THE MOLECULAR EVOLUTION OF STRIGOLACTONE PERCEPTION IN PLANTS

by

CAITLIN ELIZABETH CONN

(Under the Direction of David C. Nelson)

ABSTRACT

Root parasitic plants in the Orobanchaceae germinate in response to chemicals exuded by plants, including the hormones strigolactones. Strigolactones are important for beneficial symbioses with fungi and for many aspects of plant development; however, strigolactones are only known to stimulate germination of parasites. Soon after parasitic plants germinate, they can cause irreversible damage to hosts, which often include important crops. Thus, strigolactone-responsive seed germination is a promising target for the development of parasite control strategies. However, little was known about strigolactone perception in parasites until recently.

We identified candidate genes for roles in host-responsive seed germination in parasites based on signaling systems in the model non-parasitic plant *Arabidopsis thaliana*. In *A. thaliana*, *KAI2* responds to smoke-derived germination stimulants called karrikins, and probably an unidentified endogenous signal called KAI2 ligand (KL). Its homolog *D14* is required for strigolactone responsiveness in *A. thaliana* and other model non-parasites. We discovered extensive duplication of *KAI2* in parasite genomes. We functionally characterized *KAI2* paralogs from parasites by testing them as transgenes in *A. thaliana* null mutants, and we found that some parasite *KAI2* confer strigolactone-responsive seed germination. Other *KAI2* paralogs from parasites function more similarly to Arabidopsis *KAI2* (*AtKAI2*), although one responds preferentially to karrikin, and others to KL. Thus, gene duplication and sub- and neofunctionalization have likely

shaped *KAI2* genes in parasites, some of which enable host-responsive seed germination.

We next investigated the molecular basis for differences in likely ligand preference among different *KAI2* genes. We modified sites that we identified as potential specificity-determining positions (SDPs) in AtKAI2, and we tested AtKAI2 variants for responsiveness to various signals. Although we did not reconstitute strigolactone responsiveness, we made AtKAI2 more specific for karrikin or KL through different sets of targeted modifications. We also functionally characterized two *KAI2* paralogs from lettuce and found that they have likely subfunctionalized relative to *AtKAI2*, similarly to strigolactone-unresponsive *KAI2* from parasites. Together, our results illustrate the complex evolutionary history of *KAI2*. They also provide valuable information for the fight against parasitic plant infestations, which threaten food security in resource-limited parts of the world.

INDEX WORDS: strigolactone, karrikin, Orobanchaceae, parasitic plants, gene duplication, neofunctionalization, convergent evolution

THE MOLECULAR EVOLUTION OF STRIGOLACTONE PERCEPTION IN PLANTS

by

CAITLIN ELIZABETH CONN

BS, The Pennsylvania State University, 2011

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

THE MOLECULAR EVOLUTION OF STRIGOLACTONE PERCEPTION IN PLANTS

by

CAITLIN ELIZABETH CONN

Major Professor:	David C. Nelson
Committee:	Kelly Dyer
	Richard Meagher
	Robert Schmitz
	Chung-Jui Tsai

Electronic Version Approved:

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

DEDICATION

This dissertation is dedicated to my role model, my sister, and my dear friend, Nicole Goshorn Balich. Without her example of brilliance, determination, and courage, I never would have realized my own potential. Without her support, friendship, and love, I never would have become the person I am today.

ACKNOWLEDGEMENTS

I thank my advisor, David Nelson, for pushing me beyond my comfort zone and helping me to realize my potential as a scientist. I thank my committee for sharing their diverse expertise with me and for taking a sincere interest in my work. I thank the many collaborators, mentors, and teachers who have inspired me and helped me to grow. In particular, I am grateful to Margaret Varner, Andrew Meiser, Staci Baker, Katrina Robbins, Paulette Onestak, Robert Everly, Don Wilhour, Blair Hedges, and Stephen Schaeffer for contributing to my early love of science, my pursuit of the answers to my most pressing questions, and ultimately, my decision to work toward a doctorate in genetics. I also thank Stephanie Dennison, Elisabeth Rochel, and Carla Hass for their patience and kindness in teaching me molecular biology techniques.

I am indescribably grateful to my family and friends for their support, their faith in my ability as a scientist and student, and their love of the natural world, which they instilled in me at an early age. I am especially thankful to my parents, William and Janet Conn; my sister, Nicole Goshorn Balich; my grandparents, William and Anabel Conn and Barbara Gilson; my cousins, Neal and Leah Bryner; and my aunts and uncles, Joan and Rod Bryner, Jim and Joyce Gilson, and Rebecca and Michael Trego. I thank my creative, kind, and supportive partner Kimberly Kellett, who is a brilliant scientist and a fearless explorer. I also thank Daniel Frailey for his words of wit and wisdom, and for his dear friendship. Finally, I thank the many unmentioned friends and colleagues who have enriched my life in and out of graduate school.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Molecular evolution underlies phenotypic adaptation

An important consequence of natural selection is the adaptation of species to novel or changing environments. A shift in environmental conditions is often followed by a shift in the frequencies of phenotypes that are under selection from the environment. A classic example of this process of natural selection comes from the peppered moth, which can have primarily light or dark wings. Following the Industrial Revolution, an increase in the frequency of the dark-winged forms was observed in Britain. This shift in phenotypic frequency is linked to environmental change; as resting surfaces such as tree trunks grew darker with pollution, moths with dark wings were better camouflaged from predators and thus enjoyed survival and reproductive advantages. In Britain, darker forms of dozens of other Lepidopteran species increased in frequency as the country industrialized, although the peppered moth was the most thoroughly studied (Kettlewell 1955). Indeed, the peppered moth provides one of the best-known examples of a species evolving in the face of environmental change.

Phenotypic adaptations, such as darkly colored wings, are relatively simple to detect. However, the genetic aspect of environmental adaptation can be more difficult to ascertain. How does molecular evolution contribute to phenotypic evolution? How do genes evolve new or different functions without deleterious consequences? Nonsynonymous mutations are an important contributor to adaptive evolution, as they change at least the primary structure of the encoded protein. However, beneficial mutations are generally quite rare (Graur and Li 2000). According to the renowned

geneticist and evolutionary biologist Susumu Ohno, duplicate genes – or paralogs – are an important substrate for the accumulation of nonsynonymous mutations. In theory, as long as one paralog retains its original function, others may be freer to diverge in sequence and in function (Ohno 1970). Indeed, gene duplication has played a role in numerous adaptations of species to their environments. For example, duplicate genes encode antifreeze glycoproteins in Antarctic cod, help yeast adapt to nutrient limitation, and likely confer resistance to antimicrobial drugs in *Plasmodium falciparum* (Kondrashov 2012). Genes undergoing adaptive evolution are often under positive Darwinian selection, which in some cases can be inferred if the nonsynonymous substitution rate exceeds the synonymous substitution rate (Yang and Bielawski 2000). In summary, gene duplication and positive Darwinian selection are two processes associated with adaptive molecular evolution.

Signaling systems: key players in plant adaptive evolution

Signaling pathways enable organisms to sense and respond to their environments; thus, they play a crucial role in adaptive evolution. Environmental stimuli include chemical cues, some of which are involved in species interactions, including those between parasites/pathogens and hosts (e.g. Cook et al. 1966, Gerardo et al. 2006). One analysis of molecular evolution in distantly related eukaryotic taxa shows that rapidly evolving gene families that have undergone large copy number changes are enriched for several gene ontology categories, including chemoreception (Demuth and Hahn 2009). Signaling systems are particularly important for perception of the surrounding environment in plants, which are sessile and must therefore adapt to (rather than escape) new or changing conditions. For instance, diverse angiosperm species process chemical cues from the environment that promote seed germination. Suitable germination conditions vary among plants with different physiological needs. For

example, obligate parasitic plants in the Orobanchaceae must form a physical connection to a host plant's roots and steal its nutrients to survive (Westwood et al. 2010). Many obligate parasitic weeds in this family, including *Striga, Orobanche*, and *Phelipanche,* germinate in response to host hormones called strigolactones (SLs; Bouwmeester et al. 2003), which were first discovered in cotton root exudates (Cook et al. 1966). These parasite seedlings are only equipped with enough nutrients to survive on their own for a few days (Matusova et al. 2005), so a germination response to host hormones is likely to be adaptive.

While an environment with high plant density might be ideal for the germination of parasitic plants, competition for light and soil resources is likely to be heavy. Many non-parasitic plant species germinate in post-fire environments (Van Staden et al. 2000), in which competition is likely to be reduced (Nelson et al. 2010). For diverse angiosperms, including monocot and dicot species from at least three different continents, smoke-derived compounds called karrikins (KARs) are germination stimulants (Flematti et al. 2004, Daws et al. 2007, Stevens et al. 2007). Karrikins also stimulate seed germination in the model non-parasitic plant *Arabidopsis thaliana* (Nelson et al. 2009).

The karrikin and strigolactone signaling pathways

Although they stimulate seed germination in separate groups of plants, SLs and KARs share partial structural similarity – specifically, a butenolide moiety – and both classes of compounds require the F-box protein MAX2 for perception in Arabidopsis (Nelson et al. 2011). Despite these similarities in the SL and KAR signaling systems in Arabidopsis, each class of compounds has a unique receptor. The α/β hydrolase D14 is a receptor for SLs (Hamiaux et al. 2012, Kagiyama et al. 2013, Nakamura et al. 2013, Zhao et al. 2013), which influence various aspects of plant development as endogenous

phytohormones (reviewed in Ruyter-Spira et al. 2013). Perhaps most notably, SLs inhibit shoot branching, as evidenced by the extra axillary branches present in rice and Arabidopsis *d14* mutants (Arite et al. 2009, Waters et al. 2012). Strigolactones are also exuded from plant roots into the rhizosphere. There, they promote hyphal branching of arbuscular mycorrhizal fungi (AMF), which enter into mutualistic symbiotic relationships with plants (Akiyama et al. 2005). In Arabidopsis, *d14* mutants have short petioles and large rosettes prior to bolting, and short stature and extra axillary branches as adults. Additionally, *d14* mutants are insensitive to treatment with exogenous SLs, which enhance light sensitivity in wild-type seedlings (Waters et al. 2012).

Interestingly, the receptor for KARs in Arabidopsis is a D14 homolog called KAI2 (Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Zhao et al. 2013). In addition to KARs, KAI2 likely responds to an endogenous regulator of germination and development, which is currently referred to as KAI2 ligand (KL). Although KL has yet to be identified, the phenotypes of *kai2* mutants provide evidence for its existence. Besides insensitivity to KARs, *kai2* mutants have high seed dormancy, low light sensitivity, and altered rosette morphology (distinct from *d14* mutants). These KAR-independent phenotypes collectively indicate that *kai2* mutants cannot perceive an endogenous regulator of these processes (Waters et al. 2012). Additionally, *KAI2* orthologs from three different plant species rescue only KAR-independent *kai2* mutant phenotypes as transgenes, suggesting that they are specific for KL. One of these species is the basal land plant *Selaginella moellendorffii*, and the other two are parasitic angiosperms from the Orobanchaceae family. The presence of KL-sensitive KAI2 orthologs in such evolutionarily distant taxa suggests that KL response may be the ancestral function of KAI2 (Conn and Nelson 2016, Waters et al. 2015). This hypothesis is further supported by the conservation of *KAI2* in plants that are unlikely to encounter fire often; for example, KAR responsiveness in Arabidopsis may simply be due to structural similarity

between KL and KAR (Waters et al. 2012). However, because of its broad distribution – including in some fire-prone environments – KAR-responsive seed germination may be adaptive rather than coincidental in Arabidopsis (Lamont and He 2017). In rice, the *KAI2* ortholog *D14L* plays a critical role in symbiosis with AMF, with which more than 80% of plants associate. Thus, another possible ancestral function for KAI2 may be in this mutualistic relationship between plants and fungi (Gutjahr et al. 2015). Finally, in addition to KAR and KL, KAI2 in Arabidopsis also responds to an enantiomer of the SL mimic GR24$^{5\text{-deoxystrigol}}$ (GR24^{5DS}); however, it does not respond to SLs or SL mimics themselves (Scaffidi et al. 2014).

Despite similarities in the SL and KAR response systems and the ability of Arabidopsis KAI2 to respond to multiple signals, these two distinct pathways maintain specificity for their respective signals. The mutant phenotypes in *kai2* plants are expressed from seed germination to rosette development, while *d14* mutant phenotypes are apparent from seedling development through maturity. Unsurprisingly, the transcript abundance of *KAI2* is about 100-fold higher than that of *D14* in imbibed seed, but *D14* transcript abundance slightly surpasses *KAI2* in seedlings (Waters et al. 2012). However, expression differences are not sufficient to explain the specificity of the *D14* and *KAI2* signaling systems. Expressing *KAI2* under the control of the *D14* promoter fails to rescue *d14* mutant phenotypes, and the same is true of *D14* expression under the *KAI2* promoter in *kai2* mutants (Waters et al. 2015).

Rather than expression changes, differences in the ligand-binding pockets of D14 and KAI2 likely confer specificity for SL and KAR/KL, respectively. Rice and Arabidopsis D14 and their ortholog in petunia (DAD2) degrade a racemic mix of synthetic SL (GR24) *in vitro*, dependent upon a functional catalytic triad (Hamiaux et al. 2012, Nakamura et al. 2013, Zhao et al. 2013). Isothermal titration calorimetry (ITC) also indicates interaction between Arabidopsis D14 and GR24. According to a crystal

structure of rice D14, SL is likely to fit in the ligand-binding pocket (Kagiyama et al. 2013). However, due to an amino acid replacement between D14 and KAI2, D14 cannot bind KAR as KAI2 can (Zhao et al. 2013). Indeed, binding between D14 and a SL breakdown product has been observed by crystallography (Nakamura et al. 2013, Zhao et al. 2013). On the other hand, KAI2 has a smaller ligand-binding pocket than D14, at least partially due to a tyrosine residue (site 124) that projects into the ligand-binding pocket (Bythell-Douglas et al. 2013). Other amino acids in KAI2 probably contribute to a different ligand-binding pocket shape as well, such as 193 and 194, which affect the middle of the pocket. Molecular docking indicates that KAI2 can accommodate KAR as a ligand, but cannot accommodate GR24 in the way that D14 can (Zhao et al. 2013). Interaction between KAI2 and KAR_1 has been observed in an ITC experiment (Kagiyama et al. 2013) and by equilibrium microdialysis, nuclear magnetic resonance (NMR) spectroscopy, and co-crystallization of KAI2 and KAR_1 (Guo et al. 2013). Taken together, these biochemical data indicate specificity of D14 for SL and KAI2 for KAR.

Additionally, interactions between D14 or KAI2 and downstream signaling partners likely confer specificity in these two signaling pathways. Downstream of *MAX2*, paralogous heat-shock proteins (SMAX1 and its SMXL paralogs) enter the karrikin and strigolactone pathways. Based in part on phenotypes of *smax1*/*smxl* single mutants and *smax1*/*smxl max2* double mutants, it appears that *SMXL6*, *SMXL7*, and *SMXL8* repress SL signaling, while *SMAX1* represses KAR signaling (Stanga et al. 2013, Soundappan et al. 2015). Biochemical data also support a role for SMXL6 – 8 in the SL signaling pathway. Treatment with *rac*-GR24 results in ubiquitination and degradation of SMXL6, SMXL7, and SMXL8 in protoplasts in a D14- and MAX2-dependent manner (Wang et al. 2015). This mechanism of SL-induced degradation of SMXL proteins was originally proposed in rice, in which D53 is an ortholog of SMXL6 – 8 (Jiang et al. 2013, Zhou et al.

2013, Soundappan et al. 2015). As of yet, no biochemical data have been published on the mechanistic role of SMAX1 in the KAR signaling pathway.

The question of why D14 and KAI2 interact with different protein partners has recently been investigated. One amino acid residue on the surface of the D14 protein likely confers specificity for interaction partners. Amino acid replacement at this site renders D14 nonfunctional. This site (169 in D14, 168 in KAI2) is highly conserved as proline in D14 and as serine in KAI2. Because of its position in the protein and its strict conservation as one amino acid in D14 and as another in KAI2, site 169/168 is likely a specificity-determining position in these homologous proteins (Chevalier et al. 2014). Other sites in D14 that influence protein-protein interactions have been identified through crystallography. The interface between D14 and MAX2/D3 has been identified in the crystal structure of these two proteins complexed with ASK1 (Yao et al. 2016), which forms part of an E3 ligase complex with MAX2 (Nelson et al. 2011). However, this interface is also highly conserved in KAI2, indicating that it likely interacts with MAX2 similarly to D14 (Bythell-Douglas et al. 2017). Thus, these residues are unlikely to influence the different specificities of the D14 and KAI2 signaling systems.

The similarity and specificity of the D14 and KAI2 signaling systems raise questions about their evolutionary origins. The *KAI2* lineage extends back to early plant evolution. Orthologs of *KAI2* but not *D14* are present in charophytes (green algae), and although the function of these *KAI2* orthologs is unknown, some green algae produce SLs (Delaux et al. 2012). The moss *Physcomitrella patens* has 11 orthologs of *KAI2*, some of which respond transcriptionally to SL treatment (Lopez-Obando et al. 2016). According to numerous phylogenetic analyses, *D14* arose later in plant evolution via a duplication of *KAI2* in vascular plants (e.g. Delaux et al. 2012, Waters et al. 2012, Waters et al. 2015, Lopez-Obando et al. 2016). The origin of *D14,* coupled with duplication of D53/SMAX1 in early land plants, may have enabled diversification of the

KAI2 and D14 signaling pathways (Waldie et al. 2014). Another possibility is that a major *KAI2* duplication occurred earlier in land plant evolution, producing a clade of divergent *D14/DLK2/KAI2* (*DDK*) genes. (*DLK2* is another homolog of *D14* and *KAI2*, and its function is currently unknown; Waters et al. 2012). According to this model of evolution, *KAI2* has been highly conserved throughout plant evolution, but the *DDK* lineage has not. Additional gene duplications and evolutionary diversification of *DDK* genes may have resulted in a set of *KAI2* homologs that vary in sequence, function, and copy number across different plant taxa (Bythell-Douglas et al. 2017). These two evolutionary scenarios disagree somewhat on the origin of *D14* and the fate of *KAI2* duplicates. However, both models of *KAI2* evolution illustrate the potential for evolutionary innovation after *KAI2* duplication.

Regarding the functional diversification of *KAI2* or *D14* duplicates, one particularly interesting taxonomic group is the Orobanchaceae. As noted above, parasitic weeds in this family germinate in response to host root exudates, which can include a mixture of SL structural variants and other related compounds (reviewed in Xie et al. 2010). However, the mechanism of SL perception in parasite seeds was unknown until recently. KAI2 and D14 were obvious candidates for SL receptors in parasite seeds. If KAI2 enables host-responsive seed germination in parasitic weeds, then presumably its ligand specificity has switched from KARs (and probably KL) to SLs. On the other hand, if D14 mediates this response to host hormones in parasite seeds, then it must have evolved a role in seed germination. Genetic resources for parasitic plants of the Orobanchaceae are limited; most publicly available sequence data are from incomplete transcriptome assemblies or an EST database (Westwood et al. 2012, Yoshida et al. 2010). Furthermore, parasitic weeds are tightly regulated in the United States, and growing them in a laboratory setting requires a special permit. Despite the challenges of studying SL perception in parasite seeds, this investigation is critical for the development

of new parasite control strategies. It also has the potential to increase our understanding of fundamental processes of molecular evolution, such as gene duplication, neofunctionalization, and adaptation to a parasitic lifestyle.

Parasitic plants of the Orobanchaceae

Parasitic plants are diverse in morphology and in habitat, but they are all united by one key feature: a haustorium. The haustorium is a specialized invasive structure with which parasitic plants invade the roots or shoots of a host plant and establish a vascular connection. Parasitism has likely evolved at least 12 times in plants, including once in the Orobanchaceae family. The plants in this family display a broad range of host dependence, from the non-parasitic genus *Lindenbergia* to obligate holoparasites that no longer photosynthesize and thus require a host for survival. Intermediate between these two extremes are facultative hemiparasites, which carry out photosynthesis and thus do not require a host, and obligate hemiparasites, which can photosynthesize to some extent but rely on a host during at least one life stage (Westwood et al. 2010). Host range and preference are variable in this family, and sometimes within a single genus or species. Although host/parasite compatibility is not well understood in this system, the composition of host root exudates appears to be an important contributor to host recognition, at least in *Orobanche* and *Phelipanche* species (Fernández-Aparicio et al. 2009).

The Orobanchaceae family includes some of the most destructive parasitic weeds in the world, which attack crop hosts and cause billions of dollars of damage annually. *Striga* species are obligate hemiparasites that usually parasitize grass crops and are tremendously harmful to smallholder farmers in sub-Saharan Africa. *Orobanche* and *Phelipanche* species are obligate holoparasites that attack a variety of important crops and can be found in parts of Europe, Africa, and Asia (Westwood et al. 2010).

Parasitic weeds are extremely difficult to control, in part because they produce large numbers of seed that maintain long-term viability. For example, *Striga* can produce ~100,000 seeds per plant, which can remain dormant in the soil for decades (Scholes and Press 2008). After germinating in response to a nearby host, parasite seedlings attach to the host plant and begin to damage it before becoming visible above the soil surface (Sauerborn et al. 2007). This renders control strategies that require visual identification of the parasite – such as hand weeding, and in some cases, herbicide application – largely ineffective.

Another control strategy is crop rotation, in which non-host or resistant host crops are grown in certain years. However, parasitic weeds of the Orobanchaceae are often generalists with broad host ranges, at least in the obligate holoparasitic genera *Orobanche* and *Phelipanche* (Schneeweiss 2007). Host-switching events have been described in this family; for example, the recently weedy species *Orobanche cumana* and *O. foetida* recently switched from wild to cultivated hosts. Although they are still considered specialists, both parasite species germinate in response to the root exudates of a relatively broad range of plants (Fernández-Aparicio et al. 2009). Thus, the possibility of parasitic weeds undergoing host range expansions and parasitizing crops that were once considered non-hosts should not be ignored. Furthermore, the cultivation of resistant hosts has preceded the discovery of new parasite races with higher virulence. In *O. cumana*, at least six races of parasite have been identified, which vary in their ability to overcome host resistance (discussed in Pérez-Vich et al. 2004). Multiple races of *Striga gesnerioides* have also been discovered that differ in the number and identity of cowpea cultivars that they can parasitize (Li et al. 2009).

A variety of other control strategies exist as well, but none is currently sufficient to solve the problem of parasitic weeds in resource-limited parts of the world. However, the unique adaptation in parasites of host-responsive seed germination has become an

attractive target for the development of new control strategies. Perhaps by manipulating parasite seed germination, parasitic weeds could be prevented from ever reaching their hosts in the first place. Indeed, suicidal seed germination, in which a parasite germinates in the absence of a host and therefore dies, has been investigated as a potential control strategy for decades (Zwanenburg et al. 2016). However, little was known about the mechanism of host-responsive seed germination in parasitic weeds until recently.

A framework for the genetic analysis of non-model parasitic plants

An understanding of how parasitic weeds germinate in response to a host is critical to the development of control strategies that target seed germination. What are the genetic components of the strigolactone signaling system in parasitic weeds? As described previously, studies of parasitic plant genetics are challenging for several reasons, including a lack of genetic resources and the strict regulation of parasitic weeds. Nevertheless, we set out to investigate host-responsive seed germination in parasitic weeds. We collected *KAI2* and *D14* sequences from parasitic plant sequence databases and performed next-generation shotgun genome sequencing for five parasite species. We compared *KAI2* and *D14* evolution in parasites and non-parasites and found that *KAI2* (but not *D14*) has undergone extensive duplication in parasitic weeds. Cross-species complementation indicated that some *KAI2* from parasitic plants function similarly to *KAI2* in Arabidopsis, while others confer SL responses in seed germination. Thus, gene duplication and subsequent neofunctionalization likely enabled host-responsive seed germination in parasitic weeds.

We combined several computational tools to address the question of how SL responsiveness evolved in some *KAI2* paralogs from parasites. We identified amino acid replacements that may affect ligand specificity in KAI2 proteins; however, we were unable to obtain a complete picture of how SL responsiveness may have evolved.

Finally, we used computational and cross-species complementation methods to investigate *KAI2* duplication in non-parasitic plants. We found no evidence of convergent molecular evolution in *KAI2* paralogs following independent duplication events. We functionally characterized two *KAI2* copies from lettuce and found that they have subfunctionalized, relative to Arabidopsis *KAI2*, similarly to some parasite *KAI2* paralogs. However, because lettuce and the Orobanchaceae are closely related, we cannot cite this as convergent evolution of function in *KAI2* duplicates. Other than in lettuce and parasitic weeds, the functional diversity of *KAI2* paralogs remains largely unexplored in angiosperms. Perhaps Arabidopsis, with its single *KAI2* gene, is not an ideal model for studying this signaling system. By exploring *KAI2* duplicates in diverse non-parasitic plants, we may develop a more thorough understanding of the adaptive evolution of duplicate genes. More importantly, by continuing to study *KAI2* in parasitic weeds, we may be able to develop more effective control strategies for these devastating threats to food security.

CHAPTER 2

NEOFUNCTIONALIZATION OF DUPLICATE GENES CONTRIBUTES TO HOST-RESPONSIVE SEED GERMINATION IN PARASITIC WEEDS[1]

[1] This work is published in Conn, C. E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543.

Abstract

Parasitic weeds in the Orobanchaceae family attack important crops around the world and cause billions of dollars of damage every year. In field conditions, the seeds of parasitic weeds germinate in response to nearby host plants. Specifically, parasite seeds perceive strigolactones (SLs), which are phytohormones that signal to beneficial arbuscular mycorrhizal fungi in the soil. To the seeds of a parasitic weed, SLs indicate that a potential host is nearby. The mechanism of SL perception in parasite seeds has not been elucidated; however, this information is crucial to the development of more effective control strategies for parasitic weed infestations.

In the model plant *Arabidopsis thaliana*, which is non-parasitic, two homologous α/β hydrolases are involved in SL signaling and in seed germination, respectively. Arabidopsis does not germinate in response to SLs, but these hormones play other roles in development. D14 is a SL receptor in Arabidopsis. Its homolog KAI2 is necessary for normal seed germination, as well as seedling light sensitivity and rosette development. KAI2 also mediates responses to smoke-derived germination cues called karrikins (KARs). KARs share partial structural similarity with SLs but are only known to stimulate seed germination in some non-parasitic plants.

Here, we take a candidate gene approach to elucidate the mechanism of SL signaling in parasitic weeds. We analyze *D14* and *KAI2* evolution and find that *KAI2* copy number is significantly higher in parasitic plants than in non-parasitic relatives. Next, we use cross-species complementation to study the function of parasite *KAI2* genes in an Arabidopsis *kai2* null mutant background. As transgenes, some parasite *KAI2* confer germination responses to SLs. Thus, they likely enable parasite seeds to detect a nearby host and germinate in response. The non-parasite SL receptor D14 is thought to have arisen in an ancient KAI2 duplication event; thus, SL sensitivity likely evolved convergently in non-parasite D14 and in some KAI2 from parasitic weeds.

Introduction

Parasitism has likely evolved at least a dozen times in plants. Parasitic plants invade host plant tissues, establish vascular connections, and steal resources at the host plant's expense. One particularly diverse and destructive family of parasitic plants is the Orobanchaceae. Parasitic weeds in this family attack the roots of numerous crop species and cause billions of dollars of yield losses each year (Westwood et al. 2010). The Orobanchaceae includes plants that span a broad range of host dependence. The non-parasite *Lindenbergia philippensis* has no need for a host plant, and facultative hemiparasites such as *Triphysaria versicolor* carry out photosynthesis themselves but can parasitize a host if one becomes available. On the other hand, *Striga hermonthica* is an obligate hemiparasite, meaning it can photosynthesize to some extent but is also dependent upon a host. Plants in the *Orobanche* and *Phelipanche* genera are obligate holoparasites, which do not photosynthesize and thus require a host for survival (Westwood et al. 2010, Westwood et al. 2012).

Species in the *Orobanchaceae* vary greatly in host range, as demonstrated by the obligate holoparasitic genera *Orobanche* and *Phelipanche*. Nonweedy *Orobanche* and *Phelipanche* species, which parasitize wild hosts, tend to be specialists with narrow host ranges. However, species that have become agricultural weeds, such as *Phelipanche aegyptiaca*, often have broader host ranges (Schneeweiss 2007). Two *Orobanche* species have recently become weedy, due to the adaptation of the parasites to new, cultivated hosts. How parasitic plants undergo host range expansions is not well understood, but recognition of the root exudates of host plants appears to be an important component of host/parasite compatibility (Fernández-Aparicio et al. 2009). Specifically, the seeds of parasitic weeds germinate in response to host-derived strigolactones (SLs; reviewed in Xie et al. 2010), which were first discovered as stimulants of seed germination for *Striga lutea* (Cook et al. 1966). Strigolactones also

signal to beneficial arbuscular mycorrhizal fungi in the rhizosphere, which interact with the roots of plants in mutualistic symbioses (Akiyama et al. 2005). Different plants can produce different types and relative amounts of SL structural variants (e.g. Awad et al. 2006, Jamil et al. 2011, Xie et al. 2013). Generalist parasites often germinate in response to root exudates from a wide variety of plant species, while specialists sometimes germinate almost exclusively in response to the exudates from their specific host (Fernández-Aparicio et al. 2009). Thus, recognition of host-derived SLs likely contributes to host range in parasitic weeds.

In agricultural systems, the battle against parasitic weed infestations is extremely challenging. Host-responsive seed germination, high fecundity, and long-term seed viability (Scholes and Press 2008) are aspects of parasitic plant biology that make them difficult to control. Additionally, parasitic weeds attach to and damage hosts before emerging through the soil surface (Sauerborn et al. 2007); thus, when parasitic weeds can be visually identified, host plants have often already been harmed. Unlike control strategies that target visible parasites, manipulation of parasite seed germination may be effective at preventing parasites from reaching their hosts in the first place. For example, suicidal seed germination occurs when a parasite germinates too far from a host root to survive, and germination is inhibited in some parasites when SL concentrations are too high. Thus, crops that over-export SLs into the soil may be able to stimulate suicidal germination of faraway parasites, and inhibit the germination of parasite seeds that are nearby (Joel 2000). To maximize the effectiveness of control strategies that target parasite seed germination, an understanding of how parasite seeds perceive SLs is necessary. The mechanism of SL perception in parasite seeds has only recently been explored (Conn et al. 2015, Toh et al. 2015, Tsuchiya et al. 2015), but SL signaling has been studied thoroughly in the model non-parasite *Arabidopsis thaliana*.

In Arabidopsis, the α/β-hydrolase D14 is a SL receptor (Hamiaux et al. 2012, Kagiyama et al. 2013, Nakamura et al. 2013, Zhao et al. 2013). SLs do not stimulate seed germination in Arabidopsis (Scaffidi et al. 2014), but *d14* mutant phenotypes indicate that they influence aspects of development such as inhibition of shoot branching and morphology of rosette leaves (Waters et al. 2012). The D14 homolog KAI2 is a receptor for KAR in Arabidopsis (Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Zhao et al. 2013). In addition to KAR insensitivity, *kai2* mutants have high seed dormancy, low light sensitivity as seedlings, and altered rosette morphology that is different from that of *d14* mutants. These phenotypes suggest that KAI2 may also perceive an endogenous regulator of germination and development (Waters et al. 2012), which is currently called KAI2 ligand (KL; Conn and Nelson 2016). Importantly, while *KAI2* is required for normal seed germination in Arabidopsis, *D14* is not (Waters et al. 2012).

We set out to elucidate the evolution and molecular mechanism of SL-responsive seed germination in parasitic weeds. Parasitic plants are unique because their seed germination (controlled by KAI2 in Arabidopsis) is responsive to SLs (perceived by D14 in Arabidopsis). Thus, we began our search for the SL receptor in parasite seeds by focusing on two candidate genes: *KAI2* and *D14*. We hypothesized that either A) *KAI2* controls seed germination in parasitic weeds, as it does in Arabidopsis, but its ligand preference has changed from KAR and KL to SL, or B) *D14* is SL-responsive in parasitic weeds, as it is in Arabidopsis, but it has evolved a role in seed germination. To address these hypotheses, we collected *KAI2* and *D14* sequences from parasitic plant species represented in sequence databases. We also used next-generation shotgun genome sequencing to obtain *KAI2* and *D14* sequence data for three additional parasitic weed species. We found that parasitic plants have extra copies of *KAI2* but not *D14*. Phylogeny-based selection tests revealed that some parasite *KAI2* paralogs are

relatively fast-evolving. As transgenes in an Arabidopsis *kai2* null mutant background, some fast-evolving parasite *KAI2* (divergent *KAI2*, or *KAI2d*) confer germination responses to SLs. Thus, they likely control host-responsive seed germination in parasitic weeds (Conn et al. 2015). Currently, the SL receptor D14 is thought to have originated from an ancient duplication of KAI2 (Waters et al. 2012). Therefore, the evolution of SL responsiveness in D14 and in parasite KAI2d is an interesting case of convergent evolution between independently duplicated genes.

Methods

Collection of *KAI2* and *D14* sequences.

Collection from existing databases. The AtKAI2 protein sequence was used as a TBLASTN query against the following databases: Phytozome (Goodstein et al. 2012), GenBank (Benson et al. 2005), the 1000 Plants Initiative (1KP; http://onekp.com/), the SOL Genomics Network (Fernandez-Pozo et al. 2015), the Parasitic Plant Genome Project (PPGP; Westwood et al. 2012), and the *Striga hermonthica* EST database (Yoshida et al. 2010). Complete and incomplete predicted coding sequences were included in the BLAST hits. In some cases, fragments of coding sequences could be assembled together using DNASTAR SeqMan or by eye. Hits with ambiguous orthology were reciprocally BLASTed against the Arabidopsis Information Resource (Berardini et al. 2015). Only those for which Arabidopsis *KAI2* or *D14* was the top match were retained for further analysis.

Based on *KAI2* and *D14* sequences from PPGP, primers were designed to amplify these genes from *Phelipanche aegyptiaca* and *Striga hermonthica*. *D14* was successfully amplified and sequenced for both species. Five *KAI2* orthologs were identified in *Phelipanche aegyptiaca*, and all were amplified and sequenced. Four of five

KAI2 orthologs identified from PPGP data in *Striga hermonthica* were amplified and sequenced (Table 2.1). Where PPGP and Sanger sequencing data disagreed, PPGP sequences were discarded and Sanger sequences were retained. Primer sequences are provided in Table 2.2.

Next-generation sequencing and assembly of KAI2 and D14. Five parasite species were chosen for next-generation sequencing (NGS) and assembly of genes of interest: the nonweedy obligate holoparasite *Conopholis americana*, which specializes on oak hosts (Baird and Riopel 1986); the weedy obligate parasite species *Orobanche cernua*, *O. cumana*, and *O. minor* (Parker 2013); and the facultative hemiparasite wildflower *Triphysaria versicolor* (Westwood et al. 2010). Parasitic plant tissue was collected from Callaway Gardens in Georgia (*C. americana*), from Israel (*O. cernua* and *O. cumana*), and from Virginia (*O. minor*). Seeds of *T.versicolor* were donated by John Yoder at the University of California, Davis. Tissue was harvested after *T. versicolor* germinated and was grown in a laboratory setting. The CTAB protocol (Doyle and Doyle 1990) was modified (http://www.cilr.uq.edu.au/UserImages/File/Plant%20Genomic%20DNA%20 Extraction%20by%20CTAB%20_2__Fiona.pdf) and used to extract genomic DNA (gDNA) from all five parasite species. DNA was fragmented to ~200 base pairs and used to generate a gDNA library, following the protocol of Schmitz et al. (2013). Paired-end (*O. cernua, O. cumana, O. minor*) or single-end (*C. americana*, *T. versicolor*) 101-base pair sequencing was done on an Illumina HiSeq2000.

NGS data were assembled with SOAPdenovo2 (Luo et al. 2012) with a k-mer value of 43. Assemblies were searched for *KAI2* and *D14* orthologs by TBLASTn with an Arabidopsis KAI2 query. A custom Perl script was developed to extend incomplete *KAI2* and *D14* contigs. This Perl script identifies reads in FASTQ files that overlap with the 5' and 3' ends of incomplete contigs. Contigs and overlapping reads were then assembled

using DNASTAR SeqMan NGen. Some sequences could not be extended with this custom Perl script, including some single exons that could not be joined together with intron sequence. Some of these were pieced together by PCR with primers matching the 5' end of the first exon and the 3' end of the second. These PCR products were then Sanger sequenced to verify coding sequences. Twenty-nine genes of interest that were identified in our NGS data were amplified and Sanger sequenced (Table 2.1). Sanger sequence data replaced NGS data if the two clearly disagreed on a particular gene sequence. All primer sequences used to amplify parasite *KAI2* and *D14* are listed in Table 2.2.

Finally, a gDNA assembly of the facultative hemiparasite *Phtheirospermum japonicum* was generated by Satoko Yoshida (RIKEN, Japan) as described in Conn et al. 2015. This assembly was mined for *KAI2* and *D14* orthologs, as described above. Without access to the raw data files for *P*. *japoncium*, the custom Perl script was not used to extend any sequences from this species.

Sequence alignment and filtering. *KAI2* and *D14* sequences were translated, aligned by eye, and trimmed at the ends to minimize gaps and regions where alignment was ambiguous. Sequences were discarded if they met any of the following three criteria: 1) less than 85% of the predicted protein sequence is represented, 2) frameshift mutations are present, and 3) amino acid replacements are present at catalytic triad sites. After sequences were filtered, 91% of *KAI2* sequences and 98% of *D14* sequences were predicted to be full-length coding sequences. The final *KAI2* alignment has 756 nucleotides, and the final *D14* alignment has 771 nucleotides (both published in Conn et al. 2015).

