Exploring the genetic basis of germination specificity in the parasitic plants *Orobanche cernua* and *O. cumana*

Hailey Larose

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

Doctor of Philosophy
In
Plant Pathology, Physiology and Weed Science

James Westwood (Chair)
Amy Brunner
David Haak
Dorothea Tholl

February 16th, 2018
Blacksburg, VA

Keywords: Parasitic plants, germination stimulant, germination, *Orobanche cumana, Orobanche cernua,* Orobanchaceae, strigolactone, dehydrocostus lactone, *KAI2, D14*

Exploring the genetic basis of germination specificity in the parasitic plants *Orobanche cernua* and *O. cumana*

Hailey Larose

ABSTRACT (Academic)

Seeds of the root parasitic plants of the genus *Orobanche* germinate specifically in response to host-derived germination signals, which enables parasites to detect and attack preferred hosts. The best characterized class of germination stimulants is the strigolactones (SLs), although some species respond to non-SL compounds, such as dehydrocostus lactone (DCL). Recent work indicates that SLs are perceived by members of the *KARRIKIN-INSENSITIVE2* (*KAI2)* gene family, and suggests that within parasitic Orobanchaceae the *KAI2* genes have undergone duplication and specialization. The "diverged" clade of these genes, termed *KAI2d*, has been shown to bind SL germination stimulants in model system assays, but the precise role for *KAI2d* in regulating germination specificity in a parasitic plant has not been demonstrated. To address this issue, we used genetic and genomic approaches involving two closely related species, *Orobanche cernua* and *O. cumana*, which differ primarily in host range and stimulant preference. *Orobanche cernua* parasitizes tomato (and other Solanaceous crops) and responds to orobanchol, the major SL from tomato roots, whereas *O. cumana* specifically parasitizes sunflower and responds to DCL. Crosses between *O. cernua* and *O. cumana* produced hybrid populations that segregate for stimulant specificity, creating a tractable genetic system. *Orobanche cernua* contains four *KAI2d* genes (numbered *OrceKAI2d1-4*), while *O. cumana* contains six genes (*OrcuKAI2d1-6)*. The DNA from 94 F_2 hybrids was genotyped to identify the *KAI2d* gene composition and these were correlated with germination phenotype. The pattern of segregation indicated that the *KAI2d* genes are linked, but pointed to *OrceKAI2d2* as a likely orobanchol receptor. Response to DCL was associated with inheritance of all *O. cumana KAI2d* genes together. Each *KAI2d* gene was expressed in the *Arabidopsis thaliana kai2* mutant background and tested for ability to recover the mutant phenotype when exposed to SLs (including orobanchol, 5-deoxystrigol and GR24) or DCL. One *O. cernua* gene, *OrceKAI2d2,* responded to all SLs, but not DCL in this system. No DCL-specific *KAI2* genes were identified. In summary, we have identified the likely SL receptor in *O. cernua*, and show evidence that the

DCL receptor is either not a *KAI2d* protein, or uses *KAI2d* in combination with other signaling pathway components.

ABSTRACT (Public)

The mechanisms by which parasitic plants of the family Orobanchaceae detect their hosts is a long-standing mystery in plant science. For over half a century it has been known that seeds of parasitic plants will lie dormant until they detect a host-derived germination stimulant. Upon perception of an appropriate germination stimulant, the parasite seeds will send out a radical that has approximately 72 hours to reach a host root before the limited nutrients within the seed are exhausted. The practical impact of this plant signaling regulation is profound, as the parasites in this family include some of the most destructive weeds in the world, including broomrapes (*Orobanche* and *Phelipanche* species) and witchweeds (*Striga* species). Scientists have sought to understand the signaling mechanisms in order to produce crop plants that don't produce/exude the signal or to create chemicals that can mimic stimulants and artificially trigger parasite seed germination. Our goal was to further the understanding of the parasite germination mechanism by determining the genes involved in parasite host specificity in *Orobanche,* of which most members germinate in response to strigolactones (SLs).

Recent work indicates that SLs are perceived by members of the *KARRIKIN-INSENSITIVE2* (*KAI2)* gene family and suggests that within parasitic Orobanchaceae the *KAI2* genes have undergone duplication and specialization. The "diverged" clade of these genes, termed *KAI2d*, has been shown to bind SL germination stimulants in model system assays, but the precise role for *KAI2d* in regulating germination specificity in a parasitic plant has not been demonstrated. To this end we used two closely related species that differ in their germination stimulant and host preferences. *Orobanche cernua* which like most members of Orobanchaceae responds to a SL, and *O. cumana* which has switched to responding to a novel germination stimulant, dehydrocostus lactone (DCL). Through genetic and genomic studies of these two species, we demonstrated that one *O. cernua* gene, *OrceKAI2d2,* responded to all SLs, but not DCL in this system. No DCL-specific *KAI2* genes were identified. In summary, we have identified the likely SL receptor in *O. cernua*, and show evidence that the DCL receptor is either not a *KAI2d* protein, or uses *KAI2d* in combination with other signaling pathway components.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. James Westwood, for accepting me into his lab, and allowing me to integrate into his research team. Dr. Westwood has been beyond patient and supportive as I have transitioned through many life changes. His door has always been open, creating an environment where we can freely discuss research and ask questions. Dr. Westwood has taught me to think critically, creatively and thoroughly. Most importantly, he has always listened to my ideas and theories with an open mind and allowed me to try any experiment I proposed provided I could justify it. This allowed me to gain confidence as a research scientist.

I am grateful to my committee members Drs. Amy Brunner, David Haak and Dorothea Tholl for their feedback, advice and guidance over the past four years. Their unique backgrounds and thorough thinking have taught me to look at my research from different angles. I thank you all for dedicating your time and effort into advising and supporting me through my PhD program.

I would like to thank my collaborators from Ne'we Yaar Research Institute in Israel, including Hanan Eizenburg, Kobi Tadmor and Dina Plakhine, for their constant support and consultation on this project. I'd also like to thank them for hosting me for three months at their research center and the guidance and skills they taught me for working with parasitic plants.

I'd also like to thank my friends and family, for without them, this journey would not have been possible. My family has been instrumental in my success, supporting me through every academic milestone, and assuring me I could achieve my goals. Without my lab mate, and best friend, Kristen Clermont, this journey would have been incomplete. I'd also like to thank my friend, and father of my son, Jeremy Larose, for his support in allowing me to pursue my dreams. He's always available to help with our son when I have to work late, or travel for work. Lastly, I thank my son, Benjamin Larose. I began this program when he was ten months old and have missed many moments of his life. He has tolerated my absences, and flourished despite of it. I look forward to beginning this new chapter of our lives together and continuing to watch you grow.

Attributions

Chapter 2: The *Orobanche cumana* x *Orobanche cernua* genetic system provides insight into the regulation of germination specificity in a parasitic plant.

Hanan Eizenberg, PhD is a research scientist at Newe Ya'ar Research Center in Israel. Dr. Eizenberg is a co-author on this paper, and contributed to the project directions and editorial comments on the manuscript.

David Nelson, PhD is Associate Professor of Genetics, UC Riverside. Dr. Nelson is a co-author on this paper and contributed to project directions and editorial comments on the manuscript. Dr. Nelson advised on the complementation assay, and provided the Gateway destination vector, pKAI2pro-GW.

Dina Plakhine, PhD is a research scientist at Newe Ya'ar Research Center in Israel. Dr. Plakhine is a co-author on this paper, and contributed to the project directions and editorial comments on the manuscript. Dr. Plakhine generated the hybrid lines used in this study, and advised on germination assays of the parasites.

Yaakov Tadmor, PhD is a research scientist at Newe Ya'ar Research Center in Israel. Dr. Tadmor is a co-author on this paper, and co-principle investigator on this grant. Dr. Tadmor contributed to the project directions and editorial comments on the manuscript.

James Westwood, PhD is a Professor of Plant Pathology, Physiology and Weed Sciences at Virginia Tech. Dr. Westwood is a co-author on this paper, and co-principle investigator on this grant. Dr. Westwood helped to design the project directions and assisted in writing the manuscript.

Nathan Wycoff, MS was a Masters student at Virginia Tech in the Department of Statistics. Mr. Wycoff developed the mathematical model used to measure genotype against phenotype of our hybrid lines in this study. Mr. Wycoff provided editorial comments on the manuscript.

Chapter 3: Transcriptomic insights into *Orobanche cernua* and *Orobanche cumana*.

Hanan Eizenberg, PhD is a research scientist at Newe Ya'ar Research Center in Israel. Dr. Eizenberg is a co-author on this paper, and contributed to the project directions and editorial comments on the manuscript.

Dina Plakhine, PhD is a research scientist at Newe Ya'ar Research Center in Israel. Dr. Plakhine is a co-author on this paper, and contributed to the project directions and editorial comments on the manuscript. Dr. Plakhine generated the tissue for RNA-sequencing.

Yaakov Tadmor, PhD is a research scientist at Newe Ya'ar Research Center in Israel. Dr. Tadmor is a co-author on this paper, and co-principle investigator on this grant. Dr. Tadmor contributed to the project directions and editorial comments on the manuscript.

James Westwood, PhD is a Professor of Plant Pathology, Physiology and Weed Sciences at Virginia Tech. Dr. Westwood is a co-author on this paper, and co-principle investigator on this grant. Dr. Westwood helped to design the project directions and assisted in writing the manuscript.

Funding

Chapter 1: This work was supported by BARD award No. US-4616-13 to JW and YT. This work was also supported by a BARD Graduate Student Fellowship award no. GS-33-2016 to HL.

Chapter 2: This work was supported by BARD award No. US-4616-13 to JW and YT.

Table of Contents

List of Figures

List of Tables

Chapter 1

1.1 Background

1.1.1 Agricultural problem and research significance

The root parasitic weeds of the family Orobanchaceae are amongst the most globally destructive agricultural pests, reducing crop yields and resulting in economic loss (Hegenauer et al., 2017; Lumba et al., 2017). Some examples include the root parasitic weeds of the genera *Orobanche* and *Phelipanche*, which are major constraints to the production of legume and vegetable crops in Mediterranean, Eastern European and Middle East regions. Their distribution is expanding into Australia, the USA and South America (Parker, 2013). The root parasitic weeds of the genus *Striga* severely impact cereal and legume crops throughout Africa and parts of Asia (Yoder & Scholes, 2010)

The severity of crop destruction is a result of many compounding factors. First is the location of the parasite. These root-parasitic plants spend most of their life cycle underground, emerging for the purpose of reproduction. They survive by forming physical connections to the roots of their host plant through a structure called a haustorium. It is through this haustorium that they acquire nutrients for growth and reproduction. Consequently, by the time the parasites are visually detected in the field, the damage to the crop has been done. Once above ground, they can set and release hundreds of thousands of seeds per plant. As this occurs year after year, the seed bank in a field accumulates to high levels. There exist very few effective methods to control these parasites in the field, and even fewer for managing seed bank populations. Research is needed to better understand the biology of these parasites so that more effective control methods can be devised.

1.1.2 *Orobanche cumana* and *Orobanche cernua*

Two obligate holoparasitic weeds of the genera *Orobanche* are the root parasites *O. cumana* and *O. cernua*. These root parasites are both completely dependent on a host for nutrient acquisition. *Orobanche cumana* is closely related to *O. cernua*, and was previously named *O. cernua* ssp. *cumana*, but is now considered a distinct species. This split was justified through comparing morphological features, host-species preference and fatty acid profiles of seeds (Pujadas-Salvà &

Velasco, 2000). While it is often difficult to distinguish the two species under field conditions, major differences between the species are summarized as follows:

1.1.3 *Orobanche cumana*

Orobanche cumana can grow to approximately 50cm in height, and is characterized by small flowers that are narrowly tubular, markedly down curved and white to pale-blue in color (Parker, 2013). It has evolved to specialize on cultivated sunflower (*Helianthus annuus* L.) and ranges from Spain to China (Joel et al., 2011; Parker, 2012). *Orobanche cumana* is unique among the genus *Orobanche* because it germinates in response to dehydrocostus lactone (DCL), while the majority of *Orobanche* species respond to stimulants belonging to the strigolactone (SL) class of phytohormones (Joel et al., 2011).

1.1.4 *Orobanche cernua*

Orobanche cernua is morphologically similar to *O. cumana* with slight differences in flower morphology; flowers are slightly less narrowly tubular, less bent, more deeply colored, and clustered more tightly together (Chris Parker, 2012). Due to the morphological similarities between *O. cumana* and *O. cernua*, the range of *O. cernua* is difficult to pinpoint. It exhibits a more southern distribution, extending into North Africa and Southern Asia, having recently expanded into Eastern and Western Africa (Parker, 2013). The development of a set of simple sequence repeat (SSR) markers for *O. cumana* can assist in species identification as some SSRs are present only in one species (Pineda-Martos et al., 2014). Like most members of the Orobanchaceae, *O. cernua* germinates in response to SLs, mainly parasitizing members of the *Solanaceae*; particularly tomato (*Lycopersicon esculentum* Mill.), tobacco (*Nicotiana tabacum*) and eggplant (*Solanum melongena* L.) (Parker 2013).

1.1.5 Germination and stimulant perception

While the seeds of most plants will germinate more or less when exposed to appropriate temperature, humidity, oxygen, water and light levels, seeds of root parasitic plants of the genera Orobanchaceae require the perception of a host-derived germination signal exuded from host plant roots. This stimulant dependency restricts seed germination to the vicinity of host roots and prevents seed germination in the absence of a preferred host (Westwood et al., 2010). This restriction is crucial; the seedling will die if its radicle, which can grow only a few millimeters, does not reach the host before the limited resources in the seed are exhausted (Westwood et al.,

2010). For most members of Orobanchaceae the host-derived germination signal is a SL. *Orobanche* species are able to identify their preferred hosts by recognizing specific compositions of host-derived stimulants.

There are three main physiological blocks to germination of *Orobanche* seeds: after-ripening, conditioning (a period of exposure to water at appropriate temperature), and the perception of the host-derived germination stimulant. These steps must be alleviated sequentially for germination to proceed. Once germinated and in contact with a host root, the parasite radicle develops a haustorium that establishes vascular connections with the host and subsequently obtains all nutrition needed for the development of the parasite shoot, flowers and seeds.

1.1.6 *Orobanche* seed anatomy

Major consideration regarding the germination of a parasite seed is the location of the stimulant receptor. It has been suggested that the perisperm cells, which are of maternal origin, are the sites that perceive the host-derived germination stimulant (Joel et al., 2012, Plakhine et al., 2012; Tsuchiya et al., 2015). An ultra-structure analysis of an *Orobanche* seed revealed that these seeds contain an inner coat that is water impermeable due to cutinized walls (Joel et al., 2012). Swelling of the endothelium during the imbibition period leads to an opening at the micropyle through which water may enter. The cells that surround the micropyle are perisperm cells that have direct contact with the embryo (indicated by asterisks in Figure 1).

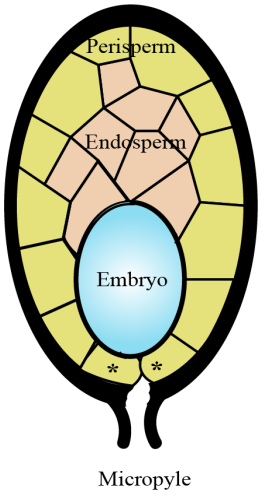


Figure 1.1 Structure of *Orobanche* seed. * indicates perisperm cells that are thought to be the site of stimulant perception. (Diagram based on Plakhine et al., 2012, Joel et al., 2012)

Another hypothesis is that the stimulant could reach the embryo by going between or through the perisperm cells, which would make the receptor located in the embryo. However, a study by Plakhine et al. (2012) provided genetic evidence that stimulant perception is located within the maternally derived perisperm cells, rather than the embryo. By crossing *O. cumana* and *O. cernua* they observed development of a spontaneous germination (germination in the absence of a stimulant) phenotype appearing in the F_3 generation. The absence of spontaneous germination in the F_1 and F_2 generation, and its appearance in F_3 suggests that the genes are located in maternal tissue, rather than the embryo. Given that the perisperm is retained tissue with maternal origin, it is likely that the stimulant receptors are located here (Plakhine et al., 2012). Tsuchiya et al, 2015 further demonstrated this idea using imaging of a fluorescence turn-on probe, an artificial SL termed Yoshimulactone Green (YLG). Fluorescence of this probe occurs when the YLG-SL is hydrolyzed by a receptor into fluorescein and the biologically active D-ring. Live imaging of *Striga* seeds exposed to YLG showed fluorescence initiating near the micropyle, before extending upwards through the seed tip (Tsuchiya et al., 2015).

1.1.7 *O. cumana* and *O. cernua* hybrid families

Host specificity for *O. cumana* and *O. cernua* is primarily determined by sensitivity to a germination stimulant. Each species responds to the root exudates of its preferred host, but the reciprocal exposures have no effect. Specifically, *O. cernua* will not germinate in response to sunflower exudates, nor will *O. cumana* to tomato exudates (Plakhine et al., 2012). However, once germinated these two parasites can penetrate, make a vascular connection, and grow to maturity on either tomato or sunflower.

One of the major challenges in trying to elucidate the germination mechanism of parasitic Orobanchaceae is the lack of genetic resources for these parasites, such as populations of a single species that differ markedly in their germination specificity. As a substitute for such a system, crosses of *O. cumana* and *O. cernua* could help circumvent this problem. The recent evolutionary divergence of these species from each other enables *O. cumana* and *O. cernua* to interbreed and produce fertile hybrid offspring. In fact, crossing these species resulted in F_3 hybrid lines that segregated for stimulant specificity (Plakhine et al., 2012), creating a tractable genetic system that we can use to investigate the germination mechanism.

Another limitation to research on *Orobanche* species is the lack of a protocol to generate transgenic parasites. The related species *Phelipanche aegyptiaca* has been transformed

(Fernández-Aparicio et al., 2011) but *O. cumana* and *O. cernua* have proved to be much more difficult to culture. Without the capacity to knock out genes or express transgenes in the parasites, a genetic system based on phenotypically segregating hybrids is an attractive option.

1.2 Strigolactones and karrikin as plant hormones

1.2.1 Biological roles of SLs

Strigolactones play a role in many biological functions across diverse organisms. To date, most naturally occurring germination stimulants of the parasitic plant family Orobanchaceae, are SLs (Yoneyama et al., 2013). SLs have also been shown to serve as a host recognition mechanism for arbuscular mycorrhizal (AM) fungi, which form symbiotic associations with the roots of more than 80% of terrestrial plants (Akiyama et al., 2005). More recently, SLs were classified as phytohormones, involved in regulating shoot and root architecture in response to nutrient access (Al-Babili & Bouwmeester, 2015; Waters, 2017). The wide breadth of SL involvement in different kingdoms has led to extensive research across many disciplines, yielding the identification of many components of the SL biosynthesis and perception pathways.

1.2.2 Structural requirements of SLs:

Significant research has gone into the chemistry of germination stimulants, highlighting the importance of stereochemistry for bioactivity and describing the functionality for the various forms of SLs (Zwanenburg et al., 2016). Currently all characterized naturally derived SLs have the same structural features: an "ABC" tricyclic ring system connected to a five-membered butenolide "D-ring" by an enol-ether bridge (Zwanenburg & Pospíšil, 2013) (Figure 1.2). The C-D ring moiety is the essential structure for conferring germination stimulation activity in parasitic plants (Zwanenburg et al., 2016).

Unlike *O. cernua* and the majority of Orobanchaceae, *O. cumana* responds to DCL, a guaianolide sesqiterpene lactone derived from the cytosolic mevalonate biosynthesis pathway (Joel et al., 2011) (Figure 1.2). Since the two closely related species of *O. cumana* and *O. cernua* differentially respond to either SL or DCL respectively, it was speculated that the evolution of stimulant specificity may have involved a single mutation that changed the binding site of the receptor, and may be coded by two alleles of the same gene (Joel at al. 2011).

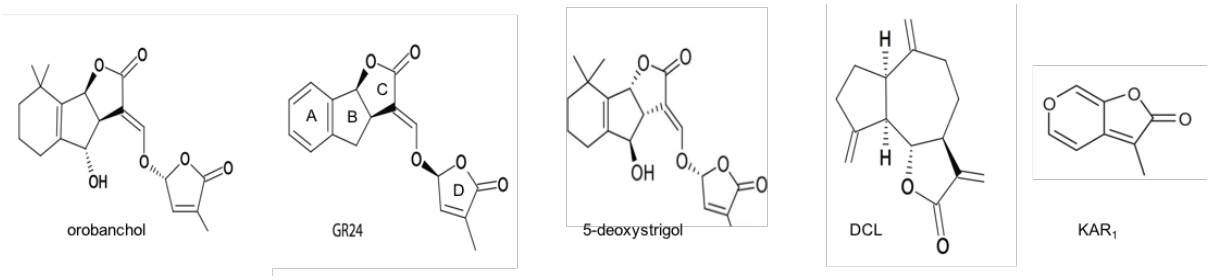


Figure 1.2 Chemical structure examples of strigolactones, DCL and Karrikins.

1.2.3 Biosynthesis pathway of strigolactones

Since SLs were recognized as phytohormones in 2008, SL research has increased rapidly (Gomez-Roldan et al., 2008; Umehara et al., 2008). Researchers have elucidated many genes involved in the biosynthesis and signaling/perception pathway by characterizing branching mutants of several plant species including: thale cress (*Arabidopsis thaliana)* ***max*** (more axillary growth) mutants, rice (*Oryza sativa*) *d* (dwarf) mutants, pea (*Pisum sativum*) *rms* (ramosus) mutants and petunia (*Petunia hybrida*) *dad* (decreased apical dominance) mutants (reviewed in (De Cuyper et al., 2017; Lumba et al., 2017; Waters et al., 2017). The history of these genes being first characterized in different plant species has led to many synonymous names for genes. In this work, we will use the *Arabidopsis* nomenclature (indicated in bold at first use).

Some key genes involved in the SL biosynthesis process of these species have been identified. In the proposed pathway ***MAX3*** (*RMS5, D17/HTD1 and DAD3*) encodes *CAROTENOID CLEAVAGE DIOXYGENASE 7* (*CCD7*), ***MAX*** *4* (*RMS1, D10*, and *DAD1*) encodes another class of CCDs designated *CCD8*. ***MAX1*** (*2 PsMAX1, 5 OsMAX1, PhMAX1*) encodes a cytochrome P450 (Arite et al., 2007; Booker et al., 2005; Lin et al., 2009; Snowden et al., 2005). *CCD7* and *CCD8* catalyze sequential carotenoid cleavage reactions, and *MAX1* is a cytochrome P450 involved downstream in the conversion of carlactone to carlactonoic acid, the precursor to all known natural occurring SLs (Gomez-Roldan et al., 2008).

1.2.4 Perception pathway of strigolactones:

In contrast to SL biosynthesis, the mechanisms by which the various root parasitic plants respond to SL germination signals has been more difficult to resolve. The hypothesized model for SL perception in the root parasitic plants shares commonalities between the gibberellin (GA), auxin and jasmonic acid (JA) perception pathways (Morffy et al., 2016) . Consequently, it was proposed that the receptor could be a distinct protein, or the F-box protein (Figure 1.3).

Much like the above listed hormone signaling pathways, SL signaling is mediated through ubiquitin-mediated degradation of target proteins (Figure 1.3). The SL receptor was shown to be an α/β-hydrolase termed *DWARF-14* or ***D14***. *D14* acts not only as a receptor for SL's, but also as a hydrolase that catalyzes the cleavage of many natural and synthetic SLs (Nakamura et al., 2013; Yao et al., 2016; Zhao et al., 2015). This cleavage is necessary to induce the conformational change of *D14*, exposing the correct interface for *MAX2* interaction. This hydrolysis by the Ser97-His247-Asp218 catalytic triad releases the ABC-rings of the SL, and retains the D-ring as a covalently linked intermediate molecule (CLIM) (Yao et al., 2016).

