem anatomy, exudation, and transport of organic nutrients in cucurbits. *Plant Physiol* 7**,** 183-225.

David-Schwartz R, R.S., Townsley B, Machuka J, Sinha N (2008). Long-distance transport of mRNA via parenchyma cells and phloem across the host-parasite junction in *Cuscuta*. *New Phytol* 179**,** 1133-1141.

Davis, C., Anderson, W., and Wurdack, K. (2005). Gene transfer from a parasitic flowering plant to a fern. *Proc R Soc Lond B Biol Sci* 272**,** 2237 - 2242.

Davis, C.C., and Wurdack, K.J. (2004). Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science* 305**,** 676-678.

Doering-Saad, C., Newbury, H.J., Bale, J.S., and Pritchard, J. (2002). Use of aphid stylectomy and RT-PCR for the detection of transporter mRNAs in sieve elements. *J Exp Bot* 53**,** 631-637.

Doi, Y., Teranaka, M., Yora, K., and Asuyama, H. (1967). Mycoplasma- or PLT Group-like Microorganisms Found in the Phloem Elements of Plants Infected with Mulberry Dwarf, Potato Witches' Broom, Aster Yellows, or Paulownia Witches' Broom. *Ann Phytopath Soc Japan* 33**,** 259-266.

Dörr, I. (1972). Contact of *Cuscuta* hyphae with sieve tubes of its host plants. *Protoplasma* 75**,** 167-184.

Dorr, I., and Kollmann, R. (1995). Symplasmic sieve element continuity between *Orobanche* and its host. *Bot Acta* 108**,** 47-55.

Fisher, D.B., and Frame, J.M. (1984). A guide to the use of the exuding-stylet technique in phloem physiology. *Planta* 161**,** 385-393.

Fratianne, D.G. (1965). The interrelationship between the flowering of dodder and the flowering of some long and short day plants. *Am J Bot* 52**,** 556-562.

Gal-On, A., Naglis, A., Leibman, D., Ziadna, H., Kathiravan, K., Papayiannis, L., Holdengreber, V., Guenoune-Gelbert, D., Lapidot, M., and Aly, R. (2009). Broomrape can acquire viruses from its hosts. *Phytopathology* 99**,** 1321-1329.

Gallagher, K.L., and Benfey, P.N. (2009). Both the conserved GRAS domain and nuclear localization are required for SHORT-ROOT movement. *Plant J* 57**,** 785-797.

Gaupels, F., Buhtz, A., Knauer, T., Deshmukh, S., Waller, F., Van Bel, A., Kogel, K.-H., and Kehr, J. (2008). Adaptation of aphid stylectomy for analyses of proteins and mRNAs in barley phloem sap. *J Exp Bot* 59**,** 3297 - 3306.

Griffitts, A.A., Cramer, C.L., and Westwood, J.H. (2004). Host gene expression in response to Egyptian broomrape (*Orobanche aegyptiaca*). *Weed Sci* 52**,** 697-703.

Ham, B.-K., Brandom, J.L., Xoconostle-Cazares, B., Ringgold, V., Lough, T.J., and Lucas, W.J. (2009). A Polypyrimidine Tract Binding Protein, Pumpkin RBP50, Forms the Basis of a Phloem-Mobile Ribonucleoprotein Complex. *Plant Cell* 21**,** 197-215.

Hannapel, D.J. (2010). A Model System of Development Regulated by the Long-distance Transport of mRNA. *J Integr Plant Biol* 52**,** 40-52.

Harada, T. (2010). Grafting and RNA transport via phloem tissue in horticultural plants. *Sci Hortic-Amsterdam* 125**,** 545-550.

Haupt, S., Cowan, G.H., Ziegler, A., Roberts, A.G., Oparka, K.J., and Torrance, L. (2005). Two plant viral movement proteins traffic in the endocytic recycling pathway. *Plant Cell* 17**,** 164-181.

Haupt, S., Oparka, K.J., Sauer, N., and Neumann, S. (2001). Macromolecular trafficking between *Nicotiana tabacum* and the holoparasite *Cuscuta reflexa*. *J Exp Bot* 52**,** 173-177.

Haywood, V., Yu, T.S., Huang, N.C., and Lucas, W.J. (2005). Phloem long-distance trafficking of GIBBERELLIC ACID-INSENSITIVE RNA regulates leaf development. *Plant J* 42**,** 49-68.

Hosford, R.M.J. (1967). Transmission of plant viruses by dodder. *Bot Rev* 42**,** 387-406.

Huang, N.-C., and Yu, T.-S. (2009). The sequences of Arabidopsis GA-INSENSITIVE RNA constitute the motifs that are necessary and sufficient for RNA long-distance trafficking. *Plant J* 59**,** 921-929.

Hyun, T.K., Uddin, M.N., Rim, Y., and Kim, J.Y. (2011). Cell-to-cell trafficking of RNA and RNA silencing through plasmodesmata. *Protoplasma* 248**,** 101-116.

Imlau, A., Truernit, E., and Sauer, N. (1999). Cell-to-cell and long-distance trafficking of the green fluorescent protein in the phloem and symplastic unloading of the protein into sink tissues. *Plant Cell* 11**,** 309-322.

Irving, L.J., and Cameron, D.D. (2009). You are what you eat: Interactions between root parasitic plants and their hosts. *Adv Bot Res* 50**,** 87-138.

Joel, D.M., Losner-Goshen D (1994). "Early host-parasite interaction: models and observations of host root penetration by the haustorium of Orobanche". (Proceedings of the 3rd International Workshop on Orobanche and Related Striga Research: Amsterdam: Royal Tropical Institute).

Kamińska, M., and Korbin, M. (1999). Graft and dodder transmission of phytoplasma affecting lily to experimental hosts. *Acta Physiol Plant* 21**,** 21-26.

Kawakami, S., Watanabe, Y., and Beachy, R.N. (2004). Tobacco mosaic virus infection spreads cell to cell as intact replication complexes. *P Natl Acad Sci USA* 101**,** 6291-6296.

Kim, J.Y., Rim, Y., Wang, L., and Jackson, D. (2005). A novel cell-to-cell trafficking assay indicates that the KNOX homeodomain is necessary and sufficient for intercellular protein and mRNA trafficking. *Gene Dev* 19**,** 788-793.

Kim, M., Canio, W., Kessler, S., and Sinha, N. (2001). Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* 293**,** 287-289.

Kragler, F. (2010). RNA in the phloem: A crisis or a return on investment? *Plant Sci* 178**,** 99-104.

Kuijt, J. (1983). "Tissue compatibility and the haustoria of parasitic angiosperms," in *Vegetative Compatibility Responses in Plants,* ed. R. Moore. (Waco, TX: Baylor University), 1-12.

Lee, J.-Y., Yoo, B.-C., Rojas, M.R., Gomez-Ospina, N., Staehelin, L.A., and Lucas, W.J. (2003). Selective trafficking of non-cell-autonomous proteins mediated by NtNCAPP1. *Science* 299**,** 392-396.

Lee, K.B. (2007). Structure and development of the upper haustorium in the parasitic flowering plant *Cuscuta japonica* (Convolvulaceae). *Am J Bot* 94**,** 737-745.

Li, C., Gu, M., Shi, N., Zhang, H., Yang, X., Osman, T., Liu, Y., Wang, H., Vatish, M., Jackson, S., and Hong, Y. (2011). Mobile FT mRNA contributes to the systemic florigen signalling in floral induction. *Sci Rep* 1.

Li, C., Zhang, K., Zeng, X., Jackson, S., Zhou, Y., and Hong, Y. (2009). A cis Element within Flowering Locus T mRNA Determines Its Mobility and Facilitates Trafficking of Heterologous Viral RNA. *J Virol* 83**,** 3540-3548.

Lough, T., and Lucas, W. (2006). Integrative plant biology: role of phloem long-distance macromolecular trafficking. *Annu Rev Plant Biol* 57**,** 203 - 232.

Lu, X.-Y., and Huang, X.-L. (2008). Plant miRNAs and abiotic stress responses. *Biochem Biophys Res Commun* 368**,** 458-462.

Lucas, W.J., Bouche-Pillon, S., Jackson, D.P., Nguyen, L., Baker, L., Biao, D., and Hake, S. (1995). Selective trafficking of KNOTTED1 homeodomain protein and its mRNA through plasmodesmata. *Science* 270**,** 1980-1983.

Lucas, W.J., Ham, B.-K., and Kim, J.-Y. (2009). Plasmodesmata - bridging the gap between neighboring plant cells. *Trends Cell Biol* 19**,** 495-503.

Lucas, W.J., and Lee, J.-Y. (2004). Plasmodesmata as a supracellular control network in plants. *Nat Rev Mol Cell Biol* 5**,** 712-726.

Lucas, W.J., Yoo, B.-C., and Kragler, F. (2001). RNA as a long-distance information macromolecule in plants. *Nat Rev Mol Cell Biol* 2**,** 849-857.

Marcone, C., Ragozzino, A., and Seemuller, E. (1997). Dodder transmission of alder yellows phytoplasma to the experimental host *Catharanthus roseus* (periwinkle). *Eur J Forest Pathol* 27**,** 347-350.

Moore, M.J. (2005). From birth to death: The complex lives of eukaryotic mRNAs. *Science* 309**,** 1514-1518.

Mower, J., Stefanovic, S., Hao, W., Gummow, J., Jain, K., Ahmed, D., and Palmer, J. (2010). Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol* 8**,** 150.

Mower, J.P., Stefanovic, S., Young, G.J., and Palmer, J.D. (2004). Plant genetics: gene transfer from parasitic to host plants. *Nature* 432**,** 165-166.

Mudge, K., Janick, J., Scofield, S., and Goldschmidt, E.E. (2009). "A History of grafting," in *Horticultural Reviews*. John Wiley & Sons, Inc.), 437-493.

Park, J.M., Manen, J.F., and Schneeweiss, G.M. (2007). 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., and Riches, C.R. (1993). *Parasitic weeds of the world: biology and control.* Wallingford, Oxon: CAB International.

Richardson, A.O., and Palmer, J.D. (2007). Horizontal gene transfer in plants. *J Exp Bot* 58**,** 1-9.

Roney, J., Khatibi, P., and Westwood, J. (2007). Cross-species translocation of mRNA from host plants into the parasitic plant dodder. *Plant Physiol* 143**,** 1037 - 1043.

Rosin, F.M., Hart, J.K., Horner, H.T., Davies, P.J., and Hannapel, D.J. (2003). Overexpression of a Knotted-like homeobox gene of potato alters vegetative development by decreasing gibberellin accumulation. *Plant Physiol* 132**,** 106-117.

Ruiz-Medrano, R., Xoconostle-Cazares, B., and Lucas, W. (1999a). Phloem long-distance transport of CmNACP mRNA: implications for supracellular regulation in plants. *Development* 126**,** 4405 - 4419.

Ruiz-Medrano, R., Xoconostle-Cazares, B., and Lucas, W.J. (1999b). Phloem long-distance transport of CmNACP mRNA: implications for supracellular regulation in plants. *Development* 126**,** 4405-4419.

Ryan, C.A. (1990). Protease Inhibitors in Plants: Genes for Improving Defenses Against Insects and Pathogens. *Annual Review of Phytopathology* 28**,** 425-449.

Sharkey, P.J., and Pate, J.S. (1976). Translocation from leaves to fruits of a legume, studied by a phloem bleeding technique: Diurnal changes and effects of continuous darkness. *Planta* 128**,** 63-72.

Stegemann, S., and Bock, R. (2009). Exchange of genetic material between cells in plant tissue grafts. *Science* 324**,** 649-651.

Swarbrick, P.J., Huang, K., Liu, G., Slate, J., Press, M.C., and Scholes, J.D. (2008). Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant *Striga hermonthica*. *New Phytol* 179**,** 515-529.

Tomilov, A.A., Tomilova, N.B., Wroblewski, T., Michelmore, R., and Yoder, J.I. (2008). Trans-specific gene silencing between host and parasitic plants. *Plant J* 56**,** 389-397.

Turgeon, R., and Wolf, S. (2009). Phloem Transport: Cellular Pathways and Molecular Trafficking. *Annu Rev Plant Biol* 60**,** 207-221.

Vachev, T., Ivanova, D., Minkov, I., Tsagris, M., and Gozmanova, M. (2010). Trafficking of the potato spindle tuber viroid between tomato and *Orobanche ramosa*. *Virology* 399**,** 187-193.

Van Dorst, H.J.M., and Peters, D. (1974). Some biological observations on pale fruit, a viroid-incited disease of cucumber. *Eur J Plant Pathol* 80**,** 85-96.

Vaughn, K.C. (2003). Dodder hyphae invade the host: a structural and immunocytochemical characterization. *Protoplasma* 220**,** 189-200.

Vaughn, K.C. (2006). Conversion of the searching hyphae of dodder into xylic and phloic hyphae: A cytochemical and immunocytochemical investigation. *Intl J Plant Sci* 167**,** 1099-1114.

Waigmann, E., Lucas, W.J., Citovsky, V., and Zambryski, P. (1994). Direct functional assay for tobacco mosaic virus cell-to-cell movement protein and identification of a domain involved in increasing plasmodesmal permeability. *Proc Natl Acad Sci U S A* 91**,** 1433-1437.

Waigmann, E., and Zambryski, P. (1995). Tobacco mosaic virus movement protein-mediated protein transport between trichome cells. *Plant Cell* 7**,** 2069-2079.

Wang, Y., and Ding, B. (2010). Viroids: Small probes for exploring the vast universe of RNA trafficking in plants. *J Integr Plant Biol* 52**,** 28-39.

Westwood, J.H., Depamphilis, C.W., Das, M., Fernandez-Aparicio, M., Honaas, L., Timko, M.P., Wafula, E.K., Wickett, N.J., and Yoder, J.I. (2012). The Parasitic Plant Genome Project: new tools for understanding the biology of *Orobanche* and *Striga*. *Weed Sci* 60**,** 295-306.

Westwood, J.H., and Foy, C.L. (1998). "Arabidopsis thaliana can be a model host for Orobanche research", in: *Current Problems of Orobanche Research, Proceedings of the Fourth International Workshop on Orobanche.* (eds.) K. Wegmann, L.J. Musselman & D.M. Joel. (Albena, Bulgaria: Institute for Wheat and Sunflower "Dobroudja").

Westwood, J.H., Roney, J.K., Khatibi, P.A., and Stromberg, V.K. (2009). RNA translocation between parasitic plants and their hosts. *Pest Man Sci* 65**,** 533-539.

Westwood, J.H., Yoder, J.I., Timko, M.P., and Depamphilis, C.W. (2010). The evolution of parasitism in plants. *Trends Plant Sci* 15**,** 227-235.

Xoconostle-Cazares, B., Xiang, Y., Ruiz-Medrano, R., Wang, H.-L., Monzer, J., Yoo, B.-C., Mcfarland, K.C., Franceschi, V.R., and Lucas, W.J. (1999). Plant paralog to viral movement protein that potentiates transport of mRNA into the phloem. *Science* 283**,** 94-98.

Yoshida, S., Maruyama, S., Nozaki, H., and Shirasu, K. (2010). Horizontal gene transfer by the parasitic plant *Striga hermonthica*. *Science* 328**,** 1128.

Table 4.1. Macromolecules demonstrated to transfer between hosts and parasitic plants.

Material	Examples	Parasite	Host(s)	Reference
mRNA	Many	*Cuscuta*	Tomato Pumpkin Alfalfa	(Roney, J *et al.*, 2007; David-Schwartz R, 2008)
siRNA	GUS silencing	*Triphysaria*	Lettuce	(Tomilov *et al.*, 2008)
	M6PR silencing	*Phelipanche*	Tomato	(Huang & Yu, 2009a)
Viruses	Many	*Cuscuta*		(Hosford, 1967; Birschwilks, M. *et al.*, 2006)
	ToMV, PVY, TYLCV, CMV	*Phelipanche*	Tomato	(Gal-On *et al.*, 2009)
Viroids	HSVd	*Cuscuta*	Cucumber	(van Dorst & Peters, 1974)
	PSTVd	*Phelipanche*	Tomato	(Vachev *et al.*, 2010)
Phytoplasmas	Yellows disease	*Cuscuta*	Alder to Periwinkle Lily to Periwinkle	(Marcone *et al.*, 1997; Kamińska & Korbin, 1999)

Protein	GFP	*Cuscuta*	Tobacco, *Arabidopsis*	(Haupt *et al.*, 2001a; Birschwilks, M. *et al.*, 2006)
	GFP	*Phelipanche*	Tomato	(Aly *et al.*, 2011)
DNA	*nad1B-C*	Rafflesiaceae	Vitaceae	(Davis & Wurdack, 2004)
	atp1, *atp6*, *matR*	*Cuscuta*	*Plantago*	(Mower *et al.*, 2004; Mower *et al.*, 2010)
	atp1	Orobanchaceae, Convolvulaceae	*Plantago*	(Mower *et al.*, 2004)
	atp1	*Pilostyles*	Legumes	(Barkman *et al.*, 2007)
		Rafflesiaceae	Vitaceae	
		Cytinus	*Helianthemum*	
		Mitrastema	*Fagus*	
	rps2	*Orobanche* and *Phelipanche*	unknown	(Park *et al.*, 2007)
	ShContig9483	*Striga*	grass	(Yoshida *et al.*, 2010)
	nad1B-C, *matR*	Santalales	*Botrychium virginianum*	(Davis *et al.*, 2005)

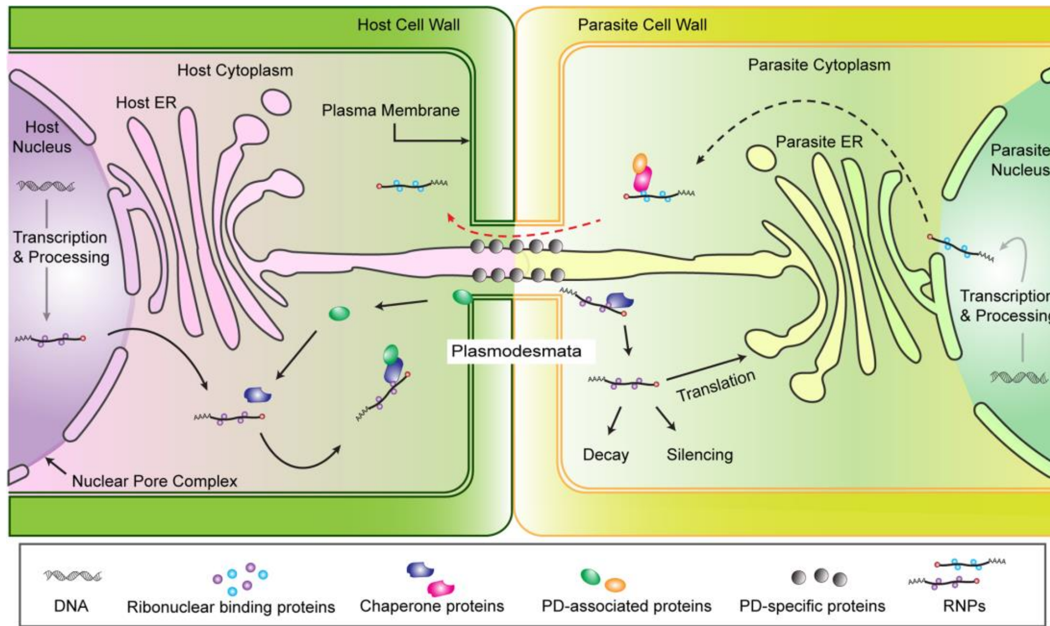


Figure 1.1. Hypothesized model of RNA trafficking from host to *Cuscuta* via plasmodesmata.

Mature RNAs (mRNA, miRNA or siRNA) associate with RNA binding proteins and are targeted for export from the nucleus and post-transcriptional regulation. The ribonuclearprotein complex (RNP) includes the RNA molecule plus proteins that help facilitate RNP export through the nuclear pore complex (NPC) into the cytoplasm (Lucas *et al.*, 2001; Moore, 2005) . The RNPs in the cytoplasm may be translated into protein, degraded or translocated into adjacent cells. Selective mechanisms for PD transport of non-cell-autonomous proteins (NCAPs) or RNPs proposed by Lucas et al. (2001, 2009) suggest chaperone proteins and/or PD-associated proteins carry NCAPs or RNPs to a PD docking protein which dilates the PD channel to transport the RNP to the cytoplasm of neighboring cells (Kim *et al.*, 2005). Nothing is known about the fate of a host RNP after it reaches the parasite cell, but formal possibilities include translation, degradation, or modulation of parasite gene expression (e.g., silencing). The existence of similar mechanisms that would allow RNP trafficking from parasite to host are expected in parasite cells (Red dashed arrow).

Chapter 2: *De novo* assembly and characterization of the parasitic plant *Cuscuta pentagona* stem and haustoria transcriptomes

Gunjune Kim[1], Megan LeBlanc[1], Eric Wafula[2], Claude dePamphilis[2], James Westwood[1].

[1]Department of Plant Pathology, Physiology and Weed Science, Virginia Tech, Blacksburg, Virginia, USA

[2]Department of Biology and the Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA

Abstract

Recently, large-scale RNA exchange has been demonstrated between the parasitic plant *Cuscuta pentagona* and its hosts. *Cuscuta* are destructive species in agriculture and the RNA mobility with host plants may provide insights into the key parasitic organ, the haustorium. The exact mechanism of haustorium function remains poorly understood despite decades of research. In order to develop suitable resources for understanding the process of *Cuscuta* parasitism, we have generated Illumina sequence data from the host stem, the haustorial region (attachment point combining host and parasite tissue), and *Cuscuta* stem near the point of host attachment. From these sequences a *de novo* assembly of the *Cuscuta* transcriptome was generated using Trinity software. Differential expression analysis was conducted to compare the transcriptomes of the haustorial region and *Cuscuta* stem. A total of 81,704 *Cuscuta* unigenes were generated. Using *Cuscuta* unigenes and differential expression data generated from host-parasite interface, we have endeavored to address a longstanding question of the origin of the *Cuscuta* haustorium based on correlating haustorium upregulated transcripts to known root- or shoot-specific transcription factors in *Arabidopsis*. We conclude that genes expressed in the *Cuscuta* haustorium are primarily derived from those functioning in roots, although *Cuscuta* haustorial gene expression contains shoot-specific transcription factor homologs as well.

Keywords: RNA-Seq, transcriptome, haustorium, root, shoot, *Cuscuta pentagona*, *de novo* assembly,

Introduction

Parasitic plants are unique compared to other plant species in that they have evolved feeding structures called the haustoria. Through haustoria, parasitic plants make contact with hosts, penetrate the epidermis, and establish vascular connections to acquire water and nutrients. Parasitic plants can be categorized as stem or root parasites and also can be classified as hemiparasite or holoparasite depending upon photosynthetic ability and host dependence (Westwood *et al.*, 2010)..

Cuscuta (commonly called dodder) has a vegetative body that lacks roots or expanded leaves, and consists primarily of a yellow-orange stem that entwines around above-ground organs of host plants (Parker & Riches, 1993; Press & Graves, 1995). *Cuscuta* is also a troublesome species in agricultural fields and is able to parasitize a wide host range, causing significant losses in crops such as tomato, alfalfa, soybean, etc. (Smith *et al.*, 2013). The fact that *Cuscuta* itself is anatomically simple, yet is able to connect with a wide range of hosts make *Cuscuta* a model system to study host-parasite connections. Recent studies showing mRNA movement between *Cuscuta* and its hosts further demonstrates the unique capacities of the *Cuscuta* haustorium (Roney *et al.*, 2007; Kim G *et al.*, 2014)

Haustorial development has been reported in detail for *Cuscuta australis* and *Cuscuta japonica* (Lee, Kyu Bae, 2007). When the *Cuscuta* stem contacts the host, parasite cortical parenchyma cells between the stele and the epidermis undergo dedifferentiation to initiate a haustorium. The epidermal cells of the *Cuscuta* stem then produce fingerlike projections in response to host contact, light quality, and chemical cues, and these digitate cells are reported to have dynamic metabolic activity during growth through host cells (Lee, K. B., 2007; Haidar &

Boss, 2009; Furuhashi *et al.*, 2011). The digitate cells undergo cell division and rapid growth towards host tissue to become searching hyphae cells, eventually differentiating to develop into xylic or phloic hyphae cells.

All parasitic plants have xylem connections with their hosts, but species may differ substantially in their symplastic connections with the host. Parasitic plants can extract host resources through apoplastic transger, plasmodesmata connections, direct phloem-phloem connections or combination of these ways (Smith *et al.*, 2013). As mentioned, *Cuscuta* haustoria develop thin cells called searching hyphae, which elongate through the host tissue until they reach the vascular system of the host. These searching hyphae have ability to identify the host cells they meet so that *Cuscuta* hyphae cells are able to differentiate into xylem or phloem elements, depending on which vascular tissue they contact first (Dörr, 1990; Vaughn, 2006). The hyphae can grow between and through host cells and make direct parasite-host connections via plasmodesmata (PD) (Vaughn, 2003). The types of PD connection can be simple or branched and form complete structures with desmotubules characteristic of typical PD (Birschwilks *et al.*, 2006).

Unlike PD connections, the anatomical evidence of phloem-phloem connections between host and *Cuscuta* is not well demonstrated. However, dye experiments suggest convincing evidence of open phloem continuity. Both carboxyfluorescein, a symplastic marker, and a phloem promoter-driven green fluoresecent protein (GFP) transfer from host to *Cuscuta* via the phloem of established haustoria (Haupt, S. *et al.*, 2001; Birschwilks *et al.*, 2006; Birschwilks *et al.*, 2007).

Previous studies have speculated on the origin of *Cuscuta* haustoria in terms of whether the structure originated from root or shoot tissue (Lee, Kyu Bae, 2007; Alakonya *et al.*, 2012). However, the identity of *Cuscuta* haustoria is still uncertain. Recently, a number of studies using *de novo* assembly and transcriptome analysis in *Cuscuta* as well as other parasitic plant species have sought to characterize haustoria to answer this question by looking at up or down regulated gene expression profiles in haustorial formation (Wickett *et al.*, 2011; Westwood *et al.*, 2012; Honaas *et al.*, 2013; Jiang *et al.*, 2013; Ranjan *et al.*, 2014). The rationale for this approach is based on the hypothesis that haustorial genes are evolutionarily derived from genes with ancestral function in other tissues. For example, finding a haustorial gene with a homologous gene known to have a specific function in a root suggests a root origin of haustoria.

In this study we have generated a *Cuscuta* transcriptome assembly using tissues of *Cuscuta* stem (2.5cm) near the point of attachment and haustoria (mixed with host tissue but informatically distinguishable as *Cuscuta*). This work was driven by research into host-parasite mRNA mobility (Kim et al., 2014), but also provides a system in which to explore *Cuscuta* gene expression during parasitism. To address the question of the origin of haustorial genes we have compared upregulated *Cuscuta* transcripts found in haustoria to transcription factors with known expression patterns that are specific for either shoot or root tissue of *Arabidopsis* (Czechowski *et al.*, 2004). The differential expression analysis by comparing haustoria and *Cuscuta* stem tissue indicated 522 *Cuscuta* haustorium upregulated transcripts. Among those, we have identified thirteen root specific transcription factor related transcripts. Also, we have detected a few haustorium induced transcripts homologous to shoot-specific transcription factors. From our results, it shows that more root specific transcription factor related transcripts found in haustorial upregulated data than shoot-specific related transcripts. This may indicate function of haustoria

may follow traditional views of haustorium being lateral adventitious roots however, upregulated transcripts similar to shoot-specific transcription factors detected in haustorium also illustrates that the origin of the haustorium is more complex than we think.

Materials and Methods

Plant material

Arabidopsis thaliana (L.) Heynh. and tomato (*Solanum lycopersicum* L. cv. Rutgers) were grown in a growth chamber at 18-20°C with 12-h per day light cycle, illuminated (200 µmol $m^{-2}s^{-1}$) with metal halide (400W, GE multi-vapor lamp) and spot-gro (65W, Sylvania) lamps. *S*eeds of *Cuscuta pentagona* (Engelm.) were scarified by soaking in concentrated sulfuric acid for 45 min, rinsed with water and dried. Seeds were placed in potting medium at the base of four-week-old *Arabidopsis* and two-week-old tomato plants and allowed to germinate and attach to hosts. *Cuscuta* seedlings germinated and established connections with host plants and were allowed to grow and spread on host plants for an additional three weeks. Tissue was collected from *Cuscuta* stems (2.5 cm, near region of attachment), the region of attachment, and host stems (1 cm, above the region of attachment). To ensure uniform, robust connections, tissue was only collected from attachments with a minimum of two *Cuscuta* coils around the host stem and at least 10 cm of *Cuscuta* shoot beyond the attachment region. All three sections of tissue were harvested at the same time and material from three to five attachments were pooled for RNA extraction. To avoid any chance of cross-contamination, tissues were harvested far enough from the point of connection to ensure collection of *Cuscuta* and host stems, and were subsequently rinsed for 10 min in 70% ethanol to remove any surface contamination.

RNA isolation and sequencing

Total RNA was isolated using RNeasy plant mini kit (Qiagen), including a DNase I treatment to remove residual DNA. RNA quality was verified by Agilent Bioanalyzer 2100, and only samples with RNA integrity numbers above 8.0 were used for sequencing. One replicate each of tissues generated with *Arabidopsis* and tomato hosts were sequenced using the Genome Analyzer II (Illumina, Inc.) at the Virginia Bioinformatics Institute. Libraries were generated from poly(A) RNA, with cDNA prepared using random hexamer primers. Each library was sequenced in one full lane to yield 75 bp paired end reads, with two additional control lanes. A second replicate of *Arabidopsis* with *Cuscuta* was sequenced on a HiSeq 2000 (Illumina, Inc.) at the McGill University and Génome Québec Innovation Centre. This produced 100 bp paired end reads, using one full lane per library. The higher number of mobile transcripts identified by the full-lane HiSeq 2000 samples as compared to the GAII platform demonstrates the importance of sequencing depth in characterizing transcripts in mixed tissue samples.

Sequence pre-processing and filtering

In total, about 1.6 billion raw reads from hosts and parasite tissues were analyzed, beginning with quality control using the trim sequences tools (quality score limit: 0.05, trim ambiguous nucleotides: < 2) in CLC genomics workbench (version 5.0.1). Additionally, relatively short reads (< 50 bp) were discarded. Potential human and microbial contaminants were removed by mapping reads to cDNA sequences of human (GRCh37) (Flicek *et al.*, 2013), fungal (http://fungidb.org/fungidb/) (Stajich *et al.*, 2012), and bacterial (http://patricbrc.vbi.vt.edu/portal/portal/patric/Home) (Gillespie *et al.*, 2011) databases using the Map Reads to Reference function (Mismatch cost=1, Insertion cost=1, Deletion cost=1, Length fraction=1, Similarity fraction=0.97). Reads were then filtered against plant transcriptomes to identify highly conserved sequences that would not be expected to distinguish between *Cuscuta*

and its hosts. Species used were: *Brachypodium distachyon, Brassica rapa, Carica papaya, Chlamydomonas reinhardtii, Cucumis sativus, Glycine max, Gossypium hirsutum, Hordeum vulgare, Lotus japonica, Manihot esculenta, Medicago truncatula, Mimulus guttatus, Oryza sativa, Physcomitrella patens, Populus trichocarpa, Prunus persica, Ricinus communis, Selaginella moellendorffi, Setaria italic, Solanum tuberosum, Sorghum bicolor, Triticum aestivum, Vitis vinifera and Zea mays* (www.plantgdb.org, Data retrieved on April, 2012) using same parameter set with Map Reads to Reference function.

Filteration of parasite mRNA reads and *de novo* assembly of the*Cuscuta* transcriptome

All libraries were assumed to contain a mixture of host and parasite sequences and since a *Cuscuta* genome was not available, we assembled a *Cuscuta* transcriptome *de novo* for use in identifying mobile transcripts from *Cuscuta* (Martin & Wang, 2011). We started with 438 million reads from *Cuscuta* growing on *Arabidopsis* and tomato (from which host reads had been filtered out) and used Trinity (version r2013-02-25) (Grabherr *et al.*, 2011) with default parameters (k-mer=25) to assemble contigs. Transcriptome assembly was performed with MAGYK server (Molecular Plant Science at Virginia Tech) using 250 GB memory with 12 core nodes. The assembly command line is: Trinity.pl --seqType fq --JM 250G --Left Cp_all_1.fastq –Right Cp_all_2.fastq –CPU 12.

The primary contigs generated from Trinity were filtered using ESTscan (Iseli *et al.*, 1999) to detect coding regions. Then a clustering algorithm, UCLUST (Edgar, 2010), was applied to consolidate representative sequences from among similar sequences. The post-processed *Cuscuta* unigenes were further analyzed using BLASTX (E-value = 1e-5) against NCBI nr database. The results of plant hits and no hits of *Cuscuta* unigenes were compared to *Arabidopsis* and tomato genomes, as well as the Convolvulaceae database using BLASTn (E-

value = 1e-10) to resolve ambiguous sequences and remove potential host contamination in contig level. The final post-process filtering resulted in a total of 81,794 *Cuscuta* unigenes, The mean transcript length was 1,203 bp and the N50 was 2,039.

<u>Differential expression analysis for *Cuscuta* haustorium induced transcripts</u>

To assess transcript abundance in the haustorium, FPKM values (fragments per kilobase of transcript per million mapped reads) (Trapnell *et al.*, 2010) were calculated using the RNA-seq analysis function in CLC genomics workbench based on mapping reads to the assembled *Cuscuta* transcriptome. *Cuscuta* reads from two *Arabidopsis*-*Cuscuta* (GAIIx and HiSeq2000) interface and tomato-*Cuscuta* interface were mapped to *de novo* contigs as well as reads from three *Cuscuta* stem tissues alone were mapped for estimation of transcript expression. Then differential expression analysis was performed between three interface libraries and three *Cuscuta* tissues alone using the Bioconductor package EdgeR. (Robinson *et al.*, 2010)

<u>GO enrichment analysis for up-regulated *Cuscuta* unigenes in haustorial tissue</u>

The transcripts/unigenes in the haustorial tissue defined as overexpressed in differential analysis ($p < 0.05$) were assigned gene ontology (GO) terms using InterProScan version 46.0 (Jones *et al.*, 2014). GO-terms and GO slim terms of 522 significantly upregulated in hasutorium were generated by using a background of all GO terms of expressed *Cuscuta* transcripts with Fisher's exact test ($p < 0.05$).

Results

<u>*Cuscuta* transcriptome *de novo* assembly, optimization test and evaluation of quality</u>

A total of 1.6 billion RNA-seq reads were generated to obtain sufficient reads for a *de novo* assembly of transcriptome. To maximize the sequencing coverage we initially sequenced

using one library per lane (paired-end reads) using Illumina GAIIx and HiSeq2000 platforms. GAIIx produced an average of 50-60 million raw reads and HiSeq2000 produced 300-400 million reads per lane. The raw reads were processed initially by removing low-quality sequences, adapter and primer sequences using the CLC genomics workbench trimming tool and host (*Arabidopsis* and tomato) sequences were removed by mapping filtered reads to both host transcriptome/genome references. The host filtered reads were further cleaned by mapping to available bacteria, fungi, plant and human sequences to eliminate potential contaminations and identify highly conserved sequences. Also, duplicated reads and reads shorter than 35bp for GAIIx (76bp) and 50bp for HiSeq2000 (100bp) respectively after preprocessing were discarded before assembly. Cleaned reads from each of the libraries were combined and sorted to distinguish unpaired data from data with forward and reverse reads. In total about 435 million paired reads were put into the Trinity *de novo* assembler (version r2013-02-25) with default setting (k-mer: 25) to generate a *Cuscuta* reference transcriptome (Grabherr *et al.*, 2011).

Before we combined all preprocessed reads and made a decision on processing all reads with Trinity *de novo* assembler, we did an experiment evaluating which assembler would provide the best statistical value for *de novo* assembly with our reads. The big challenge for plant transcriptome *de novo* assembly is mis-assembly because of the huge amount of paralogs and various allelic differences in plants (Duan *et al.*, 2012). In order to choose the optimal assembly strategy, we performed *de novo* assembly with three different state-of art de Bruijn assemblers, Trinity (k-mer 25) (Grabherr *et al.*, 2011), CLC bio (k-mer 25) (http://www.clcbio.com) and Soapdenovo-Trans (k-mer 29) (Xie *et al.*, 2014) with the cleaned individual reads from three interface and three *Cuscuta* tissues. Trinity is described to be efficient in generating full-length transcripts and producing many independent de Bruijn graphs to reconstruct transcripts with

alternative splice isoforms. Additionally, Trinity is simple and intuitive, and only uses a single k-mer so requires no parameter modification (Grabherr *et al.*, 2011). CLC bio *de novo* assembler is a commercial brand that many authors have used to publish plant *de novo* transcriptomes, and the software is user friendly and requires no command-line scripts (McKain *et al.*, 2012; Dorn *et al.*, 2013). Soapdenovo-Trans assembler was developed by a BGI (Beijing Genomics Institute) team and the assembler was built on error correcting model algorithm that was used in Trinity assembler. This assembler has been tested with known transcriptomes of rice and mouse to deliver high contig length, lower redundancy and faster execution (Xie *et al.*, 2014). For each of the *de novo* assemblers, we performed six assemblies as follows: TriI_1 (Trinity with Interface tissue 1), TriI_2, TriI_3, TriP_1 (Trinity with Parasite tissue 1), TriP_2, TriP_3, CLCI_1 (CLC bio with Interface tissue 1), CLCI_2, CLCI_3, CLCP_1 (CLC bio with Parasite tissue 1), CLCP_2, CLCP_3, SoapI_1 (Soapdenovo-Trans with Interface tissue 1), SoapI_2, SoapI_3, SoapP_1 (Soapdenovo-Trans with Parasite tissue 1), SoapP_2, and SoapP_3 (table 2.1).

