TWO TO TANGO: ALLOPOLYPLOIDY AND ROOT NODULE SYMBIOSIS IN *GLYCINE* SUBGENUS *GLYCINE*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Adrian Federico Powell

January 2017

TWO TO TANGO: ALLOPOLYPLOIDY AND ROOT NODULE SYMBIOSIS IN *GLYCINE* SUBGENUS *GLYCINE*

Adrian Federico Powell, Ph.D.

Cornell University 2017

Polyploidy (or whole genome duplication) and root nodule symbioses with bacteria (termed 'rhizobia') have both been important phenomena in the evolutionary history of the legume family (Leguminosae). Recently, it has been proposed that polyploidy may have played a critical role in the development or refinement of nodulation. Given the varied potential impacts of polyploidy, effects on biotic interactions are plausible. However, direct studies of the interactions between these phenomena in symbiotic, nodule-forming species are lacking.

In this dissertation, using a complex of recently formed allopolyploids in *Glycine* subgenus *Glycine*, the perennial relatives of soybean, we examined (1) the root metabolites and symbiotic signaling capacity of multiple allopolyploid species relative to the diploid progenitor species that hybridized to form each allopolyploid, (2) the nodulation-related responses of allopolyploids and diploid progenitors to rhizobia and (3) the transcriptome-level responses to inoculation in allopolyploid *G. dolichocarpa* (T2) and its diploid progenitors. These objectives were pursued using a variety of approaches including root metabolite profiling, inoculation trials, and RNA sequencing.

We found that, while there were no common transgressive patterns in the root metabolite profiles of allopolyploids in the complex, one of the progenitors of T2 had distinctive root metabolite and exudate profiles; profiles of symbiotic signaling metabolites were also altered in

the allopolyploid. In addition, T2 demonstrated an enhanced symbiotic capacity when inoculated with rhizobia compared to its diploid progenitors. Finally, transcriptomic analyses revealed patterns of non-additive expression in T2, with evidence of transcriptional responses to inoculation that were distinct from one or both progenitors and that highlighted differences in hormonal signaling responses.

Taken together, our findings highlight effects of hybridity, and the properties of diploid progenitors, in determining the symbiotic capabilities of allopolyploids in *Glycine* subgenus *Glycine*. However, the combination of features from progenitors can enhance symbiotic responses in allopolyploids in ways that deviate from expectations of additivity. Such effects provide potential explanations for the apparent abilities of allopolyploids in *Glycine* subgenus *Glycine* to colonize novel habitats beyond their native ranges.

BIOGRAPHICAL SKETCH

Adrian Federico Powell is a plant biologist whose work focuses on using variety of metabolomic, transcriptomic and genomic approaches to explore the diversity of plant biotic interactions. He completed his B.Sc. degree at Trent University in 2008 and an M.Sc. at the same institution in 2010. Prior to commencing his studies at Cornell University, he also conducted research at the RIKEN Plant Science Center in Yokohama, Japan. Upon completion of his dissertation, he will continue his training in bioinformatics as a postdoctoral researcher at the Boyce Thompson Institute.

ACKNOWLEDGEMENTS

First, I want to acknowledge my dissertation committee for supporting me throughout this endeavor. To Jeff Doyle, my main dissertation advisor: thank you for providing me a fantastic opportunity for exploration and learning on this journey, for the innumerable insightful discussions and for your patience and continuing belief in me and my abilities – without your support none of this would have been possible. To André Kessler and Dan Buckley, thank you for your keen interest in teaching, for your willingness to share time and resources, and for your enthusiasm for my work. Many thanks to all members of the Doyle and Kessler labs over the years, with special thanks to Sue Sherman-Broyles. At Cornell, I would also like to thank Steve Vanek, Lynn Johnson and the Cornell Statistical Consulting Unit, and Alejandra Gandolfo for their guidance and willingness to help.

Throughout my time at Cornell I have been supported by NSERC, NSF, 3CPG, the Presidential Life Sciences Fellowship, the Barbara McClintock Award, the SIPS Section of Plant Biology and teaching assistantships through the Office of Undergraduate Biology. I am grateful for the support of all these programs, institutions and organizations.

Finally, I wish to thank all my friends at Cornell and elsewhere for making my time in graduate school a wonderful experience. Thanks also to my family members in Argentina and North America for their support. To Norma Pattuelli and Daniel Powell, my mother and father: thank you for all your support throughout my many years of study; I have only made it this far because of your constant love and encouragement. Finally, I would like to thank my partner, Elizabeth Fox, for all her support and love.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The use of nitrogen fertilizers in agriculture has increased greatly since the 1960s, contributing to substantial yield gains in that time, and this use is expected to increase globally over the next several decades (Sutton and Bleeker 2013). However, fertilizer use presents a number of challenges. Nitrogen fertilizers contribute to atmospheric pollution through the release of NO_x, which leads to the formation of ground level ozone, and N_2O, which is itself a potent greenhouse gas (Reay et al. 2012; Winiwarter et al. 2013). Nitrogen fertilizer use in agricultural systems is also linked to algal blooms and eutrophication (Diaz and Rosenberg 2008; Sutton et al. 2011). In addition, the fossil fuels required in the production and use of nitrogen fertilizers are environmentally costly (Jensen et al. 2012). Given the estimated costs, initiatives and opportunities to manage better or reduce the use and impact of nitrogen fertilizers are highly desirable (Beatty and Good 2011; Sutton and Bleeker 2013).

Legumes offer opportunities to address these challenges, because they engage in a process known as nodulation, which involves symbiosis with phylogenetically diverse soil bacteria collectively called rhizobia; these bacteria fix atmospheric nitrogen and make it available to the plant host (van Hameren et al. 2013). Legumes thus have the capacity to be an important source of nitrogen in sustainable agricultural systems (Crews and Peoples 2004). Knowledge of nodulation is being developed further and applied to agricultural practices through a variety of projects, including the N2Africa program funded by the Bill and Melinda Gates Foundation (Sprent 2012).

How the nodulation symbiosis originated remains an important evolutionary question (Doyle 2011, 2016). Several researchers, noting the strong correlation between nodulation and a whole genome duplication (WGD) event approximately 50 Ma in an ancestor of papilionoid legumes, have suggested that polyploidy played a significant role in either the origin or the

refinement of this key symbiosis (Young et al. 2011; Kim et al. 2013b; Li et al. 2013). The ability of polyploidy to generate evolutionary novelty is widely accepted (Freeling and Thomas 2006; Madlung 2013; Wendel 2015) because whole genome duplication affects every aspect of plant biology, from genetics and epigenetics to physiology and ecology (e.g., Coate et al. 2012; Ramsey and Ramsey 2014; Yoo et al. 2014). Much remains to be discovered about the impact of polyploidy on biotic interactions (Soltis et al. 2010), though recent work has shown that polyploidy can have a substantial effect on mycorrhizal symbioses (Těšitelová et al. 2013). However, in the case of nodulation, existing studies primarily make inferences from patterns of retention and expression of key genes that have been retained in duplicate for over 50 Myr (Young et al. 2011; Li et al. 2013), and more direct studies of the interaction between these phenomena are still to be conducted. In this dissertation, we present direct studies addressing this interaction.

Our study system is the perennial *Glycine* subgenus *Glycine* allopolyploid complex, which comprises several 'triads', each consisting of an allopolyploid and its two diploid progenitor species (Bombarely et al. 2014; Sherman-Broyles et al. 2014). With this complex, we have a model system for studying how nodulation is affected in polyploids. Specifically, we use a three-pronged approach, investigating plant signaling, studying differences in rhizobial species responses, and profiling transcriptomes of inoculated and uninoculated roots in several allopolyploid triads. This three-pronged approach relates directly to the three primary objectives of this dissertation: (1) to evaluate symbiotic signaling capacity in the T1, T2 (*G. dolichocarpa*) and T5 allopolyploid triads, compared to their respective diploid progenitors, by assessing biosynthesis and exudation of root metabolites; (2) to determine whether the allopolyploid species T2 exhibits differential symbiotic responses to inoculation with rhizobia, assessed for

characters such as ability to form nodulating associations, frequency of root hair deformation, and estimates of nodule mass and nodule number; (3) to study differences in gene expression following inoculation in T2 relative to its diploid progenitors.

Chapter 2 provides a critical introduction to, and review of, key considerations concerning potential interactions between nodulation-related signaling and polyploidy. In Chapter 3, we catalogue the presence of signaling metabolites in the root tissues and exudates of the three allopolyploids and their respective diploid progenitors, while also examining gene expression data relating to biosynthesis of flavonoids in T2 (Objective 1). Chapter 4 is a study testing for enhanced symbiotic capabilities and responses in inoculation trials of the allopolyploid T2 compared to its diploid progenitors (Objective 2). In Chapter 5, we examine overall transcriptomic responses of T2 to inoculation, and contrast these responses with those of its diploid progenitors (Objective 3). In the final chapter (Chapter 6), we summarize the research findings, highlight their contribution to the literature and their broader implications, and propose directions for future work.

CHAPTER 2

THE IMPLICATIONS OF POLYPLOIDY FOR THE EVOLUTION OF SIGNALING IN RHIZOBIAL NODULATION SYMBIOSIS

ABSTRACT

Nodulation is a symbiotic interaction between soil bacteria and plant hosts, most notably between rhizobia and legumes. This interaction is important for plant hosts, since it enables them to access atmospheric nitrogen made available by the bacteria. The establishment of this symbiosis requires several levels of communication and signaling between the plant host and the bacterial symbiont. The evolution of nodulation, and of the signaling mechanisms involved, appears to have involved recruitment of genes from pre-existing processes. Polyploidy, or whole genome duplication, has the capacity to provide duplicate genes that can be recruited for novel functions and genomic analyses have recently given rise to speculation that polyploidy has played an important role in the evolution of nodulation. Assessing the implications of polyploidy for nodulation enables predictions about possible outcomes and, while several challenges exist for studying the effects of polyploidy on nodulation signaling, testing of these predictions in a variety of study systems will yield insights into the evolutionary consequences of polyploidy for nodulation symbioses.

INTRODUCTION

Polyploidy, or whole genome duplication (WGD), and nitrogen-fixing symbioses have both been instrumental in contributing to the evolution and diversity of plants. Nitrogen-fixing symbioses involve interactions between host plants and bacterial symbionts, generally enabling the plants to acquire nitrogen that would be inaccessible otherwise (Vance 2001; White et al. 2007). Such symbioses, particularly those termed 'rhizobial' interactions that involve nodule-forming interactions with bacteria ('rhizobia') from numerous genera of *Alphaproteobacteria* and *Betaproteobacteria*, constitute an important aspect of the diversity of the legume family (Leguminosae) (Sprent 2009). WGD events, in turn, are also believed to have had an important role in the evolution and diversification of plants (Madlung 2013; Weiss-Schneeweiss et al. 2013). Given the importance of both polyploidy and nodulation, interest in the intersection between the two phenomena has increased in recent years.

Rhizobial nodulation symbiosis provides a model for understanding the implications of polyploidy for biotic signaling and interactions, given that many aspects of the signaling between plants and rhizobia have been elucidated. Similarly, a basic understanding of the potential evolutionary effects of polyploidy has been developed. Thus, it is possible to make predictions about the effects of polyploidy on interactions between plants and rhizobia by examining particular elements that relate to signaling. Genomic studies have also provided perspectives on the possible role of polyploidy in the evolution and refinement of nodulation symbiosis and signaling. Along with these recent advances and current opportunities, there are also challenges to the study of connections between these two phenomena.

Much of the seminal, critical work in understanding the signaling mechanisms underlying interactions between plants and rhizobia was conducted using limited sets of plant and rhizobial

genotypes, often under laboratory conditions. Such experimental systems are necessary, and they have yielded important results. However, we are also interested in rhizobial interactions at the level of *species*; in our case, our focus is primarily on the plant species that serve as hosts. Ultimately, in considering the interactions for a plant species, we would like to understand the changes in interactions that can occur at the level of species over time, as well as over spatial variability, which includes environmental interactions.

Researchers have elucidated numerous positive and negative determinants of rhizobial symbioses (e.g., Broughton et al. 2000; Keen and Staskawicz 1988; Oldroyd et al. 2011). Many of these involve signaling components, such as secondary metabolite signals in the rhizosphere (see below). The status of each determinant will yield a proximal outcome and, while some sets of determinants act in a sequential manner where the outcome of one determinant is not possible without a prior outcome, each determinant and its outcomes exist in a complex network of interacting determinants. The fitness consequences, benefits, and costs of the interactions will also interact with the other variables to modulate probabilities of outcomes and, ultimately, to determine the range of symbiotic interactions. Nevertheless, each signaling determinant or variable can be visualized as defining a 'signaling space' (Figure 2.1). This space represents the total hypothetical set of rhizobial genotypes for which the given variable under consideration in a particular legume host *genotype* will yield an outcome that permits symbiotic nodulation. This can also be conceptualized for a host plant *species*, where the species signaling space will comprise the rhizobial genotype by plant genotype (G_{plant} x $G_{rhizobium}$) combinations that will yield outcomes conducive to symbiosis. However, each variable, considered singly, may enable conducive outcomes with different sets of rhizobial genotypes, so as additional variables are considered and the signaling spaces for each variable are overlain, the possible symbiotic

interactions will generally become reduced and, as additional non-signaling variable spaces are included, the realized or actual symbiotic space, for a host plant species or particular genotype, is reached. The signaling components that serve as determinants in rhizobial interactions, and the potential implications of polyploidy for these determinants, can be understood in the context of the signaling spaces conceived in this manner. These spaces (notably the overall symbiotic space) will also be variable over space (environment) and time (species evolution).

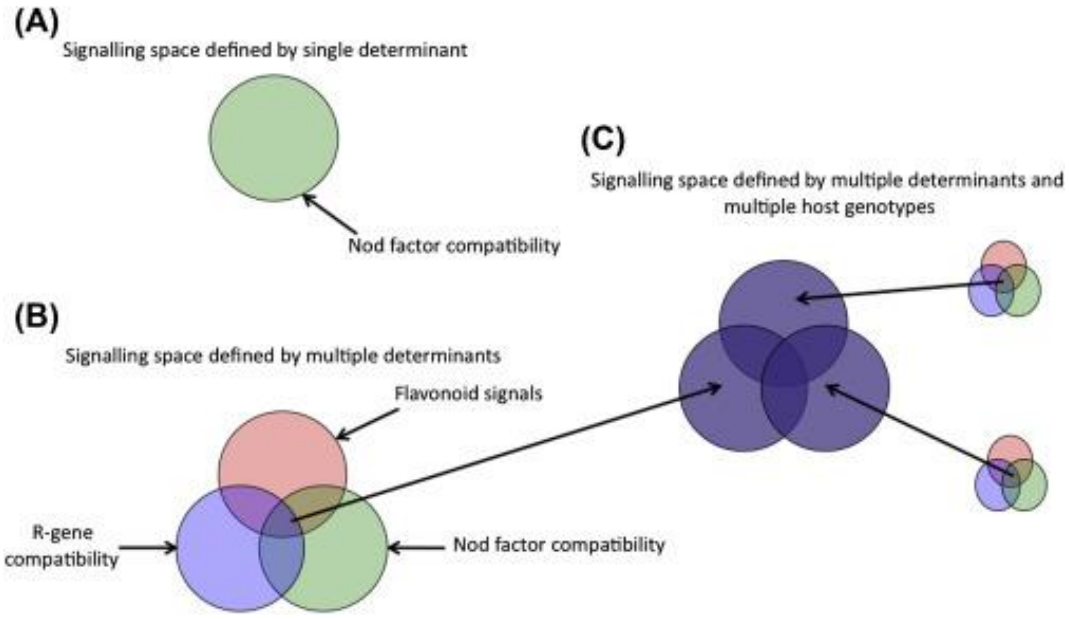


Figure 2.1. Visualization of the signaling spaces defined by determinants of symbiotic nodulation. Each determinant of nodulation can be understood as having a direct outcome that is needed for symbiosis. The direct outcomes of these determinants act to define the host-symbiont combinations that are effective (functional). (A) For any single host plant genotype, any single signaling determinant (e.g., flavonoid stimulation of Nod factor production, Nod factor binding to NFR genes, etc.) delineates a 'signaling space', which is the subset of all possible rhizobial genotypes for which the determinant will lead to a direct outcome that is necessary or conducive to nodulation. (B) As the signaling spaces of sequentially acting determinants are overlain (intersect), the composite signaling space will either remain the same or become more restricted. (C) This conception of the signaling space can be applied not only to a single plant host genotype, but also to a host species, where the species signaling space is defined as the set (union) of rhizobial genotypes with which the sum of host plant genotypes that make up the species can interact. As with the signaling space of a single plant genotype, one or multiple determinants can be considered.

NODULATION SIGNALING: DETERMINANTS OF SYMBIOTIC INTERACTIONS

Many aspects of the evolution of nodulation and nodulation signaling appear to be marked by the recruitment of pre-existing genetic capacities, including those related to developmental programs and symbiotic interactions, notably signaling components involved in the evolutionarily older fungal mycorrhizal mutualism and in microbial defense responses (Mathesius 2003). The recruitment or co-option of these programs ('exaptation' as defined by Gould and Vrba (1982)) also indicates that the concept of 'deep homology' is relevant to understanding the evolution of signaling in rhizobial symbiosis. This term describes cases where, even when an organ or trait (i.e., character) may have evolved independently in several lineages and is therefore considered non-homologous, the independent origins have involved recruitment of homologous programs or components (Shubin et al. 2009). Such deep homology and recruitment have been proposed for several aspects of rhizobial nodulation-related signaling (Doyle 2011; Soyano and Hayashi 2014).

Successful establishment of rhizobial nitrogen-fixing nodule-forming symbioses requires numerous steps, in which signaling plays a critical role in various forms and at various stages, including initial communication between the plant host and rhizobia, signal transduction and root hair responses, bacterial invasion, infection thread formation and growth, and nodule organogenesis (Oldroyd and Downie 2008). Many of the mechanisms and key steps have been elucidated, particularly in the root hair mediated rhizobial symbioses that are our primary focus (Figure 2.2). Variations in plant metabolites and molecules employed in rhizosphere signaling between plant and bacteria, the plant receptors for perception of bacterial signals, and additional signaling components required for nodulation can serve to enhance or lessen symbiotic interactions.

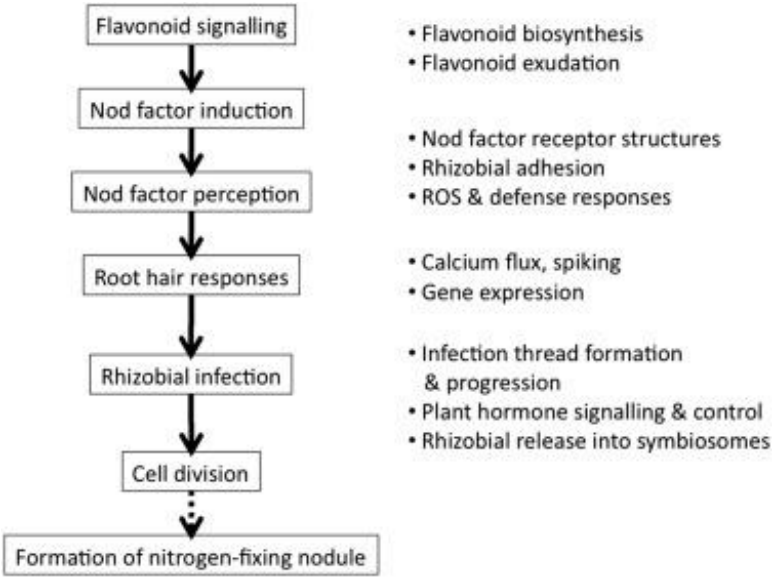


Figure 2.2. Summary schematic of steps required in the formation of rhizobial nodulation symbioses via root hair infection. Additional components involved at each stage are noted on the right. The interaction is initiated through the exudation of plant secondary metabolites, which lead to Nod factor induction in rhizobia. The Nod factors are perceived by the plant via Nod factor receptors. This is followed by signal transduction and a variety of responses, including root hair curling and formation of the bacterial infection thread. The infection proceeds to reach the site of cortical cell division, where rhizobia are released into symbiosomes.

The plant metabolites involved in establishing successful symbiotic relationships have been well-characterized through the use of various model papilionoid legume systems, including *Lotus japonicus, Medicago truncatula,* and *Glycine max* (soybean), along with corresponding rhizobial symbionts (Oldroyd 2013; Oldroyd and Downie 2008). For nodulation symbiosis to be established, the initial requirement is for plants to exude compounds that are able to trigger the induction of *nod* genes in the rhizobia. Compatible compounds, notably flavonoids, interact with the transcriptional regulator NodD in the bacteria, enhancing the DNA binding affinity of NodD to a consensus sequence in the promoter region of *nod* genes (termed the *nod* box), activating transcription of the *nod* genes (Peck et al. 2006). Noninducing flavonoids are competitive inhibitors, indicating that they also bind to NodD and can enhance *nod* box binding affinity of

NodD, but without leading to *nod* gene transcription (Peck et al. 2006). While evidence has been documented of *nod* gene induction by compounds such as aldonic acids (Gagnon and Ibrahim 1998), jasmonates (Mabood et al. 2006) and betaines (Phillips et al. 1992), flavonoids constitute the primary class of compounds implicated in *nod* gene induction in legumes (Cesco et al. 2010; Oldroyd 2013).

Induction of *nod* genes, in turn, leads to the production of Nod factors (lipochitooligosaccharides) that are secreted by rhizobial bacteria and perceived by the plant, triggering subsequent symbiotic responses. Perception of the compatible Nod factors leads to signal transduction, transcriptional activation of the early nodulation (ENOD) genes, and physiological responses including root hair curling and infection thread formation (Ferguson et al. 2010). The backbone of Nod factors is composed of *N*-acetylglucosamines, with an *N*-linked acyl chain on the terminal nonreducing residue (D'Haeze and Holsters 2002; Cullimore and Gough 2013). Substantial diversity is possible in the structure of Nod factors synthesized and secreted by rhizobia, and this is largely due to variability in the acyl group, as well as substituents at one or both of the nonreducing- and reducing- terminal residues (D'Haeze and Holsters 2002). Different species and strains of rhizobia have varying capacities for synthesizing Nod factors with distinctive structures (D'Haeze and Holsters 2002). In general, broad host range rhizobia produce a correspondingly broad array of Nod factors; narrow host range rhizobia produce a more restricted and specific set (Cullimore and Gough 2013). For example, the promiscuous strain NGR234 synthesizes a diverse family of Nod factors with various combinations of substitutions, encoded by a series of *nod* genes (Price et al. 1992). Following the induction and release of Nod factors by rhizobia in response to host plant compounds, the host plants then perceive the Nod factors, primarily through Nod factor receptors (NFRs). Here,

again, there is a requirement for compatibility between the Nod factor and the plant receptors. Diversity of Nod factor structures, with their decorations, and NFRs have been implicated in the specificity of interactions in a variety of legume genera (Cullimore and Gough 2013).

NFRs were identified and characterized in *L. japonicus* by Madsen et al. (2003) and Radutoiu et al. (2003). Radutoiu et al. (2003) cloned *NFR1* in *L. japonicus* and showed that it was necessary for the earliest responses to Nod factors. NFR5 was also identified as essential for nodulation in *L. japonicus* (Madsen et al. 2003). It has been proposed that the NFR5 and NFR1 proteins form a heterodimer and function together in Nod factor perception (Arrighi et al. 2006; Oldroyd 2013). The extracellular domains of NFR1 and NFR5 were found to be similar to the LysM (lysin motif) domains of previously-identified chitinases and peptidoglycan-binding proteins (Gust et al. 2012; Madsen et al. 2003; Radutoiu et al. 2003). Diverse LysM proteins are thus involved in binding a variety molecules that are structurally similar to Nod factors, including the microbe-associated molecular patterns (MAMPs) such as fungal chitin and bacterial peptidoglycan (Gust et al. 2012). Op den Camp et al. (2011a) also found that, in the only known instance of a non-legume that nodulates with rhizobial bacteria, *Parasponia andersonii*, a LysM Nod factor receptor protein involved in mycorrhizal association has been recruited for nodulation symbiosis. Their data suggest that the *PaNFP* gene identified serves a dual function in both mycorrhizal (Myc factor) signal reception and Nod factor reception, which is plausible given the apparent structural similarity of Nod factors to Myc factors as well (Gust et al. 2012).

The importance of NFRs in determining specificity between legume hosts and rhizobial partners has been demonstrated through transformation studies. The expression of the *NFR1* and *NFR5* genes from *L. japonicus* in *M. truncatula* and *L. filicaulis* enabled nodulation of these

latter species with *M. loti*, which commonly nodulates *L. japonicus* but not the other species (Radutoiu et al. 2007). While nodulation capability was transferred, the developmental activation in *M. truncatula* resulted in the indeterminate nodules typical for that species, as opposed to the determinate nodules formed by *L. japonicus*. Using domain swap and mutant experiments, the critical role of the LysM domains of the NFRs in determining specificity was also demonstrated (Radutoiu et al. 2007). Overexpression of receptors can lead to increased nodule numbers (Indrasumunar et al. 2011), suggesting that the amount or density of receptors can also lead to variation in the amount of interaction with Nod factor molecules, and the consequent response, that is possible. While most evidence for the role of Nod factor receptors in nodulation was obtained through studies of mutants, more recently, Broghammer et al. (2012) obtained direct evidence showing binding of NFR1 and NFR5 to Nod factors. In addition to Nod factor receptors, lectin nucleotide phosphohydrolases (LNPs) have also been identified as proteins that bind to Nod factors at the plant surface and are thus also thought to mediate symbiotic interactions with rhizobia (De Hoff et al. 2009; Etzler et al. 1999).

Toll-interleukin receptor/nucleotide-binding site/leucine-rich repeat (TIR-NBS-LRR) resistance genes (*R* genes) have also been implicated in determining rhizobial interactions in cultivars of *Glycine max* (Yang et al. 2010). *Rfg1* and *Rj2* alleles are known to confer the specificity for nodulation with particular strains of *Bradyrhizobium japonicum* and *Sinorhizobium fredii* (e.g., Nakano et al. 1997)(see summary in (Hayashi et al. 2012a)), the characterization of these alleles as *R* genes is a more recent development (Yang et al. 2010). The particular mechanisms of action or the roles of these *R* genes in inhibiting nodulation remain unresolved, although it appears that failed infection thread growth is involved, which may be due to plant defense responses by perception of bacterial effectors, leading to recognition of

incompatible rhizobia as parasites rather than mutualists (Yang et al. 2010). Another gene, *Rj4*, which is known to restrict nodulation of *G. max* with specific incompatible strains has recently been found to encode a thaumatin-like protein; it is a member of the pathogenesis-related protein family 5, which is implicated plant host resistance (Hayashi et al. 2014). Proteins, including nodulation outer proteins, are also secreted by rhizobia through type III and type IV secretion systems and these proteins have additional potential for both positive and negative effects on symbiosis, which further emphasizes the importance of plant defense in determining rhizobial interactions and also highlights the likely evolutionary connections between pathogenic and mutualistic symbioses (Deakin and Broughton 2009).

In addition to root hair responses, symbiotic responses are also triggered in cortical cells, leading to the expression of additional genes required for nodule organogenesis (Oldroyd and Downie 2008). Plant hormones have been implicated in several aspects of signaling and regulation of nodulation. Cytokinins and cytokinin-related signaling, in particular, appear to have important roles in the development and regulation of nodule formation. A histidine kinase cytokinin receptor in cortical cells is necessary for signaling subsequent to the root hair cell responses following initial contact with rhizobia. RNA interference of the cytokinin receptor homolog in *M. truncatula* (*MtCRE1*) resulted in strongly reduced nodulation, including decreased progression of rhizobial infection and formation of nodule primordia (Gonzalez-Rizzo et al. 2006). Gain-of-function mutation of the orthologous gene in *L. japonicus* led to spontaneous nodule organogenesis (Tirichine et al. 2007). Experiments in *L. japonicus* overexpressing cytokinin oxidase, the primary enzyme for degradation of cytokinins, also led to both a decrease in the number of nodules formed per plant and an increase in the number of lateral roots (Lohar et al. 2004). In contrast to this positive role in nodule formation, recent work

also suggests that cytokinins may be implicated in the shoot-derived inhibition of nodulation that is part of the autoregulation of nodulation (AON) pathway in legumes (Sasaki et al. 2014). AON is another process that provides feedback regulation of nodule number, and genes in this pathway have been identified from mutants of several species, including soybean, possessing increased nodule numbers (Miyahara et al. 2008; Searle et al. 2003). Other hormones have also been found to have roles in the control of nodule development and nodule number, including ethylene, which is a negative regulator of nodule number (Gresshoff et al. 2009; Lohar et al. 2009; Penmetsa et al. 2008).

In the context of the evolution of nodulation, the genes involved in many of the processes underlying nodulation, including signaling, often appear to have additional, related functions within a given plant species (Yokota and Hayashi 2011). This confirms the long-held hypothesis that, in many cases, genes with pre-existing functions have been recruited to play roles in nodulation (Doyle 1994). A notable example of this type of recruitment is apparent in the early stages of nodulation signaling, where the genes involved in the common symbiosis pathway were recruited from the older arbuscular mycorrhizal symbiosis (Soyano and Hayashi 2014). These genes include those encoding nuclear cation channels (*POLLUX*, *CASTOR/DMI1*), a plasma membrane localized LRR receptor kinase (*SYMRK/DMI2*), and a calcium-dependent and calmodulin-dependent protein kinase (*CCaMK*/*DMI3*), among others (Ané et al. 2004; Charpentier et al. 2008; Endre et al. 2002; Lévy et al. 2004; Mitra et al. 2004; Stracke et al. 2002).

WHOLE GENOME DUPLICATION AS A PROCESS ALTERING PLANT SIGNALING, BIOTIC INTERACTIONS AND ECOLOGICAL AND GEOGRAPHICAL RANGES

Polyploidy is a prominent process in the evolutionary history of flowering plants, having been suggested to underlie their origin and diversification (De Bodt et al. 2005; Vanneste et al. 2014). Wood et al. (2009) estimated that 15% of all speciation events in flowering plants have involved polyploidy. Moreover, it is now recognized that all flowering plants have a polyploid past (Jiao et al. 2011), and that the genomes of most modern species show evidence of more than one WGD—for example, the soybean genome shows evidence of the two polyploidy events shared with all flowering plants, a whole genome triplication shared with eudicots (Jiao et al. 2011), a WGD shared with many legumes (Cannon et al. 2015), and a 5-10 MYA polyploidy event unique to the genus *Glycine* (Egan and Doyle 2010). Thus, it is indisputable that polyploidy has produced successful lineages and has played a significant role in plant evolution.

Recruitment of duplicated genes can provide a means of evolving novel or distinct functions. For example, recent work on nodulation suggests that gene duplication has given rise to the GRAS transcription factors NSP1 and RAM1, which function downstream of the common symbiosis pathway (Gobbato et al. 2012). RAM1 has a role primarily in the mycorrhizal symbiosis, while NSP1 has a role in nodulation and does not appear to have any function in plant response to mycorrhizal signaling (Gobbato et al. 2012). In the case of the gene family of *MtLYK3-LjNFR1a*, several duplication events are believed to have preceded the evolution of nodulation, the origin of legumes and thus any polyploidy events in the legumes, nevertheless contributing to the evolutionary trajectory toward nodulation (De Mita et al. 2014).

If the duplication and subsequent evolution of single genes can contribute to the evolution of novel traits, including diversification of symbiotic interactions and components of

nodulation signaling, duplication of the entire set of genes as well as regulatory sequences by polyploidy has the potential to greatly amplify these processes, providing duplicates across entire biosynthetic, signaling and regulatory networks. Evolution is driven by mutation, and polyploidy can be thought of as a super-mutation, leading to evolutionary novelty; Freeling and Thomas (2006) have suggested that the retention and recruitment of duplicated genes, specifically by polyploidy, is a major driver of morphological novelty and complexity.

While gene duplicates are often lost following a WGD event, several models suggest ways in which natural selection can lead to retention of both paralogous copies, such as neofunctionalization and subfunctionalization (Innan and Kondrashov 2010). In the case of nodulation signaling, if a gene with a pre-existing function were duplicated, one copy could be recruited for a novel function in nodulation, while the other copy could maintain the original function. This provides an example of neofunctionalization. Alternately, if a gene developed both nodulation-related and non-nodulation-related functions and was subsequently duplicated, the two functions could be divided between the two resultant copies, and subsequently each paralogue could evolve for specialization in each function. This process is referred to as subfunctionalization.

Quantitative dosage effects will also play a role in determining the initial retention of certain types of gene duplicates; WGD can provide an opportunity for retention of genes involved in pathways where selection for specific stoichiometry means that single gene duplications are less likely to be favored, while the maintenance of stoichiometric relationships occurring through a WGD would be favored (Birchler and Veitia 2012). The classes of genes that are subject to dosage balance in this manner often include transcription factors and the components of signaling pathways (Blanc and Wolfe 2004; Maere et al. 2005; Paterson et al.

2006). In the context of nodulation, retention of duplicates in both of these cases could contribute to the evolution of novel signaling or the enhancement and modulation of existing signaling mechanisms.

Autopolyploidy and allopolyploidy are two endpoints on a spectrum of genetic and taxonomic possibilities for WGD. Autopolyploidy is typically defined as a WGD event occurring involving genetically similar genomes, often from within the same species. In contrast, allopolyploidy involves a WGD event coupled with hybridization between two or more progenitors, often from distinct species. Both result in doubled genomes and the concomitant effects specifically related to an increase in DNA content (nucleotypic effects, e.g., Bennett 1972), while allopolyploidy also involves potential effects of genome merger, such as heterosis (Washburn and Birchler 2014). In an allopolyploid's complex genome, genetic loci that were orthologous in the diploid progenitors are termed homoeologues (Wendel and Doyle 2005). Allopolyploids thus at least initially have the opportunity to deploy the homoeologous gene copies from both progenitors, often with variations in homoeologue expression bias (Chen 2010; Grover et al. 2012; Yoo et al. 2013). WGD can result in transgressive, novel expression patterns and homoeologue silencing in natural polyploids (Yoo et al. 2014). The changes to patterns of epigenetic regulation and gene expression are referred to as 'genomic shock' (McClintock 1984). Furthermore, unequal expression of homoeologous copies is thought to lead to fractionation (loss of one copy in each homoeologue gene pair); bias in expression of homoeologous copies can be a legacy of diploid progenitors, with one transcriptome dominant over the other (Buggs et al. 2014). It is hypothesized that in allopolyploids this process can lead to biased fractionation across the duplicated genome (Garsmeur et al. 2014; Schnable et al. 2011). In addition to

fractionation, biases in gene conversion can also alter the expected additive pattern of homoeologous copies in allopolyploids (Flagel et al. 2012).

Polyploidy can thus lead to substantial chromosomal rearrangement or loss, genetic and epigenetic changes, and gene expression changes that, while in many cases deleterious, have the capacity to generate increased variability and potential for adaptation. Allopolyploidy, in particular, can confer advantages in changing environments, including variable climate and habitat disturbance (Combes et al. 2012; Fawcett and Van de Peer 2010; Stebbins 1985). Furthermore, polyploids are often found to show an increased ability to colonize new habitats and alter their ranges. Evidence indicates that polyploidy can broaden environmental distributions and that polyploids can have larger ranges than their diploid progenitors (Green et al. 2011, 2013; Hijmans et al. 2007; Lowry and Lester 2006), as would be expected of plants with a greater capacity for colonization, although it may not be a general tendency (Harbert et al. 2014; Martin and Husband 2009). Pandit et al. (2011) conducted a comprehensive study and meta-analysis of invasive species in the world, and found that polyploidy correlated strongly with invasiveness. This could be due to alterations that affect polyploids in numerous ways and incorporate a variety of effects of polyploidy including phenotypic plasticity, ecological tolerance and niche differentiation (te Beest et al. 2012). Of particular interest here, te Beest et al. (2012) highlighted the importance of biotic interactions and, in particular, changes to biotic signaling for apparent increased invasiveness and colonization ability of polyploids. Connecting invasiveness of polyploids in this manner suggests that polyploids can possess broadened signaling spaces, compared to related diploids. Given the emerging awareness of the significance of the microbiome in plant and animal health and ecology (Berendsen et al. 2012; Cho and Blaser 2012; Ezenwa et al. 2012), it is possible that a polyploid's ability to occupy and colonize

new habitats or geographic ranges may be partly due to its ability to interact with diverse microbes. In the case of legumes, this could include compatible rhizobial signaling interactions available to the plant, coupled with the supply and demand for available soil nitrogen.

Given the many genetic, physiological and ecological changes that can result from polyploidy, researchers have begun examining the effects of polyploidy on biotic interactions. While the effects of polyploidy on biotic interactions are not yet fully elucidated (Soltis et al. 2010), recent research has yielded progress. Several studies have established that polyploidy has the potential to generate differences in, and diversification of, biotic interactions. In pollination interactions, *Heuchera grossulariifloria* polyploids have suites of pollinators that differ in the relative visitation rates and contributions to seed set of particular pollinators between diploid and autopolyploid plants (Segraves and Thompson 1999; Thompson and Merg 2008; Thompson et al. 2004). Thompson and Merg (2008) note that these differential visitation rates of pollinators on diploids and polyploids are likely due to chemical signaling cues employed by the pollinators, in addition to any morphological differences, since certain diploid and polyploid populations studied were morphologically indistinguishable; explanations in terms of chemical signaling cues suggest alterations to the signaling space in the polyploids. Damage from herbivory also varies between diploids and polyploids in a number of study systems (Hull-Sanders et al. 2009a; Münzbergová 2006; Nuismer and Thompson 2001). Differences in performance of the insect herbivore *Spodoptera exigua* on *Solidago gigantea* (giant goldenrod) plants of differing cytotypes were also attributed to changes in secondary chemistry (Hull-Sanders et al. 2009a, b), indicating further modulation of the interaction space.

That symbiotic interactions with soil microorganisms can also be affected by polyploidy has been shown in relation to arbuscular mycorrhizal (AM) associations. Studies of the

Gymnadenia conopsea group (Orchidaceae) found that plants of different ploidy levels formed associations with distinct suites of mycorrhizal fungi (Těšitelová et al. 2013). In this work, the segregation of mycorrhizal symbionts by ploidy level was most strongly observed in the comparison of diploids and derived autopolyploids growing in close proximity at sites containing mixtures of plants of several ploidy levels. This research suggests the potential for differences in signaling between plants of varying ploidy and mycorrhizal symbionts in that the interactions with different suites of symbionts may be indicative of signaling specificity and differentiation of signaling spaces. Subsequent work has confirmed interactive effects of ploidy and AM inoculation in other genera, affecting plant growth and mycorrhizal parameters (Sudová et al. 2014), which also suggests effects of polyploidy on signaling and interaction spaces. Differing growth responses to particular symbionts between autopolyploids and diploids also suggest that divergent selection can lead to differentiation of signaling subsequent to polyploid formation. However, the effects of polyploidy can vary with taxon. In a study by Sudová et al. (2010), differences in mycorrhizal growth responses were not observed between diploids and polyploids for *Campanula gentilis* and *Pimpinella saxifraga*, though the growth of *Aster amellus* diploids was positively responsive to mycorrhizal inoculation, while hexaploids were either not responsive or negatively responsive. In these experiments, ploidy did not have any effect on mycorrhizal parameters, including extraradical mycelial growth and the extent of mycorrhizal colonization (Sudová et al. 2010).

This research indicates the potential for plant polyploidy to alter interactions between plants and various types of symbionts, notably soil microorganisms, and suggests that alterations in signaling may be connected to the differences observed. The next section explores the

potential implications of polyploidy for nodulation signaling and symbiosis, with attention devoted to effects on particular components of signaling.

POTENTIAL IMPLICATIONS OF POLYPLOIDY FOR NODULATION SIGNALING

The visualization of symbiotic and signaling spaces introduced earlier is useful in considering effects of polyploidy, since polyploidy can alter these spaces by affecting diverse processes in plants, leading to a variety of possible outcomes. If we consider a scenario with two diploid progenitors with distinct but overlapping hypothetical symbiotic spaces, we can predict a variety of outcomes in the symbiotic space of an allopolyploid, including additive, parental, and transgressive patterns (Figure 2.3). The following sections will explore a number of effects through which polyploidy can alter these hypothetical symbiotic spaces. We will consider alterations to metabolite signaling in exudates, signal reception, as well as additional signaling mechanisms and effects on signaling.

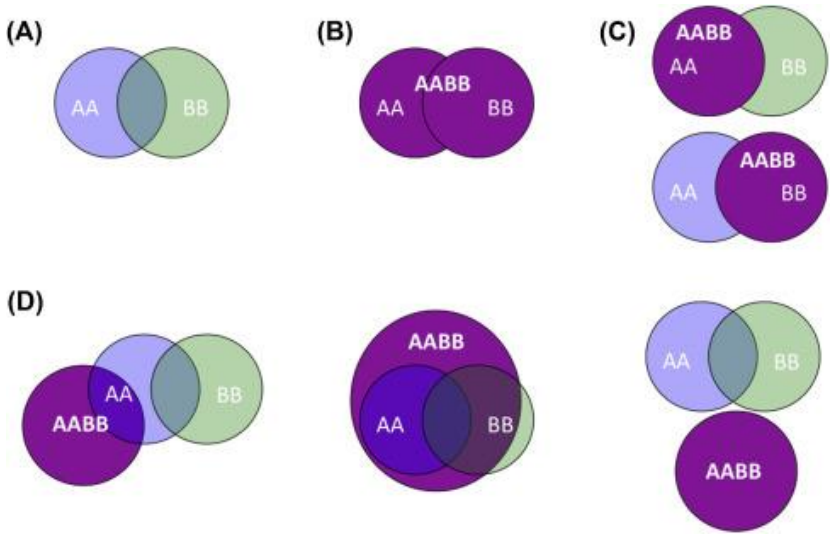


Figure 2.3. Visualizations of species-level signaling spaces (defined by multiple determinants) and outcomes of allopolyploidy. (A) The composite signaling space illustrated in Fig. 1C is shown for two diploid species (genome designations AA and BB), representing the rhizobial genotypes for which interactions are possible with plant genotypes (A_1A_1, A_1A_2, A_2A_2, … ; B_1B_1, B_1B_2, B_2B_2,…) that make up each host species. In this case, the two species have partly

overlapping, partly non-overlapping signaling spaces. When an allopolyploid is formed from the two diploid progenitors, a variety of outcomes are possible for the signaling space of the allopolyploid, including: (B) a signaling space that comprises the summed signaling spaces of both diploids, (C) a signaling space that is the same as one diploid or the other (possibly due to dominance of one homoeologous genome over the other), or (D) some combination of the signaling spaces of the diploid progenitors, including potential transgressive (non-parental) signaling spaces.

Flavonoids

Polyploidy can alter the biosynthesis of secondary metabolites. These effects include increased accumulation and concentration of secondary metabolites (e.g., Hull-Sanders et al. 2009a; Lavania et al. 2012), including changes in flavonoid pathways (Griesbach and Kamo 1996). Qualitative variations in flavonoid profiles between ploidy levels, including novel compounds in polyploids that are not found in progenitor diploids, have also been documented (Levy and Levin 1971, 1974). Historically, such qualitative differences in flavonoid profiles, particularly when additive, were often used as evidence for determining the parentage of allopolyploids (Mears 1980).

The synthesis of novel, more diverse flavonoids, and novel combinations of flavonoids, can occur as a result of diversification of genes affecting the biosynthetic processes by neofunctionalization or through the effects of hybridity (including differential regulation or novel regulatory patterns) in an allopolyploid. Qualitative differences in exuded flavonoids can alter symbiotic outcomes because, as noted above, there is specificity between particular flavonoid compounds and the response of *nod* gene induction in particular strains of rhizobial bacteria (Aoki et al. 2000; Long 1996). A given flavonoid can have distinct inducing and inhibiting effects on Nod factor synthesis across various species and strains of rhizobia (Kape et al. 1991; Kosslak et al. 1987; Kosslak et al. 1990; Yokoyama 2008). Similarly, polyploidy-induced alterations to gene expression that affect enzymes involved in the synthesis of signaling

compounds could have quantitative effects on symbiotic interactions, increasing or decreasing the magnitude and intensity of the signaling. It is also conceivable that, in terms of absolute exudation on a per-plant basis, changes to the biomass of biosynthetic tissues in polyploids could also increase or decrease the total amount of flavonoids produced. *In vitro*, increases in flavonoid concentration can lead to increased Nod factor synthesis (e.g., Yokoyama 2008), which in turn can lead to different calcium responses (Shaw and Long 2003). In optimized hydroponic systems, exogenous application of inducing flavonoids stimulates nodulation (Novak et al. 2002). Such induction responses can increase with flavonoid concentration over a certain range, beyond which, in certain cases, the response decreases, potentially due to toxicity effects (Bolaños-Vásquez and Werner 1997; Novak et al. 2002; Pueppke et al. 1998). In general, any change in biosynthesis of particular flavonoids that affects the profile of exuded compounds would be expected to lead to alterations in the magnitude of induction or inhibition of Nod factor synthesis, which would alter the probabilities of outcomes within the signaling space defined by flavonoid signaling.

In the context of signaling spaces, one can propose a broad, hypothetical signaling space delineated solely by the exudation of flavonoids and their capacity for inducing or inhibiting Nod factor production. The size of the signaling space can initially be taken as all rhizobial genotypes for which a plant's exudate yields a non-zero probability of Nod factor induction. The relative size of the signaling space under consideration can be increased or decreased depending on setting a biologically relevant Nod factor induction threshold. The exclusion or inclusion of a rhizobial strain from a particular signaling space established by a given Nod factor induction threshold will be determined by the total Nod factor inducing capability of the mixture of compounds in the exudate of a plant. This could be calculated by summing the products of the

total amount of each compound released by the plant and taken in by a given bacterium, and the Nod factor inducing or inhibiting effect of each compound. Additional terms are also required to account for interactive effects between the compounds that will change the total induction or inhibition capacity of the exudate for any particular rhizobial genotype. In this framework, the effect of quantitative or qualitative changes in flavonoid profiles due to polyploidy could be approximated by comparing the concentrations of compounds and inductive capabilities of the compounds in the polyploids with those of the diploids.

While changes in the biosynthesis of flavonoids due to polyploidy are critical to consider, in the context of nodulation, it is the exudation of flavonoids from the root into the rhizosphere that affects the potential for signaling with rhizobial bacteria. This requires consideration of ATP-binding cassette (ABC) and MATE transporters. ABC-type transporters have been implicated in the exudation of flavonoids from legume root cells and evidence suggests they could have a variable affinity for different flavonoid compounds (Sugiyama et al. 2007). Competitive inhibition assays using soybean plasma membrane vesicles found that the isoflavonoid daidzein, as well as biochanin A and formononetin, inhibited genistein transport, while the flavonone naringenin and the flavonol kaempferol did not, suggesting the latter were not readily transported (Sugiyama et al. 2007). Genistin, the glucoside of genistein, also inhibited genistein transport, but appeared to do so to a lesser degree than the daidzein, biochanin A and formononetin (Sugiyama et al. 2007). MATE transporters have not been reported as being involved in exudation from the root or transport across the plasma membrane, but several MATE transporters have been found that are involved in transport of flavonoids into the vacuole, leading to suggestions that MATE transporters may also play a role in flavonoid exudation (Gomez et al. 2009; Marinova et al. 2007; Zhao and Dixon 2009). Given the importance of

transporters in exudation and varying capacities for transporting different flavonoids, simply synthesizing novel compounds alone would not necessarily translate into their exudation from the plant root tissues; different transporters may be required for effective exudation of particular flavonoids. The observation that mutations to individual transporters can lead to changes throughout the biosynthetic networks involved in flavonoid production (e.g., Zhao et al. 2011) suggests that changes to the number and specificities of transporters could have a profound influence on rhizosphere signaling. Thus, for synthesis of novel or more diverse flavonoids in polyploids to affect nodulation signaling, the capacity to transport the novel array of compounds must also be present among the transport proteins that participate in the exudation of flavonoid compounds from the root tissues to the rhizosphere. In a polyploid, the genes encoding the necessary transporters may be available due to their presence in the diploid progenitors; alternately, novel transport capacities could theoretically be developed by mutation (creation of new alleles) or by neofunctionalization of duplicated genes.

