The Transcriptional Regulation of Host Recognition and Prehaustorium Development in *Triphysaria versicolor*

By

DANIEL BEN STEELE

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Plant Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

John Yoder, Chair

_____

Stacey Harmer

_____

Daniel Runcie

Committee in Charge

2021

ACKNOWLEDGEMENTS

I could not have completed this dissertation without the many people who have helped along the way. First, I would like to thank my advisor, John Yoder. His constant support and encouragement made my time in graduate school an exceedingly positive and fulfilling experience. I valued his faith in me to work independently and pursue the research directions I was passionate about while providing guidance throughout the process. I am extremely grateful for his mentorship throughout my academic journey.

I would also like to thank Yaxin Wang who taught me how to be a grounded and effective scientist and provided countless help on my projects. Her positive, hard-working spirit was an inspiration and delight to be around. I am also tremendously grateful for all the support provided by other Yoder lab members including Pradeepa Gunathilake, Tristan Tran, Jingyanshan Li, Carly Tyer, and Maylin Murdock.

As part of my dissertation committee, Stacey Harmer and Daniel Runcie provided a significant amount of help with my research direction and the writing of this dissertation. I am sincerely grateful for all their support. The Designated Emphasis in Biotechnology program led by Denneal Jamison-McClung at UC Davis provided the unique opportunity of an internship at a Biotechnology company. This internship, together with other resources, made this program a very rewarding experience.

Our collaboration with the Parasitic Plant Genome Project (PPGP) was critical in facilitating my research through providing both genomic resources and support on specific projects. The lab of Jim Westwood provided substantial help with the *Phelipanche aegyptiaca* research in Chapter 4. Claude DePamphilis's lab gave crucial support in bioinformatics and

discussing project directions. I also want to thank Satoko Yoshida who was not part of the PPGP but graciously provided genomic and transcriptomic resources for multiple parasitic plants.

I am also sincerely grateful to all my sources of funding as my work could not have happened without their support. The UC Davis Plant Sciences Department GSR made a huge difference in covering part of my tuition and stipend. I want to thank the National Science Foundation and the US-Israel BARD Fund for both providing funding to our lab. The Jastro Shields Graduate Research awards I received were also a great help in covering research and travel related costs. Thank you to all that supported my research.

Finally, I would like to thank all my friends and family who supported me through this process. Most of all though, my wife Annie. Your encouragement throughout my academic journey and in all my pursuits enables any success I achieve. I could not ask for a better partner in life and thank you for everything you do.

ABSTRACT

Parasitic plants have evolved the ability to recognize neighboring plants and invade their tissues to acquire resources. We hypothesize that the mechanisms used by parasitic plant roots to recognize neighboring plant roots are a specialized adaptation of a general phenomenon that happens in most plants. In Chapter 2, I analyzed the transcriptional regulation of host recognition and prehaustorium development in *Triphysaria versicolor* using RNASeq. Promoter elements enriched in host-responsive genes were identified and evaluated using a fluorescent reporting construct. Seven elements showed clear tissue specific regulation in *Triphysaria*. The spatial regulation of five elements was strongly conserved in the two non-parasitic plants *Arabidopsis* and *Mimulus*. One element enhanced transcription of our fluorescent reporter in response to the host derived compound DMBQ. In Chapter 3, I show the TvQR1 promoter contains DMBQ responsive and tissue specific elements with conserved activities in two non-parasitic plants. The lack of up-regulation in the endogenous QR1 genes of these non-parasites suggests *Triphysaria* has co-opted the expression of QR1 for haustorium formation using conserved cis-elements. Lastly, as described in Chapter 4, I tested a Host Induced Gene Silencing approach to control the parasitic weed *Phelipanche aegyptiaca.* This method yielded a modest reduction in transcripts for four lipid biosynthesis genes but did not sufficiently suppress parasite growth for agricultural application. Together, this work contributes to the understanding of plant-plant interactions through studying the transcriptional regulation of host recognition in a parasitic plant and its conserved aspects in non-parasitic plants. It also yields possible applications for engineering patterns of gene expression in plants and gives insights into controlling parasitic weeds.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE **Page**

LIST OF TABLES

TABLE **Page**

CHAPTER 1:

INTRODUCTION

Parasitic plants have evolved the ability to recognize nearby plants and directly acquire resources from them. The ability of plant roots to recognize and respond to their neighbors is not unique to parasitic plants and is likely mediated by root secreted chemicals for both parasites and non-parasites [1,2]. The co-option of genes for new functions through changes in regulation is thought to be the major source for establishing new traits during evolution [3]. Thus, the initiation of haustorium development may have been co-opted from an earlier root-root recognition mechanism from its autotrophic progenitor.

The evolution of a parasitic lifestyle has occurred at least 12 times independently and represents roughly 1% of all flowering plants [4]. This includes parasites with varying degrees of host reliance. Facultative parasites are able to live and reproduce without a host but can connect with neighboring plants when presented the opportunity. Alternatively, obligate parasites require a host to complete their life cycle. Parasitic plants can be further ordered by their ability to photosynthesize. Hemiparasites retain the capability to photosynthesize while holoparasites have lost this function and acquire nearly all their fixed carbon from the host plant [5]. Among the families containing parasitic plants, only Orobanchaceae contains examples of all these classes of parasitic plants, including some non-parasitic plant species (e.g. *Lindenbergia)* [4].

The multifunctional organ used by parasitic plants to attach, invade, and connect with the host's vasculature is known as a haustorium [6]. A key evolutionary step in the transition to parasitism is the origin of haustoria. These likely occurred first as lateral haustoria, similar to those in facultative hemiparasites such as *Triphysaria* and *Phtheirospermum* [4]. These haustoria form laterally on the side of roots near the elongation zone and allow the root tip to continue

growing and potentially form additional haustoria [7,8]. The clear phenotypic response of haustoria in recognition of a host plant, makes parasitic plants an excellent model for studying plant-plant interactions.

Triphysaria versicolor is a non-weedy annual wildflower, common along the Pacific Coast of North America. It is used as a model parasite for Orobanchaceae as research can occur without restrictions and risk of spreading invasive plant material. Being a facultative hemiparasite that can live without a host plant makes *Triphysaria* easier to cultivate in a laboratory setting. The ability to precisely trigger haustorium development through the addition of chemical Host Inducing Factors (HIFs) greatly facilitates the study of plant-plant recognition and the developmental events of haustorium development in this model parasite. A highly efficient root transformation system has been established for *Triphysaria* which enables the introduction of new genetic material for evaluation [9]. The transcriptomic and genomic data created by the Parasitic Plant Genome Project (PPGP) are critical resources in studying the genetics of parasitic plants and our studies build upon their foundational work.

The haustoria of *Triphysaria* appear as hemispherical or spherical swellings with a localized proliferation of root hairs [10]. If a host root is not present, the haustorium development is halted and does not continue to develop structures for invading host tissues and acquiring resources such as intrusive cells or a xylem bridge. These partially developed haustoria that are induced without a host, will be referred to as “prehaustoria” in this work to fit with the currently suggested naming convention [11]

In the root parasitic plants of Orobanchaceae, haustoria are triggered by the perception of chemical HIFs from host root exudates[6]. Similar mechanisms are proposed to regulate the interactions of non-parasitic plant roots as well [12]. Phenolic and quinone compounds have been

identified as active HIFs through various in vitro applications [13–16]. One of the few HIFs identified from host plants is 2,6-dimethoxybenzoquinone (DMBQ) which was isolated from sorghum root extracts[17]. There are likely more HIFs present in host root exudates that have yet to be identified [18]. In the current model of haustorium development in Orobanchaceae, HIFs are redox cycled between quinone and phenolic states which generates highly reactive semiquinone intermediates that activate a redox sensitive signaling pathway [19–21]. TvQR1 is a gene from *Triphysaria* that encodes a quinone oxidoreductase that generates a semiquinone and is necessary for efficient haustorium development in *Triphysaria*.

Chapter 3 of this work examines patterns of spatial and temporal transcription from natural variants of the TvQR1 promoter in response to host root exposure. Using fluorescent reporter constructs, I showed that the TvQR1 promoter was upregulated in transgenic roots of *Triphysaria* as well as transgenic roots of *Arabidopsis* and *Mimulus*. However the endogenous *Arabidopsis* and *Mimulus* QR1 genes were not upregulated in response to DMBQ while the endogenous *Triphysaria* gene was. This shows the differential regulation of QR1 between *Triphysaria* and the non parasites *Arabidopsis* and *Mimulus* results from cis-element differences in the relevant promoters.

When a neighboring plant is recognized and haustorium development is initiated, a suite of developmental genes and pathways are activated which transition the parasite root to form this novel structure. Many of the genes in Orobanchaceae that are differentially expressed during haustorium development are present in other non-parasitic plants [22]. This includes the genes TvQR1, TvQR2, TvPirin, and PjYUCC3 which were shown to be necessary for haustorium development using RNA interference (RNAi) in either *Triphysaria versicolor* or *Phtheirospermum japonicum* [21,23–25]. It is largely unknown how these genes are regulated in

non-parasitic plants in response to host exudates. These genes may have been co-opted for haustorium development by parasitic plants by changing their regulation in the course of evolution.

The co-option of genes and pathways for new functions is thought to be the major source for establishing new traits during evolution [3]. The abundance, location, and timing of gene expression during development is often controlled by cis-regulatory elements which are sequences of DNA typically near a gene that serve as binding sites for regulatory proteins. Changing these cis-regulatory elements can co-opt the function of a gene for a new purpose by adjusting where and when it is expressed. This change in cis-elements has the potential to not interrupt the role already served by the gene by leaving the established elements in place. In contrast, changes to a coding sequence may have much higher pleiotropic effects as this affects any cell expressing that gene and may alter its function [26].

Chapter 2 of this work describes my search for cis-regulatory elements located in the promoters of genes responsive to host root factors. Several sequence motifs were identified that led to the expression of a fluorescent marker in specific root tissues, notably the root tip and epidermal cells. One motif conferred responsiveness in the transgenic plant to the host derived compound DMBQ.

Parasitic plants can be major agricultural pests, with those in Orobanchaceae being among the most damaging. Species of *Striga* alone are estimated to cause one billion US dollars in crop losses annually [27]. Egyptian Broomrape (*Phelipanche aegyptiaca*) causes severe damage to a wide range of economically important crops in the families Solanaceae, Fabaceae, Brassicaceae, Cucurbitaceae, Apiaceae and Asteraceae [28]. *P. aegyptiaca* is particularly limiting in processing tomato production [29]. These weeds can produce copious amounts of

tiny, long-lived seeds that lie dormant in the soil until the exudate of a host root triggers germination. Currently there are few methods to control these parasites with those available often being costly and resource intensive [30].

In Chapter 4, I describe my experiments using Host Induced Gene Silencing to reduce the negative impact of the parasitic plant *Phelipanche aegyptiaca* on tomato. Tomatoes were transformed with an RNAi construct that targeted four *Phelipanche* genes simultaneously via RNAi, transformed tomato, and then assayed Phelipanche growth on transgenic tomatoes. While transcript levels in attached Phelipanche plants were reduced, their growth was not sufficiently suppressed for commercial application.

Studying the transcriptional regulation of host recognition in a parasitic plant will give insights into the genes and functions involved during a plant-plant interaction and the development of a novel organ. The mechanisms regulating these transcriptional changes can suggest the evolutionary changes that lead to parasitism. The cis-elements controlling this gene regulation, could potentially be applied in other non-parasitic plant systems for engineering new expression patterns. The knowledge underlying the basic biology of haustorium development, could also yield new mechanisms for controlling parasitic weeds. Applying this knowledge by altering the gene expression of parasitic plants through host derived RNA interference has shown promise as a means of engineering resistance [31–35].

References:

1. Depuydt S. Arguments for and against self and non-self root recognition in plants. Front Plant Sci. 2014;5: 614.

2. Wang N, Kong C, Wang P, Meiners SJ. Root exudate signals in plant–plant interactions. Plant, Cell & Environment. 2021. pp. 1044–1058. doi:10.1111/pce.13892

3. True JR, Carroll SB. Gene co-option in physiological and morphological evolution. Annu Rev Cell Dev Biol. 2002;18: 53–80.

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

5. Irving LJ, Cameron DD. You are what You Eat: Interactions Between Root Parasitic Plants and Their Hosts. 2009.

6. Yoshida S, Cui S, Ichihashi Y, Shirasu K. The Haustorium, a Specialized Invasive Organ in Parasitic Plants. Annu Rev Plant Biol. 2016;67: 643–667.

7. Goldwasser Y, Westwood JH, Yoder JI. The Use of Arabidopsis to Study Interactions between Parasitic Angiosperms and Their Plant Hosts. Arabidopsis Book. 2002;1: e0035.

8. Cui S, Wakatake T, Hashimoto K, Saucet SB, Toyooka K, Yoshida S, et al. Haustorial Hairs Are Specialized Root Hairs That Support Parasitism in the Facultative Parasitic Plant Phtheirospermum japonicum. Plant Physiol. 2016;170: 1492–1503.

9. Bandaranayake PCG, Yoder JI. Factors affecting the efficiency of Rhizobium rhizogenes root transformation of the root parasitic plant Triphysaria versicolor and its host Arabidopsis thaliana. Plant Methods. 2018. doi:10.1186/s13007-018-0327-2

10. Jamison DS, Yoder JI. Heritable variation in quinone-induced haustorium development in the parasitic plant Triphysaria. Plant Physiol. 2001;125: 1870–1879.

11. Furuta KM, Xiang L, Cui S, Yoshida S. Molecular dissection of haustorium development in Orobanchaceae parasitic plants. Plant Physiol. 2021. doi:10.1093/plphys/kiab153

12. Wang N-Q, Kong C-H, Wang P, Meiners SJ. Root exudate signals in plant-plant interactions. Plant Cell Environ. 2021;44: 1044–1058.

13. Lynn DG, Steffens JC, Kamut VS, Graden DW, Shabanowitz J, Riopel JL. Isolation and characterization of the first host recognition substance for parasitic angiosperms. Journal of the American Chemical Society. 1981. pp. 1868–1870. doi:10.1021/ja00397a062

14. Steffens JC, Lynn DG, Kamat VS, Riopel JL. Molecular Specificity of Haustorial Induction in Agalinis purpurea (L.) Raf. (Scrophulariaceae)*. Annals of Botany. 1982. pp. 1–7. doi:10.1093/oxfordjournals.aob.a086332

15. Albrecht H, Yoder JI, Phillips DA. Flavonoids promote haustoria formation in the root parasite triphysaria versicolor. Plant Physiol. 1999;119: 585–592.

16. Cui S, Wada S, Tobimatsu Y, Takeda Y, Saucet SB, Takano T, et al. Host lignin composition affects haustorium induction in the parasitic plants Phtheirospermum japonicum and Striga hermonthica. New Phytol. 2018;218: 710–723.

17. Chang M, Lynn DG. The haustorium and the chemistry of host recognition in parasitic angiosperms. J Chem Ecol. 1986;12: 561–579.

18. Wang Y, Murdock M, Lai SWT, Steele DB, Yoder JI. Kin Recognition in the Parasitic Plant Triphysaria versicolor Is Mediated Through Root Exudates. Frontiers in Plant Science. 2020. doi:10.3389/fpls.2020.560682

19. Smith CE, Ruttledge T, Zeng Z, O'Malley RC, Lynn DG. A mechanism for inducing plant development: the genesis of a specific inhibitor. Proc Natl Acad Sci U S A. 1996;93: 6986–6991.

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

21. Bandaranayake PCG, Filappova T, Tomilov A, Tomilova NB, Jamison-McClung D, Ngo Q, et al. A single-electron reducing quinone oxidoreductase is necessary to induce haustorium development in the root parasitic plant Triphysaria. Plant Cell. 2010;22: 1404–1419.

22. Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, et al. Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol. 2015;32: 767–790.

23. Bandaranayake PCG, Tomilov A, Tomilova NB, Ngo QA, Wickett N, dePamphilis CW, et al. The TvPirin gene is necessary for haustorium development in the parasitic plant Triphysaria versicolor. Plant Physiol. 2012;158: 1046–1053.

24. Ishida JK, Wakatake T, Yoshida S, Takebayashi Y, Kasahara H, Wafula E, et al. Local Auxin Biosynthesis Mediated by a YUCCA Flavin Monooxygenase Regulates Haustorium Development in the Parasitic Plant Phtheirospermum japonicum. Plant Cell. 2016;28: 1795–1814.

25. Ishida JK, Yoshida S, Shirasu K. Quinone oxidoreductase 2 is involved in haustorium development of the parasitic plant Phtheirospermum japonicum. Plant Signal Behav. 2017;12: e1319029.

26. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2011;13: 59–69.

27. Spallek T, Mutuku M, Shirasu K. The genus Striga: a witch profile. Mol Plant Pathol. 2013;14: 861–869.

28. Parker C. The Parasitic Weeds of the Orobanchaceae. Parasitic Orobanchaceae. 2013. pp. 313–344. doi:10.1007/978-3-642-38146-1_18

29. Eizenberg H, Goldwasser Y. Control of Egyptian Broomrape in Processing Tomato: A Summary of 20 Years of Research and Successful Implementation. Plant Dis. 2018;102: 1477–1488.

30. Eizenberg H, Plakhine D, Ziadne H, Tsechansky L, Graber ER. Non-chemical Control of Root Parasitic Weeds with Biochar. Front Plant Sci. 2017;8: 939.

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

32. Bandaranayake PCG, Yoder JI. Trans-Specific Gene Silencing of Acetyl-CoA Carboxylase in a Root-Parasitic Plant. Molecular Plant-Microbe Interactions®. 2013. pp. 575–584. doi:10.1094/mpmi-12-12-0297-r

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

34. Aly R, Dubey NK, Yahyaa M, Abu-Nassar J, Ibdah M. Gene silencing of CCD7 and CCD8 in Phelipanche aegyptiaca by tobacco rattle virus system retarded the parasite development on the host. Plant Signal Behav. 2014;9: e29376.

35. Dubey NK, Eizenberg H, Leibman D, Wolf D, Edelstein M, Abu-Nassar J, et al. Enhanced Host-Parasite Resistance Based on Down-Regulation of Phelipanche aegyptiaca Target Genes Is Likely by Mobile Small RNA. Frontiers in Plant Science. 2017. doi:10.3389/fpls.2017.01574

CHAPTER 2:

THE TRANSCRIPTIONAL REGULATION OF HOST-RESPONSIVE GENES IN *TRIPHYSARIA VERSICOLOR*

ABSTRACT:

The roots of Orobanchaceae parasitic plants recognize specific molecules present in the root exudates of neighboring plants to identify potential host roots. When parasite roots are exposed to specific host root factors, they initiate the development of haustoria, which are used by the parasite to attach, invade and acquire resources from host tissues. Root-root recognition systems are not specific to parasitic plants and have been well documented in non-parasitic interactions as well. I hypothesize that genes responsible for host recognition and haustorium initiation in parasitic plants may have evolved from root-root recognition systems functioning in their autotrophic precursors. The genes and processes involved in host recognition and prehaustorium development in *Triphysaria versicolor* were analyzed using RNASeq. The genome of *Triphysaria versicolor* was sequenced and annotated to obtain promoter sequences. The promoters of genes co-expressed in the data set were evaluated for an enrichment of cis-elements to identify potential regulators of host responsive genes. Promoter sequences for these genes were obtained through the sequencing and annotation of the *Triphysaria versicolor* genome. Candidate cis-elements were evaluated for regulatory activity using a fluorescent reporting construct. Seven motifs were identified that resulted in tissue specific patterns of expression in *Triphysaria*. One motif enhanced transcription of a reporter gene in transgenic roots exposed to the haustoria inducing factor DMBQ. I then transformed these motifs into *Arabidopsis* and *Mimulus* and observed similar patterns of transcriptional regulation. These

regulatory motifs, which are conserved across plant families, have possibly been recruited by parasitic lineages to regulate the expression of parasite genes in response to nearby host roots.

INTRODUCTION:

The ability of plants to recognize and respond to neighboring plants has been well documented [1]. Subterranean plant-plant interactions are mediated by species or genotype dependent root-root recognition systems. These root-root interactions can affect phenotypes including changes in morphology, flowering timing, and exudate profiles [2,3]. This type of recognition may be an ancient phenomenon as it occurs in phylogenetically diverse taxa including angiosperms, cycads, and likely conifers [4]. Studying this activity is challenging due to the difficulty of phenotyping often subtle below-ground changes highlighting the need for a robust model.

Evidence suggests that root secreted chemicals facilitate root-root recognition [2,3]. The root parasitic plants of Orobanchaceae present a robust response to detecting host root exudates through the formation of haustoria. These host inducing factors are typically phenolic molecules such as 2,6-dimethoxybenzoquinone (DMBQ) [5]. This same chemical has been shown to facilitate neighboring root recognition in the non-parasitic parasitic plant *Arabidopsis* [6]. The morphologically obvious development of haustoria in response to host root secreted chemicals serves as an excellent model for studying mechanisms of plant-plant interactions.

Identifying the gene regulation involved in a developmental or physiological process can contribute to understanding its mechanism. Cis-elements play a major role in regulating genes. They are composed of specific DNA sequences, typically non-coding, that contain binding sites for transcription factors or other regulatory molecules that enable, enhance, or suppress the spatio-temporal transcription of genes [7]. Cis-elements typically reside in the 5’ upstream, intronic, or 3’ downstream regions of genes with the highest proportion in the 5’ region [8–11]. The specific genomic location of a cis-element may be important for its regulatory activity but,

for the purposes of this work, we will define a cis-element just by its sequence. Experimentally identified cis-elements in plants have a median observed length of 8 base pairs and are infrequently > 30 base pairs[12]. Changes in cis-regulatory elements are hypothesized to contribute more to phenotypic divergence than coding sequence changes [7]. This is due to the higher pleiotropic effects associated with protein changes that affect all cells expressing the gene, while a change in the cis-regulatory region may only affect a tissue or specific condition under which the gene is expressed. There are likely many factors that contribute to the evolution of new phenotypes with cis-element changes being a frequent contributor.

While parasitic plants have evolved from autotrophic ancestors, the mechanism is still unknown [13]. Haustorium development in Orobanchaceae may have evolved through altering the expression of autotrophic plant genes and pathways by changing their cis-element regulation. The initiation of haustoria may have also been co-opted from an earlier root-root recognition mechanism as this phenomenon is widespread in autotrophic plants. My hypothesis is that cis-elements used to regulate genes mediating root-root interactions in autotrophic plants have been recruited by the Orobanchaceae for parasitic functions.

METHODS:

Plant material:

Seeds for *Triphysaria versicolor* were collected from an open pollinated population growing in pasture land located in Napa, CA (GPS: 38.226214, -122.270785). The seeds were sterilized by immersing in 70% ethanol for 5 minutes, removing the ethanol, adding 50% commercial bleach (6.15% sodium hypochlorite) with 1% Triton x100 and gently shaking for 30 minutes. The seeds were then washed 8 times with distilled water and plated on 0.25X Hoagland media [14] with 0.75% (w/v) Phytagel(Sigma-Aldrich: P8169) [15]. The plates were then vernalized at 4°C for at least three days. To initiate germination, the seed plates were placed at 16°C with fluorescent grow lights set to 12 hours light and 12 hours dark. When *Triphysaria* seedlings were 10-14 days old, they were used for either transformation or genotyping analyses.

Seeds for *Mimulus guttatus* were collected from an open pollinated population growing near a stream at the North Table Mountain Ecological Preserve (GPS: 39°36'39.2"N 121°33'33.8"W). *Arabidopsis thaliana* seeds were a wild-type accession of Col-0. Seeds from both non-parasites were treated with the same methods as *Triphysaria* except the ethanol incubation was reduced to 1 minute and the bleach treatment to 5 minutes and 10 minutes, respectively for *Mimulus* and *Arabidopsis*.

Induction of host recognition and prehaustoria formation:

For the RNASeq, ten to fourteen day old *Triphysaria* seedlings were transferred to square plastic plates containing 0.25X Hoagland nutrient media [14], 0.75% (w/v) sucrose and 0.75% (w/v) Phytagel(Sigma-Aldrich: P8169). The square plates were placed for 3 days at 22°C in a vertical position with a fluorescent light set to 16 hour light and 8 hour dark. The seedlings were treated with 1 mL of *Arabidopsis* root exudate applied directly to the roots. *Arabidopsis* root

exudate was obtained using a previous method and standardized to 100 μM total phenolic content [16]. The two control time points (0.5 and 24 hours "C") were treated with H2O. The plates were left horizontal for 2 hours to allow the inducer to settle into the media and for gravity to elicit a tactile response in the seedlings. Roughly 0.5 cm of root tip were collected from 20-25 plants per time point and pooled in a single 2 mL tube for RNA isolation. Three replicates of each time point were collected. To minimize the influence of circadian rhythm on gene expression, we staggered the treatment of *Arabidopsis* root exudate to collect each time point at roughly the same time. For instance, the 24 hour time point was induced 23 hours before the 1 hour time point so the roots could be collected at the same time.

The induction and root collection of other Orobanchaceae species was performed the same as *Triphysaria*. The induction of transgenic roots was performed similarly but with 30 μM DMBQ instead of root exudate. These prehaustoria were evaluated 24 hours after DMBQ treatment.

RNA isolation and reverse transcription:

Two stainless steel balls were added to the pooled roots from each collection in round bottom 2 mL tubes. The tubes were placed in a liquid nitrogen chilled plastic block and ground for one minute using a paint shaker. This method facilitated the thorough grinding of all root material in a high throughput manner. RNA was isolated from the ground material using the Qiagen RNeasy plant mini kit (74904) following manufacturer's instructions with the optional RNase-Free DNase Set (79254). RNA was reverse transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems,Cat.# 4368814).

Quantitative PCR (qPCR):

Common target and reference gene primers were established for *Triphysaria* and the other Orobanchaceae studied in this work. Eight potential reference genes were identified as having a stable and high level of expression across our RNASeq samples. Primers for both targets and reference genes were chosen in conserved regions of the coding sequence using genomic sequences from *Phtheirospermum japonicum* and *Striga asiatica* and transcriptomic sequences from the PPGP. Multiple sets of primers for each gene were evaluated and those chosen all had a single dissociation peak and a linear amplification efficiency > 90% and <110% using a 10-fold dilution series of cDNA from multiple Orobanchaceae species. The program Normfinder was used to identify the two genes, PTB and PAB2, as having the most stable expression relative to all of the genes assessed [17]. The relative expression pattern obtained when using these reference genes matched the pattern obtained from the RNASeq quantification for three separate genes, QR1, BAG, and Pirin-like, Supplemental Figure S1. The agreement between these different quantification methods for the same sample of RNA, suggests both are accurate at estimating the relative abundance of transcripts across samples.

The cDNA (diluted 15-20 times) was combined with gene specific primers (300 nM) with the Power SYBR Green PCR Master Mix (Thermo Scientific Cat.# 4368577). The PCR was run in 96 well plates (MicroAmp Optical 96-Well Reaction Plate, Thermo Scientific™, 4306737) in an Applied Biosystems 7300 Real-Time PCR Machine. The program used was: 1 cycle 95°C for 10 minutes; 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds, followed by a dissociation step transitioning from 95°C to 60°C to 95°C. The Ct values for the transcripts of each gene were obtained from the mean of three technical replicates. The delta-delta Ct

method ($2^{-\Delta\Delta Ct}$) was used to calculate the levels of target gene transcripts relative to the geometric mean of two internal reference genes, PTB and PAB2.

RNA sequencing:

Strand-specific and barcode indexed RNA-seq libraries were generated from 500 ng total RNA each, after poly-A enrichment, using the Kapa mRNA-seq Hyper kit (Kapa Biosystems-Roche, Basel, Switzerland) following the instructions of the manufacturer. The fragment size distribution of the libraries was verified via micro-capillary gel electrophoresis on a LabChip GX system (PerkinElmer, Waltham, MA). The libraries were quantified by fluorometry on a Qubit instrument (LifeTechnologies, Carlsbad, CA) and pooled in equimolar ratios. The pool was quantified by qPCR with a Kapa Library Quant kit (Kapa Biosystems) and sequenced on an Illumina HiSeq 4000 (Illumina, San Diego, CA) with paired-end 150 bp reads. The library preparations and sequencing were carried out at the UC Davis Genome Center DNA Technologies and Expression Analysis Core, supported by NIH Shared Instrumentation Grant 1S10OD010786-01.

RNASeq data processing:

The adaptor sequences and low quality regions were trimmed from the raw fastq reads for each sample using the program scythe (v. 0.991) and sickle (v. 1.33) respectively with default parameters. The program Salmon (v. 0.14.1) was used to map and quantify the processed reads using the refined transcriptome described below. In Salmon, an index for the transcriptome was created using the "-quasi" and "-k 31" options. The reads were then mapped using this index and different library types. The library type that gave the highest mapping rate was chosen, in our case "-l IU". The resulting TPM quantification for the transcripts in each sample was used moving forward.

High molecular weight DNA isolation for *Triphysaria* genome:

The methods used to isolate high molecular weight DNA from *Triphysaria* can be found in Supplemental Method 1.

10X genomics linked-read genome sequencing:

High molecular weight genomic DNA was isolated from frozen above-ground material of *Triphysaria versicolor* using a modified CTAB method, described in detail in Supplemental Method 1. The genomic DNA was then size-selected to enrich for fragments greater than 40 kb on a BluePippin instrument (Sage Science, Beverly, MA). The DNA was then treated in a PreCR Repair Mix reaction (NEB, Ipswich, MA). The DNA was loaded onto a 10X Genomics Chromium Genome Chip at a concentration of 1 ng/ul and processed according to the instructions of the manufacturer (10X Genomics, Pleasanton, CA). The resulting random priming products were sheared to a peak size of 350 bp on an E220 Focused Ultrasonicator (Covaris, Woburn, MA) before sequencing library preparation 10X Genomics Chromium Genome Kit V2. The library fragment size distribution was verified via micro-capillary gel electrophoresis on a Bioanalyzer 2100 (Agilent, Santa Clara, CA). The library was sequenced with paired-end 150 bp reads on an Illumina Hiseq X sequencer (Illumina, San Diego, CA).

Genome assembly:

The genomic sequencing generated 713.5 million paired-end reads, which were input to supernova (v2.1.1) for assembly using default parameters [18]. The estimated genome size by supernova was 2.02Gb. The raw sequencing data represented 53X coverage of the genome. After removing the GEM barcodes, the effective coverage was 39X. About 39% of the nonduplicated reads provided the phasing information. The final haploid assembly contained ~1.9G bases in 235998 scaffolds, which included 8.14% uncharacterized ‘N’ bases. The GC content of the

assembly is 39.88%. The N50 of the assembly is ~23Kb. The completeness of the genome was assessed to be 87.4% (S:62.2%, D:25.2%) by BUSCO, using the embryophyte_odb9 dataset and Arabidopsis as Augustus training species [19].

Genome annotation:

The haploid assembly was used for annotation using MAKER (v3.01.02-beta) [20]. Transcriptome sequences for *Triphysaria versicolor* generated using previous RNASeq data were provided as EST evidence. The protein sequences from *Mimulus guttatus* were provided as protein homology evidence. The annotation process produced 62730 predicted genes.

Refining the transcriptome and genome:

The trimmed read files were then assembled into transcriptomes using the "De Novo Assembly" function with default parameters in CLC Genomics Workbench (v. 12.0.3). Different combinations of samples were assembled together to maximize the chance of assembling as many unique transcripts as possible. These newly made transcriptomes were combined with those created by the Parasitic Plant Genome project into a single file. A reciprocal best hit blast approach was used to identify the best pair of transcript and genomic coding sequence. The program Blast2GO (v.5.2.5) was used with blastn at an e-value $\leq$ 1.0E-25 to obtain the reciprocal best hit blast pairs which resulted in a refined set of 23378 genes. This greatly reduced the redundancy and fragmentation of our transcriptomes and is refined for genes expressed in our RNASeq data set. The refined set of genomic coding sequences represent the most likely sources of each transcript which allows us to best associate a promoter region with each transcript. To test if I retained most of the expressed genes from this strict refinement, the RNASeq reads were mapped to both the full and refined transcriptomes. A minimal loss in the mapping rate was observed when using the refined version of the transcriptome.

Differential expression analyses:

The TPM values from Salmon were loaded into Rstudio using the tximport package. The transcript levels between our RNASeq samples were compared using the package DESeq2 (v. 1.24.0). To compare the overall similarities and differences between the expression values of our samples, their variance was plotted using a PCA after a variance stabilizing transformation of the TPM values. This transformation also normalizes for library size. To identify differentially expressed transcripts I first filtered out those with low expression across our data set (those with a total TPM ≤ 24 across all 24 of our samples, 95.2% of the transcripts remained after this step). Transcripts were considered differentially expressed if there was at least a 2 fold change in expression (with an adjusted of p-value ≤ 0.01) when compared with the 0.5 hour water treated control time point.

The expression levels of our 1619 genes with differential expression were normalized by dividing the TPM of each time point by the total of all time points for each transcript. This normalization shows the relative change of each transcript across our data set. These relative expression profiles were grouped by similarity using a Self-Organizing Map (SOM) created from the kohonen package in R [21]. Using an SOM with 16 groups in a 4 by 4 hexagonal pattern yielded groups of transcripts with distinct expression changes without too much overlap between the groups. This level of group separation was judged by eye. All graphs were generated using ggplot2 in R [22].

Predicting functions:

Blast2GO was used to predict the functions and provide GO term annotations for each transcript. BLASTX was used with the NCBI nr database (updated on 12/17/2019) to identify 20 hits for each transcript with an e-value $\leq$ 1.0E-5. A word size of 6 was used with the remaining

parameters set to the default. The functional annotation pipeline in Blast2GO was used with default parameters to predict functions of each transcript through a summary of the best hits.

GO enrichment:

Blast2GO was used to identify the Gene Ontology (GO) terms over and under represented in groups of differentially expressed genes between our time points. Transcripts up and down regulated between our 7 time points when compared to the 0.5 hour control were evaluated for an enrichment of GO terms using a two-tailed Fisher's Exact test and the whole set of transcripts as the reference. In R, the over-represented GO terms from the enrichment test of each time point compared to the control were filtered for a False Discovery Rate adjusted p-value (FDR) ≤ 0.05. All enriched GO terms that fit this criteria were collected and the FDR from each time point obtained. GO terms were selected for Figure 2 to summarize the processes enriched during our developmental RNASeq. A heatmap of the log2 transformed FDR values was generated using ggplot2.

Motif identification:

Promoters from groups of differentially expressed genes were evaluated for an enrichment of known and de novo motifs. All promoters analyzed were 500 bps in length. Promoters were grouped based on the SOM or being differentially expressed at a certain time point. Known motifs were identified using the AME program in the MEME suite (v.5.0.5) [23]. The promoters from genes non-differentially expressed from our data set were used as control reference sequences. The average odds scoring and fisher exact test method were selected. The known motif databases for plants include the Plant Transcription Factor Database (v4.0) and the JASPAR 2020 Core Plants non-redundant set were utilized for enrichment [24]. From the

resulting enriched motifs, specific candidates were selected with low p-values and as representatives of repeatedly enriched motifs.

To identify de novo motifs, the same groups of promoter sequences mentioned above were used in the oligo-diff program from the web-based Regulatory Sequence Analysis Tool (RSAT)[25]. Oligomer lengths of 6, 7, and 8 were used for each enrichment. The resulting de novo motifs were combined into a single file and validated for enrichment using the AME program. This RSAT file served as the known set of motifs and was queried for enrichment against the differentially expressed programs from which they were originally identified. This analysis confirmed using a different algorithm that these motifs were enriched and ranked the level of enrichment for later selection. Motifs with low enrichment p-values and at least some sequence complexity were chosen for further analyses.

The presence of a larger conserved motif between *Triphysaria* promoters was checked for those identified through both approaches. The specific sequences and locations for each motif in the promoter sequences was determined using the FIMO program from the MEME suite. These sites were filtered based on their similarity to the original motif and the expression of the promoter's transcript. These refined sites were extended by 10 bps on either side and run through MEME to identify any enrichment of a larger motif sequence.

Orobanchaceae expression comparison:

Genes commonly up-regulated across multiple Orobanchaceae species were identified through qPCR and by comparing expression data sets. The qPCR was performed as described above across nine members of Orobanchaceae. Expression data were obtained for three parasites exposed to host factors. *Striga gesnerioides* and *Phelipanche aegyptiaca* RNASeq data was provided by the PPGP while a *Phtheirospermum japonicum* microarray data set was provided by

Satoko Yoshida. To identify genes commonly up-regulated across these parasites, the data were filtered for transcripts having at least a 2 fold increase in expression after host recognition. Next, an orthologous sequence for each parasite transcript was identified using both a receptacle best hit blast (RHBH) and best hit blast (BHB) approaches. Both parasite transcriptomes were queried against the other and the single best hit for each transcript was obtained using blastn, an e-value of 1.0E-10, and word size of 11 in Blast2GO. For the RHBH, only those transcripts that matched with one another from both data sets were considered orthologous. For the BHB, the single best hit from the parasite data set for *Triphysaria* was considered orthologous. Both of these approaches were used to identify genes commonly up-regulated with more weight given to the strict RBHB method.

Candidate promoter selection:

Specific candidate promoter sequences were identified for each selected motif. These genes were chosen based on the p-value of the motif site, the promoter's transcript exhibiting a robust expression change in our data set, the transcripts being common up-regulation across multiple Orobanchaceae, and having an interesting functional annotation.

Phylogenetic footprinting:

Phylogenetic footprinting was used to identify if candidate motifs existed within a larger conserved sequence that may be important for their function. Orthologous promoter sequences from *Phtheirospermum japonicum, Striga asiatica*, *Mimulus guttatus, Sesamum indicum,* and *Antirrhinum majus* were identified from the coding sequence of *Triphysaria* for each candidate gene. The promoters from these genes were then aligned in the program Geneious and compared for conserved regions. Extra weight was given to conserved promoter sites between the three

parasitic plants. The non-parasite sequences were often useful to refine the conserved sites to manageable sizes for later evaluation.

Predicting transcription factor binding sites:

The potential transcription factor binding sites in each TRE motif were predicted using the Plant Transcription Factor Database (PlantTFDB v.5.0) and filtered for a p-value $\leq$ 1.0E-5.

Fluorescent reporting construct:

The promoter reporting construct pDS-MRC was created using the pCAMBIA-0380 binary plant expression vector. The individual components of the T-DNA shown in Figure 6 were amplified using PCR with flanking sequences overlapping the adjacent fragment and column purified before assembly. The template for the Omega-NLS-tdTomato sequences came from the addgene vector #61628 [26]. The sequence for mClover3 came from the addgene vector #74252 [27]. The 35S promoter and terminator originated from the pHG8-YFP vector [28]. The completed pDS-MRC was completed through multiple rounds of Gibson assembly (NEB #E2611) that sequentially added the above-mentioned PCR amplified components into the multiple cloning site of pCAMBIA-0380.

The relative positions of regulatory components and number of motif copies were informed by the best practices outlined previously [29]. A 35S minimal promoter was used to assemble the basic machinery for transcription so a motif can trigger the process. The motifs in pDS-MRC were placed 60 bp upstream of the TATA box in the 35S minimal promoter to give sufficient space for regulatory proteins to interact. The 35S minimal promoter and 60 bp spacing sequence were obtained from a vector used to report the regulatory activity of an auxin motif [26]. The 10 bp spacer sequence between the tandem motif copies was randomly generated and selected for having few predicted transcription factor binding sites. The restriction enzyme sites

BspEI and AatII were placed on either side of the motif region for ligating in new sequences. These enzymes were chosen for not already being present in the vector, working in the same NEB 3.1 buffer, and creating incompatible sticky ends to minimize the empty recircularization of the vector. To clone a motif into the construct, a primer pair was ordered with the motif sequence and complementary sticky-ends for the vector. Either a single motif copy or tandem duplicate was ordered based on the motif size. The 10 bp spacer sequence was included between the tandem duplicated motif. The annealed primers with exposed sticky-ends were then ligated into the pDS-MRC vector with T4 Ligase (NEB #M0202).

The completed vectors were transformed into NEB® 5-alpha Competent E. coli (NEB# C2987) and selected for using kanamycin resistance and colony PCR. T-DNA regions from the resulting vectors were sequenced to verify their correct identity.

Plant transformation:

Triphysaria root transformations were performed following procedures established previously[30]. In short, young seedlings for each of the three species used were cut at or slightly below the root shoot junction and the cut site on the part of the plant containing the shoot was dipped in freshly grown *Rhizobium rhizogenes* strain MSU440 containing the vector of interest. These inoculated shoots were left on the horizontal surface of a plate containing 1X MS media with vitamins(PhtyoTechLabs: M519), 30 g/L sucrose, 2 g/L Phytagel(Sigma-Aldrich: P8169), and 400 uM Acetosyringone for one to two weeks depending on the growth of MSU440. The shoots were then transferred to a 0.25X Hoagland plate with 0.75% sucrose, 0.75% Phytagel, and 300 mg/L Timentin to stimulate root development and suppress the bacterial growth [30].

Fluorescent microscopy:

Dissecting: A Zeiss Stemi SV 11 dissecting scope with dsRED and GFP NB filters (equipped with an HBO 100 1007-980 light source powered by an ebq 100 isolated power supply) was used to observe the red and green fluorescence from tdTomato and mClover respectively exhibited by our transgenic roots. Transgenic roots were selected if they displayed a robust expression of mClover3. The presence of tdTomato before and after prehaustrium induction was used as a first round of screening before imaging with the confocal microscope described below.

Confocal: Roots were imaged using a Zeiss Axio inverted Observer Z1 microscope with an LSM 710 laser scanning confocal system and a Plan-Apo 10X objective. The red and green fluorescent proteins, tdTomato and mClover3, were excited using lasers at wavelengths 561 nm and 488 nm respectively. The pinhole was adjusted to 1 AU. The lasers were set to 100% to maximize the depth at which we could image inside the roots a Z stack. The digital gain was adjusted for each sample using the "Range Indicator" to avoid over-exposing the fluorescence of any image. The detection of red and green fluorescence was done with separating images to avoid the false-positive red fluorescence created by the 488 nm green excitation laser.

PI stain:

Triphysaria root tips were stained with 10 μg/ml PI in H2O to better visualize the cell files and help estimate the tissue layer identities. Roots were imaged using the same procedure described above for the Zeis confocal microscope with a 514 nm excitation wavelength

Image processing and analyses:

Images from the dissecting scope images were evaluated using the program Lightroom (from Adobe) with exposure and contrast adjusted to more easily see the red and green

fluorescence. Confocal images were processed using ImageJ [31]. The CZI files were loaded into ImageJ using the Bio-Formats plugin. To better visualize the fluorescence throughout the root, a Z Projection was created using the max intensity from the images in each Z stack. The brightness and contrast were subsequently adjusted for each image.

Statistical analysis:

A statistically significant difference between two groups of measurements was tested using a two-tailed Welch's t-test. A significant difference was obtained if the resulting p-value ≤ 0.05.

RESULTS:

The experiments in this chapter aimed to identify and functionally annotate cis-regulatory elements from the promoters of host responsive genes in parasitic plants. The approaches were to 1) identify genes upregulated in parasite roots within 24 hour of host exposure, 2) obtain promoter sequences for the host responsive genes, 3) bioinformatically identify candidate cis-regulatory elements in these promoters, and 4) functionally evaluate the regulatory activity of the cis-elements using a fluorescent reporter in the transgenic roots of parasitic and non-parasitic plants.

1. Identify parasite genes upregulated after exposure to host root exudates

Identification of upregulated transcripts by RNASeq

RNASeq was used to identify genes transcriptionally regulated soon after exposure to host exudates. This was created by treating *Triphysaria* roots with *Arabidopsis* root exudate and harvesting the root tips at eight time points after exposure, Table 1. The times selected focused on initiation and early stages of prehaustoria development. Control roots were treated with water and harvested after 0.5 and 24 hours. Representative images of *Triphysaria* root tips at each of these developmental time points is in Figure 1B.

To validate the transcript levels obtained from quantifying the RNASeq, quantitative PCR (qPCR) was performed on cDNA from the same samples of RNA used for the RNASeq. The relative transcript levels for three host-responsive genes (TvQR1, TvPirin, and TvBAG) were quantified using both the RNASeq data set and qPCR (Figure S1). Both methods estimated a similar relative transcript level across our samples.

A principal component analysis (PCA) analysis was also performed on the expression values across all of the RNASeq samples to evaluate the quality of the replicates and distinction between the time points, Figure S2. This analysis measures the variance between samples and plots the values for the two components containing the most variation. All replicates for a given time point clustered tightly with one another and were clearly distinguished from other samples.

The relatively large distance between early time points, indicates the greatest change in transcription expression occurs soon after host exudate exposure. The relative distance between timepoints continually diminishes between 3 and 24 hours, suggesting the change in transcriptional expression lessens over time. This waning of expression change continues to the point where the clusters for 12 hours and 24 hours are largely overlapping, meaning there are little differences in expression between these time points. Together, these results suggest that largest changes in transcription occur between 0.5 and 3 hours after host exudate exposure and these changes diminish to the point where the expression patterns for prehaustorium development are largely stable by 12 hours.