Multiple *KAI2* or *D14* sequences were found in the genomes of many plant species (Figure 2.1). To distinguish between alleles and paralogs, *KAI2* sequence

identity was calculated. Among species with completely sequenced genomes, the highest sequence identity between two *KAI2* paralogs is ~95% (*Glycine max* and *Glycine max 4*). Therefore, 95% was set as our sequence identity cutoff. Sequences from the same species were considered alleles instead of paralogs if they share >95% sequence identity and were obtained from incomplete assemblies. For each species, paralogous genes are distinguished by a numerical identifier (for example, the two *KAI2* paralogs *Agastache rugosa* and *Agastache rugosa 2*). Alleles share the same numerical identifier, if applicable, but are distinguished from one another by a letter suffix (for example, the alleles *Striga hermonthica KAI2d3* and *Striga hermonthica KAI2d3B*).

Phylogenetic analyses. *KAI2* sequences from the moss *Physcomitrella patens* were included in each alignment as an outgroup. The *KAI2* and *D14* alignments were each used to build a Bayesian phylogeny in MrBayes version 3.2.1 (Ronquist and Huelsenbeck 2003), under the General Time Reversible + gamma + invariance (GTR + Γ + I) model of evolution with two independent MCMC runs per tree. Trees were run for 20,000,000 generations, with sampling every 10,000 generations. The first 500,000 trees were discarded as burnin, and the remaining trees were used to build a consensus tree (Figure 2.2 and 2.3).

Phylogeny-based selection tests were implemented in Phylogenetic Analysis by Maximum Likelihood 4 (PAML4; Yang 2007) to investigate molecular evolution of *KAI2* and *D14*. Branch models were run to calculate the ratio of nonsynonymous changes per nonsynonymous site to synonymous mutations per synonymous site, or ω. ω values below 1 suggest purifying selection, while ω values greater than 1 are indicative of positive selection. To find the simplest branch model that is significantly the best fit to the data, likelihood ratio tests (LRTs) and the Akaike Information Criterion (AIC) were used.

Structural analysis. Rohan Bythell-Douglas (Imperial College London) generated homology models of parasite KAI2 proteins as described in Conn et al. 2015. He also provided calculations of ligand-binding pocket volumes. These models and measurements complement the molecular evolutionary analyses described above.

Four amino acid residues in KAI2 show interesting patterns of conservation and divergence. Three of these, sites 124, 157, and 194 in Arabidopsis KAI2, have been predicted to affect the ligand-binding pocket shape and/or interact with a ligand (Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Zhao et al. 2013). Patterns of amino acid conservation at these sites were illustrated using WebLogo (Crooks et al. 2004).

Functional analyses of *KAI2* in parasites.

Transformation of Arabidopsis kai2. D14 and *KAI2* genes were cloned from *A. thaliana*, *P. aegyptiaca*, and *S. hermonthica* into an entry vector. Clones were Sanger sequenced to confirm that no mutations were introduced during gene amplification. Correct sequences were then transferred to the destination vector pKAI2pro-GW, in which they were under the control of the *Arabidopsis KAI2* promoter. These constructs were transformed into *Agrobacterium tumefaciens* (strain C58C1), which was used to transform *Arabidopsis kai2* according to the floral dip protocol (or a modification of it) of Clough and Bent (1998). Dipped plants were harvested when siliques were brown, and their T0 seed was plated on 1/2X Murashige-Skoog media supplemented with hygromycin (15 – 25 µg/mL). The pKAI2pro-GW vector includes hygromycin resistance, so transformants (T1s) were identified as plants that developed normally on hygromycin. Single-insertion lines were identified by a 3:1 ratio of hygromycin-resistant to -sensitive plants in the T2 generation. Homozygous transgenic lines were identified in further generations and used in functional assays.

Germination assays. Homozygous transgenic plants and controls (wild type Arabidopsis in the L*er* background and *kai2-2* mutants) were potted in Fafard 3B Mix soil, randomized in flats, and grown in continuous light at approximately 22°C. Soil was supplemented with Gnatrol and Marathon to minimize insect damage. Plants were harvested in four lots per genotype when siliques were brown and were dried for 0 – 7 days. Seeds were then cleaned and either stored or used in germination assays.

For germination assays, seeds were sterilized with 70% ethanol, 0.05% Triton; rinsed with 70% and 95% ethanol; and dried in a sterile hood. Dry seeds were plated on 0.8% agar buffered with 2.6 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), pH ~5.7. Experimental plates were supplemented with KAR_1, KAR_2, *rac*-GR24, and individual GR24 stereoisomers dissolved in acetone. These treatments were diluted to a final concentration of 1 μM. Control media received an equal volume of acetone. Seed germination was defined as protrusion of the radicle through the endosperm. Germination was scored in all four seed lots per genotype.

Bryan Whittington and James Westwood (Virginia Tech) tested parasite seed in germination assays as described by Conn et al. 2015. In their assays, they included seed from *Phelipanche aegyptiaca* and *Striga hermonthica*, the two parasite species from which *KAI2* were tested as transgenes. Their results are not included here in a figure but are noted in the Results section.

Results

D14 and *KAI2* copy number vary among parasitic and non-parasitic plants. *D14* copy number is comparable in parasites and non-parasites; however, parasitic plants have a significantly higher *KAI2* copy number than non-parasites (Figure 2.1, Table 2.3). This observation led to the hypothesis that some *KAI2* in parasites may have neofunctionalized and evolved a role in host-responsive seed germination. In our

Bayesian phylogeny, *D14* from parasites do not form a monophyletic clade; instead, they are distributed among genes from non-parasitic relatives in the Lamiales. We found that ω – a rough measure of the rate of protein evolution – is higher in parasite *D14* than in orthologs from non-parasites ($\omega_P = 0.15$, $\omega_0 = 0.07$; Figure 2.2). *KAI2* from parasites group with other *KAI2* from non-parasitic Lamiids, and this large clade can be divided into several smaller clades with different rates of evolution. The conserved clade (*KAI2c*) has the slowest rate of protein evolution ($\omega_C = 0.07$) and includes *KAI2* from parasites and non-parasites. The paraphyletic intermediate clade (*KAI2i*) has an intermediate rate of protein evolution ($\omega_I = 0.10$) that is similar to that of *KAI2* outside the Lamiids ($\omega_0 = 0.11$). The divergent clade (*KAI2d*) is parasite-specific and contains the majority of *KAI2* paralogs per parasitic weed species. It has the fastest rate of molecular evolution ($\omega_D = 0.27$; Figure 2.3). Purifying selection is apparently relaxed on *D14* and *KAI2d* in parasites; however, we did not detect positive selection in any taxa when we analyzed gene sequences as a whole (Tables 2.4 – 2.7). Positive selection can act on individual codons within a gene sequence, and this possibility is addressed in Chapter 4.

The extensive duplication of *KAI2* in parasitic plants led to the hypothesis that some of these duplicates might have neofunctionalized and evolved a role in host-responsive seed germination. Structural models of KAI2 ligand-binding pockets support this hypothesis. The KAR receptor AtKAI2 has a smaller ligand-binding pocket than the SL receptor D14 (Bythell-Douglas et al. 2013, Kagiyama et al. 2013, Zhao et al. 2013). KAI2c proteins from parasites are similar to AtKAI2 in predicted ligand-binding pocket volume. KAI2d proteins from parasites vary widely in predicted ligand-binding pocket size; however, their volumes are often much larger than that of AtKAI2 and even OsD14. KAI2i are predicted to have ligand-binding pockets that are intermediate to known KAR and SL receptors (Conn et al. 2015). Thus, structural models suggest that KAI2 proteins from parasites may have evolved to accommodate different signals.

Four amino acid residues show interesting patterns of substitution in KAI2d in parasites. Three of these – sites 124, 157, and 194, in AtKAI2 – have been noted to affect the ligand-binding pocket (Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Zhao et al. 2013). In particular, site 124 is tyrosine (Y) in AtKAI2 and phenylalanine (F) in OsD14. Relative to the ligand-binding pocket of OsD14, that of AtKAI2 is essentially split into two smaller cavities by Y124. The F residue at the corresponding position in OsD14 does not have this effect (Bythell-Douglas et al. 2013). All KAI2c proteins in parasites or non-parasitic Lamiids have Y at this site, as does AtKAI2. In KAI2i proteins in parasites, this site is always F, as it is in the SL receptor Arabidopsis D14. In KAI2d proteins from parasites, the amino acid present at this site varies, and in many cases is a relatively small hydrophobic residue (Figure 2.4). In addition to predicted ligand-binding pocket volumes, these patterns of amino acid replacement suggest that different clades of parasite KAI2 proteins may respond to different signals.

We transformed *KAI2* from *Striga hermonthica* and *Phelipanche aegyptiaca* into Arabidopsis *kai2* null mutants and tested transgenic lines in seed germination. Wild type Arabidopsis (L*er* ecotype) had positive germination responses to two KAR structural variants (KAR_1 and KAR_2) and to the synthetic SL mix *rac*-GR24 (Figure 2.5A – B). The *kai2-2* mutant seed had high seed dormancy on control medium and no positive germination response to any of the chemical treatments we applied. We tested Arabidopsis *KAI2* in an Arabidopsis *kai2* null mutant background as a control, and found that it rescued the *kai2-2* mutant phenotypes of high seed dormancy and insensitivity to KAR and *rac*-GR24. As transgenes, parasite *KAI2* conferred different phenotypes. The conserved *KAI2* from *P. aegyptiaca* (*PaKAI2c*) restored a wild-type germination level on control medium but did not confer sensitivity to KAR or *rac*-GR24. This suggests that it may be responsive to the unidentified endogenous signal KL. In contrast, the conserved

KAI2 from *S. hermonthica* (*ShKAI2c*) had no noticeable effect on seed germination. Intermediate *KAI2* are not present in *P. aegyptiaca*, so we only tested one parasite *KAI2* from this clade. As a transgene, *ShKAI2i* conferred a positive germination response to KAR, which is surprising because parasitic weeds are not KAR-responsive in seed germination (Conn et al. 2015). Finally, we tested two *KAI2d* from *P. aegyptiaca* (*PaKAI2d1* and *PaKAI2d3*) and two from *S. hermonthica* (*ShKAI2d1* and *ShKAI2d2*). Three of these four *KAI2d* conferred germination responses to *rac*-GR24 (Figure 2.5B), suggesting that they may play a role in host-responsive seed germination.

Although *Arabidopsis KAI2* is not sensitive to natural SLs, it enables a positive germination response to *rac*-GR24. This is because *rac*-GR24 is a racemic mix; two components mimic natural SLs ($GR24^{5\text{-deoxystrigol}}$ and $GR24^{4\text{-deoxyorobanchol}}$), and the enantiomer of each of these components ($GR24^{\textit{ent}\text{-5-deoxystrigol}}$ and $GR24^{\textit{ent}\text{-4-deoxyorobanchol}}$) does not. Arabidopsis KAI2 responds to an enantiomer of a natural SL mimic, but not to the SL mimic itself (Scaffidi et al. 2014). Therefore, although parasite *KAI2d* transgenes conferred positive germination responses to *rac*-GR24, we could not immediately conclude that they are responsive to natural SLs. To resolve this question, we assayed transgenic lines and controls with purified stereoisomers of *rac*-GR24 (Figure 2.5A). As expected, wild-type Arabidopsis and the transgenic line carrying Arabidopsis *KAI2* had significant positive germination responses only to $GR24^{\textit{ent}\text{-5-deoxystrigol}}$. In contrast, each SL-responsive parasite *KAI2d* transgene conferred a positive germination response to the SL-mimicking components of *rac*-GR24. Interestingly, *ShKAI2d1* was responsive to all four components of *rac*-GR24, although it conferred stronger germination responses to natural SL mimics (Figure 2.5C). Overall, the results of our germination assay with purified stereoisomers of *rac*-GR24 were consistent with a role for parasite *KAI2d* in host-responsive seed germination.

Discussion

Strigolactone-responsive seed germination in parasitic weeds is a critical adaptation that helps to ensure that parasite seedlings find a host plant. Strigolactone was discovered decades ago (Cook et al. 1966), but the evolutionary and molecular mechanisms of host-responsive seed germination have remained unknown. By combining phylogenetic analyses, protein modeling, and cross-species complementation, we discovered that duplicate gene evolution likely produced this adaptation in parasitic weeds. Duplicate genes have been implicated in numerous cases of coevolution between hosts and parasites or pathogens. For example, duplication and diversification of mammalian defensins (Hughes and Yeager 1997) and plant disease resistance (R) genes (Huang et al. 2005) may influence – or be influenced by – host/pathogen coevolution. Have the hosts of parasitic weeds also experienced gene duplication and subsequent neofunctionalization? The answer to this question is unknown. However, host plants can produce different types and relative amounts of SL structural variants (Awad et al. 2006, Jamil et al. 2011, Xie et al. 2013), of which more than a dozen have been found in nature (Xie et al. 2010). The diversity in strigolactone profiles across different host plant species may be a result of selection for recognition by AMF and evasion of parasitic weeds. Supporting this hypothesis, we observed higher *KAI2* copy numbers in parasitic weeds, at least some of which tend to be generalists, than in non-weedy parasites, which can have narrower host ranges (Figure 2.6; Schneeweiss 2007). Perhaps different KAI2d paralogs have specialized to perceive different SL structural variants. This hypothesis remains untested.

Like gene duplication (and perhaps in conjunction with it), positive Darwinian selection is known to play a role in host/parasite and host/pathogen coevolution. For instance, positive selection has been linked to coevolution between the malaria-causing pathogen *Plasmodium falciparum* and primate hosts (Baum et al. 2003, Wang et al.

2003). We did not detect positive selection on any *D14* or *KAI2* genes from parasites. We did observe relaxed purifying selection on the parasite-specific *KAI2d* clade, in which SL-responsive *KAI2* are present. It is possible that positive selection has acted on specific codons in *KAI2d* genes in parasites. This possibility remains to be explored. Interestingly, we also observed relaxed purifying selection on *D14* in parasitic plants. However, we prioritized *KAI2* in protein modeling and cross-species complementation because of its extensive duplication in parasites. Future studies should elucidate the role of *D14* in parasitic plants.

Our new understanding of how parasitic plants perceive host-derived SLs will contribute to the development of more effective control strategies. First, it will improve efforts to combat parasitic weed infestations by inducing suicidal germination. Now that we know that KAI2d likely perceive host-derived SLs in parasite seeds, we can work to optimize crops that stimulate suicidal germination of parasite seeds. For example, we can now do germination assays to determine whether different KAI2d paralogs have different SL preferences. If this is the case, we may be able to choose non-host crops that produce a particular parasite's preferred SLs, thus maximizing suicidal germination in parasite seeds. Furthermore, we are now a step closer to doing *in vitro* screens for small molecules that activate *KAI2d* signaling. The identification of additional components of the SL signaling system in parasitic weeds will make such small molecule screens more effective.

The evolution of SL-sensitivity in *KAI2* duplicates has likely occurred outside of parasitic plants as well. The Arabidopsis SL receptor D14 likely arose from an ancient *KAI2* duplication in early land plants (Waters et al. 2012). Green algae and the moss *Physcomitrella patens* may have diverged prior to the *KAI2* duplication that produced *D14* (but see Bythell-Douglas et al. 2017), but the charophyte *Chara corallina* and *P. patens* are SL-responsive (Delaux et al. 2012, Hoffmann et al. 2014). Extra *KAI2* copies

are present in *P. patens*, some of which respond transcriptionally to SL treatment (Lopez-Obando et al. 2016). Thus, the evolution of SL sensitivity in *KAI2d* in parasites, in some *KAI2* paralogs in moss, and in *D14* in vascular plants represents multiple cases of convergent molecular evolution in *KAI2* duplicates (Figure 2.7; moss not indicated here). Whether SL-responsiveness has evolved in *KAI2* paralogs in other plant taxa remains to be determined.

Overall, this study illuminates a fascinating case of gene duplication, neofunctionalization, and convergent evolution. These processes are especially interesting in the context of a host/parasite system, in which both organisms must often continually evolve to overcome the other's adaptations (Van Valen 1973). Our work also lays the foundation for the development of more effective control methods for parasitic weed infestations. Although questions remain about additional components of the SL signaling system in parasitic weeds, we have identified a set of paralogous genes that are likely to play a central role in host-responsive seed germination.

Table 2.1. *D14* and *KAI2* sequences used in this study. Modified from Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

Gene	Order	Sequence name in phylogeny[1]	Clade[2]	Original ID	Source[3]	Verif.[4]	Family
D14	Brassicales	*Arabidopsis lyrata*		Arly477656	Phytozome		Brassicaceae
		Arabidopsis thaliana		Arth3G039901	TAIR		Brassicaceae
		Brassica rapa		BrraBra036416	Phytozome		Brassicaceae
		Brassica rapa 2		BrraBra031957	Phytozome		Brassicaceae
		Capsella rubella		Carubv10014401m	Phytozome		Brassicaceae
		Eutrema salsugineum		Thhalv10021292m	Phytozome		Brassicaceae
		Carica papaya		Capaevmmodelsupercontig13233	Phytozome		Cariacaceae
	Cucurbitales	*Cucumis sativus*		Cucsa3956701	Phytozome		Cucurbitaceae
	Fabales	*Glycine max*		Glyma17g353601	Phytozome		Fabaceae
		Glycine max 2		Glyma0092s002401	Phytozome		Fabaceae
		Medicago truncatula		Medtr1g0233801	Phytozome		Fabaceae
		Phaseolus vulgaris		Phvulv091011045m	Phytozome		Fabaceae
	non-parasitic Lamiales	*Byblis gigantea*		scaffold-GDZS-2012582-2010455-Byblis_gigantea	1KP		Byblidaceae
		Agastache rugosa		scaffold-PUCW-2010996-Agastache_rugosa	1KP		Lamiaceae
		Ajuga reptans		scaffold-UCNM-2056057-Ajuga_reptans	1KP		Lamiaceae
		Marrubium vulgare		scaffold-EAAA-2007328-Marrubium_vulgare	1KP		Lamiaceae
		Melissa officinalis		scaffold-TAGM-2047258-Melissa_officinalis	1KP		Lamiaceae
		Micromeria fruticosa		scaffold-WHNV-2043304-Micromeria_fruticosa	1KP		Lamiaceae
		Nepeta cataria		scaffold-FUMQ-2017076-Nepeta_cataria	1KP		Lamiaceae
		Oxera neriifolia		scaffold-GNPX-2005621-Oxera_neriifolia	1KP		Lamiaceae
		Oxera pulchella		scaffold-RTNA-2020810-Oxera_pulchella	1KP		Lamiaceae
		Pogostemon sp.		scaffold-GETL-2013627-Pogostemon_sp.	1KP		Lamiaceae
		Prunella vulgaris		scaffold-PHCE-2004407-Prunella_vulgaris	1KP		Lamiaceae
		Salvia sp.		scaffold-EQDA-2057038-Salvia_spp.	1KP		Lamiaceae
		Thymus vulgaris		scaffold-IYDF-2003375-Thymus_vulgaris	1KP		Lamiaceae
		Paulownia fargesii		scaffold-UMUL-2013773-Paulownia_fargesii	1KP		Paulowniaceae
		Mimulus guttatus		Migumgv1a027102m	Phytozome		Phrymaceae
		Vitex agnus-castus		scaffold-DMLT-2007660-Vitex_agnus_castus	1KP		Verbenaceae
	Orobanchaceae (including parasitic Lamiales)	*Conopholis americana*			Conn et al.		Orobanchaceae
		Lindenbergia philippensis		scaffold-WUZV-2006074-Lindenbergia_philippensis-flower	1KP		Orobanchaceae
		Orobanche cernua			Conn et al.	✓	Orobanchaceae
		Orobanche cumana			Conn et al.	✓	Orobanchaceae
		Orobanche fasciculata		scaffold-PHOQ-2000177-Orobanche_fasciculata-stem	1KP		Orobanchaceae
		Orobanche minor			Conn et al.	✓	Orobanchaceae
		Phelipanche aegyptiaca			PPGP, Conn et al.	✓	Orobanchaceae
		Ptheirospermum japonicum			Conn et al.		Orobanchaceae
		Striga hermonthica			PPGP, Conn et al.	✓	Orobanchaceae
		Triphysaria versicolor			PPGP		Orobanchaceae
	Malpighiales	*Manihot esculenta*		Maescassava41013999m	Phytozome		Euphorbiaceae
		Ricinus communis		Rico30170m013744	Phytozome		Euphorbiaceae
		Populus trichocarpa		POPTR0002s119701	Phytozome		Salicaceae
		Populus trichocarpa 2		POPTR0014s016801	Phytozome		Salicaceae
	Myrtales	*Eucalyptus grandis*		EugrEucgrF004781	Phytozome		Myrtaceae

		Eucalyptus grandis 2		EugrEucgrF022541	Phytozome		Myrtaceae
	Ranunculales	*Aquilegia caerulea*		Aquca042001931	Phytozome		Ranunculaceae
	Rosales	*Malus domestica*		MadoMDP0000888050	Phytozome		Rosaceae
		Prunus persica		Prpeppa010005m	Phytozome		Rosaceae
	Sapindales	*Citrus clementina*		Ciclclementine09018899m	Phytozome		Rutaceae
		Citrus clementina 2		Ciclclementine09018310m	Phytozome		Rutaceae
		Citrus clementina 3		Ciclclementine09030520m	Phytozome		Rutaceae
		Citrus sinensis		Cisiorange11g024681m	Phytozome		Rutaceae
		Citrus sinensis 2		Cisiorange11g045774m	Phytozome		Rutaceae
	Solanales	*Solanum lycopersicum*		TomatoSGNU572523	SGN		Solanaceae
KAI2	Brassicales	*Arabidopsis lyrata*		Arly944104	Phytozome		Brassicaceae
		Arabidopsis thaliana		Arth4G374701	TAIR		Brassicaceae
		Brassica rapa		BrraBra010600	Phytozome		Brassicaceae
		Brassica rapa 2		BrraBra017826	Phytozome		Brassicaceae
		Brassica rapa 3		BrraBra004765	Phytozome		Brassicaceae
		Capsella rubella		Carubv10005485m	Phytozome		Brassicaceae
		Eutrema salsugineum		Thhalv10025969m	Phytozome		Brassicaceae
		Carica papaya		Capaevmmodelsupercontig657	Phytozome		Cariacaceae
	Cucurbitales	*Cucumis sativus*		Cucsa3671301	Phytozome		Cucurbitaceae
		Cucumis sativus 2		Cucsa1029001	Phytozome		Cucurbitaceae
	Fabales	*Glycine max*		Glyma01g398501	Phytozome		Fabaceae
		Glycine max 2		Glyma05g219301	Phytozome		Fabaceae
		Glycine max 3		Glyma17g179101	Phytozome		Fabaceae
		Glycine max 4		Glyma11g054401	Phytozome		Fabaceae
		Glycine max 5		Glyma17g178901	Phytozome		Fabaceae
		Lotus japonicus		BT134978.1	NCBI		Fabaceae
		Medicago truncatula		Medtr8g0991901	Phytozome		Fabaceae
		Medicago truncatula 2		Medtr5g0161401	Phytozome		Fabaceae
		Phaseolus vulgaris		Phvulv091016531m	Phytozome		Fabaceae
		Phaseolus vulgaris 2		Phvulv091024520m	Phytozome		Fabaceae
		Phaseolus vulgaris 3		Phvulv091007379m	Phytozome		Fabaceae
	Funariales	*Physcomitrella patens*		Phpa1s8817V64	Phytozome		Funariaceae
		Physcomitrella patens 2		Phpa1s11314V63	Phytozome		Funariaceae
		Physcomitrella patens 3		Phpa1s11627V63	Phytozome		Funariaceae
		Physcomitrella patens 4		Phpa1s150116V61	Phytozome		Funariaceae
		Physcomitrella patens 5		Phpa1s201138V61	Phytozome		Funariaceae
		Physcomitrella patens 6		Phpa1s4312V61	Phytozome		Funariaceae
		Physcomitrella patens 7		Phpa1s74232V61	Phytozome		Funariaceae
		Physcomitrella patens 8		Phpa1s91156V61	Phytozome		Funariaceae
		Physcomitrella patens 9		Phpa1s95125V61	Phytozome		Funariaceae
		Physcomitrella patens 10		Phpa1s18847V61	Phytozome		Funariaceae
	Gentianales	*Coffea arabica*	C	CoarSGNU609432	Phytozome		Rubiaceae
	non-parasitic Lamiales	*Byblis gigantea*	C	scaffold-GDZS-2114812-Byblis_gigantea	1KP		Byblidaceae
		Agastache rugosa	C	scaffold-PUCW-2003367-Agastache_rugosa	1KP		Lamiaceae
		Agastache rugosa 2	I	scaffold-PUCW-2000958-Agastache_rugosa	1KP		Lamiaceae
		Ajuga reptans	C	scaffold-UCNM-2054960-Ajuga_reptans	1KP		Lamiaceae
		Ajuga reptans 2	I	scaffold-UCNM-2055996-Ajuga_reptans	1KP		Lamiaceae
		Marrubium vulgare	I	scaffold-EAAA-2061153-Marrubium_vulgare	1KP		Lamiaceae
		Melissa officinalis	C	scaffold-TAGM-2045102-Melissa_officinalis	1KP		Lamiaceae
		Micromeria fruticosa	C	scaffold-WHNV-2042236-Micromeria_fruticosa	1KP		Lamiaceae
		Oxera neriifolia	C	scaffold-GNPX-2014520-Oxera_neriifolia	1KP		Lamiaceae
		Oxera neriifolia 2	I	scaffold-GNPX-2010479-Oxera_neriifolia	1KP		Lamiaceae
		Oxera neriifolia 3	I	scaffold-GNPX-2022199-Oxera_neriifolia	1KP		Lamiaceae
		Oxera pulchella	C	scaffold-RTNA-2088906-Oxera_pulchella	1KP		Lamiaceae

		Oxera pulchella 2	I	scaffold-RTNA-2089199-Oxera_pulchella	1KP		Lamiaceae
		Pogostemon sp.	I	scaffold-GETL-2020484-Pogostemon_sp.	1KP		Lamiaceae
		Prunella vulgaris	C	scaffold-PHCE-2069891-Prunella_vulgaris	1KP		Lamiaceae
		Prunella vulgaris 2	I	scaffold-PHCE-2069246-Prunella_vulgaris	1KP		Lamiaceae
		Rosmarinus officinalis	C	scaffold-FDMM-2039813-Rosmarinus_officinalis, scaffold-FDMM-2011341-Rosmarinus_officinalis	1KP		Lamiaceae
		Salvia sp.	C	scaffold-EQDA-2004006-Salvia_spp.	1KP		Lamiaceae
		Scutellaria montana	C	scaffold-ATYL-2122589-Scutellaria_montana	1KP		Lamiaceae
		Solenostemon scutellarioides	C	scaffold-BAHE-2035595-Solenostemon_scutellarioides	1KP		Lamiaceae
		Thymus vulgaris	C	scaffold-IYDF-2076185-Thymus_vulgaris	1KP		Lamiaceae
		Thymus vulgaris 2	I	scaffold-IYDF-2016771-Thymus_vulgaris	1KP		Lamiaceae
		Paulownia fargesii	C	scaffold-UMUL-2014265-Paulownia_fargesii	1KP		Paulowniaceae
		Paulownia fargesii 2	I	scaffold-UMUL-2006662-Paulownia_fargesii	1KP		Paulowniaceae
		Mimulus guttatus	C	Migumgv1a011843m	Phytozome		Phrymaceae
		Mimulus guttatus 2	C	Migumgv1a011794m	Phytozome		Phrymaceae
		Mimulus guttatus 3	I	Migumgv1a008430m	Phytozome		Phrymaceae
		Vitex agnus-castus	C	scaffold-DMLT-2101555-Vitex_agnus_castus	1KP		Verbenaceae
		Vitex agnus-castus 2	I	scaffold-DMLT-2101140-Vitex_agnus_castus	1KP		Verbenaceae
		Vitex agnus-castus 3	I	scaffold-DMLT-2018459-Vitex_agnus_castus	1KP		Verbenaceae
	Orobanchaceae (including parasitic Lamiales)	*Conopholis americana KAI2c*	C	scaffold-FAMO-2094892-Conopholis_americana	1KP		Orobanchaceae
		Lindenbergia philippensis KAI2c	C	LiPhGnB1_98152	1KP		Orobanchaceae
		Lindenbergia philippensis KAI2i	I	LiPhGnB1_81152	1KP		Orobanchaceae
		Orobanche cernua KAI2c	C		Conn et al.	✓	Orobanchaceae
		Orobanche cernua KAI2d1	D		Conn et al.	✓	Orobanchaceae
		Orobanche cernua KAI2d2	D		Conn et al.	✓	Orobanchaceae
		Orobanche cernua KAI2d3	D		Conn et al.		Orobanchaceae
		Orobanche cumana KAI2c	C		Conn et al.	✓	Orobanchaceae
		Orobanche cumana KAI2d1	D		Conn et al.	✓	Orobanchaceae
		Orobanche cumana KAI2d2	D		Conn et al.	✓	Orobanchaceae
		Orobanche cumana KAI2d3	D		Conn et al.	✓	Orobanchaceae
		Orobanche cumana KAI2d4	D		Conn et al.	✓	Orobanchaceae
		Orobanche cumana KAI2d5	D		Conn et al.	✓	Orobanchaceae
		Orobanche cumana KAI2d6	D		Conn et al.		Orobanchaceae
		Orobanche fasciculata KAI2c	C	scaffold-VTOK-2017963-Orobanche_fasciculata-flower_buds	1KP		Orobanchaceae
		Orobanche fasciculata KAI2d	D	scaffold-PHOQ-2095967-Orobanche_fasciculata-stem	1KP		Orobanchaceae
		Orobanche minor KAI2c	C		Conn et al.	✓	Orobanchaceae
		Orobanche minor KAI2d1	D		Conn et al.		Orobanchaceae
		Orobanche minor KAI2d2	D		Conn et al.	✓	Orobanchaceae
		Orobanche minor KAI2d3	D		Conn et al.	✓	Orobanchaceae
		Orobanche minor KAI2d4	D		Conn et al.	✓	Orobanchaceae
		Orobanche minor KAI2d5	D		Conn et al.	✓	Orobanchaceae
		Phelipanche aegyptiaca KAI2c	C		PPGP, Conn et al.	✓	Orobanchaceae
		Phelipanche	D		PPGP,	✓	Orobanchaceae

		aegyptiaca KAI2d1			Conn et al.		
		Phelipanche aegyptiaca KAI2d2	D		PPGP, Conn et al.	✓	Orobanchaceae
		Phelipanche aegyptiaca KAI2d3	D		PPGP, Conn et al.	✓	Orobanchaceae
		Phelipanche aegyptiaca KAI2d4	D		PPGP, Conn et al.	✓	Orobanchaceae
		Ptheirospermum japonicum KAI2c	C		Conn et al.		Orobanchaceae
		Ptheirospermum japonicum KAI2i	I		Conn et al.		Orobanchaceae
		Ptheirospermum japonicum KAI2d1	D		Conn et al.		Orobanchaceae
		Ptheirospermum japonicum KAI2d2	D		Conn et al.		Orobanchaceae
		Ptheirospermum japonicum KAI2d3	D		Conn et al.		Orobanchaceae
		Ptheirospermum japonicum KAI2d4	D		Conn et al.		Orobanchaceae
		Ptheirospermum japonicum KAI2d5	D		Conn et al.		Orobanchaceae
		Striga hermonthica KAI2c	C		PPGP, Conn et al.	✓	Orobanchaceae
		Striga hermonthica KAI2cB	C	ShContig8304	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2i	I		PPGP, Conn et al.	✓	Orobanchaceae
		Striga hermonthica KAI2iB	I	ShContig7528	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d1	D		PPGP, Conn et al.	✓	Orobanchaceae
		Striga hermonthica KAI2d2	D		PPGP, Conn et al.	✓	Orobanchaceae
		Striga hermonthica KAI2d3	D		PPGP		Orobanchaceae
		Striga hermonthica KAI2d3B	D	ShContig8325	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d4	D	ShContig9933	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d5	D	ShContig8521	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d6	D	ShContig7135	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d7	D	ShContig9645	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d8	D	ShContig3725	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d9	D	ShContig176	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d10	D	ShContig1509	*Sh*EST		Orobanchaceae
		Striga hermonthica KAI2d11	D	ShContig9269	*Sh*EST		Orobanchaceae
		Triphysaria versicolor KAI2c	C		Conn et al.	✓	Orobanchaceae
		Triphysaria versicolor KAI2i	I		PPGP, Conn et al.		Orobanchaceae
		Triphysaria versicolor KAI2d1	D		PPGP, Conn et al.		Orobanchaceae
		Triphysaria versicolor KAI2d2	D		PPGP, Conn et al.		Orobanchaceae
		Triphysaria versicolor KAI2d3	D		PPGP, Conn et al.		Orobanchaceae
	Malpighiales	*Manihot esculenta*		Maescassava41013852m	Phytozome		Euphorbiaceae
		Manihot esculenta 2		Maescassava41028107m	Phytozome		Euphorbiaceae
		Manihot esculenta 3		Maescassava41013869m	Phytozome		Euphorbiaceae
		Manihot esculenta 4		Maescassava41033950m	Phytozome		Euphorbiaceae
		Manihot esculenta 5		Maescassava41032611m	Phytozome		Euphorbiaceae
		Manihot esculenta 6		Maescassava41023769m	Phytozome		Euphorbiaceae
		Ricinus communis		Rico29970m001027	Phytozome		Euphorbiaceae
		Ricinus communis 2		Rico29970m001033	Phytozome		Euphorbiaceae
		Ricinus communis 3		Rico29579m000205	Phytozome		Euphorbiaceae
		Populus trichocarpa		POPTR0007s102001	Phytozome		Salicaceae
		Populus trichocarpa 2		POPTR0005s188501	Phytozome		Salicaceae
	Ranunculales	*Aquilegia caerulea*		Aquca042000871	Phytozome		Ranunculaceae
		Aquilegia caerulea 2		Aquca04200861	Phytozome		Ranunculaceae
		Aquilegia caerulea 3		Aquca042000851	Phytozome		Ranunculaceae
	Rosales	*Malus domestica*		MadoMDP0000127844	Phytozome		Rosaceae
		Prunus persica		Prpeppa009957m	Phytozome		Rosaceae
	Sapindales	*Citrus clementina*		Ciclclementine09018593m	Phytozome		Rutaceae
		Citrus sinensis		Cisiorange11g046596m	Phytozome		Rutaceae
	Solanales	*Nicotiana benthamiana*	C	NibeSGNU513314	SGN		Solanaceae

		Nicotiana tabacum	C	NitaSGNU448341	SGN		Solanaceae
		Nicotiana tabacum 2	C	NitaSGNU447980	SGN		Solanaceae
		Nicotiana tabacum 3	I	NitaSGNU444704	SGN		Solanaceae
		Solanum lycopersicum	C	TomatoSGNU573177	SGN		Solanaceae
		Solanum lycopersicum 2	C	TomatoSGNU588241	SGN		Solanaceae
		Solanum lycopersicum 3	I	TomatoSGNU573176	SGN		Solanaceae
		Solanum tuberosum	C	SotuSGNU276257	SGN		Solanaceae
	Vitales	*Vitis vinifera*		ViviGSVIVT01000162001	Phytozome		Vitaceae

[1]Putative alleles are designated by a "B" suffix. [2]Lamiid *KAI2* only; C, conserved; I, intermediate; D, divergent. [3]Conn et al., Conn et al. 2015; 1KP, 1000 Plants Initiative, (onekp.com); Phytozome (Goodstein et al. 2012); NCBI, National Center for Biotechnology Information (Benson et al. 2005); SGN, Sol Genomics Network (Fernandez-Pozo et al. 2015); PPGP, Parasitic Plant Genome Project, (Westwood et al. 2012); ShEST, *Striga hermonthica* EST database (Yoshida et al. 2010); TAIR, The Arabidopsis Information Resource (Berardini et al. 2015). 1KP and PPGP data are from *de novo* transcriptome assemblies. Phytozome, SGN, and TAIR data are from assembled genomes. NCBI entry for *L. japonicus* is an mRNA sequence. [4]Coding sequence verified by Sanger sequencing.