Similarly, quantitative increases in the biosynthesis of flavonoids would not necessarily lead to greater exudation, but would depend on the density of transporters available for exudation. This would, in turn, depend in a complex manner on the number of transporters per cell and the external cell surface area. This is not expected to be constant between polyploids and diploids, since cell size and, therefore, surface-to-volume ratios, are often altered by WGD (Coate et al. 2012; Lavania et al. 2012; Levin 1983). Alterations to transcriptome size and differential translational regulation will also have effects on the number of flavonoid transporters on the root surface; both types of changes have been documented in polyploids (Coate et al. 2014; Coate and Doyle 2010, 2015) (Coate and Doyle, 2010; Coate et al., 2014; Coate and

Doyle, 2015). These aspects will also be relevant in the following section on genetic diversity and gene expression.

Although flavonoids can be important in the attraction of rhizobia, they are not necessarily the primary compounds that serve this purpose. Early studies on the chemotaxis of rhizobia found that with some rhizobial species, such as *Rhizobium leguminosarum* and *Sinorhizobium meliloti*, chemotaxis occurred in response not only to flavonoids but also to nutrients, specifically organic acids (Aguilar et al. 1988; Armitage et al. 1988; Caetano-Anolles et al. 1988). In these species, flavonoids that were *nod* gene inducers were also found to be chemotactic attractants. However, in *Bradyrhizobium japonicum*, chemotaxis occurred in response to nutrients but not flavonoids (Barbour et al. 1991; Kape et al. 1991). In this case, it appeared that chemotactic molecules were distinct from *nod* gene inducers. More recent work has examined varying chemotaxis of *B. japonicum* strains to nutrient attractants (Althabegoiti et al. 2008), and also explores the bacterial genes and receptors involved in perception of chemotactic signals in various rhizobial species (Miller et al. 2007; Webb et al. 2014; Yost et al. 1998). Given the chemotactic importance of non-flavonoid compounds, the impacts of polyploidy on rhizobial interactions may extend beyond the effects on flavonoid biosynthesis and exudation to the synthesis and exudation of organic acids and other nutrients as well, when considering the contribution of chemotaxis as a variable defining the signaling space.

Nod factor receptors

In the context of nodulation signaling, polyploids have the potential to possess increased diversity of Nod factor receptors, thereby enabling perception of a broader array of rhizobial signals. As fixed hybrids, allopolyploid species can, at least initially, combine Nod factor receptors from different diploid species. Neofunctionalization and further diversification of Nod

factor receptors could increase the capacity to perceive a broader array of bacterial Nod factors having diverse structural properties. A general prediction is that greater Nod factor receptor diversity should be associated with a greater range of bacterial symbiont partners with which root hair curling is elicited (Indrasumunar et al. 2010; Radutoiu et al. 2007). Thus, one hypothesis is that the allopolyploids, as fixed hybrids, will have equal or greater symbiont ranges (interaction spaces) relative to their diploid progenitors. However, tests of selection and differences in expression may reveal early fractionation, neutral evolution and loss of particular gene copies, in which case the allopolyploid may not have an increased diversity of receptor genes and the expected concomitant increased symbiotic range. Further, polyploidy-related modulation of the expression of plant receptors could also have quantitative effects on symbiotic interactions, increasing or decreasing the magnitude and intensity of the signaling.

Resistance genes

R genes encode R proteins, which play an important role in plant defense against pathogens. This contribution to resistance can be achieved by several potential mechanisms, including through the R protein binding directly to pathogen avirulence molecules (effectors) or by monitoring cellular targets of effectors (according to the ‘guard hypothesis’; Jones and Dangl 2006). Tandem duplications and segmental genomic duplications have contributed to *R* gene family expansion, though paralogous genes may also be lost following duplication in some cases, and duplicated genes have subsequently diverged due additional accumulated mutations (Meyers et al. 2005). Thus, both genome duplication and fixed hybridity (in allopolyploids) have the capacity to increase diversity and create novel combinations of *R* genes, which in turn would affect potential biotic interactions, including nodulation. In particular, variations on the allelic combinations of *Rfg1* and *Rj2* in *G. max*, mentioned earlier, could enhance or reduce the ability

of a polyploid to form nodules with particular bacterial species or strains. The general predictions for *R* genes in polyploids stand in opposition to those for the Nod factor receptors, because greater diversity of *R* genes would be expected to lead to a more restricted symbiont range, given that particular alleles at the locus serve to prevent nodulation with certain symbionts. Thus, when considering interaction outcomes defined by *R* genes, one could predict that allopolyploid fixed hybrids, for example, would have an equal or more restricted signaling space than their diploid progenitors, unless the allopolyploids silence *R* genes from one progenitor. That *R* genes may belong to the general class of 'duplication resistant' genes (Paterson et al. 2006) is plausible given preferential NB-LRR gene loss in *G. max* following polyploidy (Ashfield et al. 2012). The general lability of *R* genes is suggested by extreme copy number variation in several plant taxa (Zhang et al. 2010).

Additional avenues for effects of polyploidy on signaling mechanisms via plant growth and cytokinins

Polyploidy often has effects on plant growth. In general, wild autopolyploids, along with induced neopolyploids, tend to have larger organs than related diploids (Ramsey and Schemske 2002; Ramsey and Ramsey 2014). Such changes to mature plant size and plant growth rates can be related to alterations to nutrient demand and uptake. Early studies found greater tolerance of *Nicotiana* polyploids for growth in nutrient-poor calcareous soils (Noguti et al. 1940); more recent work also found greater tolerance to calcium treatments in polyploid *Solidago gigantea* (Schlaepfer et al. 2010) and greater salinity tolerance in tetraploid *Brassica rapa* (Meng et al. 2011). Effects of ploidy resulting in increased or decreased nutrient uptake efficiency, however, appear to be species-specific, in the cases of potassium and sulfate (Cacco et al. 1976). If polyploidy results in alterations to nitrogen uptake and demand, such effects would be likely to

yield changes in rhizosphere signaling. Flavonoid and metabolite exudation into the rhizosphere is changed by high nitrogen treatments: nitrogen deficiency can trigger the biosynthesis of flavonoids and nitrate application decreases flavonoid accumulation, including the nodulation signaling isoflavonoids daidzein and genistein in soybean (Cho and Harper 1991; Lea et al. 2007). In addition to effects on secondary metabolite accumulation, changes in plant growth and nutrient regulation are also connected to differences in hormonal concentrations and signaling (Depuydt and Hardtke 2011; Kiba et al. 2011; Krouk et al. 2011). Cytokinin levels, as noted above, have the potential to affect the number of nodules formed (Lohar et al. 2004; Sasaki et al. 2014); thus certain changes to hormonal signaling in polyploid plants can have additional effects on symbiotic interactions. Furthermore, cytokinins have the capacity to repress a number of transporters of nutrients, including nitrate, ammonium, sulfate and phosphate (Sakakibara et al. 2006), while cytokinin transport and biosynthesis can also be regulated by nitrate availability (Takei et al. 2001; Takei et al. 2004). Nitrogen availability relative to plant demand, then, has the potential for interacting effects on cytokinin signaling and flavonoid biosynthesis, both of which have implications for signaling and interactions between legumes and rhizobia. If there were changes in the response of a polyploid to the relative quantities of signaling molecules, such as cytokinins, the densities of receptors for these signals, or the sensitivity of receptors to the signals would lead to alterations, as noted above, in nodule formation, in addition to altered plant growth. Altered cytokinin signaling in the polyploid would also have the potential to interact with nitrogen transport, thereby affecting flavonoid biosynthesis and possibly flavonoid signaling into the rhizosphere.

Op den Camp et al. (2011b) examined the duplication of genes in the cytokinin signaling pathway through a WGD event in the papilionoid legume lineage to determine if this duplication

could lead to evolution of nodulation-related functions. The authors identified a gene pair of cytokinin response regulators that were duplicated at the time of the WGD event, and their data appear to indicate that the two genes have differing expression patterns and responses to Nod factors, with MtRR9 having the strongest induction in response to symbiotic signals, while MtRR11 showed nonsymbiotic expression. They suggested that the WGD event in the papilionoids had an effect on nodulation signaling through the duplication of genes in the cytokinin pathway.

Although the potential impacts of polyploidy on nodulation symbioses and signaling can be explored to make predictions, extensive studies to match those conducted with pollinators, herbivores and AM fungi remain necessary. Nevertheless, recent genomic studies have generated interest in the interface between polyploidy and rhizobial interactions. These studies have yielded insights into the evolutionary connections between polyploidy and nodulation, with genetic data relating specifically to signaling processes.

EVIDENCE OF ROLES FOR POLYPLOIDY IN THE EVOLUTION OF NODULATION AND RHIZOBIAL SIGNALING

Nitrogen-fixing symbioses have evolved in four plant orders (Cucurbitales, Rosales, Fagales and Fabales) that comprise the nitrogen-fixing clade (NFC) of the Fabid group of rosid eudicots; the instances of nodulation in this clade are thought to be nonhomologous (Doyle 2011). Prominent within this group is the legume family, where symbioses with rhizobial bacteria evolved (Figure 2.4); with the exception of *Parasponia* in the order Rosales, the other nitrogen-fixing symbioses in the NFC involve plants that are nodulated by Gram-positive soil bacteria in the genus *Frankia* (Pawlowski and Sprent 2007). The Cucurbitales, Rosales and

Fagales each contain a relatively small number of taxa that form nodulating associations. In contrast, rhizobial nodulation symbioses in legumes are widespread and extremely diverse, such that Gram-negative bacteria from eleven genera interact with a wide taxonomic range of legumes (Pawlowski and Sprent 2007). There is also ample diversity of structure and growth in legumes (e.g., indeterminate nodules that maintain a persistent meristem or determinate nodules that do not), method of infection (e.g., rhizobia enter the plant tissues through root hairs, or by epidermal or crack entry) and nutrient transport (e.g., nitrogen is exported from the nodule as amides or ureides) among the symbioses between legumes and rhizobia (Sprent 2007). Here, we review recent genomic studies that have suggested connections between the evolution of nodulation in legumes and WGD events, and have led to speculation about the contribution of polyploidy to the development and refinement of nodulation interactions (Li et al. 2013; Young et al. 2011).

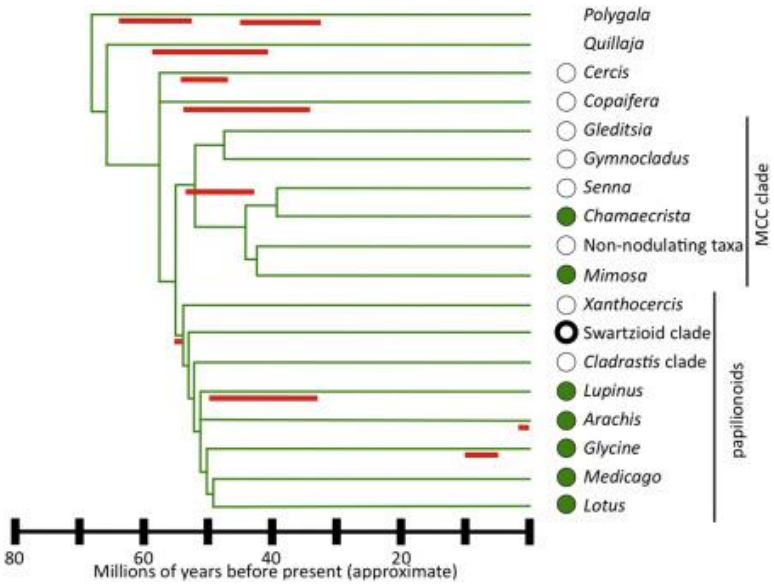


Figure 2.4: Summary legume phylogeny adapted from Cannon et al. (2015), showing selected taxa (for more detailed figure see Doyle, 2011). Filled circles indicate that many species in the clade nodulate, while empty circles indicate that no members of the clade are known to nodulate. The circle for the Swartzioid clade is meant to indicate that some members of the clade nodulate. Timing of hypothesized genome duplication events is indicated using red bars. *Polygala* and

Quillaja are non-legume outgroups, neither of which nodulates; these genera and legumes are members of the order Fabales, one of several orders of the NFC, where nodulation occurs sporadically. Nodulation in legumes occurs only in the MCC clade and in the papilionoids, and may have originated in their common ancestor. However, this hypothesis is not parsimonious given the absence of nodulation from most genera in the MCC clade outside of the core mimosoids (represented here by *Mimosa*, but including over 3,000 species, nearly all of which nodulate), and from some of the "basal" lineages of papilionoids (here represented by *Xanthocercis*, the *Cladastris* clade, and some members of the Swartzioid clade). It is more parsimonious to hypothesize multiple independent origins of nodulation, one or more times each in the MCC and papilionoid clades (Doyle, 2011). Independent origins of nodulation are known.

The exact role of WGD events in the evolution of nodulation has been difficult to ascertain. In part, this is due to the difficulty of determining the timing of WGD events and in pinpointing instances of the novel evolution of nodulation. The ability to distinguish between ancestral and derived states is necessary to understand the evolution of any trait. In the case of many legume genera, the states are not known for nodulation (Doyle 2011). For example, the genus *Senna* does not nodulate (Sprent 2009). If none of its ancestors were able to nodulate, lack of nodulation would be an ancestral ('primitive') character; but if its ancestors were able to nodulate, but this ability has been lost, the absence of nodulation would be a derived state. Thus, it is at present unknown how many independent origins of nodulation there have been in the legume family (Doyle 2011). Similarly, the number and timing of polyploidy events in the legumes has not been known until recently (Cannon et al. 2010; Cannon et al. 2015). Moreover, when WGD events are older than a few million years, the ancestor or ancestors of the polyploid typically are not known, because diploid progenitors are extinct. Finally, it is difficult to reconstruct ancestral states for gene families due to processes including gene birth and death, concerted evolution, and the evolution of gene expression.

Despite these difficulties, progress has been made in assessing the role of polyploidy in the evolution of nodulation. Researchers have proposed the unique evolution of a still-unknown

'predisposition' or 'precursor' for nodulation in a common ancestor of the NFC (Soltis et al. 1995). A recent modeling study updates this hypothesis, reaffirming that this predisposition evolved a single time, over 100 MYA, and suggesting that subsequent evolution led to 'actualization' that enabled formation of symbiotic associations in nodule structures (Werner et al. 2014). In turn, this actualization was followed by additional processes of refinement, elaboration and diversification in different lineages. Elements of signaling cascades are clearly candidates for the evolution of the precursor and steps in the process of actualization leading to effective nodulation symbioses.

Constraining the analysis to rhizobial symbiosis in legumes, establishing timing of WGD events and instances of actualization, or evolution of nodulation proper, has also been difficult. While fossil evidence exists for AM symbioses (Remy et al. 1994; Taylor et al. 1995), there are no known fossil nodules (Sprent 2007). The most recent common ancestor of legumes existed around 60 MYA, and the early evolution of the family was rapid, with the major lineages all established within 10-15 MY (Bruneau et al. 2008; Lavin et al. 2005). Nodulation has been rarely lost in two major lineages—over 95% of papilionoids nodulate, and a comparable percentage of mimosoids do as well—with several other legume groups also able to nodulate. Given uncertainties in legume phylogeny, incomplete data on nodulation, and an inability to discriminate loss of nodulation from ancestral inability to nodulate (discussed above), we still do not know how many times nodulation was gained or lost in the family. However, focusing only on papilionoids, we know that the common papilionoid ancestor experienced a WGD event (Cannon et al. 2015), and, given our current understanding of legume phylogeny (Cardoso et al. 2012; Cardoso et al. 2013), it seems most likely that this ancestor did not nodulate. Thus, the timing and potential coincidence of this WGD event with the advent or refinement of rhizobial

symbioses has led to speculation that the polyploidy event in the papilionoid lineage had a critical role in the evolution of legume nodulation.

Recent evidence, primarily derived from genomic and transcriptomic data on gene retention and expression, has been used to support the proposition that the WGD event shared by the papilionoid legumes enabled the refinement and specialization of various nodulation-related processes (e.g., Kim et al. 2013b; Li et al. 2013; Young et al. 2011). The evidence supporting the role of this WGD in the development of nodulation centers on the retention of gene duplicates originating from the papilionoid WGD where one member of each pair is currently involved in nodulation-related processes. This, coupled with differential expression across tissues and functions, suggests that the WGD contributed genetic material that served in the development and/or refinement of nodulation.

For example, in their analysis of the *Medicago truncatula* genome, Young et al. (2011) found that the nodulation-related genes *NFP*, the Nod factor receptor involved in rhizobial signal perception, and *ERN1* both have duplicates (*LYR1* and *ERN2*, respectively) that date to the papilionoid WGD, and have contrasting expression patterns, suggesting functional specialization. While *NFP* and *ERN1* are implicated in and expressed during nodulation, *LYR1* and *ERN2* are associated with mycorrhizal infection. This is in contrast to *Parasponia andersonii*, the non-legume nodulating with rhizobia: *P. andersonii* possesses a single *NFP* orthologue that functions in both rhizobial and mycorrhizal interactions (Op den Camp et al. 2011a). This suggests that the signaling receptors *NFP* and *LYR1* developed their specialized functions following duplication via the papilionoid WGD.

Li et al. (2013), working with the sequenced genomes of four papilionoid legumes, found further examples of apparent preferential gene retention following duplication during the

papilionoid WGD, where a member of a pair of duplicate genes had a role in nodulation, indicating the polyploidy event 'might have induced the emergence of critical symbiotic genes and increased the complexity of the symbiotic signaling pathway'. They found genes in the flavonoid biosynthesis pathway that were retained in duplicate, and suggest that 'more abundant and diverse flavonoids would be synthesized as a result of the polyploidy event and that the enrichment of flavonoids might be adaptive for the complex signaling required for legume–Rhizobium symbiosis' (Li et al. 2013). Some caution is warranted in interpreting this evidence; while possession of duplicated genes encoding an enzyme may be found to lead to increased enzyme concentrations and biosynthetic activity, functional divergence of the homoeologous gene copies would be required for synthesizing a greater diversity of flavonoids. In addition to retention of duplicates involved in flavonoid biosynthesis, this work also highlighted retention of duplicates in other ontology terms that include processes involved in nodulation. These included signaling receptors, nodule organogenesis (particularly gene ontology (GO) terms including hormone-mediated signaling pathways and cytokinin-mediated signaling pathways), infection thread formation and growth, and nutrient transport, exchange and metabolism. These hypotheses provide fertile ground for further, more detailed studies on individual gene families and genes.

As noted above, however, any simple account of a connection between polyploidy and the evolution of nodulation based solely on the relationship of nodulation to the papilionoid WGD event is complicated by several additional considerations. The presence of nodulation in the genus *Chamaecrista*, a legume that does not share the papilionoid WGD, also indicates that this particular WGD event was not necessary for the appearance of nodulation, even in legumes (Cannon et al. 2010). On this basis, Cannon et al. (2010) suggested that polyploidy did not

necessarily predate nodulation in all legumes and, consequently, there can be no necessary role ascribed to the papilionoid-specific WGD. A recent, more comprehensive phylogenomic study, however, found evidence of an additional polyploidy event at the base of the Mimosoid-Cassiinae-Caesalpinieae (MCC) clade, which includes the genus *Chamaecrista* (Cannon et al. 2015). Thus, nodulation in the MCC clade may yet be connected to an MCC-specific WGD event. However, just as there are non-nodulating taxa in the papilionoid subfamily that experienced the papilionoid-specific WGD event, there are also non-nodulating taxa that shared the MCC-specific event (e.g., *Gleditsia*, *Gymnocladus*, *Senna*); indeed, outside the mimosoid clade, most members of the MCC do not nodulate. Moreover, evidence was also found for separate polyploidies in the history of the Cercideae and Detarieae (Cannon et al. 2015), which are groups in which there is no known nodulation (Sprent 2009). The presence of these non-nodulating lineages that have undergone WGD events in their evolutionary history reinforces the observation that polyploidy does not inevitably lead to nodulation in legumes, despite the presence of the nodulation precursor. It remains unclear, though, whether these non-nodulating taxa represent multiple losses (preceded by a single origin of nodulation) or whether there were, conversely, multiple independent gains of nodulation. According to the models of Werner et al. (2014), the 'actualization' of nodulation occurred independently on numerous occasions in the MCC-clade; however, this modeling study used a very different tree than that which has been used in other legume phylogenies and does not provide definitive information on where nodulation originated in the clade. Even assuming the evolution of a shared predisposition for nodulation in all members of the Fabales, it appears that subsequent polyploidy events do not necessarily lead to nodulation, given evidence of independent polyploidy events in the non-nodulating *Polygala* and *Quillaja* lineages (Cannon et al. 2015).

The current evidence indicates that no single WGD event appears to have been necessary for the evolution of rhizobial signaling mechanisms and symbiosis throughout the legume family. If there were a necessary connection between the two events, the interpretation would require acknowledgement of the distinct, lineage-specific WGD events in nodulating legumes, and the apparent absence of a single WGD event in the most recent common ancestor of all nodulating lineages. While it is plausible that polyploidy can, and likely did, provide raw materials for the elaboration or refinement of nodulation signaling and symbioses (as it does for all other plant processes), it is also still uncertain that WGD is sufficient for the evolution of rhizobial nodulation in all legumes or even in papilionoids. There are, nevertheless, several examples of critical genes involved in papilionoid nodulation, including some encoding signaling functions, that arose as duplicates due to WGD in the evolutionary history of the legume family. Additional tests of the effects of more recent polyploidy events in nodulating legume lineages could help to clarify the potential of WGD events to enhance or yield further refinements.

CURRENT AND FUTURE STUDIES OF POLYPLOIDY AND NODULATION: CHALLENGES AND OPPORTUNITIES

Studies using genomic data to examine the connection between polyploidy and the origin or refinement of nodulation have highlighted a need for additional evidence and direct experimentation. Several early inoculation studies attempted to examine the relationship between polyploidy and nodulation in a restricted number of study systems (Weir 1961; Leps et al. 1980). These studies, however, were primarily conducted in synthetically generated autopolyploids and preceded the substantial advances in understanding of the signaling mechanisms underlying nodulation, many of which have occurred since the 1980s (e.g., reviewed in Long 1996;

Broughton et al. 2000; Ferguson et al. 2010; Oldroyd 2013). Moreover, substantial progress was also made over the course of these decades in advancing the study of polyploidy through the use of molecular techniques (Soltis et al. 2010). Thus, there is now a great opportunity to engage again in direct study of the interaction between these phenomena, given both our enhanced understanding of the molecular mechanisms underlying the evolutionary context and relevance of polyploidy for nodulation indicated by recent genomic studies. To this end, our research has turned to experiments testing relationships between polyploidy and nodulation, flavonoid exudation, receptor diversity, and transcriptome-level responses, primarily using *Glycine* subgenus *Glycine* as our study system, which includes a complex of recently formed allopolyploids (Sherman-Broyles et al. 2014).

Along with the opportunities for advancing this area of research, several challenges exist for any future work on the connection between polyploidy and nodulation. These include issues related to variability of the effects of polyploidy at different time-scales and the importance of assessing the progenitor contributions to the formation of polyploids, given the sensitivity of rhizobial interactions to genotype-by-genotype-by-environment (G x G x E) interactions. Each of these issues is broadly relevant to the study of polyploids and their ecological interactions, but here we consider particular aspects that will be important in assessing the implications for nodulation symbioses and signaling.

Studying effects at several time-scales: natural paleopolyploids, neopolyploids and synthetics

Polyploidy can have variable effects at different time-scales and this is a challenge for developing any general understanding of the effects polyploidy can have on nodulation. While ancient polyploidy events and the genomes of ancient paleopolyploid have been the source of data pertaining to the evolution of nodulation, direct evaluations are generally not possible, since

the progenitors of these polyploids are not known or available (e.g., Young et al. 2011). Natural, recently formed polyploids, though, often present opportunities to study both the extant progenitors and the polyploids to which they gave rise (Hegarty et al. 2013). Even here, however, caution is required in making inferences, since the formation of polyploids may involve radical genetic and epigenetic changes (e.g., 'genomic shock' of McClintock (1984)) and selection has had the capacity and time to alter rhizobial symbioses in the polyploid plants, such that the effects of selection would be combined with any effects due to polyploidy (Hegarty et al. 2013). In addition, the diploid progenitors and their symbiotic interactions will also have evolved and, in their current ecological responses, may not be accurate representatives of the plants that originally gave rise to the polyploid. In general, even when comparison of natural polyploids and diploid progenitors is possible, such a comparison necessarily confounds the near-term effects of polyploidy, which include effects of changes to cell size and division, development and gene dosage, with longer term effects, including altered responses to natural selection (Ramsey and Ramsey 2014). These issues complicate any direct comparisons that could be attempted between extant polyploids and presumed diploid progenitors. Not coincidentally, studies of natural polyploids must necessarily involve consideration of species rather than individual genotypes, since the relevant genotypes are unknown and unsampled if indeed they still exist.

When available, synthetic polyploids present an excellent opportunity to study the immediate effects of polyploidy. The effect of genome doubling itself can be studied, removing from consideration subsequent effects of natural selection in an ecologically relevant context, and can also be used to separate effects of genome duplication from those of genome merger by examining both autopolyploids and allopolyploids (Hegarty et al. 2013). With nodulation-related plant signaling mechanisms, as outlined above, enhanced exudation of a broader diversity of

signaling compounds or the expression of Nod factor receptors from distinct diploid progenitors might be such immediate outcomes of polyploidy. However, constraints to the inferences that can be made from the study of synthetic polyploids are precisely due to the isolation from effects at longer time-scales noted above, when natural selection has shaped the symbiotic interactions. Interactions between host plant and rhizobial symbiont can exist along a continuum from mutualism to parasitism (Denison and Kiers 2004); with respect to changes in nodulation-related signaling, interactions that are initially enabled or made possible by polyploidy may later be selected against if they are adversely impact fitness. While the expression and silencing of one member of some classes of progenitor genes (duplication resistant genes) appears to be predictable, stochastic alterations are also observed over the early generations of synthetically formed allopolyploids (e.g., Gaeta et al. 2007; Gaeta et al. 2009; Madlung and Wendel 2013; Wang et al. 2004). Thus, a polyploid's biosynthesis of signaling molecules, receptors and other elements of rhizobial signal transduction pathways could also change rapidly in the generations following its formation. These issues, coupled with the fact that the environmental range (and thus the adaptive context for rhizobial symbiosis) is often altered over time for polyploids relative to their diploid progenitors (te Beest et al. 2012), suggest that a combination of studies examining both synthetic and natural (i.e., non-synthetic) polyploids, including established and very recently formed neopolyploids, will be needed in order to develop a complete picture of the implications of polyploidy for nodulation.

Challenges of genotype-by-genotype interactions

The definition of 'species' continues to be debated (e.g., De Queiroz 2007), but for our present purposes it can be defined as a collection of genotypes, each with different potentials for genetic interaction (in addition to gene flow, the ability to independently assort and recombine to

produce diversity). Consideration of rhizobial nodulation in a species necessarily requires consideration of interactions between genotypes of the host plants (G_{plant}) by the genotypes of the rhizobia ($G_{rhizobia}$) by the environment (E). The variability of G_{plant} x $G_{rhizobia}$ and G_{plant} x $G_{rhizobia}$ x E interactions has been observed in a number of cases (e.g., Heath et al. 2012; Heath and Tiffin 2007). This variability is relevant when studying polyploids for several reasons. Polyploids may be formed multiple times by various combinations of diploid progenitor genotypes, and may incorporate further variation from its progenitors by gene flow after formation (Soltis and Soltis 2009). Even so, formation will involve only a subset of genotypes from within the diploid progenitor population. Thus, the variability of these interactions in a set of plant genotypes within a progenitor species has the ability to affect the symbiotic outcomes in the polyploid produced from them. In a natural context, the symbiotic capacity of a polyploid will depend on the genotypes it has 'sampled' and incorporated from the progenitor population. When a polyploid arises multiple times, the resulting polyploids may interbreed to form a single polyploid species; however, if polyploids from different origins are separated geographically or genetically, the result can be multiple reproductively isolated polyploid species (Soltis and Soltis 2009). Isolated polyploid species arising from multiple origins would produce low diversity in each individual polyploid species; multiple origins with sampling of a significant part of the diversity of the diploid progenitors, followed by recombination (including independent assortment) could produce a single, much more diverse, polyploid species that would be more likely both to sample more of the diploid diversity (though the sum of diversity across all of the polyploid taxa might be the same in both cases) and to produce much more recombinational diversity. In turn, these differences in diversity will affect the resultant symbiotic signaling capabilities and interaction spaces of the polyploid species. In such cases, comparisons of

symbiotic interactions should ideally be made between the polyploids and the diploid progenitor genotypes that contributed to its formation, rather than other genotypes of the diploid species, though this would be difficult in all but the most recently formed natural polyploids and would depend on sampling.

Allopolyploids thus involve the complications of interactive effects due to hybridity and the combination of two diploid progenitor genotypes. Allopolyploid complexes nevertheless provide opportunities to understand the emergent effects of polyploidy on nodulation. While repeated formation of allopolyploids, with different combinations of diploid progenitors, can provide additional complexity, it also provides opportunities to examine a naturally replicated system, where emergent properties can be observed across independent instances of polyploid formation among closely related species. An additional challenge arises, however, given the combination of the effects of hybridity and genome doubling in allopolyploids. Nevertheless, this challenge can be addressed in natural systems by relating levels of hybridity with particular effects to determine whether a possible effect of allopolyploidy correlates strongly with the level of hybridity (i.e., the genetic distance or divergence between progenitor species). Effects that are strongly associated with allopolyploids, but do not correlate well with hybridity levels are thus less likely to be due to hybridity and may be attributable to the phenomenon of genome doubling in the allopolyploids (and subsequent selection acting on them). In the context of rhizobial symbiosis, untangling such effects is important, since there is evidence that both hybridity and polyploidy can, for example, affect biosynthesis of flavonoid compounds (e.g., Levy and Levin 1971, 1974; Orians 2000; Schwarze 1959). Thus, changes in rhizobial interactions observed between an allopolyploid and its diploid progenitors may be attributable either to hybridity, genome doubling, or both.

CONCLUSION

In recent years, researchers have expressed interest in the possibility of engineering biological nitrogen fixation in cereals, due to the potential value of reducing nitrogen fertilizer application (Beatty and Good 2011; Oldroyd and Dixon 2014). Such potential applications have stimulated additional interest in understanding the evolution of nodulation in legumes and other NFC species. Analyses of genomic data from nodulating legumes known to have experienced whole genome duplication have suggested an important role for polyploidy in the evolution of nodulation in such species. While several challenges exist to the direct evaluation of the connections between polyploidy and nodulation, the opportunities for increased understanding are great. Many of the signaling mechanisms underlying nodulation have now been elucidated. The methods available for accurately identifying and assessing polyploids have improved. Thus, future research involving the combined study of diverse nodulating systems will yield insights into the emergent implications of polyploidy for rhizobial signaling and symbiotic interactions with plants.

CHAPTER 3

EFFECTS OF ALLOPOLYPLOIDY AND RHIZOBIAL INOCULATION ON ROOT METABOLITE PROFILES, DIVERSITY, EXUDATION AND BIOSYNTHETIC GENE EXPRESSION IN *GLYCINE* SUBGENUS *GLYCINE*

INTRODUCTION

Polyploidy (whole genome duplication; WGD) has been a prevalent phenomenon in the evolutionary history of plants. All flowering plants have polyploidy in their history and polyploidy is believed to have been involved in their diversification (Jiao et al. 2011). Polyploidy can generate evolutionary novelty (e.g., Freeling and Thomas 2006; Madlung 2013), and WGD affects many aspects of plant biology (Coate et al. 2012; Lavania et al. 2012; Coate et al. 2013; Chaudhary et al. 2009; Hao et al. 2013; Marchant et al. 2016; Visger et al. 2016). Novelty through polyploidy has also been linked to invasiveness and the ability to colonize new habitats (te Beest et al. 2012). However, relatively little is known about the impact of polyploidy on biotic interactions (Soltis et al. 2010), but recent work has explored the effects of polyploidy on herbivory, pollination and mycorrhizal symbioses (e.g., Nuismer and Thompson 2001; Thompson and Merg 2008; Těšitelová et al. 2013).

Nodulation in legumes is a specialized symbiotic relationship, established between plants and a diverse set of nitrogen-fixing bacteria ('rhizobia') to meet the critical challenge of nitrogen acquisition (Herridge 2008; Vance 2001, 1998). How the nodulation symbiosis originated remains an important evolutionary question (Doyle 2011), but several workers, noting the strong correlation between nodulation and a WGD event approximately 55 to 60 million years ago (MYA) in the common ancestor of the papilionoid legumes, have suggested that polyploidy played a significant role in either the origin or the refinement of this key symbiosis (Young et al. 2011; Li et al. 2013; De Mita et al. 2014; Kim et al. 2013b). Such studies addressing the role of polyploidy in the evolution of nodulation are primarily those that make inferences from patterns of retention and expression of genes involved in nodulation that have been retained in duplicate for over 50 MY. With the exception of several pre-genomic studies that found differences in

nodule number and effectiveness, primarily in autopolyploids (Weir 1961; Leps et al. 1980; Stalker et al. 1994), little research has addressed the intersection of these two phenomena directly or examined underlying mechanisms, using recently formed polyploids and their known diploid progenitors.

Li et al. (2013) identified a number of nodulation-related processes through which polyploidy may have affected the development and evolution of this symbiosis. One such process relates to the signaling and synthesis of flavonoids. Flavonoids are important for rhizobial symbiosis since they are exuded by the plant and they induce Nod factor production in compatible bacteria; Nod factors are, in turn, perceived by the plant host and initiate a series of symbiotic responses in the plant (Oldroyd 2013). As the initial signal from hosts to symbionts, the types and abundances of flavonoids produced by plant hosts have the potential to affect the diversity of rhizobia with which the host has the capacity to interact (e.g., Peck et al. 2006; Kobayashi and Broughton 2008). Li et al. (2013) found that genes for a number of enzymes involved in flavonoid biosynthesis had been duplicated and retained following the papilionoid WGD event. The authors suggested that more abundant and more diverse flavonoids would be produced due to polyploidization, and that this could be adaptive in the context of rhizobial symbiosis (Li et al. 2013). This is also related to a general expectation that the diversity of metabolites synthesized can increase in allopolyploids through combined effects of genome doubling and hybridity (e.g., Mears 1980; Levin 1983; Orians 2000).

On this basis, one would predict that the root metabolite profiles of allopolyploids would contain more abundant, more diverse, or novel flavonoid compounds compared to their diploid progenitors. If diploid progenitors produce different suites of compounds, an allopolyploid with an additive metabolite profile could have enhanced signaling capabilities, having the ability to

engage in a different set of interactions, potentially with a broader, more diverse suite of symbionts (Powell and Doyle 2015). Over time, this allopolyploid could also develop a metabolite profile consisting of a unique subset of the compounds exuded by both diploid progenitors. Alternately, the allopolyploid could exhibit transgressive profiles including novel compounds, which also might result in broader or novel interactions. Experimental evidence has shown that accumulation of natural products can vary in polyploids relative to diploids (Hull-Sanders et al. 2009a; Lavania et al. 2012) and other past studies have dealt with accumulation of flavonoids by polyploids, primarily in above-ground tissues (e.g., Levy and Levin 1971; Murray and Williams 1976; Murray and Williams 1973). However, relevant analyses have not been conducted on species that form nodulation symbioses.

The metabolites synthesized by a host plant delineate the broadest potential symbiotic signaling capacity of the plant, and the symbiotic signaling capacity of a species is the aggregate symbiotic signaling capacity of the individual members of the species. Here, the concept of symbiotic signaling capacity is used in a manner similar to the constitutive information landscape described by Kessler (2015) for herbivores and other interacting partners, though here the chemical information landscape, or symbiotic signaling capacity, creates a type of boundary excluding certain species or genotypes of mutualist rhizobia and other bacteria, while enabling compatible symbionts to interact with the host. Furthermore, in the case of signaling between legumes and rhizobia, while the flavonoids synthesized in the root establish the broadest set of interactions (interpretable as a fundamental information space), the compounds present in the exudates under different growing conditions represent a narrower, realized information space, contributing, ultimately, to the realized set of interactions in the rhizosphere.

The allopolyploid complex of *Glycine* subgenus *Glycine* provides a model system for studying the intersection of nodulation and polyploidy, since members of the complex are known to form nodulation symbioses (Pueppke 1988; Pueppke and Broughton 1999; Brockwell et al. 1998). The complex includes several independently-formed allopolyploids that resulted from crosses between known, extant diploid progenitors within the past several hundred thousand years (Bombarely et al. 2014). The diploid species are found primarily in Australia, with some species' ranges extending to Papua New Guinea (Doyle et al. 2002; Doyle et al. 2004). Some of the allopolyploids, in contrast, are found in diverse locations including Taiwan, the Philippines, and the Ryukyu Islands of Japan, in addition to their native ranges in Australia and Papua New Guinea (Doyle et al. 2002; Doyle et al. 2004). Conveyance by birds has been suggested as a potential mechanism of dispersal (Hymowitz et al. 1990).

Here, we characterize the flavonoid and nodulation-related metabolite profiles of roots and root exudates in members of the *Glycine* subgenus *Glycine* complex, asking several questions relating to the effects of allopolyploidy. First, do allopolyploids in the complex have metabolite profiles that differ from their diploid progenitors qualitatively or quantitatively? Second, are there common transgressive patterns among the allopolyploid metabolite profiles, or are there differences between allopolyploid species? If they do not clearly differ from one or both of their progenitors, or from each other, it is unlikely that the allopolyploid species are capable of a different set of interactions with rhizobia, while novel interactions are possible to the degree that metabolite profiles differ. Where differing profiles were observed, we explored the gene expression that might underlie the differences observed, and also analyzed the degree to which distinctive features related to key signaling compounds were also present in root exudates.

MATERIALS AND METHODS

Plant materials

For flavonoid analysis, the study group consisted of three allopolyploid species and their four diploid progenitor species. The allopolyploids include *G. dolichocarpa* 'T2', *G. tomentella* 'T1', and *G. tomentella* 'T5'. *G. syndetika* 'D4' and *G. tomentella* 'D3' are the diploid progenitors of *G. dolichocarpa* 'T2'; *G. tomentella* 'D1' and *G. tomentella* 'D3' are the progenitors of *G. tomentella* 'T1'; *G. clandestina* (abbreviated as 'CLA' herein) and *G. tomentella* 'D1' are the progenitors of *G. tomentella* 'T5'. The accessions used in this study for each species were selected to sample across the estimated geographic ranges and genetic diversity of these species in Australia and Papua New Guinea (Table 3.1) (Doyle et al. 2002; Harbert et al. 2014). This set of samples also includes an accession collected outside of this range; the T2 accession G1854 was collected in Taiwan.

Table 3.1. Accessions of species in *Glycine* subgenus *Glycine* used in this study. The table gives the accession G (CSIRO Perennial Glycine Collection) numbers, the provenance (locality and country, as well as the Australian state where relevant) and the datasets in which the accession was included.

Species	Accession	Provenance	Root Metabolites	Exudates	RNA-Seq
D1	G1156	Charleville, Queensland, Australia	Yes	Yes	-
D1	G1157	Surat, Queensland, Australia	Yes	Yes	-
D1	G1316	Texas, Queensland, Australia	Yes	Yes	-
D1	G1858	Condamine, Queensland, Australia	Yes	Yes	-
D1	G2312	Injune, Queensland, Australia	Yes	Yes	-
D1	G2734	Mitchell, Queensland, Australia	Yes	Yes	-
D3	G1364	Nadzab, Papua New Guinea	Yes	Yes	Yes
D3	G1403	Yalkula, Queensland, Australia	Yes	Yes	Yes
D3	G1686	Mareeba, Queensland, Australia	Yes	Yes	-
D3	G1749	Lockhart River, Queensland, Australia	Yes	Yes	-
D3	G1820	Bloomsbury, Queensland, Australia	Yes	Yes	Yes
D4	G1300	Mt. Garnet, Queensland, Australia	Yes	Yes	Yes
D4	G1772	Mt. Surprise, Queensland, Australia	-	Yes	-
D4	G1775	Greenvale, Queensland, Australia	Yes	Yes	-
D4	G1784	Pentland, Queensland, Australia	-	Yes	-

Table 3.1 (Continued)

D4	G2073	Charters Towers, Queensland, Australia	Yes	Yes	Yes
D4	G2321	Bogantungan, Queensland, Australia	Yes	Yes	Yes
D4	G2471	Townsville, Queensland, Australia	-	Yes	-
G. clandestina	G1126	Iandra, New South Wales, Australia	Yes	Yes	-
G. clandestina	G1253	Captains Flat, New South Wales, Australia	Yes	Yes	-
G. clandestina	G1731	Rainbow Beach, Queensland, Australia	Yes	Yes	-
G. clandestina	G2757	Dunolly, Victoria, Australia	Yes	Yes	-
G. clandestina	G2940	Ajana, Western Australia, Australia	Yes	Yes	-
T1 (= D1xD3)	G1133	Brampton Island, Queensland, Australia	Yes	Yes	-
T1 (= D1xD3)	G1136	Delungra, New South Wales, Australia	-	Yes	-
T1 (= D1xD3)	G1260	Mackay, Queensland, Australia	Yes	Yes	-
T1 (= D1xD3)	G1288	Grafton, New South Wales, Australia	Yes	Yes	-
T1 (= D1xD3)	G1763	Cooktown, Queensland, Australia	Yes	Yes	-
T2 (= D4xD3)	G1134	Brampton Island, Queensland, Australia	Yes	Yes	Yes
T2 (= D4xD3)	G1188	Lindeman Island, Queensland, Australia	Yes	Yes	Yes
T2 (= D4xD3)	G1393	Kalunga, Queensland, Australia	Yes	Yes	Yes
T2 (= D4xD3)	G1854	Taiwan	Yes	Yes	-

Table 3.1 (Continued)

T2 (= D4xD3)	G2320	Capella, Queensland, Australia	Yes	Yes	-
T2 (= D4xD3)	G2809	Whitsunday Islands, Queensland, Australia	Yes	Yes	-
T5 (=D1x*G.clandestina*)	G1487	Station Creek, New South Wales, Australia	Yes	Yes	-
T5 (=D1x*G.clandestina*)	G1655	Narrabri, New South Wales, Australia	-	Yes	-
T5 (=D1x*G.clandestina*)	G1656	Baradine, New South Wales, Australia	Yes	Yes	-
T5 (=D1x*G.clandestina*)	G1741	Narrabri, New South Wales, Australia	Yes	Yes	-
T5 (=D1x*G.clandestina*)	G1969	Gilgandra, New South Wales, Australia	Yes	Yes	-
T5 (=D1x*G.clandestina*)	G1972	Gilgandra, New South Wales, Australia	Yes	Yes	-

Root tissue analysis

For both root tissue and exudate analysis, seeds were surface sterilized by submersion in 40% ethanol for five minutes and 40% bleach for five minutes, followed by five rinses with sterilized water. The seeds were then nicked with a razor blade and placed in petri plates with water. The petri plates were moved to a dark drawer for three days to ensure germination.

Root tissue was then plated on nitrogen-free Jensen's media (Somasegaran and Hoben 1994), and was either mock-inoculated with sterile water or inoculated with a culture of approximately 10^7 cells per ml of rhizobium strain NGR234, for which the cell count was determined with a Helber counting chamber (Somasegaran and Hoben 1994). For details on the methods for culturing rhizobia, refer to Powell and Doyle (2016). Seedlings were then grown for

a week in a growth chamber with a day/night cycle of 16h/8h, with day/night temperatures of 22C/18.5C, relative humidity of 60% and light intensity of 150 µmol/m^2/s. Plant root tissue was collected, frozen in liquid nitrogen and stored at -80C. The tissue was subsequently ground while frozen with a FastPrep tissue homogenizer (MP Biomedicals LLC, Santa Ana, CA, USA) and 0.9 g of 2.3 mm Zirconia/Silica beads (Biospec Products Inc., Bartlesville, OK, USA). Phenolics were then extracted by vortexing and incubating the tissue with 80% MeOH for two hours. Prepared extracts were then analyzed by HPLC on an Agilent 1100 instrument with a Gemini C18 reverse-phase column (3 µm, 150 x 4.6 mm, Phenomenex Inc., Torrance, CA, USA). The mobile phase (A) was 0.25% phosphoric acid and (B) acetonitrile, with a method starting with 5% B for 4 minutes, increasing to 60% B at 24 minutes, and a further increase to 95% at 34 minutes, holding for 1 minute. The flow rate was 0.7 ml/minute and the injection volume was of 15 µl. Standards for compounds that have previously been characterized as having relevance to the symbiosis between *Glycine max* and rhizobia were included in each sample analysis. These standards (all from Sigma, St. Louis, MO, USA, catalog numbers given) included daidzein (D7802), genistein (G6649), daidzin (42926), genistin (48756), coumestrol (27883), and formononetin (47752). All peaks were integrated and quantified at 254 nm.

For this dataset, HPLC peaks that did not reach a signal intensity area of 200 (equivalent to approximately 1.46 ng/µl of genistein) in any of the samples were excluded from these analyses as identification to class became unreliable below this limit. This resulted in a set of 18 total compounds; based on clearly identifiable UV spectra, a subset of 12 isoflavones within this set of 18 compounds was also used for certain analyses. Signal intensity was converted to concentration units of genistein equivalents, using a standard curve based on a genistein standard dilution series. For both sets of compounds, the percentage contribution of each compound to the

total amount of each set was calculated and these percentages were visualized as stacked column plots using the ggplot() function in the R package ggplot2 (Wickham 2009). The total amounts for each compound set were compared for each triad of species, comprising each allopolyploid and its diploid progenitors, and tested using a mixed model including species and treatment as fixed effects, as well as accession as a random effect and the interaction term between species and treatment, using the lme4 (Bates et al. 2015), lmerTest (Kuznetsova et al. 2015), and lsmeans (Lenth and Hervé 2016) R packages. In addition, these packages were used in testing a model for Shannon's H index values, which were calculated for each sample; analyses of variance (ANOVA) were based on all samples across all species and included species and treatment as fixed effects, the interaction term between species and treatment, and accession as a random effect. For testing statistical differences in individual compounds between species of the T2 triad, given non-normality of residuals, the Kruskal-Wallis test was used, along with the Dunn test for pairwise differences, through the PMCMR package in R (Pohlert 2014).

For assessment of unique and shared metabolites between species, the presence of a metabolite in a given species, within total set of metabolites, was assessed based on which samples had HPLC peaks integrated at the relevant retention time for each metabolite. For this analysis, if the metabolite was present in any accession of a species under either treatment, the metabolite was counted as present in the species.

For multivariate analyses, principal component analysis was conducted using the R package vegan (Oksanen et al. 2015), using the rda() function. Redundancy analyses were conducted using the rda() function with a specified model including species and treatment as explanatory variables. The ANOVA-like permutation test provided by the anova() function

within the vegan package was used, in conjunction with the rda() function, to determine significance of the species and treatment terms.

The approach to random forest presented here primarily follows the implementation presented for volatile organic compounds in the context of plant chemical ecology by Ranganathan and Borges (2010). Variable selection using random forest was conducted with the varSelRF R package (Diaz-Uriarte 2007) and also provided estimates of the prediction error rates using random forest. Mean decreases in accuracy (MDA) of classification when particular predictor variables (compounds) were excluded from the model were calculated using the importance() function of the randomForest R package (Liaw and Wiener 2002).

Exudate analysis

The protocol for exudate analysis was adapted from previously published methods (Bolaños-Vásquez and Werner 1997; Pueppke et al. 1998). For each accession, three seedlings, germinated as described above, were incubated for seven days in a 13 x 100 mm borosilicate glass tube containing a cellulose acetate membrane (Whatman OE 66, 0.2 µm, 25 mm diameter, GE Healthcare, UK) and 3 ml of 1mM CaCl2 in 5mM MES buffer at a final pH of 6.8. For all accessions, three such samples were prepared for analysis, except for G1136, G1253 and G2734, for which two samples were analysed. Following incubation, the cellulose acetate membranes were separated from the plants and buffer, rinsed five times with deionized water, and the compounds were eluted from the membranes using two sequential washes of 3mL 100% methanol and one rinse with 90% methanol. The combined eluates were dried in a rotary evaporator (Labconco, Kansas City, MO, USA) before being reconstituted in 200 µl of 80% methanol. Umbelliferone (H24003, Aldrich, St. Louis, MO, USA) was also added to each sample as an internal standard with a final concentration of 50 ng/µl.