Table 5.1. Summary statistics of transcriptome assemblies generated from *Cuscuta* haustoria (I) and stem (P) tissues.

	I_1	**I_2**	**I_3**	**P_1**	**P_2**	**P_3**
Total raw Reads	52,077,210	342,828,670	55,614,346	52,345,003	325,719,511	54,773,608
Total filtered Reads	32,120,999	206,467,316	12,063,926	50,607,345	276,563,957	46,582,940
Total Bases	2,441,195,924	18,502,209,038	895,189,866	3,846,158,220	26,855,811,933	3,514,018,878
Trinity (k-mer 25)	**TriI_1**	**TriI_2**	**TriI_3**	**TriP_1**	**TriP_2**	**TriP_3**
No. of contigs	84,873	208,395	46,245	97,879	198,392	89,768
Max contig size	7,045	13,824	3,713	6,258	14,805	5,017
Mean contig size	518	1,149	469	600	1,023	525
N50 (bp)	635	1,887	549	783	1,635	653
Total nucleotide length (bp)	43,960,694	239,535,454	21,696,047	58,687,579	203,027,434	47,135,135
Total GC count	19,283,635	96,127,262	9,911,903	25,384,417	81,546,220	20,528,622
GC %	43.87	40.13	45.69	43.25	40.17	43.55
CLC bio (k-mer 25)	**CLCI_1**	**CLCI_2**	**CLCI_3**	**CLCP_1**	**CLCP_2**	**CLCP_3**
No. of contigs	53,756	143,364	32,572	60,705	134,195	58,498
Max contig size	5,318	9,031	3,548	7,783	9,061	3,324
Mean contig size	418	488	399	457	489	410
N50 (bp)	443	550	415	501	548	429
Total nucleotide length (bp)	22,457,997	69,949,707	13,000,865	27,759,790	65,656,015	23,984,136
Total GC count	9,941,052	27,417,051	5,920,668	12,164,636	25,963,484	10,531,636
GC %	44.27	39.20	45.54	43.82	39.54	43.91
Soapdenovo-Trans (k-mer 29)	**SoapI_1**	**SoapI_2**	**SoapI_3**	**SoapP_1**	**SoapP_2**	**SoapP_3**

No. of contigs	59,027	129,812	36,110	65,605	128,429	64,922
Max contig size	4,732	9,224	2,499	6,325	10,508	4,694
Mean contig size	461	613	400	504	596	443
N50 (bp)	526	788	429	600	749	496
Total nucleotide length (bp)	27,219,452	76,637,356	14,426,176	33,091,958	76,561,183	28,758,536
Total GC count	11,901,750	30,670,997	6,583,921	14,341,659	29,654,063	12,492,145
GC %	43.73	38.51	45.64	43.34	38.73	43.44

Contigs < 200bp were not included.

For each assembler, we used identical (Trinity and CLC bio with k-mer 25) or similar (Soapdenovo-Trans with k-mer 29) k-mer settings for comparison between assemblies. For each assembly, the more reads that were used, the N50 value (statistical measure of average length of a set of sequences) (Mundry *et al.*, 2012) and total nucleotide length were larger. The largest N50 from all assemblies was 1,887 bp from the TriI_2 assembly and smallest N50 was 415 bp from CLCI_3. Overall, Trinity assemblies had much larger N50 values than CLCbio and Soapdenovo-Trans andalso had longer nucleotide length compared to other assemblies (fig.2.1). CLCbio (k-mer 25) and Soapdenovo-Trans (k-mer 29) produced similar results in number of contigs, maximum contig length, mean contig length, N50, total nucleotide length and total GC content (fig.2.1).

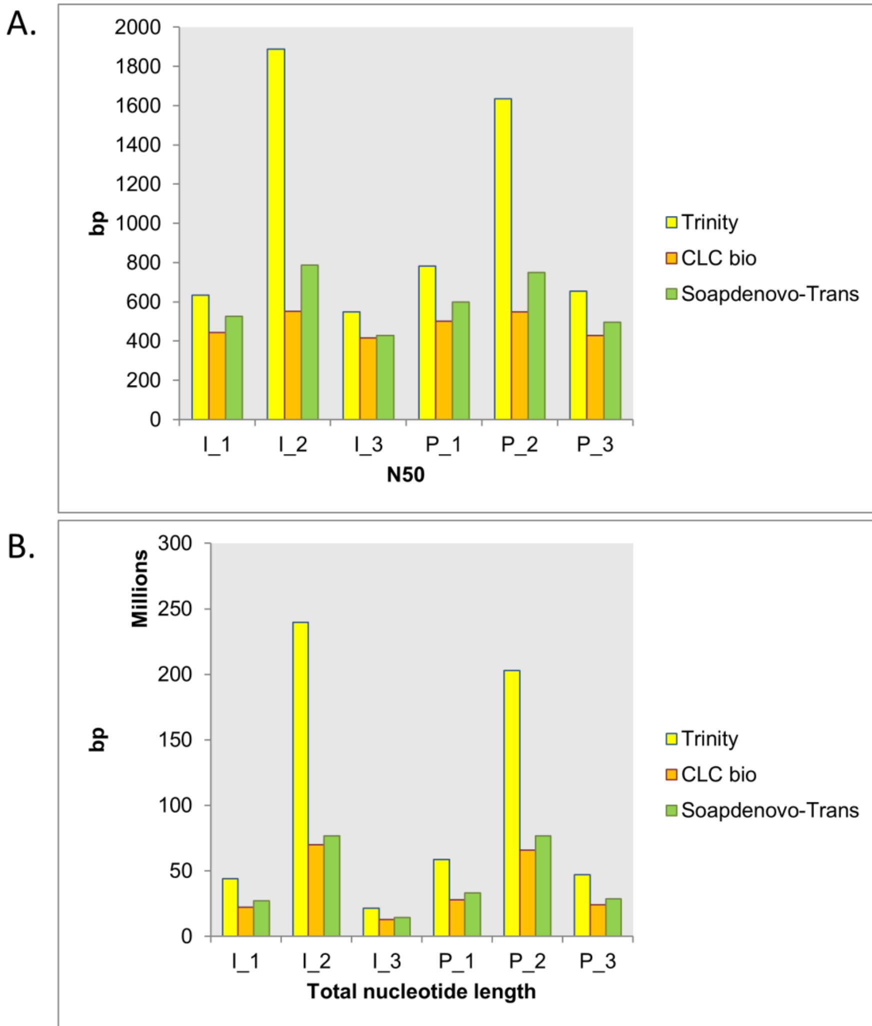


Figure 2.1. N50 (A) and total nucleotide length (B) of three assemblies. Assembly statistics of different assemblers were compared. I_1 – *Cuscuta* reads from interface of *Arabidopsis* and *Cuscuta* (GAIIx), I_2 – *Cuscuta* read from interface of *Arabdiopsis* and *Cuscuta* (HiSeq2000), I_3-*Cuscuta* reads from interface of tomato and *Cuscuta* (GAIIx), P_1 – *Cuscuta* reads from *Cusucta* stem on *Arabidopsis* (GAIIx), P_2 – *Cuscuta* reads from *Cuscuta* stem on *Arabidopsis* (HiSeq2000) and P_3 – *Cuscuta* reads from *Cuscuta* stem on tomato (GAIIx).

Overall, Trinity surpassed CLC bio and Soapdenovo-Trans with our individual cleaned reads from three interface and three parasite stem tissues in assembly continuity. Based on these results we decided to pursue *Cuscuta* assembly with Trinity. All *Cuscuta* reads were combined, resulting in Trinity assembling a total of 252,580 contigs, with N50 and average contig lengths of 1,919 bp and 1,164 bp respectively (table 2.2). The length of contigs ranged from 200 to 13,899 and large proportion fell in the 200-400 bp range (fig 2.2).

Table 2.2. Statistics of combined read sequence assembly using Trinity

Trinity *de novo* (k-mer 25)	
No. of contigs	252,580
Max contig size	13,899
Mean contig size	1,164
N50 (bp)	1,919
Total nucleotide length (bp)	294,045,182
Total GC count	97,329,494
GC %	40.19

Contigs < 200 bp were not included.

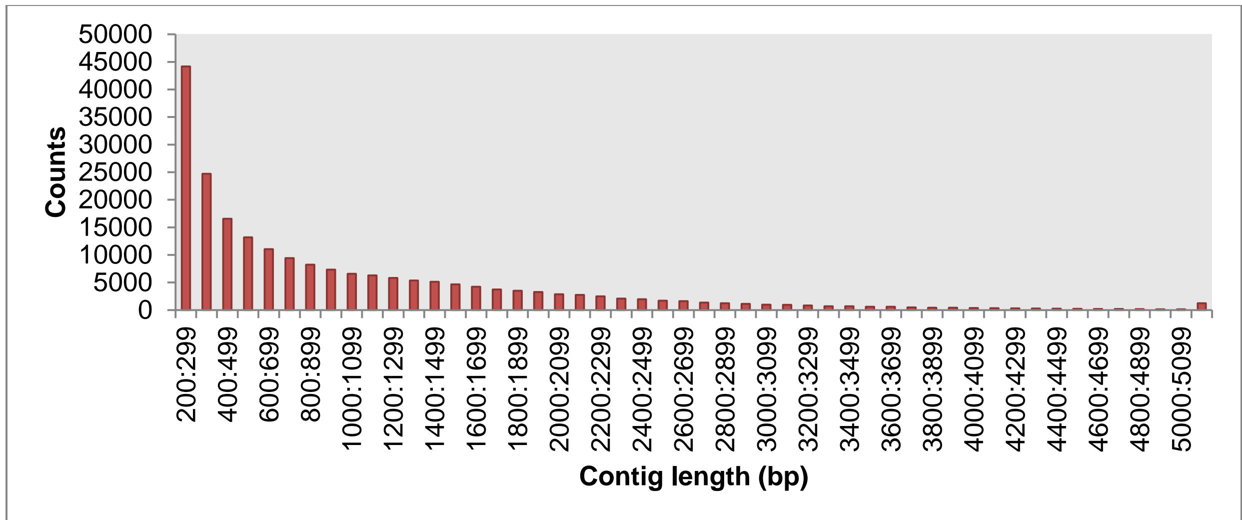


Figure 2.2. Contig length distribution of *Cuscuta* transcriptome generated in Trinity

The 252,580 primary contigs generated from Trinity were filtered using ESTscan (Iseli *et al.*, 1999) to detect 134,881 coding regions. Then a clustering algorithm, UCLUST (Edgar, 2010), was applied and resulted in 89,867 representative sequences. These post-processed *Cuscuta* unigenes were further analyzed using BLASTX (E-value = 1e-5) against NCBI nr database. This resulted in 68,257 *Cuscuta* unigenes having plant hits and 21,147 with no hits. All sequences were pooled and compared to *Arabidopsis* and tomato cDNAs and genomes to eliminate potential host contamination at the contig level. 13,059 sequences were matched to either *Arabidopsis* or tomato genome however, since these sequences are potentially similar to *Cuscuta* species we mapped them to a Convolvulaceae database using BLASTn (E-value = 1e-10). We retained 5,449 sequences that were homologous to Convolvulaceae to avoid discarding authentic *Cuscuta* genes. Although these are likely conserved genes, the read mapping parameters used in determining RNA mobility were considered sufficiently stringent to distinguish host from parasite sequences. The final post-process filtering resulted in a total of 81,794 *Cuscuta* unigenes. The mean transcript length was 1,203 bp and the N50 was 2,039.

Transcript expression abundance Gene ontology (GO) analysis across haustorium and parasite stem tissue

All Illumina reads from the three interface and parasite tissues were aligned to the 81,794 assembled *Cuscuta* unigenes and the number of differentially expressed transcripts were compared with normalized FPKM values. Differential gene expression between haustorium (Interface) and parasite (*Cuscuta* stem) tissues were compared using EdgeR with p-value cutoff equal to 0.05 and log fold change above 2. Comparison of these tissues is valuable because both contain stem tissue, but only the interface samples have haustoria, so gene expression changes between the tissues can be largely attributed to haustoria. We identified 522 significantly up-regulated haustorial associated transcripts as compared to the *Cuscuta* stem (fig 2.3, supplemental data 1). Then GO-slim terms were assigned to haustorial up-regulated transcripts and GO clusters in similar categories were viewed using REVIGO (reduce + visualized gene ontology) (fig 2.4A and B). Biological process included response to oxidative stress, negative regulation of catalytic activity, oxidation-reduction process, cell wall organization, proteolysis, carbohydrate metabolism and molecular function category included peroxidase activity and various genes related to functions in bindings (fig 2.4B)

To address the question of root or shoot origins of haustorium up-regulated genes we compared 522 *Cuscuta* haustorial upregulated unigenes to 1,400 *Arabidopsis* transcription factors that have been characterized by real-time RT-PCR as either root specific or shoot specific and that play critical roles in development of these organs (Czechowski *et al.*, 2004). This work identified 35 root-specific and 52 shoot-specific transcription factors. To find the best homologs between *Cuscuta* and *Arabidopsis* genes, the orthogroup between *Arabidopsis* CDS and *Cuscuta* putative CDS was generated using OrthoMCL. Then *Arabidopsis* root- and shoot-specific genes from the same orthogroup as haustorial up-regulated *Cuscuta* unigenes were used to sort the

Cuscuta unigenes into organ-specific categories. In total, 18 *Cuscuta* unigenes were found in orthogroups of either root- or shoot-specific *Arabidopsis* transcription factors. Among those, 13 *Cuscuta* haustorial induced transcripts were identified as root-specific transcription factors, encoding MYB domain, NAC domain, WRKY DNA-binding protein and telomerase activator 1. Five of 18 *Cuscuta* transcripts were related to shoot transcription factors, which encode MYB and NAC domain proteins (fig 2.5).

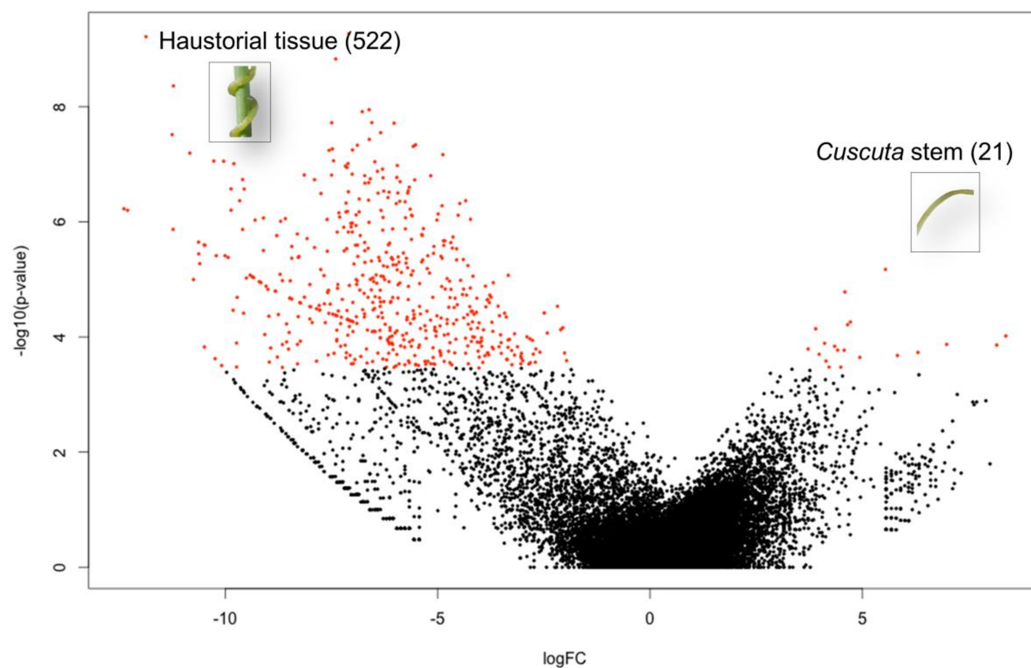


Figure 2.3. Volcano plot illustrating differential *Cuscuta* unigene expression between haustorial tissue (interface) compared to *Cuscuta* stem (parasite). Red dots indicates significantly up-regulated genes. The x-axis represents log of fold change and y-axis represents –log 10 of p-value.

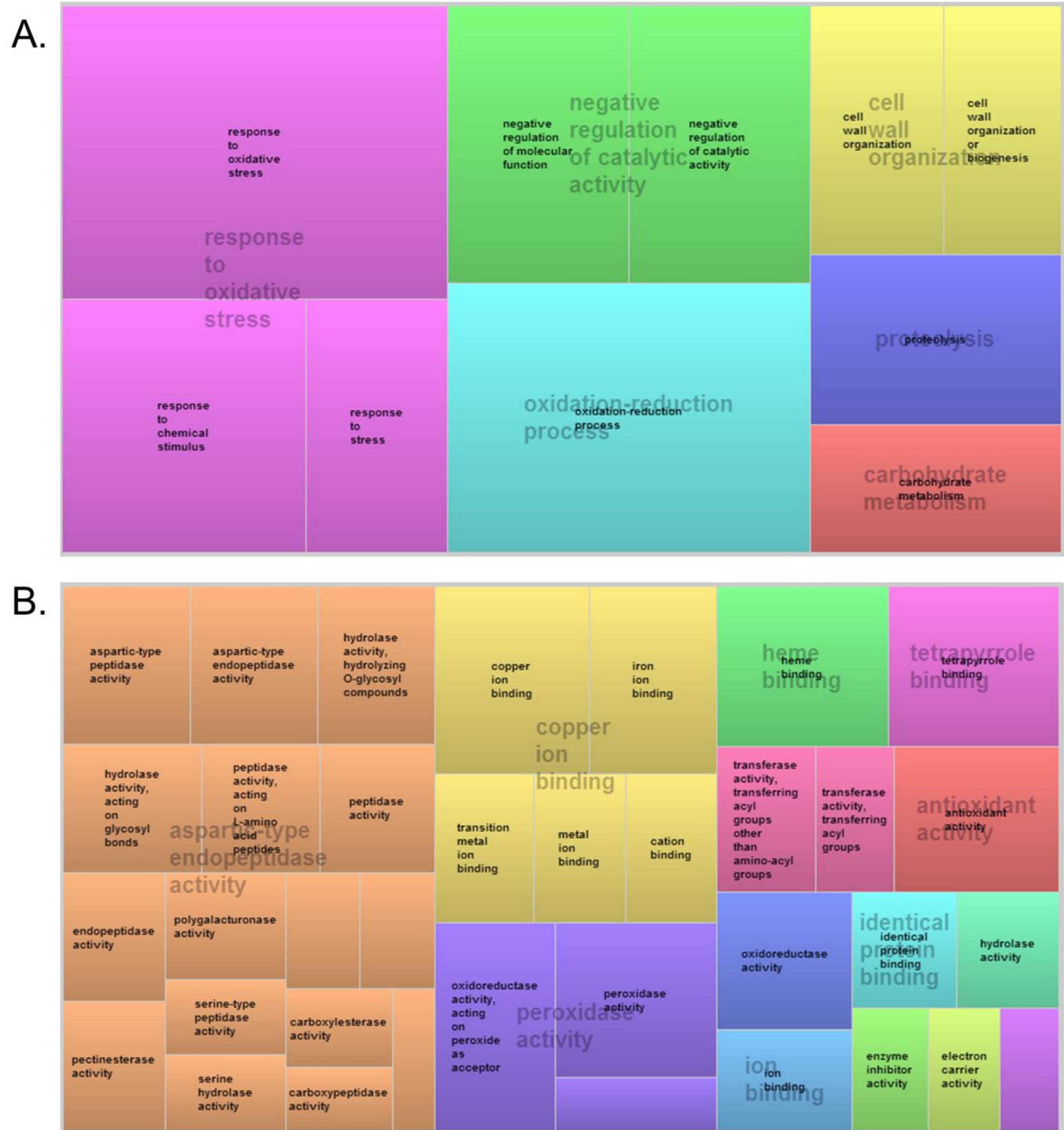


Figure 2.4. GO slim annotation for Biological Process (A) and Molecular function (B) of 522 haustorial up-regulated transcriptome using REVIGO. Color represents similar type of GO hierarchal term based on p-value assign cut-off < 0.05 (Statistical test: Fisher and Multiple test adjustment test: Yekutieli-FDR under dependency).

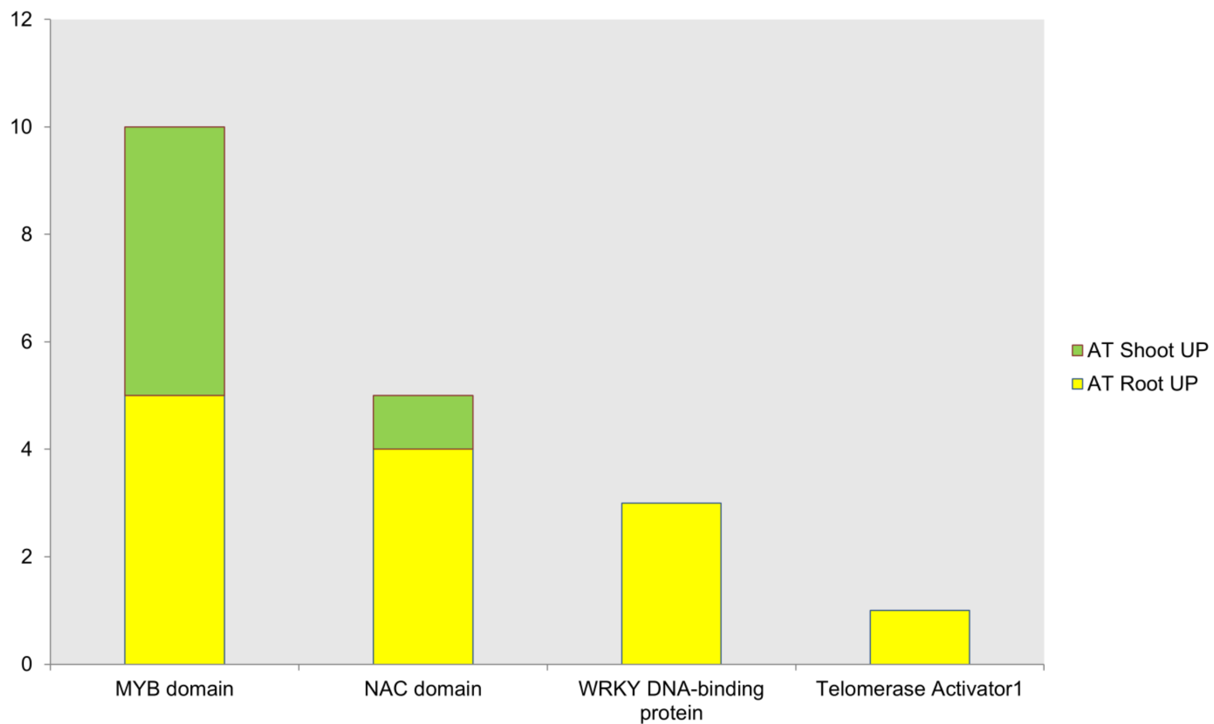


Figure 2.5. Number of *Cuscuta* haustorial associated unigenes orthologous to shoot- or root-specific ***Arabidopsis*** transcription factors that are *Cuscuta* transcripts. Yellow bars represent root up-regulated transcription factors and green bar represents shoot up-regulated transcription factors.

Discussion

Prior to the current work there were only 1,147 known Convolvulacease and 127 known *Cuscuta* plastid sequences in NCBI public database. The original purpose for generating RNA-seq libraries from host, interface and *Cuscuta* tissues was to understand mobile RNA movement in the host-*Cuscuta* system based on identification of RNA translocation in previous studies (Roney *et al.*, 2007; Westwood *et al.*, 2009). In order to confirm mobile RNAs in samples of

mixed sequence, we had to generate *de novo* assembly of a *Cuscuta* transcriptome to distinguish between host and parasite reads. Even given recent advances in bioinformatics, it is still challenging to use massive quantities of short read data from the Illumina platform to produce a reference transcriptome *de novo* that is of sufficient quality for differential expression analysis (Wang *et al.*, 2009; Surget-Groba & Montoya-Burgos, 2010; Schliesky *et al.*, 2012).

In this study, we generated a comprehensive *Cuscuta* transcriptome using the Trinity *de novo* assembler. The number of *de novo* transcriptomes of non-model species generated with high throughput sequencing is rapidly increasing because of the cost-efficiency of the technology. In the plant community, many transcriptomes have been published for non-model plants such as safflower (Huang *et al.*, 2012), rubber tree (*Hevea brasiliensis*) (Xia *et al.*, 2011), St. John's wort (He *et al.*, 2012), coconut (Fan *et al.*, 2013), Eucalyptus tree (Mizrachi *et al.*, 2010), Chili pepper (Liu *et al.*, 2013), garlic (Sun *et al.*, 2012), *Camelina sativa* (Liang *et al.*, 2013), pennycress (Dorn *et al.*, 2013) and etc. These new *de novo* transcriptomes have increased over the past 8 years and newly assembled transcriptomes are frequently being deposited into Genebank. Most of these studies have used popular and user proven assemblers such as Trinity, Oases, Soapdenovo-Trans, Trans-ABySS and CLC bio for generating transcriptome. Some studies only use one assembler with one parameter setting for their *de novo* assembly, while others try to optimize assembly with many different assemblers and various settings. Since non-reference-guided assembly contains insufficient information to confirm the quality of assembly, researchers tend to rely on assembly statistics such as N50 (a statistical measure of average length of a set of sequences. It is used widely in genomics, especially in reference to contig or supercontig lengths within a draft assembly) (Kumar & Blaxter, 2010), total length, mean length, and total output of contigs.

Since no *Cuscuta* genome has been published and publically available EST data for *Cuscuta* species was minimal (only a small number of plastid sequences) when we generated the transcriptome assembly, we adapted the assembly strategy of non-reference guided plant transcriptomes. From the available assemblers, we tested Trinity and Soapdenovo-Trans, which do not require alignment to a reference genome for discovering transcript structure but rather rely solely on RNA-seq data (Grabherr *et al.*, 2011; Xie *et al.*, 2014). Additionally, we also used the commercial brand CLC bio *de novo*.

In our optimization study we observed that Trinity outperformed CLC bio and Soapdenovo-Trans using similar parameter settings. The Trinity assembler produced much higher numbers of N50 and total contig length values. CLC bio and Soapdenovo-Trans gave similar statistical results to each other. However, it is possible that the results of this study would have been different if we had explored different K-mer settings. Trinity is already optimized for using a setting of k-mer 25. Most assembly studies reveal that choice of k-mer size in different assemblers is a huge factor in determining assembly quality. In many cases, low expressed genes are better assembled with a smaller k-mer and highly expressed genes are more effectively assembled with larger k-mer (Robertson *et al.*, 2010). Therefore, our results could have been different if we had tried many different k-mer settings for CLC bio and Soapdenovo-Trans. However, due to the time needed for assemblies and limitations of computing resource availability, we decided to proceed with the Trinity assembly for downstream analysis. In the future, choosing only one best assembler for generating transcriptome may not be necessary, as it is possible to use the best features of multiple assemblies together to produce more complete transcriptomes. Nakasugi et al. (2014) demonstrate high quality transcriptome assembly from

combining initial assemblies with multiple *de novo* assemblers in *Nicotiana benthaminan*a. (Nakasugi *et al.*, 2014).

In addition to using the *Cuscuta* transcriptome to analyze mobile RNA between hosts and parasite (Kim et al., 2014), we used the data to address questions of haustorial-specific gene expression and whether the expressed genes in haustoria can be traced to functional origins in shoots or roots. The haustorium as a whole is a unique structure with remarkable capabilities, but broken down into its component functions (i.e., intrusive growth, vascular connections, nutrient acquisition), reflects processes that occur in other plant tissues (Lee, K. B., 2007). The initiation of *Cuscuta* haustorium was well characterized in detail with *Cuscuta australis* and *Cuscuta japonica* that upper haustorium initials originate and develop from dedifferentiation of extending epidermal cells protruding from *Cuscuta* stem. As upper haustorium matures endophyte primordium (EP) develops to make contact with the hosts. The EP was not observed in all *Cuscuta* species but a similar structure to the EP was reported in *Cuscuta epilinum* and *Cuscuta Pentagona.* Anatomical and histological studies suggested that the haustorium is a modified adventitious root (Truscott, 1958; Kuijt & Toth, 1976). In 1977, Kuijt also suggested that the origin of *Cuscuta* haustoria was a modification of a lateral root (Kuijt, 1977). . A recent study of haustorial upregulated transcript, SHOOT MERISTEMLESS-like (STM), which has not been found to be involved in root development in any species, does not support the adventitious root hypothesis of haustorium origin (Alakonya *et al.*, 2012).

Our finding that haustorium induced genes have homologous transcripts to root and shoot specific transcription factors suggest a more nuanced view that hasutorium genes include not only root up-regulated transcription factors but also shoot up-regulated transcription factors (fig. 2.5). We conclude that genes expressed in *Cuscuta* haustoria reflect evolutionary origin as a

modified adventitious roots, but that genes related to shoot specific transcription factors demonstrate the composit nature of the organ. In addition, the complex structure and function of the haustorium in different stages of host infection such as initiation in upper haustorium, penetration, branching and maturation make it difficult to define its true origin based on a snapshot of gene expression. As Kuijt (1977) suggested, the acceptable view of the origin of the haustorium is modified roots in the morphological perspective, but gene expression data indicate that the derivation of haustorium is more complex.

References

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. *The Plant Cell Online*.

Aly R, Shomin H, Joel D, Steinitz B, Zelcer A, Gal-On A 2007. A new approach to parasitic weed control based on silencing of a key metabolic gene in the parasite. *9th World Congress on Parasitic Plants*. Charlottesville, VA.

Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms D, Mittapalli O. 2011. Transcriptomic signatures of ash (*Fraxinus* spp.) phloem. *PLoS ONE* **6**(1): e16368.

Bailey-Serres J. 1999. Selective translation of cytoplasmic mRNAs in plants. *Trends in Plant Science* **4**(4): 142-148.

Bailey-Serres J, Sorenson R, Juntawong P. 2009. Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends in Plant Science* **14**(8): 443-453.

Baliga BS, Pronczuk AW, Munro HN. 1969. Mechanism of cycloheximide inhibition of protein synthesis in a cell-free system prepared from rat liver. *J Biol Chem* **244**(16): 4480-4489.

Banerjee AK, Chatterjee M, Yu Y, Suh S-G, Miller WA, Hannapel DJ. 2006. Dynamics of a Mobile RNA of Potato Involved in a Long-Distance Signaling Pathway. *Plant Cell* **18**(12): 3443-3457.

Bennet CW. 1944. Studies of dodder tranmission of plant viruses. *Phytopathology* **34**: 905-932.

Birschwilks M, Haupt S, Hofius D, Neumann S. 2006. Transfer of phloem-mobile substances from the host plants to the holoparasite *Cuscuta* sp. *J. Exp. Bot.* **57**: 911-921.

Birschwilks M, Sauer N, Scheel D, Neumann S. 2007. *Arabidopsis thaliana* is a susceptible host plant for the holoparasite *Cuscuta* spec. *Planta* **226**(5): 1231-1241.

Borsics T, Lados M. 2002. Dodder infection induces the expression of a pathogenesis-related gene of the family PR-10 in alfalfa. *J Exp Bot* **53**(375): 1831-1832.

Branco-Price C, Kaiser KA, Jang CJ, Larive CK, Bailey-Serres J. 2008. Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in Arabidopsis thaliana. *Plant J* **56**(5): 743-755.

Branco-Price C, Kawaguchi R, Ferreira RB, Bailey-Serres J. 2005. Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. *Ann Bot* **96**(4): 647-660.

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. 2012. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* **335**(6068): 552-557.

Browning KS. 2004. Plant translation initiation factors: it is not easy to be green. *Biochemical Society Transactions* **32**: 589-591.

Chen F, Mackey AJ, Stoeckert CJ, Roos DS. 2006. OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Research* **34**: D363-D368.

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, et al. 2007. FT protein movement contributes to long-distance signalling in floral induction of Arabidopsis. *Science* **316**: 1030 - 1033.

Cudney DW, Orloff SB, Demason DA. 1993. Effects of Thiazopyr and Trifluralin on Dodder (Cuscuta-Indecora) in Alfalfa (Medicago-Sativa). *Weed Technology* **7**(4): 860-864.

Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK. 2004. Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J* **38**(2): 366-379.

David-Schwartz R, Runo S, Townsley B, Machuka J, Sinha N. 2008. Long-distance transport of mRNA via parenchyma cells and phloem across the host-parasite junction in *Cuscuta*. *New Phytol* **179**(4): 1133-1141.

Dawson JH. 1989. Dodder (Cuscuta Spp) Control in Established Alfalfa (Medicago-Sativa) with Glyphosate and Sc-0224. *Weed Technology* **3**(4): 552-559.

Dawson JH. 1990. Dodder (Cuscuta Spp) Control with Dinitroaniline Herbicides in Alfalfa (Medicago-Sativa). *Weed Technology* **4**(2): 341-348.

Dawson JH, Saghir AR. 1983. Herbicides Applied to Dodder (Cuscuta Spp) after Attachment to Alfalfa (Medicago-Sativa). *Weed Science* **31**(4): 465-471.

Deeken R, Ache P, Kajahn I, Klinkenberg J, Bringmann G, Hedrich R. 2008. Identification of Arabidopsis thaliana phloem RNAs provides a search criterion for phloem-based transcripts hidden in complex datasets of microarray experiments. *Plant J* **55**: 746 - 759.

Doering-Saad C, Newbury H, Couldridge C, Bale J, Pritchard J. 2006. A phloem-enriched cDNA library from Ricinus: insights into phloem function. *J Exp Bot* **57**: 3183 - 3193.

Dohm JC, Minoche AE, Holtgrawe D, Capella-Gutierrez S, Zakrzewski F, Tafer H, Rupp O, Sorensen T, Stracke R, Reinhardt R, et al. 2014a. The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature* **505**(7484): 546-+.

Dohm JC, Minoche AE, Holtgrawe D, Capella-Gutierrez S, Zakrzewski F, Tafer H, Rupp O, Sorensen TR, Stracke R, Reinhardt R, et al. 2014b. The genome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature* **505**(7484): 546-549.

Dorn KM, Fankhauser JD, Wyse DL, Marks MD. 2013. De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *The Plant Journal* **75**(6): 1028-1038.

Dörr I 1990. Sieve elements in haustoria of parasitic angiosperms. In: Behnke H-D, Sjölund R-D eds. *Sieve Elements: Comparative Structure, Induction and Development*. Berlin: Springer, 239-253.

Duan JL, Xia C, Zhao GY, Jia JZ, Kong XY. 2012. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *BMC Genomics* **13**.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19): 2460-2461.

Fan HK, Xiao Y, Yang YD, Xia W, Mason AS, Xia ZH, Qiao F, Zhao SL, Tang HR. 2013. RNA-Seq Analysis of Cocos nucifera: Transcriptome Sequencing and De Novo for Subsequent Functional Genomics Approaches. *PLoS ONE* **8**(3).

Farny NG, Hurt JA, Silver PA. 2008. Definition of global and transcript-specific mRNA export pathways in metazoans. *Genes Dev* **22**(1): 66-78.

Fennoy SL, Nong T, Bailey-Serres J. 1998. Transcriptional and post-transcriptional processes regulate gene expression in oxygen-deprived roots of maize. *Plant Journal* **15**(6): 727-735.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. 2013. Ensembl 2013. *Nucleic Acids Res* **41**(Database issue): D48-55.

Furuhashi T, Furuhashi K, Weckwerth W. 2011. The parasitic mechanism of the holostemparasitic plant Cuscuta. *Journal of Plant Interactions* **6**(4): 207-219.

Gauthier D, Murthy MRV. 1987. Efficacy of Rnase Inhibitors during Brain Polysome Isolation. *Neurochemical Research* **12**(4): 335-339.

Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, et al. 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect Immun* **79**(11): 4286-4298.

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. *Nature Biotechnology* **29**(7): 644-U130.

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, et al. 2013. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet* **45**(1): 51-58.

Haidar MA, Boss WF. 2009. Blue light induced changes in inositol 1,4,5-trisphosphate in Cuscuta campestris seedlings. *Weed Research* **49**(6): 628-633.

Hannapel DJ, Sharma P, Lin T. 2013. Phloem-mobile messenger RNAs and root development. *Front Plant Sci* **4**: 257.

Haupt S, Oparka K, Sauer N, Neumann S. 2001. Macromolecular trafficking between Nicotiana tabacum and the holoparasite Cuscuta relexa. *J Exp Bot* **52**: 173 - 177.

Haupt S, Oparka KJ, Sauer N, Neumann S. 2001. Macromolecular trafficking between *Nicotiana tabacum* and the holoparasite *Cuscuta reflexa*. *J. Exp. Bot.* **52**(354): 173-177.

Haywood V, Yu T, Huang N, Lucas W. 2005. Phloem long-distance trafficking of GIBBERELLIC ACID-INSENSITIVE RNA regulates leaf development. *Plant J* **42**: 49 - 68.

He F, Peltz SW, Donahue JL, Rosbash M, Jacobson A. 1993. Stabilization and ribosome association of unspliced pre-mRNAs in a yeast upf1- mutant. *Proc Natl Acad Sci U S A* **90**(15): 7034-7038.

He M, Wang Y, Hua WP, Zhang Y, Wang ZZ. 2012. De Novo Sequencing of Hypericum perforatum Transcriptome to Identify Potential Genes Involved in the Biosynthesis of Active Metabolites. *PLoS ONE* **7**(7).

Honaas L, Wafula E, Yang Z, Der J, Wickett N, Altman N, Taylor C, Yoder J, Timko M, Westwood J, 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 Biology* **13**(1): 9.

Hosford RMJ. 1967. Transmission of plant viruses by dodder. *Botan. Rev.* **42**: 387-406.

Huang LL, Yang X, Sun P, Tong W, Hu SQ. 2012. The First Illumina-Based De Novo Transcriptome Sequencing and Analysis of Safflower Flowers. *PLoS ONE* **7**(6).