SOM Clustering

To reduce the dimensionality of our data set and to identify patterns in the expression values, I grouped transcripts based on the similarity of their expression profiles across the time points. Transcripts with at least a 2-fold change in expression were clustered in co-expressed groups using an unsupervised machine learning technique known as Self-Organizing Map (SOM)[21]. The average expression levels of the transcripts in the 16 groups identified by the SOM can be seen as a heat map in Figure 1A. The number of transcripts in each group and representative pictures of *Triphysaria* roots at each time point are included in Figure 1A,B. The relative expression profiles for all transcripts in each SOM group can be seen in Figure S3.

The SOM identified groups of transcripts with clear peaks of expression at specific time points. Roughly half of the differentially expressed transcripts were upregulated within 0.5-3 hours after Arabidopsis exudate treatment. This indicates a burst of transcriptional activity quickly after exposure to host root exudates. The remaining SOM groups included transcripts that continued to increase expression up to the 24 hour time point, or were down-regulated at some point in time. About twice as many transcripts show patterns of up-regulation in our data set compared to those that are down-regulated.

GO term enrichment analysis

To identify biological processes functioning during prehaustorium initiation and development, I determined which Gene Ontology (GO) terms were enriched in differentially expressed transcripts. The heat map in Figure 2 shows the level of statistical significance for each term in transcripts differentially expressed at each time point. The positive and negative values show whether the enrichment occurs in transcripts up or down-regulated respectively. The GO terms in Figure 2 were selected based on their level of enrichment and function. A total of 157 GO terms were enriched in up-regulated transcripts while 36 terms were enriched in transcripts down-regulated.

Identifying genes commonly upregulated across multiple Orobanchaceae.

To identify if the genes differentially expressed in *Triphysaria* are commonly up-regulated in other Orobanchaceae parasites in response to host recognition, I performed qPCR and compared data sets from other parasites.

For the qPCR, I collected seeds from hemiparasitic Orobanchaceae from diverse sites across California with representative pictures and details in Supplemental Figure S4A. Voucher specimens for these species were submitted to the UC Davis Herbarium with the details in Figure

S4B. Roots of the nine plants were treated with Arabidopsis exudate for 1 hour, Figure 3. The phylogenetic relationships shown to the left came from a larger study of the Orobanchaceae [32].

For this analysis, I created universal reference gene and target gene primers for all nine Orobanchaceae species. The primers were designed in conserved coding regions, validated using cDNA from each plant, and were observed to be efficient and specific during PCR. The heat map in Figure 3 shows the fold change in transcript level after this treatment for six host responsive genes. The genes were chosen based on having a known association with haustorium development in Orobanchaceae or as differentially expressed in our RNASeq.

Overall, the hemiparasites up-regulate these genes more than the closely related non-parasite *Lindenbergia*. Three genes, BAG, bHLH93 and ERF86, were significantly upregulated in multiple parasites. BAG, but not bHLH93 or ERF86, was also upregulated in *Lindenbergia*. Three genes; QR1, QR2, and Pirin-like were up-regulated in some parasites but not others. The regulation of QR1 was species specific and significantly upregulated in *T. versicolor* but not other *Triphysaria* genera.

In addition to the qPCR described above, I compared expression of our differentially expressed genes with data sets from several other Orobanchaceae parasites shown in Figure S5. Genes were determined to be up-regulated in the other data sets if their expression increased by at least 2-fold after exposure to host factors. The orthologs for each gene were determined through a reciprocal best hit BLAST approach. The genes with orthologs up-regulated in multiple parasite data sets can be seen in Figure S5.

2. Identification of promoter sequences

Genome sequence for *Triphysaria*

With the goal of identifying promoter sequences for the differentially expressed genes identified above, we used 10X Chromium libraries followed by Illumina sequencing to obtain genomic sequences from a single *T. versicolor* plant, Table 2. This work was done in collaboration with the Bioinformatics and DNA Technologies Cores at the UC Davis Genome Center. The resulting genome had an N50 of 23,077 bp, meaning half of the genome is contained on fragments equal to or larger than 23 kbp. A BUSCO (Benchmarking Universal Single-Copy Orthologs) score of 87.4% indicates that most single copy orthologs are represented in the dataset. The *Triphysaria* genome is of appropriate quality to identify promoter regions of most genes.

Using the gene sequences for the closely related plant *Mimulus guttatus* as a guide, we predicted 62730 genes to be present in our *Triphysaria* genome with the program Maker [20]. This number is likely an overestimation due to factors such as the heterozygosity between alleles causing them to be counted as separate genes, the fragmentation in assembly, and possible alternative splicing of transcripts.

To reduce this redundancy and noise, the number of genes and transcripts was reduced through a reciprocal best hit BLAST approach between the annotated CDS regions of the *Triphysaria* genome and a combined set of transcriptomes created from our RNASeq data and those developed by the PPGP. This yielded the single best transcript-gene match and enabled a link between the expression of a transcript and a specific promoter region. After this refinement, the effective gene/transcript number used in further analyses was 23378. When comparing the mapping rate of RNASeq reads to the full transcriptome and our refined version, there was a

minimal loss in mapping frequency which suggests the redundancy was reduced while capturing the majority of expressed transcripts in our data set.

For this study the promoter was defined as the region upstream of the start codon. This includes the 5' untranslated region (UTR) as it may contain regulatory elements of interested. All of the motif identifying analyses moving forward use the 500 bp of promoter sequence as this region immediately up from the start codon is thought to contain the highest frequency of active cis-regulatory elements[8–10].

3. Identify candidate cis-elements from the promoters of host responsive genes

Candidate cis-elements were identified from the promoters of host-responsive genes using three methods: an enrichment of known cis-element motifs, an enrichment of de novo motifs, and phylogenetic footprinting. Promoters were grouped using different degrees of co-expression before the enrichment approaches.

Enrichment of known motifs

Using transcription factors databases from the PlantTFBD and JASPAR, I searched for enriched known motifs in groups of promoter sequences with the AME program from the MEME suite. This yielded a large set of mostly redundant motifs with some clear trends in enrichment. I observed strong and repeated levels of over-represented CAMTA/FAR1, WRKY and HSF motifs. The CAMTA/FAR1 and WRKY motifs were enriched in promoters of genes up-regulated within an hour of exudate treatment. Specifically these were SOM 8 for CAMTA/FAR1 and SOMs 3, 7, 8, and 12 for WRKY. The HSF motifs were enriched in genes up-regulated throughout the time course with the highest enrichment being at 3 hours, SOMs 13 and 14.

To determine if there was an association between the enriched motifs and their predicted corresponding transcription factors, I analyzed the expression profiles of transcription factors across our data set. The expression profiles for differentially expressed transcription factors grouped by their annotated families are shown in Figure S6. Twenty eight total transcription factor families had distinct changes in expression during the initiation and development of prehaustoria with the most frequent change being rapid up-regulation after host exposure.

Of note, several WRKY and HSF transcription factors were up-regulated at time points matching with their enriched motifs. For instance, the HSF transcription factors peak in expression at 1 hour while the genes with enriched HSF motifs in their promoters peak at 3 hours. The association of enriched motifs with the expression of their predicted corresponding transcription factors, suggests they may play roles in prehaustorium initiation.

Enrichment of de novo motifs

The oligo-diff program from the Regulatory Sequence Analysis Tools group was used to identify de novo motifs in the promoters of host responsive genes. These motifs were validated as enriched using the AME program from the MEME suite. Motifs with high levels of enrichment and non-repetitive sequences were chosen as candidates.

Selection of representative genes with promoter motifs

The motifs identified represent a frequency of nucleotides at specific positions. To obtain a specific nucleotide sequence for the enriched motif, I identified representative genes with a robust expression change in our RNASeq data set as well as those from three other Orobanchaceae parasites mentioned above. I also considered the functional annotation in the selection of gene representatives.

Phylogenetic footprinting

Phylogenetic footprinting was used to identify conserved promoter sequences between parasites and non-parasites. Promoters from genes with known associations with haustoria formation in Orobanchaceae parasites, such as QR1 and QR2, and those with robust changes in expression across multiple Orobanchaceae were compared for conserved sequences. Phylogenetic footprinting was also used to determine if a sequence larger than the enriched motifs was conserved among parasites and/or non-parasites.

For this analysis, I used promoter sequences from three parasites (*T. versicolor, P. japonicum* and *S. asiatica*) and three closely related non-parasites (*M. guttatus, Sesamum indicum,* and *Antirrhinum majus*). Promoters from the orthologous genes in these species were obtained using BLASTn and the coding sequence from *Triphysaria*. Promoters were aligned using Geneious to identify regions of conservation. The promoter sequence in *Triphysaria* surrounding the location of each motif that was well conserved among the three parasites was chosen for evaluation in the fluorescent constructs discussed below. If the use of these promoters did not refine the element under 70 bps, then the non-parasitic promoter sequences were included.

Conclusions

I identified 34 sequences conserved in the promoters of host responsive parasite genes (Table S1). I next evaluated how the presence of these motifs affected the spatial and temporal expression of a reporter construct under the regulation of a minimal promoter.

4. Functional analysis

pDS-MRC construction

The pDS-MRC construct is a T-DNA based binary vector constructed using pCAMBIA-0380 to evaluate the regulatory activity of promoter motifs in transgenic roots, Figure 4, S7. In a set of pilot experiments, I evaluated several fluorescent marker proteins for their brightness in transformed *Triphysaria* roots and found the red and green fluorescent proteins tdTomato and mClover3 were among the brightest. The plasmid constitutively expresses mClover3 using a 35S promoter to select transformed roots and outline cells during imaging.

The regulatory activities of candidate motifs were reported using the red fluorescent protein tdTomato. A minimal 35S promoter was used to assemble the machinery for transcription and allow the candidate motif to initiate the process. Motifs were cloned as tandem (with a 10 bp spacer) or single copies at a location 60 bps upstream of the TATA box in the 35S minimal promoter. The relative spacing of these elements was determined from the best practice outline in a review on plant synthetic promoters [29].

Additionally, a nuclear localizing signal was used to target tdTomato to the nucleus. Sequestering the red fluorescence to the nucleus allowed more ready visualization of which cells were expressing the protein as each cell is represented by a single dot of red fluorescence. The Omega sequence from Tobacco Mosaic Virus was used to enhance translation of both tdTomato and mClover3.

As a negative control, the level of tdTomato fluorescence was assessed for *Triphysaria* roots transformed with an empty version of pDS-MRC that did not contain a cloned motif. Minimal red fluorescence was observed in these roots, Figure S8,9 .

Seven functional motifs

I cloned each of 34 candidate motifs into the reporter construct and then transformed these into *Triphysaria* roots using *Agrobacterium rhizogenes.* The level of tdTomato expression was assessed before and 24 hours after the induction of haustoria using DMBQ, Table S1. Seven motifs facilitated a robust expression of tdTomato across multiple independently transformed roots, Table 3. The remaining motifs did not show evidence of enhancing expression. These seven motifs were named “Triphysaria Regulatory Elements” or “TRE”.

Spatial expression in *Triphysaria* roots

The TRE motifs fell into two groups based on the localization of tdTomato fluorescence; roots containing TRE 1-3 expressed tdTomato predominantly near the root tip while TRE 4-7 resulted in tdTomato expression in more mature root parts, Figure 5. This pattern in localization was mirrored by the pattern of temporal expression for their source genes, Figure S10. The three genes containing the tip expressed motifs all peak in transcript levels early in our RNASeq. All four of the genes that contained mature root localized motifs (mostly epidermal and root hair), peak in transcript level at our last time point, 24 hours. The tissue types of *Triphysaria* roots were identified through propidium iodide staining which outlined the cell files, Figure S11.

The first two motifs, TRE1 and 2, both regulated expression almost exclusively in the columella, Figure 5A-F. TRE1 tends to give a higher expression level in the younger columella cells and potentially the meristematic region when compared to TRE2.

TRE3 led to tdTomato being expressed in the root cap, a layer of ground tissue, and some in a vascular associated tissue. In the promoter of a 10-hydroxygeraniol dehydrogenase (10HGO) gene in *Triphysaria*, this motif was copied at least four times. These copies were also conserved in *Phtheirospermum* and *Striga* promoters. Intriguingly, the coding sequence for this gene was

also copied in *Triphysaria* many times without intron sequences. These intron-less copies of 10HGO were also found in the parasites *Phtheirospermum* and *Striga* but not the non parasites *Arabidopsis*, *Mimulus*, *Antirrhinum*, or Tomato. The intron-less gene copies of 10HGO were compared between the parasites and suggested some of these copies have occurred independently in each parasite as the copies were more similar to those within a parasite and not between parasites.

The motifs TRE4-6 all predominantly facilitate expression in epidermal or hair cells. TRE4 is mostly expressed in both mature epidermis and root hairs with some lower expression in a vascular associated tissue and young columella cells, Figure 5J-L. The level of tdTomato in the vasculature and tips looked to increase when imaged 24 hours after DMBQ exposure.

The other two epidermal utilized motifs, TRE5 and 6 were both specific to hair cells, Figure 5M-R. These motifs enabled expression in both haustorial hairs and typical root hairs. The full promoters of the genes these motifs originated from were cloned into the construct pDS-PRC and show almost identical patterns of root hair specific expression, Supplemental Figure S12. The refined motifs for both TRE5 and 6 (just 18 bp), were able to recapitulate the tissue specific expression of their full promoter.

TRE7 was predominantly localized at the sites of lateral roots. This is consistent with TRE7 originating from the gene LRP1 (Lateral root primordia 1) which has been shown with a full promoter to be highly expressed in developing lateral roots [33,34]. Part of this tissue localization is now refined to an element only 36 bp in size.

Spatial expression in non-parasite roots

I next transformed the TRE reporting constructs into two non-parasitic plants *Mimulus guttatus* and *Arabidopsis thaliana*. *Mimulus* is fairly closely related to *Triphysaria* as both are in the same order Lamiales. *Arabidopsis* is less related and classified as a Rosid while *Triphysaria* and *Mimulus* are Asterids. These groups represent two distinct branches of the eudicot evolutionary tree which are estimated to have shared a common ancestor roughly 128 million years ago[35].

The expression patterns of tdTomato under the influence of TRE motifs in *Triphysaria* was similar in *Arabidopsis* and *Mimulus*, except for TRE3 and TRE7 which did not express robustly in the non-parasites, Figures 6,7. The motifs TRE1 and 2 directed expression to the root tip and the columella as they did in *Triphysaria*. In *Mimulus*, TRE1 was also expressed in the lateral root cap and in some cells lining young vascular tissues, Figures 6A-C. In *Arabidopsis*, TRE1 was more localized to younger columella cells, Figure 6 G-I. When transformed with TRE2 constructs, both *Mimulus* and *Arabidopsis* exhibited expression in the root cap overall, which includes both columella and the lateral root cap, Figure 6D-F, J-L.

The three *Triphysaria* motifs that directed expression to epidermal and hair tissues, TRE4-6, exhibited similar patterns in the non-parasites, Figure 7. For TRE4, both *Mimulus* and *Arabidopsis* had expression throughout the mature root but at a relatively higher level in the epidermis, including root hairs, and vasculature. To a lesser extent expression was observed at the outer root tip. The motifs TRE5 and 6 were root hair specific in both non-parasites which is consistent with the pattern observed in *Triphysaria*.

DMBQ responsiveness

The regulation of expression in response to DMBQ was also assessed for these motifs in *Triphysaria*. Only the TRE4 motif enhanced the expression of tdTomato after 0.5 hours of DMBQ treatment, Figure 8. The motifs TRE1-3 were not clearly up-regulated after DMBQ treatment.

DISCUSSION:

I discovered seven motifs from the promoters of host exudate responsive genes with regulatory activity in the parasitic plant *T. versicolor*. The spatial regulation of these motifs was highly conserved in two non-parasitic plants and likely represent elements used by other eudicots as well.

Triphysaria transcriptome during early haustorium development

One of the foundations on which my work was based is the RNASeq data set which contains the transcriptional response of *Triphysaria* roots exposed to *Arabidopsis* root exudates. For Orobanchaceae parasitic plants, this recognition results in the triggering of haustorium development. For non-parasitic plants, the responses to nearby root exudates can vary widely and have significant effects on development [3].

Our expression data builds off previous work in collaboration with the PPGP and adds a higher level of detail for initiation and prehaustoria development in *Triphysaria*. The raw, assembled, and quantified data are publicly available and will be a valuable resource for those investigating parasitic plants and those generally interested in chemical communications between plants.

The RNASeq analysis showed that more genes were up regulated than down regulated. Furthermore, these changes mostly occurred immediately after host factor exposure. When I analyzed the functional processes enriched in these differentially expressed genes, I also found a much higher level of over represented GO terms than under represented.

The GO terms “geraniol dehydrogenase activity” and “lignin biosynthetic process” were enriched in transcripts upregulated early after exudate exposure and downregulated later within 24 hour. The earliest enriched GO terms include processes responsible for responding to stimuli

including signal transduction, transcriptional regulation and hormone mediated activities. For instance, calcium is a well known signaling molecule and developmental regulator while the phosphorelay signal transduction system has established roles in cytokinin and ethylene signaling[36]. Another cytokinin related process enriched at early time points of prehaustorium development is “cytokinin dehydrogenase activity”, which is known to degrade cytokinin molecules.

The GO terms related to oxidation-reduction (redox) reactions are enriched in up-regulated transcripts within an hour of exposure. These include the general terms “oxidoreductase activity” and more specific redox related terms such a “iron ion binding” and “heme binding”. Shortly after this burst of redox related activities there is an enrichment of processes such as “response to hydrogen peroxide”, “response to reactive oxygen species”, “unfolded protein binding” and “response to heat”. This immediate and robust response is diminished by 12 hours. Another redox responsive term, “antioxidant activity”, increased its representation from 3 hours to 24 hours. This suggests we are able to see a progression of related activities in response to a burst of redox activity after host recognition.

Processes related to cell wall modifications, including “cell wall”, “pectin catabolic process”, “polysaccharide catabolic process”, and “peroxidase activity”, are enriched from about 3 to 24 hours. Similarly enriched are terms referring to the outside of the cell including “extracellular region” and “cell periphery”. The enrichment of these processes is likely related to the rapid increase in cell size during prehaustoria development.

Lastly, some terms are enriched in down-regulated transcripts. These include “Response to karrikin” which decreases by 0.5 hours after host exudate exposure. Terms relating to the cell-

cell connections including “cell-cell junction assembly” and “Casparian strip” are strongly enriched in transcripts down-regulated 12 hours after exposure.

Triphysaria genome

The above analyses identified *Triphysaria* transcripts up and down regulated within 24 hours of exposure to host exudates. A genomic sequence of *Triphysaria* was needed to identify the specific promoter sequences regulating these expression changes. The genome sequence we generated for *Triphysaria* using 10X Chromium libraries and Illumina sequencing was of an appropriate quality for our analyses. However, the highly polymorphic structure of *Triphysaria* led to a higher level of fragmentation compared to other published genomes. The incorporation of longer sequences through another platform such as Pac-Bio or Oxford Nanopore would likely improve the continuity of our genome.

Motif discovery

The methods I used for motif discovery and validation utilized both our RNASeq expression data and genomic sequence to identify promoter motifs enriched in host responsive genes. There are many methods and programs that can analyze our type of data for the goal of identifying patterns of expression and enriched sequences but due to a lack of consensus, a custom combination of programs and steps was developed to find enriched motifs.

There was a fair amount of judgement required at certain steps through our methods, for instance, choosing the threshold genes are considered co-expressed, how well a motif needs to match a genomic site to be present, and what level of conservation is significant during phylogenetic footprinting. These decisions leave room to improve our methods as it is often unclear what the best parameters are. The more experiment data that could be integrated into the identification and refinement of motifs, the more successful it would likely be. A method that

models the complex combination of motifs in promoters and interacting regulatory proteins could be effective in predicting the combinations and spacing of motifs that may be required for the regulation at specific times or locations.

To lessen the impact of setting strict thresholds, I started the motif discovery methods with promoters from genes with varying degrees of expression similarity. The goal was to increase the chance of finding a group of promoters with a high enough enrichment signal to rise above the noise. These included promoters from genes up-regulated at any time point in our data set to those specifically grouped into a single SOM cluster and multiple groups in between.

I also utilized three routes to identify enriched motifs to not rely too heavily on a single set of assumptions and decisions. These included identifying an enrichment of known motifs, de novo motifs, and conserved motifs through phylogenetic footprinting. All three routes successfully identified motifs with regulatory activity at a roughly similar ratio compared to the total number of motifs tested from each route. The number of motifs tested from the known and de novo enrichment approaches (5 and 7 motifs respectively) were not large numbers so it is difficult to make conclusions on the relative success rate of one technique over another. Each route did identify functional motifs however and thus show promise for future use.

Our methodology mostly identified motifs with tissue specific activity and only one motif with a temporal response to DMBQ. It is possible that the identification processes through the enrichment approaches and phylogenetic footprinting could be bias for tissue specific motifs due to a common characteristic of that type of regulatory element. The vector used to recapitulate the motifs activity may not be ideal for this type of element due to the motif location, number of repeats, motif spacing, or use of the 35S minimal promoter for instance. It is also possible the motifs that regulate expression in response to a stimulus may be more complex and require a

combination of motifs to function. Regardless of these possible issues, our methodology was effective in identifying motifs with regulatory activity in specific tissues.

TRE motif responsive to DMBQ

The TRE4 motif enhanced the expression level of tdTomato transcripts in response to DMBQ. To our knowledge, this is the first promoter motif identified that is responsive to DMBQ. The TRE4 motif was identified from the promoter of a peroxidase gene and immediate up-regulated expression 0.5 hours after DMBQ treatment. This function and timing of regulation suggests it could be involved in the burst of redox related activities that occur at the same time. The genomic distribution and conservation in DMBQ response should be further explored to determine if this motif is part of a conserved response to neighboring root chemicals or a function specific to parasites. This is of particular interest as it has been suggested that *Arabidopsis* can use DMBQ to sense the roots of neighboring plants [6].

Tissue specificity of TREs

Three TRE motifs identified had clear activity at the root tip. The genes they were identified from all peaked in transcriptional expression by 1 hour after exudate treatment. The sequence for TRE1 came from a homolog of calmodulin-binding protein 25. These proteins act as secondary messengers by binding to calcium and interacting with downstream targets to perpetuate calcium signal transduction pathways [37]. A calcium binding GO term was enriched in our RNASeq and calcium has been shown to be involved in triggering haustorium development in the closely related parasite *Phtheirospermum* [38]. TRE2 was identified from a BAG family molecular chaperone regulator gene which has shown to be involved in both abiotic and biotic stress responses [39]. The tissue specific expression of both TRE1 and 2 suggests these genes are expressed and function in the columella.

TRE3 originated from a 10-hydroxygeraniol dehydrogenase (10HGO) promoter. promoter. This gene is a zinc-dependent oxidoreductase involved in the biosynthesis of 10-oxogeranial which is a crucial step in iridoid biosynthesis [40]. The GO term “geraniol dehydrogenase activity” was enriched at 1 hour in up-regulated genes and enriched in down-regulated genes between 6-24 hours. Oxidoreductases have known roles in haustorium development and the expression of this specific type of oxidoreductase seems to be dynamically regulated during host recognition and prehaustorium development.

To summarize the genomic observations for this gene, the de novo motif responsible for identifying the TRE3 sequence was copied multiple times and conserved in the promoter of 10HGO, this gene was copied multiple time in the parasites without its introns, some of this copying look to have occurred independently in the parasites and seems to be specific to the parasites when compared to a sample of non-parasite genomes.

The second group of TRE motifs, TRE4-7, all had expression in mature root tissues. The genes these motifs originated from all increased in transcriptional expression up to the last time point in our data set, suggesting a role in tissues developing later in our data set such as root hairs. TRE4 was found from a de novo motif in a lignin-forming anionic peroxidase gene. Peroxidases are heme-containing glycoproteins with known roles in removing hydrogen peroxide, oxidizing toxic reductants, the metabolism of lignin in cell walls, and have been shown to release HIFs from *Triphysaria* roots [16,41]. TRE4 was the only motif tested to be confirmed as DMBQ responsive. The expression of TRE4 was also found to be predominantly localized in the epidermis and root hairs. Together, these observations fit with the motif regulating genes in response to external chemical stimuli.

Both TRE5 and 6 were found to be expressed specifically in root and prehaustorial hairs. These motifs were independently identified through phylogenetic footprinting of a pistil-specific extensin-like protein (PELP) and a non-classical arabinogalactan protein 30 (AGP30). Both genes were chosen for their host induced expression that increased across our RNASeq data set. These genes are types of structural glycoproteins that form crosslinked networks with components in cell walls and have been found to be important for cell expansion including root hair growth [42]. The hair specific expression of both motifs supports the role of these specific genes in root hair development.

The motifs for TRE5 and 6 convergently resembled a conserved root hair specific motif known as RHE [43]. TRE5 matched the consensus motif presented for RHE, while TRE6 was close but had some clear differences as well. The conserved regulation of TRE6 in *Arabidopsis* and *Mimulus* root hairs suggests the root hair specific motif RHE should be amended to include more sequences to encompass more of the variation present in root hair specific motifs such as TRE6.

TRE7 was identified from the gene Lateral root primordia 1 (LRP1) which was found to be highly expressed in early lateral roots and functions in root elongation [33,34]. Our refined motif was able to recapitulate some of this tissue specific regulation with the lateral root sites clearly showing tdTomato expression in *Triphysaria*. This was one of the few motifs without conserved regulatory activity in the non-parasitic plants which suggests *Triphysaria* uses a slightly different motif to control the location of LRP1 expression.

Transcription factors

The number of transcription factors from each family that were identified from the refined set of *Triphysaria* coding sequences are shown in Figure S13. A few cases stand out where some transcription factor families in *Triphysaria* look to have amplified when compared to *Mimulus* and *Arabidopsis*. For instance, the FAR1 family in *Triphysaria* has more than eight times as many copies as *Arabidopsis*, four fold more than *Mimulus*, and more copies than 98% of the plants present in the Plant Transcription Factor Database [24]. Interestingly, a known FAR1 site was equally similar to TRE1 as the CAMTA site used to identify it. Thus, a FAR1 transcription factor may be the regulator of TRE1. Additionally, FAR1 transcription factors did not change in expression in our RNASeq. This expansion of FAR1 in *Triphysaria* could relate to a different stage of parasitism than what was studied, not require a change in transcription to serve its functional role, or perhaps serve a purpose outside of parasitism as *Triphysaria* can live autotrophically as well.

The predicted binding sites of known transcription factors were identified in the TRE sequences using the PlantTFDB, Figure S14. All seven TRE sequences had at least 1 significant match (p-value $\leq$ 1e-5) with a known transcription factor binding motif. TRE3 contained up to 7 significant predicted sites. These predictions serve as a starting point for future research on the transcription factors that parasitic and non-parasitic plants use to bind the TRE sequences and regulate transcriptional expression.

Translational opportunities of plant responsive motifs

The root specific patterns of transcription conferred by five different sequence motifs were highly conserved in both parasitic and non-parasitic plants. One of these, motif TRE4, was specifically upregulated when roots were exposed to the host root factor DMBQ. The ability to

control where and when a gene is expressed with a minimal motif sequence has clear applications in synthetic biology to more readily engineer the expression dynamics of a desired gene. My reporting construct could facilitate this goal by replacing the tdTomato gene with a candidate gene of choice. Once small regulatory elements are identified, they can be deleted or replaced in planta using a CRISPR system.

FIGURES:

A.

SOM	# of transcripts	0.5hr C	0.5hr	1hr	3hr	6hr	12hr	24hr	24hr C
8	107								
12	65								
11	171								
16	79								
10	198								
15	49								
13	35								
14	105								
6	105								
9	51								
1	51								
5	39								
2	186								
4	144								
3	47								
7	187								

B.

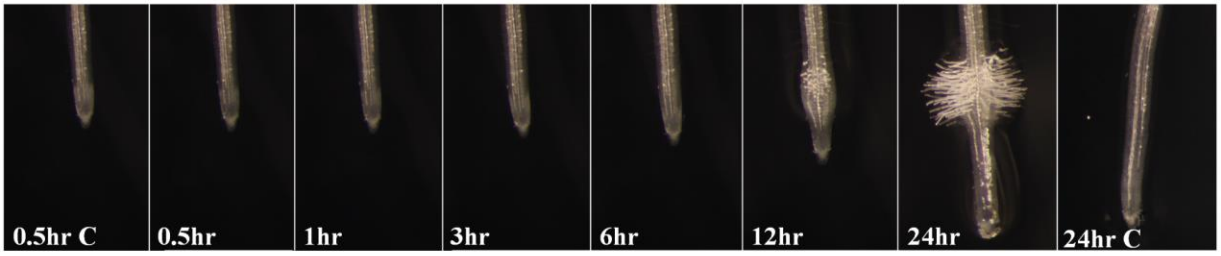


Figure 1: Clustering co-expressed transcripts.
A. Differentially expressed transcripts were clustered into 16 groups using a Self-Organizing Map (SOM) based on their expression similarity across the RNASeq data set. The average transcript level for each group across the 8 time points is shown as a heat map. The number of transcripts in each SOM group is shown to the right. B. Representative images of *T. versicolor* root morphology at each developmental time point in the RNASeq.

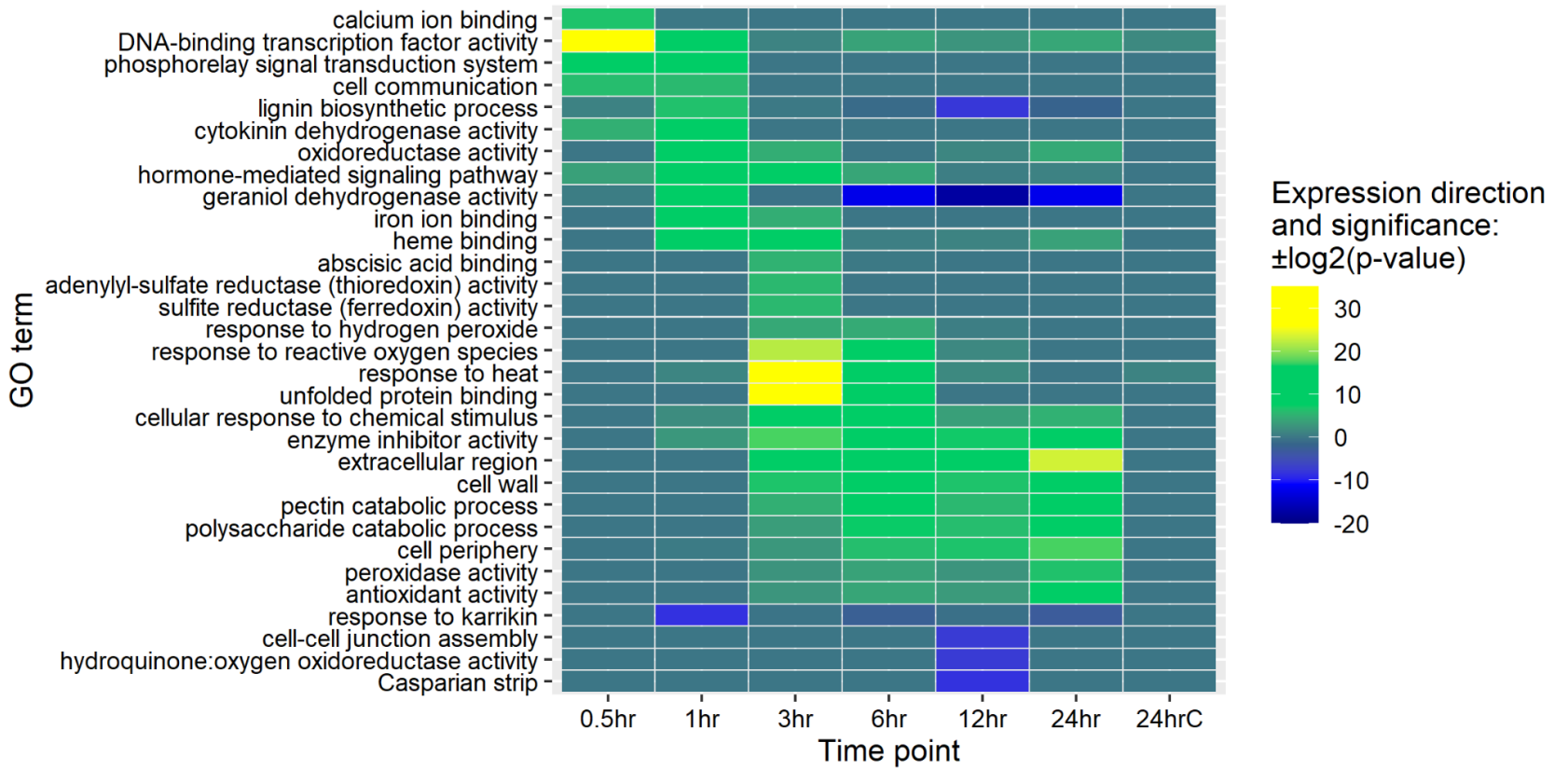


Figure 2: Gene ontology (GO) term enrichment during prehaustorium development

The GO terms shown were selected based on their level of enrichment and functional attribute. The heat map indicates the level of statistical significance each term is enriched in transcripts differential expressed at each time point. Positive (yellow) and negative (blue) values indicate an enrichment in up-regulated or down-regulated transcripts respectively.

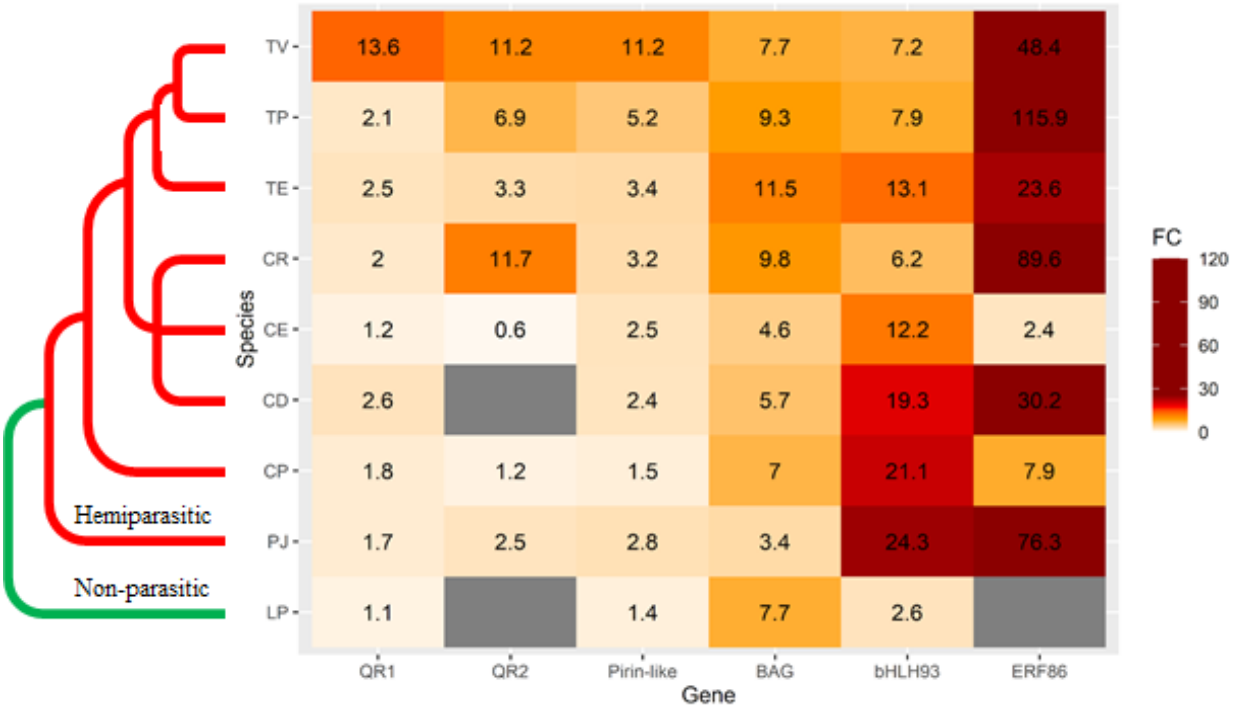


Figure 3: Heat map of qPCR fold change values for host responsive transcripts from multiple hemiparasitic and one non-parasitic member of Orobanchaceae. The transcripts were chosen based on their host responsive activity in the *Triphysaria* RNASeq data and functional annotation. Roots from each species were treated with water for 0.5 hours or *Arabidopsis* exudate for 1 hour before collection. The numbers in the heat map represent the fold change in transcript level. Grey boxes represent data not obtained due to the common primers not working for that specific target. The phylogenetic relationships between these species is shown. QR1: Quinone Reductase 1, QR2: Quinone Reductase 2, BAG: BAG family molecular chaperone regulator 2, bHLH93: Basic helix-loop-helix protein 93, ERF86: Ethylene-responsive Transcription Factor 86. TV: *Triphysaria versicolor*, TP: *Triphysaria pusilla*, TE: *Triphysaria eriantha*, CR: *Castilleja rubicundula*, CE: *Castilleja exserta*, CD: *Castilleja densiflora*, CP: *Cordylanthus pilosus*, PJ: *Phtheirospermum japonicum*, and LP: *Lindenbergia philippensis*.

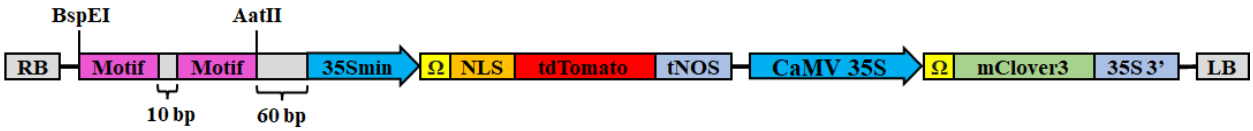


Figure 4: Motif reporting vector

pDS-MRC is a T-DNA based plasmid to evaluate the regulatory activity of promoter motifs using the red fluorescence protein tdTomato. tdTomato was localized to nuclei to aid in identifying cells. The green fluorescent protein mClover3 served as a selectable marker for transformation and to outline the cells during imaging. Motifs were cloned as either a single copy or tandem duplication (with a 10 bp spacer sequence) into the construct. The sizes of each element in the diagram are not to scale. RB: Right border, 35Smin: Minimal CaMV 35S promoter, Ω: Omega translational enhancer, NLS: Nuclear localizing signal, tNOS: NOS terminator, 35S 3': CaMV 35S terminator, LB: Left border.

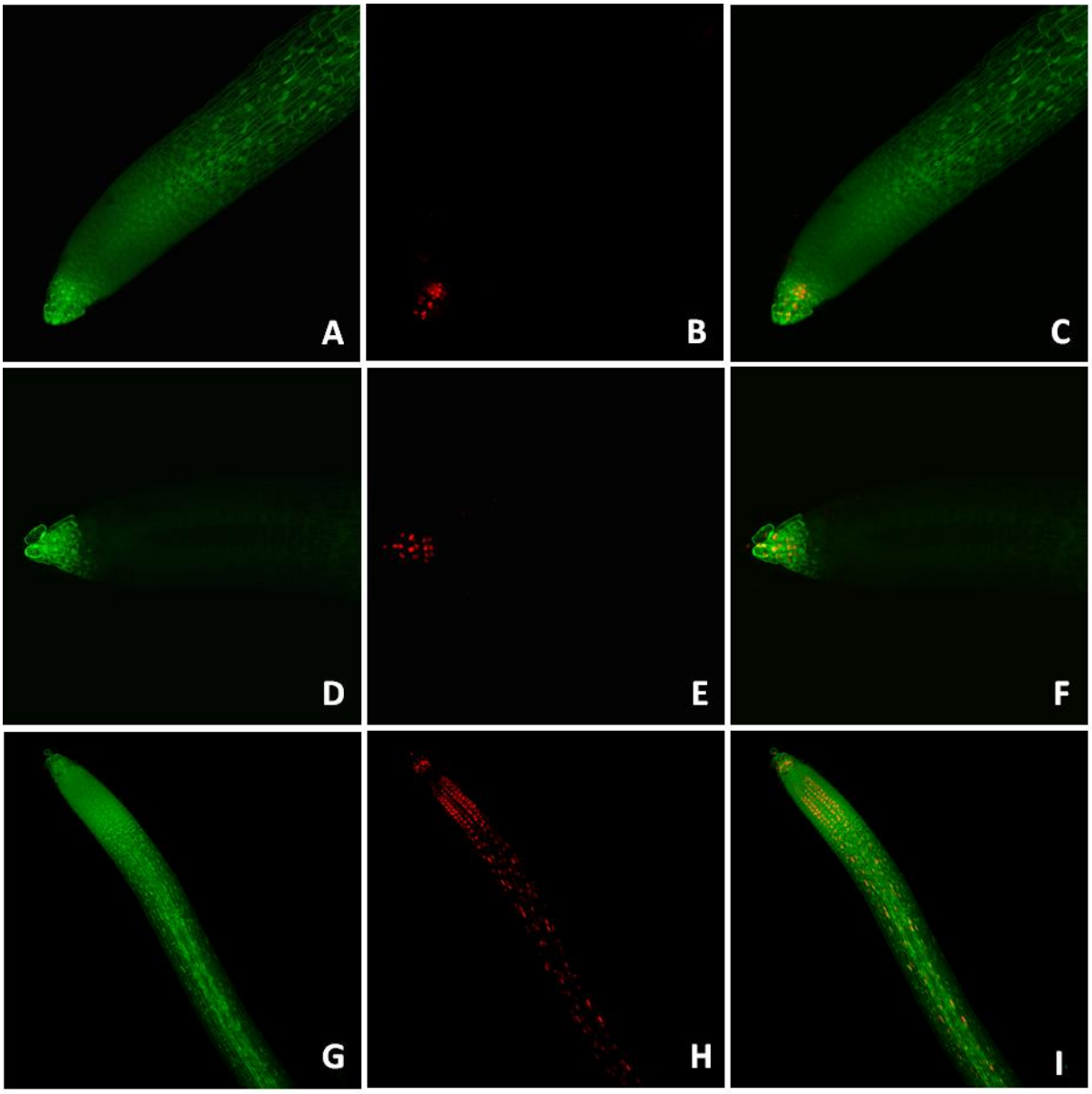


Figure 5: Tissue specific expression of TRE motifs in Triphysaria
Confocal images of *T. versicolor* roots transformed with pDS-MRC containing the active TRE motifs shown in Table 1. Green, red, and green/red overlay images are presented for each root. A-C: TRE1, D-F: TRE2, G-I: TRE3, J-L: TRE4, M-O: TRE5, P-R: TRE6, S-U: TRE7.

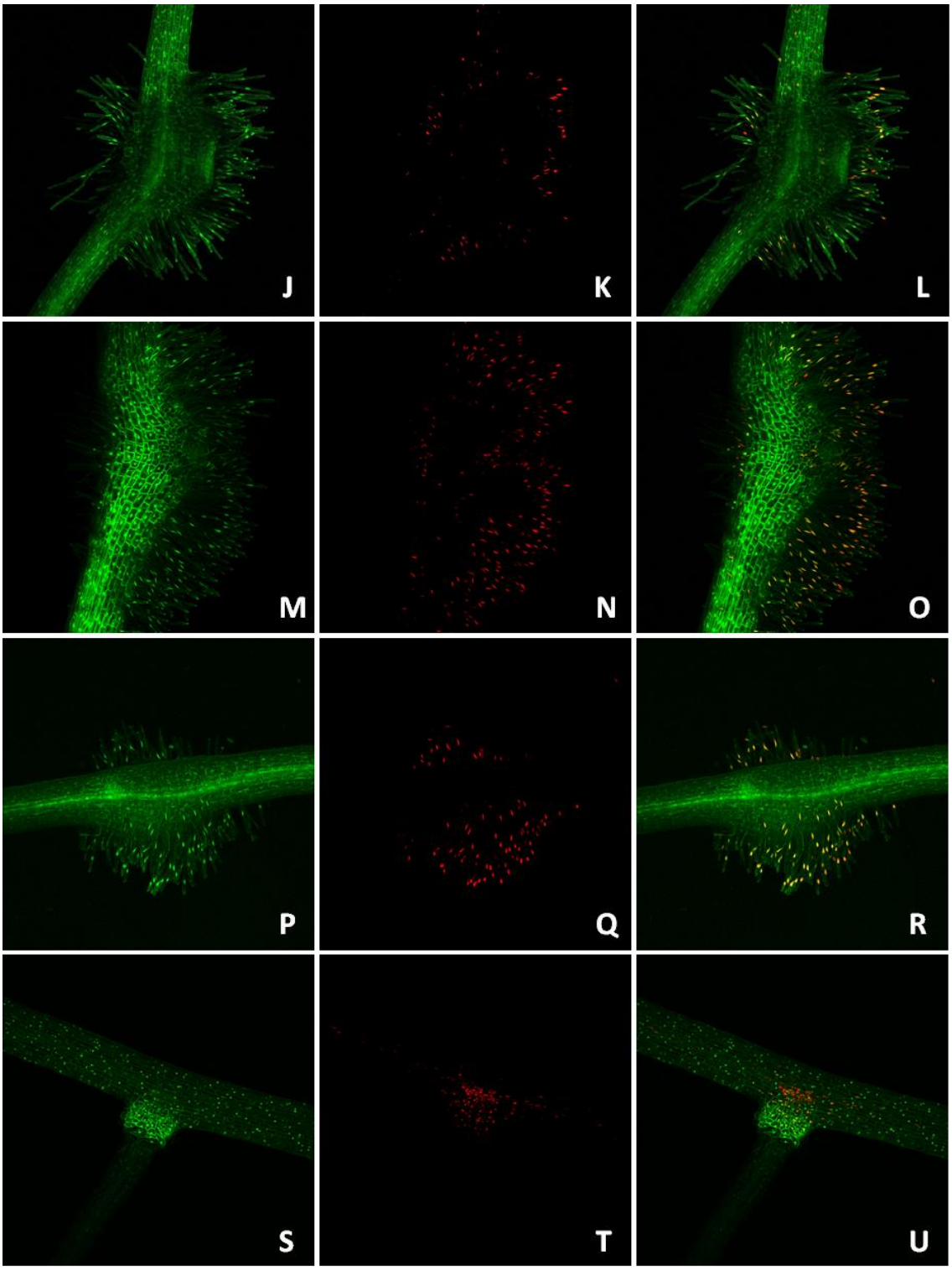


Figure 5 continued: Tissue specific expression of TRE motifs in *Triphysaria*
Confocal images of *T. versicolor* roots transformed with pDS-MRC containing the active TRE motifs shown in Table 1. Green, red, and green/red overlay images are presented for each root. A-C: TRE1, D-F: TRE2, G-I: TRE3, J-L: TRE4, M-O: TRE5, P-R: TRE6, S-U: TRE7.