A 10-µl aliquot of each sample was analysed for secondary metabolites by liquid chromatography-mass spectrometry (LC-MS) on a Thermo Scientific TSQ Quantum Access LC-MS system equipped with a Gemini C18 reverse-phase column (3 µm, 150 x 2, Phenomenex Inc., Torrance, CA, USA). The method used a mobile phase of (A) 0.1% formic acid and (B) ACN, linear increasing B from 5% to 75% over 20 minutes, holding for 5 minutes, then increasing to 95% at 26 minutes until 30 minutes, with a flow rate of 200 µl/minute. The identity of the compounds in the exudates was determined by comparison with authentic standards for each. Exudate concentration values in units of ng IS per g fresh weight (FWT) were calculated. For statistical testing, values for each sample were log-transformed, with a value of 0.001 added to each concentration prior to log-transformation, in order to improve the normality of residuals. A mixed-effect model for individual compounds in each allopolyploid triad was used including species as a fixed effect and accession as a random effect, and Tukey's HSD was used for pairwise comparisons in cases where species was a significant factor.

Expression analysis

RNA-Seq libraries were prepared using pooled root tissue from multiple plants from three accessions for each species in the T2 triad, and inoculated and mock-inoculated samples were grown for each accession as described above. Total RNA was extracted using the AllPrep DNA/RNA/miRNA Universal Kit (QIAGEN, Hilden, Germany), with on-column DNase treatment (QIAGEN, Valencia, CA, USA). Illumina single-end libraries were made and multiplexed with the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina Inc., San Diego, CA, USA). The Illumina HiSeq 2500 platform at the Cornell University Biotechnology Resource Center's sequencing facility (http://www.biotech.cornell.edu/brc/genomics-facility) was used for sequencing. Fastq-mcf (Aronesty 2013) was used to trim sequences and to remove short reads,

with a minimum quality setting of 30 and a minimum read length of 50. Tophat2 (Kim et al. 2013a) was then used to map the reads to the *G. max* genome (2.75_Wm82.a2.v1). The command samtools view -h input.bam | grep -w "@SQ\|@PG\|NH:i:1" was used to filter for uniquely mapped reads. The command htseq-count (Anders et al. 2014) then yielded counts of reads mapped to each exon. Separate analyses were conducted for mock-inoculated samples and inoculated samples. Pairwise comparisons of the T2, D3 and D4 samples were conducted using DESeq2 (Love et al. 2014), examining specific genes involved in the biosynthesis of flavonoids, which were identified from the literature (Gutierrez-Gonzalez et al. 2010; Zhang et al. 2014). Sequences for genes in the biosynthetic pathway leading to the synthesis of isoflavones were obtained from Gutierrez-Gonzalez et al. (2010), and these were used as queries in BLAST searches in Soybase (www.soybase.org) against the Wm82.a2.v1 transcript sequences, and the top matches for each of the following genes were selected for study: *PAL* (Glyma.03G181600, Glyma.03G181700, Glyma.19G182300), *CH4* (Glyma.02G236500, Glyma.14G205200), *4CL* (Glyma.01G232400, Glyma.11G010500), and *CHR* (Glyma.14G005700). Another chalcone reductase gene was examined; the chalcone reductase gene sequence described in Zhang et al. (2014) was also used, and gene expression of the best match, Glyma.16G219500, was examined.

RESULTS

The overall relative quantities of metabolites in the root tissues showed variation across diploid and allopolyploid species. The various compounds observed made up different proportions of the profiles of the diploid species although percent allocation to certain metabolites was broadly similar across the species. This was the case both when the set of 18 compounds quantified were examined (Figure 3.1), and when only isoflavones were considered

(Figure 3.2). In particular, among the diploid species, the compound formononetin appeared to make a substantial contribution only in D4, as well as to a lesser degree in *G. clandestina*, which is also an 'A-genome' species and is thus closely related to D4. Across all samples, genistin, and to a lesser extent daidzein, were particularly abundant.

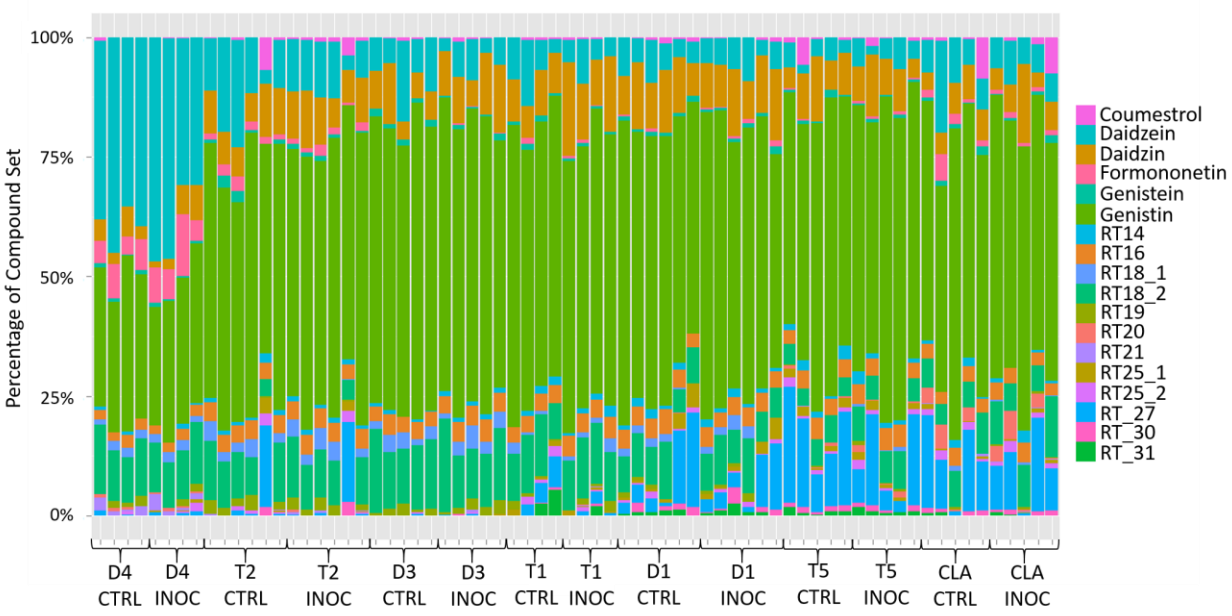


Figure 3.1. 100% stacked column chart showing the proportion in each sample of each compound in the 18-compound set. Sample species and treatment groups are labeled below the plot (CTRL = mock-inoculated samples; INOC = samples inoculated with rhizobia).

Figure 3.2. 100% stacked column chart showing the proportion of each isoflavone in each sample. Sample species and treatment groups are labelled below the plot (CTRL = mock-inoculated samples; INOC = samples inoculated with rhizobia).

In terms of shared and unique metabolites, in the T1 and T5 triads, all 18 metabolites were found to be present in both diploid progenitors and the allopolyploids. However, in the T2 triad, while 15 compounds were shared among all three species, one compound was present in all accessions of D3 but not in any D4 accessions and another was present in D4 but not D3. In both cases, the metabolites that were unique to one of the progenitors were also present in the allopolyploid. The last of these 18 compounds was not present in any of these three species.

Shannon's *H* values were calculated for each sample as an assessment of the diversity of compounds present in each profile. Shannon's *H* index takes into account not only the presence or absence of compounds, but also the evenness of their distribution within a sample. Species was a significant factor in the analysis of compound diversity when considering the 18-compound set across all samples ($F_{6, 28} = 2.78$, $P < 0.05$), while treatment ($F_{1, 28} = 0.55$, $P > 0.05$)

and the interaction of treatment and species ($F_{6, 28} = 0.81$, $P > 0.05$) were not significant (Figure 3.3). In pairwise comparisons, D3 and D4, the diploid progenitors of T2, were significantly different from each other (Tukey's HSD, $P < 0.05$). *G. clandestina* was also significantly different from D3 (Tukey's HSD, $P < 0.05$), though this comparison is not directly relevant to natural allopolyploidy in *Glycine*. When considering the isoflavone set, as with the 18-compound set, species was a significant factor ($F_{6, 28} = 8.95$, $P < 0.001$), but treatment ($F_{1, 28} = 0.59$, $P > 0.05$) and the interaction term ($F_{6, 28} = 0.71$, $P > 0.05$) were not (Figure 3.4). For the isoflavone set, D4 was significantly different from all other species except T2; T2 was also different from T2 and D4 (Tukey's HSD, $P < 0.05$).

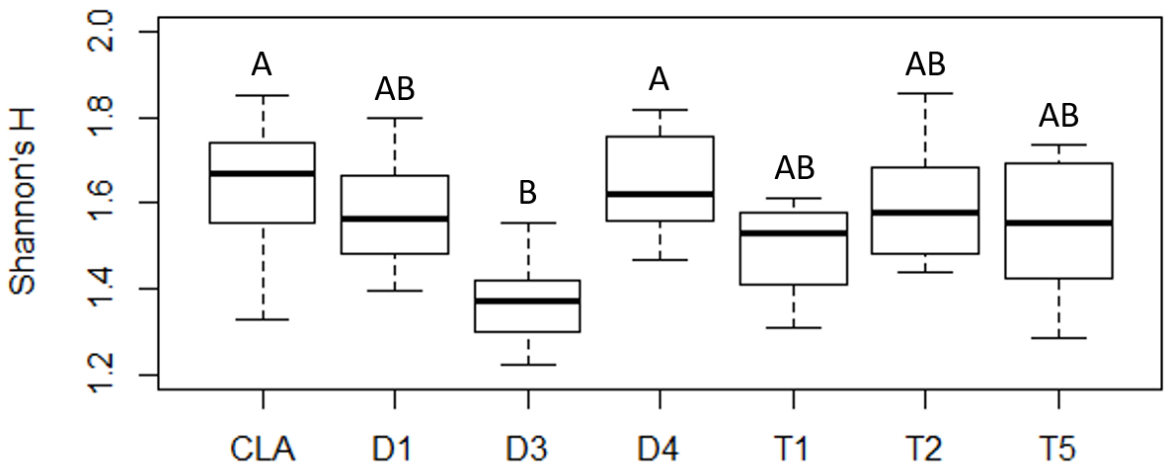


Figure 3.3. Shannon's H diversity index values across all species, calculated based on the set of 18 compounds in root tissues. Boxplots show the median and interquartile range, with whiskers indicating minimum and maximum values. Species sharing the same letter group are not significantly different (p > 0.05).

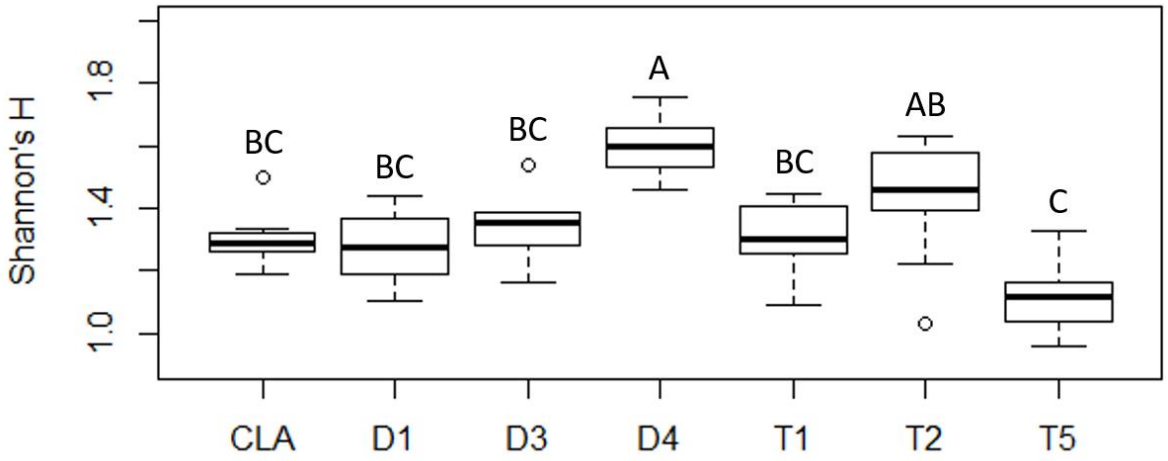


Figure 3.4. Shannon's H diversity index values across all species, calculated based on the set of isoflavones in root tissues. Boxplots show the median and interquartile range, with whiskers indicating minimum and maximum values. Species sharing the same letter group are not significantly different ($p > 0.05$).

For the set of all compounds analyzed in the study, as well as for the subset of identified isoflavones, total amounts present in roots were also calculated and compared between species in each triad (Figures 3.5-3.10). With the T1 triad, neither factor nor the interaction term was significant in either the total compound set or the isoflavone set. Inoculation treatment was significant in explaining differences in the total isoflavones for the T2 triad ($p=0.045$). In the T5 triad, total amounts in the set of 18 compounds were significantly affected by treatment ($p=0.033$), and there was a significant interaction between treatment and species ($p=0.021$). For total isoflavones in the T5 triad, the interaction term was also significant ($p=0.031$). In general, no significant differences were observed in comparisons between species within each triad. The

significant interaction terms for the T5 triad indicate a differential response to inoculation among the three species, which appears to be due to the decrease in T5 under inoculation (Figure 3.9 and Figure 3.10).

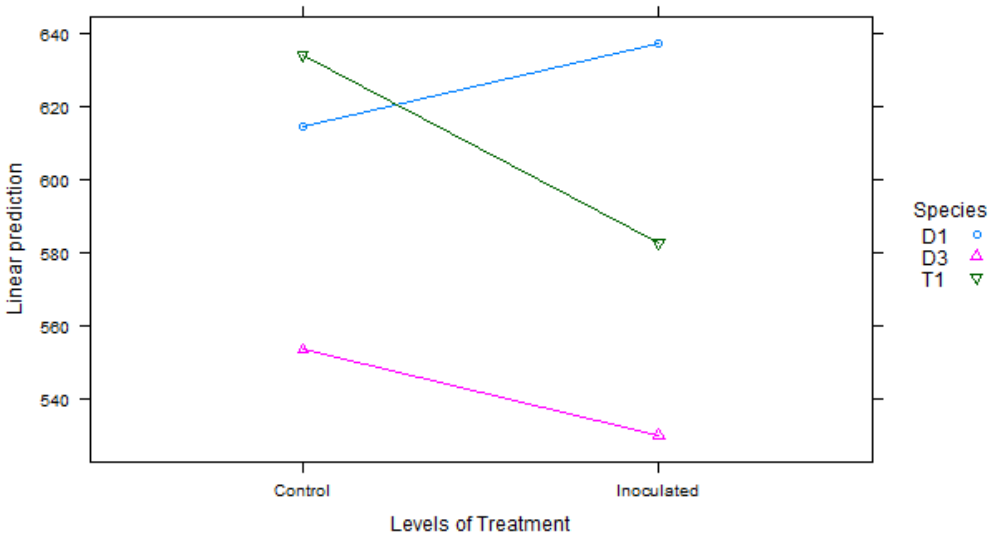


Figure 3.5. Interaction plot showing least-square means model predictions of total compound amounts for the 18-compound set in the T1 triad.

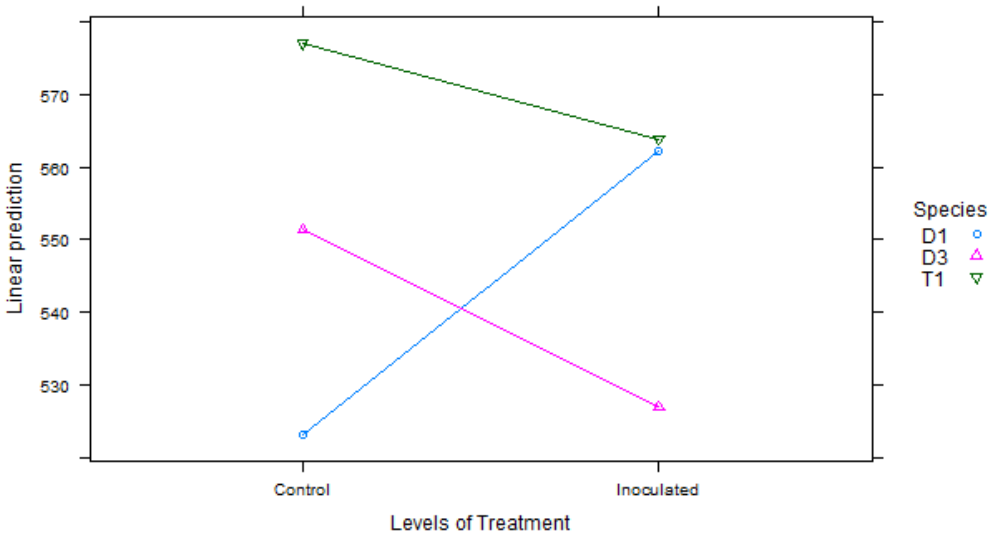


Figure 3.6. Interaction plot showing least-square means model predictions of total isoflavone amounts in the T1 triad.

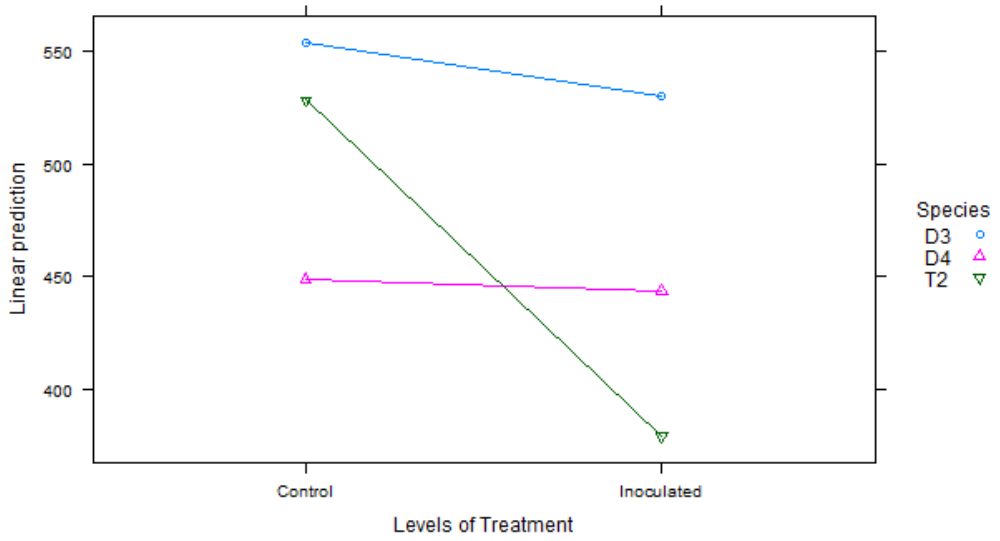


Figure 3.7. Interaction plot showing least-square means model predictions of total compound amounts for the 18-compound set in the T2 triad.

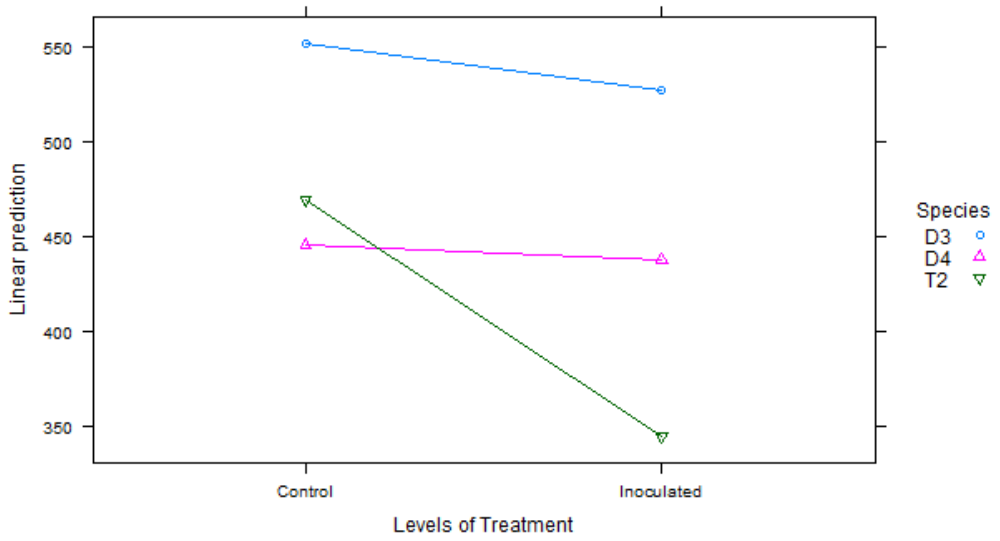


Figure 3.8. Interaction plot showing least-square means model predictions of total isoflavone amounts in the T2 triad.

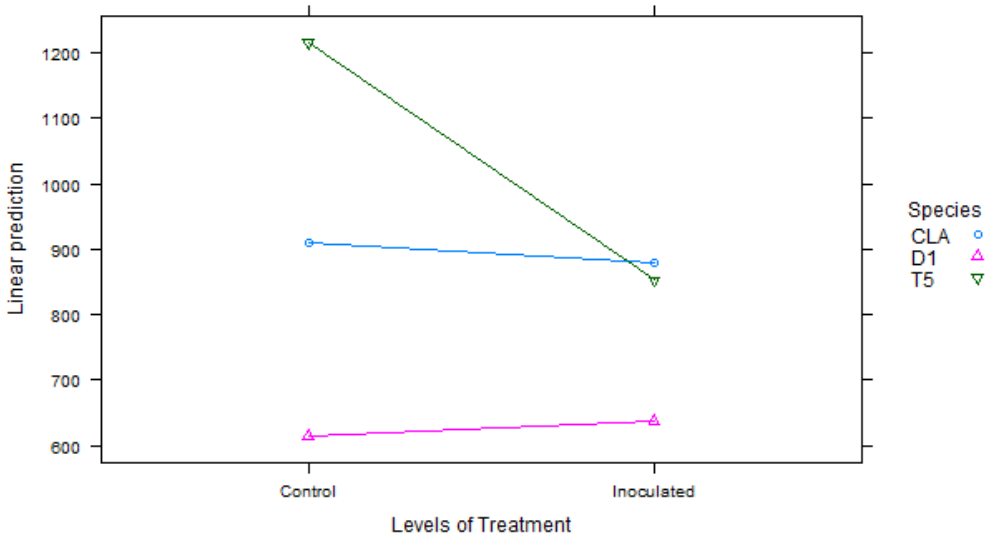


Figure 3.9. Interaction plot showing least-square means model predictions of total compound amounts for the 18-compound set in the T5 triad.

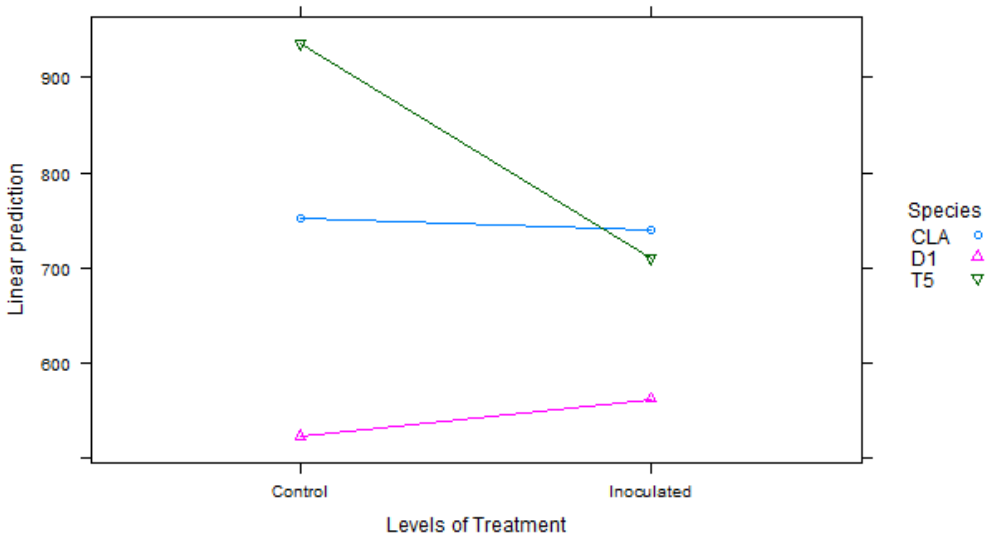


Figure 3.10. Interaction plot showing least-square means model predictions of total isoflavone amounts in the T5 triad.

Principal component analyses

Overall root metabolite profiles were used for principal component analysis of all samples (Figure 3.11). From this analysis, D4 and *G. clandestina* appear to have the most distinctive profiles of all species studied. In addition, profiles for each allopolyploid triad were analyzed by principal component analyses, with control samples and inoculated samples analyzed separately (Figures 3.12-3.14). In the case of the T2 allopolyploid triad, principal components (PCs) 1 and 2 accounted for 37.4% and 33.7% of the total variance, respectively, for the control samples, and 38.0% and 31.4% for PC1 and PC2 for the inoculated samples. Here, T2 appears to form an intermediate group between D3 and D4. For the T5 allopolyploid triad control samples, PC1 accounted for 41.9% of the total variance while PC2 accounted for 18.9%. In T5 inoculated samples, PC1 accounted for 43.3% and PC2 accounted for 18.4% of the variance. Lastly, PC1 and PC2 accounted for 40.5% and 30.5% of the total variance in the T1 allopolyploid triad control samples and 38.3% and 21.4% in the PCA of the T1 triad inoculated samples.

In the redundancy analysis, using the 18-compound dataset, species was a significant factor in explaining the differences between samples in the T2 triad ($p=0.002$), while inoculation treatment was not ($p=0.517$). While the allopolyploids in the other triads (T1 and T5) seemed to be placed primarily between the samples of the two corresponding diploids, species was not a significant explanatory factor for these triads (T1: $p=0.357$; T5: $p=0.067$). Treatment did not have a discernible, significant effect in either the T1 ($p=0.603$) or T5 ($p=0.792$) triads.

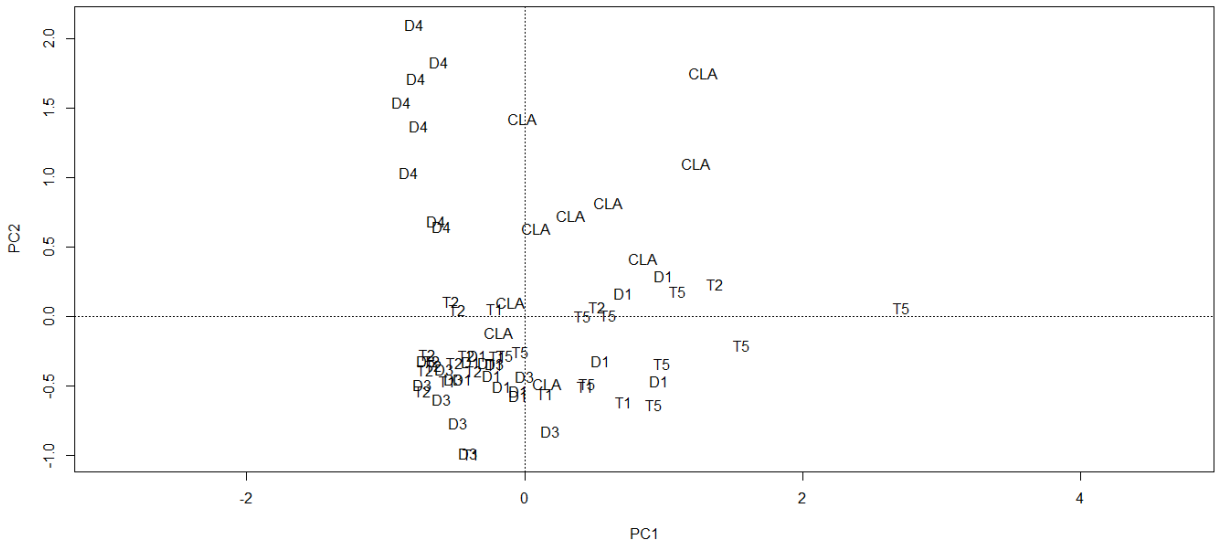


Figure 3.11. PCA plot of all samples by species. PCA was based on the complete 18-compound dataset.

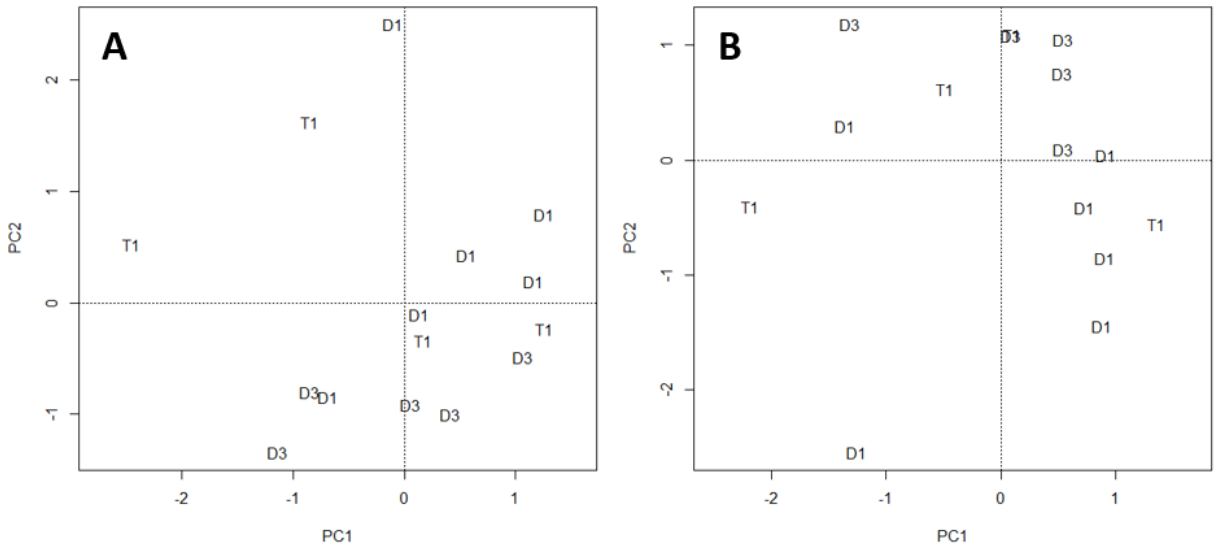


Figure 3.12. PCA plots of T1 triad samples by species for (A) control and (B) inoculated samples. PCAs were based on the complete 18-compound dataset.

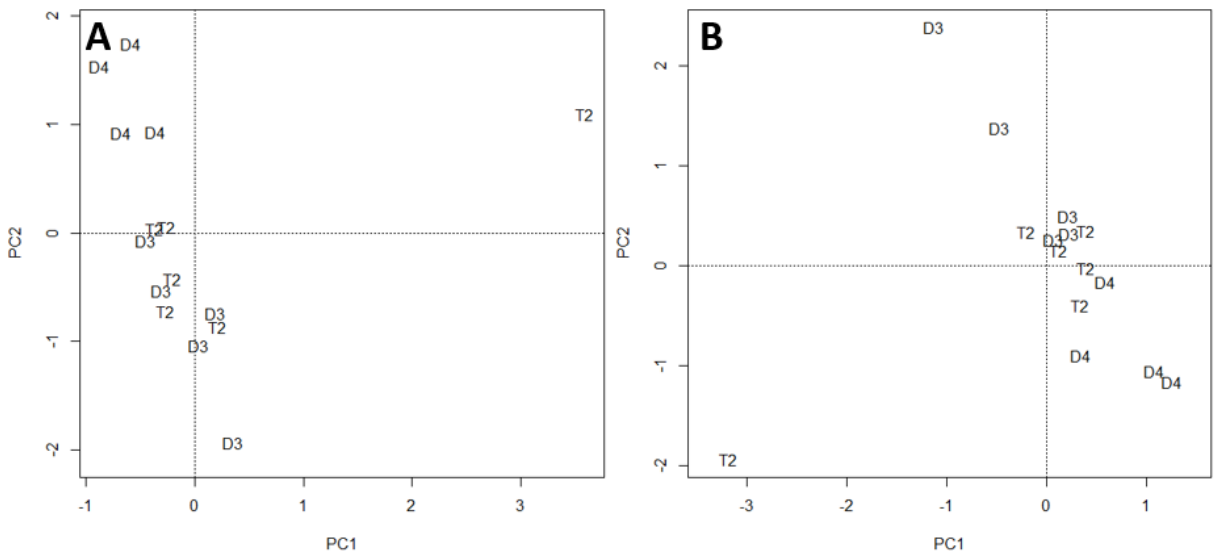


Figure 3.13. PCA plots of T2 triad samples by species for (A) control and (B) inoculated samples. PCAs were based on the complete 18-compound dataset.

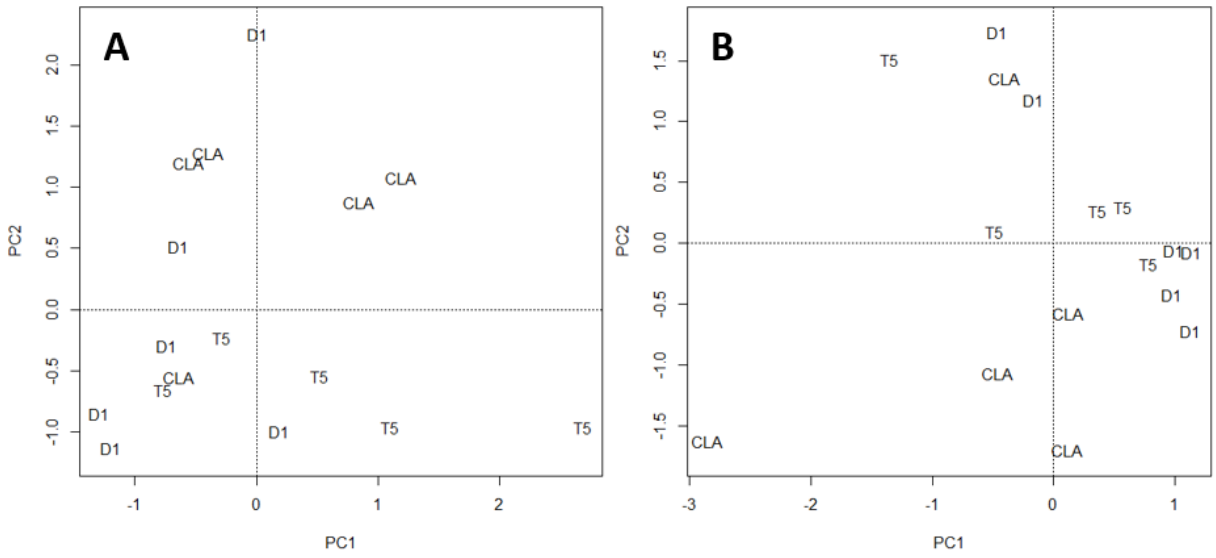


Figure 3.14. PCA plots of T5 triad samples by species for (A) control and (B) inoculated samples. PCAs were based on the complete 18-compound dataset.

Random forest on T2 triad tissue metabolites

To confirm the results of the PCA and redundancy analysis, as well as to identify key compounds in the roots that are primarily responsible for the distinctiveness of the T2 triad profiles, we examined the root metabolite data using random forest. The estimates of prediction error rates for distinguishing samples of a given species from other samples in an allopolyploid triad show that low error rates were obtained for D3 and D4 in the T2 triad, while the control *G. clandestina* samples in the T5 triad and the inoculated D1 samples, also in the T5 triad comparison, had comparably low estimated prediction error rates (Table 3.2). In most cases, the allopolyploid samples were associated with a higher estimated prediction error than either of the relevant diploids (particularly in relation to the error rates at random when information from the groups is not used).

Table 3.2. Error rates generated using the varSelRFBoot() function with 200 bootstrap samples. Random error denotes the minimal error obtained if information from the metabolites in the 18-compound set is not used. Low RF error rates showing improvement on the random error indicate that the metabolite profile of a given species can be identified as distinctive from other samples in the triad by random forest.

Classification	Random Error – Control	RF Error Rate – Control	Random Error – Inoculated	RF Error Rate – Inoculated
T1 Triad				
D1 versus the rest	0.400	0.397	0.400	0.322
D3 versus the rest	0.333	0.153	0.333	0.210
T1 versus the rest	0.267	0.390	0.267	0.430
T2 Triad				
D4 versus the rest	0.267	0.047	0.267	0.062
D3 versus the rest	0.333	0.069	0.333	0.058
T2 versus the rest	0.400	0.449	0.400	0.238
T5 Triad				
D1 versus the rest	0.375	0.216	0.375	0.043
CLA versus the rest	0.313	0.098	0.313	0.223
T5 versus the rest	0.313	0.298	0.313	0.447

Given the results in table 3.2, indicating that the most distinctive metabolite profiles are present in the species of the T2 triad, we sought to identify the components of the profiles in the T2 triad that contributed most to their distinctiveness. The MDA values in the T2 triad suggest that formononetin levels are the key distinguishing feature for D3 samples in relation to the other samples in the T2 triad, and formononetin was also the compound with the highest MDA values in the T2 samples (Figures 3.15-3.17). Formononetin also tended to have relatively high MDA values for the D4 samples. In the control samples, daidzein also had a high MDA value in D4. Based on these MDA values, the metabolite formononetin appears to be an important feature distinguishing the three species in the T2 triad.

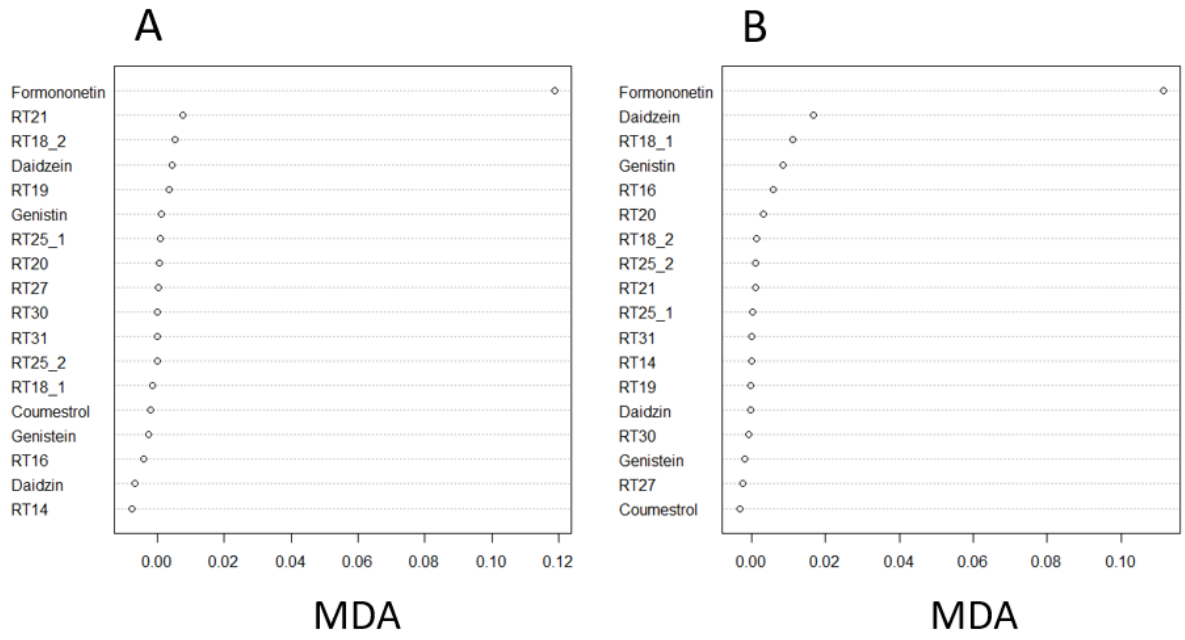


Figure 3.15. Mean decrease in accuracy (MDA) when individual compounds are removed from the model for D3 versus the rest of the samples (from D4 and T2), when (A) control and (B) inoculated samples are analyzed.

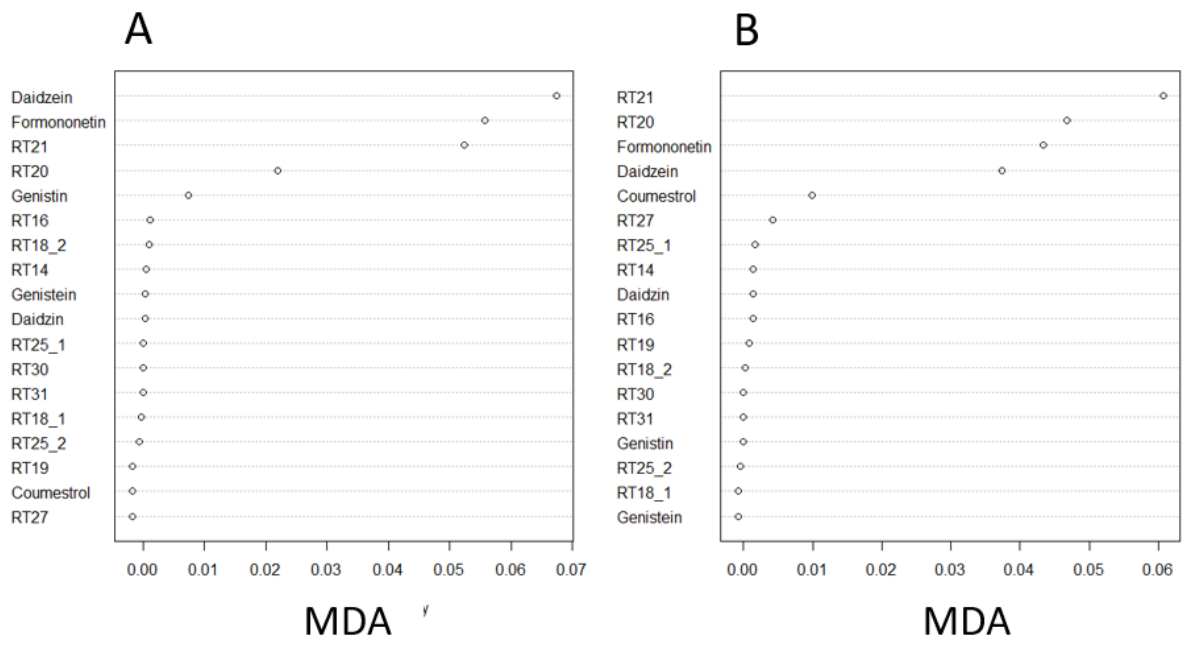


Figure 3.16. Mean decrease in accuracy (MDA) when individual compounds are removed from the model for D4 versus the rest of the samples (from D3 and T2), when (A) control and (B) inoculated samples are analyzed.

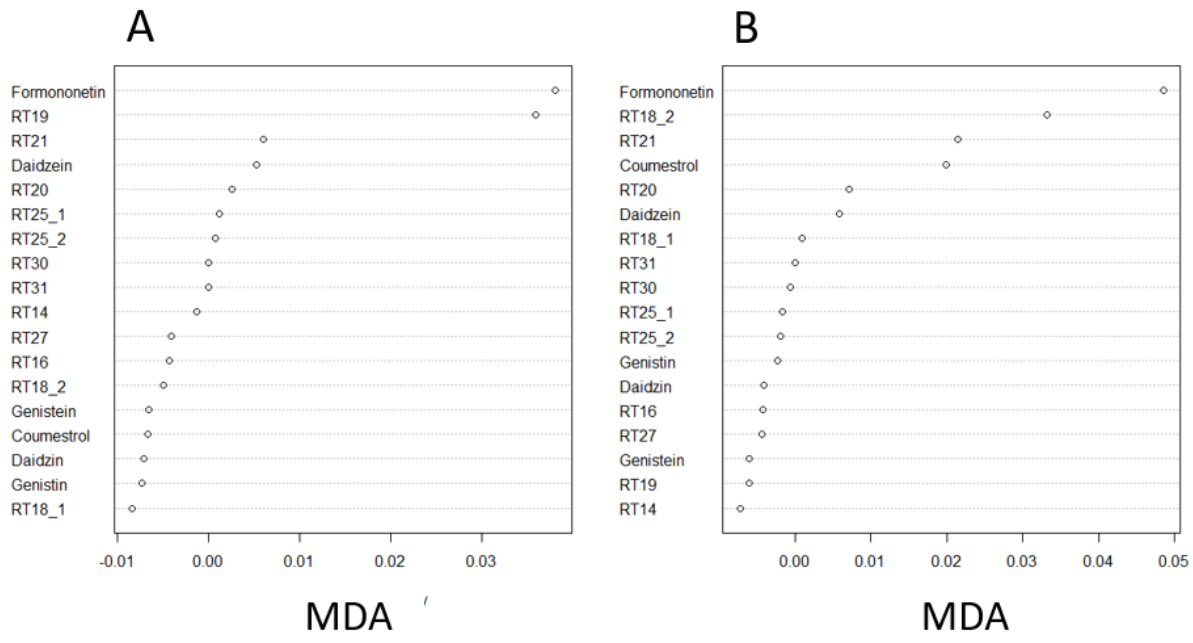


Figure 3.17. Mean decrease in accuracy (MDA) when individual compounds are removed from the model for T2 versus the rest of the samples (from D3 and D4), when (A) control and (B) inoculated samples are analyzed.

Analyses of root exudates

For root exudates, four key signaling compounds were analyzed and quantified: daidzein, genistein, coumestrol and formononetin; daidzin and genistin were not quantified because they were not observed in any exudate samples in this study. In the T1 triad, there were no differences between species in any of the four compounds ($p > 0.05$), while in the T5 triad, there was a significant difference in the levels of formononetin ($p = 0.0037$), with the levels in *G. clandestina* and T5 exudates being greater than those in D1 (Figure 3.18).

Given the apparent differences in the percentages in root metabolite profiles of the T2 triad and the relative importance of formononetin as a distinguishing feature between the species, the T2 triad exudates were also analyzed and the predominant components were quantified. In

the root tissues, formononetin was significantly different between species ($p=0.0016$), present at greater levels in the D4 samples compared D3, with T2 as an intermediate (Figure 3.19); a similar pattern was present in the exudates, where D4 exuded substantial amounts of formononetin, while both T2 and D3 exuded negligible levels of the compound (Figure 3.18). Daidzein was also detected at greater levels in D4 roots than in D3 or T2 ($p=0.022$), but levels in the exudates were not significantly different between the three species ($p>0.05$) (Figure 3.18; Figure 3.19). Genistin was not detected in the exudates, though genistin constituted a greater percentage of compounds in the roots of D3 samples compared to D4 samples.