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, et al. 2009. The genome of the cucumber, *Cucumis sativus* L. *Nat Genet* **41**(12): 1275-1281.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**(8): 1534-1550.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**(5924): 218-223.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**(4): 789-802.

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.

Jackson S, Hong Y. 2012. Systemic movement of FT mRNA and a possible role in floral induction. *Frontiers in Plant Science* **3**.

Jiang L, Wijeratne AJ, Wijeratne S, Fraga M, Meulia T, Doohan D, Li Z, Qu F. 2013. Profiling mRNAs of two Cuscuta species reveals possible candidate transcripts shared by parasitic plants. *PLoS ONE* **8**(11): e81389.

Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**(9): 1236-1240.

Juntawong P, Bailey-Serres J. 2012. Dynamic Light Regulation of Translation Status in Arabidopsis thaliana. *Front Plant Sci* **3**: 66.

Juntawong P, Girke T, Bazin J, Bailey-Serres J. 2014. Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc Natl Acad Sci U S A* **111**(1): E203-212.

Juntawong P, Sorenson R, Bailey-Serres J. 2013. Cold shock protein 1 chaperones mRNAs during translation in Arabidopsis thaliana. *Plant J* **74**(6): 1016-1028.

Kawaguchi R, Bailey-Serres J. 2002. Regulation of translational initiation in plants. *Current Opinion in Plant Biology* **5**(5): 460-465.

Kawaguchi R, Bailey-Serres J. 2005. mRNA sequence features that contribute to translational regulation in Arabidopsis. *Nucleic Acids Research* **33**(3): 955-965.

Kawaguchi R, Girke T, Bray EA, Bailey-Serres J. 2004. Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana. *Plant J* **38**(5): 823-839.

Khandal D, Samol I, Buhr F, Pollmann S, Schmidt H, Clemens S, Reinbothe S, Reinbothe C. 2009. Singlet oxygen-dependent translational control in the tigrina-d.12 mutant of barley. *Proceedings of the National Academy of Sciences of the United States of America* **106**(31): 13112-13117.

Kim G, LeBlanc M, Wafula E, dePamphilis C, JH W. 2014. Genomic-Scale Exchange of mRNA between a Parasitic Plant and its Hosts *Science* **345**(August).

Kim M, Canio W, Kessler S, Sinha N. 2001. Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* **293**: 287-289.

Kragler F. 2010. RNA in the phloem: A crisis or a return on investment? *Plant Science* **178**(2): 99-104.

Kuijt J. 1969. *The Biology of Parasitic Flowering Plants*. Berkeley: University of California Press.

Kuijt J. 1977. Haustoria of phanerogamic parasites. *Ann. Rev. Phytopathol.* **17**: 91-118.

Kuijt J, Toth R. 1976. Ultrastructure of angiosperm haustoria: A review. *Ann Bot* **40**(170): 1121-1130.

Kumar S, Blaxter ML. 2010. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* **11**.

LeBlanc M, Kim G, Patel B, Stromberg V, Westwood J. 2013. Quantification of tomato and Arabidopsis mobile RNAs trafficking into the parasitic plant Cuscuta pentagona. *New Phytologist* **200**(4): 1225-1233.

LeBlanc ML, Kim G, Westwood JH. 2012. RNA trafficking in parasitic plant systems. *Frontiers in Plant Science* **3**: 203.

Lee KB. 2007. Structure and development of the upper haustorium in the parasitic flowering plant Cuscuta japonica (Convolvulaceae). *Am J Bot* **94**(5): 737-745.

Lee KB, Lee CD. 1989. The structure and development of the haustorium of *Cuscuta australis*. *Can. J. Bot.* **67**: 2975-2982.

Li C, Gu M, Shi N, Zhang H, Yang X, Osman T, Liu Y, Wang H, Vatish M, Jackson S, et al. 2011. Mobile FT mRNA contributes to the systemic florigen signalling in floral induction. *Sci Rep* **1**: 73.

Liang C, Liu X, Yiu SM, Lim BL. 2013. De novo assembly and characterization of Camelina sativa transcriptome by paired-end sequencing. *BMC Genomics* **14**.

Liu MJ, Wu SH, Chen HM, Wu SH. 2012. Widespread translational control contributes to the regulation of Arabidopsis photomorphogenesis. *Mol Syst Biol* **8**: 566.

Liu SQ, Li WS, Wu YM, Chen CM, Lei JJ. 2013. De Novo Transcriptome Assembly in Chili Pepper (Capsicum frutescens) to Identify Genes Involved in the Biosynthesis of Capsaicinoids. *PLoS ONE* **8**(1).

Lu KJ, Huang NC, Liu YS, Lu CA, Yu TS. 2012. Long-distance movement of Arabidopsis FLOWERING LOCUS T RNA participates in systemic floral regulation. *RNA Biol* **9**(5): 653-662.

Marcone C, Ragozzino A, Seemuller E. 1997. Dodder transmission of alder yellows phytoplasma to the experimental host Catharanthus roseus (periwinkle). *European Journal of Forest Pathology* **27**(6): 347-350.

Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet* **12**(10): 671-682.

Matsuura H, Ishibashi Y, Shinmyo A, Kanaya S, Kato K. 2010. Genome-wide analyses of early translational responses to elevated temperature and high salinity in Arabidopsis thaliana. *Plant Cell Physiol* **51**(3): 448-462.

Matsuura H, Takenami S, Kubo Y, Ueda K, Ueda A, Yamaguchi M, Hirata K, Demura T, Kanaya S, Kato K. 2013. A computational and experimental approach reveals that the 5'-proximal region of the 5'-UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in Arabidopsis. *Plant Cell Physiol* **54**(4): 474-483.

McKain MR, Wickett N, Zhang Y, Ayyampalayam S, McCombie WR, Chase MW, Pires JC, dePamphilis CW, Leebens-Mack J. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *Am J Bot* **99**(2): 397-406.

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

Moeller JR, Moscou MJ, Bancroft T, Skadsen RW, Wise RP, Whitham SA. 2012. Differential accumulation of host mRNAs on polyribosomes during obligate pathogen-plant interactions. *Mol Biosyst* **8**(8): 2153-2165.

Mower J, Stefanovic S, Hao W, Gummow J, Jain K, Ahmed D, Palmer J. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol.* **8**(1): 150.

Mower J, Stefanovic S, Young G, Palmer J. 2004. Plant Genetics: gene transfer from parasitic to host plants. *Nature* **432**: 165-166.

Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PG. 2012. Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. *PLoS ONE* **7**(2): e31410.

Mustroph A, Juntawong P, Bailey-Serres J. 2009. Isolation of Plant Polysomal mRNA by Differential Centrifugation and Ribosome Immunopurification Methods. *Plant Systems Biology* **553**: 109-126.

Nadler-Hassar T, Rubin B. 2003. Natural tolerance of Cuscuta campestris to herbicides inhibiting amino acid biosynthesis. *Weed Research* **43**(5): 341-347.

Nakasugi K, Crowhurst R, Bally J, Waterhouse P. 2014. Combining Transcriptome Assemblies from Multiple De Novo Assemblers in the Allo-Tetraploid Plant Nicotiana benthamiana. *PLoS ONE* **9**(3).

Nicolai M, Roncato MA, Canoy AS, Rouquie D, Sarda X, Freyssinet G, Robaglia C. 2006. Large-scale analysis of mRNA translation states during sucrose starvation in arabidopsis cells identifies cell proliferation and chromatin structure as targets of translational control. *Plant Physiol* **141**(2): 663-673.

Park SH, Chung PJ, Juntawong P, Bailey-Serres J, Kim YS, Jung H, Bang SW, Kim YK, Do Choi Y, Kim JK. 2012. Posttranscriptional control of photosynthetic mRNA decay

under stress conditions requires 3' and 5' untranslated regions and correlates with differential polysome association in rice. *Plant Physiol* **159**(3): 1111-1124.

Parker C, Riches CR. 1993. *Parasitic Weeds of the World: Biology and Control*. Wallingford: CAB International.

Peirce GJ. 1893. On the Structure of the Haustoria of Some Phanerogamic Parasites. *Annals of Botany* **os-7**(3): 291-324.

Press MC, Graves JD, eds. 1995. *Parasitic Plants*. London: Chapman & Hall.

Pscroph. A parasitic plant database.

Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, Hanson J. 2009. Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol* **150**(3): 1356-1367.

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*.

Ribeiro DM, Araujo WL, Fernie AR, Schippers JHM, Mueller-Roeber B. 2012. Translatome and metabolome effects triggered by gibberellins during rosette growth in Arabidopsis. *Journal of Experimental Botany* **63**(7): 2769-2786.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al. 2010. De novo assembly and analysis of RNA-seq data. *Nat Methods* **7**(11): 909-912.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1): 139-140.

Roney JK, Khatibi PA, Westwood JH. 2007. Cross-species translocation of mRNA from host plants into the parasitic plant dodder. *Plant Physiol.* **143**(2): 1037-1043.

Rosado A, Li RX, van de Ven W, Hsu E, Raikhel NV. 2012. Arabidopsis ribosomal proteins control developmental programs through translational regulation of auxin response factors. *Proceedings of the National Academy of Sciences of the United States of America* **109**(48): 19537-19544.

Ross J. 1996. Control of messenger RNA stability in higher eukaryotes. *Trends Genet* **12**(5): 171-175.

Sato S, Tabata S, Hirakawa H, Asamizu E, Shirasawa K, Isobe S, Kaneko T, Nakamura Y, Shibata D, Aoki K, et al. 2012. The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* **485**(7400): 635-641.

Schepetilnikov M, Dimitrova M, Mancera-Martinez E, Geldreich A, Keller M, Ryabova LA. 2013. TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. *EMBO J* **32**(8): 1087-1102.

Schliesky S, Gowik U, Weber APM, Brautigam A. 2012. RNA-seq assembly - are we there yet? *Frontiers in Plant Science* **3**.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* **15**(2): 121-132.

Slaymaker DH, Hoppey CM. 2006. Reduced polysome levels and preferential recruitment of a defense gene transcript into polysomes in soybean cells treated with the syringolide elicitor. *Plant Science* **170**(1): 54-60.

Smith JD, Mescher MC, De Moraes CM. 2013. Implications of bioactive solute transfer from hosts to parasitic plants. *Current Opinion in Plant Biology* **16**(4): 464-472.

Sormani R, Delannoy E, Lageix S, Bitton F, Lanet E, Saez-Vasquez J, Deragon JM, Renou JP, Robaglia C. 2011. Sublethal cadmium intoxication in Arabidopsis thaliana impacts translation at multiple levels. *Plant Cell Physiol* **52**(2): 436-447.

Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, et al. 2012. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res* **40**(Database issue): D675-681.

Sun XD, Zhou SM, Meng FL, Liu SQ. 2012. De novo assembly and characterization of the garlic (Allium sativum) bud transcriptome by Illumina sequencing. *Plant Cell Reports* **31**(10): 1823-1828.

Surget-Groba Y, Montoya-Burgos JI. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* **20**(10): 1432-1440.

Tcherkezian J, Brittis PA, Thomas F, Roux PP, Flanagan JG. 2010. Transmembrane Receptor DCC Associates with Protein Synthesis Machinery and Regulates Translation. *Cell* **141**(4): 632-644.

Tholl D. 2006. Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr Opin Plant Biol* **9**(3): 297-304.

Tomilov AA, Tomilova NB, Wroblewski T, Michelmore R, Yoder JI. 2008. Trans-specific gene silencing between host and parasitic plants. *Plant Journal* **56**(3): 389-397.

Torres M, Tomilov A, Tomilova N, Reagan R, Yoder J. 2005. Pscroph, a parasitic plant EST database enriched for parasite associated transcripts. *BMC Plant Biol.* **5**(1): 24.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals

unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**(5): 511-U174.

Tripodi G. 1970. Localization of tryptophan rich proteins and β-glycerophosphatase activity inCuscuta haustorial cells. *Protoplasma* **71**(1-2): 191-196.

Truscott FH. 1958. On the Regeneration of New Shoots from Isolated Dodder Haustoria. *American Journal of Botany* **45**(3): 169-177.

van Dorst HJM, Peters D. 1974. Some biological observations on pale fruit, a viroid-incited disease of cucumber. *Netherlands Journal of Plant Pathology* **80**(3): 85-96.

Vaughn KC. 2002. Attachment of the parasitic weed dodder to the host. *Protoplasma* **219**(3-4): 227-237.

Vaughn KC. 2003. Dodder hyphae invade the host: a structural and immunocytochemical characterization. *Protoplasma* **220**: 189-200.

Vaughn KC. 2006. Conversion of the searching hyphae of dodder into xylic and phloic hyphae: A cytochemical and immunocytochemical investigation. *Intl J Plant Sci* **167**(6): 1099-1114.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**(1): 57-63.

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**(2): 295-306.

Westwood JH, Roney JK, Khatibi PA, Stromberg VK. 2009. RNA translocation between parasitic plants and their hosts. *Pest Man. Sci.* **65**(5): 533-539.

Westwood JH, Yoder JI, Timko MP, dePamphilis CW. 2010. The evolution of parasitism in plants. *Trends Plant Sci.* **15**(4): 227-235.

Wickett NJ, Honaas LA, Wafula EK, Das M, Huang K, Wu B, Landherr L, Timko MP, Yoder J, Westwood JH, et al. 2011. Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Current Biology* **21**(24): 2098-2104.

Williams AJ, Werner-Fraczek J, Chang IF, Bailey-Serres J. 2003. Regulated phosphorylation of 40S ribosomal protein S6 in root tips of maize. *Plant Physiol* **132**(4): 2086-2097.

Xia ZH, Xu HM, Zhai JL, Li DJ, Luo HL, He CZ, Huang X. 2011. RNA-Seq analysis and de novo transcriptome assembly of Hevea brasiliensis. *Plant Molecular Biology* **77**(3): 299-308.

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**(12): 1660-1666.

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**(7378): 520-524.

Article appears courtesy of *Science* and is made available as an electronic reprint with the permission of the journal. The paper can be found at the following URL on the Science website: http://www.sciencemag.org

Systematic or multiple reproduction or distribution to multiple locations via electronic or other means is prohibited and is subject to penalties under law.

Chapter 3: Genomic-Scale Exchange of mRNA between a Parasitic Plant and its Hosts

Gunjune Kim[1], Megan L. LeBlanc[1], Eric K. Wafula[2], Claude W. dePamphilis[2], and James H. Westwood[1]*

[1]Department of Plant Pathology, Physiology and Weed Science, Virginia Tech, Blacksburg, VA 24060 USA.

[2]Department of Biology and the Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802 USA.

Abstract

Movement of RNAs between cells of a single plant is well documented, but cross-species RNA transfer is largely unexplored. *Cuscuta pentagona* (dodder) is a parasitic plant that forms symplastic connections with its hosts and takes up host mRNAs. We sequenced transcriptomes of *Cuscuta* growing on *Arabidopsis* and tomato hosts to characterize mRNA transfer between species and found that mRNAs move in high numbers and in a bidirectional manner. The mobile transcripts represented thousands of different genes, and nearly half the expressed transcriptome of *Arabidopsis* was identified in *Cuscuta*. These findings demonstrate that parasitic plants can exchange large proportions of their transcriptomes with hosts, providing potential mechanisms for RNA-based interactions between species and horizontal gene transfer.

One Sentence Summary: Transcriptome sequencing of *Cuscuta* and host tissues reveals that the parasite exchanges a large number of transcripts with its hosts.

Main text

Cuscuta species (dodders) are parasitic plants that obtain water and nutrients from their plant hosts using specialized organs termed haustoria. The haustoria of *Cuscuta* develop from the stem of the parasite where it coils around the host, penetrating host tissues and ultimately forming vascular connections (*1, 2*). These connections allow transfer of not only water and nutrients into the parasite, but macromolecules including mRNAs (*3-5*) and proteins (*6*), and even pathogens such as viruses (*7, 8*), viroids (*9*), and phytoplasmas (*10*). Here we characterize the scope and directionality of mRNA movement.

Messenger RNA trafficking between cells regulates plant development (*11, 12*), with potential for controlling processes such as leaf shape (*13, 14*), time of flowering (*15*), tuber formation (*16, 17*), and root growth (*18*). Small RNAs can also act systemically to influence plant development (*19*), and a construct encoding a silencing RNA and expressed in a host plant can silence a *Cuscuta* gene (*20*). Although this last example is from an artificial construct, it supports the idea that RNA movement between separate plant individuals can function as a type of organismal communication (*21*). We used transcriptomics to investigate the RNA transfer between *Cuscuta* and its hosts.

Cuscuta species parasitize a wide range of broad-leaved plants (often simultaneously) and are destructive to crops such as tomato (*Solanum lycopersicum*) (*22*). We grew *Cuscuta* on *Arabidopsis thaliana* and tomato hosts because the sequenced genomes of these species facilitates confident identification of host and parasite transcripts from mixed RNA populations. We harvested three distinct regions for each parasite-host association for analysis (fig. 3.1A). The *Cuscuta* haustorium grows toward the center of the host stem and does not spread systemically inside the host (*2*), so the tissues harvested did not risk inclusion of endophytic

haustorial tissue that could lead to cross contamination of samples. Furthermore, the growth habit of *Cuscuta* allows unattached stem regions to be easily collected separate from the interface regions where haustoria bind tightly to host tissues (figs. 3.1A, 3.S1).

To identify host and parasite mobile transcriptomes, we sequenced cDNA libraries derived from each of the three tissues. The rate of RNA mobility was expected to be low, so the first libraries were sequenced with a full lane of Illumina GAIIx (one run each of *Cuscuta* growing on *Arabidopsis* and tomato hosts). Second and third biological replicates of *Cuscuta* with *Arabidopsis* were sequence with a full lane and one-sixth lane of the higher output HiSeq 2000 platform, respectively. This yielded over 1.6 billion high-quality reads, which were subjected to various controls by which we filtered out contaminating reads, poor quality reads, and trimmed away adapters and primers. Reads were then identified as host or parasite, or too highly conserved to assign (fig. 3.S2). The parasite reads were used to reconstruct a *Cuscuta* transcriptome assembly.

Reads from each library were stringently mapped to host and parasite transcriptomes to estimate RNA movement between the species. *Arabidopsis* read proportions in parasite tissue averaged 1.1% of total mapped reads across the three sequencing runs, while host stems contained 0.6% *Cuscuta* reads (fig. 3.1B). Read mapping in the tomato-*Cuscuta* association suggested somewhat lower rates of transfer, but the pattern was similar to that with *Arabidopsis* with the exception of interface tissue where the greater mass of the tomato stem likely resulted in a higher proportion of reads. Bidirectional mobility in transcript movement is consistent with the long-established ability of *Cuscuta* to transmit viruses between plants bridged by the parasite (*7*), and suggests that *Cuscuta* is capable of transmitting mRNAs between different plants.

Independent confirmation of mobility was shown by RT-PCR amplification and subsequent sequencing of selected transcripts. Mobility of 24 tomato transcripts into *Cuscuta* has been documented this way (*3, 23*), so here we analyzed *Arabidopsis* transcripts moving into *Cuscuta*, and *Cuscuta* transcripts moving into *Arabidopsis* and tomato hosts (fig. 3.S3). Such confirmation is not practical for all mobile transcripts, but the output of read mapping itself produced a compelling picture of RNA transfer (fig. 3.2). The read sequences and coverage from parasite stem tissue closely matched those of the interface tissue, with the exception that mobile mRNAs in the parasite occurred in fully spliced mature form; Introns were only found in libraries derived from host stem or interface tissues.

The diversity of transcripts represented by the mobile reads was determined by high stringency mapping of reads from the three species (*Arabidopsis*, tomato and *Cuscuta*) to their combined reference sequences. The criterion for transcript mobility was set using fragment counts where one fragment represents either a matched pair of reads or a single unpaired read. The threshold for mobile transcripts was set at a mean of four fragments per transcript because this level was found to produce positive RT-PCR confirmation (fig. 3.S3), while eight fragments per transcript was considered strong evidence of mobility. The greatest number of mobile transcripts originated from *Arabidopsis* hosts, with 45% (9,518) of the genes in the expressed *Arabidopsis* transcriptome found in *Cuscuta*, and most of these (5,983) showed strong evidence of mobility (Table 3.1). In contrast, tomato hosts produced substantially fewer mobile transcripts than *Arabidopsis*, with 347 (1.6% of total expressed) detected in the parasite. Part of the difference between tomato and *Arabidopsis* transcript mobility into *Cuscuta* may be attributed to the single sample of tomato-*Cuscuta* sequenced and the lack of deep sequencing from a full lane of HiSeq 2000 data, but even allowing for these differences there appear to be differences in

RNA transfer to the parasite from different host species.

With respect to movement from parasite to host, 8,655 *Cuscuta* unigenes were classified as mobile into *Arabidopsis* stem and 5,973 unigenes showed strong evidence of mobility (Table 1). This is 24% of the 35,614 unigenes expressed in *Cuscuta*. Tomato host uptake of *Cuscuta* transcripts was again lower than that of *Arabidopsis*, with 288 unigenes showing evidence of mobility. The rates of transcript movement between *Cuscuta* and the two hosts were consistent in both directions, with a much freer exchange occurring between *Cuscuta* and *Arabidopsis* than between *Cuscuta* and tomato, suggesting that mechanisms regulating haustorial selectivity may be host-specific.

We asked whether mobile and non-mobile RNAs have distinctive properties that provide insight into mechanisms of mobility. One characteristic common to mobile transcripts was high abundance, as measured by FPKM, in the interface region, and this was especially pronounced for the *Arabidopsis* interaction with *Cuscuta* (fig. 3.3A). FPKM (Fragments Per Kilobase per Million mapped reads (*24*)) was used because it normalizes fragment counts to transcript length and depth of transcriptome sequencing to better estimate transcript levels. The patterns were similar for mobile and non-mobile transcripts in the tomato-*Cuscuta* interaction, although tomato non-mobile transcripts spanned the spectrum from low to high abundance (fig. 3.3B). This indicates that one aspect of transcript mobility is related to their high abundance in the cells near the host-parasite boundary, but this is not the only factor influencing mobility as evidenced by the many transcripts with similar expression levels yet differing mobility.

To consider whether transcript mobility is associated with gene function, we used OrthoMCL to generate orthologous clusters of mobile and non-mobile gene classes that were common to all three species (*25*) (fig. 3.3C). Assigning genes from these clusters to gene

ontology terms led to the identification of terms enriched among mobile and non-mobile classes (Table S2). Restricting the list to those terms that were only enriched for multiple species (e.g., transcripts from both *Arabidopsis* and *Cuscuta*) yielded smaller sets of terms that may reflect core mobile and non-mobile categories (fig. 3.3D). These results demonstrate that mobility can be correlated with gene function, but the mechanistic basis for such correlations remains obscure. For example, a large proportion of mobile transcripts are assigned to the Response to Stimulus term; it is possible that these transcripts are specifically targeted for intercellular mobility, but also possible that characteristics of transcript accumulation or localization in the cytoplasm makes them especially prone to host-parasite exchange.

Further evidence for selective mobility of RNAs comes from plots of transcript abundance in the interface region vs. abundance of the same transcripts in the parasite stem (fig. 3.4). The plot of *Arabidopsis* to *Cuscuta* mobile transcripts showed that levels of most transcripts in the parasite were about one hundredth of those in the interface tissues, indicating that most transcripts follow the same dynamics of movement (fig. 3.4A). However, some host RNAs appear to move more readily into the parasite and occurred at FPKM levels in the parasite nearly equal to those in the interface (seen as outliers above the main group in Figure 3.4). The tomato-*Cuscuta* data showed a more dispersed pattern of mobilities that support the idea that dynamics of movement differ between tomato and *Arabidopsis* hosts (fig. 3.4B).

An unresolved question regarding *Cuscuta* haustoria is the precise route used to acquire material from the host. Substantial physiological evidence points to symplastic connections consistent with direct transfer between phloem tissues of host and parasite (e.g., *8*), but no open phloem connections have been observed (*26*). Rather, *Cuscuta* haustorial cells share plasmodesmata with hosts across chimeric cell walls (*2, 8*) and these have been implicated in

host-parasite mobility of RNA (*4*). The long distance movement of RNAs in the parasite suggest phloem involvement (*4, 5*), but our data indicate that the situation is complex. We compared transcripts moving from *Arabidopsis* into *Cuscuta* to published phloem transcriptome data from *Arabidopsis* and four other species (*27-31*), finding significant associations between the data sets (Table S3). Further analysis using the subset of *Arabidopsis* transcripts with especially high mobility into *Cuscuta* (i.e. those significantly above the mass of data points in fig. 3.4A) indicated correlations with the more robust data sets (*Arabidopsis* and ash), but do not demonstrate a linkage between phloem-associated transcripts and high-mobility into *Cuscuta* (Table S4). Our data also indicate that *Cuscuta* acquires transcripts such as the RUBISCO small subunit and other plastid-encoded mRNAs that are not considered part of an authentic phloem transcriptome (*32*). Taken together, these data suggest that host-parasite RNA exchange includes RNAs known to occur in phloem, but also many RNAs from other cells.

We can only speculate about the significance of large-scale mRNA movement between individuals of different species. For example, some specific mRNAs transmit information long distances in plants (*13, 14, 18, 33*), and these same information molecules could help the parasite track host physiological status or – in the other direction – use its own mRNA to manipulate the host to facilitate parasitism. However, it is not known whether mobile mRNAs act through translation into protein or through another mechanism, so it is unclear whether mRNAs could even function across widely different species. Host mRNAs disappear within several hours inside *Cuscuta* (*5*), but this could be due to processes such as translation into protein or degradation for nucleotide recycling. In this regard it would be interesting to know whether *Cuscuta* can distinguish its own transcripts from those of its hosts.

This widespread exchange of mRNA raises the possibility of horizontal gene transfer (HGT). Given what appears to be a constant exchange of mRNA between *Cuscuta* and its hosts, it is not surprising that cases of HGT involving *Cuscuta* are relatively prevalent (*34-38*). Although most documented cases of HGT in parasitic plants suggest a mechanism involving direct transfer of DNA, at least one case of HGT into a parasitic plant (*Striga hermonthica*) exhibits evidence of an RNA intermediate in the mechanism (*39*). The ability of one *Cuscuta* plant to bridge many different host individuals raises the possibility that this parasite could mediate RNA exchange across different individuals and even across hosts of different species.

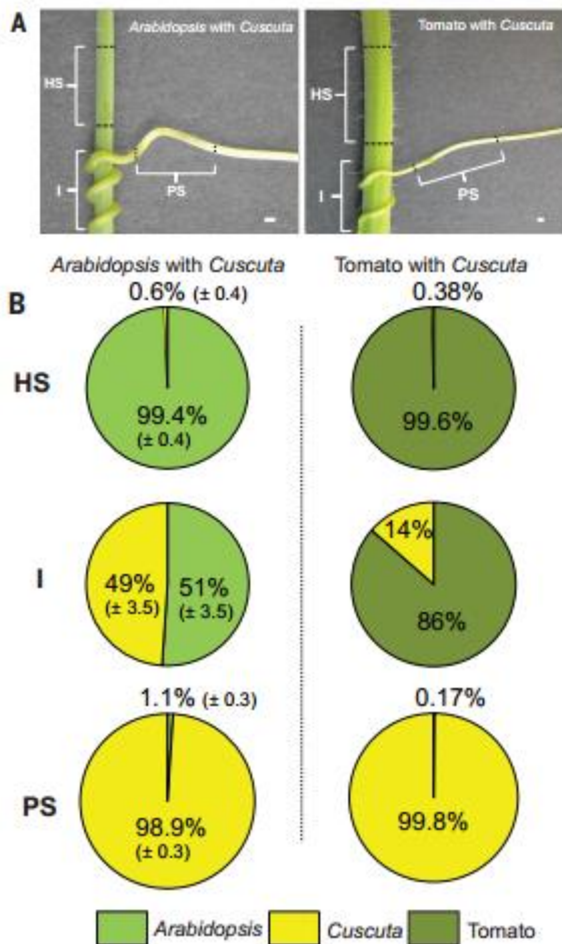


Figure 3.1. Transcriptome compositions of host and parasite tissues at and near the region of haustorial attachment.

(A) Tissues analyzed were the host stem above the region of attachment (HS), interface region where parasite is connected to the host (I), and the parasite stem near the region of attachment (PS). Scale bars represent 1 mm. (B) Pie charts show the proportions of reads mapped to host and parasite transcriptomes in each tissue. *Arabidopsis* with *Cuscuta* data are means (±SE) of three separate sequencing runs; Tomato with *Cuscuta* data are from one run.

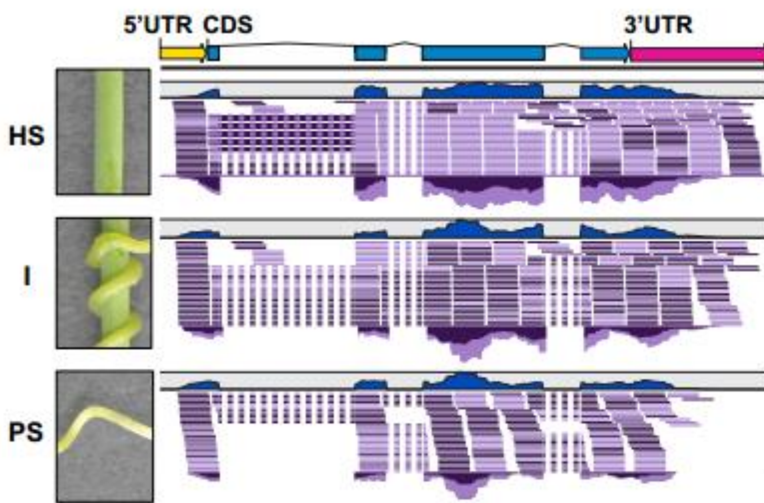


Figure 3.2. Example of read assemblies of an *Arabidopsis* gene, *Translationally Controlled Tumor Protein* (*AtTCTP*), in host stem (HS), interface (I) and parasite stem (PS) tissues.

Intron sequences were not found in sequences derived from parasite tissue. The gene model at top indicates coding sequence as gold bars and introns as line bridges. Red and green lines indicate forward and reverse paired reads, respectively, as mapped to the gene model.

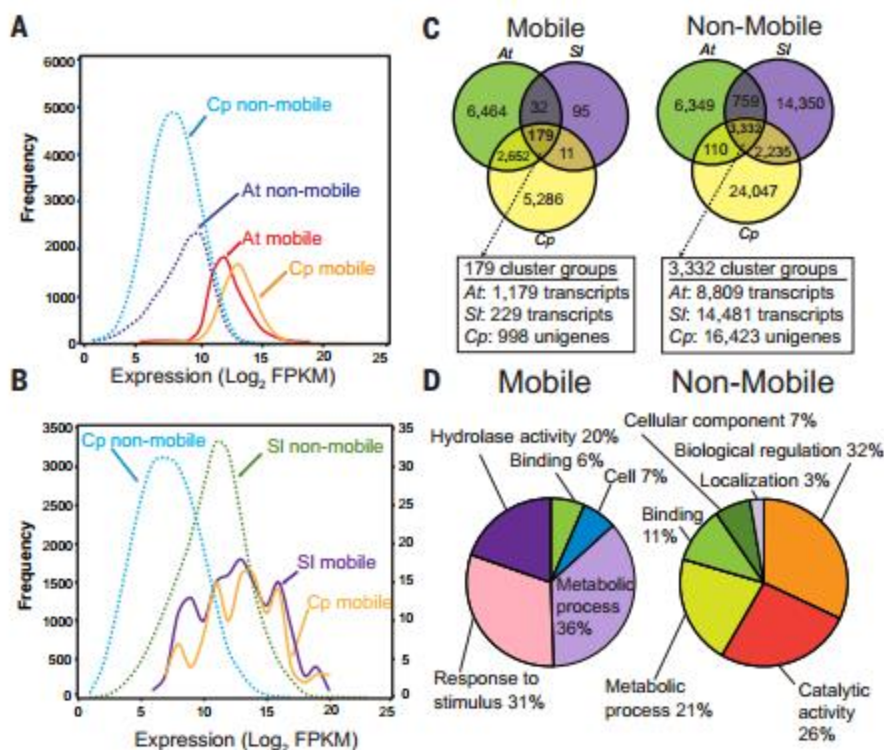


Figure 3.3. Properties of mobile and non-mobile RNAs.

(A) Distribution of transcript expression levels in interface tissue as related to mobility in *Arabidopsis*-*Cuscuta* associations. (B) Same as A, but for mobility in tomato-*Cuscuta* associations. (C) Venn diagrams showing common sets of transcripts that were either mobile or non-mobile out of *Arabidopsis*, tomato or *Cuscuta.* Numbers are orthologous clusters as determined by OrthoMCL. (D) Pie charts showing GO slim terms as proportions of sets of 11 mobile and 23 non-mobile GO terms that were enriched for multiple species. The full lists of GO slim terms for these data sets and all terms significantly over-represented and under-represented in each of the three species are given in table S2.

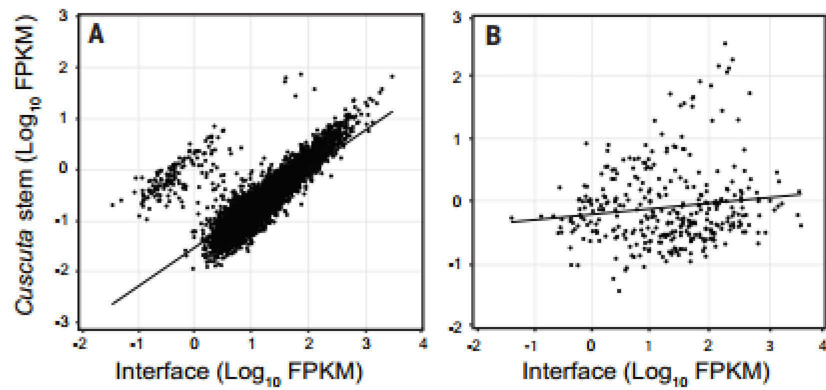


Figure 3.4. Scatterplots of transcript abundance (FPKM) in parasite stem vs. host-parasite interface.

(A) 9,518 *Arabidopsis* transcripts identified as mobile into *Cuscuta* (Table 1). (B) 347 tomato transcripts identified as mobile into *Cuscuta*. Lines are linear regressions of the data.

Table 3.1. Numbers of genes/unigenes with mobile transcripts from hosts into *Cuscuta* or *Cuscuta* into hosts.

The numbers represent transcript reference (TAIR10/ITAG2.4/*Cuscuta* unigenes) sequences as categorized by number of fragments detected in self and non-self tissues. A threshold of four fragments per gene was used to determine transcript detection, with four fragments detected in non-self tissues considered evidence for mobility, and eight fragments providing strong evidence for mobility.

Mobility category	*Arabidopsis-Cuscuta*		*Tomato-Cuscuta*	
	Arabidopsis genes	*Cuscuta* unigenes	Tomato genes	*Cuscuta* unigenes
Total mobile	9,518	8,655	347	288
Mobile (> 8 fragments)	5,983	5,973	147	116
Mobile (> 4 fragments)	3,535	2,682	200	172
Non-mobile	11,874	26,960	21,848	26,717
Total expressed genes/unigenes	21,392	35,614	22,194	27,005

Materials and methods

Plant material

Arabidopsis thaliana (L.) Heynh. and tomato (*Solanum lycopersicum* L. cv. Rutgers) were grown in a growth chamber at 18-20°C with 12-h per day light cycle, illuminated (200 μmol $m^{-2}s^{-1}$) with metal halide (400W, GE multi-vapor lamp) and spot-gro (65W, Sylvania) lamps. Seeds of *Cuscuta pentagona* (Engelm.) were scarified by soaking in concentrated sulfuric acid for 45 min, rinsed with water and dried. Seeds were placed in potting medium at the base of four-week-old *Arabidopsis* and two-week-old tomato plants and allowed to germinate and attach to hosts. *Cuscuta* seedlings germinated and established connections with host plants and were allowed to grow and spread on host plants for an additional three weeks. Tissue was collected as illustrated in Figure 3.1A from *Cuscuta* stems (2.5 cm, near region of attachment), the region of attachment, and host stems (1 cm, above the region of attachment). To ensure uniform, robust connections, tissue was only collected from attachments with a minimum of two *Cuscuta* coils around the host stem and at least 10 cm of *Cuscuta* shoot beyond the attachment region. All three sections of tissue were harvested at the same time and material from three to five attachments were pooled for RNA extraction. Tissues were harvested to avoid any chance of cross-contamination between *Cuscuta* and hosts, and were subsequently rinsed for 10 min in 70% ethanol to remove any surface contamination.

RNA isolation and sequencing

Total RNA was isolated using RNeasy plant mini kit (Qiagen), including a DNase I treatment to remove residual DNA. RNA quality was verified by Agilent Bioanalyzer 2100, and only samples with RNA integrity numbers above 8.0 were used for sequencing. One replicate each of tissues generated with *Arabidopsis* and tomato hosts were sequenced using the Genome Analyzer II (Illumina, Inc.) at the Virginia Bioinformatics Institute. Libraries were generated

from poly(A) RNA, with cDNA prepared using random hexamer primers. Each library was sequenced in one full lane to yield 75 bp paired end reads, with two additional control lanes. A second replicate of *Arabidopsis* with *Cuscuta* was sequenced on a HiSeq 2000 (Illumina, Inc.) at the McGill University and Génome Québec Innovation Centre. This produced 100 bp paired end reads, using one full lane per library. The third replicate of *Arabidopsis* with *Cuscuta* was sequenced on a HiSeq 2000 instrument at Beckman Coulter, Inc. This produced 100 bp paired end reads, with the three libraries multiplexed with nine others such that each library represented one sixth of the lane. The higher number of mobile transcripts identified by the full-lane HiSeq 2000 samples as compared to the GAII platform and multiplexed HiSeq 2000 samples demonstrates the importance of sequencing depth in characterizing transcripts in mixed tissue samples.