Table 2.2. Primers used to amplify parasite *D14* and *KAI2* genes. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

Gene name	GW[1]	forward primer	reverse primer
Orobanche cernua D14		CGTTTACCGATAATCACCCATC	GAAAACTCGAACTCCACAATTATCC
Orobanche cernua KAI2c		GACCATTGTCCTTGCCCACG	GAGCTCAGTTGCGGAAGGTG
Orobanche cernua KAI2d1		GCGATCTATACTCCAACGTTCTATTC	CGACGTAAGCTACAGTCGTAACAC
Orobanche cernua KAI2d2		CCCAGAATTCTACTCAAAACAAATC	GACTAAACATCGATCATTAACGGTG
Orobanche cernua KAI2d3		CCGTCACAACGTCCGAGTCC	GAGCAGCACCGGAATCGTGAC
Orobanche cumana D14		CAGACTTGACTTCAGCCACAGG	CTCGAACTCCACAATTATCCCC
Orobanche cumana KAI2c		GACCATTGTCCTTGCCCACG	GAGCTCAGTTGCGGAAGGTG
Orobanche cumana KAI2d1		GTCACAACGTCCGAGTCGTGG	GAGCAGCACCGGAATCGTGC
Orobanche cumana KAI2d2		CTATTTAAAGCGATATAGACTCC	GACTAAACATTTTCTCACAATCG
Orobanche cumana KAI2d3		CTCCAAYGTTCTATTCAAACAAATCC	GAAATAACACAAAACCCTCCCG
Orobanche cumana KAI2d4		GAGTAACACAAGTAGACAGACTGG	CAATACAACACAAACCCTCCC
Orobanche cumana KAI2d5		CTAACTCAAAACGAATCCTGAACC	CGTCAAAGAGACTAAACATCGATCAC
Orobanche minor D14		CGGTTAAATAATTGGTAAATGAC	CCACGTGGCCATCGACTCTTC
Orobanche minor KAI2c		GTCTCCACGTAATTCCCATCC	CATTTCATACAGGTGCAAGATTTCG
Orobanche minor KAI2d1		CATGAACCGTATAGTTGGACTTGGAC	GGAGAAGCACCGGAATCGTG
Orobanche minor KAI2d2		GAGCTCCACAAATTCGGTTCATAAG	GTTTAGTATTACAACAAACCCTTCCG
Orobanche minor KAI2d3		CCAGTGCATCGTACGGGTAAG	GTAGTAGACGCTCTTGCCGCACAC
Orobanche minor KAI2d4		GCACACATGCATTGAGAGGAG	GTTTTCTCATAGTCTCACCATTTTC
Orobanche minor KAI2d5		CATGGGTAGCATAGTTGGTGC	TCTCAACCATCAACGATATCG
Phelipanche aegyptiaca D14	✓	ATGGGGCACAGACTG	TCACGGTGGTAACGC
Phelipanche aegyptiaca KAI2c	✓	ATGGGAATCGCCCAAG	TCAGGCAGCGATATT
Phelipanche aegyptiaca KAI2d1	✓	ATGAGCCCATTAGGAG	TCAGGCATCAGCAAT
Phelipanche aegyptiaca KAI2d2	✓	ATGAGCTCAGTTGGAC	TTAGGCATCAGCAAT
Phelipanche aegyptiaca KAI2d3	✓	ATGAACATTAACAGAG	TCAATTAGCATCTGC
Phelipanche aegyptiaca KAI2d4	✓	ATGACCACAATTGGA	TCAAACAATATCGTG
Striga hermonthica D14	✓	ATGGTGCAGAGTCTT	TCAGCGGGGCAAGGC
Striga hermonthica KAI2c	✓	ATGGGTCTCGCCCAAG	TCAATTTGGGCGTGC
Striga hermonthica KAI2d1	✓	ATGGGCACAGTCGGAG	TCAGTTATCCACGAT
Striga hermonthica KAI2d2	✓	ATGAACAAAGTTGGA	TTAGATGTCCTGCCG
Striga hermonthica KAI2d3	✓	ATGAGCACAGTCGGG	TCAATCATCTGCTAT
Striga hermonthica KAI2i	✓	ATGAATAGAGTGGAG	TCAAAACCTCCGGCC
Triphysaria versicolor KAI2c		CATCGGATCCGGGCTACAAACC	CATTGATGTGGTCTATCTAATACAG

[1]Gateway attB1 and attB2 sequences were appended to the 5' ends of these primers for cloning.

Table 2.3. Average *D14* and *KAI2* copy number in parasitic plants, related non-parasites, and other dicots. Only species in which at least one *KAI2* and at least one *D14* were found are included. Copy number is the mean ± standard error of the mean for each taxonomic group. N is the number of species included in each count. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

Species class	*D14*	*KAI2*	N
parasitic Orobanchaceae	1.0 ± 0 (max. 1)	5.6 ± 1.2 (max. 13)	9
non-parasitic Lamiales	1.0 ± 0 (max. 1)	1.8 ± 0.19 (max. 3)	16
other dicots	1.3 ± 0.13 (max. 3)	2.2 ± 0.34 (max. 6)	19

Table 2.4. Branch models assessing the evolution of *D14*. *p*, number of parameters in the model; lnL, loglikelihood; κ, transition: transversion ratio; ω, dN/dS ratios for clades indicated in Figure 2.2. The two ω model was the simplest model that provided the best fit to the data ($p<0.01$; see Tables 2.6 and 2.7). From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

Model	Hypothesis	*p*	lnL	κ	ω_0	ω_P	ω_N
One ω	$\omega_0=\omega_P=\omega_N$	121	-21328.06	1.80	0.080	$=\omega_0$	$=\omega_0$
Two ω	$\omega_0\neq\omega_P$	122	-21310.40	1.80	0.073	0.146	$=\omega_0$
Three ω	$\omega_0\neq\omega_P\neq\omega_N$	123	-21310.40	1.80	0.073	0.146	0.073

Table 2.5. Branch models assessing the evolution of *KAI2*. *p*, number of parameters in the model; lnL, loglikelihood; κ, transition: transversion ratio; ω, dN/dS ratios for clades indicated in Figure 2.3. The four ω model was the simplest model that provided the best fit to the data ($p<0.01$; see Tables 2.6 and 2.7). From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

Model	Hypothesis	*p*	lnL	κ	ω_0	ω_C	ω_I	ω_D
One ω	$\omega_0=\omega_C=\omega_I=\omega_D$	279	-48570.61	1.84	0.128	$=\omega_0$	$=\omega_0$	$=\omega_0$
Two ω	$\omega_0\neq\omega_D$	280	-48330.77	1.83	0.094	$=\omega_0$	$=\omega_0$	0.267
Four ω	$\omega_0\neq\omega_C\neq\omega_I\neq\omega_D$	282	-48307.91	1.83	0.107	0.070	0.100	0.267

Table 2.6. Likelihood ratio tests of branch models. df, degrees of freedom; 2ΔlnL, likelihood ratio test statistic. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

Gene	Model	Null model	Df	2ΔlnL	P
D14	Two ω: $\omega_0 \neq \omega_P$	One ω: $\omega_0 = \omega_P = \omega_N$	1	35.32	2.84×10^{-9}
D14	Three ω: $\omega_0 \neq \omega_P \neq \omega_N$	Two ω: $\omega_0 \neq \omega_P$	1	0	1
KAI2	Two ω: $\omega_0 \neq \omega_D$	One ω: $\omega_0 = \omega_C = \omega_I = \omega_D$	1	479.68	<0.0001
KAI2	Four ω: $\omega_0 \neq \omega_C \neq \omega_I \neq \omega_D$	Two ω: $\omega_0 \neq \omega_D$	2	45.72	<0.0001

Table 2.7. Akaike Information Criterion tests of branch models. *p*, number of parameters in the model; lnL, log likelihood of model; AIC, Akaike Information Criterion score for model; ΔAIC, difference between AIC for model and lowest AIC for gene; Prob (model), model probability. For each gene, the sum of model probabilities may deviate slightly from 1 due to rounding. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

Gene	Model	Hypothesis	*p*	lnL	AIC	ΔAIC	Prob (model)
D14	One ω	$\omega_0=\omega_P=\omega_N$	121	-21328.06	42898.12	33.32	$4.25\text{x}10^{-8}$
D14	Two ω	$\omega_0\neq\omega_P$	122	-21310.40	42864.80	0	0.73
D14	Three ω	$\omega_0\neq\omega_P\neq\omega_N$	123	-21310.40	42866.80	2	0.27
KAI2	One ω	$\omega_0=\omega_C=\omega_I=\omega_D$	279	-48570.61	97699.22	519.4	$1.64\text{x}10^{-113}$
KAI2	Two ω	$\omega_0\neq\omega_D$	280	-48330.77	97221.54	41.72	$8.72\text{x}10^{-10}$
KAI2	Four ω	$\omega_0\neq\omega_C\neq\omega_I\neq\omega_D$	282	-48307.91	97179.82	0	1

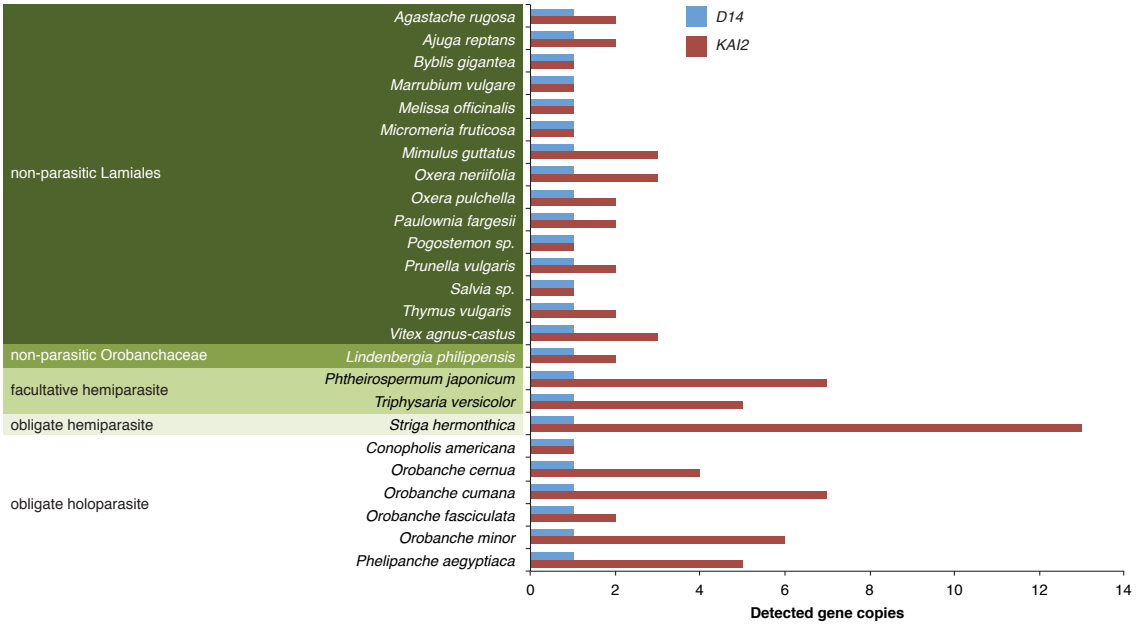

Figure 2.1. *KAI2* duplication in the Lamiales. *D14* (blue) and *KAI2* (red) copy number in parasitic plants and related non-parasites are shown. Only species in which at least one *KAI2* and at least one *D14* were found are included. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

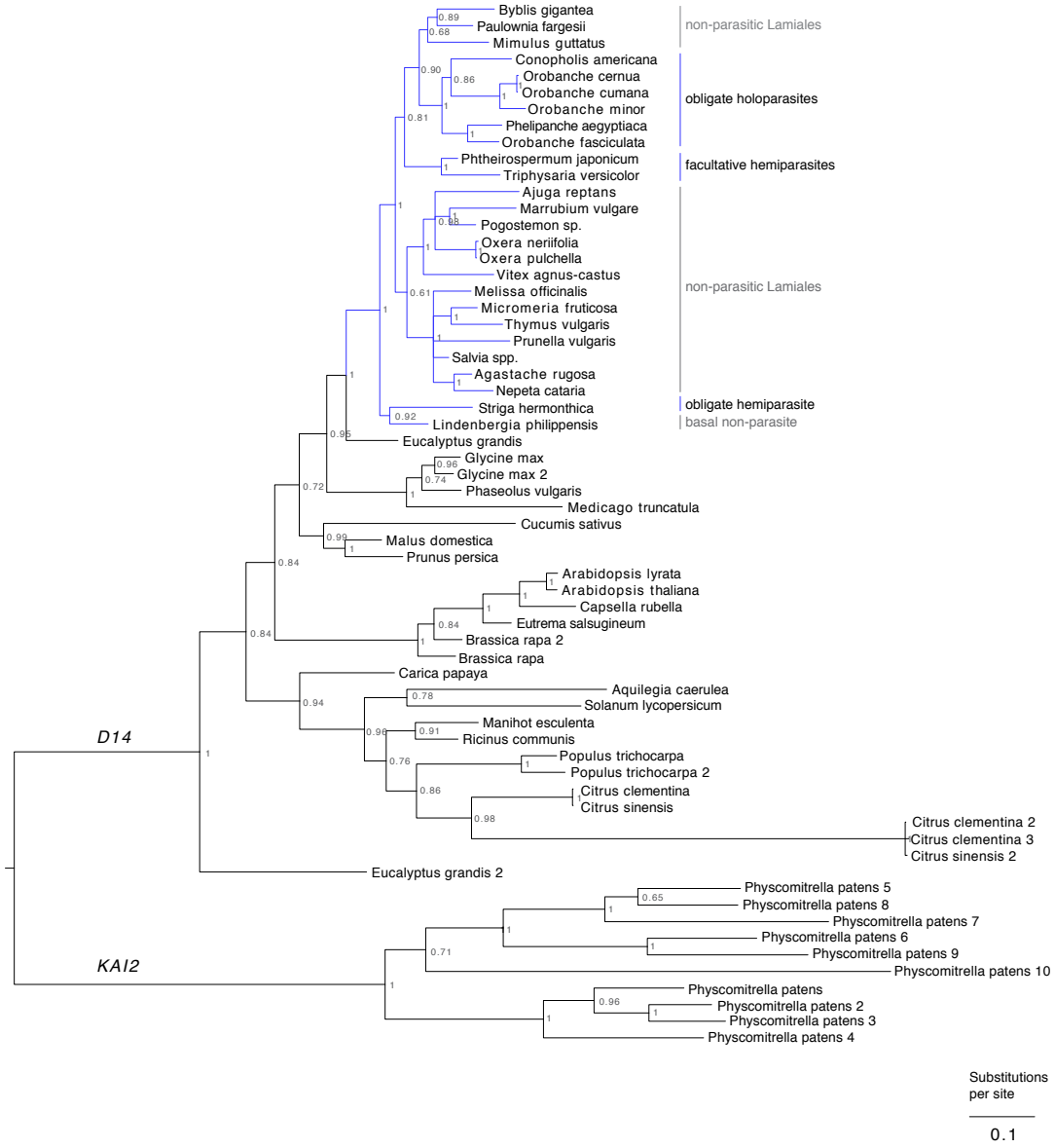

Figure 2.2. Bayesian phylogeny of *D14* in dicots. Nucleotide sequences were used to reconstruct the evolutionary history of *D14*. *KAI2* from the moss *Physcomitrella patens* were included as an outgroup because *D14* likely arose via an ancient *KAI2* duplication. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

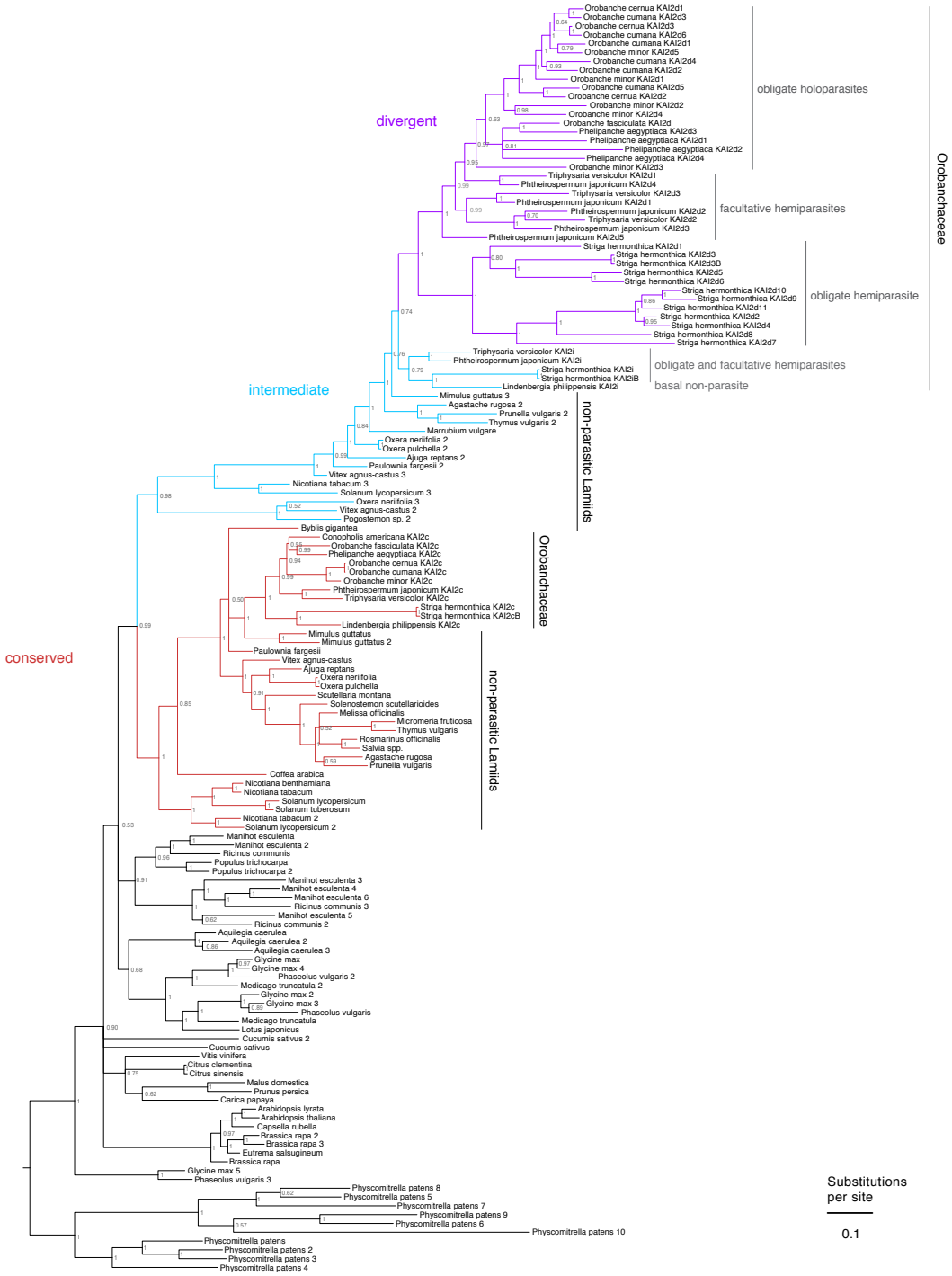

Figure 2.3. Bayesian phylogeny of *KAI2* in dicots. Nucleotide sequences were used to reconstruct the evolutionary history of *KAI2*. Sequences from *Physcomitrella patens* were included as an outgroup because all other taxa in the tree are dicots. Conserved, intermediate, and divergent clades are labeled. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

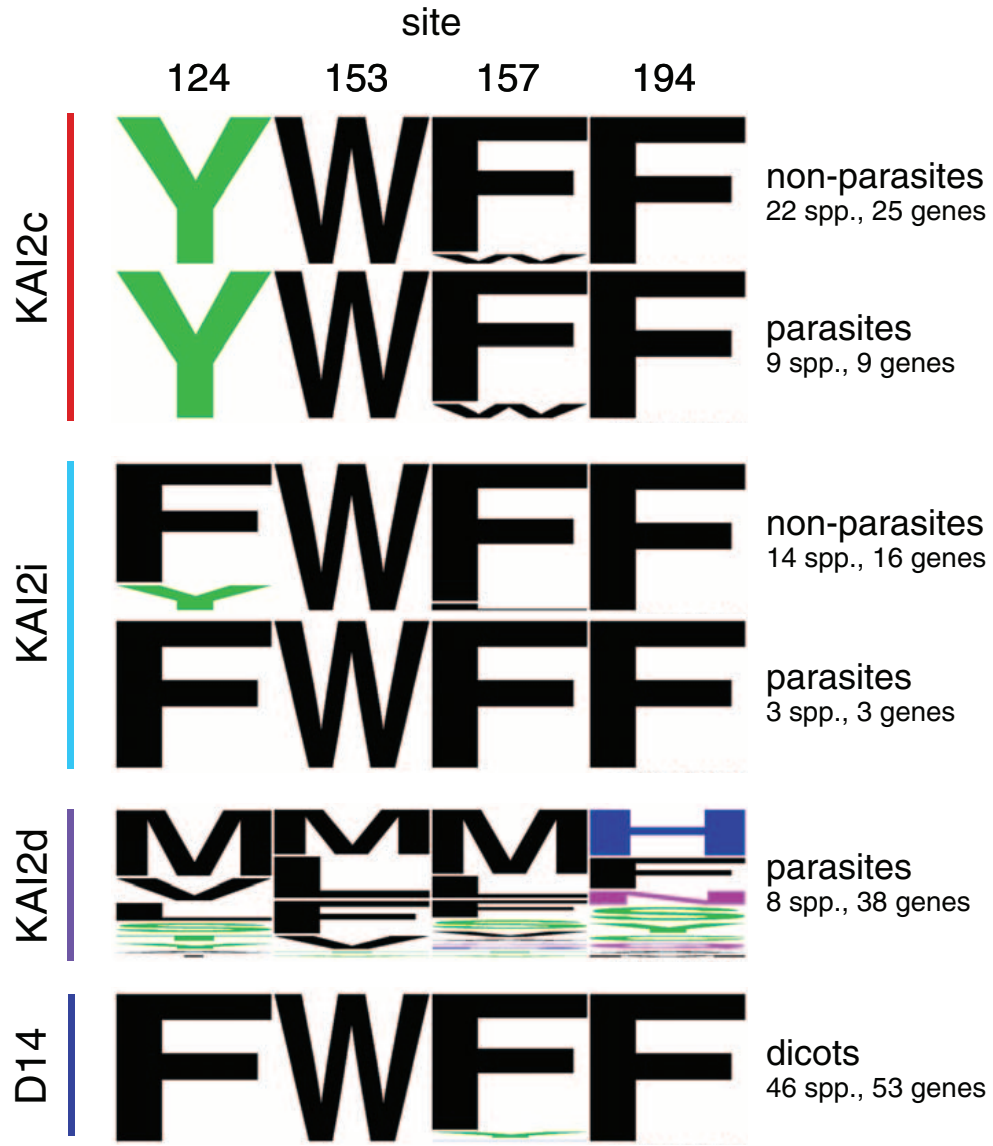

Figure 2.4. Amino acid conservation at four sites in the KAI2 protein. All four sites have high conservation outside of the parasite-specific KAI2d clade and high divergence within in. At least three of these sites – 124, 157, and 194 – are predicted to contribute to the ligand-binding pocket. Site numbers refer to the Arabidopsis KAI2 protein. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

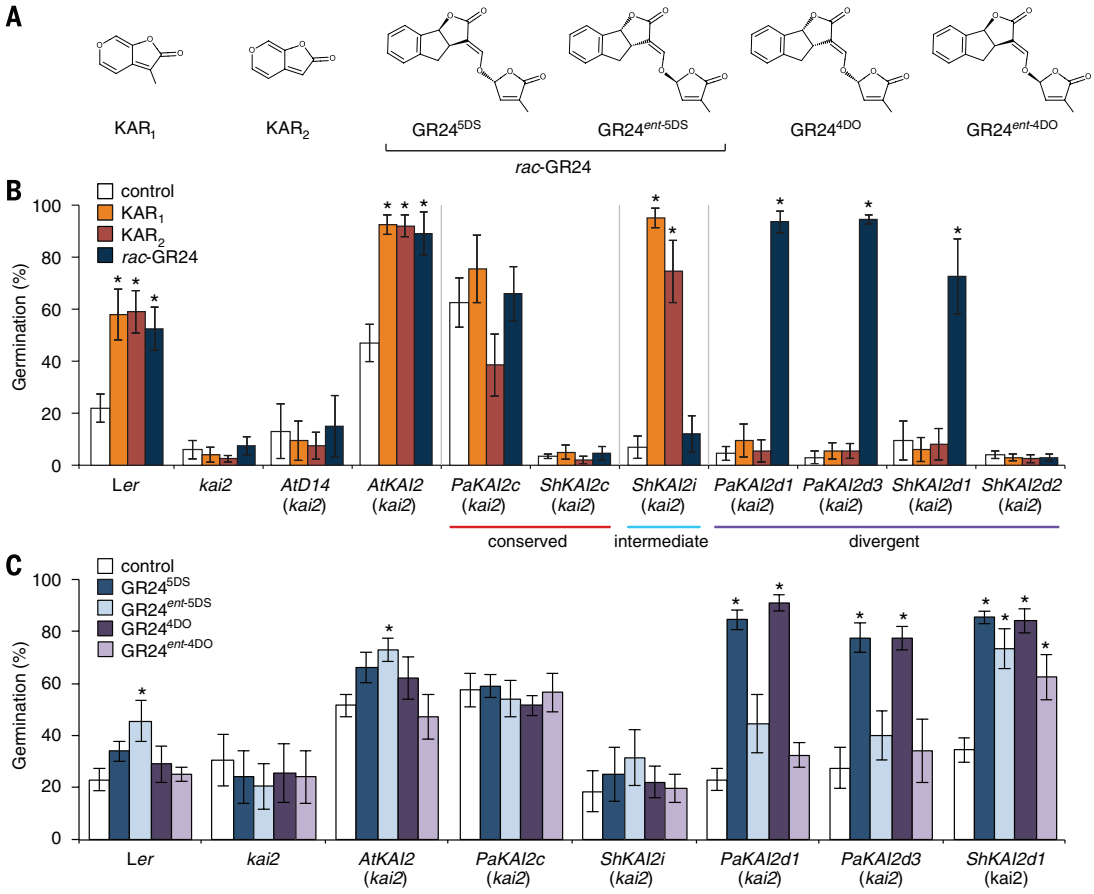

Figure 2.5. Germination responses conferred by parasite *KAI2* transgenes. Transgenes were put into an *Atkai2-2* null mutant background under the control of the *AtKAI2* promoter. Four replicates of 50 seeds were assayed for each line and treatment, and germination was averaged across replicates. *P < 0.01, according to the Tukey-Kramer HSD test comparing treated seeds to acetone control for each genotype. A) Treatments in germination assays included KAR_1, KAR_2, *rac*-GR24, and its four separate stereoisomers. B) Seed germination after six days in long-day light conditions at 21°C, with 1 µM treatment or acetone control. Treatments included the two KAR variants and *rac*-GR24. C) Seed germination after five days in long-day light conditions at 21°C, with 1 µM treatment or acetone control. Treatments included the four stereoisomers from *rac*-GR24. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

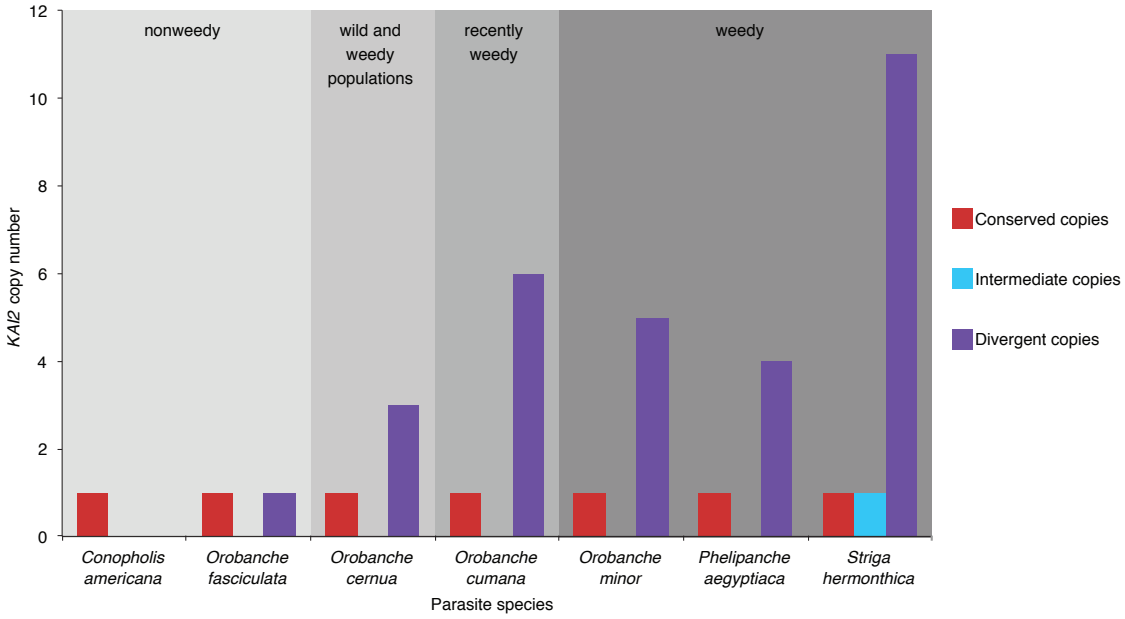

Figure 2.6. *KAI2* copy number in seven parasitic plant species. Weedy parasites, which tend to have broader host ranges in the *Orobanche* and *Phelipanche* (Schneeweiss et al. 2007), often have higher *KAI2* copy number due to an expansion of *KAI2d*. Information on weediness was obtained from Baird and Riopel 1986 (*C. americana*), Reuter 1986, Welsh et al. 1993, Brotherson et al. 2005 (*O. fasciculata*), and Parker 2013 (all other species).

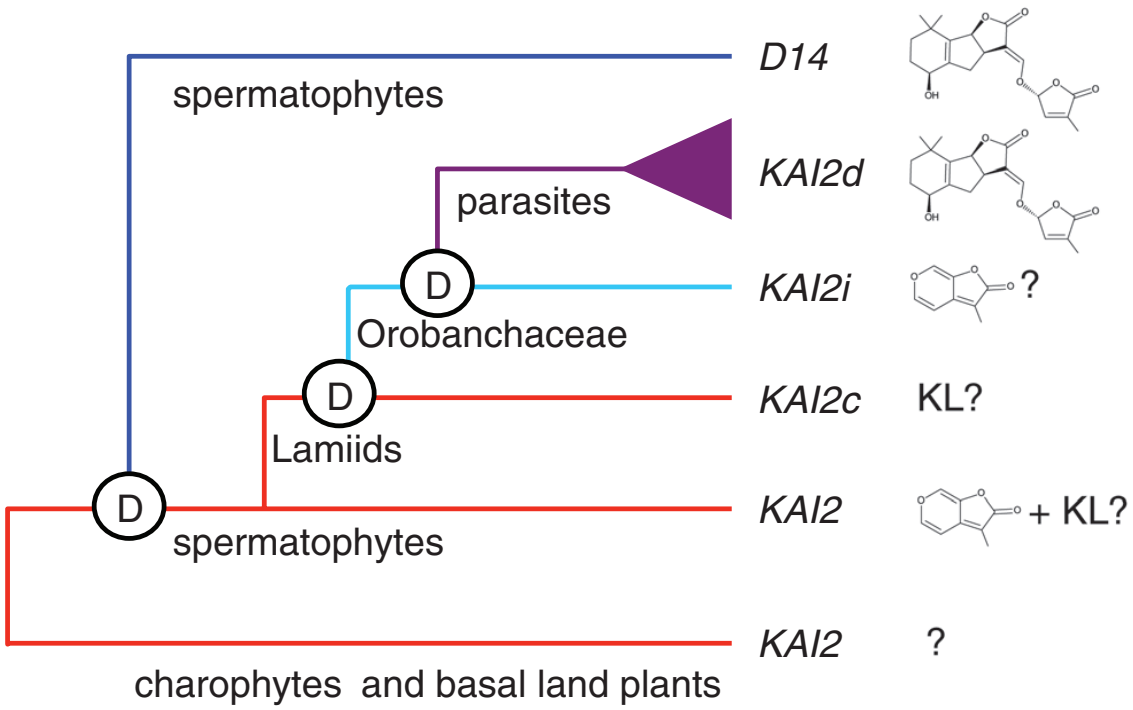

Figure 2.7. Model for the evolution of *KAI2*, including major duplication events. The ancestral function of *KAI2* in basal plants is unknown. SL responsiveness likely evolved convergently in the ancient *KAI2* duplicate *D14* and in *KAI2d* in parasitic plants. Relative to *AtKAI2*, which likely perceives KL and KAR, *KAI2c* and *KAI2i* in parasites appear to have undergone subfunctionalization. From Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543. Reprinted with permission from AAAS.

CHAPTER 3

PARASITIC WEEDS PROVIDE NEW EVIDENCE FOR AN ENDOGENOUS LIGAND OF KAI2[2]

[2] This work is published in Conn, C.E., Nelson D.C. (2016) Evidence that KARRIKIN-INSENSITIVE2 (KAI2) receptors may perceive an unknown signal that is not karrikin or strigolactone. *Frontiers in Plant Science* 6:1219. DOI: 10.3389/fpls.2015.01219.

Abstract

Germination is an irreversible transition in the life of a plant, and therefore germination in response to favorable environmental conditions is likely to be adaptive. For example, parasitic weeds in the Orobanchaceae family germinate in response to phytohormones in the soil known as strigolactones, which indicate that a potential host is likely nearby. On the other hand, many non-parasitic plants germinate in response to compounds derived from burning plant material called karrikins, which are abundant in post-fire environments with reduced competition. The α/β-hydrolase KAI2 has undergone extensive duplication in parasitic weeds, and some paralogs respond to strigolactones. A single KAI2 responds to karrikins in the model non-parasite *Arabidopsis thaliana*. Thus, KAI2 has evolved to perceive favorable germination conditions in different plant species.

The phenotypes of Arabidopsis *kai2* mutants suggest that KAI2 may also perceive an endogenous regulator of germination and development, which has yet to be identified and is referred to here as KAI2 ligand (KL). Cross-species complementation analyses indicate that one *KAI2* paralog from the basal vascular plant *Selaginella moellendorffii* and one from the parasitic weed *Phelipanche aegyptiaca* also respond to KL. To further investigate KL responsiveness in parasitic weeds, we tested *KAI2* paralogs from two species as transgenes in Arabidopsis *kai2* seedlings. We found that strigolactone-unresponsive *KAI2* from parasites preferentially respond to KL or to karrikin; thus, they appear to have subfunctionalized, relative to Arabidopsis *KAI2.* The existence of KL has been controversial, but our data on transgenic seedling development provide compelling new evidence for its existence. Our work here supports strong conservation of function in *KAI2* orthologs from physiologically diverse plants and hints at a possible ancestral function for *KAI2* in the transduction of an endogenous regulatory signal.

Introduction

Seed germination: a point of no return. Seed germination is a critical transition in the life of a plant. After germination, a plant must develop, grow, and reproduce with little control of its surroundings. Therefore, seed dormancy is an important adaptation that helps to restrict germination to times when environmental conditions are favorable (Bewley 1997). For different plant species, different sets of environmental conditions might be ideal for seed germination and seedling survival.

Parasitic plants attach to host plants and steal nutrients at the host's expense. The Orobanchaceae family includes some tremendously destructive agricultural weeds, many of which are obligate parasites that depend upon a host in order to survive and/or reproduce (Westwood et al. 2010). An ideal environment for the germination of these parasite seeds might thus be expected to have a high density of potential host plants. Strigolactones (SLs) are plant hormones that are abundant in the soil when plant density is high. SLs are exuded from plant roots into the rhizosphere, where they communicate with beneficial arbuscular mycorrhizal fungi (AMF; Akiyama et al. 2005). Because they are phytohormones, SLs in the soil are also an indication that a potential host plant is likely to be nearby. Parasitic weeds in the Orobanchaceae have evolved a germination response to SLs, and this adaptation helps to ensure that parasite seedlings will find a host (Hirsch et al. 2003). Thus, SLs are an indication of ideal germination conditions for parasites.

For many non-parasitic plant species, an environment with high plant density might be unfavorable for seed germination. When plants are abundant, competition among autotrophs for light and soil resources may be as well. An environment with low plant density – and therefore low competition – could be more ideal for the germination of non-parasites. For example, seedlings might be likely to thrive in post-fire

environments because competitors have been destroyed. Indeed, flowering, seed release, and seed germination are stimulated by fire or smoke in different plant species (reviewed in Nelson et al. 2012). One class of smoke-derived germination stimulants comprises the karrikins (KARs), which promote seed germination in diverse plant species, including monocots and dicots (Daws et al. 2007, Stevens et al. 2007) from at least three different continents (Flematti et al. 2004). KARs also stimulate seed germination in the model plant *Arabidopsis thaliana* (Nelson et al. 2009), in which both SL and KAR signaling have been studied extensively.

The SL and KAR signaling systems in Arabidopsis. Interestingly, SLs and KARs share partial structural similarity; both classes of compounds are characterized by a butenolide moiety (Flematti et al. 2004). Furthermore, the receptors for SLs and KARs in Arabidopsis – D14 and KAI2, respectively – are homologous α/β hydrolases (Hamiaux et al. 2012, Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Nakamura et al. 2013, Zhao et al. 2013). The signaling pathways for SLs and KARs converge on an F-box protein MAX2 (Nelson et al. 2011), and paralogous SMAX1/SMXL proteins then enter these signaling systems and help to confer specificity for SL or KAR signals (Stanga et al. 2013, Soundappan et al. 2015). Thus, in Arabidopsis, the SL and KAR signaling systems have similar components, and share at least one protein in common.

Despite these similarities, SLs and KARs cannot substitute for one other. In Arabidopsis, SLs do not stimulate seed germination (Scaffidi et al. 2014), but they are implicated in several developmental processes in this species and other model non-parasites (reviewed in Ruyter-Spira et al. 2013). On the other hand, KARs stimulate germination and enhance seedling light sensitivity in Arabidopsis (Nelson et al. 2009, Nelson et al. 2010). Along the same lines, *D14* and *KAI2* are not interchangeable in Arabidopsis. Different mutant phenotypes are observed in *Atd14* mutants than in *Atkai2*

mutants. Early in life, *Atd14* mutants are phenotypically normal, but later in life they develop extra axillary branches, similar to an SL-deficient mutant. While axillary branch number decreases in an SL-deficient mutant upon application of exogenous SLs, it does not in *Atd14*; thus, their branching phenotype is likely due to SL insensitivity. In contrast, *Atkai2* mutant phenotypes are apparent beginning with seed germination. In addition to KAR insensitivity, high seed dormancy, low light sensitivity in seedlings, and altered rosette morphology are all characteristic of *Atkai2* mutants (Waters et al. 2012). These mutant phenotypes in *Atkai2* plants cannot all be explained by KAR insensitivity alone.