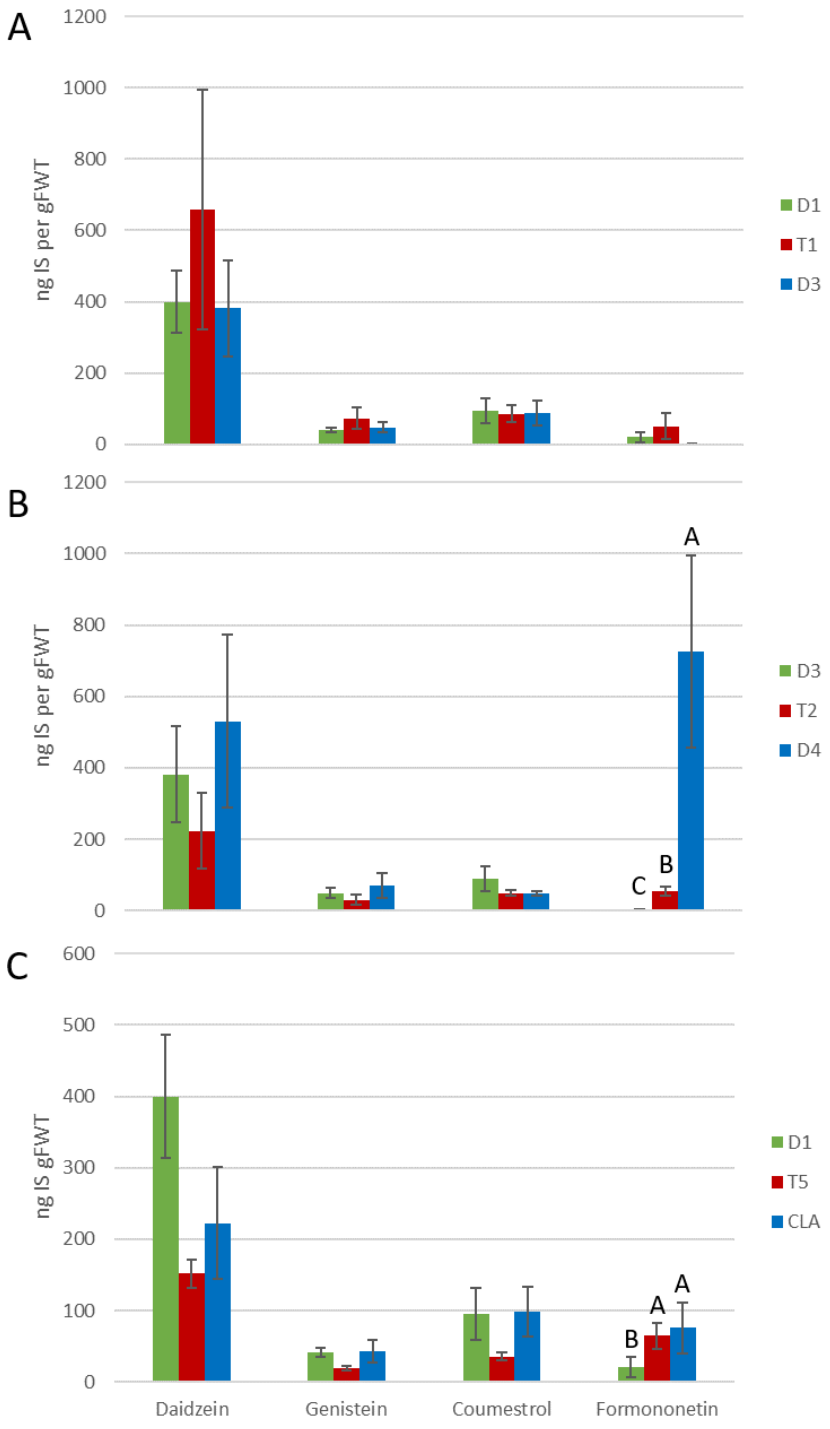


Figure 3.18. Concentrations of daidzein, genistein, coumestrol and formononetin in exudates of species from the (A) T1, (B) T2 and (C) T5 triads. With compounds for which significant differences were detected between species, values that the same letter group are not significantly different ($p>0.05$). Where no letter grouping is shown, no significant differences were detected between species for these compounds.

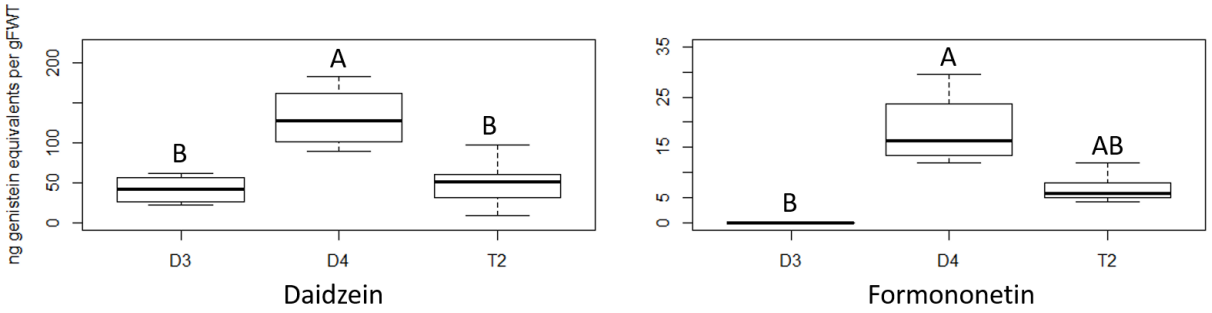


Figure 3.19. Concentrations of daidzein and formononetin in root tissues of species in the T2 triad. Boxplots show the median and interquartile range, with whiskers indicating minimum and maximum values. For each compound, species sharing the same letter group are not significantly different ($p > 0.05$).

Expression Analyses

Based on the differences in root metabolite profiles for D3, D4 and T2, we also looked at differential expression of several genes in the flavonoid biosynthetic pathway (Figure 3.20). For a series of genes in the biosynthetic pathway leading to isoflavones, including those encoding phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumaroyl:CoA-ligase (4CL), and isoflavone synthase (IFS), there was general upregulation in each species following inoculation, with greatest upregulation in D3, with considerably less in D4, and intermediate upregulation in T2. Analyzing control and inoculated samples for the three species separately, expression of chalcone reductase (*CHR*), a gene encoding an enzyme involved in isoflavone biosynthesis at the step differentiating the daidzein and genistein pathways, was examined for between species differences. One gene encoding CHR (Glyma.16G219500) showed differential expression in pairwise comparisons between the three species. Expression of this chalcone reductase gene was significantly higher in D4 relative to D3 (Wald test p-value < 0.001) and T2 (Wald test p-value = 0.005) in inoculated samples, and while p-values were greater than 0.05 in

control samples, expression showed a similar pattern when compared to D3 (Wald test p-value = 0.055) or T2 (Wald test p-value = 0.137). Another *CHR* paralogue (Glyma.14G005700), showed a similar pattern, with D4 expression greater than D3 (Wald test p-value = 0.088) and T2 (Wald test p-value = 0.101) in control samples, though Wald test p-values were substantially greater in comparisons between inoculated samples.

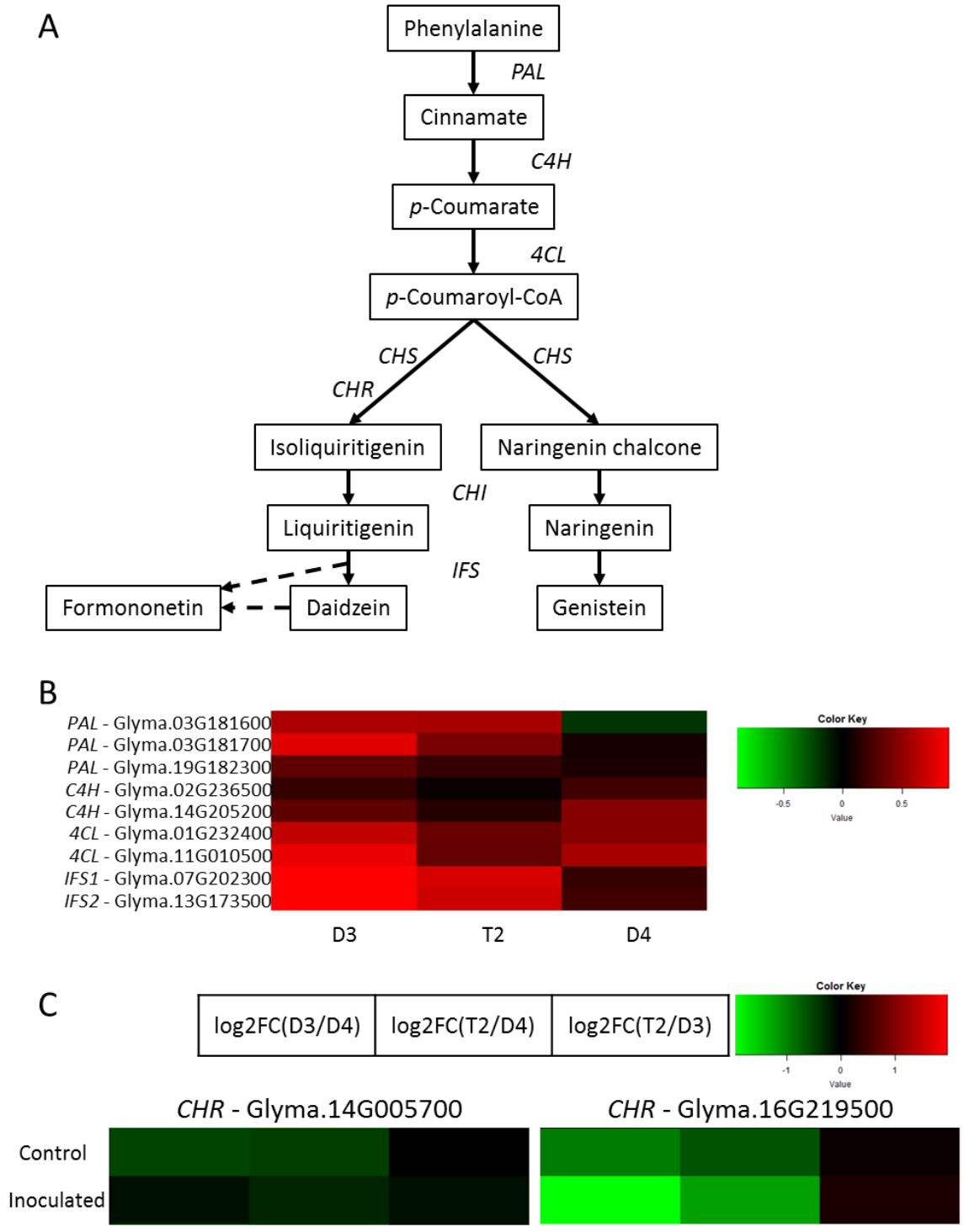


Figure 3.20. Expression data for genes encoding enzymes in (A) the biosynthetic pathway leading to isoflavones, including (B) comparison of differential expression of key genes in the pathway in terms of log fold changes between inoculated and mock-inoculated samples of each species and (C) log fold change in expression between species for genes encoding chalcone reductase.

DISCUSSION

In this study, we examined several aspects connected to metabolite presence, biosynthesis and exudation in diverse accessions of three natural allopolyploid species, along with their diploid progenitors. Given that these allopolyploids were not synthetically generated, they integrate the effects of polyploidy and hybridity, as well as selection and evolution (Buggs et al. 2014) in the context of nodulation symbiosis. Ample evidence exists to suggest that enrichment of flavonoid biosynthesis can occur following polyploidization (e.g., Mears 1980; Levin 1983; Orians 2000). Previous observations of duplicate gene retention and differential gene expression suggested that polyploidy might lead to qualitative and quantitative enrichment of flavonoid exudates involved in rhizobial signaling (Li et al. 2013). Here, we studied connections between root metabolite profiles, flavonoid exudation, and differential gene expression of candidate genes involved in flavonoid biosynthesis. Our findings relate to, and have implications for, several areas of research: (1) the effects of allopolyploidy and inoculation treatment on overall profiles of symbiotic signaling metabolites, (2) biosynthesis and exudation of signaling compounds in *Glycine*, and (3) differences in colonization ability in species of *Glycine* subgenus *Glycine*.

Effects of allopolyploidy and inoculation treatment on metabolite profiles

In previous studies using members of *Glycine* subgenus *Glycine*, daidzein was the phenolic compound with the highest concentrations in roots and genistein was also among the predominant compounds in root phenolic profiles (Lin et al., 2005; Lin et al., 2010). This is comparable to our results here; however, we have sampled a broader variety of species than these earlier studies, which focused on *G. dolichocarpa* (T2), *G. tabacina*, and *G. tomentella* (T4) (Lin et al. 2005). Furthermore, we examined the effects of both allopolyploidy and inoculation treatment on root metabolite profiles.

Inoculation treatment did not appear to have an effect when considering overall profiles or compound diversity, but did affect total levels of compounds in some cases. To date, the most comprehensive root metabolome analyses in the genus *Glycine* have been conducted using *G. max* (soybean). While our study focused largely on metrics for overall profiles and total amounts, studies of soybean root hairs have found differences in a variety of particular metabolites between inoculated and control samples. Sixteen (iso)flavonoids were regulated at one or several time points following inoculation (Brechenmacher et al. 2010). However, these regulated compounds did not include the key nodulation-regulated soybean compounds daidzein and genistein, nor did they include formononetin, daidzin, genistin or coumestrol. The symbiotic significance of the regulated compounds identified by Brechenmacher et al. (2010) is thus not entirely clear, and the relevance of focusing exclusively on root hairs in the context of isoflavones is not certain given that evidence related to gene expression of isoflavone synthase, the enzyme responsible for biosynthesis of isoflavones, indicates that, while expression is greater in roots than in other tissues, localized expression occurs in root epidermal cells that are not root hairs, in addition to occurring in root hair cells (Subramanian et al. 2004). Nevertheless, most of the inoculation-responsive compounds identified by Brechenmacher et al. (2010) were upregulated by inoculations, though several were downregulated as well, and increased *IFS* gene expression in response to inoculation (Subramanian et al. 2006) further suggests that total biosynthesis can be enhanced by exposure to rhizobia. This last evidence is consistent with the findings here, in that species in the T2 triad showed general upregulation of genes encoding enzymes in the isoflavone biosynthetic pathways in response to inoculation. However, this contrasts with the apparent decrease observed here in total isoflavones, particularly in T2. Thus,

while T2 and T5 showed evidence of responsiveness in total levels of root compounds, this cannot directly be accounted for by the patterns of gene expression observed.

Given that species was not a significant factor for explaining metabolite profiles in two of the triads (T1 and T5; Figure 3.12, Figure 3.14), there does not appear to be any common, repeated transgressive effect of ploidy in the allopolyploids of this complex. Furthermore, while we cannot discount the possibility that the species in the T1 and T5 triads have different symbiotic interactions based on their metabolite profiles, we did not find evidence of a species-level difference between the profiles and, as such, we cannot infer any capacities for differing symbiotic signaling and interactions. In the case of the T2 triad, however, there is differentiation between the metabolite profiles of the three species. This suggests that species in this triad may differ in their interactions with potential symbionts. As T2 appears to have a broadly intermediate metabolite profile, this further confirms the absence of any common transgressive effect due to ploidy, and suggests that any change in the allopolyploid T2 is related to hybridity and the particular characteristics of its diploid progenitors, D3 and D4. This hybridity effect is also suggested by the combination, in T2, of metabolites present in D3 and D4. From the random forest results (Table 3.2), it appears that T2 is not differentiated from its diploid progenitors; instead, the difference between species in the T2 triad is mainly due to the difference between the diploid progenitors D3 and D4. Also, the diversity of compounds observed showed no clear difference between diploids and the various allopolyploids examined, although D3 and D4 differed again while T2 appeared to be intermediate. Thus, while its progenitors are likely to have differing capacities for symbiotic interactions, T2's symbiotic capabilities, as inferred from root metabolite profiles, would not be expected to differ as greatly from either diploid progenitor

as the progenitors differ from each other. If rhizobial interactions are specified by flavonoids, T2 would be expected to combine the repertoires of its progenitors.

Previous research on the effects of hybridity and polyploidy have yielded variable results. Hybridity in and of itself can lead to a variety of outcomes including production of all metabolites in progenitors, failure to express all metabolites from the progenitors or the production of novel metabolites, as well as a spectrum of quantitative variation (Orians 2000). Orians (2000) highlights several factors that affect the outcomes: parental taxa, hybrid origin, ploidy level, chemical class, and the genetics of expression. Evidence indicating that ploidy level is relevant for flavonoid-related outcomes was gathered as early as the 1970s with several studies showing that novel, transgressive metabolite profiles could be obtained in polyploids (Levy and Levin 1971, 1974; Levy 1976; Levin 1983), and more recent work has also shown associations between increased ploidy and greater accumulation of secondary metabolites (Lavania et al. 2012). The results obtained here, however, suggest a role for hybridity in determining metabolite profiles, rather than ploidy-related effects leading to transgressive patterns, and, while inoculation treatment appeared to affect overall levels of compounds, species and allopolyploidy did not.

Biosynthesis and exudation of signaling compounds in Glycine

Profiles of flavonoids present in root tissues are relevant in determining symbiotic signaling capabilities in that the types and quantities of flavonoids contribute to determining the potential range of rhizobia with which a host species can form symbiotic interactions (Kobayashi and Broughton 2008; Liu and Murray 2016). Nod factor signaling and specificity and, by extension, flavonoid signaling also continues to be important to the development of symbiosis once the rhizobia are in direct contact with, or inside, the root tissues, in the processes of

infection thread formation and progression (e.g., Walker and Downie 2000; Den Herder et al. 2007; Oldroyd and Downie 2008), and isoflavones also have a role as regulators of auxin transport in the process of nodulation (Subramanian et al. 2006). Thus, the metabolites found in root tissues are important to consider. However, exudation of metabolites is also critical since presence of flavonoids in the rhizosphere can serve as the initial signal to potential symbionts. In the context of signaling, we are thus also interested in how the levels of specific compounds observed in the root tissues relate to levels in the plant exudates, whether there is evidence of differential transport or whether what is observed in the exudates is purely a reflection of biosynthesis in the roots.

In general, the predominant compounds that we observed in the exudates of the various diploid and allopolyploid species have also been found in the analysis of exudates of *G. max* and *G. soja*, where the primary compounds were daidzein, genistein, and coumestrol (Pueppke et al., 1998). The abundance of daidzein and genistein in root exudates of perennial *Glycine* species presented here coincides well with their role as *nod* gene inducers in the typical rhizobial symbionts of *G. max*, and contrasts with predominant exudates of other legumes that preferentially interact with other species of rhizobia (Bolaños-Vásquez and Werner, 1997).

An isoflavone of particular interest is formononetin, which was present at high levels in both roots and root exudates of D4, absent in D3, and present at low levels in T2 (Figures 3.18-3.19). While daidzein and genistein are the common primary *nod* gene inducers in the symbioses between *G. max* and rhizobia, formononetin is believed to be an inhibitor of *nod* gene induction in *B. japonicum* and *E. fredii*; it is structurally similar to daidzein, but does not induce *nod* gene transcription to the same degree as daidzein (Kosslak et al. 1987; Kosslak et al. 1990; Pueppke et al. 1998). There is evidence that this is also true in the wide host-range symbiont, NGR234

(Bassam et al. 1988; Le Strange et al. 1990). Experiments with NodD transcriptional activators have found that flavonoids that were not *nod* gene inducers for a given species competitively bound to the transcriptional activator, but only the species-specific inducers both bound to NodD and led to *nod* gene expression in compatible rhizobia (Peck et al. 2006). Thus, one might expect that the formononetin present in the exudates of D4 plants could be a competitive inhibitor to nodulation with certain types of rhizobia.

Genistein, along with its glucoside genistin, is formed from naringenin, while daidzein and formononetin are formed from liquiritigenin. The branch point in flavonoid synthesis between these two pathways is the biosynthetic step when p-Coumaroyl-CoA is converted to either 4,2',4'-trihydroxychalcone or 4,2',4',6'-tetrahydroxychalcone (Figure 3.20). The increased relative amounts of both daidzein and formononetin in D4 thus suggest a shift from the biosynthetic branch involving genistein, as observed in most species studied here, to the second branch favored in D4. This is supported by the comparisons of transcript expression for chalcone reductase, the enzyme that differentiates the biosynthesis pathway leading to synthesis of 4,2',4'-trihydroxychalcone, liquiritigenin, and daidzein from 4,2',4',6'-tetrahydroxychalcone, naringenin and, subsequently, genistein (and genistin). Thus, based on the relative amounts of genistin, daidzein and formononetin in the roots of D3, T2 and D4, along with greater expression of chalcone reductases in D4 relative to D3 and T2, it appears that D4 upregulates the biosynthesis of not only formononetin but daidzein as well, at the expense, in relative terms, of the pathway leading to genistein and genistin. In the case of D4, then, any overall inhibition of *nod* gene induction due to formononetin synthesis and exudation could possibly be offset by the increased allocation to daidzein as well.

However, the substantially greater amounts of formononetin in D4 exudates compared with D3 and T2 were not matched by any concomitant increase in the daidzein (or genistein) present in the exudates of D4 over D3 and T2. The statistically intermediate levels of formononetin in the exudates of T2 are also consistent with the patterns of intermediacy observed in the results for overall root tissue profiles such as, for example, the principal component analyses. Furthermore, while the pattern of formononetin levels in the exudates matches that present in the root tissues, the fact that the relative levels of daidzein in the exudates are not reflected in the root tissue levels indicates that there may be differential regulation of transport and exudation of the daidzein in the three species. The influence of formononetin leads to the prediction that T2 should have enhanced symbiotic responses relative to D4 with strains in which formononetin has an inhibitory effect, and this is indeed what was observed in inoculation experiments involving the T2 triad and NGR234, as well as certain *B. japonicum* strains (Powell and Doyle 2016).

Given this, it is not only synthesis, but also differential exudation and transport of flavonoids into exudates that are important for determining realized symbiotic signaling. Much remains unknown about the mechanisms underlying transport and exudation of flavonoids in legumes. An ATP-binding cassette transporter of isoflavones was characterized in *G. max* (Sugiyama et al. 2007). Competition assays established that daidzein and genistein, as well as several other aglycones including formononetin, were secreted, while the glycosylated forms, such as daidzin and genistin, were not as readily recognized by this transport protein, which is consistent with the absence of daidzin and genistin in the exudates we examined. Transport activity also appeared to be independent of nitrogen nutrition, in that in vitro ATP-dependent transport of genistein into soybean root-derived plasma membrane vesicles was unchanged with

nitrogen addition. However, in terms of the relative inhibition of genistein transport in the competition experiments, both daidzein and formononetin had similar effects, suggesting that there was no differential transport of daidzein and formononetin (Sugiyama et al. 2007). From evidence relating to the characterized isoflavone transporter in *G. max*, then, we find a plausible explanation for the absence of glucosides observed in our study of the exudates, but evidence supporting a mechanistic explanation for differential exudation of daidzein and formononetin is still lacking.

Symbiotic signaling, allopolyploidy and implications for colonization

The differing potentials in symbiotic signaling capacity hold implications for colonization ability. As noted above, certain allopolyploids in *Glycine* subgenus *Glycine* appear to have colonized habitats such as Taiwan, islands in the South Pacific, and the Ryukyu Islands of Japan, while the diploids appear confined to Australia and Papua New Guinea (Harbert et al., 2014; Sherman-Broyles et al., 2014). Given that there were no clear patterns in our data differentiating the allopolyploids, as a group, from the diploid species, we do not have evidence to support any general effect of allopolyploidy on metabolites that could mediate a difference in colonization between ploidy levels. It is interesting to note, however, that of the three allopolyploids included in the present study, only T2 has been found at locations outside of Australia and Papua New Guinea (Taiwan). The range of T1 overlaps largely with the ranges of its diploid progenitors, and the range of T5 seems to be a small fraction of the area where the ranges of its diploid progenitor species overlap (Harbert et al., 2014; Sherman-Broyles et al., 2014). We have provided evidence of differences between the three allopolyploids in their root metabolite profiles and exudates, and the contrasts in ranges and colonization among the allopolyploids can be interpreted in the context of this evidence.

These three allopolyploids are similar in having formed within the last several hundred thousand years (Bombarely et al., 2014), but differ in the breadth of the cross involved in their formation. T1 is formed from closely related genomes, whereas T2 and T5 involved crosses between more divergent species. This could be relevant to the differences between the root metabolite profiles of these three allopolyploids, given that D4 and *G. clandestina* profiles are noticeably different from both D1 and D3. However, this alone cannot account for the apparent difference in colonization in T2, since both T2 and T5 resulted from hybridizations between an A-genome species (D4 and *G. clandestina*, respectively) and the equally distantly diverged species D3 (in the case of T2) and D1 (in T5), yet T5 had significantly less isoflavone diversity when compared to T2. Thus, the particular effects of hybridity on the metabolite profiles and exudates must be examined to account for differential symbiotic interactions and, ultimately, colonization.

For T1 and T5, we have little evidence suggesting clear differences between their metabolite profiles and those of their diploid progenitors (or, for that matter, between the profiles of the diploid progenitors, considered apart from the allopolyploid), though a stronger case can be made for distinctiveness in T5. On the basis of these allopolyploids having less distinctive profiles relative to their progenitors, we would not expect a differential symbiotic signaling capacity, enabling novel or varying symbiotic interactions, for any of the three species within each of these triads.

In the T2 triad, though, D3 and D4 differ in terms of metabolite profiles and there is differential exudation of formononetin in D4. T2 shows intermediacy, or perhaps it would be better to say additivity, between these two diploid progenitors. It thus has greater isoflavone diversity than T5 (the allopolyploid with the smallest range), suggesting a possible connection

between isoflavone diversity and range or colonization ability, and potential progenitor-specific increase in diversity in an allopolyploid, similar to the outcome predicted by Li et al. (2013). The interpretation of this outcome warrants caution, since increased diversity in T2 is likely contributed by the formononetin synthesis present in D4, and formononetin may not be an enhancer of symbiosis in certain interactions with specific types of rhizobia. Nevertheless, the particular combination of the properties of D3 and D4 can lead to a different potential capacity for interactions with the hypothetical spectrum of possible nodulating symbionts for allopolyploid T2. With the specific set of rhizobia tested to date, we have found that T2 tends to have an enhanced capacity for symbiotic interactions with rhizobia (Powell and Doyle, 2016). In turn, ecological experiments have provided evidence that the capacity for interaction with compatible nodulating rhizobia may contribute to the ability of certain plant hosts to become established in new habitats (Stanton-Geddes and Anderson, 2011). Thus, the metabolite profiles present in T2 could contribute, in part, to the ability of the species to occupy different geographic areas than its D4 diploid progenitor. The contrast between T2 and D3, however, is less clear and it is likely that other factors are contributing to enhanced interactions with rhizobia and enhanced colonization ability in T2 as compared to D3. We are currently exploring additional subsequent symbiotic responses in this allopolyploid triad at the level of transcriptional regulation that may contribute and differentiate the response of T2 from D3. Clearly, a causative relationship between isoflavone diversity and symbiont diversity is a hypothesis worth testing with field sampling in the native ranges of these species.

Conclusion

In this study, we found an absence of any transgressive metabolite profiles in allopolyploid species and there were no clear effects of inoculation treatment on metabolites in

three different allopolyploid species triads. However, in the T2 triad, distinctive profiles in diploid progenitors D3 and D4 yielded differences in the metabolite profile of the allopolyploid T2 (*G. dolichocarpa*). Differences relating to the levels of formononetin in the root tissues were reflected in the exudates studied for D3, D4 and T2. These differences will likely affect the potential range of rhizobia with which these species engage. Relatedly, we observed differential accumulation of transcripts encoding a key enzyme, chalcone reductase, in the flavonoid biosynthetic pathway.

CHAPTER 4

ENHANCED RHIZOBIAL SYMBIOTIC CAPACITY IN AN ALLOPOLYPLOID SPECIES OF *GLYCINE* (LEGUMINOSAE)

INTRODUCTION

Polyploidy is a prevalent phenomenon in the evolutionary history of plants. A whole-genome duplication (WGD) event is believed to have occurred in the common ancestor of all seed plants and a second WGD in the common ancestor of all flowering plants (Jiao et al. 2011), with many more events scattered among angiosperm lineages (Wendel 2015). Polyploidy can have a variety of phenotypic effects in plants, often yielding increased stature, larger organs, as well as altered biochemistry and other anatomical properties that have the potential to affect competition and environmental tolerance (Levin 2002; Ramsey and Ramsey 2014). Given these types of phenotypic changes, researchers have explored diverse evidence for ecological effects including niche differentiation, increased colonization ability, and enhanced ecological tolerances and ranges of polyploids beyond those of their diploid progenitors (Lowry and Lester 2006; McIntyre 2012; Theodoridis et al. 2013), though conclusions appear to vary depending on the system studied and have been historically conflicting, particularly concerning changes to ecological tolerance (Stebbins 1985; Petit and Thompson 1999; Martin and Husband 2009). Nevertheless, Pandit et al. (2011); (2014) found that invasive plants tend to be polyploid, which may be related to competitive advantages of polyploidy in particular environments. te Beest et al. (2012) suggested that changes in biotic interactions, as well as changes to nutrient uptake and physiology, can be factors contributing to the ability of polyploids to colonize novel habitats and to extend their ranges beyond those of their diploid progenitors.

Several studies have examined the relationship between polyploidy and enhanced ecological tolerance through altered nutrient uptake and performance under diverse soil conditions (te Beest et al. 2012). Increased salinity tolerance in polyploids has been found in a wide range of species (e.g.,Meng et al. 2012; Meng et al. 2011; Yan et al. 2015). Polyploid *Solidago gigantea* also

exhibited greater tolerance to certain calcium treatments (Schlaepfer et al. 2010). Additional species-specific effects of ploidy have been found in relation to phosphate and sulfate uptake, when evaluated in wheat, sugar beet, and tomato (Cacco et al. 1976). Nutrient relations have the potential to be an influential determinant of ecological tolerance in polyploids. Nutrient status, in turn, can be affected by biotic interactions, notably those with soil microorganisms, and recent research has examined effects of polyploidy on mycorrhizal symbioses. In the orchid *Gymnadenia conopsea*, Těšitelová et al. (2013) found differences in mycorrhizal interactions between ploidy levels, and Sudová et al. (2014) observed interactive effects of ploidy and mycorrhizal inoculation on root mass, shoot phosphorus concentrations, and extraradical mycelium length in *Aster amellus*.

In addition to mycorrhizal interactions, nodulation is an important phenomenon affecting plant nutrient relations, notably for plants in the legume family (Leguminosae) (Sprent 2007; White et al. 2007). Nodulation is a symbiotic interaction between plant hosts and diverse nitrogen-fixing bacteria (collectively termed rhizobia) where the rhizobia are maintained in an organ known as a nodule, typically on the plant root, and fix atmospheric nitrogen, rendering it available to the plant host in exchange for carbon assimilates from the plant (e.g., White et al. 2007; Libault 2014). Recently, on the basis of genomic studies, it has been proposed that polyploidy had a role in the evolution of rhizobial nodulation and in the diversification of legumes (Young et al. 2011; Li et al. 2013). However, there is an absence of recent direct studies examining the potential interactions between polyploidy and nodulation.

Allopolyploids are the complex products of hybridity, genome doubling, and divergence subsequent to their origin (e.g., Buggs et al. 2014). In the case of allopolyploids, several competing predictions can be made about effects on rhizobial symbiotic interactions (Powell and

Doyle 2015). As fixed hybrids, recent allopolyploids combine the genomes, and thus the nodulation gene repertoires, of their diploid progenitors. If the ability to form an effective symbiotic partnership with particular bacterial genotypes is determined by host genes that facilitate symbiosis, then a simple expectation would be that the allopolyploid could form effective symbioses with the full range of bacterial genotypes hosted effectively by either of its diploid progenitors. In contrast, if the range of functional host-symbiont combinations is determined by genes that limit infection by bacteria, an allopolyploid might have fewer potential partners than either diploid. Alternately, transgressive effects could lead to capabilities for partnerships with unique rhizobial symbionts that cannot form effective symbioses with either diploid progenitor species.

The polyploid complex of *Glycine* subgenus *Glycine* provides a useful study system for examining these phenomena. As members of the legume family, these wild perennial relatives of the cultivated soybean, *G. max*, are capable of forming nodulation symbioses (Pueppke 1988; Pueppke and Broughton 1999). The complex consists of approximately 30 species, mostly diploids, but including eight allopolyploid species formed from extant diploid genomes in various combinations (Sherman-Broyles et al. 2014). Allopolyploids in this complex are estimated to have been formed within the last half million years (Bombarely et al. 2014), thereby providing an opportunity to study impacts of allopolyploidy, coupled with subsequent natural selection, on nutrient relations and nodulation interactions. Within *Glycine* subgenus *Glycine*, the allopolyploid *G. dolichocarpa* (a member of the "*G. tomentella* complex" designated T2) was formed, with multiple origins, from the diploid progenitors *G. syndetika* (D4) and *G. tomentella* s.l. (D3) (Doyle et al. 2002; 2004). Five of the eight allopolyploid species in *Glycine* subgenus *Glycine*, among them T2, have expanded beyond their native Australian ranges, having

colonized numerous islands in the Pacific Ocean as far north as the Ryukyu Islands of Japan (e.g., T2 has colonized Taiwan), whereas the diploid species are all confined to Australia and, in the case of D3, Papua New Guinea (Doyle et al. 2002; 2004). T2 shows many differences from its diploid progenitors in such traits as photosynthesis-related phenotypes, photoprotection, gene transcription and translation (Coate et al. 2012; 2013; 2014; Ilut et al. 2012); any of these could help account for its abilities as a colonizer. Similarly, rhizobial symbiosis, tolerance of various nutrient conditions, and the interaction between these two may also contribute to T2's apparent increased ability for colonization.

In this study, we tested whether the allopolyploid T2 has a greater capacity for rhizobial interaction and differential responses to inoculation and nitrogen treatment than do its two diploid progenitor species. We assayed "symbiotic capacity"—the capability to form effective rhizobial symbioses, which includes early signaling interactions, as well as nodule formation. We ask several questions regarding how recent allopolyploidy, coupled with subsequent natural selection, affects responses to nitrogen and rhizobial inoculation: (1) Is there evidence of an increased range and probability of potential rhizobial interactions in T2 when compared with D3 or D4? (2) What effect do nitrogen, inoculation and species have on total nodule mass per plant and on nodule number per plant? (3) Is plant performance and competitive ability, as measured by fresh plant biomass, different for the three species? For each species, how is plant growth affected by nitrogen treatment and rhizobial inoculation? If responses to nitrogen levels and rhizobial symbionts (particularly where nitrogen is limiting) are contributing to T2's apparent colonization ability, we expect either transgressive responses with potential symbionts and enhanced performance under the treatment conditions tested or the additive combination of parental traits and capabilities that maximize beneficial symbiotic interactions. If T2 possesses a

more restricted and reduced symbiotic capacity relative to either of its diploid progenitors, then it is less likely that such interactions contribute to any unique colonization ability in T2.

MATERIALS AND METHODS

Study system

For the preparation of experimental rhizobial cultures, lyophilized samples of bacterial strains NGR234, USDA110, USDA138, USDA191, and USDA257 were obtained from the USDA-ARS Rhizobium Germplasm Resource Collection. All lyophilized samples were revived and isolated according to protocols supplied by the USDA-ARS Rhizobium Germplasm Resource Collection. The strains that were used in the experiment were all symbionts that were not isolated from any member of *Glycine* subgenus *Glycine*, nor do they originate from areas where such plants were present, such that they represent equally novel potential partners for each of the three plant species in this study. USDA191 and USDA257, as representative strains of *Ensifer* (*Sinorhizobium*) *fredii*, were isolated in mainland China; USDA191 was isolated from soil in Shanghai and USDA257 was isolated from *G. soja* (a member of *Glycine* subgenus *Soja*) in Henan (Chen et al. 1988). The two strains of *Bradyrhizobium japonicum*, USDA110 and USDA138, were included to test further the symbiotic range of these species, as these were isolated from *G. max* in the United States (Cassman et al. 1981; Kaneko et al. 2002). *Ensifer fredii* strain NGR234, a broadly nodulating strain also included in this study, was isolated from *Lablab purpureus* in Papua New Guinea (Trinick 1980).

T2 plant accessions used were G1134, G1188, and G1393; D3 accessions were G1364, G1403, and G1820; and D4 accessions were G1300, G1772, G1775, G2073, and G2321. For inoculation trials, seeds of multiple accessions per species were also used from diploid progenitors D3 (accessions G1364, G1403, G1686, and G1820) and D4 (accessions G1300,

G1772, G1775, G2073, and G2321), while four accessions were used from T2 (accessions G1134, G1188, G1393, and G2809). The accessions used in this study were chosen to sample geographic and genetic diversity of each species (Brown et al. 2002; Doyle et al. 2002). The accessions represent samples collected and propagated by selfing by the Commonwealth Scientific and Industrial Research Organization (CSIRO).

Root hair deformation assay

For the root hair deformation assay, three seedlings per accession per treatment were evaluated. Rhizobial inoculum was prepared by growing separate rhizobial strains in arabinose gluconate (AG) media for 5 d for *E. fredii* and 7 d for *B. japonicum* at 30°C and bacterial cell concentrations in media were determined with a Helber counting chamber (Somasegaran and Hoben 1994) and an Olympus CX41RF microscope (Olympus, Tokyo, Japan) and 40× objective. Cultures of approximately 10^9 bacterial cells per mL were used to inoculate seedlings. Seeds were surface-sterilized for 5 min in 40% ethanol, followed by 5 min in 20% commercial bleach. The seeds were then washed in sterile water three times and left to soak in sterile water for 2 h. Under sterile conditions, the seeds were nicked and placed in petri plates with 8 mL of sterile water. Five-day-old seedlings were transferred to microcentrifuge tubes. Cultures were pelleted and washed twice with sterile water. Seedlings were inoculated with 200 µL of bacterial culture containing 10^9 bacterial cells per mL. Sample tubes were coded and labeled, such that the researcher counting root hair deformations was blind to the identity of each plant sample. Seedlings were incubated with the bacteria for 60 h. Root hairs were then observed using phase contrast optics as described by Somasegaran and Hoben (1994), using the Olympus CX41RF microscope and 20× objective. Root hairs were observed over the entire length of one side of the root. All deformation counts were summed as a percentage of total root hairs observed, and these

were evaluated following exposure to rhizobia, as well as in water-treated control samples; the ratio of the percentage deformation in treatment roots to the mean percentage in control roots for each accession was then calculated.

Seedling inoculation trials

Seeds of the three species were sterilized and nicked as described already. The seeds in petri plates were then stored in the dark for 7 d. Following germination, the seedlings were separated from their seed coats under sterile conditions and were moved to growth tubes containing 10 mL of Jensen's media each. N-free and N-added Jensen's media was prepared according to Somasegaran and Hoben (1994). The N-added media contained 0.1% KNO_3. Equal numbers of plants per accession were randomly allocated to the N-free or N-added tubes (3 to 5 per accession per nitrogen treatment per inoculation treatment). The tubes were then inoculated with rhizobial treatment or control media.

For inoculation treatments, liquid cultures of each rhizobial strain were grown to stationary phase, as described. Each treatment plant seedling received 10^9 cells by pipetting, applied directly to the roots. Each control plant was mock-inoculated with an equal volume (0.5 mL) of sterile media. The plants were placed in a growth room under 125 µmol light, with 16 h days/8 h nights at 22.5°C for 10 wk. All plants received an equal volume of additional water under sterile conditions at 2, 6, and 8 wk postinoculation. At the end of 10 wk, all plants were removed from media, plant tissue was weighed (aboveground and belowground tissue were separated at the hypocotyl, and nodule number and mass were assessed. Due to variation within species for ability to nodulate, we assessed the effect on plant biomass of exposure to rhizobia regardless of whether the plant was able to form nodules. In the context of colonization, we are primarily interested in the comparison of overall growth responses between species with exposure to

rhizobia; less growth in one species due to reduced nodulation is relevant and appropriate when considering colonization ability, and this approach has been used previously (e.g., Parker et al. 2006). While acetylene reduction assays are most accurate in determining levels of nitrogen fixation, for the purposes of this study a standard scoring method for effectiveness of nodules, assessing red coloration in dissected nodules as described by Somasegaran and Hoben (1994) was used.

Statistical analyses

For the root hair assays, the ratios of root hair deformation in treatment plants to controls were analyzed using a model including rhizobial inoculation treatment and plant species as fixed effects, with accession included as a random effect, along with the interaction of plant species and rhizobial treatment.

To assess the propensity for nodulation of the three species, data on the presence or absence of nodules on each plant was evaluated with a logistic regression model. The model was applied to data for the trial with the strain NGR234, using species and nitrogen as fixed effects and accession as a random effect. The interaction between nitrogen and species was also included in the model. Pairwise differences were assessed using Tukey's honestly significant difference (HSD) test. The data for tests using other strains were not amenable for several statistical tests due to the absence of any nodulation in numerous cases.

Additional statistical tests were conducted on the data from the NGR234 trial. A normal mixed-effects model was used to evaluate significant factors affecting plant fresh biomass. Species, nitrogen, and rhizobial inoculation (i.e., whether a sample was inoculated with NGR234 or was mock-inoculated) were included as fixed effects, along with plant accession as a random effect, and the interactions between the fixed effects were also included.

Hurdle models were run for total root nodule mass per plant and number of nodules per plant, due to the substantial number of zero values in these data sets. The first part of each hurdle model consisted of a logistic regression for zero and nonzero values. For both of these hurdle models, the results of the logistic regressions were identical to the logistic model described above. The second part of each hurdle model involved a normal ANOVA model fit on the nonzero subsets of each dataset. To improve the normality of the residuals, the models were run on the log-transformed values for both data sets. For both steps of each hurdle model, nitrogen treatment and species were included as fixed effects and accession was included as a random effect. The interaction term between species and nitrogen was also included. Tukey's HSD test was used to examine pairwise differences between groups. All analyses described were conducted using the R packages lme4 (Bates et al. 2015) and lsmeans (Lenth and Hervé 2016).

RESULTS

Root hair deformation

In assessing root hair deformations in response to rhizobial exposure, numerous types of deformation phenotypes were observed, including curling, branching, wiggling, and bulging, as has been previously noted in studies using *G. max* and *B. japonicum* (Duzan et al. 2004). In roots treated with rhizobia, accession averages ranged from near zero to approximately 11% of root hairs deformed. In general, root hair deformation was greatest in T2 (Figure 4.1). For the overall model, plant species was a statistically significant predictor of root hair response ($F_{2, 8} = 8.87$, $P < 0.01$). Pairwise differences between species showed that the greater root hair deformation in T2 was significantly different from both D3 (Tukey's HSD, $P < 0.01$) and D4 (Tukey's HSD, $P < 0.05$).

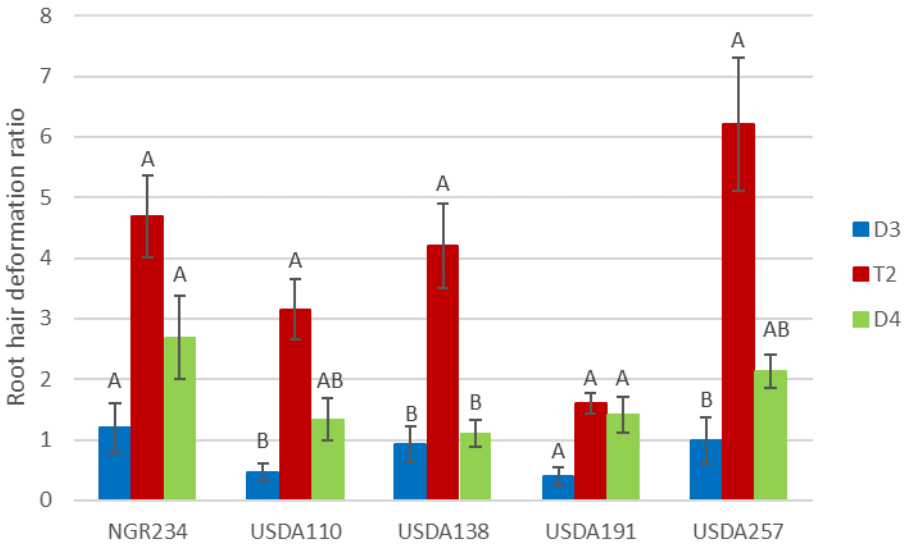


Figure 4.1. Root hair deformation ratios (deformation percentage in treatment plants divided by mean accession percentage in control plants) for D3, T2, and D4 species. Values represent means ± SE of all samples for a species (N = 9, 9, 15 for D3, T2, and D4, respectively, for each rhizobial treatment). For each rhizobial treatment, letters above the bars indicate statistical significance; categories sharing a letter are not significantly different from each other (Tukey's HSD, $P > 0.05$), while those not sharing a letter are significantly different (Tukey's HSD, P □ 0.05).

In the overall model, treatment (i.e., the bacterial strain with which roots were incubated) was a significant factor ($F_{4, 142} = 9.00$, $P < 0.01$) and the interaction of plant species and treatment was also significant ($F_{8, 142} = 2.91$, $P < 0.01$), indicating that effect of species varied depending on the particular bacterial strain treatment. While our interest was not primarily in the variable effects of different bacterial treatments, we also explored unique effects of each strain on the three plant species. Models tested for each treatment strain separately showed that, in the case of USDA138, the curling ratio of T2 was significantly greater than those of D3 and D4 (Tukey's HSD, $P < 0.05$), while there was no significant pairwise difference between D3 and D4. With both USDA110 and USDA257, T2 was greater than D3 (Tukey's HSD, $P < 0.05$); no other pairwise differences were statistically significant. For NGR234 and USDA191, no significant differences were found despite T2 having numerically higher ratios in both cases.

Although differences between species, specifically between the allotetraploid and its progenitors, were of primary interest, variation at the level of accession within plant species was accounted for in the statistical models used here. Substantial differences can be observed between accessions for particular rhizobial treatments, and these differences can be observed for all three species (Appendix 1). Such detailed data are useful in understanding the correlations between root hair deformations and other explanatory variables connected to differences in signaling. Thus, despite differences within species, the root hair deformation assays indicate that differences in the responses to inoculation exist between species, and that T2 appears to be more responsive to rhizobial treatment than either of its diploid progenitors (particularly D3).

Inoculation trials

Root hair deformation occurs in the earliest stages of the establishment of rhizobial symbiosis, and as such, the root hair response can serve as a proximal indicator of symbiotic signaling. However, since the signaling that gives rise to root hair deformation and curling constitutes the earliest stages of symbiotic interaction, it does not necessarily follow that these will translate directly into differences in nodule formation and concomitant contributions to plant growth. We therefore also assessed nodule formation following inoculation.

We calculated the total proportion of plants for each accession and species that formed nodules following inoculation with one of the five rhizobial strains (Figure 4.2). Qualitatively, T2 appeared to have a greater capacity for nodulation, with instances of nodulation observed in T2 plants for all of the inoculation treatments with each of the five strains, while this was not the case with D3 and D4. Furthermore, with strains where nodules were observed on only one diploid species, the diploid that formed nodules was not consistent across strains; in the case of USDA110, nodules were found on some D3 plants but no D4 plants, while with USDA138 the

opposite was true. The overall proportion of nodulated plants was also greater for T2, compared with D3 and D4, for each of the strains tested, although overall proportions of nodulation were low for all three plant species with certain strains (e.g., USDA138 and USDA257). No nodules were observed on mock-inoculated control plants of any of the three *Glycine* species. The qualitatively greater capacity for nodulation with diverse strains observed for T2 suggests some potential for both a broader range of interactions with novel symbiotic partners and a greater propensity to form nodules with particular strains.

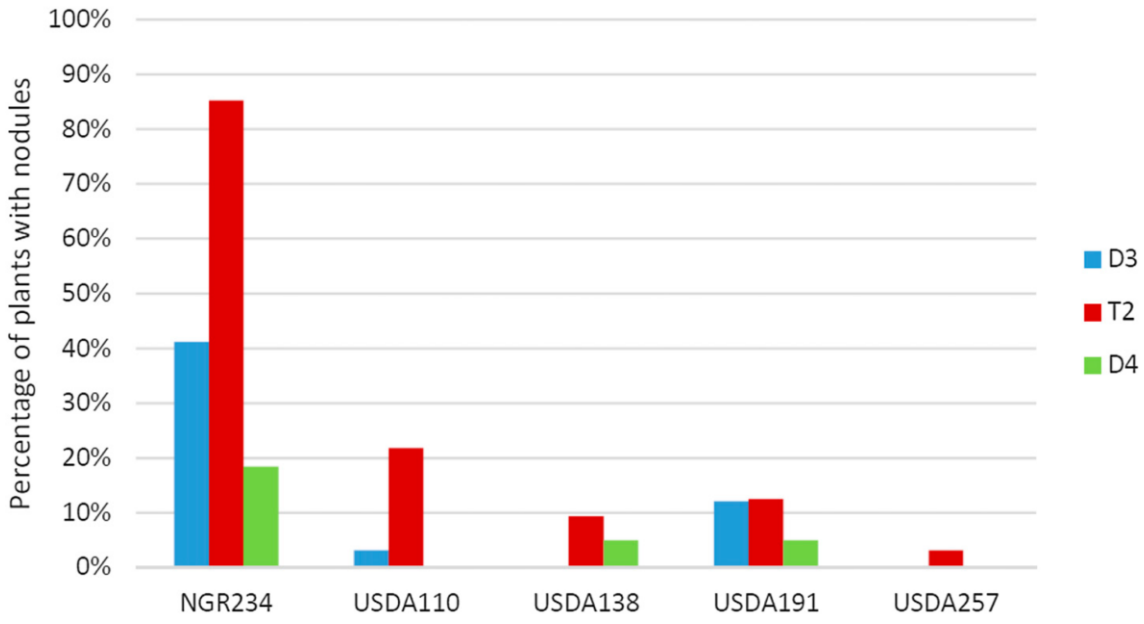


Figure 4.2. Percentage of plants with nodules for D3, T2, and D4 species following inoculation with rhizobia NGR234, USDA110, USDA138, USDA191, and USDA257.