Given the importance of sequence data for making conclusions about RNA mobility, we obtained records from each of the sequencing centers on how samples were handled and sequenced. For all centers the libraries were prepared using robotic pipetting systems in which the samples only contact disposable plastics, avoiding risk of cross-contamination. With respect to the sequencing process, the centers flush their instruments with NaOH between runs to negate any carryover from previous runs. Regardless, carryover was not possible for the sequences generated at the Virginia Bioinformatics Institute because they were the first samples (aside from PhiX standards) sequenced on a new instrument in 2010. At the Génome Québec Innovation Centre (2012), the *Arabidopsis* host stem and interface tissue samples were run in separate lanes of one flow cell, while the *Cuscuta* stem tissue sample was prepared and run on a different flow cell and sequencing machine two weeks later. Illumina TruSeq adapters, which contain a unique six base sequence were used to correlate reads with specific libraries. At Beckman Coulter

(2014) the three libraries were prepared with unique barcodes (Illumina TruSeq adapters and indexes) and multiplexed with nine other samples in two lanes to reduce lane-to-lane variation. Only one other plant (rice) was sequenced on this machine in the previous two months. Considering these differences in sequencing location, time and specific protocol, it seems impossible that sequencing center contamination can explain the consistent patterns of *Cuscuta, Arabidopsis, and tomato* RNA mixtures observed in the data.

Sequence pre-processing and filtering

In total, about 1.6 billion raw reads from hosts and parasite tissues were analyzed, beginning with quality control using the trim sequences tools (quality score limit: 0.05, trim ambiguous nucleotides: <2) in CLC genomics workbench (version 5.0.1). Additionally, relatively short reads (< 50 bp) were discarded. Potential human and microbial contaminants were removed by mapping reads to cDNA sequences of human (GRCh37) (*40*), fungal (http://fungidb.org/fungidb/) (*41*), and bacterial (http://patricbrc.vbi.vt.edu/portal/portal/patric/Home) (*42*) databases using the Map Reads to Reference function (Mismatch cost=1, Insertion cost=1, Deletion cost=1, Length fraction=1, Similarity fraction=0.97) (fig. 3.S2). Reads were then filtered against plant transcriptomes to identify highly conserved sequences that would not be expected to distinguish between *Cuscuta* and its hosts. Species used were: *Brachypodium distachyon, Brassica rapa, Carica papaya, Chlamydomonas reinhardtii, Cucumis sativus, Glycine max, Gossypium hirsutum, Hordeum vulgare, Lotus japonica, Manihot esculenta, Medicago truncatula, Mimulus guttatus, Oryza sativa, Physcomitrella patens, Populus trichocarpa, Prunus persica, Ricinus communis, Selaginella moellendorffi, Setaria italic, Solanum tuberosum, Sorghum bicolor, Triticum*

aestivum, Vitis vinifera and Zea mays (www.plantgdb.org, Data retrieved on April, 2012) using same parameter set with Map Reads to Reference function.

Identification of mRNA movement between hosts and parasite

All libraries were assumed to contain a mixture of host and parasite sequences. For confirmation of mRNA movement from hosts to parasite, we furthered mapped pre-processed reads to 1,147 and 127 known sequences of Convolvulaceae and *Cuscuta*, respectively (from GenBank), to eliminate this source of highly conserved sequences between hosts and parasite. All remaining non-*Cuscuta* reads were aligned to *Arabidopsis* genome (TAIR10 cDNA representative model) and tomato genome (ITAG2.4 cDNA) (*43*) to identify host sequences in each of the libraries. Gene detection and expression levels were determined by mapping reads against the host reference genome using the RNA-Seq Analysis function (Mismatch cost=2, Insertion cost=1, Deletion cost=1, Length fraction=1, Similarity fraction=0.97, Maximum number of hits allowed = 1) in CLC genomics workbench using FPKM (fragments per kilobase of exon per million mapped reads). Because all libraries contained mixtures of host and parasite RNA, FPKM values were calculated using combined reference transcriptomes for each species. The cutoff level for mobile transcripts was set at a mean of four fragments (one fragment being equal to one set of paired reads or a single unpaired read). This level was supported empirically by the specific amplification of the *Arabidopsis At5TE15240* transcript from *Cuscuta* tissue, which was detected at the one-fragment level in our transcriptome analysis (fig. 3.S3 and the RT-PCR Validation section below).

Since a *Cuscuta* genome was not available, we assembled a *Cuscuta* transcriptome *de novo* for use in identifying mobile transcripts from *Cuscuta* (fig. 3.S2) (*44*). We started with 438 million reads from *Cuscuta* growing on *Arabidopsis* and tomato (from which host reads had been

filtered out) and used Trinity with default parameters (k-mer=25) to assemble contigs. The primary contigs were filtered using ESTscan (*45*) to detect coding regions. Then a clustering algorithm, UCLUST (*46*), was applied to consolidate representative sequences from among similar sequences. The post-processed *Cuscuta* unigenes were further analyzed using BLASTX (E-value = 1e-5) against NCBI nr database. The results of plant hits and no hits of *Cuscuta* unigenes were compared to *Arabidopsis* and tomato genomes, as well as the Convolvulaceae database using BLASTn (E-value = 1e-10) to resolve ambiguous sequences and remove potential host contamination in contig level. The final post-process filtering resulted in a total of 81,794 *Cuscuta* unigenes, which were used for mRNA mobility analysis. The mean transcript length was 1,203 bp and the N50 was 2,039 (fig. 3.S4, Table S1). The *Cuscuta* reads from various tissues were mapped to these unigenes to identify mobile, putatively mobile, and non-mobile *Cuscuta* unigenes as described above for host transcript movement.

<u>Estimating RNA mobility from read counts</u>

The paired-end and orphan reads generated from pre-processing and filtering were mapped to *Arabidopsis* genome (TAIR10 cDNA representative model), tomato genome (ITAG 2.4 cDNA) and the *Cuscuta* unigenes generated from *de novo* assembly. This mapping was performed with parameters set to optimized counting of uniquely mapped reads using CLC genomics workbench, Map Reads to Reference (Mismatch cost=2, Insertion cost=1, Deletion cost=1, Length fraction=1, Similarity fraction=0.97, Maximum number of hits allowed = 1). Non-specific matched reads were ignored in the counting and uniquely mapped read counts were used to generate the proportion of reads attributable to hosts or parasite in each of the libraries (fig. 3.1B).

RT-PCR validation

RT-PCR was used to validate and confirm *in silico* data of the mobile and non-mobile transcriptomes between hosts and parasite. A set of transcripts spanning a range of high and low read abundance were selected from the mobile transcript data, and primers were designed to specifically amplify these transcripts (Table S5). GoTaq Master Mix (Promega) was used to initially confirm gene specific primer sets. Then SuperScript III one-step RT-PCR with Platinum Taq (Invitrogen) was used to amplify the transcripts. To further verify the identity of PCR products, the bands were directly sequenced with forward and reverse primers. Since previous research confirmed many tomato mobile transcripts, we only designed primers from *Arabidopsis* and *Cuscuta* mobile RNAs.

Analyses of mobile RNA properties

Transcript expression levels ($\log_2$ FPKM) for *Arabidopsis*, tomato and *Cuscuta* in the interface tissue were sorted and grouped into discrete intervals. Frequencies of these groups were then plotted for mobile and non-mobile transcript categories

Sets of common mobile (724) vs. non-mobile (736) cluster groups from all three species were generated using OrthoMCL (*25*). The transcripts/unigenes in the common set of mobile and non-mobile cluster groups were assigned gene ontology (GO) terms using InterProScan version 46.0 (*45*). Then significantly over-represented and under-represented terms were determined by comparison of species specific mobile and non-mobile GO terms to a background of all GO terms of expressed transcripts in a species using Fisher's exact test ($p < 0.05$). The over-represented GO terms (with Bonferroni correction for multiple comparisons, $p < 0.01$) were then used to identify GO terms common to at least two species to find common themes in

mobility. GO Slim terms were assigned to the fine grained GO terms in mobile vs. non-mobile to provide a broader context for specific terms as presented in a pie chart (fig. 3.4D).

Comparison with phloem transcriptomes

Phloem transcriptomes from *Arabidopsis*, ash (*Fraxinus* spp.), castor bean (*Ricinus communis*), cucumber (*Cucumis sativus*) and watermelon (*Ditrullus lanatus*) were obtained from published data (*27-31*). To establish comparable data sets from ash, castor bean, cucumber and watermelon, BLASTx (1e-5) was used to identify a set of transcripts with confident *Arabidopsis* homologs. These were then compared to mobile *Arabidopsis* transcripts from our study to determine whether host-to-parasite mobility was associated with phloem localization of transcripts. Associations were evaluated using the Fisher-Freeman-Halton chi-square test in SPSS version 22 (IBM Corp.). The high- mobility class was determined by dividing the mean FPKM values of mobile transcripts in the *Cuscuta* stem by the FPKM values of the transcript in the interface. Transcripts with a ratio over 0.1 were classified as high-mobility because this represented a ten-fold increase over the mean ratio of 0.01 for all mobile transcripts. Transcripts not meeting the four fragment threshold for mobility were classified as non-mobile.

Figure 3.S1. Overview of parasite attached to host plants.

(A) *Arabidopsis*. (B) Tomato. Arrows indicate shoots of *Cuscuta*.

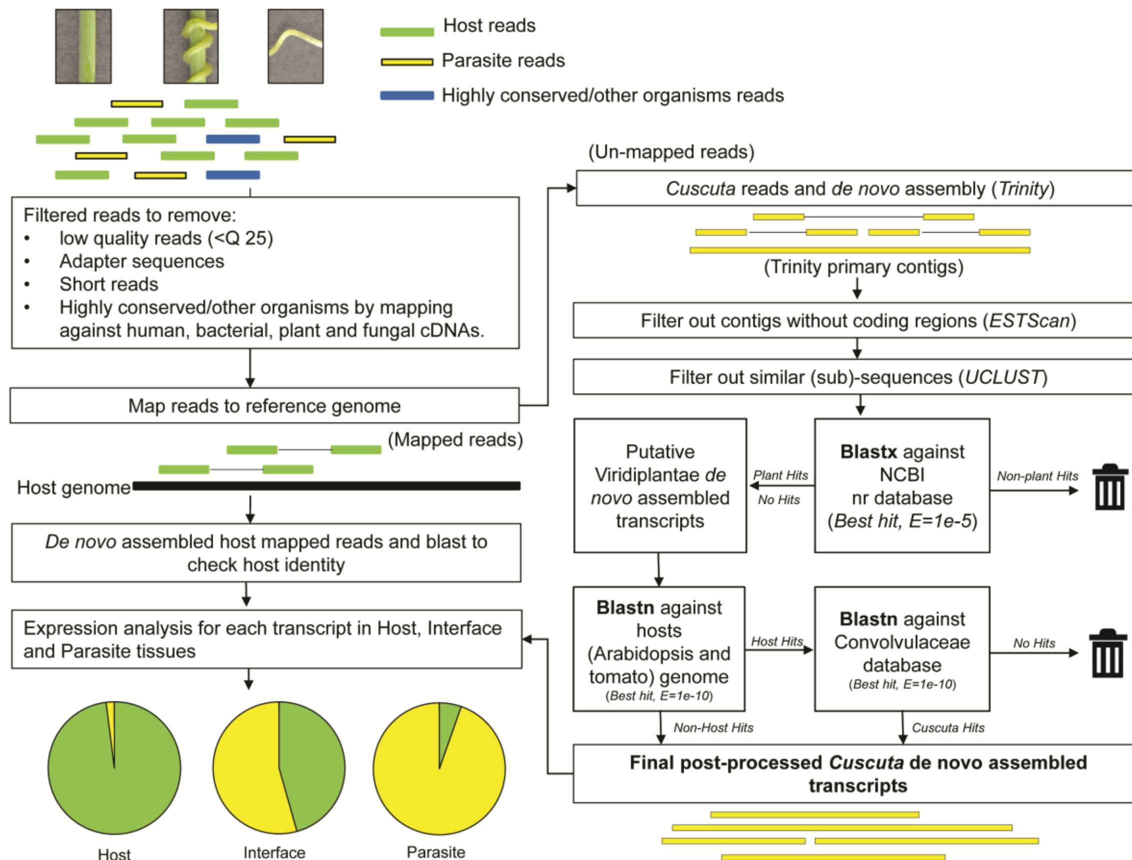
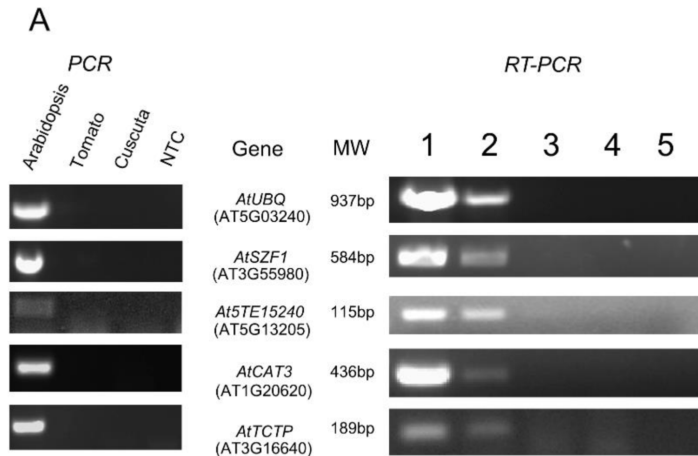


Figure 3.S2. Pipeline of read filtering, mapping, and assembly to distinguish *Cuscuta* transcripts from host transcripts and other contaminants.



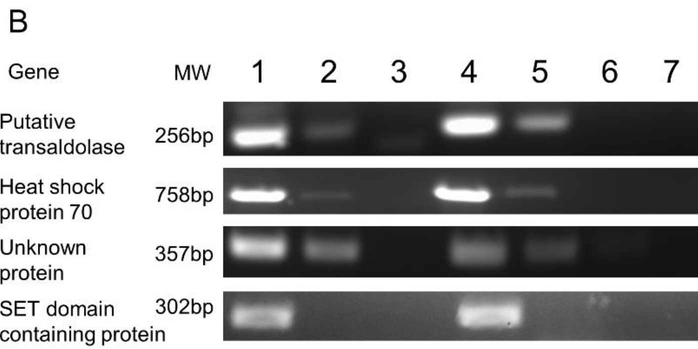


Figure 3.S3. Confirmation of mobility of selected *Arabidopsis* and *Cuscuta* transcripts. (**A**) PCR (left) of *Arabidopsis*, tomato and *Cuscuta* DNA shows specificity of primers for *Arabidopsis*. NTC, no template control. RT-PCR (right) amplifies bands from RNA extracted from *Arabidopsis* and *Cuscuta* growing on *Arabidopsis*, but not from tomato or *Cuscuta* growing on tomato. Lanes are 1, *Arabidopsis* stem; 2, *Cuscuta* growing on *Arabidopsis*; 3, tomato stem; 4, *Cuscuta* growing on tomato; 5, no template control. (**B**) RT-PCR confirming movement of mRNA into hosts for three *Cuscuta* genes, putative transaldolase, heat shock protein 70, and an unknown protein. Primers amplify a band from *Cuscuta* growing on *Arabidopsis* or tomato and from the stem of each host taken near the point of parasite attachment, but no amplification is detected on non-parasitized hosts or no-template control (NTC). A fourth gene, a SET domain containing protein, was among those not detected in the hosts by RNA-seq. For this gene RT-

PCR produced bands only from *Cuscuta* RNA and not associated hosts. Lanes are 1, *Cuscuta* growing on *Arabidopsis*; 2, *Arabidopsis* stem near *Cuscuta* attachment; 3, *Arabidopsis* without *Cuscuta*; 4, *Cuscuta* growing on tomato; 5, tomato stem near *Cuscuta* attachment; 6, tomato without *Cuscuta*; 7, no template control.

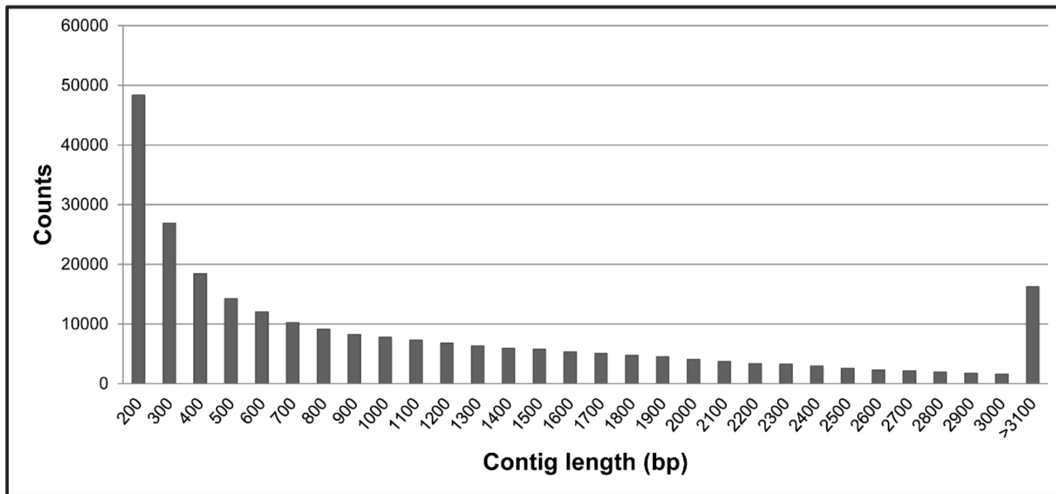


Figure 3.S4. Summary of *Cuscuta* contig length distribution for the *Cuscuta* transcriptome. Contigs below 200bp was not included.

Table 3.S1.

Summary of read sequence assembly for the *Cuscuta* transcriptome. Contigs <200 bp were not included.

	Count (n)	**Total bases (Mbp)**	**Length range (bp)**	**Mean length (bp)**	**N50 (bp)**	**GC (%)**
Clean reads	273,244,032	25,098,868,335	75-100	91.86		
Trinity Contigs	252,580	294,045,182	200-13,899	1,164	1,919	40.10%
Trinity Unigenes	81,794	119,746,844	200-9,820	1,203	2,039	41.96%

Table 3.S2.

Summaries of Gene Ontology (GO) over-representation and under-representation of mobile and non-mobile transcripts from *Arabidopsis*, tomato and *Cuscuta*. Data are organized in sub-tables by mobility class and species of transcript origin. Over-represented (OR) and under-represented (UR) GO terms were determined by Bonferroni corrected p-value < 0.01.

3.S2.1 Gene Ontology enrichment for overrepresentation of mobile transcripts that occur in at least two species. Table provides: GO accession; GO categories - (BP) Biological Process, (CC) Cellular Component, and (MF) Molecular Function; hierarchy GO term; description; numbers in study; numbers in population; Fisher's p-value; Bonferroni correction p-value; *Arabidopsis* core mobile GO accession matched to mobile *Solanum lycopersicum* GO accession; *Arabidopsis* core mobile GO accession matched to mobile *Cuscuta*.

GO accession	GO cat.	Description	Numbers in study	Numbers in pop	p_ fisher	p_ bonferroni	Sl match	Cp match
GO:0005618	CC	cell wall	67	403	3.5E-26	1.5E-24	-	Cuscuta
GO:0030312	CC	external encapsulating structure	67	407	5.7E-26	2.4E-24	-	Cuscuta
GO:0003824	MF	catalytic activity	656	9638	6.8E-105	1.6E-103	-	Cuscuta
GO:0016817	MF	hydrolase activity, acting on acid anhydrides	127	839	3.9E-45	9.3E-44	-	Cuscuta
GO:0016787	MF	hydrolase activity	266	3468	1E-41	2.5E-40	-	Cuscuta
GO:0000166	MF	nucleotide binding	131	2267	2.6E-11	6.3E-10	-	Cuscuta
GO:0050896	BP	response to stimulus	310	4057	3.5E-49	2.5E-47	-	Cuscuta
GO:0006950	BP	response to stress	185	2320	3E-30	2.1E-28	-	Cuscuta
GO:0006629	BP	lipid metabolic process	67	841	2.1E-11	1.5E-09	-	Cuscuta
GO:0009719	BP	response to endogenous stimulus	73	1068	1.6E-09	1.10E-07	-	Cuscuta
GO:0009607	BP	response to biotic stimulus	49	638	3.00E-08	2.20E-06	Tomato	-

3.S2.2 Gene Ontology enrichment for overrepresentation of non-mobile transcripts that occur in at least two species. Table provides: GO accession; GO categories - (BP) Biological Process, (CC) Cellular Component, and (MF) Molecular Function; hierarchy GO term; description; numbers in study; numbers in population; Fisher's p-value; Bonferroni correction p-value; *Arabidopsis* core mobile GO accession matched to mobile *Solanum lycopersicum* GO accession; *Arabidopsis* core mobile GO accession matched to mobile *Cuscuta*.

GO accession	**GO cat.**	**Description**	**Numbers in study**	**Numbers in pop**	**Fisher uncorrected**	**p_ bonferroni**	**Sl match**	**Cp match**
GO:0006350	BP	transcription	625	1923	1.3E-14	9.9E-13	Tomato	Cuscuta
GO:0060255	BP	regulation of macromolecule metabolic process	644	2060	1.9E-12	1.5E-10	Tomato	Cuscuta
GO:0065007	BP	biological regulation	1186	4188	5.6E-12	4.3E-10	Tomato	Cuscuta
GO:0010468	BP	regulation of gene expression	624	2001	6.3E-12	4.8E-10	Tomato	Cuscuta
GO:0050794	BP	regulation of cellular process	977	3375	1.2E-11	9.2E-10	Tomato	Cuscuta
GO:0050789	BP	regulation of biological process	1053	3697	3.7E-11	2.9E-09	Tomato	Cuscuta
GO:0019222	BP	regulation of metabolic process	664	2210	2.6E-10	2.00E-08	Tomato	Cuscuta
GO:0006464	BP	protein modification process	451	1474	3.9E-08	3.00E-06	Tomato	Cuscuta
GO:0043412	BP	macromolecule modification	491	1636	6.9E-08	5.30E-06	Tomato	Cuscuta

GO:0003677	MF	DNA binding	981	2892	1.6E-27	4.8E-26	Tomato	Cuscuta
GO:0003700	MF	transcription factor activity	742	2173	2.1E-21	6.5E-20	Tomato	Cuscuta
GO:0030528	MF	transcription regulator activity	792	2417	4.8E-19	1.5E-17	Tomato	Cuscuta
GO:0000166	MF	nucleotide binding	711	2267	7.3E-14	2.3E-12	Tomato	Cuscuta
GO:0016787	MF	hydrolase activity	997	3468	2.5E-11	7.9E-10	Tomato	Cuscuta
GO:0005215	MF	transporter activity	411	1477	2.20E-04	6.90E-03	Tomato	Cuscuta
GO:0016740	MF	transferase activity	1006	3321	5.5E-16	1.7E-14	Tomato	-
GO:0016772	MF	transferase activity, transferring phosphorus-containing groups	568	1887	4.5E-09	1.40E-07	Tomato	-
GO:0016301	MF	kinase activity	493	1641	5.9E-08	1.80E-06	Tomato	-
GO:0016818	MF	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	259	837	1.50E-05	4.80E-04	Tomato	-
GO:0016817	MF	hydrolase activity, acting on acid anhydrides	259	839	1.80E-05	5.50E-04	Tomato	-
GO:0016462	MF	pyrophosphatase activity	257	832	1.80E-05	5.70E-04	Tomato	-
GO:0017111	MF	nucleoside-triphosphatase activity	244	792	3.40E-05	1.00E-03	Tomato	-
GO:0016020	CC	membrane	1074	4068	5.80E-06	2.70E-04	-	Cuscuta

3.S2.3 Gene Ontology enrichment for 1,179 mobile *Arabidopsis* transcripts into *Cuscuta*. Table provides: GO accession; enrichment - (OR) Over-Represented, (UR) Under-Represented; hierarchy GO term; ratio in study; ratio in population; uncorrected p-value; Bonferroni correction p-value; Holm's p-value; Sidak test p-value.

GO accession	Enrichment	GO Term	ratio_in_study	ratio_in_pop	p_uncorrected	p_bonferroni	p_holm	p_sidak
GO:0006096	OR	glycolysis	192/5946	215/20550	1.64E-10	9.38E-07	9.18E-07	9.14E-07
GO:0009805	OR	coumarin biosynthetic process	42/5946	53/20550	2.17E-11	1.24E-07	1.24E-07	1.21E-07
GO:0022626	OR	cytosolic ribosome	158/5946	189/20550	1.35E-10	7.70E-07	7.59E-07	7.50E-07
GO:0006833	OR	water transport	122/5946	144/20550	1.06E-10	6.07E-07	6.04E-07	5.92E-07
GO:0050832	OR	defense response to fungus	171/5946	335/20550	1.65E-10	9.41E-07	9.20E-07	9.17E-07
GO:0009630	OR	gravitropism	106/5946	152/20550	1.10E-10	6.29E-07	6.24E-07	6.13E-07
GO:0050897	OR	cobalt ion binding	36/5946	46/20550	4.74E-11	2.71E-07	2.71E-07	2.64E-07
GO:0005507	OR	copper ion binding	125/5946	200/20550	1.60E-10	9.12E-07	8.95E-07	8.89E-07
GO:0016558	OR	protein import into peroxisome matrix	58/5946	93/20550	9.50E-11	5.43E-07	5.40E-07	5.29E-07
GO:0006635	OR	fatty acid beta-oxidation	110/5946	169/20550	1.06E-10	6.08E-07	6.04E-07	5.93E-07
GO:0006098	OR	pentose-phosphate shunt	112/5946	189/20550	1.35E-10	7.70E-07	7.59E-07	7.50E-07
GO:0006364	OR	rRNA processing	155/5946	246/20550	1.60E-10	9.13E-07	8.96E-07	8.90E-07
GO:0006972	OR	hyperosmotic response	168/5946	246/20550	1.60E-10	9.13E-07	8.96E-07	8.90E-07

GO:0030003	OR	cellular cation homeostasis	95/5946	158/20550	1.59E-10	9.10E-07	8.93E-07	8.87E-07
GO:0009697	OR	salicylic acid biosynthetic process	107/5946	205/20550	1.62E-10	9.24E-07	9.05E-07	9.01E-07
GO:0030243	OR	cellulose metabolic process	85/5946	144/20550	1.06E-10	6.07E-07	6.04E-07	5.92E-07
GO:0019761	OR	glucosinolate biosynthetic process	112/5946	167/20550	1.45E-10	8.30E-07	8.18E-07	8.09E-07
GO:0009750	OR	response to fructose	109/5946	142/20550	1.52E-10	8.70E-07	8.55E-07	8.48E-07
GO:0006816	OR	calcium ion transport	93/5946	119/20550	1.53E-10	8.76E-07	8.61E-07	8.54E-07
GO:0019344	OR	cysteine biosynthetic process	162/5946	210/20550	1.63E-10	9.32E-07	9.13E-07	9.09E-07
GO:0048767	OR	root hair elongation	108/5946	155/20550	1.65E-10	9.46E-07	9.25E-07	9.22E-07
GO:0010155	OR	regulation of proton transport	52/5946	77/20550	6.27E-11	3.59E-07	3.58E-07	3.50E-07
GO:0010218	OR	response to far red light	65/5946	97/20550	1.02E-10	5.84E-07	5.81E-07	5.69E-07
GO:0019684	OR	photosynthesis, light reaction	85/5946	144/20550	1.06E-10	6.07E-07	6.04E-07	5.92E-07
GO:0009637	OR	response to blue light	76/5946	123/20550	1.60E-10	9.13E-07	8.95E-07	8.90E-07
GO:0016820	OR	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	92/5946	166/20550	1.31E-10	7.49E-07	7.39E-07	7.30E-07
GO:0005730	OR	nucleolus	178/5946	286/20550	1.66E-10	9.48E-07	9.26E-07	9.24E-07
GO:0009684	OR	indoleacetic acid biosynthetic process	77/5946	99/20550	7.41E-11	4.24E-07	4.22E-07	4.13E-07
GO:0000502	OR	proteasome complex	43/5946	53/20550	2.17E-11	1.24E-07	1.24E-07	1.21E-07
GO:0080129	OR	proteasome core complex assembly	105/5946	126/20550	1.43E-10	8.20E-07	8.08E-07	7.99E-07
GO:0043161	OR	proteasome-mediated ubiquitin-dependent protein catabolic	71/5946	123/20550	1.60E-10	9.13E-07	8.95E-07	8.90E-07

		process						
GO:0019288	OR	isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway	136/5946	228/20550	1.31E-10	7.48E-07	7.38E-07	7.29E-07
GO:0005768	OR	endosome	151/5946	252/20550	1.66E-10	9.49E-07	9.28E-07	9.25E-07
GO:0035304	OR	regulation of protein dephosphorylation	79/5946	134/20550	7.35E-11	4.20E-07	4.19E-07	4.09E-07
GO:0009902	OR	chloroplast relocation	67/5946	104/20550	1.15E-10	6.60E-07	6.54E-07	6.43E-07
GO:0010287	OR	plastoglobule	46/5946	68/20550	1.20E-10	6.85E-07	6.77E-07	6.68E-07
GO:0034976	OR	response to endoplasmic reticulum stress	159/5946	335/20550	1.65E-10	9.41E-07	9.20E-07	9.17E-07
GO:0042538	OR	hyperosmotic salinity response	85/5946	157/20550	1.22E-10	6.95E-07	6.86E-07	6.78E-07
GO:0009695	OR	jasmonic acid biosynthetic process	86/5946	132/20550	1.01E-10	5.77E-07	5.74E-07	5.63E-07
GO:0034440	OR	lipid oxidation	115/5946	180/20550	1.46E-10	8.34E-07	8.21E-07	8.13E-07
GO:0008135	OR	translation factor activity, nucleic acid binding	74/5946	123/20550	1.60E-10	9.13E-07	8.95E-07	8.90E-07
GO:0042254	OR	ribosome biogenesis	82/5946	122/20550	1.54E-10	8.80E-07	8.64E-07	8.57E-07
GO:0009646	OR	response to absence of light	30/5946	36/20550	1.81E-11	1.03E-07	1.03E-07	1.01E-07
GO:0006569	OR	tryptophan catabolic process	55/5946	71/20550	1.10E-10	6.27E-07	6.22E-07	6.11E-07
GO:0009749	OR	response to glucose	56/5946	82/20550	1.14E-10	6.50E-07	6.44E-07	6.33E-07
GO:0006487	OR	protein N-linked glycosylation	81/5946	101/20550	5.62E-11	3.21E-07	3.20E-07	3.13E-07
GO:0006636	OR	unsaturated fatty acid biosynthetic process	49/5946	70/20550	7.78E-11	4.45E-07	4.43E-07	4.33E-07
GO:0016117	OR	carotenoid biosynthetic process	65/5946	104/20550	1.15E-10	6.60E-07	6.54E-07	6.43E-07

GO:0015995	OR	chlorophyll biosynthetic process	80/5946	111/20550	1.21E-10	6.89E-07	6.81E-07	6.72E-07
GO:0009963	OR	positive regulation of flavonoid biosynthetic process	74/5946	103/20550	1.03E-10	5.90E-07	5.86E-07	5.75E-07
GO:0006568	OR	tryptophan metabolic process	61/5946	91/20550	1.19E-10	6.80E-07	6.72E-07	6.63E-07
GO:0009694	OR	jasmonic acid metabolic process	95/5946	154/20550	1.47E-10	8.38E-07	8.24E-07	8.17E-07
GO:0009610	OR	response to symbiotic fungus	30/5946	37/20550	6.91E-11	3.95E-07	3.94E-07	3.85E-07
GO:0044242	OR	cellular lipid catabolic process	127/5946	241/20550	1.64E-10	9.39E-07	9.20E-07	9.16E-07
GO:0042744	OR	hydrogen peroxide catabolic process	58/5946	71/20550	1.10E-10	6.27E-07	6.22E-07	6.11E-07
GO:0010119	OR	regulation of stomatal movement	35/5946	46/20550	4.74E-11	2.71E-07	2.71E-07	2.64E-07
GO:0000325	OR	plant-type vacuole	37/5946	43/20550	1.34E-11	7.68E-08	7.68E-08	7.48E-08
GO:0006888	OR	ER to Golgi vesicle-mediated transport	65/5946	102/20550	1.68E-10	9.61E-07	9.61E-07	9.39E-07
GO:0006546	OR	glycine catabolic process	42/5946	53/20550	2.17E-11	1.24E-07	1.24E-07	1.21E-07
GO:0015994	OR	chlorophyll metabolic process	113/5946	171/20550	1.46E-10	8.36E-07	8.23E-07	8.15E-07
GO:0005982	OR	starch metabolic process	127/5946	220/20550	1.64E-10	9.39E-07	9.19E-07	9.16E-07
GO:0007033	OR	vacuole organization	49/5946	57/20550	2.23E-11	1.28E-07	1.27E-07	1.24E-07
GO:0043269	OR	regulation of ion transport	90/5946	149/20550	1.26E-10	7.18E-07	7.08E-07	7.00E-07
GO:0019760	OR	glucosinolate metabolic process	130/5946	199/20550	1.46E-10	8.37E-07	8.24E-07	8.16E-07
GO:0009853	OR	photorespiration	114/5946	157/20550	1.22E-10	6.95E-07	6.86E-07	6.78E-07
GO:0022627	OR	cytosolic small ribosomal subunit	65/5946	95/20550	1.48E-10	8.46E-07	8.32E-07	8.25E-07
GO:0022625	OR	cytosolic large ribosomal subunit	90/5946	112/20550	1.57E-10	8.99E-07	8.82E-07	8.76E-07

GO:0009850	OR	auxin metabolic process	88/5946	128/20550	1.10E-10	6.28E-07	6.23E-07	6.12E-07
GO:0032880	OR	regulation of protein localization	33/5946	40/20550	3.08E-11	1.76E-07	1.76E-07	1.71E-07
GO:0007034	OR	vacuolar transport	94/5946	129/20550	1.36E-10	7.79E-07	7.68E-07	7.59E-07
GO:0009851	OR	auxin biosynthetic process	80/5946	110/20550	1.09E-10	6.21E-07	6.17E-07	6.05E-07
GO:0051049	OR	regulation of transport	104/5946	172/20550	1.57E-10	8.98E-07	8.81E-07	8.75E-07
GO:0000956	OR	nuclear-transcribed mRNA catabolic process	73/5946	104/20550	1.15E-10	6.60E-07	6.54E-07	6.43E-07
GO:0019829	OR	cation-transporting ATPase activity	44/5946	64/20550	4.79E-11	2.74E-07	2.73E-07	2.67E-07
GO:0015399	OR	primary active transmembrane transporter activity	115/5946	217/20550	1.19E-10	6.78E-07	6.71E-07	6.61E-07

3.S2.4 Gene Ontology enrichment for 229 mobile *Solanum lycopersicum* transcripts into *Cuscuta*. Table provides: GO accession; enrichment - (OR) Over-Represented, (UR) Under-Represented; hierarchy GO term; ratio in study; ratio in population; uncorrected p-value; Bonferroni correction p-value; Holm's p-value; Sidak test p-value.

GO accession	Enrichment	GO Term	ratio_in_study	ratio_in_pop	p_uncorrected	p_bonferroni	p_holm	p_sidak
GO:0008152	OR	metabolic process	62/93	5128/12161	2.29E-06	0.00072	0.000711	0.000702
GO:0003824	OR	catalytic activity	64/93	5619/12161	0.0000137	0.00431	0.00418	0.0042
GO:0006629	OR	lipid metabolic process	13/93	368/12161	4.19E-06	0.00132	0.00129	0.00128

GO:0016491	OR	oxidoreductase activity	31/93	1138/12161	1.27E-10	3.99E-08	3.98E-08	3.89E-08
GO:0055114	OR	oxidation-reduction process	30/93	1000/12161	1.41E-10	4.43E-08	4.4E-08	4.32E-08
GO:0005506	OR	iron ion binding	13/93	243/12161	3.7E-08	0.0000116	0.0000115	0.0000113
GO:0016705	OR	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	13/93	350/12161	2.42E-06	0.000759	0.000747	0.00074
GO:0020037	OR	heme binding	8/93	307/12161	0.0000214	0.00673	0.0065	0.00656
GO:0008610	OR	lipid biosynthetic process	11/93	127/12161	5.49E-06	0.00172	0.00168	0.00168
GO:0009607	OR	response to biotic stimulus	13/93	30/12161	2.88E-06	0.000903	0.000886	0.00088
GO:0006952	OR	defense response	8/93	50/12161	0.0000379	0.0119	0.0114	0.0116

3.S2.5 Gene Ontology enrichment for 998 mobile *Cuscuta* unigenes into *Arabidopsis* or *Solanum lycopersicum* hosts. Table provides: GO accession; enrichment - (OR) Over-Represented, (UR) Under-Represented; hierarchy GO term; ratio in study; ratio in population; uncorrected p-value; Bonferroni correction p-value; Holm's p-value; Sidak test p-value.

