Characterization of the Transcription Factor and Transcriptionally Active Protein Gene Families in Cowpea (*Vigna unguiculata* L.)

Vikram Alexander Misra
Columbia, Maryland

B.S., B.A. University of Maryland, Baltimore County, 2007

A Dissertation presented to the Graduate Faculty of the University of Virginia
in Candidacy for the Degree of Doctor of Philosophy

Department of Biology

University of Virginia
December, 2013

Abstract

Cowpea, *Vigna unguiculata* (L.) Walp., is an important food and forage legume in the semi-arid tropics. Among the greatest constraints of crop yield in cowpea are root-parasitic angiosperms such as *Striga gesnerioides*, which drain the cowpea plant of much needed nutrients. Thus, improved resistance to parasitic weeds, among other stresses, is among the most important goals of cowpea breeding programs. Plant response to such biotic and abiotic stresses involves complex regulatory networks; transcription factors (TFs) play a central role in these networks. Here, I have identified and classified the TF repertoire of cowpea, and demonstrated that some TF families have members whose expression changes when cowpea shows resistance to *Striga* parasitism, thereby demonstrating that these TFs have a role in cowpea resistance against parasitism. Moreover, I have demonstrated that these same TF families can have members whose expression is changed when cowpea shows susceptibility. In addition, to identify the role of sugar transporters in host-parasite interactions, I identified and classified sugar transporters in the genomes of three parasitic angiosperms and determined that in general, sugar transporter expression increases throughout the life cycles of parasitic weeds. Further, I discuss a possible ramification of the characterization of cowpea TFs, namely the possibility that cowpea and *Striga gesnerioides* are competing for sugars and that cowpea TFs may be involved with this competition.

Acknowledgments

This work would never have been realized without the much-needed communication between me and several faculty members and fellow graduate students. First and foremost, my graduate advisor, Professor Michael P. Timko, always ensured that I would grow as a scientist, giving me the advice and encouragement that enabled me to succeed. I will always appreciate that the questions he asked made me think actively about my work. I will always be proud to have studied in his laboratory. The Timko lab was a sanctuary for me, for I worked with so many wonderful people. From Paul Rushton and Xianfeng Chen, to Bhavani Gowda and Marta Sears, to Joseph Opoku and Kan Huang, from Karolina Mellor and Hai Liu, to Monica Fernandez-Aparicio Ruiz, Yu Wang and Tatyana Kotova, everyone made my time in the Timko lab a wonderful experience.

I would like to thank the researchers from other labs who worked with and communicated with our lab. Aaron Mackey, Mark Lawson and Stephen Turner from the Bioinformatics Core were quite enthusiastic and always more than happy to communicate and to help. I would also like to thank Deborah Triant from the Taylor Lab, with whom I would discuss bioinformatics. These discussions were very helpful, and Deb was very kind when engaging in these discussions.

I am also proud to have worked with my committee members, past and present. I will never forget that Lei Li, Martin Wu, Herman Wijnen, Stefan Bekiranov and Ben Blackmann all helped me in a major way, challenging me with questions that drove me to think about the approaches that I took with my research, and giving me valuable advice.

Dedication

This work would never have been possible without my father, Sunil Kumar Misra, who sacrificed every day and every night to make sure I did everything needed and more to become a scholar and a good human being.

This work is also dedicated to my mother, Lisa Bockmiller Misra, who encouraged me incessantly.

This work is also in memory of my grandmother, Chandra Kala Misra, to whom I promised that this day would come.

This work celebrates the loving memory of my grandmother, Chandra Kala Misra, who would have been extremely happy on this occasion.

My grandfather, Dwarika Nath Misra, has a pioneering spirit and has led a life of sacrifice so that we as a family could rise.

Table of Contents

List of Figures

List of Tables

Chapter one

General Introduction

Cowpea, *Vigna unguiculata* (L.) Walp., is an important food and forage warm-season legume that grows in semi-arid tropical regions, where it provides a major source of dietary protein (Singh, 2005; Timko et al., 2007a). Cowpea crop yields are reduced by many stresses, especially parasitic weeds (Timko et al., 2007a; Alonge et al., 2005a; Alonge et al., 2005b; Cardwell et al., 1995). Despite its importance, however, cowpea has not been researched as much as other plants (Timko & Singh, 2008). It is therefore important that cowpea be compared at the genomic level to other legumes. Such comparative analyses are important when investigating how a plant defends itself against the barrage of stresses and pests it faces in its native environment. Some of the more important investigations into plant defense have involved the study of transcription factors, proteins that regulate genetic expression (MacQuarrie et al., 2011) and often lie at the crossroads of many important signaling pathways (Broun et al., 2004; Lindemose et al., 2013), including a variety of defense pathways that plants use (Gfeller et al., 2010; Pieterse et al., 2012).

History and Uses of Cowpea

Much of cowpea evolution occurred in Africa, where there are two centers of diversity: one in the eastern and southern regions, and the other in the western and central regions (Baudoin & Marechal, 1985; Huynh et al., 2013). Cowpea may have been introduced to India in the Neolithic period (Pant et al., 1982) and was likely in southern Europe since the 8th century BC (Tosti & Negri 2002; Ng & Marechal 1985). Cowpea was most likely introduced to the New World when Spanish explorers arrived in the Americas in the 16th century (Purseglove, 1968).

Today, though it is also grown in India, southern Europe, and in the Americas, most cowpea production takes place in West and Central Africa, mostly by poor subsistence farmers (Singh, 2005; Timko et al., 2007a).

Cowpea is important for its nutritional content; it is high in protein and carbohydrates (Nielsen et al.,1993; Hall et al., 2003). Cowpea leaves are also eaten (Nielsen et al. (1997)), and they are shown to be high in phosphorus, ascorbic acid and protein, especially when dried (Ahenkora et al., 1998). Furthermore, its stems are shown to be an effective fodder for livestock (Singh et al., 2003), with a crude protein content of 13% to 17% (Tarawali et al., 1997). Moreover, cowpea as a crop can restore nitrogen to soils (Elowad & Hall, 1987), and can be effective as a companion crop to cereals (Oseni (2010)).

Besides its nutritional content, cowpea is grown because it is tolerant of harsh conditions, such as drought and poor soil quality (Hall et al. 2002; Hall 2004).

Taxonomy and Evolutionary Origin of Cowpea

Taxonomy

Cowpea, a dicotyledonous plant, belongs to the Fabaceae family (syn. Leguminoseae), in the subfamily Papilionoideae (syn. Faboideae), tribe Phaseolae, in the subtribe Phaseolinae; in the genus *Vigna*, the cowpea lies in the Catiang section (Verdcourt, 1970; Marechal et al., 1978). The *Vigna* genus is quite diverse; besides cowpea, the blackgram (*Vigna mungo*), adzuki bean (*Vigna angularis*), and groundnut (*Vigna subterranea*) are also members of the *Vigna* genus (Timko & Singh, 2008; Ajibade et al., 2000). Even within the species *Vigna unguiculata* are several subspecies, including yardlong bean (ssp. *sesquipedalis*), wild cowpea (ssp.

dekindtiana) and the southern pea (ssp. *unguiculata*) (Ng & Marechal, 1985; Ajibade et al., 2000).

Evolutionary History of Legumes

Using Bayesian phylogenetic analysis, Lavin et al. (2005) found that almost 60 million years ago (Mya), the Fabaceae family diverged into three subfamilies: the Papilionoideae, the Mimosoideae and the Caesalpinoideae. Most cultivated legumes are in two clades of the Papilionoideae: the Galegoids (cool-season legumes), a clade which is estimated to have originated approximately 51 Mya (Lavin et al., 2005) and includes lentil (Lens sp.) and alfalfa (*Medicago sativa*); or the Phaseoloids (syn. Millettoids) (warm-season legumes), which includes soybean (*Glycine max*), common bean (*Phaseolus vulgaris*) and cowpea (*Vigna unguiculata*) (Young & Bharti, 2012). The Phaseoloid clade is estimated to have originated approximately 45 Mya (Lavin et al. 2005). A summarizing schematic of the phylogenetic relationship between examples of Galegoid and Phaseoloid legumes is shown in Figure 1.

Knowing this evolutionary history of legumes is important for comparative analyses because organisms in the same family share homologous genes and gene order; in otherwords, there is greater synteny (more specifically, a greater conserved synteny) between closely related organisms (Abrouk et al., 2010). For example, a strong synteny exists between cowpea and its fellow warm-season legume soybean (Lucas et al., 2013b), whereas a weak synteny exists between cowpea and the non-legume *Arabidopsis thaliana* (Pottorff et al. 2012). Therefore, what is learned about one legume may inform what may exist in another legume.

Cowpea Domestication

There is disagreement on the location where cowpea was first domesticated (Timko & Singh, 2008a). Vaillancourt & Weeden (1992) analyzed polymorphisms in chloroplast DNA and proposed that Nigeria was the origin of cowpea domestication. However, Coulibaly et al. (2002) analyzed amplified fragment length polymorphism (AFLP) profiles to suggest that cowpea was first domesticated in the northeastern region of Africa. Ba et al. (2004) used analysis of random amplified polymorphic DNA (RAPD) to propose that domesticated cowpea is more closely related to the varieties of wild cowpea (*Vigna unguiculata* ssp. unguiculata var. spontanea) from West Africa than to wild cowpea from the Eastern or southern regions of Africa. From this, Ba et al. (2004) suggest that domesticated cowpea arose from wild populations in the northern part of Africa.

Goals of breeding programs. Cowpea breeding has the ultimate goal of maximizing cowpea grain yield and quality (Timko et al., 2007a). In the process, cowpea breeding programs aim to breed for resistance to the myriad stresses and pests that can harm the cowpea plant (Timko et al., 2007a). Cowpea must endure stresses such as soil salinity (Sobhanian et al., 2011) and toxic metals (Kopittke et al., 2007; Kopittke et al., 2011), and can be devastated by a host of bacterial, viral, and fungal diseases, insects, and nematodes (Timko et al., 2007b; Singh 2005; Roberts et al. 1996, 1997; Das et al. 2010). Some of the most destructive threats to cowpea are parasitic weeds. These parasitic weeds, such as *Striga gesnerioides* (Timko et al., 2007b), can cause significant losses in crop yield (Alonge et al., 2005a;

Alonge et al., 2005b), and in some cases, up to 100% crop yield loss (Cardwell & Lane, 1995).

Despite the importance of the cowpea, the cowpea has received relatively little attention from researchers (Timko & Singh, 2008). It is therefore important that researchers use genomics and bioinformatics to compare cowpea to other legumes.

Model Legumes and their Relationship to Cowpea

Research in the model legumes species *Lotus japonicus*, *Glycine max* and *Medicago truncatula* (barrel medic) was pivotal in understanding other legumes such as common bean and cowpea; synteny analysis is particularly helpful as it can be used to gauge how well two legumes can be compared.

Lotus japonicus

Lotus japonicus is a wild legume that belongs to the Galegeae clade (cool-season legumes) of the Fabaceae (Young & Bharti, 2012). This legume is useful because of its diploidy (n=6), its small genome size (472.1 megabases (Mb)) (Ito et al. 2000) and its short life cycle (2-3 months) (Sato & Tabata, 2006). In early genomic analyses of *Lotus japonicus*, expressed sequence tags (ESTs) were created from cDNA libraries (Asamizu et al. 2000). In other projects, sequencing in *Lotus japonicus* was done using a combination of bacterial artificial chromosomes (BACs) and transformation-competent artificial chromosomes (TACs) (Liu et al. 1999; Nakamura et al. 2002; Kaneko et al. 2003; Asamizu et al. 2003; Kato et al. 2003). In 2008, the Lotus japonicus genome was published (Sato et al. 2008). At that time, the sequences determined represented two thirds of the genome (315 Mb), and linkage mapping at that time could only anchor 130 Mb to chromosomes (Sato et al.

2008). Today, there exists a newer genome assembly for *Lotus japonicus* (ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r2.5/ pseudomolecule/), which contains 268 Mb of anchored pseudomolecules (Young & Bharti, 2012).

Lotus japonicus is shown to be highly syntenic with asparagus bean (*Vigna unguiculata* ssp. Sesquipedialis), a subspecies of cowpea (Xu et al. 2011); therefore, it is possible to use *Lotus japonicus* as a model legume when doing a comparative analysis with cowpea.

Medicago truncatula

Medicago truncatula (barrel medic) belongs to the Galegeae clade, and is closely related to alfalfa (*Medicago sativa*) (Young & Bharti, 2012). *Medicago truncatula* is a desirable model legume for many reasons. It has a small genome size (500-550 Mb), and simple diploidy (Cook 1999, Choi et al. 2004). Moreover, *Medicago truncatula* is prolific in seed production, its generation time is short, and its transformation happens with relative ease (Cook 1999, Young & Udvardi, 2009).

One of the first *Medicago* genome sequencing projects, the Medicago Genome Initiative (MGI), involved the creation of expressed sequence tags (ESTs) (Bell et al. 2001). Other *Medicago* sequencing projects involved the use of BAC libraries (Roe & Kupfer, 2004; Young et al. 2005).

One problem arose when sequencing groups used different automated annotation pipelines on the *Medicago* sequences found (Town, 2006; Spannagl et al., 2007). Several sequencing and informatics centers attempted to generate whole-genome datasets, and in the process produced multiple predicted protein sequence sets and several gene models; this was confusing for the average user (Town, 2006).

Bioinformaticians from the groups that were sequencing and annotating the *Medicago truncatula* genome came together to form the International *Medicago* Genome Annotation Group (IMGAG; www.medicago.org/genome/IMGAG/) (Town, 2006). Afterwards, the IMGAG group created a pipeline for annotating *Medicago* that would be used across several sequencing and informatics centers.

Young et al. (2011) produced the most recent version of the *Medicago* genome to date using a combination of BAC clones and 40x Illumina whole genome shotgun (WGS). This version, named Mt3.5, covers approximately 94% of all genes in the barrel medic, and its high coverage enabled Young et al. (2011) to find high levels of synteny between *Medicago truncatula*, *Lotus japonicus* and *Glycine max*.

Muchero et al. (2009) used EST-derived single nucleotide polymorphisms (SNPs) to find that *Medicago* has 82% macrosynteny with cowpea. Muchero et al. (2011) used the syntenic relationships between the two legumes to study response to drought and to *Macrophomina phaseolina*, a fungal pathogen, and hence identify sources of cowpea resistance. Therefore, the barrel medic can be effectively used as a model legume when doing a comparative analysis with cowpea.

Glycine max

The soybean (*Glycine max*) is a legume in the Phaseoloid clade (warm-season legumes). For years, soybean has been widely studied because of its agricultural importance (Ferguson & Gresshoff, 2009). However, there was a time when soybean was not much of a model legume due to its molecular disadvantages to *Medicago truncatula* and *Lotus japonicus*, including the large genome size of soybean (Ferguson & Gresshoff, 2009) of 1,115 Mb (Arumuganathan & Earle, 1991).

In that time, extensive genomic work on soybean was being done. According to a review by Chan et al., (2012), such work spanned more than ten years, and started with the creation of physical and linkage maps using such methods as combining populations (Cregan et al. 1999; Marek et al. 2001; Song et al. 2004; Choi et al. 2007; Hyten et al. 2010a), genomic BAC and BIBAC libraries (Wu et al. 2004, 2008; Shoemaker et al. 2008), and integrating information from available maps (Hyten et al. 2010b).

A landmark achievement occurred when Schmutz et al. (2010) sequenced the soybean genome by integrating physical and genetic maps with the WGS approach. This not only allowed for more efficient gene discovery in soybean, but also allowed for easier comparative analyses between soybean and other legumes (Ferguson & Gresshoff, 2009).

Afterwards, soybean became an important model legume (Ferguson & Gresshoff, 2009), especially in the study of cowpea. Das et al. (2008) used a soybean genome array to find and validate single feature polymorphisms (SFPs) in cowpea. Moreover, Muchero et al. (2009) found that there exists 85% macrosynteny between soybean and cowpea.

The availability of data on soybean also makes it possible to study cowpea defense against biotic and abiotic stresses. For example, a soybean genome array made it possible for Das et al. (2010) to compare the transcriptome profiles of resistant and susceptible cowpea interactions with root-knot nematode, and to find that in the resistant genotype, more genes were down-regulated and fewer genes were up-regulated than in the susceptible genotype. In another example, Muchero et al.

(2011) used of syntenic relationships between cowpea and soybean (alongside genic SNPs and cowpea synteny with Medicago) to find sources of cowpea resistance to the fungal pathogen *Macrophomina phaseolina*. Moreover, Lucas et al. (2013a) used regions of the soybean genome syntenic with cowpea to find candidate genes for heat tolerance in cowpea. Soybean is thus an indispensable resource as a model legume when doing a comparative analysis with cowpea.

Phaseolus vulgaris

Phaseolus vulgaris (common bean) has great potential to be a model organism (Blair & Ishitani, 2009). It has a relatively small genome and is a diploid organism (2n=22) (Gepts et al. 2008; McClean et al. 2008). Moreover, since it lies in the same tribe as cowpea (Phaseoloids/Millettoids), the common bean has great potential as a model legume for cowpea.

Most of the early genetic maps made for *Phaseolus vulgaris* for a long time were low-density linkage maps that did not cover the genome well (McClean et al. 2008). Several EST collections for common bean were available though (Hernandez at al. 2004; Ramirez et al. 2005; Melotto et al. 2005; McClean et al. 2008).

BAC-end libraries were pivotal in coming to a more complete understanding of the *Phaseolus vulgaris* genome. Combinations of such libraries have been used to create more complete physical maps of around 9x – 12x coverage (Gepts et al. 2008; Schlueter et al. 2008; Córdoba et al. 2010).

Today, a genome assembly for the common bean is available on Phytozome (www.phytozome.net/commonbean.php; Goodstein et al. 2012). This assembly was made using 454 reads, paired end reads, a genetic map from over 7000 markers and

BAC and fosmid-end sequences; the coverage is more than twice the coverage in previous genomic maps such as the map in Schlueter et al. (2008) (www.phytozome.net/commonbean.php).

Phaseolus vulgaris has extensive synteny with other legumes such as *Medicago truncatula*, *Lotus japonicus* and *Arachis sp.* (Hougaard et al. 2008; Galeano et al. 2009), although it shares greater synteny with its fellow millettoid soybean (McClean et al. 2010; Galeano et al. 2009). Common bean can thus be an invaluable resource in comparative analyses with cowpea.

Vigna Unguiculata

Cowpea is a diploid legume whose genome is estimated to be approximately 620 Mbp, which is relatively small compared to other legumes (Arumuganathan & Earle, 1991; Timko et al., 2008). Compared to other legumes, not as much research has been done on cowpea. Early genetic work on cowpea was done with the goal of determining relationships between cowpea and closely related legumes (Fatokun et al., 1992, 1993; Kaga et al., 1996; Ajibade et al., 2000), or was done with the goal of finding genetic markers linked to cowpea resistance against *Striga gesnerioides* (Ouedraogo et al., 2001; Ouedraogo et al., 2002). Since a genome assembly was not available at the time, very important aspects of cowpea resistance to parasitic weeds were left unseen.

Genomic resources for cowpea. Only a comprehensive understanding of the cowpea genome could remedy this problem. An important step forward in gaining this comprehensive understanding was when the gene-rich space of cowpea was sequenced and analyzed (Timko et al., 2008). From this data, 3700

microsatellite (SSR) markers were predicted, and 1000 SSRs derived from the gene space sequences (GSS) were mapped (Timko et al., 2008). After this, Xu et al. (2011) mapped 1,801 other GSS-derived SSRs in asparagus bean (*Vigna unguiculata* ssp. *sesquipedalis*), which had an estimated mean density of 0.79 SSR/kb.

Other sequencing efforts have played an instrumental role in providing genomic information on cowpea. Among them were several EST sequencing projects, including more than 300,000 sequences in the NCBI public database, which include sequences by Coetzer et al. (2010), and 17 EST libraries containing 183,118 ESTs representing approximately 30,000 unigenes; sequences from these 17 EST libraries are also made available on HarvEST: Cowpea 1.27 assemblies P07 and P12 (Wanamaker & Close, 2011). From the EST sequences, Ehlers et al. (2012) showed that a high-throughput genotyping platform (Illumina 1536 GoldenGate Assay) produced 1375 successful SNPs. Lucas et al. (2011) used the Illumina 1536 GoldenGate Assay, a high-throughput genotyping platform, to genotype almost 1300 individuals from 13 crossing populations of cowpea to produce a consensus genetic map of cowpea that contained 1,107 EST-derived SNP markers. In addition, Ehlers et al. (2012) use that platform to produce 1375 SNPs, and show that the most dependable mapped SNPs from their Illumina assay, numbering to 1052, were converted to a different assay, the KBiosciences KASPar assay (Chen et al., 2010; (KBiosciences Ltd, Hoddesdon, Hertfordshire, United Kingdom).

The access to cowpea GSS and EST sequences allowed for researchers to perform analyses that required a comprehensive understanding of the cowpea genome, such

as the syntenic relationships between cowpea and related legumes (Timko et al., 2008; Muchero et al., 2009; Muchero et al., 2011).

In addition to GSS and EST sequences, three BAC libraries of cowpea were assembled: one with 6X coverage (MP Timko, UVA), one 10X (DR Cook, UC Davis) and one with 17X coverage (TJ Close & MC Luo, UC Riverside) (Close, 2008). Moreover, two sets of BAC end sequences, one consisting of 50,120 sequences covering 36.7 Mbp of data (DR Cook, UC Davis & NCBI) and one set of 30,000 sequences from UC Riverside (TJ Close & MC Luo). From this, a physical map of the cowpea genotype IT97K-499-35 (M Luo, UC Davis; T Close, UC Riverside) has been made, which represents the DNA fingerprinting of 60,000 BACs from two libraries (HindIII and MboI), and thus an assembly with 18x genome coverage consisting of 73,728 clones, approximately 30,000 BESs with an average size of 150kb (Luo et al., 2003). The final map represents 43,717 BACs with 11x genome coverage, with clones assembled into contigs using fingerprinted contig (FPC) software (Soderlund et al., 2000); this information is available at http://phymap.ucdavis.edu/cowpea/ (Ehlers et al., 2009).

The current version of the whole genome sequence for cowpea, version 0.03 for cowpea genotype IT97K-499-35 (Close et al., 2011), was assembled from several sources: a SOAP de novo assembly with 67X coverage, 39 Gb of Illumina GAII paired-end sequences (70-130) using TrueSeq chemistry, 250,000 GSS sequences with an average length of 609 bases, 30,000 BAC-end sequences with an average length of 673 bases, and BLASTN hits for 97% of EST-derived consensus sequences from assembly P12 from www.harvest-blast.org (Wanamaker & Close, 2011). From the

high coverage of the SOAP de novo assembly to the combination of GSS, EST and BAC end sequences (BES); the cowpea assembly version v0.03 is the most comprehensive cowpea genomic assembly to date (Close et al., 2011).

The cowpea assembly is thus one product of the many advances in genomics and bioinformatics, from more advanced sequencing platforms, to more bioinformatics resources. As a consequence, researchers can use this assembly to do any study on cowpea, including studies on cowpea TFs.

Genomics and Bioinformatics of TFs

The fields of genomics and bioinformatics have greatly advanced in the past two decades; genomes take a short time to sequence, and the functions of genes in an organism can be determined more efficiently. The sharp increase in the number and variety of genomic resources, including TF databases, available is due to the increasing availability and variety of genomic resources (Martinez, 2011; Meyer et al., 2012), which is in turn due to more advanced sequencing platforms such as Roche 454 (Marguiles et al., 2005) and Illumina replacing traditional Sanger sequencing (Jackson et al., 2011; Martinez, 2011).

Genomic and Bioinformatics Resources

In earlier years of genomic research, it was not guaranteed that one would have genomic information for an organism being investigated, so researchers relied more on sequence similarity searches such as BLAST (Altschul et al., 1990) and FASTA (Pearson & Lipman, 1988; Pearson, 1990). In addition, signature databases such as Pfam (Sonnhammer et al., 1997; Punta et al., 2012) and Interpro (Apweiler et al., 2001; Hunter et al., 2009) served as tools to classify protein-coding sequences into

families to facilitate the annotation of genes (reviewed by Martinez, 2011). The genomic databases that were available were few. In plant research, among the only databases available was The Arabidopsis Information Resource (TAIR) (Huala et al., 2001), which, due to the value of *Arabidopsis thaliana* as a model plant (Meinke et al., 1998), proved to be a pivotal resource for researchers.

These resources, along with more advanced sequencing techniques, set the stage not only for other genomes to be sequenced more efficiently and with greater coverage, but also for comparisons between organisms at the genomic level.

Recently, comparative genomic resources such as GreenPhylDB (Conte et al., 2008; Rouard et al., 2011), Phytozome (Goodstein et al., 2012) and PLAZA (Proost et al., 2009; Van Bel et al., 2012), have been created to provide resources such as genome sequences and annotation data. Comparing organisms on a genomic scale greatly facilitates analyses such as predicting the functions of certain genes in an organism, or determining how a certain gene has changed over time using phylogenetic reconstruction (reviewed in Bachhawat, 2006).

Thus, genomic and bioinformatics resources, from signature databases to comparative genomic databases, allow for the efficient discovery of protein-coding gene families (Martinez, 2011); this can lead to the production of dedicated databases, databases with comprehensive data on various types of gene families (Martinez, 2011); TF databases are a prime example.

TF Identification

When most creators of TF databases search for the TF repertoire in an organism, they compare DNA or protein sequences from the genome of a given organism to

sequences with known DNA-binding domains in a database (Martinez, 2011; Vaquerizas et al., 2012; reviewed by Charoensawan et al., 2010a). This is done using genomic resources such as the ones discussed above, and has the goal of computationally identifying TFs (Charoensawan et al., 2010a).

After this, curation of the resulting set of TFs is done with the goal of eliminating genes unlikely to be TFs (Vaquerizas et al., 2012); this improves specificity by differentiating the true positives from the false negatives (Charoensawan et al., 2010a). When done manually, curation can involve literature curation or benchmarking against other databases manually (Charoensawan et al., 2010a). When done automatically, curation involves using automatic pairwise sequence comparison (Charoensawan et al., 2010). After curation, the TF sets are ready for analysis for potential roles in a variety of biological processes, including defense.

TFs and Their Roles in Abiotic and Biotic Stress Response

Defense Systems

In the face of myriad stresses, pests and diseases, plants have developed complex defense systems. Plants can develop constitutive defenses that are active throughout the life of the plant, as well as inducible defenses that are only active when the plant is under stress (Mithofer & Boland, 2012). Constitutive defenses include thorns (Mithofer & Boland, 2012), waxy cuticles (Dominguez et al., 2011) and toxins (reviewed by Wittstock & Gershenzon, 2002). At the genetic level, constitutive defenses appear in the form of resistance genes that work independently of inducible signaling pathways to inhibit the effects of an infection (Carr et al., 2010).

Should these constitutive defenses fail, plants can use various kinds of inducible defenses (reviewed by Mithofer & Boland, 2012; Carr et al., 2010; Pieterse et al., 2012), defenses that are more active when infected or stressed (Carr et al., 2010; Ballare, 2011). Many of these inducible defenses are complex signaling pathways that lead to a response to the pathogen; such signaling pathways include the jasmonic acid, salicylic acid, and ethylene pathways (Pieterse et al., 2012; Pieterse et al., 2009; Pozo et al., 2005; van Loon et al., 2006a; Loake & Grant, 2007). Transcription factors play an integral part of these pathways (reviewed in Van Verk et al., 2009; Gfeller et al., 2010; Pieterse et al., 2012; Gutterson & Reuber, 2004).