A logistic regression model was used to test whether species and nitrogen treatment affected the probability of a plant forming any nodules when inoculated with NGR234. These statistical tests were not possible with other strains due to the preponderance of zero values, which led to violations of model assumptions. For the NGR234 treatment, species was a significant

explanatory factor (likelihood ratio test, $X^2 = 8.13$, df = 2, $P < 0.05$). The predicted probability of having nodules was greatest for T2, followed by D3 and D4, and the pairwise difference between T2 and D4 was statistically significant (Tukey's HSD, $P < 0.05$). However, there was no effect of nitrogen treatment (likelihood ratio test, $X^2 = 0.76$, df = 1, $P > 0.05$), nor was there a significant effect of the interaction between species and nitrogen (likelihood ratio test, $X^2 = 2.53$, df = 2, $P > 0.05$). Growth of putatively effective nodules when inoculated with NGR234, characterized by visible red coloration when opened, was also confirmed on eight T2 plants and three D4 plants, while, though nodules were observed on D3 plants, putatively effective nodules were not observed.

NGR234 nodule growth

The first component of our hurdle models (i.e., the first "hurdle") consists of a logistic regression model on nodule presence and absence data. For both nodule mass and nodule number, the logistic regression components of both hurdle models were identical to those of the logistic regression discussed already. The second step of each of the hurdle models involved taking only the subset of the data for which nonzero values occurred (i.e., plants for which nonzero values were obtained for total nodule mass per plant and nodule number per plant) and testing for differences between groups. The log-transformed total nodule mass per plant and nodule number per plant were evaluated and compared for those plants that had at least some nodules present. For the relevant subset of total nodule mass data, species was highly significant ($F_{2, 43} = 5.77$, $P < 0.01$). T2 had the greatest nodule mass, followed by D4 and then D3. In pairwise comparisons, T2 had a significantly greater total nodule mass per plant when compared with D3 (Tukey's HSD, $P < 0.05$), but D4 was not significantly different from either D3 or T2 (Figure 4.3A). The interaction between nitrogen treatment and species was not significant ($F_{2, 43}$

= 1.98, $P > 0.05$), and nitrogen treatment was also not statistically significant ($F_{1, 43} = 3.69$, $P = 0.062$). In the subset of log-transformed nodule number data, neither species ($F_{2, 6} = 0.65$, $P > 0.05$), nitrogen ($F_{1, 37} = 0.60$, $P > 0.05$), nor the interactions between species and nitrogen ($F_{2, 37} = 1.67$, $P > 0.05$) were significant (Figure 4.3B). Species and nitrogen were also not significant in the model excluding the interaction term. These results suggest that since, among nodulated plants, there was no difference in the number of nodules, T2 achieved the greater total nodule mass by producing nodules that were also individually larger, which corresponds with observations of a greater proportion of larger nodules on T2, followed by D4 and then D3 (observations not shown).

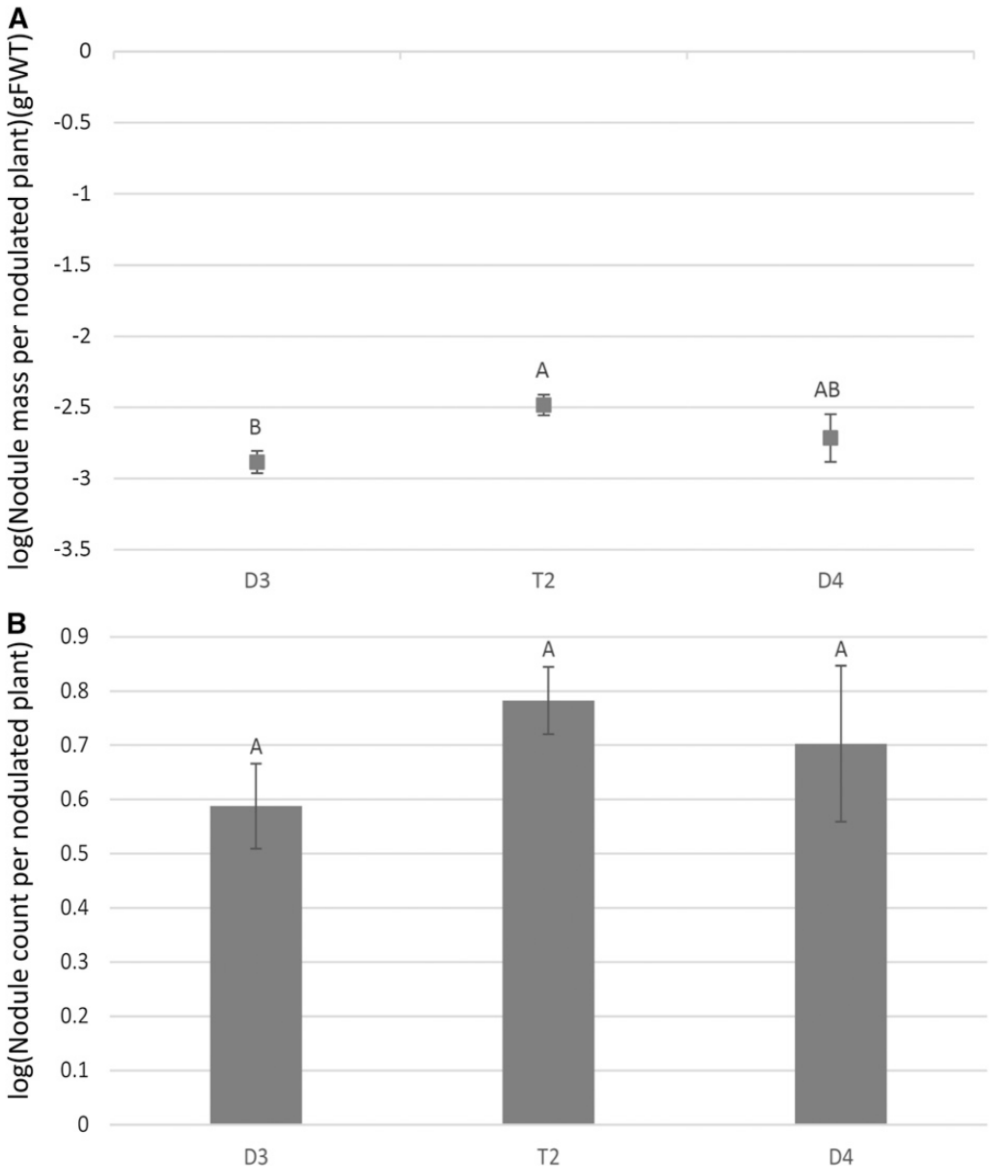


Figure 4.3. Nodule growth measurements for plants with nodules present following inoculation with NGR234. (A) Log-transformed total nodule mass per plant. (B) Log-transformed nodule count per plant. Values represent means ± SE (N = 13, 29 and 7 for D3, T2, and D4, respectively). Letters above the bars indicate statistical significance; categories sharing a letter are not significantly different from each other (Tukey's HSD, $P > 0.05$).

Plant fresh biomass

For assessment of biomass, we focused on the NGR234 treatment, since a greater proportion of nonzero values were obtained with this strain for the variables of interest. Fresh total plant

biomass was assessed for each plant species inoculated with NGR234 (Figure 4.4A). When considering all plants inoculated with rhizobial strain NGR234, species ($F_{2, 10} = 29.62$, $P < 0.01$), nitrogen treatment ($F_{1, 179} = 64.04$, $P < 0.01$) and inoculation treatment (i.e., inoculation with NGR234 or mock-inoculation) ($F_{1, 179} = 9.19$, $P < 0.01$) were significant predictors of plant total fresh biomass. There were also significant interactions between species and nitrogen treatment ($F_{2, 179} = 5.08$, $P < 0.01$) and between inoculation and nitrogen treatment ($F_{1, 179} = 10.81$, $P < 0.01$). Results for aboveground and belowground tissue, analyzed separately, were slightly different from those obtained for total plant biomass (Figure 4.4B, 4.4C). For aboveground tissue, species was a significant factor ($F_{2, 10} = 17.82$, $P < 0.01$), nitrogen treatment was significant ($F_{1, 180} = 83.83$, $P < 0.01$), and the interactions between species and nitrogen ($F_{2, 180} = 10.97$, $P < 0.01$) and between nitrogen and rhizobial inoculation ($F_{1, 180} = 22.68$, $P < 0.01$) were significant. For belowground tissue, species ($F_{2, 10} = 31.63$, $P < 0.01$), nitrogen treatment ($F_{1, 179} = 19.40$, $P < 0.01$) and rhizobial inoculation ($F_{1, 179} = 48.24$, $P < 0.01$) were all significant factors, and the interactions between species and rhizobial inoculation was also significant ($F_{2, 179} = 10.18$, $P < 0.01$). This significant interaction term indicates that the effect of inoculation on plant growth was dependent on the species. For belowground tissue, when a model was analyzed excluding inoculated plants that did not form nodules, the species, nitrogen, and inoculation terms remained significant, but the interaction between species and inoculation was no longer significant, suggesting that the differing growth response of species to inoculation was in large part due to the differing proportions of plants that nodulated for each species.

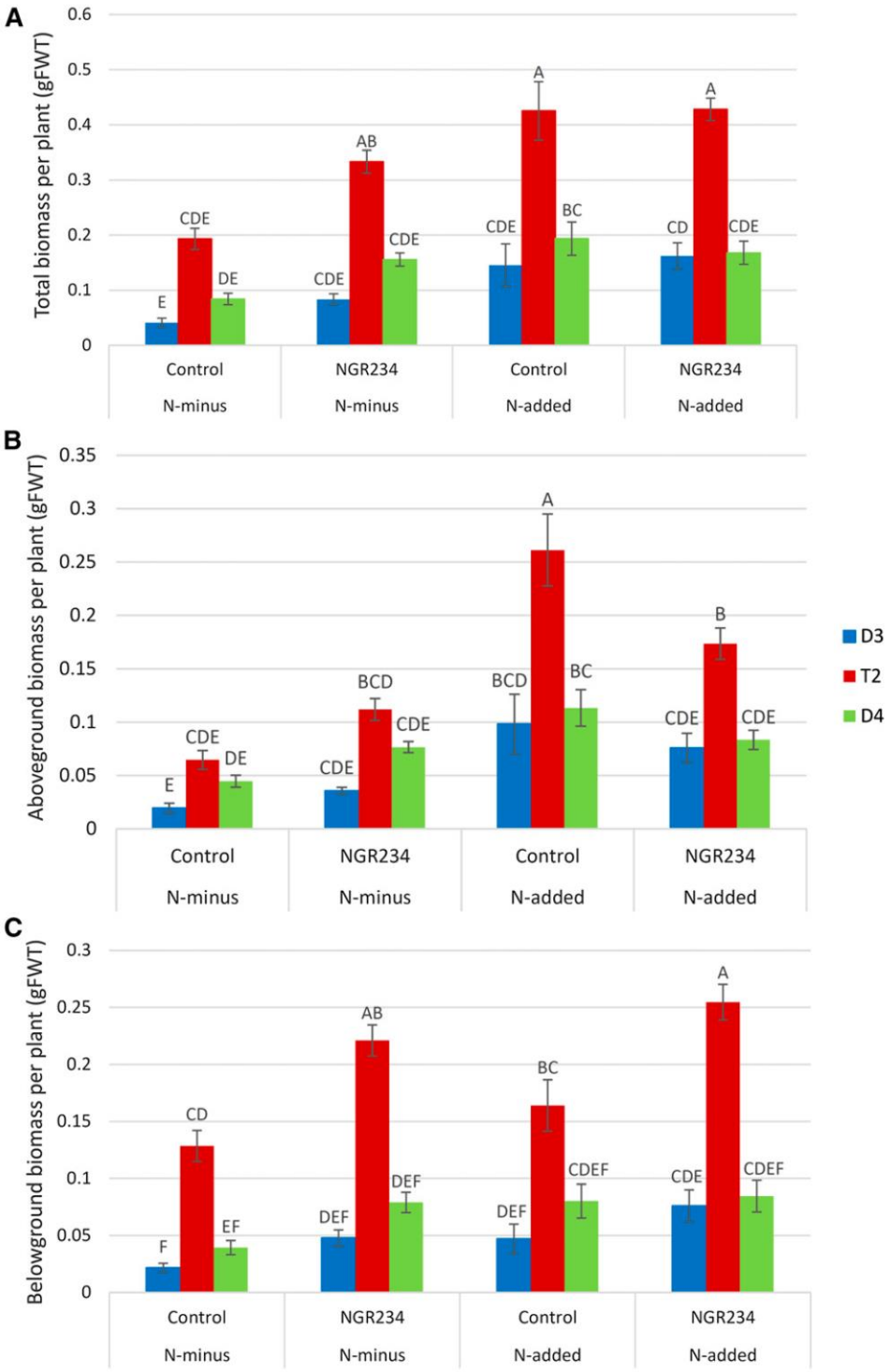


Figure 4.4. Fresh biomass as affected by plant species, nitrogen treatment, and inoculation with NGR234. (A) Total biomass. (B) Aboveground biomass. (C) Belowground biomass. Values represent means of all samples for each species ± SE (N = 15-19 per species per treatment group). Letters above the bars indicate statistical significance; categories sharing a letter are not significantly different from each other (Tukey's HSD, $P > 0.05$).

Using Tukey's HSD test for pairwise comparisons, T2 mock-inoculated, nitrogen-free plants were not significantly different from the D3 and D4 mock-inoculated, nitrogen-free plants (Tukey's HSD, $P > 0.05$). However, for the remaining treatment groups (inoculated and nitrogen-free, mock-inoculated with added nitrogen, inoculated with added nitrogen), T2 showed a significantly greater total fresh biomass than in the corresponding D3 and D4 samples (Tukey's HSD, $P < 0.05$). When compared with the low-nitrogen mock-inoculated controls, the biomass change (BC) for T2 was greater when exposed to NGR234 (BC = 0.140, 95% confidence interval [CI] = 0.082–0.198), while corresponding increases in D3 and D4 were not as large (D3 BC = 0.042, 95% CI = 0.016–0.068; D4 BC = 0.072, 95% CI = 0.039–0.104). Similarly, again in comparison with the low-nitrogen mock-inoculated controls, T2 with added nitrogen (without NGR234 treatment) increased its biomass (BC = 0.232, 95% CI = 0.114–0.351), while the increase appeared to be reduced in the diploids (D3 BC = 0.104, 95% CI = 0.020–0.188; D4 BC = 0.109, 95% CI = 0.043–0.176). With respect to T2's responses in comparison with the mock-inoculated controls, it is also notable that the source of the nitrogen did not greatly affect biomass; both rhizobial treatment and nitrogen elicited similar increases. However, the results also indicate that much of the gain by T2 from exposure to NGR234 occurs in belowground tissue (Figure 4.4C); this result is a further illustration of the importance of the rhizobial treatment to belowground growth indicated by the significant interaction between species and rhizobial treatment noted above. Thus, from the fresh biomass responses, T2 appears to make greater gains, in comparison with its diploid progenitors, when exposed to any source of additional nitrogen.

DISCUSSION

There are few precedents for predicting whether an allopolyploid, in permanently combining two differentiated genomes, should have a broader or a narrower range of symbionts, relative to its diploid progenitors, with which it can form effective rhizobial symbioses. Compelling arguments can be made for either prediction.

If the range of symbiotic partners in diploid progenitors were determined by plant genes and alleles that limit bacterial infection and nodulation to specific bacterial strains, allopolyploids might be expected to have a reduced capability for symbiotic partnerships compared to either progenitor. Not surprisingly, given the likely recruitment of at least some nodulation functions from pre-existing defense responses (Oldroyd 2013), there are analogies with polyploid responses to pathogen resistance and defense. In modeling studies, Oswald and Nuismer (2007) found that autopolyploids had the potential for increased resistance when multiple alleles and multiple genes confer resistance. In empirical studies, the modulation of pathogen resistance responses is well recognized as an important key in determining nodulation outcomes in legumes (Gourion et al. 2015). In *G. max*, in particular, a set of host genes have been associated with restriction of interactions with specific rhizobial strains (Hayashi et al. 2012a; Hayashi et al. 2014; Yang et al. 2010). For example, Yang et al. (2010) identified *Rfg1* and *Rj2* as allelic variants of a gene coding for a Toll-interleukin receptor/nucleotide-binding site/leucine-rich repeat (TIR-NBS-LRR) resistance (R) proteins, each leading to restricted nodulation with strains of *B. japonicum* and *E. fredii*.

Our results, however, indicate that the allopolyploid T2 possesses an enhanced capability for rhizobial interaction, suggesting that factors other than those that could limit symbiotic capacity are likely operating to determine symbiotic ranges. A plausible expectation would be that the allopolyploid, by combining genes that facilitate symbiotic interaction from diploid progenitors,

would combine the capabilities for symbiotic interactions present in both progenitors, with the additional possibility of novel, transgressive interaction capabilities as well. T2 appears to be capable of root hair responses that are greater than those of either parent. This responsiveness of T2 may be due to variations in the types and quantities of signaling compounds in root exudates between T2, D3 and D4. In the most common type of legume rhizobial symbiosis, plants exude compounds, primarily flavonoids, that trigger Nod factor synthesis in the bacteria, whose subsequent reception by the plant leads to induction of root hair curling and deformation before bacterial infection and nodule organogenesis. The signaling between the rhizobial strains and legume species contributes to the specificity of the interactions (Oldroyd 2013; Oldroyd and Downie 2008). Hybridity and polyploidy, in general, can lead to the production and accumulation of compounds, including flavonoids, at greater or lower levels than in progenitors (Griesbach and Kamo 1996; Lavania et al. 2012). In addition, novel compounds synthesized in allopolyploids and not present in either progenitor have been historically well documented, with reconstruction and complementation of biosynthetic pathways of related genomes as a possible mechanism (e.g., Levy and Levin 1971, 1974; Dhawan and Lavania 1996). Our results suggest that T2 may exude greater quantities of Nod-factor-inducing compounds than D3 does and reduced levels of inhibiting compounds when compared with D4 (A. F. Powell and J. J. Doyle, unpublished manuscript). Different types or amounts of flavonoids in the exudates of T2 could elicit increased Nod factor production by rhizobia and result in a greater root hair response. Such phenomena have been observed, for example, in *Medicago sativa* and its symbiont *E. meliloti,* where the flavonoid, luteolin, binds with the NodD protein and induces Nod factor synthesis, whereas other noninducing flavonoids bind competitively to NodD (Peck et al. 2006). Quantitative differences are also important: varying the amount of Nod factor to which a plant

root is exposed subsequently causes differential early nodulation responses (Shaw and Long 2003). Alternatively, when considering T2 and its diploid progenitors, differences between the potential plant hosts in the density and type of symbiotic signaling receptors, including the Nod factor receptors NFR1 and NFR5 in *Glycine*, may also modulate the plant response (e.g., Radutoiu et al. 2007). Available evidence indicates that T2 plants express both homoeologues of key receptors and thus have the capacity to respond to Nod factor signals recognizable by both progenitor species (A. F. Powell and J. J. Doyle, unpublished manuscript). The ability to perceive a greater diversity of signals may thus be contributing to the enhanced symbiotic capacity observed in the allopolyploid.

Nodulation and colonization

Research with members of the *Glycine* subgenus *Glycine* allopolyploid complex aims at understanding anatomical, biochemical, and genetic differences between allopolyploids and their diploid progenitors (e.g., Coate et al. 2012). We are interested in these differences both intrinsically and because we seek to understand how several of these recently evolved allopolyploids have been able to colonize new habitats and ranges beyond Australia and Papua New Guinea while no subgenus *Glycine* diploids have done so in their several million years of existence. Several studies have been dedicated to understanding traits that could contribute to this apparent increased ability for colonization. T2 has higher photosynthetic rates on a per cell basis than do D3 or D4 (Coate et al. 2012), potentially due to higher expression of Calvin cycle enzymes (Ilut et al. 2012). Coate et al. (2013) also showed that T2 is better able to manage light stress, when compared with its diploid progenitors. Whether the superiority of T2 in any of these traits is related to its colonizing ability is unknown. Harbert et al. (2014) conducted an ecological niche modeling study on the Australian ranges of four subgenus *Glycine* allopolyploids and their

diploid progenitors; one goal of their study was to determine whether climate niches of allopolyploids in their native Australian ranges predicted greater colonizing ability. Harbert et al. (2014) found that all four allopolyploid species, including T2, were ecologically intermediate relative to their diploid progenitors, but otherwise found no consistent patterns. Although some of the allopolyploid species had larger ecological and geographical potential ranges than their diploid progenitors, others, including T2, did not. Thus, the Australian climate niche of T2 did not predict a greater colonizing ability outside of that continent (Harbert et al. 2014), suggesting that other traits are likely responsible.

Biotic interactions (e.g., pollination ecology, disease resistance, mutualisms such as mycorrhizal or nodulation symbioses) represent another class of relevant traits. The effects of polyploidy on biotic interactions are thought to play a role in the enhanced invasiveness of polyploid plants (Pandit et al. 2011; Pandit et al. 2014; te Beest et al. 2012). A focus on the potential impact of nodulation in this context is warranted, given that studies have shown that nodulation can affect the range expansion and occupation of novel habitats by legume species. For example, Stanton-Geddes and Anderson (2011) found that the range of *Chamaecrista fasciculata* was likely limited by the presence of compatible rhizobia. Subsequent experiments showed that reduced plant performance and nodulation was primarily due a reduced availability of compatible rhizobia at the limits of the range (Stanton-Geddes and Anderson 2011). Similarly, Parker et al. (2006) found that seedling growth, measured by plant mass, for an invasive legume species was greatest when transplanted seedlings were inoculated with compatible rhizobia when compared with uninoculated transplants, suggesting that a dearth of compatible symbionts affected plant performance. These examples demonstrate that availability of compatible rhizobial partners can affect the ability of a species to expand its range and to colonize new areas,

suggesting that the enhanced symbiotic capacity observed in T2 could contribute to the species' apparent increased ability for colonization relative to its diploid progenitors, D3 and D4.

To mimic colonization of a previously unoccupied region, we exposed T2, D3, and D4 to rhizobial strains that are novel partners for these species (i.e., these strains were not isolated from any of these species, nor were they collected from areas where accessions of *Glycine* subgenus *Glycine* are native). We studied symbiotic capacity, which is taken to encompass early signaling responses as well as nodule formation and growth, and found that T2 appears to show overall greater early signaling and nodulation responses with a broader range of rhizobial strains. Furthermore, with NGR234, T2 showed a greater percentage of nodulated plants relative to one diploid progenitor and greater nodule mass per nodulated plant compared with the second progenitor. It is also worth noting that inoculation results for the NGR234 strain represent a special case, where at least some plants from accessions of all species formed nodules, consistent with the unusually broad host range of this strain (Pueppke and Broughton 1999). NGR234's ability to be a broad-range symbiont has been attributed to its ability to produce an exceptionally diverse array of Nod factors, with a variety of structures and substituents (Price et al. 1992; Price et al. 1996; D'Haeze and Holsters 2002) and in greater amounts, for example 40 times as much as USDA257 (Pueppke and Broughton 1999; Relić et al. 1994). Nod factor structure, in turn, is believed to be at least partly responsible for determining the specificity of interactions between rhizobia and legume hosts (Oldroyd and Downie 2008). Thus, the use of NGR234 enables an assessment of T2's ability to form nodulating associations beyond its responses to a single more typical bacterial strain and emphasizes its increased symbiotic capacity relative to its diploid progenitors.

Glycine, nodulation, and polyploidy

Although there have been a few studies of nodulation in subgenus *Glycine*, none are directly comparable with our results in testing an allopolyploid species with its diploid progenitors. Pueppke (1988) tested individuals from three perennial *Glycine* species (none of those studied here) with several rhizobial strains; although ploidy was not reported, we now know that one of these was an allopolyploid, and the remaining two were diploids, but were not progenitors of the allopolyploid (A. F. Powell, unpublished data). Interestingly, the allopolyploid was able to nodulate with a variety of non-native strains of rhizobia, including USDA138, USDA257, and USDA191, while the diploids were not. This difference is consistent with our findings, and suggests that the ability of allopolyploid *Glycine* to nodulate with a broader range of rhizobia than diploids may not be confined to T2. Interestingly, when tested against strains isolated from members of subgenus *Glycine* in Australia, Pueppke (1988) observed no such difference in nodulation capability between the diploid and allopolyploid accessions tested. It would be interesting to test whether this observation, with its implications for colonization, would be mirrored in comparisons of allopolyploids and their diploid progenitors, as we have done here. In another study, Pueppke and Broughton (1999) showed that plants from the same three species could form nodules with NGR234, but did not report frequencies of nodule formation. Other than these studies, there have only been a few descriptive studies of nodulation of subgenus *Glycine* species by *Bradyrhizobium* spp., where *Glycine* plants were used as "trap plants" to identify resident bacteria (Brockwell et al. 1998); ploidy levels were not examined or reported in this study.

Outside of *Glycine*, the literature on nodulation ability of polyploids and their diploid progenitors is sparse and has generally involved synthetic autopolyploids. Among the few

studies of synthetic polyploids and nodulation published before the development of a molecular understanding of nodulation processes, one obtained results similar to those presented here. Leps et al. (1980) found that autotetraploid alfalfa (*Medicago sativa*) had higher rates of nitrogen fixation from rhizobia compared with the diploid in the first 10 d of growth, after which there was no difference (there was also no difference between the tetraploid and octoploid levels of nitrogen fixation). Thus, there was a greater early gain from nodulation in the tetraploid relative to the diploid, which was also reflected in a higher plant nitrogen content, similar to the polyploid advantage suggested in the current study. However, the *Medicago* study was not conducted on allopolyploids nor in a context relevant to colonization of novel habitats. In contrast to our findings, Weir (1961) observed that synthetic autopolyploids of *Trifolium pratense* tended to have fewer nodules than did corresponding diploids when tested against a variety of rhizobial strains. Working with autotetraploids of a peanut species (*Arachis villosa*), Stalker et al. (1994) found that there was no significant difference in nodule number per plant between autotetraploids and the progenitor diploid genotype, but did find that, in field and greenhouse experiments, two natural allopolyploid species, *A. monticola* and *A. hypogaea* (peanut, groundnut), tended to have more nodules per plant, nodule mass, and nitrogenase activity when compared with a broad variety of diploid species. Thus, while a number of earlier studies focused on synthetic autopolyploids, our results provide novel insights into the interactions between nodulation and allopolyploidy.

Polyploidy, plant growth, and nutrient responses

In studying nutrient responses and interactions with rhizobia, we have found differences between the allopolyploid T2 and its diploid progenitor that can plausibly contribute to T2's success as a colonizer. The qualitative breadth of symbiotic interactions observed here, with T2

combining and going beyond ("transgressing") the symbiotic capabilities of its diploid progenitors, along with its greater root hair responses to rhizobial exposure, suggest a greater capacity for interactions with a range of rhizobial partners. Yet rhizobia can vary substantially in the benefits that they confer on their hosts (Sachs et al. 2010; Barrett et al. 2015). For this reason, we also sought to examine parameters connected to plant fitness. Biomass is a component of plant fitness (Violle et al. 2007), and as such, biomass can serve as a measure of competitive ability (Weiher et al. 1999).

Assessing plant growth when exposed to NGR234 and to added nitrogen is helpful in understanding T2's colonization potential. The growth responses of T2 to the treatments were striking and set it apart from its diploid progenitors, while the growth response of the allopolyploid to added nitrogen was also informative. These growth responses indicate that T2 could maximize the contribution of available soil nitrogen or exposure to the rhizobial symbiont to a greater degree than D3 or D4, potentially aiding its establishment in a new habitat. The statistically significant interaction effect of species and rhizobial inoculation on belowground biomass is also of interest in this context since T2 showed a strong response to inoculation while D3 and D4 did not. This response is undoubtedly related to the higher proportion of nodulated plants in the T2 sample, coupled with the greater total nodule mass per plant, leading to a stronger effect of nitrogen fixation. Furthermore, for a stable mutualism to persist, both partners must gain in fitness and, indeed, in inoculation experiments with single strains, the fitness gains of both partners are positively correlated (Friesen 2012). For the low-nitrogen samples, T2 was the only species to show a total biomass gain when exposed to NGR234, which, coupled with the fact that the total nodule mass was also greatest for this symbiosis, is consistent with a positive relationship between the partners and suggests that the mutualism would be capable of persisting.

Previous work with T2 had suggested that T2 individuals are more robust and grow more rapidly and to greater sizes than do D3 and D4 (A. F. Powell, personal observation); here, we provide quantitative evidence for these observations. Similarly, the general significant effect of additional nitrogen on biomass was also to be expected, since comparisons of nitrogen addition and deprivation normally show that greater biomass is produced with greater amounts of available nitrogen (e.g., in *G. max*: Gan et al. 2003). However, the interactions observed between nitrogen and species are of particular interest, since it appears that T2 is able to maximize use of additional nitrogen to a degree that neither of its diploid progenitors can. Previous studies of nutrient relations in polyploids have also found increased ability to manage uptake of nutrients. For example, in *Arabidopsis thaliana*, autotetraploids showed greater potassium uptake and content in leaves than diploids, as well as increased salinity tolerance and greater fitness under saline conditions (Chao et al. 2013). Similarly, as noted above, increased salinity tolerance was also found in autopolyploid *Brassica rapa* (Meng et al. 2011), and polyploid *Solidago gigantea* was also found to have greater tolerance to calcium treatments (Schlaepfer et al. 2010).

The primary findings based on plant biomass under rhizobial and nitrogen treatments indicate that, while T2 is generally larger than its diploid progenitors irrespective of treatment, its growth also responds to rhizobial inoculation while D3 and D4 do not, when comparing plants with low added nitrogen. Furthermore, it responds to increased available nonsymbiotically fixed nitrogen to a greater degree than its D3 progenitor. The strong effect of inoculation on T2 belowground tissue growth was also notable.

Nodulation in the context of polyploidy and biotic interactions

Our study is consistent with a number of studies that have found that polyploidy can affect a variety of biotic interactions (e.g., Arvanitis et al. 2007; Arvanitis et al. 2010; Halverson et al.

2008). Additionally, Thompson et al. (2004) posed several questions for polyploidy researchers pertaining to plant and animal interactions, and our study also intersects with analogous questions. For example, Thompson et al. (2004) were interested in whether polyploidy could make pathogen host shifts more or less likely. Our work explores the possibility that rhizobial symbiont shifts to novel hosts may be more likely for an allopolyploid, relative to its diploid progenitors, thereby making a persistent symbiosis more likely in a context of colonization. Thompson et al. (2004) also noted that such shifts can often be explained by the distribution of chemical compounds; it has long been known that polyploidy may affect plant chemistry (e.g., Levin 1983). Our studies on interactions between members of *Glycine* subgenus *Glycine* and rhizobial symbionts have tested predictions of allopolyploids (Powell and Doyle 2015) and shown qualitative and quantitative changes in the synthesis and exudation of chemical constituents in polyploids relative to their diploid progenitors (A. F. Powell and J. J. Doyle, unpublished data).

Our study suggests that allopolyploids and diploids have different likelihoods of interacting with particular strains of rhizobia and thus could have the potential to recruit diverse, distinct rhizobia, though direct evidence of such recruitment will require additional field studies. Similar observations were previously made by Thompson and Merg (2008) in the context of plant–pollinator interactions, where diploids and autopolyploids showed differential visitation and pollination by various pollinators, and preferential visitation of some pollinator to diploids or tetraploids also varied by study site. Similar non-uniform effects have also been observed in relation to herbivory; autopolyploid *Heuchera grossulariifolia* had different responses to herbivores than did its diploid cytotype, being attacked less by one herbivore species but more by two others, when compared with the diploid (Nuismer and Thompson 2001). Such non-uniform

effects may be expected in nodulation interactions in *Glycine* subgenus *Glycine*, though our data, with several strains tested, tended to show a generally enhanced symbiotic capacity in the allopolyploid.

As in the current study, within-species variability in the ability to nodulate with particular strains of rhizobia has been observed previously for *G. max* and *G. soja*, where certain cultivars and accessions formed nodules with NGR234 and USDA257, whereas others did not (Pueppke and Broughton 1999). This variability between accessions within species has particular relevance in the case of rhizobial interactions, confirming the general expectation of variable genotype-by-genotype interactions and responses (e.g., Heath 2010; Heath et al. 2012). Our data show this type variability. For example, levels of nodulation with NGR234 varied among D4 accessions: for some accessions no plants produced nodules, whereas one, the genetically distinctive G2073 (Bombarely et al. 2014; S. Sherman-Broyles [Cornell University] and J. J. Doyle, unpublished data) had a high proportion of nodulating plants (Appendix 2). In the context of allopolyploidy, such variation in nodulation ability within diploid progenitors is important since the allopolyploid, even if it originates multiple times, may only sample a portion the genetic variation within a given diploid progenitor.

Additional questions remain to be addressed in terms of what other interacting factors exist that might alter the effects of polyploidy on nodulation symbioses. Studies of herbivory on polyploid plants have found that factors such as population size, habitat, and whether a herbivore is a generalist or specialist can alter the effect of polyploidy on the biotic interaction and determine the predictability of such effects (Hull-Sanders et al. 2009b; Münzbergová 2006). Analogously, in the context of nodulation and polyploidy, it is known that the outcomes of rhizobial symbioses can be changed by the level of nitrate or the activity of herbivores (Heath et

al. 2010; Heath and Lau 2011). Our studies here, involving addition of nitrogen, are a first step in understanding how biotic and abiotic factors can affect nodulation in the *Glycine* subgenus *Glycine* allopolyploid complex, and how this in turn might affect colonizing ability.

Conclusions

Taken together, the results of interactions with rhizobia demonstrated by T2, as compared with D3 and D4, support the hypothesis that T2 has a broader, enhanced capacity for engaging in rhizobial symbiosis. This more robust nodulation response could explain why T2, and not the diploid species that contributed their genomes to its formation, has populations outside of Australia and Papua New Guinea. Our current work is directed toward understanding the mechanisms by which these symbiotic capabilities occur, including assays of the types and quantities of flavonoid signaling compounds exuded by diploid and allopolyploid *Glycine*, variations in the Nod factor receptors that each species possesses, and the molecular and physiological bases of early responses to rhizobial signaling that lead to nodulation (A. F. Powell, unpublished data). The perennial *Glycine* complex provides several independent examples of colonization by allopolyploid species (Sherman-Broyles et al. 2014) and thus allows tests of the generality of our hypotheses. By elucidating the interactive effects between polyploidy, rhizobial symbiosis and nutrient relations, as well as the mechanisms underlying these interactions, we will gain a more complete understanding of how biotic interactions can contribute to enhanced colonization ability in polyploid species.

CHAPTER 5

NON-ADDITIVE TRANSCRIPTOMIC RESPONSES TO INOCULATION WITH RHIZOBIA IN ALLOPOLYPLOID *GLYCINE DOLICHOCARPA*

INTRODUCTION

Nodule-forming interactions between soil bacteria and plant hosts, referred to as root nodule symbiosis, are a prominent feature of legume biology, with the capacity for such interactions being widespread across the family (Leguminosae). The nitrogen-fixing nodules formed in such symbioses are a key source of nitrogen for many legumes and can account for an important addition of nitrogen to specific biological systems (Herridge 2008). Nodulation symbioses are diverse among legumes, including various determinate and indeterminate growth types, ureide and amide exporting mechanisms, as well as alternative mechanisms such as the stem nodulation found in *Sesbania rostrata* (Sprent 2009). The bacteria (termed 'rhizobia') that interact with legumes in such symbioses are also diverse; while many are in the *Alphaproteobacteria* genera of *Rhizobium*, *Sinorhizobium*, *Bradyrhizobium*, *Betaproteobacteria*, in genera such as *Burkholderia* and *Cupriavidus*, are also known to form nodulation symbioses. Many of the signaling components and molecular mechanisms required for the establishment of rhizobial symbioses have been elucidated (e.g., Oldroyd 2013). The signaling between partners involves Nod factors, lipochitooligosaccharides synthesized by rhizobia in response to Nod factor-inducing compounds (often flavonoids) in a host plant's root, which are then perceived via plant receptors; variation in these several signaling components is believed to contribute to the diversity and specificity of interactions (Radutoiu et al. 2007).

With recent genomic studies, whole genome duplication (WGD, polyploidy) has been implicated as a phenomenon that may have had a role in the evolution and refinement of nodulation in legumes (Young et al. 2011; Li et al. 2013). With the publication of the *Medicago truncatula* genome, evidence of a WGD shared by many papilionoid legumes was coupled with evidence of duplicate gene retention and expression patterns to suggest that key genes involved

in nodulation signaling evolved from genes duplicated by this WGD. The authors stated that "the WGD early in papilionoid evolution allowed the emergence of critical components in Nod factor signaling and contributed to the complexity of rhizobial nodulation observed in this clade"(Young et al. 2011). Polyploidy can certainly have varied effects, including those at genetic, biochemical, organismal and ecological levels. In particular, changes at the level of gene expression can impact phenotypes and, ultimately, responses to abiotic and biotic factors. Patterns of nonadditive expression in allopolyploids have been observed in numerous cases, such as for example in *Spartina anglica* (Chelaifa et al. 2010) and *Senecio cambrensis* (Hegarty et al. 2012), suggesting that substantial expression-level variability can occur. Expression of multiple parental homoeologues can also be a source of variation, as expression of differing copies of enzymes, receptors or transcription factors, for example, and the degree to which homoeologous copies are expressed can lead to alterations in biological function. Thus, parental legacies and homoeologue expression biases can lead to numerous potential expression-level outcomes that, in turn, alter plant traits (e.g., Yoo et al. 2013; Yoo et al. 2014; Buggs et al. 2014; Hegarty et al. 2013).

Next-generation sequencing approaches have also provided insights directly into the transcriptome-level responses of legumes to inoculation with rhizobia. Libault et al. (2010), for example, were able to isolate inoculated root hairs and compare responses over time, while Hayashi et al. (2012b) examined early responses to inoculation in the zone of nodulation and de Carvalho et al. (2013) examined whole soybean roots at later time points. Recently, laser capture microdissection has also been coupled with transcriptomic analysis to examine expression in specific tissue regions (e.g., Jardinaud et al. 2016; Roux et al. 2014). However, while transcriptomic studies have been used effectively to study nodulation and polyploidy as distinct

phenomena, there have been few studies examining both in the same system. In the present study, we examine the transcriptional responses of a recent allopolyploid to rhizobial inoculation and compare them to those of its diploid progenitors. Thus, we are inquiring into the effects of recent allopolyploidy on nodulation responses. We also examine the relevance of homoeologue usage and parental legacies to expression-level responses to nodulation.

The *Glycine* subgenus *Glycine* allopolyploid complex presents an opportunity to study the intersection of nodulation and the effects of recent allopolyploidy. The complex consists of approximately 20 species, including several diploid species that have formed allopolyploids in various combinations within the last several hundred thousand years (Bombarely et al. 2014; Sherman-Broyles et al. 2014). The capacity for nodulation in a number of the species in this complex has been documented (Pueppke 1988; Pueppke and Broughton 1999; Brockwell et al. 1998), and recent work has examined rhizobial interaction in relation to allopolyploidy in the system (Powell and Doyle 2016). Transcriptomic approaches have been used in the complex to examine responses to light stress, as well as effects of polyploidy on photosynthesis and transcriptome size (Coate et al. 2013; Coate and Doyle 2010; Ilut et al. 2012).

Here, we used RNA-Seq to examine transcriptional responses to inoculation in the allopolyploid *G. dolichocarpa* (referred to throughout as T2), and its diploid progenitors *G. tomentella* (D3) and *G. syndetika* (D4). We are interested in determining the transcriptional patterns of the allopolyploid when compared to the progenitors, both transcriptome-wide and when considering specific sets of genes. Previous work found that T2 appears to be broadly more responsive in early symbiotic signaling when inoculated with rhizobia than D3 and D4, has a greater propensity to form more nodules than either diploid, and tends to form a greater nodule mass than the D3 progenitor when considering nodulated plants (Powell and Doyle 2016). Thus,

we expect that a transcriptomic analysis will provide insights into distinctive mechanisms that contribute to these responses to rhizobia in T2. Furthermore, in considering the regulation of expression levels in T2, we are interested in patterns of parental dominance, additivity and transgressive variation and how these relate to the response to rhizobial inoculation. Lastly, transcriptomic data can be used to assess relative homoeologue expression within the allopolyploid and, thus, examine how the D3 and D4 homoeologues of key nodulation-related genes are deployed in response to inoculation. Through such analyses, we are able to identify transcriptional responses and processes implicated in interactions with rhizobia that set the allopolyploid apart from one or both of its diploid progenitors.

MATERIALS AND METHODS

Plant material

Three accessions for each of the three species were used to generate transcriptomes: G1364, G1403, and G1820 for *G. tomentella* (D3); G1300, G2073, and G2321 for *G. syndetika* (D4); G1134, G1188, and G1393 for *G. dolichocarpa* (T2). Seeds were surface sterilized, nicked and placed in petri plates for germination under sterile conditions according to the protocol outlined in Powell and Doyle (2016). Seedlings were germinated for seven days before being transferred, in a laminar flow hood, to plates of Jensen's nitrogen-free media (Somasegaran and Hoben 1994). Following the transfer to media, the seedlings were either mock-inoculated with sterilized water (also referred to as 'control' samples) or inoculated with a liquid culture of the rhizobial strain NGR234. Lyophilized NGR234 was obtained from the USDA-ARS Rhizobium Germplasm Resource Collection and revived according to provided protocols. The liquid culture was grown at 30°C for five days in AG media (Somasegaran and Hoben 1994). Prior to

inoculation, the bacterial culture was pelleted and washed three times, before being resuspended in sterilized water. The culture used to inoculate the seedlings had a concentration of approximately 10^7 bacterial cells per mL as determined with a Helber counting chamber (Somasegaran and Hoben 1994). Each seedling was inoculated with 1 ml of bacterial culture. Seedlings in media plates were then placed in a growth chamber with a day:night cycle of 16h:8h and temperatures of 22°C during the day and 18.5°C at night. Relative humidity was set at 60% and the light was 150 µmol/m^2/s. After seven days, plant root tissue was harvested, frozen in liquid nitrogen and stored at -80°C.

RNA sequencing

Total RNA was extracted from pooled root tissue from multiple individuals per accession and RNA-seq libraries were constructed for each accession using the pooled tissue. For each of the three accessions per species, two libraries were constructed: one from the control, mock-inoculated tissue and another from the tissue inoculated with NGR234. In this design, three samples were obtained under each treatment, for each species. Total RNA was extracted using the AllPrep DNA/RNA/miRNA Universal Kit (QIAGEN, Hilden, Germany) with on-column DNase treatment (QIAGEN, Valencia, CA, USA). Single-end libraries were constructed and multiplexed using the TruSeq Stranded mRNA LT Sample Prep Kit (Illumina, San Diego, CA, USA).

The 18 libraries in this experiment were 6-plex sequenced. Sequencing was conducted using the Illumina HiSeq 2500 platform at the Cornell University Biotechnology Resource Center's sequencing facility (http://www.biotech.cornell.edu/brc/genomics-facility).

RNA-seq data processing

Fastq-mcf (Aronesty 2013) was used to process reads; sequence reads were trimmed (with a minimum quality cutoff of 30) and short reads were removed (with a minimum read length of 50). Cleaned reads were mapped to the G. max genome (2.75_Wm82.a2.v1) using TopHat2 (Kim et al. 2013a). Uniquely mapped reads were obtained using the command samtools view -h input.bam | grep -w "@SQ\|@PG\|NH:i:1" and the htseq-count command (Anders et al. 2014) was used to count reads mapped to each exon.

Homoeologue mapping

Homoeologues were identified and mapped using the approach developed in Bombarely et al. (2014). In order to separate reads in T2 samples based on diploid progenitor origin, we generated reference transcriptomes for the D3 and D4 progenitors using the RNA-Seq root samples described above, along with additional sequence data from paired-end DNA-Seq samples that were also generated from root tissue, using Samtools and Mpileup (Li et al. 2009), Gffread (Trapnell et al. 2012), and Bowtie2 (Langmead and Salzberg 2012) to generate a reference transcriptome. Reads from T2 samples were then mapped to a concatenated reference comprising both the D3 and D4 reference transcriptomes using HISAT2 (Kim et al. 2015). Reads were then separated based on preferential mapping to each of the diploid reference transcriptomes using the SeparateHomeolog2Sam (https://github.com/aubombarely/GenoToolBox/blob/master/SeqTools/SeparateHomeolog2Sam). The separated homoeologue-specific reads were then mapped back to the reference soybean genome and reads were counted using htseq-count (Anders et al. 2014). For estimates of homoeologue usage percentages for the overall transcriptome, genes with zero counts for both homoeologous copies were excluded from the

calculation for each accession; of the 56044 genes in the transcriptome, following homoeologue separation and mapping, an average of 37599 were used for mock-inoculated samples and an average of 36313 were used for inoculated samples.

Expression-level analyses

The R package DESeq2 (Love et al. 2014) was used to identify differentially expressed genes (DEGs) between control and inoculated samples within species, using a false discovery rate (FDR) of 0.05 to control for multiple comparisons. The approach to analysis of expression-level dominance was adapted from Yoo et al. (2013). Expression levels were analyzed separately in mock-inoculated samples and inoculated samples using the edgeR package (Robinson et al. 2010), comparing across species. Accession samples for each species were first normalized, dispersions were estimated and differential expression was assessed between the allopolyploid T2 and each diploid progenitor using exactTest() (Anders et al. 2013; Yoo et al. 2013). Analysis was restricted to genes that had at least 1 read per million in at least three samples in each comparison (Anders et al. 2013). Differential expression between species was determined with a false discovery rate (FDR) < 0.05 using the BH adjustment (Benjamini and Hochberg 1995). Genes with differential expression in the allopolyploid were then placed into one of twelve possible categories of differential expression patterns as defined by Rapp et al. (2009) and Yoo et al. (2013).

For the analysis of known nodulation-related genes, a list of such genes was obtained from the Schmutz et al. (2010) soybean genome paper. The corresponding gene IDs in the 2.75_Wm82.a2.v1 soybean genome were obtained and this resulted in a set of 47 genes used for ordination and fold-change analysis, including genes encoding symbiotic signaling receptors, transcription factors and nodulins. For the expression-level analysis based on the approach of

Yoo et al. (2013) described above, several genes did not meet the cutoff for inclusion in the analysis; as such, 32 genes were used in comparisons between control samples and 33 genes were used in comparisons between inoculated samples.

Weighted gene co-expression network analysis

Expression networks were constructed, and coexpression modules were identified, using the R package WGCNA (Langfelder and Horvath 2008, 2012) and count data normalized using the rlog() function of the DESeq2 package (Love et al. 2014). Applying the approximate scale-free topology criterion (Langfelder and Horvath 2008; Zhang and Horvath 2005), we selected a threshold power of $\beta = 5$. A maximum block size of 20000 was specified in the blockwisemodules() function. We conducted the signed network construction using all inoculated and mock-inoculated samples, and module assignment was achieved using the Dynamic Tree Cut function. We examined module eigengene expression in the inoculated samples in relation to trait data obtained from Powell and Doyle (2016), including root hair deformation ratio, total nodule mass per plant, total number of nodules per plant, and percent of plants with nodules per accession, and identified hub genes based on correlations between gene significance for the root hair trait and module membership.