An endogenous ligand for KAI2. KARs are not endogenous signals, and *Atkai2* plants have noticeable mutant phenotypes in the absence of KAR treatment. Thus, it has been hypothesized that KAI2 is also a receptor for an endogenous regulator of seed germination and development (Waters et al. 2012). This endogenous signal has yet to be identified. We refer to it here as KAI2 ligand (KL).

One argument against the existence of KL is that *KAI2* may be constitutively active at a low level, and KAR treatment may simply enhance its activity. The fact that *kai2* mutant phenotypes are directly opposite the effects of KAR treatment supports this hypothesis. However, several lines of evidence suggest that this is not the case. First, if KAR treatment only enhances constitutive *KAI2* activity, then overexpression of *KAI2* should have the same effect as KAR treatment. However, at the seedling stage, *KAI2* overexpression lines are not significantly different from wild type plants in the absence of KAR treatment (Waters and Smith 2013). Second, KAI2 proteins have a conserved catalytic triad, and amino acid replacement at one of these sites has been demonstrated to destroy KAI2 function (Waters et al. 2015). Third, KAI2 is likely a promiscuous receptor, as it can respond to GR24$^{ent\text{-}5\text{-deoxystrigol}}$, an enantiomer of an SL mimic (Scaffidi et al. 2014, Waters et al. 2015). Perhaps the response of AtKAI2 to KAR is merely a

consequence of its ability to perceive multiple, structurally related signals. Finally, a *KAI2* ortholog from the basal vascular plant *Selaginella moellendorffii* partially rescues seedling light insensitivity and altered rosette morphology in *Atkai2* mutants; however, it does not confer KAR sensitivity. Thus, it is likely specific for KL (Waters et al. 2015). Taken together, data from Arabidopsis indicate that KAI2 and its potential ligands – KAR and KL – have roles in seed germination and in early development.

Evidence for KL from parasitic plants. As in Arabidopsis, *KAI2* in parasitic weeds enable germination in favorable conditions. As noted above, parasitic weeds germinate in response to SLs rather than KARs, and KAI2 is not known to be a SL receptor in any non-parasitic plants. However, in parasites, *KAI2* has undergone extensive duplication, and some paralogs have evolved sensitivity to SLs. These SL-responsive *KAI2* copies in parasites are relatively fast-evolving, and are thus called divergent *KAI2*, or *KAI2d*. Parasite *KAI2* with intermediate (*KAI2i*) and conserved (*KAI2c*) evolutionary rates are not SL-responsive. When expressed as a transgene in an *Atkai2* mutant background, one *KAI2i* from the parasitic weed *Striga hermonthica* confers a germination response to KARs. This is surprising, because parasitic weeds themselves are not KAR-responsive in seed germination. A *KAI2c* gene from another parasitic weed, *Phelipanche aegyptiaca*, rescues the seed dormancy phenotype of *Atkai2* mutants but does not restore KAR sensitivity. Thus, this *PaKAI2c* paralog may be specific for KL. Another slow-evolving paralog from *S. hermonthica*, *ShKAI2c*, does not appear to function in seed germination, at least as a transgene in *Atkai2* mutants (Conn et al. 2015). Because only one *KAI2i* transgene and one *KAI2c* transgene from parasitic plants confer germination phenotypes in *Atkai2* mutants, the function of these slower-evolving *KAI2* paralogs from parasites remains unclear.

Here, we test *Atkai2* lines carrying parasite *KAI2* transgenes at the seedling stage. We assay light sensitivity and KAR response and obtain results similar to those of Conn et al. 2015. As in seed germination, *KAI2d* genes from parasites do not restore karrikin responsiveness, and have only a very slight effect on light sensitivity. One *KAI2i* transgene again confers KAR sensitivity, and partially restores wild type seedling light sensitivity in *Atkai2* plants. While *KAI2d* and *KAI2i* from parasites may retain slight sensitivity to KL, their responses to SL and KAR, respectively, are much more striking. On the other hand, as in seed germination, the *KAI2c* transgene from *P. aegyptiaca* rescues only the KAR-independent phenotype of light insensitivity in *Atkai2* mutants. Additionally, *KAI2c* from *S. hermonthica* restores light sensitivity to *Atkai2* mutants but does not confer KAR response. Thus, although *ShKAI2c* had no noticeable effect on seed germination, at the seedling stage it functions similarly to the conserved *KAI2* from *P. aegyptiaca* (Conn and Nelson 2016). The apparent specificity for KL among *KAI2c* paralogs from parasitic plants provides compelling evidence for the existence of this endogenous signal. Furthermore, the conservation of KL sensitivity in evolutionarily distant and physiologically diverse plants suggests that the ancestral role of KAI2 may have been response to this endogenous signal.

Methods

Methods are in accordance with those reported by Conn and Nelson (2016). Transgenic lines were generated by Conn et al. 2015. Seedling light sensitivity assays were performed on 10-day-old seedlings grown in short day conditions (8 h white light, ~36 µE : 16 h dark) at 21°C. Seedlings were grown on solid 1/2X Murashige and Skoog medium, adjusted to a pH of ~5.7 and supplemented with 1 µM treatments or an equivalent volume of acetone as a control. Seedlings were photographed, and their hypocotyls and cotyledons were measured using ImageJ (http://rsb.info.nih.gov/ij/).

Tissue was harvested and frozen from 10-day-old seedlings. RNA was extracted and used to make cDNA, and qRT-PCR was performed according to the method of Stanga et al. 2013. Primers used for qRT-PCR are presented in Table 3.1. Statistical tests were done in JMP (SAS Institute).

Results

Multiple *KAI2* from parasites are KL-responsive at the seedling stage. Light insensitivity in seedlings is one of the hallmark characteristics of Arabidopsis *kai2* null mutants. This is reflected in the long hypocotyls and small cotyledons observed in *Atkai2* seedlings (Sun and Ni 2011, Waters et al. 2012). In our seedling light sensitivity assays, we tested wild type and *Atkai2-2* mutant seed, as well as two transgenic lines carrying wild type *AtKAI2* in an *Atkai2-2* null mutant background. These three control genotypes behaved as expected, with a full restoration of light sensitivity conferred to *Atkai2* mutants by the *AtKAI2* transgene in at least one of two lines tested. We also tested two transgenic lines per parasite *KAI2* transgene. The conserved *KAI2* from *Phelipanche aegyptiaca* (*PaKAI2c*) rescued the KAR-independent mutant phenotypes of *Atkai2-2*, as it did in seed germination. In one of two transgenic lines tested, *PaKAI2c* fully restored wild type hypocotyl length and cotyledon size, and in the other, rescue of these phenotypes was nearly complete. In contrast to its apparent ineffectiveness in seed germination, the conserved *KAI2* from *Striga hermonthica* (*ShKAI2c*) behaved similarly to *PaKAI2c* at the seedling stage. In both transgenic lines tested, *ShKAI2c* at least partially rescued hypocotyl length and cotyledon size, although it only enabled a full restoration of the hypocotyl length phenotype, which occurred in one of two transgenic lines. The intermediate *KAI2* from *S. hermonthica* (*ShKAI2i*) appeared to weakly improve light sensitivity in *Atkai2-2* mutants, although it had not rescued seed dormancy in the

absence of KAR treatment in germination assays (Conn et al. 2015). However, *ShKAI2i* could not fully restore either light insensitivity phenotype in the transgenic lines that we tested. Finally, some transgenic lines carrying *KAI2d* from parasites appeared to have slightly greater light sensitivity than *Atkai2* mutants, but these transgenic lines still had long hypocotyls and small cotyledons. Of all the transgenes tested, *KAI2d* had the least impact on hypocotyl length and cotyledon size (Figure 3.1).

Our results were not unexpected, given the phenotypes conferred by parasite *KAI2* transgenes at the seed germination stage. However, a few transgenes – most notably *ShKAI2c* and *ShKAI2i* – appeared to be at least weakly KL-responsive at the seedling stage, while they conferred no KL response in seed germination (Conn et al. 2015). This observation raised the question of whether parasite *KAI2* transgenes can respond to a greater diversity of signals at later stages in plant development.

KAR responsiveness is similar in seed germination and seedling development. To determine whether parasite *KAI2* respond to more or different signals in seedling development than in seed germination, we tested our transgenic lines for responses to KAR and the synthetic strigolactone GR24$^{5\text{-deoxystrigol}}$ (GR24^{5DS}). We also included the enantiomer of this synthetic strigolactone, GR24$^{\textit{ent}\text{-5-deoxystrigol}}$ (GR24$^{\textit{ent}\text{-5DS}}$), which is perceived by KAI2 in Arabidopsis (Scaffidi et al. 2014). As in seed germination, *ShKAI2i* was the only KAR-responsive transgene from parasites. However, all lines tested were responsive to GR24^{5DS} and GR24$^{\textit{ent}\text{-5DS}}$ (Figure 3.2). Sensitivity to GR24^{5DS} was not surprising; all seedlings had a functional D14, which can mediate responses to strigolactones and synthetic mimics in seedling development (Waters et al. 2012, Scaffidi et al. 2014). However, sensitivity to GR24$^{\textit{ent}\text{-5DS}}$ was unexpected, especially in *Atkai2* null mutants. Furthermore, of all six parasite *KAI2* tested, only one had conferred

a germination response to GR24$^{ent\text{-}5DS}$, making the responsiveness of all transgenic lines observed here quite surprising.

Taken together, the behavior of *Atkai2* mutants and all 12 transgenic lines (two per parasite *KAI2* transgene) suggest a KAI2-independent mechanism of GR24$^{ent\text{-}5DS}$ perception in Arabidopsis. This signal was previously thought to be mediated by *KAI2* and not *D14* in Arabidopsis. However, Scaffidi et al. (2014) show a small, statistically insignificant response of *Atkai2* mutants to GR24$^{ent\text{-}5DS}$ and *ent*-5-deoxystrigol in seedling development. This slight response in *Atkai2* mutants is not observed in *Atd14 kai2* double mutants (Scaffidi et al. 2014). Thus, AtD14 may be able to confer some responsiveness to this enantiomer of an SL mimic. Because *Atkai2* mutants responded to both GR24^{5DS}and GR24$^{ent\text{-}5DS}$ at the seedling stage, we cannot draw conclusions about the responsiveness of parasite *KAI2* transgenes to these signals.

Conserved and intermediate *KAI2* from parasites rescue *Atkai2* rosette morphology. Later in plant development, *Atkai2* mutants have altered morphology, characterized by long petioles and small leaves (Waters et al. 2012, Waters et al. 2015). This phenotype is observed in the absence of KAR treatment and is thus likely attributable to KL insensitivity. As in our seedling light sensitivity assay, conserved parasite *KAI2* transgenes fully or almost fully restored wild type rosette morphology. The intermediate *KAI2* from *S. hermonthica* partially restored the wild type phenotype. Transgenic lines carrying parasite *KAI2d* were indistinguishable from *Atkai2* null mutants at this stage (Figure 3.3). These results are consistent with KL sensitivity in parasite *KAI2c*, and to a lesser extent in parasite *KAI2i*.

Transgene expression level correlates with observed KL responsiveness. We observed different degrees of apparent KL responsiveness among different *KAI2* transgenes.

Furthermore, *KAI2d* transgenes had little to no noticeable effect on *Atkai2* mutants from the seedling stage on. To determine whether transgene expression level could explain the degree of KL responsiveness observed, we measured expression of each transgene itself and of *DLK2*, which is a reporter gene for *KAI2* signaling (Waters et al. 2012). Interestingly, expression of these two genes was correlated with the degree of KL responsiveness that we observed among our transgenic lines (Figure 3.4). Thus, it is possible that differences in transgene expression rather than in KL sensitivity explain the range of phenotypes conferred by conserved, intermediate, and divergent *KAI2* transgenes. However, while expression of *KAI2d* transgenes was low at the seedling stage, parasite *KAI2d* conferred strong germination responses to SL analogs in seed germination (Conn et al. 2015). Thus, a low level of transgene expression is unlikely to fully explain the failure of parasite *KAI2d* to rescue *Atkai2* mutant phenotypes at the seedling stage and beyond.

Discussion

In general, parasite *KAI2* transgenes functioned similarly at the seed germination and seedling stages. However, in seed germination, only one representative of the conserved *KAI2* clade appeared to be KL-responsive. In seedlings, both conserved *KAI2* from parasites that we tested fully or almost fully restored the KAR-independent phenotypes of *Atkai2* mutants. While *ShKAI2i* – and perhaps, to a lesser extent, *KAI2d* paralogs –had slight KL sensitivity in seedling development, KL response does not appear to be the primary function of *KAI2* from the intermediate or divergent clades.

As in seed germination, *ShKAI2i* conferred KAR sensitivity in seedlings. Parasite *KAI2d* were not KAR responsive, and transgenic lines carrying them were almost indistinguishable from *Atkai2* mutants in the absence of exogenous chemical treatments. Thus, our results here support the model of gene duplication and subsequent evolution

put forth by Conn et al. (2015). The *KAI2c* and *KAI2i* paralogs from parasites appear to have subfunctionalized, relative to *AtKAI2*. While *KAI2c* and *KAI2i* are primarily responsive to KL and to KAR, respectively, *AtKAI2* is sensitive to both signals. Although our seedling data provide little to no new information on the function of *KAI2d* from parasites, our results here are consistent with minimal KL sensitivity among these fast-evolving parasite *KAI2*.

The *ShKAI2c* and *ShKAI2i* transgenes appear to be increasingly KL responsive throughout development, at least as a plant transitions from the seed germination to the seedling stage. Interestingly, this trend has also been observed for a *KAI2* ortholog from *Selaginella moellendorrfii* (*SmKAI2a*). Although *SmKAI2a* has little noticeable impact on the seed germination of *Atkai2* mutants, it partially restores seedling light responses. Later in development, *SmKAI2a* largely rescues the rosette morphology phenotype of *Atkai2* mutants. Waters et al. (2015) suggest that this could be because developmental processes in *S. moellendorffii* and Arabidopsis become more homologous as plant development progresses. However, parasitic plants and Arabidopsis undergo much more similar developmental processes, making this an unlikely explanation for the phenotypes conferred by parasite transgenes. An alternative explanation might be that transgenes from parasitic plants and *S. moellendorffii* cannot interact with other components of the KL signaling system as effectively as *AtKAI2* can. Perhaps the abundance of interacting proteins is a limiting factor in seed germination, but becomes less and less important in subsequent developmental stages. This hypothesis is currently unaddressed.

Together with data from Waters et al. (2015), our results here provide compelling evidence for the existence of KL. The KAR-independent phenotypes of *Atkai2* mutants were once one of the most convincing arguments for the existence of this unidentified endogenous regulator. Now, *KAI2* transgenes from evolutionarily and physiologically

diverse plant taxa indicate that KL is a highly conserved signal. The identification of KL is a daunting challenge; our only hypothesis regarding its identity is that it likely includes a butenolide moiety, as do the known ligands of KAI2 and D14 (Hamiaux et al. 2012, Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Nakamura et al. 2013, Zhao et al. 2013, Scaffidi et al. 2014, Waters et al. 2015). Because of its effects on seed germination and seedling development, the identification of KL will be an important discovery with agricultural applications. For example, KL or synthetic analogs could be used to stimulate the germination of weeds, which could subsequently be destroyed prior to the planting of crops.

The conservation of KL sensitivity across distantly related plant taxa hints that KL response might be the ancestral function of *KAI2*. This gene is present throughout much of the plant kingdom, including some green algae (Delaux et al. 2012), but its function is largely unknown in basal plants. At least one species of green algae is SL-responsive (Delaux et al. 2012), and some *KAI2* paralogs in the moss *Physcomitrella patens* respond transcriptionally to SL treatment (Lopez-Obando et al. 2016). Thus, SL response is a candidate ancestral function of *KAI2*. However, the angiosperm SL receptor D14 is absent in green algae (Delaux et al. 2012) and has been thought to be derived from a duplication of KAI2 sometime during the evolution of vascular plants or spermatophytes (e.g. Delaux et al. 2012, Waters et al. 2012, Waters et al. 2015). However, according to a new, extensive phylogeny of *KAI2* and its homologs throughout the plant kingdom, a highly conserved set of *KAI2* orthologs is present in land plants. Duplications of these conserved *KAI2* have led to functional diversification, exemplified by SL-responsive homologs in angiosperms (*D14*) and in more basal land plants, including moss. This phylogeny suggests a new classification scheme for *KAI2* homologs, in which most SL-responsive *KAI2* from moss are closer relatives to angiosperm *D14* than *KAI2*. Thus, rather than SL perception being the ancestral function

of *KAI2*, it may have arisen in *KAI2* duplicates in basal plants, whose divergence was previously thought to predate the origin of SL responsiveness in *KAI2* homologs (Bythell-Douglas et al. 2017).

The presence of other, more highly conserved *KAI2* genes throughout the plant kingdom suggests an ancestral function other than SL responsiveness. The *KAI2* ortholog in rice was recently shown to mediate interactions with arbuscular mycorrhizal fungi, with which ~80% of plants associate; thus, this widespread mutualism could be an ancestral function of *KAI2* (Gutjahr et al. 2015). However, the conservation of KL responsiveness in *KAI2* from a basal vascular plant, Arabidopsis, and parasitic weeds suggests that KL signaling is the ancestral function of *KAI2*. Identifying KL and determining its function in plants with different developmental trajectories (e.g. parasitic plants, moss) are important next steps in the study of this endogenous signal. This information will provide valuable new insight into the evolution of *KAI2* and the signaling pathways involved in seed germination and seedling development.

Table 3.1. Primers used for qRT-PCR. From Conn, C.E., Nelson, D.C. (2016) Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Frontiers in Plant Science* 6: 1219. DOI: 10.3389/fpls.2015.01219.

Gene name	Forward primer	Reverse primer
AtKAI2	TGATCTCTGCTTCTCCGAGATACG	CCACGCTTTGTAGTTGCTTCGG
PaKAI2c	CGTTGGCCATTCTGTCTCGG	CGTCTTGTTCGAATCCTCCAAAG
ShKAI2c	CCAAGCTTGTCACAGTCGCCGGCTC	ATGGTCTGGAGAATGCTGAGGG
ShKAI2i	CATCATCCGCCCTGACCTCTTCC	CCTGCCAGCTGCTCCACGTCC
PaKAI2d1	CGTCGGACACTCTTTGTCTGCC	CGCCCTTCCCGATTCCAGTAAC
PaKAI2d3	TGTATCTACGTCGGCCACTCTCTG	TCCTCCTTTGTAATCAGCCGAG
ShKAI2d1	CTCTCGTCCATGGTTGCGGC	GTTTTCCTCCATGGCGGCTTGC
DLK2	GCTGCTTCTCCAAGGTATATAA	GAAATCAACCGCCCAAGCT
CACS	GGAGAAGAGAGGGCCTTGCTTACAA	TTAGCTGGGCGAGATTTCATTTCTG

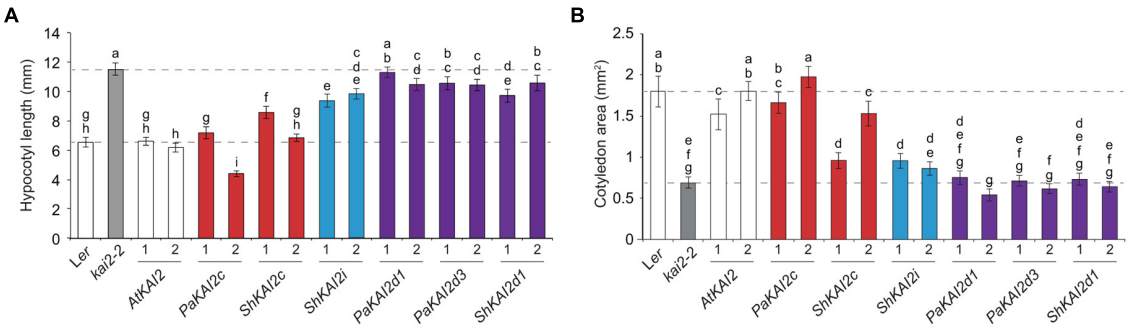

Figure 3.1. Seedling light responses conferred by parasite *KAI2* transgenes. (A) Hypocotyl length and (B) cotyledon area were measured in seedlings grown in short-day white light conditions for ten days. Transgenes were from *Arabidopsis thaliana* (*AtKAI2*), *Phelipanche aegyptiaca* (*PaKAI2*), and *Striga hermonthica* (*ShKAI2*) and were under the control of the *AtKAI2* promoter in *Atkai2-2* null mutants. Two transgenic lines per transgene were tested. Mean ±99% CI, $n = 45$. Tukey–Kramer HSD test, $*p < 0.01$. From Conn, C.E., Nelson, D.C. (2016) Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Frontiers in Plant Science* 6: 1219. DOI: 10.3389/fpls.2015.01219.

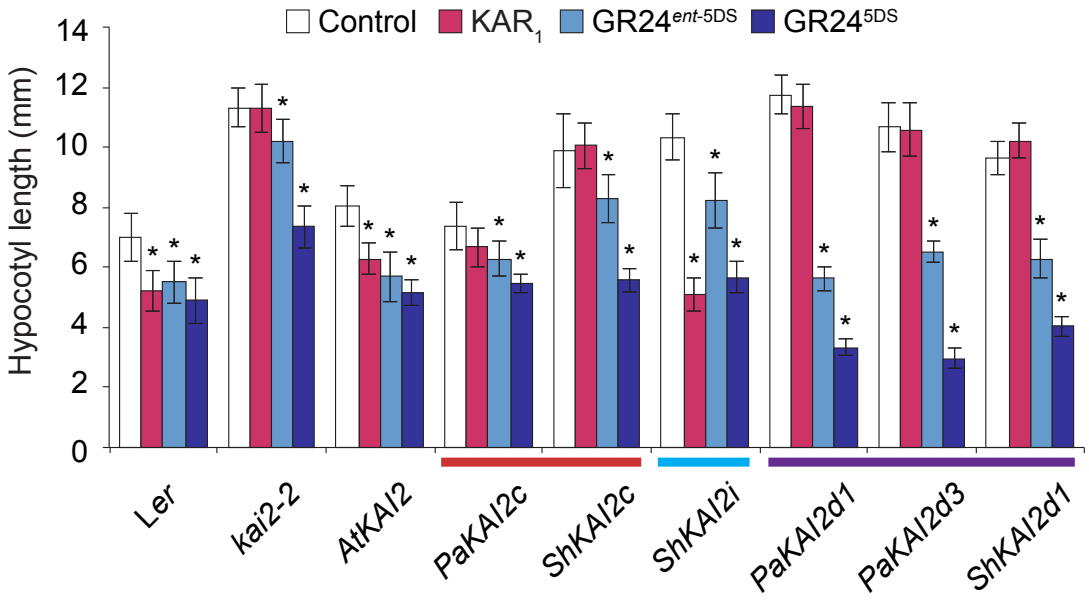

Figure 3.2. Responses to karrikin and strigolactone enantiomers conferred by parasite *KAI2* transgenes. Hypocotyl length was measured in seedlings grown in short-day white light conditions for ten days. Transgenic lines were produced as described in Figure 3.1. Treatments included karrikin 1 (KAR_1), the natural strigolactone mimic $GR24^{5\text{-deoxystrigol}}$ ($GR24^{5DS}$), and its enantiomer $GR24^{ent\text{-}5\text{-deoxystrigol}}$ ($GR24^{ent\text{-}5DS}$). Mean ±99% CI, $n = 15$. Dunnett's test comparing treatments to control, $^*p < 0.01$. From Conn, C.E., Nelson, D.C. (2016) Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Frontiers in Plant Science* 6: 1219. DOI: 10.3389/fpls.2015.01219.

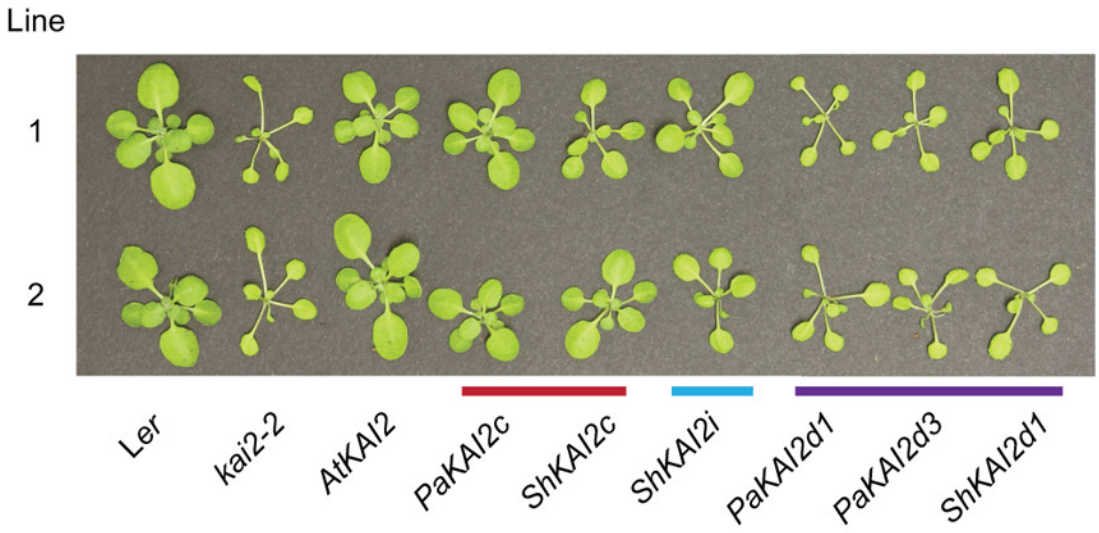

Figure 3.3. Rosettes in 31-day-old plants from transgenic lines and controls. Plants were grown in short-day white light conditions. One representative from each of two transgenic lines per transgene was photographed. From Conn, C.E., Nelson, D.C. (2016) Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Frontiers in Plant Science* 6: 1219. DOI: 10.3389/fpls.2015.01219.

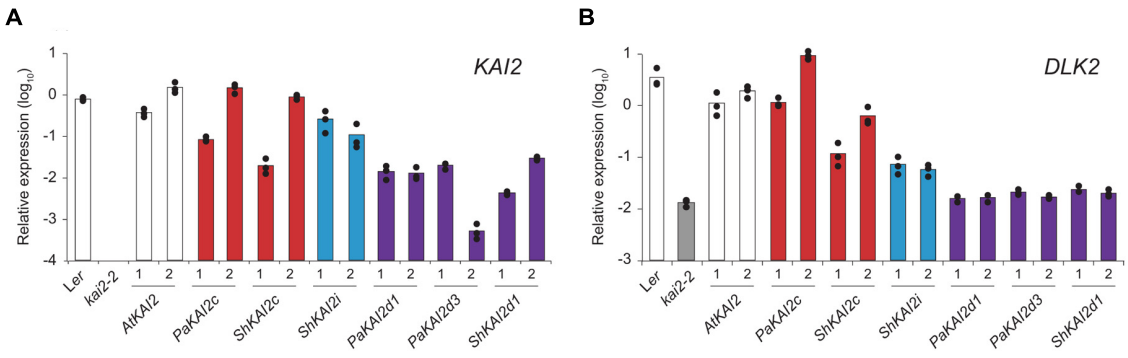

Figure 3.4. Expression of *KAI2* and *DLK2* in ten-day-old seedlings. Expression was measured relative to the *CACS* reference gene. Three biological replicates (indicated by black dots) with at least two technical replicates each were measured and averaged. A) *KAI2* expression was measured for wild-type *AtKAI2* (L*er*) or the *KAI2* transgene (each transgenic line). B) Native *DLK2* expression was measured in each line. From Conn, C.E., Nelson, D.C. (2016) Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Frontiers in Plant Science* 6: 1219. DOI: 10.3389/fpls.2015.01219.

CHAPTER 4

FROM MOSS TO PARASITIC WEEDS: AN EVOLUTIONARY AND STRUCTURAL ANALYSIS OF KAI2 EVOLUTION[3]

[3]Conn, C.E., Voges, L.R., Nelson, D.C. To be submitted to *Current Biology*.

Abstract

The *KAI2* gene is represented throughout much of the plant kingdom and plays various roles in environmental perception. It has primarily been studied in the model plant *Arabidopsis thaliana*, in which it responds to smoke-derived germination cues, and likely to an endogenous signal. Although *KAI2* exists as a single-copy gene in Arabidopsis, it has undergone frequent duplications throughout the plant phylogeny. Parasitic weeds have especially high *KAI2* copy number, and three different clades of parasite KAI2 paralogs likely respond to different signals. Here, we investigate the relationship between structure and function of KAI2 proteins, with an emphasis on fast-evolving KAI2 from parasitic weeds that enable a germination response to host-derived strigolactones. We identify predicted specificity-determining positions (SDPs) in parasite KAI2 proteins and test their function by modifying them in Arabidopsis KAI2. We partially recapitulate the preference of a strigolactone-unresponsive parasite KAI2 for an abiotic signal; however, we cannot engineer a strigolactone receptor with our targeted modifications of Arabidopsis KAI2. Thus, the evolution of strigolactone sensitivity in KAI2 is likely more complex than the few amino acid replacements we made.

Next, we survey *KAI2* in diverse non-parasite taxa and analyze the evolution of *KAI2* paralogs that have arisen from separate duplication events. Specifically, we investigate whether *KAI2* paralogs evolve convergently in distantly related plants. We also take a cross-species complementation approach to determine the function of two *KAI2* paralogs from *Lactuca sativa* (lettuce), which appear to have subfunctionalized, relative to Arabidopsis *KAI2*. This is a novel functional analysis of *KAI2* duplicates in a non-parasitic angiosperm. Together with functional data from *KAI2* duplicates in parasitic weeds, our transgenic lines provide valuable insight into gene duplication and subsequent evolution of paralogs.

Introduction

KAI2 genes perceive different signals. The *KAI2* gene encodes an α/β-hydrolase that has been implicated in the detection of different chemical signals in various plant taxa. In the model species *Arabidopsis thaliana*, a single *KAI2* gene likely responds to three different cues. First, it is a receptor for smoke-derived compounds called karrikins (KARs; Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Zhao et al. 2013), which promote seed germination in a wide variety of plant species (e.g. Flematti et al. 2004, Daws et al. 2007, Stevens et al. 2007). Arabidopsis KAI2 also responds to GR24$^{ent\text{-}5\text{-deoxystrigol}}$, which is an enantiomer of a synthetic version of the phytohormone strigolactone (SL; Scaffidi et al. 2014, Waters et al. 2015). In addition to these abiotic signals, KAI2 likely perceives an endogenous regulator of germination and development (Waters et al. 2012), which has yet to be identified but is referred to as KAI2 ligand (KL; Conn and Nelson 2016). Another α/β-hydrolase called D14 likely arose from an ancient duplication of *KAI2* (Waters et al. 2012, Bythell-Douglas et al. 2017). In Arabidopsis, D14 is a receptor for SLs (Hamiaux et al. 2012, Kagiyama et al. 2013, Nakamura et al. 2013, Zhao et al. 2013), which play various roles in development (Ruyter-Spira et al. 2013). Both KARs and SLs share a butenolide moiety (Flematti et al. 2004). Thus, in Arabidopsis, the known ligands of KAI2 and its ancient paralog D14 have partial structural similarity to one another.

While *KAI2* has been characterized most thoroughly in Arabidopsis, some recent studies have begun to uncover the functional diversity of this gene throughout the plant kingdom. In rice, *KAI2* plays an important role in symbiosis with arbuscular mycorrhizal fungi (Gutjahr et al. 2015). The moss *Physcomitrella patens* has at least 11 *KAI2* paralogs, some of which respond transcriptionally to SL treatment (Lopez-Obando et al. 2016). It is possible that *KAI2* also responds to SL in charophyte green algae, which lack

D14 but in which at least one species is SL-responsive (Delaux et al. 2012). Finally, a *KAI2* gene in the basal vascular plant *Selaginella moellendorffii* responds to KL but not KAR as a transgene in Arabidopsis (Waters et al. 2015).

Parasitic weeds in the Orobanchaceae, many of which germinate in response to host-derived SLs (Cook et al. 1966, Bouwmeester et al. 2003), have an especially interesting array of *KAI2* genes. Some *KAI2* paralogs from parasites respond to SLs; these copies are relatively fast-evolving, and are thus called divergent *KAI2*, or *KAI2d*. One *KAI2* paralog from the parasitic weed *Striga hermonthica* has an intermediate rate of evolution (*KAI2i*) and appears to preferentially respond to KARs. Slow-evolving or conserved *KAI2* (*KAI2c*) from two parasitic weed species are likely specific for KL (Conn et al. 2015, Conn and Nelson 2016). While *KAI2d* are only found in parasitic plants, *KAI2i* and *KAI2c* are present in non-parasitic Lamiids (Lamiales, Gentianales, and Solanales) as well. In summary, although tests of *KAI2* function in non-model plants are currently quite limited, the data collected to date suggest that *KAI2* duplication is a relatively common process in plants, and that paralogs often evolve specificity for a more limited set of signals than is observed for Arabidopsis KAI2.

The preference of some *KAI2* genes for a single signal raises the question of why Arabidopsis *KAI2* responds to so many. Arabidopsis has not been considered to be a fire-following species (Nelson et al. 2012), so perhaps its responsiveness to smoke-derived KAR – and to GR24$^{ent\text{-}5\text{-deoxystrigol}}$ – is coincidental (Conn and Nelson 2016). However, because of the broad distribution of Arabidopsis, smoke-responsive seed germination may indeed be adaptive (Lamont and He 2017). Another possibility is that KAR responsiveness evolved in a fire-following ancestor of Arabidopsis and has not yet been lost from species that occur outside of fire-prone environments. Support for this hypothesis comes from the close relationship between Arabidopsis and the fire follower *Brassica tournefortii* (Morffy et al. 2016), which is KAR-responsive in seed germination

(Nelson et al. 2009). Regardless of the reason why KAI2 perceives multiple signals in Arabidopsis, functional data from other species suggest that this receptor is unusually promiscuous in Arabidopsis.

For KAI2 orthologs in other plant species that perceive a narrower range of signals, the adaptive significance of likely ligand specificity is easy to imagine. For example, the SL specificity of some parasite KAI2 likely helps to ensure that parasitic weeds germinate only in the presence of a host (Hirsch et al. 2003, Conn et al. 2015, Toh et al. 2015, Tsuchiya et al. 2015). On the other hand, the KAR specificity of a *KAI2* paralog from the parasitic weed *Striga hermonthica* is somewhat puzzling. Parasitic weeds themselves are not KAR-responsive in seed germination (Conn et al. 2015); indeed, germination in the absence of a host is suicidal for obligate holoparasites (Zwanenburg et al. 2016). However, non-parasitic relatives of *S. hermonthica* occur in fire-prone environments and are stimulated to germinate by smoke (Moreira et al. 2010). At least one of these non-parasitic relatives, *Thymus vulgaris*, has a *KAI2* gene that groups phylogenetically with the KAR-responsive *KAI2* from *S. hermonthica* (Conn et al. 2015). The function of *KAI2* paralogs in *T. vulgaris* has not been tested. However, the smoke-responsive seed germination of this species (Moreira et al. 2010) and the phylogenetic position of one of its *KAI2* genes suggest a mechanism of *KAI2*-mediated germination after fires. With regard to parasitic weeds, then, it is possible that a *KAI2* paralog evolved KAR specificity in a fire-following ancestor. This hypothesis is similar to one described above regarding KAR responsiveness in Arabidopsis.

If a *KAI2* paralog evolved specificity for KAR in a smoke-responsive ancestor of parasitic weeds, then this paralog might be expected to undergo pseudogenization in parasites over time. Curiously, hemiparasites in the Orobanchaceae, which retain photosynthetic potential, have *KAI2* paralogs that belong to the same phylogenetic clade as the KAR-responsive one from *S. hermonthica*. On the other hand, obligate

holoparasites, which have lost photosynthetic potential, appear to have lost the *KAI2* paralog that would belong to this clade. This pattern of gene retention only in photosynthetically competent parasites suggests a function for these paralogs other than KAR signaling. Perhaps they also play a role in light response or another process related to photosynthesis.

Finally, the adaptive significance of KL specificity in some *KAI2* genes is unclear, largely because KL is currently unidentified and therefore not well understood. However, the conservation of KL sensitivity among *KAI2* genes from such evolutionarily distant plants as *S. moellendorffii* (Waters et al. 2015) and parasitic weeds (Conn et al. 2015, Conn and Nelson 2016) suggests that this signal is vitally important. If this is the case, it is not at all surprising that diverse plant species have *KAI2* paralogs that are specialized for KL responsiveness.

Investigating the structural basis of likely ligand specificity. The question of why different *KAI2* genes might have evolved specificity for different signals is relatively simple to address. While the selective pressures that have shaped *KAI2* cannot be identified with certainty, hypotheses can easily be developed based on the ecological and physiological characteristics of different plant species. For instance, specificity for KAR is likely adaptive in some autotrophic plant species that thrive in post-fire environments with reduced competition (Nelson et al. 2010), and, as discussed previously, specificity for SLs is probably beneficial in parasitic weeds, which require a host soon after germination in order to survive (Matusova et al. 2005). The question of how *KAI2* genes might have evolved specificity for different signals is less straightforward. Structural differences have been noted in KAI2 proteins that respond to different signals; for example, SL-responsive KAI2 paralogs in parasitic weeds are predicted to have larger ligand-binding pockets than paralogs that respond to KAR or KL. Furthermore, several

amino acid sites that are predicted to influence the ligand binding are highly conserved outside of the clade containing SL-responsive parasite *KAI2* and highly divergent within it (Conn et al. 2015). These structural differences among KAI2 proteins that respond to different signals are noteworthy, but are they sufficient to explain differences in likely ligand preference? To thoroughly address this question, specificity-determining positions (SDPs) in KAI2 should be predicted by sophisticated computational analyses and functionally tested by modification of the KAI2 protein.