GO accession	**Enrichment**	**GO Term**	**ratio_in_study**	**ratio_in_pop**	**p_uncorrected**	**p_bonferroni**	**p_holm**	**p_sidak**
GO:0009116	OR	nucleoside metabolic process	39/3221	74/11379	1.17E-05	0.0213	0.0207	0.0207

GO:0005198	OR	structural molecule activity	149/3221	283/11379	1.96E-10	3.57E-07	3.57E-07	3.48E-07
GO:0019752	OR	carboxylic acid metabolic process	102/3221	246/11379	8.00E-06	0.0146	0.0142	0.0142
GO:0005622	OR	intracellular	126/3221	262/11379	2.42E-10	4.41E-07	4.38E-07	4.30E-07
GO:0006886	OR	intracellular protein transport	66/3221	117/11379	2.94E-10	5.34E-07	5.30E-07	5.21E-07
GO:0003735	OR	structural constituent of ribosome	135/3221	247/11379	1.72E-10	3.13E-07	3.13E-07	3.05E-07
GO:0006412	OR	translation	130/3221	243/11379	2.14E-10	3.90E-07	3.89E-07	3.80E-07
GO:0005840	OR	ribosome	135/3221	245/11379	2.34E-10	4.26E-07	4.24E-07	4.15E-07
GO:0043234	OR	protein complex	199/3221	482/11379	5.95E-10	1.08E-06	1.06E-06	1.06E-06
GO:0015992	OR	proton transport	31/3221	55/11379	1.43E-05	0.026	0.0252	0.0253
GO:0046983	UR	protein dimerization activity	17/3221	139/11379	6.73E-06	0.0122	0.012	0.0119
GO:0003964	UR	RNA-directed DNA polymerase activity	10/3221	322/11379	2.19E-10	3.99E-07	3.98E-07	3.89E-07
GO:0006278	UR	RNA-dependent DNA replication	10/3221	322/11379	2.19E-10	3.99E-07	3.98E-07	3.89E-07
GO:0003723	UR	RNA binding	124/3221	596/11379	2.05E-05	0.0372	0.0361	0.0363
GO:0006260	UR	DNA replication	22/3221	376/11379	3.17E-10	5.76E-07	5.70E-07	5.62E-07
GO:0006807	UR	nitrogen compound metabolic process	277/3221	1331/11379	4.66E-10	8.48E-07	8.37E-07	8.26E-07
GO:0006139	UR	nucleobase-containing compound metabolic process	205/3221	1137/11379	4.72E-10	8.58E-07	8.46E-07	8.36E-07
GO:0003676	UR	nucleic acid binding	398/3221	1714/11379	2.96E-07	0.000539	0.000527	0.000526

3.S2.6 Gene Ontology enrichment for 8,809 non-mobile *Arabidopsis* transcripts. Table provides: GO accession; enrichment - (OR) Over-Represented, (UR) Under-Represented; hierarchy GO term; ratio in study; ratio in population; uncorrected p-value; Bonferroni correction p-value; Holm's p-value; Sidak test p-value.

GO accession	Enrichment	GO Term	ratio_in_study	ratio_in_pop	p_uncorrected	p_bonferroni	p_holm	p_sidak
GO:0005739	OR	mitochondrion	1496/11091	2507/20550	1.59E-09	8.79E-06	7.83E-06	8.57E-06
GO:0003700	OR	sequence-specific DNA binding transcription factor activity	828/11091	1295/20550	4.55E-10	2.51E-06	2.32E-06	2.45E-06
GO:0003677	OR	DNA binding	972/11091	1592/20550	3.67E-09	2.03E-05	1.80E-05	1.98E-05
GO:0008283	OR	cell proliferation	121/11091	160/20550	1.61E-08	8.89E-05	7.84E-05	8.67E-05
GO:0048451	OR	petal formation	67/11091	70/20550	1.11E-10	6.13E-07	6.11E-07	5.98E-07
GO:0048453	OR	sepal formation	67/11091	70/20550	1.11E-10	6.13E-07	6.11E-07	5.98E-07
GO:0006275	OR	regulation of DNA replication	123/11091	137/20550	2.19E-10	1.21E-06	1.18E-06	1.18E-06
GO:0010389	OR	regulation of transition of mitotic cell	61/11091	68/20550	3.31E-10	1.83E-06	1.83E-06	1.73E-06
GO:0006270	OR	DNA replication initiation	59/11091	66/20550	9.90E-10	5.47E-06	4.90E-06	5.33E-06
GO:0051567	OR	histone H3-K9 methylation	132/11091	178/20550	3.57E-08	0.000197	0.000173	0.000192
GO:0051726	OR	regulation of cell cycle	193/11091	287/20550	5.19E-06	0.0287	0.0245	0.028
GO:0034968	OR	histone lysine methylation	163/11091	239/20550	8.01E-06	0.0443	0.0376	0.0432

GO:0006260	OR	DNA replication	217/11091	284/20550	1.95E-10	1.08E-06	1.05E-06	1.05E-06
GO:0006261	OR	DNA-dependent DNA replication	152/11091	202/20550	5.78E-10	3.19E-06	2.92E-06	3.11E-06
GO:0000724	OR	double-strand break repair via homologous recombination	58/11091	68/20550	5.98E-08	0.000331	0.00029	0.000322
GO:0016444	OR	somatic cell DNA recombination	30/11091	32/20550	1.17E-06	0.00647	0.00558	0.00631
GO:0051225	OR	spindle assembly	40/11091	46/20550	3.79E-06	0.0209	0.0179	0.0204
GO:0006259	OR	DNA metabolic process	467/11091	731/20550	3.91E-08	0.000216	0.00019	0.000211
GO:0006310	OR	DNA recombination	164/11091	238/20550	2.92E-06	0.0162	0.0139	0.0158
GO:0048449	OR	floral organ formation	109/11091	135/20550	1.59E-10	8.77E-07	8.64E-07	8.55E-07
GO:0006302	OR	double-strand break repair	86/11091	109/20550	6.86E-08	0.000379	0.000332	0.00037
GO:0047134	OR	protein-disulfide reductase activity	56/11091	57/20550	6.34E-11	3.51E-07	3.51E-07	3.42E-07
GO:0005783	UR	endoplasmic reticulum	155/11091	502/20550	3.08E-10	1.70E-06	1.62E-06	1.66E-06
GO:0016020	UR	membrane	1720/11091	4552/20550	1.04E-09	5.76E-06	5.15E-06	5.61E-06
GO:0016125	UR	sterol metabolic process	60/11091	166/20550	4.67E-06	0.0258	0.022	0.0252
GO:0009751	UR	response to salicylic acid	165/11091	456/20550	2.72E-10	1.50E-06	1.44E-06	1.46E-06
GO:0009409	UR	response to cold	159/11091	578/20550	3.02E-10	1.67E-06	1.59E-06	1.63E-06
GO:0009639	UR	response to red or far red light	123/11091	367/20550	3.20E-10	1.77E-06	1.67E-06	1.73E-06
GO:0009753	UR	response to jasmonic acid	145/11091	460/20550	3.29E-10	1.82E-06	1.72E-06	1.77E-06
GO:0046686	UR	response to cadmium ion	47/11091	461/20550	3.40E-10	1.88E-06	1.77E-06	1.83E-06

GO:0009723	UR	response to ethylene	112/11091	338/20550	3.48E-10	1.92E-06	1.81E-06	1.88E-06
GO:0009737	UR	response to abscisic acid	176/11091	567/20550	3.58E-10	1.98E-06	1.86E-06	1.93E-06
GO:0009651	UR	response to salt stress	158/11091	767/20550	4.78E-10	2.64E-06	2.44E-06	2.58E-06
GO:0007623	UR	circadian rhythm	46/11091	137/20550	1.70E-06	0.00938	0.00807	0.00914
GO:0019288	UR	isopentenyl diphosphate biosynthetic process, methylerythritol 4-phosphate pathway	48/11091	228/20550	1.86E-10	1.03E-06	1.01E-06	1.00E-06
GO:0009941	UR	chloroplast envelope	100/11091	528/20550	2.65E-10	1.47E-06	1.40E-06	1.43E-06
GO:0009570	UR	chloroplast stroma	108/11091	578/20550	3.02E-10	1.67E-06	1.59E-06	1.63E-06
GO:0003723	UR	RNA binding	240/11091	679/20550	4.04E-10	2.23E-06	2.08E-06	2.18E-06
GO:0009507	UR	chloroplast	1379/11091	3369/20550	7.89E-10	4.36E-06	3.94E-06	4.25E-06
GO:0000166	UR	nucleotide binding	1314/11091	3029/20550	8.11E-10	4.48E-06	4.04E-06	4.37E-06
GO:0008150	UR	biological_process	10165/11091	19045/20550	1.41E-09	7.81E-06	6.97E-06	7.62E-06
GO:0000956	UR	nuclear-transcribed mRNA catabolic process	12/11091	104/20550	1.48E-10	8.18E-07	8.09E-07	7.97E-07
GO:0048522	UR	positive regulation of cellular process	284/11091	724/20550	4.10E-10	2.27E-06	2.11E-06	2.21E-06
GO:0006396	UR	RNA processing	272/11091	845/20550	4.85E-10	2.68E-06	2.47E-06	2.61E-06
GO:0005515	UR	protein binding	946/11091	2363/20550	6.81E-10	3.76E-06	3.42E-06	3.67E-06
GO:0055114	UR	oxidation-reduction process	603/11091	1409/20550	6.13E-10	3.39E-06	3.09E-06	3.30E-06
GO:0005768	UR	endosome	43/11091	252/20550	2.75E-10	1.52E-06	1.45E-06	1.48E-06
GO:0015031	UR	protein transport	308/11091	1116/20550	4.90E-10	2.71E-06	2.49E-06	2.64E-06

GO:0005525	UR	TP binding	92/11091	236/20550	3.74E-06	0.0207	0.0207	0.0177
GO:0009414	UR	response to water deprivation	112/11091	403/20550	3.95E-10	2.19E-06	2.04E-06	2.13E-06
GO:0003824	UR	catalytic activity	3480/11091	7366/20550	1.05E-09	5.83E-06	5.21E-06	5.68E-06
GO:0055085	UR	transmembrane transport	275/11091	684/20550	3.64E-10	2.01E-06	1.89E-06	1.96E-06
GO:0006811	UR	ion transport	441/11091	1036/20550	4.93E-10	2.72E-06	2.51E-06	2.65E-06
GO:0005886	UR	plasma membrane	1068/11091	2824/20550	8.10E-10	4.48E-06	4.04E-06	4.37E-06
GO:0008270	UR	zinc ion binding	586/11091	1227/20550	7.04E-06	0.0389	0.0331	0.0379
GO:0005622	UR	intracellular	376/11091	850/20550	7.46E-09	4.12E-05	3.65E-05	4.02E-05
GO:0009738	UR	abscisic acid-activated signaling pathway	54/11091	209/20550	1.81E-10	1.00E-06	9.84E-07	9.78E-07
GO:0008152	UR	metabolic process	4305/11091	9820/20550	1.09E-09	6.05E-06	5.40E-06	5.90E-06
GO:0005524	UR	ATP binding	758/11091	1758/20550	7.09E-10	3.92E-06	3.56E-06	3.82E-06
GO:0005737	UR	cytoplasm	1207/11091	2977/20550	7.86E-10	4.35E-06	3.93E-06	4.24E-06
GO:0016772	UR	transferase activity, transferring phosphorus-containing groups	648/11091	1391/20550	1.26E-08	6.94E-05	6.13E-05	6.76E-05
GO:0016301	UR	kinase activity	541/11091	1177/20550	1.68E-08	9.26E-05	8.17E-05	9.03E-05
GO:0005575	UR	cellular_component	10360/11091	19415/20550	4.54E-10	2.51E-06	2.32E-06	2.45E-06
GO:0016740	UR	transferase activity	1245/11091	2612/20550	8.25E-10	4.56E-06	4.11E-06	4.44E-06
GO:0009058	UR	biosynthetic process	1641/11091	4238/20550	1.01E-09	5.57E-06	4.99E-06	5.43E-06
GO:0030170	UR	pyridoxal phosphate binding	22/11091	88/20550	3.85E-08	0.000213	0.000187	0.000207

GO:0005794	UR	golgi apparatus	264/11091	842/20550	3.60E-10	1.98E-06	1.99E-06	1.86E-06
GO:0009617	UR	response to bacterium	131/11091	475/20550	3.45E-10	1.91E-06	1.80E-06	1.86E-06
GO:0016021	UR	integral component of membrane	335/11091	732/20550	6.73E-06	0.0372	0.0317	0.0363
GO:0006631	UR	fatty acid metabolic process	141/11091	427/20550	3.47E-10	1.92E-06	1.80E-06	1.87E-06
GO:0006499	UR	N-terminal protein myristoylation	202/11091	470/20550	1.65E-06	0.0091	0.00783	0.00887
GO:0048046	UR	apoplast	104/11091	348/20550	2.16E-10	1.20E-06	1.16E-06	1.17E-06
GO:0005618	UR	cell wall	175/11091	541/20550	4.33E-10	2.39E-06	2.22E-06	2.33E-06
GO:0006508	UR	proteolysis	262/11091	777/20550	4.85E-10	2.68E-06	2.47E-06	2.61E-06
GO:0006457	UR	protein folding	131/11091	324/20550	9.07E-07	0.00501	0.00433	0.00489
GO:0016192	UR	vesicle-mediated transport	128/11091	548/20550	3.19E-10	1.76E-06	1.67E-06	1.72E-06
GO:0008565	UR	protein transporter activity	38/11091	127/20550	4.94E-08	0.000273	0.000239	0.000266
GO:0050665	UR	hydrogen peroxide biosynthetic process	11/11091	76/20550	5.82E-11	3.22E-07	3.22E-07	3.14E-07
GO:0016926	UR	protein desumoylation	14/11091	79/20550	6.81E-11	3.77E-07	3.76E-07	3.67E-07
GO:0010228	UR	vegetative to reproductive phase transition of meristem	89/11091	281/20550	2.38E-10	1.31E-06	1.27E-06	1.28E-06
GO:0005215	UR	transporter activity	477/11091	1101/20550	4.65E-10	2.57E-06	2.37E-06	2.51E-06
GO:0005975	UR	carbohydrate metabolic process	522/11091	1763/20550	6.37E-10	3.52E-06	3.21E-06	3.43E-06
GO:0043169	UR	cation binding	1069/11091	2335/20550	7.34E-10	4.06E-06	3.68E-06	3.96E-06
GO:0009506	UR	plasmodesma	194/11091	807/20550	3.96E-10	2.19E-06	2.04E-06	2.13E-06
GO:0050832	UR	defense response to fungus	106/11091	335/20550	2.97E-10	1.64E-06	1.56E-06	1.60E-06

GO:0010200	UR	response to chitin	146/11091	416/20550	3.90E-10	2.16E-06	2.02E-06	2.10E-06
GO:0005507	UR	copper ion binding	55/11091	200/20550	1.78E-10	9.84E-07	9.66E-07	9.59E-07
GO:0006886	UR	intracellular protein transport	262/11091	1011/20550	4.69E-10	2.60E-06	2.39E-06	2.53E-06
GO:0017111	UR	nucleoside-triphosphatase activity	320/11091	772/20550	4.30E-10	2.38E-06	2.20E-06	2.32E-06
GO:0016887	UR	ATPase activity	180/11091	462/20550	4.83E-10	2.67E-06	2.46E-06	2.60E-06
GO:0006810	UR	transport	1154/11091	3013/20550	7.41E-10	4.10E-06	3.71E-06	3.99E-06
GO:0042626	UR	ATPase activity, coupled to transmembrane movement of substances	59/11091	165/20550	2.96E-06	0.0164	0.014	0.016
GO:0009646	UR	response to absence of light	6/11091	36/20550	5.48E-06	0.0303	0.0258	0.0295
GO:0009793	UR	embryo development ending in seed dormancy	183/11091	473/20550	3.14E-10	1.73E-06	1.64E-06	1.69E-06
GO:0000394	UR	RNA splicing, via endonucleolytic cleavage and ligation	45/11091	140/20550	2.22E-07	0.00123	0.00107	0.0012
GO:0016051	UR	carbohydrate biosynthetic process	248/11091	831/20550	4.40E-10	2.44E-06	2.25E-06	2.37E-06
GO:0009505	UR	plant-type cell wall	89/11091	238/20550	2.72E-07	0.00151	0.00131	0.00147
GO:0009790	UR	embryo development	208/11091	534/20550	3.37E-10	1.86E-06	1.75E-06	1.81E-06
GO:0006511	UR	ubiquitin-dependent protein catabolic process	99/11091	354/20550	3.08E-10	1.70E-06	1.61E-06	1.66E-06
GO:0009416	UR	response to light stimulus	417/11091	1113/20550	4.67E-10	2.58E-06	2.39E-06	2.52E-06
GO:0016773	UR	phosphotransferase activity, alcohol group as acceptor	502/11091	1098/20550	2.00E-08	0.00011	9.73E-05	0.000108
GO:0009536	UR	plastid	1405/11091	3443/20550	8.72E-10	4.82E-06	4.34E-06	4.70E-06
GO:0042542	UR	response to hydrogen peroxide	86/11091	256/20550	2.60E-10	1.44E-06	1.38E-06	1.40E-06

GO:0009644	UR	response to high light intensity	81/11091	217/20550	9.44E-07	0.00522	0.00451	0.00509
GO:0009408	UR	response to heat	117/11091	289/20550	4.22E-06	0.0233	0.0199	0.0227
GO:0001510	UR	RNA methylation	30/11091	178/20550	2.82E-10	1.56E-06	1.49E-06	1.52E-06
GO:0006633	UR	fatty acid biosynthetic process	63/11091	177/20550	9.76E-07	0.0054	0.00466	0.00526
GO:0016818	UR	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	342/11091	809/20550	4.17E-10	2.30E-06	2.14E-06	2.24E-06
GO:0016817	UR	hydrolase activity, acting on acid anhydrides	346/11091	819/20550	4.18E-10	2.31E-06	2.15E-06	2.25E-06
GO:0006139	UR	nucleobase-containing compound metabolic process	1127/11091	2645/20550	7.82E-10	4.32E-06	3.90E-06	4.21E-06
GO:0005829	UR	cytosol	244/11091	1448/20550	5.01E-10	2.77E-06	2.55E-06	2.70E-06
GO:0022857	UR	transmembrane transporter activity	372/11091	857/20550	7.94E-10	4.39E-06	3.96E-06	4.28E-06
GO:0009873	UR	ethylene-activated signaling pathway	26/11091	109/20550	2.00E-10	1.10E-06	1.08E-06	1.08E-06
GO:0006955	UR	immune response	163/11091	538/20550	2.91E-10	1.61E-06	1.53E-06	1.57E-06
GO:0003924	UR	TPase activity	34/11091	114/20550	2.22E-07	0.00123	0.00123	0.00107
GO:0016614	UR	oxidoreductase activity, acting on CH-OH group of donors	73/11091	193/20550	8.12E-06	0.0449	0.0381	0.0438
GO:0006629	UR	lipid metabolic process	640/11091	1464/20550	6.16E-10	3.40E-06	3.10E-06	3.32E-06
GO:0043231	UR	intracellular membrane-bounded organelle	7018/11091	13495/20550	1.04E-09	5.76E-06	5.15E-06	5.61E-06
GO:0005730	UR	nucleolus	65/11091	286/20550	2.24E-10	1.24E-06	1.20E-06	1.21E-06
GO:0048519	UR	negative regulation of biological process	481/11091	1116/20550	4.90E-10	2.71E-06	2.49E-06	2.64E-06

GO:0035556	UR	intracellular signal transduction	195/11091	519/20550	3.36E-10	1.86E-06	1.75E-06	1.81E-06
GO:0048767	UR	root hair elongation	28/11091	155/20550	1.96E-10	1.09E-06	1.06E-06	1.06E-06
GO:0009611	UR	response to wounding	73/11091	326/20550	2.37E-10	1.31E-06	1.27E-06	1.28E-06
GO:0006970	UR	response to osmotic stress	174/11091	820/20550	4.34E-10	2.40E-06	2.22E-06	2.34E-06
GO:0008610	UR	lipid biosynthetic process	416/11091	952/20550	5.26E-10	2.91E-06	2.67E-06	2.84E-06
GO:0000502	UR	proteasome complex	4/11091	53/20550	1.83E-10	1.01E-06	9.91E-07	9.85E-07
GO:0016558	UR	protein import into peroxisome matrix	13/11091	93/20550	7.68E-11	4.25E-07	4.24E-07	4.14E-07
GO:0005777	UR	peroxisome	41/11091	165/20550	1.51E-10	8.32E-07	8.23E-07	8.11E-07
GO:0009880	UR	embryonic pattern specification	7/11091	44/20550	3.29E-07	0.00182	0.00158	0.00177
GO:0005773	UR	vacuole	93/11091	527/20550	3.41E-10	1.89E-06	1.78E-06	1.84E-06
GO:0006098	UR	pentose-phosphate shunt	39/11091	189/20550	1.36E-10	7.54E-07	7.49E-07	7.35E-07
GO:0015995	UR	chlorophyll biosynthetic process	12/11091	111/20550	1.71E-10	9.43E-07	9.27E-07	9.19E-07
GO:0009695	UR	jasmonic acid biosynthetic process	26/11091	132/20550	1.81E-10	1.00E-06	9.83E-07	9.77E-07
GO:0009072	UR	aromatic amino acid family metabolic process	44/11091	234/20550	2.11E-10	1.17E-06	1.14E-06	1.14E-06
GO:0000096	UR	sulfur amino acid metabolic process	53/11091	342/20550	2.15E-10	1.19E-06	1.15E-06	1.16E-06
GO:0006733	UR	oxidoreduction coenzyme metabolic process	58/11091	255/20550	2.32E-10	1.28E-06	1.24E-06	1.25E-06
GO:0006364	UR	rRNA processing	42/11091	246/20550	2.60E-10	1.44E-06	1.38E-06	1.40E-06
GO:0044272	UR	sulfur compound biosynthetic process	92/11091	480/20550	3.15E-10	1.74E-06	1.65E-06	1.70E-06
GO:0019748	UR	secondary metabolic process	188/11091	594/20550	3.97E-10	2.20E-06	2.05E-06	2.14E-06

GO:0008652	UR	cellular amino acid biosynthetic process	118/11091	562/20550	3.98E-10	2.20E-06	2.05E-06	2.14E-06
GO:0009117	UR	nucleotide metabolic process	212/11091	701/20550	4.57E-10	2.53E-06	2.33E-06	2.46E-06
GO:0006464	UR	cellular protein modification process	1066/11091	2395/20550	7.46E-10	4.13E-06	3.74E-06	4.02E-06
GO:0006636	UR	unsaturated fatty acid biosynthetic process	13/11091	70/20550	1.65E-09	9.10E-06	8.11E-06	8.88E-06
GO:0006546	UR	glycine catabolic process	8/11091	53/20550	5.96E-09	3.29E-05	2.92E-05	3.21E-05
GO:0006766	UR	vitamin metabolic process	25/11091	98/20550	1.72E-08	9.49E-05	8.37E-05	9.25E-05
GO:0009108	UR	coenzyme biosynthetic process	54/11091	159/20550	3.72E-07	0.00206	0.00179	0.00201
GO:0010207	UR	photosystem II assembly	49/11091	175/20550	2.15E-10	1.19E-06	1.15E-06	1.16E-06
GO:0004721	UR	phosphoprotein phosphatase activity	61/11091	173/20550	7.50E-07	0.00414	0.00358	0.00404
GO:0016791	UR	phosphatase activity	107/11091	271/20550	1.99E-06	0.011	0.00944	0.0107
GO:0022891	UR	substrate-specific transmembrane transporter activity	292/11091	706/20550	4.09E-10	2.26E-06	2.11E-06	2.21E-06
GO:0005802	UR	trans-golgi network	36/11091	221/20550	2.44E-10	1.35E-06	1.35E-06	1.30E-06
GO:0006486	UR	protein glycosylation	30/11091	219/20550	1.97E-10	1.09E-06	1.06E-06	1.06E-06
GO:0009853	UR	photorespiration	21/11091	157/20550	1.54E-10	8.54E-07	8.42E-07	8.32E-07
GO:0010090	UR	trichome morphogenesis	37/11091	126/20550	2.57E-08	0.000142	0.000125	0.000139
GO:0042538	UR	hyperosmotic salinity response	40/11091	157/20550	1.54E-10	8.54E-07	8.42E-07	8.32E-07
GO:0006979	UR	response to oxidative stress	205/11091	590/20550	3.46E-10	1.91E-06	1.80E-06	1.87E-06
GO:0009269	UR	response to desiccation	4/11091	38/20550	2.57E-08	0.000142	0.000125	0.000139
GO:0016787	UR	hydrolase activity	1214/11091	2580/20550	7.67E-10	4.24E-06	3.84E-06	4.14E-06

GO:0019252	UR	starch biosynthetic process	56/11091	186/20550	1.60E-10	8.84E-07	8.71E-07	8.62E-07
GO:0009637	UR	response to blue light	28/11091	123/20550	1.19E-10	6.57E-07	6.54E-07	6.41E-07
GO:0009743	UR	response to carbohydrate	106/11091	386/20550	3.22E-10	1.78E-06	1.68E-06	1.74E-06
GO:0010182	UR	sugar mediated signaling pathway	31/11091	102/20550	1.92E-06	0.0106	0.00911	0.0103
GO:0070301	UR	cellular response to hydrogen peroxide	8/11091	75/20550	1.02E-10	5.63E-07	5.61E-07	5.48E-07
GO:0071215	UR	cellular response to abscisic acid stimulus	58/11091	222/20550	2.63E-10	1.46E-06	1.40E-06	1.42E-06
GO:0048573	UR	photoperiodism, flowering	39/11091	150/20550	1.63E-10	9.01E-07	8.88E-07	8.79E-07
GO:0009451	UR	RNA modification	145/11091	359/20550	2.50E-07	0.00138	0.0012	0.00135
GO:0009615	UR	response to virus	66/11091	179/20550	5.12E-06	0.0283	0.0241	0.0276
GO:0008233	UR	peptidase activity	186/11091	451/20550	4.53E-08	0.000251	0.00022	0.000244
GO:0009744	UR	response to sucrose	65/11091	206/20550	3.44E-10	1.90E-06	1.79E-06	1.85E-06
GO:0042170	UR	plastid membrane	21/11091	86/20550	2.72E-08	0.000151	0.000133	0.000147
GO:0015934	UR	large ribosomal subunit	31/11091	152/20550	2.20E-10	1.22E-06	1.18E-06	1.19E-06
GO:0003735	UR	structural constituent of ribosome	108/11091	369/20550	2.58E-10	1.43E-06	1.37E-06	1.39E-06
GO:0006412	UR	translation	114/11091	383/20550	2.76E-10	1.53E-06	1.46E-06	1.49E-06
GO:0010287	UR	plastoglobule	10/11091	68/20550	1.47E-10	8.14E-07	8.06E-07	7.94E-07
GO:0009535	UR	chloroplast thylakoid membrane	73/11091	310/20550	2.58E-10	1.43E-06	1.37E-06	1.39E-06
GO:0009062	UR	fatty acid catabolic process	63/11091	213/20550	2.70E-10	1.49E-06	1.43E-06	1.45E-06
GO:0006812	UR	cation transport	306/11091	720/20550	7.67E-10	4.24E-06	3.83E-06	4.13E-06

GO:0022625	UR	cytosolic large ribosomal subunit	10/11091	112/20550	1.85E-10	1.02E-06	1.00E-06	9.98E-07
GO:0042254	UR	ribosome biogenesis	21/11091	122/20550	1.90E-10	1.05E-06	1.03E-06	1.02E-06
GO:0005840	UR	ribosome	107/11091	415/20550	2.68E-10	1.48E-06	1.42E-06	1.44E-06
GO:0004722	UR	protein serine/threonine phosphatase activity	43/11091	138/20550	6.88E-08	0.00038	0.000333	0.000371
GO:0009165	UR	nucleotide biosynthetic process	116/11091	355/20550	2.29E-10	1.26E-06	1.22E-06	1.23E-06
GO:0006913	UR	nucleocytoplasmic transport	54/11091	187/20550	1.96E-10	1.08E-06	1.06E-06	1.05E-06
GO:0007165	UR	signal transduction	689/11091	1630/20550	6.70E-10	3.71E-06	3.37E-06	3.61E-06
GO:0019761	UR	glucosinolate biosynthetic process	33/11091	167/20550	1.99E-10	1.10E-06	1.08E-06	1.07E-06
GO:0031348	UR	negative regulation of defense response	78/11091	271/20550	2.47E-10	1.36E-06	1.32E-06	1.33E-06
GO:0009862	UR	systemic acquired resistance, salicylic acid mediated signaling pathway	78/11091	241/20550	2.53E-10	1.40E-06	1.35E-06	1.36E-06
GO:0042742	UR	defense response to bacterium	94/11091	376/20550	2.63E-10	1.45E-06	1.39E-06	1.42E-06
GO:0009867	UR	jasmonic acid mediated signaling pathway	73/11091	274/20550	3.05E-10	1.68E-06	1.60E-06	1.64E-06
GO:0010363	UR	regulation of plant-type hypersensitive response	81/11091	367/20550	3.20E-10	1.77E-06	1.67E-06	1.73E-06
GO:0000165	UR	MAPK cascade	62/11091	201/20550	3.22E-10	1.78E-06	1.68E-06	1.74E-06
GO:0006612	UR	protein targeting to membrane	84/11091	377/20550	3.92E-10	2.17E-06	2.02E-06	2.11E-06
GO:0009697	UR	salicylic acid biosynthetic process	68/11091	205/20550	2.03E-09	1.12E-05	1.00E-05	1.09E-05
GO:0043900	UR	regulation of multi-organism process	35/11091	124/20550	1.02E-08	5.62E-05	4.97E-05	5.48E-05
GO:0009595	UR	detection of biotic stimulus	28/11091	102/20550	6.86E-08	0.000379	0.000332	0.00037

GO:0009625	UR	response to insect	19/11091	72/20550	2.40E-06	0.0132	0.0114	0.0129
GO:0010310	UR	regulation of hydrogen peroxide metabolic process	65/11091	178/20550	3.32E-06	0.0184	0.0157	0.0179
GO:0005774	UR	vacuolar membrane	72/11091	462/20550	2.60E-10	1.44E-06	1.38E-06	1.40E-06
GO:0006568	UR	tryptophan metabolic process	18/11091	91/20550	2.11E-10	1.17E-06	1.14E-06	1.14E-06
GO:0006606	UR	protein import into nucleus	41/11091	130/20550	3.17E-07	0.00175	0.00152	0.00171
GO:0048193	UR	golgi vesicle transport	45/11091	320/20550	2.36E-10	1.31E-06	1.31E-06	1.26E-06
GO:0030244	UR	cellulose biosynthetic process	25/11091	97/20550	2.71E-08	0.00015	0.000132	0.000146
GO:0009612	UR	response to mechanical stimulus	10/11091	60/20550	4.34E-09	2.40E-05	2.13E-05	2.34E-05
GO:0006084	UR	acetyl-CoA metabolic process	20/11091	87/20550	4.17E-09	2.30E-05	2.05E-05	2.25E-05
GO:0016126	UR	sterol biosynthetic process	54/11091	159/20550	3.72E-07	0.00206	0.00179	0.00201
GO:0060548	UR	negative regulation of cell death	34/11091	172/20550	1.33E-10	7.36E-07	7.31E-07	7.18E-07
GO:0019725	UR	cellular homeostasis	110/11091	306/20550	4.32E-10	2.39E-06	2.21E-06	2.33E-06
GO:0046777	UR	protein autophosphorylation	36/11091	145/20550	1.40E-10	7.73E-07	7.67E-07	7.54E-07
GO:0007155	UR	cell adhesion	17/11091	91/20550	2.11E-10	1.17E-06	1.14E-06	1.14E-06
GO:0071555	UR	cell wall organization	177/11091	447/20550	8.67E-10	4.79E-06	4.31E-06	4.67E-06
GO:0045010	UR	actin nucleation	29/11091	99/20550	7.91E-07	0.00437	0.00378	0.00426
GO:0034976	UR	response to endoplasmic reticulum stress	94/11091	335/20550	2.97E-10	1.64E-06	1.56E-06	1.60E-06
GO:0022626	UR	cytosolic ribosome	18/11091	189/20550	1.36E-10	7.54E-07	7.49E-07	7.35E-07
GO:0044237	UR	cellular metabolic process	3412/11091	8241/20550	1.18E-09	6.50E-06	5.80E-06	6.33E-06

GO:0050662	UR	coenzyme binding	123/11091	313/20550	1.69E-07	0.000934	0.000814	0.00091
GO:0010119	UR	regulation of stomatal movement	9/11091	46/20550	2.10E-06	0.0116	0.00998	0.0113
GO:0015931	UR	nucleobase-containing compound transport	44/11091	148/20550	3.46E-09	1.92E-05	1.70E-05	1.87E-05
GO:0016829	UR	lyase activity	107/11091	298/20550	5.47E-10	3.02E-06	2.77E-06	2.95E-06
GO:0009627	UR	systemic acquired resistance	76/11091	239/20550	2.11E-10	1.16E-06	1.13E-06	1.14E-06
GO:0006073	UR	cellular glucan metabolic process	123/11091	409/20550	2.80E-10	1.55E-06	1.48E-06	1.51E-06
GO:0006952	UR	defense response	467/11091	1214/20550	5.06E-10	2.80E-06	2.57E-06	2.73E-06
GO:0008219	UR	cell death	32/11091	110/20550	1.39E-07	0.000768	0.00067	0.000748
GO:0000902	UR	cell morphogenesis	206/11091	491/20550	7.31E-08	0.000404	0.000354	0.000394
GO:0051604	UR	protein maturation	18/11091	78/20550	3.42E-08	0.000189	0.000166	0.000184
GO:0000398	UR	mRNA splicing, via spliceosome	27/11091	100/20550	5.14E-08	0.000284	0.000249	0.000277
GO:0043481	UR	anthocyanin accumulation in tissues in response to UV light	30/11091	103/20550	4.35E-07	0.00241	0.00209	0.00235
GO:0042546	UR	cell wall biogenesis	56/11091	177/20550	2.04E-09	1.13E-05	1.00E-05	1.10E-05
GO:0042732	UR	D-xylose metabolic process	5/11091	35/20550	1.50E-06	0.0083	0.00715	0.00809
GO:0009220	UR	pyrimidine ribonucleotide biosynthetic process	43/11091	139/20550	4.62E-08	0.000255	0.000224	0.000249
GO:0009640	UR	photomorphogenesis	68/11091	195/20550	9.74E-08	0.000538	0.000471	0.000525
GO:0009825	UR	multidimensional cell growth	19/11091	83/20550	1.17E-08	6.46E-05	5.71E-05	6.30E-05
GO:0030833	UR	regulation of actin filament polymerization	36/11091	111/20550	5.46E-06	0.0302	0.0257	0.0294

GO:0006569	UR	tryptophan catabolic process	9/11091	71/20550	1.60E-10	8.87E-07	8.74E-07	8.65E-07
GO:0009684	UR	indoleacetic acid biosynthetic process	13/11091	99/20550	1.68E-10	9.26E-07	9.11E-07	9.03E-07
GO:0031347	UR	regulation of defense response	133/11091	530/20550	2.87E-10	1.58E-06	1.51E-06	1.54E-06
GO:0046872	UR	metal ion binding	1007/11091	2198/20550	7.06E-10	3.91E-06	3.54E-06	3.81E-06
GO:0003743	UR	translation initiation factor activity	23/11091	80/20550	7.69E-06	0.0425	0.0362	0.0415
GO:0009534	UR	chloroplast thylakoid	34/11091	211/20550	2.21E-10	1.22E-06	1.19E-06	1.19E-06
GO:0010155	UR	regulation of proton transport	12/11091	77/20550	2.01E-10	1.11E-06	1.08E-06	1.08E-06
GO:0080129	UR	proteasome core complex assembly	10/11091	126/20550	1.08E-10	6.00E-07	5.97E-07	5.84E-07
GO:0051788	UR	response to misfolded protein	35/11091	178/20550	2.82E-10	1.56E-06	1.49E-06	1.52E-06
GO:0045271	UR	respiratory chain complex I	15/11091	65/20550	5.31E-07	0.00293	0.00254	0.00286
GO:0007010	UR	cytoskeleton organization	161/11091	379/20550	7.10E-06	0.0392	0.0334	0.0383
GO:0009750	UR	response to fructose	21/11091	142/20550	1.64E-10	9.08E-07	8.94E-07	8.85E-07
GO:0009749	UR	response to glucose	18/11091	82/20550	4.47E-09	2.47E-05	2.20E-05	2.41E-05
GO:0043234	UR	protein complex	525/11091	1351/20550	5.80E-10	3.21E-06	2.93E-06	3.13E-06
GO:0006520	UR	cellular amino acid metabolic process	177/11091	815/20550	3.80E-10	2.10E-06	1.97E-06	2.05E-06
GO:0043069	UR	negative regulation of programmed cell death	33/11091	168/20550	1.33E-10	7.36E-07	7.31E-07	7.18E-07
GO:0030968	UR	endoplasmic reticulum unfolded protein response	36/11091	171/20550	2.01E-10	1.11E-06	1.08E-06	1.08E-06
GO:0009863	UR	salicylic acid mediated signaling pathway	101/11091	338/20550	3.48E-10	1.92E-06	1.81E-06	1.88E-06
GO:0052542	UR	defense response by callose deposition	12/11091	61/20550	4.66E-08	0.000258	0.000226	0.000251

GO:0007154	UR	cell communication	183/11091	434/20550	7.93E-07	0.00438	0.00379	0.00427
GO:0009658	UR	chloroplast organization	56/11091	235/20550	2.39E-10	1.32E-06	1.27E-06	1.29E-06
GO:0040007	UR	growth	211/11091	559/20550	3.59E-10	1.99E-06	1.86E-06	1.94E-06
GO:0010114	UR	response to red light	24/11091	102/20550	6.88E-10	3.81E-06	3.46E-06	3.71E-06
GO:0043248	UR	proteasome assembly	35/11091	178/20550	2.82E-10	1.56E-06	1.49E-06	1.52E-06
GO:0043161	UR	proteasome-mediated ubiquitin-dependent protein catabolic process	36/11091	123/20550	3.11E-08	0.000172	0.000151	0.000168
GO:0030001	UR	metal ion transport	218/11091	526/20550	7.08E-09	3.91E-05	3.46E-05	3.81E-05
GO:0009620	UR	response to fungus	126/11091	453/20550	3.24E-10	1.79E-06	1.69E-06	1.75E-06
GO:0009791	UR	post-embryonic development	194/11091	522/20550	3.80E-10	2.10E-06	1.96E-06	2.05E-06
GO:0006096	UR	glycolysis	15/11091	215/20550	2.25E-10	1.25E-06	1.21E-06	1.21E-06
GO:0006006	UR	glucose metabolic process	70/11091	539/20550	4.02E-10	2.22E-06	2.07E-06	2.17E-06
GO:0008324	UR	cation transmembrane transporter activity	151/11091	387/20550	3.40E-09	1.88E-05	1.67E-05	1.83E-05
GO:0016049	UR	cell growth	186/11091	495/20550	3.18E-10	1.76E-06	1.66E-06	1.71E-06
GO:0009755	UR	hormone-mediated signaling pathway	225/11091	649/20550	3.62E-10	2.00E-06	1.88E-06	1.95E-06
GO:0000303	UR	response to superoxide	8/11091	43/20550	3.52E-06	0.0195	0.0167	0.019
GO:0006635	UR	fatty acid beta-oxidation	31/11091	169/20550	1.58E-10	8.73E-07	8.61E-07	8.51E-07
GO:0007033	UR	vacuole organization	4/11091	57/20550	6.34E-11	3.51E-07	3.51E-07	3.42E-07
GO:0019829	UR	cation-transporting ATPase activity	16/11091	64/20550	3.42E-06	0.0189	0.0162	0.0184