RESULTS AND DISCUSSION

T2 exhibits overall expression-level intermediacy, while hydrogen peroxide and oxidative stress responses are differentially regulated in D3

Among the transcriptomes sampled in the present study, the T2 transcriptomes tend towards overall intermediacy in expression levels (Figure 5.1). Principal component analysis of the transcriptional profiles in DESeq2 shows the samples mainly grouping by species. The first

two principal components account for approximately 74% of the variation between samples. The T2 samples appear as intermediates, separated from the D3 and D4 samples along the second principal component. Similarly, the two dimensions of the multidimensional scaling plot generated in edgeR, based on MDS analysis of the 500 most heterogeneously expressed genes differentiating the samples, reveal similar intermediacy of T2 samples to D3 and D4 (Figure 5.2). This indicates that the greatest variability in expression between samples occurs between species. If a set of known nodulation-related genes, compiled by Schmutz et al. (2010), is considered, T2 samples appear again to be intermediate to the diploids in an ordination plot using variance-stabilized expression values (Figure 5.3). From this analysis of nodulation-related expression, it appears that T2 samples show a broadly similar, intermediate pattern as in the full transcriptome.

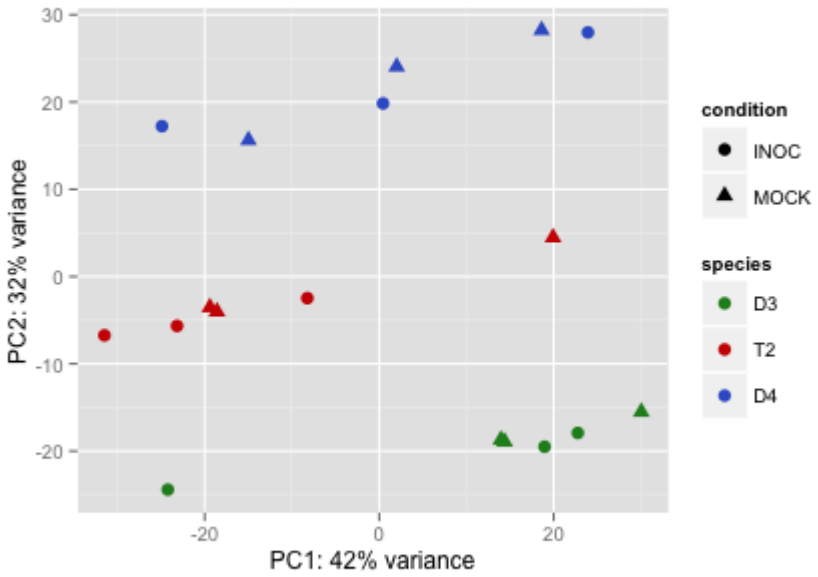


Figure 5.1. PCA of all D3, D4, and T2 samples, conducted using rlog transformed expression values for the top 500 heterogeneously expressed genes across all samples.

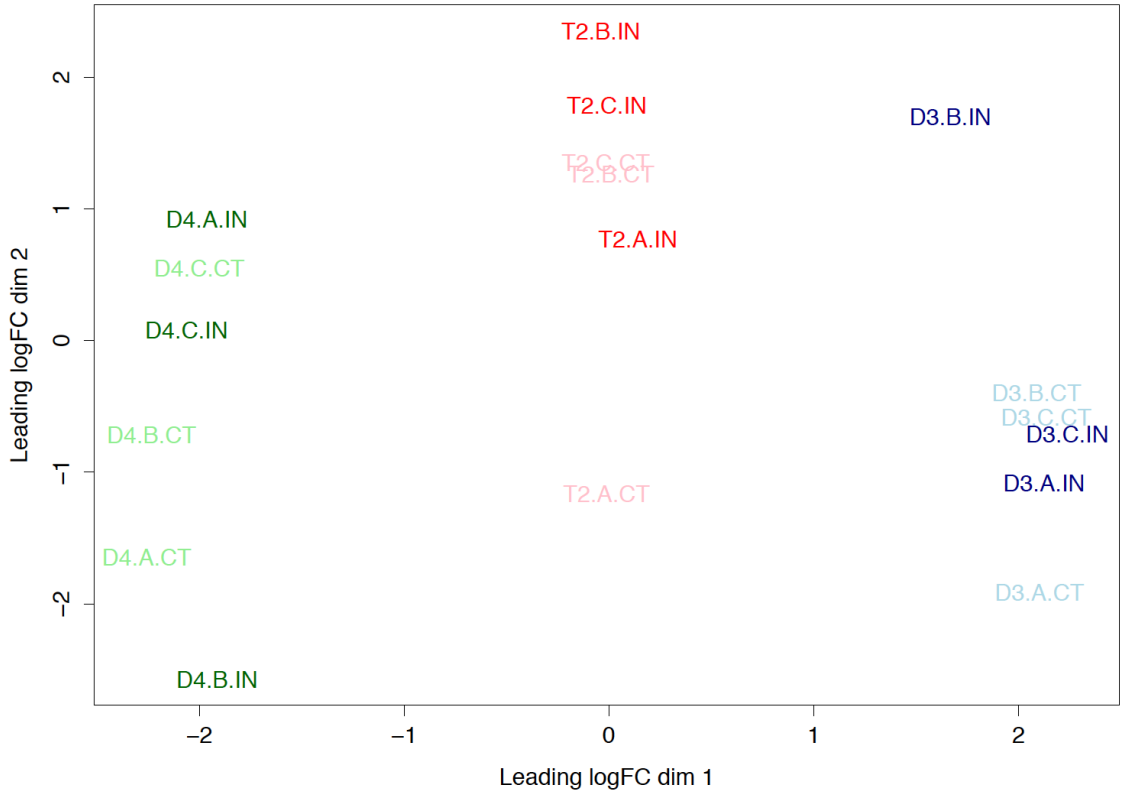


Figure 5.2. Multidimensional scaling (MDS) plot showing distances between samples, based upon log2 fold changes using the top 500 heterogeneously expressed genes. T2 samples are shown in red, D3 samples are shown in blue, and D4 samples are shown in green. Samples with the suffix '.IN' were inoculated, while samples with the suffix '.CT' were controls. FC = fold change.

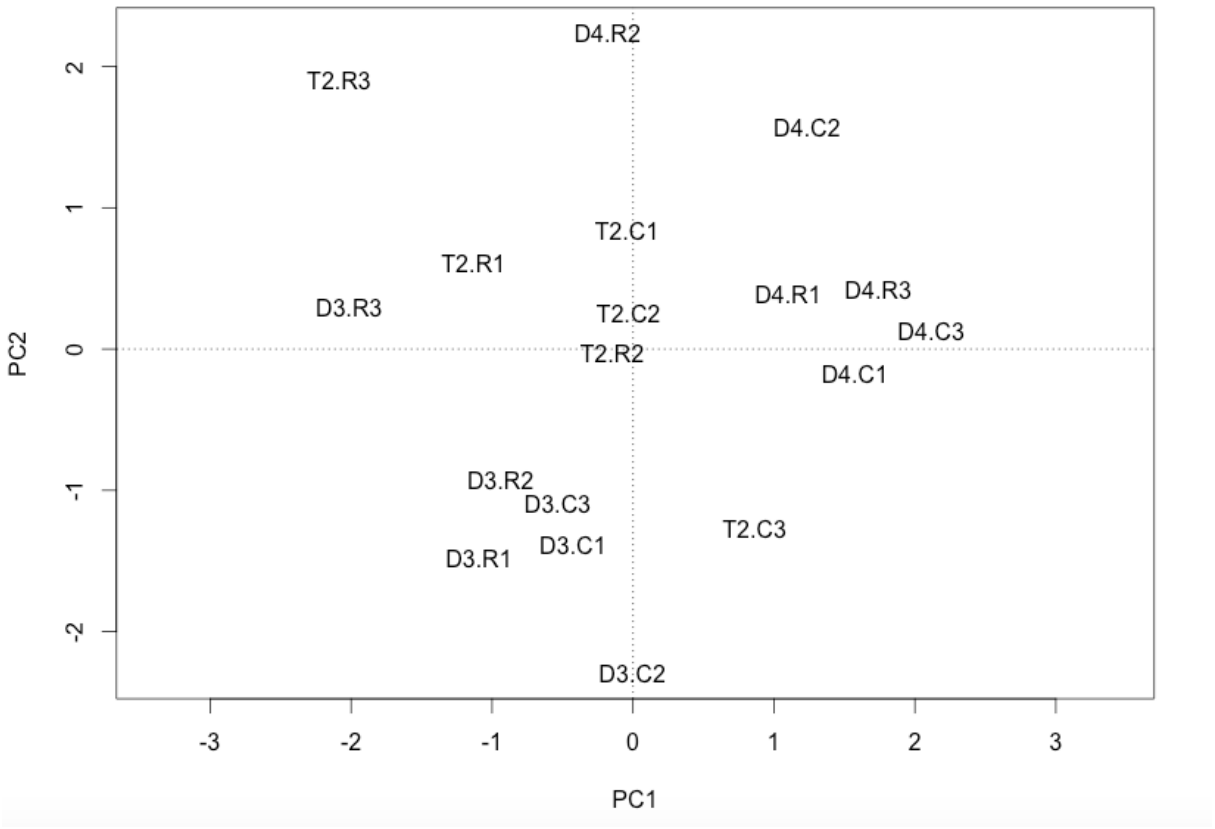


Figure 5.3. PCA of all D3, D4, and T2 samples, generated using variance-stabilized expression values for a set of pre-defined, known nodulation genes. Samples are designated by the appropriate species prefix (D3, T2, D4); samples C1-C3 are control samples in each species and samples R1-R3 are the corresponding samples inoculated with rhizobia.

In addition to this broad expression-level intermediacy, T2 shows a lesser overall transcriptional response to inoculation compared to its diploid progenitors. This differential response to inoculation in the three species is observed in the number of genes that were significantly differentially regulated following exposure to rhizobia. In total (i.e., including both upregulated and downregulated genes), T2 had 81 significantly differentially expressed genes (DEGs), while D4 and D3 had 277 and 1248 significantly DEGs, respectively (Figure 5.4). Considering differentially upregulated and downregulated genes by species, T2 showed a greater

exclusive overlap with D3, as opposed to D4 (Figure 5.5). Of the upregulated transcripts in T2, 56% overlapped exclusively with D3 upregulated transcripts, while 13% overlapped exclusively with D4 upregulated transcripts. None of the 11 downregulated transcripts in T2 were exclusively shared with D4, while two were shared exclusively with D3 and seven were shared with both diploids. The greater responses in D3 and D4 compared to T2 can also be illustrated in 'MA' plots, which show additional information on the magnitude of fold-changes of gene expression for individual genes, by plotting log ratios (M) of expression against mean average (A) expression for each gene (Figure 5.6).

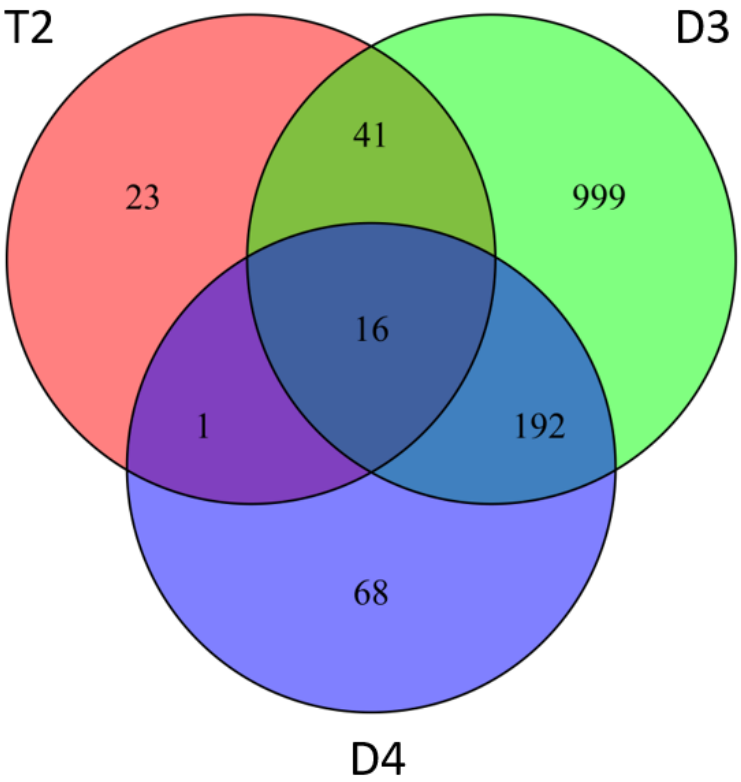


Figure 5.4. Venn diagram showing overlap between D3, T2, and D4 for all genes differentially expressed in response to inoculation in each species.

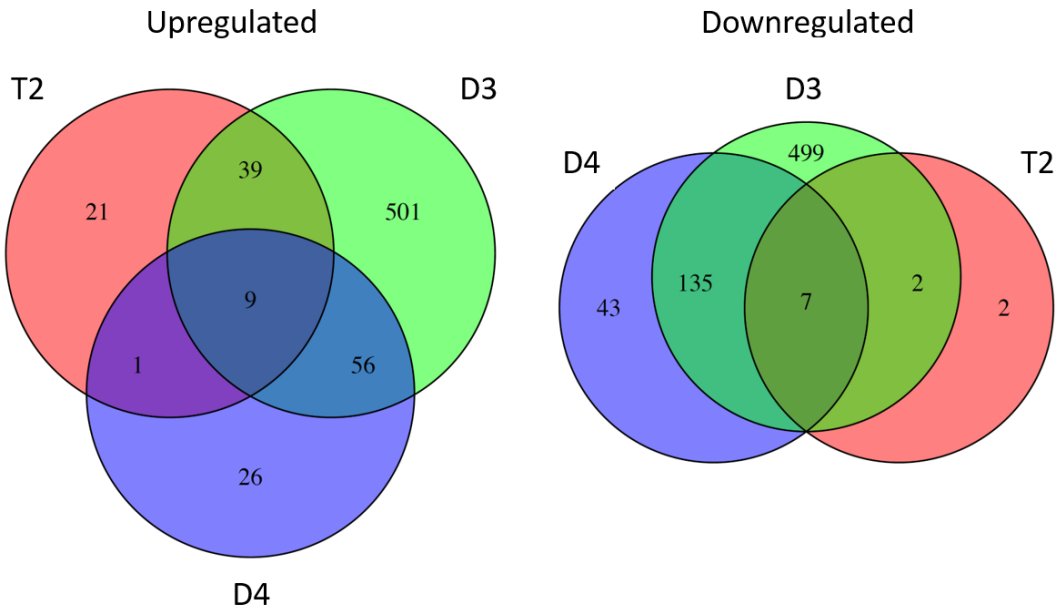


Figure 5.5. Venn diagrams showing overlap between D3, T2, and D4 for upregulated and downregulated genes in response to inoculation in each species.

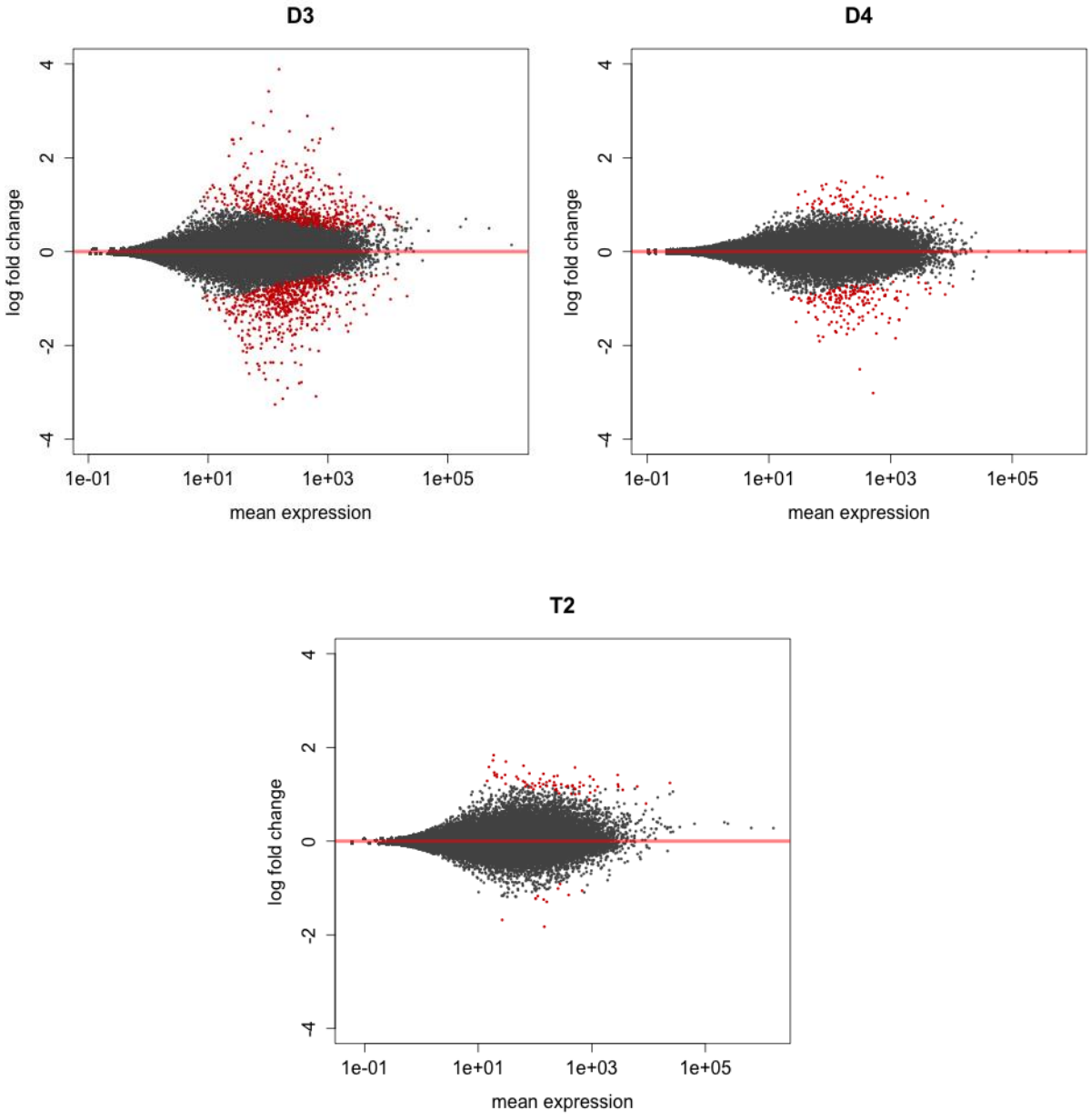


Figure 5.6. Pairwise comparisons of differential gene expression, showing average log fold change in expression between inoculated and control samples relative to mean expression, for species D3, D4 and T2. Red dots are genes that were significantly differentially expressed (FDR < 0.05), while gray dots indicate genes that were not significantly differentially expressed.

To further explore the nature of the enhanced transcriptional response in D3, we examined the GO categories of genes that were significantly overrepresented among upregulated and downregulated genes. Not surprisingly given the disparity in DEGs, in both classes, D3 had more GO categories with overrepresented genes than did either T2 or D4. In the sets of T2 upregulated and downregulated DEGs, no categories of molecular function, biological process or cellular enrichment were significantly overrepresented. Of note, however, D3 had several significantly overrepresented categories related to oxidative stress among upregulated genes (Figure 5.7). These included the biological process category related to hydrogen peroxide catabolism. These GO categories are suggestive of an enhanced defense response in D3 at the level of the transcriptome; this was also explored in terms of fold changes for individual genes. Libault et al. (2010), in their study of the *G. max* root hair transcriptomes inoculated with rhizobia, identified a set of defense-related, nodulation-responsive genes. This set was used to query the list of genes regulated only by inoculation treatment, across all samples in D3, T2 and D4. Two genes encoding peroxidases were obtained, Glyma.02G259300 and Glyma.14G053600. The fold-change patterns observed for these peroxidase genes also indicate that D3 upregulated these genes to the greatest degree, D4 was the progenitor upregulating them the least, and T2 was regulating them in a manner similar to D4, rather than D3 (Figure 5.8). Among the genes identified by Libault et al. (2010), a third defense-related gene (Glyma.17G133400) was also found to be significantly upregulated in D3, while T2 appears to show intermediate regulation compared to the two diploids; this third gene is identified as a proprotein convertase subtilisin/kexin (https://phytozome.jgi.doe.gov). One factor contributing to T2's lesser overall transcriptional response, then, is that it has an apparently reduced upregulation of defense

response transcripts when compared to D3, though it does not provide insight into T2's decreased response compared to D4.

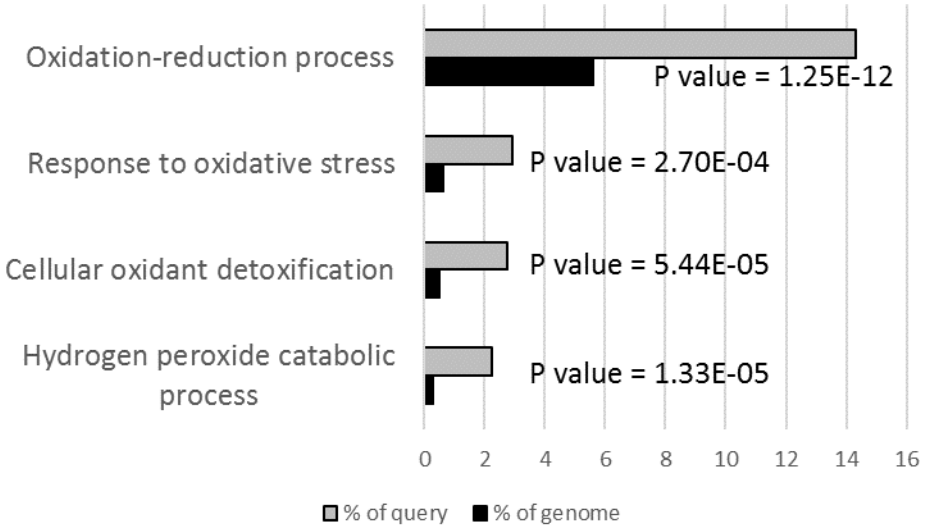


Figure 5.7. The four most highly significant GO biological process parent categories resulting from a GO overrepresentation analysis of the D3 genes upregulated by inoculation treatment. The figure illustrates this overrepresentation by showing the genes in each category as a percentage of the genome and as a percentage of the set of D3 upregulated genes.

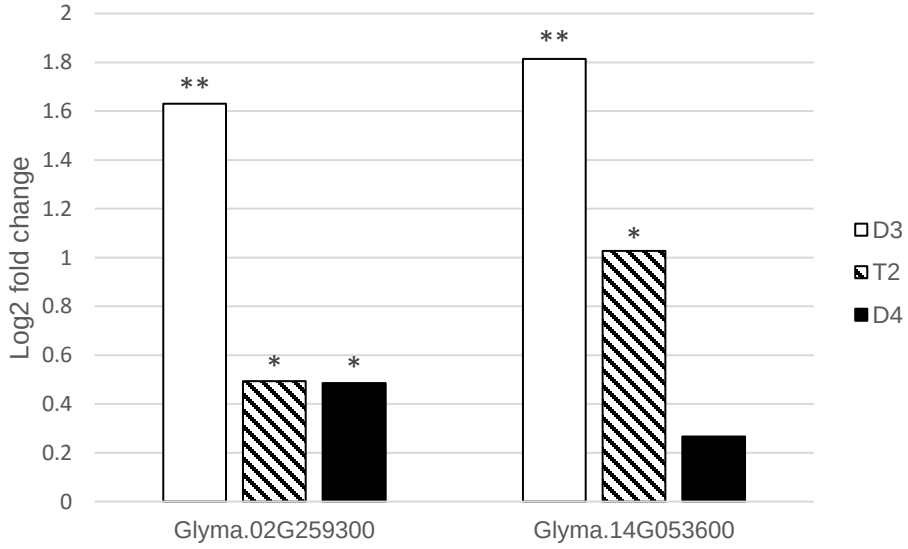


Figure 5.8. D3 shows greater upregulation of two peroxidase genes in response to inoculation than T2 or D4. Single asterisks indicate expression differences between inoculated and control

samples with an unadjusted p-value < 0.05. Double asterisks indicate expression differences with FDR < 0.05 (Benjamini-Hochberg adjustment).

The production of reactive oxygen species (ROS), including hydrogen peroxide, contributes to the hypersensitive response and is a well-characterized aspect of responses to bacterial pathogens in plants, including *G. max* (Levine et al. 1994; Petrov and Van Breusegem 2012). The result obtained here is also consistent with 3,3′-diaminobenzidine (DAB) staining assays, where D3 shows greater hydrogen peroxide evolution when exposed to a strain of *S. fredii* rhizobia than do T2 or D4 (Powell, unpublished data). Indeed, legumes show a variety of defense responses following inoculation with rhizobia, though these tend to be rapidly suppressed in the case of effective symbionts. At the transcriptional level, expression of defense-related genes is upregulated in *G. max*, *L. japonicus*, and *M. truncatula* in the early stages of response to inoculation (Libault et al. 2010; Lohar et al. 2007; Kouchi et al. 2004). The evolution of ROS (hydrogen peroxide) has also been observed upon inoculation. The rate of production of ROS decreases following application of compatible Nod factors (Shaw and Long 2003), and inoculation with mutant rhizobia that are defective in Nod factor synthesis leads to increased production of ROS in *M. sativa* (Bueno et al. 2001). However, the initial burst of hydrogen peroxide is also potentially necessary for triggering certain nodulation-related processes and controlling gene expression that is conducive to nodulation (Jamet et al. 2007; Andrio et al. 2013). This initial burst, however, appears to occur at an earlier time point than that which is sampled here; in the sampling conducted here, one would expect defense responses to have been attenuated. Further, there is some evidence that alterations to defense-related mechanisms and increased responses can also lead reductions in the size of individual nodules and total nodule mass per plant (e.g., Marie et al. 2003; Dai et al. 2008). The apparent transcriptional upregulation

of defense-related genes in D3 may therefore, in part, account for the reduced nodule mass per plant previously observed in D3 relative to T2 (Powell and Doyle 2016). D3, by maintaining a stronger defense response, could be inhibiting progression of infection and symbiosis with certain rhizobia.

From the overall transcriptional responses, then, we can observe that while the allopolyploid T2 samples are generally transcriptionally intermediate in relation to diploids D3 and D4, T2 has a reduced transcriptional response to inoculation with rhizobia when compared to either of its diploid progenitors. A similar reduced response was observed in T2 in the context of light stress (Coate et al. 2013), where the allopolyploid was not stressed to the same degree as the diploids when confronted with elevated light levels and had reduced transcriptional responses for stress-related genes (Coate et al. 2013). In the present study, the overrepresented GO categories and regulation of specific genes suggest that T2 has a reduced transcriptional defense response compared to the D3 diploid progenitor.

Implications of expression-level dominance and homoeologue usage for nodulation

In addition to overall transcriptome-level responses, we examined regulation of known nodulation-related genes in response to inoculation. Nine of the 47 nodulation genes from Schmutz et al. (2010) were among those regulated by inoculation treatment alone across all samples (FDR < 0.05). The within-species fold-changes in expression for these genes are shown in Figure 5.9. These genes revealed a variety of fold-change patterns in specific transcript accumulation between control and inoculated samples for each species. For a number of these genes, T2 showed intermediate fold-change patterns between D3 and D4, with D3 showing the greatest increase. It is particularly noteworthy that *NIN* showed this pattern of regulation, since it is a key transcriptional activator in downstream nodulation responses (Schauser et al. 1999). If

T2 is able to upregulate *NIN* to a greater degree than one of its progenitors, this suggests an increase in the degree of symbiotic response over that progenitor.

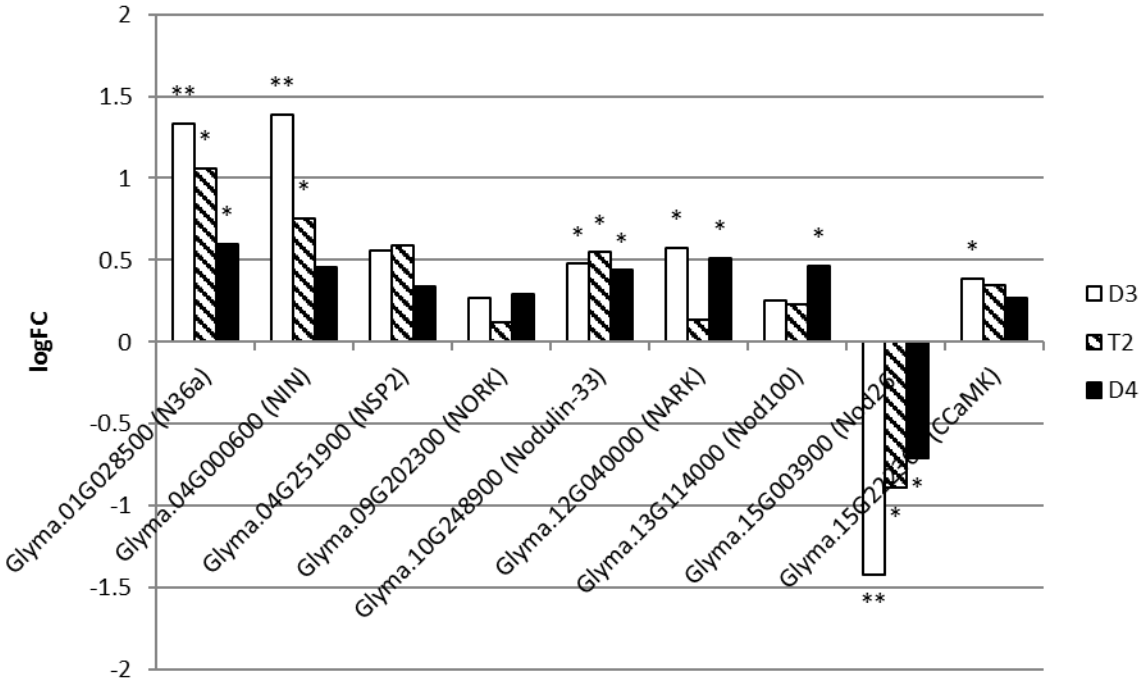


Figure 5.9. Log fold change (logFC) in expression of inoculation-responsive, known nodulation genes in inoculated relative to control samples within each species. Single asterisks indicate differences in expression between inoculated and control samples with an unadjusted p-value < 0.05. Double asterisks indicate differences in expression with a FDR < 0.05 (Benjamini–Hochberg adjustment).

As noted in the introduction, allopolyploidy raises several major questions concerning transcriptional responses and regulation of gene expression. One question relates to expression levels in the allopolyploids compared to those of the diploid progenitors, asking whether the sum of the two homoeologous copies in the polyploid is similar to that of one diploid progenitor or the other (expression-level dominance; (Grover et al. 2012; Yoo et al. 2013), versus being expressed at mid-parent or transgressive levels. Furthermore, if there is novel expression of genes or silenced expression relative to the diploid progenitors, this could have the capacity of

altering biotic interactions. Here, the meaning of 'silencing' and 'novel expression' were adopted from Yoo et al. (2013) and were evaluated in a similar manner, where silencing meant an absence of expression in the allopolyploid of genes expressed in one or both diploids and novel expression meant expression of genes in the allopolyploid that were not expressed in either diploid. There are also distinct questions about the relative expression of homoeologues from each of the diploid progenitors in the allopolyploid (referred to here as homoeologue usage). We thus examined each of these areas in relation to nodulation responses.

In the *a priori* set of nodulation genes, there was no evidence for cases of silencing or novel expression in the allopolyploid. Studies using synthetic and natural hybrids and allopolyploids have yielded evidence that there are effects on polyploids of both genome merger and genome doubling, as well as the passage of evolutionary time. While expression levels in an allopolyploid can be a result of the parental legacy of expression patterns in progenitors or can relate to truly novel changes due to allopolyploidization (Buggs et al. 2014), there is evidence that genome merger itself alters regulatory interactions and leads to substantial changes in expression levels, which are further enhanced over evolutionary time (Yoo et al. 2014). As the time since genome merger increases, expression-level dominance, as well as the number of transgressively expressed genes, also increases; this has been shown in the case of cotton (Yoo et al. 2013). Current estimates suggest that *Glycine* subgenus *Glycine* allopolyploids, including T2, were formed within the last several hundred thousand years (Bombarely et al. 2014; Sherman-Broyles et al. 2014). The absence of silencing observed in the nodulation genes studied here is understandable given the relatively recent formation of T2, and a similar general absence of silencing has previously been observed in T2 related to expression of genes involved in photosynthesis and photoprotection (Coate et al. 2013; Ilut et al. 2012).

Many of the genes did not show a difference between the species, but for the genes that showed differential expression between species in mock-inoculated samples of T2, 10 genes were identified and classified by their expression patterns, and in the samples inoculated with rhizobia, eight genes were identified and classified (Table 5.1; Table 5.2). Under both treatment conditions, for genes showing patterns of expression-level dominance, the allopolyploid tended to have more genes expressed at the level of the diploid progenitor with higher levels of expression (i.e., comparing the sum of categories II and IV with the sum of XI and IX yields 4 and 2, respectively, for control samples, and 2 and 1, respectively, for inoculated samples).

Table 5.1. Expression patterns for nodulation-related genes from Schmutz et al., (2010). The twelve categories represent possible patterns of differential expression in the allopolyploid relative to its diploid progenitors, and have been adapted from (Rapp et al. 2009). For pairwise comparisons, control ('CTRL') and inoculated ('INOC') samples were analyzed separately.

	Additivity		D4 expression-level dominance		D3 expression-level dominance		Transgressive downregulation			Transgressive upregulation			No Change
	I	XII	II	XI	IV	IX	III	VII	X	V	VI	VIII	
	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	D3 T2 D4	
CTRL	1	-	1	1	3	1	1	-	-	-	1	1	22
INOC	-	1	1	-	1	1	1	-	-	1	1	1	25

Table 5.2. Identity and description of genes placed into the twelve categories of expression patterns shown in Table 5.1.

Gene ID	Treatment	Expression Pattern	Category	Description
Glyma.07G208900	Control	Additivity	I	SAN1B - 2OG-Fe(II) oxygenase
Glyma.13G114000	Control	D3 expression-level dominance	IV	Nod100 - sucrose synthase; Glycosyl transferases
Glyma.02G270800	Control	D3 expression-level dominance	IV	NFR1
Glyma.11G063100	Control	D3 expression-level dominance	IV	NFR5
Glyma.07G209100	Control	D3 expression-level dominance	IX	SAN1A - 2OG-Fe(II) oxygenase
Glyma.12G050100	Control	D4 expression-level dominance	II	NIN2 - RWP-RK/PB1; putative TF
Glyma.06G065600	Control	D4 expression-level dominance	XI	Nodulin - SPFH/Band 7 family; integral membrane protein
Glyma.15G003900	Control	Transgressive downregulation	III	Nod26 - Major intrinsic protein
Glyma.10G248900	Control	Transgressive upregulation	VI	Nodulin-33 - haloacid dehalogenase-like hydrolase
Glyma.12G040000	Control	Transgressive upregulation	VIII	NARK
Glyma.06G065600	Inoculated	Additivity	XII	NIN2 - RWP-RK/PB1; putative TF
Glyma.02G270800	Inoculated	D3 expression-level dominance	IV	NFR1
Glyma.15G003900	Inoculated	D3 expression-level dominance	IX	Nod26 - Major intrinsic protein
Glyma.07G208900	Inoculated	D4 expression-level dominance	II	SAN1B - 2OG-Fe(II) oxygenase
Glyma.07G209100	Inoculated	Transgressive downregulation	III	SAN1A - 2OG-Fe(II) oxygenase
Glyma.14G150100	Inoculated	Transgressive upregulation	V	ENOD8 - GDSL-like Lipase/Acylhydrolase
Glyma.04G251900	Inoculated	Transgressive upregulation	VI	NSP2 - GRAS TF
Glyma.10G248900	Inoculated	Transgressive upregulation	VIII	Nodulin-33 - haloacid dehalogenase-like hydrolase

Despite the general intermediacy of T2 in overall transcriptional analyses, then, certain key nodulation genes examined here showed expression-level dominance, where the allopolyploid T2 had similar expression to that of the higher-expressing diploid progenitor, and transgressive upregulation in T2 when compared to D3 and D4. Transgressive upregulation was observed for nodulin (Nodulin-33 in both control and inoculated samples) and early nodulin genes (ENOD8 in inoculated samples), showing an enhanced capacity for expression-level responses to inoculation in T2. *NSP2* was also transgressively upregulated in inoculated T2 samples. NSP2 is a member of the GRAS family of transcription factors and is a key regulator forming a complex with NSP1 that binds to the promoters of several nodulation-induced genes essential for the symbiotic signaling response, including *NIN* and *ERN1* (Hirsch et al. 2009; Kaló et al. 2005). In terms of expression-level dominance, it is notable that, in the control samples, both Nod factor receptor genes *NFR1* and *NFR5* fell into this category, with T2 expressing at the higher level of D3, thereby potentially enabling enhanced perception of rhizobial signaling in T2 relative to D4. Thus, these cases of non-additive expression patterns indicate a broader capacity in T2 for increased expression of genes in these particular classes and in connection with nodulation interactions.

Coupled with the enhanced defense response in D3, the examination of expression patterns for nodulation-related genes in the allopolyploid compared to its progenitors therefore provides a potential partial explanation for enhanced symbiotic responsiveness of T2 to NGR234. In conjunction with its reduced overall transcriptional response, T2 appears to have a reduced defense response relative to D3, while also expressing several key nodulation-related genes at or above the level in one or both diploid progenitor species. This result further

highlights the importance of a balance and interaction between nodulation and defense responses required for functional nodulation symbioses.

We also explored homoeologue usage in relation to effects of inoculation; if a shift in homoeologue usage following inoculation were detected, this could signal mobilization and preferential involvement of one subgenome in the process of nodulation. However, little overall tendency toward D3 or D4 was observed in the T2 samples and no significant effect of inoculation was detected. Estimates for the overall transcriptome showed approximately equal D3 and D4 homoeologue usage in both control (average D3 homoeologue usage of 49.68%) and inoculated T2 samples (average D3 homoeologue usage of 49.70%) (Appendix 3). No effects of inoculation on homoeologue usage in T2 were observed on estimates of the percentage of D3 homoeologue expression; average percentages under the two treatments were not significantly different (paired t-test: p-value = 0.66) (Appendix 3). Similarly, for the known nodulation gene subset, the mock-inoculated sample average D3 usage was 49.08% and the inoculated average D3 usage was 50.02%; the difference was not significant (paired t-test: p-value = 0.59) (Appendix 3). When comparing the percentage of D3 homoeologue usage in the nodulation gene subset by treatment type against the corresponding treatment estimates for the overall transcriptome, there were also no significant differences (control paired t-test: p-value = 0.45; inoculated paired t-test: p-value = 0.65).

Thus, in relation to nodulation and responses to rhizobia, we detected no effects of inoculation on the percentage of D3 homoeologue usage either in the transcriptome, generally, or in the pre-defined nodulation-related gene set. However, allopolyploid T2 does in fact show evidence of expressing multiple homoeologous copies of key symbiotic receptors (Figure 5.10), and this holds implications for the perception of rhizobial signals and the establishment of nodule

symbioses (Powell and Doyle, 2015). We found that in inoculated and uninoculated samples, both the D3 and D4 homoeologues of *NFR1* and *NFR5* were expressed in T2. For *NFR1*, the percentages of counts from the D3 homoeologue were 59.3% and 42.1% in control and inoculated samples respectively, while for *NFR5*, the percentages of D3 counts were 59.9% and 55.2% in control and inoculated samples; these percentages were not significantly different between treatments (*NFR1* paired t-test: p-value = 0.33; *NFR5* paired t-test: p-value = 0.48). NORK is another receptor that is believed to be important in nodulation; this receptor also showed expression of both D3 and D4 homoeologues in T2. The difference in D3 usage between 45.5% in mock-inoculated and 49.5% in inoculated T2 samples was not significant (paired t-test: p-value = 0.13).

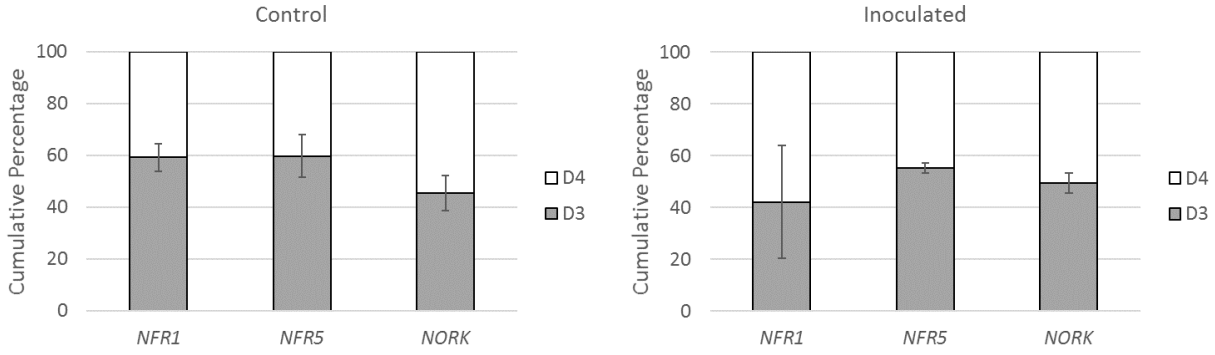


Figure 5.10. Homoeologue usage with respect to three key nodulation-related signaling receptors, in control (mock-inoculated) and inoculated samples of allopolyploid T2. Error bars indicate standard deviation.

In the examination of the percentages of D3 and D4 read mapping in T2 for these individual receptor genes, there were also no clear patterns of bias toward either progenitor (Table 5.3). Some individual samples showed significant deviations from 50% expression of

each homoeologue: *NFR1* expression in the inoculated G1393 sample deviated significantly, with a bias toward the D4 homoeologue (X^2=4.45, $p < 0.05$), *NFR5* expression in the mock-inoculated G1134 sample deviated significantly, with a bias toward the D3 homoeologue (X^2=14.54, $p < 0.05$), and *NORK* expression in the mock-inoculated G1393 sample deviated significantly, with a bias toward the D4 homoeologue (X^2=8.30, $p < 0.05$). Nevertheless, there was no clear bias across receptors; even among the examples with significant deviations, there was no consistency in the sample, accession or the direction of the bias involved.

Table 5.3. X^2 tests for deviations from 50:50 expression of homoeologues for genes encoding receptors involved in nodulation-related signal perception.

Receptor	T2 Accession	Treatment	X^2	Significance
NFR1	1134	Control	0.48	n.s.
		Inoculated	0.11	n.s.
	1188	Control	2.31	n.s.
		Inoculated	1.29	n.s.
	1393	Control	1.19	n.s.
		Inoculated	4.45	< 0.05
NFR5	1134	Control	14.55	< 0.05
		Inoculated	1.24	n.s.
	1188	Control	0.11	n.s.
		Inoculated	0.71	n.s.
	1393	Control	1.52	n.s.
		Inoculated	0.20	n.s.
NORK	1134	Control	2.81	n.s.
		Inoculated	3.06	n.s.
	1188	Control	0.0031	n.s.
		Inoculated	3.82	n.s.
	1393	Control	8.30	< 0.05
		Inoculated	1.29	n.s.

The expression of both homoeologues of these receptors and the absence of a clear pattern of usage bias has implications for the symbiotic capacity of the allopolyploid. The Nod factor receptors NFR1 and NFR5 are thought to form heterodimers and, together, are believed to

partly mediate specificity in interactions between legumes and rhizobia (Radutoiu et al. 2007); compatibility between the structure of the Nod factors and these Nod factor receptors can contribute to determining the symbiotic response. Thus, by expressing both homoeologous copies of these receptors, as we found in both mock-inoculated and inoculated samples, the allopolyploid T2 has the potential to increase its range of symbiotic partners. Similarly, NORK is an important receptor that constrains nodule formation; it appears to confer additional specificity between host and symbiont and may interact with NFR1 or NFR5 (Ried et al. 2014; Oldroyd 2013). In allopolyploid T2, as with *NFR1* and *NFR5*, both D3 and D4 homoeologous copies of *GmNORK* are expressed, thereby potentially enabling symbiotic interactions with a larger set of rhizobial genotypes. While expression of multiple copies of such symbiotic receptors is not a sufficient condition for an enhanced symbiotic range, it does constitute a necessary condition; without it, such an increase in potential symbiont diversity mediated by receptor diversity would not be possible (Powell and Doyle 2015).

Nodulation-related modules and hub genes identified by network analysis have non-additive expression patterns

To complement our examination of transcriptional responses of the allopolyploid and its two diploid progenitors using a set of known, pre-defined nodulation-related genes, we sought to examine expression patterns of modules and hub genes that were not defined *a priori* but, rather, were identified as relevant to nodulation through analysis of the transcriptome data from D3, D4, and T2. We used WGCNA to identify modules that were co-expressed across all species; the analysis identified 46 modules of co-expressed genes divided into three blocks, based on the blockwise construction of the network (Figure 5.11). However, it should be noted that the 'grey' module is reserved for genes that could not be placed into any coexpression module; here, the

grey module contained 2791 genes. Collectively, these modules contain 45697 of the 56044 genes represented in the transcriptomes (10347 were excluded from the analysis for having too many missing values or zero variance).

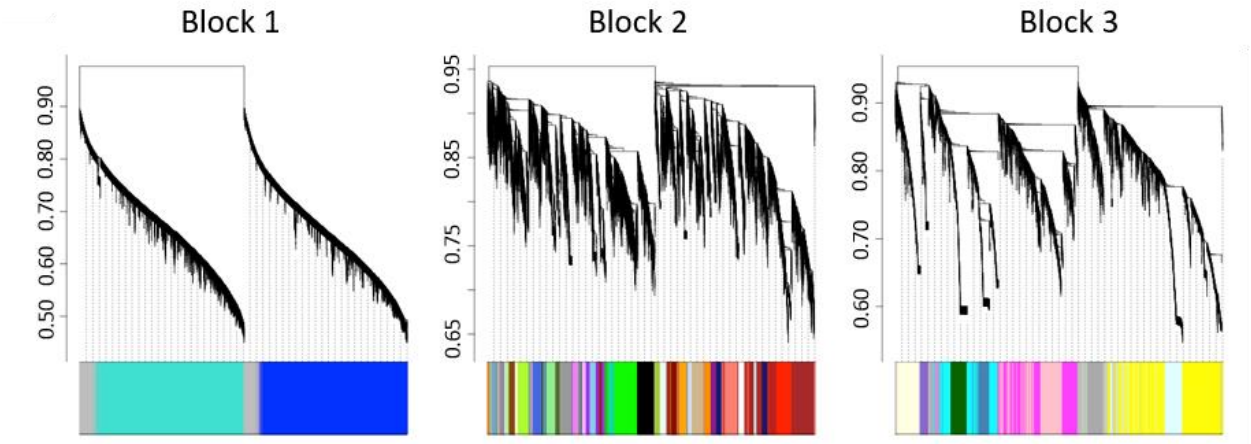


Figure 5.11. Gene dendrograms obtained by WGCNA using blockwise network construction and all inoculated and mock inoculated transcriptome samples. The colors below the dendrograms show module assignment determined by the Dynamic Tree Cut.

We then assessed the correlation between module eigengene expression levels and several traits in inoculated samples (Figure 5.12). Several modules consistently showed either positive or negative correlations across all traits. For subsequent analyses, we focused primarily on the root hair deformation trait, due to the transcriptome sampling corresponding most closely in time to the measurements of this trait, and its proximal relevance to early signaling and transcriptional responses in the nodulation symbiosis. The eigengene expression of several modules was correlated with root hair deformation. Modules with significant positive correlations ($p < 0.01$) included those designated with the color labels ‘red’ (cor. = -0.60; p-value = 0.008) and ‘brown’ (cor. = 0.72; p-value = 0.0007), while modules with significant negative correlations were ‘green’ (cor. = -0.76; p-value < 0.05) and ‘black’ (cor. = -0.66; p-value =

0.003). The ‘grey60’ (cor. = -0.70; p-value = 0.001) and ‘plum1’ (cor. = -0.62; p-value = 0.006) modules also had significant negative correlations with this trait. These thus formed an initial set of modules implicated in nodulation-related processes.