Numerous methods have been developed for identifying candidate SDPs in proteins. Many look for sites in a protein alignment where amino acid substitutions correspond to differences in protein function (Capra and Singh 2008). One simple and broadly applicable method for identifying SDPs is JDet, which requires only an amino acid alignment as input. Within JDet, the program S3Det can divide aligned amino acid sequences into predicted functional groups and identify candidate SDPs at the same time (Rausell et al. 2010, Muth et al. 2012). The functional diversity of KAI2 among plants is largely unknown, so JDet is a useful method for classifying these orthologs and paralogs into predicted functional groups. Furthermore, in their paper describing S3det, Rausell et al. (2010) discuss the distribution of SDPs in ligand-binding pockets and in protein interaction interfaces. SDPs in the former but not in the latter likely contribute to the ability of different KAI2 proteins to bind different signals. An advantage of JDet is that three-dimensional protein models can be visualized in it, and candidate SDPs identified by S3det can be highlighted (Muth et al. 2012). This can help to distinguish SDPs that may determine ligand specificity from those that might influence protein-protein interactions. Notably, the S3det program in JDet has been used to identify SDPs between KAI2 and D14, at least one of which is a surface residue that is likely involved in protein-protein interactions. This site (169 in D14) has a different amino acid identity in

KAI2; thus it may contribute to the specificity of the D14 and KAI2 signaling systems that is conferred by interactions with different protein partners (Chevalier et al. 2014).

Another method for identifying sites that may be functionally important is Phylogenetic Analysis by Maximum Likelihood (PAML). Although PAML is not a functional residue detection program, it can identify sites of adaptive evolution by finding codons with an excess of nonsynonymous mutations. To do this, PAML calculates ω, or the ratio of the nonsynonymous substitution rate (dN) to the synonymous substitution rate (dS). Positive selection is inferred if ω exceeds 1. PAML also performs ancestral state reconstruction, and although sequences are reconstructed with uncertainty, this tool is useful for tentative mapping of evolutionary transitions on a phylogeny (Yang 2007). An example of PAML's usefulness for analyzing molecular evolution comes from a study of the amino acid replacements responsible for a transition in flower color. Different species in the genus *Iochroma* have red or blue flowers, depending on which anthocyanin precursor they use to produce pigment. Blue flowers are ancestral in this genus, but red flowers have evolved and are present in some *Iochroma* species. Branch-site model selection tests in PAML indicate five positively selected sites in the DFR enzyme, which acts on blue or red pigment precursors in the anthocyanin biosynthesis pathway. Ancestral state reconstruction enables the substitutions in these sites to be mapped onto a phylogeny, and from this information, it seems that three amino acid replacements in positively selected sites shifted the specificity of DFR from blue to red pigment precursors in *Iochroma*. Later amino acid replacements at the other two sites are associated with further specialization of DFR for red pigment precursors (Smith et al. 2013). The shift in DFR specificity analyzed by Smith et al. (2013) is somewhat analogous to the transition from KAR and/or KL responsiveness in non-parasite KAI2 to SL sensitivity in paralogs from parasitic weeds. While selection on parasite *KAI2* genes as a whole has been investigated with PAML (Conn et al. 2015),

branch-site modeling and ancestral state reconstruction have not yet been employed. These additional analyses in PAML are logical next steps in the study of *KAI2* evolution in parasitic weeds. As with JDet, the sites selected by PAML can be highlighted in a model of the KAI2 protein. Then, to reduce noise, only those that likely occur in or near the ligand-binding pocket can be selected for further study.

Three-dimensional protein structures are complex, and amino acid replacement at one site may affect interactions with other sites elsewhere in the protein. Based on the assumption that amino acids that physically interact often coevolve (Göbel et al. 1994), Direct Coupling Analysis (DCA) was developed to predict contacts between amino acids, based on a protein alignment input (Morcos et al. 2011). Recently, DCA was used to study amino acid coevolution in homo-oligomerization interfaces in nearly two thousand protein families (Uguzzoni et al. 2017). Tools like JDet and PAML may also capture sites that interact with SDPs; however, trimming the resulting site lists to only include those in or near the ligand-binding pocket may cause important coevolving sites to be missed. With a tool like DCA, sites that are predicted to interact with candidate SDPs in close proximity to the ligand-binding pocket can also be identified.

Of course, computational analyses are useful but not conclusive, and predicted SDPs must be functionally tested. For an investigation of likely ligand preference in KAI2 proteins, SDPs can be altered, and the resulting modified KAI2 proteins can be tested for responsiveness to various signals. This could be accomplished by assaying modified KAI2 *in vitro* or by generating transgenic Arabidopsis *kai2* lines carrying modified *KAI2* transgenes. In either case, many different combinations of modifications may need to be tested, making this functional assessment of candidate SDPs potentially time- and labor-intensive.

KAI2 in a broader context. As described above, the functional diversity of *KAI2* throughout the plant kingdom has only just begun to come to light. The single *KAI2* gene in Arabidopsis has been the primary focus of studies on *KAI2* function; however, many plant species, ranging from moss to parasitic weeds, have multiple copies of *KAI2* in their genomes. In parasitic plants, *KAI2* copies appear to have undergone sub- and neofunctionalization (Conn et al. 2015), which may have also occurred among the 11 *KAI2* paralogs in *P. patens*. Some *KAI2* copies in this moss species respond transcriptionally to SL treatment, and others do not (Lopez-Obando et al. 2016). Are these evolutionary processes common among *KAI2* paralogs? Do extra *KAI2* copies in non-parasitic angiosperms also evolve specificity for different signals? Currently, this question remains unaddressed.

An ideal non-parasitic angiosperm for studying the function of *KAI2* paralogs is the Grand Rapids cultivar of *Lactuca sativa* (lettuce), which is KAR-responsive in seed germination (Flematti et al. 2004). One of two KAI2 paralogs in lettuce has a phenylalanine at site 124 (site number refers to Arabidopsis KAI2), which is found in the KAR-responsive KAI2 paralog from *S. hermonthica*. The other KAI2 paralog from lettuce has tyrosine at site 124, which is characteristic of KL-responsive paralogs from parasites and of Arabidopsis KAI2 (Conn et al. 2015). The KAR-responsiveness of *L. sativa* and the amino acids at site 124 in its KAI2 copies raise the question of whether one paralog has specialized for KAR and the other for KL. The ease of obtaining and cultivating lettuce makes this a very straightforward question to address.

A recent extensive phylogenetic analysis of *KAI2* indicates that paralogs may have diversified in many other plant taxa as well. According to a new study by Bythell-Douglas et al. (2017), *KAI2* is present in basal plants, including some green algae, but *D14* is restricted to land plants. This conclusion is largely in agreement with previous phylogenetic analyses (e.g. Delaux et al. 2012, Waters et al. 2012, Lopez-Obando et al.

2016), although it places the *KAI2* duplication that generated *D14* earlier in land plant evolution. Unique to this new study is the division of most *KAI2* genes and homologs from land plants into two large clades. One is described as "eu-*KAI2*" and contains highly conserved *KAI2* orthologs from many land plant taxa, with the notable exception of hornworts (which the authors describe as a "problematic" clade). The other clade is labeled "*DDK*," and comprises *D14*, *DLK2*, and some divergent *KAI2* orthologs that were previously considered closer relatives of *KAI2* than of *D14*. (An example is some SL-responsive *KAI2* from *P. patens*, which had previously been thought to lack *D14*). Regarding *KAI2* evolution in land plants, this new study draws two important conclusions. First, it pushes the origin of *D14* (or at least of the *KAI2* duplication that eventually led to *D14*) further back in land plant evolution. Second, it describes the *eu-KAI2* clade in land plants as highly conserved, and the *DDK* clade as one of evolutionary innovation. Additional smaller, taxon-specific duplications within this phylogeny have produced an even greater array of diverse *KAI2* homologs, including some that may have lost the ability to interact with an important signaling partner. Even within the Zygnematales, which are green algae that diverged from the ancestor of land plants prior to the origin of *D14*, somewhat conserved *KAI2* genes and more divergent paralogs (*NOT KAI2*, or *NK2*) are both present (Bythell-Douglas et al. 2017). In summary, this study points to the tremendous diversity of *KAI2* and its homologs in the plant kingdom, and emphasizes the need for a greater understanding of *KAI2* function in non-model plant species.

The relationship between structure and function of KAI2 in diverse plant taxa: a preliminary study. *KAI2* paralogs in parasitic weeds have specialized to perceive different chemical signals; thus, parasites are an ideal system for studying the relationship between KAI2 structure and function. To determine how KAI2 evolves

specificity for KL, KAR, or SL, we combined several analyses to identify candidate SDPs in KAI2 proteins from parasites and non-parasitic relatives. We functionally assessed candidate sites by switching them in Arabidopsis KAI2 (AtKAI2) and transforming modified *AtKAI2* into Arabidopsis *kai2* or *d14 htl3* null mutants. (*HTL3* is another name for *KAI2*.) We assayed transgenic lines for responses to KL, KAR, and SL and found that targeted amino acid replacements partially recapitulated the KAR specificity of one parasite KAI2 protein. However, we were not able to create a SL-responsive version of KAI2 by modifying candidate SDPs.

To study *KAI2* evolution in a broader context of plant evolution, we analyzed *KAI2* paralogs derived from five separate duplication events. We identified amino acid replacements that may contribute to different functions in KAI2 paralogs or that have been acted on by positive selection. For the most part, different sites appeared to contribute to functional diversification or be positively selected after different *KAI2* duplication events. Thus, we did not observe compelling evidence of convergent molecular evolution among independently derived KAI2 paralogs. Finally, we tested the function of two *KAI2* paralogs from lettuce, and found that they appear to have evolved specificity for KL and for KAR, respectively. This subfunctionalization of *KAI2* paralogs in lettuce is similar to the evolution of SL-unresponsive *KAI2* in parasitic plants (Conn et al. 2015). Our characterization of *KAI2* from lettuce provides preliminary functional data on *KAI2* paralogs from a non-parasitic angiosperm. The commonness of *KAI2* duplication in plants and the growing list of functionally diverse paralogs suggest that Arabidopsis may not be an ideal system for studying this gene. By extending studies of *KAI2* evolution and function to more diverse plant species, we may uncover new insights into this important signaling system, and more generally, into fundamental processes of molecular evolution.

Methods

Identification of candidate SDPs. A visual analysis of amino acid conservation in the three clades of KAI2 proteins in Lamiids (conserved, intermediate, and divergent) was conducted to identify sites with high conservation within clades and high divergence across them. Sites that met this criterion and were predicted to be close to the ligand-binding pocket were selected as candidate SDPs. Although sites predicted to influence KL specificity (in conserved KAI2, or KAI2c), KAR sensitivity (in intermediate KAI2, or KAI2i), and SL responsiveness (in parasite-specific divergent KAI2, or KAI2d) were identified, only those relevant to KAI2c and KAI2i were noted in Tables 4.1 – 4.2 as KAI2cmod1 and KAI2imod2 modifications. Among these sites, 161 and 190 were further assessed for their individual effects on likely ligand specificity in KAI2. Sites 124 and 161 were also studied further for their combined influence on likely ligand preference.

To identify sites that contribute to SL responsiveness in KAI2d in parasites, a preliminary analysis combining several computational tools replaced the visual analysis of amino acid conservation. First, trimmed and aligned *KAI2* sequences from parasitic weeds were collected from Conn et al. 2015. *Striga asiatica KAI2* sequences were obtained from Satoko Yoshida, and additional *Striga hermonthica KAI2* sequences were provided by Tsuchiya et al. 2015. Predicted coding sequences were translated to generate amino acid sequences. Sequences were split into two sets of data, one specific for *Orobanche* and *Phelipanche KAI2,* and one specific for *Striga KAI2*. Sequences were aligned manually and trimmed at ends, if necessary, to maximize easily alignable sequence and minimize gaps. Nucleotide sequences were used to generate Bayesian phylogenies in MrBayes v3.2.5 (Ronquist and Huelsenbeck 2003), with two independent MCMC chains run for 20,000,000 generations. The initial 25% of trees were discarded as burn-in, and the remaining trees were used to build consensus trees. These

consensus trees were used as the basis of branch-site selection tests in PAML (Yang 2007), with the *KAI2d* clade set as foreground in each phylogeny. Positively selected sites were added to the list of candidate SDPs.

Predicted amino acid sequences in *Orobanche*/*Phelipanche*- and *Striga*-specific alignments were entered into JDet (Muth et al. 2012). Sequences were analyzed with default settings in S3Det, which divides sequences into predicted functional groups and identifies residues that may contribute to functional differences among those groups (Rausell et al. 2010). The *Orobanche*/*Phelipanche* KAI2 sequences were split into KAI2c and KAI2d; however, *Striga* KAI2 were split into three groups: KAI2c, KAI2i and some KAI2d, and the remaining KAI2d. Excluding sites that only differed between the two *Striga* groups containing KAI2d, sites selected by S3Det were added to a list of candidate SDPs.

From branch-site analyses in PAML, ancestral state reconstructions were obtained for nodes throughout the corresponding trees. Sites that differed between the last common ancestor of *KAI2d* and the last common ancestor of *KAI2c* in each tree were identified. These sites were the final additions to the preliminary list of candidate SDPs.

To filter the list of candidate SDPs, the AtKAI2 protein (PDB entry 4JYM) was visualized in PyMol in surface mode, with cavities and pockets culled. By coloring candidate target sites, those that occur near the ligand-binding pocket of AtKAI2 were identified; the list of candidate SDPs was trimmed to only include these.

Finally, Direct Coupling Analysis (DCA; Morcos et al. 2011) was used to identify sites that likely interact with candidate SDPs near the ligand-binding pocket. *Orobanche*/*Phelipanche* and *Striga*-specific KAI2 alignments were entered into the Rice University DCA webservice and analyzed with default settings. Sites predicted to interact with target ligand-binding pocket sites were added to the list of target sites if they had a

direct interaction (DI) value > 0.1. By combining candidate SDP prediction in JDET and PAML, candidate site filtration using PyMol, and interaction site identification with DCA, the list of target sites for modification in AtKAI2 was finalized.

Modification of AtKAI2. Complete (*AtKAI2cmod1* and *AtKAI2imod2*) or partial (*AtKAI2dmod6* and *AtKAI2dmod7*) coding sequences for all four AtKAI2 variants were synthesized as GBLOCKS by Integrated DNA Technologies (IDT). For *AtKAI2cmod1* and *AtKAI2imod2*, BP Clonase (Gateway, or GW) was used to insert genes into the entry vector pDONR221. In the case of *AtKAI2dmod6* and *AtKAI2dmod7*, only the 477 base pairs of *AtKAI2* including target mutations were synthesized. A pDONR221-AtKAI2 construct used by Conn et al. 2015 was linearized using primers listed in Table 4.5, which eliminate the internal 477 base pairs of *AtKAI2* coding sequence corresponding to those present in *AtKAI2dmod6* and *AtKAI2dmod7*. These 477-base-pair GBLOCKS were then inserted into the linearized construct by Gibson Assembly (New England BioLabs Inc.). To put single amino acid replacements in AtKAI2 (KAI2A161C and KAI2G190A), two halves of *KAI2cmod1* were amplified separately (one encoding the KAI2A161C replacement, the other encoding KAI2G190A). Wild-type *AtKAI2* cDNA was also amplified in two pieces using the same primers. Cycled ligation assembly was used to join each piece of wild-type *AtKAI2* cDNA with the opposite piece of *KAI2cmod1*. These modified versions of *AtKAI2*, each encoding a single amino acid replacement, were then inserted into pDONR221 with GW-compatible primers. Finally, to generate *AtKAI2Y124F/A161V*, a pDONR221-*AtKAI2* construct was linearized to exclude 140 internal base pairs of *AtKAI2* coding sequence. This internal piece was amplified with primers incorporating mutations that encode phenylalanine at site 124 and valine at site 161. The linearized pDONR221 vector and mutated internal portion of *AtKAI2* cDNA

were then joined together by cycled ligation assembly. All primers and scaffolds used are listed in Table 4.5.

Constructs were transformed into *E. coli*, confirmed by Sanger sequencing, and transferred to the destination vector pKAI2-GW, in which transgene expression is driven by the *AtKAI2* promoter (876 bp of sequence upstream of the *KAI2* start codon). pKAI2-GW constructs were transformed into *Agrobacterium tumefaciens*, and Agrobacterium-mediated transformation of *kai2-2* (*KAI2cmod1, KAI2imod2*) and *d14 htl3* (all other transgenes) mutants was done according to the floral dip method of Clough and Bent (1998). Transformants (T1s) were selected by plating on 1/2X Murashige & Skoog medium, supplemented with 15 – 25 µg/ml Hygromycin.

Plant growth and seedling assays. Plants were grown in long-day light (16 h light: 8 h dark) at 21 – 22 °C on Fafard 3B Mix soil, with Gnatrol (*Bacillus thuringiensis*) and Marathon (imidacloprid) treatments for pest control. Plants were randomized in flats to minimize the influence of maternal environment on subsequent assays. Plants were harvested when some siliques were brown. Plants were harvested into paper bags and dried at room temperature for 3 – 5 days. Seeds were then cleaned and either used or stored at -80°C.

For seedling assays, seeds were surface-sterilized in 70% ethanol with 0.05% Triton-X 100 for five minutes, subsequently rinsed with 70% and 95% ethanol, and air-dried on filter paper. Seeds were then plated on 1/2X Murashige and Skoog medium plus 0.8% agar, with pH adjusted to ~5.7. Medium was supplemented with KAR_1, KAR_2, *rac*-GR24, $GR24^{5DS}$, or $GR24^{ent\text{-}5DS}$ treatment, or with an equivalent volume of acetone as a control. KAR and GR24 treatments were diluted from 1000X stocks dissolved in acetone. Seedlings were grown for ten days in short-day white light conditions (8 h light : 16 h dark) at 21°C. Seedlings were then photographed, and hypocotyls and cotyledons

were measured in ImageJ (http://rsb.info.nih.gov/ij/). Statistical analyses were done in JMP (SAS Institute).

Transgenes in the *kai2-2* mutant background were tested in homozygous transgenic lines, with the exception of *KAI2cmod1* line 3, which was originally thought to be homozygous but in reality was not. However, only plants carrying the *KAI2cmod1* transgene were selected from line 3 for analysis. Transgenes in the *d14 htl-3* mutant background were tested at the T2 or T3 stage, at which point the transgene was still segregating.

A refined search for SDPs in KAI2. Full-length *KAI2* sequences from *Orobanche*, *Phelipanche*, and *Striga* were collected from sources described above, aligned manually, and trimmed at ends to minimize gaps and alignment ambiguity. Relative to the *AtKAI2* coding and protein sequences, codons and amino acids 6 – 269 were retained in the *Orobanche*/*Phelipanche*-specific alignment, and 8 – 268 were retained in the *Striga*-specific alignment. Nucleotide alignments were used to generate Bayesian trees, as described above.

Unlike before, an analysis of predicted protein sequences was carried out in JDet prior to PAML analyses. The *Orobanche*/*Phelipanche*- and *Striga*-specific *KAI2* alignments were translated to predicted amino acid alignments, separately entered into JDet, and analyzed with default settings in S3Det. Predicted functional residues were recorded in Table 4.1.

In preliminary analyses, *KAI2d* were used as foreground in branch-site analyses in PAML. This time, the predicted functional groups assigned by JDet were used to inform foreground choices for branch-site analyses. Positively selected sites detected by PAML in *Orobanche*/*Phelipanche*- and *Striga*-specific *KAI2* alignments were recorded in Table 4.1.

Ancestral state reconstructions were again obtained from PAML, and the ancestor of the KAI2d clade in *Orobanche*/*Phelipanche* and in *Striga* was compared with AtKAI2. Sites that differed between the KAI2d ancestor and AtKAI2 were recorded in Table 4.1.

As before, the AtKAI2 protein (PDB entry 4JYM) was modeled in PyMol, and the list of target sites was narrowed to include only those that appear to be in or near the ligand-binding pocket. DCA may require ~250 sequences for accurate results regarding protein coevolution (Morcos et al. 2011). This number is much greater than the number of sequences that we tested previously. To perform a more accurate DCA, additional *KAI2* sequences were obtained as described below and as listed in Table 4.3. These were aligned manually and trimmed. 230 sequences were present in this alignment, which was translated to a predicted protein alignment and analyzed by DCA with default settings. Sites predicted to interact with target ligand-binding pocket sites were recorded in Table 4.1 if they had one of the top 30 DI values, as suggested by Morcos et al. (2011).

Phylogenetic analysis of *KAI2* duplicates beyond parasites. Full-length *KAI2* sequences were obtained from Conn et al. (2015). *KAI2* sequences from *Physcomitrella patens* were retrieved from Lopez-Obando et al. (2016). As described previously, *KAI2* sequences from *S. asiatica* were obtained from Satoko Yoshida, and additional sequences from *S. hermonthica* were retrieved from Tsuchiya et al. (2015). Two *KAI2* sequences from *Lactuca sativa* were obtained from the Lettuce Genome Resource (https://lgr.genomecenter.ucdavis.edu/). Sequences were manually aligned and trimmed and used to build a Bayesian phylogeny, all as described above.

Based on the phylogenetic placement of *KAI2* duplicates in Figure 4.3, five taxonomic groups were chosen for further analysis of the evolution of *KAI2* paralogs.

Each of these groups – Brassicaceae, Euphorbiaceae, Fabaceae, *Physcomitrella patens*, and Lamiids (conserved and intermediate *KAI2* only) – appears to have undergone a separate duplication event, relative to other groups. Thus, a comparison of the evolution of *KAI2* duplicates in each of these groups was used to determine whether convergent molecular evolution had occurred.

First, additional *KAI2* sequences from species in the Brassicaceae, Euphorbiaceae, and Fabaceae were obtained from Phytozome (Goodstein et al. 2012) and from the 1000 Plants Initiative (1KP; http://onekp.com/) by BLASTing each database using *AtKAI2* as a query. If hits had ambiguous orthology, they were reciprocally BLASTed against The Arabidopsis Information Resource (TAIR; Berardini et al. 2015) and only retained if Arabidopsis KAI2 was the top hit. *KAI2* sequences from the Brassicaceae, Euphorbiaceae, and Fabaceae were put into family-specific alignments. *KAI2* sequences from *Physcomitrella patens* were retrieved from Lopez-Obando et al. 2016 and put into a species-specific alignment. Conserved and intermediate *KAI2* sequences from Lamiids were obtained from Conn et al. 2015, Tsuchiya et al. 2015, and Satoko Yoshida, and were put into a fifth taxon-specific alignment. Each set of sequences was aligned manually and trimmed to minimize gaps and alignment ambiguity.

Predicted protein alignments were entered into JDet and analyzed with default settings by S3Det. Sites selected by JDet as potential contributors to differences in protein function were noted in Table 4.4. Nucleotide alignments were used to build Bayesian phylogenies as described above. Functional groups predicted by JDet were set as foreground in branch-site analyses in PAML. Results of selection tests in PAML were recorded in Table 4.4. After both methods for identifying functionally important and/or positively selected sites (JDet and branch-site models in PAML) had been employed, site lists were compared across the five taxonomic groups.

Functional analysis of two *KAI2* paralogs from lettuce. The sequences of two *KAI2* paralogs from *Lactuca sativa* were obtained as described above. Both paralogs were amplified from genomic DNA (see Table 4.5 for primers), and GW cloning was used to insert each into the entry vector pDONR221. Transgenic lines were produced and screened as described above for AtKAI2 modifications. Transgenic lines were assayed with wild type (Columbia ecotype; Col), *d14*, and *d14 htl3* controls in seedling development, as described above.

Results

Some preliminary AtKAI2 modifications shift ligand preference. Although only one to three members of the conserved, intermediate, and divergent clades of *KAI2* in Lamiids have been functionally characterized, each clade is thought to preferentially respond to a different signal (Conn et al. 2015, Tsuchiya et al. 2015, Conn and Nelson 2016). Therefore, we conducted visual assessments and preliminary computational analyses of KAI2 proteins in Lamiids. Based on easily detectable patterns of amino acid conservation in KAI2c and KAI2i, we chose two and seven amino acid replacements, respectively, that we hypothesized contribute to KL and KAR preference in these clades. We switched amino acids 161 and 190 of AtKAI2 to match their identity in PaKAI2c, which appears to be KL-specific. This modified *AtKAI2* was named *KAI2cmod1*. We generated *AtKAI2imod2* to encode amino acids that match the KAR-responsive ShKAI2i at sites 96, 124, 143, 154, 161, 190, and 218 (Tables 4.1 – 4.2).

We transformed these two modified *KAI2* transgenes into an Arabidopsis *kai2-2* null mutant background (L*er* ecotype) and tested transgenic lines at the seedling stage. We found that the *KAI2cmod1* transgene conferred KL sensitivity, as all three lines carrying it had significantly shorter hypocotyls than *kai2-2* mutants. In two *KAI2cmod1*

lines, hypocotyls were significantly shortened by 1 μM KAR_1 and KAR_2 treatment (Figure 4.1A). These results supported our hypothesis that *KAI2cmod1* would be KL-responsive. However, we did not abolish KAR-responsiveness in *KAI2cmod1* as we hypothesized we would.

The average hypocotyl length on control plates was greater in all three lines carrying *KAI2imod2* than in *KAI2cmod1* lines and the L*er* control. This difference was significant between L*er* and two *KAI2imod2* lines and is indicative of diminished KL sensitivity. *KAI2imod2* was significantly responsive to KAR_1 but not to KAR_2, which enhances seedling light sensitivity more strongly than KAR_1 in wild-type Arabidopsis (Nelson et al. 2010). Thus, *KAI2imod2* appears to have higher specificity for a particular signal than AtKAI2 (Figure 4.1A).

We hypothesized that fewer than two and seven amino acid replacements in AtKAI2 may be sufficient to confer KL and KAR preference, respectively. We made two single amino acid modifications to AtKAI2, one in which only site 161 was switched to match PaKAI2c (KAI2A161C), and one in which only site 190 was switched to match PaKAI2c (KAI2G190A). We also made a version of AtKAI2 in which only sites 124 and 161 were switched to match ShKAI2i (KAI2Y124F/A161V; Tables 4.1 – 4.2). These three modified versions of *AtKAI2* were transformed into *d14 htl-3* mutants, which are in the Columbia (Col) ecotype. As transgenes, all three modified versions of *AtKAI2* conferred KL sensitivity, and hypocotyls were significantly shorter in *KAI2G190A* and *KAI2Y124F/A161V* lines than in *d14 htl3* double mutants. However, KAR responsiveness was not apparent for any of these. Unfortunately, the wild-type Arabidopsis control assayed with them also failed to respond to KAR (Figure 4.1B); however, another transgene assayed at the same time was responsive to KAR_1, suggesting that this chemical stock was at least somewhat effective. Nevertheless, without a significant response to KARs in the wild-type control, we cannot draw

conclusions about the responsiveness of *KAI2A161C*, *KAI2G190A,* or *KAI2Y124F/A161V* to KARs.

Next, we conducted a preliminary analysis of KAI2d from *Phelipanche aegyptiaca* and *Striga hermonthica* using several computational tools. By combining site selection in JDet and branch-site analysis and ancestral state reconstruction in PAML, we identified amino acids that may contribute to SL perception in KAI2d proteins in these two parasitic weeds. We modeled AtKAI2 in PyMol, and narrowed our list of candidate sites to include only those that likely occur in or near the ligand-binding pocket. We also included site 153, which is not predicted to directly connect to the ligand-binding pocket but was pointed out by Conn et al. (2015) as having high conservation outside of the KAI2d clade and high divergence within it. Finally, we searched for sites that may interact with candidate sites in or near the ligand-binding pocket by running a Direct Coupling Analysis (Morcos et al. 2011). Sites selected by these collective methods are indicated in Tables 4.1 – 4.2 as *KAI2dmod6* and *KAI2dmod7* modifications, and were switched in AtKAI2 to match their identity in the SL-responsive proteins PaKAI2d3 (KAI2dmod6) and ShKAI2d1 (KAI2dmod7). One exception is site 201, which is phenylalanine in PaKAI2d3 but was modified to leucine (a common amino acid at this site in predicted KAI2d proteins) in KAI2dmod6 (Table 4.1 – 4.2).

The *KAI2dmod6* and *KAI2dmod7* transgenes were also tested in *d14 htl-3* mutants. Hypocotyls were shortened in the line carrying *KAI2dmod7*, indicative of at least some KL sensitivity; however, *KAI2dmod6* did not appear to enhance seedling light sensitivity in this way. As with the assay of *KAI2A161C, KAI2G190A,* and *KAI2Y124F/A161V*, exogenous chemical treatments failed to enhance light sensitivity in the wild-type Col control (Figure 4.1C). Thus, we cannot comment on whether *KAI2dmod6* and *KAI2dmod7* respond to SLs.

A refined analysis of candidate SDPs reveals additional sites that may contribute to ligand specificity. We conducted a second search for SDPs in KAI2d from parasites, with the goal of optimizing a list of sites for future modification of AtKAI2. First, we entered predicted protein sequences into JDet, in which S3det divides amino acid sequences into predicted functional groups. Next, we built phylogenetic trees based on *Orobanche/Phelipanche* and *Striga*-specific nucleotide alignments. These trees approximate evolutionary relatedness based on a Bayesian analysis. Despite the differences in these two approaches, JDet and Bayesian phylogenies consistently split parasite KAI2c/*KAI2c* into one single group (group 1, black branches in Figure 4.2A). The remaining *Orobanche*/*Phelipanche* sequences were from the divergent clade, and these too formed one group in both JDet and phylogenetic analyses (group 2, orange branches; Figure 4.2B). In *Striga*, *KAI2i* and *KAI2d* formed a monophyletic clade in our Bayesian tree. Besides KAI2c, JDet created two additional groups of *Striga* KAI2. One group contained KAI2i and some KAI2d (group 2, orange branches), and the other contained the remaining KAI2d (group 3, green branches; Figure 4.2B). Interestingly, among the *Striga hermonthica KAI2* paralogs that have been functionally tested, those that confer germination responses as transgenes are clustered in group 2. Within this group are KAR-responsive KAI2i and SL-responsive KAI2d. Group 3 includes paralogs that interact with SLs *in vitro* but do not confer germination responses as transgenes (Conn et al. 2015, Toh et al. 2015, Tsuchiya et al. 2015). Thus, differences between the two groups of KAI2d proteins from *S. hermonthica* may contribute to processes other than ligand binding, such as interaction with protein partners.

JDet and branch-site analyses in PAML chose non-overlapping sets of amino acids in KAI2d in *Orobanche/Phelipanche*; similarly, there was no overlap in their site lists in *Striga* either. PAML also performs ancestral state reconstruction, and sequences at nodes of interest can be compared to sequences of proteins with known functions.

Because the *KAI2d* clade is parasite-specific and includes all known SL-responsive *KAI2* paralogs from parasites, we were specifically interested in the ancestor of this clade. According to ancestral state reconstruction, many sites differ between AtKAI2 and the ancestor of KAI2d in *Orobanche/Phelipanche* and in *Striga*; we added these sites to those selected by JDet and PAML's branch-site analyses (Table 4.1). We filtered this list of target sites so that only those predicted by PyMol to be in or near the ligand-binding pocket were analyzed further (Table 4.2). Using all 230 KAI2 sequences listed in Table 4.3, we performed a Direct Coupling Analysis (Morcos et al. 2011) to identify sites that likely interact with target sites that are close to the ligand-binding pocket. Our final list of target sites is indicated by X's in the next to last column of Table 4.2.

The target sites identified by this second analysis largely overlap with sites modified in KAI2dmod6 and KAI2dmod7. However, a few new sites were selected in our second round of analysis. Furthermore, some sites that had been selected for alteration in KAI2dmod6 and KAI2dmod7 were not detected by the more refined search for SDPs (Table 4.2). Perhaps the combination of AtKAI2 modifications suggested by these new and improved computational analyses would enable noticeable SL responses in transgenic lines. This possibility is currently untested.

Convergent molecular evolution is not apparent in *KAI2* paralogs following separate duplication events. Next, we took a broader view of *KAI2* evolution in dicots by building the Bayesian tree shown in Figure 4.3. According to this phylogeny, independent *KAI2* duplications have occurred in at least five taxonomic groups. The duplication that generated conserved and intermediate *KAI2* in Lamiids is represented in the "Asterids" clade at the top of the tree. Two *KAI2* paralogs from lettuce – which is the only non-Lamiid Asterid species in the tree – cluster with conserved *KAI2* from Lamiids. Their phylogenetic placement suggests that they arose from a lettuce-specific duplication, but

additional non-Lamiid Asterid species would need to be analyzed to determine exactly when this duplication occurred.

A separate duplication is apparent in the Fabales. *KAI2* from this order do not form a monophyletic clade; *Glycine max 5* and *Phaseolus vulgaris 3* are positioned outside of all other dicot *KAI2*. These two sequences were identified as *KAI2* orthologs by Conn et al. (2015), who used a reciprocal BLAST method to determine which sequences to include in their *KAI2* phylogeny. By this criterion, *G. max 5* and *P. vulgaris 3* are classified as *KAI2* rather than *D14* or *DLK2.* However, the recent analysis by Bythell-Douglas et al. suggests that genes considered to be *D14* or *DLK2* orthologs are part of a larger monophyletic clade of divergent *KAI2* paralogs, called *DDK*. The *DDK* clade is separate from "eu-*KAI2*" genes, which have higher conservation (Bythell-Douglas et al. 2017). Thus, even though *KAI2* is the top BLAST hit in TAIR for G*. max 5* and *P. vulgaris 3*, they may belong to *DDK* rather than eu-*KAI2*.

Independent *KAI2* duplications were also apparent in the Malpighiales and Brassicales, as well as in *P. patens.* Interestingly, *KAI2* paralogs in *P. patens* fall into different clades with different evolutionary rates, and some have transcriptional responses to SL treatment (Lopez-Obando et al. 2016). Thus, SL sensitivity may have convergently evolved not just in D14 and in parasite KAI2d, but in KAI2 from *P. patens* as well. Because of this convergent evolution of *KAI2* function, and because *KAI2* has apparently evolved preferences for ecologically relevant signals in different plant species, we hypothesized that convergent molecular evolution might have acted on the KAI2 ligand-binding pocket. Although the function of most of the *KAI2* duplicates in Figure 4.3 is unknown, we set out to determine whether convergent molecular evolution had occurred after independent *KAI2* duplication events.

To study the evolution of duplicate genes, we first collected additional *KAI2* sequences from Brassicaceae (Brassicales), Euphorbiaceae (Malpighiales), and

Fabaceae (Fabales), because these three families were well represented in Figure 4.3. We made taxon-specific nucleotide and protein alignments for five groups in which independent *KAI2* duplications were apparent: Brassicaceae, Euphorbiaceae, Fabaceae, *P. patens*, and Lamiids. We analyzed each alignment by JDet, and by building Bayesian phylogenies and using them as the basis of branch-site tests in PAML (Figures 4.4 – 4.8). Branch-site tests chose unique positively selected sites in specified sequences from each taxonomic group; however, no sites were under positive selection in multiple taxonomic groups. Similarly, JDet mostly selected unique sites in each group; however, a few sites were selected in multiple groups (Table 4.4). This apparent lack of convergent molecular evolution could be due to *KAI2* duplicates evolving specificity for different signals in different plant taxa. Functional data on *KAI2* paralogs in diverse plant species are needed to determine whether there are any consistent patterns of evolution following *KAI2* duplication events.

Two *KAI2* paralogs in lettuce may have subfunctionalized. To investigate the function of *KAI2* paralogs from a non-parasitic angiosperm, we tested two *KAI2* copies from lettuce in Arabidopsis. One (LsKAI2B) has a phenylalanine at site 124 and a valine at site 161 (site number refers to AtKAI2), which were two of the seven modifications made in the KAR-responsive *KAI2imod2* transgene. The other paralog from lettuce (LsKAI2A) matches AtKAI2 at these sites. Because lettuce is KAR-responsive in seed germination (Flematti et al. 2004), we hypothesized that *LsKAI2B* would preferentially respond to KAR, while *LsKAI2A* would be more sensitive to KL. Indeed, a transgenic line carrying *LsKAI2A* has hypocotyls that are significantly shorter than *d14 htl3* mutants but not significantly different from Col. This phenotype is consistent with sensitivity to KL. Two *d14 htl3* [*LsKAI2B*] lines had significantly longer hypocotyls on control medium than Col, but their light sensitivity increased in response to KAR_1. This response to KAR_1

treatment was significant in one of two lines tested, although only for a low concentration. The seedlings measured in this assay were not homozygous for *LsKAI2B*; thus the statistically insignificant responses to KAR treatment observed in both lines likely underrepresent the full effect of this transgene. In this assay, the wild type control (Col) failed to respond to KAR treatment, so we cannot rule out that *LsKAI2A* is KAR-responsive, or that *LsKAI2B* can respond to KAR_2 as well as KAR_1 (Figure 4.9). The lack of KAR response in Col could indicate that the KAR stocks used were ineffective; however, these same stocks were used to test lines in the L*er* background, which responded to both KAR variants as expected. Thus we can at least conclude that *LsKAI2B* prefers KAR_1, similar to *KAI2imod2* and *ShKAI2i* (Conn et al. 2015, Conn and Nelson 2016). In general, the Col ecotype seems to have higher light sensitivity than the L*er* ecotype; thus, lines in this background may simply require different conditions for assaying KL and KAR responsiveness.