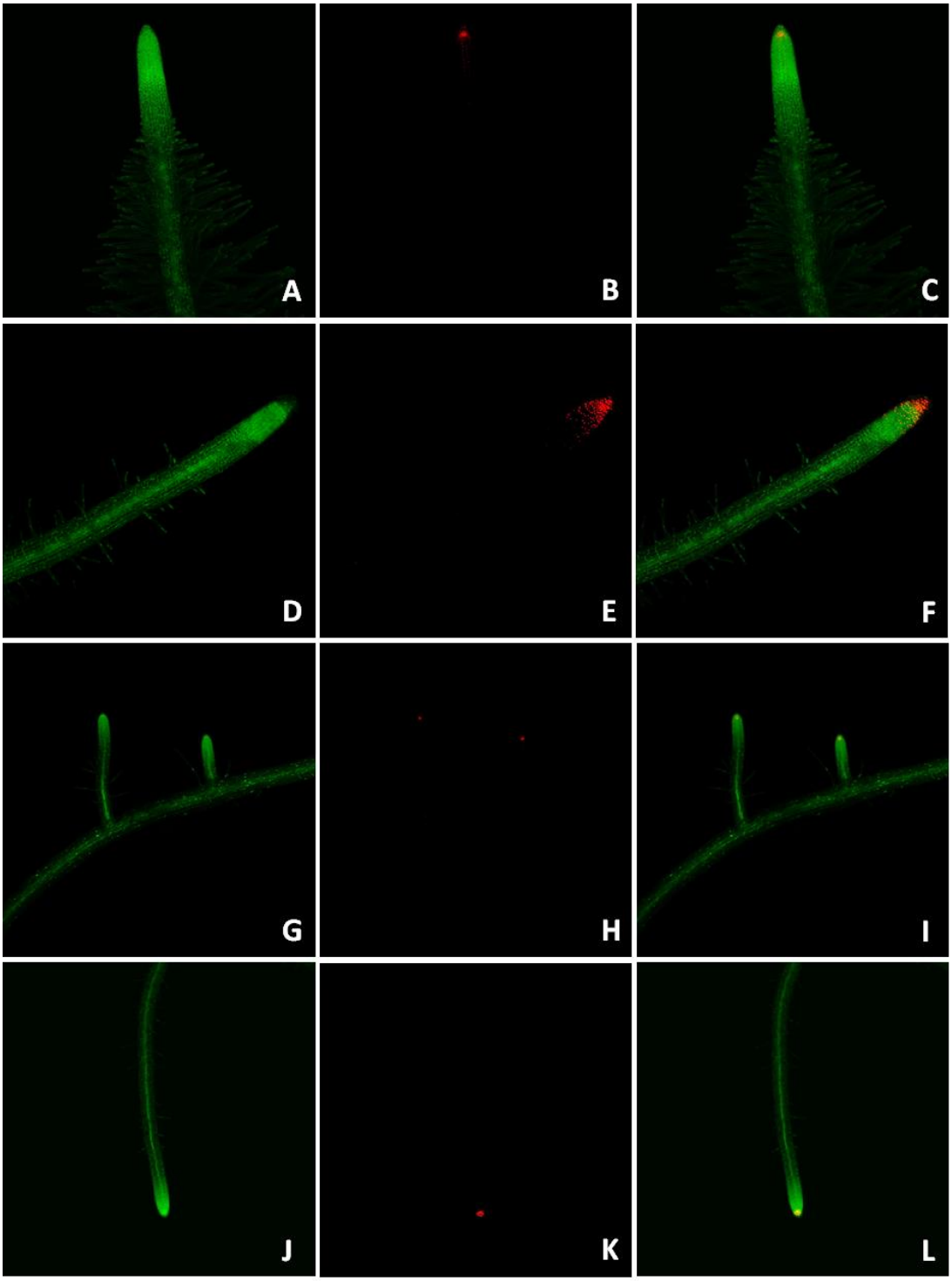


Figure 6: Root tip expression of TRE1 and 2 in the non-parasitic plants *Mimulus guttatus* and *Arabidopsis thaliana*.

Confocal images of roots from the non-parasitic plants *Mimulus guttatus* (A-F) and *Arabidopsis thaliana* (G-L) transformed with pDS-MRC containing the TRE motifs TRE1 (A-C, G-I) or TRE2 (D-F, J-L)

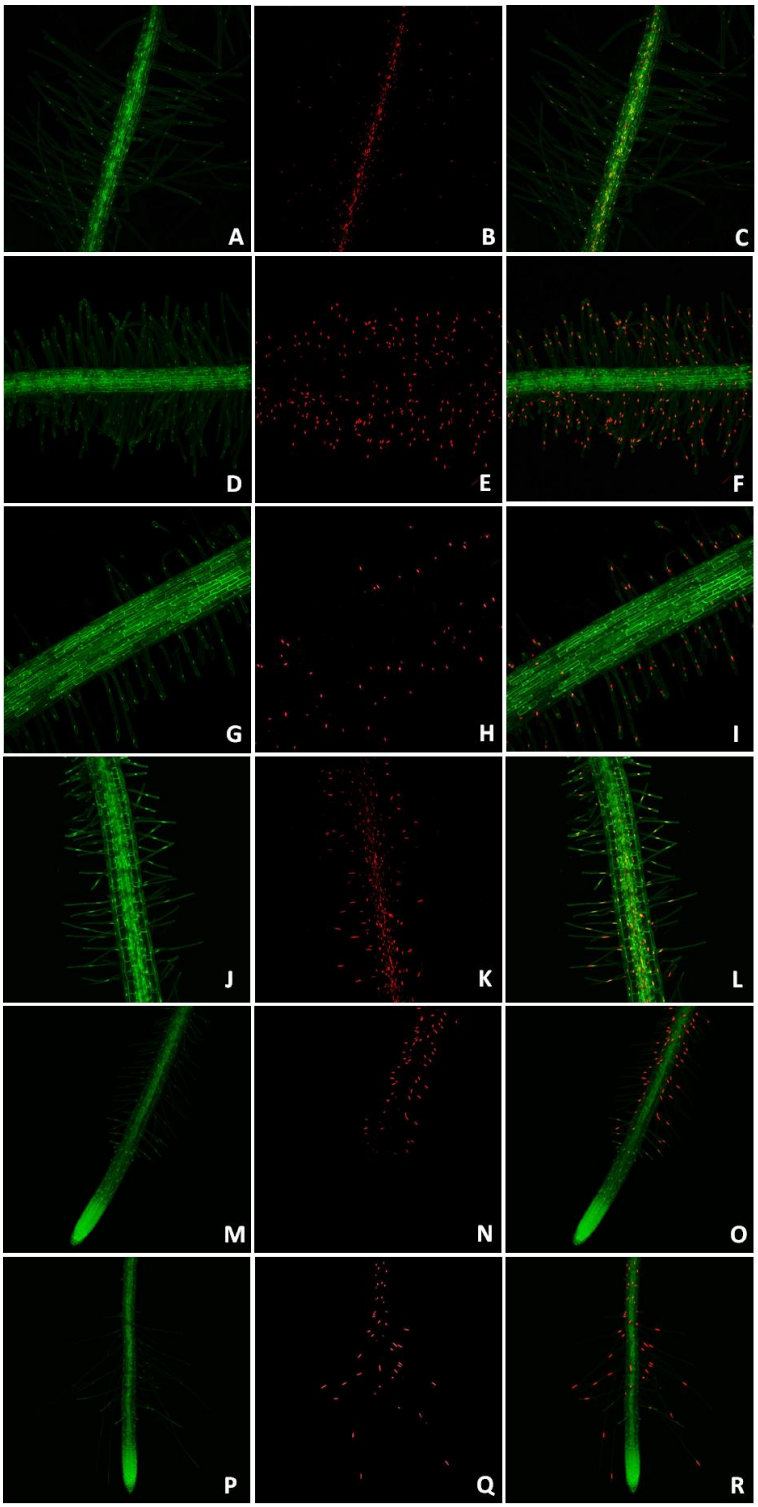


Figure 7: Epidermal and hair expression of TRE4, 5, and 6 in *Mimulus guttatus* and *Arabidopsis thaliana*.

Confocal images of roots from M. guttatus (A-I) or A. thaliana (J-R). TRE4 (A-C, J-L), TRE5 (D-F, M-O), TRE6 (G-I, P-R).

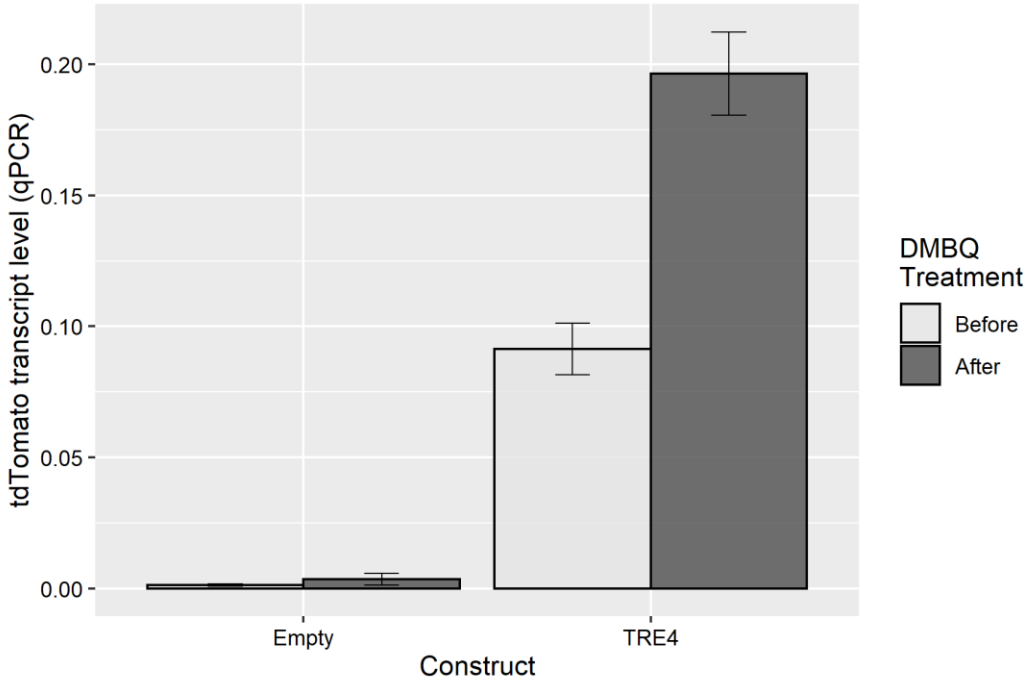


Figure 8: TRE4 is responsive to DMBQ

Quantitative PCR (qPCR) of tdTomato transcript levels in transgenic roots bearing pDS-MRC-TRE4 before and after exposure to 30μM DMBQ for thirty minutes. The relative transcript levels were determined using two independent reference genes. The bars represent the average of 2-3 independently transformed plants. The Empty vector is the full pDS-MRC construct without the motif sequences. The error bars represent the standard error for each group of measurements. The means for TRE4 are significantly different between DMBQ treatments (p-value = 0.045), while the means for the Empty construct were not significantly different between treatments (p-value = 0.436).

TABLES:

Prehaustorium RNASeq and Transcriptome	
Illumina library type:	Paired-end, 150 bp
Time points:	8
Replicates:	3
Focus:	Early development
Host inducing factor (HIF):	Arabidopsis root exudate
Refined transcript number:	23378
Average transcript size:	1575 bp
Functionally annotated transcripts:	99.98%
Differentially expressed transcripts:	1619

Table 1: RNASeq dataset and the *Triphysaria* transcriptome
Twenty four RNASeq reactions were run on RNA samples collected from *Triphysaria* roots harvested at eight timepoints after exposure to *Arabidopsis* root exudate. We identified 1619 differentially expressed (DE) genes at a threshold of a ≥ 2-fold change with an adjusted p-value ≤ 0.01. Transcripts were functionally annotated using Blast2GO and the NCBI nr database.

Triphysaria versicolor Genome	
Library type:	10X Chromium
Genome size:	2.02 Gb
Genome coverage:	39X
Scaffold number:	235998
N50:	23077 bp
L50:	10983 scaffolds
Predicted genes:	62730
BUSCO	87.4%

Table 2: *Triphysaria* versicolor genome
10X chromium sequencing was used on high molecular weight DNA isolated from a single *Triphysaria* plant. N50: the shortest contig length at which half the genome is contained on equal to or larger contigs. L50: the smallest number of contigs that contain half of the genome, BUSCO: Benchmarking Universal Single-Copy Orthologs. The BUSCO score reflects the presence of single copy genes identified.

Motif	Gene	Method	Size	Expression
TRE1	Calmodulin-binding protein 25-like	Known site	15	Columella with concentration near initials.
TRE2	BAG family molecular chaperone regulator 2	Phylo-foot	34	Columella
TRE3	Hydroxygeraniol dehydrogenase	De novo	71	Root cap, ground tissue, some vascular.
TRE4	Lignin-forming anionic peroxidase	De novo	26	DMBQ responsive, epidermis, some vascular and columella.
TRE5	Pistil-specific extensin-like protein	Phylo-foot	18	Root hairs
TRE6	Non-classical arabinogalactan protein 30	Phylo-foot	18	Root hairs
TRE7	LATERAL ROOT PRIMORDIUM 1-like	Phylo-foot	36	Lateral root junction

Table 3: Motifs with regulatory activity in *Triphysaria*

Seven motifs with clear regulatory activity in *Triphysaria* roots are listed with details about the gene whose promoter yielded the motif, how the motif was identified, the size of the motif and the expression patterns observed. TRE: Triphysaria Regulatory Element.

SUPPLEMENTAL:

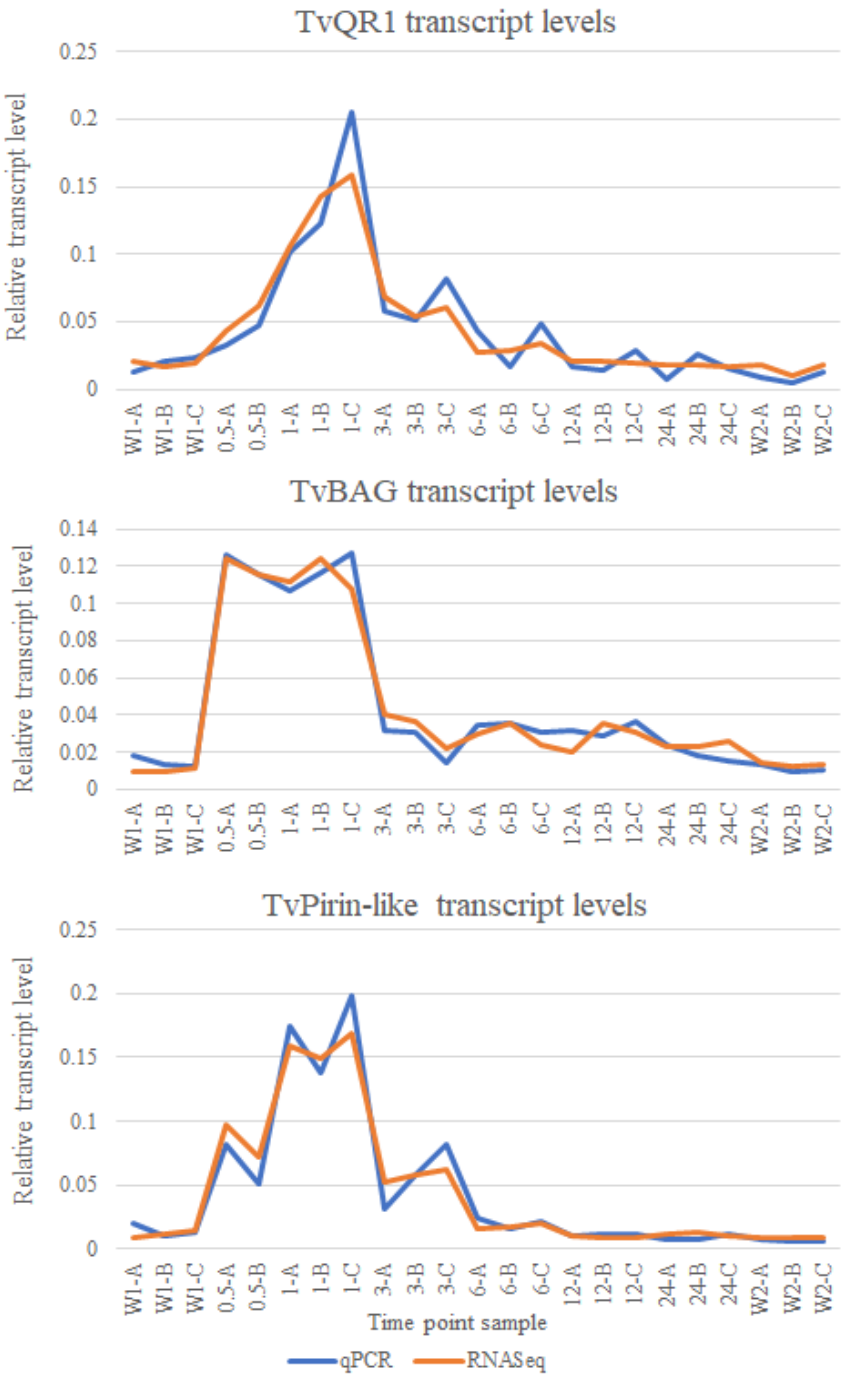


Figure S1: Validation of transcript quantification using RNASeq and qPCR. The relative transcript levels for three host-responsive genes are shown from the RNASeq data set and qPCR. The relative transcript levels for each gene were obtained by dividing the absolute transcript level for each sample by the total of all absolute transcript levels for that gene.

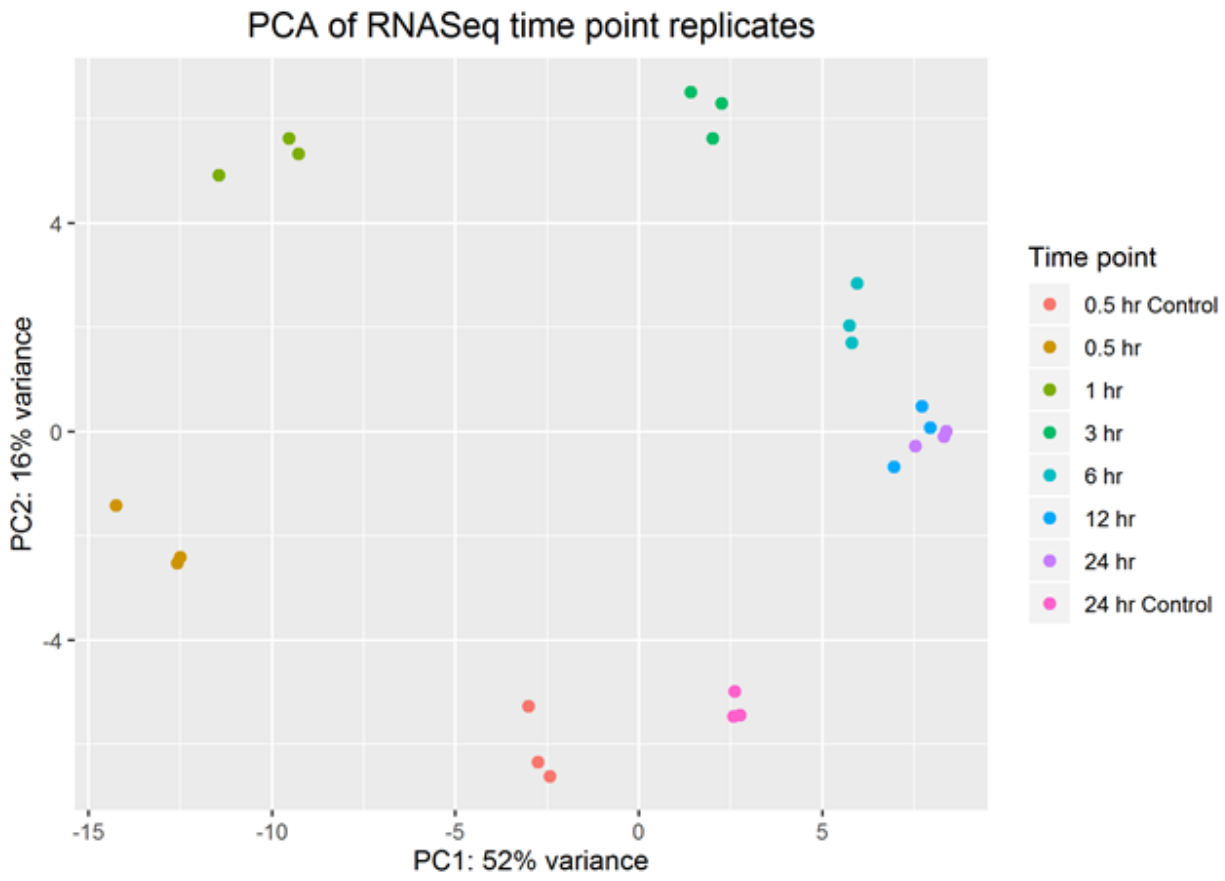


Figure S2: Principal component analysis of RNASeq samples. The variance between the RNASeq expression values for all 24 samples (8 time points and 3 replicates each) were estimated and graphed using DESeq2.

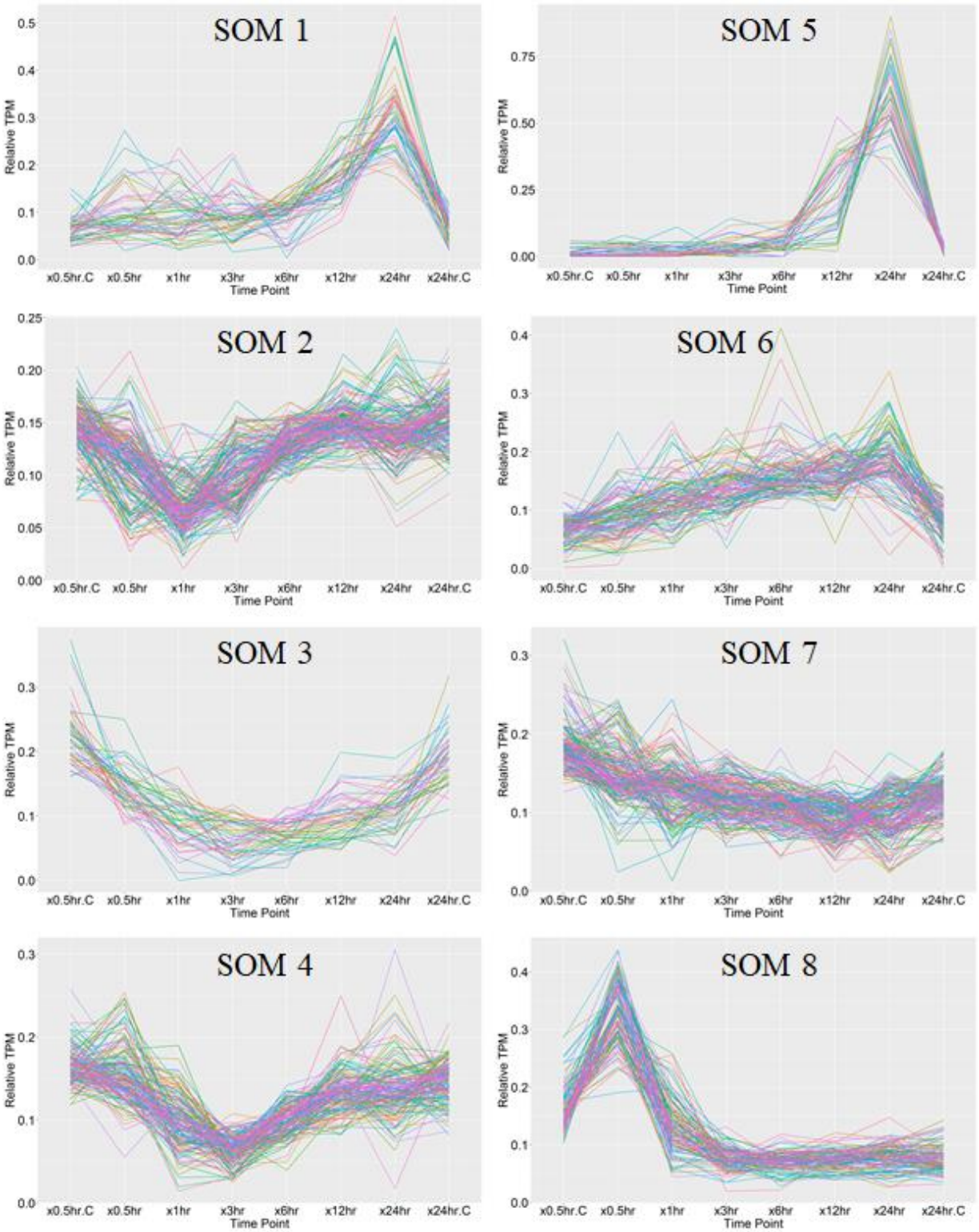


Figure S3: Relative expression profiles for all transcripts in each self-organizing map (SOM) created group. The SOM placed all differentially expressed transcripts into one of 16 different groups based on the similarity of their expression profiles across our eight time points.

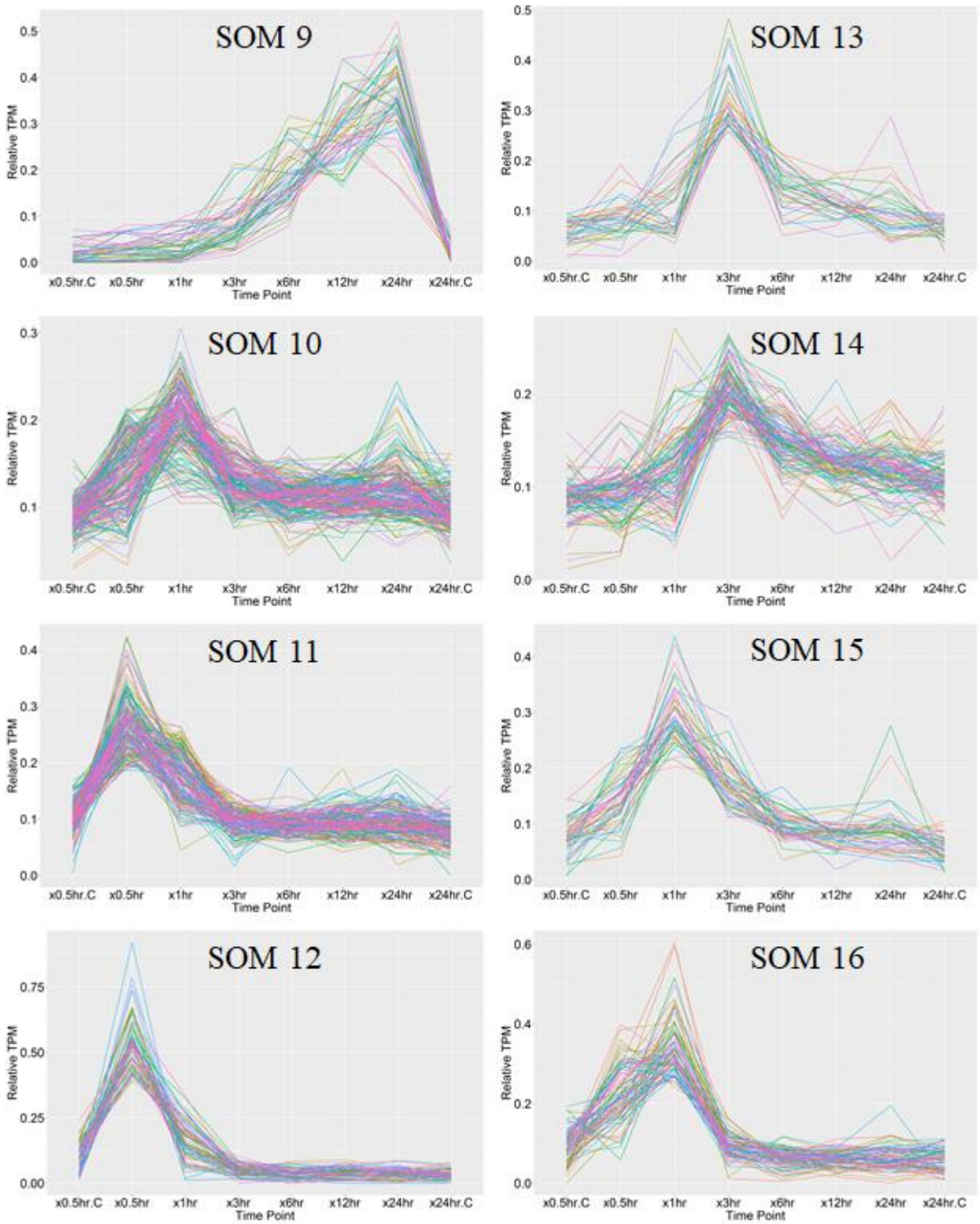
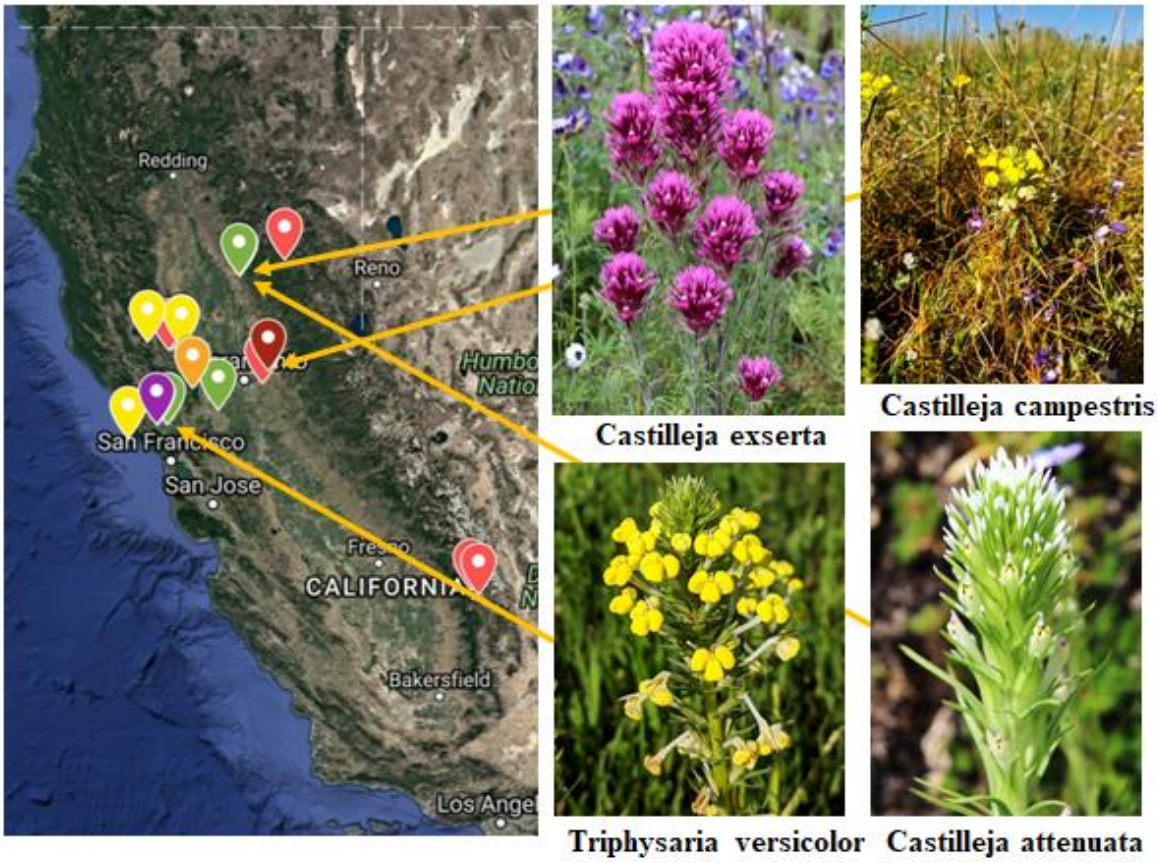


Figure S3 continued: Relative expression profiles for all transcripts in each self-organizing map (SOM) created group. The SOM placed all differentially expressed transcripts into one of 16 different groups based on the similarity of their expression profiles across our eight time points.

A



B

Symbiota ID	Catalog Number	Scientific Name	Decimal Lat.	Decimal Long.
3793957	DAV351143	Castilleja minor	38.8521472	-122.3973639
3793989	DAV396710	Pedicularis groenlandica	36.4468194	-118.5947639
3793949	DAV397799	Cordylanthus pilosus	38.663399	-121.191775
3793963	DAV396833	Triphysaria versicolor subsp. Faucibarbata	38.784182	-123.007948
3793979	DAV351394	Castilleja rubicundula	38.868784	-122.402197
3793967	DAV396834	Triphysaria versicolor subsp. Faucibarbata	38.126244	-122.592946
3793973	DAV397760	Bellardia trixago	38.129525	-122.599826
3793976	DAV397896	Parentucellia viscosa	38.129922	-122.599761
3793984	DAV350129	Castilleja densiflora	38.126	-122.595
3793971	DAV350113	Castilleja campestris	38.535023	-121.262605
3793987	DAV350036	Castilleja attenuata	38.53465	-121.261722
3793992	DAV350208	Castilleja exserta	39.598838	-121.5533

Figure S4: Hemiparasitic Orobanchaceae collections in California. A. Map showing the seed collection sites across California from different genera of hemiparasitic Orobanchaceae. The images shown to the right were taken at the collection sites. B. Details for the voucher specimens submitted to the University of California Davis herbarium for each of the species listed.

Gene	Pj mircoarray	StHe BC3	StHe BC4	OrAe BC5	Castillejineae qPCR
Ethylene-responsive transcription factor ERF109-like	✔			✔	
Pirin-like protein	✔				✔
Transcription factor bHLH93-like	✔				✔
Quinone Reductase 2 (QR2)	✔	✔			
Trihelix transcription factor GT-3b-like	✔	✔	✔		
Gibberellin 2-beta-dioxygenase-like	✔	✔	✔		
Lignin-forming anionic peroxidase-like	✔	✔	✔		
Probable carboxylesterase 6		✔			
Non-classical arabinogalactan protein 30		✔	✔		
Uncharacterized protein		✔		✔	
Uncharacterized protein 2		✔		✔	
Putative pectate lyase 2		✔		✔	
Aquaporin NIP1-1-like		✔	✔	✔	
Pistil-specific extensin-like protein			✔		
Calmodulin-binding protein 25-like			✔	✔	
Scarecrow-like protein 14			✔	✔	
Gibberellin receptor GID1B-like			✔	✔	
Lateral Root Primordium 1-like			✔	✔	
KAI2			✔	✔	
17.3 kDa class II heat shock protein-like			✔	✔	
10-hydroxygeraniol oxidoreductase				✔	
AP2/ERF and B3 domain-containing transcription factor RAV1-like				✔	
Ethylene-responsive transcription factor ERF086		✔		✔	✔
BAG family molecular chaperone regulator 2-like					✔
Quinone Reductase 1 (QR1)					✔
Zinc finger protein 5					
Heat shock protein 83					
Ethylene-responsive transcription factor ERF003-like					

Figure S5: Genes with expression changes across multiple parasitic Orobanchaceae whose promoters were used in phylogenetic footprinting to identify conserved motifs. The chart shows the frequency at which orthologs for each of the listed genes were also up-regulated in other Orobanchaceae parasite data sets. The Phtheirospermum japonicum (Pj) microarray data was provided by Satoko Yoshida while the Striga hermonthica (StHe) and Orobanche aegyptiaca (OrAe) expression data was provided by the Parasitic Plant Genome Project.

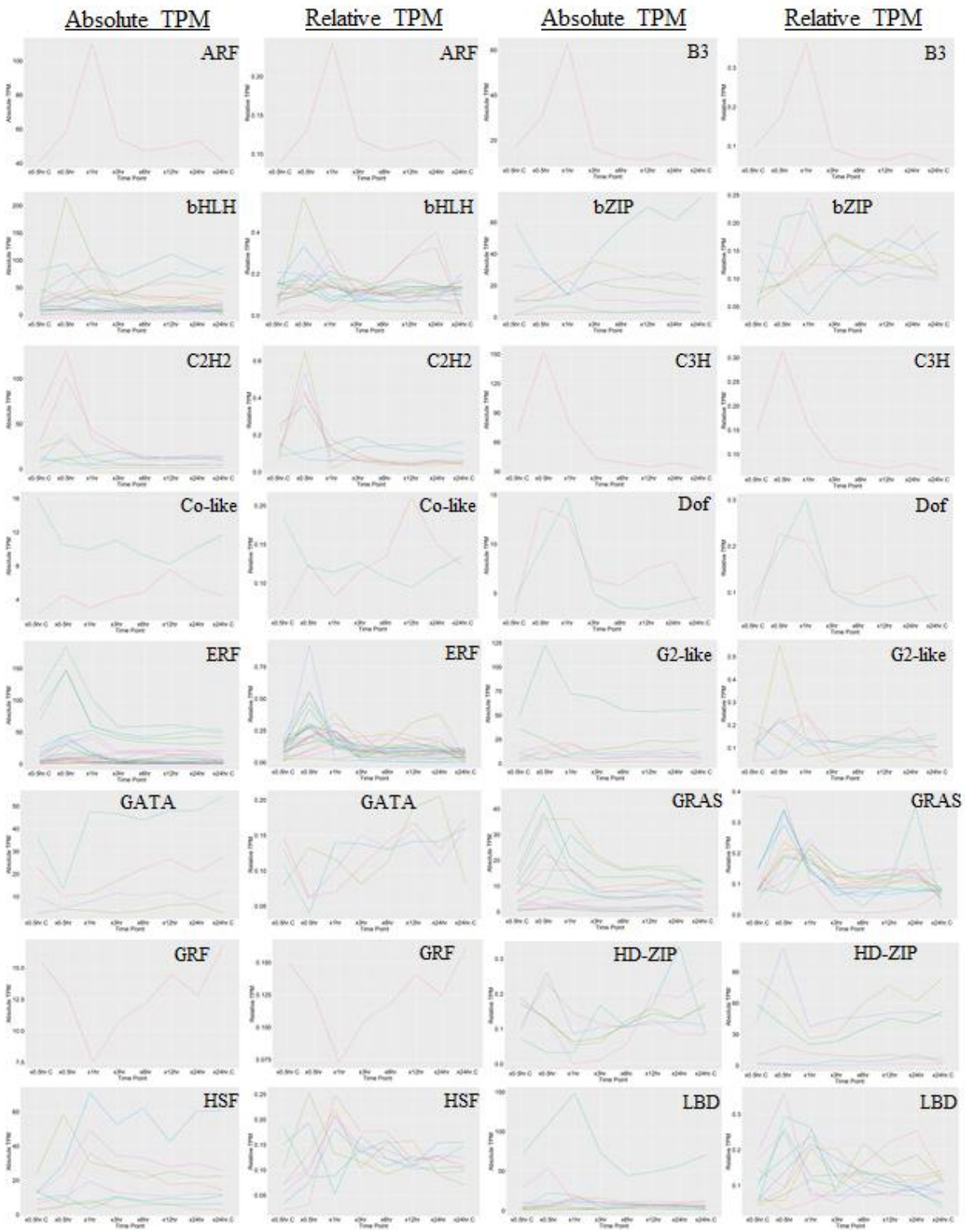


Figure S6: Absolute and relative expression profiles of differentially expressed transcripts from each transcription factor (TF) family. TFs were annotated using the Plant Transcription Factor Database v5 and graphed if they had at least a 2-fold change in expression.

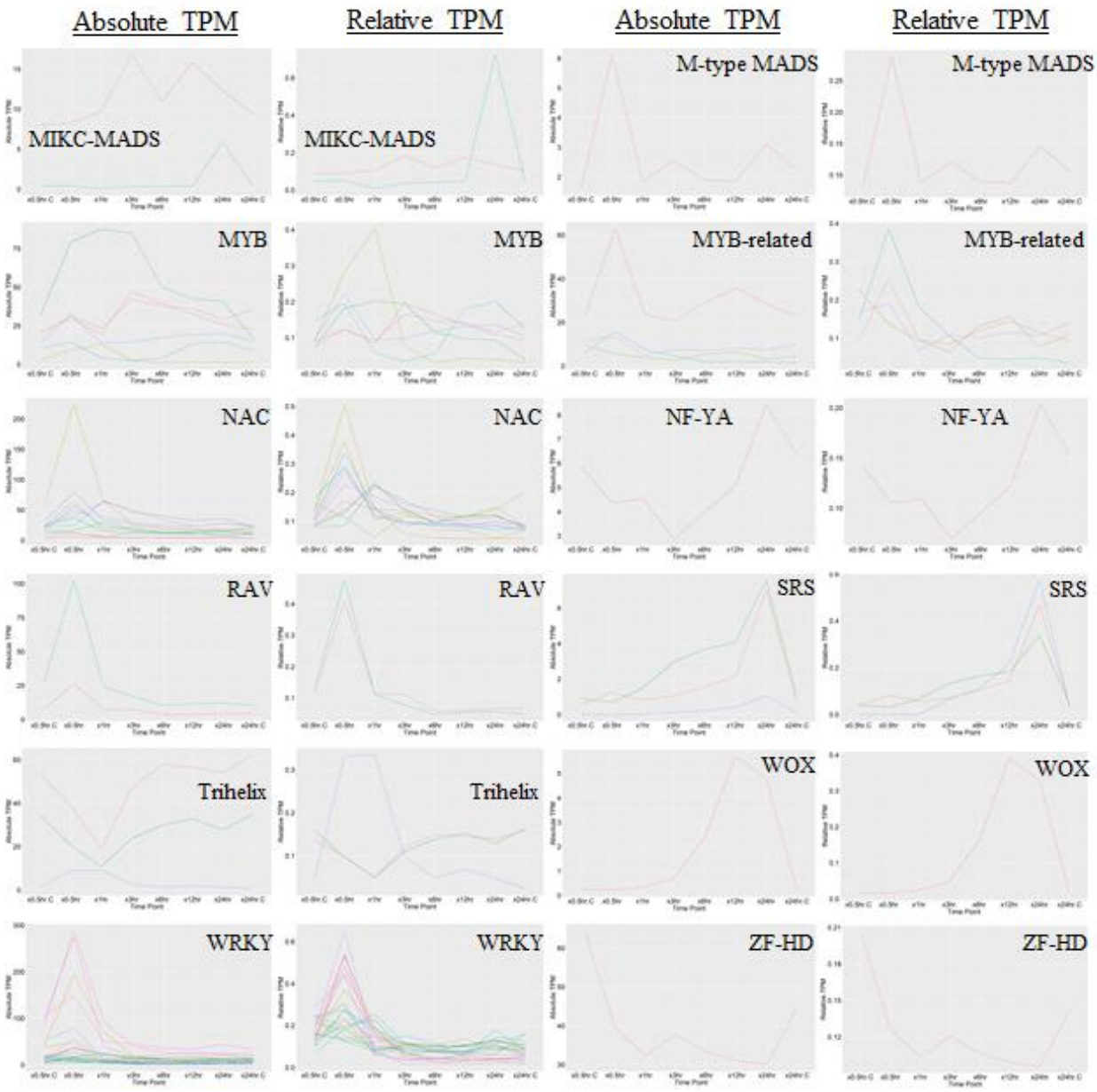


Figure S6 continued: Absolute and relative expression profiles of differentially expressed transcripts from each transcription factor (TF) family. TFs were annotated using the Plant Transcription Factor Database v5 and graphed if they had at least a 2-fold change in expression in our data set.