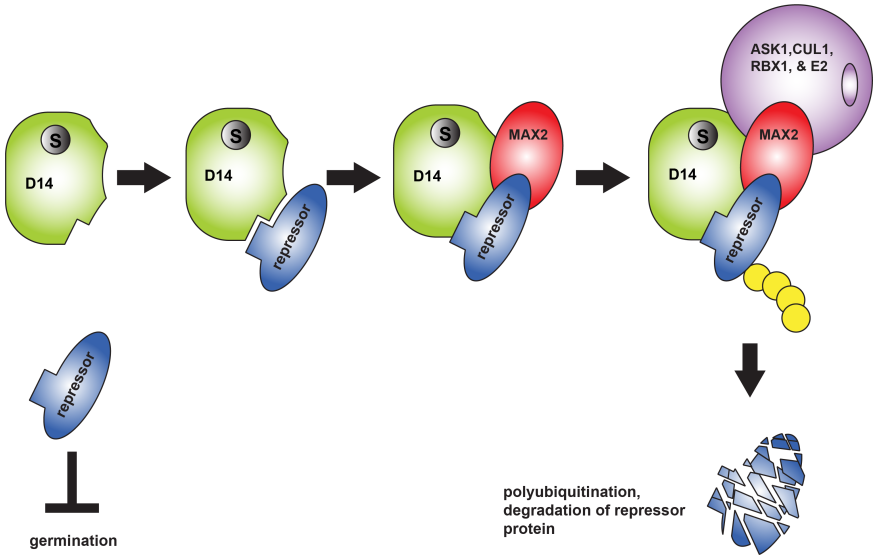


Figure 1.3 Hypothetical model of SL perception in *Orobanche.*

The action of cleaving the SL is required, as the D-ring itself is not sufficient to induce *D14-MAX2-ASK1* interaction (Yao et al., 2016). The open, unbound state of *D14* exposes a large open pocket (420 Å), which is compatible with bulky molecules such as SLs. However, following the appropriate conformational change and binding to the CLIM, the pocket shifts in size to 80 Å (Yao et al., 2016). The closed-state of *D14* contains CLIM within the binding pocket and a collapsed lid exposing three helices compatible with binding *MAX2* (Yao et al., 2016). Once bound with *MAX2,* which is part of an SCF^{MAX2} complex, a repressor of germination from the *SMXL* gene family is recruited, and the entire complex is destroyed through proteasome mediated degradation (Nelson et al., 2012). This would allow for subsequent transcription of downstream genes required for germination (Figure 1.3).

1.2.5 Divergence of karrakin receptor, *KAI2,* in parasitic plants

Because members of Orobanchaceae generally germinate in response to SLs, it was initially hypothesized that the signal perception mechanism may have repurposed the SL perception pathway into detecting host derived germination stimulants. However, two recently published papers have provided convincing evidence that the neo-functionalization of a *D14*-related receptor, KARRIKIN-INSENSITIVE-2 (***KAI2***), may be responsible for stimulant perception (Conn et al., 2015; Tsuchiya et al., 2015).

Karrikins (KARs) are hormones that are derived from smoke and are a germination stimulant for over 1,200 species worldwide, inducing germination after a fire (Nelson et al. 2008). They are composed of an A and a B ring, and have structural similarity to the D ring of SLs (Figure 1.2). They are thought to be a substrate for *KAI2.*

Like *D14, KAI2* is an alpha-beta hydrolase, containing a catalytic triad at positions Ser95-His246-Asp217. Both proteins have been crystalized revealing remarkably similar structures including a double layer V-shaped helical fold containing a substrate-binding cavity, which is notably smaller in *KAI2* (Zhao et al., 2013). The differential size of the binding cavity is attributed to the size of the hormones perceived by the receptors. The *KAI2* perception pathway contains similar interacting proteins to that of *D14. KAI2* interacts with *MAX2* and *SMAX1/SMLX2* of the *SMXL* gene family (Figure 1.3).

It is currently unknown whether *KAI2* behaves like *D14* by hydrolyzing KARs, and whether hydrolysis is necessary for the induction of the conformational change of *KAI2* for *MAX2* interaction. While the binding pocket of *KAI2* is insufficient in size to hold a molecule the size of a SL, KARs also contain the D-ring moiety (without the enol-ether connection) that is capable of fitting within the cavity (Figure 1.2), hinting at the possibility of hydrolysis without the necessity of a released intermediate (Zhou et al., 2013, Scaffoldi et al., 2014).

Conn et al. (2015) proposed that the *KAI2* family has expanded and *diverged* in parasite species to perceive host derived germination stimulants and this may contribute to host specificity. They investigated *KAI2* and *D14* in ten species that represent the full range of parasitism in Orobanchaceae and observed that *KAI2*, but not *D14*, is present at higher copy numbers in parasitic species than in nonparasitic relatives. They classified the *KAI2* orthologs into three different phylogenetic clades: *KAI2c* (conserved), *KAI2i* (intermediate) and *KAI2d*

(divergent), and found that the *KAI2d* was the fastest evolving clade and contained the largest number of *KAI2* orthologs.

Homology modeling predicted that the ligand binding pockets of *KAI2d* most resemble *D14* with less conservation in the amino acids surrounding and inside the binding pocket as compared to the other two phylogenetic clades. To further test their hypothesis, they performed cross-species complementation assays where they inserted *KAI2d* transgenes from *Phelipanche aegyptiaca* and *Striga hermonthica* into a *kai2-2 A. thaliana* mutant background, and tested the ability of the transgenes to rescue the delayed germination phenotype when exposed to stimulant KARs and the synthetic SL, GR24. They concluded that the *KAI2* orthologs *D14* and *KAI2d* underwent convergent evolution for species-specific SL recognition.

A paper by Tsuchiya et al. (2015) demonstrated that *HTL/KAI2* gene was responsible for germination stimulant perception in *Striga hermonthica* using a modified SL that gives off a fluorescent signal when cleaved, named Yoshimulactone Green (YLG). By isolating all twelve *HTL* genes present within *S. hermonthica*, they measured the ability of each protein to cleave the YLG molecule, and demonstrated that a group of ten *HTL* genes could bind to SL at varying affinities. They also demonstrated YLG perception in *S. hermonthica* seeds using live imaging technology, noting maximum fluorescence at the micropyle end of the seed around four hours after exposure to YLG, with a minimum for six hours of exposure needed for efficient germination (Tsuchiya et al., 2015).

Toh et al. (2015) followed up on the work by Tsuchiya et al. (2015) by testing the ability of eleven *ShHTL* genes to complement the *htl* mutant phenotype in *Arabidopsis thaliana* mutant. Through this assay they demonstrated that a subset of six *HTL* genes was sufficient for *Striga* germination with preferences to certain SL variants. Additionally, they crystalized a representative of the responsive subclade and noted substantial increase in the binding cavity size compared to *AtKAI2.* By comparing *Striga HTL* genes, they concluded that the change of certain key amino acids allows for a binding cavity large enough to properly bind a SL molecule (Toh et al., 2015).

Most recently, Yao et al., (2017) demonstrated that *ShHTL7,* the most sensitive *ShHTL* to GR24, interacted with *AtASK1-ShMAX2* and *AtSMAX1* in a GR24 dependent manner through pull down assays. They also demonstrated through size exclusion chromatography assays that *ShHTL7* was capable of cleaving GR24 into the appropriate CLIM molecule (Yao et al., 2017).

The KAR perception pathway mirrors the SL perception pathway in that both require the F-box protein *MAX2* and an α/β-hydrolase receptor. Binding of KAR to *KAI2* recruits the F-Box protein *MAX2* and a repressor, which is polyubiquinated and degraded by the SCF^{MAX-2} complex. *SMAX1* was identified as KAR-inducing genes (Nelson et al., 2011). Further studies have shown that *SMAX1* belongs to a gene family containing *SMAX-LIKE 2-8*, which retains functional redundancy and may contribute differently to the SL and KAR pathways (Soundappan et al., 2015; Stanga et al., 2013). It is predicted that *SMAX1* and *SMXL2* act as downstream repressors in the KAR pathway while *SMXL6-8* maintain functional redundancy and act as a downstream repressor in the SL pathway (Stanga et al., 2013; Morffy et al., 2016).

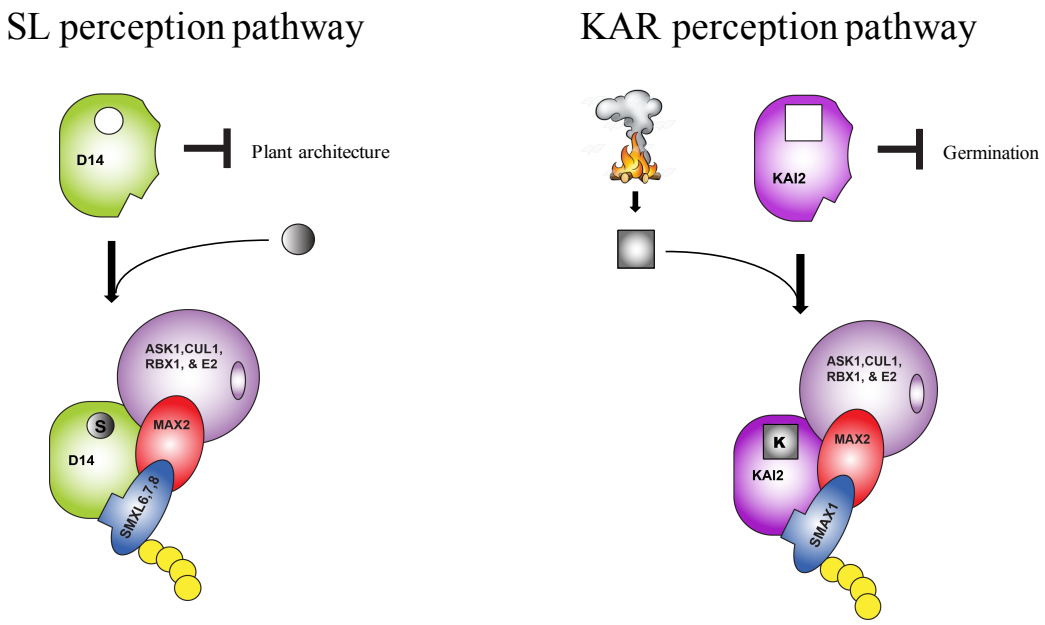


Figure 1.4: Hypothetical model of SL and KAR perception.

1.2.6 *SMXL* gene family, the repressor of germination

The repressor proteins functioning downstream of *MAX2* was originally implicated as *DWARF53* (D53) in rice (Jiang et al., 2013; Zhao et al., 2013). Following this discovery, the gene families SUPPRESSOR OF MAX2-1 and SMAX-LIKE (*SMAX1/SMXL2-8*) proteins were identified as the *D53* homolog in *Arabidopsis* (Soundappan et al., 2015; Stanga et al., 2013; Stanga et al., 2016; Wallner et al., 2017). A screen for genetic suppressors of the enhanced seed dormancy phenotype of *max2* in *Arabidopsis* identified the *suppressor of max2 1 (smax1)* mutant. *smax1* restores the seed germination and seedling photomorphogenesis phenotypes of *max2* but does not

affect the branching phenotypes of *max2.* The authors concluded that during seed germination and seedling growth, *SMAX1* played an important role downstream of *MAX2* in KAR/SL signaling, but is not necessary for all *MAX2*-dependent responses (Stanga et al., 2013). *SMAX1* belongs to a gene family containing *SMX-LIKE2-8.* While *SMAX1* was identified to play a role in KAR/SL signaling in seed germination, it was also noted that the *smax1* seedlings were still responsive to KAR and GR24 application, indicating the potential of functional redundancy amongst the *SMXL2-8* gene family (Stanga et al., 2013).

This gene family can be clustered into three sub-clades based on protein homology: sub-clade one containing *SMAX1* and *SMXL2*, sub-clade two containing *SMXL3,4,5* and subclade three containing *SMXL6,7,8* (Stanga et al., 2013; Wallner et al., 2017). Each sub-clade has been demonstrated to have involvement in different aspects of SL and KAR plant regulation. Current studies have provided a multitude of genetic evidence that pinpoints *SMXL* gene expression to different regions of plant development, including expression only in seed development (KAR-related) or in axillary buds (SL-related), as well as biochemical studies showing a direct *MAX2*-dependent proteolysis of a *D14/SMXL6-7* complex (Jiang et al., 2013; Soundappan et al., 2015; Stanga et al., 2013; Stanga et al., 2016; Umehara et al., 2015; Wang et al., 2015; Zhao et al., 2013).

Within the *SMXL* family, sub-clade one (*SMAX1, SMXL2*) has been demonstrated to mediate KAR/KL responses while sub-clade three (*SMXL6-8*) has been demonstrated to mediate SL responses (Stanga et al., 2013; Soundappan et al., 2015; Stanga et al., 2016). Recently, sub-clade two (*SMXL4-6*) was shown to act independently of KAR/KL and SL signaling downstream of *MAX2* in early phloem development (Wallner et al., 2017). Assuming that parasites have repurposed the SL or KAR/KL pathway to detect exogenously exuded hormones from a host plant, it is of further interest to explore putative repressors as a mechanism for conferring stimulant specificity.

1.3 Hypothesis and objectives

We hypothesized that the *KAI2* gene family mediates germination specificity in *O. cumana* and *O. cernua*. We explored this hypothesis using genetic and functional genomic approaches. To overcome the lack of a good genetic system for studying germination specificity and the lack of a transformation protocol for these parasitic plants, we developed a genetic system by generating

hybrids of *O. cumana* x *O. cernua* that segregated in stimulant response. We identified *O. cumana* and *O. cernua KAI2* genes and genes of known associated proteins, and correlated *KAI2d* genes with stimulant-response phenoptypes of segregating hybrid populations. We explored the functional ability of each *KAI2* gene to respond DCL and various SLs through cross-species complementation studies. Lastly, we used our species to explore gene regulation through dormancy relief pre-and post-germination stimulant to assess the similarities.

1.4 References

Akiyama, K., Matsuzaki, K. I., & Hayashi, H. (2005). Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature*, *435*(7043), 824–827. https://doi.org/10.1038/nature03608

Al-Babili, S., & Bouwmeester, H. J. (2015). Strigolactones, a Novel Carotenoid-Derived Plant Hormone. *Annual Review of Plant Biology*, *66*(1), 161–186. https://doi.org/10.1146/annurev-arplant-043014-114759

Arite, T., Iwata, H., Ohshima, K., Maekawa, M., Nakajima, M., Kojima, M., … Kyozuka, J. (2007). DWARF10, an *RMS1/MAX4/DAD1* ortholog, controls lateral bud outgrowth in rice. *Plant Journal*, *51*(6), 1019–1029. https://doi.org/10.1111/j.1365-313X.2007.03210.x

Booker, J., Sieberer, T., Wright, W., Williamson, L., Willett, B., Stirnberg, P., … Leyser, O. (2005). *MAX1* encodes a cytochrome P450 family member that acts downstream of *MAX3/4* to produce a carotenoid-derived branch-inhibiting hormone. *Developmental Cell*, *8*(3), 443–449. https://doi.org/10.1016/j.devcel.2005.01.009

Conn, C. E., Bythell-douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J. H., … Nelson, D. C. (2015). Parasitic Plants. *Science (New York, N.Y.)*, *2*(6247), 7–11.

De Cuyper, C., Struk, S., Braem, L., Gevaert, K., De Jaeger, G., & Goormachtig, S. (2017). Strigolactones, karrikins and beyond. *Plant Cell and Environment*, *40*(9), 1691–1703. https://doi.org/10.1111/pce.12996

Fernández-Aparicio, M., Rubiales, D., Bandaranayake, P. C. G., Yoder, J. I., & Westwood, J. H. (2011). Transformation and regeneration of the holoparasitic plant Phelipanche aegyptiaca. *Plant Methods*, *7*(1), 1–10. https://doi.org/10.1186/1746-4811-7-36

Gomez-Roldan, V., Fermas, S., Brewer, P. B., Puech-Pags, V., Dun, E. A., Pillot, J. P., … Rochange, S. F. (2008). Strigolactone inhibition of shoot branching. *Nature*, *455*(7210), 189–194. https://doi.org/10.1038/nature07271

Hegenauer, V., Körner, M., & Albert, M. (2017). Plants under stress by parasitic plants. *Current Opinion in Plant Biology*, *38*, 34–41. https://doi.org/10.1016/j.pbi.2017.04.006

Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., … Li, J. (2013). DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature*, *504*(7480), 401–405. https://doi.org/10.1038/nature12870

Joel, D. M., Bar, H., Mayer, A. M., Plakhine, D., Ziadne, H., Westwood, J. H., & Welbaum, G. E. (2012). Seed ultrastructure and water absorption pathway of the root-parasitic plant Phelipanche aegyptiaca (Orobanchaceae). *Annals of Botany*, *109*(1), 181–195. https://doi.org/10.1093/aob/mcr261

Joel, D. M., Chaudhuri, S. K., Plakhine, D., Ziadna, H., & Steffens, J. C. (2011). Dehydrocostus lactone is exuded from sunflower roots and stimulates germination of the root parasite Orobanche cumana. *Phytochemistry*, *72*(7), 624–634. https://doi.org/10.1016/j.phytochem.2011.01.037

Lin, H., Wang, R., Qian, Q., Yan, M., Meng, X., Fu, Z., … Wang, Y. (2009). DWARF27, an Iron-Containing Protein Required for the Biosynthesis of Strigolactones, Regulates Rice Tiller Bud Outgrowth. *The Plant Cell Online*, *21*(5), 1512–1525. https://doi.org/10.1105/tpc.109.065987

Lumba, S., Holbrook-Smith, D., & McCourt, P. (2017). The perception of strigolactones in

vascular plants. *Nature Chemical Biology*, *13*(6), 599–606. https://doi.org/10.1038/nchembio.2340
Morffy, N., Faure, L., & Nelson, D. C. (2016). Smoke and Hormone Mirrors: Action and Evolution of Karrikin and Strigolactone Signaling. *Trends in Genetics*, *32*(3), 176–188. https://doi.org/10.1016/j.tig.2016.01.002
Nakamura, H., Xue, Y. L., Miyakawa, T., Hou, F., Qin, H. M., Fukui, K., … Asami, T. (2013). Molecular mechanism of strigolactone perception by DWARF14. *Nature Communications*, *4*, 1–10. https://doi.org/10.1038/ncomms3613
Nelson, D. C., Flematti, G. R., Ghisalberti, E. L., Dixon, K. W., & Smith, S. M. (2012). Regulation of Seed Germination and Seedling Growth by Chemical Signals from Burning Vegetation. *Annual Review of Plant Biology*, *63*(1), 107–130. https://doi.org/10.1146/annurev-arplant-042811-105545
Nelson, D. C., Scaffidi, A., Dun, E. A., Waters, M. T., Flematti, G. R., Dixon, K. W., … Smith, S. M. (2011). F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences*, *108*(21), 8897–8902. https://doi.org/10.1073/pnas.1100987108
Parker, C. (2012). Parasitic Weeds: A World Challenge. *Weed Science*, *60*(2), 269–276. https://doi.org/10.1614/WS-D-11-00068.1
Parker, C. (2013). The Parasitic Weeds of the Orobanchaceae BT - Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.) (pp. 313–344). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38146-1_18
Pineda-Martos, R., Velasco, L., & Pérez-Vich, B. (2014). Identification, characterisation and discriminatory power of microsatellite markers in the parasitic weed *Orobanche cumana*. *Weed Research*, *54*(2), 120–132. https://doi.org/10.1111/wre.12062
Plakhine, D., Tadmor, Y., Ziadne, H., & Joel, D. M. (2012). Maternal tissue is involved in stimulant reception by seeds of the parasitic plant Orobanche. *Annals of Botany*, *109*(5), 979–986. https://doi.org/10.1093/aob/mcs016
Pujadas-Salvà, A. J., & Velasco, L. (2000). Comparative studies on *Orobanche cernua L.* And *O. Cumana Wallr.* (Orobanchaceae) in the Iberian Peninsula. *Botanical Journal of the Linnean Society*, *134*(4), 513–527. https://doi.org/10.1006/bojl.2000.0346
Snowden, K. C., Simkin, A. J., Janssen, B. J., Templeton, K. R., Loucas, H. M., Simons, J. L., … Klee, H. J. (2005). The Decreased apical dominance1/Petunia hybrida CAROTENOID CLEAVAGE DIOXYGENASE8 gene affects branch production and plays a role in leaf senescence, root growth, and flower development. *The Plant Cell*, *17*(3), 746–59. https://doi.org/10.1105/tpc.104.027714
Soundappan, I., Bennett, T., Morffy, N., Liang, Y., Stanga, J. P., Abbas, A., … Nelson, D. C. (2015). *SMAX1-LIKE/D53* Family Members Enable Distinct *MAX2*-Dependent Responses to Strigolactones and Karrikins in Arabidopsis. *The Plant Cell*, *27*(11), 3143–3159. https://doi.org/10.1105/tpc.15.00562
Stanga, J. P., Morffy, N., & Nelson, D. C. (2016). Functional redundancy in the control of seedling growth by the karrikin signaling pathway. *Planta*, *243*(6), 1397–1406. https://doi.org/10.1007/s00425-015-2458-2
Stanga, J. P., Smith, S. M., Briggs, W. R., & Nelson, D. C. (2013). SUPPRESSOR OF MORE AXILLARY GROWTH2 1 Controls Seed Germination and Seedling Development in Arabidopsis. *Plant Physiology*, *163*(1), 318–330.

https://doi.org/10.1104/pp.113.221259
Toh, S., Holbrook-smith, D., Stogios, P. J., Onopriyenko, O., Lumba, S., Tsuchiya, Y., … Mccourt, P. (2015). Structure-function analysis identifies highly sensitive strigolancone receptors in Striga. *Plant Science*, *350*(6257), 203–208.
Tsuchiya, Y., Yoshimura, M., Sato, Y., Kuwata, K., Toh, S., Holbrook-Smith, D., … Hagihara, S. (2015). Probing strigolactone receptors in *Striga hermonthica* with fluorescence. *Science*, *349*(6250), 864–868. https://doi.org/10.1126/science.aab3831
Umehara, M., Cao, M., Akiyama, K., Akatsu, T., Seto, Y., Hanada, A., … Yamaguchi, S. (2015). Structural requirements of strigolactones for shoot branching inhibition in rice and arabidopsis. *Plant and Cell Physiology*, *56*(6), 1059–1072. https://doi.org/10.1093/pcp/pcv028
Umehara, M., Hanada, A., Yoshida, S., Akiyama, K., Arite, T., Takeda-Kamiya, N., … Yamaguchi, S. (2008). Inhibition of shoot branching by new terpenoid plant hormones. *Nature*, *455*(7210), 195–200. https://doi.org/10.1038/nature07272
Wallner, E. S., López-Salmerón, V., Belevich, I., Poschet, G., Jung, I., Grünwald, K., … Greb, T. (2017). Strigolactone- and Karrikin-Independent SMXL Proteins Are Central Regulators of Phloem Formation. *Current Biology*, *27*(8), 1241–1247. https://doi.org/10.1016/j.cub.2017.03.014
Wang, L., Wang, B., Jiang, L., Liu, X., Li, X., Lu, Z., … Li, J. (2015). Strigolactone Signaling in Arabidopsis Regulates Shoot Development by Targeting *D53-Like SMXL* Repressor Proteins for Ubiquitination and Degradation. *The Plant Cell*, *27*(11), 3128–3142. https://doi.org/10.1105/tpc.15.00605
Waters, M. T. (2017). From little things big things grow: Karrikins and new directions in plant development. *Functional Plant Biology*, *44*(4), 373–385. https://doi.org/10.1071/FP16405
Waters, M. T., Gutjahr, C., Bennett, T., & Nelson, D. C. (2017). Strigolactone Signaling and Evolution. *Annual Review of Plant Biology*, *68*(1), 291–322. https://doi.org/10.1146/annurev-arplant-042916-040925
Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends in Plant Science*, *15*(4), 227–235. https://doi.org/10.1016/j.tplants.2010.01.004
Yao, R., Ming, Z., Yan, L., Li, S., Wang, F., Ma, S., … Xie, D. (2016). DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature*, *536*(7617), 469–473. https://doi.org/10.1038/nature19073
Yao, R., Wang, F., Ming, Z., Du, X., Chen, L., Wang, Y., … Xie, D. (2017). *ShHTL7* is a non-canonical receptor for strigolactones in root parasitic weeds. *Cell Research*, *27*(6), 838–841. https://doi.org/10.1038/cr.2017.3
Yoder, J. I., & Scholes, J. D. (2010). Host plant resistance to parasitic weeds; recent progress and bottlenecks. *Current Opinion in Plant Biology*, *13*(4), 478–484. https://doi.org/10.1016/j.pbi.2010.04.011
Yoneyama, K., Ruyter-Spira, C., & Bouwmeester, H. (2013). Induction of Germination. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.), *Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies* (pp. 167–186).
Zhao, L. H., Edward Zhou, X., Wu, Z. S., Yi, W., Xu, Y., Li, S., … Eric Xu, H. (2013). Crystal structures of two phytohormone signal-transducing α/β hydrolases: Karrikin-signaling *KAI2* and strigolactone-signaling DWARF14. *Cell Research*, *23*(3), 436–439.

https://doi.org/10.1038/cr.2013.19
Zhao, L. H., Zhou, X. E., Yi, W., Wu, Z., Liu, Y., Kang, Y., … Xu, H. E. (2015). Destabilization of strigolactone receptor DWARF14 by binding of ligand and E3-ligase signaling effector DWARF3. *Cell Research*, *25*(11), 1219–1236. https://doi.org/10.1038/cr.2015.122
Zwanenburg, B., & Pospíšil, T. (2013). Structure and activity of strigolactones: New plant hormones with a rich future. *Molecular Plant*, *6*(1), 38–62. https://doi.org/10.1093/mp/sss141
Zwanenburg, B., Pospíšil, T., & Ćavar Zeljković, S. (2016). Strigolactones: new plant hormones in action. *Planta*, *243*(6), 1311–1326. https://doi.org/10.1007/s00425-015-2455-5

Chapter 2*

The *Orobanche cumana* x *Orobanche cernua* genetic system provides insight into the regulation of germination specificity in a parasitic plant.