Discussion

Throughout the plant kingdom, the diversity of *KAI2* has been largely unexplored. This gene has been studied in a few select taxa, including the model angiosperm *Arabidopsis thaliana*, the basal moss *Physcomitrella patens*, and parasitic weeds of the Orobanchaceae. Even in these species, the function of *KAI2* is still not completely understood. We know the most about *KAI2* in Arabidopsis, but one of three signals that it perceives, the endogenous regulator KL, has yet to be identified. In *P. patens*, some *KAI2* are known to respond transcriptionally to SL, but what processes these genes control in moss is unclear. In parasitic plants, *KAI2* paralogs have been found to respond to different chemical cues, including SLs; however, how parasite KAI2d have evolved sensitivity to these host hormones is not known.

The work presented here addresses the question of how molecular evolution in duplicate genes enables responsiveness to different signals. *KAI2* paralogs in parasitic weeds provide an ideal system for this investigation because of their apparent preferences for KL, KAR, or SL. By identifying candidate SDPs in parasite SL-responsive KAI2d and testing their function in Arabidopsis, we took another step forward in understanding the evolution of host-responsive seed germination. Although our targeted amino acid mutations did not make AtKAI2 an SL responder, we conducted a second round of SDP prediction that may have improved our list of target sites. A functional analysis of these sites should be the focus of future work.

We did successfully identify amino acid replacements that contribute to KAR specificity, as evidenced by the phenotypes conferred by *KAI2imod2*. Although we may not have completely eliminated KL sensitivity by modifying AtKAI2, we now have an understanding of how specialization for one of these two signals can evolve. Interestingly, two of the seven amino acid replacements that we made in *KAI2imod2* are present in one *KAI2* paralog from lettuce (*LsKAI2B*). Both *KAI2imod2* and *LsKAI2B* can confer a significant response to KAR_1 but not KAR_2, at least as transgenes in seedlings. In contrast, wild type Arabidopsis is more sensitive to KAR_2 at this stage (Nelson et al. 2010). Of all the KAR structural variants, KAR_1 is the most abundant and often has the greatest effect on seed germination among KAR-responsive plant species (Nelson et al. 2012), and over 60 different plants respond positively to KAR_1 in seed germination (Chiwocha et al. 2009). The commonness of KAR_1 and its stronger effect on seed germination than other KARs suggest that preference for KAR_1 may be adaptive. Perhaps *LsKAI2B* and *ShKAI2i* are more responsive to KAR_1 than to KAR_2 because specialization for KAR_1 benefits fire-followers (which, for *S. hermonthica*, would have to be a fire-following ancestor). If this is the case, the preference of Arabidopsis for KAR_2 could be further evidence that its KAR sensitivity is coincidental. Alternatively,

Arabidopsis may not be under strong selective pressure to maintain a KAR_1 preference. As evidenced by *KAI2imod2*, evolving higher sensitivity to one KAR variant over another requires no more than seven amino acid replacements. Substitutions at sites 124 and 161, to phenylalanine and valine, respectively, are the only ones in common between KAI2imod2 and LsKAI2B, relative to AtKAI2. We hypothesized that these alone might be sufficient to diminish KL responsiveness and confer a preference for KAR_1, but KAI2Y124F/A161V appeared to function similarly to wild type AtKAI2. Thus, although we have begun to identify the molecular basis of likely ligand preference in KAI2, many questions remain unanswered.

Like lettuce, many non-parasitic plants outside of the Lamiids have multiple *KAI2* genes; however, we still know very little about the function of *KAI2* paralogs in plants. Our analyses of molecular evolution after independent *KAI2* duplications showed no clear patterns of convergence. For example, sites under positive selection in intermediate *KAI2* in the Lamiids did not appear to be positively selected in *KAI2* paralogs from other duplication events. Despite this apparent lack of convergent molecular evolution, *KAI2* duplicates have convergently evolved in function several times. The SL-responsive genes *D14* and parasite *KAI2d* arose from separate *KAI2* duplications, and the proteins they encode have different amino acids at several key positions near the ligand-binding pocket. Nevertheless, the ligand-binding pockets of D14 and KAI2d resemble each other on a larger scale (Conn et al. 2015). The SL-responsive KAI2 paralogs in *P. patens* are more of a mystery. While some may have extra volume in their ligand-binding pockets, similar to D14, others that may be involved in SL signaling do not resemble D14 (Lopez-Obando et al. 2016). This suggests that KAI2 proteins can undergo different structural changes but evolve sensitivity to the same signals. Thus, although *KAI2* duplicates do not exhibit convergent molecular evolution, they may evolve similar functions.

Investigating the functional diversity of *KAI2* in non-model plants is not a trivial task, especially when multiple copies that may be functionally redundant are present. Our analyses of KAI2 duplicates in JDet suggest that paralogs may often have different functions, but what processes they control and which signals they respond to are unknown. Although a complete survey of *KAI2* function in plants is impractical, the commonness of *KAI2* duplication suggests that Arabidopsis may not be an ideal model for studying this signaling system. In addition to the function of *KAI2* paralogs, little is known about the mechanisms of *KAI2* duplication or about when and where paralogs are expressed. Future work should address these questions in representatives of diverse plant taxa.

Another focus of future work should be the molecular mechanism of SL perception in parasites. Among SL-responsive KAI2d from *S. hermonthica*, some preferences for different SL structural variants have been observed *in vitro* (Tsuchiya et al. 2015). While the structural basis for differences in SL preference is unknown, this information would lay the groundwork for studies of host/parasite specificity. Have KAI2d paralogs in generalists evolved to perceive the SL profiles of a wide range of hosts? Have paralogs been lost in specialists because perceiving extra hosts is maladaptive? Answering these questions requires extensive sampling of parasite species and populations; however, the answers may provide valuable information about genetic contributors to host range.

Overall, this work explores gene duplication and subsequent evolution in a signaling system found throughout much of the plant kingdom. Our investigation of how parasite KAI2 paralogs evolved preferences for different signals addresses broader questions about patterns of molecular evolution in the context of parasitism. Our analysis of independent *KAI2* duplications investigates whether convergent evolution occurs on the molecular level, as it sometimes does on the functional level among *KAI2* paralogs.

Finally, our characterization of two *KAI2* copies from lettuce provides new functional data about *KAI2* paralogs in a non-parasitic angiosperm. Many questions about *KAI2* evolution and function remain unanswered, but this study provides a starting point for future investigations of the diversity of *KAI2* signaling.

Table 4.1. Overview of candidate specificity-determining positions in KAI2. Sites identified by preliminary analyses are identified as KAI2cmod1, KAI2imod2, KAI2dmod6, and KAI2dmod7 modifications. Results of a refined analysis of candidate SDPs in KAI2d from parasites are indicated in the right-most 8 columns. Foreground specifications in *Striga*-specific analyses are explained in Figure 4.2B.

		Sites identified by preliminary analyses				Sites identified by refined analysis							
Site number (AtKAI2)	Site ID (AtKAI2)	KAI2cmod1 modification	KAI2imod2 modification	KAI2dmod6 modification	KAI2dmod7 modification	Jdet (*Orobanche, Phelipanche*)	Branch-site model (*Orobanche, Phelipanche*)	Ancestral state reconstruction (*Orobanche* and *Phelipanche*)	Jdet (*Striga*)	Branch-site model (*Striga*)	Ancestral state reconstruction (*Striga*)	Ligand-binding pocket site?*	DCA
1	M												
2	G												
3	V												
4	V												
5	E												
6	E							X					
7	A												
8	H												
9	N												
10	V												
11	K							X			X		
12	V												
13	I							X			X		
14	G												
15	S												
16	G												
17	E												
18	A							X			X		
19	T												
20	I							X			X		
21	V												
22	L												
23	G												
24	H												
25	G												
26	F									***X*** (foreground B)			
27	G												
28	T												
29	D												
30	Q												
31	S												
32	V												
33	W												
34	K												

35	H							X					
36	L												
37	V												
38	P												
39	H												
40	L												
41	V												
42	D												
43	D							X			X		
44	Y												
45	R												
46	V												
47	V								X		X		
48	L												
49	Y								X*				
50	D												
51	N												
52	M									X (foreground B)			
53	G												
54	A												
55	G												
56	T												
57	T												
58	N								X**				
59	P												
60	D												
61	Y						X						
62	F												
63	D												
64	F												
65	D							X			X		
66	R												
67	Y												
68	S							X					
69	N							X			X		
70	L								X**				
71	E												
72	G												
73	Y												
74	S							X			X		
75	F							X			X		
76	D												
77	L					X*			X*				
78	I												
79	A												
80	I												
81	L												
82	E												
83	D							X			X		

84	L					X		X			X		
85	K							X			X		
86	I										X		
87	E					X		?			X		
88	S							?			X		
89	C												
90	I												
91	F							X			X		
92	V												
93	G												
94	H												
95	S												
96	V		L	L	L			X			X	X	
97	S												
98	A												
99	M								X**				
100	I					X		X			X		
101	G				A						X		X (121)
102	V							X			X		
103	L							X			X		X (198)
104	A												
105	S												
106	L							X			X		
107	N			L				X			X		
108	R									X (foreground B)			
109	P												
110	D									***X*** (foreground B)			
111	L												
112	F												
113	S							X			X		
114	K												
115	I							X			X		
116	V			I				X			X		
117	M								X*				
118	I												
119	S												
120	A												
121	S				T						X	X	X (101, 171)
122	P												
123	R												
124	Y		F	S	M			X			X	X	X (196)
125	V						X	X			X		
126	N												
127	D							X	X		X		
128	V							X			X		
129	D												
130	Y												
131	Q							X			X		
132	G												
133	G												

134	F						X					X	
135	E												
136	Q												
137	E							X					
138	D									X (foreground B)	X		
139	L						X				X (foreground B only)	X	
140	N							X			X		
141	Q												
142	L												
143	F		A			X		X		X (foreground AB)	X		
144	E							X			X		
145	A												
146	I							X		***X*** (foreground B)	X		
147	R							X			X		
148	S					X		X			X		
149	N												
150	Y						X						
151	K												
152	A							X			X		
153	W			L	L			X			X		
154	C		V					X			X		X (157)
155	L							X			X		
156	G												
157	F			M				X				X	X (154)
158	A												
159	P									X (foreground B)			
160	L												
161	A	C	V		M			X			X	X	
162	V							X					
163	G												
164	G												
165	D								X**				
166	M										X		
167	D												
168	S												
169	I							X			X		
170	A												
171	V										?		X (121)
172	Q												
173	E												
174	F												
175	S												
176	R												

177	T												
178	L												
179	F									***X*** (foreground B)			
180	N									X (foreground B)			
181	M												
182	R									X (foreground B)			
183	P												
184	D												
185	I												
186	A									***X*** (foreground B)			
187	L												
188	S									***X*** (foreground B)			
189	V										?		
190	G	A	F	V			X	X			?		
191	Q					X		X			?		
192	T									X (foreground B)	?		
193	I												
194	F			H	H		X	X				X	
195	Q							X			X		
196	S			Y	L			X			X	X	X (124)
197	D												
198	M				L				X		X	X	X (201)
199	R												
200	Q							X			X		
201	I			L				X	X		X		X (198)
202	L												
203	P							X			X		
204	F							X			X		
205	V												
206	T												
207	V												
208	P												
209	C												
210	H												
211	I												
212	L							X			X		
213	Q							X					
214	S												
215	V			C				X			X		
216	K												

217	D												
218	L		M	W			X				X	X	
219	A			V				X			X	X	
220	V												
221	P												
222	V							X					
223	V							X			X		
224	V												
225	S							X			X		
226	E												
227	Y												
228	L												
229	H												
230	A						X	X		X (foreground A)	X		
231	N								X*		X		
232	L												
233	G												
234	C							X			X		
235	E							X			X		
236	S												
237	V							X					
238	V												
239	E												
240	V												
241	I							X			X		
242	P												
243	S						X	X			X		
244	D							X			X		
245	G												
246	H												
247	L									X (foreground B)			
248	P												
249	Q							X			X		
250	L												
251	S												
252	S							X					
253	P												
254	D							X			X		
255	S							X			X		
256	V							X			X		
257	I												
258	P												
259	V												
260	I							X			X		
261	L												
262	R												
263	H												

264	I												
265	R							X					
266	N							X			X		
267	D												
268	I												
269	A												
270	M												
						*KAI2d largely match AtKAI2		? indicates that reconstruction was done in area of the alignment that contains gaps; thus, results are unreliable.	*many KAI2d match AtKAI2 **site differs primarily in group of some KAI2d paralogs, and is largely the same in KAI2c, KAI2i, and the KAI2d that group with KAI2i		? indicates that reconstruction was done in area of the alignment that contains gaps; thus, results are unreliable.	*All target sites are indicated with gray shading. Those predicted to influence the ligand-binding pocket are marked with an X.	Sites in parentheses are predicted to interact with sites indicated by X.

Table 4.2. Condensed overview of candidate SDPs in KAI2. Only sites chosen as candidate targets are shown. Those identified by preliminary analyses are indicated as KAI2cmod1, KAI2imod2, KAI2dmod6, or KAI2dmod7 modifications. Sites picked by a refined analysis are also indicated, alongside information on how they were chosen.

		Sites identified by preliminary analyses				Sites identified by refined analysis	
Site number (AtKAI2)	Site ID (AtKAI2)	KAI2cmod1 modification	KAI2imod2 modification	KAI2dmod6 modification	KAI2dmod7 modification	Target site?	Identification method
96	V		L	L	L	X	ASR, PyMol
101	G				A	X	DCA
103	L					X	DCA
107	N			L			
116	V			I			
121	S				T	X	ASR (*Striga* only), PyMol, DCA
124	Y		F	S	M	X	ASR, PyMol, DCA
134	F					X	BS (*Orobanche*/*Phelipanche* only), PyMol
139	L					X	BS (*Orobanche*/*Phelipanche* only), ASR (*Striga* only), PyMol
143	F		A				
153	W			L	L		
154	C		V			X	DCA
157	F			M		X	ASR (*Orobanche*/*Phelipanche* only), PyMol, DCA
161	A	C	V		M	X	ASR
171	V					X	DCA
190	G	A	F	V			
194	F			H	H	X	BS (*Orobanche*/*Phelipanche* only), ASR (*Orobanche*/*Phelipanche* only), PyMol
196	S			Y	L	X	ASR, PyMol, DCA
198	M				L	X	JDet (*Striga* only), ASR (*Striga* only), PyMol, DCA
201	I			L		X	DCA
215	V			C			
218	L		M	W		X	BS (*Orobanche*/*Phelipanche* only), ASR (*Striga* only), PyMol
219	A			V		X	ASR, PyMol
							ASR, ancestral state reconstruction; BS, branch-site modeling; DCA, direct coupling analysis

Table 4.3. Information on sequences used for phylogenetic analyses and JDet. Sequences were obtained from Conn et al. 2015, Tsuchiya et al. 2015, Lopez-Obando et al. 2016, Satako Yoshida, the 1000 Plants Initiative (www.onekp.com/), the Lettuce Genome Resource (https://lgr.genomecenter.ucdavis.edu/), and Phytozome (Goodstein et al. 2012).

Taxon in tree	Trees containing taxon	Order	Family	Source	Sequence ID
Acacia argyrophylla	Figure 4.6	Fabales	Fabaceae	1KP	ZCDJ-2012140
Agastache rugosa	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Agastache rugosa 2	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Ajuga reptans	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Ajuga reptans 2	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Apios americana	Figure 4.6	Fabales	Fabaceae	1KP	NXOH-2051076
Apios americana 2	Figure 4.6	Fabales	Fabaceae	1KP	NXOH-2051629
Aquilegia caerulea	Figure 4.3	Ranunculales	Ranunculaceae	Conn et al. 2015	
Aquilegia caerulea 2	Figure 4.3	Ranunculales	Ranunculaceae	Conn et al. 2015	
Aquilegia caerulea 3	Figure 4.3	Ranunculales	Ranunculaceae	Conn et al. 2015	
Arabidopsis lyrata	Figure 4.3, Figure 4.4	Brassicales	Brassicaceae	Conn et al. 2015, Phytozome 12	944104CDS
Arabidopsis thaliana	Figure 4.3, Figure 4.4	Brassicales	Brassicaceae	Conn et al. 2015, Phytozome 12	At4G37470.1CDS
Arabis alpina	Figure 4.4	Brassicales	Brassicaceae	1KP	TZWR-2043906
Astragalus membranaceus	Figure 4.6	Fabales	Fabaceae	1KP	HJMP-2003953
Astragalus membranaceus 2	Figure 4.6	Fabales	Fabaceae	1KP	HJMP-2074695
Astragalus propinquus	Figure 4.6	Fabales	Fabaceae	1KP	MYMP-2007503
Astragalus propinquus 2	Figure 4.6	Fabales	Fabaceae	1KP	MYMP-2060717
Bauhinia tomentosa	Figure 4.6	Fabales	Fabaceae	1KP	JETM-2024684
Bituminaria bituminosa	Figure 4.6	Fabales	Fabaceae	1KP	TVSH-2008472
Boechera stricta	Figure 4.4	Brassicales	Brassicaceae	Phytozome 12	Bostr.30440s0001.1CDS
Brassica nigra	Figure 4.4	Brassicales	Brassicaceae	1KP	IPWB-2015482
Brassica nigra 2	Figure 4.4	Brassicales	Brassicaceae	1KP	IPWB-2016980
Brassica rapa	Figure 4.3, Figure 4.4	Brassicales	Brassicaceae	Phytozome 12	Brara.H01722.1CDS
Brassica rapa 2	Figure 4.3, Figure 4.4	Brassicales	Brassicaceae	Phytozome 12	Brara.K01217.1CDS
Brassica rapa 3	Figure 4.3, Figure 4.4	Brassicales	Brassicaceae	Phytozome 12	Brara.E00360.1CDS
Byblis gigantea	Figure 4.3, Figure 4.8	Lamiales	Byblidaceae	Conn et al. 2015	
Capsella grandiflora	Figure 4.4	Brassicales	Brassicaceae	Phytozome 12	Cagra.1232s0005.1CDS
Capsella rubella	Figure 4.3, Figure 4.4	Brassicales	Brassicaceae	Conn et al. 2015, Phytozome 12	Carubv10005485mCDS
Carica papaya	Figure 4.3	Brassicales	Cariacaceae	Conn et al. 2015	
Cercis canadensis	Figure 4.6	Fabales	Fabaceae	1KP	RKFX-2041665
Cercis canadensis 2	Figure 4.6	Fabales	Fabaceae	1KP	RKFX-2046260

Citrus clementina	Figure 4.3	Sapindales	Rutaceae	Conn et al. 2015	
Citrus sinensis	Figure 4.3	Sapindales	Rutaceae	Conn et al. 2015	
Cochlearia officinalis	Figure 4.4	Brassicales	Brassicaceae	1KP	CSUV-2062621
Codariocalyx motorius	Figure 4.6	Fabales	Fabaceae	1KP	SUAK-2038671
Codariocalyx motorius 2	Figure 4.6	Fabales	Fabaceae	1KP	SUAK-2040388
Coffea arabica	Figure 4.3, Figure 4.8	Gentianales	Rubiaceae	Conn et al. 2015	
Conopholis americana KAI2c	Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Copaifera officinalis	Figure 4.6	Fabales	Fabaceae	1KP	RKLL-2012448
Copaifera officinalis 2	Figure 4.6	Fabales	Fabaceae	1KP	RKLL-2065454
Croton tiglium	Figure 4.5	Malpighiales	Euphorbiaceae	1KP	VVPY-2064675
Cucumis sativus	Figure 4.3	Cucurbitales	Cucurbitaceae	Conn et al. 2015	
Cucumis sativus 2	Figure 4.3	Cucurbitales	Cucurbitaceae	Conn et al. 2015	
Desmanthus illinoensis	Figure 4.6	Fabales	Fabaceae	1KP	XOOE-2029459
Draba hispida	Figure 4.4	Brassicales	Brassicaceae	1KP	GTSV-2001102
Draba oligosperma	Figure 4.4	Brassicales	Brassicaceae	1KP	LAPO-2002358 + LAPO-2042725
Draba ossetica	Figure 4.4	Brassicales	Brassicaceae	1KP	LJQF-2058545
Draba sachalinensis	Figure 4.4	Brassicales	Brassicaceae	1KP	BXBF-2009756
Euphorbia mesembryanthemifolia	Figure 4.5	Malpighiales	Euphorbiaceae	1KP	LSLA-2054901
Euphorbia pekinensis	Figure 4.5	Malpighiales	Euphorbiaceae	1KP	PXYR-2003722
Eutrema salsugineum	Figure 4.3, Figure 4.4	Brassicales	Brassicaceae	Conn et al. 2015, Phytozome 12	Thhalv10025969mCDS
Gleditsia sinensis	Figure 4.6	Fabales	Fabaceae	1KP	VHZV-2009289
Gleditsia sinensis 2	Figure 4.6	Fabales	Fabaceae	1KP	VHZV-2016835
Gleditsia triacanthos	Figure 4.6	Fabales	Fabaceae	1KP	GEHT-2058807
Glycine max	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Glyma.01G191200.1CDS
Glycine max 2	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Glyma.05G102800.1CDS
Glycine max 3	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Glyma.17G164500.1CDS
Glycine max 4	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Glyma.11G051000.1CDS
Glycine max 5	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Glyma.17G164400.1CDS
Glycyrrhiza glabra	Figure 4.6	Fabales	Fabaceae	1KP	PEZP-2012713
Glycyrrhiza glabra 2	Figure 4.6	Fabales	Fabaceae	1KP	PEZP-2012714
Glycyrrhiza glabra 3	Figure 4.6	Fabales	Fabaceae	1KP	PEZP-2068435
Glycyrrhiza glabra 4	Figure 4.6	Fabales	Fabaceae	1KP	PEZP-2012715
Glycyrrhiza lepidota	Figure 4.6	Fabales	Fabaceae	1KP	JTQQ-2015263
Glycyrrhiza lepidota 2	Figure 4.6	Fabales	Fabaceae	1KP	JTQQ-2069055
Glycyrrhiza lepidota 3	Figure 4.6	Fabales	Fabaceae	1KP	JTQQ-2069351
Gompholobium polymorphum	Figure 4.6	Fabales	Fabaceae	1KP	VLNB-2014376
Gompholobium polymorphum 2	Figure 4.6	Fabales	Fabaceae	1KP	VLNB-2014518
Gymnocladus dioicus	Figure 4.6	Fabales	Fabaceae	1KP	QZXQ-2092549
Lactuca sativa A	Figure 4.3	Asterales	Asteraceae	Lettuce Genome Resource	Lsat_1_v4_lg_4:361607540..361608739
Lactuca sativa B	Figure 4.3	Asterales	Asteraceae	Lettuce Genome Resource	Lsat_1_v4_lg_4:361560640..361561729

Lathyrus sativus	Figure 4.6	Fabales	Fabaceae	1KP	KNMB-2013129
Lathyrus sativus 2	Figure 4.6	Fabales	Fabaceae	1KP	KNMB-2013897
Lindenbergia philippensis KAI2c	Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Lindenbergia phillipensis KAI2i	Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Lotus japonicus	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015	
Lupinus angustifolius	Figure 4.6	Fabales	Fabaceae	1KP	TTRG-2096913
Lupinus polyphyllus	Figure 4.6	Fabales	Fabaceae	1KP	CMFF-2014779
Lupinus polyphyllus 2	Figure 4.6	Fabales	Fabaceae	1KP	CMFF-2071837
Malus domestica	Figure 4.3	Rosales	Rosaceae	Conn et al. 2015	
Manihot grahamii	Figure 4.5	Malpighiales	Euphorbiaceae	1KP	XNLP-2015130
Manihot esculenta	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Conn et al. 2015, Phytozome 12	Manes.13G114700.1 CDS
Manihot esculenta 2	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Conn et al. 2015, Phytozome 12	Manes.12G112700.1 CDS
Manihot esculenta 3	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Conn et al. 2015, Phytozome 12	Manes.12G112600.1 CDS
Manihot esculenta 4	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Conn et al. 2015, Phytozome 12	Manes.13G114300.1 CDS
Manihot esculenta 5	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Conn et al. 2015, Phytozome 12	Manes.13G114200.1 CDS
Manihot esculenta 6	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Conn et al. 2015, Phytozome 12	Manes.13G114500.1 CDS
Manihot esculenta 7	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Phytozome 12	Manes.13G114400.1 CDS
Marrubium vulgare	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Medicago truncatula	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Medtr4g095310.1CDS
Medicago truncatula 2	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Medtr5g016150.1CDS
Melissa officinalis	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Micromeria fruticosa	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Mimulus guttatus	Figure 4.3, Figure 4.8	Lamiales	Phrymaceae	Conn et al. 2015	
Mimulus guttatus 2	Figure 4.3, Figure 4.8	Lamiales	Phrymaceae	Conn et al. 2015	
Mimulus guttatus 3	Figure 4.3, Figure 4.8	Lamiales	Phrymaceae	Conn et al. 2015	
Nicotiana benthamiana	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	
Nicotiana tabacum	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	
Nicotiana tabacum 2	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	
Nicotiana tabacum 3	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	

Orobanche cernua KAI2c	Figure 4.2B, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cernua KAI2d1	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cernua KAI2d2	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cernua KAI2d3	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cumana KAI2c	Figure 4.2B, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cumana KAI2d1	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cumana KAI2d2	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cumana KAI2d3	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cumana KAI2d4	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cumana KAI2d5	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche cumana KAI2d6	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche fasciculata KAI2c	Figure 4.2B, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche fasciculata KAI2d	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche minor KAI2c	Figure 4.2B, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche minor KAI2d1	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche minor KAI2d2	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche minor KAI2d3	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche minor KAI2d4	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Orobanche minor KAI2d5	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Oxera neriifolia	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Oxera neriifolia 2	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Oxera neriifolia 3	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Oxera pulchella	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	

Oxera pulchella 2	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Paulownia fargesii	Figure 4.3, Figure 4.8	Lamiales	Paulowniaceae	Conn et al. 2015	
Paulownia fargesii 2	Figure 4.3, Figure 4.8	Lamiales	Paulowniaceae	Conn et al. 2015	
Phaseolus vulgaris	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Phvul.003G237900.1CDS
Phaseolus vulgaris 2	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Phvul.002G013800.1CDS
Phaseolus vulgaris 3	Figure 4.3, Figure 4.6	Fabales	Fabaceae	Conn et al. 2015, Phytozome 12	Phvul.003G237600.1CDS
Phelipanche aegyptiaca KAI2c	Figure 4.2B, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Phelipanche aegyptiaca KAI2d1	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phelipanche aegyptiaca KAI2d2	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phelipanche aegyptiaca KAI2d3	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phelipanche aegyptiaca KAI2d4	Figure 4.2B, Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phtheirospermum japonicum KAI2c	Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Phtheirospermum japonicum KAI2d1	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phtheirospermum japonicum KAI2d2	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phtheirospermum japonicum KAI2d3	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phtheirospermum japonicum KAI2d4	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phtheirospermum japonicum KAI2d5	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Phtheirospermum japonicum KAI2i	Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Physcomitrella patens D14L-B	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-C	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-D	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-E	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-F	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-G	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-H	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-I	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-J	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	

Physcomitrella patens D14L-K	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Physcomitrella patens D14L-L	Figure 4.3, Figure 4.7	Funariales	Funariaceae	Lopez-Obando et al. 2016	
Pogostemon sp.	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Populus trichocarpa	Figure 4.3	Malpighiales	Salicaceae	Conn et al. 2015	
Populus trichocarpa 2	Figure 4.3	Malpighiales	Salicaceae	Conn et al. 2015	
Prunella vulgaris	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Prunella vulgaris 2	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Prunus persica	Figure 4.3	Rosales	Rosaceae	Conn et al. 2015	
Ricinus communis	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Phytozome 12, 1KP	29970.m001027 CDS, PAZJ-2063133
Ricinus communis 2	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Phytozome 12	29970:456307..461956
Ricinus communis 3	Figure 4.3, Figure 4.5	Malpighiales	Euphorbiaceae	Phytozome 12	29970.m001030 CDS
Rosmarinus officinalis	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Salvia sp.	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Scutellaria montana	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Senna hebecarpa	Figure 4.6	Fabales	Fabaceae	1KP	KZED-2063215 + KZED-2062441
Sinapis alba	Figure 4.4	Brassicales	Brassicaceae	1KP	VMNH-2014473
Solanum lycopersicum	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	
Solanum lycopersicum 2	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	
Solanum lycopersicum 3	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	
Solanum tuberosum	Figure 4.3, Figure 4.8	Solanales	Solanaceae	Conn et al. 2015	
Solenostemon scutellarioides	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Striga asiatica KAI2c1	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2c2	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d1	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d2	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d3	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d4	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	

Striga asiatica KAI2d5	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d6	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d7	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d8	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d9	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d10	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d11	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d12	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d13	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d14	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d15	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d16	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2d17	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Satoko Yoshida	
Striga asiatica KAI2i	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 1	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 2	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 3	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 4	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 5	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 6	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 7	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 8	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 9	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	

Striga hermonthica HTL 10	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica HTL 11	Figure 4.2C, Figure 4.3	Lamiales	Orobanchaceae	Tsuchiya et al. 2015	
Striga hermonthica KAI2c	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d1	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d2	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d3	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d4	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d5	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d6	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d7	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d8	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d9	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d10	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2d11	Figure 4.2C, Figure 4.3	Lamiales	Lamiaceae	Conn et al. 2015	
Striga hermonthica KAI2i	Figure 4.2C, Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Thymus vulgaris	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Thymus vulgaris 2	Figure 4.3, Figure 4.8	Lamiales	Lamiaceae	Conn et al. 2015	
Trifolium pratense	Figure 4.6	Fabales	Fabaceae	Phytozome 12	Tp57577_TGAC_v2_mRNA24708CDS
Triphysaria versicolor KAI2c	Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Triphysaria versicolor KAI2d1	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Triphysaria versicolor KAI2d2	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Triphysaria versicolor KAI2d3	Figure 4.3	Lamiales	Orobanchaceae	Conn et al. 2015	
Triphysaria versicolor KAI2i	Figure 4.3, Figure 4.8	Lamiales	Orobanchaceae	Conn et al. 2015	
Vitex agnus-castus	Figure 4.3, Figure 4.8	Lamiales	Verbenaceae	Conn et al. 2015	
Vitex agnus-castus 2	Figure 4.3, Figure 4.8	Lamiales	Verbenaceae	Conn et al. 2015	
Vitex agnus-castus 3	Figure 4.3, Figure 4.8	Lamiales	Verbenaceae	Conn et al. 2015	
*Vitis vinifera**	Figure 4.3	Vitales	Vitaceae	Phytozome 12	GSVIVG01000162001

Wisteria floribunda	Figure 4.6	Fabales	Fabaceae	1KP	RMWJ-2052403 + RMWJ-2054193
Xanthocercis zambesiaca	Figure 4.6	Fabales	Fabaceae	1KP	ZSSR-2023522
Xanthocercis zambesiaca 2	Figure 4.6	Fabales	Fabaceae	1KP	ZSSR-2108674
*Vitis vinifera**: Phytozome mislabels coding sequence as intron; obtain genomic sequence for full gene sequence					

Table 4.4. Results of branch-site selection tests and JDet analyses in *KAI2* paralogs from independent duplication events. Foreground specifications are explained in Figure 4.4 – 4.8.

Site # (Arabidopsis)	Site ID (Arabidopsis)	JDet (Brassicaceae)	Branch-site model (Brassicaceae)	JDet (Euphorbiaceae)	Branch-site model (Euphorbiaceae)	JDet (Fabaceae)	Branch-site model (Fabaceae)	JDet (*Physcomitrella patens*)	Branch-site model (Physcomitrella patens)	JDet (Lamiid conserved and intermediate clades)	Branch-site model (Lamiid conserved and intermediate clades)
1	M										
2	G										
3	V										
4	V										
5	E										
6	E										
7	A										
8	H										
9	N										
10	V										
11	K										
12	V			X							
13	I										
14	G										
15	S			X							
16	G										
17	E										
18	A										
19	T										***X*** (foreground B)
20	I					X					
21	V										
22	L										
23	G									X	
24	H										
25	G				X (foreground C)						
26	F										
27	G										
28	T									X	
29	D										
30	Q										
31	S										
32	V										
33	W										
34	K										
35	H			X				X			
36	L										
37	V										
38	P										

39	H							X			
40	L										
41	V										
42	D			X				X			
43	D										***X*** (foreground A)
44	Y			X							
45	R			X							***X*** (foreground A, foreground AB)
46	V										
47	V										***X*** (foreground A, foreground AB)
48	L										
49	Y										
50	D										
51	N										
52	M										
53	G										
54	A										
55	G										
56	T										
57	T										
58	N										
59	P										
60	D								X (foreground A)		
61	Y										
62	F										
63	D										
64	F										
65	D										
66	R										
67	Y										
68	S										
69	N										
70	L									X	
71	E										
72	G										
73	Y					X					
74	S										
75	F										
76	D										
77	L									X	
78	I										
79	A										
80	I										
81	L										
82	E										

83	D										
84	L										
85	K										
86	I										
87	E										
88	S										
89	C										
90	I							X			
91	F										
92	V										
93	G										
94	H										
95	S										
96	V					X				X	
97	S										
98	A										
99	M										
100	I									X	
101	G										
102	V										
103	L										
104	A										
105	S										
106	L										
107	N										
108	R					X					
109	P										
110	D										***X*** (foreground B)
111	L										
112	F										
113	S										***X*** (foreground A, foreground AB)
114	K										
115	I							X			
116	V										
117	M							X			
118	I										
119	S										
120	A										
121	S			X							
122	P										
123	R										
124	Y			X		X				X	
125	V										X (foreground A)
126	N										
127	D										X (foreground A)

128	V										
129	D										
130	Y										
131	Q			X	***X*** (foreground B, foreground AC)						
132	G										
133	G										
134	F										
135	E					X					***X*** (forground B)
136	Q		X								
137	E										
138	D										
139	L										
140	N							X			
141	Q										
142	L									X	
143	F										
144	E										
145	A			X							
146	I										
147	R										
148	S										
149	N										
150	Y				X (foreground C)						
151	K										
152	A										
153	W										
154	C										
155	L										
156	G										
157	F										
158	A										
159	P									X	
160	L										
161	A		X			X					
162	V										
163	G										
164	G										
165	D										
166	M									X	
167	D										
168	S										
169	I								X (foreground A, foreground AB)		
170	A									X	
171	V										
172	Q										

173	E										
174	F										
175	S										
176	R										
177	T										
178	L										
179	F										
180	N							X			
181	M										
182	R										
183	P										
184	D										
185	I										
186	A										
187	L										
188	S										
189	V										
190	G										
191	Q										
192	T										
193	I										
194	F										
195	Q		X							X	
196	S					X					
197	D										
198	M										
199	R										
200	Q						X (foreground AB)				
201	I										
202	L										
203	P										
204	F						***X*** (foreground A, foreground AB)			X	
205	V										
206	T										
207	V										
208	P										
209	C										
210	H										
211	I										
212	L	X									
213	Q										
214	S										
215	V							X		X	
216	K										
217	D										
218	L										
219	A										

220	V										
221	P										
222	V										
223	V										
224	V										
225	S										
226	E							X			
227	Y										
228	L										
229	H								X (foreground AB)		
230	A										***X*** (foreground B, foreground AB)
231	N										
232	L										
233	G										
234	C										
235	E										
236	S										
237	V							X			
238	V										
239	E										
240	V										
241	I	X									
242	P	X									***X*** (foreground A, foreground AB)
243	S										
244	D			X	***X*** (foreground AC), X (foreground C)						
245	G										
246	H										
247	L										
248	P										
249	Q										
250	L										
251	S										
252	S									X	
253	P										
254	D									X	
255	S							[redacted]			***X*** (foreground A, foreground AB)
256	V										
257	I										
258	P										
259	V										

260	I										
261	L										
262	R										
263	H										
264	I										
265	R										***X*** (foreground A, foreground AB)
266	N		X								
267	D										
268	I										
269	A										***X*** (foreground A, foreground AB)
270	M										
Total sites selected		3	4	11	4	8	2	11	3	16	14

Bold, italic: prob(ω>1) > 0.99
Regular: prob(ω>1) > 0.95

Table 4.5. Oligos used to generate entry clones containing transgenes.