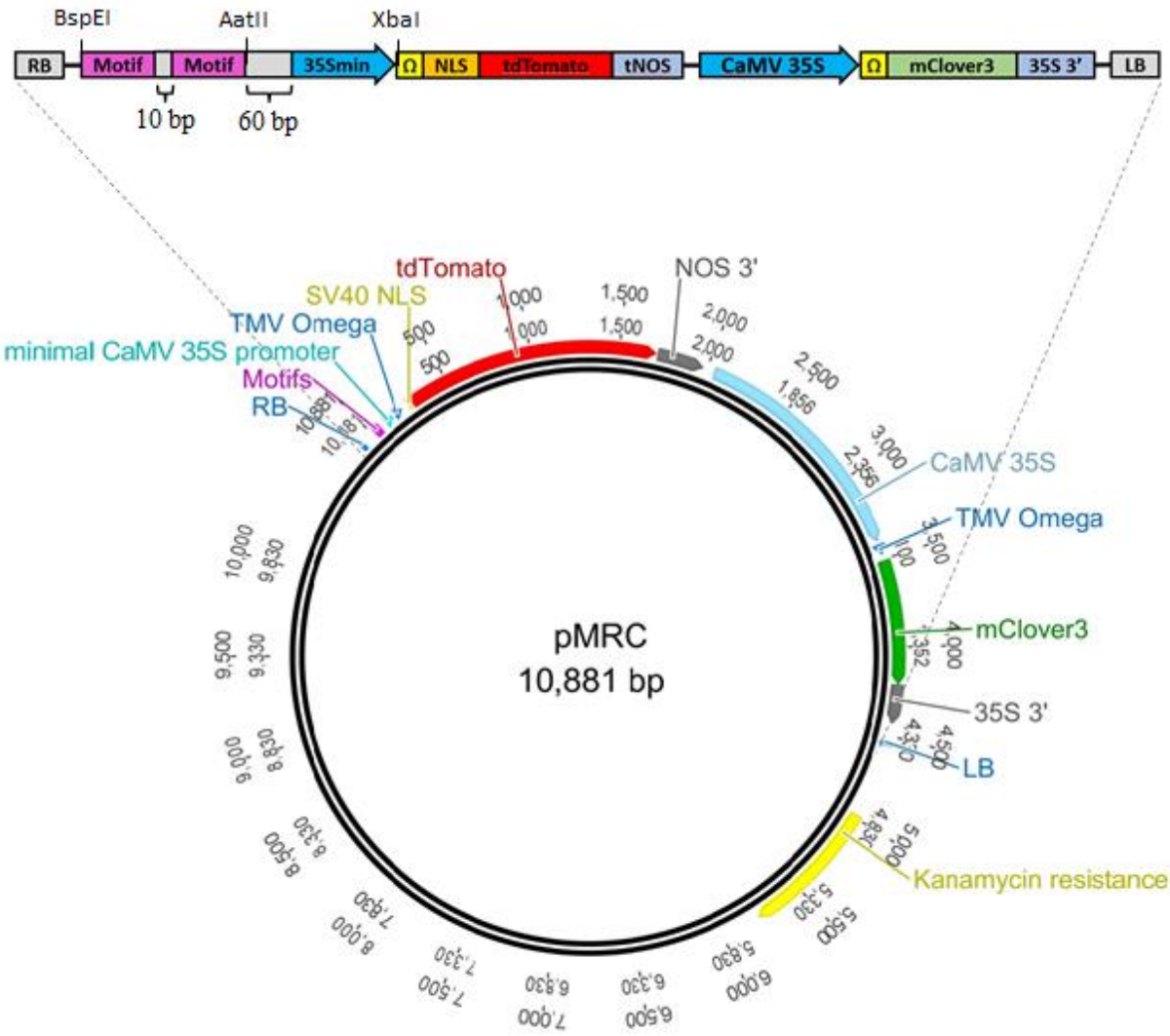


Figure S7: Diagram of the motif reporting construct pMRC. The circular image of pMRC was generated through the program Geneious and represent the exact location and size of each element. The linear diagram is easier to read but the elements are not proportional in size. RB: Right border, LB: Left border, 35Smin: Minimal 35S promoter, Ω: TMV Omega, NLS: Nuclear localization signal, tNOS: NOS terminator, CaMV 35S: Full 35S promoter, and 35S 3': 35S terminator.

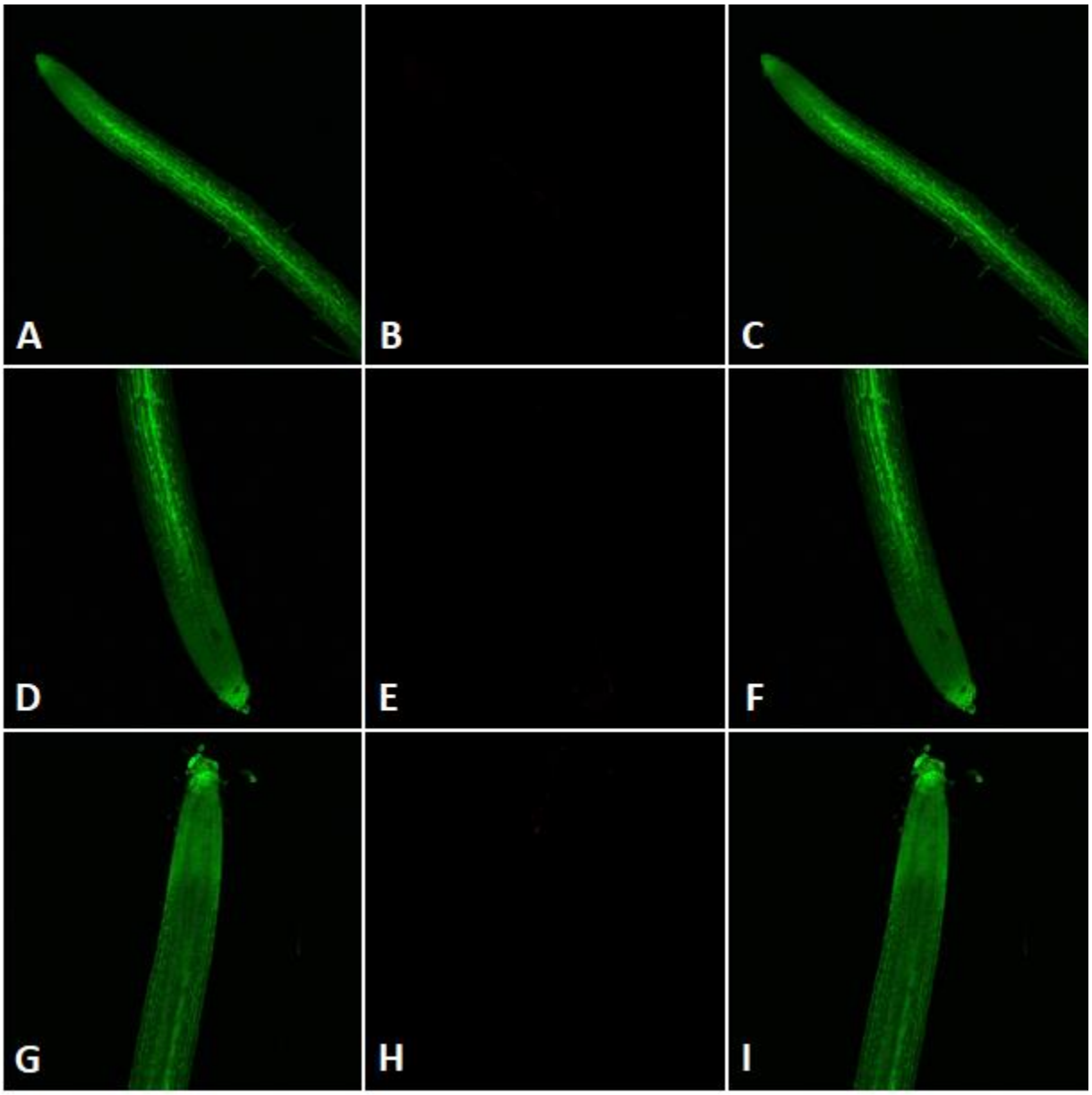


Figure S8: Confocal images of root tips transformed with pDS-MRC-Empty.
Confocal images of *T. versicolor* roots transformed with pDS-MRC-Empty which did not contain a motif. Green, red, and green/red overlay images are presented for each root. Three independently transformed roots are shown. A-C: Root1, D-F: Root2, G-I: Root3.

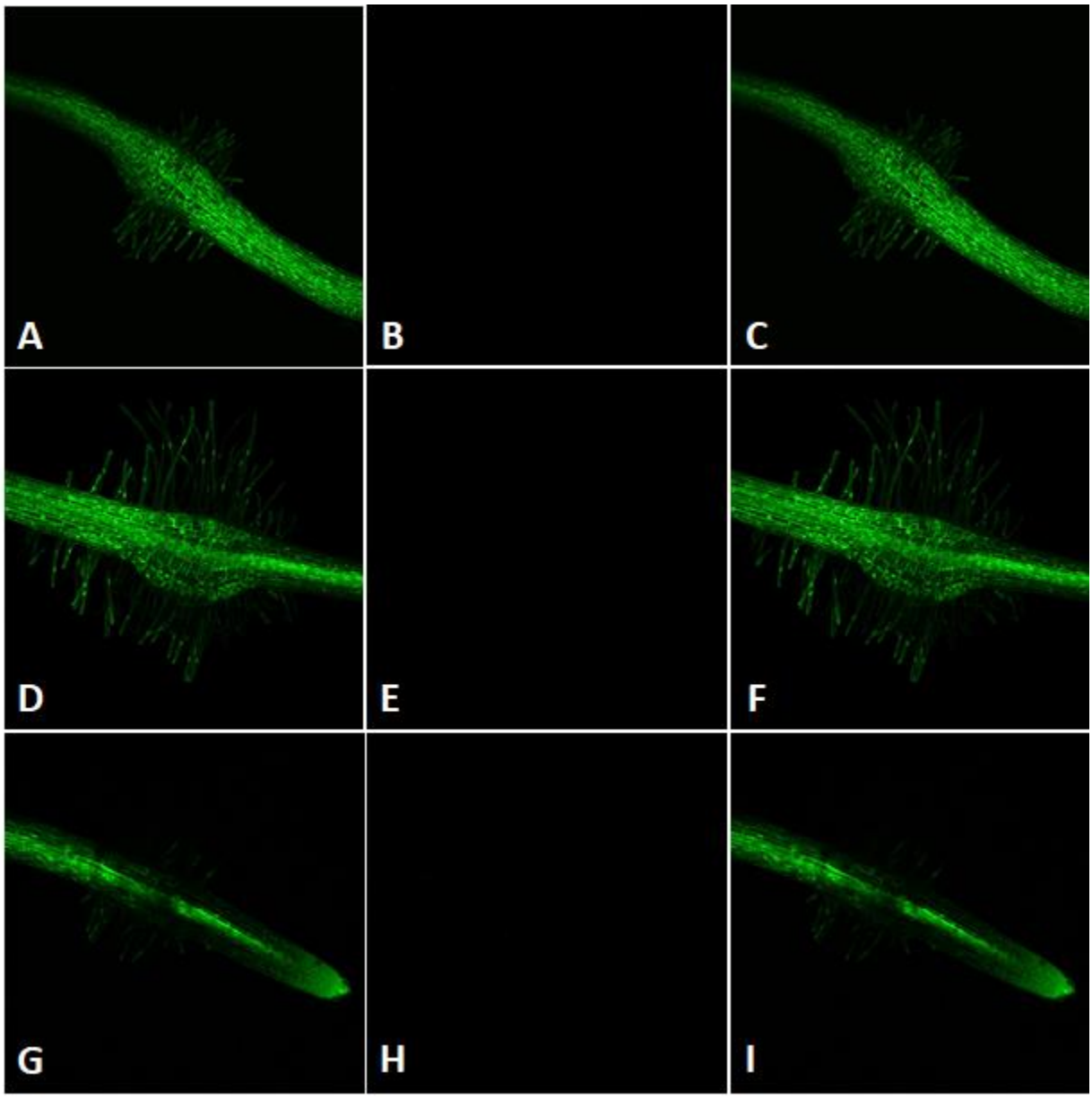


Figure S9: Confocal images of haustoria on roots transformed with pDS-MRC-Empty. Confocal images of *T. versicolor* roots transformed with pDS-MRC-Empty which did not contain a motif. Green, red, and green/red overlay images are presented for each root. Three independently transformed roots are shown. A-C: Root1, D-F: Root2, G-I: Root3.

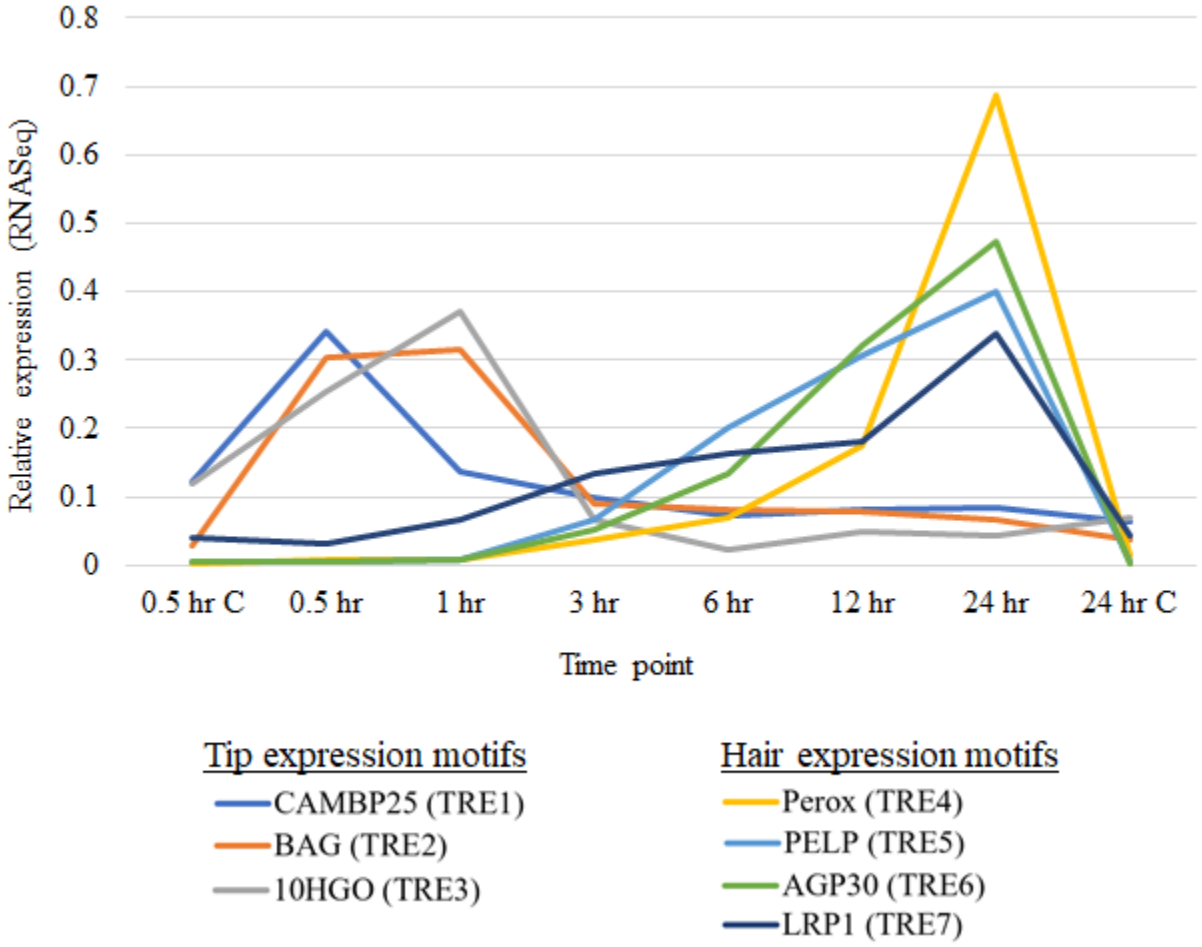


Figure S10: The relative expression patterns of the seven genes whose promoters yielded the TRE motifs. This shows the temporal expression patterns for each gene are associated with the spatial expression of their promoter motifs. The transcript levels were obtained from the RNASeq data set.

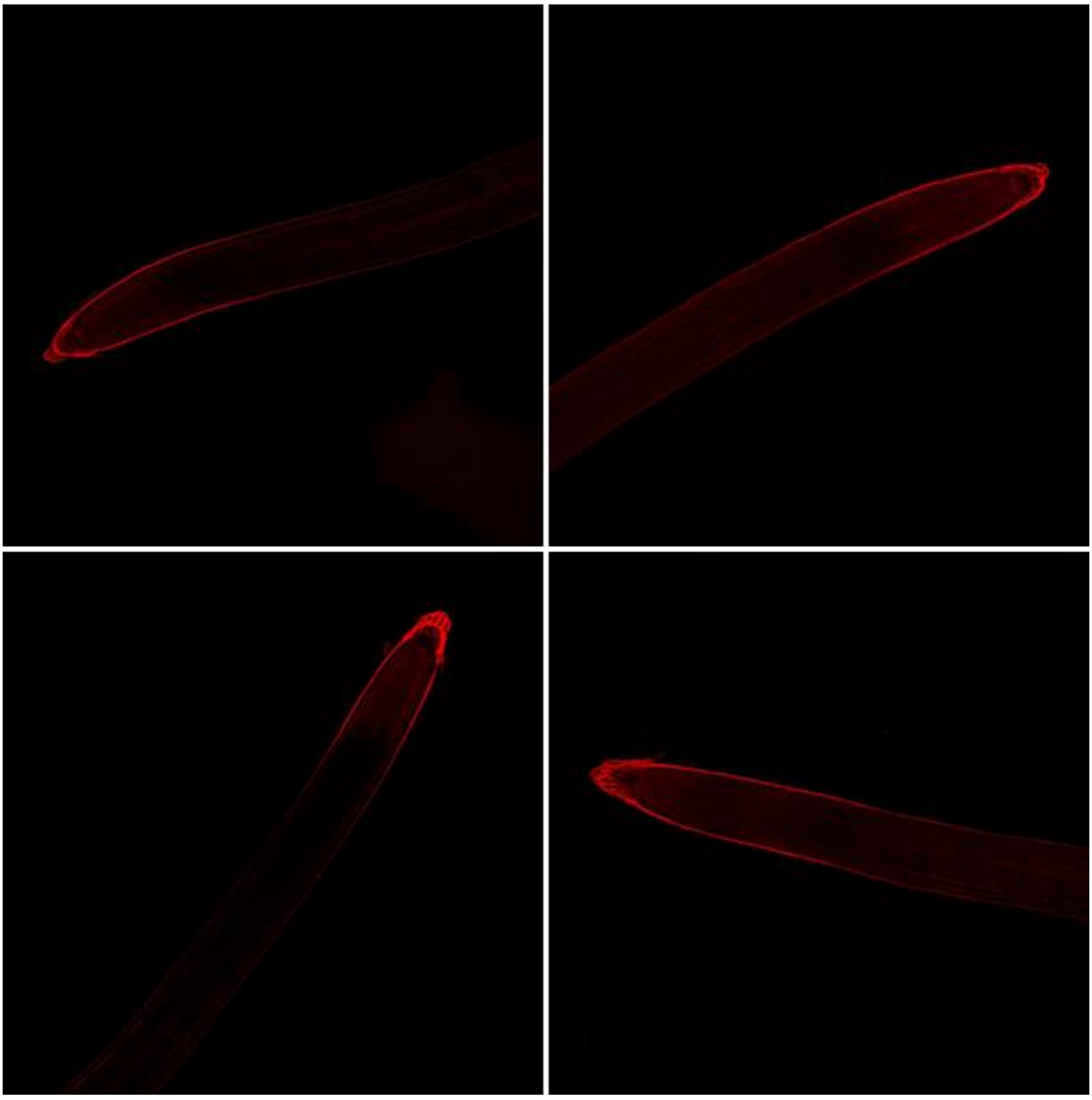


Figure S11: Confocal images of wild type *Triphysaria* roots stained with propidium iodide (PI). Four independent roots of *Triphysaria* were stained with 10 μg/ml PI in H2O.

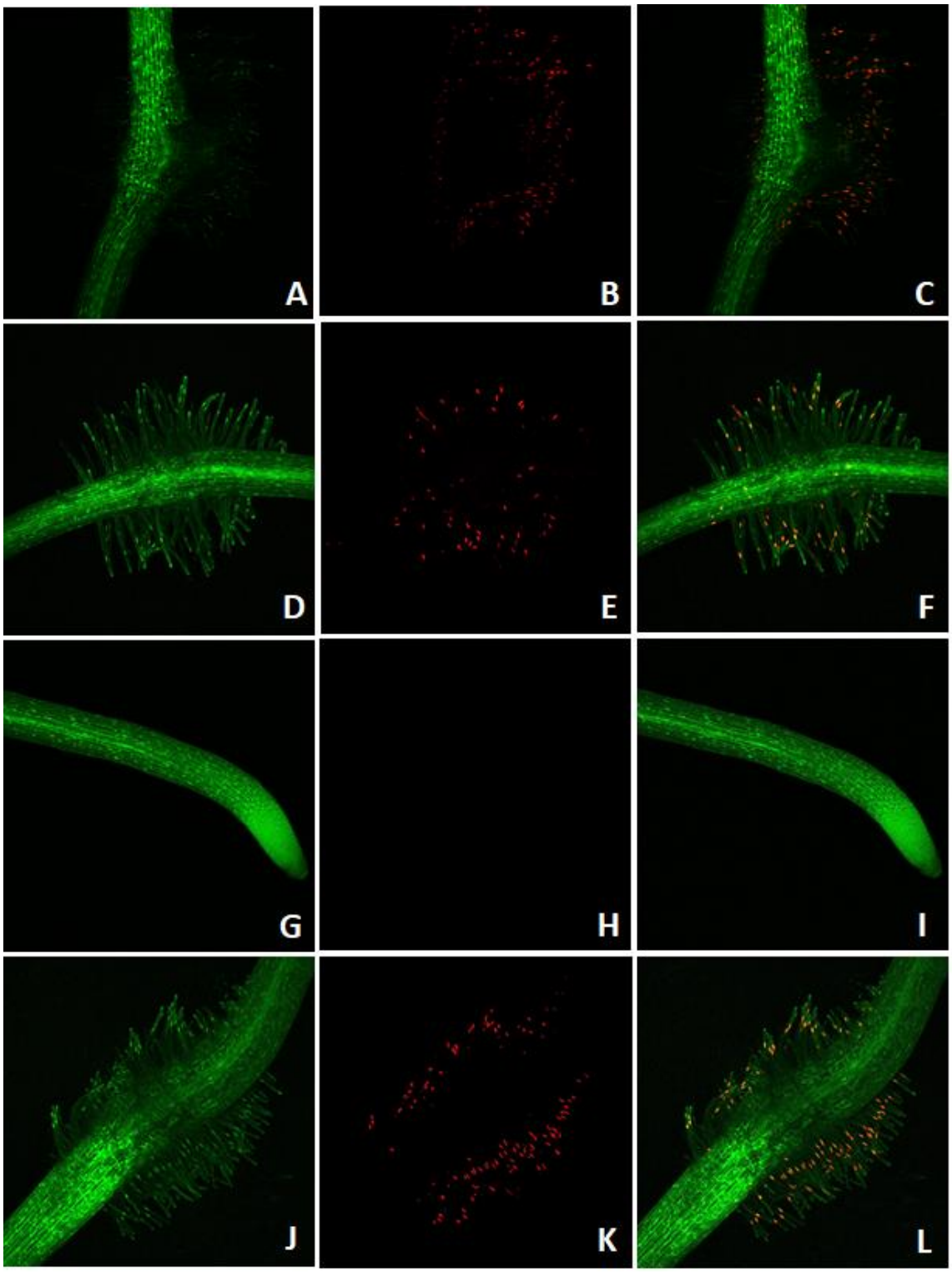


Figure S12: Confocal images of *Triphysaria* roots transformed with full promoters for PELP and AGP30. Green, red, and green/red overlay images are presented for each root. A-I: pDS-PRC-PELP, J-L: pDS-PRC-AGP30.

TF Family	AT	MG	TV	SA
Whirly	4	2	**10**	3
FAR1	26	50	**211**	64
GRF	9	6	**14**	6
NF-X1	2	1	**2**	1
RAV	7	3	**6**	7
WOX	18	10	**19**	13
GRAS	37	44	**71**	40
HSF	25	23	**35**	25
M-type_MADS	70	48	**72**	30
DBB	14	12	**17**	9
NF-YA	21	10	**14**	13
B3	77	61	**84**	48
BBR-BPC	17	9	**12**	9
CPP	9	6	**8**	8
HD-ZIP	58	56	**74**	50
MYB	168	117	**151**	98
ERF	139	116	**145**	104
ARR-B	21	17	**21**	14
WRKY	90	68	**81**	76
ARF	37	27	**32**	19
BES1	14	11	**13**	11
MYB_related	97	110	**130**	83
ZF-HD	18	23	**27**	17
SBP	30	24	**28**	21
NAC	138	105	**116**	100
G2-like	64	58	**63**	50
Dof	47	29	**31**	30
LBD	50	64	**65**	48
bHLH	225	156	**157**	129

TF Family	AT	MG	TV	SA
CO-like	22	7	**7**	6
HRT-like	2	1	**1**	0
NF-YC	21	11	**11**	10
SRS	16	6	**6**	10
STAT	4	1	**1**	1
VOZ	3	4	**4**	1
bZIP	127	87	**80**	72
GATA	41	33	**30**	37
TALE	33	26	**23**	17
Trihelix	34	37	**32**	33
C3H	66	58	**50**	41
C2H2	116	99	**82**	98
HB-other	11	16	**13**	15
E2F/DP	16	13	**10**	11
Nin-like	17	16	**12**	10
MIKC_MADS	76	56	**37**	18
AP2	30	32	**20**	19
CAMTA	10	8	**5**	7
LSD	12	5	**3**	4
EIL	6	9	**5**	5
TCP	33	29	**15**	26
GeBP	23	10	**5**	27
LFY	1	2	**1**	2
HB-PHD	3	5	**2**	2
YABBY	8	9	**3**	7
NF-YB	27	67	**18**	15
S1Fa-like	4	4	**0**	0
SAP	1	2	**0**	3
NZZ/SPL	1	0	**1**	0

Figure S13: The number of transcription factor genes belonging to each family from the genomes of *Arabidopsis thaliana* (AT), *Mimulus guttatus* (MG), *Triphysaria versicolor* (TV), and *Striga asiatica* (SA). Genes were annotated as transcription factors from specific families using the Plant Transcription Factor Database (PlantTFDB4.0).

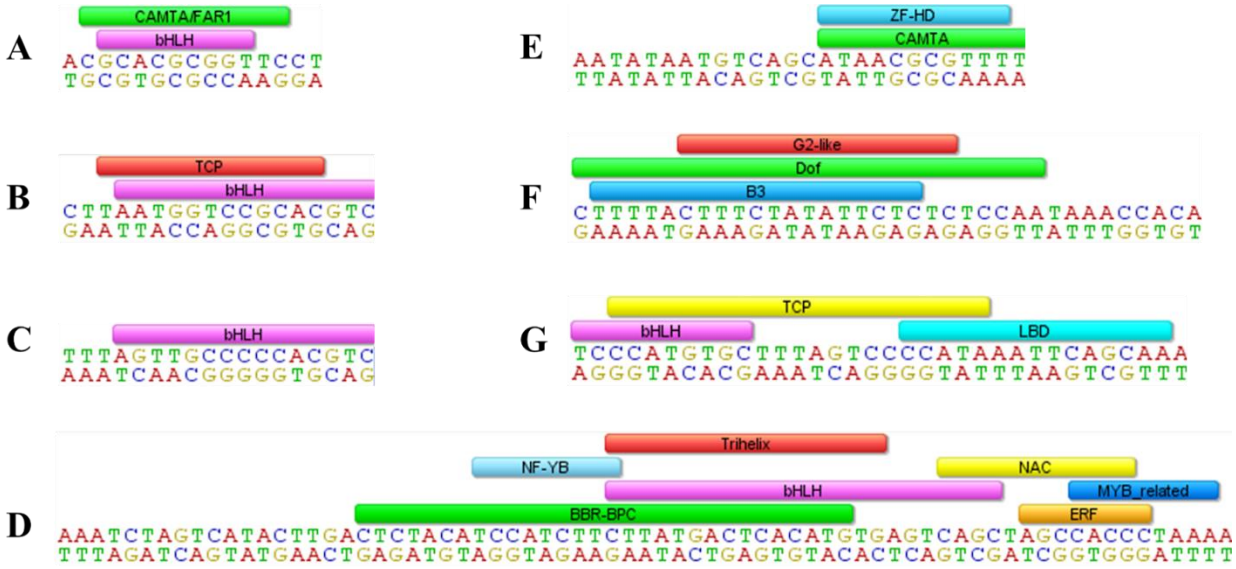


Figure S14: Predicted transcription factor binding sites in TRE sequences. A: TRE1, B: TRE5, C: TRE6, D: TRE3, E: TRE4, F: TRE7, G: TRE2. Potential sites were predicted using the "Binding Site Prediction" tool from the PlantTFDB4.0 with a p-value ≤ 1e-5.

Motif Name	Functional annotation	Gene	Method of Indentification	Promoter set enriched	Sequence evaluated	Motif copies
CAMTA (TRE1)	calmodulin-binding protein 25-like	maker-272262-exonerate_protein2genome-gene-0.0	Known site	SOM 8	ACGCACGCGGTTCCT	2X
WRKY	zinc finger protein 5	maker-694515-exonerate_protein2genome-gene-0.0	Known site	SOM 3, 7, 8, 12	TTGTTGACCTTTGCT	2X
HSF	heat shock protein 83	maker-628163-exonerate_protein2genome-gene-0.6	Known site	SOM 13,14	TTCTGGAGCCTTCT	2X
REF6	-	-	Known site	SOM Combined 1 hours	TGCTCTGTTTTT	2X
Catcat	scarecrow-like protein 14	maker-835790-exonerate_protein2genome-gene-0.0	Known site	SOM Combined 24 hours	CATCATCATCATCATCATCATCATCATC	1X
ERF3	ethylene-responsive transcription factor ERF003-like	maker-389-exonerate_est2genome-gene-0.16	De novo site	All Up 0.5 hours	GTAGGACCCAATCCCAATATTAATTGCCAGATTGAAGGCCGCAAAGCAGTGCTTACATAAAC	1X
WRKY Palindrome	probable carboxylesterase 6	maker-604109-exonerate_protein2genome-gene-0.0	De novo site	SOM Combined 0.5-3 hours	TTTTTGACTAGTCAAAAA	2X
Perox (TRE4)	lignin-forming anionic peroxidase-like	maker-683613-exonerate_protein2genome-gene-0.11	De novo site	SOM Combined 0.5-1 hours	AATATAATGTCAGCATAACGCGTTTT	2X
10HGO (TRE3)	10-hydroxygeraniol oxidoreductase	maker-691476-exonerate_protein2genome-gene-1.3	De novo site	SOM Combined 1-3 hours	AAATCTAGTCATACTTGACTCTACATCCATCTTCTTATGACTCACATGTGAGTCAGCTAGCCACCCTAAAA	1X
RAV1	AP2/ERF and B3 domain-containing transcription factor RAV1-like	maker-697030-exonerate_protein2genome-gene-0.5	De novo site	All Up 0.5 hours	ATGAATCTTCCTAGTTCCTTACCATCTTAGGGTTGACCCACACCCA	1X
HSF2	17.3 kDa class II heat shock protein-like	maker-419-exonerate_protein2genome-gene-1.4	De novo site	All Up 0.5 hours	TCCCTCCAGAAACTTCTC	2X
Cluster_5_1_6	aquaporin NIP1-1-like	maker-472966-exonerate_protein2genome-gene-0.0	De novo site	SOM Combined 1 hours	ATCAATATCCATTGATGGG	2X
GA2OX	gibberellin 2-beta-dioxygenase-like	maker-749-exonerate_est2genome-gene-0.1	Phylogenetic footprinting	-	TCTCTTCCATGCACTAGC	2X
Uncharacterized	PREDICTED: uncharacterized protein LOC105965876	maker-590154-exonerate_protein2genome-gene-0.0	Phylogenetic footprinting	-	AAAATCGTGCCCACACGACAGCCTACGACCAAGGAACACCTT	1X
Pirin-like_1	pirin-like protein	maker-696221-exonerate_protein2genome-gene-0.0	Phylogenetic footprinting	-	CTTTGACCCT	2X
Pirin-like_2	pirin-like protein	maker-696221-exonerate_protein2genome-gene-0.0	Phylogenetic footprinting	-	GATCTCAAGTAACCTCAGCCAGGCTGGCCCCAC	2X
bHLH93	transcription factor bHLH93-like	maker-652433-exonerate_protein2genome-gene-0.0	Phylogenetic footprinting	-	CAGAGTCAATGTTTCCAGAGCCGTCTGA	2X
QR2	probable NAD(P)H dehydrogenase (quinone) FQR1-like 1	maker-690263-exonerate_protein2genome-gene-0.0	Phylogenetic footprinting	-	ATTACATCTATGATGACCATAGCGTGACATCAGCA	2X
BAG (TRE2)	BAG family molecular chaperone regulator 2-like	maker-691423-exonerate_protein2genome-gene-0.50	Phylogenetic footprinting	-	TCCCATGTGCTTTAGTCCCCATAAATTCAGCAAA	2X
PELP (TRE5)	pistil-specific extensin-like protein	maker-492137-exonerate_protein2genome-gene-0.1	Phylogenetic footprinting	-	CTTAATGGTCCGCACGTC	2X
AGP30_1	non-classical arabinogalactan protein 30	maker-496909-exonerate_est2genome-gene-0.1	Phylogenetic footprinting	-	CCATCACGTT	2X
AGP30_2 (TRE6)	non-classical arabinogalactan protein 30	maker-496909-exonerate_est2genome-gene-0.1	Phylogenetic footprinting	-	TTTAGTTGCCCCCACGTC	2X
GID1B	gibberellin receptor GID1B-like	maker-472035-exonerate_protein2genome-gene-0.1	Phylogenetic footprinting	-	TTTTGTCGTCTCTCTCGTCTCCACC	2X
LRP1_1	protein LATERAL ROOT PRIMORDIUM 1-like	maker-1689-exonerate_protein2genome-gene-2.10	Phylogenetic footprinting	-	ACAACACCACCACAAAACACTCA	2X
LRP1_2 (TRE7)	protein LATERAL ROOT PRIMORDIUM 1-like	maker-1689-exonerate_protein2genome-gene-2.10	Phylogenetic footprinting	-	CTTTTACTTTCTATATTCTCTCTCCAATAAACCACA	2X
KAI2_1	probable esterase KAI2	maker-694617-exonerate_protein2genome-gene-0.9	Phylogenetic footprinting	-	ACATGTGGTTCTCGTCAT	2X
KAI2_2	probable esterase KAI2	maker-694617-exonerate_protein2genome-gene-0.9	Phylogenetic footprinting	-	ACATGTGTGATGTGTCATGTGTGGCTCTCGTCTT	2X
TriHe_1	trihelix transcription factor GT-3b-like	maker-500901-exonerate_protein2genome-gene-0.4	Phylogenetic footprinting	-	GGAAAATAGGTTTCCGCGTAGAGTGATGAGACAGACAGG	2X
TriHe_2	trihelix transcription factor GT-3b-like	maker-500901-exonerate_protein2genome-gene-0.4	Phylogenetic footprinting	-	GTCAGCATGACGTCGTATGACACAATATGAACAAATGTGGATCGCCCAAAAAAAAAGAA	1X
ERF109_1	ethylene-responsive transcription factor ERF109-like	maker-690378-exonerate_protein2genome-gene-0.17	Phylogenetic footprinting	-	ATAACTCCCCATTTTGACTTTCTCATTTC	2X
ERF109_2	ethylene-responsive transcription factor ERF109-like	maker-690378-exonerate_protein2genome-gene-0.17	Phylogenetic footprinting	-	ACTCCGTGTAAAACTAGTGTGTAAGCAGCGCAAAGACTGCTGAGCTGGCAACAC	1X
ERF109_3	ethylene-responsive transcription factor ERF109-like	maker-690378-exonerate_protein2genome-gene-0.17	Phylogenetic footprinting	-	TCCCATGTTCCTAAATTCCCCACGTGCCACAACC	2X
Unchar2	3beta-hydroxysteroid-dehydrogenase/ decarboxylaseisoform 2	maker-1049-exonerate_protein2genome-gene-0.8	Phylogenetic footprinting	-	TCAAAACATGTGAAATCACCAAACACGACTATTTGTCTGCCAGAATTTACTAATTCCAG	1X
Pectate_Ly	putative pectate lyase 2	maker-697087-exonerate_protein2genome-gene-0.11	Phylogenetic footprinting	-	CAATAGTCAATTATAACCTTCACCTCGTGCGATCACGTCTATAATC	1X

Supplemental Table 1: Details on the motifs identified and evaluated for regulatory function in this study.

Supplemental Methods 1:
High Molecular Weight Genomic DNA Isolation –
Optimized for Triphysaria versicolor

Reagents:

- DNA extraction buffer: 100 mM Tris (pH 8), 20 mM EDTA (pH 8), 2% CTAB, 1 M NaCl
- RNase A (10 mg/ml stock)
- Phenol:Chloroform:Isoamyl Alcohol (25:24:1)
- Chloroform:Isoamyl Alcohol (24:1)
- 100% ethanol
- 70% ethanol
- EB or TE

Notes:

- Use wide-bore tips for all pipetting steps.
- Take care to mix gently at all times.
- Centrifuge at 20 C.

1. Ground ~100 mg of liquid nitrogen frozen tissue for 1 minute with 2 steel balls in a rounded bottom 2 ml microcentrifuge tube using a paint shaker. Keep everything frozen throughout this process with a chilled microcentrifuge tube block.
2. Preheat 1 ml of extraction buffer and add 10 ul RNase A (10 mg/ml stock).
3. Add 1 ml preheated extraction buffer to 2 ml tube, gently mix and get all powder into the liquid. Incubate at 65°C for 10 minutes. Mix tubes a few times during incubation period by gentle inverting. Remove the steel balls during this stage as well.
4. Spin the tube at 13k g for 30 seconds.
5. Carefully remove the aqueous layer without disturbing the Pellet. Place into a clean 2 mL microcentrifuge tube.
6. Add 1 mL of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) to the aqueous layer.
7. Invert the tube gently 25 times to mix.
8. Spin the tube at 13k g for 10 minutes.
9. Carefully remove the aqueous layer, do not disturb the interface. Place into a clean 2 mL microcentrifuge tube.
10. Add an equal volume of Chloroform:Isoamyl Alcohol (24:1) to the aqueous layer.
11. Invert tube gently 25 times to mix.
12. Spin the tube at 13k g for 10 minutes.
13. Carefully remove the aqueous layer, do not disturb the interface. Place into a clean 2 mL microcentrifuge tube.
14. Add 0.3X volume of ethanol (100%) to the aqueous layer. This high-salt, low-ethanol mixture precipitates the excess polysaccharides while gDNA remains in the solution.

15. Invert tube gently 20 times to mix.
16. Spin tube at 13k g for 15 minutes.
17. Carefully remove the supernatant without disturbing the polysaccharide pellet. Place supernatant into a clean 2 mL microcentrifuge tube.
18. Add 1.7X volume of ethanol (100%) to the supernatant. The gDNA can be seen as a white stringy precipitate forming at the interface.
19. Spool with hooked pastuer pipette by mixing interface until the layers combine. gDNA should be a visible precipitate around the hook. Invert tube with unspooled gDNA multiple times to finish the mixing.
Note: I recover the leftover unspooled gDNA for other future purposes but if HMW gDNA is all you are interested in then you can only proceed forward in the next steps with the spooled gDNA.
20. Dunk hook into a tube with 1 ml 70% ethanol a few times and let sit for 10 minutes.
21. Rinse hook with 1 ml 70% ethanol.
22. Dunk hook into a new tube with 1 ml 70% ethanol a few times and let sit for 10 minutes.
23. Rinse hook into new tube with 70% ethanol and wash/scrape spooled DNA off hook.
Spooled DNA:
24. Spin tube with spooled DNA at 2.5k g for 1 minute.
25. Pour off supernatant and save in a clean tube.
26. Add 1 mL of 70% ethanol and let sit for 5 minutes.
27. Spin the tube at 2.5k g for 2.5 minutes.
28. Pour off supernatant and save in a clean tube.
29. Add 1 mL of 70% ethanol.
30. Spin the tube at 13k g for 3 minutes. Proceed forward at step 38.
Left-over gDNA (recover is desired for other purposes):
31. Spin tube with leftover unspooled at 13k g for 5 minute.
32. Pour off supernatant and save in a clean tube.
33. Add 1 mL of 70% ethanol.
34. Spin the tube at 13k g for 2.5 minutes.
35. Pour off supernatant and save in a clean tube.
36. Add 1 mL of 70% ethanol.
37. Spin the tube at 13k g for 2.5 minutes.
Continue for all gDNA:
38. Carefully remove the supernatant; do not disturb the pellet. Save supernatant in a clean tube. Quick spin to gather the residual ethanol at the bottom of the tube and carefully remove with a pipette tip.
39. Let pellet air dry for 5 min at room temperature, taking care not to over dry.
40. Resuspend the gDNA pellets in 50 µL EB. Incubate at 4ºC for at least a few hours or overnight to resuspend. Store at 4ºC for use within one week, or store at -80ºC for long-term storage.

REFERENCES:

1. Brooker RW. Plant–plant interactions and environmental change. New Phytologist. 2006. pp. 271–284. doi:10.1111/j.1469-8137.2006.01752.x

2. Depuydt S. Arguments for and against self and non-self root recognition in plants. Front Plant Sci. 2014;5: 614.

3. Wang N, Kong C, Wang P, Meiners SJ. Root exudate signals in plant–plant interactions. Plant, Cell & Environment. 2021. pp. 1044–1058. doi:10.1111/pce.13892

4. Gorelick R, Marler TE. Kin recognition by roots occurs in cycads and probably in conifers. Commun Integr Biol. 2014;7: e28009.

5. Bandaranayake PCG, Yoder JI. Haustorium Initiation and Early Development. Parasitic Orobanchaceae. 2013. pp. 61–74.

6. Fuller AW, Young P, Pierce BD, Kitson-Finuff J, Jain P, Schneider K, et al. Redox-mediated quorum sensing in plants. PLoS One. 2017;12: e0182655.

7. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2011;13: 59–69.

8. Korkuc P, Schippers JHM, Walther D. Characterization and identification of cis-regulatory elements in Arabidopsis based on single-nucleotide polymorphism information. Plant Physiol. 2014;164: 181–200.

9. Lin Z, Wu W-S, Liang H, Woo Y, Li W-H. The spatial distribution of cis regulatory elements in yeast promoters and its implications for transcriptional regulation. BMC Genomics. 2010;11: 581.

10. Guo Y, Curtis Jamison D. The distribution of SNPs in human gene regulatory regions. BMC Genomics. 2005. doi:10.1186/1471-2164-6-140

11. Turco G, Schnable JC, Pedersen B, Freeling M. Automated conserved non-coding sequence (CNS) discovery reveals differences in gene content and promoter evolution among grasses. Front Plant Sci. 2013;4: 170.

12. Reineke AR, Bornberg-Bauer E, Gu J. Evolutionary divergence and limits of conserved non-coding sequence detection in plant genomes. Nucleic Acids Res. 2011;39: 6029–6043.

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

14. Hoagland DR, Arnon DI. The Water-culture Method for Growing Plants Without Soil. D. R. Hoagland and D. I. Arnon. 1938.

15. Jamison DS, Yoder JI. Heritable variation in quinone-induced haustorium development in the parasitic plant Triphysaria. Plant Physiol. 2001;125: 1870–1879.

16. Wang Y, Murdock M, Lai SWT, Steele DB, Yoder JI. Kin Recognition in the Parasitic Plant Triphysaria versicolor Is Mediated Through Root Exudates. Frontiers in Plant Science. 2020. doi:10.3389/fpls.2020.560682

17. Andersen CL, Jensen JL, Ørntoft TF. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. Cancer Research. 2004. pp. 5245–5250. doi:10.1158/0008-5472.can-04-0496

18. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. Genome Res. 2017;27: 757–767.

19. Seppey M, Manni M, Zdobnov EM. BUSCO: Assessing Genome Assembly and Annotation Completeness. Methods Mol Biol. 2019;1962: 227–245.

20. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12: 491.

21. Wehrens R. Self-Organizing Maps. Chemometrics with R. 2011. pp. 67–78. doi:10.1007/978-3-642-17841-2_5

22. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media; 2009.

23. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43: W39–49.

24. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research. 2017. pp. D1040–D1045. doi:10.1093/nar/gkw982

25. Nguyen NTT, Contreras-Moreira B, Castro-Mondragon JA, Santana-Garcia W, Ossio R, Robles-Espinoza CD, et al. RSAT 2018: regulatory sequence analysis tools 20th anniversary. Nucleic Acids Res. 2018;46: W209–W214.

26. Liao C-Y, Smet W, Brunoud G, Yoshida S, Vernoux T, Weijers D. Reporters for sensitive and quantitative measurement of auxin response. Nat Methods. 2015;12: 207–10, 2 p following 210.

27. Bajar BT, Wang ES, Lam AJ, Kim BB, Jacobs CL, Howe ES, et al. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. Sci Rep. 2016;6: 20889.

28. Fernández-Aparicio M, Rubiales D, Bandaranayake PC, Yoder JI, Westwood JH. Transformation and regeneration of the holoparasitic plant Phelipanche aegyptiaca. Plant Methods. 2011;7: 36.