*This chapter is formatted to conform to New Phytologist submission requirements.

Author contributions: Hybrid lines were generated by DP. Parental RNA was isolated by HE and YT and hybrid genomic DNA was isolated by HL. NW developed the statistical model, with input from HL. DN provided the *kai2* construct and advised on the complementation assay. HL did all other laboratory work, bioinformatics analyses, and wrote the manuscript with input from JW.

The *Orobanche cumana* x *Orobanche cernua* genetic system provides insight into the regulation of germination specificity in a parasitic plant.

[1]Larose, H., [2]D. Plakhine, [3]N. Wycoff, [2]D.M Joel, [2]H. Eizenberg, [2]Y. Tadmor, [4]D. Nelson and [1]J. Westwood

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

[2]Newe Ya'ar Research Center, A.R.O., Israel

[3]Department of Statistics, Virginia Tech, USA

[4]Department of Botany and Plant Sciences, University of California, Riverside, USA

Abstract

- Strigolactones (SLs) are the most studied class of germination stimulants for Orobanchaceae seeds, and recent work implicates that SLs are perceived by members of the *KARRIKIN-INSENSITIVE2* (*KAI2)* gene family. Parasitic plants appear to have undergone duplication and specialization of *KAI2* genes, resulting in a group of divergent *KAI2* (*KAI2d*) genes that are thought to be responsible for enabling parasite seeds to identify specific host plants. However, this concept has not been fully tested in the context of parasitic plants that differ in germination specificity.
- We used genetic and genomic approaches to investigate the role of *KAI2d* genes in the germination specificity in *Orobanche cernua* and *O. cumana*, closely related species that differ in germination stimulant specificity. Whereas *O. cernua* parasitizes tomato and responds to the SL orobanchol (Oro), *O. cumana* parasitizes sunflower and responds to dehydrocostus lactone (DCL). Crosses between the two species produced hybrids that segregate for stimulant specificity, creating a tractable genetic system. Each *KAI2d* gene was also assayed for functionality in a *kai2 arabidopsis* mutant background.
- *Orobanche cernua* contains four *KAI2d* genes (*OrceKAI2d1-4*), while *O. cumana* contains six genes (*OrcuKAI2d1-6*). Analysis of hybrid lines indicated that the *KAI2d* genes appear to be linked, but the *O. cernua* gene, *OrceKAI2d2,* was associated with SL response. This was

confirmed by the response of *OrceKAI2d2* to SLs in the heterologous assay. In contrast, only weak evidence was found for the correlation of an *O. cumana KAI2d* gene with response to DCL. Interestingly, one *O. cumana* gene, *OrcuKAI2d5* (the homolog of *OrceKAI2d2*), was responsive to SLs.

- This work fits with the current model of SL signaling perception in parasitic plants with respect to *O. cernua*, with *OrceKAI2d2* as a primary receptor for Oro. However, the data do not explain how *O. cumana* avoids germinating in response to SLs despite having at least one *KAI2d* gene that responds to Oro, nor how *O. cumana* is able to detect the non-SL stimulant DCL. We conclude that additional genes are involved in regulating stimulant perception in parasite seeds.

Key words: Parasitic plants, germination stimulant, germination, *Orobanche cernua, Orobanche cumana*, strigolactone, dehydrocostus lactone, *KAI2*.

2.1 Introduction

Parasitic plants cause major agricultural, and resulting economical, damage to many areas across the world, including the Mediterranean, Asian, Eastern European and African regions (Parker, 2013). This damage results it an annual loss of approximately 1$ billion US dollars, and affects the food supply of over a hundred million people (Hegenauer et al., 2017).

Some of the most severe agricultural parasitic plants belong to a class called root parasites and include *Striga, Phelipanche and Orobanche spp*., and their severity can be attributed to numerous factors (Parker, 2012). First, these parasitic plants spend the majority of their lifecycle below ground, emerging solely for reproduction, prohibiting visual detection until the damage has already been done to the crop. Second, once above ground, each plant is able to release upwards of hundreds of thousands of microscopic seeds, which are easily transmissible to both local and neighboring plots. These seeds can build up in the soil and persist for years, causing plots to be unusable for growing host crops. Third, there is a lack of effective strategies to control these parasites, before germination or pre- and post-emergence. This is in part due to the fact that the parasite and the host share a physical connection, making selective chemical treatments difficult (Westwood et al., 2010).

Root parasitic plants of the genera Orobanchaeae have evolved mechanisms to ensure that their seeds germinate only in the presence of an acceptable host. The first requirement for *Orobanche* seed germination is a period termed conditioning, which is when the seeds imbibe in water, causing the seed to swell, and an opening of the micropyle (Joel et al., 2012). Once conditioning is complete, these seeds require the perception of a host-derived germination stimulant. For most members of Orobanchaceae this stimulant is a natural byproduct of plant development, strigolactones (SL). Parasitic plants have evolved a way to detect specific SL variants, singly or in combinations, from hosts. This signal perception is crucial, because once germinated, the parasite has only about 72 hours to make a physical connection with a host-root to form the haustoria feeding structure before the limited nutrients stored within the seed are exhausted (Westwood et al., 2010).

Much work has gone into understanding the mechanism by which parasites are able to detect the various SLs in their environment, and the key SL receptors have been identified as members of a gene family called *KAI2* (KARRIKIN-INSENSITIVE-2) (Conn et al., 2015; Toh

et al., 2015; Tsuchiya et al., 2015; Yao et al., 2017). The *KAI2* orthologs from parasitic plants have been classified into three major phylogenic clades: conserved, intermediate and diverged, designated *KAI2c*, *KAI2i* and *KAI2d*, respectively (Conn et al., 2015). Genes of the conserved clade presumably respond to a yet unknown karrikin-like ligand endogenous to plants, the intermediate clade responds to karrikins (KARs), and the diverged clade, which is thought to be the fastest evolving clade, responds to SLs. Numerous studies have demonstrated that a specific group of *KAI2d* genes, termed *ShHTL4-9,* are sufficient for inducing germination in *Striga hermonthica* through a variety of biochemical assays (Toh et al., 2015; Tsuchiya et al., 2015; Yao et al., 2017).

However, not all parasitic plants respond to SLs as germination stimulants. Notably, *Orobanche cumana,* a major agricultural parasite on cultivated sunflower *(Helianthus annuus),* responds to dehydrocostus lactone (DCL) rather than SLs. Its closely related relative, *O. cernua,* responds to the SL orobanchol (Oro), and parasitizes *Solanaceous* crops, like tomato (*Solanum lycopersicon*) and eggplant (*Solanum melongenae*). Both *O. cernua* and *O. cumana* respond to the synthetic strigolactone, GR24. Host specificity in these species is regulated primarily at the level of germination signals, because once germinated, each species can grow on either tomato or sunflower hosts. These species present an interesting contrast in that they have been considered to be two forms of the same species, differing primarily in host preference (Parker & Riches, 1993), but are now treated as different species. Within species of parasitic plants, such pronounced variation in germination specificity has not been observed, so the *O. cernua* / *O. cumana* complex offers a unique system in which to explore germination mechanisms.

Here we describe experiments using the *O. cernua* x *O. cumana* genetic system to identify the key components underlying germination specificity. Transcriptomes of each species were sequenced to identify candidate stimulant receptors, revealing differences in the *KAI2d* gene family. Hybrid lines differing in stimulant response phenotype were genotyped and certain *KAI2d* genes were loosely correlated with stimulant response. Cloning each *KAI2d* gene and expressing it in a heterologous system indicated one form that recognizes SLs, but no receptor for DCL was identified. Taken together, these results indicate that *KAI2d* proteins are involved in parasitic plant stimulant perception to SLs, but the currently model fails to completely explain DCL stimulant specificity.

2.2 Methods

2.2.1 Seed sources

Orobache cernua Loefl. seeds were collected in tomato fields in the Upper Galilee, Israel (1994), and *O. cumana* Wallr. seeds were collected in sunflower fields in the Lower Galilee, Israel (1998). Using these seeds, *O. cernua* and O. *cumana* were then grown each year in a net-house at Newe Ya'ar Research Center for seed production for seed stock replenishing.

2.2.2 Hybrid lines

Crosses of *Orobanche cernua* and *O. cumana* are described in Plakhine et al (2012). These lines were selfed to produce the F_3 generation, at which point the stimulant response phenotype segregates (Plakhine et al., 2012). Presumably, segregation is seen in the F_3 generation due expression of the stimulant perception mechanism in the maternally derived perisperm tissue of the seeds (Plakhine et al., 2012). The lines were classified into one of five phenotypic categories, responding to: DCL and GR24, Oro and GR24, DCL, Oro and GR24, GR24 only, or DCL and Oro.

2.2.3 Germination bio-assay

Seeds were surface sterilized according to Plakhine et al., 2012. Between 30-50 seeds were sown on 6-mm glass fiber discs (Watman GFA). Six such discs were placed in a petri dish (lined with a later of filter paper wetted with 1mL of water). The petri dishes were stored at 23C for seven days. During conditioning, seeds were checked once daily for spontaneous germination for seven days. After conditioning, the discs were blotted to remove water and transferred to a new petri dish containing 1 p.p.m DCL, Oro or GR24. The petri dishes were then placed in 23C for ten days and scored for germination.

2.2.4 Tissue collection for RNA-sequencing

Orobanche cernua and *O. cumana* seeds were surface sterilized according to Plakhine et al., 2012. For conditioning, seeds were placed in a petri dish (lined with a later of filter paper wetted with 300 µl of water) and stored in the dark at 22C for 1, 3 or 5 days, and pooled for RNA extraction. For conditioned, seeds were placed in a petri dish (lined with a layer of filter paper wetted with 300 µl of water) and stored in the dark at 22C for seven days. For stimulated, seeds

were conditioned for seven days and exposed to either species-specific stimulant, Oro or DCL, at 10-8, 10-7 M, respectively, or the synthetic germination stimulant, GR24 10-6 M, for four or eight hours, and pooled for RNA extraction.

2.2.5 De novo transcriptome assembly of O. cernua and O. cumana

Raw read quality was assessed using FastQC (Andrews, 2014). Prior to assembly, raw reads were trimmed to remove poor quality reads and Illumina adaptor sequences using Trimmomatic (Bolger et al., 2014). A minimum length of 50bp after trimming was required to retain the read. The reads that retained their paired-end mate were used for sequencing. The Trinity software package (version 2.4.0) was used for de novo transcriptome assembly of *O. cernua* and *O. cumana* transcriptomes using default parameters (Haas et al., 2014). Transcriptomes were assembled for each species using the processed PE reads from all sequenced stages of that species. Raw reads were mapped back to the de novo transcriptomes using Bowtie2 software and had a mapping efficiency of greater than ninety percent for both species (Langmead & Salzberg, 2012). The Core Eukaryotic Genes Mapping Approach (CEGMA) pipeline was used to estimate the completeness of the transcriptome assemblies. The CEGMA pipeline contains a set of 458 highly conserved proteins demonstrated to be present within virtually all eukaryotic organisms, which is searched within our transcriptomes to measure which genes were properly captured and assembled (Parra, Bradnam, & Korf, 2007).

2.2.6 KAI2 gene identification

KAI2 and *D14* gene sequences were identified for each species through BLAST tblastx search of *Arabidopsis thaliana KAI2d* sequences using default parameters with an e-value of $< 10^{-5}$ (Altschul et al., 1997). All contig hits were extracted from each transcriptome and aligned against the *O. cumana* and *O. cernua KAI2d* genes identified by Conn, et al. (Conn et al., 2015). Two additional *KAI2* genes per species were found, termed *OrcuKAI2i1-2* and *OrceKAI2i1-2*, by analyzing the additional *KAI2* hits for correct motifs. Sequences were validated by Reverse Transcriptase (RT) PCR amplication from seven-day conditioned *O. cumana* and *O. cernua* RNA and Sanger sequenced.

2.2.7 Genomic DNA preparation and sequencing

F_2 hybrid tissue for the targeted sequence capture assay was collected based on segregation ratios seen in F_3 seed germination. Floral tissue was collected from a total of 94 F_2 hybrids. Genomic

DNA was extracted using a Cetyltrimethylammonium bromide (CTAB)-based extraction method. Floral tissue was ground under liquid nitrogen using a mortar and pestle and mixed with 500uL of 2X CTAB buffer (2% CTAB, 100mM Tris-Cl, 20mM EDTA, 1.4M NaCl, and 0.1% beta-mercaptoethanol, pH 8.0) and incubated at 60C for twenty minutes to an hour. After incubation, the sample was mixed with 500uL of chloroform and centrifuged to separate phases. The upper, aqueous phase was transferred to a fresh reaction tube and genomic DNA was precipitated using isopropanol stored at 4C, followed by centrifugation to pellet the gDNA. The DNA pellet was rinsed with 100% ethanol, and dried for one hour. Pelleted DNA was resuspended with 1X Tris-EDTA buffer (10mM Tris-HCL, 0.1mM EDTA, pH 8.0) overnight at 4C. The quality and concentration of the genomic DNA was checked using a Nanodrop One (ThermoFisher) and gel electrophoresis.

2.2.8 Primer design and testing

Three primer sets capable of amplifying all *OrceKAI2d1-4* and *OrcuKAI2d1-6* genes were used to produce polymerase chain reaction (PCR) products sufficient for Illumina sequencing. Primers that selectively amplify sets of *KAI2* genes were designed based on regions with common alignments. These were termed 'Universal Primers' and used to efficiently amplify *O. cumana KAI2d1-6* and *O. cernua KAI2d1-4* genes from parental lines and hybrids. Each primer was diluted to a working concentration of 10µM, pooled, and tested on parental genomic DNA to assure amplification of all parental *KAI2d* genes using iProof High-Fidelity DNA Polymerase (BioRad #172530). Primer sequences are provided in Supplemental Table 1.

2.2.9 MiSeq library preparation and sequencing

Each F_2 hybrid genomic DNA sample was diluted to a final concentration of 10ng per µL, with 1 µL used in each 50µL reaction, and amplified with the pooled-Universal Primers using iProof High-Fidelity DNA Polymerase. Libraries were prepared using a Tn5 transposase which simultaneously fragments DNA to sizes less than 1000bp while ligating Nextera sequencing primers and indexing barcodes to each sample (Adey et al., 2010). DNA libraries were loaded to MiSeq using Illumina MiSeq Reagent Kit v3 (600 cycles) (Illumina, MS-102-3003) to generate 300bp paired-end reads. Upon completion of MiSeq run, over 30 million raw reads were generated. Data were split into fastq files based on indexing sequencing of each sample in preparation for analyses. Cutadaptor was used to trim Tn5 adaptor sequences (Martin, 2011).

2.2.10 KAI2d gene assignment

Variant calling was done through GATK to determine the population of SNPs for each reference gene and 'marker SNPs' were identified (i.e SNPs that were true between orthologs and not due to allelic diversity) (McKenna et al., 2010). Raw reads were aligned to reference genes using BBMap software allowing up to one mismatch with zero gaps or substitutions (Bushnell, 2016). To assign whether a gene was present, the alignments were checked manually for every sample using IGV Genome Browser and results were recorded in Figure 2.2 (J. T. Robinson et al., 2011).

2.2.11 Phenotypic assignment of F_2 hybrid germination response

F_2 hybrids were classified as responding to either: 1) GR24 and Oro, 2) GR24 and DCL 3) GR24, Oro and DCL 4) GR24 only or 5) Oro and DCL. A hybrid line was considered responsive to a stimulant if more than five percent germination was observed when exposed to a final following concentration DCL at 10-7 M, Oro 10-8 M or GR24 at 10-6 M. Germination was counted after 7 days exposure to stimulant. Spontaneous germination was accounted for by measuring germination of seeds exposed to pure water. Maximum possible germination for a given line was represented by germination in response to GR24, accounting for spontaneous germination.

2.2.12 Statistical evaluation and germination modelling

To check if the germination response of the 94 F_2 hybrids could be explained through the presence or absence of a single *KAI2d,* or a combination of *KAI2d* genes, we fit a Nested Generalized Linear Mixed-Effects Model (GLMM) using R (R Development Core Team, 2017) and JAGS (Plummer, 2003) on the resulting genotypes obtained through the targeted sequence capture assay. Through the GLMM, the number of seeds germinating under each experimental condition was modeled as a binomial random variate, with germination probability modeled through the Probit function as the additive effect of each *KAI2d* gene. Random effects were fit for Petri dishes nested within parent plants in order to account for Petri dish to Petri dish variation as well as unmeasured characteristics of parent plants, genetic and otherwise, termed 'plant effects'.

A Bayesian approach was taken in parameter estimation. Bayesian statistics is most well known as a rigorous method to combine subjective prior information with data to perform inference, but it also allows for exact inference in cases where asymptotic methods had been

used classically (Hoff, 2009). In our case, we seek minimally informative prior distributions to conduct an objective analysis, and follow the advice of (Gelman, 2011) to this end. Marginal normal distributions were selected for fixed effects and marginal half normal for the random effects, both with mean 0 and a variance of 100. Priors for random effects variances are known to be more sensitive than priors for fixed effects (Fong et al., 2010) so we conducted a sensitivity analysis on the random effects prior by changing the prior variance hyper parameter to 1 and to 100 and found that it did not affect any of this article's conclusions.

Presence of *OrcuKAI2d3* and *OrcuKAI2d5* were found to be highly correlated ($\hat{\rho} = 0.98$), and therefore only their combined effect could be analyzed (the combined gene is marked as present if either *OrcuKAI2d3* or *OrcuKAI2d5* or both is present). Disambiguation of the combined effect would require further experimentation. JAGS Markov Chain Monte Carlo (MCMC) was found to converge by visual assessment of two chains: 400,000 sampling iterations were run after a 10,000 iteration burn in with a thinning rate of 100 to achieve 4,000 samples from the posterior for each quantity for each chain, and 8,000 in total posterior samples for each parameter.

In order to assess significance, both practical and statistical, of model parameters, we provide symmetric 95% posterior credible intervals, as well as full posterior distributions approximated by MCMC in Supplementary Materials (Supplementary Figure 2.3, Supplementary Table 2.3). Interpretation is of model parameters in the Probit space is difficult beyond an intuition that positive effects indicate that gene presence corresponds to higher germination rates, negative values the opposite, and magnitudes determine the strength of association. A gene is found to have a significant relationship if its posterior credible interval contains a value far from zero.

2.2.13 Functional complementation of A. thaliana kai2

KAI2 coding sequences were amplified from seven-day conditioned RNA of *O. cernua* and *O. cumana* through reverse-transcription PCR, and cloned into pENTR/D-TOPO (ThermoFisher). Primer sequences are listed in table Supplementary Table 3.4. Entry cloned were verified using Sanger sequencing and transferred into pKAI2pro-GW (a Gateway compatible vector containing the *Arabidopsis KAI2* promotor) through Gateway recombination (Conn et al., 2015). Destination vectors were transformed into *Agrobacterium tumefaciens* strain GV3101. *Arabidopsis kai2-2* mutants were transformed through agrobacterium mediated transformation

by floral dip method (Clough & Bent, 1999). Transformed plants (T_1), were selected on 0.5x Murashige-Skoog media supplemented with hygromycin (25ug/mL) according to the protocol developed by Harrison, et al. (2006) (Harrison et al., 2006). The seed of T_1 transgenic lines showing a segregation ratio of 3:1 hygromycin resistance were used in germination assays.

2.2.14 Arabidopsis thaliana growth and selection

Plants were grown in 10-hours light/14-hours dark, at 22C for two weeks, then transferred to 24 hours continuous light, at 22C. Plants were harvested into paper bags when siliques were brown and the stem was green to maintain primary dormancy, and dried at room temperature for three days before seed harvest. Seeds were stored at -80C until used for germination assay to maintain primary dormancy.

2.2.15 Arabidopsis thaliana germination assay

Seeds were surface sterilized for 2 minutes in 50% (v/v) bleach with 0.1% sodium dodecyl sulfate (w/v), rinsed with sterile dionized water 3 times, resuspended in 95% EtOH and immediately dried on sterile filter paper. Germination assays were performed with surface sterilized seed plated on 2-(N-morpholino)ethanesulfonic acid (MES) media (pH 5.7), with 0.8% (w/v) BactoAgar, supplemented with either 0.1% acetone or 1 μM concetrations of the following stimulants: GR24, DCL, GR24, 5-deoxystrigol (5DS) or Oro.

Plated seed was grown under twenty-four hour continuous light at 22C. Germination was scored every 24 hours for up to five days or until germination of the control exceeded 70%, whichever came first. Germination was defined as complete protrusion of the radical through the endosperm. Between four and ten independent seed lines were tested for each transgene, and greater than 50% of the lines must show a significant response to a stimulant for the transgene to be considered as conferring a response.

2.2.16 Statistical evaluation of germination assay

Response to a stimulant was measured by mean germination percent. For each stimulant, mean germination percent was calculated by averaging the germination percent of three replicates. The software JMP (JMP, Version 13.0, SAS Institute Inc., Cary, NC, 1989-2007) was used to calculate the standard error (SE) and compare all means using Tukey-Kramer HSD.

2.2.17 Data availability

All raw reads from the transcriptome sequencing will be deposited at NCBI.

2.3 Results

2.3.1 Transcriptome assembly

The transcriptomes of *O. cernua* and *O. cumana* were sequenced at three stages prior to germination: during conditioning, at completion of conditioning, and after treatment with germination stimulants. The final stage was further divided into treatments with species-specific (DCL or Oro) vs. general stimulant (GR24). The transcriptomes of *O. cernua* and *O. cumana* were assembled and produced over 200,000 contigs per species. The transcriptomes of each species were roughly equivalent in sequencing depth and number of predicted ESTs, with 103,570 ESTs identified for *O. cernua* and 110,019 for *O. cumana* (Table 1). The CEGMA pipeline was used to estimate transcriptome completeness, and each transcriptome assembled 99% percent of CEGMA proteins. Transcriptome statistics are summarized in Table 1.

Table 2.1 *Orobanche cernua* and *O. cumana* transcriptome statistics.

	O. cernua	***O. cumana***
Raw Reads	391,787,707	433,836,952
Number of assembled contigs	204,992	216,881
N50	1,631	1,572
Mean contig length	940.75	895.6
Total assembled bases	192,845,756	194,239,216
Mapping % of input reads (Bowtie2)	97.56%	97.70%
Predicted ESTs	103,570	110,019

2.3.2 Identification of putative stimulant receptor genes

Transcriptomes of *O. cernua and O. cumana* were searched to identify transcripts of genes involved in SL perception. Single copies of *D14* and *MAX2* were identified within each species' transcriptome. The amino acid sequences of *D14* were found to be identical between the species, with just one synonymous single nucleotide polymorphism (SNP). *MAX2* sequences revealed two non-synonymous SNPs between the species, which resulted in two amino acid changes

(Supplemental Figure 2.1). However, comparison within a subset of F_2 hybrids showed that neither amino acid change correlated with the hybrid response to stimulant. Next, the *KAI2d* genes that were reported by Conn et al. (2015) as being evolved to respond to SLs were detected, identifying four genes from *O. cernua* and six genes from *O. cumana* (designated *OrceKAI2d1-4 and OrcuKAI2d1-6*, respectively). Subsequently, two previously unreported *KAI2* genes from each species were detected and named *OrceKAI2i1-2 and OrcuKAI2i1-2*. All of the *KAI2* genes were expressed within the assembled transcriptomes (Supplemental Figure 2.2).

To validate the transcriptome assemblies, each *KAI2d* and *KAI2i* gene was cloned and sequenced from RNA to ascertain the expressed form of the gene and from genomic DNA obtain intron sequences (fasta sequences provided in Supplementary File 1). All *KAI2* genes share the same structure of two exons and one intron, with coding regions ranging in size from 810 to 825 bp. Sequences of *D14* and conserved *KAI2* (*KAI2c* as defined by Conn et al., 2015) genes were also verified and included in the analysis of relationships among the orthologous pairs from each species (Figure 2.1).