	Oligo					
Trans-gene	**Oligo description**	**Oligo #**	**Used with**	**Used in**	**Template(s)**	**Oligo sequence**
KAI2dmod6, KAI2dmod7	forward primer for amplification of GBLOCKS	1	2	amplification of GBLOCKS	477 base-pair fragments of *KAI2dmod6, KAI2dmod7*	CGATTTGATTGCAATCTTG GAAGATC
	reverse primer for amplification of GBLOCKS	2	1			CAGCCAAGATTGGCGTGA AGATAC
	forward primer for linearizing pDONR221-*AtKAI2*	3	4	vector linearization	pDONR221-*AtKAI2*	GTATCTTCACGCCAATCTT GGCTG
	reverse primer for linearizing pDONR221-*AtKAI2*	4	3			GATCTTCCAAGATTGCAAT CAAATCG
KAI2A161C, KAI2G190A	forward *KAI2* primer with GW sequence	5	6	amplification of *KAI2cmod1*, *AtKAI2* in two separate pieces	*KAI2cmod1*, *AtKAI2* cDNA	GGGGACAAGTTTGTACAAA AAAGCAGGCTATGGGTGT GGTAGAAGAAGCTCAC
	reverse internal *KAI2* primer	6	5			CTGAATTCTTGAACGGCGA TGG
	forward internal *KAI2* primer	7	8			CAGAACACTCTTCAATATG CGTCCCG
	reverse *KAI2* primer with GW sequence	8	7			GGGGACCACTTTGTACAA GAAAGCTGGGTTCACATA GCAATGTCATTAC
	scaffold for joining pieces of *KAI2cmod1* with *AtKAI2* cDNA	9	n/a	cycled ligation assembly	Products of 5/6 and 7/8	CCATCGCCGTTCAAGAATT CAGCAGAACACTCTTCAAT ATGCGTCCC
KAI2Y124F/A161V	forward *KAI2* primer for linearizing pDONR221-*AtKAI2*	10	11	vector linearization	pDONR221-*AtKAI2*	ACATGGACTCCATCGCCG TTC
	reverse *KAI2* primer for linearizing pDONR221-*AtKAI2*	11	10			ATCATGACGATTTTGGAGA AGAGATC
	forward internal *KAI2* primer encoding Y124F amino acid replacement	12	13	amplification of part of *AtKAI2* encoding two amino acid replacements		CTCTGCTTCTCCGAGATTC GTAAACGATG
	reverse internal *KAI2* primer encoding A161V amino acid replacement	13	12			CGCCACCGACGACGAGTG GAG
	scaffold for joining *AtKAI2* portion encoding Y124F and A161V with linearized vector	14	15	cycled ligation assembly to circularize pDONR221-*KAI2Y124F/A161V*	products of 10/11 and 12/13	GATCTCTTCTCCAAAATCG TCATGATCTCTGCTTCTCC GAGATACGTAAACGATG
		15	14			CTCCACTCGCCGTCGGTG GCGACATGGACTCCATCG CCGTTC
LsKAI2A	forward *LsKAI2A* primer with GW sequence	16	17	transgene amplification, Gateway cloning	lettuce genomic DNA	GGGGACAAGTTTGTACAAA AAAGCAGGCTATGGGAGT CGTAGAACAAGCTCAC
	reverse *LsKAI2A* primer with GW sequence	17	16			GGGGACCACTTTGTACAA GAAAGCTGGGTTTACACAA CAATATCACCCCG
LsKAI2B	forward *LsKAI2B* primer with GW sequence	18	19	transgene amplification, Gateway cloning	lettuce genomic DNA	GGGGACAAGTTTGTACAAA AAAGCAGGCTATGGGATC TGTTGTCGAACAAG
	reverse *LsKAI2B* primer with GW sequence	19	18			GGGGACCACTTTGTACAA GAAAGCTGGGTTTACACA GCTATATTACAAC

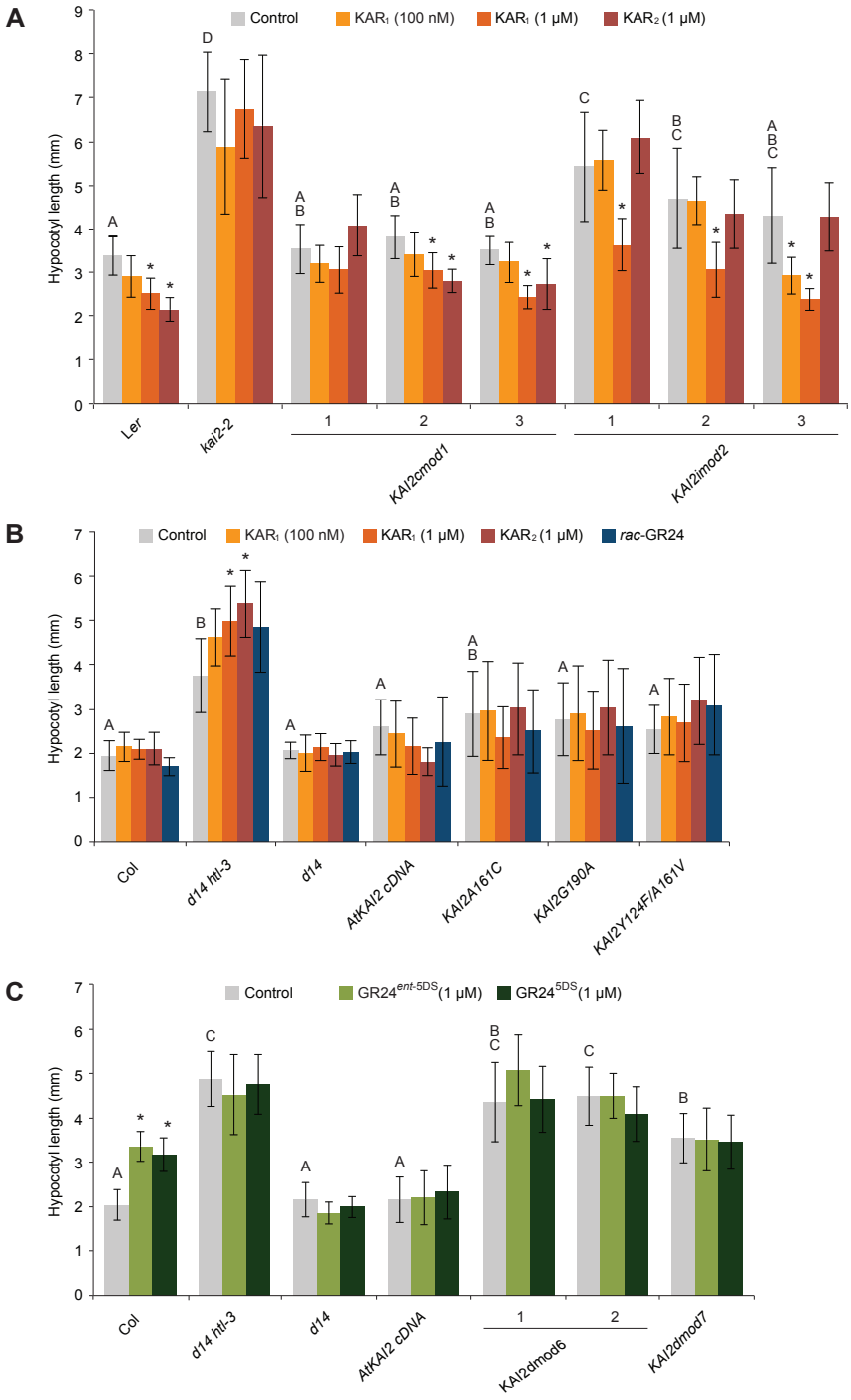

Figure 4.1. Seedling light response phenotypes conferred by modified *AtKAI2* transgenes. Hypocotyl length was measured in ten-day-old seedlings grown in short-day white light conditions. Transgenes under the control of the *AtKAI2* promoter were tested in *kai2-2* (A) or *d14 htl-3* (B – C) mutants in the L*er* or Col ecotype, respectively. Mean ± 99% confidence intervals (n = 14 – 15) are shown for 1 – 3 independent transgenic lines per transgene. Significant differences within each line between chemical treatments and control are indicated by * (Dunnett's test, $p < 0.05$). Letters indicate the results of Tukey-Kramer HSD test comparing lines on control treatment.

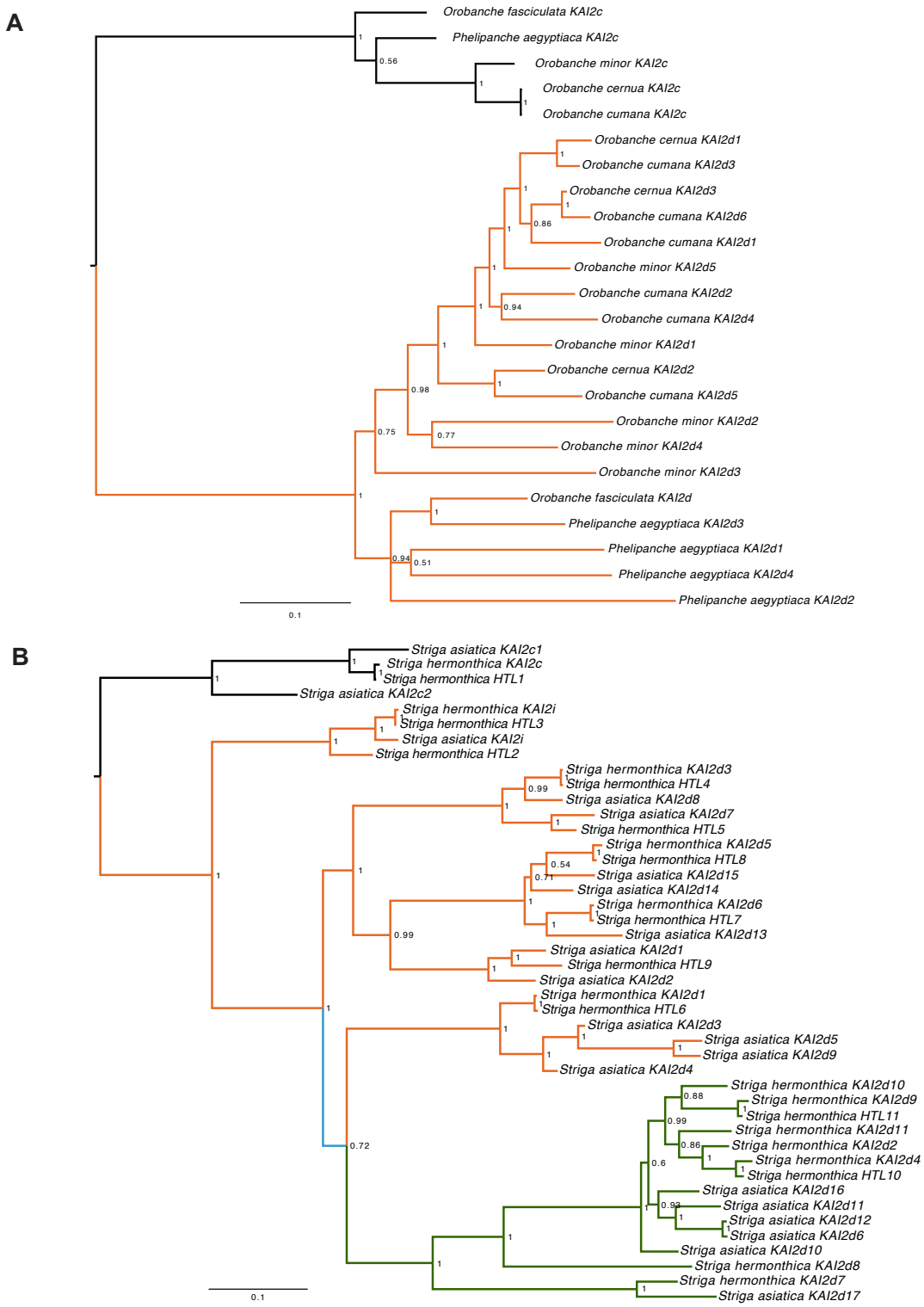

Figure 4.2. Classification of *KAI2* from parasitic weeds by phylogeny and predicted function. *KAI2* gene sequences from *Orobanche* and *Phelipanche* (A) and from *Striga* (B) were used to build phylogenies in MRBAYES3 (posterior probabilities are shown at nodes). The nucleotide alignments used to generate these trees were

translated to predicted protein alignments and analyzed by JDet. (A) The predicted amino acid sequences of *KAI2* in the black clade formed one group in JDet, and those of *KAI2* in the orange clade formed another. Orange branches and all operational taxonomic units (OTUs) in the orange clade were set as foreground in a branch-site analysis in PAML. (B) The predicted amino acid sequences of *KAI2* in the black, orange, and green clades each formed a separate group in JDet. Three branch-site tests were run in PAML. In the first, the orange clade was set as foreground (foreground A); in the second, the green clade was set as foreground (foreground B); and in the third, the orange + green + blue clade was set as foreground (foreground AB).

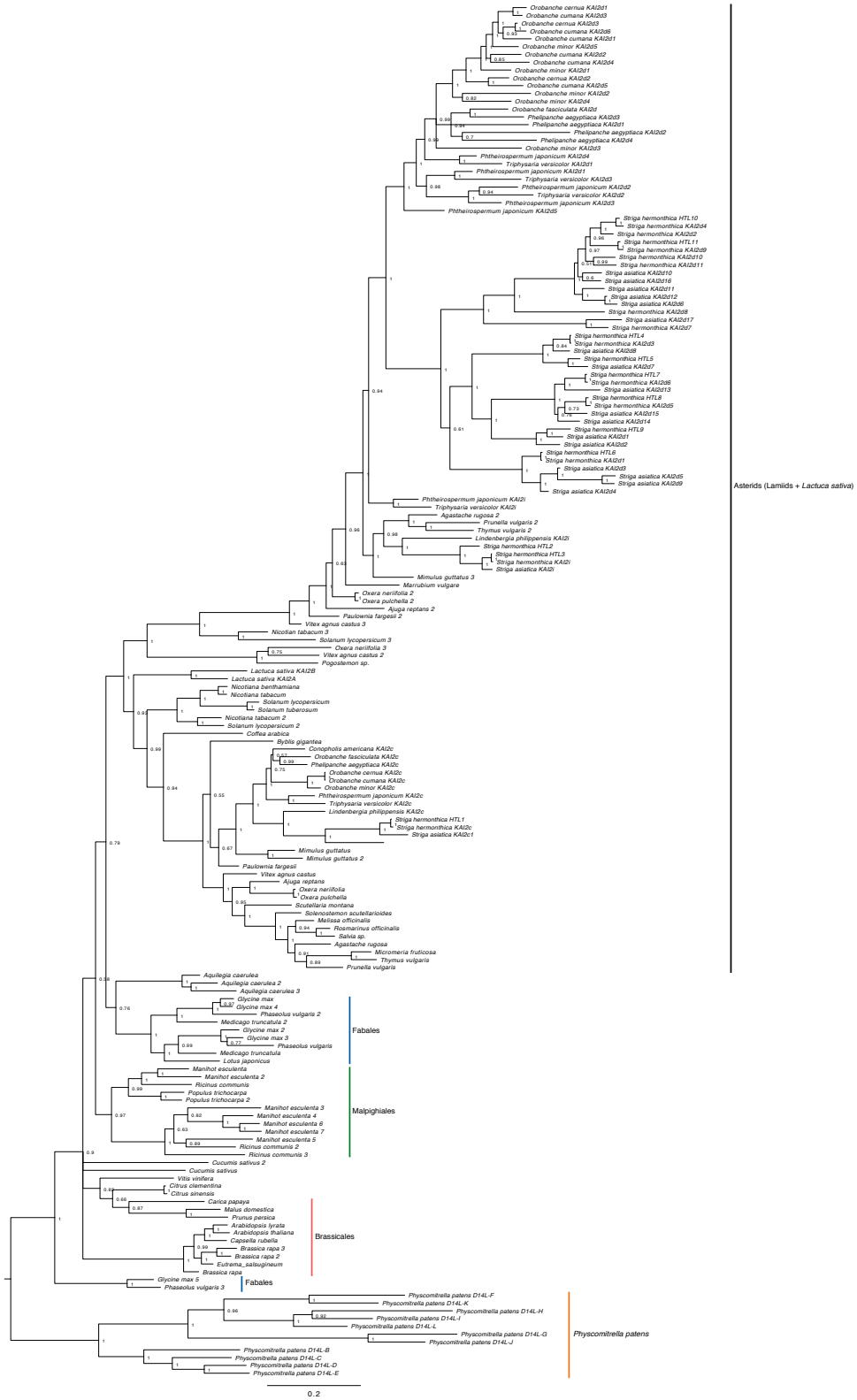

Figure 4.3. Bayesian phylogeny of *KAI2* in dicots. Sequences from the moss *Physcomitrella patens* were used as an outroup. Five taxonomic groups in which *KAI2* duplication occurred are labeled. Posterior probabilities are indicated at nodes.

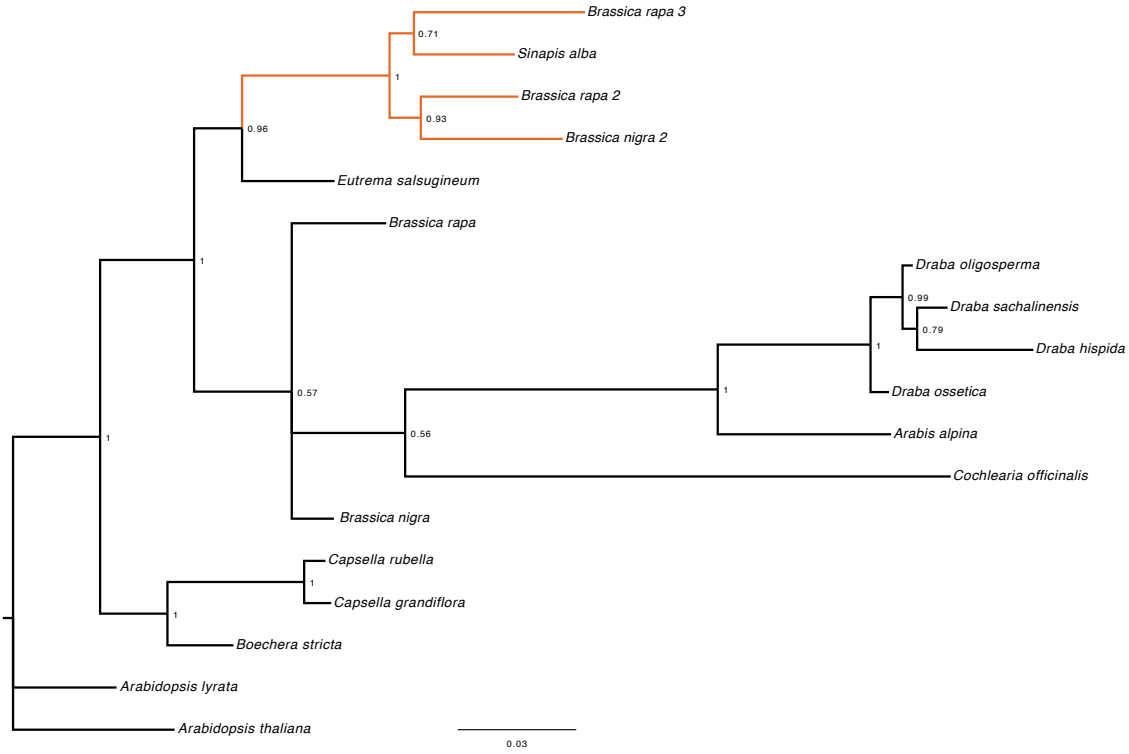

Figure 4.4. Classification of *KAI2* from Brassicaceae by phylogeny and predicted function. Posterior probabilities are indicated at nodes of the Bayesian phylogeny. The two predicted functional groups assigned by JDet are indicated by black and orange. The orange clade was set as foreground in a branch-site test for positive selection in PAML.

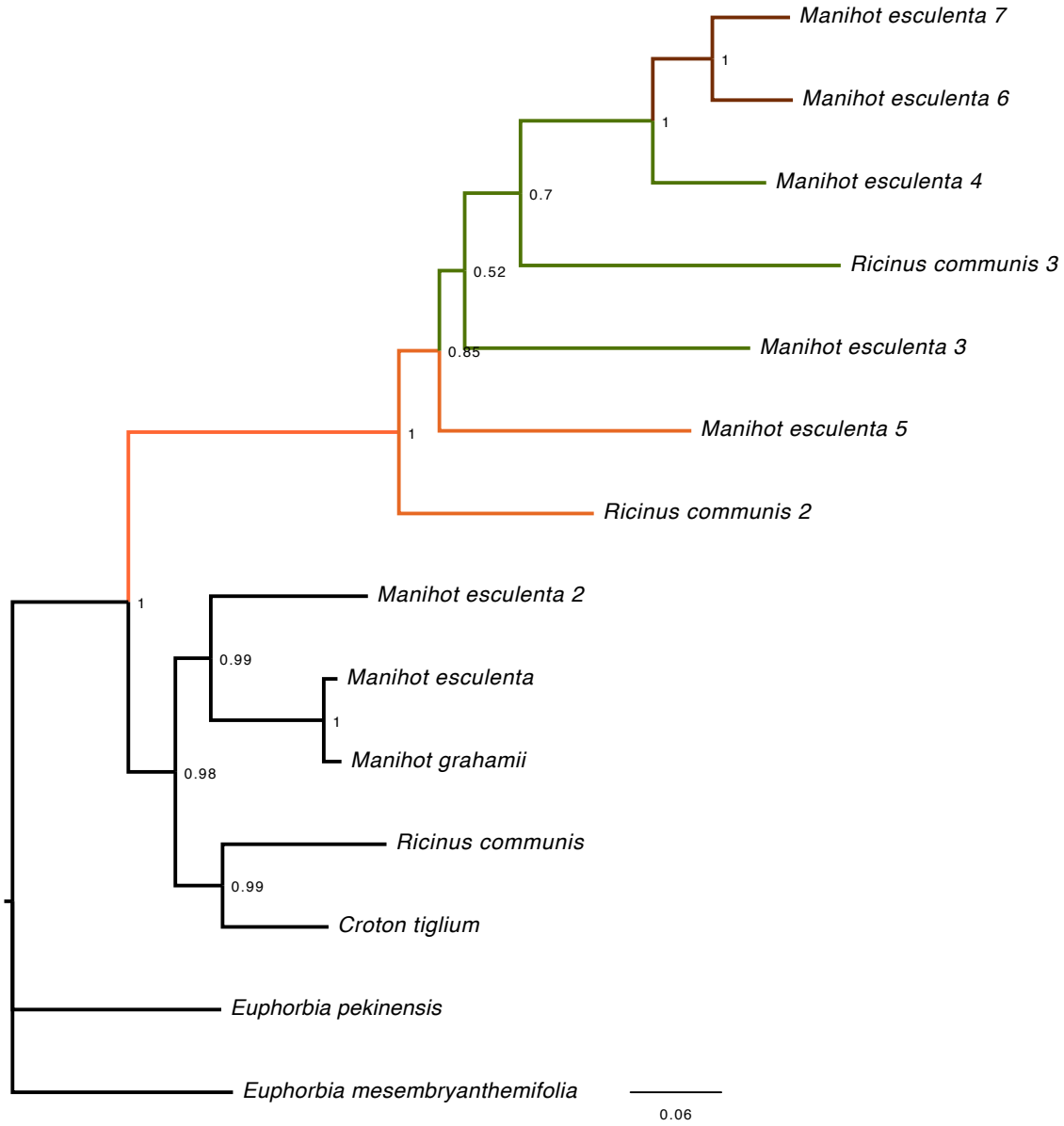

Figure 4.5. Classification of *KAI2* from Euphorbiaceae by phylogeny and predicted function. Posterior probabilities are indicated at nodes of the Bayesian phylogeny. The four predicted functional groups assigned by JDet are indicated by black, orange, green, and purple. Each colored clade was set as foreground in a branch-site test for positive selection in PAML (orange, foreground A; green, foreground B; purple, foreground C). A fourth branch-site model was assessed in which the three colored clades together were set as foreground (foreground A – C).

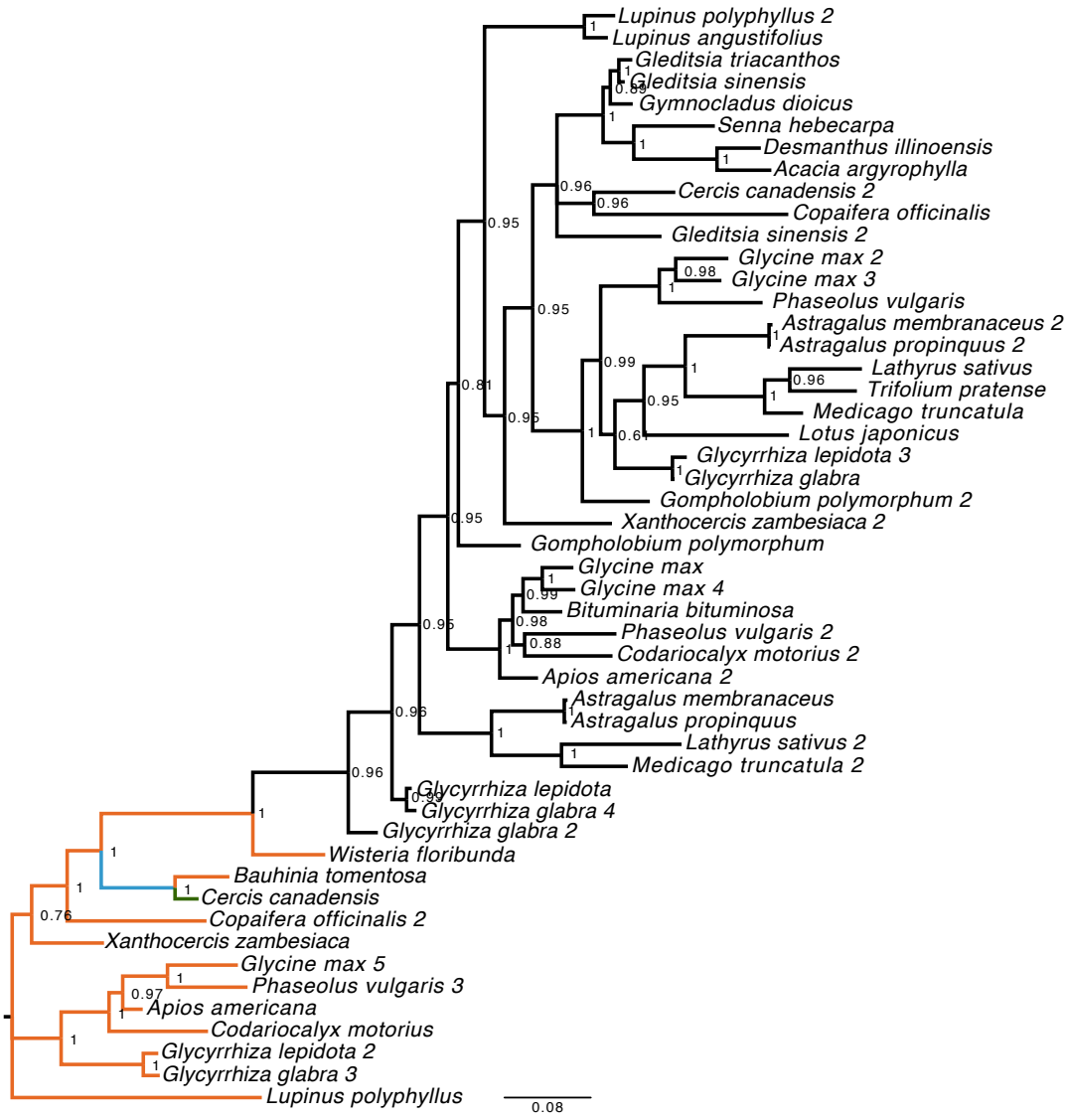

Figure 4.6. Classification of *KAI2* from Fabaceae by phylogeny and predicted function. Posterior probabilities are indicated at nodes of the Bayesian phylogeny. The three predicted functional groups assigned by JDet are indicated by black, orange, and green. The orange clade and green OTU (*C. canadensis*) were each set as foreground in two separate branch-site tests of positive selection in PAML (orange, foreground A; green, foreground B). A third branch-site model was assessed in which the orange, green, and blue portion of the tree was set as foreground (foreground AB).

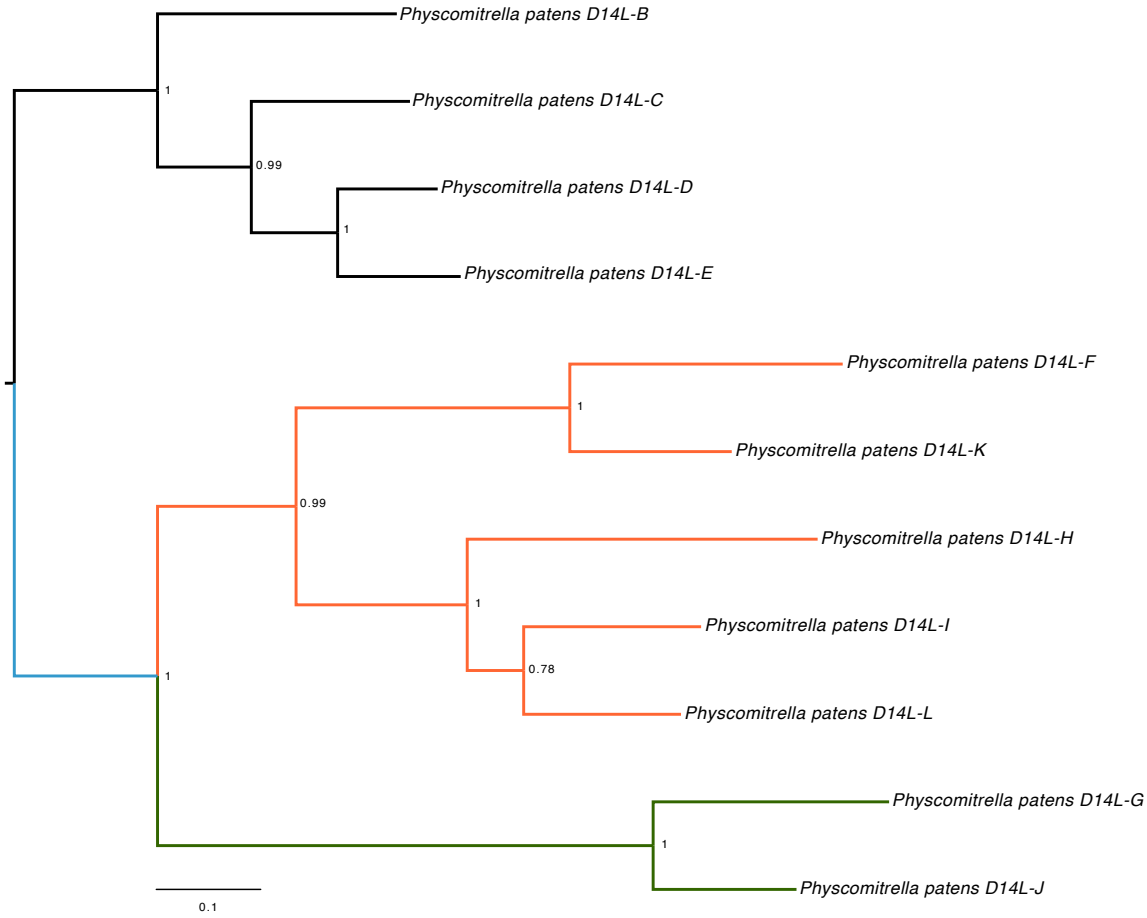

Figure 4.7. Classification of *KAI2* from *Physcomitrella patens* by phylogeny and predicted function. Posterior probabilities are indicated at nodes of the Bayesian phylogeny. The three predicted functional groups assigned by JDet are indicated by black, orange, and green. Orange and green clades were each set as foreground in two separate branch-site tests of positive selection in PAML (orange, foreground A; green, foreground B). A third branch-site model was assessed in which the orange, green, and blue portion of the tree was set as foreground (foreground AB).

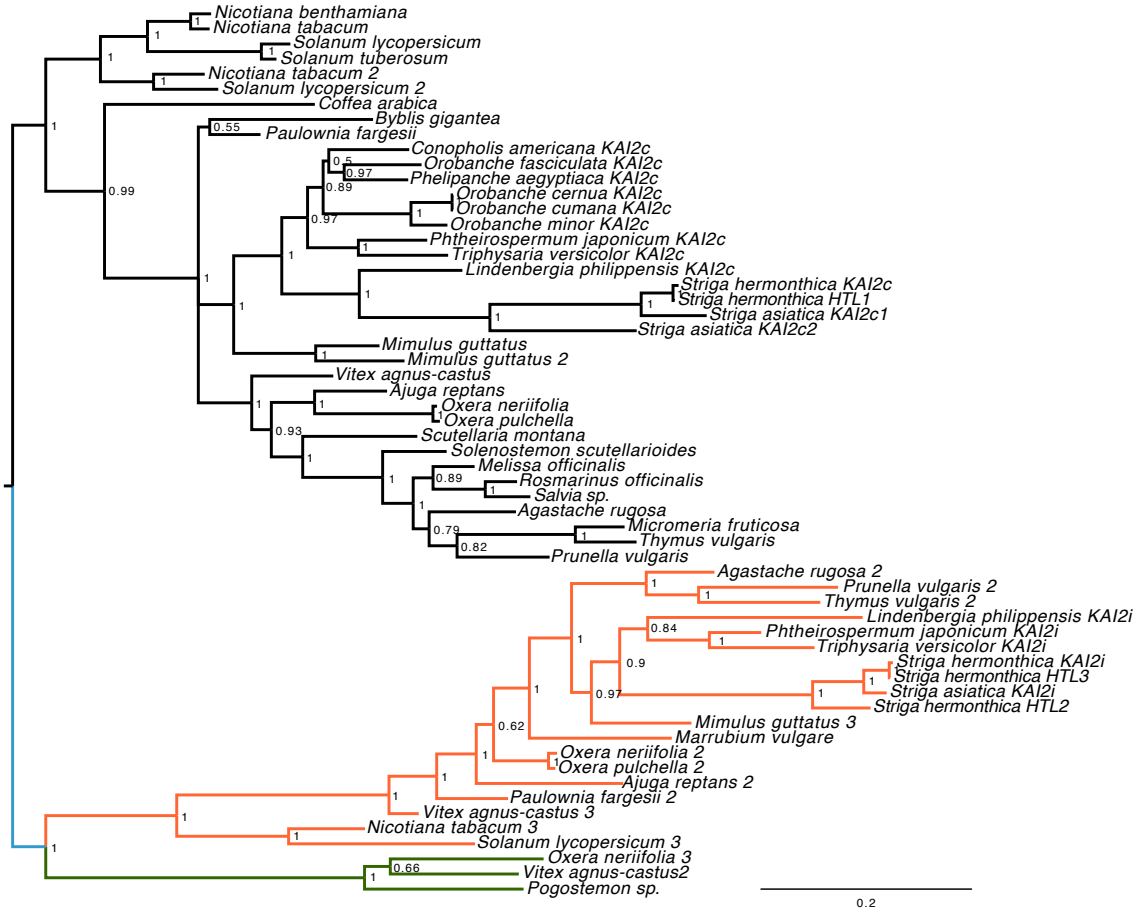

Figure 4.8. Classification of *KAI2* from the conserved and intermediate clades in Lamiids by phylogeny and predicted function. Posterior probabilities are indicated at nodes of the Bayesian phylogeny. The three predicted functional groups assigned by JDet are indicated by black, orange, and green. The orange and green clades were each set as foreground in separate branch-site tests of positive selection in PAML (orange, foreground A; green, foreground B). A third branch-site model was assessed in which the orange, green, and blue portion of the tree was set as foreground (foreground AB).

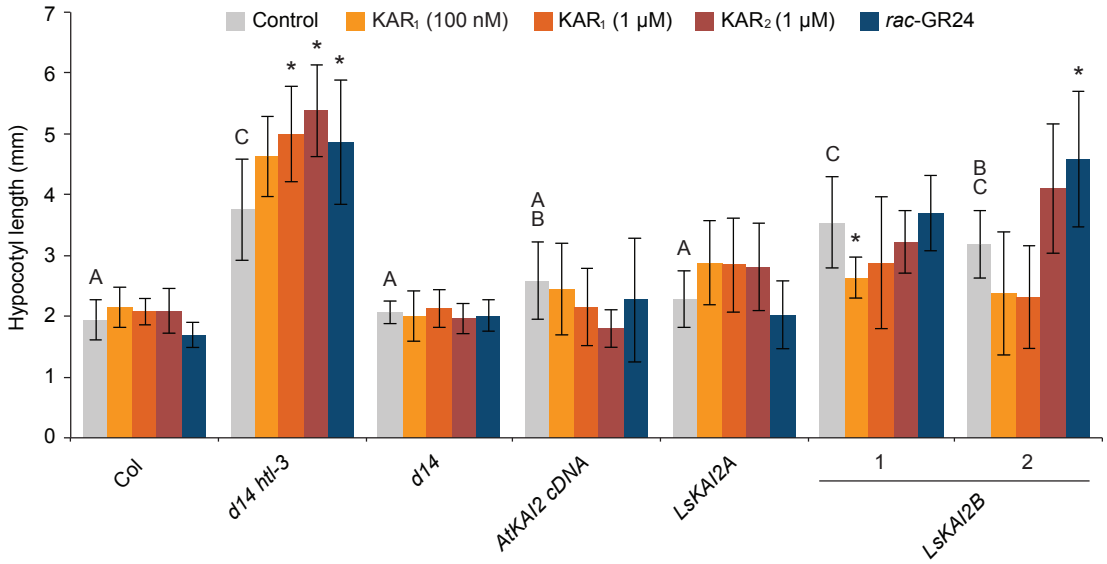

Figure 4.9 Seedling light response phenotypes conferred by modified lettuce *KAI2* transgenes. Hypocotyl length was measured in ten-day-old seedlings grown in short-day white light conditions. Transgenes under the control of the *AtKAI2* promoter were tested in *d14 htl-3* mutants in the Col ecotype. Mean ± 99% confidence intervals (n = 9 – 15) are shown for 1 – 2 independent transgenic lines per transgene. Significant differences within each line between chemical treatments and control are indicated by * (Dunnett's test, $p < 0.05$). Letters indicate the results of Tukey-Kramer HSD test comparing lines on control treatment.

CHAPTER 5

CONCLUSIONS

Thesis summary

Parasitic weeds in the Orobanchaceae family attack diverse crop hosts and cause billions of dollars of losses each year (Westwood et al. 2010). The seeds of many parasitic weeds germinate in response to host-derived hormones called strigolactones (SLs; Cook et al. 1966; reviewed in Xie et al. 2010), which improves their chance of quickly finding a nearby host (Hirsch et al. 2003). Soon after parasite seeds germinate, they attach to and damage their hosts (Sauerborn et al. 2007). Therefore, parasite seed germination is an attractive target for the development of new control strategies. However, little was known about how parasites perceive SLs until recently. The biology of parasitic plants, including parasitic weeds of the Orobanchaceae, is described in more detail in Chapter 1.