TFs

TFs are molecular agents that regulate the transcription of certain genes by interacting with specific DNA sequences (MacQuarrie et al. 2011; Latchman, 1997). TFs have been shown to regulate many types of processes in plants, from floral development (Shore & Sharrocks, 1995) to defense against abiotic and biotic stresses (Singh et al., 2002). It is therefore imperative that researchers understand TFs and their actions (Riechmann & Ratcliffe, 2000; Mitsuda & Ohme-Takagi, 2009). TFs form complex networks, at the transcriptional and at the post-transcriptional levels (Riechmann & Ratcliffe, 2000; Lindemose et al., 2013). Such networks control many processes, including metabolic pathways (Broun, 2004). Within these networks, TFs not only bind to and regulate multiple genes but can also be regulated by a number of factors (Broun, 2004; MacQuarrie et al., 2011). Such factors include other TFs (Vom Endt et al., 2002; Mol et al., 1998), phosphorylation (Gu et al., 2000; Moyano et al., 1996), proteolysis (Gray et al., 2001; Hardtke & Deng, 2000), cis-

acting elements (Izawa et al., 1994; Sakamoto et al., 1996) and epigenetic modifications such as methylation and deacetylation (Ramon-Maiques et al., 2007; Kim et al., 2008; Zeng et al., 2010; Sanchez & Zhou, 2011).

The ability of TFs to regulate genes and be regulated by other factors makes TFs an important component of defense responses in plants. Therefore, to investigate the defense mechanisms of any plant, one must understand TFs, their functions and distributions as part of a comprehensive understanding of such defense mechanisms. Moreover, one must understand TFs and their distributions in plants that are related to the plant being investigated, so as to facilitate an investigation of the functions of any TFs found in the plant being investigated.

TF Evolution

Throughout evolution, the distribution of TFs has changed to fit the needs of any organisms that have arisen. When the plant and animal lineages arose, not only did plants start to have more TFs, but plants also began to have TFs as a greater percentage of their genomes than in animals (reviewed in Riechmann & Ratcliffe, 2000; http://www.arabidopsis.org/; Rubin et al., 2000). TF families in plants expanded at a much higher rate than in animals (Shiu et al., 2005). Shiu et al. (2005) find, for example, that 14 TF families are larger in *Arabidopsis* than in humans where as only 4 TF families are larger in humans. Moreover, plants evolved to have TF families that were either absent or less abundant in animals (Charoensawan et al., 2010b). Several TF families are known to be unique in plants (including AP2-EREBP, ARF, Aux/IAA, B3, C2C2-YABBY, Dof, EIL, GARP superfamily, LFY, NAC, SBP, TCP, and WRKY) (reviewed in Yanagisawa, 1998; Riechmann & Ratcliffe, 2000;

Ohme-Takagi, 2009; Charoensawan et al., 2010b; Yamasaki et al., 2013). Three families, MYB, MADS and bZIP, have been identified as being more abundant in plants (reviewed in Riechmann & Ratcliffe, 2000). Only one TF family, C2H2, was more abundant in animals than plants (Reviewed in Riechmann & Ratcliffe, 2000). Lang et al. (2010) found that a dramatic increase in the abundance and variety of TFs occurred throughout plant evolution. In the transition from algae to land plants approximately 500 Mya (million years ago), the first land plants (or their ancestor) had 21 new transcription-associated proteins (TAPs; i.e., TFs and other transcriptional regulators (TRs)), including 16 TF families; in addition, an expansion of 44 TAP families occurred (Lang et al., 2010). Among the families that arose were NAC and WRKY (Lang et al., 2010; Yamasaki et al., 2013). It has been proposed that NAC and WRKY belong to the same superfamily (Babu et al., 2006; Olsen et al., 2005; Welner et al., 2012) and that the NAC family arose from the WRKYs (Yamasaki et al., 2013).

Lang et al. (2010) found that when vascular plants arose 470 Mya, three TAP families arose and 3 TAP families expanded. After that, the rise of angiosperms approximately 210 Mya saw the expansion of 23 TAP families (including 18 TF families) and the rise of three TF families (Lang et al., 2010). Since then, only 5 TAP families expanded (3 in monocots and 2 in dicots) and only 2 TF families arose in dicots (Lang et al., 2010).

Within the context of legumes, Lang et al. (2010) found no significant difference in TF distribution among legumes, whereas legumes had fewer MADS TFs than plants in the Brassicales order such as *Arabidopsis thaliana.*

These results, and the results of analyses showing the high amount of synteny between legumes and the lower levels of synteny between *Arabidopsis* and legumes (Xu et al., 2011; Muchero et al., 2009; Muchero et al., 2011; Lucas et al., 2013b), will be helpful in comparing the TFs of different legumes (including cowpea), because analogous TFs (with possibly similar functions) in related legumes will be relatively easy to find. The need for such comparisons is one reason why the TF resources that have become available over the years are especially useful.

TF Resources

The methods of TF identification and dataset curation have become much more efficient over the past two decades. According to a review by Charoensawan et al. (2010a), the discovery of TFs and the subsequent creation of TF databases in the early years of genomic research were slow and curation was manual (reviewed in Charoensawan et al., 2010a). Among the earliest TF databases was TRANSFAC, a database of eukaryotic TFs whose information is literature-curated, including its information on TF binding sites and composite elements (Knuppel et al., 1994; Wingender et al., 1996; Matys et al., 2006). For years, the few TF databases that were available were curated manually (reviewed by Charoensawan et al., 2010a). Then in the early 2000s, researchers began to use combinations of automatic and manual curation to create databases such as AtTFDB (now AGRIS) (Davuluri et al., 2003); this advent of the use of automatic curation (albeit alongside manual curation) coincided with a sharp increase in the number and variety of TF databases available (reviewed in Charoensawan et al., 2010a). Thereafter, almost all TF

databases used a combination of automatic and manual curation (reviewed in Charoensawan et al., 2010a).

Of all the TF databases made, only a few of the most recent ones are dedicated to legumes. These databases usually focus on *Lotus japonicus*, *Medicago truncatula* and soybean, including LegumeTFDB (Mochida et al., 2010) and SoyDB (Wang et al., 2010).

A comprehensive TF repertoire in cowpea, on the other hand, has not been produced. The only repertoire of cowpea TFs is on PlantTFDB v2.0 (Zhang et al., 2011), and only 478 TFs on that database are from cowpea (48 families). This data is primarily based on RefSeq data and ESTs, and not on a whole genome assembly (Zhang et al. 2011). Given the importance of cowpea, and given the importance of TFs in plant defense, a cowpea assembly such as version 0.03 discussed above (Close et al., 2011) would provide the information needed for a much more comprehensive analysis of TFs in cowpea. This assembly would lead to a better analysis of TFs involved with defense, and is thus needed to better understand how cowpea responds to parasitic weeds.

TFs and Plant Defense

Numerous families of TFs have been shown to have pivotal roles in defense of plants against a wide array of abiotic and biotic stresses (reviewed in Singh et al., 2002). An understanding of how plants use TFs to defend against stresses can help researchers make a hypothesis about how cowpea may use TFs to defend against stresses such as drought and parasitic weeds.

Some common families involved with stress. In responding to the many stresses in their environment, plants can rely on several TF families. Some of the more common TF families are reviewed here.

AP2-EREBP. The AP2-EREBP TF family (AP2-EREBP standing for APETALA2 – Ethylene-responsive element binding protein (Jofuku et al., 1994; Ohme-Takagi & Shinshi, 1995)) is a large family whose main characteristic is the presence of one or two AP2 DNA binding domains (reviewed by Riechmann & Meyerowitz, 1998). Jofuku et al. (1994) first discovered the AP2 domain in *Arabidopsis thaliana*; the EREBP domains, which Ohme-Takagi & Shinshi (1995) first discovered, were found to be closely related but distinct in sequence (Weigel, 1995).

The AP2-EREBP family has many functions in plants (reviewed in Riechmann & Meyerowitz, 1998); to understand these functions, one can divide the family into the AP2 and EREBP sub-families (Riechmann & Meyerowitz, 1998). The AP2 subfamily mainly functions in controlling floral development (Komaki et al., 1988; Krogan et al., 2012) while the EREBP subfamily primarily functions in defense mechanisms (Zhou et al., 1997).

AP2-EREBP members have been shown to be involved in multiple stress responses in plants. In earlier research of TF roles in plant defense, *Arabidopsis thaliana*, AP2-EREBP TFs have been shown to activate CRT/DRE genes (CRT = C-repeat element, DRE = drought-responsive elements), which confer cold resistance and drought (Liu et al., 1998; Jaglo-Ottosen et al., 1998; Haake et al., 2002). More recently, the up-regulation of soybean AP2-EREBPs in drought conditions has been shown to confer

drought resistance in transgenic experiments (Chen et al., 2007; Zhang et al., 2009; reviewed in Ku et al., 2013).

Activation of some AP2-EREBP TFs has also been shown to enhance resistance to biotic stresses. Overexpression of the soybean GmERF3 in transgenic tobacco plants was shown to increase resistance to the bacterial pathogen *Ralstonia solanacearum* and the fungal pathogen *Alternaria alternata* (Zhang et al., 2009).

NAC. The NAC (NAM/ATAF/CUC2) family consists of a well-conserved 150-amino acid long domain that is usually at the N-terminus, and usually contains a variable transcription regulatory (TR) domain, although several other variations have been found (Olsen et al., 2005; Ernst et al., 2004; Jensen et al., 2010; reviewed by Puranik et al., 2012).

The NAC family consists of TFs with a variety of functions (reviewed in Olsen et al., 2005), from developmental processes (Aida et al., 1997; Sablowski & Meyerowitz, 1998) to lateral root formation (Xie et al., 2000) and defense against abiotic and biotic stresses (Hegedus et al., 2003; reviewed in Olsen et al., 2005).

NAC TFs have a role in responding to a variety of abiotic stresses. In *Arabidopsis*, NAC TFs have been shown to increase resistance to drought (Tran et al., 2004; Wu et al., 2009). In soybean, Pinheiro et al., (2009) found several NAC TFs that were activated by osmotic stress and salinity. In chickpea, Peng et al. (2010) found that the CarNAC1 TF is activated when chickpea plants are faced with multiple stresses, including wounding, drought, salinity and cold.

NAC TFs are also important in responding to biotic stresses. For example, Xia et al. (2010) characterized a NAC TF in wheat, TaNAC4, which was expressed more often

when the wheat plant was in a resistant interaction with a pathotype of stripe rust (*Puccinia striiformis* f. sp. Tritici Westend (PST)) than when the wheat plant was in a susceptible interaction with a different pathotype of stripe rust.

In an interesting example of a TF having a negative role in plant defense, Wu et al. (2009) found that in *Arabidopsis*, overexpression of ATAF1 (a NAC TF) resulted in greater susceptibility to the fungal pathogen *Botrytis cinerea*, possibly because the fungus promotes the production of reactive oxygen species.

WRKY. The distinguishing feature of the WRKY TFs is a highly conserved 60 amino acid domain called the WRKY domain, which is characterized by the conserved N-terminal sequence WRKYGQK followed by a zinc-finger motif (Rushton et al., 1995; Eulgem et al., 2000).

WRKY TFs have been shown to be integral components of abiotic and biotic stress response pathways. Zhou et al. (2008) showed that numerous WRKY TFs increase in expression when the plant responds to different abiotic stresses; 22 WRKY TFs responded to salinity, 24 responded to drought, and 8 responded to low temperatures (Zhou et al., 2008).

WRKY TFs can play interesting roles in defense against biotic stresses. Some WRKY TFs can activate defense responses, whereas other WRKY TFs repress defenses against biotic stresses (reviewed in Eulgem & Somssich, 2007). In *Arabidopsis thaliana*, Hu et al. (2012) found three WRKY TFs whose functions not only overlap, but also are synergistic in activating a basal defense response to the bacterial pathogen *Pseudomonas syringae*. In soybean, Kang et al. (2009) found that the GmWRKY1 TF is induced after pathogenic infection by *Pseudomonas* bacteria.

On the other hand, a few WRKY TFs have been found to repress defense responses to biotic stresses. In *Arabidopsis*, WRKY11 and WRKY17 were found to repress basal defense responses to *Pseudomonas syringae* (Journot-Catalino et al., 2006). Moreover, two WRKY TFs in barley have been shown to repress basal resistance to powdery mildew (Shen et al., 2007; Eckey et al., 2004).

TFs and signaling networks for plant defense. Usually, plant defense mechanisms take the form of complex signaling networks in the host plant that can respond to most pathogens (Eulgem & Somssich, 2007; Katagiri, 2004). Several families of TFs are important components of such signaling networks (Eulgem, 2005; Somssich, 2006), including the jasmonic acid (JA), salicylic acid (SA) and ethylene (ETH) pathways (Gutterson & Reuber, 2004). These pathways are important mechanisms that plants use to respond to multiple stresses such as wounding (Gfeller et al., 2010; Glauser et al., 2008) and bacterial pathogens (reviewed by Pieterse et al., 2012).

In the JA and SA pathways, defense responses can be activated or repressed by different members of the WRKY TF family (Eulgem & Somssich, 2007; Bostock, 2005; Kim et al., 2006). In the ETH pathway, a TF in the EIL family (EIL standing for Ethylene insensitive 3 (EIN3)-like (Guo & Ecker, 2004)) has been shown to target ERF1 (ERF standing for Ethylene response factor), a TF in the AP2-EREBP superfamily (Solano et al., 1998).

Crosstalk. One of the more important functions of TFs is to serve as a means of crosstalk between signaling pathways. ERF TFs are a prime example; they have been shown to integrate the jasmonate and ethylene pathways (Lorenzo et al.,

2003). Moreover, although the SA and ethylene/JA pathways usually act antagonistically to each other (Leslie & Romani, 1988; Spoel et al., 2003; Glazebrook, 2005; Thaler et al., 2012; Derksen et al., 2013), two AP2-EREBP TFs, namely Pti4 and AtERF1, can be induced by both pathways (Gu et al., 2000; Onate-Sanchez & Singh, 2002).

WRKY proteins are another example of TFs that integrate defense pathways. Li et al. (2004) found that an *Arabidopsis* WRKY TF, WRKY70, negatively regulates responses in the JA pathway while positively regulating responses in the SA pathway. Qiu et al. (2007) found that in rice plants undergoing attack by pathogens, OsWRKY13 integrates the JA and SA pathways to facilitate a defense response. Many other TF families, including bHLH (Chinnusamy et al., 2003) and C2H2 (Pereira et al., 2011; Tian et al., 2010), play important roles in defending plants against many pathogens and stresses. Thus one must reach a comprehensive understanding of TFs in order to better understand the roles of TFs in defense. Such an understanding is most efficiently reached using genomic and bioinformatics analyses.

TFs in Cowpea

Within the context of cowpea, unfortunately, such an understanding cannot be effectively reached because the only genomic-wide analyses of TFs were done when Timko et al. (2008) identified 5,888 cowpea GSS sequences as being homologous to TFs and transcription associated factors (TAFs), and when Zhang et al. (2011) used information mostly available on public databases to identify 478 cowpea TFs. As discussed previously, new cowpea genomic resources (Wanamaker & Close, 2011),

including a new and much more complete cowpea assembly (Close et al., 2011), have been produced. Thus a TF repertoire based almost solely on public resources (Zhang et al., 2011) or on GSS sequences (Timko et al., 2008) is entirely insufficient to understand cowpea TFs and their potential roles in such functions as defense against parasitic plants.

Thus, chapter two presents a comprehensive TF repertoire in cowpea and a comparison of selected TF families between cowpea and common bean. Since cowpea and common bean are phylogenetically close (Lavin et al., 2005), it is expected that TF families in cowpea will show significant similarities to common bean. Chapter three explores one possible ramification of research on the cowpea TF repertoire, namely the identification of cowpea TFs possibly involved with defense against the parasitic plant *Striga gesnerioides*.

After investigations of sugar transporters in parasitic plants in chapters four and five, the implications and ramifications of research on the cowpea TF repertoire are discussed.

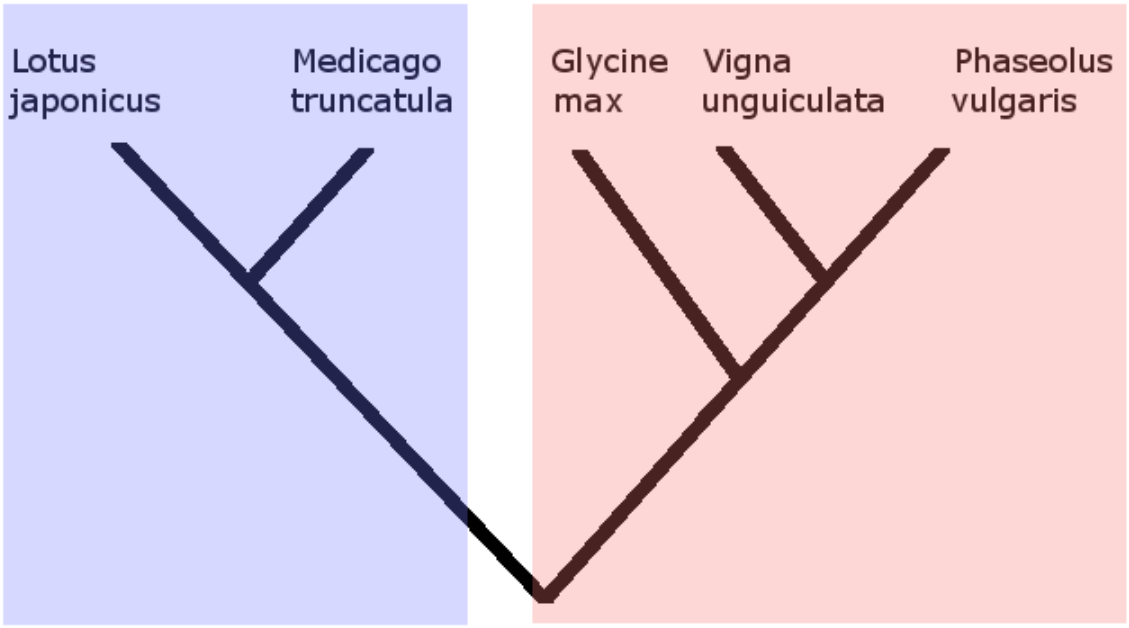


Figure 1. Summarizing schematic of examples of Galegoid (blue) and Phaseoloid (red) legumes. Figure summarized from Figure 3 from Lavin et al. (2005).

References

1. Abrouk, M., Murat, F., Pont, C., Messing, J., Jackson, S., Faraut, T., . . . Salse, J. (2010). Palaeogenomics of plants: synteny-based modelling of extinct ancestors. *Trends in Plant Science, 15*(9), 479-487. doi: http://dx.doi.org/10.1016/j.tplants.2010.06.001

2. Ahenkora, K., Dapaah, H. A., & Agyemang, A. (1998). Selected nutritional components and sensory attributes of cowpea (Vigna unguiculata [L.] Walp) leaves. Plant Foods for Human Nutrition, 52(3), 221-229.

3. Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., & Tasaka, M. (1997). Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. The Plant Cell Online, 9(6), 841-857.

4. Ajibade, S., Weeden, N., & Chite, S. (2000). Inter simple sequence repeat analysis of genetic relationships in the genus Vigna. Euphytica, 111(1), 47-55.

5. Alonge, S. O., Lagoke, S. T. O., & Ajakaiye, C. O. (2005a). Cowpea reactions to Striga gesnerioides I. effect on growth. Crop Protection, 24(6), 565-573. doi: http://dx.doi.org/10.1016/j.cropro.2004.08.014

6. Alonge, S. O., Lagoke, S. T. O., & Ajakaiye, C. O. (2005b). Cowpea reactions to Striga gesnerioides II. Effect on grain yield and nutrient composition. Crop Protection, 24(6), 575-580. doi: http://dx.doi.org/10.1016/j.cropro.2004.08.015

7. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. J Mol Biol, 215(3), 403-410. doi: 10.1016/s0022-2836(05)80360-2

8. Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Birney, E., Biswas, M., . . . Zdobnov, E. M. (2001). The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res, 29(1), 37-40.

9. Arumuganathan, K., & Earle, E. (1991). Nuclear DNA content of some important plant species. Plant molecular biology reporter, 9(3), 208-218.

10. Asamizu, E., Kato, T., Sato, S., Nakamura, Y., Kaneko, T., & Tabata, S. (2003). Structural analysis of a Lotus japonicus genome. IV. Sequence features and mapping of seventy-three TAC clones which cover the 7.5 Mb regions of the genome. DNA research, 10(3), 115-122.

11. Asamizu, E., Nakamura, Y., Sato, S., & Tabata, S. (2000). Generation of 7137 non-redundant expressed sequence tags from a legume, Lotus japonicus. DNA research, 7(2), 127-130.

12. Ba, F. S., Pasquet, R. S., & Gepts, P. (2004). Genetic diversity in cowpea [Vigna unguiculata (L.) Walp.] as revealed by RAPD markers. [Article]. Genetic Resources and Crop Evolution, 51(5), 539-550. doi: Doi 10.1023/B:Gres.0000024158.83190.4e

13. Babu, M. M., Iyer, L. M., Balaji, S., & Aravind, L. (2006). The natural history of the WRKY–GCM1 zinc fingers and the relationship between transcription factors and transposons. Nucleic Acids Res, 34(22), 6505-6520.

14. Bachhawat, A. K. (2006). Comparative genomics. Resonance, 11(8), 22-40. doi: 10.1007/bf02855776

15. Baker, R. F., Leach, K. A., & Braun, D. M. (2012). SWEET as Sugar: New Sucrose Effluxers in Plants. [Article]. Mol Plant, 5(4), 766-768. doi: 10.1093/mp/sss054

16. Ballare, C. L. (2011). Jasmonate-induced defenses: a tale of intelligence, collaborators and rascals. Trends Plant Sci, 16(5), 249-257. doi: 10.1016/j.tplants.2010.12.001

17. Baudoin, J., & Maréchal, R. (1985). Genetic diversity in Vigna. SR Singh and K. 0. Rachie, eds., Cowpea research, production and utilization. Wiley, New York, 3-11.

18. Bell, C. J., Dixon, R. A., Farmer, A. D., Flores, R., Inman, J., Gonzales, R. A., . . . Weller, J. W. (2001). The Medicago Genome Initiative: a model legume database. Nucleic Acids Res, 29(1), 114-117.

19. Blair, M., & Ishitani, M. (2009). Common bean: a model food legume for international agriculture. Grain Legumes, 53(8-9).

20. Boldt, K., Pörs, Y., Haupt, B., Bitterlich, M., Kühn, C., Grimm, B., & Franken, P. (2011). Photochemical processes, carbon assimilation and RNA accumulation of sucrose transporter genes in tomato arbuscular mycorrhiza. J Plant Physiol, 168(11), 1256-1263.

21. Bostock, R. M. (2005). Signal crosstalk and induced resistance: straddling the line between cost and benefit. Annu. Rev. Phytopathol., 43, 545-580.

22. Broun, P. (2004). Transcription factors as tools for metabolic engineering in plants. Curr Opin Plant Biol, 7(2), 202.

23. Bush, D. R. (1999). Sugar transporters in plant biology. Current Opinion in Plant Biology, 2(3), 187-191. doi: http://dx.doi.org/10.1016/S1369-5266(99)80034-X

24. Buttner, M. (2010). The Arabidopsis sugar transporter (AtSTP) family: an update. [Review]. Plant Biology, 12, 35-41. doi: 10.1111/j.1438-8677.2010.00383.x

25. Carnegie Institution of Washington Department of Plant Biology and the National Center for Genome Resources (NCGR): The Arabidopsis Information Resource, TAIR. URL http://www.arabidopsis.org/

26. Carr, J. P., Lewsey, M. G., & Palukaitis, P. (2010). Signaling in induced resistance. Adv Virus Res, 76, 57-121. doi: 10.1016/s0065-3527(10)76003-6

27. Chan, C., Qi, X., Li, M.-W., Wong, F.-L., & Lam, H.-M. (2012). Recent Developments of Genomic Research in Soybean. Journal of Genetics and Genomics, 39(7), 317-324. doi: http://dx.doi.org/10.1016/j.jgg.2012.02.002

28. Charoensawan, V., Wilson, D., & Teichmann, S. A. (2010a). Genomic repertoires of DNA-binding transcription factors across the tree of life. Nucleic Acids Res, 38(21), 7364-7377.

29. Charoensawan, V., Wilson, D., & Teichmann, S. A. (2010). Lineage-specific expansion of DNA-binding transcription factor families. [Research Support, Non-U.S. Gov't]. Trends Genet, 26(9), 388-393. doi: 10.1016/j.tig.2010.06.004

30. Chen, W., Mingus, J., Mammadov, J., Backlund, J., Greene, T., Thompson, S., & Kumpatla, S. (2010). KASPar: a simple and cost-effective system for SNP genotyping. Paper presented at the Final program, abstract and exhibit guide of the XVIII international conference on the status of plant and animal genome research, San Diego, CA.

31. Chen, M., Wang, Q.-Y., Cheng, X.-G., Xu, Z.-S., Li, L.-C., Ye, X.-G., . . . Ma, Y.-Z. (2007). < i> GmDREB2</i>, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochemical and Biophysical Research Communications, 353(2), 299-305.

32. Chinnusamy, V., Ohta, M., Kanrar, S., Lee, B. H., Hong, X., Agarwal, M., & Zhu, J. K. (2003). ICE1: a regulator of cold-induced transcriptome and

freezing tolerance in Arabidopsis. Genes Dev, 17(8), 1043-1054. doi: 10.1101/gad.1077503

33. Choi, H. K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J. H., . . . Cook, D. R. (2004). A sequence-based genetic map of Medicago truncatula and comparison of marker colinearity with M. sativa. [Comparative Study Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. Genetics, 166(3), 1463-1502.

34. Choi, I.-Y., Hyten, D. L., Matukumalli, L. K., Song, Q., Chaky, J. M., Quigley, C. V., . . . Cregan, P. B. (2007). A Soybean Transcript Map: Gene Distribution, Haplotype and Single-Nucleotide Polymorphism Analysis. Genetics, 176(1), 685-696. doi: 10.1534/genetics.107.070821

35. Close, T.J. (2008). Cowpea (Vigna unguiculata) genomic resources. Paper presented at the Plant and animal genome XVI conference, San Diego, CA, USA W.

36. Close T. J., Luo M., Lonardi S., Alpert M., Duma D., Bozdag S., Farmer A., May G. D., Jackson S. A., Ma, Y., Gu, Y.Q., Wu, J.,You, F.M., Diop, N.N., Wanamaker S., Fenton R. D., Timko M. P., Roberts P. A. and Ehlers J. D. 2011. Genome Resources For Cowpea Genotype IT97K-499-35. Plant and Animal Genomes XIX Conference, Town and Country Convention Center, San Diego, CA.