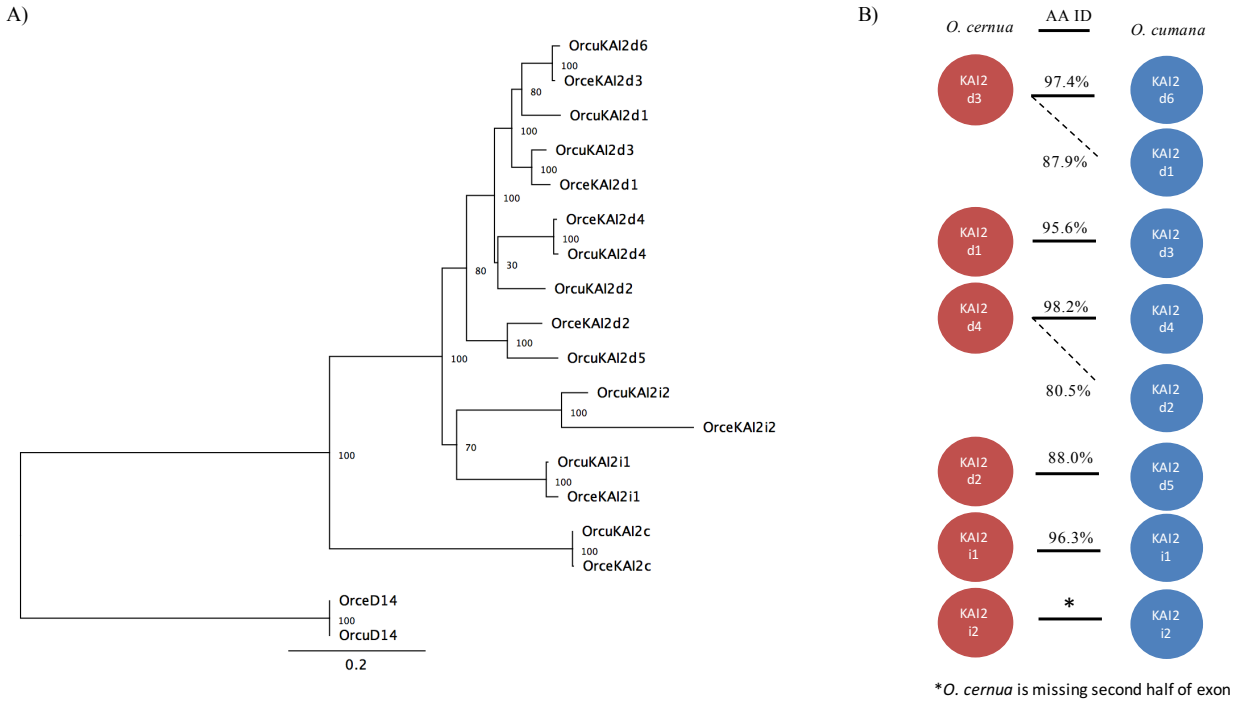


Figure 2.1 A) Relationship among *O. cernua* and *O. cumana KAI2* and *D14* genes based on coding sequences. RAxML bootstrapping values shown, using *D14* as the outgroup. B) Schematic depiction of *O. cernua* and *O. cumana KAI2* ortholog relationships. Percent amino acid identity (AA ID) is indicated between closest orthologues. s

2.3.3 Targeted sequence capture assay

Genomic DNA was extracted from parental lines and 94 *O. cernua x O. cumana* F_2 hybrid individuals segregating for stimulant specificity and used for *KAI2d* genotyping. Due to the potentially high number of *KAI2d* genes in the hybrids, PCR primer sets were developed that were capable of amplifying all *OrceKAI2d1-4* and *OrcuKAI2d1-6* genes, and the resulting products were bar-coded and sequenced using Illumina 300bp paired-end read technology. This yielded over 30 million raw reads for all 94 F_2 hybrid individuals, and these were used to produce a *KAI2d* genotype for each hybrid. The hybrid phenotypes were categorized by responsiveness to DCL, Oro, or both, but recognizing that all lines germinated in response to GR24. Two additional phenotype categories found in just five hybrid lines were response to either Oro and DCL but not GR24, or to GR24 only. In general, lines responding to DCL contained *KAI2d* genes from *O. cumana*, while lines responding to Oro contained *KAI2d* genes from *O. cernua.* However, there were multiple exceptions to this, and that the inheritance of the *KAI2d* genes usually were inherited in blocks, making it difficult to identify a single gene as being responsible for conferring specificity to Oro or DCL (Figure 2.2). Correlation analysis of *KAI2d* genotypes of the F_2 hybrid individuals showed strong associations between these groups and suggest that *KAI2d* genes are linked in the *O. cernua* and *O. cumana* genomes (Figure 2.3).

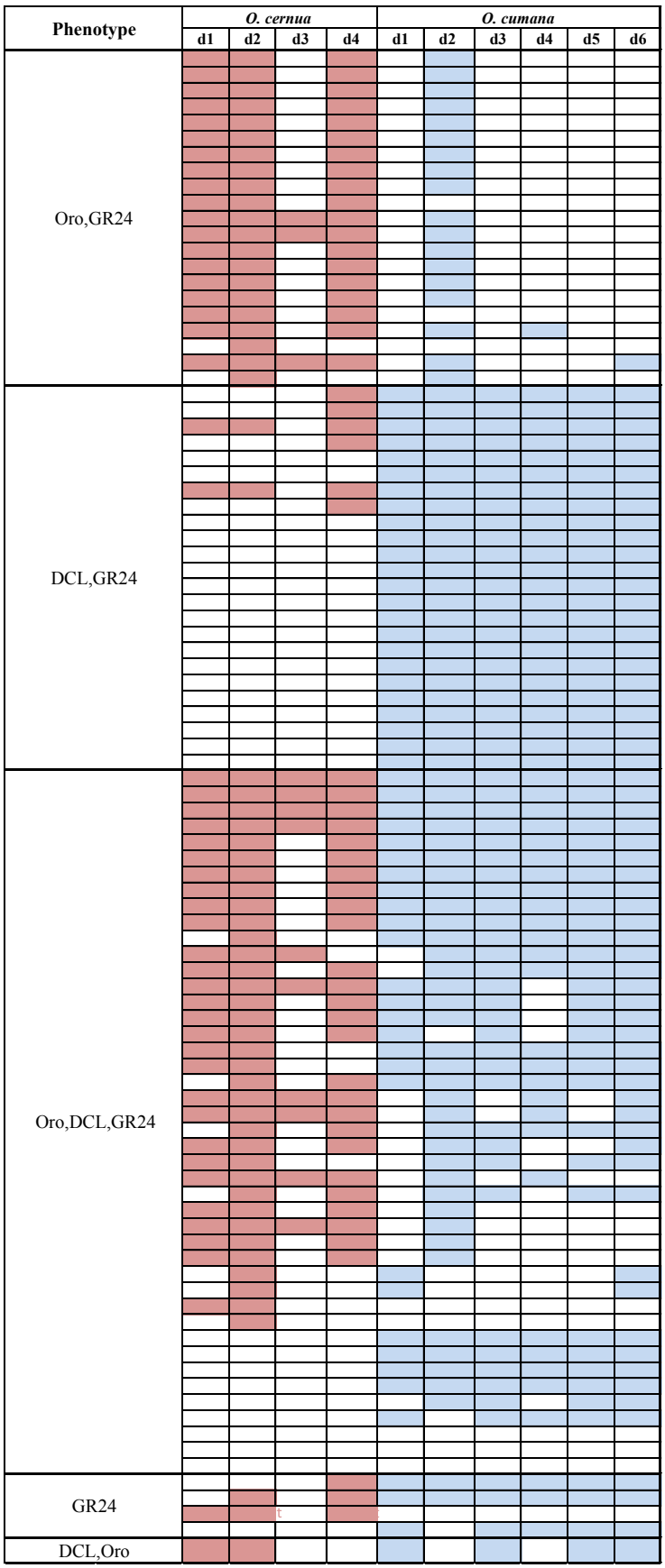

Phenotype
O. cernua
O. cumana
d1
d2
d3
d4
d1
d2
d3
d4
d5
d6
Oro,GR24
DCL,GR24
Oro,DCL,GR24
GR24
DCL,Oro

Figure 2.2 *O. cernua* and *O. cumana* F_2 hybrid *KAI2d* gene presence and absence for each category of stimulant response: F_2 hybrid lines responding to DCL and GR24, F_2 hybrid lines responding to Oro and GR24, F_2 hybrid lines responding GR24 only, F_2 hybrid lines responding DCL, Oro and GR24 and F_2 hybrid lines responding DCL and Oro only.

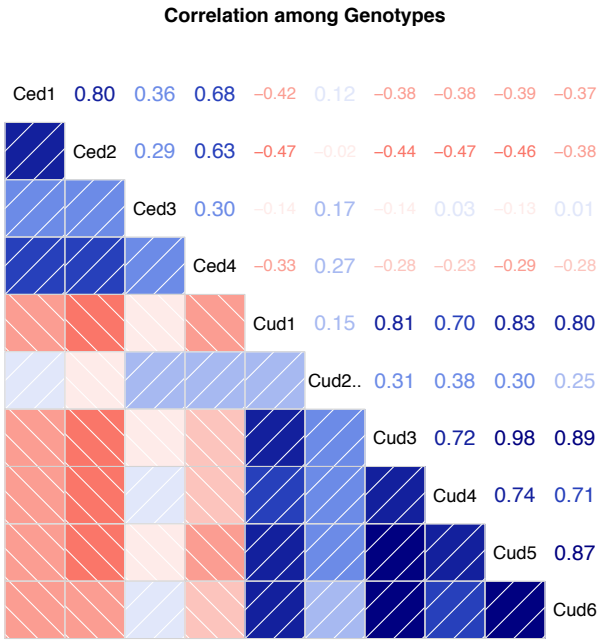


Figure 2.3 Corrgram showing the correlation amongst *KAI2d* genotypes of the 94 F_2 hybrids. Darker blue indicates gene pairs are more highly correlated while darker red indicates less likelihood of correlation.

2.3.4 Germination modelling

Given the lack of clear causality in the hybrid line analysis, a statistical approach was used to correlate gene presence in the hybrid lines to their germination rates, which provides a more nuanced phenotype than simply germinated or not. Hybrid seed populations germinated at differential levels, with some lines having a response greater than 80% to a stimulant, with others as low as 10%. While both are considered responsive to that stimulant by our criteria, it begs the question whether the *KAI2d* genes are providing an additive effect on germination response. To this end, a Nested Generalized Linear Mixed-Effects Model (GLMM) was fit to the germination responses and genotypes of the 94 F_2 hybrids using R (R Development Core Team, 2017) and JAGS (Plummer, 2003) in response to each stimulant.

To assess significance, both practical and statistical, of model parameters, symmetric 95% posterior credible intervals were calculated, as well as full posterior distributions approximated by MCMC in Supplementary Materials (Supplementary Figure 2.3, Supplementary Table 2.2) for each gene's effect on germination when exposed to each stimulant. Interpretation of model parameters in the Probit space is difficult. However, intuition dictates that positive effects indicate that gene presence corresponds to higher germination rates, negative values the opposite, and magnitudes determine the strength of association. A gene is found to have a significant relationship if its posterior credible interval contains value far from zero.

For the DCL response, the only gene with a significant credible interval is *OrcuKAI2d6* (Figure 2.4a), and even then, the posterior still places a fair amount of mass around 0, indicating that it is possible that there is no correlation. For the Oro response, the model strongly indicates that *OrceKAI2d2* co-occurs with increased germination (credible interval does not contain 0) and, to a lesser extent, *OrceKAI2d1* presence is associated with higher germination (Figure 2.4b). Other genes were not found to be correlated with higher or lower germination rates in this study (Supplemental Figure 2.3).

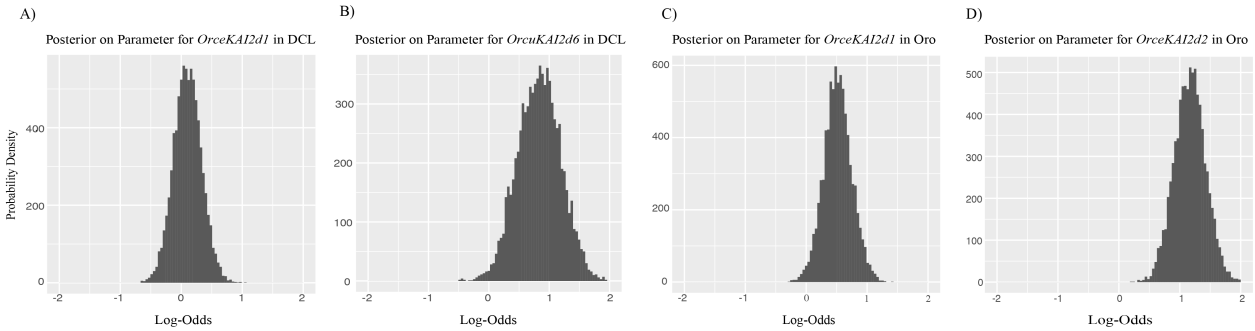


Figure 2.4 Posterior distribution charts for A) *OrceKAI2d1* response to DCL. A) *OreKAI2d1* response to DCL. B) *OrcuKAI2d6* response to DCL. C) *OrceKAI2d1* response to Oro D) *OrceKAI2d2* response to Oro. Higher values of density indicate that it is more probable that the truth lies in that region. Sign of log-odds indicates direction of correlation, and magnitude indicates strength of correlation. The Posterior for *OrceKAI2d2* is most removed from zero, and so seems to have the strongest correlation. Panel A shown to illustrate a posterior distribution of a gene with limited to no contribution to germination response.

2.3.5 Cross-species complementation assay

Each of the 16 *O. cernua* and *O. cumana KAI2* genes was aligned to *AtKAI2* and *AtD14* and evaluated for the presence of amino acid residues necessary for interacting with *AtMAX2*

(Bythell-Douglas et al., 2017; Zhao et al., 2015). All *KAI2* genes (including conserved, intermediate and diverged) from both species were cloned into a construct in fusion with the native *Arabidopsis KAI2* promoter and transformed into the *Arabidopsis kai2-2* mutant background. The resulting plants were assayed for their ability to recover the delayed germination phenotype when exposed to SLs (Oro, 5DS and GR24) and DCL. *OrceKAI2i2* (which results in a truncated protein) and an empty vector control (EVC) were included as negative controls. Four to ten independent lines were tested for each *KAI2d* and *KAI2i* construct (Supplemental Figure 2.4). Only one *O. cernua KAI2d* gene, *OrceKAI2d2,* was able to recover the mutant phenotype when exposed to SLs. The homolog of this gene in *O. cumana, OrcuKAI2d5,* was also able to recover the mutant phenotype in response to strigolactones (Figure 2.5). Within *O. cumana,* both *OrcuKAI2i1* and *OrcuKAI2i2* were also able to recover the mutant phenotype in response to SLs, whereas the *O. cernua* homologs showed no response (Figure 2.5). No genes were found to consistently and specifically recover the mutant phenotype in response to DCL (Table 2.2, Supplemental Figure 2.4).

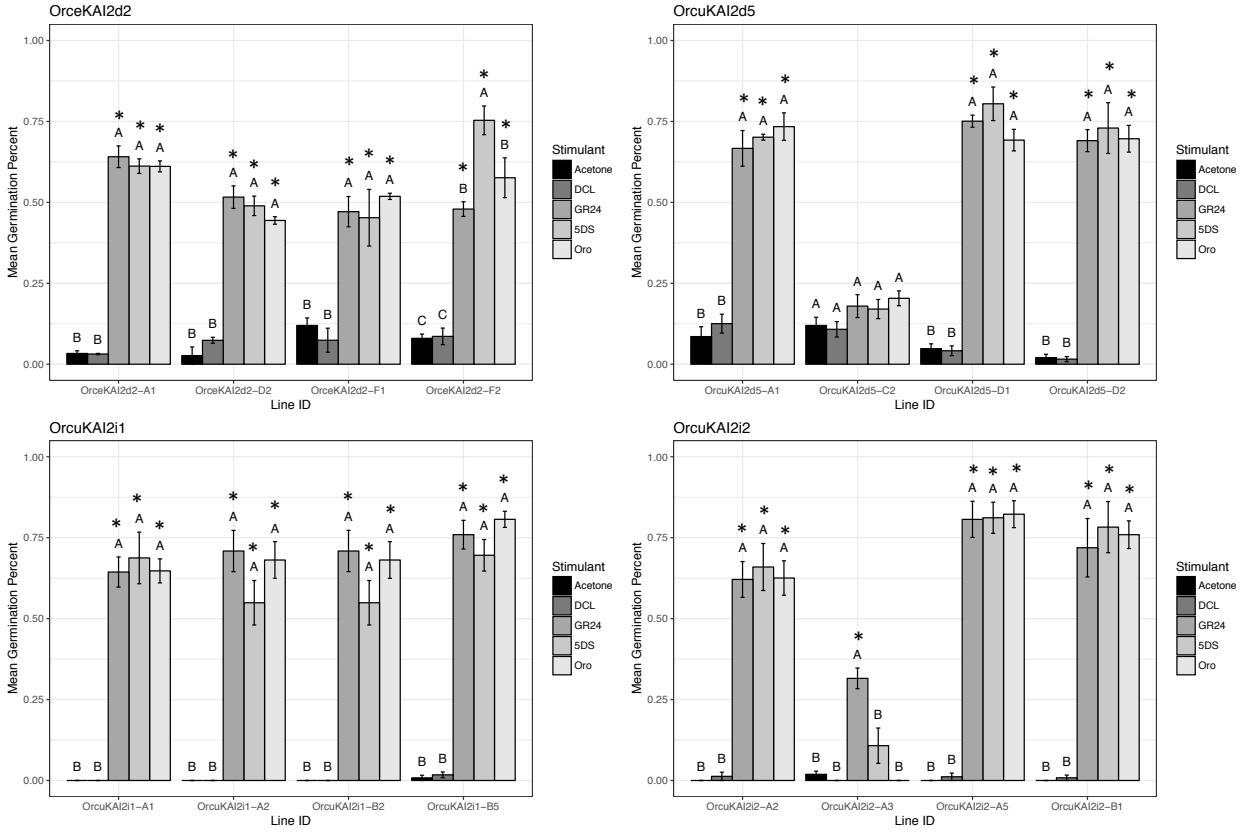


Figure 2.5 Germination response of Arabidopsis lines transformed with *KAI2d* genes from *O. cernua and O. cumana*. A) *OrceKAI2d2*. B) *OrcuKAI2d5*. C) *OrcuKAI2i1*. D) *OrcuKAI2i2*. Each

line represents a unique transformation event as was exposed to acetone (control), DCL, GR24, 5DS or Oro. Tukey-Kramer HSD test was used to determine significance, * $P < 0.05$. SE bars shown.

Table 2.2 Summary of response of each transgene to DCL and SL germination stimulants. Percent of germinating lines shown, based on significant differences between the specific stimulant and acetone control as determined by the Tukey-Kramer HSD test (See Supplemental Figure 2.4). For each *KAI2d* and *KAI2i* gene, 4-10 lines were evaluated.

Gene	DCL	GR24	Orobanchol	5-Deoxystrigol
	Percentage of transgenic lines responding			
OrceKAI2c	0	0	0	0
OrceKAI2i1	0	0	0	0
OrceKAI2i2	0	0	0	0
OrceKAI2d1	0	0	0	0
OrceKAI2d2	0	100	100	100
OrceKAI2d3	0	0	0	0
OrceKAI2d4	0	0	0	0
OrcuKAI2c	0	0	0	0
OrcuKAI2i1	0	100	100	100
OrcuKAI2i2	0	100	75	75
OrcuKAI2d1	0	0	0	0
OrcuKAI2d2	0	0	0	0
OrcuKAI2d3	10	20	10	20
OrcuKAI2d4	0	0	0	0
OrcuKAI2d5	0	75	75	75
OrcuKAI2d6	0	0	0	0

2.4 Discussion

Perception of SLs has been the subject of intense study in recent years (Reviewed in De Cuyper et al., 2017; Lumba et al., 2017; Waters et al., 2017). The main proteins involved in binding SLs are *D14* and its close homolog *KAI2/HTL,* and interacting proteins *MAX2* and the *SMXL* gene family. The interaction of these with a SL are thought to regulate specificity of SL perception and mediate downstream development or germination. Therefore, these receptors were the primary targets for investigation into the germination specificity of *O. cernua* and *O. cumana*.

MAX2 and *D14* were identified within each species and both were ruled out as putative stimulant receptors or involved in conferring stimulant specificity in *O. cernua* and *O. cumana.* Next, the *KAI2* genes that were reported by Conn et al. (2015) as being evolved to respond to SLs were found within the *O. cernua* and *O. cumana* transcriptomes. These comprise four genes from *O. cernua* and six genes from *O. cumana* (designated *OrceKAI2d1-4 and OrcuKAI2d1-6*, respectively). Examining the relationships amongst the orthologous pairs from each species, indicated two additional *KAI2d* genes within *O. cumana,* which could explain *O. cumana*'s ability to respond to a non-SL hormone, DCL (Figure 2.1). This hypothesis was not supported by subsequent data, which raises the question of whether these genes function in *O. cumana*.

We also found two *KAI2i* genes from each species. Conn et al (2015) did not identify *KAI2i* genes from the sequences available at the time, and we propose here that these fit the "intermediate" category in that they substantially differ in sequence from the other *KAI2d* genes (Figure 2.1) and show distinct substrate specificity in the complementation assay (Table 2.2).

Fifteen of these *KAI2* genes are predicted to encode functional proteins as they contain a full-length coding sequence, including the correct Ser95-His246-Asp217 catalytic triad needed for substrate hydrolysis (Bythell-Douglas et al., 2017; Yao et al., 2016; Zhao et al., 2015). The only exception is *OrceKAI2i2,* which is missing over two hundred and fifty bases from the second exon and is not functional in the complementation assay.

To test whether the genomic presence of *KAI2d* genes control the germination specificity in *O. cernua* and *O. cumana* species, we genotyped *O. cumana* x *O. cernua* F_2 hybrid individuals that segregate for stimulant specificity. The objective was to identify which *KAI2d* genes were present in each F_2 hybrid individual and relate that to germination response to a given stimulant in order to discern which *KAI2d* genes are responsible for a given stimulant response. The pattern of segregation indicated that the *KAI2d* genes are linked, but pointed to *OrceKAI2d2* as a

likely Oro receptor. Response to DCL was associated with inheritance of all *O. cumana KAI2d* genes together (Figure 2.2, Figure 2.3).

We observed that the F_3 hybrid lines did not germinate at uniform levels, and with our threshold set at 5% germination to be considered responsive, there was variation in the strength of germination response to a given stimulant. For example, seeds of one line may germinate at 40% in response to Oro, but at 90% in response to GR24. To capture this complexity, a statistical model was fit to the hybrids that examined the effect of each *KAI2d* gene, as well as the variance due to random effects, on germination rates of the F_2 hybrids in response to each stimulant. The model strongly indicated that *OrceKAI2d2* and (to a lesser extent) *OrceKAI2d1* presence tends to co-occur with increased germination rates in response to Oro (Figure 2.4a, Supplemental Table 2.2). The analysis also indicated that *OrcuKAI2d6* may be associated with DCL perception. No other genes were found to be correlated with increased or decreased germination rates (Supplemental Figure 2.3), but in all cases, there was significant plant-to-plant as well as Petri dish-to-Petri dish variation. The "plant effect" suggests that there may be additional genetic influences on germination rates for DCL and Oro, such as additional genes in the stimulant signaling pathway. The model demonstrated that composition of *KAI2d* genes is more accurate in predicting germination rates for Oro perception than they for predicting DCL perception (Supplemental Table 2.3).

To test the functional ability of the *KAI2d* genes from *O. cernua* and *O. cumana* to respond to DCL and SL's, we used a cross-species complementation assay where we inserted the *KAI2d* genes into a *kai2 Arabidopsis* mutant background. While *Arabidopsis* does not germinate in the presence of SLs, *Arabidopsis kai2-2* mutant (Landsberg ecotype) shows a slight delay in germination compared to wild-type Landsberg (Conn et al., 2015). We identified only one *O. cernua* gene, *OrceKAI2d2,* able to recover the mutant phenotype when treated with SLs in the model plant system. The homolog in *O. cumana, OrcuKAI2d5,* was also able to recover the mutant phenotype in response to all strigolactones, strongly indicating that these genes have the ability to bind and transduce a SL signal. Additionally, the *O. cumana* genes *OrcuKAI2i1* and *OrcuKAI2i2* complemented the mutant phenotype in response to SLs (Figure 2.5), while the *O. cernua* homologs *OrceKAI2i1* and *OrceKAI2i2* showed no response. The lack of response to SLs of *OrceKAI2di2* was expected due to its truncated nature, but it was surprising that *OrceKAI2i1*

did not respond. No genes from either species was able to confer a response to DCL in the model system.