GO:0009218	UR	pyrimidine ribonucleotide metabolic process	44/11091	144/20550	1.73E-08	9.58E-05	8.45E-05	9.34E-05
GO:0009648	UR	photoperiodism	40/11091	157/20550	1.54E-10	8.54E-07	8.42E-07	8.32E-07
GO:0045087	UR	innate immune response	161/11091	533/20550	3.26E-10	1.80E-06	1.70E-06	1.76E-06
GO:0045088	UR	regulation of innate immune response	97/11091	408/20550	2.66E-10	1.47E-06	1.41E-06	1.44E-06
GO:0016616	UR	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	56/11091	166/20550	1.73E-07	0.000958	0.000835	0.000934
GO:0006833	UR	water transport	13/11091	144/20550	1.17E-10	6.47E-07	6.43E-07	6.30E-07
GO:0010027	UR	thylakoid membrane organization	51/11091	196/20550	1.93E-10	1.07E-06	1.04E-06	1.04E-06
GO:0051707	UR	response to other organism	403/11091	1240/20550	5.14E-10	2.84E-06	2.61E-06	2.77E-06
GO:0006984	UR	ER-nucleus signaling pathway	40/11091	179/20550	1.90E-10	1.05E-06	1.03E-06	1.03E-06
GO:0000160	UR	phosphorelay signal transduction system	41/11091	137/20550	1.33E-08	7.35E-05	6.49E-05	7.16E-05
GO:0016070	UR	RNA metabolic process	542/11091	1497/20550	5.49E-10	3.04E-06	2.78E-06	2.96E-06
GO:0009694	UR	jasmonic acid metabolic process	35/11091	154/20550	1.84E-10	1.02E-06	9.96E-07	9.90E-07
GO:0009698	UR	phenylpropanoid metabolic process	59/11091	168/20550	1.10E-06	0.00608	0.00524	0.00592
GO:0010072	UR	primary shoot apical meristem specification	8/11091	43/20550	3.52E-06	0.0195	0.0167	0.019
GO:0071369	UR	cellular response to ethylene stimulus	38/11091	126/20550	7.45E-08	0.000412	0.000361	0.000402
GO:0048589	UR	developmental growth	179/11091	430/20550	2.51E-07	0.00139	0.00121	0.00135
GO:0009630	UR	gravitropism	28/11091	152/20550	2.20E-10	1.22E-06	1.18E-06	1.19E-06
GO:0009086	UR	methionine biosynthetic process	23/11091	100/20550	4.38E-10	2.42E-06	2.24E-06	2.36E-06

GO:0022627	UR	cytosolic small ribosomal subunit	20/11091	95/20550	1.37E-10	7.58E-07	7.52E-07	7.39E-07
GO:0032502	UR	developmental process	1493/11091	3141/20550	8.78E-10	4.85E-06	4.36E-06	4.73E-06
GO:0008237	UR	metallopeptidase activity	21/11091	77/20550	2.78E-06	0.0154	0.0132	0.015
GO:0005911	UR	cell-cell junction	196/11091	809/20550	4.17E-10	2.30E-06	2.14E-06	2.24E-06
GO:0016043	UR	cellular component organization	1089/11091	2846/20550	8.10E-10	4.48E-06	4.03E-06	4.36E-06
GO:0006406	UR	mRNA export from nucleus	7/11091	59/20550	1.52E-10	8.43E-07	8.32E-07	8.21E-07
GO:0005635	UR	nuclear envelope	7/11091	50/20550	5.30E-09	2.93E-05	2.60E-05	2.86E-05
GO:0006796	UR	phosphate-containing compound metabolic process	960/11091	2372/20550	7.27E-10	4.02E-06	3.64E-06	3.92E-06
GO:0043085	UR	positive regulation of catalytic activity	27/11091	120/20550	1.33E-10	7.34E-07	7.29E-07	7.15E-07
GO:0044267	UR	cellular protein metabolic process	1302/11091	3093/20550	8.79E-10	4.86E-06	4.36E-06	4.74E-06
GO:0000023	UR	maltose metabolic process	46/11091	151/20550	5.56E-09	3.07E-05	2.73E-05	3.00E-05
GO:0010319	UR	stromule	3/11091	36/20550	1.60E-08	8.86E-05	7.81E-05	8.63E-05
GO:0007568	UR	aging	47/11091	135/20550	8.14E-06	0.045	0.0382	0.0439
GO:0009963	UR	positive regulation of flavonoid biosynthetic process	15/11091	103/20550	1.00E-10	5.56E-07	5.55E-07	5.42E-07
GO:0009579	UR	thylakoid	54/11091	325/20550	3.23E-10	1.79E-06	1.69E-06	1.74E-06
GO:0006753	UR	nucleoside phosphate metabolic process	213/11091	703/20550	4.82E-10	2.67E-06	2.46E-06	2.60E-06
GO:0035304	UR	regulation of protein dephosphorylation	26/11091	134/20550	1.40E-10	7.75E-07	7.68E-07	7.56E-07
GO:0006397	UR	mRNA processing	46/11091	151/20550	5.56E-09	3.07E-05	2.73E-05	3.00E-05
GO:0005198	UR	structural molecule activity	136/11091	482/20550	3.39E-10	1.88E-06	1.77E-06	1.83E-06

GO:0006623	UR	protein targeting to vacuole	9/11091	117/20550	1.51E-10	8.33E-07	8.23E-07	8.12E-07
GO:0006790	UR	sulfur compound metabolic process	117/11091	551/20550	3.64E-10	2.01E-06	1.88E-06	1.96E-06
GO:0009308	UR	amine metabolic process	69/11091	220/20550	2.20E-10	1.22E-06	1.18E-06	1.18E-06
GO:0016117	UR	carotenoid biosynthetic process	20/11091	104/20550	1.48E-10	8.18E-07	8.09E-07	7.97E-07
GO:0044249	UR	cellular biosynthetic process	1431/11091	3749/20550	7.94E-10	4.39E-06	3.96E-06	4.28E-06
GO:0009814	UR	defense response, incompatible interaction	100/11091	360/20550	3.07E-10	1.70E-06	1.61E-06	1.66E-06
GO:0009250	UR	glucan biosynthetic process	88/11091	301/20550	2.98E-10	1.65E-06	1.57E-06	1.61E-06
GO:0008047	UR	enzyme activator activity	21/11091	78/20550	1.78E-06	0.00982	0.00844	0.00957
GO:0046914	UR	transition metal ion binding	822/11091	1752/20550	1.26E-09	6.97E-06	6.22E-06	6.80E-06
GO:0006605	UR	protein targeting	176/11091	746/20550	3.54E-10	1.96E-06	1.84E-06	1.91E-06
GO:0008380	UR	RNA splicing	84/11091	276/20550	2.53E-10	1.40E-06	1.35E-06	1.36E-06
GO:0010638	UR	positive regulation of organelle organization	46/11091	144/20550	1.26E-07	0.000699	0.000611	0.000682
GO:0030003	UR	cellular cation homeostasis	41/11091	158/20550	1.78E-10	9.83E-07	9.65E-07	9.58E-07
GO:0070838	UR	divalent metal ion transport	27/11091	201/20550	2.20E-10	1.22E-06	1.18E-06	1.19E-06
GO:0006972	UR	hyperosmotic response	44/11091	246/20550	2.60E-10	1.44E-06	1.38E-06	1.40E-06
GO:0016820	UR	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	59/11091	166/20550	2.13E-06	0.0118	0.0101	0.0115
GO:0009832	UR	plant-type cell wall biogenesis	30/11091	127/20550	2.48E-10	1.37E-06	1.32E-06	1.33E-06
GO:0019344	UR	cysteine biosynthetic process	25/11091	210/20550	2.08E-10	1.15E-06	1.12E-06	1.12E-06

GO:0009850	UR	auxin metabolic process	24/11091	128/20550	1.58E-10	8.73E-07	8.61E-07	8.51E-07
GO:0016874	UR	ligase activity	166/11091	446/20550	3.27E-10	1.81E-06	1.71E-06	1.76E-06
GO:0043090	UR	amino acid import	15/11091	72/20550	1.49E-08	8.23E-05	7.27E-05	8.02E-05
GO:0043226	UR	organelle	7081/11091	13704/20550	1.04E-09	5.75E-06	5.14E-06	5.60E-06
GO:0000302	UR	response to reactive oxygen species	119/11091	372/20550	3.04E-10	1.68E-06	1.60E-06	1.64E-06
GO:0005747	UR	mitochondrial respiratory chain complex I	12/11091	61/20550	4.66E-08	0.000258	0.000226	0.000251
GO:0006487	UR	protein N-linked glycosylation	10/11091	101/20550	6.12E-11	3.38E-07	3.38E-07	3.30E-07
GO:0042793	UR	transcription from plastid promoter	13/11091	72/20550	5.42E-10	3.00E-06	2.75E-06	2.92E-06
GO:0009073	UR	aromatic amino acid family biosynthetic process	19/11091	99/20550	1.68E-10	9.26E-07	9.11E-07	9.03E-07
GO:0015996	UR	chlorophyll catabolic process	12/11091	54/20550	3.07E-06	0.017	0.0146	0.0166
GO:0010388	UR	cullin deneddylation	32/11091	101/20550	7.82E-06	0.0433	0.0368	0.0422
GO:0000338	UR	protein deneddylation	32/11091	101/20550	7.82E-06	0.0433	0.0368	0.0422
GO:0050897	UR	cobalt ion binding	6/11091	46/20550	1.32E-08	7.29E-05	6.44E-05	7.11E-05
GO:0007030	UR	golgi organization	8/11091	178/20550	2.82E-10	1.56E-06	1.56E-06	1.49E-06
GO:0016310	UR	phosphorylation	546/11091	1215/20550	6.20E-10	3.43E-06	3.12E-06	3.34E-06
GO:0010431	UR	seed maturation	6/11091	38/20550	1.67E-06	0.00925	0.00795	0.00902
GO:0045595	UR	regulation of cell differentiation	43/11091	126/20550	8.79E-06	0.0486	0.0413	0.0474
GO:0048037	UR	cofactor binding	156/11091	429/20550	3.82E-10	2.11E-06	1.97E-06	2.06E-06
GO:0050790	UR	regulation of catalytic activity	119/11091	293/20550	4.87E-06	0.0269	0.023	0.0262

GO:0009059	UR	macromolecule biosynthetic process	649/11091	1526/20550	5.57E-10	3.08E-06	2.81E-06	3.00E-06
GO:0006094	UR	gluconeogenesis	13/11091	165/20550	1.51E-10	8.32E-07	8.23E-07	8.11E-07
GO:0010498	UR	proteasomal protein catabolic process	48/11091	203/20550	1.52E-10	8.39E-07	8.29E-07	8.18E-07
GO:0010817	UR	regulation of hormone levels	139/11091	413/20550	2.42E-10	1.34E-06	1.29E-06	1.31E-06
GO:0000271	UR	polysaccharide biosynthetic process	184/11091	557/20550	3.33E-10	1.84E-06	1.73E-06	1.79E-06
GO:0010218	UR	response to far red light	19/11091	97/20550	1.04E-10	5.76E-07	5.74E-07	5.61E-07
GO:0030076	UR	light-harvesting complex	1/11091	27/20550	2.56E-08	0.000142	0.000125	0.000138
GO:0016168	UR	chlorophyll binding	4/11091	30/20550	7.65E-06	0.0423	0.0359	0.0412
GO:0009765	UR	photosynthesis, light harvesting	4/11091	30/20550	7.65E-06	0.0423	0.0359	0.0412
GO:0051603	UR	proteolysis involved in cellular protein catabolic process	113/11091	439/20550	2.33E-10	1.29E-06	1.25E-06	1.26E-06
GO:0019760	UR	glucosinolate metabolic process	43/11091	199/20550	2.61E-10	1.44E-06	1.38E-06	1.40E-06
GO:0015935	UR	small ribosomal subunit	36/11091	117/20550	5.29E-07	0.00293	0.00253	0.00285
GO:0009610	UR	response to symbiotic fungus	5/11091	37/20550	4.30E-07	0.00238	0.00206	0.00232
GO:0009725	UR	response to hormone	584/11091	1424/20550	5.24E-10	2.90E-06	2.66E-06	2.82E-06
GO:0009653	UR	anatomical structure morphogenesis	419/11091	948/20550	1.14E-09	6.31E-06	5.63E-06	6.15E-06
GO:0043067	UR	regulation of programmed cell death	90/11091	390/20550	3.96E-10	2.19E-06	2.04E-06	2.13E-06
GO:0009805	UR	coumarin biosynthetic process	8/11091	53/20550	5.96E-09	3.29E-05	2.92E-05	3.21E-05
GO:0015075	UR	ion transmembrane transporter activity	232/11091	578/20550	3.02E-10	1.67E-06	1.59E-06	1.63E-06
GO:0034220	UR	ion transmembrane transport	61/11091	167/20550	7.28E-06	0.0403	0.0342	0.0392

GO:0030529	UR	ribonucleoprotein complex	151/11091	509/20550	3.07E-10	1.69E-06	1.61E-06	1.65E-06
GO:0006888	UR	ER to Golgi vesicle-mediated transport	13/11091	102/20550	1.64E-10	9.08E-07	9.08E-07	8.93E-07
GO:0008654	UR	phospholipid biosynthetic process	144/11091	396/20550	2.80E-10	1.55E-06	1.48E-06	1.51E-06
GO:0015994	UR	chlorophyll metabolic process	24/11091	171/20550	2.01E-10	1.11E-06	1.08E-06	1.08E-06
GO:0009532	UR	plastid stroma	108/11091	578/20550	3.02E-10	1.67E-06	1.59E-06	1.63E-06
GO:0019538	UR	protein metabolic process	1538/11091	3784/20550	9.13E-10	5.05E-06	4.53E-06	4.92E-06
GO:0008135	UR	translation factor activity, nucleic acid binding	34/11091	123/20550	4.89E-09	2.70E-05	2.40E-05	2.63E-05
GO:0050776	UR	regulation of immune response	98/11091	410/20550	2.92E-10	1.62E-06	1.54E-06	1.57E-06
GO:0009415	UR	response to water	116/11091	409/20550	2.80E-10	1.55E-06	1.48E-06	1.51E-06
GO:0009696	UR	salicylic acid metabolic process	72/11091	212/20550	5.75E-09	3.18E-05	2.82E-05	3.10E-05
GO:0034620	UR	cellular response to unfolded protein	37/11091	173/20550	2.56E-10	1.41E-06	1.36E-06	1.38E-06
GO:0006779	UR	porphyrin-containing compound biosynthetic process	27/11091	141/20550	1.27E-10	7.04E-07	7.00E-07	6.87E-07
GO:0033014	UR	tetrapyrrole biosynthetic process	29/11091	145/20550	1.40E-10	7.73E-07	7.67E-07	7.54E-07
GO:0009657	UR	plastid organization	66/11091	311/20550	2.85E-10	1.57E-06	1.50E-06	1.53E-06
GO:0006816	UR	calcium ion transport	10/11091	119/20550	1.06E-10	5.87E-07	5.85E-07	5.72E-07
GO:0048523	UR	negative regulation of cellular process	346/11091	775/20550	1.39E-07	0.000766	0.000669	0.000747
GO:0016197	UR	endosomal transport	5/11091	35/20550	1.50E-06	0.0083	0.00715	0.00809
GO:0009526	UR	plastid envelope	103/11091	533/20550	3.26E-10	1.80E-06	1.70E-06	1.76E-06
GO:0000272	UR	polysaccharide catabolic process	12/11091	54/20550	3.07E-06	0.017	0.0146	0.0166

GO:0006461	UR	protein complex assembly	205/11091	591/20550	4.70E-10	2.60E-06	2.40E-06	2.53E-06
GO:0006807	UR	nitrogen compound metabolic process	1387/11091	3533/20550	9.12E-10	5.04E-06	4.53E-06	4.92E-06
GO:0018377	UR	protein myristoylation	202/11091	470/20550	1.65E-06	0.0091	0.00783	0.00887
GO:0009069	UR	serine family amino acid metabolic process	38/11091	270/20550	2.28E-10	1.26E-06	1.22E-06	1.23E-06
GO:0006414	UR	translational elongation	7/11091	45/20550	1.85E-07	0.00102	0.000889	0.000995
GO:0031090	UR	organelle membrane	178/11091	755/20550	4.47E-10	2.47E-06	2.29E-06	2.41E-06
GO:0009966	UR	regulation of signal transduction	47/11091	147/20550	1.07E-07	0.000591	0.000516	0.000576
GO:0009987	UR	cellular process	4695/11091	10658/20550	1.11E-09	6.12E-06	5.46E-06	5.96E-06
GO:0031365	UR	N-terminal protein amino acid modification	203/11091	474/20550	9.11E-07	0.00504	0.00435	0.00491
GO:0009902	UR	chloroplast relocation	14/11091	104/20550	1.48E-10	8.18E-07	8.09E-07	7.97E-07
GO:0034660	UR	ncRNA metabolic process	95/11091	388/20550	3.57E-10	1.97E-06	1.85E-06	1.92E-06
GO:0009266	UR	response to temperature stimulus	269/11091	899/20550	4.20E-10	2.32E-06	2.16E-06	2.26E-06
GO:0010243	UR	response to organonitrogen compound	152/11091	431/20550	3.12E-10	1.72E-06	1.63E-06	1.68E-06
GO:0016462	UR	pyrophosphatase activity	333/11091	800/20550	4.23E-10	2.34E-06	2.17E-06	2.28E-06
GO:0006778	UR	porphyrin-containing compound metabolic process	42/11091	205/20550	1.90E-10	1.05E-06	1.03E-06	1.02E-06
GO:0019318	UR	hexose metabolic process	82/11091	576/20550	3.81E-10	2.11E-06	1.97E-06	2.05E-06
GO:0071446	UR	cellular response to salicylic acid stimulus	103/11091	340/20550	1.89E-10	1.05E-06	1.02E-06	1.02E-06
GO:0034614	UR	cellular response to reactive oxygen species	25/11091	105/20550	3.61E-10	1.99E-06	1.87E-06	1.94E-06
GO:0019684	UR	photosynthesis, light reaction	39/11091	144/20550	2.17E-10	1.20E-06	1.17E-06	1.17E-06

GO:0005982	UR	starch metabolic process	61/11091	220/20550	2.20E-10	1.22E-06	1.18E-06	1.18E-06
GO:0043269	UR	regulation of ion transport	36/11091	149/20550	2.50E-10	1.38E-06	1.33E-06	1.35E-06
GO:0000325	UR	plant-type vacuole	5/11091	43/20550	9.54E-09	5.28E-05	4.67E-05	5.14E-05
GO:2000377	UR	regulation of reactive oxygen species metabolic process	68/11091	194/20550	1.36E-07	0.000753	0.000657	0.000734
GO:0007034	UR	vacuolar transport	14/11091	129/20550	1.01E-10	5.57E-07	5.55E-07	5.43E-07
GO:0034599	UR	cellular response to oxidative stress	29/11091	114/20550	1.04E-09	5.77E-06	5.16E-06	5.62E-06
GO:0006163	UR	purine nucleotide metabolic process	79/11091	217/20550	2.17E-07	0.0012	0.00104	0.00117
GO:0030243	UR	cellulose metabolic process	36/11091	144/20550	2.17E-10	1.20E-06	1.17E-06	1.17E-06
GO:0030054	UR	cell junction	196/11091	811/20550	4.39E-10	2.43E-06	2.25E-06	2.36E-06
GO:0009851	UR	auxin biosynthetic process	19/11091	110/20550	1.23E-10	6.78E-07	6.74E-07	6.61E-07
GO:0030163	UR	protein catabolic process	54/11091	251/20550	1.64E-10	9.06E-07	8.92E-07	8.83E-07
GO:0006007	UR	glucose catabolic process	59/11091	460/20550	3.29E-10	1.82E-06	1.72E-06	1.77E-06
GO:0016071	UR	mRNA metabolic process	135/11091	363/20550	5.13E-10	2.84E-06	2.61E-06	2.77E-06
GO:0032549	UR	ribonucleoside binding	940/11091	2137/20550	6.65E-10	3.68E-06	3.34E-06	3.58E-06
GO:0019752	UR	carboxylic acid metabolic process	550/11091	1810/20550	6.33E-10	3.50E-06	3.19E-06	3.41E-06
GO:0034440	UR	lipid oxidation	35/11091	180/20550	2.28E-10	1.26E-06	1.22E-06	1.23E-06
GO:0016879	UR	ligase activity, forming carbon-nitrogen bonds	135/11091	339/20550	1.53E-07	0.000846	0.000738	0.000825
GO:0055035	UR	plastid thylakoid membrane	73/11091	312/20550	2.93E-10	1.62E-06	1.54E-06	1.58E-06
GO:0005839	UR	proteasome core complex	2/11091	24/20550	3.67E-06	0.0203	0.0174	0.0198

GO:0051246	UR	regulation of protein metabolic process	106/11091	297/20550	4.95E-10	2.74E-06	2.52E-06	2.67E-06
GO:0016042	UR	lipid catabolic process	81/11091	248/20550	1.96E-10	1.09E-06	1.06E-06	1.06E-06
GO:0032268	UR	regulation of cellular protein metabolic process	101/11091	282/20550	8.33E-10	4.61E-06	4.15E-06	4.49E-06
GO:0042651	UR	thylakoid membrane	77/11091	326/20550	2.37E-10	1.31E-06	1.27E-06	1.28E-06
GO:0044262	UR	cellular carbohydrate metabolic process	255/11091	724/20550	4.10E-10	2.27E-06	2.11E-06	2.21E-06
GO:0006637	UR	acyl-CoA metabolic process	21/11091	90/20550	3.70E-09	2.05E-05	1.82E-05	2.00E-05
GO:0009060	UR	aerobic respiration	4/11091	34/20550	3.59E-07	0.00199	0.00172	0.00193
GO:0009962	UR	regulation of flavonoid biosynthetic process	24/11091	119/20550	1.06E-10	5.87E-07	5.85E-07	5.72E-07
GO:0016482	UR	cytoplasmic transport	165/11091	664/20550	4.34E-10	2.40E-06	2.22E-06	2.34E-06
GO:0006221	UR	pyrimidine nucleotide biosynthetic process	46/11091	148/20550	2.74E-08	0.000152	0.000133	0.000148
GO:0010608	UR	posttranscriptional regulation of gene expression	51/11091	164/20550	3.41E-09	1.88E-05	1.68E-05	1.84E-05
GO:0010038	UR	response to metal ion	96/11091	575/20550	3.60E-10	1.99E-06	1.86E-06	1.94E-06
GO:0042744	UR	hydrogen peroxide catabolic process	5/11091	71/20550	1.60E-10	8.87E-07	8.74E-07	8.65E-07
GO:0019395	UR	fatty acid oxidation	34/11091	174/20550	1.71E-10	9.44E-07	9.27E-07	9.20E-07
GO:0009311	UR	oligosaccharide metabolic process	78/11091	234/20550	5.11E-10	2.83E-06	2.60E-06	2.76E-06
GO:0009629	UR	response to gravity	29/11091	160/20550	1.32E-10	7.28E-07	7.23E-07	7.09E-07
GO:0006164	UR	purine nucleotide biosynthetic process	41/11091	126/20550	1.45E-06	0.00804	0.00693	0.00784
GO:0009314	UR	response to radiation	471/11091	1188/20550	4.94E-10	2.73E-06	2.52E-06	2.66E-06
GO:0006625	UR	protein targeting to peroxisome	14/11091	95/20550	1.37E-10	7.58E-07	7.52E-07	7.39E-07

GO:0051179	UR	localization	48/11091	150/20550	6.01E-08	0.000332	0.000291	0.000324
GO:0072593	UR	reactive oxygen species metabolic process	23/11091	168/20550	1.33E-10	7.36E-07	7.31E-07	7.18E-07
GO:0006497	UR	protein lipidation	213/11091	486/20550	6.06E-06	0.0335	0.0285	0.0327
GO:0006091	UR	generation of precursor metabolites and energy	128/11091	515/20550	2.87E-10	1.58E-06	1.51E-06	1.54E-06

3.S2.7 Gene Ontology enrichment for 14,481 non-mobile *Solanum lycopersicum* transcripts. Table provides: GO accession; enrichment - (OR) Over-Represented, (UR) Under-Represented; hierarchy GO term; ratio in study; ratio in population; uncorrected p-value; Bonferroni correction p-value; Holm's p-value; Sidak test p-value.

GO accession	**Enrichment**	**GO Term**	**ratio_in_study**	**ratio_in_pop**	**p_uncorrected**	**p_bonferroni**	**p_holm**	**p_sidak**
GO:0006464	OR	cellular protein modification process	1079/11932	1082/12161	1.71E-06	0.00406	0.00405	0.00396
GO:0005506	UR	iron ion binding	223/11932	243/12161	3.18E-08	7.58E-05	7.57E-05	7.39E-05
GO:0016705	UR	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	329/11932	350/12161	2.81E-06	0.0067	0.00668	0.00654
GO:0020037	UR	heme binding	288/11932	307/12161	5.47E-06	0.013	0.013	0.0127
GO:0016829	UR	lyase activity	145/11932	159/12161	1.78E-06	0.00424	0.00422	0.00413
GO:0010333	UR	terpene synthase activity	25/11932	31/12161	2.07E-05	0.0494	0.0491	0.0481

GO:0009607	UR	response to biotic stimulus	24/11932	30/12161	1.70E-05	0.0405	0.0402	0.0394
GO:0055114	UR	oxidation-reduction process	951/11932	1000/12161	3.30E-10	7.86E-07	7.86E-07	7.66E-07
GO:0016491	UR	oxidoreductase activity	1086/11932	1138/12161	1.15E-09	2.74E-06	2.74E-06	2.67E-06

3.S2.8 Gene Ontology enrichment for 16,423 non-mobile *Cuscuta* unigenes from plants grown with *Arabidopsis* and *Solanum lycopersicum* hosts. Table provides: GO accession; enrichment - (OR) Over-Represented, (UR) Under-Represented; hierarchy GO term; ratio in study; ratio in population; uncorrected p-value; Bonferroni correction p-value; Holm's p-value; Sidak test p-value.

GO accession	**Enrichment**	**GO Term**	**ratio_in_study**	**ratio_in_pop**	**p_uncorrected**	**p_bonferroni**	**p_holm**	**p_sidak**
GO:0016740	OR	transferase activity	1512/7083	2168/11379	7.20E-10	1.30E-06	1.28E-06	1.27E-06
GO:0016772	OR	transferase activity, transferring phosphorus-containing groups	1008/7083	1392/11379	5.36E-10	9.66E-07	9.53E-07	9.42E-07
GO:0016779	OR	nucleotidyltransferase activity	372/7083	438/11379	3.48E-10	6.28E-07	6.22E-07	6.12E-07
GO:0006139	OR	nucleobase-containing compound metabolic process	857/7083	1137/11379	5.46E-10	9.85E-07	9.71E-07	9.61E-07
GO:0006260	OR	DNA replication	333/7083	376/11379	2.46E-10	4.44E-07	4.43E-07	4.33E-07
GO:0015074	OR	DNA integration	165/7083	178/11379	2.11E-10	3.81E-07	3.80E-07	3.71E-07
GO:0003964	OR	RNA-directed DNA polymerase activity	294/7083	322/11379	2.91E-10	5.26E-07	5.23E-07	5.12E-07
GO:0006278	OR	RNA-dependent DNA replication	294/7083	322/11379	2.91E-10	5.26E-07	5.23E-07	5.12E-07
GO:0003723	OR	RNA binding	437/7083	596/11379	5.25E-09	9.47E-06	9.28E-06	9.23E-06

GO:0003676	OR	nucleic acid binding	1180/7083	1714/11379	1.12E-09	2.03E-06	1.99E-06	1.98E-06
GO:0006259	OR	DNA metabolic process	579/7083	665/11379	3.09E-10	5.58E-07	5.53E-07	5.44E-07
GO:0006807	OR	nitrogen compound metabolic process	966/7083	1331/11379	5.30E-10	9.57E-07	9.44E-07	9.33E-07
GO:0043234	UR	protein complex	231/7083	482/11379	3.36E-10	6.05E-07	6.00E-07	5.90E-07
GO:0006886	UR	intracellular protein transport	39/7083	117/11379	2.64E-10	4.76E-07	4.73E-07	4.64E-07
GO:0006096	UR	glycolysis	16/7083	54/11379	1.48E-06	0.00266	0.00258	0.0026
GO:0006412	UR	translation	89/7083	243/11379	1.86E-10	3.35E-07	3.35E-07	3.27E-07
GO:0005840	UR	ribosome	85/7083	245/11379	2.36E-10	4.26E-07	4.25E-07	4.15E-07
GO:0003735	UR	structural constituent of ribosome	87/7083	247/11379	2.62E-10	4.72E-07	4.70E-07	4.60E-07
GO:0015992	UR	proton transport	17/7083	55/11379	3.84E-06	0.00693	0.00671	0.00676
GO:0005622	UR	intracellular	116/7083	262/11379	3.01E-09	5.43E-06	5.33E-06	5.29E-06
GO:0015031	UR	protein transport	60/7083	157/11379	1.13E-09	2.05E-06	2.01E-06	1.99E-06
GO:0006006	UR	glucose metabolic process	21/7083	69/11379	1.18E-07	0.000213	0.000208	0.000208
GO:0005198	UR	structural molecule activity	104/7083	283/11379	2.28E-10	4.12E-07	4.11E-07	4.01E-07
GO:0030117	UR	membrane coat	11/7083	42/11379	2.90E-06	0.00523	0.00506	0.00509

Table 3.S3.

Correlation of *Arabidopsis* transcript mobility into *Cuscuta* with reported phloem-associated transcripts from *Arabidopsis* and other species.

Category	***Arabidopsis* to *Cuscuta***	***Arabidopsis* phloem**	**Ash phloem**	**Castor bean phloem**	**Cucumber phloem**	**Watermelon phloem**
	Total in category	*Number of phloem transcripts/ESTs used in comparison**				
	33,602	2,415	10,558	146	821	1,206
		Number of matches to transcripts in Arabidopsis-Cuscuta mobility categories				
Mobile	9,518	2,080	6,498	99	659	603
Not mobile	11,874	302	3,773	45	150	561
Not detected	11,210	33	435	2	12	64
χ^2 Value		4,101	11,560	146	1,123	523

$P > \chi^2$	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

* Sets of *Arabidopsis* homologs for comparison were developed using best BLAST hits (1e-5) from phloem transcriptomes reported from four plant species in addition to *Arabidopsis* (Doering-Saad *et al.*, 2006; Deeken *et al.*, 2008; Huang *et al.*, 2009; Bai *et al.*, 2011; Guo *et al.*, 2013).

Table 3.S4.

Correlation of specific categories of *Arabidopsis*-to-*Cuscuta* transcript mobility with reported phloem-associated transcripts from *Arabidopsis* and other species. High-mobility transcripts in the parasite system were identified from those with PS:I FPKM ratios ten-fold higher than that of the mobile transcript mean.

Category	***Arabidopsis* to *Cuscuta***	***Arabidopsis* phloem**	**Ash phloem**	**Castor bean phloem**	**Cucumber phloem**	**Watermelon phloem**
	Total in category	*Number of phloem transcripts/ESTs used in comparison**				
	9,518	2,080	6,498	99	659	603
		Number of matches to transcripts in Arabidopsis-Cuscuta mobility categories				
High mobility	167	1	25	1	0	5
Other mobile	9,351	2,079	6,473	98	659	598
χ^2 Value		45.965	222.932	0.322†	12.645	3.198

$P > \chi^2$	<0.0001	<0.0001	0.479	0.165	0.042

*Transcripts corresponding to *Arabidopsis*-to-*Cuscuta* mobile from Table S3 were used in comparison. †Caution: One cell in this contingency table had expected counts less than 5.

Table 3.S5.

Primers used in amplifying transcripts from *Arabidopsis* and *Cuscuta*.

Gene	**Accession**	**Forward primer**	**Reverse primer**
AtUBQ	AT5G03240	TGGCCGTACCTTGGCCGATT	AGGCCCGTTACAAGCCCAAACG
AtSZF1	AT3G55980	GTCCCCTGTCCCGAGTTCCGT	ATGGCGGTTGTCTCACCGGG
At5TE15240	AT5G13205	TGACGAAGGAGTTCGAGATGACGGA	ACTCAAGAAGGCTATGCTAAGGAGGT
AtCAT3	AT1G20620	AGCGCGTACAACGCCCCATT	ACGCGTGGACAACATCCGGG
AtTCTP	AT3G16640	TGCTCCCCAGGCTCAGTGAC	TCGCTGCGACCGACCACAGT
Cp_Putative transaldolase	Cp_57330	AGCTGTAATTGATGCTTACCTG	GCTTCTTCGCACCTTTCTTC
Cp_Heat shock protein 70	Cp_156474	CCCCGGAAGCGTCAGCAGTG	CACCAGCCAAGCGTCAGGCA
Cp_Unknown protein	Cp_48979	ACACGGCCAAAGGGTGAGCG	TGTCTCTGCAGCCGGCAAGG
Cp_SET domain containing protein	Cp_210341	TCTCAACTCCACTCACCACC	TTGGGCAGAGAGAGCGAAAG

References

1. J. Dawson, L. Musselman, P. Wolswinkel, I. Dörr, *Rev. Weed Sci.* 6, 265-317 (1994).
2. K. C. Vaughn, *Protoplasma* 220, 189-200 (2003).
3. J. K. Roney, P. A. Khatibi, J. H. Westwood, *Plant Physiol.* 143, 1037-1043 (2007).
4. R. David-Schwartz, S. Runo, B. Townsley, J. Machuka, N. Sinha, New Phytol 179, 1133-1141 (2008).
5. M. LeBlanc, G. Kim, B. Patel, V. Stromberg, J. Westwood, *New Phytol.* 200, 1225-1233 (2013).
6. S. Haupt, K. Oparka, N. Sauer, S. Neumann, *J Exp Bot* 52, 173 - 177 (2001).
7. R. M. J. Hosford, *Botan. Rev.* 42, 387-406 (1967).
8. M. Birschwilks, S. Haupt, D. Hofius, S. Neumann, *J. Exp. Bot.* 57, 911-921 (2006).
9. H. J. M. van Dorst, D. Peters, *Eur. J. Plant Pathol.* 80, 85-96 (1974).
10. M. Kaminska, M. Korbin, *Acta Physiol Plant* 21, 21-26 (1999).
11. W. J. Lucas *et al.*, *J. Integ. Plant Biol.* 55, 294-388 (2013).
12. C. G. N. Turnbull, R. M. Lopez-Cobollo, *New Phytol.* 198, 33-51 (2013).
13. V. Haywood, T.-S. Yu, N.-C. Huang, W. J. Lucas, *Plant J* 42, 49-68 (2005).
14. M. Kim, W. Canio, S. Kessler, N. Sinha, *Science* 293, 287-289 (2001).
15. C. Li *et al.*, *Sci. Rep.* 1, 73 (2011).
16. A. K. Banerjee, T. Lin, D. J. Hannapel, *Plant Physiol.* 151, 1831-1843 (2009).
17. D. J. Hannapel, *J. Integ. Plant Biol.* 52, 40-52 (2010).
18. M. Notaguchi, S. Wolf, W. J. Lucas, *J. Integ. Plant Biol.* 54, 760-772 (2012).
19. A. Molnar *et al.*, *Science* 328, 872-875 (2010).
20. A. Alakonya *et al.*, *Plant Cell* 24, 3153-3166 (2012).
21. P. Sarkies, E. A. Miska, *Science* 341, 467-468 (2013).

22. Y. Goldwasser, W. T. Lanini, R. L. Wrobel, *Weed Science* 49, 520-523 (2001).

23. J. H. Westwood, J. K. Roney, P. A. Khatibi, V. K. Stromberg, *Pest Man. Sci.* 65, 533-539 (2009).

24. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat Meth* 5, 621-628 (2008).