37. Coetzer, N., Gazendam, I., Oelofse, D., & Berger, D. K. (2010). SSHscreen and SSHdb, generic software for microarray based gene discovery: application to the stress response in cowpea. Plant Methods, 6, 10. doi: 10.1186/1746-4811-6-10

38. Conte, M. G., Gaillard, S., Lanau, N., Rouard, M., & Perin, C. (2008). GreenPhylDB: a database for plant comparative genomics. Nucleic Acids Res, 36(Database issue), D991-998. doi: 10.1093/nar/gkm934

39. Cook, D. R. (1999). Medicago truncatula--a model in the making! Curr Opin Plant Biol, 2(4), 301-304.

40. Cordoba, J. M., Chavarro, C., Schlueter, J. A., Jackson, S. A., & Blair, M. W. (2010). Integration of physical and genetic maps of common bean through BAC-derived microsatellite markers. BMC Genomics, 11, 436. doi: 10.1186/1471-2164-11-436

41. Coulibaly, S., Pasquet, R. S., Papa, R., & Gepts, P. (2002). AFLP analysis of the phenetic organization and genetic diversity of Vigna unguiculataL. Walp. reveals extensive gene flow between wild and domesticated

types. Theor Appl Genet, 104(2-3), 358-366. doi: 10.1007/s001220100740

42. Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., Kahler, A. L., . . . Specht, J. E. (1999). An Integrated Genetic Linkage Map of the Soybean Genome The mention of trade or manufacturer names is made for information only and does not imply an endorsement, recommendation, or exclusion by USDA-ARS. Crop Sci., 39(5), 1464-1490. doi: 10.2135/cropsci1999.3951464x

43. Das, S., Bhat, P. R., Sudhakar, C., Ehlers, J. D., Wanamaker, S., Roberts, P. A., . . . Close, T. J. (2008). Detection and validation of single feature polymorphisms in cowpea (Vigna unguiculata L. Walp) using a soybean genome array. BMC Genomics, 9, 107. doi: 10.1186/1471-2164-9-107

44. Das, S., Ehlers, J., Close, T., & Roberts, P. (2010). Transcriptional profiling of root-knot nematode induced feeding sites in cowpea (Vigna unguiculata L. Walp.) using a soybean genome array. BMC Genomics, 11(1), 480.

45. Davuluri, R. V., Sun, H., Palaniswamy, S. K., Matthews, N., Molina, C., Kurtz, M., & Grotewold, E. (2003). AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics, 4, 25. doi: 10.1186/1471-2105-4-25

46. Derksen, H., Rampitsch, C., & Daayf, F. (2013). Signaling cross-talk in plant disease resistance. Plant Science, 207(0), 79-87. doi: http://dx.doi.org/10.1016/j.plantsci.2013.03.004

47. Doidy, J., Grace, E., Kuhn, C., Simon-Plas, F., Casieri, L., & Wipf, D. (2012). Sugar transporters in plants and in their interactions with fungi. [Review]. Trends Plant Sci, 17(7), 413-422. doi: 10.1016/j.tplants.2012.03.009

48. Domínguez, E., Cuartero, J., & Heredia, A. (2011). An overview on plant cuticle biomechanics. Plant Science, 181(2), 77-84. doi: http://dx.doi.org/10.1016/j.plantsci.2011.04.016

49. Eckey, C., Korell, M., Leib, K., Biedenkopf, D., Jansen, C., Langen, G., & Kogel, K. H. (2004). Identification of powdery mildew-induced barley genes by cDNA-AFLP: functional assessment of an early expressed MAP kinase. Plant Mol Biol, 55(1), 1-15. doi: 10.1007/s11103-004-0275-2

50. Ehlers, J., Luo, M.-C., Close, T.J., & Roberts, P.A. (2009) Physical map of cowpea. http://phymap.ucdavis.edu/cowpea/ Accessed on August 5, 2013.

51. Ehlers, J. (2012). Modern Cowpea Breeding to Overcome Critical Production Constraints in Africa. Paper presented at the Plant and Animal Genome XX Conference (January 14-18, 2012).

52. Elowad, H. O., & Hall, A. E. (1987). Influences of early and late nitrogen fertilization on yield and nitrogen fixation of cowpea under well-watered and dry field conditions. Field Crops Research, 15(3), 229-244.

53. Ernst, H. A., Olsen, A. N., Skriver, K., Larsen, S., & Lo Leggio, L. (2004). Structure of the conserved domain of ANAC, a member of the NAC family of transcription factors. [Article]. Embo Reports, 5(3), 297-303. doi: 10.1038/sj.embor.7400093

54. Eulgem, T. (2005). Regulation of the< i> Arabidopsis</i> defense transcriptome. Trends Plant Sci, 10(2), 71-78.

55. Eulgem, T., Rushton, P. J., Robatzek, S., & Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. [Review]. Trends Plant Sci, 5(5), 199-206.

56. Eulgem, T., & Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review]. Curr Opin Plant Biol, 10(4), 366-371. doi: 10.1016/j.pbi.2007.04.020

57. Fatokun, C.A., Menancio-Hautea, D.I., Danesh, D., Young, N.D. (1992) Evidence for orthologus seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132: 841–846.

58. Fatokun, C.A., Danesh, D., Young, N.D., and Stewart, E.L. (1993) Molecular taxonomic relationships in the genus Vigna based on RFLP analysis. Theor Appl Genet 86: 97–104.

59. Ferguson, B., & Gresshoff, P. (2009). Soybean as a model legume. Grain Legumes, 53(7).

60. Fotopoulos, V., Gilbert, M. J., Pittman, J. K., Marvier, A. C., Buchanan, A. J., Sauer, N., . . . Williams, L. E. (2003). The monosaccharide transporter gene, AtSTP4, and the cell-wall invertase, Atbetafruct1, are induced in Arabidopsis during infection with the fungal biotroph Erysiphe

cichoracearum. Plant Physiol, 132(2), 821-829. doi: 10.1104/pp.103.021428

61. Galeano, C., Fernandez, A., Gomez, M., & Blair, M. (2009). Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (Phaseolus vulgaris L.). BMC Genomics, 10(1), 629.

62. Geiger, D. (2011). Plant sucrose transporters from a biophysical point of view. Mol Plant, 4(3), 395-406. doi: 10.1093/mp/ssr029

63. Gepts, P., Aragão, F. J., de Barros, E., Blair, M. W., Brondani, R., Broughton, W., . . . Lariguet, P. (2008). Genomics of Phaseolus beans, a major source of dietary protein and micronutrients in the tropics Genomics of tropical crop plants (pp. 113-143): Springer.

64. Gfeller, A., Dubugnon, L., Liechti, R., & Farmer, E. E. (2010). Jasmonate biochemical pathway. Sci Signal, 3(109), cm3. doi: 10.1126/scisignal.3109cm3

65. Glauser, G., Grata, E., Dubugnon, L., Rudaz, S., Farmer, E. E., & Wolfender, J.-L. (2008). Spatial and temporal dynamics of jasmonate synthesis and accumulation in Arabidopsis in response to wounding. Journal of Biological Chemistry, 283(24), 16400-16407.

66. Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. Annu. Rev. Phytopathol., 43, 205-227.

67. Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., . . . Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. Nucleic Acids Res, 40(Database issue), D1178-1186. doi: 10.1093/nar/gkr944

68. Gray, W. M., Kepinski, S., Rouse, D., Leyser, O., & Estelle, M. (2001). Auxin regulates SCFTIR1-dependent degradation of AUX/IAA proteins. Nature, 414(6861), 271-276.

69. Gu, Y.-Q., Yang, C., Thara, V. K., Zhou, J., & Martin, G. B. (2000). Pti4 is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. The Plant Cell Online, 12(5), 771-785.

70. Guo, H., & Ecker, J. R. (2004). The ethylene signaling pathway: new insights. Curr Opin Plant Biol, 7(1), 40.

71. Gutterson, N., & Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. Curr Opin Plant Biol, 7(4), 465.

72. Haake, V., Cook, D., Riechmann, J. L., Pineda, O., Thomashow, M. F., & Zhang, J. Z. (2002). Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. Plant Physiol, 130(2), 639-648. doi: 10.1104/pp.006478

73. Hall, A., Ismail, A., Ehlers, J., Marfo, K., Cisse, N., Thiaw, S., & Close, T. (2002). Breeding cowpea for tolerance to temperature extremes and adaptation to drought. Challenges and Opportunities for Enhancing Sustainable Cowpea Production, 14-21.

74. Hall, A. E. (2004). Breeding for adaptation to drought and heat in cowpea. European Journal of Agronomy, 21(4), 447-454.

75. Hall, A. E., Cisse, N., Thiaw, S., Elawad, H. O., Ehlers, J. D., Ismail, A. M., . . . Murdock, L. L. (2003). Development of cowpea cultivars and germplasm by the Bean/Cowpea CRSP. Field Crops Research, 82(2), 103-134.

76. Hardtke, C. S., & Deng, X.-W. (2000). The cell biology of the COP/DET/FUS proteins. Regulating proteolysis in photomorphogenesis and beyond? Plant Physiol, 124(4), 1548-1557.

77. Harrison, M. J. (1996). A sugar transporter from Medicago truncatula: altered expression pattern in roots during vesicular-arbuscular (VA) mycorrhizal associations. The Plant Journal, 9(4), 491-503.

78. Hegedus, D., Yu, M., Baldwin, D., Gruber, M., Sharpe, A., Parkin, I., . . . Lydiate, D. (2003). Molecular characterization of Brassica napus NAC domain transcriptional activators induced in response to biotic and abiotic stress. Plant Mol Biol, 53(3), 383-397.

79. Hernandez, G.-U., Ramirez, M.-U., Blair, M.-C., COLOMBIA, S., Lara, M.-U., & Blanco, L.-U. (2004). Comparative Analysis of Common Bean (Phaseolus Vulgaris) Nodule, Root, Pod, and Leaf Expressed Sequence Tag (Ests) Libraries: a Platform for" phaseomics" Research. Paper presented at the International Conference on Legume Genomics and Genetics.

80. Hougaard, B. K., Madsen, L. H., Sandal, N., de Carvalho Moretzsohn, M., Fredslund, J., Schauser, L., . . . Tabata, S. (2008). Legume anchor markers link syntenic regions between Phaseolus vulgaris, Lotus

japonicus, Medicago truncatula and Arachis. Genetics, 179(4), 2299-2312.

81. Hu, Y., Dong, Q., & Yu, D. (2012). < i> Arabidopsis</i> WRKY46 coordinates with WRKY70 and WRKY53 in basal resistance against pathogen< i> Pseudomonas syringae</i>. Plant Science, 185, 288-297.

82. Huala, E., Dickerman, A. W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., . . . Rhee, S. Y. (2001). The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. [Article]. Nucleic Acids Res, 29(1), 102-105. doi: 10.1093/nar/29.1.102

83. Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., . . . Yeats, C. (2009). InterPro: the integrative protein signature database. Nucleic Acids Res, 37(Database issue), D211-215. doi: 10.1093/nar/gkn785

84. Huynh, B. L., Close, T. J., Roberts, P. A., Hu, Z., Wanamaker, S., Lucas, M. R., Chiulele, R., Cisse, N., David, A., Hearne, S., Fatokun, C., Diop, N. N., & Ehlers, J. D. (2013). Gene Pools and the Genetic Architecture of Domesticated Cowpea. *The Plant Genome, 6* (2).

85. Hyten, D. L., Cannon, S. B., Song, Q., Weeks, N., Fickus, E. W., Shoemaker, R. C., . . . Cregan, P. B. (2010). High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics, 11(1), 38.

86. Hyten, D. L., Choi, I. Y., Song, Q. J., Specht, J. E., Carter, T. E., Shoemaker, R. C., . . . Cregan, P. B. (2010). A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. [Article]. Crop Science, 50(3), 960-968. doi: 10.2135/cropsci2009.06.0360

87. Ito, M., Miyamoto, J., Mori, Y., Fujimoto, S., Uchiumi, T., Abe, M., . . . Fukui, K. (2000). Genome and chromosome dimensions of Lotus japonicus. Journal of Plant Research, 113(1112), 435-442. doi: Doi 10.1007/Pl00013952

88. Izawa, T., Foster, R., Nakajima, M., Shimamoto, K., & Chua, N.-H. (1994). The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. The Plant Cell Online, 6(9), 1277-1287.

89. Jackson, S. A., Iwata, A., Lee, S. H., Schmutz, J., & Shoemaker, R. (2011). Sequencing crop genomes: approaches and applications. New Phytol, 191(4), 915-925. doi: 10.1111/j.1469-8137.2011.03804.x

90. Jackson, S., Schmutz, J., & McClean, P. (2012) *Phaseolus vulgaris* (common bean). URL http://phytozome.net/commonbean.php

91. Jaglo-Ottosen, K. R., Gilmour, S. J., Zarka, D. G., Schabenberger, O., & Thomashow, M. F. (1998). Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. Science, 280(5360), 104-106.

92. Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O'Shea, C., & Skriver, K. (2010). The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. Biochemical Journal, 426(2), 183-196. doi: 10.1042/bj20091234

93. Joel, D. M. (2000). The long-term approach to parasitic weeds control: manipulation of specific developmental mechanisms of the parasite. Crop Protection, 19(8–10), 753-758. doi: http://dx.doi.org/10.1016/S0261-2194(00)00100-9

94. Jofuku, K. D., Den Boer, B., Van Montagu, M., & Okamuro, J. K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. The Plant Cell Online, 6(9), 1211-1225.

95. Journot-Catalino, N., Somssich, I. E., Roby, D., & Kroj, T. (2006). The transcription factors WRKY11 and WRKY17 act as negative regulators of basal resistance in Arabidopsis thaliana. Plant Cell, 18(11), 3289-3302. doi: 10.1105/tpc.106.044149

96. Kaneko, T., Asamizu, E., Kato, T., Sato, S., Nakamura, Y., & Tabata, S. (2003). Structural analysis of a Lotus japonicus genome. III. Sequence features and mapping of sixty-two TAC clones which cover the 6.7 Mb regions of the genome. DNA research, 10(1), 27-33.

97. Kang, S. G., & Do, K. S. (2009). Identification of a pathogen-Induced glycine max transcription factor GmWRKY1. The Plant Pathology Journal, 25(4), 381-388.

98. Kaga, A., Tomooka, N., Egawa, Y., Hosaka, K., and Kamijima, O. (1996) Species relationships in the subgenus Ceratotropis (genus Vigna) as revealed by RAPD analysis. Euphytica 88: 17–24.

99. Katagiri, F. (2004). A global view of defense gene expression regulation–a highly interconnected signaling network. Curr Opin Plant Biol, 7(5), 506-511.

100. Kato, T., Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., & Tabata, S. (2003). Structural analysis of a Lotus japonicus genome. V. Sequence features and mapping of sixty-four TAC clones which cover the 6.4 Mb regions of the genome. DNA research, 10(6), 277-285.

101. Kazusa DNA Research Institute (2012). Lotus japonicus genomic assembly [data file]. Retrieved from ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r2.5/pseudomolecule/

102. Kim, K.-C., Fan, B., & Chen, Z. (2006). Pathogen-induced Arabidopsis WRKY7 is a transcriptional repressor and enhances plant susceptibility to Pseudomonas syringae. Plant Physiol, 142(3), 1180-1192.

103. Kim, K.-C., Lai, Z., Fan, B., & Chen, Z. (2008). Arabidopsis WRKY38 and WRKY62 transcription factors interact with histone deacetylase 19 in basal defense. The Plant Cell Online, 20(9), 2357-2371.

104. Kim, S., Kang, J. Y., Cho, D. I., Park, J. H., & Kim, S. Y. (2004). ABF2, an ABRE-binding bZIP factor, is an essential component of glucose signaling and its overexpression affects multiple stress tolerance. Plant J, 40(1), 75-87. doi: 10.1111/j.1365-313X.2004.02192.x

105. Knuppel, R., Dietze, P., Lehnberg, W., Frech, K., & Wingender, E. (1994). TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins. J Comput Biol, 1(3), 191-198.

106. Komaki, M., Okada, K., Nishino, E., & Shimura, Y. (1988). Isolation and characterization of novel mutants of Arabidopsis thaliana defective in flower development. Development, 104(2), 195-203.

107. Kopittke, P. M., Asher, C. J., Kopittke, R. A., & Menzies, N. W. (2007). Toxic effects of Pb2+ on growth of cowpea (Vigna unguiculata). Environmental Pollution, 150(2), 280-287. doi: http://dx.doi.org/10.1016/j.envpol.2007.01.011

108. Kopittke, P. M., Blamey, F. P., McKenna, B. A., Wang, P., & Menzies, N. W. (2011). Toxicity of metals to roots of cowpea in relation to their binding strength. Environ Toxicol Chem, 30(8), 1827-1833. doi: 10.1002/etc.557

109. Krogan, N. T., Hogan, K., & Long, J. A. (2012). APETALA2 negatively regulates multiple floral organ identity genes in Arabidopsis by recruiting the co-repressor TOPLESS and the histone deacetylase HDA19. Development, 139(22), 4180-4190. doi: 10.1242/dev.085407

110. Ku, Y.-S., Au-Yeung, W.-K., Yung, Y.-L., Li, M.-W., Wen, C.-Q., Liu, X., & Lam, H.-M. (2013). Drought Stress and Tolerance in Soybean.

111. Kühn, C., Quick, W., Schulz, A., Riesmeier, J., Sonnewald, U., & Frommer, W. (1996). Companion cell-specific inhibition of the potato sucrose transporter SUT1. Plant, Cell & Environment, 19(10), 1115-1123.

112. Lalonde, S., & Frommer, W. B. (2012). SUT Sucrose and MST Monosaccharide Transporter Inventory of the Selaginella Genome. Frontiers in plant science, 3, 24. doi: 10.3389/fpls.2012.00024

113. Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D. M., Correa, L. G., . . . Rensing, S. A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. [Comparative Study Research Support, Non-U.S. Gov't]. Genome Biol Evol, 2, 488-503. doi: 10.1093/gbe/evq032

114. Latchman, D. S. (1997). Transcription factors: an overview. [Review]. Int J Biochem Cell Biol, 29(12), 1305-1312. doi: http://dx.doi.org/10.1016/S1357-2725(97)00085-X

115. Lavin, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. Syst Biol, 54(4), 575-594.

116. Leslie, C. A., & Romani, R. J. (1988). Inhibition of ethylene biosynthesis by salicylic Acid. Plant Physiol, 88(3), 833-837.

117. Li, J., Brader, G., & Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. Sci Signal, 16(2), 319.

118. Lindemose, S., O'Shea, C., Jensen, M. K., & Skriver, K. (2013). Structure, Function and Networks of Transcription Factors Involved in Abiotic Stress Responses. [Review]. International Journal of Molecular Sciences, 14(3), 5842-5878. doi: 10.3390/ijms14035842

119. Liu, L. S., White, M. J., & MacRae, T. H. (1999). Transcription factors and their genes in higher plants - Functional domains, evolution and regulation.

[Review]. European Journal of Biochemistry, 262(2), 247-257. doi: 10.1046/j.1432-1327.1999.00349.x

120. Liu, Q., Kasuga, M., Sakuma, Y., Abe, H., Miura, S., Yamaguchi-Shinozaki, K., & Shinozaki, K. (1998). Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought-and low-temperature-responsive gene expression, respectively, in Arabidopsis. The Plant Cell Online, 10(8), 1391-1406.

121. Loake, G., & Grant, M. (2007). Salicylic acid in plant defence—the players and protagonists. Curr Opin Plant Biol, 10(5), 466-472. doi: http://dx.doi.org/10.1016/j.pbi.2007.08.008

122. Lorenzo, O., Piqueras, R., Sánchez-Serrano, J. J., & Solano, R. (2003). ETHYLENE RESPONSE FACTOR1 integrates signals from ethylene and jasmonate pathways in plant defense. The Plant Cell Online, 15(1), 165-178.

123. Lucas, M. R., Diop, N.-N., Wanamaker, S., Ehlers, J. D., Roberts, P. A., & Close, T. J. (2011). Cowpea–soybean synteny clarified through an improved genetic map. The Plant Genome, 4(3), 218-225.

124. Lucas, M. R., Ehlers, J. D., Huynh, B. L., Diop, N. N., Roberts, P. A., & Close, T. J. (2013a). Markers for breeding heat-tolerant cowpea. [Article]. Molecular Breeding, 31(3), 529-536. doi: 10.1007/s11032-012-9810-z

125. Lucas, M. R., Huynh, B. L., da Silva Vinholes, P., Cisse, N., Drabo, I., Ehlers, J. D., . . . Close, T. J. (2013b). Association Studies and Legume Synteny Reveal Haplotypes Determining Seed Size in Vigna unguiculata. Front Plant Sci, 4, 95. doi: 10.3389/fpls.2013.00095

126. Luo, M.-C., Thomas, C., You, F. M., Hsiao, J., Ouyang, S., Buell, C. R., . . . Dvorak, J. (2003). High-throughput fingerprinting of bacterial artificial chromosomes using the snapshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics, 82(3), 378-389.

127. MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. Trends in Genetics, 27(4), 141-148. doi: http://dx.doi.org/10.1016/j.tig.2011.01.001

128. Marek, L. F., Mudge, J., Darnielle, L., Grant, D., Hanson, N., Paz, M., . . . Shoemaker, R. C. (2001). Soybean genomic survey: BAC-end

sequences near RFLP and SSR markers. Genome, 44(4), 572-581. doi: 10.1139/g01-052

129. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., . . . Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. [10.1038/nature03959]. Nature, 437(7057), 376-380. doi: http://www.nature.com/nature/journal/v437/n7057/suppinfo/nature03959_S1.html

130. Martinez, M. (2011). Plant protein-coding gene families: emerging bioinformatics approaches. [Review]. Trends Plant Sci, 16(10), 558-567. doi: 10.1016/j.tplants.2011.06.003

131. Matys, V., Kel-Margoulis, O. V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., . . . Wingender, E. (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res, 34(Database issue), D108-110. doi: 10.1093/nar/gkj143

132. McClean, P., Mamidi, S., McConnell, M., Chikara, S., & Lee, R. (2010). Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. BMC Genomics, 11(1), 184.

133. McClean, P. E., Lavin, M., Gepts, P., & Jackson, S. A. (2008). Phaseolus vulgaris: a diploid model for soybean Genetics and Genomics of Soybean (pp. 55-76): Springer.

134. Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). Arabidopsis thaliana: A Model Plant for Genome Analysis. Science, 282(5389), 662-682. doi: 10.1126/science.282.5389.662

135. Melotto, M., Monteiro-Vitorello, C. B., Bruschi, A. G., & Camargo, L. E. (2005). Comparative bioinformatic analysis of genes expressed in common bean (Phaseolus vulgaris L.) seedlings. Genome, 48(3), 562-570.

136. Meyer, C. A., Tang, Q. Z., & Liu, X. S. (2012). Minireview: Applications of Next-Generation Sequencing on Studies of Nuclear Receptor Regulation and Function. [Review]. *Molecular Endocrinology, 26*(10), 1651-1659. doi: 10.1210/me.2012-1150

137. Mithofer, A., & Boland, W. (2012). Plant Defense Against Herbivores: Chemical Aspects. In S. S. Merchant (Ed.), Annual Review of Plant Biology, Vol 63 (Vol. 63, pp. 431-450). Palo Alto: Annual Reviews.

138. Mitsuda, N., & Ohme-Takagi, M. (2009). Functional analysis of transcription factors in Arabidopsis. [Review]. Plant Cell Physiol, 50(7), 1232-1248. doi: 10.1093/pcp/pcp075

139. Mochida, K., & Shinozaki, K. (2010). Genomics and Bioinformatics Resources for Crop Improvement. [Review]. Plant and Cell Physiology, 51(4), 497-523. doi: 10.1093/pcp/pcq027

140. Mol, J., Grotewold, E., & Koes, R. (1998). How genes paint flowers and seeds. Trends Plant Sci, 3(6), 212-217.

141. Moyano, E., Martínez-Garcia, J. F., & Martin, C. (1996). Apparent redundancy in myb gene function provides gearing for the control of flavonoid biosynthesis in antirrhinum flowers. The Plant Cell Online, 8(9), 1519-1532.

142. Muchero, W., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., . . . Close, T. J. (2009). A consensus genetic map of cowpea [Vigna unguiculata (L) Walp.] and synteny based on EST-derived SNPs. Proc Natl Acad Sci U S A, 106(43), 18159-18164. doi: 10.1073/pnas.0905886106

143. Muchero, W., Ehlers, J. D., Close, T. J., & Roberts, P. A. (2011). Genic SNP markers and legume synteny reveal candidate genes underlying QTL for Macrophomina phaseolina resistance and maturity in cowpea [Vigna unguiculata (L) Walp.]. BMC Genomics, 12, 8. doi: 10.1186/1471-2164-12-8

144. Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Sato, S., & Tabata, S. (2002). Structural analysis of a Lotus japonicus genome. II. Sequence features and mapping of sixty-five TAC clones which cover the 6.5-Mb regions of the genome. DNA research, 9(2), 63-70.

145. Nehls, U., Grunze, N., Willmann, M., Reich, M., & Kuster, H. (2007). Sugar for my honey: Carbohydrate partitioning in ectomycorrhizal symbiosis. [Review]. Phytochemistry, 68(1), 82-91. doi: 10.1016/j.phytochem.2006.09.024

146. Ng, N. Q., & Marechal, R. (1985). Cowpea taxonomy, origin and germplasm. In S. R. Singh & K. O. Rachie (Eds.), Cowpea Research, Production and Utilization (pp. 11-21). Chichester, NY: John Wiley and Sons, Ltd.

147. Nielsen, S., Ohler, T., & Mitchell, C. (1997). Cowpea leaves for human consumption: production, utilization, and nutrient composition. Advances in cowpea research, 326.

148. Nielsen, S. S., Brandt, W. E., & Singh, B. B. (1993). Genetic-Variability for Nutritional Composition and Cooking Time of Improved Cowpea Lines. Crop Science, 33(3), 469-472. doi: 10.2135/cropsci1993.0011183X003300030010x

149. Ohme-Takagi, M., & Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. The Plant Cell Online, 7(2), 173-182.

150. Olsen, A. N., Ernst, H. A., Leggio, L. L., & Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. Trends Plant Sci, 10(2), 79-87. doi: 10.1016/j.tplants.2004.12.010

151. Onate-Sanchez, L., & Singh, K. B. (2002). Identification of Arabidopsis ethylene-responsive element binding factors with distinct induction kinetics after pathogen infection. Plant Physiol, 128(4), 1313-1322. doi: 10.1104/pp.010862

152. Oseni, T. O. (2010). Evaluation of sorghum-cowpea intercrop productivity in savanna agro-ecology using competition indices. Journal of Agricultural Science, 2(3), P229.

153. Ouédraogo, J., Maheshwari, V., Berner, D., St-Pierre, C.-A., Belzile, F., & Timko, M. (2001). Identification of AFLP markers linked to resistance of cowpea (Vigna unguiculata L.) to parasitism by Striga gesnerioides. TAG Theoretical and Applied Genetics, 102(6), 1029-1036.

154. Ouédraogo, J. T., Tignegre, J.-B., Timko, M. P., & Belzile, F. J. (2002). AFLP markers linked to resistance against Striga gesnerioides race 1 in cowpea (Vigna unguiculata). Genome, 45(5), 787-793.

155. Pant, K., Chandel, K., & Joshi, B. (1982). Analysis of diversity in Indian cowpea genetic resources. SABRO J, 14, 103-111.

156. Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. Methods in enzymology, 183, 63.

157. Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A, 85(8), 2444-2448.

158. Peng, H., Yu, X., Cheng, H., Shi, Q., Zhang, H., Li, J., & Ma, H. (2010). Cloning and characterization of a novel NAC family gene CarNAC1 from chickpea (Cicer arietinum L.). Mol Biotechnol, 44(1), 30-40. doi: 10.1007/s12033-009-9202-8

159. Pereira, S. S., Guimaraes, F. C., Carvalho, J. F., Stolf-Moreira, R., Oliveira, M. C., Rolla, A. A., . . . Nepomuceno, A. L. (2011). Transcription factors expressed in soybean roots under drought stress. [Research Support, Non-U.S. Gov't]. Genet Mol Res, 10(4), 3689-3701. doi: 10.4238/2011.October.21.5

160. Pieterse, C. M., Leon-Reyes, A., Van der Ent, S., & Van Wees, S. C. (2009). Networking by small-molecule hormones in plant immunity. Nature Chemical Biology, 5(5), 308-316.