In the study documented in Chapter 2, we discovered that the *KAI2* gene has undergone extensive duplication in parasitic plants. Some fast-evolving or divergent paralogs from parasites (*KAI2d*) are SL-responsive, while those with conserved and intermediate rates of evolution (*KAI2c* and *KAI2i*, respectively) function more similarly to their ortholog in the model non-parasite *Arabidopsis thaliana* (Conn et al. 2015). *KAI2* exists as a single copy in Arabidopsis, in which it mediates seed germination and seedling light sensitivity, likely by perceiving an unidentified endogenous signal called KAI2 ligand (KL; Waters et al. 2012, Waters et al. 2015, Conn and Nelson 2016). In Arabidopsis, *KAI2* also responds to smoke-derived germination cues called karrikins (KARs; Waters et al. 2012). Thus, although *KAI2* responds to different signals in

parasitic weeds and in Arabidopsis, it appears to have a conserved role in seed germination.

In Chapter 3, we further investigated the function of SL-unresponsive *KAI2c* and *KAI2i* paralogs in parasites. We found that relative to Arabidopsis *KAI2*, *KAI2c* and *KAI2i* appear to have undergone subfunctionalization. Two *KAI2c* from parasitic weeds – one from *Phelipanche aegyptiaca*, and one from *Striga hermonthica* – are sensitive to KL, and to a lesser extent the *KAI2i* gene from *S. hermonthica* is as well. However, the most notable response of the *S. hermonthica KAI2i* gene is to KAR. Taken together, data from Chapters 2 and 3 support a model of *KAI2* evolution in parasites in which multiple duplications were followed by sub- and neofunctionalization (Conn et al. 2015, Conn and Nelson 2016). The work presented in these chapters provides practical information for the fight against parasitic weeds and uncovers new insight into fundamental processes of molecular evolution.

In Chapter 4, we addressed the structural basis of differences in likely ligand preference among KAI2c, KAI2i, and KAI2d from parasites. By modifying predicted specificity-determining positions in Arabidopsis KAI2, we were able to partially recapitulate the KAR preference of KAI2i from *S. hermonthica*. However, we were not able to engineer a SL-responsive version of Arabidopsis KAI2. We also compared *KAI2* paralogs that have arisen from separate duplication events in non-parasitic taxa to determine whether convergent molecular evolution has occurred. We generated lists of amino acids that are predicted to contribute to differences in protein function in KAI2 copies, as well as codons under positive selection in *KAI2* paralogs. Across taxonomic groups with independent *KAI2* duplication events, these site lists had very little overlap, suggesting an absence of convergent molecular evolution. However, this does not preclude convergent evolution of function among *KAI2* duplicates, as SL responsiveness has evolved convergently in the *KAI2* homolog from angiosperms, *D14*; in *KAI2d* from

parasites; and in some *KAI2* paralogs in moss (Waters et al. 2012, Conn et al. 2015, Lopez-Obando et al. 2016). Overall, this dissertation work has elucidated the evolution and genetic basis of host-responsive seed germination in parasitic plants, uncovered a new story of sub- and neofunctionalization of duplicate genes, investigated the structural basis of differences in signal perception among KAI2 paralogs, and analyzed *KAI2* evolution more broadly throughout the plant kingdom.

Toward more effective control of parasitic weeds

Although questions remain as to how SL responsiveness evolved among KAI2 proteins in parasites, we have gained valuable insight into host detection in parasitic weeds. How can our results contribute to the development of new control strategies for parasitic weeds? One possible solution to parasitic weed infestations is to induce suicidal seed germination. For obligate parasites, this occurs when a seed germinates in the absence of a suitable host (reviewed in Zwanenburg et al. 2016). Ethylene gas is one way to stimulate suicidal germination in parasites. After a parasitic weed outbreak occurred in the United States in 1956, ethylene gas was successfully incorporated into a campaign to contain the outbreak (Eplee 1975). However, as with herbicide application, the use of ethylene is impractical in resource-limited parts of the world where parasitic weed infestations are the most damaging. Both ethylene and herbicides are expensive, and large investments of time and energy are required for their effective implementation (Oswald 2005). Furthermore, ethylene is ineffective against some parasitic weeds (Joel 2000). Synthetic strigolactones have been developed as stimulants of suicidal seed germination in parasites; however, they are also expensive and difficult to apply underground (Tsuchiya and McCourt 2012).

Trap and catch cropping may be more viable methods for stimulating suicidal germination of parasites because they reduce the parasite seed bank without expensive

chemical treatments. Trap crops stimulate parasite seed germination; however, after parasitic weeds germinate, they are unable to use the trap crops as a host. For example, cotton and peanut stimulate seed germination of *S*. *hermonthica*, but the parasite cannot survive on these non-hosts. Success in combating *S. hermonthica* infestation has been documented after five weeks of trap cropping. Catch crops also stimulate the germination of parasite seeds, but are susceptible hosts on which parasitic weeds can survive (Musselman 1980). To be effective control agents, catch crops must be harvested before the parasites attached to them have a chance to reproduce. Strigolactone overproduction in trap and catch crops has been suggested as a way to make these control methods more effective (López-Ráez et al. 2009). Trap and catch cropping can thus be greatly improved by our new understanding of strigolactone signaling in parasites. Over a dozen strigolactone structural variants have been identified in nature (Xie et al. 2010), and different host plants can produce these structural variants in different relative amounts (Awad et al. 2006, Jamil et al. 2011, Xie et al. 2013). The strigolactone profiles of hosts may determine how detectable they are to parasite seeds. Germination assays on parasite seeds indicate that recognition of host root exudates plays an important role in host/parasite compatibility (Fernández-Aparicio et al. 2009). Furthermore, recent cross-species complementation and biochemical data suggest that in the parasitic weed *Striga hermonthica*, different KAI2d paralogs may have preferences for different strigolactone structural variants (Toh et al. 2015, Tsuchiya et al. 2015). If the *KAI2* alleles and paralogs in a particular parasite population were sequenced, then trap and catch crops that produce the proteins' preferred strigolactones could be selected. In this way, trap and catch cropping could be improved to maximize suicidal germination of parasite seeds.

Our research also enables more efficient identification of small molecules that stimulate suicidal germination in parasite seeds. Synthetic SL analogs that stimulate

parasite seed germination have been identified; for example, GR24 and Nijmegen 1 were tested on parasite seeds directly and shown to promote germination of *Striga* and *Orobanche* (Wigchert et al. 1999). However, the strict regulation of parasitic weeds makes testing these compounds on parasite seeds difficult. Our knowledge of which genes in parasitic weeds are involved in host-responsive seed germination facilitates *in vitro* screens for activators of *KAI2* signaling. Additionally, transgenic Arabidopsis lines carrying parasite *KAI2* transgenes (generated by Conn et al. 2015, Toh et al. 2015) can now be used to test potential stimulants of suicidal parasite germination. In summary, the work presented in this thesis opens up new possibilities for testing control strategies that target seed germination in parasitic weeds.

Additional future directions

The research described here also raises questions about the evolution of *KAI2* signaling, specifically within parasitic plants and more broadly throughout the plant kingdom. First, parasitic plants vary in host range, and weedy parasites – which tend to have broader host ranges, at least in *Orobanche* and *Phelipanche* (Schneeweiss 2007) – also often have higher *KAI2* copy number (see Chapter 2). However, whether *KAI2* amplification and host range expansion are related is currently unknown. To address this question, sequence capture and next-generation sequencing could be used to assess *KAI2* copy number and sequence variation in a larger number of weedy and non-weedy parasites with different host ranges. If *KAI2* amplification indeed facilitates host range expansion, then generalist parasites would be expected to have higher *KAI2* copy number than specialists. The results of this proposed study could then guide future work, such as an investigation of *KAI2* allelic diversity or expression of *KAI2* paralogs in generalists and specialists. Although these experiments would be time- and labor-

intensive, they could provide important information on the genetic basis of host range in parasitic weeds.

Another pressing question regards the unidentified endogenous regulator KL. In Chapter 3, we added to a growing body of evidence for the existence of KL. But what is the identity of this unknown KAI2 ligand? Is it a single signal, or does KAI2 respond to multiple endogenous signals? Consistent with the known ligands of KAI2 and its homolog D14 (Hamiaux et al. 2012, Bythell-Douglas et al. 2013, Guo et al. 2013, Kagiyama et al. 2013, Nakamura et al. 2013, Zhao et al. 2013), a butenolide moiety is expected in KL (Conn and Nelson 2016). Currently, nothing else is known about its identity. However, an elegant system for detection of KL signaling was recently devised by fusing the promoter of *DLK2*, a marker for *KAI2* signaling, with firefly luciferase. Treatment with KARs induces luminescence in this system, as does leaf extract from Arabidopsis, which presumably contains KL. Thus, this assay provides new evidence for the existence of KL, and could eventually be used to identify this mystery compound (Sun et al. 2016).

Finally, the evolution and function of *KAI2* paralogs in non-parasitic angiosperms should be studied more thoroughly. Other than the two *KAI2* copies in lettuce – which may have arisen from a duplication event shared with parasitic plants – almost nothing is known about how *KAI2* paralogs function in angiosperms outside of the Orobanchaceae family. Is specialization for different signals a common feature of *KAI2* duplicates, as we observed in parasitic plants? Are *KAI2* paralogs expressed at different developmental stages and/or in different tissues? The answers to these questions about sub- and neofunctionalization of duplicate genes will provide new insight into a signaling system that is highly conserved throughout the plant kingdom.

Parasitic plants in a broader ecological and cultural context

Parasitic weeds cause tremendous agricultural damage, particularly in parts of the world where resources are limited (Westwood et al. 2010). Parasitic weeds often have broad host ranges (Schneeweiss 2007) and damage their hosts soon after germination (Sauerborn et al. 2007). They have numerous other adaptations that make them difficult to control as well, including high seed production, long seed viability (Scholes and Press 2008), and the ability to adapt to new hosts or overcome host resistance (Haussmann and Hess 2001, Pérez-Vich et al. 2004). The threat presented by parasitic weeds should not be understated, and the need for more effective and inexpensive control strategies is dire.

Despite the severity of the parasitic weed problem, some parasitic plants are important parts of natural ecosystems. Outside of agricultural contexts, parasitic plants can enhance community diversity and be considered keystone species. For example, parasitic plants can enable non-hosts to thrive if their hosts are dominant plant species. Parasitic plants also generate nutrient-rich litter and may therefore impact microbes in the rhizosphere. Finally, birds nest in some parasitic plants, and some herbivores consume parasitic plants as a food source (Press and Phoenix 2005). Thus, non-weedy parasites can influence the composition of the communities in which they are found.

Despite the threat that the weedy Orobanchaceae pose to agriculture, some parasitic plants are also valuable to humans. People have cultivated parasitic plants for food and medicine, including some species from the Orobanchaceae. For example, *Cistanche deserticola* is used in traditional Japanese and Chinese medicine (Xu et al. 2009). *Orobanche crenata* is part of a recipe that is hundreds of years old, and it appears to be consumed regularly in part of Italy to this day (Renna et al. 2015). Humans and parasitic plants have apparently interacted for centuries, and some parasites have served important purposes in different cultures.

In conclusion, parasitic plants illustrate important biological concepts, such as gene duplication, neofunctionalization, and adaptive evolution. The search for more effective control strategies for parasitic weed infestations should be a top priority for parasitic plant researchers. But in the process of translating basic research into practical applications, the rich evolutionary history of parasitic plants should not be ignored. Furthermore, non-weedy parasitic plants should be recognized as important contributors to biodiversity. By learning more about the evolution, physiology, and ecology of parasites, we will develop more effective control methods for weed infestations and make new discoveries about the remarkable biology of parasitic plants.

REFERENCES

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

Arite, T., Umehara, M., Ishikawa, S., Hanada, A., Maekawa, M., Yamaguchi, S., Kyozuka, J. (2009) *d14*, a Strigolactone-Insensitive Mutant of Rice, Shows an Accelerated Outgrowth of Tillers. *Plant & Cell Physiology* 50(8): 1416 – 1424.

Awad, A.A., Sato, D., Kusumoto, D., Kamioka, H., Takeuchi, Y., Yoneyama, K. (2006) Characterization of Strigolactones, Germination Stimulants for the Root Parasitic Plants *Striga* and *Orobanche*, Produced by Maize, Millet and Sorghum. *Plant Growth Regulation* 48: 221. DOI: https://doi.org/10.1007/s10725-006-0009-3.

Baird, W.V., Riopel, J.L. (1986) Life History Studies of Conopholis americana (Orobanchaceae). *The American Midland Naturalist* 116(1): 140 – 151.

Baum, J., Thomas, A.W., Conway, D.J. (2003) Evidence for Diversifying Selection on Erythrocyte-Binding Antigens of *Plasmodium falciparum* and *P. vivax*. *Genetics* 163(4): 1327 – 1336.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. (2005) GenBank. *Nucleic Acids Research* 33: D34 – D38.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., Huala, E. (2015) The Arabidopsis Information Resource: Making and Mining the "Gold Standard" Annotated Reference Plant Genome. *genesis* 53: 474 – 485.

Bewley, J.D. (1997) Seed Germination and Dormancy. *The Plant Cell* 9: 1055 – 1066.

Bouwmeester, H.J., Matusova, R., Zhongkui, S., Beale, M.H. (2003) Secondary metabolite signaling in host-parasitic plant interactions. *Current Opinion in Plant Biology* 6: 358 – 364.

Brotherson, J.D., Simmons, B.T., Ball, T., Anderson, W.R. (2005) Nutrient Relationships Between *Orobanche fasciculata* Nutt. and its Host *Artemisia pygmaea* Gray in the Uinta Basin of Utah, USA. *Western North American Naturalist* 65(2): 242 – 247.

Bythell-Douglas, R., Rothfels, C.J., Stevenson, D.W.D., Graham, S.W., Wang, G.K-S., Nelson, D.C., Bennett, T. (2017) The complex origins of strigolactone signaling in land plants. *In review.* Accessed from http://www.biorxiv.org/content/early/2017/01/25/102715. DOI: https://doi.org/10.1101/102715.

Bythell-Douglas, R., Waters, M.T., Scaffidi, A., Flematti, G.R., Smith, S.M. (2013) The Structure of the Karrikin-Insensitive Protein (KAI2) in *Arabidopsis thaliana*. *PLOS ONE* 8(1): e54758. DOI: https://doi.org/10.1371/journal.pone.0054758.

Capra, J. A., Singh, M. (2008) Characterization and prediction of residues determining protein functional specificity. *Bioinformatics* 24(13): 1473 – 1480.

Chevalier, F., Nieminen, K., Sánchez-Ferrero, J.C., Rodríguez, M.L., Chagoyen, M., Hardtke, C.S., Cubas, P. (2014) Strigolactone Promotes Degradation of DWARF14, an α/β Hydrolase Essential for Strigolactone Signaling in *Arabidopsis*. *The Plant Cell* 26(3): 1134 – 1150.

Chiwocha, S.D.S., Dixon, K.W., Flematti, G.R., Ghisalberti, E.L., Merritt, D.J., Nelson, D.C., Riseborough, J.-A. M., Smith, S.M., Stevens, J.C. (2009) Karrikins: A new family of plant growth regulators in smoke. *Plant Science* 177: 252 – 256.

Clough, S.J., Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* 16(6): 735 – 743.

Conn, C.E., Bythell-Douglas, R., Neumann, D., Yoshida, S., Whittington, B., Westwood, J.H., Shirasu, K., Bond, C.S., Dyer, K.A., Nelson, D.C. (2015) Convergent

evolution of strigolactone perception enabled host detection in parasitic plants. *Science* 349(6247): 540 – 543.

Conn, C.E., Nelson, D.C. (2016) Evidence that KARRIKIN-INSENSITIVE2 (KAI2) Receptors may Perceive an Unknown Signal that is not Karrikin or Strigolactone. *Frontiers in Plant Science* 6: 1219. DOI: 10.3389/fpls.2015.01219.

Cook, C.E., Whichard, L.P., Turner, B., Wall, M.E., Egley, G.H. (1966) Germination of Witchweed (Striga lutea Lour.): Isolation and Properties of a Potent Stimulant. *Science* 154(3753): 1189 – 1190.

Crooks, G.E., Hon, G., Chandonia, J.-M., Brenner, S.E. (2004) WebLogo: A Sequence Logo Generator. *Genome Research* 14: 1188 – 1190.

Daws, M.I., Davies, J., Pritchard, H.W., Brown, N.A.C., van Staden, J. (2007) Butenolide from plant-derived smoke enhances germination and seedling growth of arable weed species. *Plant Growth Regulation* 51: 73 – 82.

Delaux, P.-M., Xie, X., Timme, R.E., Puech-Pages, V., Dunand, C., Lecompte, E., Delwiche, C.F., Yoneyama, K., Bécard, G., Séjalon-Delmas, N. (2012) Origin of strigolactones in the green lineage. *New Phytologist* 195: 857 – 871.

Demuth, J.P., Hahn, M.W. (2009) The life and death of gene families. *BioEssays* 31: 29 – 39.

Doyle, J.J., Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus* 12: 13 – 15.

Eplee, R.E. (1975) Ethylene: A Witchweed Seed Germination Stimulant. *Weed Science* 23(5): 433 – 436.

Fernández-Aparicio, M., Flores, F., Rubiales, D. (2009) Recognition of root exudates by seeds of broomrape (*Orobanche* and *Phelipanche*) species. *Annals of Botany* 103: 423 – 431.

Fernandez-Pozo, N., Menda, N., Edwards, J.D., Saha, S., Tecle, I.Y., Strickler, S.R., Bombarely, A., Fisher-York, T., Pujar, A., Foerster, H., Yan, A., Mueller, L.A. (2015) The Sol Genomics Network (SGN)–from genotype to phenotype to breeding. *Nucleic Acids Research* 43(D1): D1036 – D1041.

Flematti, G.R., Ghisalberti, E.L., Dixon, K.W., Trengrove, R.D. (2004) A Compound from Smoke That Promotes Seed Germination. *Science* 305(5686): 977.

Gerardo, N.M., Jacobs, S.R., Currie, C.R., Mueller, U.G. (2006) Ancient Host-Pathogen Associations Maintained by Specificity of Chemotaxis and Antibiosis. *PLOS Biology* 4(8): e235. DOI: 10.1371/journal.pbio.0040235.

Göbel, U., Sander, C., Schneider, R., Valencia, A. (1994) Correlated Mutations and Residue Contacts in Proteins. *Proteins* 18(4): 309 – 317.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1): D1178–D1186.

Graur, D., Li, W.-H. (2000) *Fundamentals of Molecular Evolution*. Sunderland, MA: Sinauer Associates, Inc.

Guo, Y., Zheng, Z., La Clair, J.J., Chory, J., Noel, J.P. (2013) Smoke-derived karrikin perception by the α/β-hydrolase KAI2 from *Arabidopsis*. *PNAS* 110(20): 8284 – 8289.

Gutjahr, C., Gobbato, E., Choi, J., Riemann, M., Johnston, M.G., Summers, W., Carbonnel, S., Mansfield, C., Yang, S.-Y., Nadal, M., Acosta, I., Takano, M., Jiao, W.-B., Schneeberger, K., Kelly, K.A., Paskowski, U. (2015) Rice perception of symbiotic arbuscular mycorrhizal fungi requires the karrikin receptor complex. *Science* 350(6267): 1521 – 1524.

Hamiaux, C., Drummond, R.S.M., Janssen, B.J., Ledger, S.E., Cooney, J.M., Newcomb, R.D., Snowden, K.C. (2012) DAD2 Is an alpha/beta Hydrolase Likely to Be Involved in the Perception of the Plant Branching Hormone, Strigolactone. *Current Biology* 22: 2032 – 2036.

Haussmann, B.I.G., Hess, D.E. (2001) Variability within and among *Striga* species, and stability of resistance. In Kroschel, J. (Ed.), *A Technical Manual for Parasitic Weed Research and Extension* (104 – 105). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Hirsch, A.M., Bauer, W.D., Bird, D.M., Cullimore, J., Tyler, B., Yoder, J.I. (2003) Molecular signals and receptors: controlling rhizosphere interactions between plants and other organisms. *Ecology* 84(4): 858 – 868.

Hoffmann, B., Proust, H., Belcram, K., Labrune, C., Boyer, F.-D., Rameau, C., Bonhomme, S. (2014) Strigolactones Inhibit Caulonema Elongation and Cell Division in the Moss *Physcomitrella patens*. *PLOS ONE* 9(6): e99206. DOI: 10.1371/journal.pone.0099206.

Huang, S., van der Vossen, E.A.G., Kuang, H., Vleeshouwers, V.G.A.A., Zhang, N., Borm, T.J.A., van Eck, H.J., Baker, B., Jacobsen, E., Visser, R.G.F. (2005) Comparative genomics enabled the isolation of the R3a late blight resistance gene in potato. *The Plant Journal* 42: 251 – 261.

Hughes, A.L., Yeager, M. (1997) Coordinated Amino Acid Changes in the Evolution of Mammalian Defensins. *Journal of Molecular Evolution* 44: 675 – 682.

Jamil, M., Charnikhova, T., Cardoso, C., Jamil, T., Ueno, K., Verstappen, F., Asami, T., Bouwmeester, H.J. (2011) Quantification of the relationship between strigolactones and *Striga hermonthica* infection in rice under varying levels of nitrogen and phosphorous. *Weed Research* 51(4): 373 – 385.

Jiang, L., Liu, X., Xiong, G., Liu, H., Chen, F., Wang, L., Meng, X., Liu, G., Yu, H., Yuan, Y., Yi, W., Zhao, L., Ma, H., He, Y., Wu, Z., Melcher, K., Qian, Q., Xu, H.E., Wang, Y., Li, J. (2013) DWARF 53 acts as a repressor of strigolactone signalling in rice. *Nature* 504: 401 – 405.

Joel, D.M. (2000) The long-term approach to parasitic weeds control: manipulation of specific developmental mechanisms of the parasite. *Crop Protection* 19: 753 – 758.

Kagiyama, M., Hirano, Y., Mori, T., Kim, S.-Y., Kyozuka, J., Seto, Y., Yamaguchi, S., Hakochima, T. (2013) Structures of D14 and D14L in the strigolactone and karrikin signaling pathways. *Genes to Cells* 18: 147 – 160.

Kettlewell, H.B.D. (1955) Selection experiments on industrial melanism in the *Lepidoptera*. *Heredity* 9(3): 323 – 342.

Kondrashov, F.A. (2012) Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B* 279(1749): 5048 – 5057.

Lamont, B.B., He, T. (2017) Fire-Proneness as a Prerequisite for the Evolution of Fire-Adapted Traits. *Trends in Plant Science* 22(4): 278 – 288.

Li, J., Lis, K.E., Timko, M.P. (2009) Molecular genetics of race-specific resistance of cowpea to *Striga gesnerioides* (Willd.). *Pest Management Science* 65(5): 520 – 527.

Lopez-Obando, M., Conn, C.E., Hoffmann, B., Bythell-Douglas, R., Nelson, D.C., Rameau, C., Bonhomme, S. (2016) Structural modelling and transcriptional responses highlight a clade of *PpKAI2-LIKE* genes as candidate receptors for strigolactones in *Physcomitrella patens*. *Planta* 243: 1441 – 1453.

López-Ráez, J.A., Matusova, R., Cardoso, C., Jamil, M., Charnikhova, T., Kohlen, W., Ruyter-Spira, C., Verstappen, F., Bouwmeester, H. (2009) Strigolactones:

ecological significance and use as a target for parasitic plant control. *Pest Management Science* 65(5): 471 – 477.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W., Wang, J. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1:18. DOI: 10.1186/2047-217X-1-18.

Matusova, R., Rani, K., Verstappen, F.W.A., Franssen, M.C.R., Beale, M.H., Bouwmeester, H.J. (2005) The Strigolactone Germination Stimulants of the Plant-Parasitic *Striga* and *Orobanche* spp. Are Derived from the Carotenoid Pathway. *Plant Physiology* 139(2): 920 – 934.

Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T., Weigt, M. (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *PNAS* 108(49): E1293 – E1301.

Moreira, B., Tormo, J., Estrelles, E., Pausas, J.G. (2010) Disentangling the role of heat and smoke as germination cues in Mediterranean Basin flora. *Annals of Botany* 105: 627 – 635.

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.

Musselman, L.J. (1980) The Biology of *Striga*, *Orobanche*, and Other Root-Parasitic Weeds. *Annual Reviews of Phytopathology* 18: 463 – 489.

Muth, T., García-Martín, J.A., Rausell, A., Juan, D., Valencia, A., Pazos, F. (2012) JDet: interactive calculation and visualization of function-related conservation patterns in multiple sequence alignments and structures. *Bioinformatics* 28(4): 584 – 586.

Nakamura, H., Xue, Y.-L., Miyakawa, T., Hou, F., Qin, H.-M., Fukui, K., Shi, X., Ito, E., Ito, S., Park, S.-H., Miyauchi, Y., Asano, A., Totsuka, N., Ueda, T., Tanokura, M., Asami, T. (2013) Molecular mechanism of strigolactone perception by DWARF14. *Nature Communications* 4: 2613. DOI: 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: 107 – 130.

Nelson, D.C., Flematti, G.R., Riseborough, J.-A., Ghisalberti, E.L., Dixon, K.W., Smith, S.M. (2010) Karrikins enhance light responses during germination and seedling development in *Arabidopsis thaliana. PNAS* 107(15): 7095 – 7100.

Nelson, D.C., Riseborough, J.-A., Flematti, G.R., Stevens, J., Ghisalberti, E.L., Dixon, K.W., Smith, S.M. (2009) Karrikins Discovered in Smoke Trigger Arabidopsis Seed Germination by a Mechanism Requiring Gibberellic Acid Synthesis and Light. *Plant Physiology* 149: 863 – 873.

Nelson, D.C., Scaffidi, A., Dun, E.A., Waters, M.T., Flematti, G.R., Dixon, K.W., Beveridge, C.A., Ghisalberti, E.L., Smith, S.M. (2011) F-box protein MAX2 has dual roles in karrikin and strigolactone signaling in *Arabidopsis thaliana*. *PNAS* 108(21): 8897 – 8902.

Ohno, S. (1970) *Evolution by Gene Duplication*. Springer-Verlag Berlin Heidelberg.

Oswald, A. (2005) *Striga* control – technologies and their dissemination. *Crop Protection* 24(4): 333 – 342.

Parker, C. (2013) The Parasitic Weeds of the Orobanchaceae. In Joel, D.M., Gressel, J., and Musselman, L.J. (Eds.), *Parasitic Orobanchaceae* (313-344). Heidelberg: Springer-Verlag Berlin.

Pérez-Vich, B., Akhtouch, B., Knapp, S.J., Leon, A.J., Velasco, L., Fernández-Martínez, J.M., Berry, S.T. (2004) Quantitative trait loci for broomrape (*Orobanche cumana* Wallr.) resistance in sunflower. *Theoretical and Applied Genetics* 109(1): 92 – 102.

Press, M.C., Phoenix, G.K. (2005) Impacts of parasitic plants on natural communities. *New Phytologist* 166(3): 737 – 751.

Rausell, A., Juan, D., Pazos, F., Valencia, A. (2010) Protein interactions and ligand binding: From protein subfamilies to functional specificity. *PNAS* 107(5): 1995 – 2000.

Renna, M., Serio, F., Santamaria, P. (2015) Crenate broomrape (*Orobanche crenata* Forskal): prospects as a food product for human nutrition. *Genetic Resources and Crop Evolution* 62: 795 – 802.

Reuter, B.C. (1986) The Habitat, reproductive ecology and host relations of *Orobanche fasciculata* Nutt. (Orobanchaceae) In Wisconsin. *Bulletin of the Torrey Botanical Club* 113(2): 110 – 117.

Ronquist, F., Huelsenbeck J.P. (2003) MRBAYES 3: Bayesian phylogenetic Inference under mixed models. *Bioinformatics* 19: 1572 – 1574.

Ruyter-Spira, C., Al-Babili, S., van der Krol, S., Bouwmeester, H. (2013) The biology of strigolactones. *Trends in Plant Science* 18(2): 72 – 83.

Sauerborn, J., Müller-Stöver, D., Hershenhorn, J. (2007) The role of biological control in managing parasitic weeds. *Crop Protection* 26(3): 246 – 254.

Scaffidi, A., Waters, M.T., Sun, Y.K., Skelton, B.W., Dixon, K.W., Ghisalberti, E.L., Flematti, G.R., Smith, S.M. (2014) Strigolactone Hormones and Their

Stereoisomers Signal through Two Related Receptor Proteins to Induce Different Physiological Responses in Arabidopsis. *Plant Physiology* 165(3): 1221 – 1232.

Schmitz, R.J., Schultz, M.D., Urich, M.A., Nery, J.R., Pelizzola, M., Libiger, O., Alix, A., McCosh, R.B., Chen, H., Schork, N.J., Ecker, J.R. (2013) Patterns of population epigenomic diversity. *Nature* 495(7440): 193 – 198.

Schneeweiss, G.M. (2007) Correlated evolution of life history and host range in the nonphotosynthetic parasitic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). *Journal of Evolutionary Biology* 20: 471 – 478.

Scholes, J.D., Press, M.C. (2008) *Striga* infestation of cereal crops – an unsolved problem in resource limited agriculture. *Current Opinion in Plant Biology* 11(2): 180 – 186.

Smith, S.D., Wang, S., Rausher, M.D. (2013) Functional Evolution of an Anthocyanin Pathway Enzyme during a Flower Color Transition. *Molecular Biology and Evolution* 30(3): 602 – 612.

Soundappan, I., Bennett, T., Morffy, N., Liang, Y., Stanga, J.P., Abbas, A., Leyser, O., 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.

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: 318 – 330.

Stevens, J.C., Merritt, D.J., Flematti, G.R., Ghisalberti, E.L., Dixon, K.W. (2007) Seed germination of agricultural weeds is promoted by the butenolide 3-methyl-2*H*-furo[2,3-*c*]pyran-2-one under laboratory and field conditions. *Plant and Soil* 298: 113 – 124.

Sun, X.-D., Ni, M. (2011) HYPOSENSITIVE TO LIGHT, an Alpha/Beta Fold Protein, Acts Downstream of ELONGATED HYPOCOTYL 5 to Regulate Seedling De-Etiolation. *Molecular Plant* 4(1): 116 – 126.

Sun, Y. K., Flematti, G.R., Smith, S.M., Waters, M.T. (2016) Reporter Gene-Facilitated Detection of Compounds in *Arabidopsis* Leaf Extracts that Activate the Karrikin Signaling Pathway. *Frontiers in Plant Science* 7: 1799. DOI 10.3389/fpls.2016.01799.

Toh, S., Holbrook-Smith, D., Stogios, P.J., Onopriyenko, O., Lumba, S., Tsuchiya, Y., Savchenko, A., McCourt, P. (2015) Structure-function analysis identifies highly sensitive strigolactone receptors in *Striga*. *Science* 350(6257): 203 – 207.

Tsuchiya, Y., McCourt, P. (2012) Strigolactones as small molecule communicators. *Molecular Biosystems* 8: 464 – 469.

Tsuchiya, Y., Yoshimura, M., Sato, Y., Kuwata, K., Toh, S., Holbrook-Smith, D., Zhang, H., McCourt, P., Itami, K., Kinoshita, T., Hagihara, S. (2015) Probing strigolactone receptors in *Striga hermonthica* with fluorescence. *Science* 349(6250): 864 – 868.

Uguzzoni, G., Lovis, S.J., Oteri, F., Schug, A., Szurmant, H., Weigt, M. (2017) Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis. *PNAS* 114(13): E2662 – E2671.

Van Staden, J., Brown, N.A.C., Jäger, A.K., Johnson, T.A. (2000) Smoke as a germination cue. *Plant Species Biology* 15: 167 – 178.

Van Valen, L. (1973) A New Evolutionary Law. *Evolutionary Theory* 1: 1 – 30.

Waldie, T., McCulloch, H., Leyser, O. (2014) Strigolactones and the control of plant development: lessons from shoot branching. *The Plant Journal* 79: 607 – 622.

Wang, H.-Y., Tang, H., Shen, C.-K. J., Wu, C.-I. (2003) Rapidly Evolving Genes in Human. I. The Glycophorins and Their Possible Role in Evading Malaria Parasites. *Molecular Biology and Evolution* 20(11): 1795 – 1804.

Wang, L., Wang, B., Jiang, L., Liu, X., Li, X., Lu, Z., Meng, X., Wang, Y., Smith, S.M., 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.

Waters, M.T., Nelson, D.C., Scaffidi, A., Flematti, G.R., Sun, Y.K., Dixon, K.W., Smith, S.M. (2012) Specialisation within the DWARF14 protein family confers distinct responses to karrikins and strigolactones in *Arabidopsis*. *Development* 139: 1285 – 1295.

Waters, M.T., Scaffidi, A., Moulin, S.L.Y., Sun, Y.K., Flematti, G.R., Smith, S.M. (2015) A *Selaginella moellendorffii* Ortholog of KARRIKIN INSENSITIVE2 Functions in Arabidopsis Development but Cannot Mediate Responses to Karrikins or Strigolactones. *The Plant Cell* 27(7): 1925 – 1944.

Waters, M.T., Smith, S.M. (2013) KAI2- and MAX2-Mediated Responses to Karrikins and Strigolactones Are Largely Independent of HY5 in *Arabidopsis* Seedlings. *Molecular Plant* 6(1): 63 – 75.

Welsh, S.L., Atwood, N.D., Goodrich, S., Higgings, L.C. (1993) *A Utah flora*. Provo, UT: Brigham Young University Press.

Westwood, J.H., dePamphilis, C.W., Das, M., Fernández-Aparicio, M., Honaas, L.A., Timko, M.P., Wafula, E.K., Wickett, N.J., Yoder, J.I. (2012) The Parasitic Plant Genome Project: New Tools for Understanding the Biology of *Orobanche* and *Striga*. *Weed Science* 60: 295 – 306.

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.

Wigchert, S.C.M., Kuiper, E., Boelhouwer, G.J., Nefkens, G.H.L., Verkleij, J.A.C., Zwanenburg, B. (1999) Dose-Response of Seeds of the Parasitic Weeds *Striga* and *Orobanche* toward the Synthetic Germination Stimulants GR 24 and Nijmegen 1. *Journal of Agricultural and Food Chemistry* 47: 1705 – 1710.

Xie, X., Yoneyama, K., Kisugi, T., Uchida, K., Ito, S., Akiyama, K., Hayashi, H., Yokota, T., Nomura, T., Yoneyama, K. (2013) Confirming Stereochemical Structures of Strigolactones Produce by Rice and Tobacco. *Molecular Plant* 6(1): 153 – 163.

Xie, X., Yoneyama, K., Yoneyama, K. (2010) The Strigolactone Story. *Annual Review of Phytopathology* 48: 93 – 117.

Xu, R., Chen, J., Chen, S.-L., Liu, T.-N., Zhu, W.-C., Xu, J. (2009) *Cistanche deserticola* Ma cultivated as a new crop in China. *Genetic Resources and Crop Evolution* 56(1): 137 – 142.

Yang, Z. (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* 24(8): 1586 – 1591.

Yang, Z., Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* 15(12): 496 – 503.

Yao, R., Ming, Z., Yan, L., Li, S., Wang, F., Ma, S., Yu, C., Yang, M., Chen, L., Chen, L., Li, Y., Yan, C., Miao, D., Sun, Z., Yan, J., Sun, Y., Wang, L., Chu, J., Fan, S., He, W., Deng, H., Nan, F., Li, J., Rao, Z., Lou, Z., Xie, D. (2016) DWARF14 is a non-canonical hormone receptor for strigolactone. *Nature* 536(7617): 469 – 473.

Yoshida, S., Ishida, J.K., Kamal, N.M., Ali, A.M., Namba, S., Shirasu, K. (2010) A full-length enriched cDNA library and expressed sequence tag analysis of the parasitic weed, *Striga hermonthica*. *BMC Plant Biology* 10:55. DOI: 10.1186/1471-2229-10-55.

Zhao, L.-H., Zhou, X.E., Wu, Z.-S., Yi, W., Xu, Y., Li, S., Xu, T.-H., Liu, Y., Chen, R.-Z., Kovach, A., Kang, Y., Hou, L., He, Y., Xie, C., Song, W., Zhong, D., Xu, Y.,

Wang, Y., Li, J., Zhang, C., Melcher, K., Xu, H.E. (2013) Crystal structures of two phytohormone signal-transducing α/β hydrolases: karrikin-signaling KAI2 and strigolactone-signaling DWARF14. *Cell Research* 23: 436 – 439.

Zhou, F., Lin, Q., Zhu, L., Ren, Y., Zhou, K., Shabek, N., Wu, F., Mao, H., Dong, W., Gan, L., Ma, W., Gao, H., Chen, J., Yang, C., Wang, D., Tan, J., Zhang, X., Guo, X., Wang, J., Jiang, L., Liu, X., Chen, W., Chu, J., Yan, C., Ueno, K., Ito, S., Asami, T., Cheng, Z., Wang, J., Lei, C., Zhai, H., Wu, C., Wang, H., Zheng, N., Wan, J. (2013) D14-SCF^{D3}-dependent degradation of D53 regulates strigolactone signalling. *Nature* 504: 406 – 410.

Zwanenburg, B., Mwakaboko, A.S., Kannan, C. (2016) Suicidal germination for parasitic weed control. *Pest Management Science* 72(11): 2016 – 2025.