25. S. Fischer *et al.*, in *Curr. Protoc. Bioinformatics*. (John Wiley & Sons, Inc., 2002).

26. K. C. Vaughn, *Intl J Plant Sci* 167, 1099-1114 (2006).

27. R. Deeken *et al.*, *Plant J* 55, 746 - 759 (2008).

28. X. Bai *et al.*, *PLoS ONE* 6, e16368 (2011).

29. C. Doering-Saad, H. Newbury, C. Couldridge, J. Bale, J. Pritchard, *J Exp Bot* 57, 3183 - 3193 (2006).

30. S. Guo *et al.*, *Nature genetics* 45, 51-58 (2013).

31. S. Huang *et al.*, *Nature genetics* 41, 1275-1281 (2009).

32. R. Ruiz-Medrano, B. Xoconostle-Cazares, W. Lucas, *Development* 126, 4405 - 4419 (1999).

33. D. J. Hannapel, P. Sharma, T. Lin, *Fron. Plant Sci.* 4, (2013).

34. J. Mower, S. Stefanovic, G. Young, J. Palmer, *Nature* 432, 165-166 (2004).

35. J. Mower *et al.*, *BMC Biol.* 8, 150 (2010).

36. L. Jiang *et al.*, *PLoS ONE* 8, e81389 (2013).

37. Y. Zhang *et al.*, *BMC Evol. Biol.* 13, 48 (2013).

38. D. Zhang *et al.*, *BMC Plant Biology* 14, 19 (2014).

39. S. Yoshida, S. Maruyama, H. Nozaki, K. Shirasu, *Science* 328, 1128 (2010).

40. P. Flicek *et al.*, *Nucleic Acids Res* 39, D800-D806 (2011).

41. J. E. Stajich *et al.*, *Nucleic Acids Res* 40, D675-D681 (2012).

42. J. J. Gillespie *et al.*, *Infect Immun* 79, 4286-4298 (2011).

43. S. Sato *et al.*, *Nature* 485, 635-641 (2012).

44. J. A. Martin, Z. Wang, *Nat. Rev. Genet.* 12, 671-682 (2011).

45. C. Iseli, C. Jongeneel, P. Bucher, in *Proc. Int. Conf. Intell. Syst. Mol. Biol.* (1999), pp. 138-148.

46. R. C. Edgar, *Bioinformatics* 26, 2460-2461 (2010).

Acknowledgments

Data are archived at NCBI SRA (BioProject ID: PRJNA257158). This research was supported by NSF award IOS-0843372 to JHW, and NSF Plant Genome Research Program awards DBI-0701748 and IOS-1238057 to JHW and CWD, with additional support to JHW from the U.S. Department of Agriculture Hatch Project no. 135997. We are grateful to T. Lanini for supplying *C. pentagona* seeds. Additional data are in supplement.

Chapter 4: Summary

An overview of host-*Cuscuta* RNA trafficking

The best-known examples of RNA trafficking between host and parasite are the movement of viruses through *Cuscuta*. The ability of *Cuscuta* to transmit viruses through bridges between infected and non-infected hosts has been known for over 60 years (Bennet, 1944). Also 57 viruses have been reported as transmitted through *Cuscuta* (Hosford, 1967). Recently, Birschwilks el al., (2006) demonstrated the transfer of potato virus Y isolate N (PVY^N) through *Cuscuta* bridge between two tobacco hosts and observed the high virus levels in donor and recipient hosts, but almost no virus detection in the *Cuscuta* bridge itself, suggesting that virus did not multiply during transmission.

It was previously demonstrated in the Westwood lab that tomato (*Solanum lycopersicum*) and pumpkin (*Curcubita maxima*) mRNAs were translocated into *Cuscuta* by using reverse transcriptase (RT)-PCR to amplify host RNAs extracted from *Cuscuta* tissue grown on these hosts (Roney et al., 2007). A tomato microarray was used to hybridize RNA from *Cuscuta* grown on tomato, resulting in the identification of 474 putatively mobile host transcripts, from which a sample of 24 transcripts were validated by RT-PCR and sequencing. Initially, because only a relativlty small set of transcripts were detected as mobile, it was interpreted that RNA translocation was selective rather than being the result of diffusion or leakage of host genes to parasite.

However, our recent research focussed on analyzing the interaction between *Cuscuta* and its tomato and *Arabidopsis* hosts using a genomics approach shows majority large number of *Arabidopsis* transcripts are found in *Cusucta* and that the movement of RNAs is bidirectional between host and parasite (LeBlanc *et al.*, 2013; Kim G *et al.*, 2014). This study has reshaped the

understanding of RNA trafficking between host and parasite. Transcriptome analyses of *Arabidopsis*, tomato, and *Cuscuta* with precisely harvested host stem, interface and parasite stem using RNA-seq provides more sensitive detection of mobile transcripts than does the microarray approach and also indicates gene expression

Previous studies indicated that the direction of RNA translocation was predominantly from hosts to parasite, although it was hypothesized that *Cuscuta* may send its own RNAs to the hosts (David-Schwartz *et al.*, 2008). We have observed that RNA movement in the "reverse direction" from *Cuscuta* to *Arabidopsis* was not even dramatically lower than host to parasite movement. This shows host and *Cuscuta* interacting with each other by allowing its own RNAs to move and by accepting foreign RNAs into their system.

One other question regarding RNA translocation is to understand the route by which they are moving. The key organ that makes connection between hosts and parasite is the haustorium, which forms open symplastic connections between its cells and those of the host. It has been demonstrated that proteins, RNAs and viruses are exchanged between hosts and parasite (LeBlanc *et al.*, 2012).

It has been speculated that *Cuscuta* takes up host RNAs through one of several potential routes: Through transfer cells via apoplast movement, through direct phloem-phloem connections, cell-to-cell plasmodesmata connections, or combinations of these. Our data suggest RNA movement between hosts and parasite may occur via a combination of these routes, but regardless it appears that most of RNAs move by an unregulated bulk flow process, which would consist of RNAs being taken up in the flow of solutes moving from the host plants to parasite.

In addition to bulk flow of mRNA from host to parasite, our research also suggests that RNA movement takes place in the fully spliced form. Roney et al., (2007) demonstrated that a

tomato transctipt, *LeMOB1* (encoding a gene of unknown function), that was mobile into *Cuscuta* was detected in tomato in both fully spliced and non-spliced forms. However, only the fully spliced form was detected in *Cuscuta* (Roney *et al.*, 2007). Furthermore, our recent RNA-seq data shows that host introns were only found in transcripts derived from host tissue, but in the parasite only the mature spliced transcripts were detected.

We have detected similar amounts of RNA moving between *Arabidopsis* and *Cuscuta* but when we compared the rates of transcript movement between *Cuscuta* with two different hosts, *Arabidopsis* and tomato, more extensive exchange was identified with *Arabidopsis*-*Cuscuta* connections, suggesting that the control of translocation of RNAs through haustoria may be host-specific. This idea is further supported by the experiment comparing *Cuscuta* infection of different hosts including *Arabidopsis*, tomato, sugarbeet, and *Medicago* (Kim and Westwood, unpublished data).

Once again, the *Arabidopsis*-*Cuscuta* connection showed much more efficient RNA exchange compared to the *Cuscuta*-tomato, -sugarbeet and -*Medicago* connection. Also, with sugarbeet and *Medicago* hosts we observed surprisingly small numbers of RNAs translocating from host to parasite. This suggests that *Arabidopsis* and *Cuscuta* form a unique connection to allow exchange of RNA.

The specific fate of mobile RNAs in another organism remains uncertain. To address this question we designed an experiment to assay translation of mobile RNA using a genomics approach involving extraction of ribosome-bound RNAs. Global analysis of translatome and ribosome footprints will provide us with some clues as to the fate of trafficked genes. If translocated RNA is not translated for production of proteins, then most likely it would be used as a resource for nucleotides following degradation.

Many studies have been performed on how mobile RNAs or mobile proteins can influence distant cells. Some RNAs that translocate have been shown to change morphology of leaf shape, potato tuber formation and influencing flowering time (Kim *et al.*, 2001; Haywood *et al.*, 2005; Banerjee *et al.*, 2006; Lu *et al.*, 2012). However, these genes were demonstrated in systemic RNA movement and by grafting systems involving only one species, and there is little evidence that parasitic plants will change their shape and morphology because of the host they are contacting. Some studies have suggested that *Cuscuta* flowering time depends on host flowering time, which raising a question that host FT RNA and protein may influence parasite flowering time (Corbesier *et al.*, 2007). Although this has not been proven, if a transcription factor RNA translocates and produces a functional protein in specific destination cells, it is possible that a low rate of transcript movement could trigger a new developmental program (Kragler, 2010).

Another important question from transcriptome data was whether there are distinctive properties of RNA that make it mobile or non-mobile. One of the characteristics we saw from our data was that mobile transcripts are highly expressed, especially in the interface region, however, this was not the only property to make them mobile because some similarly-expressed transcripts in the interface were not translocated.

One of the obvious and logical questions that arose after we found large-scale movement was what is the function of mobile RNA? We tried to find a common theme for mobile transcript function however, predicted gene functions fell into broad categories of gene ontology and did not present a clear picture. Because of the relatively small number of host mobile RNAs detected in the assays of multiple host, we search for common functions of these mobile RNAs. The list of host mobile genes we have found from *Medicago* with *Cuscuta* shows pectinesterase, pectate

lyase, and pectinesterase inhibitor, all of which function in cell wall modification. We also found that cell wall modification-related genes are mobile in *Arabidopsis* to *Cuscuta* and tomato to *Cuscuta*. Knowing that the majority of mobile RNAs are abundantly expressed near the point of attachment property, we speculate that genes related to cell wall modification are more prone to being picked up by parasite. Cell wall modification genes are also found in the parasite haustorium and this is a recurring theme in the interface between host and parasite. However, we should distinguish host cell wall modification transcripts which were observed as mobile in at least three species versus parasite cell wall modification genes where we don't know whether these parasite transcripts are also mobile. It will be very interesting to explore whether parasite cell wall modification-related transcripts are trafficking to host cells to loosen host cell walls.

One potentially-promising area of host-parasite communication is post transcriptional gene silencing. There is some evidence from the facultative root parasites, *Triphysaria versicolor* and *Phelipanche aegyptiaca* showing that parasite-specific gene silencing factors generated in transgenic hosts successfully entered into the parasites to suppress the parasite target genes (Aly *et al.*, 2007; Tomilov *et al.*, 2008). Furthermore, a recent study demonstrates silencing RNAi construct of *Cuscuta KNOX1* transcription factor, STM (SHOOT MERISTEMLESS-like) gene, express in tobacco hosts effectively suppressed its target in *Cuscuta pentagona* (Alakonya *et al.*, 2012).

Overall, it is still not clear what is the true mechanism of RNA translocation in haustorium function between host and parasite. Taken together, evidence for the mostly bulk-flow bidirectional RNA movement in *Arabidopsis*-*Cuscuta* system and the observation that trafficked RNA disappears eight hours after detachment of *Cuscuta* from tomato hosts (LeBlanc *et al.*, 2013) suggest that host and parasite exchange and process information through mobile

transcripts and also raises a interesting question that whether *Cuscuta* can distinguish self or non-self transcripts.

In conclusion, if RNA translocation between hosts and parasite is important, parasites evolved to adapt this interaction to benefit themselves, and in response the host would have to develop resistance mechanisms to fight against the RNA exchange. This work highlights RNA trafficking as a potentially important interaction between hosts and *Cuscuta*.

References

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. *The Plant Cell Online*.

Aly R, Shomin H, Joel D, Steinitz B, Zelcer A, Gal-On A 2007. A new approach to parasitic weed control based on silencing of a key metabolic gene in the parasite. *9th World Congress on Parasitic Plants*. Charlottesville, VA.

Banerjee AK, Chatterjee M, Yu Y, Suh S-G, Miller WA, Hannapel DJ. 2006. Dynamics of a Mobile RNA of Potato Involved in a Long-Distance Signaling Pathway. *Plant Cell* **18**(12): 3443-3457.

Bennet CW. 1944. Studies of dodder tranmission of plant viruses. *Phytopathology* **34**: 905-932.

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, et al. 2007. FT protein movement contributes to long-distance signalling in floral induction of Arabidopsis. *Science* **316**: 1030 - 1033.

Haywood V, Yu T, Huang N, Lucas W. 2005. Phloem long-distance trafficking of GIBBERELLIC ACID-INSENSITIVE RNA regulates leaf development. *Plant J* **42**: 49 - 68.

Hosford RMJ. 1967. Transmission of plant viruses by dodder. *Botan. Rev.* **42**: 387-406.

Kim G, LeBlanc M, Wafula E, dePamphilis C, JH W. 2014. Genomic-Scale Exchange of mRNA between a Parasitic Plant and its Hosts *Science* **345**(August).

Kim M, Canio W, Kessler S, Sinha N. 2001. Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* **293**: 287-289.

Kragler F. 2010. RNA in the phloem: A crisis or a return on investment? *Plant Science* **178**(2): 99-104.

LeBlanc M, Kim G, Patel B, Stromberg V, Westwood J. 2013. Quantification of tomato and Arabidopsis mobile RNAs trafficking into the parasitic plant Cuscuta pentagona. *New Phytologist* **200**(4): 1225-1233.

LeBlanc ML, Kim G, Westwood JH. 2012. RNA trafficking in parasitic plant systems. *Frontiers in Plant Science* **3**: 203.

Lu KJ, Huang NC, Liu YS, Lu CA, Yu TS. 2012. Long-distance movement of Arabidopsis FLOWERING LOCUS T RNA participates in systemic floral regulation. *RNA Biol* **9**(5): 653-662.

Roney JK, Khatibi PA, Westwood JH. 2007. Cross-species translocation of mRNA from host plants into the parasitic plant dodder. *Plant Physiol.* **143**(2): 1037-1043.

Tomilov AA, Tomilova NB, Wroblewski T, Michelmore R, Yoder JI. 2008. Trans-specific gene silencing between host and parasitic plants. *Plant Journal* **56**(3): 389-397.

Appendix

Evaluating the fate of host mobile RNA in *Cuscuta* by transcriptome, translatome and ribosome profiling

Abstract

Next generation sequencing technology is making the characterization of the genome, transcriptome, and translatome, as well as novel approaches such as ribosome footprinting possible in any organism. Genome-scale bidirectional mobile RNA exchange between *Arabidopsis* and the parasitic plant, *Cuscuta,* in a previous study prompted us to start investigating the potential function of non-autonomous host mobile RNAs using transcriptome, translatome, and ribosome profiling. The method also enables the study the fate of *Cuscuta* RNAs induced in response to host attachment. This is the first example of applying a translation assay to RNA trafficking in a host-parasite interaction. Here we describe an optimized protocol for generating transcriptome, translatome and ribosome profiling data from single source tissue of *Cuscuta* and the potential application of –omics methods for evaluating post-transcriptional regulation of both host and *Cuscuta* RNAs.

Keywords: RNA-Seq, transcriptome, translatome, ribosome footprinting, Sucrose density gradient, *Cuscuta pentagona*, mobile RNAs, ultracentrifugation.

Introduction

The obligate stem parasite, lespedeza dodder (*Cuscuta pentagona*), relies entirely on a host plant for water and nutrients, which are translocated through a special structure called a haustorium that connects *Cuscuta* searching hyphae to host vascular tissues. Previous research showed that *Cuscuta* takes up not only water and carbon compounds via haustoria, but also mRNAs (Roney *et al.*, 2007; David-Schwartz *et al.*, 2008; LeBlanc *et al.*, 2012), small RNAs (Alakonya *et al.*, 2012), viruses (Hosford, 1967; Birschwilks *et al.*, 2006), viroids (van Dorst & Peters, 1974) , phytoplasmas (Marcone *et al.*, 1997), DNAs (Mower *et al.*, 2004; Mower *et al.*, 2010), and proteins (Haupt, S *et al.*, 2001; Birschwilks *et al.*, 2006). In fact, we have recently discovered that exchange of mRNA between *Cuscuta* and Arabidopsis occurs on a massive scale (Kim et al., 2014), with 45% (9,518) of host expressed genes showed strong evidence of mRNA translocation into *Cuscuta*. This study underscores the importance of understanding the fate and function of mobile RNAs in cross-species interactions.

We have discovered significant bidirectional *Arabidopsis*, tomato, and *Cuscuta* RNA trafficking. There are several specific examples of potential functions of long distance mobile mRNA. The *Gibberellic acid-insensitive* (GAI) transcriptmoves long–distance in the phloem and regulates leaf development (Haywood *et al.*, 2005). StBEL-5 transcription factor RNA is produced in the leaf and translocates into the stolon tip of potato via phloem to promote tuber formation (Hannapel *et al.*, 2013). A PFP-LeT6 fusion transcript moves from stock to scion, influencing leaf morphology in tomato (Kim *et al.*, 2001). The flowering locus T (FT) protein (Jackson & Hong, 2012) and also the FT-RNA traffic systemically in *Arabidopsis* and may regulate floral initiation (Li *et al.*, 2011; Lu *et al.*, 2012). All these examples illustrate that long-distance trafficking of non-cell autonomous RNAs play a significant role in regulating plant

development and growth within plants. Lu et al. (2012) suggested that trafficking not only of FT RNA but protein may play a role in integrating photoperiodic signals. These findings raise questions about how the process of RNA translocation is regulated between parasitic plants and their hosts having the effects of the foreign transcripts on the recipient plants. The work presented here lays the groundwork for understanding the route of mobile RNAs, their stability, and the ultimate fate of non-autonomous RNAs in the host-*Cuscuta* system.

After host RNAs translocate into *Cuscuta*, they could be translated and or be degraded (LeBlanc *et al.*, 2013). We have described that the levels of a Cathepsin D Proteinase inhibitor (*SIPI*) transcript of tomato taken up by *Cuscuta* gradually decreased in over eight hours after the *Cuscuta* stem was detached from the host. This could reflect translation-associated turnover of the transcript. However, there is no evidence yet available to indicate that proteins result from translation of non-autonomous RNAs. It is also possible that mobile RNAs bound to chaperone proteins play regulatory roles to influence gene expression in *Cuscuta*. In order to address these questions, we propose to understand the fate of mobile RNAs by generating transcriptome, translatome, and ribosome profiling data for *Arabidopsis* RNAs found in *Cuscuta* tissue. *Arabidopsis* mobile RNAs indicate evidence of the largest-scale movement into *Cuscuta* and *Arabidopsis* has the best-annotated genome, facilitating host-parasite transcripts informatically.

Demonstrating that individual mobile genes undergo translation is a challenging task and it is risky to select only a few genes for assaying translation, so is not efficient in terms of time, energy, and resources. Also, the fact that past studies reveal that the proportion of individual genes undergoing translation ranges between 10% to 95% under relatively stable conditions, indicates that detection of translation of a given gene is more likely with a global-scale approach (Branco-Price *et al.*, 2005; Kawaguchi & Bailey-Serres, 2005). The post-transcriptional

regulation of individual genes can be changed by both biotic and abiotic factors such as hypoxia (Branco-Price *et al.*, 2005; Branco-Price *et al.*, 2008), dehydration (Kawaguchi *et al.*, 2004; Kawaguchi & Bailey-Serres, 2005; Park *et al.*, 2012), light (Juntawong & Bailey-Serres, 2012; Liu *et al.*, 2012), heat (Matsuura *et al.*, 2013), sucrose starvation (Nicolai *et al.*, 2006; Rahmani *et al.*, 2009), excess cadmium (Sormani *et al.*, 2011), cold (Park *et al.*, 2012; Juntawong *et al.*, 2013), saline stress (Matsuura *et al.*, 2010), phytohormones (Ribeiro *et al.*, 2012; Rosado *et al.*, 2012; Schepetilnikov *et al.*, 2013), singlet oxygen (Khandal *et al.*, 2009) and pathogens (Moeller *et al.*, 2012). Among these comprehensive analyses of complex situations, several from Ingolia et al. (2012) and the Bailey-Serres group (Bailey-Serres *et al.*, 2009; Juntawong *et al.*, 2014), used genome-wide profiling of ribosome footprints in yeast and in *Arabidopsis* using next generation sequencing to establish a precise experimental approach to understanding regulation of protein synthesis and measuring gene expression at the translational level. Therefore, we decided to adapt their protocol to study mobile RNA translation in the *Arabidopsis*-*Cuscuta* system.

Translatomes can be isolated through extracting RNAs that are associated with ribosomes and polysomes by ultra-speed centrifugation of the sample in a sucrose density gradient. mRNA undergoes translation in three phases: initiation, elongation and termination. The initiation step involves the mRNA recruiting small (40S) and large (60S) subunits of ribosomes for launching peptidyl chain elongation. In this process, a polysome can be made as a result of serial initiation of many ribosomes bound to a single mRNA. The next step is polypeptide elongation. After the elongation phase is complete by reaching a stop codon, the polypeptide chain is released and the 80S ribosome dissociates from strand of mRNA. In most cases, detecting mRNA associated with

polysomes is a good indication of translation (Bailey-Serres, 1999; Kawaguchi & Bailey-Serres, 2002; Browning, 2004).

To assess translation of mRNAs, it is important to determine the quantity of the polysomal RNAs compared to the total transcript amount. In order to obtain enough polysomal host RNAs that have translocated into *Cuscuta*, we isolated polysomal RNAs from *Cuscuta* stems that had made good haustorial connections with the host. Polysomal RNA extraction buffer was prepared with heparin, which inhibits RNase activity. Magnesium chloride was used in the buffer for stabilizing the two subunits of the ribosome complex. Also, cycloheximide and chloramphenicol was included in the buffer to prevent run-off of cytosolic and organellar ribosomes, respectively. Ultra-centrifugation of samples in sucrose continuous density gradient was used to fractionate mRNAs bound with polysomes and monosomes from free mRNA and other less dense molecules (Mustroph *et al.*, 2009). This protocol has been validated with many different plants including maize, tobacco, tomato and Arabidopsis for the evaluation of translational regulation (Fennoy *et al.*, 1998; Williams *et al.*, 2003; Kawaguchi *et al.*, 2004; Branco-Price *et al.*, 2005; Slaymaker & Hoppey, 2006).

In addition to assaying the translatome, we also conducted ribosome profiling based on next generation sequencing of ribosome-protected mRNA fragments. Ingolia et al. (2011) developed this method and revealed that translating ribosomes protect mRNA from nuclease digestion, resulting in RNA fragments ("footprints") with sizes about 30 nt. The combination of deep sequencing technology and the method to map the positions of ribosomes on mRNA in an *in vitro* translation assay results in a global ribosome profiling (Ribo-seq), which makes it feasible to understand regulation and dynamics of translation *in vivo* (Ingolia *et al.*, 2012). The short fragment RNA (similar to miRNA-seq) reads protected by ribosomes can be mapped to

genomes for understanding post-transcriptional regulation of genes in a high-throughput way. The yeast and mammal research communities have already validated this method to characterize the association between transcriptomes and ribosome profiling of their species (Ingolia *et al.*, 2009; Ingolia *et al.*, 2011; Brar *et al.*, 2012). In plants, as mentioned previously, the Bailey-Serres' group demonstrated that comparison of total RNA (transcriptome), polysomal RNA (translatome) and ribosome profiling (footprinting) shows various aspects of posttranscriptional and translational control under both optimal and low-oxygen conditions in *Arabidopsis*. (Juntawong *et al.*, 2014).

Based on our previous mobile RNA data analysis we initially have generated from two replicates of *Arabidopsis*-*Cuscuta* system which resulted in massive scale of RNA translocation bidirectionally, we thought it would be reasonable to assay translation of mobile RNAs in high throughput way instead of using tagged *Arabidopsis* lines that Bailey-serres group generated for translation assay. Due to the laborious, time-consuming, and problematic biochemical assay for generating translatome and ribosome profiling data, the goal for this chapter is to show validation of our protocol and indicate initial sequence statistics of our recently received transcriptome, translatome and ribosome profiling sequencing results. Future analysis of this data should reveal whether the host mobile RNAs undergo translation in *Cuscuta*, which should be detectable as mapping of host RNAs from the translatome and ribosome profiling sequences in *Cuscuta*. In the case that there is no detection of host mobile RNAs in the translatome and ribosome profiling data, an argument may be made that host transcripts are not translated in the parasite. In any event, obtaining the *Cuscuta* transcriptome, translatome, and ribosome profiling data by itself will provide significant information about parasite transcript regulation. This

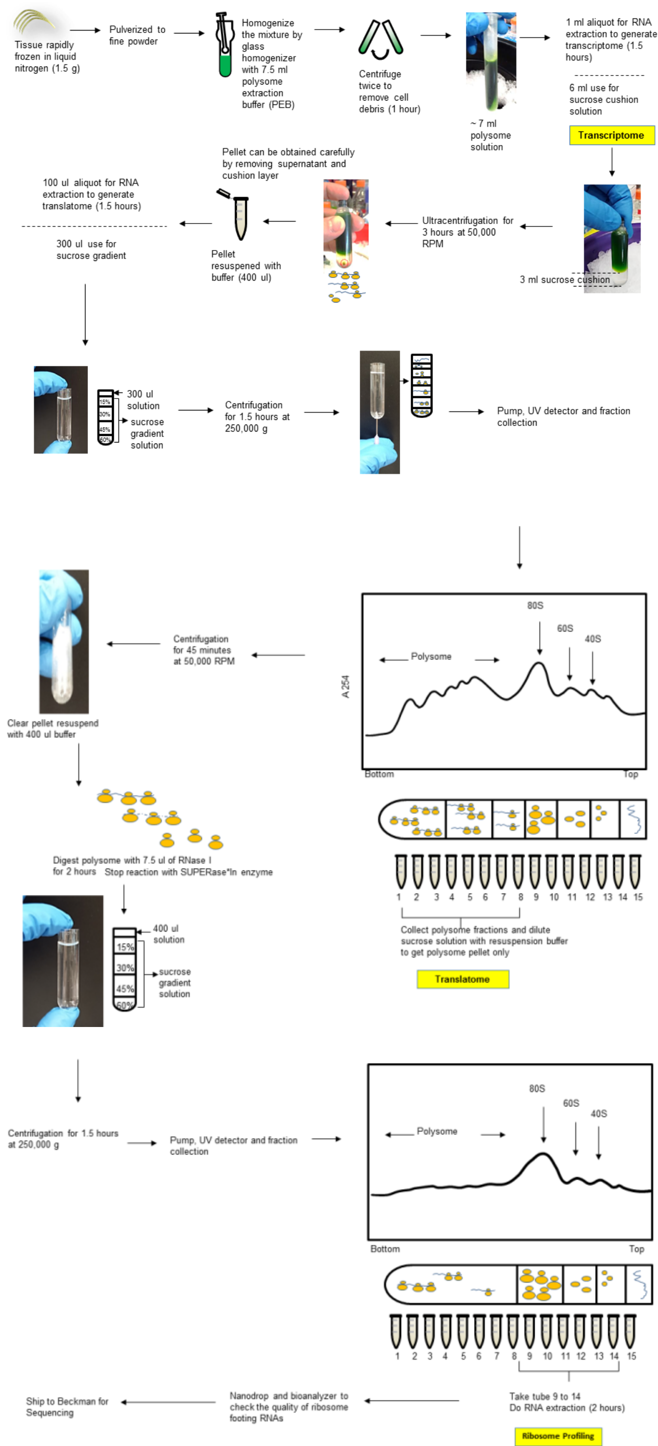
experiment offers a new approach to genome-wide analysis of RNA pools and is the first application of this technology in a host-parasite system.

We decided to assay the entire *Cuscuta* stem from near the point of *Arabidopsis*-host attachment to the dodder apical tip because of uncertainty as to where the translation may be occurring (i.e. near the point of attachment vs. in the shoot meristem). The relatively low cost and high efficiency of next generation (RNA-Seq) technology allowed us to generate three replicates from host mobile transcriptome, translatome, and ribosome profiling in *Cuscuta* tissue.

Results and discussion

Experimental strategy to validating translatome and ribosome footprinting protocol

Extraction of translating RNAs in the cell can be accomplished by sucrose gradient centrifugation and fractionation (polysome gradients), which allows the separation of ribosome-free mRNAs from ribosome-bound mRNAs. Figure A.1 shows the entire pipeline of RNA extraction for transcriptome, translatome, and ribosome profiling from a single source tissue.



Tissue rapidly frozen in liquid nitrogen (1.5 g)
Pulverized to fine powder
Homogenize the mixture by glass homogenizer with 7.5 ml polysome extraction buffer (PEB)
Centrifuge twice to remove cell debris (1 hour)
~ 7 ml polysome solution
1 ml aliquot for RNA extraction to generate transcriptome (1.5 hours)
6 ml use for sucrose cushion solution
Transcriptome
3 ml sucrose cushion
Ultracentrifugation for 3 hours at 50,000 RPM
Pellet can be obtained carefully by removing supernatant and cushion layer
Pellet resuspened with buffer (400 ul)
100 ul aliquot for RNA extraction to generate translatome (1.5 hours)
300 ul use for sucrose gradient
300 ul solution
sucrose gradient solution
Centrifugation for 1.5 hours at 250,000 g
Pump, UV detector and fraction collection
80S
60S
40S
Polysome
A 254
Bottom
Top
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Collect polysome fractions and dilute sucrose solution with resuspension buffer to get polysome pellet only
Translatome
Centrifugation for 45 minutes at 50,000 RPM
Clear pellet resuspend with 400 ul buffer
Digest polysome with 7.5 ul of RNase I for 2 hours Stop reaction with SUPERase*In enzyme
400 ul solution
15%
30%
45%
60%
sucrose gradient solution
Centrifugation for 1.5 hours at 250,000 g
Pump, UV detector and fraction collection
80S
60S
40S
Polysome
Bottom
Top
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Take tube 9 to 14
Do RNA extraction (2 hours)
Ribosome Profiling
Nanodrop and bioanalyzer to check the quality of ribosome footing RNAs
Ship to Beckman for Sequencing

Figure A.1. Schematic diagram of the transcriptome, translatome and ribosome footprinting pipeline.

Because we are initially assessing translation of translocated host RNAs in *Cuscuta* tissue, we modified a method used by the Bailey-Serres group (Juntawong *et al.*, 2014) to carry out prepare the transcriptome, translatome, and ribosome footprinting. The tissue consisting of entire *Cuscuta* stem was carefully excised from near the point of attachment to the apical tip of *Cuscuta* stem and frozen in liquid nitrogen. More than 1.5 g of sample tissue was pulverized under liquid nitrogen and polysomal extraction buffer (lysis buffer) was added to the sample. Since optimized protocol generated from Juntawong et al. (2014) showed RNasin, a strong RNAse inhibitor, alone was not sufficient to protect RNAs from RNAse activity, we included an additional RNAse inhibitor, heparin. The efficiency of heparin has been investigated in the protocol of extracting brain polysomes and demonstrated increased production of the polysome fraction (Gauthier & Murthy, 1987). In a pilot experiment, we observed a large difference in polysomal RNA yield using extraction buffer containing heparin compared to samples prepared without heparin.

Tissue in polysomal extraction buffer was transferred to centrifuge tubes for centrifugation to separate solution from cell debris. The centrifugation (15,000 x g) was performed twice. Part of the filtered sample solution was used for total RNA preparation, which was carried out with Trizol Reagent including DNase I treatment to remove DNA contamination for transcriptome analysis. The rest of the sample was transferred to an ultracentrifuge tube containing a 1.75 M sucrose cushion and centrifuged (170,000 x g) to obtain pellets comprising

ribosome subunits, ribosomes and polysomes. The pellet was resuspended in a buffer containing magnesium chloride (two ribosome subunit stabilizer), cycloheximide (inhibition of further translation) (Baliga *et al.*, 1969), and chloramphenicol (organelle—specific translation inhibitor) (Ross, 1996). The resuspended solution was loaded onto a 15-60% sucrose gradient. During ultracentrifugation, mRNAs disperse into the different fractions of the gradient based on their molecular size and number of ribosome subunits (i.e., monosomes and polysomes). The resulting sucrose density gradient was fractionated by drawing contents from the bottom of the tube and continuously recording the relative ribosome absorbance at 254nm to differentiate each fraction. A BioRad Econo peristaltic pump and LP BioRad UV detector with recorder was used to record and fractionate the dispersed gradient solutions. The identity of polysome fractions in the gradient was confirmed using a presence and absence of 25mM EDTA control experiment in the resuspension buffer since EDTA dissociates ribosomes from mRNA (He *et al.*, 1993). We monitored polysomes with 35 mM of $MgCl_2$ containing resuspension buffer, 35 mM of $MgCl_2$ containing buffer only, polysomes with 25mM of EDTA containing resuspension buffer and 25 mM of EDTA containing buffer only. In the presence of $MgCl_2$, as in Figure A.2A (green line), the peak of polysome fractions was more prominent compared to 40S, 60S and 80S fractions. In the presence of EDTA, the 60S and 40S fraction peak was more prominent and monosome and polysome fraction peaks decreased substantially. Also, we observed that control samples containing only $MgCl_2$ or EDTA resuspension buffer showed low baseline noise from the sucrose gradient solution. Figure A.2B demonstrates clearer confirmation of monosomes and polysome dissociation by subtracting buffer only baselines from the corresponding samples treated with or without EDTA. The stability of EDTA treated and non-treated polysomes has been monitored and demonstrated with UV profile of absorbance at 254 nm in previous

experiments (Farny *et al.*, 2008; Tcherkezian *et al.*, 2010). This control experiment provide us information about potential polysome degradation during the extraction procedure and guided decisions on how many fractions to collect for extracting polysomal RNA for the translatome assay (fig. A.2).

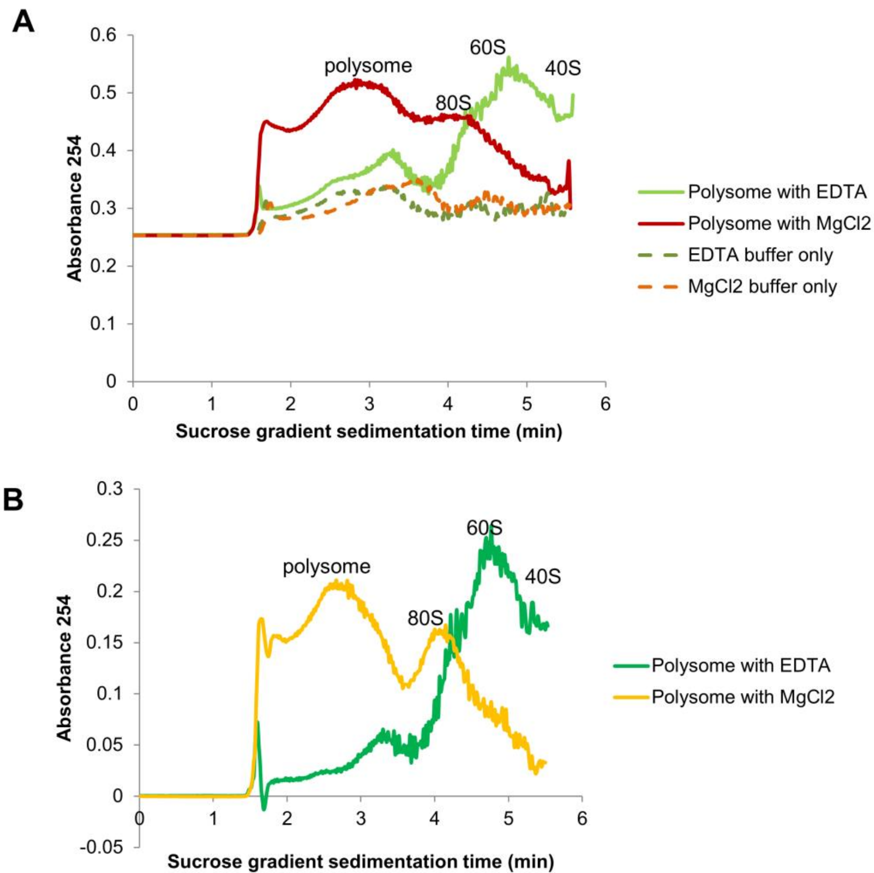


Figure A.2. Identification of polysome, ribosome and ribosome subunits using EDTA control experiment to dissociate the polysomes from mRNA. A. In the presence of EDTA (light green), dissociation of ribosome occurs with monosome and polysome but in the presence of $MgCl_2$ (red), stabilization of polysome was observed. The dotted line represents EDTA (dark green)

and $MgCl_2$ (light red) buffer-only measurements, respectively. B. Buffer baseline was subtracted from panel A to better illustrate the significant differences between polysome treated with EDTA (green) verses $MgCl_2$ (orange) containing buffers.

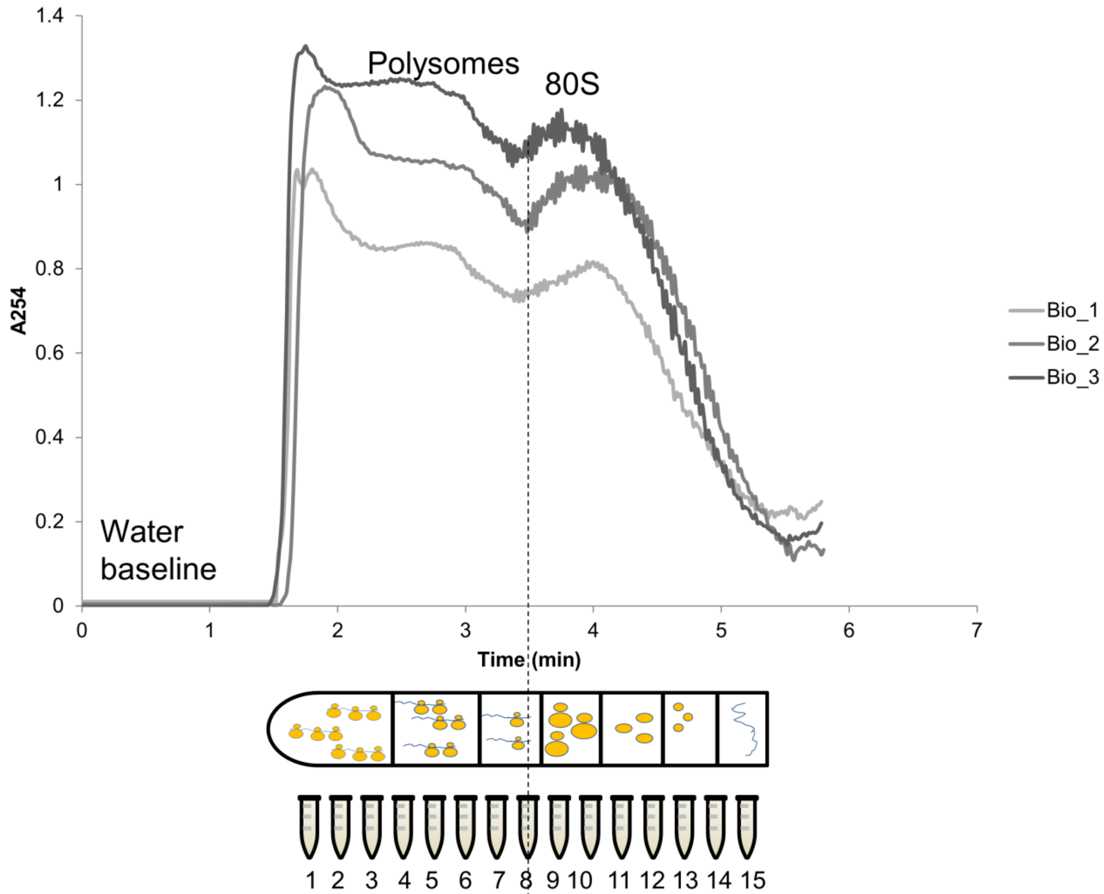


Figure A.3. Polysome extraction from three biological replicates of host and *Cuscuta* tissue by sucrose density gradient. Tubes 1-8 that contained ribosomes, polysomes, and monosomes were collected and treated with nuclease digestions. The entire swing rotor tube contained almost 4.5 ml of polysomes in sucrose gradient buffer solution. The tube was punctured from the bottom with a syringe needle that as connected via a sterile tube to a UV detector for measuring

polysome material. Each collection tube took approximately 20 seconds (about 15 drops of gradient solution) and A254 were detected using LP BioRad UV monitor recorder.