Multiple groups have demonstrated that parasitic plants have an expanded *KAI2* gene family compared to their non-parasitic relatives (Conn et al., 2015; Toh et al., 2015; Tsuchiya et al., 2015). The consensus is that this expansion has evolved to allow the parasites to detect unique compounds exuded from host roots. Within *O. cernua* and *O. cumana,* every *KAI2* gene is expressed, and at variable levels, pre- and post- germination stimulant exposure. Yet, out of a total of 16 *KAI2* genes between the two species, only four demonstrated SL perception when expressed in an *Arabidopsis kai2* mutant background. Tsuchiya et al., (2015) demonstrated that ten out of twelve identified *Striga KAI2* genes containing the catalytic triad could hydrolyze the SL agonist YLG and GR24. Toh et al., (2015) went on to further demonstrate that six *KAI2* genes out of the twelve *KAI2* in *Striga* could recover germination in *Arabidopsis kai2* mutants when exposed to SL stimulants, and concluded that the additional *ShHTL* genes must perform alternative functions in *Striga*.

The heterologous system requires that introduced *Orobanche KAI2* proteins interact properly with *Arabidopsis* components of the signaling complex. Thus, we examined each *KAI2d* and *KAI2i* gene in silico at the amino acid level for the correct domains required for interacting with *AtMAX2* and excepting *OrceKAI2i2,* found to all contain the correct domains (Bythell-Douglas et al., 2017; Zhao et al., 2015). In *Arabidopsis, AtKAI2* directly interacts with the *SMAX1* and *SMXL2* repressor proteins, however the sites at which the proteins physically interact is currently unclear. Yao et al. (2017) demonstrated that the *Striga KAI2d/HTL* receptor *ShHTL7*, could form a complex with *AtSMAX1* in vitro (Yao et al., 2017). Consequently, we propose that the *OrceKAI2* and *OrcuKAI2* genes are capable of functioning within *Arabidopsis*.

This work fits with current model of SL signaling perception in parasitic plants with respect to *O. cernua*. All data point to *OrceKAI2d2* as a primary receptor for Oro, with the other three *KAI2d* genes functioning in some other way (see Toh 2015). However, this work also raises two fundamental questions about how germination signaling is regulated. One is the failure to explain how *O. cumana* does not germinate in response to SLs despite having six *KAI2d* genes, of which at least one responds to Oro, and two *KAI2i* genes that both respond to Oro in a heterologous assay. Has *O. cumana* deactivated its SL perception pathway in regards to germination? Based on our data, this seems to be the case.

The other question is how *O. cumana* can respond to DCL as a germination stimulant when none of its *KAI2d* genes directly bind DCL in our assay. While there are two *KAI2d* genes unique to *O. cumana*, *OrcuKAI2d1* and *OrcuKAI2d2,* neither showed a response to DCL in our cross-species complementation assay. Failure to identify the DCL receptor could be explained by the presence of an additional DCL receptor that is not part of the *KAI2d* family. But the hybrid analysis suggests that *KAI2d* genes (or other linked genetic elements) segregate with DCL responsiveness. It is also possible that the complementation assay has limitations, for instance the *Arabidopsis* signaling machinery may not respond to DCL or the parasite transgenes in the same way as the *O. cumana* signaling machinery. In any case, we conclude that additional genes beyond the *KAI2* gene family are involved in regulating stimulant perception in parasite seeds.

2.5 References

Adey, A., Morrison, H. G., Asan, Xun, X., Kitzman, J. O., Turner, E. H., … Shendure, J. (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology*, *11*(12), R119. https://doi.org/10.1186/gb-2010-11-12-r119

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST : a new generation of protein database search programs, *25*(17), 3389–3402.

Andrews, S. (2014). FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bushnell, B. (2016). BBMap short read aligner. University of California, Berkeley, California.

Bythell-Douglas, R., Rothfels, C. J., Stevenson, D. W. D., Graham, S. W., Wong, G. K.-S., Nelson, D. C., & Bennett, T. (2017). Evolution of strigolactone receptors by gradual neo-functionalization of *KAI2* paralogues. *BMC Biology*, *15*(1), 52. https://doi.org/10.1186/s12915-017-0397-z

Clough, S. J., & Bent, A. F. (1999). Floral dip : a simplified method for Agrobacterium-mediated transformation of *Arabidopsis thaliana*, *16*(October 1998), 735–743.

Conn, C. E., Bythell-douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J. H., … Nelson, D. C. (2015). Parasitic Plants. *Science (New York, N.Y.)*, *2*(6247), 7–11.

De Cuyper, C., Struk, S., Braem, L., Gevaert, K., De Jaeger, G., & Goormachtig, S. (2017). Strigolactones, karrikins and beyond. *Plant Cell and Environment*, *40*(9), 1691–1703. https://doi.org/10.1111/pce.12996

Fong, Y., Rue, H., & Wakefield, J. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics*, *11*(3), 397–412. https://doi.org/10.1093/biostatistics/kxp053

Gelman, A. (2011). Induction and deduction in Bayesian data analysis. *Rationality, Markets and Morals*, *2*, 67–78. Retrieved from http://www.rmm-journal.de/%5Cnhttp://www.rmm-journal.com/htdocs/volume2.html

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., … Pochet, N. (2014). Reference generation and analysis with Trinity. *Nat Protocols* (Vol. 8). https://doi.org/10.1038/nprot.2013.084.De

Harrison, S. J., Mott, E. K., Parsley, K., Aspinall, S., Gray, J. C., & Cottage, A. (2006). A rapid and robust method of identifying transformed *Arabidopsis thaliana* seedlings following floral dip transformation, *7*, 1–7. https://doi.org/10.1186/1746-4811-2-19

Hegenauer, V., Körner, M., & Albert, M. (2017). Plants under stress by parasitic plants. *Current Opinion in Plant Biology*, *38*, 34–41. https://doi.org/10.1016/j.pbi.2017.04.006

Hoff, P. D. (2009). A first course in Bayesian statistical methods. Springer Science & Business Media.

Joel, D. M., Bar, H., Mayer, A. M., Plakhine, D., Ziadne, H., Westwood, J. H., & Welbaum, G. E. (2012). Seed ultrastructure and water absorption pathway of the root-parasitic plant *Phelipanche aegyptiaca* (Orobanchaceae). *Annals of Botany*, *109*(1), 181–195. https://doi.org/10.1093/aob/mcr261

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2, *9*(4), 357–

360. https://doi.org/10.1038/nmeth.1923
Lumba, S., Holbrook-Smith, D., & McCourt, P. (2017). The perception of strigolactones in vascular plants. *Nature Chemical Biology*, *13*(6), 599–606. https://doi.org/10.1038/nchembio.2340
Martin, M. (2011). Sequencing Reads. *EMBnet.journal*, *17*(1), 10–12.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110
Parker, C. (2012). Parasitic Weeds: A World Challenge. *Weed Science*, *60*(2), 269–276. https://doi.org/10.1614/WS-D-11-00068.1
Parker, C. (2013). The Parasitic Weeds of the Orobanchaceae BT - Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.) (pp. 313–344). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38146-1_18
Parker, C., & Riches, C. R. (1993). Parasitic Weeds of the World: Biology and Control. Wallingform: CAB International.
Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, *23*(9), 1061–1067. https://doi.org/10.1093/bioinformatics/btm071
Plakhine, D., Tadmor, Y., Ziadne, H., & Joel, D. M. (2012). Maternal tissue is involved in stimulant reception by seeds of the parasitic plant *Orobanche*. *Annals of Botany*, *109*(5), 979–986. https://doi.org/10.1093/aob/mcs016
Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, 20–22. https://doi.org/10.1.1.13.3406
Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative Genome Viewer. *Nature Biotechnology*, *29*(1), 24–6. https://doi.org/10.1038/nbt.1754.Integrative
Team, R. D. C. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org/.
Toh, S., Holbrook-smith, D., Stogios, P. J., Onopriyenko, O., Lumba, S., Tsuchiya, Y., … Mccourt, P. (2015). Structure-function analysis identifies highly sensitive strigolancone receptors in *Striga*. *Plant Science*, *350*(6257), 203–208.
Tsuchiya, Y., Yoshimura, M., Sato, Y., Kuwata, K., Toh, S., Holbrook-Smith, D., … Hagihara, S. (2015). Probing strigolactone receptors in *Striga hermonthica* with fluorescence. *Science*, *349*(6250), 864–868. https://doi.org/10.1126/science.aab3831
Waters, M. T., Gutjahr, C., Bennett, T., & Nelson, D. C. (2017). Strigolactone Signaling and Evolution. *Annual Review of Plant Biology*, *68*(1), 291–322. https://doi.org/10.1146/annurev-arplant-042916-040925
Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends in Plant Science*, *15*(4), 227–235. https://doi.org/10.1016/j.tplants.2010.01.004
Yao, R., Ming, Z., Yan, L., Li, S., Wang, F., Ma, S., … Xie, D. (2016). DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature*, *536*(7617), 469–473. https://doi.org/10.1038/nature19073

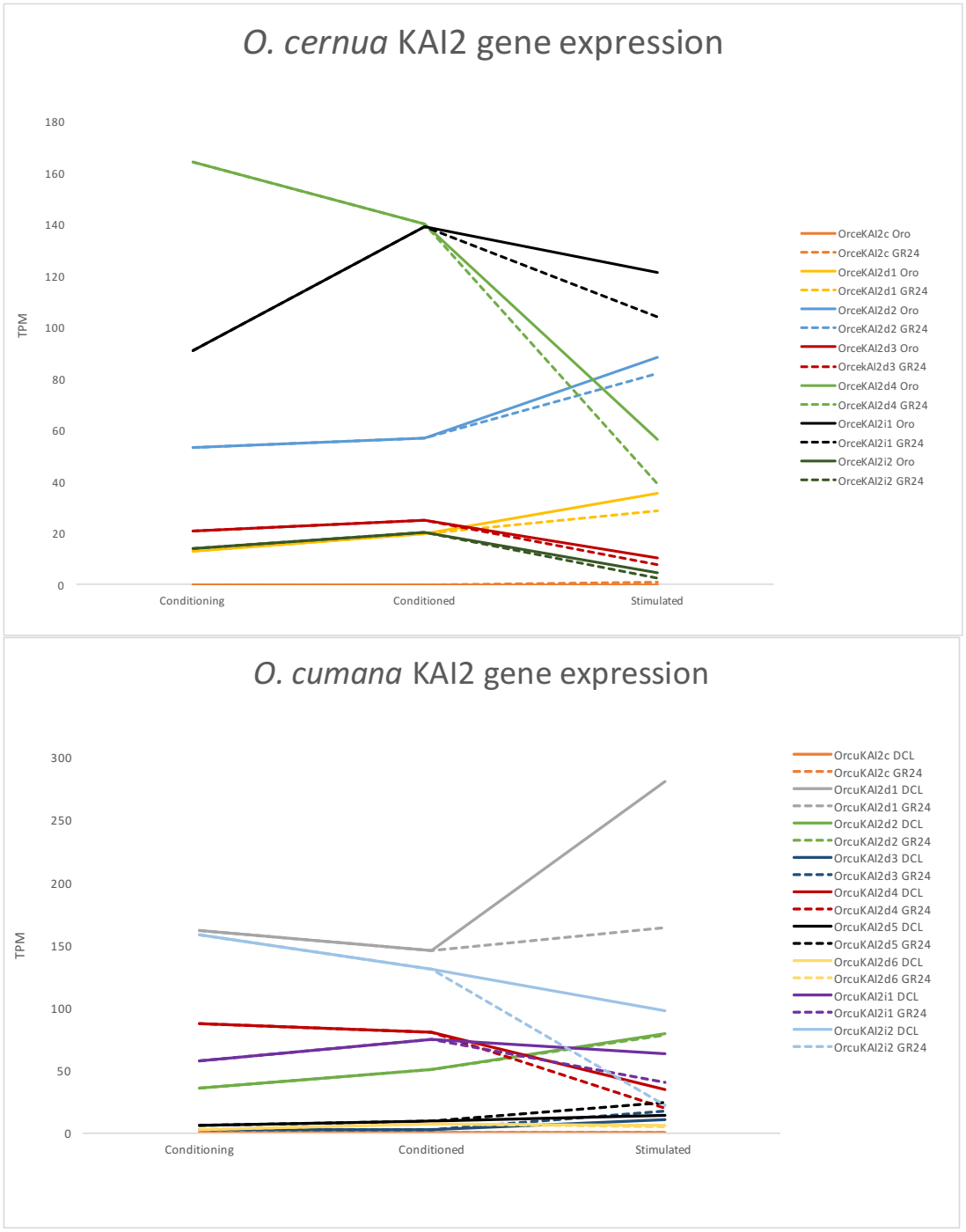
Yao, R., Wang, F., Ming, Z., Du, X., Chen, L., Wang, Y., … Xie, D. (2017). *ShHTL7* is a non-canonical receptor for strigolactones in root parasitic weeds. *Cell Research*, *27*(6), 838–841. https://doi.org/10.1038/cr.2017.3

Zhao, L. H., Zhou, X. E., Yi, W., Wu, Z., Liu, Y., Kang, Y., … Xu, H. E. (2015). Destabilization of strigolactone receptor DWARF14 by binding of ligand and E3-ligase signaling effector DWARF3. *Cell Research*, *25*(11), 1219–1236. https://doi.org/10.1038/cr.2015.122

2.6 Supplemental Materials

```
OrceMAX2        ----------------------MAVATTTLTDLPDVIVSNIIAAVCDVRSRNSAALVCRKW
OrcuMAX2        VKLPLSSRLTKIGESDLDLLPMAVATTTLTDLPDVIVSNIIAAVCDVRSRNSAALVCRKW
                                      ************************************

OrceMAX2        YVLERATRSSLCLRGNLRDLFMLPTCFQSVSHLDLSLLSPYGHPLTSASDPDPALIAHLL
OrcuMAX2        YVLERATRSSLCLRGNLRDLFMLPTCFQSVSHLDLSLLSPYGHPLTSASDPDPALIAHLL
                ************************************************************

OrceMAX2        RHALPSVTSLTLYARNPSTIQLIAPQWPNLEHLKLVRWHQRPQTDDAGDELKILISECGQ
OrcuMAX2        RHALPSVTSLTLYARNPSTIQLIAPQWPNLEHLKLVRWHQRPQTDDAGDELKILISECGQ
                ************************************************************

OrceMAX2        LKSLDLSAFYCWTDDVPLALEFCPTFASILTCLNLLNSSFSEGFKSDEVKVITKACPNLR
OrcuMAX2        LKSLDLSAFYCWTDDVPLALEFCPTFASILTCLNLLNSSFSEGFKSDEVKVITKACPNLR
                ************************************************************

OrceMAX2        EFRAACMFDPRYIGCVGDEALVSVSVNCPKLAILHLADTSALSSARGDFDMEHQVLTQED
OrcuMAX2        EFRAACMFDPRYIGCVGDEALVSVSVNCPKLAILHLADTSALSSARGDFDMEHQVLTQED
                ************************************************************

OrceMAX2        ARINAATLIEVFSGLPRLEELAIDVSVNVRDSGPALEVLKSKCPGLRSLKLGQFHGISLP
OrcuMAX2        ARINAATLIEVFSGLPRLEELAIDVSVNVRDSGPALEVLKSKCPGLRSLKLGQFHGISSP
                ************************************************************ *

OrceMAX2        VGSKLDGVALCHGLKSLSIRNVSDLSDMGLIAIGRGCCRLAKFEVHGCRKLTVRGLRTMA
OrcuMAX2        VGSKLDGVALCHGLKSLSIRNVSDLSDMGLIAIGRGCCRLAKFEVHGCRKLTVRGLRTMA
                ************************************************************

OrceMAX2        SLLHRTLVDVRISCCKSLGAVQSLQALEPLQDRIERLHIDCIWDCTTDELDETNDDDCFD
OrcuMAX2        SLLHRTLVDVRISCCKSLGAVQSLQALEPLQDRIERLHIDCIWDCTTDELDETNDDDCFD
                ************************************************************

OrceMAX2        LKSSDQGGVLNSYQPDEHTAQEWTGTDYDYDYDGMTHAIKKRKCSHDQNPSYFGMVVNSN
OrcuMAX2        LKSSDQGGVLNSYQPDEHTAQEWTGTDYDYDYDGMTHAIKKRKCSHDQNPSYFGMVVNSN
                ************************************************************

OrceMAX2        GSENVNAYGERVWDRLQCLSLSVPVGQLLNPLVSAGLENCPNLEEIRIKIEGDCRVLPKP
OrcuMAX2        GSENVNAYGERVWDRLQCLSLWVPVGQLLNPLVSAGLENCPNLEEIRIKIEGDCRVLPKP
                ******************** ************************************

OrceMAX2        TVREFGLSTLVIYPSLSKMHLDCGDIIGYTHTAPSGQMDLSLWERFCLIGIGNLSLTELD
OrcuMAX2        TVREFGLSTLVIYPSLSKMHLDCGDIIGYTHTAPSGQMDLSLWERFCLIGIGNLSLTELD
                ************************************************************

OrceMAX2        YWPPQDRDVNQRTLSLPAAGLLQQCFGLRKLFIHGTAHEHFMMFLLRIPDLRDVQLREDY
OrcuMAX2        YWPPQDRDVNQRTLSLPAAGLLQQCFGLRKLFIHGTAHEHFMMFLLRIPDLRDVQLREDY
                ************************************************************

OrceMAX2        YPAPENDMSTEMRADSCSRFEVALNGRQISD
OrcuMAX2        YPAPENDMSTEMRADSCSRFEVALNGRQISD
                ******************************
```

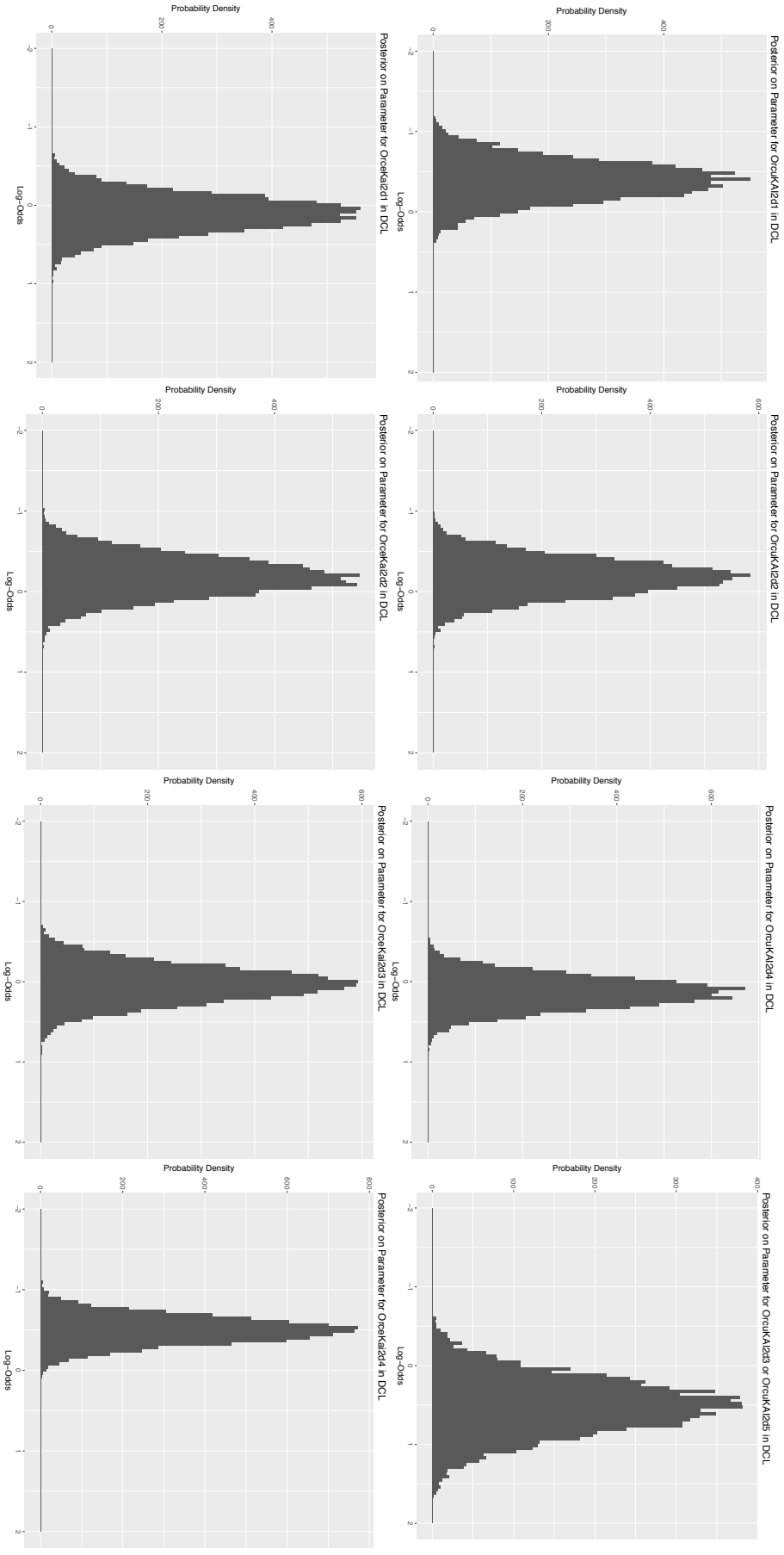
Supplemental Figure 1: CLUSTAL format alignment by MAFFT showing *OrceMAX2* and *OrcuMAX2* amino acid alignments. * indicates matching amino acid.

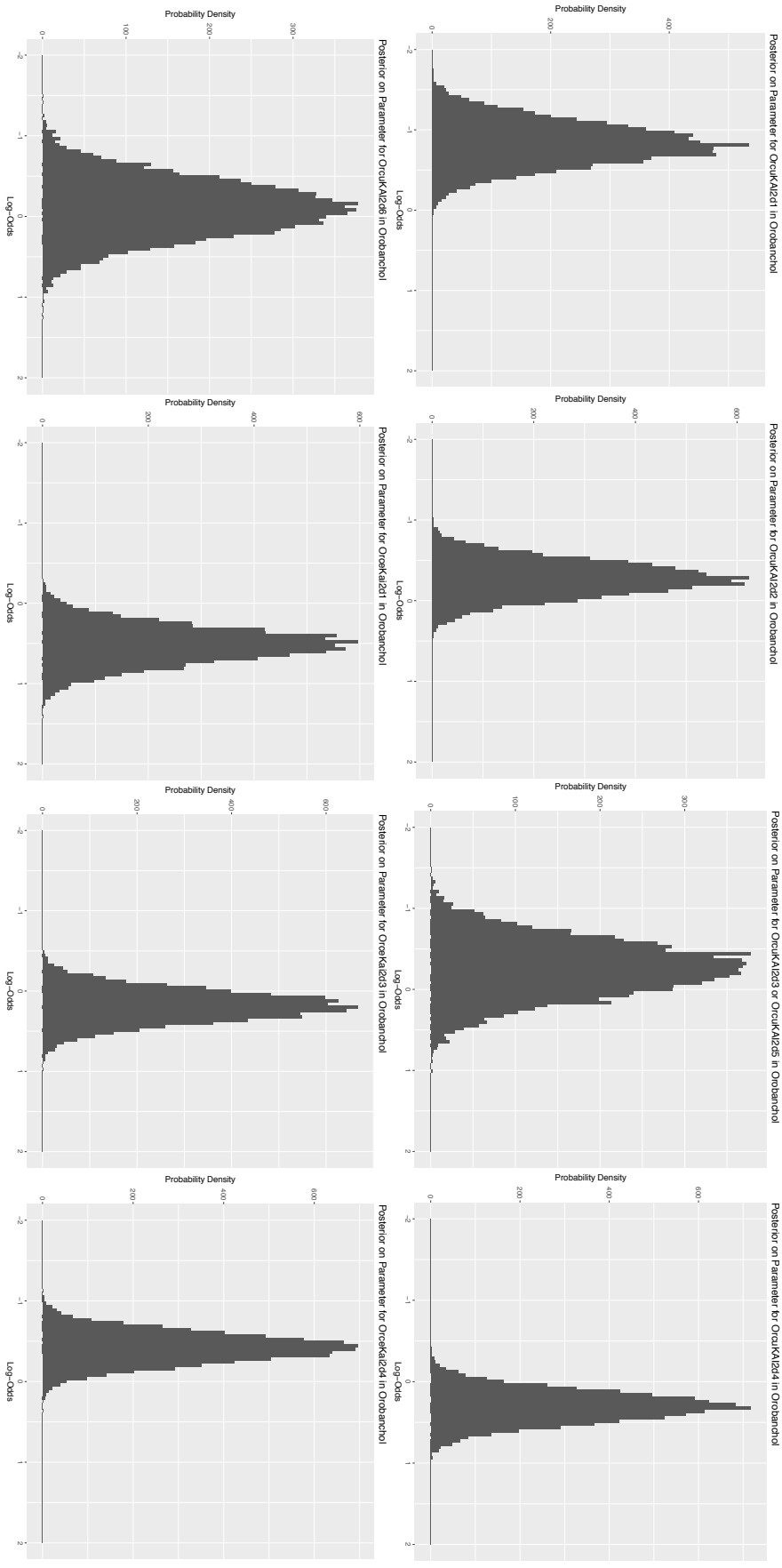


Supplemental Figure 2.1 *KAI2d* gene expression from each transcriptome stage. For each species, solid line indicates expression in response to species-specific stimulant (Oro or DCL) and dotted line indicates expression in response to GR24.