29. Hehl R. Plant Synthetic Promoters: Methods and Protocols. Humana; 2018.

30. Bandaranayake PCG, Yoder JI. Factors affecting the efficiency of Rhizobium rhizogenes root transformation of the root parasitic plant Triphysaria versicolor and its host Arabidopsis thaliana. Plant Methods. 2018. doi:10.1186/s13007-018-0327-2

31. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9: 671–675.

32. Tank DC, Olmstead RG. From annuals to perennials: phylogeny of subtribe Castillejinae (Orobanchaceae). Am J Bot. 2008;95: 608–625.

33. Krichevsky A, Zaltsman A, Kozlovsky SV, Tian G-W, Citovsky V. Regulation of root elongation by histone acetylation in Arabidopsis. J Mol Biol. 2009;385: 45–50.

34. Singh S, Yadav S, Singh A, Mahima M, Singh A, Gautam V, et al. Auxin signaling modulates LATERAL ROOT PRIMORDIUM1 (LRP1) expression during lateral root development in Arabidopsis. Plant J. 2020;101: 87–100.

35. Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. New Phytol. 2017;214: 1338–1354.

36. Pekárová B, Szmitkowska A, Dopitová R, Degtjarik O, Žídek L, Hejátko J. Structural Aspects of Multistep Phosphorelay-Mediated Signaling in Plants. Mol Plant. 2016;9: 71–85.

37. Ranty B, Aldon D, Galaud J-P. Plant calmodulins and calmodulin-related proteins: multifaceted relays to decode calcium signals. Plant Signal Behav. 2006;1: 96–104.

38. Laohavisit A, Wakatake T, Ishihama N, Mulvey H, Takizawa K, Suzuki T, et al. Quinone perception in plants via leucine-rich-repeat receptor-like kinases. Nature. 2020;587: 92–97.

39. Arif M, Li Z, Luo Q, Li L, Shen Y, Men S. The BAG2 and BAG6 Genes Are Involved in Multiple Abiotic Stress Tolerances in Arabidopsis Thaliana. International Journal of Molecular Sciences. 2021. p. 5856. doi:10.3390/ijms22115856

40. Sandholu AS, Mujawar SP, Ramakrishnan K, Thulasiram HV, Kulkarni K. Structural studies on 10-hydroxygeraniol dehydrogenase: A novel linear substrate-specific dehydrogenase from Catharanthus roseus. Proteins. 2020;88: 1197–1206.

41. Yoshida K, Kaothien P, Matsui T, Kawaoka A, Shinmyo A. Molecular biology and application of plant peroxidase genes. Applied Microbiology and Biotechnology. 2003. pp. 665–670. doi:10.1007/s00253-002-1157-7

42. Mishler-Elmore JW, Zhou Y, Sukul A, Oblak M, Tan L, Faik A, et al. Extensins: Self-Assembly, Crosslinking, and the Role of Peroxidases. Front Plant Sci. 2021;12: 664738.

43. Hwang Y, Choi H-S, Cho H-M, Cho H-T. Tracheophytes Contain Conserved Orthologs of a Basic Helix-Loop-Helix Transcription Factor That Modulate Genes. Plant Cell. 2017;29: 39–53.

CHAPTER 3:

NATURAL VARIATION IN THE TVQR1 PROMOTER: TRANSCRIPTIONAL REGULATION, MITES, AND HAUSTORIUM DEVELOPMENT

ABSTRACT:

Biochemical and functional studies have shown the TvQR1 gene from *Triphysaria versicolor* encodes an enzyme that catalyzes the reduction of DMBQ to the semiquinone form that activates the haustorium development pathway. In *Triphysaria*, TvQR1 is transcriptionally up-regulated immediately upon exposure to host root exudates or purified DMBQ. In contrast, the endogenous QR1 genes in *Arabidopsis* and *Mimulus* were not upregulated in response to DMBQ. The regulation of TvQR1 was recapitulated using a 652 bp promoter sequence and visualized in transgenic roots with a red fluorescent protein. When the TvQR1 promoter reporter was transformed into roots of the non-parasites, the fluorescent reporter was upregulated following DMBQ exposure but the endogenous genes were not. Together, this suggests *Triphysaria* has co-opted the function of TvQR1 for host response and haustorium formation through the use of conserved cis-elements.

INTRODUCTION:

Many of the genes associated with parasitism are present in other non-parasitic plants [1–5]. Non-parasitic plants do not regulate at least some of these genes the same way in response to host exudates [6]. This implies that parasitic plants have co-opted autotrophic plant genes by changing their regulation during the evolution of haustorium development. The co-option of genes and pathways for new functions is thought to be the major source for establishing new traits during evolution[7]. To date, little is known for how parasitic plants regulate their haustoria associated genes or how new expression patterns may have been co-opted for parasitic functions through changes in cis-regulatory sequences.

The gene Quinone Reductase 1 (TvQR1) is one of the earliest genes involved in haustorium initiation in *Triphysaria versicolor* [1]. TvQR1 is an NADPH-dependent oxidoreductase that reduces quinones by a single-electron to create free radical semiquinones. TvQR1 was discovered from cDNAs isolated from *T. versicolor* roots treated with DMBQ and is likely a single copy gene [6]. TvQR1 is thought to trigger haustorium formation through a signaling mechanism initiated by redox cycling of host derived phenolic compounds, such as 2,6-dimethoxybenzoquinone (DMBQ), between quinone and hydroquinone states. The transcriptional expression of TvQR1 is up-regulated as early as 30 minutes after exposure with DMBQ or host root exudates. This change in TvQR1 expression does not require the translation of new proteins meaning it is a primary response gene to the presence of nearby hosts[1].

Triphysaria versicolor is a highly polymorphic, obligate outcrossing plant. When the transcript sequence of TvQR1 was compared between *Triphysaria versicolor* plants, it was shown to contain a higher level of variability than another gene necessary for haustorium development, Pirin [8]. Our lab has also observed variation in the size and sequence content of

the promoter region for TvQR1. This variation creates the opportunity for new expression patterns or functions to arise.

A common source of genomic variation with implications for gene expression and genome evolution are transposable elements. Miniature inverted-repeat transposable elements (MITEs) are a group with high copy number and enrichment near genes[9,10]. MITEs are widespread in plant and animal genomes and can account for up to 8-10% of the total genome in plants[11,12]. MITEs are small (usually < 800 bp) deletion derivatives of autonomous DNA transposons and lack their own transposase [13,14]. MITEs retain common features of autonomous DNA transposons including terminal inverted-repeats (TIR) flanked by a target site duplication (TSD). Because MITEs lack a transposase and origin of replication, their transposition requires a transposase derived from an autonomous element[15].

The polymorphisms induced by the insertion or excision of MITEs have been shown to affect the expression of nearby genes in a variety of ways. Typically, the insertion of a MITE near a gene lowers its expression and is often attributed to an interruption of a sequence at the insertion site or the introduction of new regulatory sequences within the MITE itself [16–18]. The change in distance between promoter sequences and the gene they regulate caused by the MITE insertion could also affect the expression of nearby genes by disrupting the interactions of regulatory proteins. MITEs can also trigger the silencing of genes through the generation of sRNAs [17,19]. These sRNAs can travel to other sites and cause the silencing of sequences similar to the insertion site [20]. The genetic and epigenetic changes induced by MITEs are thought to play an important role in establishing new patterns of gene regulation and genome evolution overall [17,19,21].

METHODS:

Plant material:

Methods regarding the seeds and plants for *Triphysaria versicolor, Mimulus guttatus,* and *Arabidopsis thaliana* were performed as described in Chapter 2.

Genotyping TvQR1 promoter:

Triphysaria seedlings were grown on square plates as described in chapter 2. To genotype the TvQR1 promoter alleles of individual seedlings, a piece of leaf was taken from each plant and placed it in a 2 mL tube with multiple silica beads. The samples were then placed in a liquid nitrogen cooled plastic block and ground for 1 minute using a paint shaker. A DNA extraction buffer (100mM Tris (pH8), 20 mM EDTA (pH8), 2% CTAB, 1M NaCl, 1% 2-Mercaptoethanol) was then added to the ground leaf sample and vortexed. This solution was used immediately for the PCR with 0.25 ul being added to each reaction. The high-fidelity DNA polymerase Q5 (NEB #M0491) was used for the PCR with an annealing temperature of 66°C and an extension time of 2:30. Completed PCRs were run on a 1% agarose gel (w/v) and evaluated by size. Primers at the 5' end of the promoter and within the coding sequence of TvQR1 exon 3 were used to genotype the promoter alleles. A clear size difference allowed for a clear differentiation between most alleles. Two of the most common alleles were cloned, sequenced, and names TvQR1-P1 and TvQR1-P2. There were some additional and less frequent TvQR1 alleles that were near the size of the TvQR1-P2. To confirm the presence of the TvQR1-P2 allele, the PCR was repeated with a forward primer in the MITE sequence. A total of 855 *Triphysaria* plants were genotyped for the TvQR1 promoter. Plants genotyped as either P1/- or P2/- means the P1 and P2 were the only observed alleles on the genotyping agarose gel for those

plants. The P1/- genotype can then represent P1/P1 or P1/other null allele(not detected in the PCR).

Comparing promoter sequences:

The promoter fragments obtained through PCR were sequenced, aligned and compared in Geneious to determine their level of similarity. A dot-plot was used to study the structure of the promoter sequences and facilitated the discovery of the MITE in TvQR1-P2 by identifying its inverted repeat sequences.

Expression vectors:

The promoter reporting construct pDS-PRC is almost identical to the pDS-MRC construct from Chapter 2 except the motif expression cassette (between the unique restriction sites BspEI and XbaI) was replaced with the TvQR1-P1 promoter. For this study, the promoter was defined as the 652 bp sequence beginning after the start codon of TvQR1 and ending at the coding sequence of the next gene upstream, a PPR gene. To clone the TvQR1 promoter into pDS-PRC, the vector was linearized using the enzymes BspEI and XbaI. The TvQR1 promoter was amplified using PCR with primers that added part of the vector insertion site to the ends of the product. Gibson assembly (NEB #E2611) was then used to clone the TvQR1 promoter PCR with the linearized pDS-PRC.

The artificial promoter fragments were created using overlapping PCR. First the sequence content of the MITE from TvQR1-P2 was randomized and synthesized to create the scrambled version of the MITE. The specific scrambled sequence was chosen for having few predicted transcription factor binding sites (PlantTFBD4.0) and minimal secondary structure (RNAFold server - ViennaRNA Web Services). Each individual piece of the artificial promoter construct was PCR amplified using printers that added part of the planned adjoining fragment to its ends.

These products with overlapping DNA sequences were added to a new PCR with primers flanking the entire planned region. These flanking primers also added pieces of the vector insertion site mentioned above to facilitate a Gibson assembly of the completed promoter into the pDS-PRC vector.

Plant transformation:

As described in Chapter 2.

Fluorescent microscopy:

Confocal microscopy was used to assess the location and relative level of fluorescence in transgenic *Triphysaria* roots before and after DMBQ treatment. When referring to transgenic root images, the expression of tdTomato was estimated based on the qualitative visual level of red fluorescence. The remaining procedures and details are as described in Chapter 2.

Image processing and analyses:

As described in Chapter 2.

RNA isolation and cDNA synthesis:

As described in Chapter 2.

Quantitative PCR (qPCR):

Primers for endogenous versions of QR1 for *Arabidopsis* and *Mimulus* were designed using the closest orthologous sequence identified using NCBI blastn. Four sets of potential reference gene primers were obtained from previous works for both non-parasites. All primers were evaluated as gene specific and efficient using the same methods as Chapter 2. Two sets of reference gene primers were chosen for each plant: ACT2 and PDF2 for *Arabidopsis* and EF1A and GAP for *Mimulus.* The geometric mean of these two genes was used to normalize the data and correct for different amounts of cDNA.

The DMBQ induced change in TvQR1 transcript levels were evaluated in ten *Triphysaria* plants for each genotype. Three to twelve root tips were collected from each plant before and 1 hour after DMBQ exposure. The remaining qPCR procedures were performed as described in Chapter 2.

Prehaustoria induction and evaluation:

Prehaustoria were induced in *Triphysaria* the same as described in Chapter 2 except that wild type and transgenic seedlings were treated with 10 and 30 uM DMBQ respectively. Wild-type plants were treated with a lower level of DMBQ to better evaluate the differences between plants while the transformed plants received a higher dose to maximize the expression of our transgenic constructs. Prehaustoria were evaluated 24 hours after exposure to DMBQ using a dissecting microscope. The formation of prehausoria was judged based on the presence of localized hair proliferation and swelling just behind the tip of the root. If a root was lacking any additional hair development or swelling it was said to have not formed a prehaustoria. The proportion of roots forming haustoria was determined by the ratio of roots on a single plant that formed prehaustoria compared to the total number of root tips. The proportion of prehaustoria formation was calculated by inducing 5-40 roots per *Triphysaria* plant (16 root tips on average) with DMBQ.

Graphs:

All figures were generated using Microsoft Excel, PowerPoint or in R using the package ggplot2 [22]. In the box plots, the boxes represent the upper and lower quartiles, the whiskers represent the minimum and maximum data points (excluding outliers), and the midline in the box is the median.

Statistical analysis:

To determine if groups were significantly different, two types of tests were performed depending on the number of groups. When evaluating the difference between just two groups, a two-tailed Welch's t-test was performed. A significant difference was obtained if the resulting p-value ≤ 0.05. If more than two groups were compared, then our data was fit to a linear model using the lm() function in R. An analysis of variance (ANOVA) was then performed on our model using aov(). A Tukey HSD test was performed on this ANOVA using a confidence level of 95% to determine if groups were significantly different. Groups determined to be significantly different using this method were indicated with different letters above each group.

Genomic MITE characterization:

The location of MITEs in the Triphysaria genome (discussed in Chapter 2) were annotated using the program MITE-Hunter [23]. Additional repetitive sequences were predicted de novo using RepeateModeler [24]. The repetitive sequences were compared with a BLAST database of plant proteins and the sequences with significant hits, along with 50bp flanking sequences, were excluded.

The locations of MITEs similar to the TvQR1-P2 MITE were selected and refined based on having an SW score > 1500. The window function of the program BEDtools was used to identify genes with TvQR1-P2 like MITEs within 500, 1000, 2000, 5000, or 10000 base pairs upstream from their start codon [25]. The mean expression level for genes with MITEs at each of these distances were obtained from the RNASeq data set in Chapter 2. The values graphed in Figure 9 represent the mean of the mean expression levels for each group of genes. The 95% confidence interval for each group of means was determined using the summarySE function in R.

RESULTS:

Transcriptional response of TvQR1 to host exudate

The expression profile of TvQR1 across prehaustorium development in *Triphysaria* can be seen in Figure 1A. The transcript level of TvQR1 immediately increases after exposure to *Arabidopsis* root exudate, peaks at 1 hour, and returns to a basal level of expression by 6 hours before any noticeable morphological changes occur. This expression profile matches previously shown expression changes in TvQR1 induced with DMBQ in *Triphysaria* as determined by qPCR [1].

The fold change in QR1 transcript levels before and 1 hour after exposure to *Arabidopsis* root exudate were determined using qPCR in nine species of Orobanchaceae, Chapter 2 Figure 3. Only *Triphysaria versicolor* exhibited a robust upregulation of QR1 in response to host exudate. Notably, the other two parasites in the genus *Triphysaria* had just a small change in QR1 expression which suggests the strong response of TvQR1 to HIFs is fairly unique to *T. versicolor.*

TvQR1 promoter alleles

For this study, the promoter region was defined as the region immediately upstream of the TvQR1 start codon until the coding sequence of the next upstream gene, a pentapeptide repeat protein (TvPPR). Multiple alleles for the TvQR1 promoter region with clear size differences were identified from our single population of *Triphysaria versicolor* seeds. These alleles were identified using primers in the coding sequences of the TvQR1 and TvPPR genes flanking this promoter region. These primers reliably amplified 1-2 promoter alleles from most *Triphysaria* plants.

Two of the most frequent promoter alleles were cloned, sequenced, and named TvQR1-P1 and TvQR1-P2 (Figure 2). When the sequences were analyzed for repeated regions using a dot plot, a set of inverted repeats was discovered in TvQR1-P2 not found in TvQR1-P1. Upon further analysis, the size, target site duplication, and inverted repeat sequences matched those of a Stowaway-like MITE. A 290 bp region at the beginning of the promoter is starkly more conserved than the rest of the sequence which showed very little similarity. In the TvQR1-P2, the MITE occurs roughly in the middle of this conserved region, 152 bps upstream of the coding sequence.

In terms of allele frequencies, 51.2% of the 855 genotyped *Triphysaria* plants had at least one of these two alleles. Of this group, 41.5% contained only the P1 allele, 5.4% contained P2 and 4.3% contained both P1 and P2. The other 48.8% of plants contained either a different TvQR1 promoter allele (23.6%) or were lacking a promoter PCR product entirely (25.2%). These additional alleles were not characterized in this study.

Spatial regulation of TvQR1-P1 in Triphysaria versicolor

Given the established role of TvQR1 in haustorium development, I investigated the spatio-temporal regulation of the most common promoter allele TvQR1-P1 in response to the host derived molecule DMBQ. This 652 bp promoter sequence was used to drive the expression of the red fluorescent protein tdTomato to visually report the transcriptional regulation of the TvQR1-P1, Figure 3A.

Before DMBQ exposure, tdTomato was predominantly localized to the columella, lateral root cap, epidermis, and center of the root in mature regions, Figure 3B1-6. This central expression is likely either endodermis, pericycle, or a vascular associated tissue.

In the DMBQ induced prehaustoria, tdTomato was mostly present in the epidermis, including the haustorial hairs, Figure 3B7-9. The prehaustorium was younger tissue than the mature root section which is why the center of the prehaustorium does not show tdTomato expression.

To get a better sense for the temporal change in tdTomato levels regulated by TvQR1-P1, dissecting microscope images were taken every hour after DMBQ exposure to create time lapse images of prehaustorium development, Figure S1. Here the expression of tdTomato is seen almost exclusively in the root tip, before DMBQ treatment (0 hrs). At 6 hrs post DMBQ treatment, the region where the prehaustorium is forming starts to visibly show tdTomoto expression. Soon after, expression in the epidermis begins to visibly increase as well. The levels of tdTomato continually increases across this 24 hr window with particularly high levels at the tip, prehaustorium, and epidermis overall.

This delay in observable fluorescence after the increase in transcript level seen in Figure 1 is expected given the time it takes for translation or other intermediate processes to occur. Due to this delay, qPCR is a more effective means at assessing the temporal characteristics of transcriptional regulation than the change in visual fluorescence.

Temporal and spatial regulation of TvQR1-P1 in *Arabidopsis thaliana* and *Mimulus guttatus*.

TvQR1 is upregulated in *Triphysaria* roots after exposure to host root exudates leading to the formation of haustoria. I wanted to determine whether QR1 orthologs in plants that do not form haustoria are similarly upregulated by host root exudates. I first evaluated transcriptional regulation of the QR1 genes endogenous to two autotrophic plants *Mimulus guttatus* and

Arabidopsis thaliana. *Mimulus* and *Triphysaria* are relatively closely related with both being in the order Lamiales. *Arabidopsis* is less related and is estimated to have shared a common ancestor with *Triphysaria* than *Mimulus* roughly 128 million years ago[26].

The transcript levels for the QR1 gene endogenous to each species was evaluated using qPCR. Figure 4A shows the fold change in transcript levels after 1 hour of DMBQ exposure. There is a rapid upregulation in the number of TvQR1 transcripts in *Triphysaria*. However, there was no change in the number of AtQR1 or MgQR1 transcripts after DMBQ exposure. This shows the endogenous gene in *Triphysaria* is regulated differently than the endogenous genes in *Arabidopsis* or *Mimulus*, Figure 4A.

I similarly measured the levels of tdTomato transcripts in *Arabidopsis* or *Mimulus* roots transgenic for the TvQR1-P1 reporter, Figure 4B. The tdTomato transcript driven by the TvQR1-P1 were upregulated after DMBQ exposure while the endogenous AtQR1 and MgQR1 genes were not, Figure 4B.

The spatial regulation of TvQR1-P1 in *Triphysaria* was remarkably similar when transformed into *Mimulus* and *Arabidopsis*, Figure 5. In all three species, the TvQR1-P1 was strongly active in the columella, mature epidermis including root hairs, and the center of the root. To a lesser extent, TvQR1-P1 was also active in young epidermis and the lateral root cap of *Arabidopsis*. The clearest difference between the autotrophic plants was the strong level of TvQR1-P1 expression in the lateral root cap of *Mimulus*, Figure 5H.

Transcriptional expression of TvQR1 in plants genotyped for the promoter

Quantitative PCR (qPCR) was performed on genotyped root tips before and one hour after exposure to DMBQ, Figure 6. Before DMBQ treatment, the transcript level of TvQR1 was

low for all genotyped *Triphysaria* roots. After DMBQ treatment, the transcript level of TvQR1 was significantly higher only in roots with the P1/- genotype. Roots with either the P1/P2 or P2/- genotypes did not significantly up regulate TvQR1 in response to DMBQ.

Haustoria formation of TvQR1 genotyped plants

The proportion of *Triphysaria* roots that form haustoria in response to DMBQ was assessed with plants genotyped for the P1 and P2 alleles. *Triphysaria* with either the P1/P2 or P2/- genotypes formed significantly less haustoria 24 hours after DMBQ treatment, Figure 7.

Artificial promoter constructs

To determine the possible influence of the MITE insertion on DMBQ induced expression of TvQR1, I created artificial promoter constructs with different types of MITEs and insertion locations. These insertions test whether the MITE's effect on expression could be caused by three distinct types of promoter changes: the introduction of sequences within the MITE, the direct interruption of a sequence at the insertion site, or the increase in distance between promoter sequences caused by the insertion. When a change results in a reduced DMBQ responsiveness, then it could be responsible for the lower expression from the TvQR1-P2 allele.

The first construct (PRC-1) represents the uninterrupted TvQR1-P1 allele which is DMBQ responsive. The second construct (PRC-2) imitated the TvQR1-P2 allele and has the MITE inserted into a similar location, Figure 8A. This construct represented all three types of promoter changes and its DMBQ responsiveness was diminished, Figure 8B. The third construct (PRC-3) has a scrambled version of the MITE in the same location as PRC-2 to remove the potential effect of a MITE specific sequence. This construct was also reduced in DMBQ

upregulation. The fourth construct (PRC-4) has the intact TvQR1-P1 with the MITE sequence shifting the promoter upstream from tdTomato. This construct removes the effect of interrupting a promoter sequence but still had a reduced DMBQ response. The last construct (PRC-5) shifted the intact TvQR1-P1 promoter upstream using a scrambled MITE sequence. Similar to the other artificial promoter constructs, PRC-5 also had a diminished DMBQ response.

The constructs PRC2-5 exhibited a lower overall expression level and reduced upregulation in response to DMBQ. The only shared promoter change for all four of these constructs was an increase in distance between promoter sequences and the coding sequence. This suggests the change in spacing between promoter sequences is sufficient to interrupt the typical DMBQ induced regulatory activity of TvQR1-P1. This does not rule out the possibility of the other two types of changes also having effects, but that their effect was not large enough to see through the influence of the change in distance between sequences.

Genome characterization of the TvQR1-P2 like MITEs

To determine if the expression effect associated with the MITE in TvQR1-P2 is a trend with other genes as well, we annotated the locations of all MITE sequences in our genome. A total of 665748 potential MITEs were identified with an average size of 645.8 bp, Figure 9A. The total sites these MITEs occupied, accounted for up to 6.28% of the *Triphysaria* genome. Of the total sites, 383 belonged to the TvQR1-P2 like MITE.

The expression levels of genes with the TvQR1-P2 like MITE in their promoter were summarized to estimate if this MITE has an overall effect on transcriptional expression. The average transcript level of genes containing the TvQR1-P2 like MITE within 1000 bps of the start codon and the level from all genes in our RNASeq data set are shown in Figure 9B. A lower

level of transcriptional expression was observed for genes with the TvQR1-P2 like MITE in their promoter.

DISCUSSION:

In this study, I characterized the spatio-temporal regulation of the TvQR1 promoter before and after prehaustorium induction in *Triphysaria*. The high level of TvQR1-P1 activity at the epidermis and root tip fits with the predicted role of TvQR1 in modifying host derived compounds encountered by the parasite roots. The relatively high expression in the root cap and early epidermal cells could result from these cell types originating from the same initial cells [27]. The expression of TvQR1-P1 in the prehaustorium region also supports the established role in haustorium development for *Triphysaria*.

The DMBQ responsiveness and spatial regulation controlled by the TvQR1-P1 is remarkably conserved between all three plants tested. This defined a relatively small region (652 bps) that contains conserved cis-elements that regulate the above activites. The two non-parasites, *Arabidopsis* and *Mimulus,* represent two major branches of the eudicot phylogenetic tree, Rosids and Asterids respectively. The conservation in DMBQ response and spatial regulation across evolutionary time, suggests these elements may be utilized by other eudicots as well.

The use of DMBQ in neighboring root recognition has been supported for *Arabidopsis* in addition to parasitic plants [28]. The likely receptor for DMBQ has been identified in *Arabidopsis* as well [29]. Upon treatment with DMBQ, *Arabidopsis* roots exhibit a spike in Ca2+ and ROS followed by an arrest of further growth [28,29]. Our work identifies a region of cis-regulatory elements used by *Arabidopsis* and *Mimulus* to repond to DMBQ. The conserved spatial regulation at the epidermis and root tip for all three plants, fits with a potential role that these cis-elements regulate genes that respond to neighboring root derived compounds and that may participate in the regulation of root development.

The lack of DMBQ response for orthologous QR1 genes in *Arabidopsis* and *Mimulus* suggests these non-parasites do not use the same cis-elements to regulate QR1 as *Triphysaria.* In addition, when the response of QR1 to host exudate was measured in other hemiparasitic plants, it was found that *Triphysaria versicolor* was unique in its up-regulation of QR1, even among other species of *Triphysaria*, (see Chapter 2 Figure 1). In another hemiparasitic member of Orobanchaceae, a different quinone reductase gene (QR2) was found to be involved in haustorium development [5]. Together, these observations suggest the expression of QR1 in *Triphysaria versicolor* has been co-opted for haustorium development through the use of conserved cis-elements.

Through our experiments, we also showed how natural variation between TvQR1 promoter alleles can impact a parasite's ability to respond to a host derived compound. The allele TvQR1-P2 was associated with a dominant suppression of TvQR1's host induced up-regulation and a reduced ability for *Triphysaria* to form haustoria in response to DMBQ.

The TvQR1 promoter alleles identified through this study are most likely from the same loci as only one or two alleles were present for each plant and only one TvQR1 locus was discovered through a 10X genomic sequencing of *Triphysaria* from Chapter 1. Plants that only amplified a single allele by PCR were genotyped as either P1/- or P2/-. This was because these plants could be either homozygous (e.g. P1/P1) or heterozygous for a null allele missed by the PCR (e.g. P1/null allele). The presence of null alleles is supported by the fact that roughly a quarter of *Triphysaria* seedlings did not have a promoter amplified while a positive control gene, TUB1, confirmed the presence of amplifiable DNA. Null alleles are a known issue when genotyping wild populations [31-33].

The association of lower transcriptional expression and reduced haustorium formation observed with the TvQR1-P2 allele is consistent with TvQR1 participating in haustorium development for *Triphysaria*. The TvQR1-P2 allele being present in 19% of *Triphysaria* seedlings may be surprising if its negative effect on haustoria formation also occurs in a natural context. It is possible the effect of TvQR1-P2 on haustorium formation is stronger when grown in our controlled conditions and induced with a single host-derived compound. It is also possible the TvQR1-P2 allele has some positive selection in nature that maintains its presence even with a potential fitness cost. Annual plants have to contend with different conditions from year to year. Annual parasitic plants have the additional challenge of dealing with variable host populations between years. The variation observed in the TvQR1 alleles could be beneficial in specific contexts. These possibilities would need to be tested in a natural setting, ideally over multiple years to capture any fitness advantage conferred by annual differences in conditions such as host composition, drought, fire, pathogen pressure, or others.

One aspect of variation between the TvQR1 promoter alleles was a MITE which was found in the conserved region between these alleles. It was expected that the expression of this allele could be affected by this variation. It was not expected however, that the MITE associated reduction in expression was dominant and affected the other TvQR1 allele as well. MITE sequences are frequently targeted with sRNAs that can silence expression either post-transcriptionally through the cleavage of mRNA or transcriptionally through the recruitment of epigenetic remodeling machinery. The dsRNA structure the TvQR1 MITE is predicted to form once transcribed could result in the generation of sRNAs, Figure S3. The dominant effect on expression could be explained if the insertion of the MITE triggered the production of sRNAs that targeted the surrounding TvQR1 sequences. These sRNAs could then travel to the other

allele and trigger its silencing. This phenomenon has been observed but not explored deeply[20]. Our results suggest that MITEs may have additional effects on gene expression than what has been previously established which warrants further research.

The lower level of TvQR1 expression from the P2 allele could result from a variety of promoter changes created by the MITE insertion including the introduction of a sequence with regulatory activity, the direct interruption of a promoter sequence, or the increased distance between promoter sequences caused by the insertion. By creating artificial promoters with different combinations of these types of changes, our experiments show the change in the distance between promoter sequences resulting from the insertion can interrupt the DMBQ responsiveness of the promoter. This does not rule out the possibility of the MITE bringing in new transcription factor motifs with regulatory activity or the silencing of a MITE specific sequence, but that the change in distance between promoter elements was sufficient to reduce expression.

The identification of MITE sequences in the *Triphysaria* genome yielded a similar overall proportion to other plant genomes [11,12]. This large scale annotation of MITE locations allowed us to identify trends in their effect on transcriptional expression. Overall, genes with a TvQR1-P2 like MITE in the first 1000 bps of the promoter, had lower transcriptional expression than *Triphysaria* genes on average. This trend fits with our specific observation for the TvQR1 promoter and previous findings on MITE induced expression changes[17,18].

Our experiments show that genotypic variation created in the TvQR1 promoter by a MITE insertion may have resulted in a phenotypic change in expression and haustorium formation. Due to the prevalence, location bias, and action of MITEs potentially interrupting or creating new regulatory elements, it is thought they likely play a major role in gene and genome

evolution. The exploration of transposable elements as drivers of genomic change in parasitic plants warrants further research.

Through this work, I identified a 652 bp promoter region containing conserved cis-elements that regulate spatial expression and response to DMBQ in multiple eudicots. The function of these cis-elements may be involved in the recognition and response of roots to those of neighboring plants. Triphysaria may have used these elements to co-opt the QR1 gene for use in haustoria formation. Further research is warranted to identify the specific cis-element controlling this response, as it may yield insights into the ecology of plant-plant interactions and how parasitic plants have co-opted this communication for haustorium development during the evolution of parasitism. These refined elements may also be useful in engineering the expression of plant genes for synthetic biology.

FIGURES:

A.

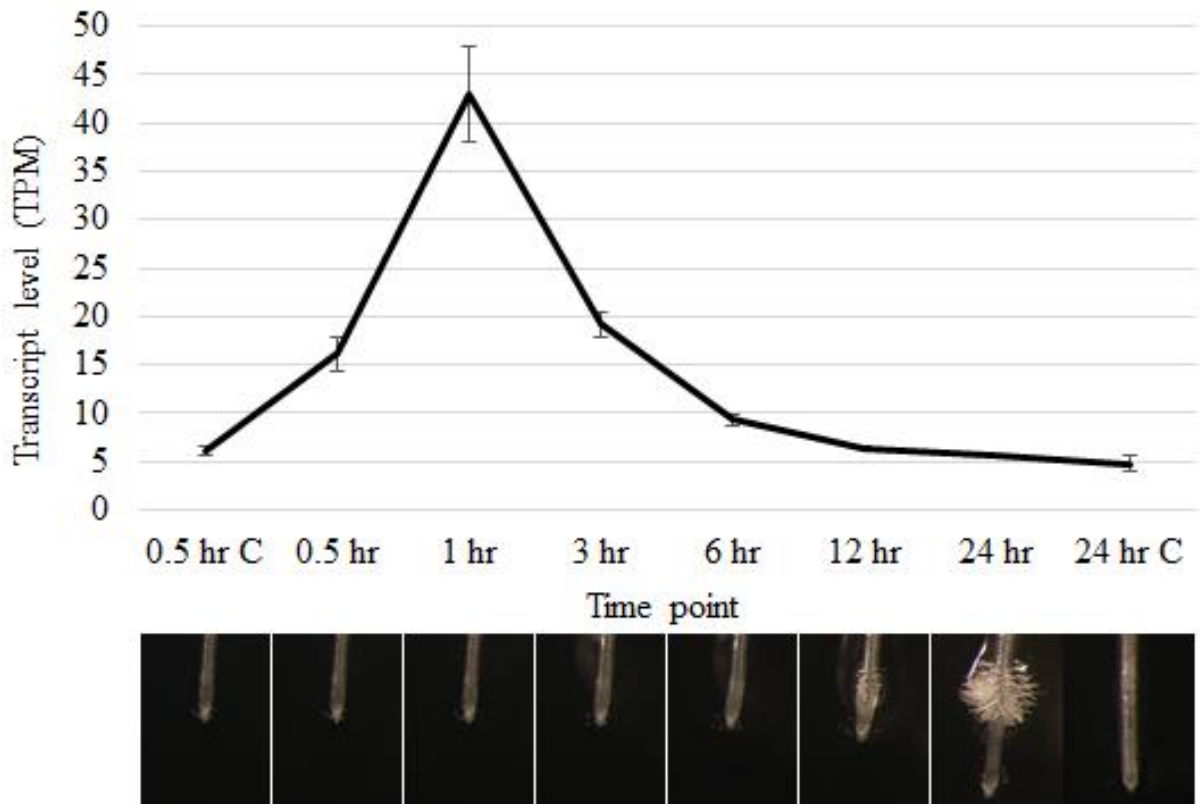


Figure 1: TvQR1 transcript levels after treatment with *Arabidopsis* root exudate.
A. Transcript levels were determined by the RNASeq experiment described in Chapter 1. The error bars represent the standard error of three replicates for each time point. Representative images of *Triphysaria versicolor* root tips are shown beneath each developmental time point.

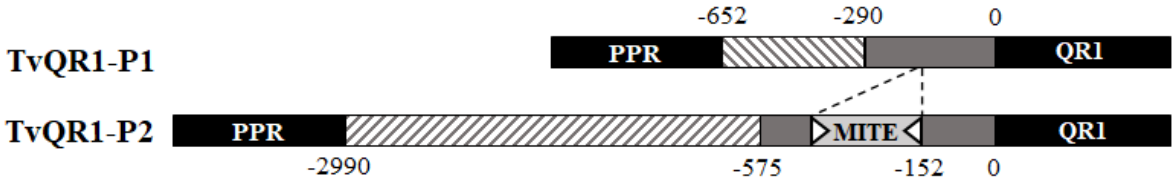
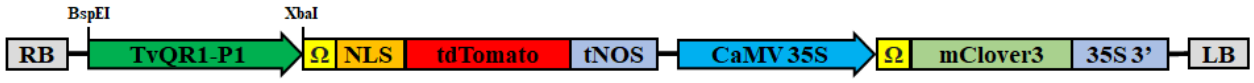


Figure 2: Diagram of two TvQR1 promoter alleles showing key features.
The TvQR1-P1 and TvQR1-P2 promoter alleles are shown with key features annotated. The location of region boundaries relative to the TvQR1 start codon are shown on either side of each promoter. The darker shaded region at the beginning of each promoter shows the region of high conservation between the sequences. The MITE is located within this conserved promoter region. The diagonal striped regions represent those with little similarity between the promoters. A pentapeptide repeat protein (PPR) gene is immediately upstream of TvQR1.

A.



B.

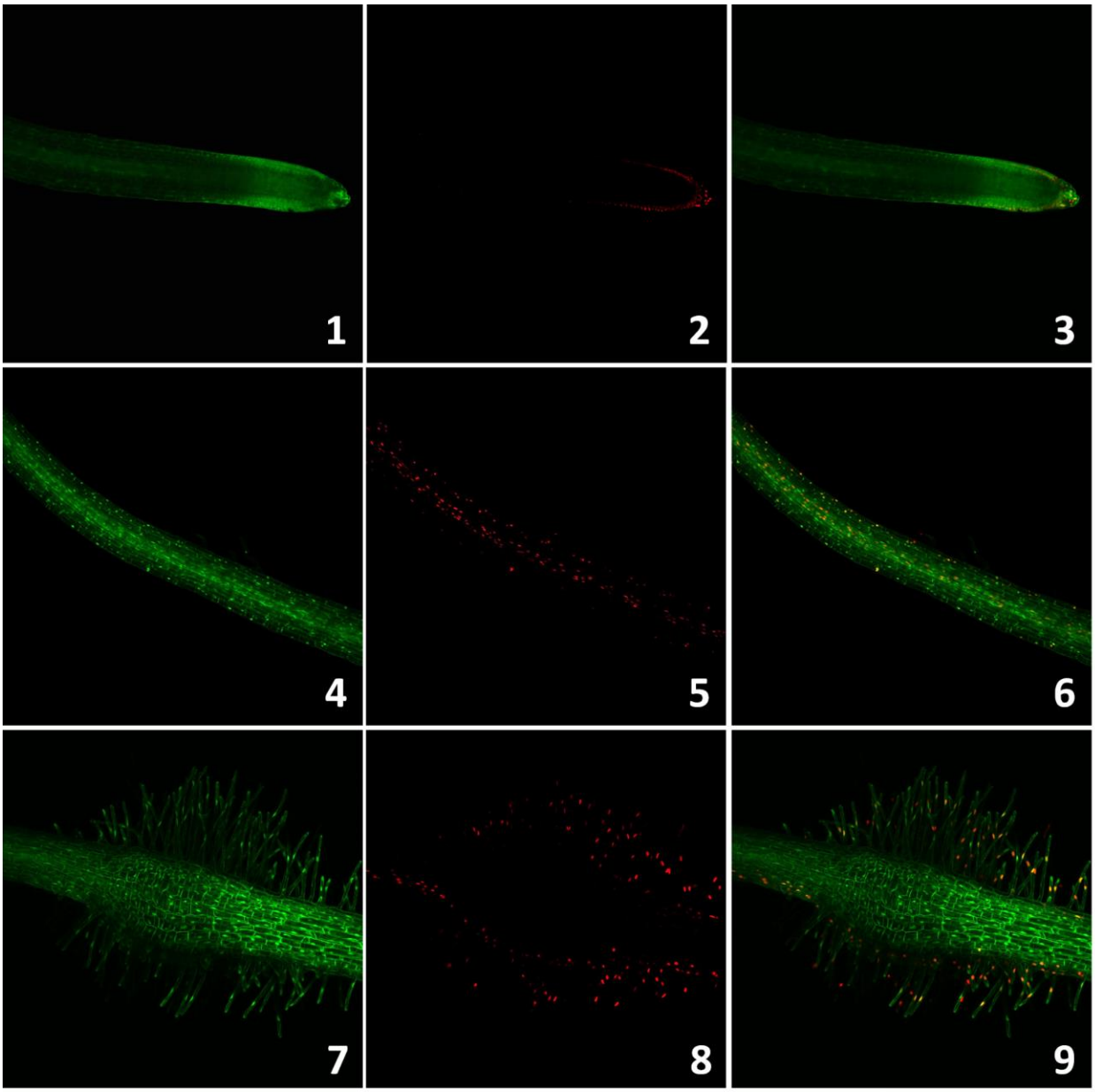


Figure 3: Fluorescent reporting system for promoter sequences. A. The promoter reporting construct pDS-PRC was used to characterize the regulatory activity of candidate promoter sequences (e.g. TvQR1-P1) in transgenic roots (Ω: TMV Omega, NLS: Nuclear Localizing Signal). B. Confocal images showing the spatial regulation of the TvQR1-P1 in *T. versicolor* root tips before DMBQ exposure(1-3), a mature root section before DMBQ exposure(4-6), and a prehaustorium 24 hours after DMBQ treatment(7-9). Red is TvQR1::tdTomato and green is the constitutively expressed 35S::mClover3 used to outline cells for imaging.

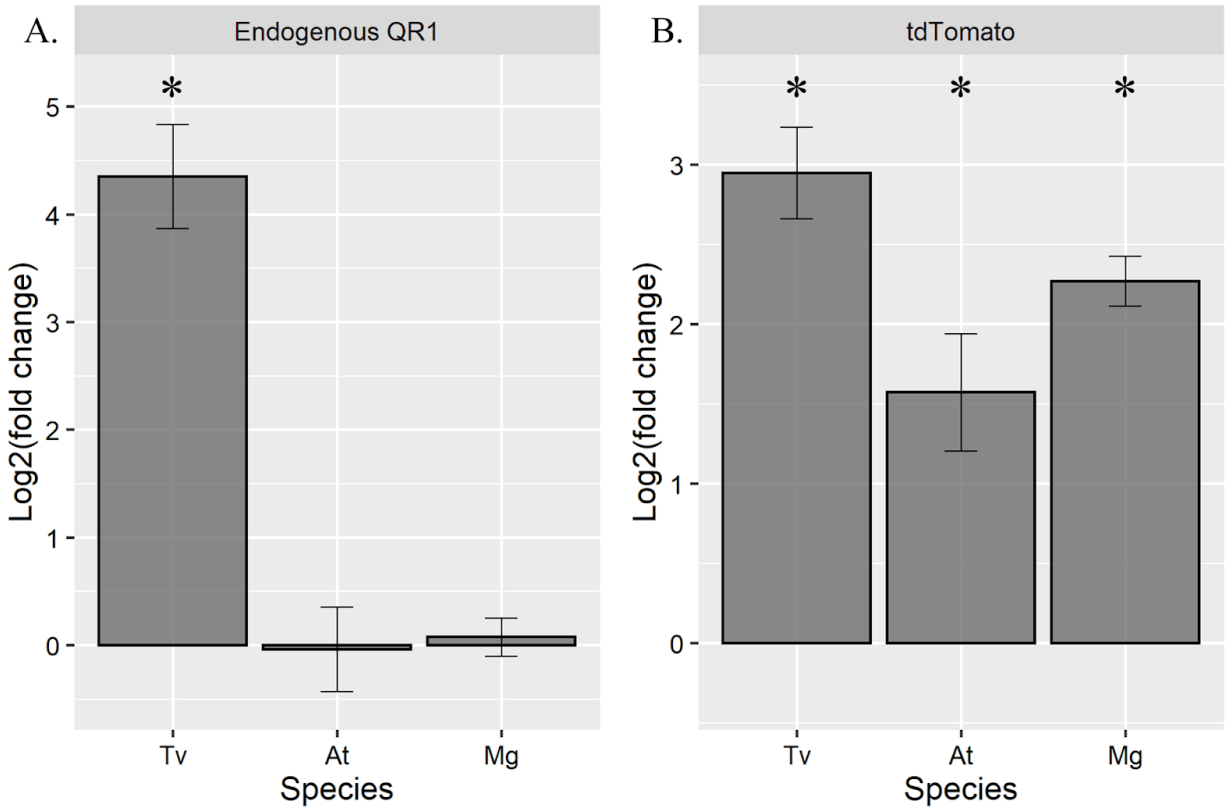


Figure 4: The DMBQ responsiveness of TvQR1-P1 in *Triphysaria, Arabidopsis,* and *Mimulus*. The pDS-PRC construct from Figure 2A was transformed into *Triphysaria versicolor* (Tv), *Mimulus guttatus* (Mg) and *Arabidopsis thaliana* (At). The transcript levels before and 1 hour after DMBQ exposure were determined by qPCR for the endogenous versions of QR1 from each species (A) and tdTomato (B). The log2 of the fold change between these time points is shown on the y-axis. The bars represent the mean of 3 independently transformed plants with the error bars showing the standard error. An * represents if the log2(fold change) values were significantly different than a mean of 0 using a Welch's t-test (p-value ≤ 0.05).