A total of 15 fractions were collected from each sucrose gradient and tubes 1 to 8 were pooled to represent the polysomal material. From this an aliquot was taken for polysomal RNA extraction and another used for ribosomal footprinting. The total sucrose gradient solution containing different sizes of molecules was about ~4.5 ml and with calculating ratio (1ml/min with 0.8 mm tube) of sample going through UV detector and collection tube, we decided to collect ~ 300ul (15 drops) in microcentrifuge tubes to make a total of 15 fractions. The polysome content of the pooled fractions (tubes 1-8) was confirmed insubsequent ribosome footprinting analysis.

As described in Figure A.3, three biological replicates demonstrated a similar pattern of polysome fraction isolation as detected with a UV monitor of the sucrose gradient sedimentation pattern. Because we started with a large amount of material (a total of 6 g combining four 1.5g samples of tissue) compared to the control experiment shown in figure A2, we detected a high abundance of polysomes at A254 nm with all three replicates. We performed RNA-seq analysis on both transcriptome (before fractionation) and translatome samples to ensure complete detection of mapped reads on both sets to investigate both the fate of host mobile RNA as well as translation of *Cuscuta* mRNAs.

After polysomal RNA isolation, the rest of the pooled polysome fractions were collected by additional ultracentrifugation (250,000 g) to extract the ribosome-protected region of the mRNAs. The collected polysomes in resuspension buffer were treated with nuclease to digest any mRNA regions not protected by ribosomes. The reaction was stopped after 2 hours and the

solution was loaded onto a sucrose density gradient for separation as described above. Figure A.4 shows the high proportion t of monosome peak affected by nuclease digestion. We also observed some undigested polysome content in the gradient. We suspect that the ratio of polysome content and digestion enzyme could be optimized to increase the efficiency of digestion. However, that a prominent monosome peak was evident compared to polysomes in figure A4 confirms successful nuclease digestion to yield sequences of ribosome footprinting. The ribosome profiling not only indicates which transcript is being translated, but also gives information of ribosome position on the transcript (Ingolia *et al.*, 2012). The comprehensive ribosome footprinting will thus further inform the potential mechanism of host mobile RNA and *Cuscuta* RNA translation.

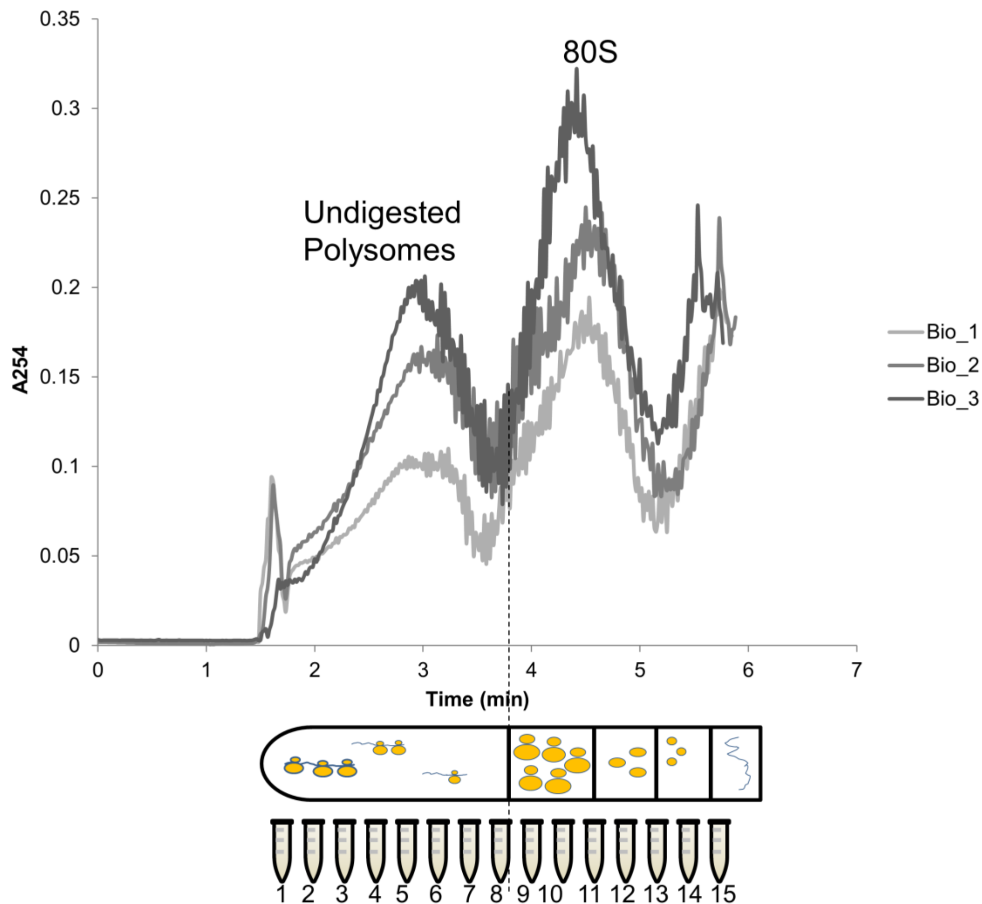


Figure A.4. Ribosome protected mRNAs were treated with RNase I (nuclease digestion) and UV monitor shows content of polysomes and 80S monosomes.

Three replicates of each of ribosome protected RNAs (~28-30 bp) were isolated with TriReagent according to manufactures protocol (Trizol and chloroform method). The RNA quality was checked with miRNA bioanalyzer and we observed RNA size ranges from 20 nt to 150 nt. Typically, it is possible to have tRNAs, small RNAs and some residual RNAs that are not efficiently treated with nuclease digestion. However, after the library was constructed by ligating adapters and other primers (~118bp additional sequences), size selection process was performed to excise out only those with ~30bp insert size libraries (about 150 bp total size). Figure A.5 shows bioanalyzer assay of pre- and post-size selection.

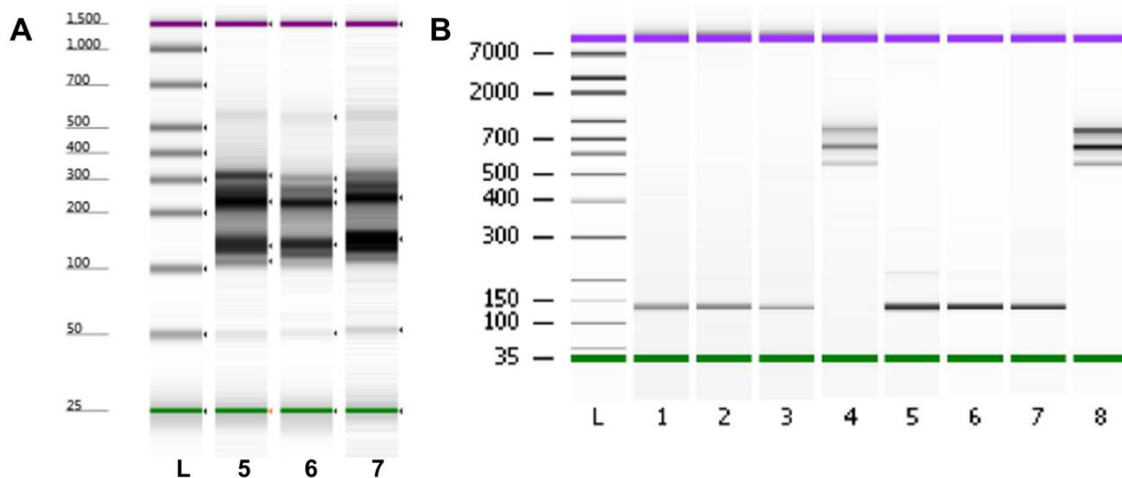


Figure A.5. Ribosome-protected mRNA library size selection quality check with bioanalyzer. A) pre-size selection gel shows various sizes of ribosome protected RNAs, tRNA and other molecules. B) post-size selection gel indicates only ribosome protected RNAs. L- DNA ladder; 1 – 1:5 diluted replicate 1; 2 – 1:5 diluted replicate 2; 3 – 1:5 diluted replicate 3; 4 - Regular RNA-seq control; 5 – replicate 1; 6 – replicate 2; 7 – replicate 3; 8 – Regular RNA-seq control.

For Illumina sequencing, more than 561 million high quality (>Q30) reads with 100bp paired –end data were generated from the transcriptome and translatome libraries and we obtained ~ 246 million high quality reads with 50bp single-end data from ribosome footprinting. These raw data will be pre-processed with pipeline that we applied in Chapter 3 (Kim et al., 2014). The host mobile RNAs and *Cuscuta* RNAs that undergo translation will be identified with bioinformatics. We anticipate that the transcriptome, translatome and ribosome footprinting approach will provide important new information regarding translation of RNAs in *Cuscuta* stem tissue.

Materials and Methods

Plant material

Arabidopsis thaliana (L.) Heynh. was grown in a growth chamber at 18-20°C with 12-h per day light cycle, illuminated (200 μmol $m^{-2}s^{-1}$) with metal halide (400W, GE multi-vapor lamp) and spot-gro (65W, Sylvania) lamps. *S*eeds of *Cuscuta pentagona* (Engelm.) were scarified by soaking in concentrated sulfuric acid for 45 min, rinsed with water and dried. Seeds were placed in potting medium at the base of four-week-old *Arabidopsis* allowed to germinate and attach to hosts. *Cuscuta* seedlings germinated and established connections with *Arabidopsis* plants and were allowed to grow and spread on host plants for an additional three weeks. About 10-15 cm from apical tip to base of *Cuscuta* seedlings were collected and used to make synchronized secondary attachments to new 4-5-weeks-old *Arabidopsis*. By this method, large numbers of uniform *Cuscuta* seedlings were generated and entire *Cuscuta* seedlings from near host attachment site were harvested and frozen into liquid N_2 immediately. Three biological samples (total 6 g of pooled *Cuscuta* tissue per each biological replicate) were harvested for total, polysomal mRNA and ribosome profiling extraction.

Preparation of polysomal RNA and fractionation

Four 1.5 g of pooled frozen *Cuscuta* tissue was pulverized to a fine powder with mortar and pestle in liquid N_2. 7.5 ml of freshly prepared polysomal extraction buffer [PEB; 200mM Tris-HCl (pH 9.0), 200 mM KCl, 25 mM EGTA, 35 mM $MgCl_2$, 1% (w/v) polyoxyethylene (23) lauryl ether (Brij-L23), 1% (v/v) Triton X-100, 1% (v/v) octylphenyl-polyethylene glycol (Igepal CA 630), 1% (v/v) Tween-20, 1% Sodium deoxycholate (DOC), 1% polyoxyethylene 10 tridecyl ether (PTE), 5 mM DTT, 2 mM phenylmethanesulfonylfluoride (PMSF), 50 μg/mL cycloheximide, 50 μg/mL chloramphenicol, 500 mg/ml Heparin, 0.5 μl/mL RNAsin (Promega)] was mixed with pulverized powder and further homogenized the mixture with glass homogenizer (Juntawong *et al.*, 2014). The homogenized mixture was placed on ice for 10 minutes and all cell debris was removed by centrifugation at 16,000 x g at 4 °C for 15 minutes. The supernatant were filtered using sterilized Miracloth (Calbiochem) and centrifuged one more time at 16,000 x g at 4 °C for 15 minutes. Four 1 ml aliquot from each filtered supernatant was mixed together and saved for isolating total RNA. The filtered supernatant was added on top of a 1.75 M of a sucrose cushion solution [400 mM Tris-HCl (pH 9.0), 200 mM KCl, 5mM EGTA, 35 mM $MgCl_2$, 5 mM DTT, 50 μg/mL cycloheximide, 50 μg/mL chloramphenicol, 0.5 μl/mL RNAsin (Promega)] and centrifuged at 213,818 x g at 4 °C for 3 hours in a type 90 Ti rotor (Beckman) to collect a polysomal pellet. After ultracentrifugation, the supernatant and sucrose cushion layer was removed and the polysomal pellet was resuspended with 400 μl of resuspension buffer [20 mM Tris-HCl (pH 8.0), 140 mM KCl, 35 mM $MgCl_2$, 50 μg/mL cycloheximide, 50 μg/mL chloramphenicol]. 400 μl of resuspended polysomal pellet solution was added on top of 15-60% (wt/v) of 2M sucrose gradient solution and centrifuged at 250,000 x g at 4 °C for 1.5 hours in a MLS-50 rotor (Beckman). Gradient fractions were evaluated and collected by using BioRad Econo peristaltic pump, LP-BioRad UV detector and LP Data-view v.1.03 recorder. Fractions

(tubes 1-8) collected with ribosome subunits, ribosomes and polysome complexes were mixed together and 500 µl of pooled samples were aliquoted for isolation of polysomal RNAs. Total and polysomal RNA was isolated using 1mL of TRIzol reagent (Invitrogen) according to the manufacture's protocol. TRIzol-chloroform extracted RNAs were cleaned and concentrated (Zymo kit) with DNase I treatment (Qiagen). The mixed fractions were further processed for ribosome profiling.

Isolation of ribosome protected RNAs

The mixed fractions containing sucrose density solution were diluted with resuspension (1:1.6 ratio) buffer and centrifuged at 250,000 x g at 4°C for 45 minutes in MLS-50 rotor (Beckman) to collect polysomal RNA pellet. The supernatant was removed and clear pellet was resuspended with 400 µl of resuspesion buffer. The solution was treated with 7.5 µl RNase I (100 U/ µl, Ambion) for 2 hours of nuclease digestion at room temperature to obtain ribosome protected RNA with gentle mixing on a nutator. The reaction was stopped by adding 10 µl of SUPERase*In (Ambion) RNase inhibitor. The digested sample was layered on top of 15-60% sucrose gradient solution and centrifuged at 250,000 x g at 4 °C for 1.5 hours in a MLS-50 rotor (Beckman). Gradient fraction was evaluated and collected once again by using BioRad econo peristaltic pump, LP-BioRad UV detector and LP Data-view v.1.03 recorder. Fractions (tubes 9-14) containing 80S monosome and the ribosomes obtained by digestion were pooled for RNA extraction using TRIzol reagent according to the manufacturer's protocol. The isolated RNA (Ribosome profiling) was also cleaned and treated with DNase I.

Library construction of transcriptome, translatome and ribosome profiling.

Each replicated samples were evaluated for quality using the Agilent Bioanalyzer 2100 Plant RNA Nano assay, as well as quantity by Invitrogen Quant-iT™ RNA Assay Kit. 250 ng of

high quality total RNAs were input into the Illumina TruSeq RNA Sample Preparation Kit v2 protocol (transcriptome and translatome)/TrueSeq small RNA sample prep (ribosome footprinting) and run in conjunction with Beckman Coulter Life Sciences' Biomek liquid handling platforms. This uses Illumina's Universal TruSeq adaptor, plus the indexes as described in the Illumina library adapter/index lists. Libraries were evaluated by Agilent 2200 TapeStation High Sensitivity D1K Tape and qPCR with the KAPA Library Quantification Kit for Illumina Sequencing. Target sized selection for next generation sequencing (Pippin Prep) was used to isolate ribosome protected mRNAs. 2nM of library sample was loaded for Cluster Generation and Sequenced on the Illumina HiSeq2500 Sequencing System at 2x100bp read length plus index read for transcriptome and translatome and 1x50bp read plus index read for ribosome footprinting. Passing Filter reads were determined and fastq files generated using Illumina CASAVA v1.8.

References

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. *The Plant Cell Online*.

Aly R, Shomin H, Joel D, Steinitz B, Zelcer A, Gal-On A 2007. A new approach to parasitic weed control based on silencing of a key metabolic gene in the parasite. *9th World Congress on Parasitic Plants*. Charlottesville, VA.

Bai X, Rivera-Vega L, Mamidala P, Bonello P, Herms D, Mittapalli O. 2011. Transcriptomic signatures of ash (*Fraxinus* spp.) phloem. *PLoS ONE* **6**(1): e16368.

Bailey-Serres J. 1999. Selective translation of cytoplasmic mRNAs in plants. *Trends in Plant Science* **4**(4): 142-148.

Bailey-Serres J, Sorenson R, Juntawong P. 2009. Getting the message across: cytoplasmic ribonucleoprotein complexes. *Trends in Plant Science* **14**(8): 443-453.

Baliga BS, Pronczuk AW, Munro HN. 1969. Mechanism of cycloheximide inhibition of protein synthesis in a cell-free system prepared from rat liver. *J Biol Chem* **244**(16): 4480-4489.

Banerjee AK, Chatterjee M, Yu Y, Suh S-G, Miller WA, Hannapel DJ. 2006. Dynamics of a Mobile RNA of Potato Involved in a Long-Distance Signaling Pathway. *Plant Cell* **18**(12): 3443-3457.

Bennet CW. 1944. Studies of dodder tranmission of plant viruses. *Phytopathology* **34**: 905-932.

Birschwilks M, Haupt S, Hofius D, Neumann S. 2006. Transfer of phloem-mobile substances from the host plants to the holoparasite *Cuscuta* sp. *J. Exp. Bot.* **57**: 911-921.

Birschwilks M, Sauer N, Scheel D, Neumann S. 2007. *Arabidopsis thaliana* is a susceptible host plant for the holoparasite *Cuscuta* spec. *Planta* **226**(5): 1231-1241.

Branco-Price C, Kaiser KA, Jang CJ, Larive CK, Bailey-Serres J. 2008. Selective mRNA translation coordinates energetic and metabolic adjustments to cellular oxygen deprivation and reoxygenation in Arabidopsis thaliana. *Plant J* **56**(5): 743-755.

Branco-Price C, Kawaguchi R, Ferreira RB, Bailey-Serres J. 2005. Genome-wide analysis of transcript abundance and translation in Arabidopsis seedlings subjected to oxygen deprivation. *Ann Bot* **96**(4): 647-660.

Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. 2012. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science* **335**(6068): 552-557.

Browning KS. 2004. Plant translation initiation factors: it is not easy to be green. *Biochemical Society Transactions* **32**: 589-591.

Corbesier L, Vincent C, Jang S, Fornara F, Fan Q, Searle I, Giakountis A, Farrona S, Gissot L, Turnbull C, et al. 2007. FT protein movement contributes to long-distance signalling in floral induction of Arabidopsis. *Science* **316**: 1030 - 1033.

Czechowski T, Bari RP, Stitt M, Scheible WR, Udvardi MK. 2004. Real-time RT-PCR profiling of over 1400 Arabidopsis transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes. *Plant J* **38**(2): 366-379.

David-Schwartz R, Runo S, Townsley B, Machuka J, Sinha N. 2008. Long-distance transport of mRNA via parenchyma cells and phloem across the host-parasite junction in *Cuscuta*. *New Phytol* **179**(4): 1133-1141.

Deeken R, Ache P, Kajahn I, Klinkenberg J, Bringmann G, Hedrich R. 2008. Identification of Arabidopsis thaliana phloem RNAs provides a search criterion for phloem-based transcripts hidden in complex datasets of microarray experiments. *Plant J* **55**: 746 - 759.

Doering-Saad C, Newbury H, Couldridge C, Bale J, Pritchard J. 2006. A phloem-enriched cDNA library from Ricinus: insights into phloem function. *J Exp Bot* **57**: 3183 - 3193.

Dorn KM, Fankhauser JD, Wyse DL, Marks MD. 2013. De novo assembly of the pennycress (Thlaspi arvense) transcriptome provides tools for the development of a winter cover crop and biodiesel feedstock. *The Plant Journal* **75**(6): 1028-1038.

Dörr I 1990. Sieve elements in haustoria of parasitic angiosperms. In: Behnke H-D, Sjölund R-D eds. *Sieve Elements: Comparative Structure, Induction and Development*. Berlin: Springer, 239-253.

Duan JL, Xia C, Zhao GY, Jia JZ, Kong XY. 2012. Optimizing de novo common wheat transcriptome assembly using short-read RNA-Seq data. *Bmc Genomics* **13**.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**(19): 2460-2461.

Fan HK, Xiao Y, Yang YD, Xia W, Mason AS, Xia ZH, Qiao F, Zhao SL, Tang HR. 2013. RNA-Seq Analysis of Cocos nucifera: Transcriptome Sequencing and De Novo for Subsequent Functional Genomics Approaches. *Plos One* **8**(3).

Farny NG, Hurt JA, Silver PA. 2008. Definition of global and transcript-specific mRNA export pathways in metazoans. *Genes Dev* **22**(1): 66-78.

Fennoy SL, Nong T, Bailey-Serres J. 1998. Transcriptional and post-transcriptional processes regulate gene expression in oxygen-deprived roots of maize. *Plant Journal* **15**(6): 727-735.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, et al. 2013. Ensembl 2013. *Nucleic Acids Res* **41**(Database issue): D48-55.

Furuhashi T, Furuhashi K, Weckwerth W. 2011. The parasitic mechanism of the holostemparasitic plant Cuscuta. *Journal of Plant Interactions* **6**(4): 207-219.

Gauthier D, Murthy MRV. 1987. Efficacy of Rnase Inhibitors during Brain Polysome Isolation. *Neurochemical Research* **12**(4): 335-339.

Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, et al. 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect Immun* **79**(11): 4286-4298.

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. *Nature Biotechnology* **29**(7): 644-U130.

Guo S, Zhang J, Sun H, Salse J, Lucas WJ, Zhang H, Zheng Y, Mao L, Ren Y, Wang Z, et al. 2013. The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. *Nat Genet* **45**(1): 51-58.

Haidar MA, Boss WF. 2009. Blue light induced changes in inositol 1,4,5-trisphosphate in <i>Cuscuta campestris</i> seedlings. *Weed Research* **49**(6): 628-633.

Hannapel DJ, Sharma P, Lin T. 2013. Phloem-mobile messenger RNAs and root development. *Front Plant Sci* **4**: 257.

Haupt S, Oparka K, Sauer N, Neumann S. 2001. Macromolecular trafficking between Nicotiana tabacum and the holoparasite Cuscuta relexa. *J Exp Bot* **52**: 173 - 177.

Haupt S, Oparka KJ, Sauer N, Neumann S. 2001. Macromolecular trafficking between *Nicotiana tabacum* and the holoparasite *Cuscuta reflexa*. *J. Exp. Bot.* **52**(354): 173-177.

Haywood V, Yu T, Huang N, Lucas W. 2005. Phloem long-distance trafficking of GIBBERELLIC ACID-INSENSITIVE RNA regulates leaf development. *Plant J* **42**: 49 - 68.

He F, Peltz SW, Donahue JL, Rosbash M, Jacobson A. 1993. Stabilization and ribosome association of unspliced pre-mRNAs in a yeast upf1- mutant. *Proc Natl Acad Sci U S A* **90**(15): 7034-7038.

He M, Wang Y, Hua WP, Zhang Y, Wang ZZ. 2012. De Novo Sequencing of Hypericum perforatum Transcriptome to Identify Potential Genes Involved in the Biosynthesis of Active Metabolites. *Plos One* **7**(7).

Honaas L, Wafula E, Yang Z, Der J, Wickett N, Altman N, Taylor C, Yoder J, Timko M, Westwood J, 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 Biology* **13**(1): 9.

Hosford RMJ. 1967. Transmission of plant viruses by dodder. *Botan. Rev.* **42**: 387-406.

Huang LL, Yang X, Sun P, Tong W, Hu SQ. 2012. The First Illumina-Based De Novo Transcriptome Sequencing and Analysis of Safflower Flowers. *Plos One* **7**(6).

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, et al. 2009. The genome of the cucumber, *Cucumis sativus* L. *Nature Genetics* **41**(12): 1275-1281.

Ingolia NT, Brar GA, Rouskin S, McGeachy AM, Weissman JS. 2012. The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. *Nat Protoc* **7**(8): 1534-1550.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science* **324**(5924): 218-223.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**(4): 789-802.

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.

Jackson S, Hong Y. 2012. Systemic movement of FT mRNA and a possible role in floral induction. *Frontiers in Plant Science* **3**.

Jiang L, Wijeratne AJ, Wijeratne S, Fraga M, Meulia T, Doohan D, Li Z, Qu F. 2013. Profiling mRNAs of two Cuscuta species reveals possible candidate transcripts shared by parasitic plants. *PLoS One* **8**(11): e81389.

Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**(9): 1236-1240.

Juntawong P, Bailey-Serres J. 2012. Dynamic Light Regulation of Translation Status in Arabidopsis thaliana. *Front Plant Sci* **3**: 66.

Juntawong P, Girke T, Bazin J, Bailey-Serres J. 2014. Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc Natl Acad Sci U S A* **111**(1): E203-212.

Juntawong P, Sorenson R, Bailey-Serres J. 2013. Cold shock protein 1 chaperones mRNAs during translation in Arabidopsis thaliana. *Plant J* **74**(6): 1016-1028.

Kawaguchi R, Bailey-Serres J. 2002. Regulation of translational initiation in plants. *Current Opinion in Plant Biology* **5**(5): 460-465.

Kawaguchi R, Bailey-Serres J. 2005. mRNA sequence features that contribute to translational regulation in Arabidopsis. *Nucleic Acids Research* **33**(3): 955-965.

Kawaguchi R, Girke T, Bray EA, Bailey-Serres J. 2004. Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana. *Plant J* **38**(5): 823-839.

Khandal D, Samol I, Buhr F, Pollmann S, Schmidt H, Clemens S, Reinbothe S, Reinbothe C. 2009. Singlet oxygen-dependent translational control in the tigrina-d.12 mutant of barley. *Proceedings of the National Academy of Sciences of the United States of America* **106**(31): 13112-13117.

Kim G, LeBlanc M, Wafula E, dePamphilis C, JH W. 2014. Genomic-Scale Exchange of mRNA between a Parasitic Plant and its Hosts *Science* **345**(August).

Kim M, Canio W, Kessler S, Sinha N. 2001. Developmental changes due to long-distance movement of a homeobox fusion transcript in tomato. *Science* **293**: 287-289.

Kragler F. 2010. RNA in the phloem: A crisis or a return on investment? *Plant Science* **178**(2): 99-104.

Kuijt J. 1977. Haustoria of phanerogamic parasites. *Ann. Rev. Phytopathol.* **17**: 91-118.

Kuijt J, Toth R. 1976. Ultrastructure of angiosperm haustoria: A review. *Ann Bot* **40**(170): 1121-1130.

Kumar S, Blaxter ML. 2010. Comparing de novo assemblers for 454 transcriptome data. *Bmc Genomics* **11**.

LeBlanc M, Kim G, Patel B, Stromberg V, Westwood J. 2013. Quantification of tomato and Arabidopsis mobile RNAs trafficking into the parasitic plant Cuscuta pentagona. *New Phytologist* **200**(4): 1225-1233.

LeBlanc ML, Kim G, Westwood JH. 2012. RNA trafficking in parasitic plant systems. *Frontiers in Plant Science* **3**: 203.

Lee KB. 2007. Structure and development of the upper haustorium in the parasitic flowering plant Cuscuta japonica (Convolvulaceae). *Am. J. Bot.* **94**(5): 737-745.

Lee KB. 2007. Structure and development of the upper haustorium in the parasitic flowering plant Cuscuta japonica (Convolvulaceae). *American Journal of Botany* **94**(5): 737-745.

Li C, Gu M, Shi N, Zhang H, Yang X, Osman T, Liu Y, Wang H, Vatish M, Jackson S, et al. 2011. Mobile FT mRNA contributes to the systemic florigen signalling in floral induction. *Sci Rep* **1**: 73.

Liang C, Liu X, Yiu SM, Lim BL. 2013. De novo assembly and characterization of Camelina sativa transcriptome by paired-end sequencing. *Bmc Genomics* **14**.

Liu MJ, Wu SH, Chen HM, Wu SH. 2012. Widespread translational control contributes to the regulation of Arabidopsis photomorphogenesis. *Mol Syst Biol* **8**: 566.

Liu SQ, Li WS, Wu YM, Chen CM, Lei JJ. 2013. De Novo Transcriptome Assembly in Chili Pepper (Capsicum frutescens) to Identify Genes Involved in the Biosynthesis of Capsaicinoids. *Plos One* **8**(1).

Lu KJ, Huang NC, Liu YS, Lu CA, Yu TS. 2012. Long-distance movement of Arabidopsis FLOWERING LOCUS T RNA participates in systemic floral regulation. *RNA Biol* **9**(5): 653-662.

Marcone C, Ragozzino A, Seemuller E. 1997. Dodder transmission of alder yellows phytoplasma to the experimental host Catharanthus roseus (periwinkle). *European Journal of Forest Pathology* **27**(6): 347-350.

Martin JA, Wang Z. 2011. Next-generation transcriptome assembly. *Nat Rev Genet* **12**(10): 671-682.

Matsuura H, Ishibashi Y, Shinmyo A, Kanaya S, Kato K. 2010. Genome-wide analyses of early translational responses to elevated temperature and high salinity in Arabidopsis thaliana. *Plant Cell Physiol* **51**(3): 448-462.

Matsuura H, Takenami S, Kubo Y, Ueda K, Ueda A, Yamaguchi M, Hirata K, Demura T, Kanaya S, Kato K. 2013. A computational and experimental approach reveals that the 5'-proximal region of the 5'-UTR has a Cis-regulatory signature responsible for heat stress-regulated mRNA translation in Arabidopsis. *Plant Cell Physiol* **54**(4): 474-483.

McKain MR, Wickett N, Zhang Y, Ayyampalayam S, McCombie WR, Chase MW, Pires JC, dePamphilis CW, Leebens-Mack J. 2012. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). *American Journal of Botany* **99**(2): 397-406.

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

Moeller JR, Moscou MJ, Bancroft T, Skadsen RW, Wise RP, Whitham SA. 2012. Differential accumulation of host mRNAs on polyribosomes during obligate pathogen-plant interactions. *Mol Biosyst* **8**(8): 2153-2165.

Mower J, Stefanovic S, Hao W, Gummow J, Jain K, Ahmed D, Palmer J. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. *BMC Biol.* **8**(1): 150.

Mower J, Stefanovic S, Young G, Palmer J. 2004. Plant Genetics: gene transfer from parasitic to host plants. *Nature* **432**: 165-166.

Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PG. 2012. Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. *PLoS One* **7**(2): e31410.

Mustroph A, Juntawong P, Bailey-Serres J. 2009. Isolation of Plant Polysomal mRNA by Differential Centrifugation and Ribosome Immunopurification Methods. *Plant Systems Biology* **553**: 109-126.

Nakasugi K, Crowhurst R, Bally J, Waterhouse P. 2014. Combining Transcriptome Assemblies from Multiple De Novo Assemblers in the Allo-Tetraploid Plant Nicotiana benthamiana. *Plos One* **9**(3).

Nicolai M, Roncato MA, Canoy AS, Rouquie D, Sarda X, Freyssinet G, Robaglia C. 2006. Large-scale analysis of mRNA translation states during sucrose starvation in arabidopsis cells identifies cell proliferation and chromatin structure as targets of translational control. *Plant Physiol* **141**(2): 663-673.

Park SH, Chung PJ, Juntawong P, Bailey-Serres J, Kim YS, Jung H, Bang SW, Kim YK, Do Choi Y, Kim JK. 2012. Posttranscriptional control of photosynthetic mRNA decay

under stress conditions requires 3' and 5' untranslated regions and correlates with differential polysome association in rice. *Plant Physiol* **159**(3): 1111-1124.

Parker C, Riches CR. 1993. *Parasitic Weeds of the World: Biology and Control*. Wallingford: CAB International.

Press MC, Graves JD, eds. 1995. *Parasitic Plants*. London: Chapman & Hall.

Rahmani F, Hummel M, Schuurmans J, Wiese-Klinkenberg A, Smeekens S, Hanson J. 2009. Sucrose control of translation mediated by an upstream open reading frame-encoded peptide. *Plant Physiol* **150**(3): 1356-1367.

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*.

Ribeiro DM, Araujo WL, Fernie AR, Schippers JHM, Mueller-Roeber B. 2012. Translatome and metabolome effects triggered by gibberellins during rosette growth in Arabidopsis. *Journal of Experimental Botany* **63**(7): 2769-2786.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, et al. 2010. De novo assembly and analysis of RNA-seq data. *Nat Methods* **7**(11): 909-912.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**(1): 139-140.

Roney JK, Khatibi PA, Westwood JH. 2007. Cross-species translocation of mRNA from host plants into the parasitic plant dodder. *Plant Physiol.* **143**(2): 1037-1043.

Rosado A, Li RX, van de Ven W, Hsu E, Raikhel NV. 2012. Arabidopsis ribosomal proteins control developmental programs through translational regulation of auxin response factors. *Proceedings of the National Academy of Sciences of the United States of America* **109**(48): 19537-19544.

Ross J. 1996. Control of messenger RNA stability in higher eukaryotes. *Trends Genet* **12**(5): 171-175.

Schepetilnikov M, Dimitrova M, Mancera-Martinez E, Geldreich A, Keller M, Ryabova LA. 2013. TOR and S6K1 promote translation reinitiation of uORF-containing mRNAs via phosphorylation of eIF3h. *EMBO J* **32**(8): 1087-1102.

Schliesky S, Gowik U, Weber APM, Brautigam A. 2012. RNA-seq assembly - are we there yet? *Frontiers in Plant Science* **3**.

Slaymaker DH, Hoppey CM. 2006. Reduced polysome levels and preferential recruitment of a defense gene transcript into polysomes in soybean cells treated with the syringolide elicitor. *Plant Science* **170**(1): 54-60.

Smith JD, Mescher MC, De Moraes CM. 2013. Implications of bioactive solute transfer from hosts to parasitic plants. *Current Opinion in Plant Biology* **16**(4): 464-472.

Sormani R, Delannoy E, Lageix S, Bitton F, Lanet E, Saez-Vasquez J, Deragon JM, Renou JP, Robaglia C. 2011. Sublethal cadmium intoxication in Arabidopsis thaliana impacts translation at multiple levels. *Plant Cell Physiol* **52**(2): 436-447.

Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, et al. 2012. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res* **40**(Database issue): D675-681.

Sun XD, Zhou SM, Meng FL, Liu SQ. 2012. De novo assembly and characterization of the garlic (Allium sativum) bud transcriptome by Illumina sequencing. *Plant Cell Reports* **31**(10): 1823-1828.

Surget-Groba Y, Montoya-Burgos JI. 2010. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res* **20**(10): 1432-1440.

Tcherkezian J, Brittis PA, Thomas F, Roux PP, Flanagan JG. 2010. Transmembrane Receptor DCC Associates with Protein Synthesis Machinery and Regulates Translation. *Cell* **141**(4): 632-644.

Tomilov AA, Tomilova NB, Wroblewski T, Michelmore R, Yoder JI. 2008. Trans-specific gene silencing between host and parasitic plants. *Plant J.* **56**(3): 389-397.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**(5): 511-U174.

Truscott FH. 1958. On the Regeneration of New Shoots from Isolated Dodder Haustoria. *American Journal of Botany* **45**(3): 169-177.

van Dorst HJM, Peters D. 1974. Some biological observations on pale fruit, a viroid-incited disease of cucumber. *Netherlands Journal of Plant Pathology* **80**(3): 85-96.

Vaughn KC. 2003. Dodder hyphae invade the host: a structural and immunocytochemical characterization. *Protoplasma* **220**: 189-200.

Vaughn KC. 2006. Conversion of the searching hyphae of dodder into xylic and phloic hyphae: A cytochemical and immunocytochemical investigation. *Intl J Plant Sci* **167**(6): 1099-1114.

Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**(1): 57-63.

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**(2): 295-306.

Westwood JH, Roney JK, Khatibi PA, Stromberg VK. 2009. RNA translocation between parasitic plants and their hosts. *Pest Man. Sci.* **65**(5): 533-539.

Westwood JH, Yoder JI, Timko MP, dePamphilis CW. 2010. The evolution of parasitism in plants. *Trends Plant Sci.* **15**(4): 227-235.

Wickett NJ, Honaas LA, Wafula EK, Das M, Huang K, Wu B, Landherr L, Timko MP, Yoder J, Westwood JH, et al. 2011. Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. *Curr. Biol.* **21**(24): 2098-2104.

Williams AJ, Werner-Fraczek J, Chang IF, Bailey-Serres J. 2003. Regulated phosphorylation of 40S ribosomal protein S6 in root tips of maize. *Plant Physiol* **132**(4): 2086-2097.

Xia ZH, Xu HM, Zhai JL, Li DJ, Luo HL, He CZ, Huang X. 2011. RNA-Seq analysis and de novo transcriptome assembly of Hevea brasiliensis. *Plant Molecular Biology* **77**(3): 299-308.

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**(12): 1660-1666.