161. Pieterse, C. M., Van der Does, D., Zamioudis, C., Leon-Reyes, A., & Van Wees, S. C. (2012). Hormonal modulation of plant immunity. Annu Rev Cell Dev Biol, 28, 489-521. doi: 10.1146/annurev-cellbio-092910-154055

162. Pinheiro, G. L., Marques, C. S., Costa, M. D. B. L., Reis, P. A. B., Alves, M. S., Carvalho, C. M., . . . Fontes, E. P. B. (2009). Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. Gene, 444(1–2), 10-23. doi: http://dx.doi.org/10.1016/j.gene.2009.05.012

163. Pottorff, M., Ehlers, J. D., Fatokun, C., Roberts, P. A., & Close, T. J. (2012). Leaf morphology in Cowpea [Vigna unguiculata (L.) Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. BMC Genomics, 13, 234. doi: 10.1186/1471-2164-13-234

164. Pozo, M. J., Van Loon, L., & Pieterse, C. M. (2004). Jasmonates-signals in plant-microbe interactions. Journal of Plant Growth Regulation, 23(3), 211-222.

165. Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., & Vandepoele, K. (2009). PLAZA: a comparative genomics resource to study gene and genome evolution in plants. Plant Cell, 21(12), 3718-3731. doi: 10.1105/tpc.109.071506

166. Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. [Research Support, Non-U.S. Gov't]. Nucleic Acids Res, 40(Database issue), D290-301. doi: 10.1093/nar/gkr1065

167. Puranik, S., Sahu, P. P., Srivastava, P. S., & Prasad, M. (2012). NAC proteins: regulation and role in stress tolerance. Trends Plant Sci, 17(6), 369-381. doi: 10.1016/j.tplants.2012.02.004

168. Purseglove, J. W. (1968). Tropical Crops: Dicotyledons, Vols. 1 & 2. Tropical Crops: Dicotyledons, Vols. 1 & 2.

169. Qiu, D., Xiao, J., Ding, X., Xiong, M., Cai, M., Cao, Y., . . . Wang, S. (2007). OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. Mol Plant Microbe Interact, 20(5), 492-499. doi: 10.1094/mpmi-20-5-0492

170. Ramírez, M., Graham, M. A., Blanco-López, L., Silvente, S., Medrano-Soto, A., Blair, M. W., . . . Lara, M. (2005). Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. Plant Physiol, 137(4), 1211-1227.

171. Ramón-Maiques, S., Kuo, A. J., Carney, D., Matthews, A. G. W., Oettinger, M. A., Gozani, O., & Yang, W. (2007). The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. Proceedings of the National Academy of Sciences, 104(48), 18993-18998. doi: 10.1073/pnas.0709170104

172. Rennie, E. A., & Turgeon, R. (2009). A comprehensive picture of phloem loading strategies. [Article]. Proc Natl Acad Sci U S A, 106(33), 14162-14167. doi: 10.1073/pnas.0902279106

173. Riechmann, J. L., & Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. Review]. Biol Chem, 379(6), 633-646.

174. Riechmann, J. L., & Ratcliffe, O. J. (2000). A genomic perspective on plant transcription factors. Curr Opin Plant Biol, 3(5), 423-434. doi: http://dx.doi.org/10.1016/S1369-5266(00)00107-2

175. Riesmeier, J. W., Willmitzer, L., & Frommer, W. B. (1994). Evidence for an essential role of the sucrose transporter in phloem loading and assimilate partitioning. The EMBO Journal, 13(1), 1.

176. Roe, B., & Kupfer, D. (2004). Sequencing gene rich regions of Medicago truncatula, a model legume Molecular Breeding of Forage and Turf (pp. 333-344): Springer.

177. Rolland, F., Baena-Gonzalez, E., & Sheen, J. (2006). Sugar sensing and signaling in plants: conserved and novel mechanisms. Annu Rev Plant Biol, 57, 675-709. doi: 10.1146/annurev.arplant.57.032905.105441

178. Rouard, M., Guignon, V., Aluome, C., Laporte, M.-A., Droc, G., Walde, C., . . . Conte, M. G. (2011). GreenPhylDB v2. 0: comparative and functional genomics in plants. Nucleic Acids Res, 39(suppl 1), D1095-D1102.

179. Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor , G. L., Miklos, Nelson, C. R., . . . Lewis, S. (2000). Comparative Genomics of the Eukaryotes. Science, 287(5461), 2204-2215. doi: 10.1126/science.287.5461.2204

180. Rushton, P. J., Macdonald, H., Huttly, A. K., Lazarus, C. M., & Hooley, R. (1995). Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of alpha-Amy2 genes. [Research Support, Non-U.S. Gov't]. Plant Mol Biol, 29(4), 691-702.

181. Sablowski, R. W., & Meyerowitz, E. M. (1998). A Homolog of< i> NO APICAL MERISTEM</i> Is an Immediate Target of the Floral Homeotic Genes< i> APETALA3/PISTILLATA</i>. Cell, 92(1), 93-103.

182. Sakamoto, A., Omirulleh, S., Nakayama, T., & Iwabuchi, M. (1996). A zinc-finger-type transcription factor WZF-1 that binds to a novel cis-acting element element of histone gene promoters represses its own promoter. Plant and Cell Physiology, 37(4), 557-562.

183. Sanchez, R., & Zhou, M. M. (2011). The PHD finger: a versatile epigenome reader. [Review]. Trends in Biochemical Sciences, 36(7), 364-372. doi: 10.1016/j.tibs.2011.03.005

184. Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., . . . Tabata, S. (2008). Genome structure of the legume, Lotus japonicus. DNA Res, 15(4), 227-239. doi: 10.1093/dnares/dsn008

185. Sato, S., & Tabata, S. (2006). Lotus japonicus as a platform for legume research. Curr Opin Plant Biol, 9(2), 128-132. doi: 10.1016/j.pbi.2006.01.008

186. Schlueter, J. A., Goicoechea, J. L., Collura, K., Gill, N., Lin, J.-Y., Yu, Y., . . . Jackson, S. A. (2008). BAC-end Sequence Analysis and a Draft Physical Map of the Common Bean (Phaseolus vulgaris L.) Genome. Tropical Plant Biology, 1(1), 40-48. doi: 10.1007/s12042-007-9003-9

187. Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., . . . Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. Nature, 463(7278), 178-183. doi: 10.1038/nature08670

188. Shen, Q. H., Saijo, Y., Mauch, S., Biskup, C., Bieri, S., Keller, B., . . . Schulze-Lefert, P. (2007). Nuclear activity of MLA immune receptors links isolate-specific and basal disease-resistance responses. Science, 315(5815), 1098-1103. doi: 10.1126/science.1136372

189. Shiu, S. H., Shih, M. C., & Li, W. H. (2005). Transcription factor families have much higher expansion rates in plants than in animals. Plant Physiol, 139(1), 18-26. doi: 10.1104/pp.105.065110

190. Shoemaker, R. C., Grant, D., Olson, T., Warren, W. C., Wing, R., Yu, Y., . . . Jackson, S. (2008). Microsatellite discovery from BAC end sequences and genetic mapping to anchor the soybean physical and genetic maps. Genome, 51(4), 294-302. doi: 10.1139/g08-010

191. Shore, P., & Sharrocks, A. D. (1995). The MADS-Box Family of Transcription Factors. European Journal of Biochemistry, 229(1), 1-13. doi: 10.1111/j.1432-1033.1995.0001l.x

192. Singh, B. (2005). Cowpea [Vigna unguiculata (L.) Walp]. Genetic resources, chromosome engineering, and crop improvement: Grain legumes, 1, 117-162.

193. Singh, B., Ajeigbe, H., Tarawali, S. A., Fernandez-Rivera, S., & Abubakar, M. (2003). Improving the production and utilization of cowpea as food and fodder. Field Crops Research, 84(1), 169-177.

194. Singh, K., Foley, R. C., & Onate-Sanchez, L. (2002). Transcription factors in plant defense and stress responses. [Research Support, Non-U.S. Gov't Review]. Curr Opin Plant Biol, 5(5), 430-436. doi: 10.1016/s1369-5266(02)00289-3

195. Solano, R., Stepanova, A., Chao, Q., & Ecker, J. R. (1998). Nuclear events in ethylene signaling: a transcriptional cascade mediated by ETHYLENE-INSENSITIVE3 and ETHYLENE-RESPONSE-FACTOR1. Genes Dev, 12(23), 3703-3714.

196. Somssich, I. E. (2006). Networks of transcriptional regulation underlying plant defense responses towards phytopathogens. In K. D. Grasser (Ed.), Regulation of Transcription in Plants (Vol. 29, pp. 266-284). Oxford: Blackwell Publishing.

197. Sobhanian, H., Aghaei, K., & Komatsu, S. (2011). Changes in the plant proteome resulting from salt stress: toward the creation of salt-tolerant crops? J Proteomics, 74(8), 1323-1337. doi: 10.1016/j.jprot.2011.03.018

198. Soderlund, C., Humphray, S., Dunham, A., & French, L. (2000). Contigs built with fingerprints, markers, and FPC V4. 7. Genome Research, 10(11), 1772-1787.

199. Song, Q. J., Marek, L. F., Shoemaker, R. C., Lark, K. G., Concibido, V. C., Delannay, X., . . . Cregan, P. B. (2004). A new integrated genetic linkage map of the soybean. Theoretical and Applied Genetics, 109(1), 122-128. doi: 10.1007/s00122-004-1602-3

200. Sonnhammer, E. L., Eddy, S. R., & Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. Proteins, 28(3), 405-420.

201. Spannagl, M., Haberer, G., Ernst, R., Schoof, H., & Mayer, K. F. (2007). MIPS plant genome information resources. Methods Mol Biol, 406, 137-159.

202. Spoel, S. H., Koornneef, A., Claessens, S. M. C., Korzelius, J. P., Van Pelt, J. A., Mueller, M. J., . . . Pieterse, C. M. J. (2003). NPR1 Modulates Cross-Talk between Salicylate- and Jasmonate-Dependent Defense Pathways through a Novel Function in the Cytosol. The Plant Cell Online, 15(3), 760-770. doi: 10.1105/tpc.009159

203. Sun, C., Palmqvist, S., Olsson, H., Borén, M., Ahlandsberg, S., & Jansson, C. (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. The Plant Cell Online, 15(9), 2076-2092.

204. Tarawali, S., Singh, B., Peters, M., & Blade, S. (1997). Cowpea haulms as fodder. Advances in cowpea research, 313.

205. Thaler, J. S., Humphrey, P. T., & Whiteman, N. K. (2012). Evolution of jasmonate and salicylate signal crosstalk. [Review]. Trends Plant Sci, 17(5), 260-270. doi: 10.1016/j.tplants.2012.02.010

206. Tian, Z. D., Zhang, Y., Liu, J., & Xie, C. H. (2010). Novel potato C2H2-type zinc finger protein gene, StZFP1, which responds to biotic and abiotic stress, plays a role in salt tolerance. [Article]. Plant Biology, 12(5), 689-697. doi: 10.1111/j.1438-8677.2009.00276.x

207. Timko, M.P., Ehlers, J.D., and Roberts P.A. (2007a) Cowpea. In: Pulses, Sugar and Tuber Crops, Genome Mapping and Molecular Breeding in Plants, Vol. 3, C. Kole (Ed.), Springer-Verlag, Berlin Heidelberg. pp. 49-67.

208. Timko, M., Gowda, B., Ouedraogo, J., Ousmane, B., Ejeta, G., & Gressel, J. (2007b). Molecular markers for analysis of resistance to Striga gesnerioides in cowpea. Integrating New Technologies for Striga

Control: Towards Ending the Witch-Hunt. Singapore: World Scientific, 115-128.

209. Timko, M. P., & Singh, B. (2008). Cowpea, a multifunctional legume. Genomics of tropical crop plants, 227-258.

210. Timko, M. P., Rushton, P. J., Laudeman, T. W., Bokowiec, M. T., Chipumuro, E., Cheung, F., . . . Chen, X. (2008). Sequencing and analysis of the gene-rich space of cowpea. BMC Genomics, 9, 103. doi: 10.1186/1471-2164-9-103

211. Tosti, N., & Negri, V. (2002). Efficiency of three PCR-based markers in assessing genetic variation among cowpea (Vigna unguiculata subsp. unguiculata) landraces. Genome, 45(2), 268-275.

212. Town, C. D. (2006). Annotating the genome of Medicago truncatula. [Research Support, U.S. Gov't, Non-P.H.S. Review]. Curr Opin Plant Biol, 9(2), 122-127. doi: 10.1016/j.pbi.2006.01.004

213. Tran, L. S., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., . . . Yamaguchi-Shinozaki, K. (2004). Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. Plant Cell, 16(9), 2481-2498. doi: 10.1105/tpc.104.022699

214. Vaillancourt, R., & Weeden, N. (1992). Chloroplast DNA polymorphism suggests Nigerian center of domestication for the cowpea, Vigna unguiculata (Leguminosae). American journal of botany, 1194-1199.

215. Van Bel, M., Proost, S., Wischnitzki, E., Movahedi, S., Scheerlinck, C., Van de Peer, Y., & Vandepoele, K. (2012). Dissecting plant genomes with the PLAZA comparative genomics platform. Plant Physiol, 158(2), 590-600. doi: 10.1104/pp.111.189514

216. van Loon, L. C., Rep, M., & Pieterse, C. M. (2006). Significance of inducible defense-related proteins in infected plants. Annu Rev Phytopathol, 44, 135-162. doi: 10.1146/annurev.phyto.44.070505.143425

217. Van Verk, M. C., Gatz, C., & Linthorst, H. J. M. (2009). Transcriptional Regulation of Plant Defense Responses. In L. C. VanLoon (Ed.), Plant Innate Immunity (Vol. 51, pp. 397-438). London: Academic Press Ltd-Elsevier Science Ltd.

218. Vaquerizas, J., Teichmann, S., & Luscombe, N. (2012). How Do You Find Transcription Factors? Computational Approaches to Compile and

Annotate Repertoires of Regulators for Any Genome. In B. Deplancke & N. Gheldof (Eds.), Gene Regulatory Networks (Vol. 786, pp. 3-19): Humana Press.

219. Verdcourt, B. (1970). Studies in the Leguminosae-Papilionoïdeae for the'Flora of Tropical East Africa': IV. Kew Bulletin, 24(3), 507-569.

220. Vom Endt, D., Kijne, J. W., & Memelink, J. (2002). Transcription factors controlling plant secondary metabolism: what regulates the regulators? Phytochemistry, 61(2), 107-114. doi: http://dx.doi.org/10.1016/S0031-9422(02)00185-1

221. Wahl, R., Wippel, K., Goos, S., Kamper, J., & Sauer, N. (2010). A novel high-affinity sucrose transporter is required for virulence of the plant pathogen Ustilago maydis. PLoS Biol, 8(2), e1000303. doi: 10.1371/journal.pbio.1000303

222. Wanamaker, S., and T.J. Close. 2011. HarvEST. HarvEST:Cowpea version 1.27. Available at http://harvest.ucr.edu (verified 27 Sept. 2011). University of California, Riverside, CA.

223. Wang, Z., Libault, M., Joshi, T., Valliyodan, B., Nguyen, H. T., Xu, D., . . . Cheng, J. L. (2010). SoyDB: a knowledge database of soybean transcription factors. [Article]. BMC Plant Biol, 10. doi: 10.1186/1471-2229-10-14

224. Weigel, D. (1995). The APETALA2 domain is related to a novel type of DNA binding domain. Plant Cell, 7(4), 388.

225. Welner, D. H., Lindemose, S., Grossmann, J. G., Mollegaard, N. E., Olsen, A. N., Helgstrand, C., . . . Lo Leggio, L. (2012). DNA binding by the plant-specific NAC transcription factors in crystal and solution: a firm link to WRKY and GCM transcription factors. Biochem J, 444(3), 395-404. doi: 10.1042/bj20111742

226. Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. [Review]. Trends in Plant Science, 15(4), 227-235. doi: 10.1016/j.tplants.2010.01.004

227. Williams, L. E., Lemoine, R., & Sauer, N. (2000). Sugar transporters in higher plants - A diversity of roles and complex regulation. [Review]. Trends Plant Sci, 5(7), 283-290. doi: 10.1016/s1360-1385(00)01681-2

228. Wingender, E., Dietze, P., Karas, H., & Knuppel, R. (1996). TRANSFAC: A database on transcription factors and their DNA binding sites. Nucleic Acids Res, 24(1), 238-241. doi: Doi 10.1093/Nar/24.1.238

229. Wippel, K., Wittek, A., Hedrich, R., & Sauer, N. (2010). Inverse pH Regulation of Plant and Fungal Sucrose Transporters: A Mechanism to Regulate Competition for Sucrose at the Host/Pathogen Interface? PLoS ONE, 5(8), e12429. doi: 10.1371/journal.pone.0012429

230. Wittstock, U., & Gershenzon, J. (2002). Constitutive plant toxins and their role in defense against herbivores and pathogens. Curr Opin Plant Biol, 5(4), 300-307. doi: http://dx.doi.org/10.1016/S1369-5266(02)00264-9

231. Wright, D., Read, D., & Scholes, J. (1998). Mycorrhizal sink strength influences whole plant carbon balance of Trifolium repens L. Plant, Cell & Environment, 21(9), 881-891.

232. Wu, C., Sun, S., Nimmakayala, P., Santos, F. A., Meksem, K., Springman, R., . . . Zhang, H.-B. (2004). A BAC- and BIBAC-Based Physical Map of the Soybean Genome. Genome Research, 14(2), 319-326. doi: 10.1101/gr.1405004

233. Wu, X., Zhong, G., Findley, S., Cregan, P., Stacey, G., & Nguyen, H. (2008). Genetic marker anchoring by six-dimensional pools for development of a soybean physical map. BMC Genomics, 9(1), 28.

234. Wu, Y., Deng, Z., Lai, J., Zhang, Y., Yang, C., Yin, B., . . . Xie, Q. (2009). Dual function of Arabidopsis ATAF1 in abiotic and biotic stress responses. Cell Res, 19(11), 1279-1290. doi: 10.1038/cr.2009.108

235. Xia, N., Zhang, G., Liu, X. Y., Deng, L., Cai, G. L., Zhang, Y., . . . Kang, Z. S. (2010). Characterization of a novel wheat NAC transcription factor gene involved in defense response against stripe rust pathogen infection and abiotic stresses. Mol Biol Rep, 37(8), 3703-3712. doi: 10.1007/s11033-010-0023-4

236. Xie, Q., Frugis, G., Colgan, D., & Chua, N.-H. (2000). Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. Genes Dev, 14(23), 3024-3036.

237. Xu, P., Wu, X., Wang, B., Liu, Y., Ehlers, J. D., Close, T. J., . . . Li, G. (2011). A SNP and SSR Based Genetic Map of Asparagus Bean (<italic>Vigna. unguiculata</italic> ssp. <italic>sesquipedialis</italic>) and Comparison with the Broader Species. PLoS ONE, 6(1), e15952. doi: 10.1371/journal.pone.0015952

238. Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K., & Yokoyama, S. (2013). DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. Trends Plant Sci, 18(5), 267-276.

239. Yanagisawa, S. (1998). Transcription factors in plants: Physiological functions and regulation of expression. [Review]. Journal of Plant Research, 111(1103), 363-371. doi: 10.1007/bf02507800

240. Yang, B., Sugio, A., & White, F. F. (2006). Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. Proceedings of the National Academy of Sciences, 103(27), 10503-10508. doi: 10.1073/pnas.0604088103

241. Young, N. D., & Bharti, A. K. (2012). Genome-enabled insights into legume biology. Annual Review of Plant Biology, 63, 283-305.

242. Young, N. D., Cannon, S. B., Sato, S., Kim, D., Cook, D. R., Town, C. D., . . . Tabata, S. (2005). Sequencing the Genespaces of Medicago truncatula and Lotus japonicus. Plant Physiol, 137(4), 1174-1181. doi: 10.1104/pp.104.057034

243. Young, N. D., Debelle, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., . . . Roe, B. A. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. [10.1038/nature10625]. Nature, 480(7378), 520-524. doi: http://www.nature.com/nature/journal/v480/n7378/abs/nature10625.html#supplementary-information

244. Young, N. D., & Udvardi, M. (2009). Translating Medicago truncatula genomics to crop legumes. Curr Opin Plant Biol, 12(2), 193-201. doi: 10.1016/j.pbi.2008.11.005

245. Zeng, L., Zhang, Q., Li, S., Plotnikov, A. N., Walsh, M. J., & Zhou, M.-M. (2010). Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. [10.1038/nature09139]. Nature, 466(7303), 258-262. doi: http://www.nature.com/nature/journal/v466/n7303/abs/nature09139.html#supplementary-information

246. Zhang, G., Chen, M., Li, L., Xu, Z., Chen, X., Guo, J., & Ma, Y. (2009). Overexpression of the soybean GmERF3 gene, an AP2/ERF type transcription factor for increased tolerances to salt, drought, and diseases in transgenic tobacco. J Exp Bot, 60(13), 3781-3796.

247. Zhang, H., Jin, J., Tang, L., Zhao, Y., Gu, X., Gao, G., & Luo, J. (2011). PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. [Research Support, Non-U.S. Gov't]. Nucleic Acids Res, 39(Database issue), D1114-1117. doi: 10.1093/nar/gkq1141

248. Zhou, J., Tang, X., & Martin, G. B. (1997). The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. The EMBO Journal, 16(11), 3207-3218.

249. Zhou, Q. Y., Tian, A. G., Zou, H. F., Xie, Z. M., Lei, G., Huang, J., . . . Chen, S. Y. (2008). Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic Arabidopsis plants. Plant Biotechnology Journal, 6(5), 486-503.

Chapter two

Identification of Transcription Factors and Transcriptionally Active Protein Gene Families in Cowpea

Abstract

Cowpea, *Vigna unguiculata* (L.) Walp., is one of the most important food and forage legumes in semi-arid tropical regions. One of the most destructive causes of crop yield loss in cowpea is *Striga gesnerioides* (witch weed). Some cultivars of cowpea show resistance to some strains of *Striga*, and susceptibility to other strains; this makes cowpea an ideal subject of study for host-parasite interactions. One important aspect of such interactions pertains to the functions of transcription factors, which regulate many regulatory pathways, including pathways involved with defense. In this study, over 3,500 TFs were identified in the cowpea genome, and three TF families related to defense, namely AP2-EREBP, NAC, and WRKY, were investigated for their potential roles in the interactions between cowpea and *Striga*. Selected families are analyzed for phylogenetic comparison between cowpea and common bean (*Phaseolus vulgaris*). Although typical in size for a diploid legume, the distribution of TFs in cowpea significantly differs from related legumes *Phaseolus vulgaris* (common bean) and soybean. For example, cowpea is over-represented in the NAC and Aux-IAA families and under-represented in MADS and SET families. A phylogenetic comparison between cowpea and common bean using selected TF families shows that for the most part, cowpea and common bean TF families are phylogenetically similar, with the notable exception of the ABI3-VP1 family.

Introduction

Cowpea is one of the most important food and forage legumes in the world, especially in sub-Saharan Africa (Ehlers & Hall, 1997; Timko & Singh, 2008). According to 2011 estimates by the Food and Agriculture Organization of the United

Nations (FAO), approximately 4.9 million metric tons of cowpea was produced across about 10.4 million hectares worldwide (faostat3.fao.org). Most of this is produced in western and central Africa by subsistence farmers (faostat3.fao.org; Singh, 2005; Timko et al., 2007; Langyintuo et al., 2003). In these regions, the cowpea fruit is an important source of protein and carbohydrates (Nielsen et al., 1993; Hall et al., 2003; Singh, 2005; Timko et al., 2007). In addition to the fruit, leaves are also eaten (Nielsen et al., 1997), and cowpea stems are an effective fodder for livestock (Singh et al., 2003). Moreover, cowpea can be used to restore nitrogen to soils (Elowad & Hall, 1987), making it an effective companion crop to cereals (Oseni, 2010; Dahmardeh et al., 2010). Furthermore, cowpea can withstand dry conditions and low quality soils relatively well (Hall et al., 2002; Hall et al., 2004). Yet cowpea faces many stresses that can decimate growth and crop yields, including abiotic stresses such as drought, heat and salinity, and biotic stresses such as nematodes, fungal and bacterial diseases (Timko et al., 2007; Singh 2005; Roberts et al. 1996, 1997; Das et al. 2010). But among the worst threats to cowpea yield, biotic or abiotic, are parasitic weeds. Weeds such as *Striga gesnerioides* (Timko et al., 2007) can reduce crop yields significantly (Alonge et al., 2005a; Alonge et al., 2005b), sometimes up to 100% yield loss (Cardwell & Lane, 1995). Yet despite the importance of cowpea and the myriad threats to cowpea that exist, cowpea has not been studied as well as other crop plants (Timko & Singh, 2008). It is thus important that researchers use bioinformatics to compare cowpea to other legumes and study how cowpea responds to environmental stresses.

In doing that, one must be able to study the regulation of gene expression, a process that is central to many biological processes at all stages of the plant life cycle. Inevitably, such study will at some point involve the study of transcription factors (TFs), factors that regulate genetic expression by interacting with specific DNA sequences (MacQuarrie et al., 2011). TFs are active in many processes in plants, from floral development (Shore & Sharrocks, 1995), to growth and development of vascular tissue (Kubo et al., 2005; Zhao et al., 2008), to defense against a plethora of abiotic and biotic stresses (Singh et al., 2002). This makes TFs important for survival in plants and thus for crop yields (Rushton et al., 2008), and thus an important subject of study for plant biologists (Richardt et al., 2007; Udvardi et al., 2007; Rushton et al., 2008).

Among the earliest genome-scale analyses of TFs in plants were analyses done in *Arabidopsis thaliana* (Riechmann et al., 2000; Guo et al., 2005). Riechmann et al. (2000) identified 29 TF families in *Arabidopsis*, over half of which were unique to plants. Five years later, Guo et al. (2005) used newer bioinformatics analysis methods and information on TF families that was not available to Riechmann et al. (2000) to find that 64 TF families existed in *Arabidopsis*. From this, Guo et al. (2005) created the Database of Arabidopsis Transcription Factors (DATF). Other early genome-scale analyses of TFs in plants included rice (Gao et al., 2006) and poplar (Zhu et al., 2007). At that time, it was found that rice had 63 of the 64 families found in *Arabidopsis* (Gao et al., 2006), and that poplar and *Arabidopsis* shared the same TF families (Zhu et al., 2007).

Shortly thereafter, new TF databases were created that contained TF repertoires from across the kingdom Plantae. Such databases include PlantTAPDB (Richardt et al., 2007), PlnTFDB (Riano-Pachon et al., 2007; Perez-Rodriguez et al., 2010), and PlantTFDB (Guo et al., 2008; Zhang et al., 2011). These databases are excellent sources of information, but for most species not only are they are not based on whole genome assemblies; they mostly rely on data already in public resources and are thus unable to provide a comprehensive analysis of legumes.

Legume TF databases are relatively recent. In these databases, it is common for model legume species such as soybean (Wang et al., 2010; Grant et al., 2010) to be studied. Among the only databases available that contains relatively comprehensive data from multiple legume species is LegumeTFDB (Mochida et al., 2010). LegumeTFDB contains whole genome assembly based data for soybean, *Medicago truncatula*, and *Lotus japonicus*, but no TF data for other legumes such as common bean or cowpea.