Supplemental Table 2.1 Universal Primer sequences used to amplify *OrcuKAI2d1-6* and *OrceKAI2d1-4* for *KAI2d* genotyping of 94 *O. cernua/O. cumana* segregating F_2 hybrids.

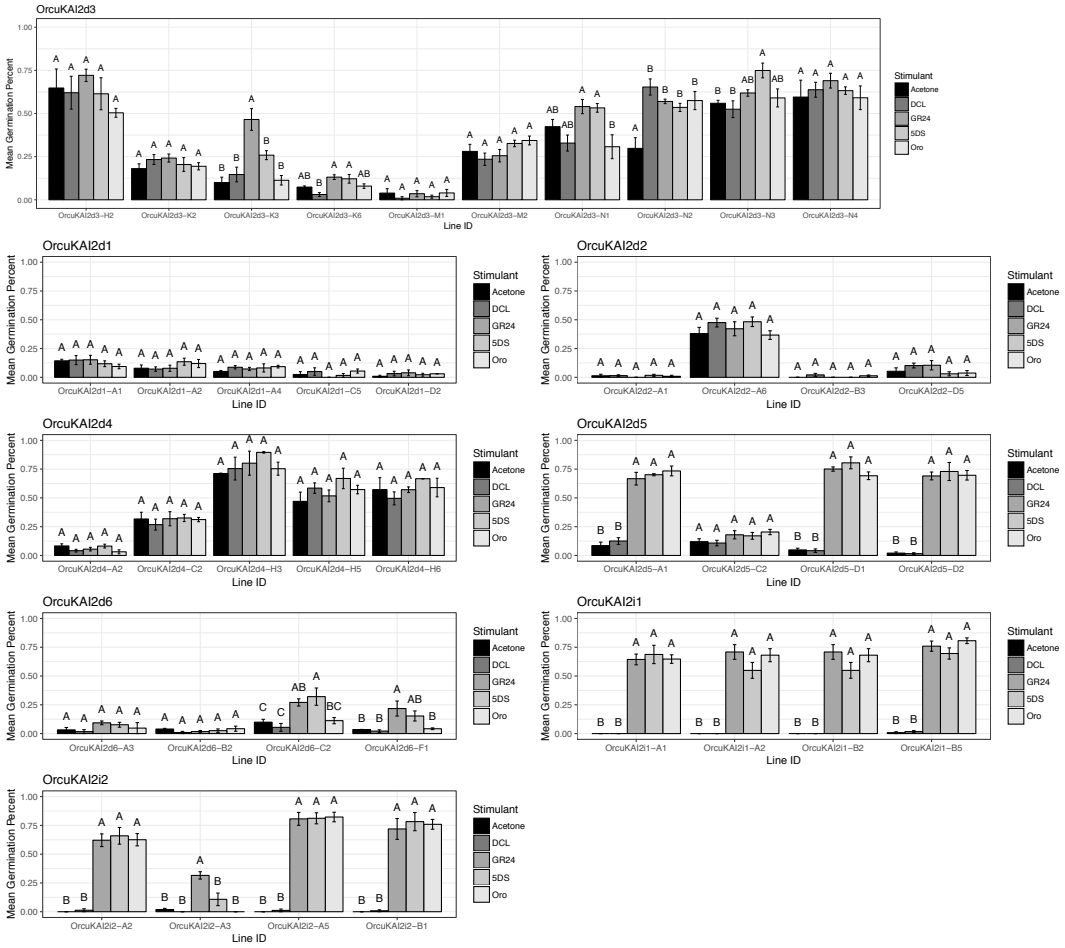
Primer ID	Sequence	Genes Amplified - *O. cumana*	Genes Amplified - *O. cernua*
R1	tcatgcatcaacaatatc	*OrcuKAI2d5, OrcuKAI2d3, OrcuKAI2d2*	
R2	tcaaccatcaacaatatc	*OrcuKAI2d4*	*OrceKAI2d4*
R3	tcaggcagcgatattata	*OrcuKAI2c*	*OrceKAI2c*
R4	tcatgcatcaatatcgtg	*OrcuKAI2d6*	*OrceKAI2d3*
F1	atgggaatcacccaag	*OrcuKAI2c*	*OrceKAI2c*
F2	atgaaccgtatagttggact	*OrcuKAI2d5*	
F3	atgagtagcatagttggtg	*OrcuKAI2d1, OrcuKAI2d3, OrcuKAI2d4, OrcuKAI2d6*	*OrceKAI2d1, OrceKAI2d4, OrceKAI2d3*
F2B	atgaacagcatagttggact		*OrceKAI2d2*
F3B	atgggtagcattgttg	*OrcuKAI2d2*	
R1B	tcatacatcagcaatatc		*OrceKAI2d2*


Posterior on Parameter for OrcuKAI2d1 in DCL
Posterior on Parameter for OrceKai2d1 in DCL
Posterior on Parameter for OrcuKAI2d2 in DCL
Posterior on Parameter for OrceKai2d2 in DCL
Posterior on Parameter for OrcuKAI2d3 in DCL
Posterior on Parameter for OrceKai2d3 in DCL
Posterior on Parameter for OrcuKAI2d4 in DCL
Posterior on Parameter for OrceKai2d4 in DCL
Posterior on Parameter for OrcuKAI2d5 or OrcuKAI2d5 in DCL
Probability Density
Log−Odds

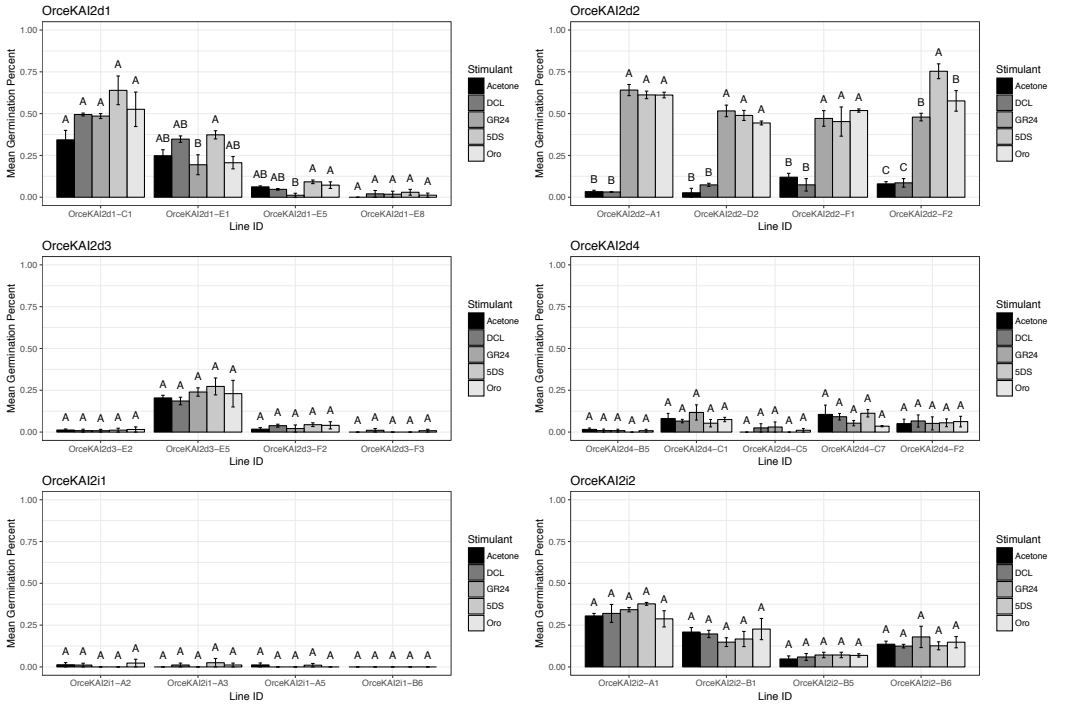

Posterior on Parameter for OrcuKAl2d1 in Orobanchol
Posterior on Parameter for OrcuKAl2d6 in Orobanchol
Posterior on Parameter for OrcuKAl2d2 in Orobanchol
Posterior on Parameter for OrceKal2d1 in Orobanchol
Posterior on Parameter for OrcuKAl2d3 or OrcuKAl2d5 in Orobanchol
Posterior on Parameter for OrceKal2d3 in Orobanchol
Posterior on Parameter for OrcuKAl2d4 in Orobanchol
Posterior on Parameter for OrceKal2d4 in Orobanchol
Probability Density
Log−Odds

Supplemental Figure 2.2 Posterior distribution charts for A) *KAI2d* genes in response to DCL. B) *KAI2d* genes in response to Oro. Higher values of density indicate that it is more probable that the truth lies in that region. Sign of log-odds indicates direction of correlation, and magnitude indicates strength of correlation. Values far removed from 0 indicates correlation of gene with germination response to a given phenotype.

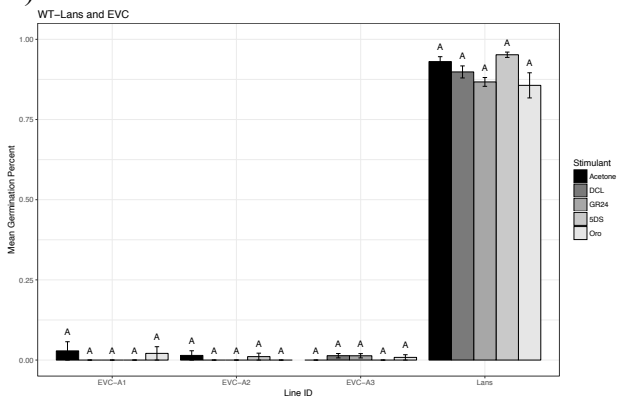
A)


OrcuKAl2d3
OrcuKAl2d1
OrcuKAl2d2
OrcuKAl2d4
OrcuKAl2d5
OrcuKAl2d6
OrcuKAl2i1
OrcuKAl2i2
Mean Germination Percent
Line ID
Stimulant
Acetone
DCL
GR24
5DS
Oro

B)


OrceKAl2d1
OrceKAl2d2
OrceKAl2d3
OrceKAl2d4
OrceKAl2i1
OrceKAl2i2
Mean Germination Percent
Line ID
Stimulant
Acetone
DCL
GR24
5DS
Oro

C)



Supplemental Figure 2.3 Complementation assay results for A*) Orobanche cernua* and B) *Orobanche cumana KAI2d* and *KAI2i* genes. C) Empty vector control (EVC) and wild-type Arabidopsis thaliana ecotype Landsberg (WT-Lans). Graphs show the germination response of each *KAI2d* and *KAI2i* gene expressed in the *Arabidopsis kai2* background after exposure to an acetone negative control or specific stimulants, DCL, GR24, 5-deoxy strigol or orobanchol. Each bar represents the mean of 3 replications and vertical lines represent SE. Means were separated using the Tukey-Kramer HSD test. Bars that do not share the same letter are different at $P < 0.05$.

Supplemental Table 2.2 Credible intervals for genes showing a significant correlation with stimulant perception in the hybrid lines. A) DCL perception; B) Oro perception.

A)

DCL Response	post_mean	2.50%	97.50%
Intercept	-1.331	-1.767	-0.898
OrceKAI2d1	0.103	-0.345	0.560
OrceKAI2d2	-0.185	-0.657	0.289
OrceKAI2d3	0.032	-0.413	0.474
OrceKAI2d4	-0.496	-0.826	-0.162
OrcuKAI2d1	-0.390	-0.870	0.090
OrcuKAI2d2	-0.177	-0.620	0.263
OrcuKAI2d3 or *OrcuKAI2d5*	0.510	-0.180	1.222
OrcuKAI2d4	0.135	-0.244	0.518
OrcuKAI2d6	0.832	0.152	1.520
Petri Dish Effect	0.613	0.527	0.709
Plant Effect	0.873	0.702	1.077

B)

Oro Response	post_mean	2.50%	97.50%
Intercept	-1.008	-1.417	-0.603
OrceKAI2d1	0.518	0.065	0.982
OrceKAI2d2	1.166	0.665	1.669
OrceKAI2d3	0.189	-0.199	0.584
OrceKAI2d4	-0.416	-0.776	-0.049
OrcuKAI2d1	-0.828	-1.354	-0.308
OrcuKAI2d2	-0.270	-0.689	0.159
OrcuKAI2d3 or *OrcuKAI2d5*	-0.260	-0.927	0.427
OrcuKAI2d4	0.309	-0.071	0.690
OrcuKAI2d6	-0.100	-0.768	0.579
Petri Dish Effect	0.591	0.502	0.691
Plant Effect	0.717	0.549	0.919

Supplemental Table 2.3 Five-fold Cross Validation, was used to evaluate model predictive abilities by comparison to the baseline model of simply guessing based on the median germination rate for a particular chemical. We find that Mean Absolute Error, or MAE (the average absolute value of the difference between the predicted and observed germination rates for each Petri dish) decreases from 0.2704 under the null model to 0.1793 when taking genotypes into account through the GLMM for Oro, while only a drop from 0.2165 under the null model to 0.1941 under the GLMM with DCL.

	DCL	Oro
Null Model	0.217	0.270
GLMM	0.194	0.179

Supplemental Table 2.4 Primer sequences used for amplifying *KAI2d* sequences for pENTR/D-TOPO (ThermoFisher) cloning.

gene	F primer	R primer
OrceKAI2c	CACCATGGGAATCACCCAAGACGCT	TCAGGCAGCGATATTATAAC
OrceKAI2d1	CACCATGAGTAGCATAGTTGGTGCC	TCATGCATCAACAATACCGA
OrceKAI2d2	CACCATGAACAGCATAGTTGGACTT	TCATACATCAGCAATATCGC
OrceKAI2d3	CACCATGAGTAGCATAGTTGGTGCC	TCATGCATCAATATCGTGAT
OrceKAI2d4	CACCATGAGTAGCATAGTTGGTGCG	TCAACCATCAACAATATCGT
OrcuKAI2c	CACCATGGGAATCACCCAAGAAGCT	TCAGGCAGCGATATTATAAC
OrcuKAI2d1	CACCATGAGTAGCATAGTTGGTGCC	ATCGTGCCCCCCGGCATACT
OrcuKAI2d2	CACCATGGGTAGCATTGTTGGTGCG	TCATGCATCAACAATATCAT
OrcuKAI2d3	CACCATGAGTAGCATAGTTGGTGCC	TCATGCATCAACAATATCGA
OrcuKAI2d4	CACCATGAGTAGCATAGTTGGTGCG	TCAACCATCAACAATATCGT
OrcuKAI2d5	CACCATGAACCGTATAGTTGGACTT	TCATGCATCAACAATATCGC
OrcuKAI2d6	CACCATGAGTAGCATAGTTGGTGCC	TCATGCATCAATATCGTGAT
OrceKAI2i1	CACCATGAACATAGTTGGAGCA	TCAGGCGTCAATGATGTC
OrceKAI2i2	CACCATGAGCACAGTTGGAGC	TCAGGCTATATCGTGTTGTAT
OrcuKAI2i1	CACCATGAACATAGTTGGAGCA	TCAGGCGTCAATGATGTC
OrcuKAI2i2	CACCATGAGCACAGTTGGAGC	TCAGGCTATATCGTGTTGTAT

Chapter 3

Transcriptomic insights into *Orobanche cernua* and *Orobanche cumana*

[1]Larose, H., [2]D. Plakhine, [2]H. Eizenberg, [2]Y. Tadmor and [1]J. Westwood.

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

[2]Newe Ya'ar Research Center, A.R.O., Israel

Author contributions: Parental RNA was isolated by DP. HE, YT and JW advised on the study design. HL did all other laboratory work, bioinformatics analyses, and wrote the manuscript with input from JW.

Abstract

Parasitic plants of the genus *Orobanche* have tightly controlled seed germination that links germination to the perception of a host-derived germination stimulant. The process requires three steps: a period of dry storage known as after-ripening, a period of imbibition in water that is termed conditioning, and the perception of a host-derived germination signal. The conditioning period must be completed before seeds will respond to the germination signal, but little is known about the conditioning process and how it prepares seeds for germination. To better understand conditioning and germination, the transcriptomes of *O. cernua* and *O. cumana* were sequenced at three stages: during conditioning, at completion of conditioning, and after treatment with germination stimulants. The final stage was further divided into stimulants that are species-specific (DCL or Oro) or universal (GR24). *De novo* assembled transcriptomes of *O. cernua* and *O. cumana* yielded over 200,000 contigs per species, resulting in 103,570 predicted ESTs for *O. cernua* and 110,019 ESTs for *O. cumana*. Gene expression analysis revealed that for each species, over 17% percent of the ESTs were unique to conditioning. Overall GO profiles and enriched GO terms were remarkably similar for both species during each stage. Approximately 25,000 orthologous pairs were identified between species, suggesting both a high level of similarity between the species and the evolution of many species-specific gene variants. Overall, this study generated two quality transcriptomes representing stages pre- and post-germination exposure from two species of *Orobanche* which differ in germination stimulant preferences.

3.1 Introduction

Root parasitic plants cause major economic loss in many areas of the Mediterranean, Eastern European, Asian and African regions, and their ranges are constantly expanding due to global travel and goods exchange (Chris Parker, 2012, 2013). *Striga spp.* alone cause an estimated crop yield loss totaling up to US $200 million in Africa, and negatively impact the food supply of upwards of a hundred million people (Rodenburg et al., 2016, Yoder & Scholes, 2010).

The severity of crop destruction can be attributed to many factors. First, is the location of the parasite. Root parasitic plants of the *Orobanche* genera spend most of their lifecycles underground, emerging solely for reproduction, and consequently evade visual detection until most of the damage has been done to the crop. Underground, they survive by forming physical connections with host plants through a feeding structure called a haustorium, which extracts water, nutrients and carbohydrates from the host (Hegenauer et al., 2017; Westwood et al., 2010). Once above ground, each plant can produce hundreds of thousands of microscopic seeds that are easily dispersed. As this cycle occurs annually, the seed bank grows and can render fields unusable for growing susceptible crops (Joel, 2013).

Few effective methods have been devised to control parasitic plants. This is due to many compounding factors. First, the parasite and host plant share a physical connection with a bi-directional exchange of materials, which creates challenges for the selective control of parasites using herbicides. Second, their primarily underground lifecycle makes detection difficult until after the parasites emerge above ground, at which time the crop damage has mostly been done. Lastly, control techniques that rely on high capital investment and technological sophistication (e.g., precision drip chemigation) are not readily transferred to small-holder substance agriculture (Yoder & Scholes, 2010). Research is needed to better understand the biology of these parasites so that more control methods can be devised.

A key feature of parasitic plant seeds is the ability to remain dormant until the root of an appropriate host is within the vicinity. This is due to the ability of these seeds to detect specific combinations and concentrations of chemicals exuded from the roots of a host plant. This tight regulation of seed germination is crucial because, once stimulated, the seed radicle has about 72 hours to make contact with a host root before the stored seed reserves are exhausted (Joel & Bar, 2013). For most members of Orobanchaceae, the germination stimulant is a strigolactone (SL)

(Yoneyama et al., 2013), however there are a few exceptions, such as *O. cumana,* which responds to a sesquiterpene lactone called dehydrocostus lactone (DCL) (Joel et al., 2011). Such differences between species can be used to understand stimulant specificity. *Orobanche cernua* parasitizes Solanaceous crops such as tomato and responds to SLs, while *O. cumana* parasitizes sunflower and responds to dehydrocostus lactone (DCL). Both species germinate in response to the synthetic strigolactone GR24. Crosses between *O. cernua* and *O. cumana* were used to investigate the basis for stimulant specificity (Chap. 2).

The dependence of parasite seed on a host stimulant has generated intense interest in understanding how parasitic plants selectively respond to specific combinations of germination stimulants (Screpanti et al., 2016). Current research efforts have focused on creating parasite-resistant crops that vary in their levels or composition of SLs so as to evade parasitic plant detection (Gobena et al., 2017). This approach shows promise, but in order to keep ahead of the evolutionary arms race, it is important to understand the genes involved in the SL perception mechanism of the parasite. Our work suggests that additional genes or processes are important in controlling stimulant perception (Larose et al. chapter 2). However, there is currently a lack of available genomic resources available for Orobanchaceae species, with no published genomes for these species, and transcriptome information limited to a few species.

Here we analyze the transcriptomes of two species of parasitic Orobanchaceae that are problematic weeds in agriculture, and that differ in their germination stimulant preferences. These transcriptomes provide a resource for investigating aspects of parasitic plant biology related to seed conditioning and germination, and provide a point of comparison between closely related species. Our analyses indicate that these species share similar molecular and biological processes, yet also show evidence of substantial evolutionary divergence.

3.2 Materials and methods:

3.2.1 Seed sources

Orobache cernua seeds were collected in tomato fields in Israel in 1994, and *O. cumana* seeds were collected in sunflower fields in Israel in 1997. *Orobanche cernua* and *O. cumana* were then grown each year in a net-house at Newe Ya'ar Research Center for seed production.

3.2.2 Tissue collection for RNA-sequencing

Tissue collection for RNA sequencing is described in Larose et al. Chapter 2.

3.2.3 De novo transcriptome assembly of O. cumana and O. cernua

Raw read quality was assessed using FastQC (Andrews, 2014). Prior to assembly, raw reads were trimmed to remove poor quality reads and Illumina adaptor sequences using Trimmomatic (Bolger et al., 2014). A minimum length of 50bp after trimming was required to retain the read. The reads that retained their paired-end mate were used for sequence assembly. The Trinity software package (version 2.4.0) was used for de novo transcriptome assembly of *O. cumana* and *O. cernua* transcriptomes using default parameters (Haas et al., 2014). Transcriptomes were assembled for each species using the processed PE reads from all sequenced stages of that species. Raw reads were aligned back to the reference transcriptomes using Bowtie2 under default parameters (Langmead & Salzberg, 2012). The CEGMA pipeline was run using tblastx parameters against the core CEGMA database at default parameters to measure transcriptome comprehension (Parra et al., 2007).

3.2.4 Transcriptome functional annotation

The PlantTribes pipeline was used to functionally annotate the *de novo* transcriptomes of *O. cumana* and *O. cernua* (Wall et al., 2008). The PlantTribes pipeline calls multiple software programs for complete functional annotation including HMMER, MAFFT and ESTScan (Finn, Clements, & Eddy, 2011; Iseli, Jongeneel, & Bucher, 1999; Katoh & Standley, 2013). The AssemblyPostProcessor pipeline and GeneFamilyClassifier pipeline were run using default parameters. ESTScan was used to calculate putative coding regions within each *de novo* transcriptome. After ESTScan prediction, the resulting ESTs were annotated against the internal 22 plant genomes database contained within PlantTribes using the BLASTP algorithm with an E-value $< 10^{-5}$ (Altschul et al., 1997).

3.2.5 Determining expressed genes

A gene was considered to be 'expressed' within the transcriptomes if the average of the two replicates had a count greater than or equal to one across any stage. Expressed genes were further divided into stages by averaging the two replicates per stage and were assigned as expressed if the average count was greater than or equal to one.

3.2.6 Gene Ontology

Gene ontology terms were assigned for each species through the PlantTribes functional annotation software (Wall et al., 2008). Overall GO annotation results were assessed using the Web Gene Ontology Annotation Tool (WEGO) (Ye et al., 2006). Gene ontology term enrichment was performed using Bioconductor's TopGO program using Fisher's exact test and ranked with three different methods, elim, weight and classic (Alexa, Rahnenführer, & Lengauer, 2006). The three different methods used within TopGO all use different approaches to ranking significant (pre-defined P-value) GO terms: 1. elim method, which assess GO term hierarchy from bottom (more specific terms) to top (most general terms), 2. weight method, which compares significance scores of connected notes of parent and child GO hierarchy to detect the most significant local terms in the GO hierarchy, and 3. Classic method, where each GO term is tested independently, not taking into account GO hierarchy (Alexa et al., 2006). Results for all three tests are presented and sorted based on the weight method.

3.2.7 Differential expression:

Bioconductor's edgeR software package was used to estimate differential expression levels (M. D. Robinson, McCarthy, & Smyth, 2009). Read counts were used to estimate gene expression levels by the software. The significance of expression was determined by the FDR value and minimum fold change. The FDR threshold was set at <0.01 and the minimum fold change required was two.