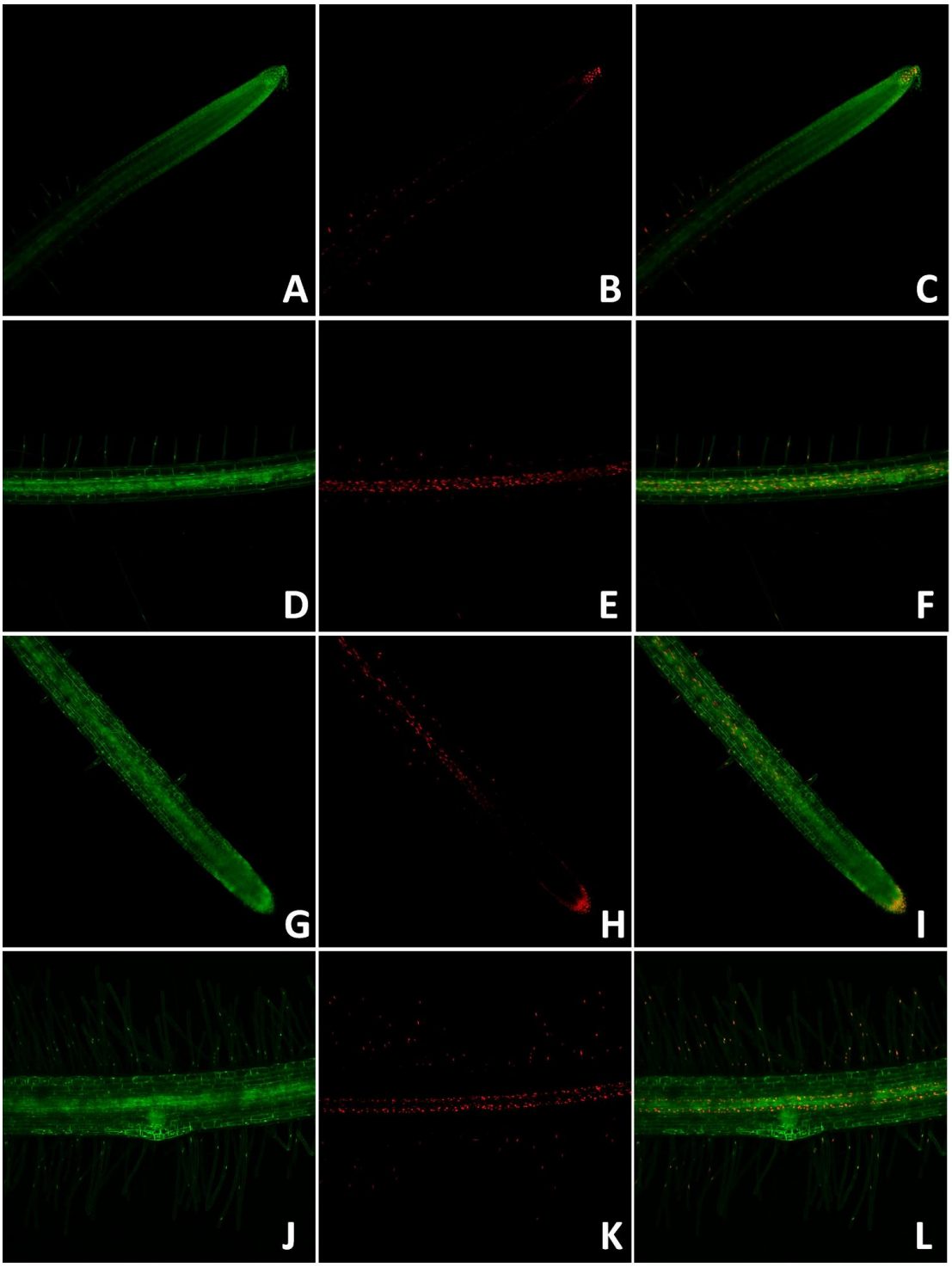


Figure 5: TvQR1-P1 fluorescent reporting construct transformed in the non-parasitic plants *Arabidopsis thaliana* and *Mimulus guttatus*. The pDS-PRC construct shown in Figure 2A, containing the *Triphysaria* TvQR1 promoter driving the expression of tdTomato, was transformed into *Arabidopsis thaliana* (A-F) and *Mimulus guttatus* (G-L). Panels A-C and G-H are root tips, while D-F and J-L are mature root sections.

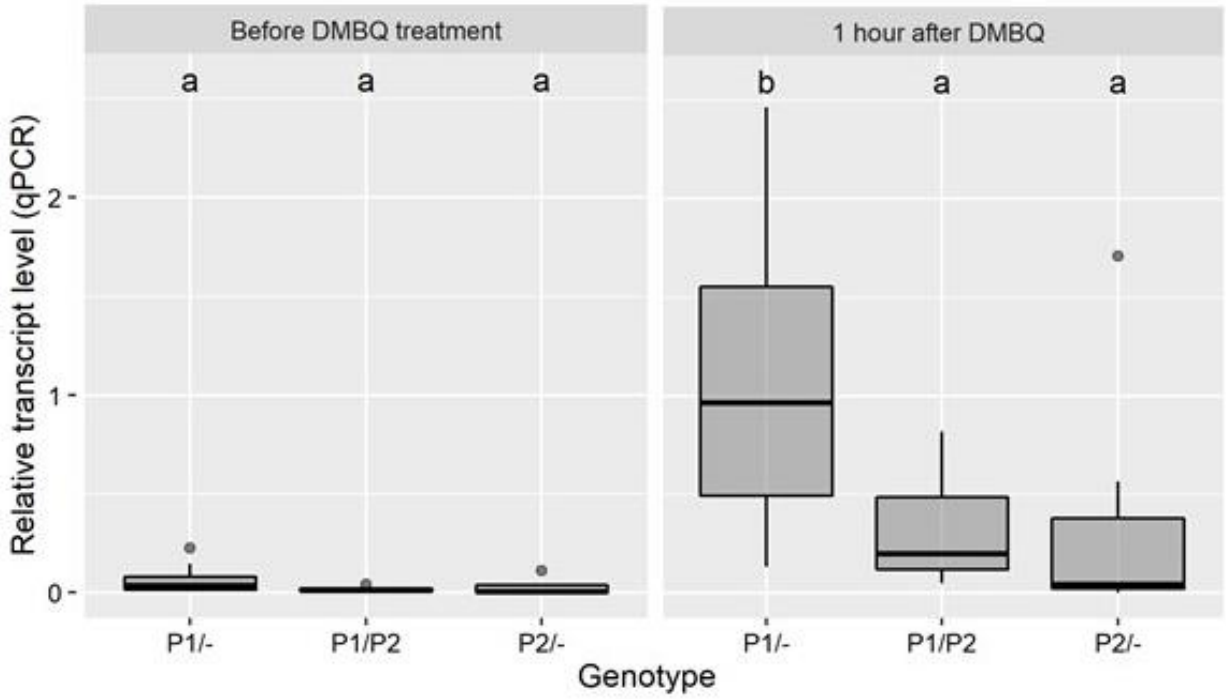


Figure 6: Endogenous TvQR1 transcript levels of genotyped *Triphysaria* plants before and after DMBQ treatment. Box plots of quantitative PCR (qPCR) data showing the relative transcript level of TvQR1 before and after DMBQ treatment in *Triphysaria* plants genotyped as TvQR1-P1 (P1/-), TvQR1-P2 (P2/-), or heterozygous for both promoter alleles (P1/P2). For each genotype, ten plants were evaluated. The letters above each box represent statistically significant groups using a TukeyHSD test with a confidence level of 95%.

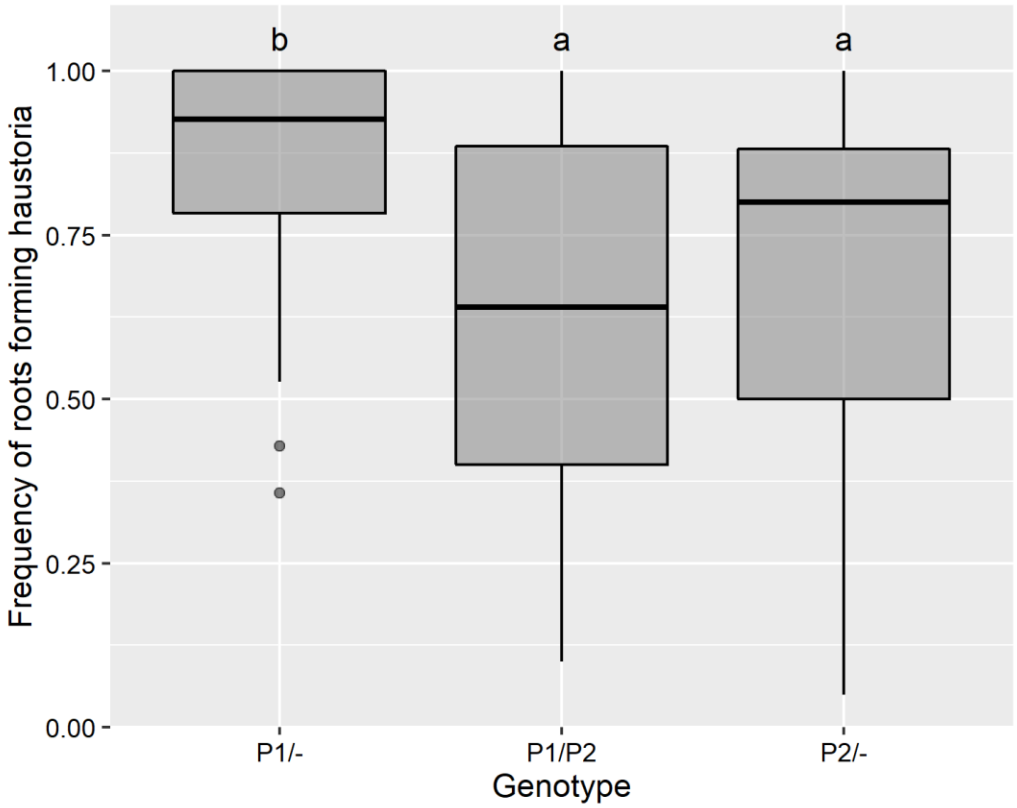


Figure 7: Frequency of haustoria formation for genotyped *Triphysaria* roots.
The frequency of haustoria formation per plant was calculated by the ratio of roots that formed haustoria to the total number of root tips present. Box plots show the distribution of these frequencies for *Triphysaria* plants genotyped as TvQR1-P1 (P1/-), TvQR1-P2 (P2/-), or heterozygous for both promoter alleles (P1/P2). The number of plants evaluated for each genotype is as follows: (P1/-)=60, (P1/P2)=24, (P2/-)=32. The letters above each box represent statistically significant groups using a TukeyHSD test with a confidence level of 95%.

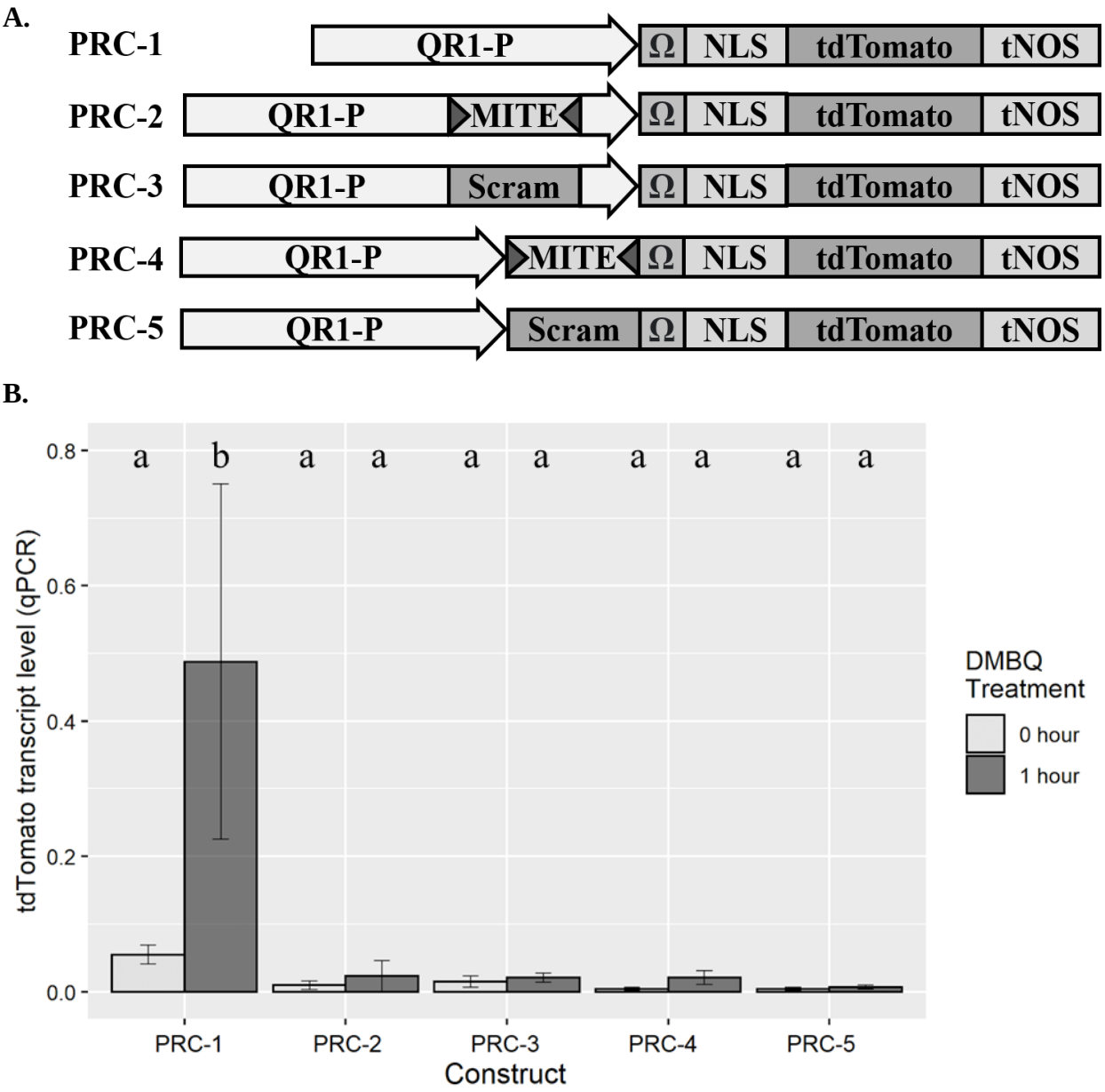


Figure 8: Artificial promoter constructs testing different effects of a MITE insertion on transcription expression. A. Diagram of five artificial promoter constructs used to test different types of MITE induced changes to a promoter sequence. MITE and scrambled MITE (Scram) sequences were inserted either within or before the TvQR1-P1. B. qPCR data showing the average tdTomato transcript level for 2-5 independently transformed roots per construct before and after DMBQ treatment. The transcript level of tdTomato is relative to two reference genes, PTB and PAB2. The error bars show the standard error of the qPCR data. The letters above each box represent statistically significant groups using a TukeyHSD test with a confidence level of 95%.

A.

Triphysaria versicolor MITE annotation summary:	
Total number of MITE locations:	Up to 665748
Proportion of Genome:	Up to 6.28%
Average MITE size:	645.8 bp
Number of TvQR1-P2 like MITE locations:	383
Size of TvQR1-P2 like MITE	267 bp

B.

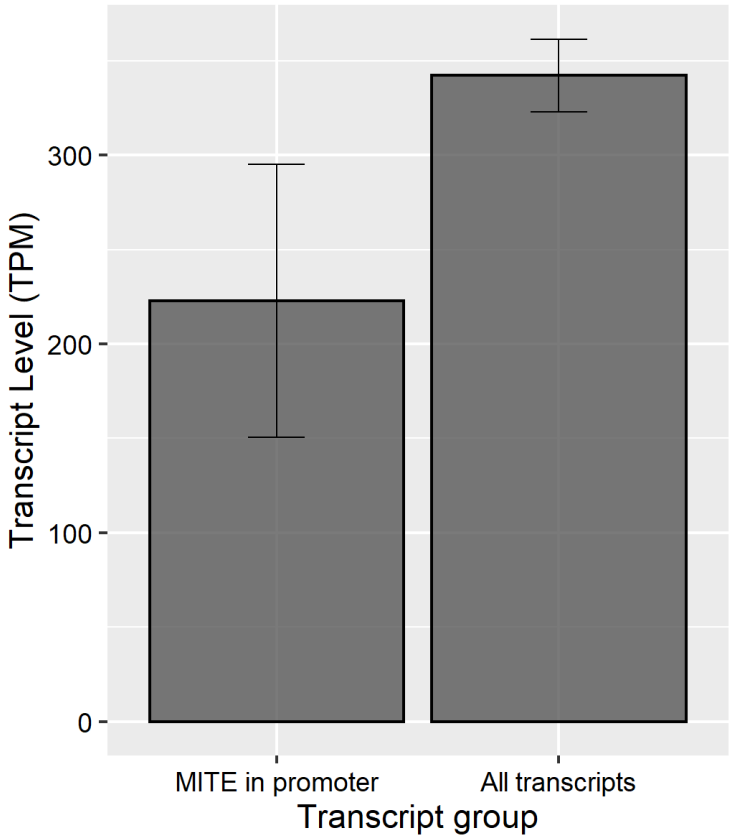


Figure 9: Genome wide effect of TvQR1-P2 like MITEs on transcriptional expression. A. Characteristics of all MITE and the TvQR1-P2 like MITE annotated in the *Triphysaria* genome. B. The x-axis shows two groups of transcripts: those with a TvQR1-P2 like MITE in their promoter (n = 76) and all of the transcripts from our RNASeq (n = 23378). The promoter was defined as 1000 bps upstream of the start codon. The y-axis represents the transcript level from our RNASeq. The transcript levels for individual genes was determined by the sum of expression across all 24 samples of our RNASeq data set described in Chapter 1. The bars represent the mean of each transcript group. The error bars show the 95% confidence intervals for the means of each group of transcripts. Using a Welch's t-test, the means are significantly different with a p-value of 0.002.

SUPPLEMENTAL:

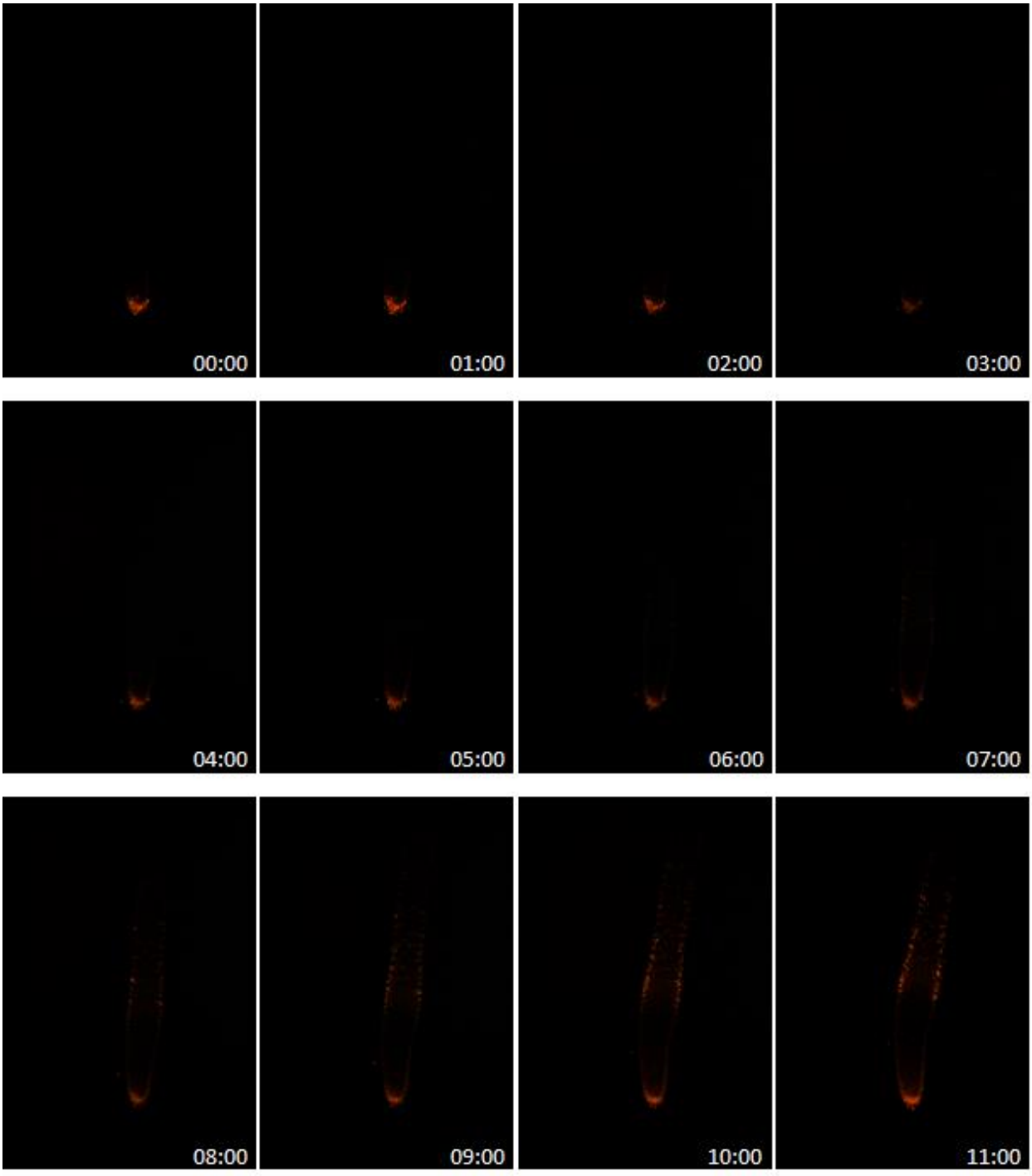


Figure S1: Time lapse images of pDS-PRC-TvQR1-P1 expression during prehaustorium development in *Triphysaria.* Prehaustoria were induced using 30 μM DMBQ. One image was taken every hour.

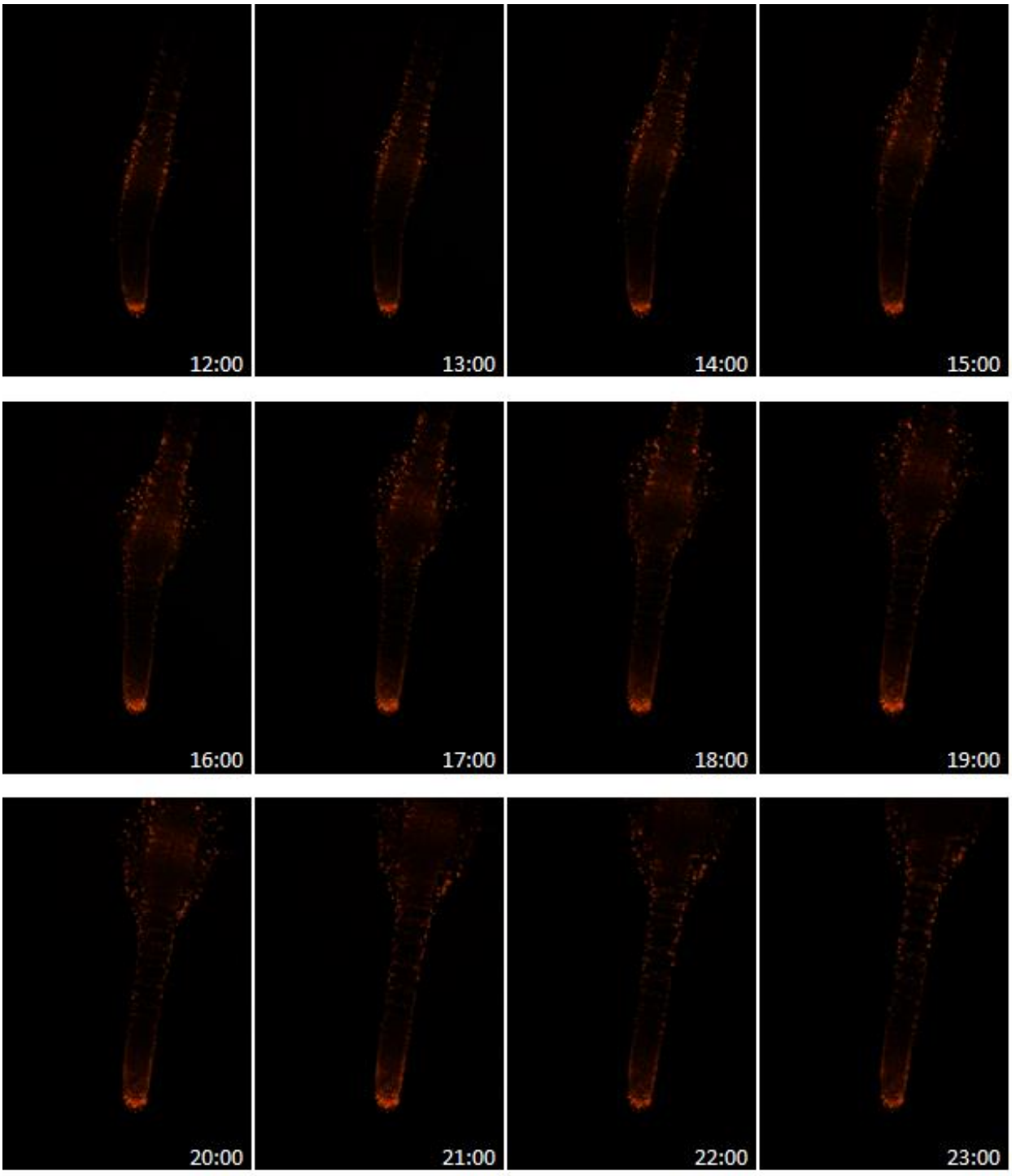
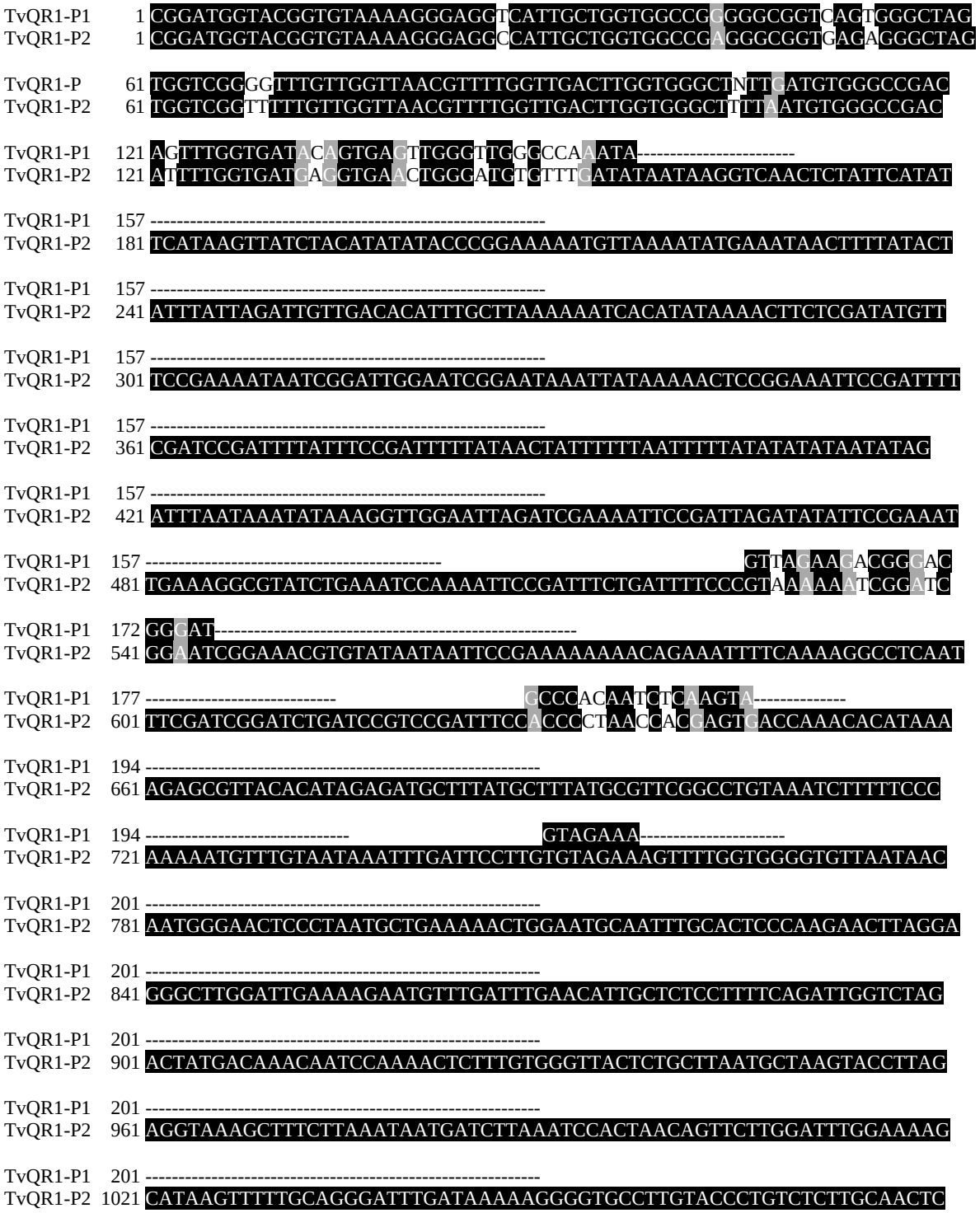
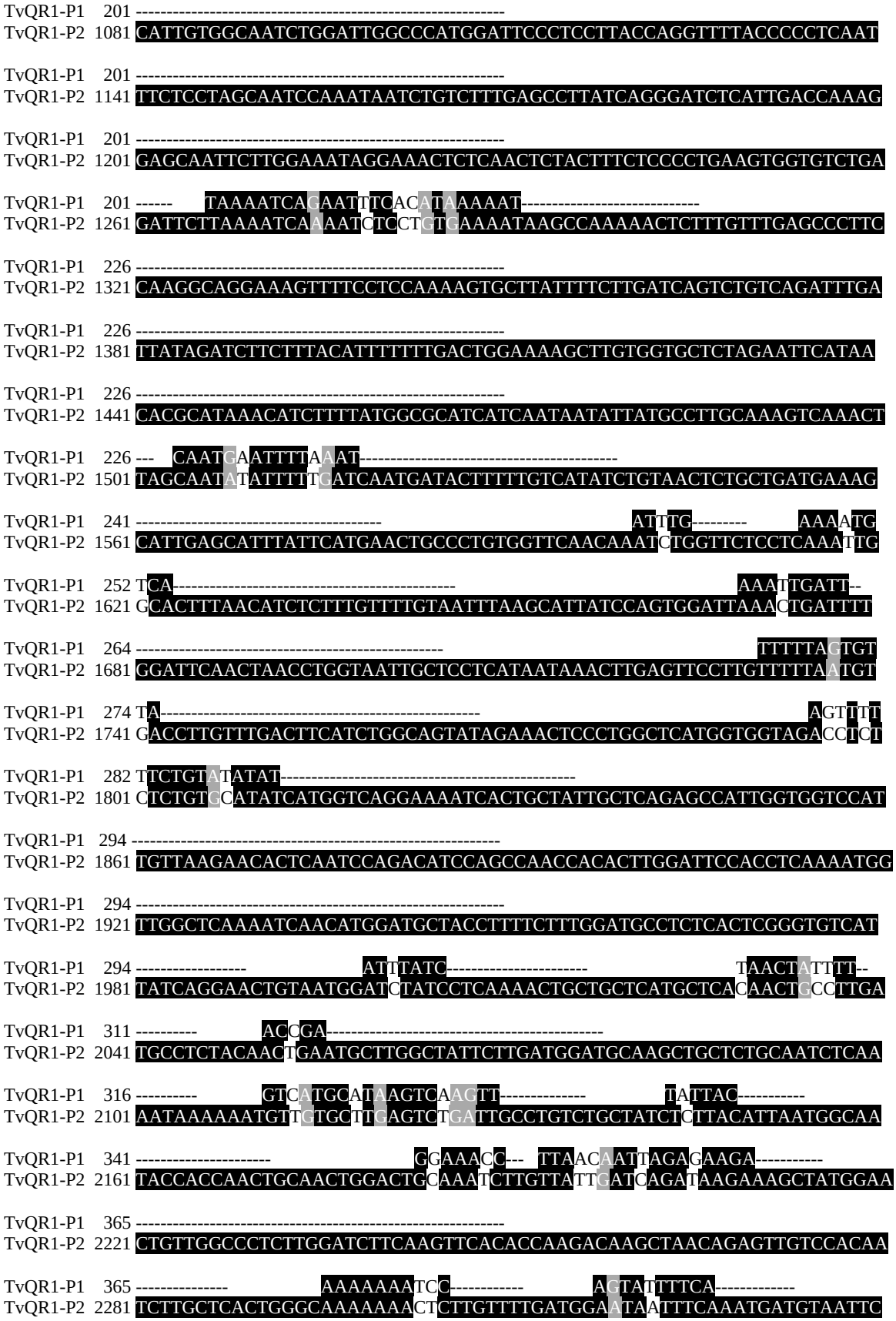
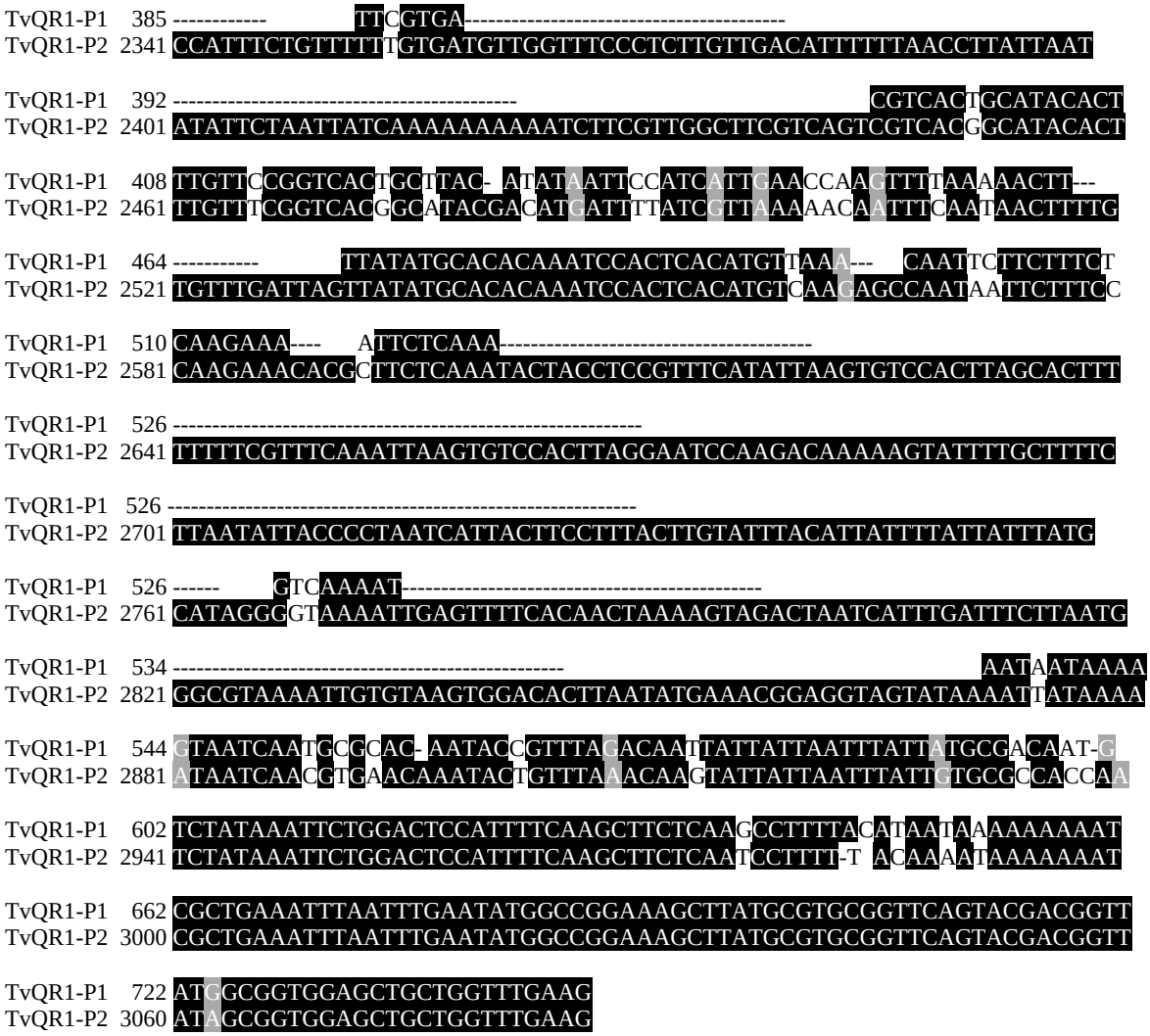


Figure S1 continued.

Figure S2: Alignment of TvQR1-P1 and TvQR1-P2 promoter alleles.
The TvQR1 promoter alleles were aligned using the MUSCLE algorithm with default parameters followed by a shaded visualization using Boxshade.




TvQR1-P1 201 ----------
TvQR1-P2 1081 CATTGTGGCAATCTGGATTGGCCCATGGATTCCCTCCTTACCAGGTTTTACCCCCTCAAT
TvQR1-P1 201 ----------
TvQR1-P2 1141 TTCTCCTAGCAATCCAAATAATCTGTCTTTGAGCCTTATCAGGGATCTCATTGACCAAAG
TvQR1-P1 201 ----------
TvQR1-P2 1201 GAGCAATTCTTGGAAATAGGAAACTCTCAACTCTACTTTCTCCCCTGAAGTGGTGTCTGA
TvQR1-P1 201 ------ TAAAATCAGAATTTCACATAAAAAT----------
TvQR1-P2 1261 GATTCTTAAAATCAAAATCTCCTCTGAAAATAAGCCAAAAACTCTTTGTTTGAGCCCTTC
TvQR1-P1 226 ----------
TvQR1-P2 1321 CAAGGCAGGAAAGTTTTCCTCCAAAAGTGCTTATTTTCTTGATCAGTCTGTCAGATTTGA
TvQR1-P1 226 ----------
TvQR1-P2 1381 TTATAGATCTTCTTTACATTTTTTTGACTGGAAAAGCTTGTGGTGCTCTAGAATTCATAA
TvQR1-P1 226 ----------
TvQR1-P2 1441 CACGCATAAACATCTTTTATGGCGCATCATCAATAATATTATGCCTTGCAAAGTCAAACT
TvQR1-P1 226 --- CAATCAATTTTAAAT----------
TvQR1-P2 1501 TAGCAATATATTTTTGATCAATGATACTTTTTGTCATATCTGTAACTCTGCTGATGAAAG
TvQR1-P1 241 ---------- ATTTG--------- AAAATG
TvQR1-P2 1561 CATTGAGCATTTATTCATGAACTGCCCTGTGGTTCAACAAATCTGGTTCTCCTCAAATTG
TvQR1-P1 252 TCA---------- AAATTGATT--
TvQR1-P2 1621 GCACTTTAACATCTCTTTGTTTTGTAATTTAAGCATTATCCAGTGGATTAAACTGATTTT
TvQR1-P1 264 ---------- TTTTTAGTGT
TvQR1-P2 1681 GGATTCAACTAACCTGGTAATTGCTCCTCATAATAAACTTGAGTTCCTTGTTTTTAATGT
TvQR1-P1 274 TA---------- AGTTTTT
TvQR1-P2 1741 GACCTTGTTTGACTTCATCTGGCAGTATAGAAACTCCCTGGCTCATGGTGGTAGACCTCT
TvQR1-P1 282 TTCTGTATATAT----------
TvQR1-P2 1801 CTCTGTGCATATCATGGTCAGGAAAATCACTGCTATTGCTCAGAGCCATTGGTGGTCCAT
TvQR1-P1 294 ----------
TvQR1-P2 1861 TGTTAAGAACACTCAATCCAGACATCCAGCCAACCACACTTGGATTCCACCTCAAAATGG
TvQR1-P1 294 ----------
TvQR1-P2 1921 TTGGCTCAAAATCAACATGGATGCTACCTTTTCTTTGGATGCCTCTCACTCGGGTGTCAT
TvQR1-P1 294 ---------- ATTTATC---------- TAACTATTTT--
TvQR1-P2 1981 TATCAGGAACTGTAATGGATCTATCCTCAAAACTGCTGCTCATGCTCACAACTGCCTTGA
TvQR1-P1 311 ---------- ACCGA----------
TvQR1-P2 2041 TGCCTCTACAACTGAATGCTTGGCTATTCTTGATGGATGCAAGCTGCTCTGCAATCTCAA
TvQR1-P1 316 ---------- GTCATGCATAAGTCAAGTT---------- TATTAC----------
TvQR1-P2 2101 AATAAAAAATGTTGTGCTTGAGTCTGATTGCCTGTCTGCTATCTCTTACATTAATGGCAA
TvQR1-P1 341 ---------- GGAAACC--- TTAACAATTAGAGAAGA----------
TvQR1-P2 2161 TACCACCAACTGCAACTGGACTGCAAATCTTGTTATTGATCAGATAAGAAAGCTATGGAA
TvQR1-P1 365 ----------
TvQR1-P2 2221 CTGTTGGCCCTCTTGGATCTTCAAGTTCACACCAAGACAAGCTAACAGAGTTGTCCACAA
TvQR1-P1 365 ---------- AAAAAAATCC---------- AGTATTTTCA----------
TvQR1-P2 2281 TCTTGCTCACTGGGCAAAAAAACTCTTGTTTTGATGGAATAATTTCAAATGATGTAATTC


TvQR1-P1 385 ---------- TTCGTGA----------
TvQR1-P2 2341 CCATTTCTGTTTTTTGTGATGTTGGTTTCCCTCTTGTTGACATTTTTTAACCTTATTAAT
TvQR1-P1 392 ---------- CGTCACTGCATACACT
TvQR1-P2 2401 ATATTCTAATTATCAAAAAAAAAATCTTCGTTGGCTTCGTCAGTCGTCACGGCATACACT
TvQR1-P1 408 TTGTTCCGGTCACTGCTTAC- ATATAATTCCATCATTGAACCAAGTTTTAAAAAACTT---
TvQR1-P2 2461 TTGTTTCGGTCACGGCATACGACATGATTTTATCGTTAAAAACAATTTCAATAACTTTTG
TvQR1-P1 464 ---------- TTATATGCACACAAATCCACTCACATGTTAAA---- CAATTCTTCTTTCT
TvQR1-P2 2521 TGTTTGATTAGTTATATGCACACAAATCCACTCACATGTCAACAGCCAATAATTCTTTCC
TvQR1-P1 510 CAAGAAA---- ATTCTCAAA----------
TvQR1-P2 2581 CAAGAAACACGCTTCTCAAATACTACCTCCGTTTCATATTAAGTGTCCACTTAGCACTTT
TvQR1-P1 526 ----------
TvQR1-P2 2641 TTTTTCGTTTCAAATTAAGTGTCCACTTAGGAATCCAAGACAAAAAGTATTTTGCTTTTC
TvQR1-P1 526 ----------
TvQR1-P2 2701 TTAATATTACCCCTAATCATTACTTCCTTTACTTGTATTTACATTATTTTATTATTTATG
TvQR1-P1 526 ------ GTCAAAAT----------
TvQR1-P2 2761 CATAGGGGTAAAATTGAGTTTTCACAACTAAAAGTAGACTAATCATTTGATTTCTTAATG
TvQR1-P1 534 ---------- AATAATAAAA
TvQR1-P2 2821 GGCGTAAAATTGTGTAAGTGGACACTTAATATGAAACGGAGGTAGTATAAAATTATAAAA
TvQR1-P1 544 GTAATCAATGCGCAC-AATACCGTTTAGACAATTATTATTAATTTATTATGCGACAAT-G
TvQR1-P2 2881 ATAATCAACGTGAACAAATACTGTTTAAACAAGTATTATTAATTTATTGTGCGCCACCAA
TvQR1-P1 602 TCTATAAATTCTGGACTCCATTTTCAAGCTTCTCAAGCCTTTTACATAATAAAAAAAAAT
TvQR1-P2 2941 TCTATAAATTCTGGACTCCATTTTCAAGCTTCTCAATCCTTTT-T ACAAAATAAAAAAAT
TvQR1-P1 662 CGCTGAAATTTAATTTGAATATGGCCGGAAAGCTTATGCGTGCGGTTCAGTACGACGGTT
TvQR1-P2 3000 CGCTGAAATTTAATTTGAATATGGCCGGAAAGCTTATGCGTGCGGTTCAGTACGACGGTT
TvQR1-P1 722 ATCGCGGTGGAGCTGCTGGTTTGAAG
TvQR1-P2 3060 ATAGCGGTGGAGCTGCTGGTTTGAAG

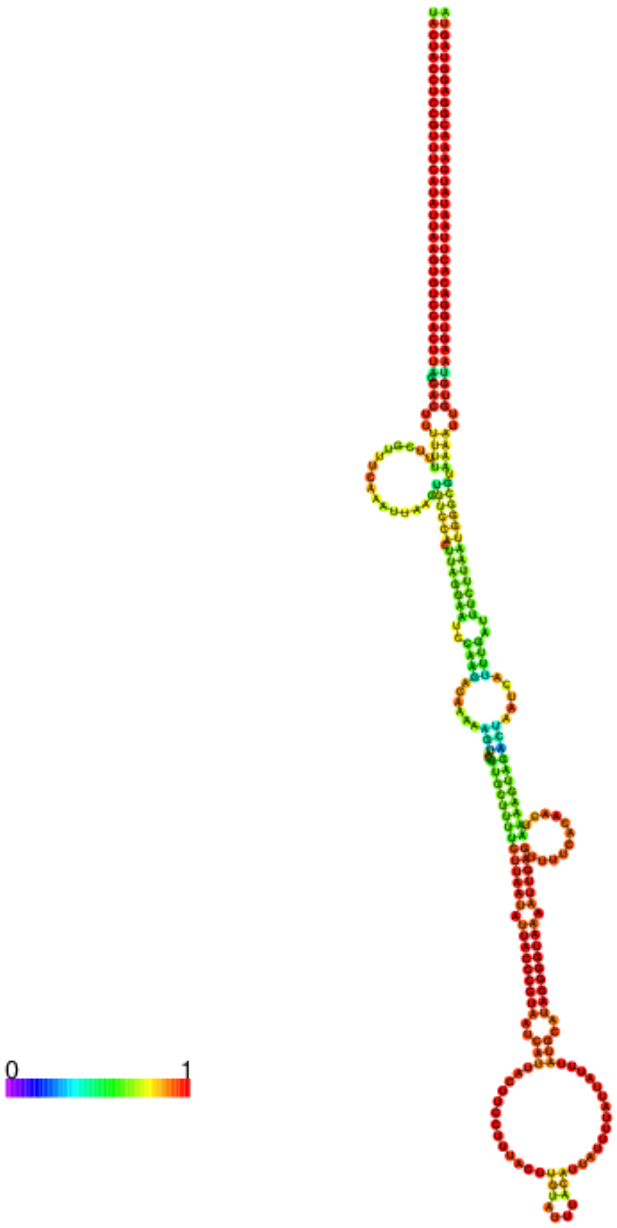


Figure S3: Model of potential secondary structure formed by the transcribed TvQR1 promoter MITE. The RNAFold Server from the ViennaRNA Web Services was used with default settings to predict the secondary folding structure the TvQR1 promoter MITE could form if transcribed. The scale bar represents the base-pair probabilities.