Cowpea is a diploid legume with a genome size of 620 Mbp, which is smaller than other legumes (Arumuganathan & Earle, 1991; Timko et al., 2008). This should make cowpea a relatively easy legume to study. Yet for years, no attempt was made to comprehensively understand the cowpea genome. Then a major step forward in gaining such an understanding occurred when the gene-rich space of cowpea was sequenced and analyzed (Timko et al., 2008). The information gained from sequencing cowpea expressed sequence tags (ESTs), when added to the information from gene space reads (GSRs) allowed for analyses such as the discovery of syntenic relationships between cowpea and related legumes (Timko et al., 2008; Muchero et

al., 2009). Recently, a cowpea genome assembly draft (v0.02) has been used to find candidate genes for cowpea resistance against the fungal pathogen *Fusarium oxysporum* (Pottorff et al., 2012). This assembly has been updated to v0.03, and can be searched using BLAST on the HarvEST database (Wanamaker & Close, 2011; http://www.harvest-blast.org/; Close et al., 2007; Close et al., 2011). The assembly that exists now includes a 67x SOAP de novo assembly, a bacterial artificial chromosome (BAC)-end sequences, and data from 97% of previously known cowpea sequences, including GSR and EST sequences; a large proportion of this comes in the form of scaffold sequences, sequences that contain two or more overlapping contigs (Pottorff et al., 2012; Close et al., 2011). Therefore, this assembly can help gain a more comprehensive understanding of the cowpea and its TFs.

Yet the only genome-wide analyses of TFs that have been done for cowpea are an incomplete repertoire in PlantTFDB (Zhang et al., 2011), and the identification of 5,888 cowpea GSR sequences homologous to TFs and transcription associated factors (TAFs) (Timko et al., 2008). Thus there still exists a significant void in cowpea research. There is a great need for a comprehensive TF repertoire in cowpea; one ramification of knowing this repertoire will be fulfilling a long-standing need to identify TFs involved in host defense against parasitic plants. This study attempts to fill this void by identifying the TF distribution of cowpea, and comparing this TF distribution with the closely related common bean (*Phaseolus vulgaris*) and soybean (*Glycine max*). In addition, selected families are studied for phylogenetic organization, and six other families are selected for comparison with common bean,

which according to Lavin et al., (2005), is even more closely related to cowpea than soybean.

Materials and Methods

Sources of Data

In all bioinformatics analyses, the cowpea genome assembly used was the Cowpea genome v0.03 (Close et al., 2011; Wanamaker & Close, 2011), which is available for BLAST searches and sequence retrieval (www.harvest-blast.org, Pottorff et al., 2012).

Version 0.03 for cowpea genotype IT97K-499-35 (Close et al., 2011), was assembled from several sources: a SOAP de novo assembly with 67X coverage, 39 Gb of Illumina GAII paired-end sequences (70-130) using TrueSeq chemistry, 250,000 GSS sequences with an average length of 609 bases, 30,000 BAC-end sequences with an average length of 673 bases, and BLASTN hits for 97% of EST-derived consensus sequences from assembly P12 from www.harvest-blast.org. Because of this combination of sources, the cowpea version v0.03 assembly is the most comprehensive cowpea genomic assembly to date (Close et al., 2011).

The cowpea assembly was annotated and translated to protein using the MAKER annotation pipeline (Cantarel et al., 2008).

Two versions of the cowpea assembly were used, each of which was based on two methods of exon detection and translation to protein (Stephen Turner, personal communication, October 29, 2013): one in which the low-complexity regions were masked via RepeatMasker (http://www.repeatmasker.org/; Tempel et al., 2012) and RepeatRunner (http://www.yandell-lab.org/software/repeatrunner.html)

before being annotated via MAKER, and one that was simply annotated via MAKER and then processed through AUGUSTUS (Stanke et al., 2004; Stanke & Morgenstern, 2005).

Schmutz et al. (2010) sequenced the soybean genome assembly version 1.0; this assembly was updated to version 1.1 and uploaded to Phytozome v9.0 on December 13, 2012.

The Jackson group at the Phaseolus Genome Sequencing Project (Jackson et al., phytozome.net/commonbean.php; DOE-JGI, 2012) sequenced the common bean genome version 1.0, whose early release was uploaded to Phytozome v8 on August 7, 2012.

Identification and Classification of TFs

For each version of the assembly, TFs were found in the cowpea genome using a workflow created by Lawson & Mackey (2011) (Figure 1) in Taverna (Missier et al. 2010; Hull et al. 2006; Oinn et al., 2006); this pipeline uses HMMER 3 (Eddy, 2009; Eddy, 2011) for Hidden Markov models and PFAM (Punta et al., 2012; Sonnhamer et al., 1997; Bateman et al., 2000) for finding protein domains. Transcription factor classification was done using rules set by Lang et al. (2010).

The TF repertoires for each version of the cowpea whole genome assembly were collapsed into a non-redundant set using the following process: Sequences with identical names were found in both versions of the assembly. It was observed that sequences with identical names had identical sequences. For example, a sequence with a name like “scaffold8708-processed-0.1” in one set would be found as identical to the sequence “scaffold8708-0.1” in the other dataset. After that,

redundant names were eliminated in Excel, and custom Perl scripts were used to retrieve the sequence set based on the non-redundant names.

Comparison of Cowpea TF Repertoire to Common Bean and Soybean. The procedure used to identify and classify TFs in cowpea was also applied to common bean and soybean. As described in the "Sources of Data" subsection, these genome sequences were provided on Phytozome (Goodstein et al., 2012).

Multiple Sequence Alignment of TF Sequences

Multiple sequence alignment was done using T-Coffee (Di Tommaso et al., 2011) algorithms Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena & Notredame, 2009; Chang et al., 2012). For both methods, the options to combine the alignment with all other known alignment methods made available on the T-Coffee web server (Notredame, tcoffee.crg.cat) was chosen in order to create a more accurate alignment based on a consensus between multiple methods (http://www.tcoffee.org/Documentation/t_coffee/t_coffee_tutorial.htm). The alignments from both methods were combined using the T-Coffee Combine algorithm (Di Tommaso et al., 2011). MEGA 5.2.2 (Tamura et al., 2011) was used for manual curation of the alignment and for visualizing the alignment.

Phylogenetic Analysis of TF Families

After alignment, Format Converter v2.0.5 at Los Alamos National Laboratory site (http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/ form.html) was used to convert FASTA sequence files into PHYLIP interleaved format. The file was

then manually adjusted so that the file was in the proper input format for ProtTest 3 (Darriba, et al. 2011) and and PhyML 3.0 (Guindon et al., 2010).

Model testing for maximum likelihood phylogeny was done using ProtTest 3 (Darriba et al., 2011) in order to find the best fitting amino acid substitution model for the aligned sequences. Phylogenetic trees were made using PhyML 3.0 (Guindon et al., 2010), using the approximate likelihood ratio test (aLRT) (Anasimova & Gascuel, 2006), and automatically selecting the better choice between nearest-neighbor interchange (NNI) and subtree pruning and regraphing (SPR), and automatically selecting the better choice between Chi-square or Shimodaira-Hasegawa (SH)-like branch supports (Guindon et al. 2010). Approximate likelihood ratio test was chosen because it has a similar accuracy to bootstrap, but much faster (Anasimova & Gascuel, 2006). Unless otherwise specified, phylogenetic trees were visualized using FigTree v1.4.0 (Rambaut, 2013; http://tree.bio.ed.ac.uk/software/figtree/).

The aLRT method. The aLRT is a parametric method that compares the likelihoods of the two best arrangements around a certain branch (Anisimova & Gascuel, 2006; Guindon et al., 2010); the aLRT does this by using a log ratio of the likelihood of the current arrangement to that of the best alternative (Guindon et al., 2010). The aLRT can be accurate, powerful and robust to mild violations of substitution models, but its greatest advantage is its great speed compared to traditional bootstrap (Anisimova et al., 2006). To minimize violations of substitution models, ProtTest 3 (Darriba et al., 2011) is used before building a maximum-likelihood phylogenetic tree. To further ensure that branches were accurate, branch supports were added

to the aLRT test. PhyML 3.0 offers a choice between chi-square parametric and the non-parametric SH-like branch supports (Guindon et al., 2010).

Deeper Analysis of AP2-EREBP, NAC and WRKY families. After alignment, the aligned set was manually curated in order to show the conserved domains for the AP2-EREBP, NAC and WRKY families. For the AP2-EREBP family, selected sequences from each clade of the *Arabidopsis* AP2-EREBP superfamily analyzed in Dietz et al. (2010) and from each clade of the rice AP2-EREBP superfamily analyzed in Sharoni et al. (2011) were used to determine the clades to which cowpea AP2-EREBP sequences belonged. To analyze the NAC family, sequences from each clade of the NAC family analyzed by Zhu et al. (2012) were used to classify cowpea NAC sequences into clades. To analyze the WRKY family, sequences from each clade of the WRKY family analyzed by Li et al. (2012) were used to classify the cowpea WRKY sequences into clades.

Results

Identification of TFs

The cowpea genomic assembly v0.03 (Close et al., 2011), which as noted in Materials and Methods is based on a combination of GSS, EST, BAC-end sequences, a 67x SOAP de novo assembly, and Illumina GAII paired-end sequences, was found to have 3573 TFs, which represents approximately 7% of the cowpea genome. The largest TF family in cowpea is PcG-FIE, with 192 members, followed by bHLH (190), MYB (180), AP2-EREBP (174), bZIP1 (170), and NAC (152). There is a similar number of TFs between cowpea and common bean (3404), and cowpea has almost half the TFs as soybean (7980) (Table 1). This finding is deceptive because in

actuality, the TF content in cowpea for most individual TF families is significantly different from common bean and soybean. Firstly, according to the pipeline by Lawson & Mackey (2011), CCAAT-HAP3 and Dicer are not represented in cowpea, but are represented in common bean and soybean. In addition, CCAAT-Dr1, Runt and TEA are represented in cowpea (with CCAAT-Dr1 and TEA having 1 member and Runt having 2 members), but not common bean or soybean.

When each cowpea TF family was compared to its counterparts in common bean and soybean with respect to percentage of their respective TF repertoires, common bean and soybean showed significant difference from cowpea (Supplementary Table S1).

In percentage of TF repertoires, 25 TF families in cowpea are 20-75% the size of their counter-parts in bean. The largest families in this category are ABI3-VP1, MADS and SET. 21 families in cowpea are 75-90% the size of their counterparts in bean. The largest of these families are bHLH, AP2-EREBP, PHD and C2H2. 17 families are similar in size (90-110%) to their counterparts in bean. The largest families in this category were MYB, MYB-related, and bZIP2. 11 families in cowpea are 10-25% larger than their bean counterparts. The largest of these are WRKY and PcG-FIE. 15 TF families in cowpea are 25-70% larger than their counter-parts in bean. The largest families in this category are bZIP1, NAC and SWI/SNF_SNF2. 12 families in cowpea were at least 90% larger, the largest being DUF246 and Aux/IAA.

Compared to soybean, in terms of percent of TF repertoire, only 1 family is 83% smaller than in soybean, the Pseudo-ARR-B family. 28 families in cowpea are 35-75% as large as in soybean; the largest of which are MADS and SET. 16 families in

cowpea are 75-90% as large as in soybean; the largest of which is bHLH. 21 cowpea families were similar (90-110%) to soybean. Several families were in this category, namely bZIP2, AP2-EREBP, PHD, MYB-related, C2H2, PcG-FIE, and MYB. 13 cowpea families are between 10 and 25% larger than their soybean counterparts, the largest being bZIP1 and WRKY. NAC, DUF246, and Aux/IAA were the largest of 14 cowpea families that were in terms of percentage of repertoire, between 25-80% larger than their soybean counterparts. 8 families in cowpea were at least 100% larger than in soybean, none of which exceeded 15 members in cowpea.

Phylogenetic Analysis of TF Families

The AP2-EREBP, NAC and WRKY families were chosen for deeper analysis due to their involvement with defense response in plants (Singh et al., 2002; Olsen et al., 2005; Van Verk et al., 2009) and because each family has easily recognizable conserved domains (Ohme-Takagi & Shinshi, 1995; Olsen et al., 2005; Eulgem et al., 2000; Rushton et al., 2010). After a multiple sequence alignment by Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena & Notredame, 2009; Chang et al., 2012) combined via the T-Coffee Combine algorithm (Di Tommaso et al., 2011), a phylogenetic analysis using maximum likelihood via PhyML 3.0 (Guindon et al., 2010) was performed on each of the three families.

AP2-EREBP. Ohme-Takagi & Shinshi (1995) first discovered the AP2-EREBP superfamily in tobacco. This superfamily is known for its roles in multiple processes, from developmental roles (Byzova et al., 1999), to defense against abiotic and biotic stresses (Nakano et al., 2006; Xu et al., 2011). The AP2-EREBP superfamily consists of three families: ERF, RAV and AP2 (Riechmann &

Meyerowitz, 1998; Kagaya et al., 1999). The ERF family has distinguishing conserved motifs such as an N-terminal AEIRD motif and a WLG (Riechmann & Meyerowitz, 1998). The RAV family has an AP2 and a B3 domain (Kagaya et al., 1999). The AP2 family usually has two AP2 domains (Riechmann & Meyerowitz, 1998). The first (N-terminal) domain, known as the R1 repeat, usually has a YEAH or WESHI at the 5' end and a YDRAA or LAALKY at the 3' end, whereas the second (C-terminal) domain, known as the R2 repeat, has a WQAR or WEAR at the 5' end and a NAVT or YDIAAI at the 3' end (Riechmann & Meyerowitz, 1998; http://compsysbio.achs.virginia.edu/tobfac/browse_family.pl?family=AP2). The AP2 family also has a conserved YLG instead of the WLG and AEIRD found in ERFs (Riechmann & Meyerowitz, 1998; http://compsysbio.achs.virginia.edu/tobfac/browse_family.pl? family=AP2). Based on the knowledge of these conserved motifs, for the AP2-EREBP TF superfamily, the separation of AP2-EREBP into its component families, the ERF, RAV and AP2 families, was based on a search against Pfam (Punta et al., 2012) and on visual inspection of the sequences in the AP2-EREBP sequence set. This separation was done so as to create more satisfactory alignments and phylogenetic trees on each of the three families in the AP2-EREBP superfamily.

ERF. A sample from a multiple sequence alignment of conserved ERF domains in cowpea along with *Arabidopsis* sequences from Dietz et al. (2010) and rice sequences from Sharoni et al. (2011) is shown in Figure 2. The *Arabidopsis* and rice sequences were chosen in order to facilitate the classification of ERF sequences into clades. The phylogenetic tree for the ERF family shown in Figure 3 is based on

this alignment, and shows that the 137 ERF cowpea TFs found in this study, when grouped using the categories from Nakano et al. (2006), namely groups I-X and VI-L (Figure 3b), shows that no group formed two separate clades (Figure 2b). When cowpea ERFs were grouped according to the grouping scheme in Dietz et al. (2010), namely groups DREB A1-A6 and ERF B1-B6, most groups formed two separate clades, with the ERF B-6 group separating into three clades (Figure 3a).

AP2. After the discovery of 34 AP2 TFs in cowpea, conserved AP2 domains in cowpea, *Arabidopsis* (Dietz et al., 2010), and rice (Sharoni et al., 2011) were alilgned (Figure 4); from this alignment, a phylogenetic tree was produced (Figure 5) using the same methods used to produce the phylogenetic tree for the ERFs.

RAV. Three RAV sequences were found in cowpea, and then aligned and made into a phylogenetic tree (Figure 6).

NAC. The NAC family is quite functionally diverse, with functions ranging from defense against biotic and abiotic stresses, to hormone signaling, to reproduction (Olsen et al., 2005). A NAC TF typically consists of five conserved regions (Ooka et al., 2003). The alignment in Figure 7 shows that cowpea usually has all five. This alignment includes NAC domains from Zhu et al. (2012) for easier identification of NAC clades.

Different plants will have NAC families with different groups, depending on whether the plant is a dicot, a monocot, a moss or a lycophyte (Zhu et al., 2012). According to the classification scheme proposed by Zhu et al. (2012), dicotyledonous plants usually have groups Ia-c, II, IIIa-c, IVa-d, Va(1), Va(2), Vb, VIa, VIc, VII, and VIII. The cowpea NAC family has all of these groups, except for group VIII (Figure 8).

WRKY. The WRKY TF family is characterized by a conserved N-terminal WRKYGQK motif and a motif resembling a zinc finger (Rushton et al., 1995; Eulgem et al., 2000). Yet variations in the conserved parts of the WRKY domain are such that within the WRKY family, there are three major groups (I-III) (Eulgem et al., 2000). Group I WRKY TFs have two WRKY domains, with their C-terminal domains being functionally distinct from the N-terminal domains (Li et al., 2012). Group II is the most variable group in terms of amino acid sequence, with five subgroups (Eulgem et al., 2000; Rushton et al., 2010; Li et al., 2012). Group III WRKYs differ in zinc finger structure from group-I and -II WRKY TFs; a group III WRKY zinc finger has a C_2-HC structure, as opposed to the C_2-H_2 in the other two WRKY groups (Eulgem et al., 2000; Rushton et al., 2010).

In Figure 9, a sample from a multiple alignment of WRKY conserved domains from cowpea, along with selected conserved WRKY domains from *Arabidopsis* and rice from each group and subgroup used in Li et al. (2012), subtly shows some of the differences between the three groups of WRKY TFs, although these differences are more explicit in the phylogenetic tree in Figure 10. Overall, WRKY TFs group neatly, with the exception of the Group I C-terminal domain group, which is separated into two clades by Group I N-terminal domain group.

Phylogenetic Comparison between Cowpea and Common Bean TF Families

Six families were chosen for phylogenetic comparison between cowpea and common bean: two families in which cowpea are under-represented compared to common bean (GRF and ABI3-VP1), two families in which the two legumes are

similar (EIL and tify), and two families in which cowpea is over-represented (HMG and Tub). The trees for these families are displayed in Figures 11-16.

Discussion

In this study, the TF repertoire was identified in cowpea and then compared to TF repertoires from common bean and soybean. The data in this study show that the TF repertoire in cowpea is similar in total number of TFs to common bean and half as large as soybean; however, when the TF families in cowpea, common bean and soybean were compared as percentages of their respective repertoires, it was found that TFs in cowpea are distributed in a manner that is significantly different from common bean and soybean. When the AP2-EREBP, NAC and WRKY families were investigated, it was found that for the most part, all three families in cowpea were organized in a similar way to corresponding families in other plants.

TFs in Cowpea and Related Legumes

From the close relationships between cowpea and other legumes, one would expect that the TF distribution in cowpea would be similar to common bean and soybean. Both cowpea and common bean are diploid plants (Arumuganathan & Earle, 1991; Gepts et al., 2008; McClean et al., 2008), whereas soybean is a paleopolyploid legume with an ancient tetraploid origin (Schmutz et al., 2010; Gill et al., 2009; Wu et al., 2004). Cowpea is more closely related to common bean than to soybean (Lavin et al., 2005), and cowpea is shown to have strong synteny with soybean (Muchero et al., 2009). Since common bean also has strong synteny with soybean (McClean et al., 2010; Galeano et al., 2009), one can expect that common bean also shares strong synteny with cowpea, and therefore has a similar TF distribution to

cowpea. The finding of TF families that significantly differed in size between cowpea and common bean contrasts with this expectation, but the finding that TF families were overall phylogenetically very similar between the two legumes is consistent with the expectation of similar TF distribution between them.

When Timko et al. (2008) sequenced and analyzed the gene-rich space of the cowpea genome, 5,888 GSR sequences homologous to TFs and TAFs were found in 62 families, which accounted for 5% of the cowpea gene space reads, which is a much smaller percentage of protein-coding genes than in soybean, where 5,671 TFs were found to account for 12.2% of all protein-coding genes in soybean (Schmutz et al., 2010).

The cowpea assembly v0.03 contains a 67X de novo assembly and Illumina reads, as well as GSS, EST, and BAC-end sequences (Close et al., 2011; Wanamaker & Close, 2011). The findings in this study thus contrast with earlier studies because in this study, 3,573 TFs were found in cowpea, which is 39% fewer than the number found in Timko et al. (2008). Moreover, the number of TFs found in this study is estimated to account for approximately 7% of all protein-coding sequences in the cowpea assembly v0.03. The contrasts in number of TFs represented may be due to the cowpea assembly v0.03 being an assembly that contains scaffold sequences, which contain overlapping contigs. It is possible that a large proportion of contigs in the GSR assembly were overlapping in sequence and therefore incorporated into the scaffolds in the cowpea assembly v0.03 (Close et al., 2011; Wanamaker & Close, 2011).

Besides the differences in numbers of TFs between this study and Timko et al. (2008), there are 41 more TF families represented in this study than in Timko et al. (2008), in part because the Lawson & Mackey (2011) pipeline in this study uses the classification scheme described in Lang et al. (2010), which includes TF families that were discovered after the publication of Timko et al. (2008).

The classification scheme in this study and the pipeline used to identify TFs (see Materials and Methods) also finds 3,404 TFs in common bean and 7,980 TFs in soybean. These findings contrast with a finding by Kalavacharla et al. (2011) that there existed 2,516 putative TFs in common bean, as well as the report by Schmutz et al. (2010) that 5,671 TFs existed in soybean. This could be because the assemblies used in this study are more recent versions of the common bean and soybean assemblies, and possibly because our TF identification scheme (see Materials and Methods) as executed by a Taverna pipeline used by Lawson & Mackey (2011) that involves Hidden Markov Models (HMMs) from HMMER 3 (Eddy, 2009; Eddy, 2011) and Pfam (Punta et al., 2012; Sonnhamer et al., 1997; Bateman et al., 2000) and custom HMMs from Lang et al. (2010), significantly differs from the methods in Kalavacharla et al. (2011), which involved a BLAST search against *Arabidopsis* proteins and the identification of sequences homologous to *Arabidopsis* TFs from PlnTFDB (Perez-Rodriguez et al., 2010); and the methods of Schmutz et al. (2010), which involved annotations using Fgenesh+ (Salamov & Solovyev, 2000) and GenomeScan (Yeh et al., 2001).

Therefore, the TF data in this study are comprehensive and will thus prove to be a valuable resource for studies in processes such as transcriptional regulation, signaling pathways, and defense mechanisms.

Phylogenetic Analysis of Selected TF Families

AP2-EREBP. Over the years, the AP2-EREBP family has been investigated in several plants, from *Arabidopsis* and rice (Nakano et al., 2006; Dietz et al., 2010; Sharoni et al., 2011), to tobacco (Rushton et al., 2008), to maize (Zhuang et al., 2010) and soybean (Zhang et al., 2008). These studies had the objectives of providing more comprehensive information on the AP2-EREBP family of the organism studied, as well as to provide functional analyses on various AP2-EREBP members (Nakano et al., 2006; Dietz et al., 2010; Sharoni et al., 2011; Rushton et al., 2008; Zhang et al., 2008; Zhuang et al., 2010). Some of these studies placed AP2, ERF and RAV into the same phylogenetic tree (Dietz et al., 2010; Zhuang et al., 2010), while some others exclusively investigate the ERF family (Nakano et al., 2006; Zhang et al., 2008). Sharoni et al. (2011) separates the AP2-EREBP into two groups: the ERF, and AP2/RAV.

ERF.

Phylogeny. As mentioned in the Results section above, two different classification schemes have emerged: one in which the ERF family is separated into twelve groups, DREB A1-A6 and ERF B1-B6 (Dietz et al., 2010); and the other in which the ERF family is separated into groups I-X and VI-L (Nakano et al., 2006). When the twelve-group DREB/ERF classification used in Dietz et al. (2010) and Sharoni et al. (2011) was used, it was found that many groups in cowpea are

polyphyletic. Using this classification scheme, Dietz et al. (2010) found polyphyletic groups in the ERF family of *Arabidopsis*, although not to the extent found in cowpea in this study. These polyphyletic groups were not found in studies of the ERF family in soybean (Zhang et al., 2008) or rice (Sharoni et al., 2011).

When the Nakano et al. (2006) classification scheme was applied, it was found that cowpea ERFs group neatly into the groups defined by Nakano et al. (2006), although in this study, the Xb-L group was absent in cowpea. This tree is similar in organization to the tree based on cowpea GSRs in Timko et al. (2008) in that groups I-IV were all in the DREB clade, and groups VI, VIII and IX were all in the ERF clade. However, in the study of ERFs based on cowpea GSRs, Timko et al. (2008) found that group V ERFs separated into two clades: one in the DREB group and the other in the ERF group. In addition, group IX ERFs in the Timko et al. (2008) study are separated into two separate clades, with the group VII ERFs between them. In this study, by contrast, all group V ERFs were found to be in the ERF group and not in DREB; and group IX are in one clade adjacent to group VII, and not split into two clades with group VII in between. These contrasts could be due in part to the use of a whole-genome assembly in this study as opposed to the GSRs used in Timko et al., (2008). Moreover, this study made use of maximum-likelihood phylogenetic trees based on alignments based on a combination of Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena & Notredame, 2009; Chang et al., 2012) in this study, unlike the trees in Timko et al. (2008), which were built using the neighbor-joining method (Saitou & Nei, 1987) based on alignments in ClustalW (Thompson et al., 1994). In this study, the combination of Expresso and PSI-Coffee was chosen because the use

of consistency-based methods of multiple sequence alignment, a well as a combination of structure-based alignments and profile-based alignments has been recommended as being more accurate than matrix-based methods (Kemena & Notredame, 2009; Notredame, 2007). Maximum likelihood was chosen in this study for the phylogenetic analysis of TFs because maximum likelihood, although computationally intensive, has strong statistical foundations and is thus more powerful than neighbor joining (Sleator, 2011).

Another contrast between the ERFs in this study and the ERFs in Timko et al. (2008) was that the ERF group VII sequences in this study contained both ERFs and DREBs, unlike in Timko et al., (2008). The rice sequences used in this study to classify cowpea ERFs were from Sharoni et al. (2011), whose classification scheme contrasts with that of Nakano et al. (2006), as explained below.

The grouping of the cowpea ERFs into the groups characterized by Nakano et al. (2006) was much neater than the grouping into DREB A1-A6 and ERF B1-B6 in Dietz et al. (2010) and Sharoni et al. (2011). More specifically, the classification in Nakano et al. (2006) was exactly consistent with the classification in Dietz et al. (2010), but contrasted with the classification in Sharoni et al. (2011). For example, a Group X ERF in rice according to Nakano et al. (2006) was classified as a DREB A5 in Sharoni et al. (2011); according to Nakano et al. (2006), Group X ERFs represent ERF group B-3 and B-4. In another example, two DREB sequences from Sharoni et al. (2011) were classified as group VII in Nakano et al. (2006), when the group-VII ERFs in Nakano et al. (2006) were classified as ERF B-2. This could explain why some group VII cowpea ERFs grouped with the DREBs in this study, unlike in Timko

et al. (2008). The details of the contrast between the classification schemes of Nakano et al. (2006) and Sharoni et al. (2011) are outlined in Supplementary Table S6.

NAC.

Phylogeny. Methods of classifying NAC TFs into groups vary greatly throughout the literature. One classification scheme by Ooka et al. (2003), which is based on NAC TFs from *Arabidopsis* and rice, shows NAC belonging to two broad cagegories (I and II). In group-I exist fifteen groups named after certain members, such as ATAF2 or Senu5. In group-II exist three groups, again named after certain members, such as ONAC003. Rushton et al. (2008) classify tobacco NACs by simply numbering clades 1 through 6, and then describing three clades unique to Solanaceae. Zhu et al. (2012) classify NACs into ten numbered groups (I-X), some of which contain several subgroups.

The cowpea NAC TFs in this study are classified according to Zhu et al. (2012), because Zhu et al. (2012) base their classification on several plants from four groups throughout the Kingdom Plantae: mono- and dicots, lycophytes and mosses. Moreover, Zhu et al. (2012) use a more accurate methodology than in other studies. The phylogenetic trees constructed in Zhu et al. (2012) to illustrate the groups of NACs are based on alignments from HMMER 3 (Eddy 1998) and PFAM (http://pfam.sanger.ac.uk), which are manually curated in BioEdit. The trees in Zhu et al. (2012) were constructed using maximum likelihood and Bayesian analysis. This contrasts to other studies (Ooka et al., 2003; Rushton et al., 2008), which use Clustal algorithms (Thompson et al., 1994; Thompson et al., 1997) for alignment and

the neighbor-joining method (Saitou & Nei, 1987) for phylogenetic trees, which are not as accurate (Kemena & Notredame, 2009; Notredame, 2007; Sleator, 2011).