3.3 Results and discussion

3.3.1 Transcriptome assembly

To get an overview of gene expression during conditioning and release from dormancy, transcriptomes of *O. cernua* and *O. cumana* were sequenced as described in Larose et al. (Chapter 2). To briefly summarize, the transcriptomes were sampled at three time points: During conditioning (pools of seeds collected at 1, 3, and 5 days after start of imbibition), at completion of conditioning (seven days after start of imbibition), and after treatment with germination stimulants (pools of seeds four and eight hours after exposure to stimulants). The final stage was divided into treatments with species-specific stimulants (orobanchol (Oro) for *O. cernua*; DCL for *O. cumana*) or the universal stimulant GR24.

3.3.2 Transcriptome functional annotation

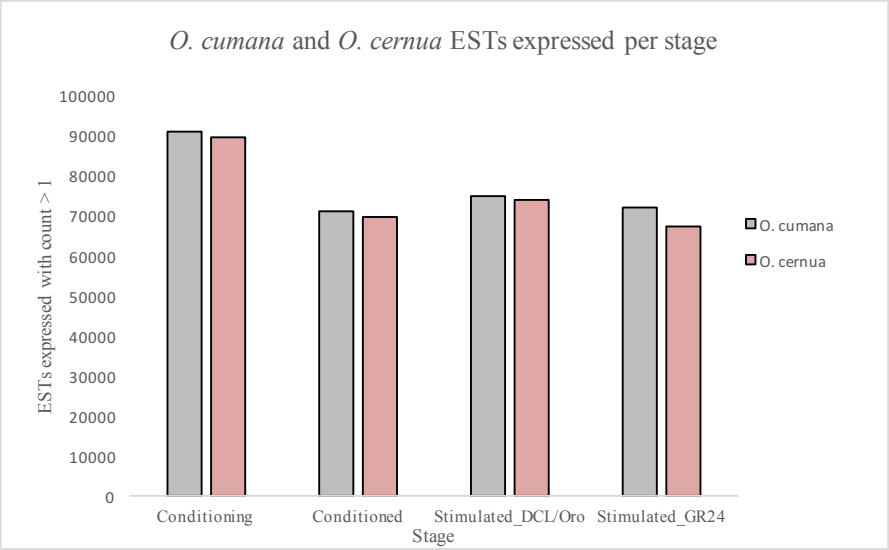
Two approaches were taken to investigate the transcriptomes of *O. cernua* and *O. cumana.* First, we explored the overall gene expression within each species independently. Second, we explored the conserved gene expression within both species by identifying orthologous pairs and examining expression patterns within each stage.

For each species, we ran the assembled transcriptome through PlantTribes, a post-processing pipeline for *de novo* transcriptome assemblies (Wall et al., 2008). It produced predicted coding regions, and their corresponding amino acid translations using ESTScan, followed by BLASTP to annotate each contig against 22 plant genome databases, resulting in a summary table for transcripts classified into orthologous plant gene family clusters with their corresponding functional annotations (Wall et al., 2008). Through this pipeline, 103,570 and 110,019 predicted coding regions (further termed ESTs) were generated for *O. cernua* and *O. cumana,* respectively (Figure 3.1a). Annotation of the predicted coding regions was performed with the BLASTP algorithm against twenty-two plant genomes and resulted in the functional annotation of 59,199 (57.2%) and 65,819 (59.8%) ESTs for *O. cernua* and *O. cumana*, respectively.

3.3.3 Determining expressed genes

To define which ESTs are expressed during conditioning, at fully conditioned, and after stimulant exposure, all raw reads were aligned to the *de novo* assembled transcriptome to estimate read counts using RSEM (Li & Dewey, 2011). An EST was considered to be expressed in a particular stage if the average of the two replicates had a count greater than one. Results are summarized in Figure 3.1. For both *O. cernua* and *O. cumana*, over fifty percent of ESTs are co-expressed across all sequenced stages of dormancy release. Interestingly, in both species around 17 percent of ESTs are unique to conditioning, a period when the seed is non-responsive to germination stimulant. This implies that seed conditioning includes many process that are unique to this stage of development (Figure 3.1b and 3.1c).

A)



B)

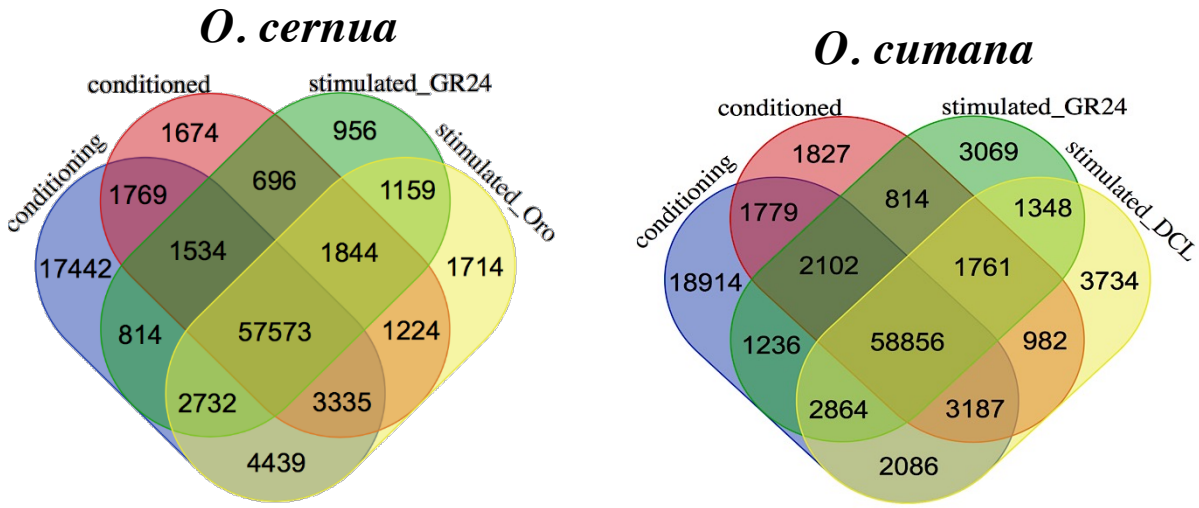


Figure 3.1 Overview of gene expression in parasitic seeds. A) Expressed ESTs per stage for *O. cernua* and *O. cumana*. Note that ESTs may not be unique to each stage. Total of 110,019 and 103,570 predicted coding regions for *O. cumana* and *O. cernua,* respectively. B) Total expressed *O. cernua* and *O. cumana* ESTs clustered by co-occurrence within stages.

3.3.4 Gene Ontology

GO terms represent a standardized way to describe the characteristics of genes and their associated biological functions. There are three main categories of GO terms: cellular processes, biological processes and molecular functions. To explore the overall events occurring within

parasitic plant seeds as it progresses through dormancy release, we assigned GO terms to our ESTs for functional classification.

A total of 59,199 and 65,819 ESTs were annotated for *O. cernua* and *O. cumana* respectively. Of the 65,819 annotated *O. cumana* contigs, a total of 15,729 (23.9%), 47,499 (72.1%), and 36,560 (55.5%) were assigned GO terms in the cellular components, molecular functions and biological process categories, respectively. Of the 59,199 *O. cernua* contigs, a total of 13,747 (23.2%), 42,501 (71.8%), and 32,329 (54.6%) were assigned GO terms in the cellular components, molecular functions and biological process categories, respectively. Both species showed remarkably similar GO term profiles (Figure 3.2). For both species, the majority of ESTs under cellular components were involved in cell (>20%), cell parts (>20%) and organelles (>15%). Within the molecular functions category, the ESTs from both species fell within binding (>70%) and catalytic activity (>59%). For biological processes, the major categories were metabolic process (>60%), cellular process (>50%) and biological regulation (>12%) (Figure 3.2).

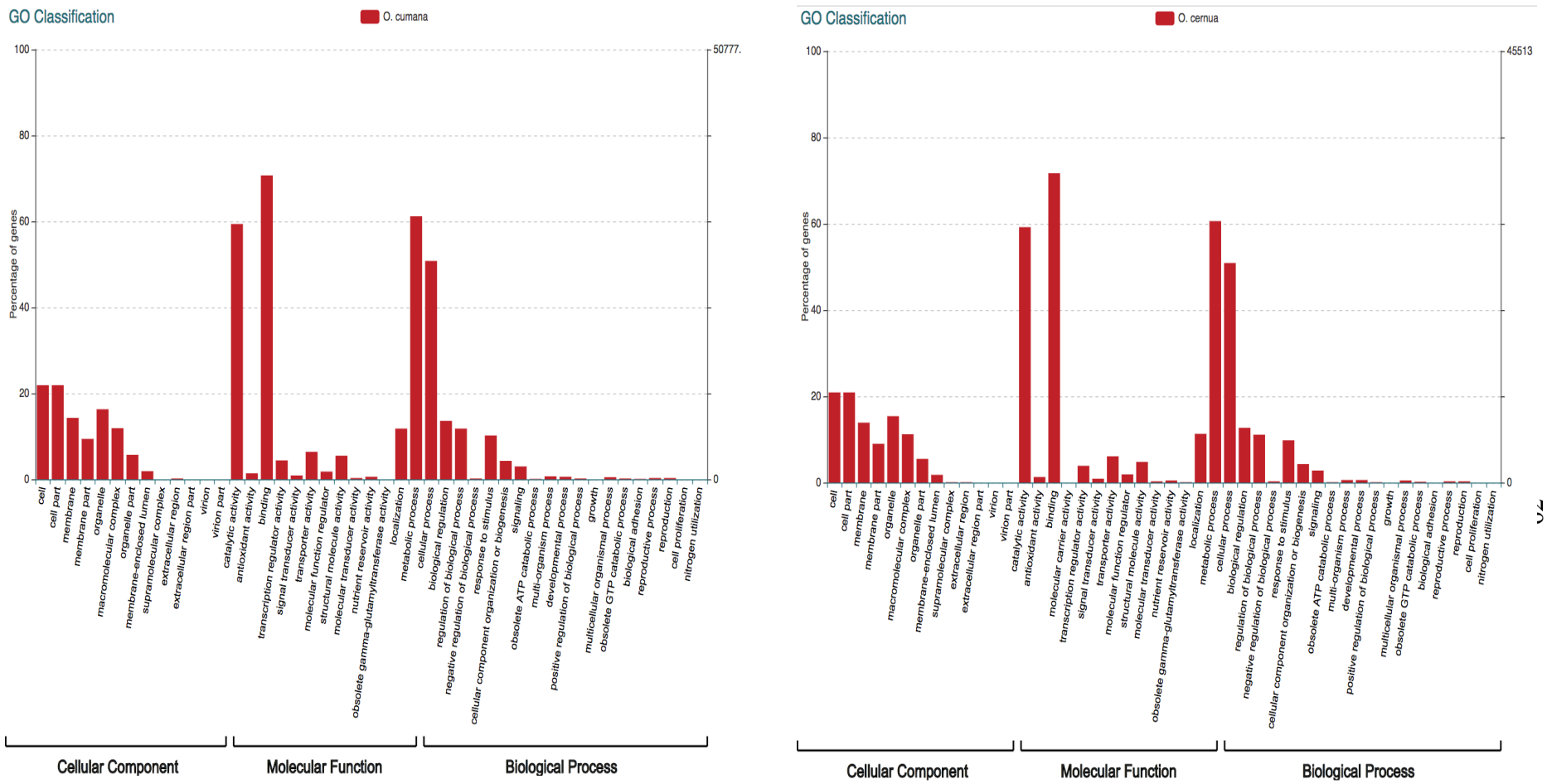


Figure 3.2 GO category distribution of *O. cumana* and *O. cernua* ESTs among the three GO categories: cellular component, molecular function and biological process.

Next, we examined each stage separately for enriched GO terms to look for specific processes abundant in each stage. Enriched GO terms were calculated for each stage of each species using Bioconductor's TopGO package in R using Fisher's exact test and sorted using the 'weigh' method, which compares significance scores of connected notes of parent and child GO hierarchy to detect the most significant local terms in the GO hierarchy (Alexa et al., 2006). For both species, across all stages the enriched GO terms are consistent in their presence, despite their arrangement according to significance. Notably, across the stages of fully conditioned, stimulated with GR24 and stimulated with species-specific stimulant, each are enriched for the biological processes of protein phosphorylation (GO:0006468), regulation of transcription (GO:0006355) and response to hormone (GO:0009725). This suggests that the seeds are undergoing changes in gene regulation, possibly in response to shifting hormone profiles, as they prepare for germination (Table 3.1).

Biological Process - *O. cernua* fully conditioned

	GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	elimFisher	weightedFisher
1	GO:0006468	protein phosphorylation	4792	4009	3743.79	20	3.00E-25	3.00E-25	5.00E-26
2	GO:0006278	RNA-dependent DNA biosynthetic process	5941	4925	4641.45	21	4.10E-24	4.10E-24	4.10E-24
3	GO:0006355	regulation of transcription, DNA-templat...	3455	2895	2699.25	32	9.00E-19	1.30E-17	3.90E-17
4	GO:0015074	DNA integration	4973	4065	3885.19	57	5.00E-12	5.00E-12	5.00E-12
5	GO:0009725	response to hormone	288	263	225	60	1.80E-09	1.80E-09	1.80E-09
6	GO:0006281	DNA repair	1576	1318	1231.26	65	1.40E-08	4.30E-07	4.70E-08

Biological Process - *O. cernua* stimulated with GR24

	GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	elimFisher	weightedFisher
1	GO:0006468	protein phosphorylation	4792	3872	3635.14	8	5.10E-19	5.10E-19	9.50E-20
2	GO:0006355	regulation of transcription, DNA-templat...	3455	2808	2620.91	19	4.20E-16	3.30E-15	4.70E-15
3	GO:0009725	response to hormone	288	260	218.47	36	2.60E-10	2.60E-10	2.60E-10
4	GO:0006351	transcription, DNA-templated	4047	3296	3069.99	5	5.20E-20	1.30E-05	2.00E-05
5	GO:0006396	RNA processing	1109	891	841.27	58	1.70E-04	1.70E-04	2.90E-05
6	GO:0006284	base-excision repair	62	59	47.03	54	5.10E-05	5.10E-05	5.10E-05

Biological Process - *O. cernua* stimulated with orobanchol

	GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	elimFisher	weightedFisher
1	GO:0006468	protein phosphorylation	4792	4092	3927.41	22	4.30E-12	4.30E-12	1.90E-12
2	GO:0006355	regulation of transcription, DNA-templat...	3455	2978	2831.63	7	1.00E-12	6.80E-12	7.90E-12
3	GO:0009725	response to hormone	288	266	236.04	38	3.20E-07	3.20E-07	3.20E-07

Biological Process - *O. cumana* fully conditioned

	GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	elimFisher	weightedFisher
1	GO:0006278	RNA-dependent DNA biosynthetic process	6208	4694	4301.65	17	< 1e-30	< 1e-30	< 1e-30
2	GO:0006468	protein phosphorylation	5347	4000	3705.05	28	4.10E-22	4.10E-22	2.30E-22
3	GO:0006355	regulation of transcription, DNA-templat...	4130	3108	2861.76	32	1.60E-19	5.40E-18	3.00E-17
4	GO:0015074	DNA integration	5409	4002	3748.01	51	1.20E-16	1.20E-16	1.20E-16
5	GO:0009725	response to hormone	289	257	200.25	55	1.80E-15	1.80E-15	1.80E-15
6	GO:0006281	DNA repair	1861	1457	1289.53	36	3.30E-19	6.50E-16	2.60E-15

Biological Process - *O. cumana* stimulated with GR24

	GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	elimFisher	weightedFisher
1	GO:0006468	protein phosphorylation	5347	4147	3811.03	15	2.00E-29	2.00E-29	6.70E-30
2	GO:0006355	regulation of transcription, DNA-templat...	4130	3203	2943.63	32	2.30E-22	7.30E-21	1.80E-20
3	GO:0006278	RNA-dependent DNA biosynthetic process	6208	4667	4424.7	52	2.50E-14	2.50E-14	2.50E-14
4	GO:0006281	DNA repair	1861	1463	1326.41	56	9.60E-14	1.40E-10	3.70E-13
5	GO:0009725	response to hormone	289	254	205.98	59	9.50E-12	9.50E-12	9.50E-12
6	GO:0006396	RNA processing	1158	923	825.36	62	1.80E-11	3.90E-08	4.20E-08

Biological Process - *O. cumana* stimulated with DCL

	GO.ID	Term	Annotated	Significant	Expected	Rank in classicFisher	classicFisher	elimFisher	weightedFisher
1	GO:0006278	RNA-dependent DNA biosynthetic process	6208	5021	4639.06	14	< 1e-30	< 1e-30	< 1e-30
2	GO:0006468	protein phosphorylation	5347	4281	3995.66	24	1.80E-23	1.80E-23	6.20E-24
3	GO:0006355	regulation of transcription, DNA-templat...	4130	3329	3086.23	31	1.80E-21	4.70E-20	9.30E-20
4	GO:0015074	DNA integration	5409	4291	4041.99	51	4.70E-18	4.70E-18	4.70E-18
5	GO:0006281	DNA repair	1861	1526	1390.67	56	1.10E-14	2.80E-13	1.40E-13
6	GO:0009725	response to hormone	289	263	215.96	60	1.40E-12	1.40E-12	1.40E-12

Table 3.1 Enriched GO Terms in biological processes for *O. cernua* (top) and *O. cumana* (bottom) for the stages of fully conditioned, stimulated with GR24 and stimulated with species-specific stimulant (Oro or DCL). Only the top six significantly enriched GO terms are shown.

3.3.5 Differential expression

Sequencing the transcriptomes at different stages of dormancy release allows us to explore which genes are differentially expressed. The mRNA profiles of each stage provide a snapshot of current gene expression, whereas the changes, either upregulated or downregulated, provide an illustration of biological processes the seeds are undergoing in preparation for germination. To identify the differentially expressed genes (DEGs) within *O. cernua* and *O. cumana,* we used Bioconductor's edgeR program with a false discovery rate (FDR) cutoff of 0.01 and fold change > 2 (Robinson et al., 2009). Comparing the gene expression between while the seed is conditioning and after it has finished conditioning, within *O. cumana* there were 505 DEGs (169 up-regulated and 336 down-regulated), and 293 DEGs (95 up-regulated and 198 down-regulated) within *O. cernua.* Comparing gene expression between when the seeds are fully conditioning conditioned to post-GR24-stimulated seeds, *O. cumana* had 2,455 DEGs (1,364 up-regulated and 1,113 down-regulated), while *O. cernua* had 2,134 DEGs (992 up-regulated and 1,142 down-regulated). Lastly, when we compare gene expression between when seeds are fully conditioned and post-stimulated seeds with species-specific stimulant (DCL or Oro), within *O. cumana* there were 832 DEGs (583 up-regulated and 249 down-regulated) and within *O. cernua* 1,945 DEGs (916 up-regulated and 1,029 down-regulated).

3.3.6 Combining transcriptomes

To investigate common themes in gene expression between the two species, we sought to find orthologous genes in *O. cernua* and *O. cumana* and asked whether their expression patterns could provide insight into parasitic plant biology relating to seed conditioning and response to specific germination stimulants. To this end, we identified orthologous pairs between *O. cernua* and *O. cumana* through reciprocal best hits (rbh) of the ESTs for each species. This allows for the identification of conserved genes between the species that are similarly expressed during each stage. Our analysis yielded 25,632 orthologous pairs with at least 90% nucleotide similarity between *O. cernua* and *O. cumana.* This represents approximately 25% of all ESTs from each species. Finding this relatively low rate of orthology between *O. cernua* and *O. cumana* suggests that either the species-specific EST datasets have many unique variants, or that the two parasites are more evolutionarily diverged than has been thought. Data from more individuals will be

needed to discern whether the apparently high number of species-specific EST variants is a result of small sampling size or truly indicates differences between the two genomes.

Examining which rbh pairs are expressed during each stage revealed that during each stage, numerous genes involved in abscisic acid synthesis and catabolism are expressed as well as genes involved in gibberellic acid synthesis. Next, we explored the differential gene expression of the rbh pairs as the seed progresses through conditioning to conditioned and conditioned to stimulated with GR24 or species-specific stimulant (DCL or Oro). As the seed progresses from conditioning to conditioned, no rbh pairs were differentially expressed in both species. However, there were 61 rbh pairs differentially expressed in *O. cernua* only, and 144 rbh pairs differently expressed in *O. cumana* only. After stimulation with GR24, there were 27 rbh pairs differentially expressed within both species, 859 rbh pairs differentially expressed in only *O. cernua*, and 20 rbh pairs differentially expressed in only *O. cumana*. Stimulation with species specific stimulant revealed 87 rbh pairs that were differentially expressed within *O. cernua* and *O. cumana*, in response to Oro or DCL, respectively. Interestingly, there were 689 rbh pairs that were differentially expressed only in *O. cernua* in response to Oro, while there were 191 differentially expressed rbh pairs unique to *O. cumana* stimulation with DCL.

3.4 Conclusions

Here we present the transcriptomes of two species of parasitic Orobanchaceae*, O. cernua* and *O. cumana*, which will be made publically available for use by other researchers. The transcriptomes represent stages pre- and post-germination exposure and in response to different germination stimulants. For each species, we examined the overall gene expression through gene ontology annotations. Both species share a remarkably similar RNA expression profile, with nearly identical biological processes for each sampled developmental time point pre- and post-germination exposure. We also found that when the seeds are conditioned or stimulated with either GR24 or species-specific stimulant, the transcriptomes are enriched in the GO biological processes categories of protein phosphorylation, regulation of transcription, and response to hormone. This suggests that the seeds are undergoing changes in gene regulation, possibly in response to shifting hormone profiles, as they prepare for germination (Figure H). Despite the similarity of processes, we noticed a relatively low rate of orthology between *O. cernua* and *O. cumana* transcriptomes, suggesting these species are more diverged than originally thought.

Further work is needed to ascertain the relationship between *O. cernua* and *O. cumana.* These data will assist researchers interested in investigating gene expression in parasitic plants.

3.5 References

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, *22*(13), 1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST : a new generation of protein database search programs, *25*(17), 3389–3402.

Andrews, S. (2014). FastQC A Quality Control tool for High Throughput Sequence Data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Conn, C. E., Bythell-douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J. H., … Nelson, D. C. (2015). Parasitic Plants. *Science (New York, N.Y.)*, *2*(6247), 7–11.

Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server : interactive sequence similarity searching, *39*(May), 29–37. https://doi.org/10.1093/nar/gkr367

Gobena, D., Shimels, M., Rich, P. J., Ruyter-Spira, C., Bouwmeester, H., Kanuganti, S., … Ejeta, G. (2017). Mutation in sorghum *LOW GERMINATION STIMULANT 1* alters strigolactones and causes *Striga* resistance. *Proceedings of the National Academy of Sciences*, *114*(17), 4471–4476. https://doi.org/10.1073/pnas.1618965114

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Philip, D., Bowden, J., … Pochet, N. (2014). Reference generation and analysis with Trinity. *Nat Protocols* (Vol. 8). https://doi.org/10.1038/nprot.2013.084.De

Hegenauer, V., Körner, M., & Albert, M. (2017). Plants under stress by parasitic plants. *Current Opinion in Plant Biology*, *38*, 34–41. https://doi.org/10.1016/j.pbi.2017.04.006

Iseli, C., Jongeneel, C. V, & Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *International Conference on Intelligent Systems for Molecular Biology* 138–148.

Joel, D. M. (2013). Seed Production and Dispersal in the Orobanchaceae. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.), Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategie*s* (pp. 143–147).

Joel, D. M., & Bar, H. (2013). The Seed and the Seedling. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.), Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies (pp. 147–163).