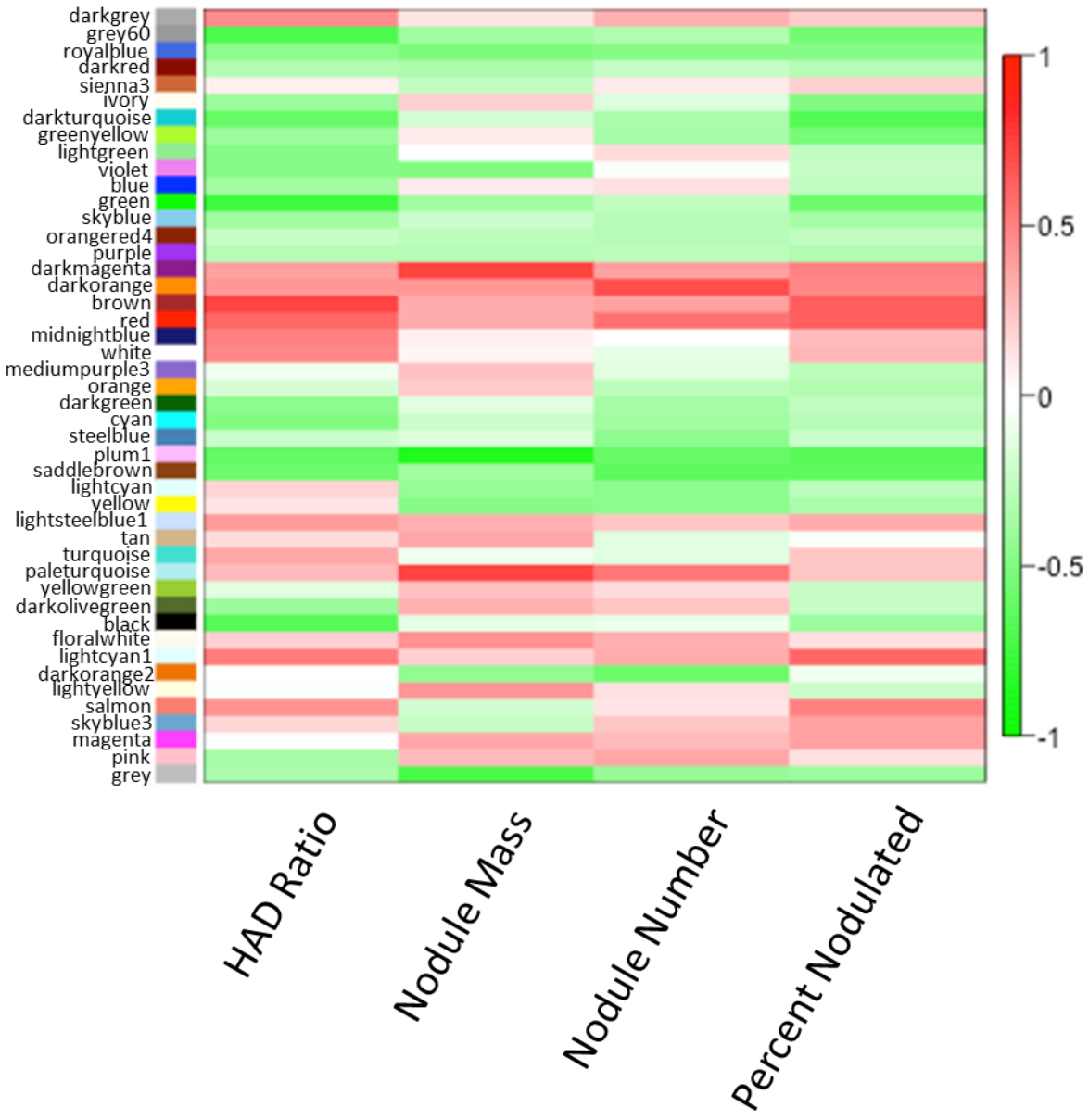


Figure 5.12. Correlation heatmap of module eigengenes with nodulation-related trait data from (Powell and Doyle 2016). Traits include the root hair deformation ratio (HAD Ratio), total

nodule mass per plant (Nodule Mass), total number of nodules per plant (Nodule Number), and percentage of plants nodulated per accession (Percent Nodulated).

Modules can be characterized and expression can be compared between species through assessment of either module eigengene expression or hub gene expression. First, in the module eigengene comparison between species for the modules of interest, species was a significant factor for the 'brown' ($p = 0.0376$), 'green' ($p = 0.006$), 'grey60' ($p = 0.00447$) and 'plum1' ($p = 0.00681$) modules. In the case of the 'brown' module, a module for which eigengene expression was positively correlated with root hair deformation, T2 eigengene expression was at the higher level of D4, while expression was lower in D3, indicating a pattern of expression in the allopolyploid similar to the higher expressing parent (Figure 5.13). With the 'green' and 'grey60' modules, the pattern was similar to the 'brown' module except that these modules were negatively correlated to the trait and D4 and T2 both expressed the eigengene at lower levels than D3 (Figure 5.13). For the 'plum1' module, a transgressive expression pattern was observed in T2; with the 'plum1' module, which was negatively correlated with the root hair trait, T2 eigengene expression was reduced relative to D3 and D4 (Figure 5.13).

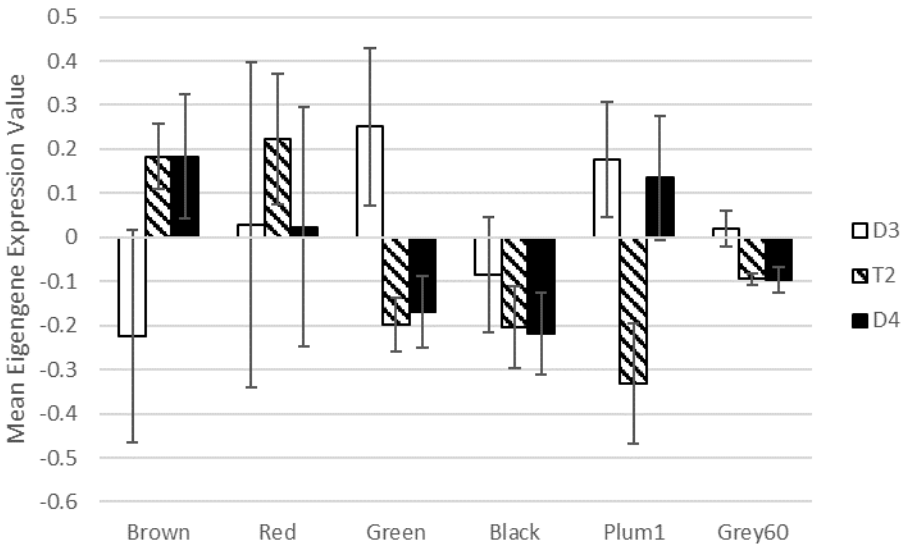


Figure 5.13. Expression of WGCNA module eigengenes, compared across species for each of the modules that were most highly correlated with the root hair deformation trait. Values shown are species means ± standard deviation.

To understand more about the composition of the modules for which species was a significant factor, GO overrepresentation tests with the 'Molecular function complete' and 'Biological process complete' annotation data sets revealed which types of transcripts were associated with each of these modules to a greater degree than expected. The 'green' and 'plum1' modules had no significantly overrepresented categories, while the 'grey60' module had the significantly overrepresented biological process and molecular function of oxygen transport (p-value: 1.08E-02) and oxygen transport activity (p-value: 8.59E-03), respectively. In contrast to the other three modules, the 'brown' module had a variety of significant categories; the top four most significantly overrepresented parent categories are shown in Figure 5.14. These included regulation of ARF (ADP-ribosylation factor) protein signal transduction (biological process; p-value: 1.76E-04), ARF guanyl-nucleotide exchange factor (ARF-GEF) activity (molecular function; p-value: 1.40E-04), Golgi vesicle transport (biological process; 2.61E-03),

and intracellular protein transport (biological process; 4.61E-02). The overrepresentation and implication of ARF-GEF proteins is relevant because auxin accumulation and auxin-related signaling are believed to be critical for stimulating cell division during nodule organogenesis; ARF-GEF proteins are key regulators of the localization of auxin PIN transporter proteins, and the localization of PIN proteins alters local auxin (Steinmann et al. 1999; Geldner et al. 2003; Mathesius 2008). The expression of the 'brown' module eigengene in T2 at the level of the high-expression parent (D4) thus indicates a greater auxin-related signaling response to inoculation, which is implicated in the process of symbiotic development and nodule formation.

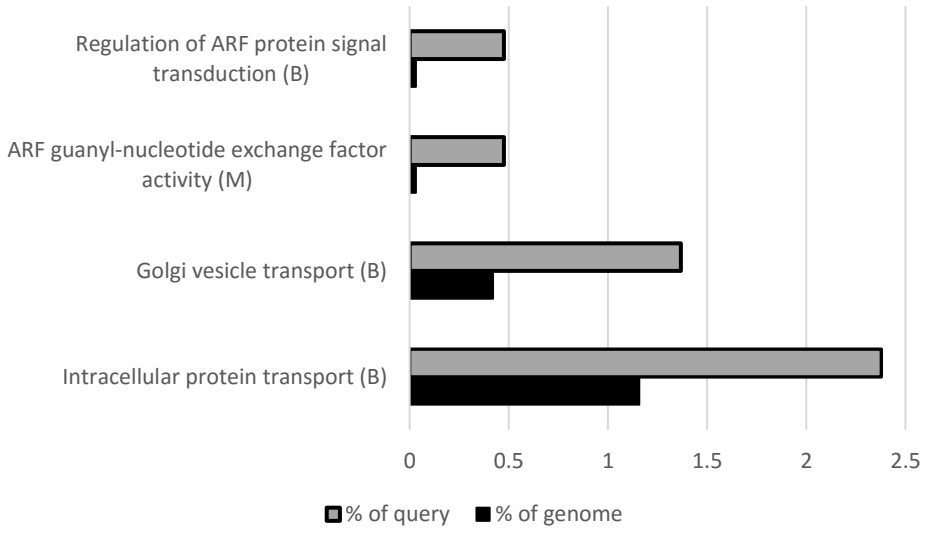


Figure 5.14. The four most highly significant GO biological process and molecular function parent categories resulting from a GO overrepresentation analysis of 'brown' module member genes. The figure illustrates this overrepresentation by showing the genes in each category as a percentage of the genome and as a percentage of the set of 'brown' module genes. For each category (B) indicates a biological process; (M) indicates a molecular function.

The second summary approach for assessing expression involves identifying hub genes for each module. For most of the modules identified as significantly related to the root hair trait, module membership of genes was highly correlated with gene significance for the trait (Figure 5.15). Genes with both high module membership (high within-module connectivity) and high

gene significance for the root hair trait were identified as hub genes of interest for the trait. With a criterion of gene significance greater than 0.7 and module membership greater than 0.7 as the minimum threshold, the ‘plum1’ and ‘grey60’ modules did not have any genes that were identified as hub genes and were thus excluded from this analysis. Three to five hub genes met the minimum threshold of gene significance and module membership for the ‘green’, ‘black’, and ‘red’ modules. With a more stringent criterion of gene significance and module membership greater than 0.8, 24 hub genes were identified in the ‘brown’ module.

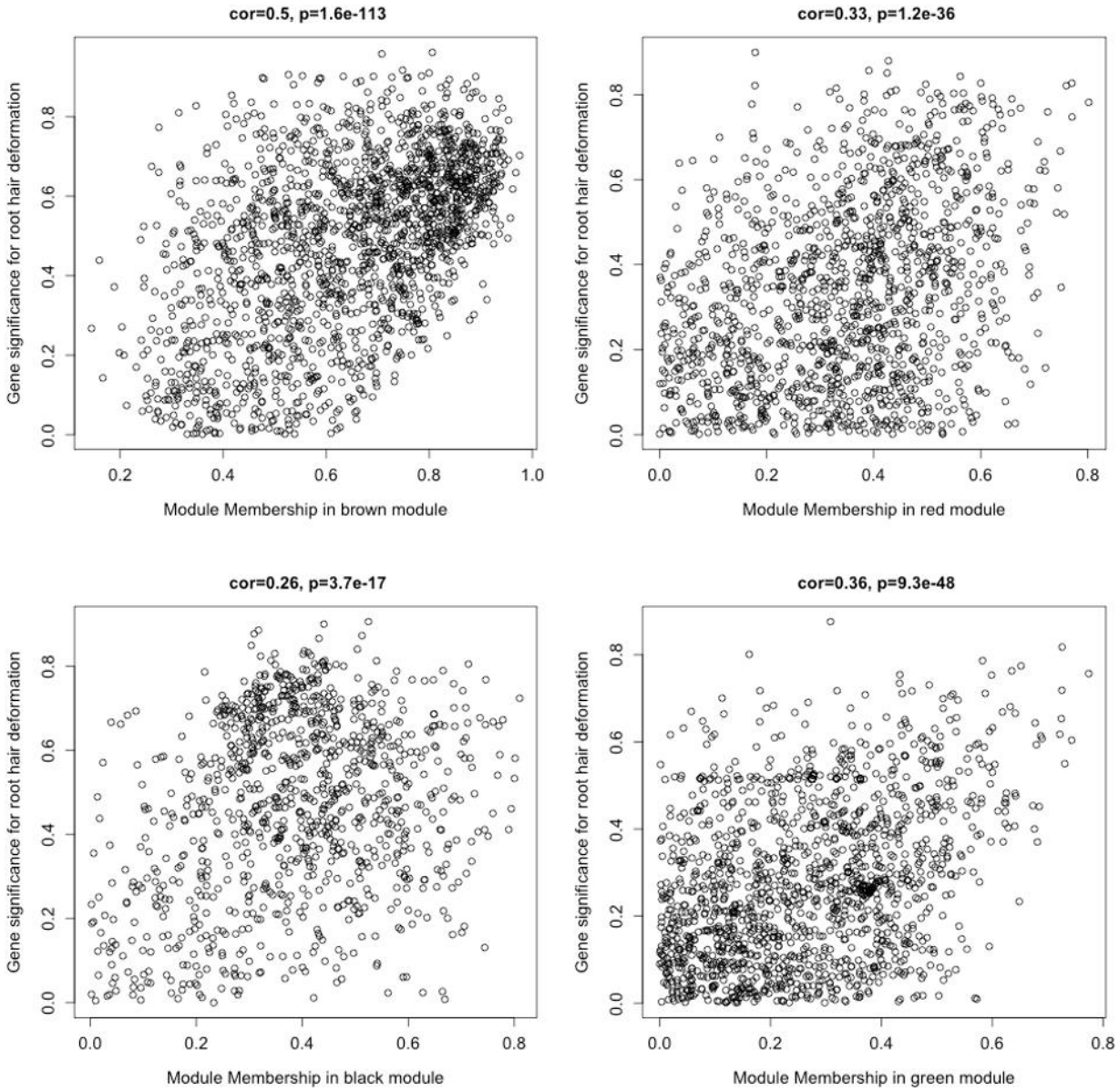


Figure 5.15. Scatterplots of module membership (x-axis) relative to gene significance (y-axis) for all genes in each of the 'brown', 'red', 'black', and 'green' modules, which were the four modules for which hub genes were obtained that met the threshold criteria for both metrics.

The hub genes identified for each module were then classified according to the relative expression categories used for the known, *a priori* set of nodulation genes (Table 5.3). In the 'red' and 'black' modules, none of the hub genes showed significant differences in pairwise comparisons, as was the case for two of the hub genes from the 'green' module, while one hub

gene in this module had a pattern of additivity. In the 'brown' module, a majority of the hub genes had a pattern of D4 expression-level dominance. This complements the eigengene expression results for this module. While the hub genes in this category do not show any obvious relationship to each other or to nodulation in terms of their identity, some of the genes appear to be implicated in ethylene signaling: Glyma.03G251700 is annotated as a signal transduction histidine kinase ethylene sensor; Glyma.04G147000 is annotated as an F-box leucine rich repeat protein. The homolog of Glyma.04G147000 in *Arabidopsis thaliana* is *EBF1*, a negative regulator involved in the ethylene response pathway (Potuschak et al. 2003). Ethylene is typically considered an inhibitor of nodulation, with ethylene-insensitivity leading to supernodulation phenotypes (Penmetsa and Cook 1997; Ferguson and Mathesius 2014) and it is possible that ethylene signaling is involved in a Nod factor-triggered negative feedback mechanism (Miyata et al. 2013). The relationship between ethylene signaling and nodulation appears to be complex, however, and there is some evidence that ethylene-insensitive mutants of *Lotus japonicus* have reduced nodule numbers (e.g., Chan et al. 2013) and, in *G. max*, there was no consistent inhibition of nodulation by ethylene, and ethylene-insensitive mutants appeared to have fewer nodules than wild-type when exposed to increased levels of an ethylene precursor (Schmidt et al. 1999). In connection with the GO analysis results for the 'brown' module discussed above, ethylene has been implicated in alterations to auxin transport during nodulation, and is coupled with changes in gene expression of the auxin efflux transporters *MtPIN1* and *MtPIN2* in *Medicago truncatula* (Prayitno et al. 2006).

Table 5.3. Expression patterns for hub genes from selected WGCNA modules.

	Additivity		D4 expression-level dominance		D3 expression-level dominance		Transgressive downregulation			Transgressive upregulation			No Change
	I	XII	II	XI	IV	IX	III	VII	X	V	VI	VIII	
Module													
Brown	-	-	14	-	-	-	-	-	-	1	-	-	8
Green	1	-	-	-	-	-	-	-	-	-	-	-	2
Red	-	-	-	-	-	-	-	-	-	-	-	-	5
Black	-	-	-	-	-	-	-	-	-	-	-	-	5

Conclusions

In this study, we observed distinctive transcriptional responses of allopolyploid T2 in response to inoculation with rhizobia. However, these distinctive responses were not initially evident in transcriptome-wide analyses. Our results indicate that an allopolyploid may have a reduced overall transcriptional response to rhizobia compared with one or both diploid progenitors, while nevertheless having an enhanced symbiotic response. Reduction of transcriptional defense-related responses can play a role in modulating the response to rhizobia and the outcome of the interaction.

Furthermore, apparent expression-level intermediacy that is revealed, for example, by analyses of the most highly heterogeneously expressed genes, can also mask other expression patterns underlying biotic responses. Here, non-additive patterns were revealed in the analyses of pre-defined nodulation-related genes and modules identified through network analysis. When analyses were constrained to these sets of genes, patterns of expression-level dominance and transgressive patterns of expression (but not homoeologue usage biased consistently toward one progenitor) were observed in T2. These provide evidence of gene regulation on the part of the allopolyploid, in response to rhizobial inoculation, that is distinct from that of one or both diploid

progenitors and deviates from expectations of additivity. The transgressive upregulation of certain nodulation-related genes and modules or their expression at levels of the higher-expressing diploid, coupled with a reduced defense response at the level of gene expression, provide evidence for enhanced symbiotic responses in allopolyploid T2 when inoculated with the NGR234 strain of rhizobia.

CHAPTER 6

CONCLUSIONS

OVERVIEW

Polyploidy and nodulation are both important phenomena in the evolutionary history of the legume family, and researchers have proposed that polyploidy may have had a role in the evolution and refinement of nodulation. The aim of this dissertation is to fill gaps in our knowledge of potential interactions between allopolyploidy and nodulation, taking advantage of recent allopolyploidy in *Glycine* subgenus *Glycine*, the perennial relatives of soybean. Specifically, we sought to examine symbiotic signaling and assess symbiotic interactions in allopolyploid and diploid progenitor species in the subgenus. In this chapter, we synthesize our findings and discuss their implications for nodulation and polyploidy research, and also outline further questions and directions for future study brought forward by our work.

SUMMARY OF RESEARCH FINDINGS

The primary objectives of the research were threefold: (1) to evaluate symbiotic signaling capacity in the T1, T2 and T5 allopolyploids, compared to their respective diploids, by assessing biosynthesis and exudation of root metabolites; (2) to determine whether allopolyploid T2 exhibits differential symbiotic responses to inoculation with rhizobia, assessed in terms of the propensity to form nodulating associations, root hair deformation, nodule mass and nodule number; (3) to study differences in gene expression following inoculation in T2 relative to its diploid progenitors. The summary of our research findings is organized around these three primary objectives.

Objective 1. Evaluate symbiotic signaling capacity in the T1, T2 and T5 allopolyploid triads

We assessed the diversity of root metabolites across the T1, T2 and T5 triads, with a focus on isoflavones because these are primary signaling molecules in nodulation symbioses. We

tested the effects of species and inoculation treatment in each of the triads. No common transgressive patterns were found across allopolyploids relative to diploid progenitors. In terms of overall profiles, the allopolyploids appeared to be intermediate to the diploids, with species having greatest influence as an explanatory factor in the T2 triad. Isoflavone diversity was also affected by species; D3 and D4 were significantly different from each other, with T2 intermediate between the two. Isoflavone diversity in T2, a widely colonizing species, was also greater than in T5, which was the allopolyploid in this study with the most restricted range. Treatment was not a significant explanatory factor, except in the case of total amounts of all metabolites and isoflavones in the T5 triad and total isoflavones in the T2 triad. Since the greatest differences between species occurred between members of the T2 triad, we focused on further aspects in these species. In terms of specific isoflavones, formononetin and daidzein showed higher levels in D4 relative to D3, with T2 also showing lower levels of daidzein and intermediate levels of formononetin. Formononetin was also found to be a key compound in distinguishing between the three species in the triad. Exudates were also assessed, with a focus on the T2 triad, since the most striking differences in root metabolites were found in that triad. We found that the difference in formononetin that was present in the root tissues was also present in the exudates, providing clear evidence that symbiotic interactions would be affected by the presence of this compound. Lastly, with T2, D3 and D4, we also examined expression of selected genes in the isoflavone biosynthetic pathway. The expression patterns of chalcone reductase genes support the proposition that the proportional increase in formononetin and daidzein in the roots of D4 is due to increased biosynthesis leading to the branch of the pathway that includes both formononetin and daidzein. The expression patterns of several other key

biosynthetic genes in the pathway also helped account for the increase in isoflavones observed in response to inoculation.

Objective 2. Determine whether allopolyploid T2 exhibits differential symbiotic responses to inoculation with rhizobia relative to its diploid progenitors

We conducted inoculation trials on several accessions of the three species in the T2 triad. With a variety of rhizobial strains, T2 showed an overall greater tendency to form nodules than the diploids, D3 and D4. Across all rhizobial strains, root hair deformation was significantly affected by species, with T2 exhibiting greater responses relative to D3. With strain NGR234, statistical analysis demonstrated that T2 had a greater propensity for plants to form nodules relative to D4, but not D3. Finally, considering nodulated plants, while no difference was detected among the three species, T2 had a significantly greater total nodule mass per plant in a pairwise comparison with D3.

Objective 3. Study differences in gene expression following inoculation in T2 relative to its diploid progenitors

In examining overall transcriptomic effects in the T2 triad, we found that T2 transcriptomes were intermediate between D3 and D4. Beneath this apparent overall intermediacy, however, substantial differences were observed, particularly in terms of the effects on gene expression of inoculation with rhizobia. First, D3 showed a greater overall transcriptomic response to inoculation compared to either T2 or D4 that appeared to be driven by an enhanced defense response. Furthermore, T2 showed non-additive patterns in expression of known nodulation-related genes. We also conducted a network analysis to identify modules of co-expressed genes and assessed modules that were significantly correlated with the root hair deformation trait. The result of this analysis highlighted the importance of auxin and ethylene

signaling, suggesting that T2 shows an upregulated auxin-related response corresponding to that of D4, which was greater than that in D3. Similarly, T2 appeared to follow D4 in having enhanced negative regulation of ethylene signaling. In general, auxin-related signaling is associated with nodule formation, while ethylene is believed to be a negative regulator of nodulation.

SPECIFIC CONTRIBUTIONS TO THE LITERATURE AND BROADER IMPLICATIONS

Here, we have added to the literature and the understanding of metabolites, particularly isoflavones, and their biosynthesis in the genus *Glycine*. We have expanded the number of species in which these topics have been examined, with most previous studies having focused on restricted sets of species, especially *G. max*. We are also the first to examine metabolites in relation to the interaction between inoculation, nodulation and allopolyploidy. We have also complemented studies of exudates in *G. max* and other legumes, by presenting the first report of exudate analysis in *Glycine* subgenus *Glycine*.

We also added to the understanding of rhizobial interactions with species in this subgenus. Previous studies have either used limited numbers of species in this complex or did not use known, well-characterized bacterial strains (Pueppke 1988; Pueppke and Broughton 1999; Brockwell et al. 1998). Furthermore, previous inoculation studies did not examine allopolyploids in relation to their relevant diploid progenitors, whereas in our study, our use of an allopolyploid triad enabled inferences relating allopolyploidy to symbiotic responses.

Our transcriptome data provides a novel data set for species in *Glycine* subgenus *Glycine*; no prior studies exist examining transcriptome responses to nodulation in this system. Our study

complements transcriptomic studies of various tissues and stages in the nodulation process in *G. max*, conducted by Hayashi et al. (2012b), de Carvalho et al. (2013), and Libault et al. (2010), supporting the importance of downregulation in defense responses for symbiosis to proceed, and adds to evidence supporting the importance of hormonal regulation in the symbiotic response, particularly the involvement of auxin and ethylene. Furthermore, previous studies have not examined transcriptional responses to inoculation in relation to polyploidy; thus, our findings show newly characterized non-additive expression patterns in the allopolyploid response to rhizobia.

In terms of broader implications, our results suggest that enhanced symbiotic capacity in *Glycine* subgenus *Glycine* allopolyploids could contribute to greater ability to colonize beyond native ranges relative to diploid species. However, enhanced symbiotic responses do not appear to be a function of ploidy alone but are, rather, dependent on the particular progenitors of the given allopolyploid. Insofar as the enhanced signaling capabilities between allopolyploids have been examined here, results are consistent with the differing ranges and colonization abilities of some perennial *Glycine* allopolyploids such as T2.

With respect to symbiotic signaling, our results highlight the necessity of considering the full suite of signaling compounds present in diploids and allopolyploids, and their relevance in the context of interactions with rhizobia. While researchers had previously suggested that synthesis of a broader diversity of flavonoids in polyploids could give rise to enhanced biotic interactions, our results emphasize that, while synthesized compounds are relevant, it is also important to consider which compounds are exuded; it is also necessary to consider that not all compounds will enhance interactions. Indeed, enhanced symbiotic interactions in polyploids may

be achieved by not synthesizing or exuding certain compounds, as they may have inhibitory effects.

DIRECTIONS FOR FUTURE RESEARCH

Our work explored various aspects related to symbiotic signaling and responses in particular allopolyploids and their diploid progenitors in *Glycine* subgenus *Glycine*. For each aspect of the study, we recommend expanding analyses to additional allopolyploids and diploids, to enable testing the generality of our findings and their applicability to other members of subgenus *Glycine*. This is especially true for the inoculation trials and the transcriptomic studies. In the case of inoculation trials, predictions for allopolyploids such as T1 and T5 can be made based on the metabolite analyses conducted here. Similarly, transcriptomic analyses of the T1 and T5 triads are currently planned, to verify whether hormonal responses in inoculation may also be different in those species.

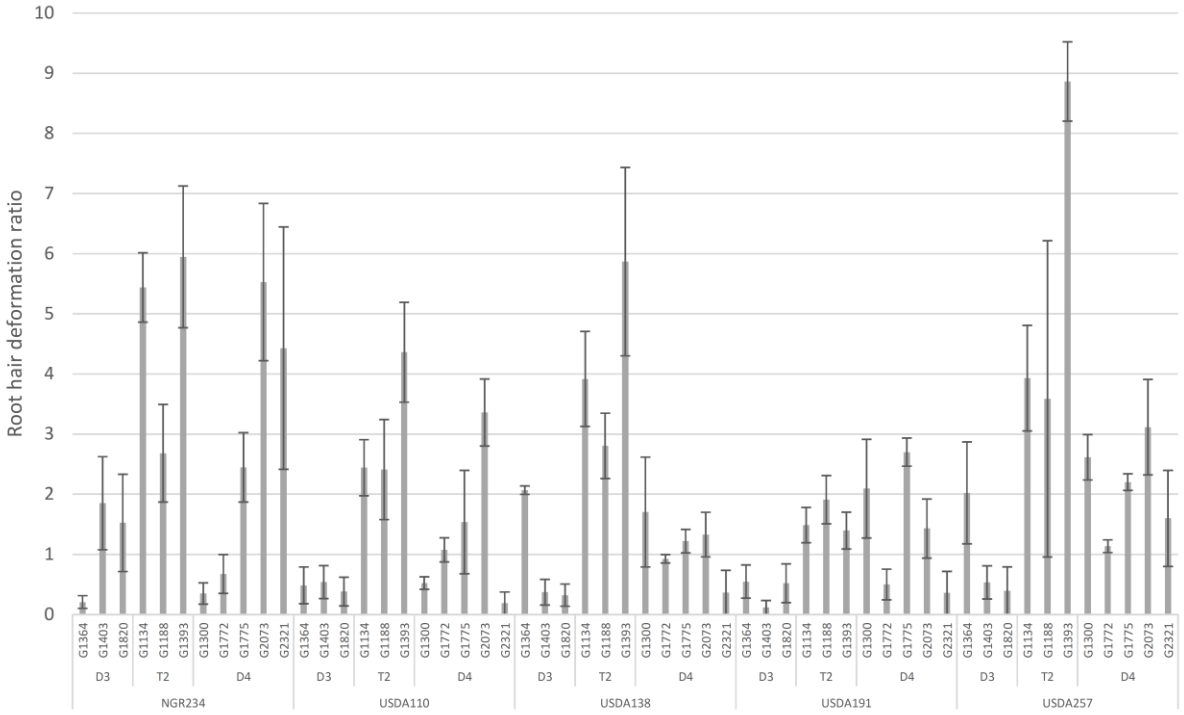
In addition to expanding the plant species tested, we are also interested in expanding the variety of rhizobia assessed. The broader the diversity of rhizobia with which tests are conducted, the more complete the assessment of realizable symbiotic capacity. In addition, field studies of rhizobial interactions with allopolyploid and diploids, both within the native Australian ranges and in colonized regions, along with reciprocal transplant experiments, would help to further elucidate the connections between symbiotic interactions with rhizobia and colonization ability.

In conjunction with this, at various levels of analysis, we wish to explore more concretely the potential role of formononetin in biotic interactions involving D4, and to compare the particular effects of the combinations of compounds found in the allopolyploids and their

diploids on bacterial strains isolated from their respective native ranges. This can be ultimately achieved by *lacz* β galactosidase experiments using mixtures of the compounds and examining their Nod factor inducing activity in the rhizobia. The importance of particular isoflavones could also be explored through knockdown of enzymes in the biosynthetic pathway leading to specific isoflavones via hairy root transformations. With further studies examining these phenomena at ecological levels of analysis, combined with additional mechanistic analyses, we will be able to build upon the results presented here and obtain a more complete understanding of the interactions between allopolyploidy and nodulation in *Glycine* subgenus *Glycine*.

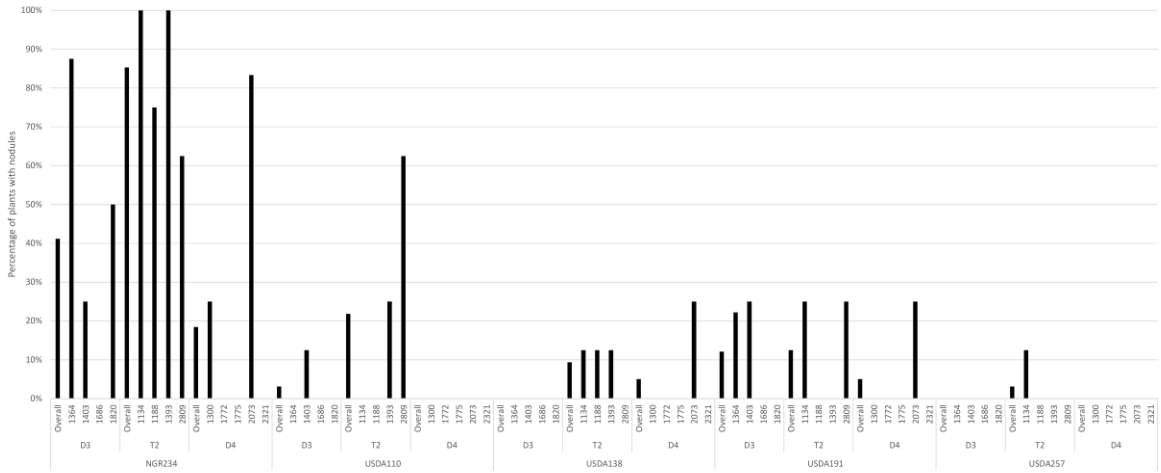
APPENDIX 1

Appendix 1. Root hair deformation ratios by accession for D3, T2, and D4 species when treated with rhizobial strains NGR234, USDA110, USDA138, USDA191 and USDA257. Values represent accession means ± standard error (N = 3 for each accession by treatment combination, except between G1188 and USDA110 with N = 2 and between G1393 and USDA257 with N = 4).



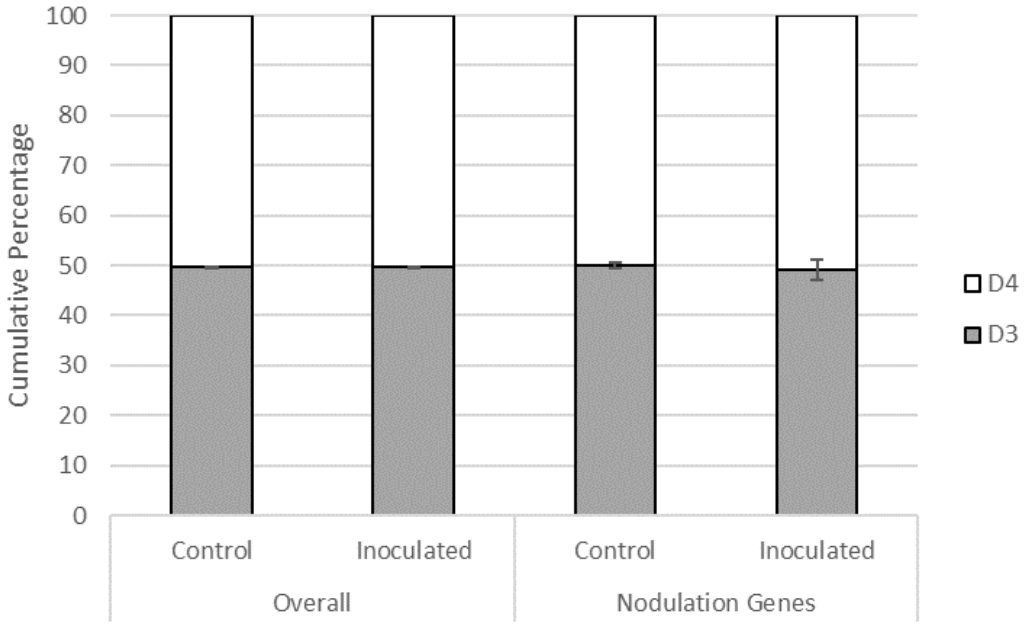
APPENDIX 2

Appendix 2. Percentage of plants with nodules by accession for D3, T2, and D4 species following inoculation with rhizobia NGR234, USDA110, USDA138, USDA191, and USDA257.



APPENDIX 3

Appendix 3. Homoeologue usage estimates for the overall transcriptome and for sets of nodulation-related genes, in control (mock-inoculated) and inoculated samples of allopolyploid T2 (n = 3). Stacked bars show mean percentages for each homoeologue. Error bars indicate standard deviation.



REFERENCES

Aguilar JMM, Ashby AM, Richards AJ, Loake GJ, Watson MD, Shaw CH (1988) Chemotaxis of *Rhizobium leguminosarum* biovar *phaseoli* towards flavonoid inducers of the symbiotic nodulation genes. Microbiology 134 (10):2741-2746

Althabegoiti MJ, López-García SL, Piccinetti C, Mongiardini EJ, Pérez-Giménez J, Quelas JI, Perticari A, Lodeiro AR (2008) Strain selection for improvement of *Bradyrhizobium japonicum* competitiveness for nodulation of soybean. FEMS Microbiology Letters 282 (1):115-123

Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD (2013) Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nature Protocols 8 (9):1765-1786

Anders S, Pyl PT, Huber W (2014) HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics:btu638

Andrio E, Marino D, Marmeys A, Segonzac MD, Damiani I, Genre A, Huguet S, Frendo P, Puppo A, Pauly N (2013) Hydrogen peroxide-regulated genes in the *Medicago truncatula–Sinorhizobium meliloti* symbiosis. New Phytologist 198 (1):179-189

Ané J-M, Kiss GB, Riely BK, Penmetsa RV, Oldroyd GE, Ayax C, Lévy J, Debellé F, Baek J-M, Kalo P (2004) *Medicago truncatula DMI1* required for bacterial and fungal symbioses in legumes. Science 303 (5662):1364-1367

Aoki T, Akashi T, Ayabe S-i (2000) Flavonoids of leguminous plants: structure, biological activity, and biosynthesis. Journal of Plant Research 113 (4):475-488

Armitage J, Gallagher A, Johnston A (1988) Comparison of the chemotactic behaviour of *Rhizobium leguminosarum* with and without the nodulation plasmid. Molecular Microbiology 2 (6):743-748

Aronesty E (2013) Comparison of sequencing utility programs. The Open Bioinformatics Journal 7 (1)

Arrighi J-F, Barre A, Amor BB, Bersoult A, Soriano LC, Mirabella R, de Carvalho-Niebel F, Journet E-P, Ghérardi M, Huguet T (2006) The *Medicago truncatula* lysin motif-receptor-like kinase gene family includes *NFP* and new nodule-expressed genes. Plant Physiology 142 (1):265-279

Arvanitis L, Wiklund C, Ehrlén J (2007) Butterfly seed predation: effects of landscape characteristics, plant ploidy level and population structure. Oecologia 152 (2):275-285

Arvanitis L, Wiklund C, Münzbergova Z, Dahlgren JP, Ehrlén J (2010) Novel antagonistic interactions associated with plant polyploidization influence trait selection and habitat preference. Ecology Letters 13 (3):330-337

Ashfield T, Egan AN, Pfeil BE, Chen NW, Podicheti R, Ratnaparkhe MB, Ameline-Torregrosa C, Denny R, Cannon S, Doyle JJ (2012) Evolution of a complex disease resistance gene cluster in diploid *Phaseolus* and tetraploid *Glycine*. Plant Physiology 159 (1):336-354

Barbour W, Hattermann D, Stacey G (1991) Chemotaxis of *Bradyrhizobium japonicum* to soybean exudates. Applied and Environmental Microbiology 57 (9):2635-2639

Barrett LG, Bever JD, Bissett A, Thrall PH (2015) Partner diversity and identity impacts on plant productivity in *Acacia*–rhizobial interactions. Journal of Ecology 103 (1):130-142

Bassam BJ, Djordjevic MA, Redmond JW, Batley M, Rolfe BG (1988) Identification of a *nodD*-dependent locus in the *Rhizobium* strain NGR234 activated by phenolic factors secreted by soybeans and other legumes. Molecular Plant-Microbe Interactions 1:161-168

Bates D, Maechler M, Bolker B, Walker S, Christensen R, Singmann H, Dai B, Grothendieck G (2015) Package "lme4". R package version 1.1-10.
Beatty PH, Good AG (2011) Future prospects for cereals that fix nitrogen. Science 333 (6041):416-417
Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B (Methodological):289-300
Bennett M (1972) Nuclear DNA content and minimum generation time in herbaceous plants. Proceedings of the Royal Society of London B: Biological Sciences 181 (1063):109-135
Berendsen RL, Pieterse CM, Bakker PA (2012) The rhizosphere microbiome and plant health. Trends in Plant Science 17 (8):478-486
Birchler JA, Veitia RA (2012) Gene balance hypothesis: connecting issues of dosage sensitivity across biological disciplines. Proceedings of the National Academy of Sciences 109 (37):14746-14753
Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. The Plant Cell 16 (7):1679-1691
Bolaños-Vásquez MC, Werner D (1997) Effects of *Rhizobium tropici*, *R. etli*, and *R. leguminosarum* bv. *phaseoli* on nod gene-inducing flavonoids in root exudates of *Phaseolus vulgaris*. Molecular Plant-Microbe Interactions 10:339-346
Bombarely A, Coate JE, Doyle JJ (2014) Mining transcriptomic data to study the origins and evolution of a plant allopolyploid complex. PeerJ 2:e391
Brechenmacher L, Lei Z, Libault M, Findley S, Sugawara M, Sadowsky MJ, Sumner LW, Stacey G (2010) Soybean metabolites regulated in root hairs in response to the symbiotic bacterium *Bradyrhizobium japonicum*. Plant Physiology 153 (4):1808-1822
Brockwell J, McIlroy R, Hebb DM (1998) The Australian collection of Rhizobium strains for temperate legumes. Catalogue 1998. CSIRO (Australia), Plant Industry, Divisional Report No. 98/1. Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia.
Broghammer A, Krusell L, Blaise M, Sauer J, Sullivan JT, Maolanon N, Vinther M, Lorentzen A, Madsen EB, Jensen KJ (2012) Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. Proceedings of the National Academy of Sciences 109 (34):13859-13864
Broughton WJ, Jabbouri S, Perret X (2000) Keys to symbiotic harmony. Journal of Bacteriology 182 (20):5641-5652
Brown A, Doyle J, Grace J, Doyle J (2002) Molecular phylogenetic relationships within and among diploid races of *Glycine tomentella* (Leguminosae). Australian Systematic Botany 15 (1):37-47
Bruneau A, Mercure M, Lewis GP, Herendeen PS (2008) Phylogenetic patterns and diversification in the caesalpinioid legumes. Botany 86 (7):697-718
Bueno P, Soto MJ, Rodríguez-Rosales MP, Sanjuan J, Olivares J, Donaire JP (2001) Time-course of lipoxygenase, antioxidant enzyme activities and H_2O_2 accumulation during the early stages of Rhizobium–legume symbiosis. New Phytologist 152 (1):91-96
Buggs RJ, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE (2014) The legacy of diploid progenitors in allopolyploid gene expression patterns. Phil Trans R Soc B 369 (1648):20130354

Cacco G, Ferrari G, Lucci G (1976) Uptake efficiency of roots in plants at different ploidy levels. The Journal of Agricultural Science 87 (03):585-589
Caetano-Anolles G, Crist-Estes D, Bauer W (1988) Chemotaxis of *Rhizobium meliloti* to the plant flavone luteolin requires functional nodulation genes. Journal of Bacteriology 170 (7):3164-3169
Cannon SB, Ilut D, Farmer AD, Maki SL, May GD, Singer SR, Doyle JJ (2010) Polyploidy did not predate the evolution of nodulation in all legumes. PLoS One 5 (7):e11630
Cannon SB, McKain MR, Harkess A, Nelson MN, Dash S, Deyholos MK, Peng Y, Joyce B, Stewart CN, Rolf M (2015) Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Molecular Biology and Evolution 32 (1):193-210
Cardoso D, De Queiroz LP, Pennington RT, De Lima HC, Fonty É, Wojciechowski MF, Lavin M (2012) Revisiting the phylogeny of papilionoid legumes: New insights from comprehensively sampled early-branching lineages. American Journal of Botany 99 (12):1991-2013
Cardoso D, Pennington R, De Queiroz L, Boatwright J, Van Wyk B-E, Wojciechowski M, Lavin M (2013) Reconstructing the deep-branching relationships of the papilionoid legumes. South African Journal of Botany 89:58-75
Cassman K, Munns D, Beck D (1981) Phosphorus nutrition of *Rhizobium japonicum*: strain differences in phosphate storage and utilization. Soil Science Society of America Journal 45 (3):517-520
Cesco S, Neumann G, Tomasi N, Pinton R, Weisskopf L (2010) Release of plant-borne flavonoids into the rhizosphere and their role in plant nutrition. Plant and Soil 329 (1-2):1-25
Chan PK, Biswas B, Gresshoff PM (2013) Classical ethylene insensitive mutants of the *Arabidopsis EIN2* orthologue lack the expected 'hypernodulation' response in *Lotus japonicus*. Journal of Integrative Plant Biology 55 (4):395-408
Chao D-Y, Dilkes B, Luo H, Douglas A, Yakubova E, Lahner B, Salt DE (2013) Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. Science 341 (6146):658-659
Charpentier M, Bredemeier R, Wanner G, Takeda N, Schleiff E, Parniske M (2008) *Lotus japonicus* CASTOR and POLLUX are ion channels essential for perinuclear calcium spiking in legume root endosymbiosis. The Plant Cell 20 (12):3467-3479
Chaudhary B, Hovav R, Flagel L, Mittler R, Wendel JF (2009) Parallel expression evolution of oxidative stress-related genes in fiber from wild and domesticated diploid and polyploid cotton (*Gossypium*). BMC Genomics 10 (1):1
Chelaifa H, Monnier A, Ainouche M (2010) Transcriptomic changes following recent natural hybridization and allopolyploidy in the salt marsh species *Spartina* × *townsendii* and *Spartina anglica* (Poaceae). New Phytologist 186 (1):161-174
Chen W, Yan G, Li J (1988) Numerical taxonomic study of fast-growing soybean rhizobia and a proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov. International Journal of Systematic and Evolutionary Microbiology 38 (4):392-397
Chen ZJ (2010) Molecular mechanisms of polyploidy and hybrid vigor. Trends in Plant Science 15 (2):57-71
Cho I, Blaser MJ (2012) The human microbiome: at the interface of health and disease. Nature Reviews Genetics 13 (4):260-270

Cho M-J, Harper JE (1991) Effect of inoculation and nitrogen on isoflavonoid concentration in wild-type and nodulation-mutant soybean roots. Plant Physiology 95 (2):435-442
Coate J, Powell A, Owens T, Doyle J (2013) Transgressive physiological and transcriptomic responses to light stress in allopolyploid *Glycine dolichocarpa* (Leguminosae). Heredity 110 (2):160-170
Coate JE, Bar H, Doyle JJ (2014) Extensive translational regulation of gene expression in an allopolyploid (*Glycine dolichocarpa*). The Plant Cell 26 (1):136-150
Coate JE, Doyle JJ (2010) Quantifying whole transcriptome size, a prerequisite for understanding transcriptome evolution across species: an example from a plant allopolyploid. Genome Biology and Evolution 2:534-546
Coate JE, Doyle JJ (2015) Variation in transcriptome size: are we getting the message? Chromosoma 124 (1):27-43
Coate JE, Luciano AK, Seralathan V, Minchew KJ, Owens TG, Doyle JJ (2012) Anatomical, biochemical, and photosynthetic responses to recent allopolyploidy in *Glycine dolichocarpa* (Fabaceae). American Journal of Botany 99 (1):55-67
Combes M-C, Cenci A, Baraille H, Bertrand B, Lashermes P (2012) Homeologous gene expression in response to growing temperature in a recent allopolyploid (*Coffea arabica* L.). Journal of Heredity 103 (1):36-46
Crews T, Peoples M (2004) Legume versus fertilizer sources of nitrogen: ecological tradeoffs and human needs. Agriculture, Ecosystems & Environment 102 (3):279-297
Cullimore J, Gough C (2013) Lipochitooligosaccharide perception and the basis of partner recognition in root endosymbioses. Molecular Microbial Ecology of the Rhizosphere: Volume 1 & 2:483-494
D'Haeze W, Holsters M (2002) Nod factor structures, responses, and perception during initiation of nodule development. Glycobiology 12 (6):79R-105R
Dai W-J, Zeng Y, Xie Z-P, Staehelin C (2008) Symbiosis-promoting and deleterious effects of NopT, a novel type 3 effector of *Rhizobium* sp. strain NGR234. Journal of Bacteriology 190 (14):5101-5110
De Bodt S, Maere S, Van de Peer Y (2005) Genome duplication and the origin of angiosperms. Trends in Ecology & Evolution 20 (11):591-597
de Carvalho GAB, Batista JSS, Marcelino-Guimarães FC, do Nascimento LC, Hungria M (2013) Transcriptional analysis of genes involved in nodulation in soybean roots inoculated with *Bradyrhizobium japonicum* strain CPAC 15. BMC Genomics 14 (1):1
De Hoff PL, Brill LM, Hirsch AM (2009) Plant lectins: the ties that bind in root symbiosis and plant defense. Molecular Genetics and Genomics 282 (1):1-15
De Mita S, Streng A, Bisseling T, Geurts R (2014) Evolution of a symbiotic receptor through gene duplications in the legume–rhizobium mutualism. New Phytologist 201 (3):961-972
De Queiroz K (2007) Species concepts and species delimitation. Systematic Biology 56 (6):879-886
Deakin WJ, Broughton WJ (2009) Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. Nature Reviews Microbiology 7 (4):312-320
Den Herder J, Vanhee C, De Rycke R, Corich V, Holsters M, Goormachtig S (2007) Nod factor perception during infection thread growth fine-tunes nodulation. Molecular Plant-Microbe Interactions 20 (2):129-137
Denison RF, Kiers ET (2004) Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. FEMS Microbiology Letters 237 (2):187-193