The differences that existed between the cowpea NAC family and the NAC family in Zhu et al. (2012) were as follows: first, group II in cowpea, although closely resembling the group II of Zhu et al. (2012), does not group between I and III like in Zhu et al. (2012). Second, group VIa in cowpea groups between groups III and IIIb. Cowpea groups IVb through IVd group together, but they lie adjacent to group Ic. Only one group VII NAC exists in cowpea, and it groups between groups Ib and Ic. Cowpea completely lacks group VIII. Finally though the rice sequence Os01g70110 (Figure 8) belongs to the monocot-only clade VIb (Zhu et al., 2012), there is a group of cowpea sequences that phylogenetically grouped with that sequence. Since cowpea is a dicot, it is more likely that either those sequences are part of group VIc, or that those sequences are part of a clade of sequences that could be unique to legumes. In this study, the sequences are labeled as VIc due to their activity upon *Striga* parasitism (see Chapter 3).

It could be argued some legume NAC families are organized differently from other dicots. Another example of a legume NAC family being organized differently from other dicots can be illustrated in the differences between Pinhiero et al. (2009) and Ooka et al. (2003); one of these differences is that clades Senu5 and TERN (which are VIa and IVc, respectively, in Zhu et al. (2012)), group much farther apart than in Ooka et al. (2003).

When the NAC families of cowpea from this study, the soybean NACs from Pinhiero et al. (2009) (whose groups are identified with their closest equivalents in Zhu et al.,

2012; please note that some groups in Zhu et al. (2012) had no equivalents in Pinhiero et al. (2009)), and the NACs from other dicots (Zhu et al., 2012) are compared from closest to furthest from the root of their respective trees, the following orders are found:

This study: II -> IVc -> IVb -> IVd -> Group Ic -> Group VII -> Group Ib -> Group Ia -> IIIc -> IIIb -> VIa -> IIIa -> IVa -> Va(1) -> Va(2) -> Vb -> VIc.

Pinhiero et al. (2009): IVc -> Ib -> Ic -> III(a) -> IVa -> Ia -> IVb -> VIa -> Va(2) -> Va(1) -> Vb.

Zhu et al. (2012) (maximum likelihood): Ia -> Ib -> Ic -> II -> IIIa -> IIIb -> IIIc -> Vb -> Va(2) -> Va(1) -> VIa -> VIb -> VIc -> IVb -> IVc -> IVd -> IVa -> VIII -> VII -> IX -> X.

The comparison above shows that the cowpea NAC TFs in this study are somewhat closer in organization to the soybean NAC TFs in Pinhiero et al. (2009) than the organization in Zhu et al. (2012); this further suggests that legume NAC families are organized differently from NAC families in other plants, including other dicots.

WRKY.

Phylogeny. As stated in the Results section, the WRKY TF family is grouped into three major groups, I – III (Eulgem et al., 2000). Since group I WRKY TFs have two WRKY domains, those who analyze WRKYs typically separate group I into N-terminal and C-terminal domains (Eulgem et al., 2000; Li et al., 2012; Rushton et al., 2008; Timko et al., 2008). Group II WRKYs are separated into five subgroups, a – e (Eulgem et al., 2000; Rushton et al., 2010), and Group III has a different zinc finger structure than the other two groups (Eulgem et al., 2000; Rushton et al., 2010).

The cowpea WRKY family in this study has all documented groups (Figure 11). Four of the subgroups of Group II in this study, namely IIa, IIb, IId and IIe, group together, as in the WRKY family based on GSRs in the Timko et al. (2008) study. However, unlike in Timko et al. (2008), the group I C-terminal group in this study forms two separate clades, separated by the group I N-terminal group. Moreover, the group IIb in Figure 11 does not separate into two distinct clades like in Timko et al. (2008). In the Timko et al. (2008) study, group IIb is split into two clades due to an artifact in the ClustalW alignment caused by truncated WRKY domains. This split did not happen in this study, probably because the alignment used in this study was based on the combination of two T-Coffee programs, namely Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena & Notredame, 2009; Chang et al., 2012). In addition, the sequences in this study were scaffold sequences, which were likely to contain two or more GSR or EST sequences, and hence would be more likely to have a complete WRKY domain.

Phylogenetic Comparison Between Cowpea and Common Bean

Overall, TF families in cowpea and common bean are similar in their phylogenetic organization. This is to be expected since cowpea and common bean are close together in the Millettoid (i.e., Phaseoloid) clade of legumes (Lavin et al., 2005). The EIL, GRF, HMG, tify and TUB families in cowpea were all similar in phylogenetic organization to common bean. As noted above, the GRF family is underrepresented in cowpea, while the HMG and TUB families are overrepresented in cowpea. Thus if a TF family in cowpea is similar in phylogenetic organization but larger in size than in common bean, it may be possible that there are instances of gene duplication

(without whole-genome duplication) in that TF family in cowpea; likewise a TF family that is larger in common bean but phylogenetically similar to cowpea may show instances of gene duplication for that TF family in common bean. Instances of such processes have been known to occur throughout plant evolution (Lang et al., 2010; Moore & Purugganan, 2005), and they may lead to the presence of sequences with new functions (Moore & Purugganan, 2005).

Among the families analyzed, the ABI3-VP1 family in cowpea showed some differences in phylogenetic organization from its counterpart in common bean. It is possible that different forms of duplication on a small scale are involved here.

ABI3-VP1. The ABI3-VP1 TFs, characterized by a B3 domain and the absence of auxin response factor and AP2 domains (Lang et al., 2010), take part in the abscisic acid (ABA) signaling pathway (Rolland et al., 2006; Nakamura et al., 2001; Shiota et al., 1998). They also take part in the auxin signaling pathway, acting both upstream and downstream of auxin (Nag et al., 2005). The ABI-VP1 family mainly functions in seed development (Bassel et al., 2006; Luerssen et al., 1998; Shiota et al., 1998). Figure 15 shows that there exists a group in cowpea that does not exist in common bean, and that there is a group in common bean that does not exist in cowpea. Here, it must be noted that there exist multiple types of B3-domain containing sequences. Romanel et al. (2009) identify five types: ABI3, which has one B3 domain; HSI, which had a B3 domain and a zf-CW domain; ARF, which can contain a B3 domain along with an auxin response factor and an Aux/IAA domain (Lang et al., 2010; Romanel et al., 2009); RAV, which contains an AP2 and a B3 domain (Kagaya et al., 1999; Romanel et al., 2009); and REM, which contains two B3

domains. By these definitions, the types of B3-domain containing sequences classified into ABI3/VP1 in this study are almost invariably ABI3 or REM. Cowpea and common bean have different proportions of each: whereas cowpea has 40 ABI3 and 18 REM sequences, common bean has 38 ABI3 and 42 REM sequences (Supplementary Table S12). This proportion is reflected in the phylogenetic tree of Figure 15; the group unique to cowpea almost exclusively has ABI3 sequences, whereas the group unique to common bean consists of virtually all REM sequences. Thus there may be scattered instances of duplication of REM sequences in common bean, possibly including the types leading to sequences with new functions (Moore & Purugganan, 2005). Also, since REM sequences are known to change more rapidly and ABI3 sequences are more conserved (Romanel et al., 2009), it can be argued that the ABI3/VP1 TF family in common bean is evolving more rapidly than in cowpea.

Given that ABI3-VP1 TFs are highly expressed in seed development (Bassel et al., 2006; Luerssen et al., 1998; Shiota et al., 1998), it may be that cowpea and common bean use B3 domain proteins differently, especially in seed development. And since ABI3 proteins can be regulated by sugar signaling as well (Rolland et al., 2006), it is also possible that sugar signaling differs between cowpea and common bean.

Limitations

TF Identification and Classification. Interestingly, some families, such as CCAAT-HAP3, are fewer in number than expected (Table 1), especially given that other organisms were found to have several CCAAT-HAP3 sequences, like the fifteen

CCAAT-HAP3 sequences found in tobacco (Rushton et al., 2008). This could be due to an artifact in the pipeline used in this study to identify and classify TFs (Lawson & Mackey, 2011): when the pipeline found a sequence with required domains for two or more TF families, the pipeline would sometimes not assign a TF family to the sequence. For example, a sequence with an NF-YB domain, a required domain of the CCAAT-HAP3 TF family, may have also contained an NF-YC domain. In this situation, the pipeline did not classify the sequence into any particular TF family. One possible course of action is to adjust the pipeline classification rules to account for the multiple domains in a sequence. This could mean, for example, treating certain domains as families, such as assigning a sequence with both NF-YB and NF-YC domains to NF-YB and NF-YC families simultaneously.

Conclusions

Overall, this study fulfills a long-standing need for a comprehensive cowpea TF repertoire based on a whole genome assembly. It was found that the number of TFs in cowpea is typical for a diploid dicotyledonous plant, but is unique in composition when compared to common bean and soybean. It was also found that the phylogenetic organization of TF families in cowpea are mostly similar to their counterparts in common bean, whether or not the number of members of certain TF families in cowpea were significantly different in size compared to their counterparts in common bean.

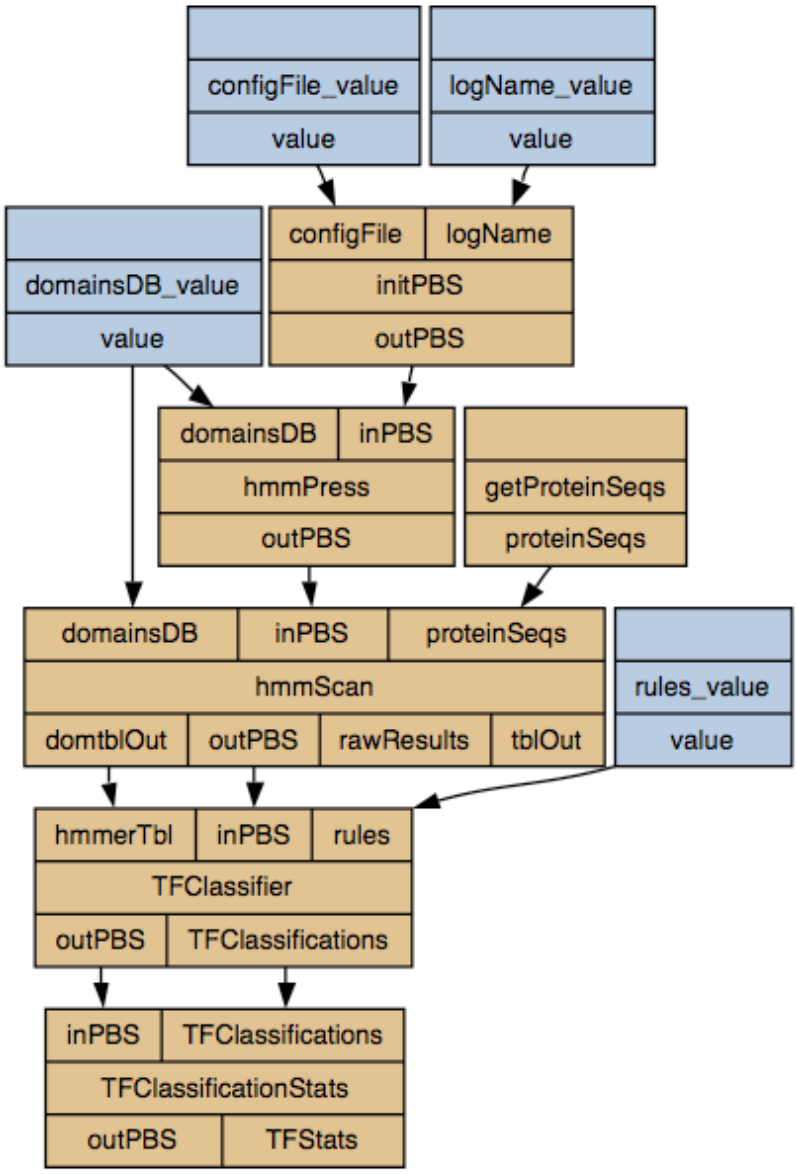


Figure 1. The TavernaPBS workflow used to classify TFs in cowpea, common bean and soybean. Figure adapted from Lawson & Mackey (2011).

Family	Cowpea	Common Bean	Soybean
ABI3/VP1	60	80	128
Alfin	4	2	13
AP2/EREBP	174	190	416
ARF	45	33	93
Argonaute	5	18	33
ARID	19	19	38
AS2/LOB	58	55	118
Aux/IAA	79	35	102
BBR/BPC	10	12	29
BES1	9	8	19
bHLH	190	206	484
bHSH	3	1	2
BSD	10	11	30
bZIP1	170	119	323
bZIP2	114	105	277
C2C2_CO-like	8	15	39
C2C2_Dof	34	45	93
C2C2_GATA	28	30	65
C2C2_YABBY	17	14	53
C2H2	108	123	236
C3H	93	92	238
CAMTA	4	9	23
CCAAT_Dr1	1	0	0
CCAAT_HAP2	13	19	57
CCAAT_HAP3	0	3	7
CCAAT_HAP5	19	16	37
Coactivator_p15	7	4	11
CPP	8	9	28
CSD	11	7	27
DBP	6	5	13
DDT	12	20	28
Dicer	0	5	15
DUF246	95	43	128
DUF296	39	48	107
DUF547	23	29	70

Table 1 (next page). Size distribution of the TFs and TAPs found in cowpea, common Bean and soybean. The estimated number of members identified in each of the 107 families of TFs found (each family identified by abbreviated name) in cowpea, common bean and soybean.

Family	Cowpea	Common Bean	Soybean
DUF632	37	22	48
DUF833	7	2	6
E2F/DP	11	14	31
EIL	8	7	16
FHA	20	23	49
GARP_ARR-B_G2	22	17	43
GARP_ARR-B_Myb	21	17	42
GARP_G2-like	12	2	5
GeBP	8	11	23
GIF	7	11	21
GNAT	62	43	110
GRAS	68	63	140
GRF	7	13	33
HB	88	82	203
HB_KNOX	15	21	46
HD-Zip_bZIP_1	10	7	19
HD-Zip_Halz	45	39	90
HMG	11	8	35
HRT	1	2	1
HSF	46	35	63
IWS1	16	18	39
Jumonji	7	9	21
LFY	5	3	9
LIM	12	18	36
LUG	6	13	25
MADS	61	93	252
MBF1	5	3	4
Med6	1	1	1
Med7	2	1	5
mTERF	44	38	71
MYB	180	184	377
MYB-related	117	103	267
NAC	152	106	250
NZZ	5	6	17
OFP	15	22	46

Table 1, continued.

Family	Cowpea	Common Bean	Soybean
PcG_FIE	192	164	408
PcG_VEFS	4	5	12
PHD	123	131	287
PLATZ	18	15	48
Pseudo_ARR-B	2	9	27
RB	3	2	8
Rcd1-like	8	2	10
Rel	0	0	1
RRN3	13	6	12
Runt	2	0	0
RWP-RK	18	15	45
S1Fa-like	5	6	13
SAP	8	9	7
SBP	26	29	74
SET	43	57	132
Sigma70-like	5	13	18
Sin3	9	10	31
Sir2	5	10	7
SOH1	1	2	3
SRS	13	14	35
SWI/SNF_BAF60b	22	27	46
SWI/SNF_SNF2	74	54	112
SWI/SNF_SWI3	8	8	10
TAZ	8	8	14
TCP	17	31	71
TEA	1	0	0
TFb2	4	1	8
tify	24	22	67
TRAF	82	72	162
Trihelix	37	46	91
TUB	27	13	45
ULT	3	1	14
VARL	1	1	6
VOZ	6	5	21
Whirly	7	4	13
WRKY	120	104	243
zf_HD	16	21	55
Total	3573	3404	7980

Table 1, continued.

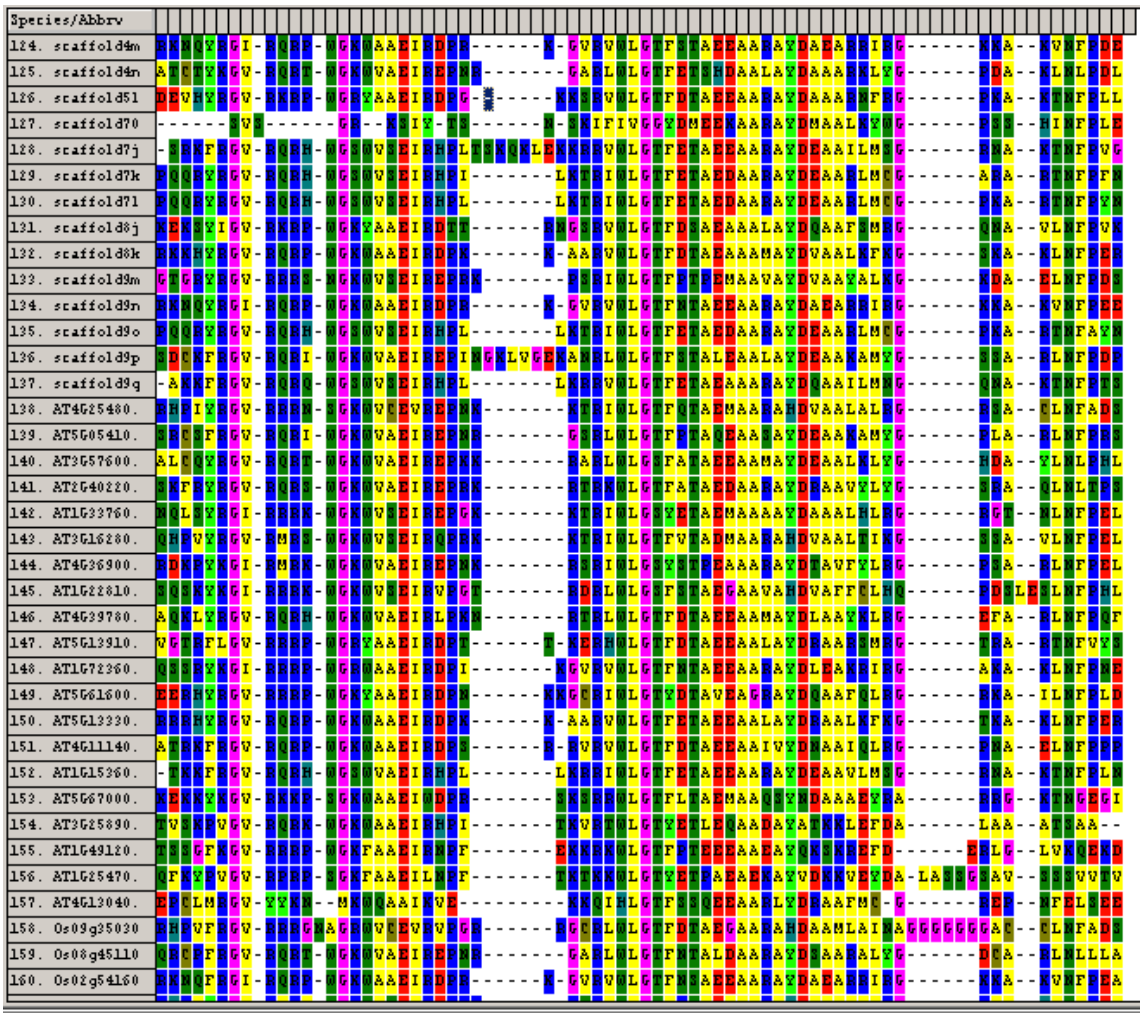


Figure 2. Multiple sequence alignment of the ERF TF conserved domains in cowpea, along with selected ERF domains from *Arabidopsis thaliana* and rice. The multiple sequence alignment was constructed using Expresso and PSI-Coffee as combined using the T-Coffee Combine algorithm (Di Tommaso et al., 2011). The sequences labeled as "C3…" or "scaffold…" are cowpea sequences.

a)

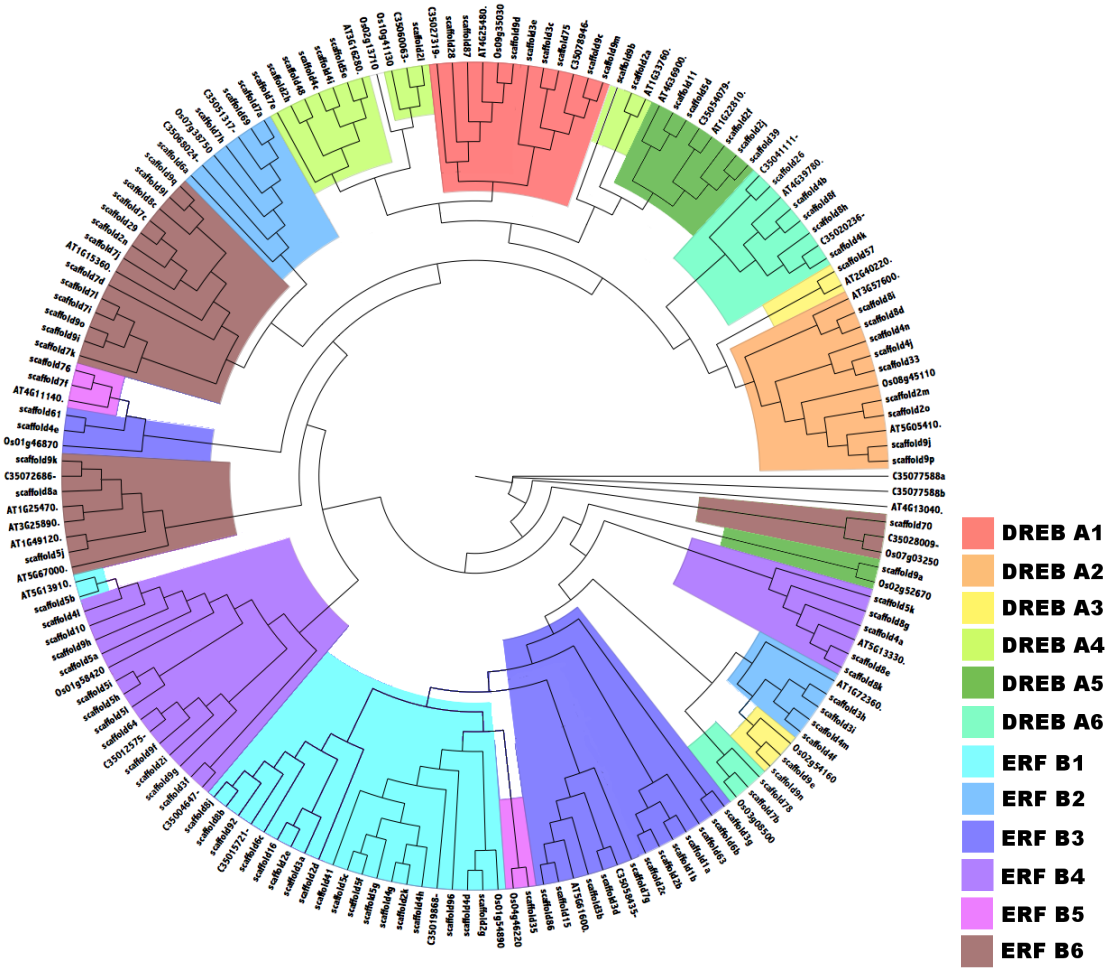


Figure 3. Cowpea ERF genes according to the classification schemes of (a) Dietz et al. (2010) and Sharoni et al. (2011), and (b) Nakano et al. (2006). Shown is a phylogenetic tree of the ERF conserved domains constructed using the maximum likelihood method via PhyML (Guindon et al., 2010) with the LG amino acid substitution model (Le & Gascuel, 2008), automatically estimated invariable sites, and 5 gamma parameters. The tree was calculated based on conserved ERF domains, as well as representative ERF domains from Arabidopsis and rice, whose selection is based on the ERF subgroup identification schemes of Dietz et al. (2010) and Sharoni et al. (2011), respectively. Since the sequence names are in PHYLIP

format, the full names can be found in Supplementary Table S1. [Note: Os07g03250 in the DREB A-5 group was not listed in any group studied in Nakano et al (2006), and Os02g13710 was listed as group III in Nakano et al. (2006) and as ERB B-7 in Sharoni et al. (2011).]

b)

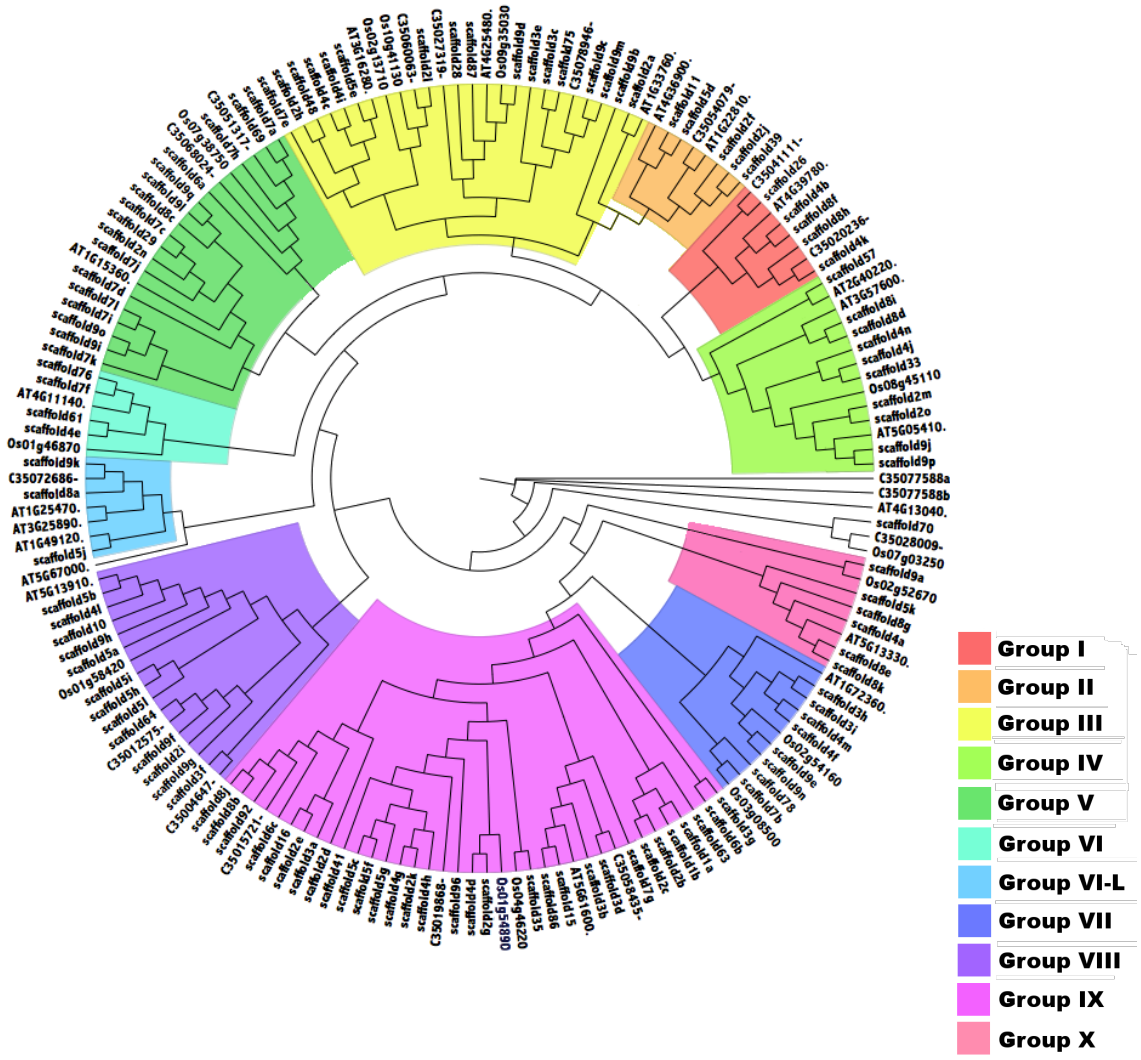


Figure 3, continued.