Joel, D. M., Chaudhuri, S. K., Plakhine, D., Ziadna, H., & Steffens, J. C. (2011). Dehydrocostus lactone is exuded from sunflower roots and stimulates germination of the root parasite *Orobanche cumana*. *Phytochemistry*, *72*(7), 624–634. https://doi.org/10.1016/j.phytochem.2011.01.037

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7 : Improvements in Performance and Usability Article Fast Track, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2, *9*(4), 357–360. https://doi.org/10.1038/nmeth.1923

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-

2105-12-323
Parker, C. (2012). Parasitic Weeds: A World Challenge. *Weed Science*, *60*(2), 269–276. https://doi.org/10.1614/WS-D-11-00068.1
Parker, C. (2013). The Parasitic Weeds of the Orobanchaceae BT - Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.) (pp. 313–344). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-38146-1_18
Parra, G., Bradnam, K., & Korf, I. (2007). CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, *23*(9), 1061–1067. https://doi.org/10.1093/bioinformatics/btm071
Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616
Rodenburg, J., Demont, M., Zwart, S. J., & Bastiaans, L. (2016). Parasitic weed incidence and related economic losses in rice in Africa. *Agriculture, Ecosystems and Environment*, *235*, 306–317. https://doi.org/10.1016/j.agee.2016.10.020
Screpanti, C., Yoneyama, K., & Bouwmeester, H. J. (2016). Strigolactones and parasitic weed management 50 years after the discovery of the first natural strigolactone strigol: status and outlook. *Pest Management Science*, *72*(11), 2013–2015. https://doi.org/10.1002/ps.4436
Wall, P. K., Leebens-Mack, J., Müller, K. F., Field, D., Altman, N. S., & Depamphilis, C. W. (2008). PlantTribes: A gene and gene family resource for comparative genomics in plants. *Nucleic Acids Research*, *36*(SUPPL. 1), 970–976. https://doi.org/10.1093/nar/gkm972
Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends in Plant Science*, *15*(4), 227–235. https://doi.org/10.1016/j.tplants.2010.01.004
Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., … Wang, J. (2006). WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research*, *34*(WEB. SERV. ISS.), 293–297. https://doi.org/10.1093/nar/gkl031
Yoder, J. I., & Scholes, J. D. (2010). Host plant resistance to parasitic weeds; recent progress and bottlenecks. *Current Opinion in Plant Biology*, *13*(4), 478–484. https://doi.org/10.1016/j.pbi.2010.04.011
Yoneyama, K., Ruyter-Spira, C., & Bouwmeester, H. (2013). Induction of Germination. In D. M. Joel, J. Gressel, & L. J. Musselman (Eds.), *Parasitic Orobanchaceae: Parasitic Mechanisms and Control Strategies* (pp. 167–186).

Chapter 4

Synthesis and Future Directions

4.1 Project summary

The mechanisms by which parasitic plants of the family Orobanchaceae detect their hosts is a long-standing mystery in plant science. For over half a century it has been known that seeds of parasitic plants will lie dormant, despite optimal germination conditions, until they detect a host-derived germination stimulant. Upon perception of an appropriate germination stimulant, the parasite seeds will send out a radical that has approximately 72 hours to reach a host root before the limited nutrients within the seed are exhausted. Consequently, it is necessary for germination to be tightly controlled within these species, because germinating in the absence of a host is lethal. The practical impact of this plant signaling regulation is profound, as the parasites in this family include some of the most destructive weeds in the world, including broomrapes (*Orobanche* and *Phelipanche* species) and witchweeds (*Striga* species). Scientists have sought to understand the signaling mechanisms in order to produce crop plants that don't produce/exude the signal or to create chemicals that can mimic stimulants and artificially trigger parasite seed germination. Our goal was to further the understanding of the parasite germination mechanism by determining the genes involved in parasite host specificity in *Orobanche*.

Summary of Contributions:

1. We have developed transcriptomes for two parasitic weed species.
   As part of this project we developed extensive datasets of the sequences of all messenger RNAs expressed in *O. cernua* and *O. cumana* seeds pre- and post-germination stimulant. These data are useful for understanding the biology of these two species and their evolutionary relationship to each other. By using these transcriptomes, we were able to identify the *D14, MAX2* and *KAI2* genes in *O. cernua* and *O. cumana.* These data will be deposited in public databases for use by other researchers.
2. We have defined key aspects of the germination stimulant perception mechanism for *O. cernua*.
   Our analyses of *O. cernua* x *O. cumana* hybrid lines shows that response to strigolactones is associated with a *KAI2d* gene from the strigolactone-responsive *O. cernua* parent, *OrceKAI2d2*. Furthermore, *OrceKAI2d2* responds to strigolactones when expressed in a model plant system, while the additional three *OrceKAI2d* genes do not. This suggests that we have identified the specific strigolactone receptor in *O. cernua*.

3. We have not identified the dehydrocostus lactone receptor in *O. cumana*.

   We used *O. cumana* in the genetic analysis because it parasitized different hosts than *O. cernua* and germinates in response to a non-strigolactone chemical, dehydrocostus lactone (DCL). Our genetic analysis suggests that response to DCL is associated with a block of *KAI2d* genes from the *O. cumana* parent, but none of the *KAI2d* genes we assayed within our model plant system showed a response to this chemical. We propose that there may be another receptor able to perceive DCL, or that other interacting proteins specific to the parasite (i.e., that are not present in our *Arabidopsis* model system assay) are involved in regulating the response.

4. We provide evidence that the germination stimulant perception mechanism in parasitic plants is yet more complicated than currently imagined.

   Recent breakthroughs suggest that host specificity is mediated through an expansion in the *KAI2d* gene families in parasites (Conn et al., 2015; Toh et al., 2015; Tsuchiya et al., 2015; Yao et al., 2017). In this way, each different *KAI2d* gene would function to detect a different host-derived chemical, and taken together would allow the parasite to identify its host the way different taste buds on the tongue allow a person to identify food. While this model may hold true in general, our observations indicate that it is insufficient by itself. First, the multiple *KAI2d* genes in each species appear to be physically linked, making it challenging to identify the specific contribution of each gene, or even determining whether other linked (non-*KAI2d*) genes are involved. Second, we built a statistical model from our data that predicts the possibility of involvement of other, non-*KAI2d* genes. Third, we found that one of the *KAI2d* genes, and two of the *KAI2i* genes in *O. cumana* responds in our model plant assay to SLs that do not stimulate germination of the parasite seed; hence, other mechanisms must keep this from triggering germination. Taken together with our inability to identify the DCL receptor, we conclude that parasitic plants may have additional pathways to detect different classes of germination stimulants, or that the *KAI2d*-mediated signaling pathway leading to germination contains additional levels of regulation.

4.2 Project Future Directions

1. The role of *KAI2d* gene expression in modulating stimulant perception.

It would be interesting to know whether gene expression levels play a role in stimulant specificity. In this project, we explored whether the presence or absence of a *KAI2d* genes could explain the segregation patterns we see within our F_3 hybrid lines, but simple presence/absence evaluations may be misleading. While evidence points to *OrceKAI2d2* as the specific SL receptor in *O. cernua*, no single *KAI2d* gene from *O. cumana* could be correlated with DCL perception. This is in part due to the co-inheritance of blocks of *KAI2d* genes within both species. Presence of a gene does not always mean the gene is expressed during the necessary stage. Assaying the expression levels of the *KAI2d* genes within the hybrids that differ in germination response may help correlate a *KAI2d* gene with DCL perception.

In addition to differences in expression levels, these *KAI2* genes may be expressed only in certain tissues, consequently preventing or limiting their interaction with potential partners. To test this theory, single cells can be removed from seed sections and the mRNA analyzed for *KAI2d* expression, in both *O. cumana* and *O. cernua* parental and hybrid lines.

2. Using structural modeling to assess whether *KAI2* a reasonable candidate receptor for DCL perception

Our experiments suggest that a block of *KAI2d* genes are associated with DCL response in *O. cumana*, yet no single gene responded to DCL in our model plant system. This is despite the presence of two additional *KAI2d* genes in *O. cumana.* A reasonable approach is to consider whether a *KAI2* gene has a binding cavity compatible with DCL. The identification of SL agonists have yielded numerous compounds that do not share significant structural similarity to SL such as 2-methoxy-1-naphthaldehyde (2-MN) that fits within the binding cavity of the rice *D14* receptor, and soporodine (SOP), which binds to *AtKAI2* (Holbrook-Smith et al., 2016; Mashita et al., 2016). Both compounds inhibit *Striga hermonthica* germination, and may indicate that the ligand-binding pockets of the parasite receptors have the ability to accommodate a wide range of structural compounds (Lumba et al., 2017).

To accomplish this, we could first model the *Orobanche KAI2d* genes against crystalized structures of *AtD14, OsD14, AtKAI2, ShHTL5* and *ShHTLiB* for insight into whether a DCL

molecule could properly bind with a *OrcuKAI2d* gene (Toh et al., 2015; Xu et al., 2016; Zhao et al., 2013). Next, a yeast-two-hybrid assay would shed insight into whether an *OrcuKAI2d* gene can form a complex with interacting partners *MAX2* or *SMAX1/SMXL2* in a DCL-dependent manner. Any interactions from a Y2H assay can be verified through affinity chromatography or co-immunoprecipitation assays.

3. Could the *SMXL* gene family be involved in regulation?

Another question is whether the germination specificity is also regulated by the repressor proteins interacting with *KAI2d* genes. Within *Arabidopsis*, the interacting repressor is a gene family termed SUPPRESSOR OF MORE AXILLARY GROWTH2 1 (*SMAX1*) and SMAX1-LIKE (*SMXL2-8*). The *SMXL* gene family has been demonstrated as the repressor directly recruited by *D14/KAI2,* with *SMAX1* and *SMXL2* involved in *KAI2/KAR* interaction, *SXML3,4,5* involved in phloem formation independent of *KAI2/D14,* and *SMXL6,7,8* involved in *D14/SL* interaction (Stanga et al., 2016; Wallner et al., 2017). Yao et al., (2017) showed that the most sensitive SL receptor in *Striga, ShHTL7,* forms a complex with *ShMAX2* and the repressor protein *AtSMAX1* in the presence of GR24 (Yao et al., 2017). It is possible that in addition to the expanding *KAI2d* gene family, the *SMXL* gene family has also expanded within parasitic plants. Additional copies or modifications of *SMXL* genes may contribute to *O. cumana's* ability to perceive DCL if they have evolved to interact specifically with a DCL-bound *KAI2* receptor. Additionally, modifications in the *KAI2d/SMXL* binding domain may inhibit the SL-bound *KAI2d* genes within *O. cumana* from interacting and triggering germination.

Given that parasite *KAI2d* genes are able to interact with *Arabidopsis SMXL* proteins, it is not likely that each parasite species has a unique form of *SMXL* protein for variant-SL perception (Yao et al., 2017). But *KAI2d* presence and absence does not directly explain all germination responses (Chap. 2, Fig. 2), indicating the necessity of additional genes for a functional response. Searches within our transcriptomes show the expression of *O. cernua* and *O. cumana* orthologs with greater than 45% amino acid identities to *AtSMAX1* and *AtSMXL2*. However, the *Orobanche* genes appear to be substantially shorter than the versions in *Arabidopsis*. Orothologs for *SMXL2-8* were not found within our transcriptomes with confident similarity. It needs to be determined whether the *O. cernua* and *O. cernua* SL-

responsive *KAI2d* genes are capable of interacting with the expressed *SMAX1* and *SMXL2* genes.

4. Identification of the DCL receptor and/or additional genes contributing to DCL perception.

To identify a candidate gene for DCL perception, we can utilize the *O. cumana x O. cernua* hybrid genetic system for genome wide association studies. This project has developed a set of recombinant inbred lines of *O. cumana x O. cernua* hybrids. As of 2017, these RILs have been taken to the F_7 generation. Hybrid lines specific to DCL or Orobanchol perception have been selected and propagated. By the F_7 generation, the RILs should be highly homozygous at most loci, which is beneficial for marker identification. The genome of *O. cumana* has been sequenced and will be released to the public in 2018.

We tried a genotyping by sequencing approach on 95 hybrid F_2 individuals with offspring that segregate for stimulant preference (Chap. 2). However due to small sample size, lack of a reference genome, and high heterozygosity, marker identification has proved challenging and de novo linkage maps have failed to yield less than 80 linkage groups.

Another strategy to identify the DCL receptor would be to use the *O. cumana x O. cernua* hybrid genetic system to generate a set of backcrosses with the *O. cernua* parental line. F_1 hybrids with strong response to DCL and no response to Oro can be back-crossed with *O. cernua* parental lines and selected for DCL response in seeds. Repeated backcrossing into the *O. cernua* parental line will create a set of plants resembling the *O. cernua* parental genotype, but with introgressed genes contributing to DCL perception from *O. cumana.* If DCL perception is a result of a single dominant gene, this process involves four rounds of backcrossing (four growing seasons) as the F_1 lines have already been developed and phenotyped. This could take longer if more genes are involved. This method would strongly benefit from the sequencing of the *O. cernua* genome since the majority of the genetic background would be of the *O. cernua* parental origin.

5. Why are there so many *KAI2d* genes in parasitic plants?

Multiple groups have demonstrated that parasitic plants have an expanded *KAI2* gene family compared to their non-parasitic relatives (Conn et al., 2015; Toh et al., 2015; Tsuchiya et al.,

2015). The consensus is that this expansion has evolved to allow the parasites to detect unique compounds exuded from host roots. Within *O. cernua* and *O. cumana,* every *KAI2* gene is expressed, and at variable levels, pre- and post- germination stimulant exposure. Yet, out of a total of 16 *KAI2* genes between the two species, only four demonstrated SL perception when expressed in an *Arabidopsis kai2* mutant background. Tsuchiya et al., (2015) demonstrated that ten out of twelve identified *Striga KAI2* genes containing the catalytic triad could hydrolyze the SL agonist YLG and GR24. Toh et al., (2015) went on to further demonstrate that six *KAI2* genes out of the twelve *KAI2* in *Striga* could moderately to highly recover germination in *Arabidopsis kai2* mutants when exposed to SL stimulants, and concluded that the additional *ShHTL* genes must perform alternative functions in *Striga*. This begs the question of what are the roles, if any, of these additional *KAI2* genes within parasitic plants?

It is worth noting that our *Arabidopsis kai2-2* mutants carrying certain *KAI2d* transgenes showed noticeable phenotypic differences in plant architecture in the T_1 generation. Notably, the *OrcuKAI2d3* and *OrcuKAI2d4* expressing plants resemble a *smxl6,7,8/max2* knockout mutant, with elongated stems and little branching (Soundappan et al., 2015). While some branching is observed from the main stem, generally, the seed pods only forms on the main stem, with seed pods presenting as skinner and elongated than wild-type or *kai2-2* mutant. For *OrcuKAI2d6,* the phenotype resembles a *d14* knockout mutant, in which the SL pathway is constitutively off (Zheng et al., 2016). The plants are short, with unregulated branching, that form short and non-viable seed pods along all stems. These phenotypes may appear in *Arabidopsis* simply because the transgene insertion site interferes with other genes, but it is also possible that these transgenes are exaggerating a role they perform in SL signaling. Currently, it is unknown whether *KAI2* functions as a receptor for other, currently unidentified, plant hormones (outside of SL/KAR) in plants. However there is evidence that a yet unidentified Karrikin-Like ligand (KL) exists within plant hormone signaling (Conn et al., 2016).

For *O. cumana* we see one *KAI2d* gene and two intermediate *KAI2* genes responding to SL in the *Atkai2* complementation assay, yet *O. cumana* does not germinate in response to SLs that we know of. Has *O. cumana* deactivated its SL perception pathway in regards to germination? *KAI2d* genes likely interact with *SMAX1/SMXL* repressor proteins after binding

a SL molecule. *Orobanche cumana* could regulate its interaction with the *SMAX1/SMXL* repressor proteins by accruing mutations in the repressor, or within the *KAI2d* receptor, preventing successful interaction, and cutting the stimulant response pathway short. Additionally, these receptors may be separated by a spatial barrier, which would ensure that receptors capable of binding and hydrolyzing SLs are contained within tissues that are not exposed to the exogenous SL germination single. Evidence for this theory exists from Plakhine et al., 2012, in which they demonstrated that phenotypes for germination response appeared first in the F_3 population, presumably due to the maternal inheritance of perisperm tissue within the seed (Plakhine et al., 2012). *Orobanche* seeds are surrounded by an impermeable cuticle, and the only site for possible stimulant entrance is a set of cells surrounding the micropyle, which are maternally derived perisperm tissue (Joel et al., 2012). It's possible that these SL responsive *KAI2d* genes are functioning as receptors within *O. cumana* for endogenous SL, or a yet unidentified karrikin-like ligand.

4.3 Overall Model

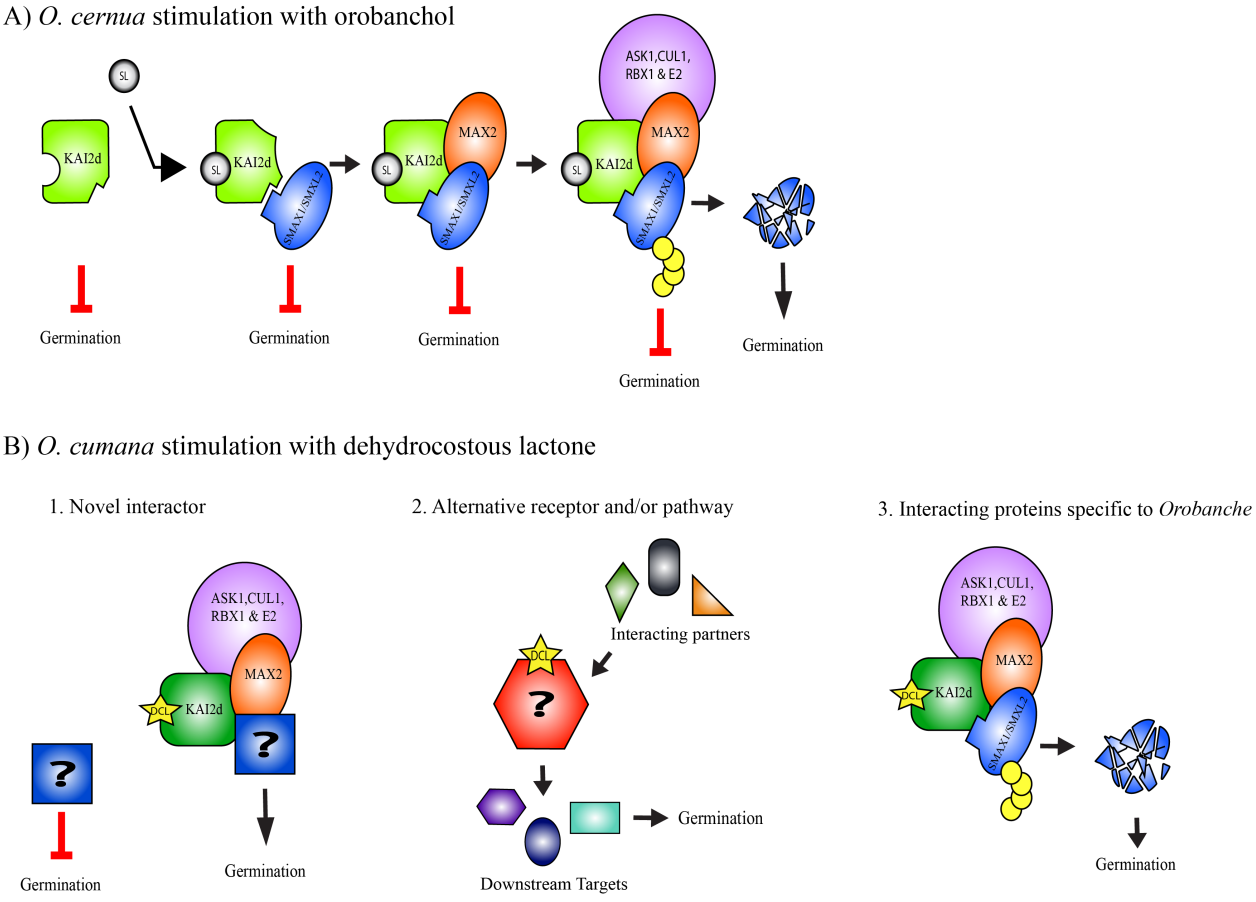


Figure 4.1 Proposed model of stimulant perception in parasitic Orobanchaceae. A) Within *O. cernua,* orobanchol is perceived by *OrceKAI2d2*. Once bound with a SL molecule, the *OrceKAI2d2* protein hydrolyses the SL, maintaining the D-ring covalently bound within the binding cavity, and releasing the ABC-ring. This hydrolysis induces a conformational change, exposing domains compatible with *MAX2* interaction. The *MAX2/KAI2* complex then recruits a repressor of the *SMXL* gene family, possibly *SMAX1* or *SMXL2*. The complex is bound by a E3 ubiquitin ligase SCF complex, polyuniquinated, and degraded. B) Within *O. cumana*, no single *KAI2d* gene was identified from our analyses as perceiving DCL. This could be due to 1) novel protein interactors within *O. cumana,* 2) an alternative receptor (non-*KAI2*) or different pathways exploited by *O. cumana,* or 3) interacting proteins unique to *O. cumana* and not present within our model system.

References

Conn, C. E., Bythell-douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J. H., … Nelson, D. C. (2015). Parasitic Plants. *Science (New York, N.Y.)*, *2*(6247), 7–11.

Holbrook-Smith, D., Toh, S., Tsuchiya, Y., & McCourt, P. (2016). Small-molecule antagonists of germination of the parasitic plant *Striga hermonthica*. *Nature Chemical Biology*, *12*(9), 724–729. https://doi.org/10.1038/nchembio.2129

Joel, D. M., Bar, H., Mayer, A. M., Plakhine, D., Ziadne, H., Westwood, J. H., & Welbaum, G. E. (2012). Seed ultrastructure and water absorption pathway of the root-parasitic plant Phelipanche aegyptiaca (Orobanchaceae). *Annals of Botany*, *109*(1), 181–195. https://doi.org/10.1093/aob/mcr261

Lumba, S., Holbrook-Smith, D., & McCourt, P. (2017). The perception of strigolactones in vascular plants. *Nature Chemical Biology*, *13*(6), 599–606. https://doi.org/10.1038/nchembio.2340

Mashita, O., Koishihara, H., Fukui, K., Nakamura, H., & Asami, T. (2016). Discovery and identification of 2-methoxy-1-naphthaldehyde as a novel strigolactone-signaling inhibitor. *Journal of Pesticide Science*, *41*(3), 71–78. https://doi.org/10.1584/jpestics.D16-028

Plakhine, D., Tadmor, Y., Ziadne, H., & Joel, D. M. (2012). Maternal tissue is involved in stimulant reception by seeds of the parasitic plant *Orobanche*. *Annals of Botany*, *109*(5), 979–986. https://doi.org/10.1093/aob/mcs016

Soundappan, I., Bennett, T., Morffy, N., Liang, Y., Stanga, J. P., Abbas, A., … Nelson, D. C. (2015). SMAX1-LIKE/D53 Family Members Enable Distinct *MAX2*-Dependent Responses to Strigolactones and Karrikins in Arabidopsis. *The Plant Cell*, *27*(11), 3143–3159. https://doi.org/10.1105/tpc.15.00562

Stanga, J. P., Morffy, N., & Nelson, D. C. (2016). Functional redundancy in the control of seedling growth by the karrikin signaling pathway. *Planta*, *243*(6), 1397–1406. https://doi.org/10.1007/s00425-015-2458-2

Toh, S., Holbrook-smith, D., Stogios, P. J., Onopriyenko, O., Lumba, S., Tsuchiya, Y., … Mccourt, P. (2015). Structure-function analysis identifies highly sensitive strigolancone receptors in Striga. *Plant Science*, *350*(6257), 203–208.

Tsuchiya, Y., Yoshimura, M., Sato, Y., Kuwata, K., Toh, S., Holbrook-Smith, D., … Hagihara, S. (2015). Probing strigolactone receptors in *Striga hermonthica* with fluorescence. *Science*, *349*(6250), 864–868. https://doi.org/10.1126/science.aab3831

Wallner, E. S., López-Salmerón, V., Belevich, I., Poschet, G., Jung, I., Grünwald, K., … Greb, T. (2017). Strigolactone- and Karrikin-Independent *SMXL* Proteins Are Central Regulators of Phloem Formation. *Current Biology*, *27*(8), 1241–1247. https://doi.org/10.1016/j.cub.2017.03.014

Xu, Y., Miyakawa, T., Nakamura, H., Nakamura, A., Imamura, Y., Asami, T., & Tanokura, M. (2016). Structural basis of unique ligand specificity of *KAI2*-like protein from parasitic weed Striga hermonthica. *Scientific Reports*, *6*(July), 1–9. https://doi.org/10.1038/srep31386

Yao, R., Wang, F., Ming, Z., Du, X., Chen, L., Wang, Y., … Xie, D. (2017). ShHTL7 is a non-canonical receptor for strigolactones in root parasitic weeds. *Cell Research*, *27*(6), 838–841. https://doi.org/10.1038/cr.2017.3

Zhao, L. H., Edward Zhou, X., Wu, Z. S., Yi, W., Xu, Y., Li, S., … Eric Xu, H. (2013). Crystal structures of two phytohormone signal-transducing α/β hydrolases: Karrikin-signaling *KAI2* and strigolactone-signaling DWARF14. *Cell Research*, *23*(3), 436–439.

https://doi.org/10.1038/cr.2013.19
Zheng, K., Wang, X., Weighill, D. A., Guo, H. B., Xie, M., Yang, Y., … Chen, J. G. (2016). Characterization of DWARF14 Genes in Populus. *Scientific Reports*, *6*(October 2015), 1–11. https://doi.org/10.1038/srep21593