Depuydt S, Hardtke CS (2011) Hormone signalling crosstalk in plant growth regulation. Current Biology 21 (9):R365-R373
Dhawan O, Lavania U (1996) Enhancing the productivity of secondary metabolites via induced polyploidy: a review. Euphytica 87 (2):81-89
Diaz-Uriarte R (2007) GeneSrF and varSelRF: a web-based tool and R package for gene selection and classification using random forest. BMC Bioinformatics 8 (1):1
Diaz RJ, Rosenberg R (2008) Spreading dead zones and consequences for marine ecosystems. Science 321 (5891):926-929
Doyle JJ (1994) Phylogeny of the legume family: an approach to understanding the origins of nodulation. Annual Review of Ecology and Systematics:325-349
Doyle JJ (2011) Phylogenetic perspectives on the origins of nodulation. Molecular Plant-Microbe Interactions 24 (11):1289-1295
Doyle JJ (2016) Chasing unicorns: Nodulation origins and the paradox of novelty. American Journal of Botany 103 (11): 1-4
Doyle JJ, Doyle JL, Brown A, Palmer RG (2002) Genomes, multiple origins, and lineage recombination in the *Glycine tomentella* (Leguminosae) polyploid complex: histone H3-D gene sequences. Evolution 56 (7):1388-1402
Doyle JJ, Doyle JL, Rauscher JT, Brown A (2004) Evolution of the perennial soybean polyploid complex (*Glycine* subgenus *Glycine*): a study of contrasts. Biological Journal of the Linnean Society 82 (4):583-597
Duzan H, Zhou X, Souleimanov A, Smith D (2004) Perception of *Bradyrhizobium japonicum* Nod factor by soybean [*Glycine max* (L.) Merr.] root hairs under abiotic stress conditions. Journal of Experimental Botany 55 (408):2641-2646
Egan AN, Doyle J (2010) A comparison of global, gene-specific, and relaxed clock methods in a comparative genomics framework: dating the polyploid history of soybean (*Glycine max*). Systematic Biology 59 (5):534-547
Endre G, Kereszt A, Kevei Z, Mihacea S, Kaló P, Kiss GB (2002) A receptor kinase gene regulating symbiotic nodule development. Nature 417 (6892):962-966
Etzler ME, Kalsi G, Ewing NN, Roberts NJ, Day RB, Murphy JB (1999) A nod factor binding lectin with apyrase activity from legume roots. Proceedings of the National Academy of Sciences 96 (10):5856-5861
Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB (2012) Animal behavior and the microbiome. Science 338 (6104):198-199
Fawcett JA, Van de Peer Y (2010) Angiosperm polyploids and their road to evolutionary success. Trends in Evolutionary Biology 2 (1):3
Ferguson BJ, Indrasumunar A, Hayashi S, Lin MH, Lin YH, Reid DE, Gresshoff PM (2010) Molecular analysis of legume nodule development and autoregulation. Journal of Integrative Plant Biology 52 (1):61-76
Ferguson BJ, Mathesius U (2014) Phytohormone regulation of legume-rhizobia interactions. Journal of Chemical Ecology 40 (7):770-790
Flagel LE, Wendel JF, Udall JA (2012) Duplicate gene evolution, homoeologous recombination, and transcriptome characterization in allopolyploid cotton. BMC Genomics 13 (1):1
Freeling M, Thomas BC (2006) Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Research 16 (7):805-814
Friesen ML (2012) Widespread fitness alignment in the legume–rhizobium symbiosis. New Phytologist 194 (4):1096-1111

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. The Plant Cell 19 (11):3403-3417
Gaeta RT, Yoo S-Y, Pires J, Doerge RW, Chen ZJ, Osborn TC (2009) Analysis of gene expression in resynthesized *Brassica napus* allopolyploids using *Arabidopsis* 70mer oligo microarrays. PLoS One 4 (3):e4760
Gagnon H, Ibrahim RK (1998) Aldonic acids: a novel family of nod gene inducers of *Mesorhizobium loti*, *Rhizobium lupini*, and *Sinorhizobium meliloti*. Molecular Plant-Microbe Interactions 11 (10):988-998
Gan Y, Stulen I, van Keulen H, Kuiper PJ (2003) Effect of N fertilizer top-dressing at various reproductive stages on growth, N_2 fixation and yield of three soybean (*Glycine max* (L.) Merr.) genotypes. Field Crops Research 80 (2):147-155
Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M (2014) Two evolutionarily distinct classes of paleopolyploidy. Molecular Biology and Evolution 31 (2):448-454
Geldner N, Anders N, Wolters H, Keicher J, Kornberger W, Muller P, Delbarre A, Ueda T, Nakano A, Jürgens G (2003) The *Arabidopsis* GNOM ARF-GEF mediates endosomal recycling, auxin transport, and auxin-dependent plant growth. Cell 112 (2):219-230
Gobbato E, Marsh JF, Vernié T, Wang E, Maillet F, Kim J, Miller JB, Sun J, Bano SA, Ratet P (2012) A GRAS-type transcription factor with a specific function in mycorrhizal signaling. Current Biology 22 (23):2236-2241
Gomez C, Terrier N, Torregrosa L, Vialet S, Fournier-Level A, Verriès C, Souquet J-M, Mazauric J-P, Klein M, Cheynier V (2009) Grapevine MATE-type proteins act as vacuolar H^+-dependent acylated anthocyanin transporters. Plant Physiology 150 (1):402-415
Gonzalez-Rizzo S, Crespi M, Frugier F (2006) The *Medicago truncatula* CRE1 cytokinin receptor regulates lateral root development and early symbiotic interaction with *Sinorhizobium meliloti*. The Plant Cell 18 (10):2680-2693
Gould SJ, Vrba ES (1982) Exaptation—a missing term in the science of form. Paleobiology 8 (01):4-15
Gourion B, Berrabah F, Ratet P, Stacey G (2015) Rhizobium–legume symbioses: the crucial role of plant immunity. Trends in Plant Science 20 (3):186-194
Green AF, Ramsey TS, Ramsey J (2011) Phylogeny and biogeography of ivies (*Hedera* spp., Araliaceae), a polyploid complex of woody vines. Systematic Botany 36 (4):1114-1127
Green AF, Ramsey TS, Ramsey J (2013) Polyploidy and invasion of English ivy (*Hedera* spp., Araliaceae) in North American forests. Biological Invasions 15 (10):2219-2241
Gresshoff PM, Lohar D, Chan P-K, Biswas B, Jiang Q, Reid D, Ferguson B, Stacey G (2009) Genetic analysis of ethylene regulation of legume nodulation. Plant Signaling & Behavior 4 (9):818-823
Griesbach R, Kamo K (1996) The effect of induced polyploidy on the flavonols of *Petunia* 'Mitchell'. Phytochemistry 42 (2):361-363
Grover C, Gallagher J, Szadkowski E, Yoo M, Flagel L, Wendel J (2012) Homoeolog expression bias and expression level dominance in allopolyploids. New Phytologist 196 (4):966-971
Gust AA, Willmann R, Desaki Y, Grabherr HM, Nürnberger T (2012) Plant LysM proteins: modules mediating symbiosis and immunity. Trends in Plant Science 17 (8):495-502

Gutierrez-Gonzalez JJ, Guttikonda SK, Tran L-SP, Aldrich DL, Zhong R, Yu O, Nguyen HT, Sleper DA (2010) Differential expression of isoflavone biosynthetic genes in soybean during water deficits. Plant and Cell Physiology 51 (6):936-948

Halverson K, Heard SB, Nason JD, Stireman III JO (2008) Differential attack on diploid, tetraploid, and hexaploid *Solidago altissima* L. by five insect gallmakers. Oecologia 154 (4):755-761

Hao GY, Lucero ME, Sanderson SC, Zacharias EH, Holbrook NM (2013) Polyploidy enhances the occupation of heterogeneous environments through hydraulic related trade-offs in *Atriplex canescens* (Chenopodiaceae). New Phytologist 197 (3):970-978

Harbert RS, Brown AH, Doyle JJ (2014) Climate niche modeling in the perennial *Glycine* (Leguminosae) allopolyploid complex. American Journal of Botany 101 (4):710-721

Hayashi M, Saeki Y, Haga M, Harada K, Kouchi H, Umehara Y (2012a) *Rj* (*rj*) genes involved in nitrogen-fixing root nodule formation in soybean. Breeding Science 61 (5):544-553

Hayashi M, Shiro S, Kanamori H, Mori-Hosokawa S, Sasaki-Yamagata H, Sayama T, Nishioka M, Takahashi M, Ishimoto M, Katayose Y (2014) A thaumatin-like protein, Rj4, controls nodule symbiotic specificity in soybean. Plant and Cell Physiology 55 (9):1679-1689

Hayashi S, Reid DE, Lorenc MT, Stiller J, Edwards D, Gresshoff PM, Ferguson BJ (2012b) Transient Nod factor-dependent gene expression in the nodulation-competent zone of soybean (*Glycine max* [L.] Merr.) roots. Plant Biotechnology Journal 10 (8):995-1010

Heath K, Stock A, Stinchcombe J (2010) Mutualism variation in the nodulation response to nitrate. Journal of Evolutionary Biology 23 (11):2494-2500

Heath KD (2010) Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. Evolution 64 (5):1446-1458

Heath KD, Burke PV, Stinchcombe JR (2012) Coevolutionary genetic variation in the legume-rhizobium transcriptome. Molecular Ecology 21 (19):4735-4747

Heath KD, Lau JA (2011) Herbivores alter the fitness benefits of a plant–rhizobium mutualism. Acta Oecologica 37 (2):87-92

Heath KD, Tiffin P (2007) Context dependence in the coevolution of plant and rhizobial mutualists. Proceedings of the Royal Society of London B: Biological Sciences 274 (1620):1905-1912

Hegarty M, Coate J, Sherman-Broyles S, Abbott R, Hiscock S, Doyle J (2013) Lessons from natural and artificial polyploids in higher plants. Cytogenetic and Genome Research 140 (2-4):204-225

Hegarty MJ, Abbott RJ, Hiscock SJ (2012) Allopolyploid speciation in action: The origins and evolution of *Senecio cambrensis*. In: Polyploidy and Genome Evolution. Springer, pp 245-270

Herridge DF (2008) Inoculation Technology For Legumes. In: Dilworth MJ, James EK, Sprent JI, Newton WE (eds) Nitrogen-fixing Leguminous Symbioses. Springer Netherlands, Dordrecht, pp 77-115. doi:10.1007/978-1-4020-3548-7_4

Hijmans RJ, Gavrilenko T, Stephenson S, Bamberg J, Salas A, Spooner DM (2007) Geographical and environmental range expansion through polyploidy in wild potatoes (*Solanum* section *Petota*). Global Ecology and Biogeography 16 (4):485-495

Hirsch S, Kim J, Muñoz A, Heckmann AB, Downie JA, Oldroyd GE (2009) GRAS proteins form a DNA binding complex to induce gene expression during nodulation signaling in *Medicago truncatula*. The Plant Cell 21 (2):545-557

Hull-Sanders HM, Johnson RH, Owen HA, Meyer GA (2009a) Effects of polyploidy on secondary chemistry, physiology, and performance of native and invasive genotypes of *Solidago gigantea* (Asteraceae). American Journal of Botany 96 (4):762-770

Hull-Sanders HM, Johnson RH, Owen HA, Meyer GA (2009b) Influence of polyploidy on insect herbivores of native and invasive genotypes of *Solidago gigantea* (Asteraceae). Plant Signaling & Behavior 4 (9):893-895

Hymowitz T, Singh R, Larkin R (1990) Long-distance dispersal: The case for the allopolyploid *Glycine tabacina* (Labill.) Benth. and *G. tomentella* Hayata in the West-Central Pacific. Micronesica: Journal of the University of Guam 23 (1):5-13

Ilut DC, Coate JE, Luciano AK, Owens TG, May GD, Farmer A, Doyle JJ (2012) A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. American Journal of Botany 99 (2):383-396

Indrasumunar A, Kereszt A, Searle I, Miyagi M, Li D, Nguyen CD, Men A, Carroll BJ, Gresshoff PM (2010) Inactivation of duplicated nod factor receptor 5 (NFR5) genes in recessive loss-of-function non-nodulation mutants of allotetraploid soybean (*Glycine max L.* Merr.). Plant and Cell Physiology 51 (2):201-214

Indrasumunar A, Searle I, Lin MH, Kereszt A, Men A, Carroll BJ, Gresshoff PM (2011) Nodulation factor receptor kinase 1α controls nodule organ number in soybean (*Glycine max* L. Merr). The Plant Journal 65 (1):39-50

Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. Nature Reviews Genetics 11 (2):97-108

Jamet A, Mandon K, Puppo A, Hérouart D (2007) H_2O_2 is required for optimal establishment of the *Medicago sativa*/*Sinorhizobium meliloti* symbiosis. Journal of Bacteriology 189 (23):8741-8745

Jardinaud M-F, Boivin S, Rodde N, Catrice O, Kisiala A, Lepage A, Moreau S, Roux B, Cottret L, Sallet E (2016) A laser dissection-RNAseq analysis highlights the activation of cytokinin pathways by Nod factors in the *Medicago truncatula* root epidermis. Plant Physiology:pp. 00711.02016

Jensen ES, Peoples MB, Boddey RM, Gresshoff PM, Hauggaard-Nielsen H, Alves BJ, Morrison MJ (2012) Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. Agronomy for Sustainable Development 32 (2):329-364

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473 (7345):97-100

Jones JD, Dangl JL (2006) The plant immune system. Nature 444 (7117):323-329

Kaló P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, Jakab J, Sims S, Long SR, Rogers J (2005) Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science 308 (5729):1786-1789

Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriguchi M, Kawashima K (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Research 9 (6):189-197

Kape R, Parniske M, Werner D (1991) Chemotaxis and *nod* gene activity of *Bradyrhizobium japonicum* in response to hydroxycinnamic acids and isoflavonoids. Applied and Environmental Microbiology 57 (1):316-319
Keen N, Staskawicz B (1988) Host range determinants in plant pathogens and symbionts. Annual Reviews in Microbiology 42 (1):421-440
Kessler A (2015) The information landscape of plant constitutive and induced secondary metabolite production. Current Opinion in Insect Science 8:47-53
Kiba T, Kudo T, Kojima M, Sakakibara H (2011) Hormonal control of nitrogen acquisition: roles of auxin, abscisic acid, and cytokinin. Journal of Experimental Botany 62 (4):1399-1409
Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12 (4):357-360
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL (2013a) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14 (4):1
Kim DH, Parupalli S, Azam S, Lee S-H, Varshney RK (2013b) Comparative sequence analysis of nitrogen fixation-related genes in six legumes. Frontiers in Plant Science 4
Kobayashi H, Broughton WJ (2008) Fine-tuning of symbiotic genes in Rhizobia: Flavonoid signal transduction cascade. In: Dilworth MJ, James EK, Sprent JI, Newton WE (eds) Nitrogen-fixing Leguminous Symbioses. Springer Netherlands, Dordrecht, pp 117-152. doi:10.1007/978-1-4020-3548-7_5
Kosslak RM, Bookland R, Barkei J, Paaren HE, Appelbaum ER (1987) Induction of *Bradyrhizobium japonicum* common *nod* genes by isoflavones isolated from Glycine max. Proceedings of the National Academy of Sciences 84 (21):7428-7432
Kosslak RM, Joshi RS, Bowen BA, Paaren HE, Appelbaum ER (1990) Strain-specific inhibition of nod gene induction in *Bradyrhizobium japonicum* by flavonoid compounds. Applied and Environmental Microbiology 56 (5):1333-1341
Kouchi H, Shimomura K, Hata S, Hirota A, Wu G-J, Kumagai H, Tajima S, Suganuma N, Suzuki A, Aoki T (2004) Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. DNA Research 11 (4):263-274
Krouk G, Ruffel S, Gutiérrez RA, Gojon A, Crawford NM, Coruzzi GM, Lacombe B (2011) A framework integrating plant growth with hormones and nutrients. Trends in Plant Science 16 (4):178-182
Kuznetsova A, Brockhoff PB, Christensen RHB (2015) Package 'lmerTest'. R package version:2.0-29
Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9 (1):1
Langfelder P, Horvath S (2012) Fast R functions for robust correlations and hierarchical clustering. Journal of Statistical Software 46 (11)
Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nature Methods 9 (4):357-359
Lavania UC, Srivastava S, Lavania S, Basu S, Misra NK, Mukai Y (2012) Autopolyploidy differentially influences body size in plants, but facilitates enhanced accumulation of secondary metabolites, causing increased cytosine methylation. The Plant Journal 71 (4):539-549

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Systematic Biology 54 (4):575-594
Le Strange KK, Bender GL, Djordjevic MA, Rolfe BG, Redmond JW (1990) The *Rhizobium* strain NGR234 *nodD*1 gene product responds to activation by the simple phenolic compounds vanillin and isovanillin present in wheat seedling extracts. Molecular Plant-Microbe Interactions 3 (4):214-220
Lea US, Slimestad R, Smedvig P, Lillo C (2007) Nitrogen deficiency enhances expression of specific MYB and bHLH transcription factors and accumulation of end products in the flavonoid pathway. Planta 225 (5):1245-1253
Lenth R, Hervé M (2016) lsmeans: least-squares means. R package version 222
Leps WT, Brill WJ, Bingham E (1980) Effect of alfalfa ploidy on nitrogen fixation. Crop Science 20 (4):427-430
Levin DA (1983) Polyploidy and novelty in flowering plants. American Naturalist:1-25
Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press,
Levine A, Tenhaken R, Dixon R, Lamb C (1994) H_2O_2 from the oxidative burst orchestrates the plant hypersensitive disease resistance response. Cell 79 (4):583-593
Lévy J, Bres C, Geurts R, Chalhoub B, Kulikova O, Duc G, Journet E-P, Ané J-M, Lauber E, Bisseling T (2004) A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science 303 (5662):1361-1364
Levy M (1976) Altered glycoflavone expression in induced autotetraploids of *Phlox drummondii*. Biochemical Systematics and Ecology 4 (4):249-254
Levy M, Levin DA (1971) The origin of novel flavonoids in *Phlox* allotetraploids. Proceedings of the National Academy of Sciences 68 (7):1627-1630
Levy M, Levin DA (1974) Novel flavonoids and reticulate evolution in the *Phlox pilosa-P. drummondii* complex. American Journal of Botany:156-167
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25 (16):2078-2079
Li Q-G, Zhang L, Li C, Dunwell JM, Zhang Y-M (2013) Comparative genomics suggests that an ancestral polyploidy event leads to enhanced root nodule symbiosis in the Papilionoideae. Molecular Biology and Evolution 30 (12):2602-2611
Liaw A, Wiener M (2002) Classification and Regression by randomForest. R News 2 (3):18-22
Libault M (2014) The carbon-nitrogen balance of the nodule and its regulation under elevated carbon dioxide concentration. BioMed Research International 2014
Libault M, Farmer A, Brechenmacher L, Drnevich J, Langley RJ, Bilgin DD, Radwan O, Neece DJ, Clough SJ, May GD (2010) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. Plant Physiology 152 (2):541-552
Lin S-J, Lay H-L, Wu S-T, Thseng F-S (2005) Contents of certain isoflavones in *Glycine dolichocarpa*, *G. tabacina* and *G. tomentella* collected in Taiwan. 藥物食品分析 13 (3):260-266
Liu C-W, Murray JD (2016) The role of flavonoids in nodulation host-range specificity: an update. Plants 5 (3):33

Lohar D, Stiller J, Kam J, Stacey G, Gresshoff PM (2009) Ethylene insensitivity conferred by a mutated *Arabidopsis* ethylene receptor gene alters nodulation in transgenic *Lotus japonicus*. Annals of Botany:mcp132
Lohar DP, Haridas S, Gantt JS, VandenBosch KA (2007) A transient decrease in reactive oxygen species in roots leads to root hair deformation in the legume–rhizobia symbiosis. New Phytologist 173 (1):39-49
Lohar DP, Schaff JE, Laskey JG, Kieber JJ, Bilyeu KD, Bird DM (2004) Cytokinins play opposite roles in lateral root formation, and nematode and Rhizobial symbioses. The Plant Journal 38 (2):203-214
Long SR (1996) Rhizobium symbiosis: nod factors in perspective. The Plant Cell 8 (10):1885
Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15 (12):1
Lowry E, Lester S (2006) The biogeography of plant reproduction: potential determinants of species' range sizes. Journal of Biogeography 33 (11):1975-1982
Mabood F, Souleimanov A, Khan W, Smith D (2006) Jasmonates induce Nod factor production by *Bradyrhizobium japonicum*. Plant Physiology and Biochemistry 44 (11):759-765
Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110 (2):99-104
Madlung A, Wendel J (2013) Genetic and epigenetic aspects of polyploid evolution in plants. Cytogenetic and Genome Research 140 (2-4):270-285
Madsen EB, Madsen LH, Radutoiu S, Olbryt M, Rakwalska M, Szczyglowski K, Sato S, Kaneko T, Tabata S, Sandal N (2003) A receptor kinase gene of the LysM type is involved in legume perception of rhizobial signals. Nature 425 (6958):637-640
Maere S, De Bodt S, Raes J, Casneuf T, Van Montagu M, Kuiper M, Van de Peer Y (2005) Modeling gene and genome duplications in eukaryotes. Proceedings of the National Academy of Sciences of the United States of America 102 (15):5454-5459
Marchant BD, Soltis DE, Soltis PS (2016) Patterns of abiotic niche shifts in allopolyploids relative to their progenitors. New Phytologist 212 (3):708-718
Marie C, Deakin WJ, Viprey V, Kopciñska J, Golinowski W, Krishnan HB, Perret X, Broughton WJ (2003) Characterization of Nops, nodulation outer proteins, secreted via the type III secretion system of NGR234. Molecular Plant-Microbe Interactions 16 (9):743-751
Marinova K, Pourcel L, Weder B, Schwarz M, Barron D, Routaboul J-M, Debeaujon I, Klein M (2007) The *Arabidopsis* MATE transporter TT12 acts as a vacuolar flavonoid/H^+-antiporter active in proanthocyanidin-accumulating cells of the seed coat. The Plant Cell 19 (6):2023-2038
Martin SL, Husband BC (2009) Influence of phylogeny and ploidy on species ranges of North American angiosperms. Journal of Ecology 97 (5):913-922
Mathesius U (2003) Conservation and divergence of signalling pathways between roots and soil microbes—the Rhizobium-legume symbiosis compared to the development of lateral roots, mycorrhizal interactions and nematode-induced galls. In: Roots: The dynamic interface between plants and the earth. Springer, pp 105-119
Mathesius U (2008) Auxin: at the root of nodule development? Functional Plant Biology 35 (8):651-668
McClintock B (1984) The significance of responses of the genome to challenge. Science (New York, NY) 226 (4676):792-801

McIntyre PJ (2012) Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. American Journal of Botany 99 (4):655-662
Mears JA (1980) Chemistry of polyploids: a summary with comments on *Parthenium* (Asteraceae-Ambrosiinae). In: Polyploidy. Springer, pp 77-101
Meng F, Pang H, Huang F, Liu L, Wang Y (2012) Tetraploid black locust (*Robinia pseudoacacia* L.) Increased salt tolerance by activation of the antioxidant system. Biotechnology & Biotechnological Equipment 26 (6):3351-3358
Meng H-B, Jiang S-S, Hua S-J, Lin X-Y, Li Y-L, Guo W-L, Jiang L-X (2011) Comparison between a tetraploid turnip and its diploid progenitor (*Brassica rapa* L.): the adaptation to salinity stress. Agricultural Sciences in China 10 (3):363-375
Meyers BC, Kaushik S, Nandety RS (2005) Evolving disease resistance genes. Current Opinion in Plant Biology 8 (2):129-134
Miller LD, Yost CK, Hynes MF, Alexandre G (2007) The major chemotaxis gene cluster of *Rhizobium leguminosarum* bv. *viciae* is essential for competitive nodulation. Molecular Microbiology 63 (2):348-362
Mitra RM, Gleason CA, Edwards A, Hadfield J, Downie JA, Oldroyd GE, Long SR (2004) A Ca^{2+}/calmodulin-dependent protein kinase required for symbiotic nodule development: gene identification by transcript-based cloning. Proceedings of the National Academy of Sciences of the United States of America 101 (13):4701-4705
Miyahara A, Hirani TA, Oakes M, Kereszt A, Kobe B, Djordjevic MA, Gresshoff PM (2008) Soybean nodule autoregulation receptor kinase phosphorylates two kinase-associated protein phosphatases in vitro. Journal of Biological Chemistry 283 (37):25381-25391
Miyata K, Kawaguchi M, Nakagawa T (2013) Two distinct *EIN2* genes cooperatively regulate ethylene signaling in *Lotus japonicus*. Plant and Cell Physiology 54 (9):1469-1477
Münzbergová Z (2006) Ploidy level interacts with population size and habitat conditions to determine the degree of herbivory damage in plant populations. Oikos 115 (3):443-452
Murray B, Williams C (1976) Chromosome number and flavonoid synthesis in *Briza* L.(Gramineae). Biochemical Genetics 14 (11-12):897-904
Murray B, Williams CA (1973) Polyploidy and flavonoid synthesis in *Briza media* L. Nature 243:87-88
Nakano Y, Yamakawa T, Ikeda M, Ishizuka J (1997) Nodulation of Rj-soybean varieties with *Rhizobium fredii* USDA 193 under limited supply of nutrients. Soil Science and Plant Nutrition 43 (4):929-932
Noguti Y, Oka H, Ôtuka T (1940) Studies on the polyploidy in *Nicotiana* induced by the treatment with colchicine. II. Growth rate and chemical analysis of diploid and its autotetraploid in *Nicotiana rustica* and *N. tabacum*. Japanese Journal of Botany 10:343-364
Novak K, Chovanec P, Škrdleta V, Kropáčová M, Lisá L, Němcová M (2002) Effect of exogenous flavonoids on nodulation of pea (*Pisum sativum* L.). Journal of Experimental Botany 53 (375):1735-1745
Nuismer SL, Thompson JN (2001) Plant polyploidy and non-uniform effects on insect herbivores. Proceedings of the Royal Society of London B: Biological Sciences 268 (1479):1937-1940

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2015) vegan: Community Ecology Package. R package version 2.2-1. http://CRAN.R-project.org/package=vegan
Oldroyd GE (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nature Reviews Microbiology 11 (4):252-263
Oldroyd GE, Dixon R (2014) Biotechnological solutions to the nitrogen problem. Current Opinion in Biotechnology 26:19-24
Oldroyd GE, Downie JA (2008) Coordinating nodule morphogenesis with rhizobial infection in legumes. Annu Rev Plant Biol 59:519-546
Oldroyd GE, Murray JD, Poole PS, Downie JA (2011) The rules of engagement in the legume-rhizobial symbiosis. Annual Review of Genetics 45:119-144
Op den Camp R, Streng A, De Mita S, Cao Q, Polone E, Liu W, Ammiraju JS, Kudrna D, Wing R, Untergasser A (2011a) LysM-type mycorrhizal receptor recruited for rhizobium symbiosis in nonlegume *Parasponia*. Science 331 (6019):909-912
Op den Camp RH, De Mita S, Lillo A, Cao Q, Limpens E, Bisseling T, Geurts R (2011b) A phylogenetic strategy based on a legume-specific whole genome duplication yields symbiotic cytokinin type-A response regulators. Plant Physiology 157 (4):2013-2022
Orians CM (2000) The effects of hybridization in plants on secondary chemistry: implications for the ecology and evolution of plant–herbivore interactions. American Journal of Botany 87 (12):1749-1756
Oswald BP, Nuismer SL (2007) Neopolyploidy and pathogen resistance. Proceedings of the Royal Society of London B: Biological Sciences 274 (1624):2393-2397
Pandit MK, Pocock MJ, Kunin WE (2011) Ploidy influences rarity and invasiveness in plants. Journal of Ecology 99 (5):1108-1115
Pandit MK, White SM, Pocock MJ (2014) The contrasting effects of genome size, chromosome number and ploidy level on plant invasiveness: a global analysis. New Phytologist 203 (2):697-703
Parker MA, Malek W, Parker IM (2006) Growth of an invasive legume is symbiont limited in newly occupied habitats. Diversity and Distributions 12 (5):563-571
Paterson AH, Chapman BA, Kissinger JC, Bowers JE, Feltus FA, Estill JC (2006) Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. Trends in Genetics 22 (11):597-602
Pawlowski K, Sprent JI (2007) Comparison between actinorhizal and legume symbiosis. In: Nitrogen-fixing actinorhizal symbioses. Springer, pp 261-288
Peck MC, Fisher RF, Long SR (2006) Diverse flavonoids stimulate NodD1 binding to *nod* gene promoters in *Sinorhizobium meliloti*. Journal of Bacteriology 188 (15):5417-5427
Penmetsa RV, Cook DR (1997) A legume ethylene-insensitive mutant hyperinfected by its rhizobial symbiont. Science 275 (5299):527-530
Penmetsa RV, Uribe P, Anderson J, Lichtenzveig J, Gish JC, Nam YW, Engstrom E, Xu K, Sckisel G, Pereira M (2008) The *Medicago truncatula* ortholog of Arabidopsis EIN2, *sickle*, is a negative regulator of symbiotic and pathogenic microbial associations. The Plant Journal 55 (4):580-595
Petit C, Thompson JD (1999) Species diversity and ecological range in relation to ploidy level in the flora of the Pyrenees. Evolutionary Ecology 13 (1):45-65

Petrov VD, Van Breusegem F (2012) Hydrogen peroxide—a central hub for information flow in plant cells. AoB Plants 2012:pls014
Phillips DA, Joseph CM, Maxwell CA (1992) Trigonelline and stachydrine released from alfalfa seeds activate NodD2 protein in Rhizobium meliloti. Plant Physiology 99 (4):1526-1531
Pohlert T (2014) The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR). R Package. http://CRAN.R-project.org/package=PMCMR
Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, Genschik P (2003) EIN3-dependent regulation of plant ethylene hormone signaling by two *Arabidopsis* F box proteins: EBF1 and EBF2. Cell 115 (6):679-689
Powell AF, Doyle JJ (2015) The implications of polyploidy for the evolution of signalling in rhizobial nodulation symbiosis. Advances in Botanical Research 75:149-190
Powell AF, Doyle JJ (2016) Enhanced rhizobial symbiotic capacity in an allopolyploid species of *Glycine* (Leguminosae). American Journal of Botany 103 (10):1771-1782
Prayitno J, Rolfe BG, Mathesius U (2006) The ethylene-insensitive sickle mutant of *Medicago truncatula* shows altered auxin transport regulation during nodulation. Plant Physiology 142 (1):168-180
Price N, RelicÂ B, Talmont F, Lewin A, Promé D, Pueppke S, Maillet F, Dénarié J, Promé JC, Broughton W (1992) Broad-host-range *Rhizobium* species strain NGR234 secretes a family of carbamoylated, and fucosylated, nodulation signals that are O-acetylated or sulphated. Molecular Microbiology 6 (23):3575-3584
Price NP, Talmont F, Wieruszeski J-M, Promé D, Promé J-C (1996) Structural determination of symbiotic nodulation factors from the broad host-range *Rhizobium* species NGR234. Carbohydrate Research 289:115-136
Pueppke S (1988) Nodulating associations among rhizobia and legumes of the genus *Glycine* subgenus *Glycine*. Plant and Soil 109 (2):189-193
Pueppke SG, Bolaños-Vásquez MC, Werner D, Bec-Ferté M-P, Promé J-C, Krishnan HB (1998) Release of flavonoids by the soybean cultivars McCall and Peking and their perception as signals by the nitrogen-fixing symbiont *Sinorhizobium fredii*. Plant Physiology 117 (2):599-606
Pueppke SG, Broughton WJ (1999) *Rhizobium* sp. strain NGR234 and *R. fredii* USDA257 share exceptionally broad, nested host ranges. Molecular Plant-Microbe Interactions 12 (4):293-318
Radutoiu S, Madsen LH, Madsen EB, Felle HH, Umehara Y, Grønlund M, Sato S, Nakamura Y, Tabata S, Sandal N (2003) Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. Nature 425 (6958):585-592
Radutoiu S, Madsen LH, Madsen EB, Jurkiewicz A, Fukai E, Quistgaard EM, Albrektsen AS, James EK, Thirup S, Stougaard J (2007) LysM domains mediate lipochitin–oligosaccharide recognition and *Nfr* genes extend the symbiotic host range. The EMBO Journal 26 (17):3923-3935
Ramsey J, Ramsey TS (2014) Ecological studies of polyploidy in the 100 years following its discovery. Phil Trans R Soc B 369 (1648):20130352
Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics:589-639
Ranganathan Y, Borges R (2010) Reducing the babel in plant volatile communication: using the forest to see the trees. Plant Biology 12 (5):735-742

Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. BMC Biology 7 (1):1
Reay DS, Davidson EA, Smith KA, Smith P, Melillo JM, Dentener F, Crutzen PJ (2012) Global agriculture and nitrous oxide emissions. Nature Climate Change 2 (6):410-416
Relić B, Perret X, Estrada-García MT, Kopcinska J, Golinowski W, Krishnan H, Pueppke S, Broughton W (1994) Nod factors of *Rhizobium* are a key to the legume door. Molecular Microbiology 13 (1):171-179
Remy W, Taylor TN, Hass H, Kerp H (1994) Four hundred-million-year-old vesicular arbuscular mycorrhizae. Proceedings of the National Academy of Sciences 91 (25):11841-11843
Ried MK, Antolín-Llovera M, Parniske M (2014) Spontaneous symbiotic reprogramming of plant roots triggered by receptor-like kinases. Elife 3:e03891
Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26 (1):139-140
Roux B, Rodde N, Jardinaud MF, Timmers T, Sauviac L, Cottret L, Carrère S, Sallet E, Courcelle E, Moreau S (2014) An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. The Plant Journal 77 (6):817-837
Sachs J, Ehinger M, Simms E (2010) Origins of cheating and loss of symbiosis in wild Bradyrhizobium. Journal of Evolutionary Biology 23 (5):1075-1089
Sakakibara H, Takei K, Hirose N (2006) Interactions between nitrogen and cytokinin in the regulation of metabolism and development. Trends in Plant Science 11 (9):440-448
Sasaki T, Suzaki T, Soyano T, Kojima M, Sakakibara H, Kawaguchi M (2014) Shoot-derived cytokinins systemically regulate root nodulation. Nature Communications 5
Schauser L, Roussis A, Stiller J, Stougaard J (1999) A plant regulator controlling development of symbiotic root nodules. Nature 402 (6758):191-195
Schlaepfer DR, Edwards PJ, Billeter R (2010) Why only tetraploid *Solidago gigantea* (Asteraceae) became invasive: a common garden comparison of ploidy levels. Oecologia 163 (3):661-673
Schmidt JS, Harper JE, Hoffman TK, Bent AF (1999) Regulation of soybean nodulation independent of ethylene signaling. Plant Physiology 119 (3):951-960
Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J (2010) Genome sequence of the palaeopolyploid soybean. Nature 463 (7278):178-183
Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences 108 (10):4069-4074
Schwarze P (1959) Untersuchungen über die gesteigerte Flavonoidproduktion in *Phaseolus*-Artbastarden (*Phaseolus vulgaris* x *Phaseolus coccineus*). Planta 54 (2):152-161
Searle IR, Men AE, Laniya TS, Buzas DM, Iturbe-Ormaetxe I, Carroll BJ, Gresshoff PM (2003) Long-distance signaling in nodulation directed by a CLAVATA1-like receptor kinase. Science 299 (5603):109-112
Segraves K, Thompson J (1999) Plant polyploidy and pollination: floral traits and insect visits to diploid and tetraploid *Heuchera grossulariifolia*. Evolution:1114-1127
Shaw SL, Long SR (2003) Nod factor elicits two separable calcium responses in *Medicago truncatula* root hair cells. Plant Physiology 131 (3):976-984

Sherman-Broyles S, Bombarely A, Powell AF, Doyle JL, Egan AN, Coate JE, Doyle JJ (2014) The wild side of a major crop: Soybean's perennial cousins from Down Under. American Journal of Botany 101 (10):1651-1665
Shubin N, Tabin C, Carroll S (2009) Deep homology and the origins of evolutionary novelty. Nature 457 (7231):818-823
Soltis DE, Buggs RJ, Doyle JJ, Soltis PS (2010) What we still don't know about polyploidy. Taxon 59 (5):1387-1403
Soltis DE, Soltis PS, Morgan DR, Swensen SM, Mullin BC, Dowd JM, Martin PG (1995) Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. Proceedings of the National Academy of Sciences 92 (7):2647-2651
Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annual Review of Plant Biology 60:561-588
Somasegaran P, Hoben HJ (1994) Handbook for rhizobia: Methods in legume-*Rhizobium* technology. Springer-Verlag, New York
Soyano T, Hayashi M (2014) Transcriptional networks leading to symbiotic nodule organogenesis. Current Opinion in Plant Biology 20:146-154
Sprent J (2009) Legume nodulation: A global perspective. Wiley-Blackwell, Chichester
Sprent J (2012) From pure to applied: legume research reports from Down Under. New Phytologist 194 (2):318-320
Sprent JI (2007) Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. New Phytologist 174 (1):11-25
Stalker H, Nickum M, Wynne J, Elkan G, Schneeweis T (1994) Evaluation of biological nitrogen fixation capacity in *Arachis* species and the possible role of polyploidy. Peanut Science 21 (1):55-60
Stanton-Geddes J, Anderson CG (2011) Does a facultative mutualism limit species range expansion? Oecologia 167 (1):149-155
Stebbins GL (1985) Polyploidy, hybridization, and the invasion of new habitats. Annals of the Missouri Botanical Garden:824-832
Steinmann T, Geldner N, Grebe M, Mangold S, Jackson CL, Paris S, Gälweiler L, Palme K, Jürgens G (1999) Coordinated polar localization of auxin efflux carrier PIN1 by GNOM ARF GEF. Science 286 (5438):316-318
Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T, Tabata S, Sandal N, Stougaard J, Szczyglowski K (2002) A plant receptor-like kinase required for both bacterial and fungal symbiosis. Nature 417 (6892):959-962
Subramanian S, Hu X, Lu G, Odelland JT, Yu O (2004) The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. Plant Molecular Biology 54 (5):623-639
Subramanian S, Stacey G, Yu O (2006) Endogenous isoflavones are essential for the establishment of symbiosis between soybean and *Bradyrhizobium japonicum*. The Plant Journal 48 (2):261-273
Sudová R, Pánková H, Rydlová J, Münzbergová Z, Suda J (2014) Intraspecific ploidy variation: A hidden, minor player in plant–soil–mycorrhizal fungi interactions. American Journal of Botany 101 (1):26-33

Sudová R, Rydlová J, Münzbergová Z, Suda J (2010) Ploidy-specific interactions of three host plants with arbuscular mycorrhizal fungi: Does genome copy number matter? American Journal of Botany 97 (11):1798-1807
Sugiyama A, Shitan N, Yazaki K (2007) Involvement of a soybean ATP-binding cassette-type transporter in the secretion of genistein, a signal flavonoid in legume-*Rhizobium* symbiosis. Plant Physiology 144 (4):2000-2008
Sutton MA, Bleeker A (2013) Environmental science: the shape of nitrogen to come. Nature 494 (7438):435-437
Sutton MA, Oenema O, Erisman JW, Leip A, van Grinsven H, Winiwarter W (2011) Too much of a good thing. Nature 472 (7342):159-161
Takei K, Sakakibara H, Taniguchi M, Sugiyama T (2001) Nitrogen-dependent accumulation of cytokinins in root and the translocation to leaf: implication of cytokinin species that induces gene expression of maize response regulator. Plant and Cell Physiology 42 (1):85-93
Takei K, Ueda N, Aoki K, Kuromori T, Hirayama T, Shinozaki K, Yamaya T, Sakakibara H (2004) *AtIPT3* is a key determinant of nitrate-dependent cytokinin biosynthesis in *Arabidopsis*. Plant and Cell Physiology 45 (8):1053-1062
Taylor TN, Remy W, Hass H, Kerp H (1995) Fossil arbuscular mycorrhizae from the Early Devonian. Mycologia:560-573
te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P (2012) The more the better? The role of polyploidy in facilitating plant invasions. Annals of Botany:mcr277
Těšitelová T, Jersáková J, Roy M, Kubátová B, Těšitel J, Urfus T, Trávníček P, Suda J (2013) Ploidy-specific symbiotic interactions: divergence of mycorrhizal fungi between cytotypes of the *Gymnadenia conopsea* group (Orchidaceae). New Phytologist 199 (4):1022-1033
Theodoridis S, Randin C, Broennimann O, Patsiou T, Conti E (2013) Divergent and narrower climatic niches characterize polyploid species of European primroses in *Primula* sect. *Aleuritia*. Journal of Biogeography 40 (7):1278-1289
Thompson JN, Merg KF (2008) Evolution of polyploidy and the diversification of plant–pollinator interactions. Ecology 89 (8):2197-2206
Thompson JN, Nuismer SL, Merg K (2004) Plant polyploidy and the evolutionary ecology of plant/animal interactions. Biological Journal of the Linnean Society 82 (4):511-519
Tirichine L, Sandal N, Madsen LH, Radutoiu S, Albrektsen AS, Sato S, Asamizu E, Tabata S, Stougaard J (2007) A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. Science 315 (5808):104-107
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols 7 (3):562-578
Trinick M (1980) Relationships amongst the fast-growing *Rhizobia* of *Lablab purpureus*, *Leucaena leucocephala*, *Mimosa* spp., *Acacia farnesiana* and *Sesbania grandiflora* and their affinities with other rhizobial groups. Journal of Applied Bacteriology 49 (1):39-53
van Hameren B, Hayashi S, Gresshoff PM, Ferguson BJ (2013) Advances in the identification of novel factors required in soybean nodulation, a process critical to sustainable agriculture and food security. Journal of Plant Biology and Soil Health 1 (1)

Vance CP (1998) Legume symbiotic nitrogen fixation: agronomic aspects. In: The rhizobiaceae. Springer, pp 509-530
Vance CP (2001) Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. Plant Physiology 127 (2):390-397
Vanneste K, Maere S, Van de Peer Y (2014) Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. Phil Trans R Soc B 369 (1648):20130353
Violle C, Navas ML, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E (2007) Let the concept of trait be functional! Oikos 116 (5):882-892
Visger CJ, Germain-Aubrey CC, Patel M, Sessa EB, Soltis PS, Soltis DE (2016) Niche divergence between diploid and autotetraploid *Tolmiea*. American Journal of Botany 103 (8):1396-1406
Walker SA, Downie JA (2000) Entry of *Rhizobium leguminosarum* bv. *viciae* into root hairs requires minimal Nod factor specificity, but subsequent infection thread growth requires *nodO* or *nodE*. Molecular Plant-Microbe Interactions 13 (7):754-762
Wang J, Tian L, Madlung A, Lee H-S, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004) Stochastic and epigenetic changes of gene expression in *Arabidopsis* polyploids. Genetics 167 (4):1961-1973
Washburn JD, Birchler JA (2014) Polyploids as a "model system" for the study of heterosis. Plant Reproduction 27 (1):1-5
Webb BA, Hildreth S, Helm RF, Scharf BE (2014) Sinorhizobium meliloti chemoreceptor McpU mediates chemotaxis toward host plant exudates through direct proline sensing. Applied and Environmental Microbiology 80 (11):3404-3415
Weiher E, Werf A, Thompson K, Roderick M, Garnier E, Eriksson O (1999) Challenging Theophrastus: a common core list of plant traits for functional ecology. Journal of Vegetation Science 10 (5):609-620
Weir J (1961) A comparison of the nodulation of diploid and tetraploid varieties of red clover inoculated with different rhizobial strains. Plant and Soil 14 (1):85-89
Weiss-Schneeweiss H, Emadzade K, Jang T-S, Schneeweiss G (2013) Evolutionary consequences, constraints and potential of polyploidy in plants. Cytogenetic and Genome Research 140 (2-4):137-150
Wendel J, Doyle J (2005) Polyploidy and evolution in plants. In: Henry RJ (ed) Plant Diversity and Evolution. CABI Publishing, Wallingford
Wendel JF (2015) The wondrous cycles of polyploidy in plants. American Journal of Botany 102 (11):1753-1756
Werner GD, Cornwell WK, Sprent JI, Kattge J, Kiers ET (2014) A single evolutionary innovation drives the deep evolution of symbiotic N_2-fixation in angiosperms. Nature Communications 5
White J, Prell J, James EK, Poole P (2007) Nutrient sharing between symbionts. Plant Physiology 144 (2):604-614
Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer, New York
Winiwarter W, Erisman JW, Galloway JN, Klimont Z, Sutton MA (2013) Estimating environmentally relevant fixed nitrogen demand in the 21st century. Climatic Change 120 (4):889-901

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences 106 (33):13875-13879
Yan K, Wu C, Zhang L, Chen X (2015) Contrasting photosynthesis and photoinhibition in tetraploid and its autodiploid honeysuckle (*Lonicera japonica* Thunb.) under salt stress. Frontiers in Plant Science 6
Yang S, Tang F, Gao M, Krishnan HB, Zhu H (2010) *R* gene-controlled host specificity in the legume–rhizobia symbiosis. Proceedings of the National Academy of Sciences 107 (43):18735-18740
Yokota K, Hayashi M (2011) Function and evolution of nodulation genes in legumes. Cellular and Molecular Life Sciences 68 (8):1341-1351
Yokoyama T (2008) Flavonoid-responsive *nodY-lacZ* expression in three phylogenetically different *Bradyrhizobium* groups. Canadian Journal of Microbiology 54 (5):401-410
Yoo M-J, Liu X, Pires JC, Soltis PS, Soltis DE (2014) Nonadditive gene expression in polyploids. Annual Review of Genetics 48:485-517
Yoo M, Szadkowski E, Wendel J (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110 (2):171-180
Yost CK, Rochepeau P, Hynes MF (1998) *Rhizobium leguminosarum* contains a group of genes that appear to code for methyl-accepting chemotaxis proteins. Microbiology 144 (7):1945-1956
Young ND, Debellé F, Oldroyd GE, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KF, Gouzy J, Schoof H (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. Nature 480 (7378):520-524
Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. Statistical Applications in Genetics and Molecular Biology 4 (1):1128
Zhang M, Wu Y-H, Lee M-K, Liu Y-H, Rong Y, Santos TS, Wu C, Xie F, Nelson RL, Zhang H-B (2010) Numbers of genes in the NBS and RLK families vary by more than four-fold within a plant species and are regulated by multiple factors. Nucleic Acids Research 38 (19):6513-6525
Zhang Z, Fu Y, Ma J, Zhang C, Wang P (2014) Isolation and characterization of soybean chalcone reductase cDNA, which encodes the key enzyme for the biosynthesis of 4, 2′, 4′-trihydroxychalcone in legumes. Molecular Breeding 34 (4):2139-2149
Zhao J, Dixon RA (2009) MATE transporters facilitate vacuolar uptake of epicatechin 3′-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. The Plant Cell 21 (8):2323-2340
Zhao J, Huhman D, Shadle G, He X-Z, Sumner LW, Tang Y, Dixon RA (2011) MATE2 mediates vacuolar sequestration of flavonoid glycosides and glycoside malonates in *Medicago truncatula*. The Plant Cell 23 (4):1536-1555