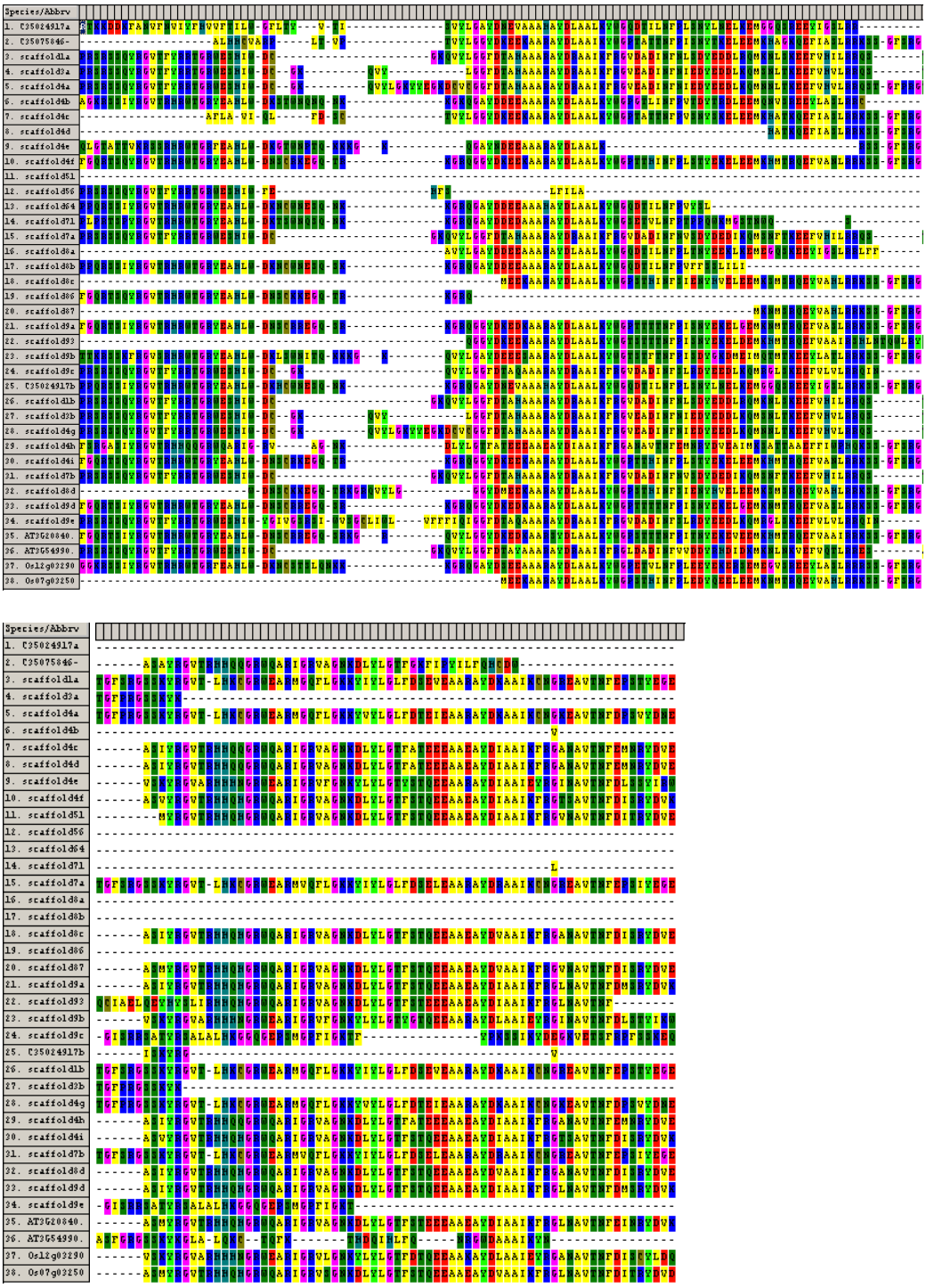


Figure 4. Multiple sequence alignment of the AP2 TF conserved domains in cowpea, along with selected AP2 domains from *Arabidopsis thaliana* (Dietz et al., 2010) and

rice (Sharoni et al., 2011). The sequences labeled as "C3…" or "scaffold…" are cowpea sequences.

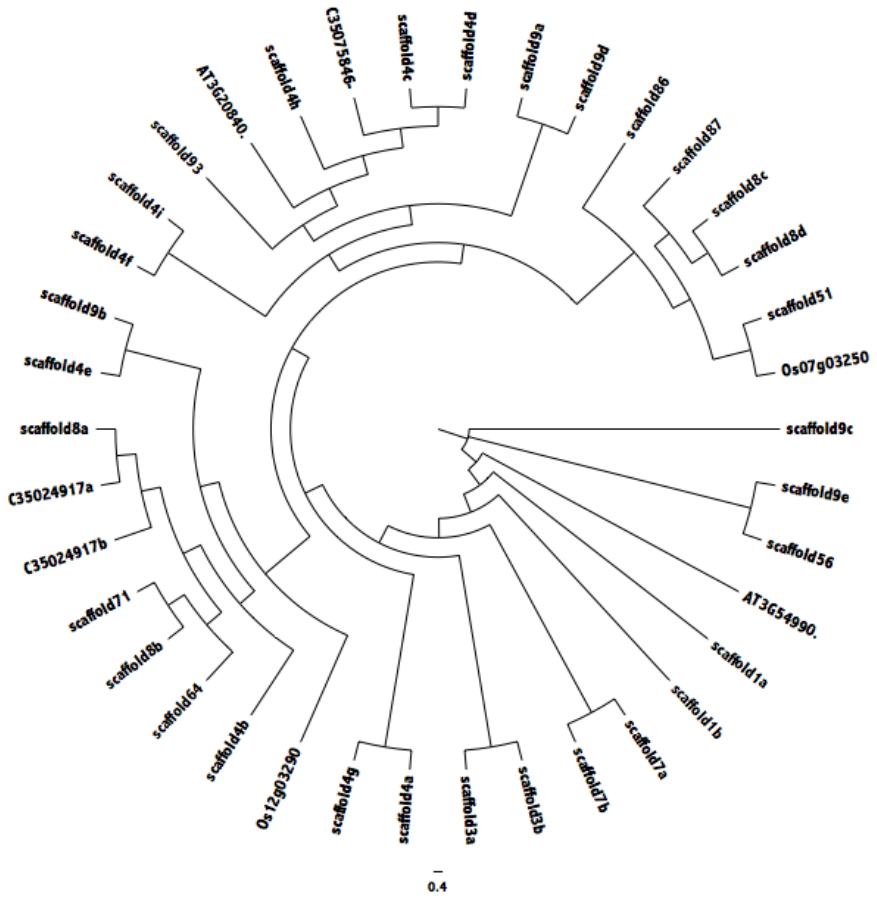


Figure 5. Cowpea AP2 genes, along with selected AP2 domains from *Arabidopsis thaliana* (Dietz et al., 2010) and rice (Sharoni et al., 2011). Shown is a phylogenetic tree of the AP2 conserved domains constructed using the maximum likelihood method via PhyML 3.0 (Guindon et al., 2010) with the VT+G amino acid substitution model (Muller & Vingron, 2000), with an automatically estimated proportion of invariable sites, and 5 gamma parameters. The tree was calculated based on conserved AP2 domains, as well as representative ERF domains from Arabidopsis

and rice, whose selection is based on the ERF subgroup identification schemes of Dietz et al. (2010) and Sharoni et al. (2011), respectively. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S2.

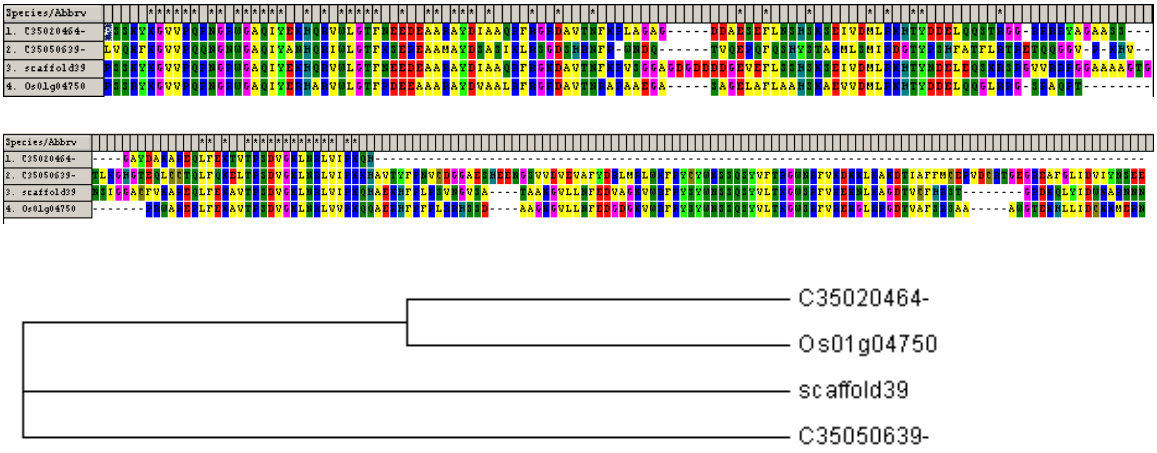


Figure 6. (A) Multiple sequence alignment of the RAV TF conserved domains in cowpea, along with a selected RAV domain from rice (Sharoni et al., 2011). (B) Phylogenetic tree of the RAV cowpea genes, as visualized in MEGA 5.2.2 (Tamura et al., 2011). The tree was generated using PhyML 3.0 (Guindon et al., 2010) with the JTT+G amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, and 5 gamma parameters. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S3.

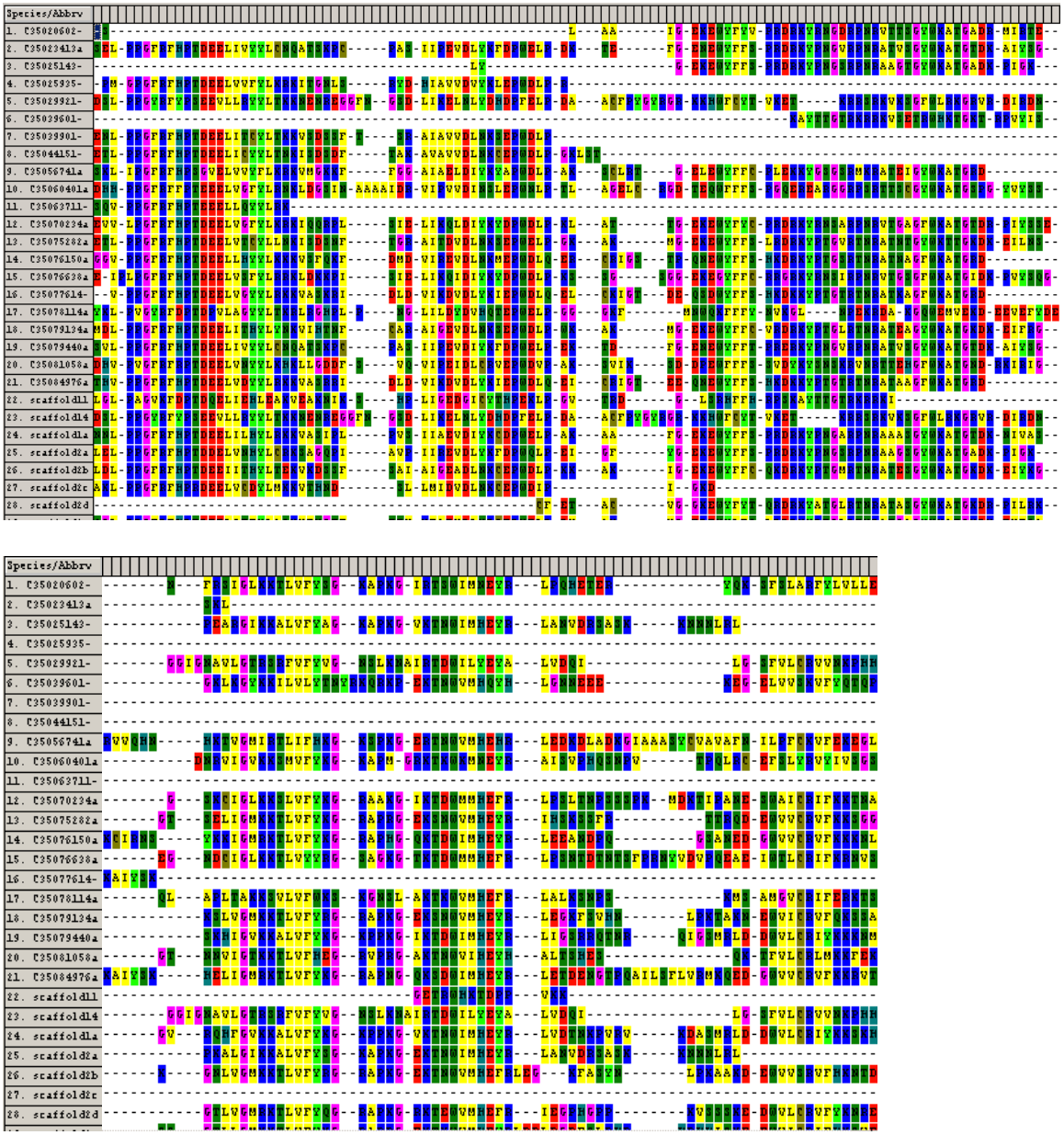


Figure 7. Multiple sequence alignment of the NAC TF conserved domains in cowpea, along with *Arabidopsis thaliana* and rice. The multiple sequence alignment was constructed using Expresso and PSI-Coffee as combined using the T-Coffee Combine algorithm (di Tommaso et al., 2011). The NAC domains used here were the domains used in Li et al. (2012). Manual curation was done in MEGA 5.2.2 (Tamura et al.,

2011). Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S4. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S3.

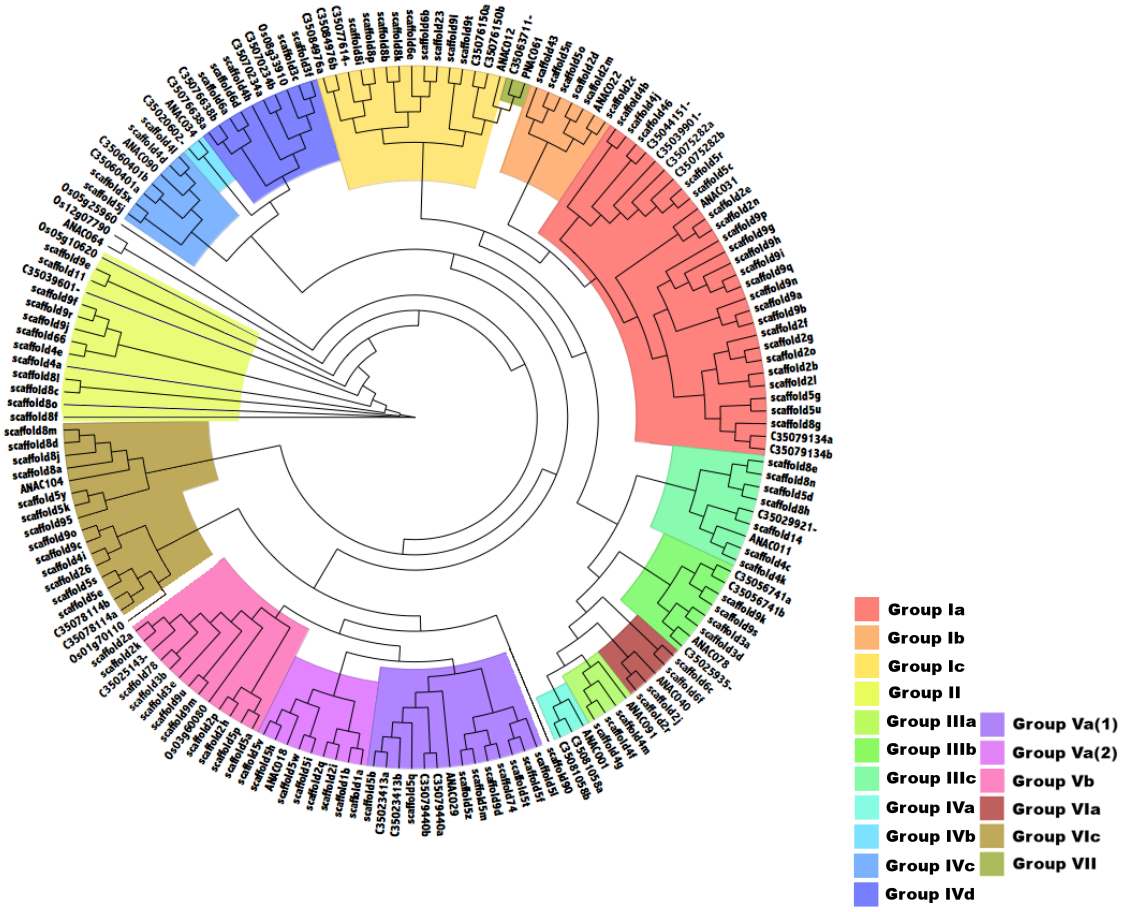


Figure 8. Cowpea NAC genes. Shown is a phylogenetic tree of the NAC conserved domains constructed using the maximum likelihood method with the JTT+G amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, and 5 gamma parameters. The tree was calculated based on NAC domain sequences of cowpea, as well as the representative

NAC domains used in the methodology of Zhu et al. (2012). Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S4.

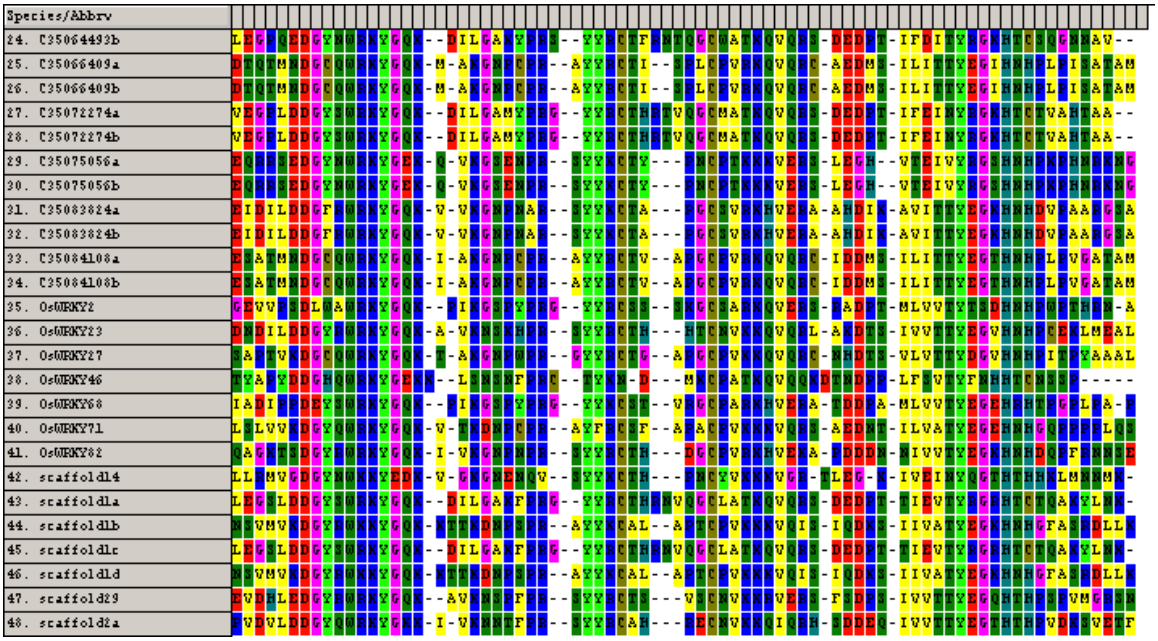


Figure 9. Multiple sequence alignment of the WRKY TF conserved domains in cowpea, along with *Arabidopsis thaliana* and rice. The multiple sequence alignment was constructed using Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena & Notredame, 2009; Chang et al., 2012) as combined using the T-Coffee Combine algorithm (di Tommaso et al., 2011). The Arabidopsis and rice WRKY domains used here were the domains used in Li et al. (2012). Manual curation was done in MEGA 5.2.2 (Tamura et al., 2011). Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S5. The sequeces labeled "C3…" or "scaffold…" are cowpea sequences.

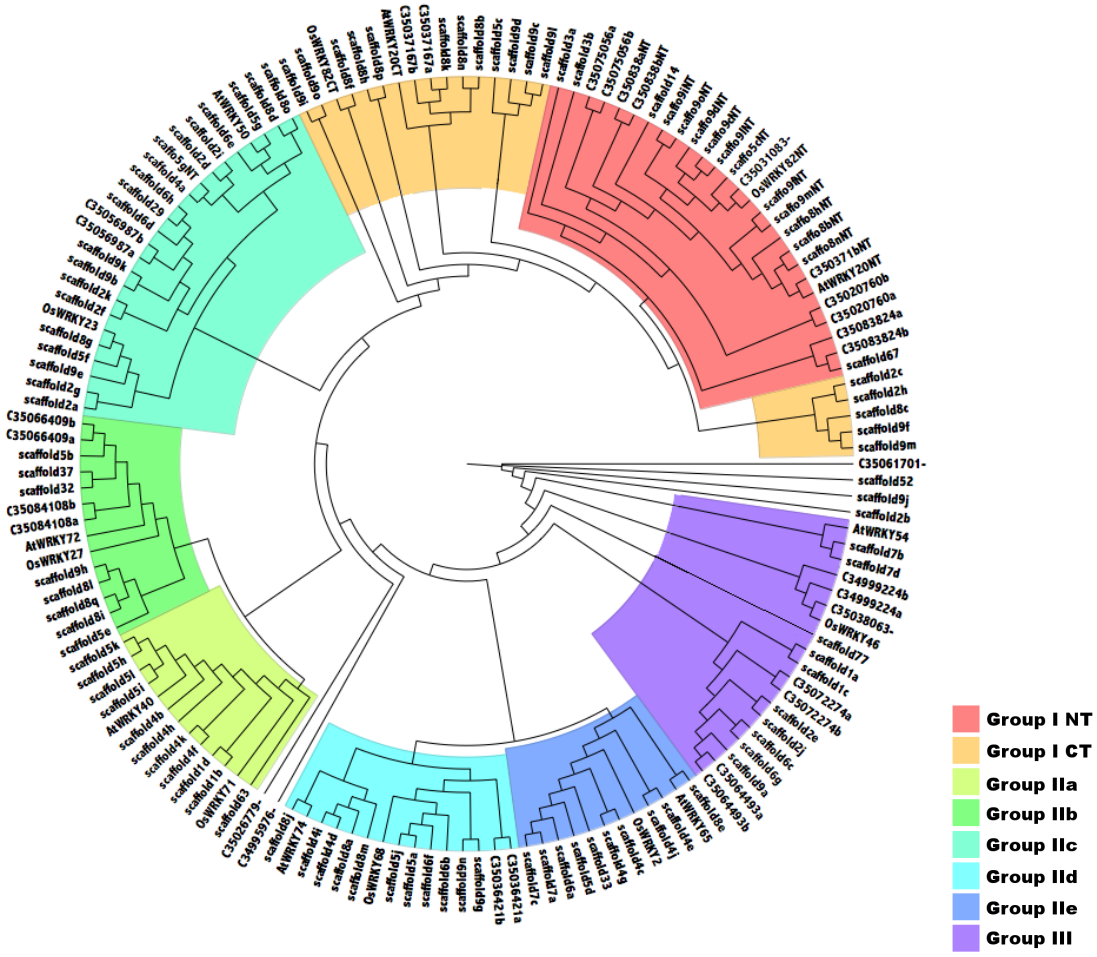


Figure 10. Cowpea WRKY genes. Shown is a phylogenetic tree of the WRKY conserved domains constructed using the maximum likelihood method with the JTT+G amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, and 5 gamma parameters. The tree was calculated based on WRKY domain sequences of cowpea, as well as selected representative Arabidopsis and rice WRKY domains; the selection of Arabidopsis and rice WRKY sequences was based on the methodology of Li et al. (2012). Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S5.

Supplementary Table S6. The grouping of *Arabidopsis thaliana* and rice (*Oryza sativa*) ERF sequences according to Nakano et al. (2006), Dietz et al. (2010) and Sharoni et al. (2011).

Sequence Used in Figure 2	Nakano et al. (2006)	Dietz et al. (2010)	Sharoni et al. (2011)
AT4G39780.	Group I (A-6)	DREB A6	N/A
AT4G36900.	Group II (A-5)	DREB A5	N/A
AT1G22810.	Group II (A-5)	DREB A5	N/A
AT4G25480.	Group III (A-1, -4, -5)	DREB A1	N/A
AT1G33760.	Group III (A-1, -4, -5)	DREB A4	N/A
AT3G16280.	Group III (A-1, -4, -5)	DREB A4	N/A
AT5G05410.	Group IV (A-2, -3)	DREB A2	N/A
AT3G57600.	Group IV (A-2, -3)	DREB A2	N/A
AT2G40220.	Group IV (A-2, -3)	DREB A3	N/A
AT1G15360.	Group V (B-6)	ERF B6	N/A
AT4G11140.	Group VI (B-5)	ERF B5	N/A
AT1G49120.	Group VI-L (B-6)	ERF B6	N/A
AT1G25470.	Group VI-L (B-6)	ERF B6	N/A
AT3G25890.	Group VI-L (B-6)	ERF B7	N/A
AT1G72360.	Group VII (B-2)	ERF B2	N/A
AT5G13910.	Group VIII (B-1)	ERF B1	N/A
AT5G61600.	Group IX (B-3)	ERF B3	N/A

AT5G13330.	Group X (B-3, -4)	ERF B4	N/A
AT5G67000.	Group Xb-L (B-6)	ERF B6	N/A
AT4G13040.	Outsider	Outsider	N/A
Os09g35030.	Group III (A-1, -4, -5)	N/A	DREB A1
Os10g41130.	Group III (A-1, -4, -5)	N/A	DREB A4
Os02g13710.	Group III (A-1, -4, -5)	N/A	ERF B7
Os08g45110.	Group IV (A-2, -3)	N/A	DREB A2
Os07g38750.	Group V (B-6)	N/A	ERF B2
Os01g46870.	Group VI (B-5)	N/A	ERF B3
Os02g54160.	Group VII (B-2)	N/A	DREB A3
Os03g08500.	Group VII (B-2)	N/A	DREB A6
Os01g58420.	Group VIII (B-1)	N/A	ERF B4
Os04g46220.	Group IX (B-3)	N/A	ERF B5
Os01g54890.	Group IX (B-3)	N/A	ERF B1
Os02g52670.	Group X (B-3, -4)	N/A	DREB A5
Os07g03250.1	Outsider	N/A	ERF B6

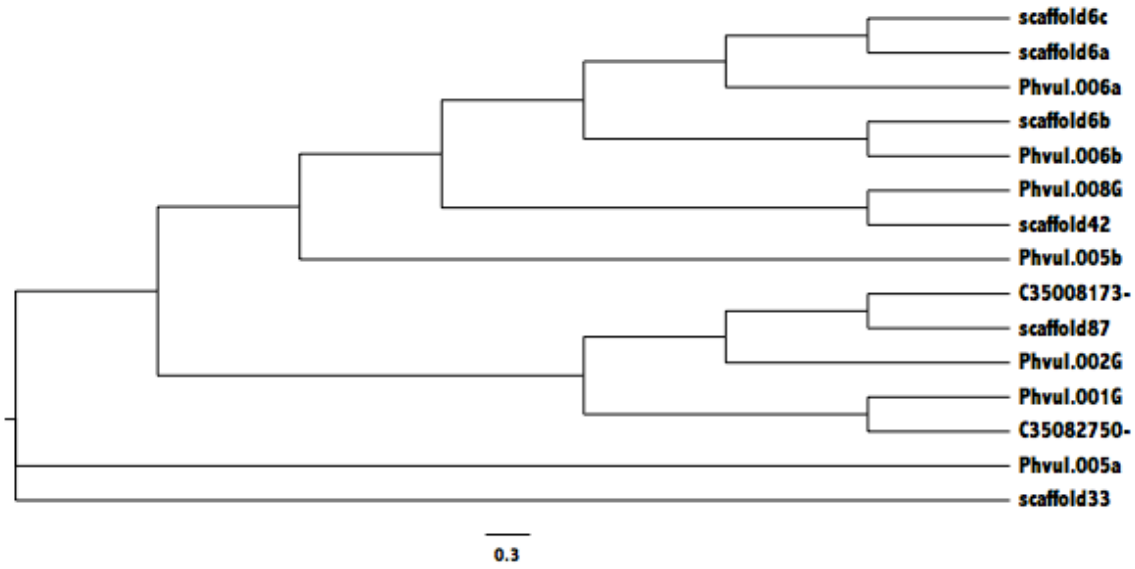


Figure 11. Phylogenetic tree of the EIL genes in cowpea and common bean. The tree was generated using PhyML (Guindon et al., 2010) with the JTT+G+F amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, estimated empirical frequencies, and 5 gamma parameters. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences, whereas sequences labeled as "Phvul…" are common bean sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S7.