REFERENCES:

1. Bandaranayake PCG, Filappova T, Tomilov A, Tomilova NB, Jamison-McClung D, Ngo Q, et al. A single-electron reducing quinone oxidoreductase is necessary to induce haustorium development in the root parasitic plant Triphysaria. Plant Cell. 2010;22: 1404–1419.

2. Bandaranayake PCG, Tomilov A, Tomilova NB, Ngo QA, Wickett N, dePamphilis CW, et al. The TvPirin gene is necessary for haustorium development in the parasitic plant Triphysaria versicolor. Plant Physiol. 2012;158: 1046–1053.

3. Yang Z, Wafula EK, Honaas LA, Zhang H, Das M, Fernandez-Aparicio M, et al. Comparative transcriptome analyses reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol Biol Evol. 2015;32: 767–790.

4. Ishida JK, Wakatake T, Yoshida S, Takebayashi Y, Kasahara H, Wafula E, et al. Local Auxin Biosynthesis Mediated by a YUCCA Flavin Monooxygenase Regulates Haustorium Development in the Parasitic Plant Phtheirospermum japonicum. Plant Cell. 2016;28: 1795–1814.

5. Ishida JK, Yoshida S, Shirasu K. Quinone oxidoreductase 2 is involved in haustorium development of the parasitic plant Phtheirospermum japonicum. Plant Signal Behav. 2017;12: e1319029.

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

7. True JR, Carroll SB. Gene co-option in physiological and morphological evolution. Annu Rev Cell Dev Biol. 2002;18: 53–80.

8. Ngo QA, Albrecht H, Tsuchimatsu T, Grossniklaus U. The differentially regulated genes TvQR1 and TvPirin of the parasitic plant Triphysaria exhibit distinctive natural allelic diversity. BMC Plant Biol. 2013;13: 28.

9. Bureau TE, Wessler SR. Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell. 1994;6: 907–916.

10. Feschotte C, Jiang N, Wessler SR. Plant transposable elements: where genetics meets genomics. Nat Rev Genet. 2002;3: 329–341.

11. Tu Z. Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, Aedes aegypti. Proc Natl Acad Sci U S A. 1997;94: 7475–7480.

12. Chen J, Hu Q, Zhang Y, Lu C, Kuang H. P-MITE: a database for plant miniature inverted-repeat transposable elements. Nucleic Acids Res. 2014;42: D1176–81.

13. Bureau TE, Ronald PC, Wessler SR. A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes. Proc Natl Acad Sci U S A. 1996;93: 8524–8529.

14. Feschotte C, Mouchès C. Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the Arabidopsis thaliana genome has arisen from a pogo-like DNA transposon. Mol Biol Evol. 2000;17: 730–737.

15. Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, et al. An active DNA transposon family in rice. Nature. 2003. pp. 163–167. doi:10.1038/nature01214

16. Morata J, Marín F, Payet J, Casacuberta JM. Plant Lineage-Specific Amplification of Transcription Factor Binding Motifs by Miniature Inverted-Repeat Transposable Elements (MITEs). Genome Biol Evol. 2018;10: 1210–1220.

17. Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in Oryza sativa. Mol Biol Evol. 2012;29: 1005–1017.

18. Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, et al. Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Res. 2009;19: 42–56.

19. Liu Y, Tahir Ul Qamar M, Feng J-W, Ding Y, Wang S, Wu G, et al. Comparative analysis of miniature inverted-repeat transposable elements (MITEs) and long terminal repeat (LTR) retrotransposons in six Citrus species. BMC Plant Biol. 2019;19: 140.

20. Gagliardi D, Cambiagno DA, Arce AL, Tomassi AH, Giacomelli JI, Ariel FD, et al. Dynamic regulation of chromatin topology and transcription by inverted repeat-derived small RNAs in sunflower. Proc Natl Acad Sci U S A. 2019;116: 17578–17583.

21. Sampath P, Murukarthick J, Izzah NK, Lee J, Choi H-I, Shirasawa K, et al. Genome-Wide Comparative Analysis of 20 Miniature Inverted-Repeat Transposable Element Families in Brassica rapa and B. oleracea. PLoS ONE. 2014. p. e94499. doi:10.1371/journal.pone.0094499

22. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media; 2009.

23. Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 2010;38: e199.

24. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117: 9451–9457.

25. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842.

26. Zeng L, Zhang N, Zhang Q, Endress PK, Huang J, Ma H. Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. New Phytol. 2017;214: 1338–1354.

27. Evert RF. Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development. John Wiley & Sons; 2006.

28. Fuller AW, Young P, Pierce BD, Kitson-Finuff J, Jain P, Schneider K, et al. Redox-mediated quorum sensing in plants. PLoS One. 2017;12: e0182655.

29. Laohavisit A, Wakatake T, Ishihama N, Mulvey H, Takizawa K, Suzuki T, et al. Quinone perception in plants via leucine-rich-repeat receptor-like kinases. Nature. 2020;587: 92–97.

30. Yoder JI. Self and cross-compatibility in three species of the hemiparasite Triphysaria (Scrophulariaceae). Environmental and Experimental Botany. 1998. pp. 77–83. doi:10.1016/s0098-8472(97)00038-5

31. Dakin EE, Avise JC. Microsatellite null alleles in parentage analysis. Heredity. 2004. pp. 504–509. doi:10.1038/sj.hdy.6800545

32. Castric V, Bernatchez L, Belkhir K, Bonhomme F. Heterozygote deficiencies in small lacustrine populations of brook charr Salvelinus Fontinalis Mitchill (Pisces, Salmonidae): a test of alternative hypotheses. Heredity . 2002;89: 27–35.

33. Montarry J, Jan P-L, Gracianne C, Overall ADJ, Bardou-Valette S, Olivier E, et al. Heterozygote deficits in cyst plant-parasitic nematodes: possible causes and consequences. Mol Ecol. 2015;24: 1654–1677.

CHAPTER 4:

HOST INDUCED GENE SILENCING TARGETING LIPID BIOSYNTHESIS IN THE PARASITIC WEED *PHELIPANCHE AEGYPTIACA*

ABSTRACT:

Root parasitic weeds are among the most internationally devastating crop pests. This chapter investigates using Host Induced Gene Silencing (HIGS) as a control strategy against parasitic weeds. We have evaluated the effectiveness of RNAi mediated silencing on the growth and development of the parasite *Phelipanche aegyptiaca* on tomato (*Solanum lycopersicum*). Hairpin RNA (hpRNA) constructs were transformed into tomato that targeted four *P. aegyptiaca* genes critical for lipid biosynthesis. Roots of both transgenic and non-transgenic host plants were then challenged with *P. aegyptiaca* and the success of parasitism was evaluated at different stages of development. While there was no reduction in the number of *P. aegyptiaca* plants attached to transgenic tomato roots, there was a decrease in their weights relative to non-transgenic tomato infections, particularly at early and late stages of parasite development. The transcript levels of all four genes were reduced to varying degrees in different tissues, most notably with a higher transcript reduction in tissue more distal to the host connection. In no case were *P. aegyptiaca* transcripts reduced more than 50%. The lack of effective resistance to *P. aegyptiaca* may result from an insufficient level of transcript silencing. Alternatively, more complicated dynamics may be at play involving dual RNAi systems associated with both the host and pathogens being plants.

INTRODUCTION:

Parasitic plants in the family Orobanchaceae are major agricultural pests, with species in the genera *Phelipanche*, *Orobanche*, and *Striga* being among the most damaging. *Striga* alone is estimated to cause one billion US dollars in crop losses annually [1]. Egyptian Broomrape (*Phelipanche aegyptiaca*) causes severe damage to a wide range of economically important crops in the families Solanaceae, Fabaceae, Brassicaceae, Cucurbitaceae, Apiaceae and Asteraceae [2]. *P. aegyptiaca* is particularly limiting in processing tomato production [3]. Currently there are few methods to control these parasites with those available often being costly and resource intensive [4].

The intimate association between hosts and parasites make them particularly difficult to control [5]. In addition, parasitic plants produce copious amounts of tiny, long-lived seeds that lie dormant in the soil until the exudate of a host root triggers germination. The emerging radicle of the parasite then attaches to the host root via a structure known as a haustorium, which subsequently penetrates the host root and establishes a vascular connection to obtain the vast majority of its nutrients [6]. The exchange of materials through this graft-like junction allows the parasite to feed, but also enables the transfer of host-produced molecules that can selectively harm the connected parasite [7]. In recent years, the prospect of genetically engineering an RNAi-mediated defense response in host plants that could be transmitted through the haustorium has been proposed [8–12].

RNA interference (RNAi) is a eukaryotic system used for silencing endogenous gene expression and that of intimately interacting organisms. This process begins with the creation of double stranded RNA that is recognized by Dicer-like enzymes and cleaved into 21-24 nt small interfering RNA (siRNA). These small RNAs then act as guides for the Argonaute proteins to

silence matching target genes. This silencing can occur at post-transcriptional, translational, or epigenetic stages of expression depending on the size of the siRNA and the specific proteins in which it associates [13]. This process has important roles in fighting viral infections and is used by both eukaryotic hosts and pathogens to manipulate one another [13]. In the parasitic plant *Cuscuta campestris*, microRNAs (miRNAs), a type of siRNA, are transferred to the host *Arabidopsis* and increase its susceptibility to the parasite [14].

RNAi can also be utilized for host resistance with a process known as Host Induced Gene Silencing (HIGS). HIGS is accomplished by engineering a host to produce RNAi silencing molecules. When the pathogen begins to acquire host resources, the silencing molecules travel to the pathogen and reduce the expression of targeted genes. This method has successfully enabled resistance to viruses, bacteria, nematodes, fungi, oomycetes, and insects through the selective silencing of parasite transcripts using host produced siRNAs [15]. This technique has been explored in parasitic plants but has yet to yield a high level of resistance, Table 1.

Early studies using the non-weedy parasite *Triphysaria versicolor* showed that RNA silencing molecules can move bi-dirctionally across the haustorial bridge from host to parasite and reduce targeted transcripts [8]. Later work with the agriculturally damaging parasites *Cuscuta pentagona* and *Phelipanche aegyptiaca* have shown some success in suppressing parasite attack but not to a level meaningful for agriculture [10–12,17]. These attempts have largely overlapping methods in targeting parasite metabolic genes and generating siRNA by expressing hpRNA or TRV sequences driven by constitutive promoters.

The 35S promoter was used in all constitutive examples except for *Striga* where another viral constitutive promoter, CMPS, was used [16]. The HIGS targeting *Cuscuta* was the only example that used a tissue specific promoter from AtSUC2, expressed in the phloem [17]. In

most cases, the siRNA was generated using short hairpin RNAs (hpRNAs) expressed in stably transformed hosts. In two examples, Agro-infiltration was used to transiently express a TRV sequence joined with parasite sequence to produce parasite specific siRNAs [11,12,17]. Both transformation systems and siRNA production methods yielded similar levels of parasite transcript silencing which in some cases lead to a degree of host resistance.

The genes targeted for silencing in all parasites except *Cuscuta pentagona* are enzymatic genes important for normal metabolism. In *C. pentagona*, two KNOX1 transcription factors were targeted to better understand the molecular mechanisms involved in haustorium development and host interactions. Unsurprisingly, when haustorium development was interrupted there was a reduction in parasite vigor that conferred a low level of host resistance [17].

In this study, we targeted four parasite genes with crucial roles in lipid biosynthesis to generate host resistance. Lipids are major structural and functional molecules in living organisms that form the foundation for cell membranes, energy storage, and many other diverse roles [18]. Reducing the expression of lipid metabolic genes can be lethal to *Arabidopsis* and even parasitic plants [9,19]. To enable crop resistance to *P. aegyptiaca*, we choose to silence four parasite genes (ACC1, LCB1, ATS2, and MCMT) that exhibit lethal phenotypes when reduced in *Arabidopsis* embryos [19–23] and *Medicago truncatula* roots. Further information on the functions of these lipid biosynthesis genes can be found in Table S1. Our results silencing these four genes in *P. aegyptiaca* highlight the complexity involved in trans-specific silencing.

MATERIALS AND METHODS

RNAi target genes

P. aegyptiaca transcript sequences that are orthologous to the RNAi targets were identified from PPGP transcriptomic data using the *Arabidopsis* mRNA sequence (ACC1: NM_103313.4, LCB1: NM_119811.3, ATS2: NM_119204.4, MCMT: NM_128573.4) and tBLASTx in the program Geneious (version R8)[25]. The *Medicago truncatula* RNAi target orthologs were identified by Pradeepa CG Bandaranayake using the *Arabidopsis* mRNA sequence and NCBI BLAST. For both *P. aegyptiaca* and *M. truncatula*, transcripts with the lowest e-value were selected as the most likely orthologs of the *Arabidopsis* genes.

RNAi Vector Construction

To evaluate the necessity of candidate genes for root development, roughly 500 base pairs of *M. truncatula* transcripts were cloned into the vector pHellsGate8-YFP in an inverted repeat conformation using a Gateway cloning strategy by Pradeepa CG Bandaranayake[24,40]. To create the *P. aegyptiaca* silencing construct, I assembled four roughly 200 bp *P. aegyptiaca* transcript sequences into one fragment using overlapping PCR. During the first reaction, primers add sequences to each of the four fragments which overlap the fragment planned to be adjacent. These four products flanked with overlapping sequences ran in a PCR for 20 cycles to build the single 4X fragment. At which point, two primers flanking the entire 4X fragment were added that built Gateway cloning sites attB1 and attB2 onto the fragment ends, and ran for an additional 15 cycles. This 4X fragment was then cloned into the destination vector pHellsGate8-YFP in an inverted repeat conformation using Gateway cloning, Figure S5. The sequences for all primers used in this study can be found in Table S2.

Plant Transformation

The *Medicago truncatula* silencing vectors were transformed into *M. truncatula* roots using *Agrobacterium rhizogenes* strain MSU440 according to a previous method by Pradeepa CG Bandaranayake[9]. The *P. aegyptiaca* silencing vectors were transformed into the tomato line T5 (*Solanum lycopersicum*) using *Agrobacterium tumefaciens* strain EHA105 at the UC Davis Transformation Facility. For each independently transformed T0 tomato line, I self-pollinated and grew out progeny to the T2 generation. To identify progeny for each independent transformant that were homozygous for the transgene, I selected those T2 progeny that all expressed YFP. Three lines with the highest YFP expression were selected for further analyses.

Polyethylene Bag Assay

Due to restrictions on working with *P. aegyptiaca* in California, the parasite was grown by the lab of Jim Westwood at Virginia Polytechnic Institute and State University. RNA samples along with data on parasite numbers and weight, and host weight were generated by the Westwood lab and sent to the Yoder lab for further investigation.

P. aegyptiaca seeds were surface-disinfested in 70 % ethanol for 40 seconds followed by immersion in a solution of 1% sodium hypochlorite with two drops of Silwet for 4 minutes. Seeds were rinsed in 0.01 M HCl solution for 4 minutes, then rinsed in sterile distilled water 4 times for 4 minutes each rinse, and finally air-dried in a laminar flow hood for 6 hr. The cleaned seeds were conditioned by spreading them evenly on sterile glass-microfiber sheets that had been moistened with sterile distilled water, in Petri dishes that were then held in the dark at 23 C for 7 d. The seeds were stimulated to germinate by blotting them to remove excess moisture, then adding filter-sterilized racemic GR24 (2 mg/L; courtesy B. Zwanenburg). After 24 h, the seeds were ready to use for inoculation. Tomato seeds were sown in potting media and grown in a

growth chamber for two weeks. Tomato seedlings were then carefully removed from their pots and roots were rinsed in tap water until free of medium. The cleaned seedlings were transferred to polyethylene bags containing moist glass-microfiber sheets and laid flat under high humidity to recover. The next day, the bags were suspended in a box to hang vertically and were covered to keep roots dark, while tomato shoots were exposed to light. Bags were kept wet initially by the addition of 10 ml of 1/8 x Hoagland solution, followed by adding 1/4 x Hoagland solution each day as needed to replenish liquid. After 10 d of growing under short day conditions at 25 C, tomato plants were ready for inoculation. The *P. aegyptiaca* seeds described above were transferred to the tomato roots using a fine brush and working under a dissecting microscope. Plants were watered by adding 1/4 x Hoagland solution each day as needed.

Soil Pot Assay

To assay the long term effects of RNAi on parasite success, the tomato plants with *P. aegyptiaca* tubercles from the polyethylene bag experiment were carefully transplanted into pots with potting medium. Plants were grown at 25 C until harvest 29 d after transplanting, at which time *P. aegyptiaca* floral shoots had emerged in most pots. The tomato plants were removed from the pots, roots rinsed of potting medium, and *P. aegyptiaca* plants carefully dissected away from hosts. Fresh weights were determined for all plants, and specific tissues were frozen in liquid N2 for subsequent RNA extraction.

The key stages of *P. aegyptiaca* development analyzed included the tubercle stage (underground purely vegetative tissue), pre-emerged (a tubercle containing a floral shoot that had not yet emerged above the soil surface), emerged (tubercle containing a floral shoot that has emerged above the soil surface) and flowering (tubercle containing a floral shoot that has emerged and at least one flower has opened). For the gene expression study, tissues were taken from flowering

parasites and divided into the basal part consisting of the tubercle adjacent to the host root, and the flower, consisting of the top 3 cm of the floral shoot.

RNA Isolation

RNA was isolated using the Qiagen RNeasy plant mini kit following manufacturer's instructions, but slightly modified for small tissue samples. *P. aegyptiaca* tubercles (0.1 g per extraction) were ground in a 2 ml glass micro tissue grinder on ice. For initial extraction, 550 ul RLT buffer was used.

Quantitative PCR (qPCR)

RNA was reverse transcribed into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems™,Cat.# 4368814). The cDNA (diluted 15-20 times) was combined with gene specific primers (300 nM) and a SYBR based master mix (ABsolute Blue SYBR Green QPCR Mix, Thermo Scientific™, Cat.# AB4323A) in 96 well plates (MicroAmp™ Optical 96-Well Reaction Plate, Thermo Scientific™, 4306737) and run in an Applied Biosystems 7300 Real-Time PCR Machine. The program used was: 1 cycle 95°C; 40 cycles of 95°C for 15 seconds, 60°C for 30 seconds, and 72°C for 30 seconds; and lastly a dissociation step transitioning from 95°C to 60°C to 95°C.

The Ct values for the transcripts of each gene were obtained from the mean of three technical replicates. The delta-delta Ct method ($2^{-\Delta\Delta Ct}$) was used to calculate the levels of target gene transcripts relative to the geometric mean of two internal reference genes, TUB1 and UBQ1. These reference genes were stability expressed across all tissue types used in the PPGP transcriptomic data and validated as stable reference genes using the RNA samples in Figure 4 and the programs Bestkeeper and Normfinder[41,42]. All qPCR primers used exhibited a single

dissociation peak and a linear amplification efficiency > 90% and <110% using a 10-fold dilution series of cDNA.

Statistical tests:

To determine the statistical significance of the transcript and biomass differences, student t-tests were performed using Microsoft Excel with two tails and unequal variance selected. The subsequent p-values were corrected for multiple testing errors using the Benjamini & Hochberg method in the “p.adjust” R package. The biological replicates for each experiment are detailed in Table S3.

RESULTS:

RNAi target genes

Twelve genes with embryo lethal phenotypes in *Arabidopsis thaliana* were identified by Pradeepa CG Bandaranayake as candidate HIGS targets in *P. aegyptiaca*. To test if these genes have vital functions in roots, she silenced orthologous *Medicago truncatula* transcripts in transformed roots. The degree of lethality was determined by the lack of transgenic roots that emerged relative to the empty vector transformed plants. When tested by a previous postdoctoral scholar in our lab, four lipid biosynthesis genes were among the most suppressive of transgenic root emergence which suggests they have important roles in root development.

Parasite silencing construct

The orthologous target gene sequences for *P. aegyptiaca* were identified from the Parasitic Plant Genome Project's (PPGP) transcriptomic data using the *Arabidopsis thaliana* mRNA sequences [25]. The expression patterns from the PPGP data show the four target genes expressed across all stages of parasite development, Figure S1. This suggests the function of these genes is needed throughout development which is consistent with their known roles as critical lipid biosynthesis genes. The *P. aegyptiaca* transcripts were aligned with orthologous *S. lycopersicum* sequences and parasite RNAi targets were identified in regions of high dissimilarity to lower the chance of unintentionally targeting tomato transcripts, Figure S2. I selected roughly 200 base pairs of 3' UTR (Untranslated Region) for ACC1, LCB1, and ATS2 and 5' UTR extending into the coding region for MCMT to generate the parasite specific siRNAs. Alignments of the target *P. aegyptiaca* sequences with the orthologous tomato sequence show the high level of dissimilarity in this region, Figure S3. These four roughly 200 base pair parasite sequences were linked together into a single fragment and cloned in an inverted repeat

conformation into the vector pHellsGate8-YFP, Figure S4. This vector was then transformed into tomato line T5 using *Agrobacterium tumefaciens*. Three independently transformed lines at the T2 generation with high YFP expression were chosen for subsequent analysis. These lines had no obvious morphological differences compared to the non-transformed line and had roughly similar weights after parasitism, Figure S5.

Parasite biomass reduced at early and late stages of development

To evaluate the effectiveness of our HIGS-based resistance, *P. aegyptiaca* seeds were inoculated onto the roots of three hpRNA expressing tomato lines (RNAi1-3) and the non-transformed control (NT) in a polyethylene bag system [26]. No obvious differences were observed for attachment success and survival of germinated *P. aegyptiaca* plants. To check if the RNAi expressing tomatoes had increased resistance to *P. aegyptiaca* at later stages of development, we transferred the infected tomatoes into soil pots. When *P. aegyptiaca* plants began to flower in most tomato pots, the parasites attached to each host were harvested, weighed, and separated into four developmental stages: tubercle, pre-emerged, emerged, and flowering. Further details describing the criteria for each developmental stage can be found in the Methods section. A reduction in the average parasite weight occurred during the tubercle and flowering stages of development for lines RNAi-1 and RNAi-2, Figure 1. Reductions in parasite biomass were observed at the emerged and flowering stages of development for line RNAi-3, however, these values were not significant ($p > 0.1$). There were no significant differences in the numbers of parasites at all developmental stages growing on RNAi or NT tomato lines, Figure S5.

Parasite tubercle transcripts are reduced

The transcript levels for *P. aegyptiaca* genes were measured using quantitative reverse-transcription PCR (qPCR). The *P. aegyptiaca* transcripts ATS2 and MCMT were reduced in

tubercles attached to RNAi tomato lines compared with the non-transformed controls by 19-52% and 46-50% respectively, Figure 2. ACC1 was only slightly reduced while LCB1 had no obvious reduction. The level of silencing in ATS2 and MCMT is comparable to previously reported HIGS-based transcript reductions in parasitic plants [9–11]. The lack of statistical significance in the transcript reduction of MCMT is likely due to the high variability observed in the control plants.

Target transcript reduction is higher in tissue more distal from siRNA source

To determine if parasite transcripts are silenced at later stages of development and how silencing changes with increased distance from the siRNA source, we collected RNA from parasite root and floral tissues from flowering *P. aegyptiaca* plants attached to RNAi line 2 or non-transformed tomatoes. Transcripts of all four target genes showed a pattern of reduction in floral tissue with ACC1 and LCB1 being statistically significant ($p < 0.05$), Figure 3. This is in contrast to the tubercle data where ACC1 and LCB1 showed little to no reduction. These results show the silencing of specific genes can be influenced by the parasite's tissue type. Intriguingly, when the levels of silencing were compared between the floral and root tissues of flowering parasites, the genes LCB1, ATS2, and MCMT were reduced more in floral tissues, Figure 3. The higher transcript reduction in floral issue compared to roots was only significant ($p < 0.05$) for one gene while the pattern still seems to be present for the other two genes. This shows the silencing of specific genes is not uniform across the parasite and can be higher in tissue farther from the siRNA source.

DISCUSSION:

Obtaining host resistance against the parasitic weed *P. aegyptiaca* has been a challenge as traditional breeding has returned limited gains [27]. In our experiment, the HIGS-based approach reduces the steady-state level of targeted parasite transcripts in a tissue dependent manner. At the tubercle stage of development, the transcripts of two genes (ATS2 and MCMT) were reduced. At the flowering stage, all four genes were reduced but only two genes (ACC1 and LCB1) had statically significant decreases. In roots, only one gene (ACC1) was reduced significantly. The different sets of genes silenced between *P. aegyptiaca* tubercle, floral, and root tissues suggests these tissues may have different levels of susceptibility to the silencing of specific genes. The tissue-specific reductions observed were unlikely due to large differences in the basal transcript levels as the expression patterns across non-transformed control tissues do not match the patterns in silencing that was observed, Figure S6. A lack of RNAi machinery needed to process siRNA and silence transcripts does not seem to be a limitation for our system as these components are expressed at all stages of parasite development, Figure S7.

The higher degree of silencing in floral tissue, which is up to 25 cms away from the source of siRNA at the host connection compared to root tissues, could result from a more effective transport of siRNA to the floral region due to differences in sink strengths. Alternatively, siRNA unloading from vascular tissues may be more effective in flowers. This pattern of higher silencing in parasite tissue more distal to the point of host attachment was also observed for HIGS targeting STM in *Cuscuta* tissues harvested at the host connection and 15 cm away [17]. *Cuscuta* is a stem parasitic plant that evolved a parasitic lifestyle separately from *Phelipanche*. The consistency in silencing patterns between these independent origins of plant parasitism may highlight a similar mechanism behind this counterintuitive result.

The effects on parasite biomass were developmentally specific. I observed lower average biomass at the earliest and latest developmental stages. Similar developmental specificity has also been observed in multiple cases for insects [28,29]. In these cases, mortality was higher for younger insects feeding on siRNA-expressing plants than those at the adult stage. The stage-specific effects observed may result from the targeted lipid metabolic genes being needed at different levels during parasite development.

In our experiments, RNAi based host resistance to parasitic plants was not obtained as shown by the lack of differences in host weights and the numbers of attached parasites between RNAi and control tomato lines. The level of transcript silencing obtained through our HIGS approach was likely insufficient to suppress the parasite's growth. This may have been due to our hairpin construct not being expressed at a high enough level or an aspect of the parasite's biology limiting the silencing effect. It would be beneficial in future HIGS studies to measure the level of hairpin expression in the host to verify it is occurring at a high level.

RNAi has many known roles in plant pathogen interactions [13]. When a viral infection occurs, viral RNAs detected by the host are used to generate siRNAs that target the degradation of viral RNA throughout the plant. Viruses have evolved an ability to overcome this defense response with proteins known as Suppressors of Gene Silencing that interfere with the host's RNAi mechanism. This interaction gets more complicated when pathogens are eukaryotic because both players have RNAi systems to regulate the expression of their genes as well as those of other interacting organisms. For example, miRNAs are used by both eukaryotic hosts and pathogens to manipulate each other for defense or invasion, respectively [13,15]. The parasitic plant *Cuscuta campestris* transfers miRNAs to the host *Arabidopsis* that increase its susceptibility to the parasite [14]. In other cases, pathogens selectively down regulate the host's

RNAi system to lower its resistance [13]. When components of the host's RNAi system are mutated, these plants are typically more susceptible to pathogens. In some cases for eukaryotic pathogens, when this host RNAi machinery is mutated, specifically DCL1 and AGO1, the trend is reversed and it increases host resistance [13]. The varying roles host RNAi components play during specific plant-pathogen interactions highlights the complexity of these trans-eukaryotic interactions. More information on how this pathway is expressed and affected in specific host-parasitic plant interactions could aid in the design of more effective HIGS.

The exchange of RNA between hosts and parasitic plants may be regulated in ways that affect the efficiency of HIGS. The transfer of mRNA through the haustoria of *Cuscuta* resembles that of graft junctions where most transcripts move in a nonspecific manner. There are examples, however, of transcripts with much higher or lower rates of transmissibility and movement across *Cuscuta* haustoria than would be predicted by expression [7]. Plasmodesmata that connect companion cells with sieve elements are major regulators of movement to and from phloem vascular tissues and mRNA movement through plasmodesmata is regulated by specific proteins that bind nucleic acids and facilitate the exchange [35]. This regulation may similarly influence the transfer of RNAs between hosts and parasitic plants. The presence of specific sequences in the mRNA may also affect its stability during transport between the host and parasite [7]. Once it is understood how specific types of RNAs can transfer from the host and move through the parasite, these features could be mimicked to enhance the movement and function of artificial RNAs for more efficient RNAi based resistance.

The previous HIGS attempts in parasitic plants enable some resistance but not at a level practical for agriculture. The lack of effective host resistance may result from an insufficient level of parasite gene reduction. In most cases, parasite transcripts are not reduced below 20% of

normal expression and are typically reduced to more intermediate levels. The majority of parasitic plant HIGS studies use constitutive viral promoters to express the siRNA-generating sequence. It has been observed that these promoters, while considered constitutive, can show significant expression variation in different tissues [36]. It is possible that the 35S promoter is not appropriate for optimal siRNA delivery. The overexpression from a 35S promoter can also cause unintended pleiotropic effects on transgenic plants. In some cases, these issues were mitigated by the use of tissue specific promoters [37]. In *Cuscuta*, the use of a phloem-specific promoter from AtSUC2, to drive the hpRNA expression, facilitated a 100-fold higher flow of mRNA between the parasite and host then what is observed to move across graft junctions [38]. This suggests tissue-specific promoters can facilitate high levels of siRNA transfer from a host to a parasitic plant.

The choice to silence metabolic genes with lethal phenotypes when mutated or targeted by an herbicide, seems logical for HIGS but have not yielded effective resistance against parasitic plants. Parasites might compensate for a reduction in biosynthetic activity by acquiring metabolites from the host. Genes with vital functions that cannot be compensated by the host, such as transcription factors or other regulatory proteins, may be more effective RNAi targets. The choice of better target genes, use of tissue specific promoters, and understanding more about the exchange of RNA between hosts and parasites may aid in further reducing parasite transcript levels.

Other artificial methods for generating siRNA, such as amiRNAs and syn-tasiRNAs, use different parts of the RNAi pathway and show promise in enabling pathogen resistance through HIGS [39]. The differences in how these RNAs are processed and utilized compared to hpRNA may enable higher levels of siRNA production and transport into the parasite. The amiRNA and

syn-tasiRNA systems also offer higher specificity of the target sequence which is of particular importance when targeting parasitic plants where sequence similarities with the host are higher than other pathogens [39].

Inducing the production of secondary siRNA within the parasite could facilitate higher levels of transcript silencing because the siRNA signal is amplified when it interacts with a target. Amplification is minimal with traditional hpRNA constructs because the products are mainly 21 nt siRNAs [17]. In contrast, amiRNAs can be engineered to produce 22 nt fragments that trigger the production of secondary siRNA through phasing and the recruitment of RDR6 [44]. When constructs expressing 22nt-amiRNA, 21nt-amiRNA, and hpRNA were directly compared, the 22nt-amiRNA triggered more widespread silencing than the other two constructs [45]. Adapting this technology to trigger an amplification of siRNA within a parasite may yield a higher level of transcript reduction facilitating effective host resistance.

Shifting the population of siRNAs to 24 nts could be effective in triggering an epigenetic form of silencing through the RNA-directed DNA methylation (RdDM) pathway [13]. The enzyme Dicer-like 3 (DCL3) is known to produce 24 nt siRNAs from dsRNA. If DCL3 was co-expressed in the host with the production of dsRNA then the epigenetic silencing of targeted genes may be triggered. The success of this and other HIGS based resistance approaches rely on the pathogen expressing certain aspects of the silencing pathway. This does not seem to be a limitation for *P. aegyptiaca* because transcriptomic data shows that all major enzymes in the RNAi pathway are expressed across development, Figure S7 [25]. Combining multiple RNAi based approaches may maximize the reduction in parasite transcripts and make it more difficult for the parasite to overcome the resistance.

Combining the RNAi strategy with other control methods will be critical in minimizing the parasite's ability to overcome the host resistance. The use of single host resistance genes in sunflower quickly breaks down as it creates high selection pressure on broomrape with the ability to overcome this single target [46]. The use of multiple control methods should prolong the efficacy of HIGS based resistance as it would be more difficult for any individual parasite to overcome multiple simultaneous controls.

Developing host resistance to parasitic plants is a paramount concern as few control methods exist. HIGS has been attempted many times against parasitic plants without generating strong host resistance. The lack of resistance obtained, compared to a similar level of transcript reduction in other systems, suggests parasitic plants may be more difficult to target using this system. The tissue specific silencing and stage specific biomass reductions observed highlight the complex dynamics involved in cross eukaryotic RNA transfer. Newer RNAi mechanisms such as amiRNAs and syn-tasiRNAs have potential to increase parasite transcript reduction and facilitate effective host resistance to parasitic plants.

FIGURES:

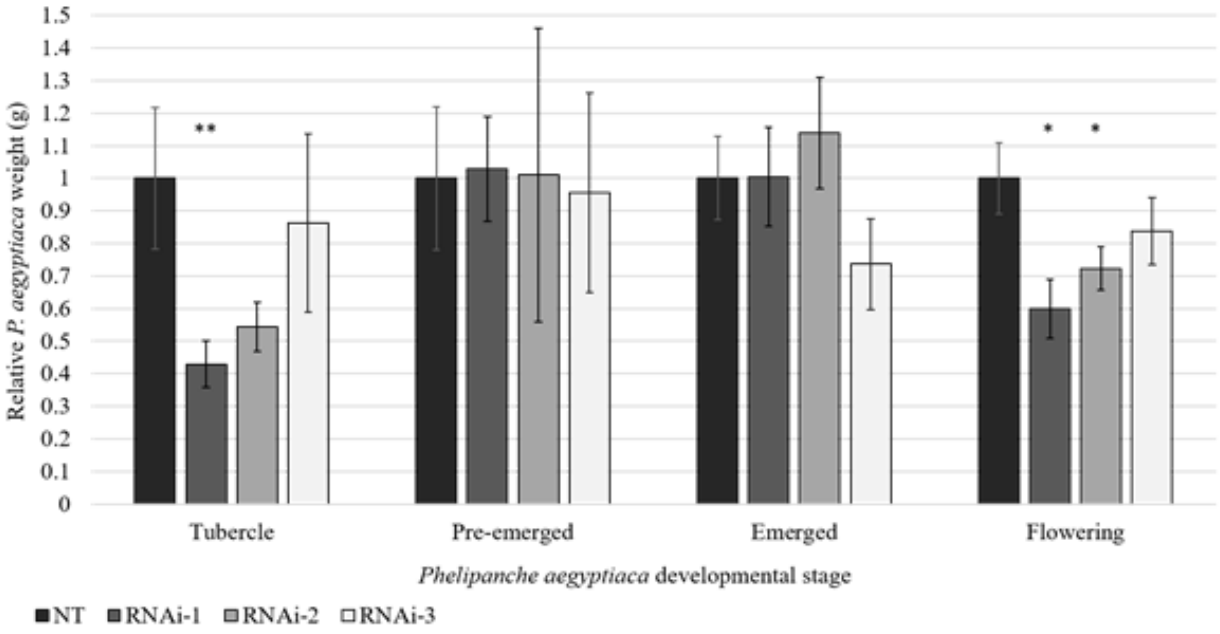


Figure 1: Stage specific growth reduction of *P. aegyptiaca*: Biomass data for four stages of parasite development growing on three independently transformed RNAi tomato lines (RNAi1-3) or non-transformed (NT) control tomato plants. A reduction in *P. aegyptiaca* biomass for tomato lines RNAi-1 and RNAi-2 during early (tubercle) and late (flowering) stages of development was observed. Data represent the mean +/- the standard error for $4 \leq n \leq 19$ samples. Values are relative to the non-transformed (NT) control plants. ** indicates p-value < 0.05, * indicates p-value ≤ 0.1.

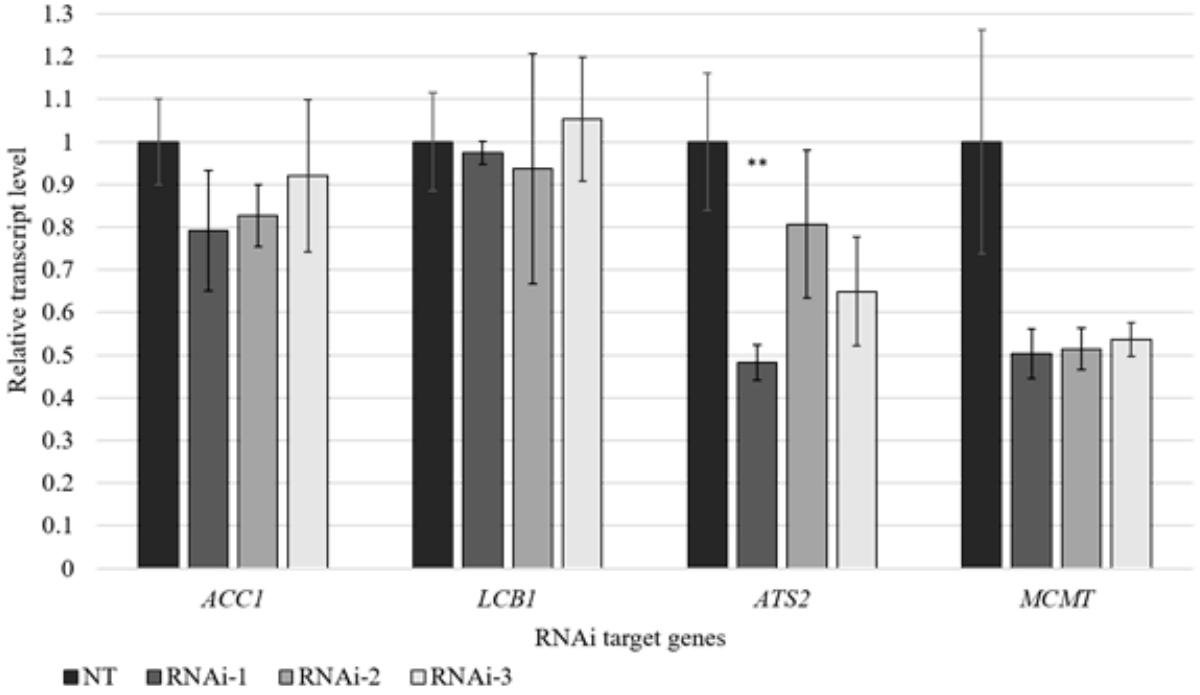


Figure 2: Reduction of target gene transcripts in *P. aegyptiaca* tubercle tissues: Transcript levels of ACC1, LCB1, ATS2, and MCMT in *P. aegyptiaca* tubercles growing on three independently transformed RNAi tomato lines (RNAi1-3) or non-transformed (NT) control tomato plants. Transcript levels were determined using qPCR. Data represent the mean +/- the standard error for $2 \leq n \leq 8$ samples. Transcript levels are relative to the non-transformed control plants. ** indicates p-value ≤ 0.05.

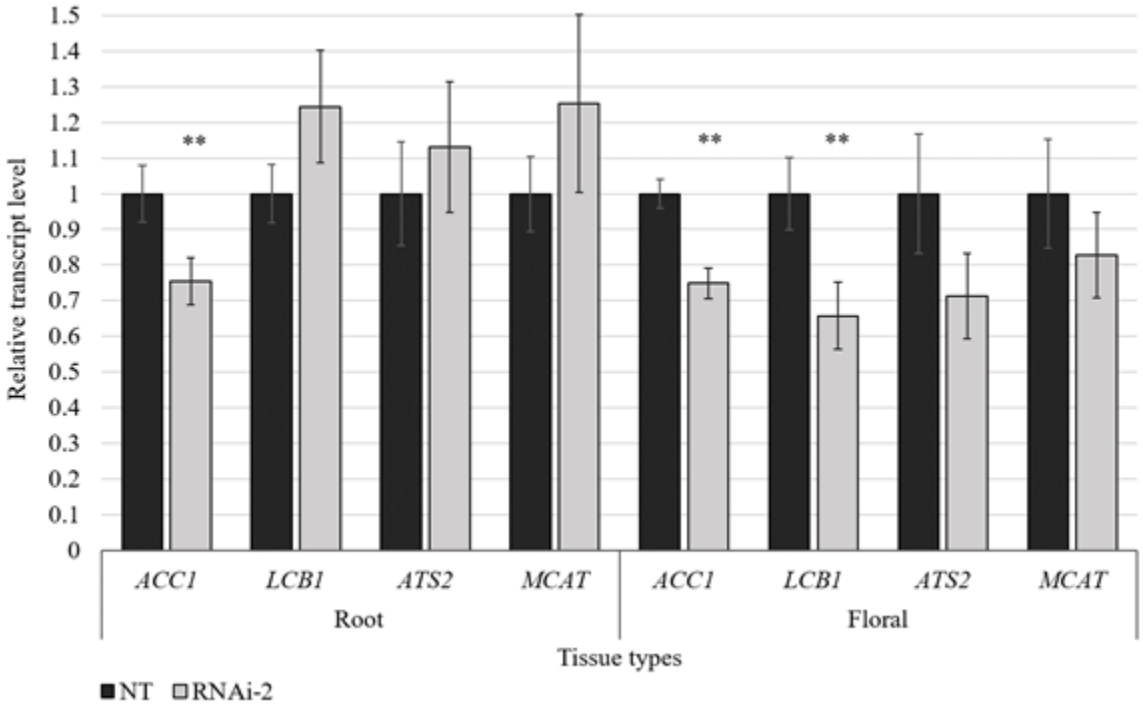


Figure 3: Transcript reduction is higher in tissue more distal to the siRNA source: Transcript levels of ACC1, LCB1, ATS2, and MCAT in *P. aegyptiaca* root and floral tissues growing on tomato line RNAi-2 and non-transformed (NT) tomatoes. Transcript levels were determined using qPCR. Data represent the mean +/- the standard error for n = 7 samples. Transcript levels are relative to the corresponding non-transformed control plant tissue. ** indicates p-value < 0.05.

TABLES:

Parasitic plant species	Host species	Transformation method	RNAi approach	Target gene(s)	Effect on parasite	Ref.
Striga asiatica	*Zea mays*	Not reported	Hairpin	*EPSP, CTase, ENR, VCL1, AdSS*	Little to no reduction in growth	[16]
Triphysaria versicolor	*Lactuca sativa, Arabidopsis thaliana*	*Agrobacterium rhizogenes*	Hairpin	*GUS*	Reduced exogenous GUS expression	[8]
Triphysaria versicolor	*Medicago truncatula*	*Agrobacterium rhizogenes*	Hairpin	*ACC1*	Root viability reduced	[9]
Cuscuta pentagona	*Nicotiana tabacum*	*Agrobacterium tumefaciens*	Hairpin	*STM, KNAT1-3*	Overall reduction in vigor	[17]
Phelipanche aegyptiaca	*Solanum lycopersicum*	*Agrobacterium tumefaciens*	Hairpin	*M6PR*	Reduction in tubercle viability	[10]
Phelipanche aegyptiaca	*Nicotiana benthamiana*	*Agrobacterium tumefaciens* infiltration	TRV sequence	*CCD7, CCD8*	Reduction in number of tubercles	[11]
Phelipanche aegyptiaca	*Solanum lycopersicum, Nicotiana benthamiana*	*Agrobacterium tumefaciens*, whole plant and infiltration	Hairpin, TRV sequence	*M6PR, ACS, PRX*	Reduction in number and weight of tubercles	[12]
Phelipanche aegyptiaca	*Solanum lycopersicum*	*Agrobacterium tumefaciens* infiltration	Hairpin	*M6PR, CWI, SUS1*	Lower level of reducing sugars in parasite shoots	[43]
Phelipanche aegyptiaca	*Solanum lycopersicum*	*Agrobacterium tumefaciens*	Hairpin	*ACC1, LCB1, ATS2, MCMT*	Stage specific reduction in weight	This Paper

Table 1: Publications on HIGS in parasitic plants: Listed in chronological order.

SUPPLEMENTAL:

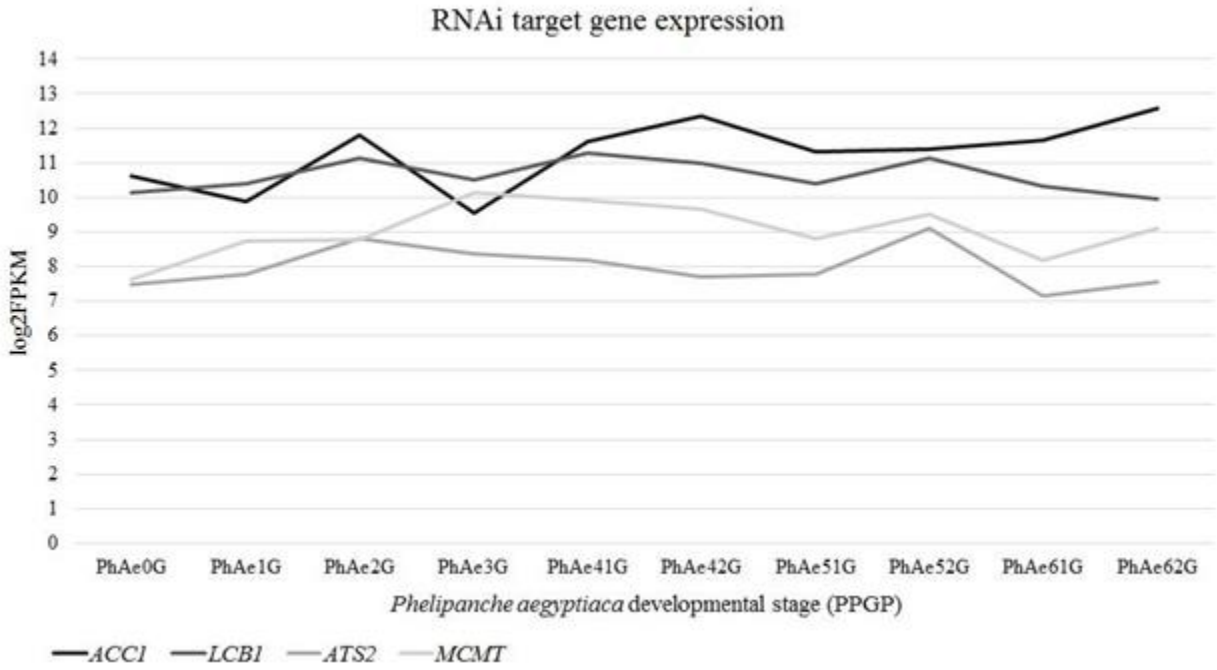


Figure S1: Expression of RNAi target gene transcripts in *P. aegyptiaca* across stages of parasite development. RNASeq data for *P. aegyptiaca* was obtained from the Parasitic Plant Genome Project (PPGP) [22]. Log2FPKM expression values show RNAi target transcripts expressed across all stages of *P. aegyptiaca* development.

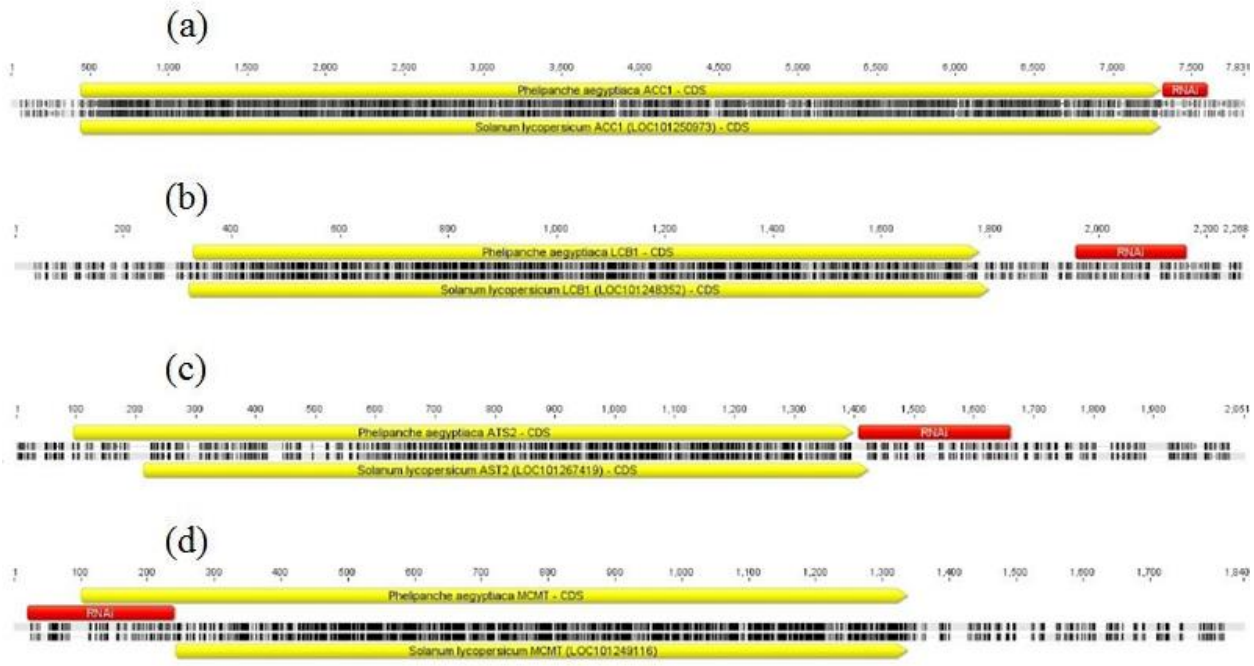


Figure S2: Alignments of *Phelipanche aegyptiaca* and *Solanum lycopersicum* orthologs of RNAi target gene transcripts. (a) *ACC1*, (b) *LCB1*, (c) *ATS2*, (d) *MCMT*. The yellow annotations show the predicted coding sequences (CDS) for each transcript and the red annotations show the location of the *P. aegyptiaca* sequences used to generate parasite specific siRNA. Nucleotides highlighted as black lines represent sequence similarity between the two species. The regions chosen for siRNA production have low levels of similarity relative to the CDS as seen by the higher frequency of black lines in the CDS region. Sequences were obtained from NCBI and the PPGP (Parasitic Plant Genome Project). Alignments were made using the program Geneious (R8).

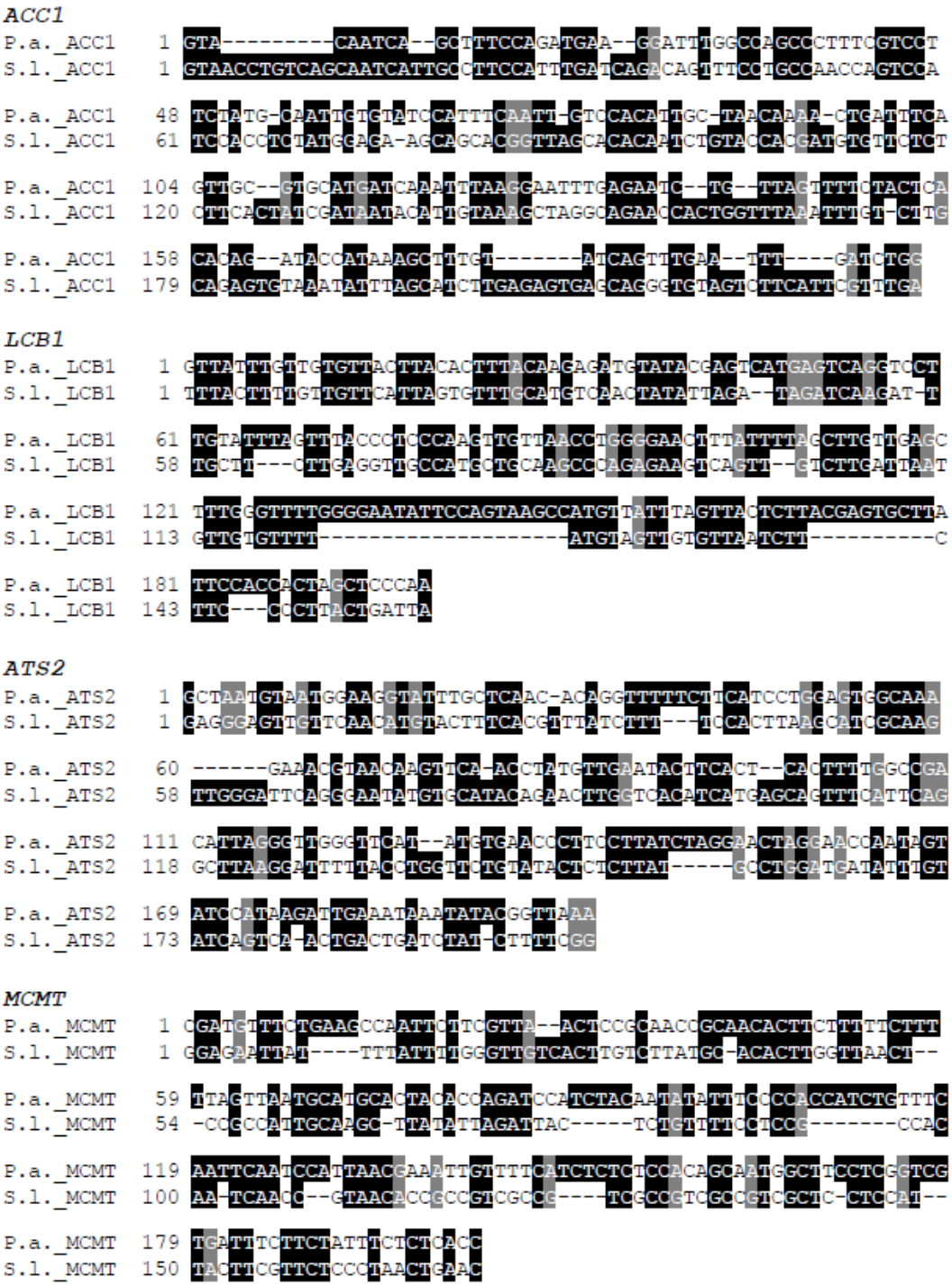


Figure S3: Alignments of *P. aegyptiaca* and *S. lycopersicum* transcript sequences in the regions used to specifically target the parasite. Nucleotides highlighted in black on both sequences at the same position are similar between the two species. The alignments show little similarity between the host and parasite in these regions. The alignments were created using the T-Coffee web server [Di Tommaso, *et al.* 2011] and visualized using the program BOXSHADE (ver. 3.2)

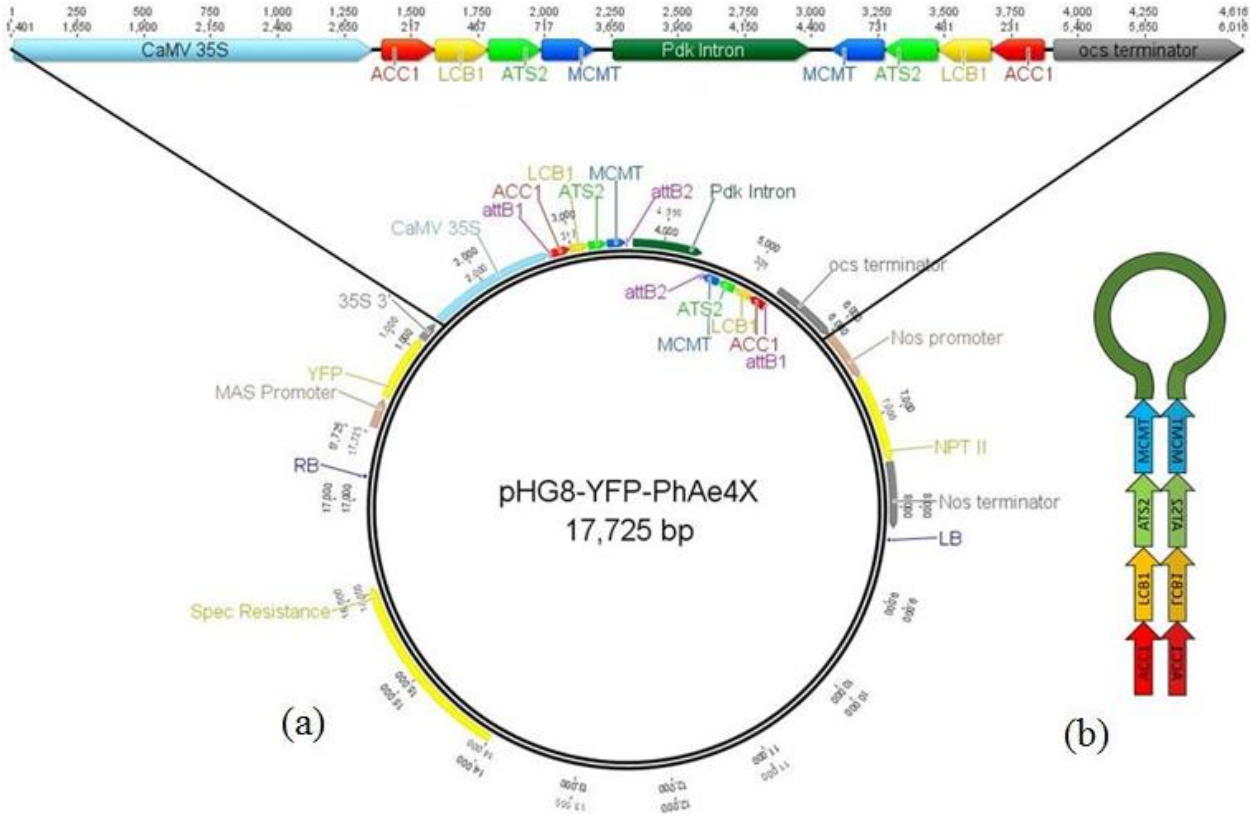


Figure S4: pHG8-YFP-4X vector map and hpRNA diagram. (a) The construct pHG8-YFP was used to produce parasite specific siRNA with an inverted repeat sequence containing *P. aegyptiaca* mRNA for four lipid biosynthetic target genes (*ACC1*, *LCB1*, *ATS2*, and *MCMT*). The diagram was created in the program Geneious (R8). (**b**) When the inverted repeat is transcribed, the complementary regions fold over to form a double stranded hairpin RNA (hpRNA).

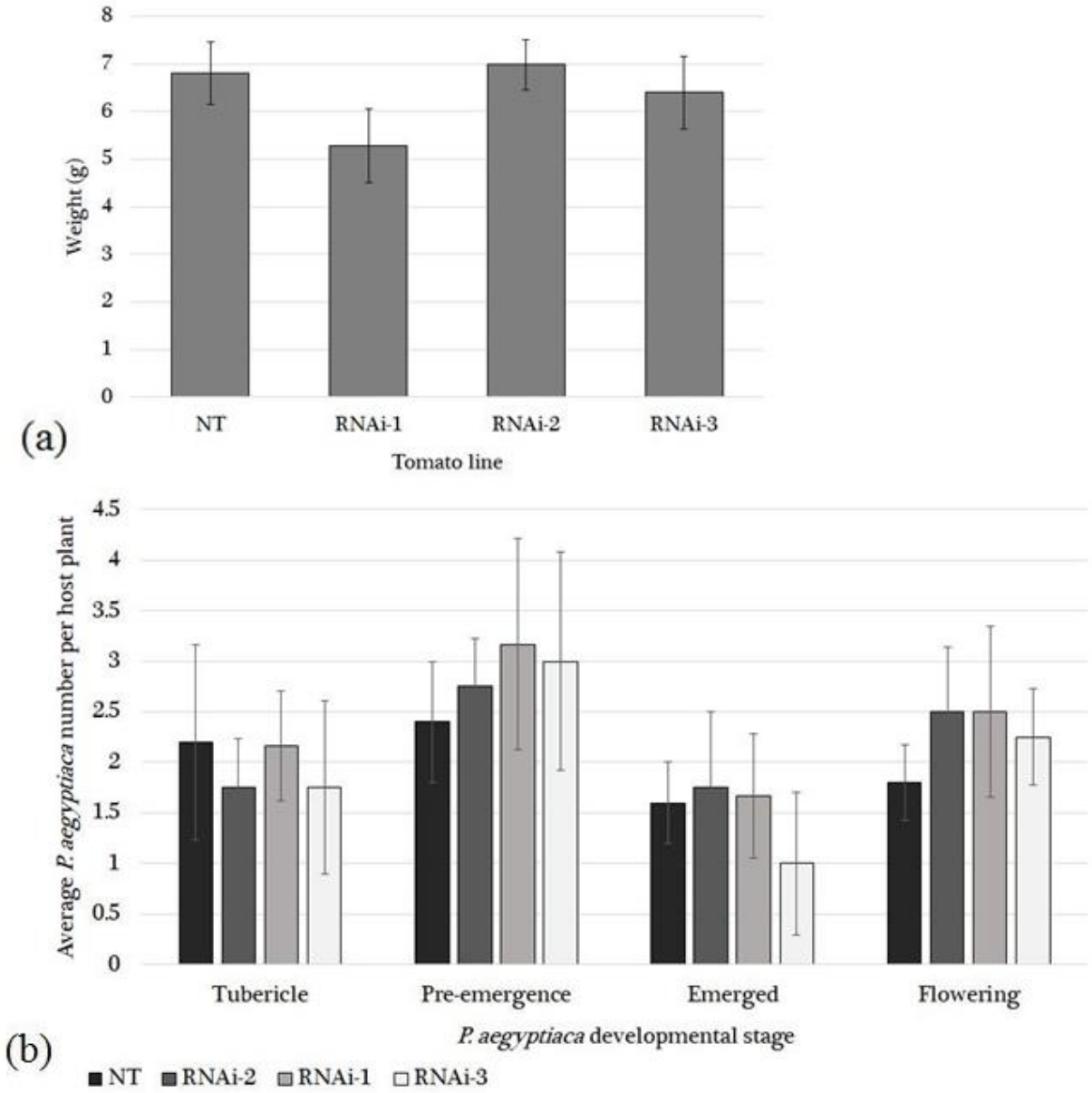


Figure S5: Host weight after parasitism and average number of parasites at each developmental stage for the four tomato lines. (a) The average host weight for the three RNAi tomato lines and the non-transformed control after parasitism. (b) The average number of *P. aegyptiaca* plants at each developmental stage growing on three RNAi tomato lines or the non-transformed control. Both sets of data show no significant differences (p-value < 0.05) between the RNAi and control tomato lines The data represents the mean +/- the standard error.

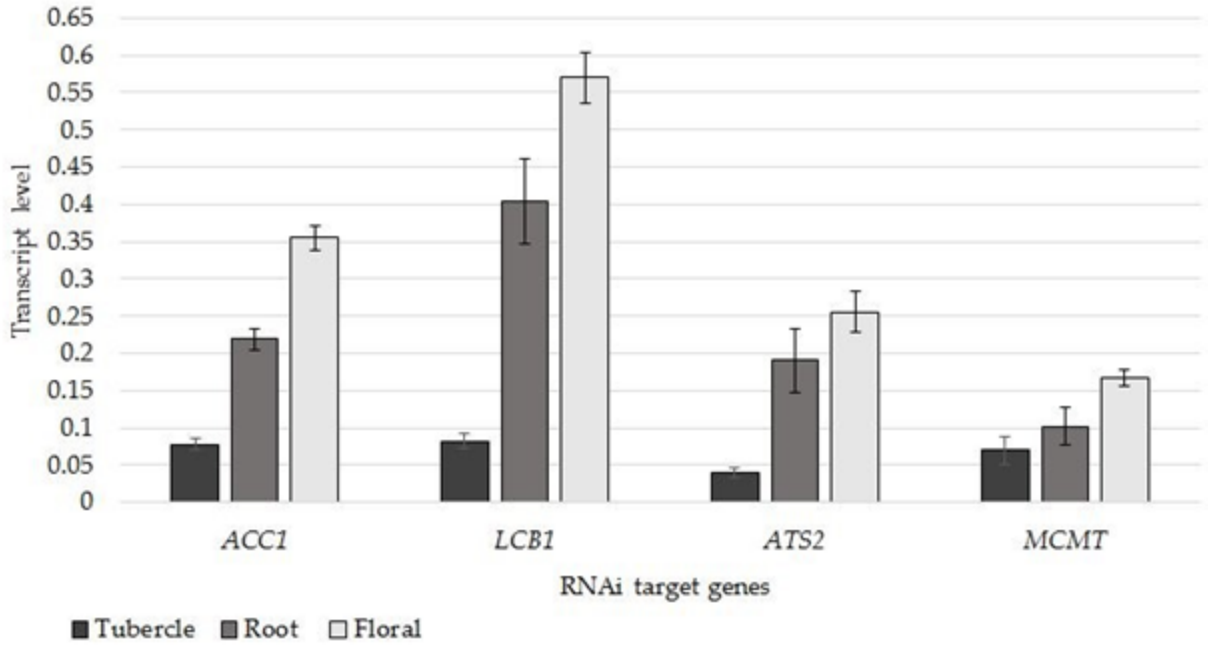


Figure S6: Transcript levels of four target genes in non-transformed controls. The transcript levels for each RNAi target gene were determined using qPCR relative to two internal reference genes, *TUB1* and *UBQ1*. Data represent the mean +/- the standard error.

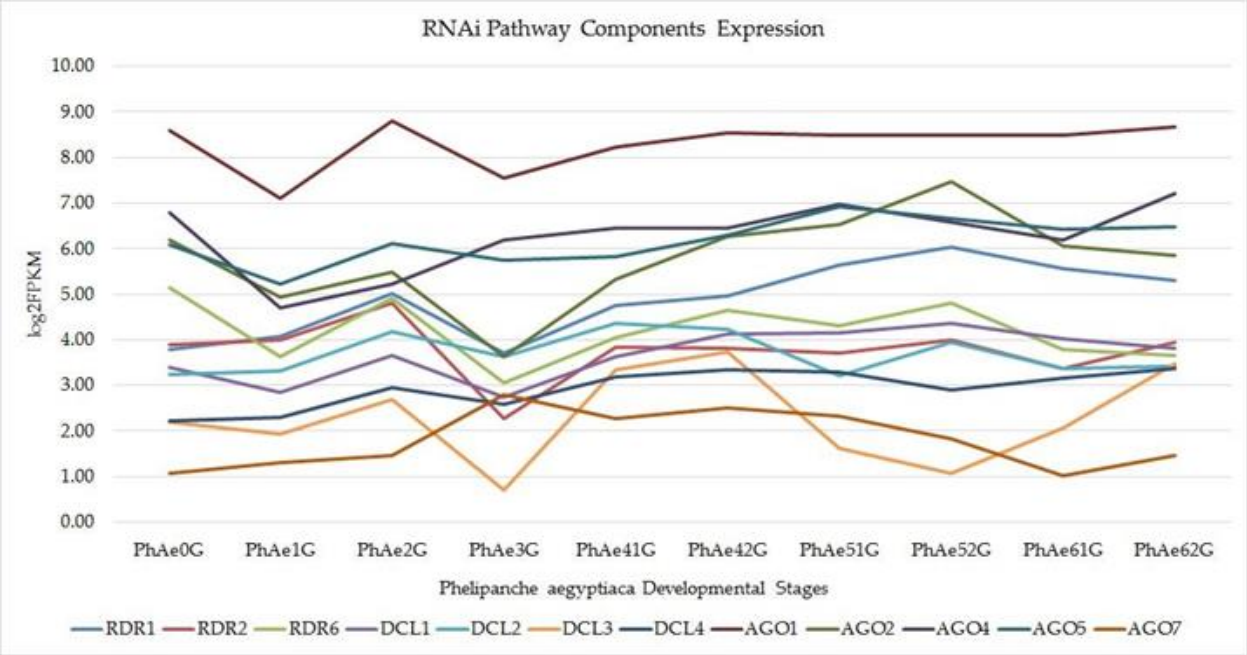


Figure S7: Expression of RNAi pathway components in *P. aegyptiaca* across stages of parasite development. RNASeq data for *P. aegyptiaca* was obtained from the Parasitic Plant Genome Project (PPGP) [22]. Log2FPKM expression values show the main RNAi components expressed across all stages of *P. aegyptiaca* development. The *P. aegyptiaca* orthologs for these genes were identified using BLASTn and the *Arabidopsis thaliana* mRNA sequence in the program Geneious (R8).

Table S1: Parasitic plant genes targeted by HIGS

Gene:	Full name:	Biological process:	Activity (Description from UniProt unless otherwise noted):
ACC1	Acetyl-CoA Carboxylase 1	Lipid biosynthesis	Catalyses the ATP-dependent carboxylation of acetyl-CoA to form malonyl-CoA in the cytosol which is a precursor for flavonoid biosynthesis and the elongation of very long chain fatty acids (Alban, Job, and Douce 2000; Roesler, Shorrosh, and Ohlrogge 1994).
LCB1	Long-chain base 1	Lipid biosynthesis	*LCB1* becomes a serine palmitoyltransferase when it forms a heterodimer with *LBC2*. This complex catalyzes the first step in sphingolipid biosynthesis which are major components of endomembranes and are linked to processes such as programed cell death and ABA-dependent guard-cell closure (Chen et al. 2006).
ATS2	Lysophosphatidic acid acyltransferase 1	Lipid biosynthesis	*ATS2* is a lysophosphatidic acid acyltransferase that catalyzes the synthesis of phosphatidic acid (PA) which is required for the formation of glycerolipids, a major structural and functional components of membranes (Yu et al. 2004).
MCMT	Malonyl CoA-ACP malonyltransferase	Lipid biosynthesis	*MCMT* converts malonyl-CoA and acyl carrier protein (ACP) to malonyl-ACP and CoA which is necessary for fatty acid biosynthesis in the plastids and the synthesis of membranes and specific enzyme cofactors in the mitochondria (Jung et al. 2019).
CTase	αCTase	Lipid biosynthesis	Part of a 4-protein Acetyl-CoA Carboxylase complex.
ENR	Enoyl-ACP reductase	Lipid biosynthesis	Catalyzes the NADPH-dependent reduction of trans-2-enoyl thioesters in mitochondrial fatty acid synthesis.
CCD7	Carotenoid cleavage dioxygenase 7	Strigolactones biosynthesis	Asymmetrically cleaves a variety of linear and cyclic carotenoids at the 9-10 double bond.
CCD8	Carotenoid cleavage dioxygenase 8	Strigolactones biosynthesis	Cleaves the C(27) 9-cis-10'-apo-beta-carotenal produced by *CCD7*.
M6PR	Mannose 6-phosphate reductase	Mannitol biosynthesis	Reduce the early photosynthetic product M6P into mannitol-1-phosphate.
ACS	1-aminocyclopropane-1-carboxylate synthase	Ethylene biosynthesis	Catalyzes the conversion of S-adenosyl-L-methionine (SAM) into 1-aminocyclopropane-1-carboxylate (ACC), a direct precursor of ethylene.
EPSP	5-enoylpyruvylshikimate 3-phosphate synthase	Chorismate biosynthesis	5-enolpyruvylshikimate-3-phosphate synthase involved in shikimic acid biosynthesis required for aromatic amino acid production.
AdSS	Adenylosuccinate synthetase	Purine nucleotide biosynthesis	Plays an important role in the de novo pathway and in the salvage pathway of purine nucleotide biosynthesis.
Prx	Peroxidase 1	Diverse processes	Removal of H_2O_2, oxidation of toxic reductants, biosynthesis and degradation of lignin, suberization, auxin catabolism, response to environmental stresses such as wounding, pathogen attack and oxidative stress.
VCL1	VACUOLELESS1	Vacuole biogenesis	Required for vacuole biogenesis and vacuole enlargment in dividing and expanding cells. Involved in the docking or fusion of prevacuolar vesicles.
GUS	β-glucuronidase	Hydrolysis of carbohydrates and Visual reporter of expression	Catalyzes the hydrolysis of a wide variety of β-glucuronide carbohydrates. Also used as visual assay for expression localization.
STM	SHOOT MERISTEMLESS	Developmental transcription factor	Required for shoot apical meristem (SAM) formation during embryogenesis.
KNAT1-3	Knotted-like 1-3	Developmental transcription factor	May play a role in meristem function, and may be involved in maintaining cells in an undifferentiated, meristematic state.

Table S2: Primers used in this study.	
Primer name:	**Sequence:**
PhAe.ACC1.F1	GTACAATCAGTTTTCCAGATGAAGG
PhAe.ACC1.R1	CCAGATCAAATTCAAACTAATACAAAGC
PhAe.LCB1.F1	AGTTATTTGTTGTGTTACTTACACTTTACAAG
PhAe.LCB1.R1	CTGGGAGCTAGTGGTGGAATAAG
PhAe.ATS2.F1	GCTAATGTAATGGAAGGTATTTGCTC
PhAe.ATS2.R1	TTTAACCGTATATTTATTTCAATCTTATGGATAC
PhAe.MCMT.F1	CGATGTTTCTGAAGCCAATTCTTCG
PhAe.MCMT.R1	GTGAGAGAAATAGAAGAAATCACGACC
PhAe.ACC1.F1.attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGTACAATCAGTTTTCCAGATGAAGG
PhAe.ACC1.R1.attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTGCCAGATCAAATTCAAACTAATACAAAGC
PhAe.LCB1.F1.attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTACAGTTATTTGTTGTGTTACTTACACTTTACAAG
PhAe.LCB1.R1.attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTGGGAGCTAGTGGTGGAATAAG
PhAe.ATS2.F1.attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCTAATGTAATGGAAGGTATTTGCTC
PhAe.ATS2.R1.attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTGTTTAACCGTATATTTATTTCAATCTTATGGATAC
PhAe.MCMT.F1.attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTACCGATGTTTCTGAAGCCAATTCTTCG
PhAe.MCMT.R1.attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTGGTGAGAGAAATAGAAGAAATCACGACC

PhAe.4X.ACC1-LCB1.F	GCTTTGTATTAGTTTGAATTTGATCTGGAGTTATTTGTTGTGTTACTTACACTTTACAAG
PhAe.4X.ACC1-LCB1.R	CTTGTAAAGTGTAAGTAACACAACAAATAACTCCAGATCAAATTCAAACTAATACAAAGC
PhAe.4X.LCB1-ATS2.F	CTTATTCCACCACTAGCTCCCAGGCTAATGTAATGGAAGGTATTTGCTC
PhAe.4X.LCB1-ATS2.R	GAGCAAATACCTTCCATTACATTAGCCTGGGAGCTAGTGGTGGAATAAG
PhAe.4X.ATS2-MCMT.F	GTATCCATAAGATTGAAATAAATATACGGTTAAACGATGTTTCTGAAGCCAATTCTTCG
PhAe.4X.ATS2-MCMT.R	CGAAGAATTGGCTTCAGAAACATCGTTTAACCGTATATTTATTTCAATCTTATGGATAC
ACC1.qPCR.F1	CGTGGGTGGTTCTCGACAGTA
ACC1.qPCR.R1	TGACTAGCTCTGGGTCGAGTC
LCB1.qPCR.F1	GGTGTGCACTGGGGAATTCAA
LCB1.qPCR.R1	TTCGTTTGTTGGCGTGGGTAA
ATS2.qPCR.F1	CGATTATTGGGTGGGCCATGT
ATS2.qPCR.R1	AGCTGCCATCTTTACTCCTAGTGC
MCMT.qPCR.F1	CTGAGCCACATGCAGATCCAG
MCMT.qPCR.R1	TTCATAGCTCTTTCTCAATCCTCTGCTT
TUB1.qPCR.F1	GCTCATTGACTCCGTGCTTGA
TUB1.qPCR.R1	CCCATCCCAGAACCAGTACCA
UBQ1.qPCR.F1	TGCATTTGGTTTTGAGGCTCAGG
UBQ1.qPCR.R1	TGGTTGCTGTGTCCACACTTC

REFERENCES:

1. Spallek, T.; Mutuku, M.; Shirasu, K. The genus Striga: a witch profile. Mol. Plant Pathol. 2013, 14, 861–869.

2. Parker, C. The Parasitic Weeds of the Orobanchaceae. Parasitic Orobanchaceae 2013, 313–344.

3. Hershenhorn, J.; Eizenberg, H.; Dor, E.; Kapulnik, Y.; Goldwasser, Y. Phelipanche aegyptiacamanagement in tomato. Weed Research 2009, 49, 34–47.

4. Eizenberg, H.; Plakhine, D.; Ziadne, H.; Tsechansky, L.; Graber, E.R. Non-chemical Control of Root Parasitic Weeds with Biochar. Front. Plant Sci. 2017, 8, 939.

5. Gressel, J.; Joel, D.M. Weedy Orobanchaceae: The Problem. Parasitic Orobanchaceae 2013, 309–312.

6. Joel, D.M. The Haustorium and the Life Cycles of Parasitic Orobanchaceae. Parasitic Orobanchaceae 2013, 21–23.

7. Westwood, J.H.; Kim, G. RNA mobility in parasitic plant – host interactions. RNA Biology 2017, 14, 450–455.

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

9. Bandaranayake, P.C.G.; Yoder, J.I. Trans-specific gene silencing of acetyl-CoA carboxylase in a root-parasitic plant. Mol. Plant. Microbe. Interact. 2013, 26, 575–584.

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

11. Aly, R.; Dubey, N.K.; Yahyaa, M.; Abu-Nassar, J.; Ibdah, M. Gene silencing of CCD7 and CCD8 in Phelipanche aegyptiaca by tobacco rattle virus system retarded the parasite development on the host. Plant Signal. Behav. 2014, 9, e29376.

12. Dubey, N.K.; Eizenberg, H.; Leibman, D.; Wolf, D.; Edelstein, M.; Abu-Nassar, J.; Marzouk, S.; Gal-On, A.; Aly, R. Enhanced Host-Parasite Resistance Based on Down-Regulation of Phelipanche aegyptiaca Target Genes Is Likely by Mobile Small RNA. Frontiers in Plant Science 2017, 8.

13. Muhammad, T.; Zhang, F.; Zhang, Y.; Liang, Y. RNA Interference: A Natural Immune System of Plants to Counteract Biotic Stressors. Cells 2019, 8.

14. Shahid, S.; Kim, G.; Johnson, N.R.; Wafula, E.; Wang, F.; Coruh, C.; Bernal-Galeano, V.; Phifer, T.; dePamphilis, C.W.; Westwood, J.H.; et al. MicroRNAs from the parasitic plant Cuscuta campestris target host messenger RNAs. Nature 2018, 553, 82–85.

15. Rosa, C.; Kuo, Y.-W.; Wuriyanghan, H.; Falk, B.W. RNA Interference Mechanisms and Applications in Plant Pathology. Annu. Rev. Phytopathol. 2018, 56, 581–610.

16. Ejeta, G.; Gressel, J. Integrating New Technologies for Striga Control 2007.

17. Alakonya, A.; Kumar, R.; Koenig, D.; Kimura, S.; Townsley, B.; Runo, S.; Garces, H.M.; Kang, J.; Yanez, A.; David-Schwartz, R.; et al. Interspecific RNA interference of SHOOT MERISTEMLESS-like disrupts Cuscuta pentagona plant parasitism. Plant Cell 2012, 24, 3153–3166.

18. Li-Beisson, Y.; Shorrosh, B.; Beisson, F.; Andersson, M.X.; Arondel, V.; Bates, P.D.; Baud, S.; Bird, D.; Debono, A.; Durrett, T.P.; et al. Acyl-lipid metabolism. Arabidopsis Book 2013, 11, e0161.

19. Dietrich, C.R.; Han, G.; Chen, M.; Berg, R.H.; Dunn, T.M.; Cahoon, E.B. Loss-of-function mutations and inducible RNAi suppression of Arabidopsis LCB2 genes reveal the critical role of sphingolipids in gametophytic and sporophytic cell viability. Plant J. 2008, 54, 284–298.

20. Yu, B.; Wakao, S.; Fan, J.; Benning, C. Loss of plastidic lysophosphatidic acid acyltransferase causes embryo-lethality in Arabidopsis. Plant Cell Physiol. 2004, 45, 503–510.

21. Baud, S.; Guyon, V.; Kronenberger, J.; Wuillème, S.; Miquel, M.; Caboche, M.; Lepiniec, L.; Rochat, C. Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in Arabidopsis. Plant J. 2003, 33, 75–86.

22. Chen, M.; Han, G.; Dietrich, C.R.; Dunn, T.M.; Cahoon, E.B. The essential nature of sphingolipids in plants as revealed by the functional identification and characterization of the Arabidopsis LCB1 subunit of serine palmitoyltransferase. Plant Cell 2006, 18, 3576–3593.

23. Jung, S.H.; Kim, R.J.; Kim, K.J.; Lee, D.H.; Suh, M.C. Plastidial and Mitochondrial Malonyl CoA-ACP Malonyltransferase is Essential for Cell Division and Its Overexpression Increases Storage Oil Content. Plant Cell Physiol. 2019, 60, 1239–1249.

24. Bandaranayake, P.C.G.; Filappova, T.; Tomilov, A.; Tomilova, N.B.; Jamison-McClung, D.; Ngo, Q.; Inoue, K.; Yoder, J.I. A single-electron reducing quinone oxidoreductase is necessary to induce haustorium development in the root parasitic plant Triphysaria. Plant Cell 2010, 22, 1404–1419.

25. Yang, Z.; Wafula, E.K.; Honaas, L.A.; Zhang, H.; Das, M.; Fernandez-Aparicio, M.; Huang, K.; Bandaranayake, P.C.G.; Wu, B.; Der, J.P.; et al. Comparative transcriptome analyses

reveal core parasitism genes and suggest gene duplication and repurposing as sources of structural novelty. Mol. Biol. Evol. 2015, 32, 767–790.

26. Aly, R. Conventional and biotechnological approaches for control of parasitic weeds. In Vitro Cellular & Developmental Biology - Plant 2007, 43, 304–317.

27. Rubiales, D.; Rojas-Molina, M.M.; Sillero, J.C. Characterization of Resistance Mechanisms in Faba Bean (Vicia faba) against Broomrape Species (Orobanche and Phelipanche spp.). Frontiers in Plant Science 2016, 7.

28. Wuriyanghan, H.; Falk, B.W. RNA Interference towards the Potato Psyllid, Bactericera cockerelli, Is Induced in Plants Infected with Recombinant Tobacco mosaic virus (TMV). PLoS One 2013, 8, e66050.

29. Khan, A.M.; Ashfaq, M.; Kiss, Z.; Khan, A.A.; Mansoor, S.; Falk, B.W. Use of recombinant tobacco mosaic virus to achieve RNA interference in plants against the citrus mealybug, Planococcus citri (Hemiptera: Pseudococcidae). PLoS One 2013, 8, e73657.

30. Cheng, W.; Song, X.-S.; Li, H.-P.; Cao, L.-H.; Sun, K.; Qiu, X.-L.; Xu, Y.-B.; Yang, P.; Huang, T.; Zhang, J.-B.; et al. Host-induced gene silencing of an essential chitin synthase gene confers durable resistance to Fusarium head blight and seedling blight in wheat. Plant Biotechnol. J. 2015, 13, 1335–1345.

31. Jahan, S.N.; Åsman, A.K.M.; Corcoran, P.; Fogelqvist, J.; Vetukuri, R.R.; Dixelius, C. Plant-mediated gene silencing restricts growth of the potato late blight pathogen Phytophthora infestans. J. Exp. Bot. 2015, 66, 2785–2794.

32. Andrade, C.M.; Tinoco, M.L.P.; Rieth, A.F.; Maia, F.C.O.; Aragão, F.J.L. Host-induced gene silencing in the necrotrophic fungal pathogenSclerotinia sclerotiorum. Plant Pathology 2016, 65, 626–632.

33. Zhu, X.; Qi, T.; Yang, Q.; He, F.; Tan, C.; Ma, W.; Voegele, R.T.; Kang, Z.; Guo, J. Host-Induced Gene Silencing of the MAPKK Gene PsFUZ7 Confers Stable Resistance to Wheat Stripe Rust. Plant Physiology 2017, 175, 1853–1863.

34. Nowara, D.; Gay, A.; Lacomme, C.; Shaw, J.; Ridout, C.; Douchkov, D.; Hensel, G.; Kumlehn, J.; Schweizer, P. HIGS: host-induced gene silencing in the obligate biotrophic fungal pathogen Blumeria graminis. Plant Cell 2010, 22, 3130–3141.

35. Lucas, W.J.; Groover, A.; Lichtenberger, R.; Furuta, K.; Yadav, S.-R.; Helariutta, Y.; He, X.-Q.; Fukuda, H.; Kang, J.; Brady, S.M.; et al. The plant vascular system: evolution, development and functions. J. Integr. Plant Biol. 2013, 55, 294–388.

36. Sunilkumar, G.; Mohr, L.; Lopata-Finch, E.; Emani, C.; Rathore, K.S. Developmental and tissue-specific expression of CaMV 35S promoter in cotton as revealed by GFP. Plant Mol. Biol. 2002, 50, 463–474.

37. Lim, C.J.; Lee, H.Y.; Kim, W.B.; Lee, B.-S.; Kim, J.; Ahmad, R.; Kim, H.A.; Yi, S.Y.; Hur, C.-G.; Kwon, S.-Y. Screening of tissue-specific genes and promoters in tomato by comparing genome wide expression profiles of Arabidopsis orthologues. Mol. Cells 2012, 34, 53–59.

38. Molnar, A.; Melnyk, C.W.; Bassett, A.; Hardcastle, T.J.; Dunn, R.; Baulcombe, D.C. Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. Science 2010, 328, 872–875.

39. Carbonell, A. Secondary Small Interfering RNA-Based Silencing Tools in Plants: An Update. Front. Plant Sci. 2019, 10, 687.

40. Helliwell, C.A.; Varsha Wesley, S.; Wielopolska, A.J.; Waterhouse, P.M. High-throughput vectors for efficient gene silencing in plants. Functional Plant Biology 2002, 29, 1217.

41. Pfaffl, M.W.; Tichopad, A.; Prgomet, C.; Neuvians, T.P. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – Excel-based tool using pair-wise correlations. Biotechnology Letters 2004, 26, 509–515.

42. Andersen, C.L.; Jensen, J.L.; Ørntoft, T.F. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. Cancer Research 2004, 64, 5245–5250.

43. Farrokhi, Z.; Alizadeh, H.; Alizadeh, H.; Mehrizi, F.A. Host-Induced Silencing of Some Important Genes Involved in Osmoregulation of Parasitic Plant Phelipanche Aegyptiaca. Mol. Biotechnol. 2019, 61, 929–937.

44. Cuperus, J.T.; Carbonell, A.; Fahlgren, N.; Garcia-Ruiz, H.; Burke, R.T.; Takeda, A.; Sullivan, C.M.; Gilbert, S.D.; Montgomery, T.A.; Carrington, J.C. Unique functionality of 22-nt miRNAs in triggering RDR6-dependent siRNA biogenesis from target transcripts in Arabidopsis. Nat. Struct. Mol. Biol. 2010, 17, 997–1003.

45. McHale, M.; Eamens, A.L.; Finnegan, E.J.; Waterhouse, P.M. A 22-nt artificial microRNA mediates widespread RNA silencing in Arabidopsis. Plant J. 2013, 76, 519–529.

46. Velasco, L.; Pérez-Vich, B.; Fernández-Martínez, J.M. Research on resistance to sunflower broomrape: an integrated vision. OCL 2016, 23, D203.

CHAPTER 5:

CONCLUSIONS

The transcription regulation of host recognition was characterized in the parasitic plant *Triphysaria*. By comparing these expression changes to non-parasitic plants, it yielded conserved regulatory elements, insights into possible mechanisms of gene co-option, and implications for plant-plant interactions overall.

By searching for enriched sequences in the promoters of host-responsive genes, I discovered seven motifs that regulate tissue specific expression in *Triphysaria* roots. The patterns of spatial regulation for five motifs were highly conserved in two non-parasitic plants *Arabidopsis* and *Mimulus*. This suggests *Triphysaria* uses conserved cis-elements to regulate the location where its host-reponsive genes are expressed. The similar use of these motifs across divergent eudicots suggests they could be useful in engineering the spatial expression of plant genes in the future.

One of the promoter motifs from *Triphysaria* was able to enhance transcription in response to the host derived compound DMBQ. The presence and use of this motif in non-parasites warrants further research to determine if it represents a conserved aspect of a plant's response to nearby plants or if parasites have co-opted its use for haustoria formation.

To evaluate the conserved regulation of a host responsive promoter, I defined a 652 bp sequence from TvQR1 which recapitulated its expression using a fluorescent reporting construct. The spatial regulation and DMBQ responsiveness was conserved in two non-parasitic plants *Arabidopsis* and *Mimulus.* Intriguingly, these non-parasites did not up-regulate their endogenous QR1 in response to DMBQ. This suggests *Triphysaria* has co-opted the expression of QR1 for haustorium initiation through the use of conserved DMBQ responsive cis-elements. Refining

these elements in the future may yield insights into the mechanism of root-root interactions among plants and how parasites have co-opted this communication for the function of parasitism.

Natural variation in the TvQR1 promoter was also found to affect transcriptional expression and haustorium formation in *Triphysaria*. A MITE transposable element was associated with these phenotypic differences and could contribute to the dominant effect of its allele. The prevalence, location bias, and action of MITEs in changing gene expression justifies further research into their role as drivers of genomic change in parasitic plants. The observed variation in TvQR1 has implications for ecology as it could affect host preference. This may be useful as the population of host plants can change from year to year and as an annual plant, *Triphysaria* must contend with these variable conditions.

Lastly, we applied our knowledge of gene expression in parasitic plants to control the agricultural weed *Phelipanche aegyptiaca* through Host Induced Gene Silencing. Our approach obtained a modest level of transcript reduction but did not sufficiently suppress the parasite for commercial applications. Other methods of inducing gene silencing, such as amiRNAs and syn-tasiRNAs could yield a higher level of transcript reduction and parasite control and thus warrant further testing.

Overall, this work contributes to the understanding of plant-plant interactions through the study of host recognition in a parasitic plant and its conserved aspects in non-parasitic plants. Possible applications of this work includes engineering specific patterns of plant gene expression with the identified motifs and improving methods of controlling parasitic weeds through RNAi based host resistance.