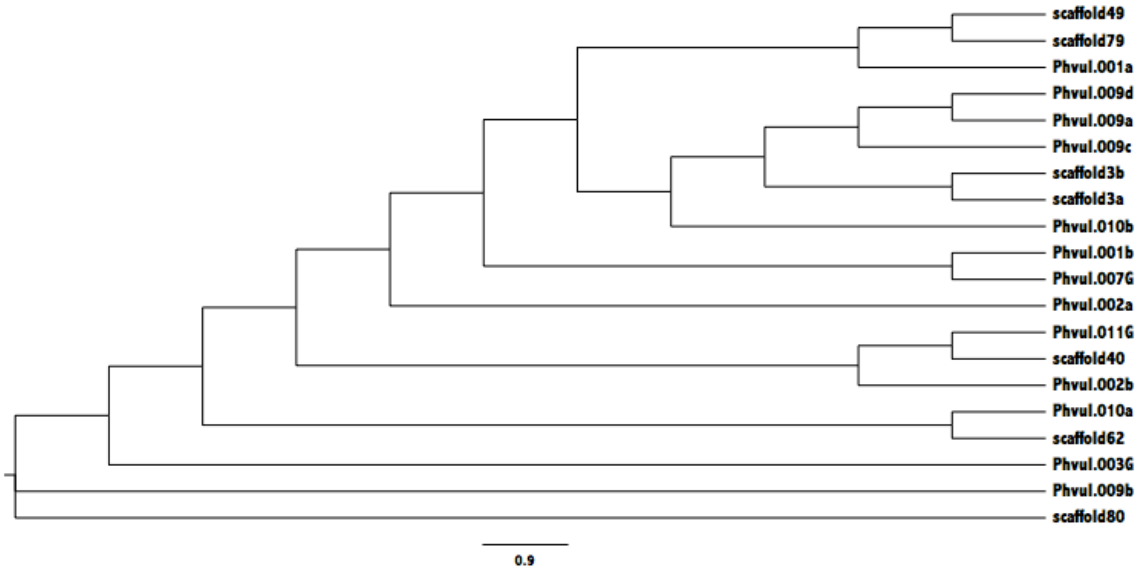


Figure 12. Phylogenetic tree of the GRF genes in cowpea and common bean. The tree was generated using PhyML (Guindon et al., 2010) with the JTT+I+G+F amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, estimated empirical frequencies, and 5 gamma parameters. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences, whereas sequences labeled as "Phvul…" are common bean sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S8.

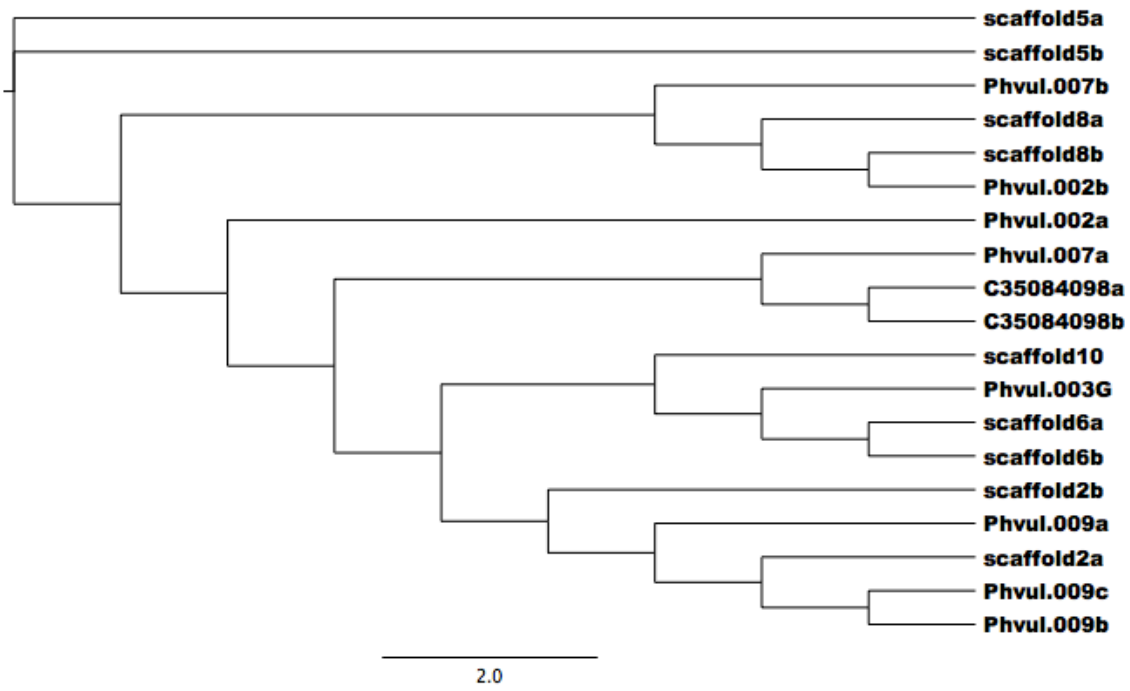


Figure 13. Phylogenetic tree of the HMG genes in cowpea and common bean. The tree was generated using PhyML (Guindon et al., 2010) with the RtREV+G+F amino acid substitution model (Dimmic et al., 2002), with an automatically estimated proportion of invariable sites, estimated empirical frequencies, and 5 gamma parameters. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences, whereas sequences labeled as "Phvul…" are common bean sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S9.

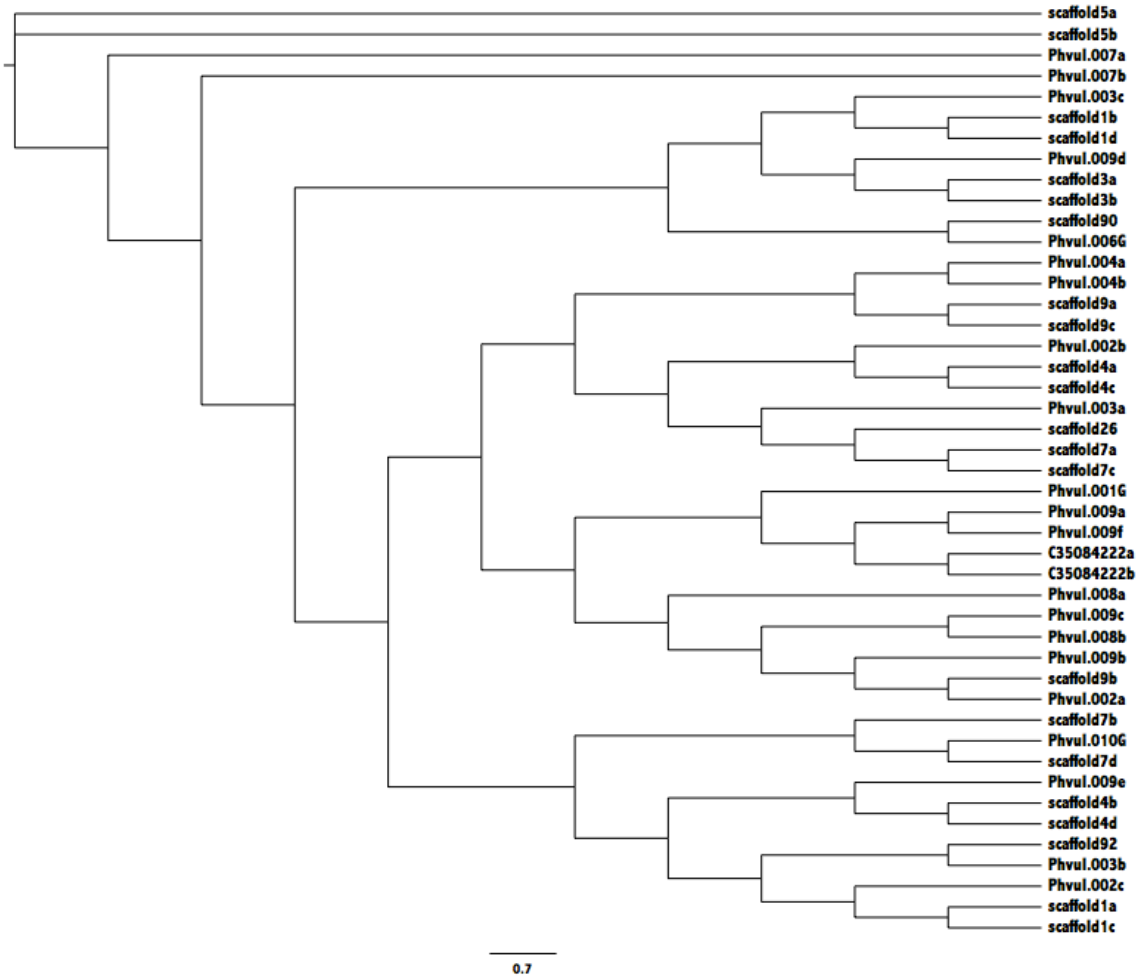


Figure 14. Phylogenetic tree of the tify genes in cowpea and common bean. The tree was generated using PhyML 3.0 (Guindon et al., 2010) with the JTT+G+F amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, estimated empirical frequencies, and 5 gamma parameters. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences, whereas sequences labeled as "Phvul…" are common bean sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S10.

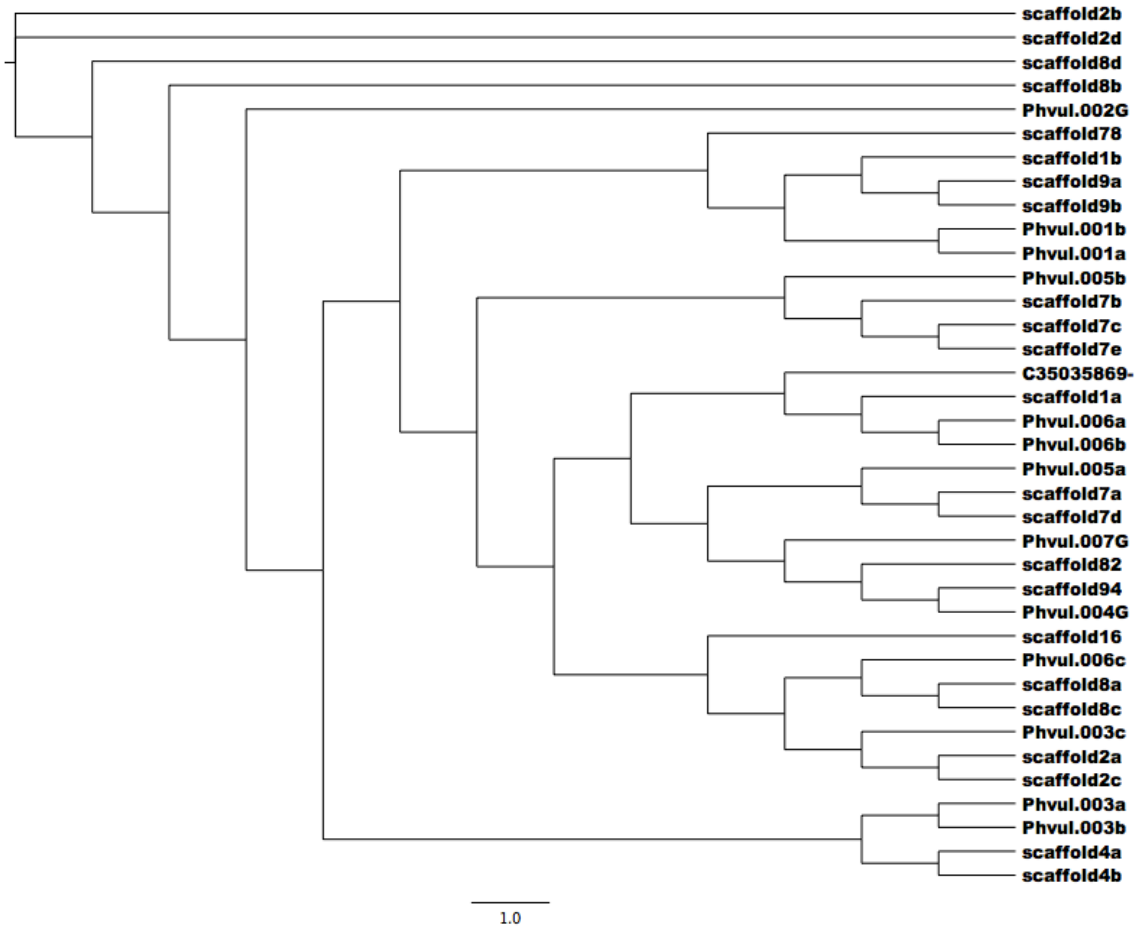


Figure 15. Phylogenetic tree of the TUB genes in cowpea and common bean. The tree was generated using PhyML (Guindon et al., 2010) with the JTT+G+F amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, estimated empirical frequencies, and 5 gamma parameters. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences, whereas sequences labeled as "Phvul…" are common bean sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S11.

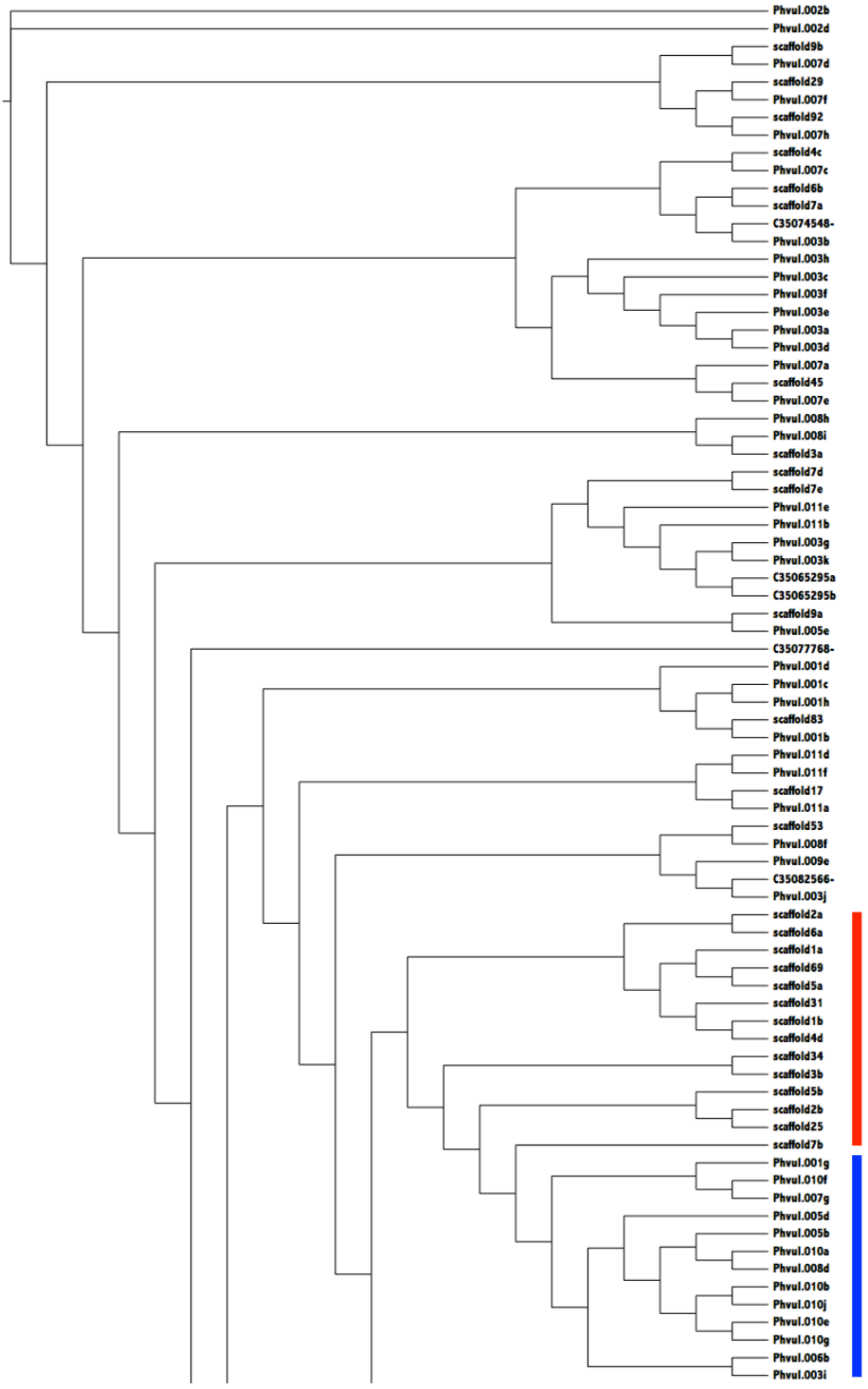


Figure 16. Phylogenetic tree of the ABI3-VP1 genes in cowpea and common bean. The group unique to cowpea is marked with a red bar, and the group unique to common bean is marked with a blue bar. The tree was generated using PhyML 3.0

(Guindon et al., 2010) with the JTT+G+F amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, estimated empirical frequencies, and 5 gamma parameters. The sequences labeled as "C3…" or "scaffold…" are cowpea sequences, whereas sequences labeled as "Phvul…" are common bean sequences. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S12.

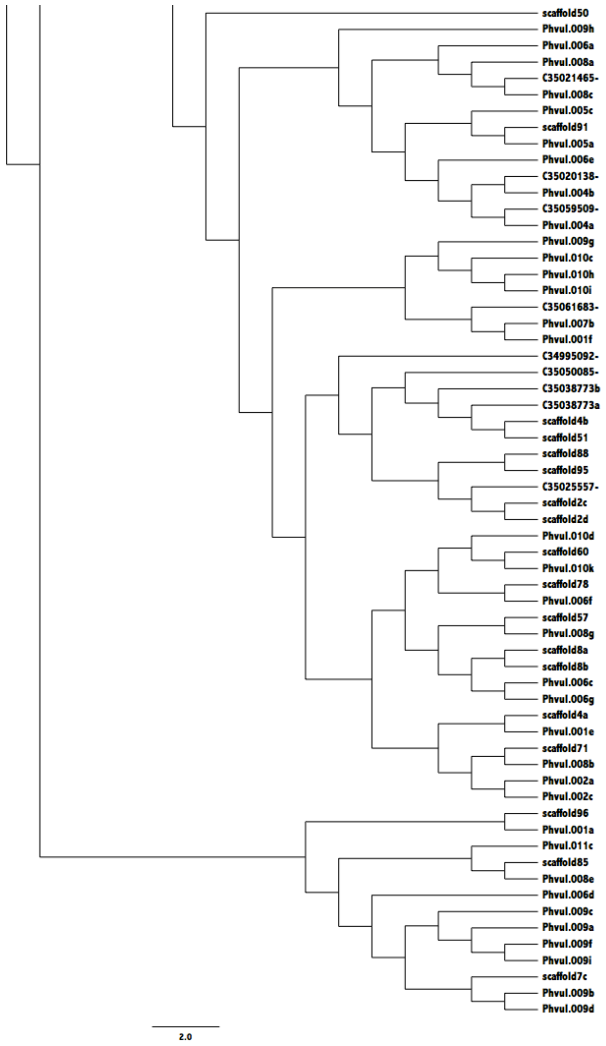


Figure 16, continued.

Supplementary Tables S1-S5, S7-S12 [Excel file]. Full names of the sequences in the phylogenetic trees whose sequences are represented in Figures 1-16.

References

Alonge, S. O., Lagoke, S. T. O., & Ajakaiye, C. O. (2005). Cowpea reactions to Striga gesnerioides I. effect on growth. Crop Protection, 24(6), 565-573. doi: http://dx.doi.org/10.1016/j.cropro.2004.08.014

Alonge, S. O., Lagoke, S. T. O., & Ajakaiye, C. O. (2005). Cowpea reactions to Striga gesnerioides II. Effect on grain yield and nutrient composition. Crop Protection, 24(6), 575-580. doi: http://dx.doi.org/10.1016/j.cropro.2004.08.015

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. [Comparative Study, Research Support, Non-U.S. Gov't]. Syst Biol, 55(4), 539-552. doi: 10.1080/10635150600755453

Andreasson, E., Jenkins, T., Brodersen, P., Thorgrimsen, S., Petersen, N. H., Zhu, S., . . . Mundy, J. (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. *EMBO J, 24*(14), 2579-2589. doi: 10.1038/sj.emboj.7600737

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., . . . Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res, 34*(suppl 2), W604-W608.

Arumuganathan, K., & Earle, E. (1991). Nuclear DNA content of some important plant species. *Plant molecular biology reporter, 9*(3), 208-218.

Atkinson, N. J., & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. [Review]. *J Exp Bot, 63*(10), 3523-3543. doi: 10.1093/jxb/ers100

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., & Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res, 28*(1), 263-266.

Bassel, G. W., Mullen, R. T., & Bewley, J. D. (2006). ABI3 expression ceases following, but not during, germination of tomato and Arabidopsis seeds. J Exp Bot, 57(6), 1291-1297. doi: 10.1093/jxb/erj101

Berri, S., Abbruscato, P., Faivre-Rampant, O., Brasileiro, A. C., Fumasoni, I., Satoh, K., . . . Piffanelli, P. (2009). Characterization of WRKY co-regulatory networks in rice and Arabidopsis. *BMC Plant Biol, 9*, 120. doi: 10.1186/1471-2229-9-120

Berrocal-Lobo, M., Molina, A., & Solano, R. (2002). Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in Arabidopsis confers resistance to several necrotrophic fungi. *Plant J, 29*(1), 23-32.

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., . . . van Lookeren Campagne, M. M. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell, 14*(8), 1737-1749.

Buttner, M., & Singh, K. B. (1997). Arabidopsis thaliana ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. *Proc Natl Acad Sci U S A, 94*(11), 5961-5966.

Byzova, M. V., Franken, J., Aarts, M. G., de Almeida-Engler, J., Engler, G., Mariani, C., . . . Angenent, G. C. (1999). Arabidopsis STERILE APETALA, a multifunctional gene regulating inflorescence, flower, and ovule development. *Genes Dev, 13*(8), 1002-1014.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., . . . Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res, 18*(1), 188-196. doi: 10.1101/gr.6743907

Cardwell, K. F., & Lane, J. A. (1995). Effect of soils, cropping system and host phenotype on incidence and severity of Striga gesnerioides on cowpea in West Africa. *Agriculture, Ecosystems & Environment, 53*(3), 253-262. doi: http://dx.doi.org/10.1016/0167-8809(94)00575-Y

Cernac, A., & Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *Plant J, 40*(4), 575-585. doi: 10.1111/j.1365-313X.2004.02235.x

Chang, J. M., Di Tommaso, P., Taly, J. F., & Notredame, C. (2012). Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinformatics, 13 Suppl 4*, S1. doi: 10.1186/1471-2105-13-s4-s1

Chen, X., Laudeman, T. W., Rushton, P. J., Spraggins, T. A., & Timko, M. P. (2007). CGKB: an annotation knowledge base for cowpea (Vigna unguiculata L.) methylation filtered genomic genespace sequences. *BMC Bioinformatics, 8*, 129. doi: 10.1186/1471-2105-8-129

Chi, Y., Yang, Y., Zhou, Y., Zhou, J., Fan, B., Yu, J.-Q., & Chen, Z. (2013). Protein–Protein Interactions in the Regulation of WRKY Transcription Factors. *Mol Plant, 6*(2), 287-300. doi: 10.1093/mp/sst026

Chuck, G., Meeley, R. B., & Hake, S. (1998). The control of maize spikelet meristem fate by the APETALA2-like gene indeterminate spikelet1. *Genes Dev, 12*(8), 1145-1154.

Close T. J., Luo M., Lonardi S., Alpert M., Duma D., Bozdag S., Farmer A., May G. D., Jackson S. A., Ma, Y., Gu, Y.Q., Wu, J.,You, F.M., Diop, N.N., Wanamaker S., Fenton R. D., Timko M. P., Roberts P. A. and Ehlers J. D. 2011. Genome Resources For Cowpea Genotype IT97K-499-35. Plant and Animal Genomes XIX Conference, Town and Country Convention Center, San Diego, CA.

Close, T., Wanamaker, S., Roose, M. L., & Lyon, M. (2007). HarvEST. In D. Edwards (Ed.), *Plant Bioinformatics* (Vol. 406, pp. 161-177): Humana Press.

Dahmardeh, M., Ghanbari, A., Syasar, B., & Ramrodi, M. (2009). Intercropping maize (Zea mays L.) and cow pea (Vigna unguiculata L.) as a whole-crop forage: Effects of planting ratio and harvest time on forage yield and quality. *Journal of Food, Agriculture and Environment, 7*(2), 505-509.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics, 27*(8), 1164-1165.

Das, S., Ehlers, J., Close, T., & Roberts, P. (2010). Transcriptional profiling of root-knot nematode induced feeding sites in cowpea (Vigna unguiculata L. Walp.) using a soybean genome array. *BMC Genomics, 11*(1), 480.

de Pater, S., Greco, V., Pham, K., Memelink, J., & Kijne, J. (1996). Characterization of a zinc-dependent transcriptional activator from Arabidopsis. *Nucleic Acids Res, 24*(23), 4624-4631.

Dimmic, M. W., Rest, J. S., Mindell, D. P., & Goldstein, R. A. (2002). rtREV: an amino acid substitution matrix for interference of retrovirus and reverse transcriptase phylogeny. *J Mol Evol,* 55(1), 65-73. doi: 10.1007/s00239-001-2304-y

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., . . . Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res, 39*(suppl 2), W13-W17.

Dietz, K. J., Vogel, M. O., & Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma, 245*(1-4), 3-14. doi: 10.1007/s00709-010-0142-8

Doidy, J., Grace, E., Kuhn, C., Simon-Plas, F., Casieri, L., & Wipf, D. (2012). Sugar transporters in plants and in their interactions with fungi. [Review]. *Trends Plant Sci, 17*(7), 413-422. doi: 10.1016/j.tplants.2012.03.009

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics, 14*(9), 755-763.

Eddy, S. R. (2009). *A new generation of homology search tools based on probabilistic inference.* Paper presented at the Genome Inform.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput Biol, 7*(10), e1002195.

Ehlers, J., & Hall, A. (1997). Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Research, 53*(1), 187-204.

Elliott, R. C., Betzner, A. S., Huttner, E., Oakes, M. P., Tucker, W. Q., Gerentes, D., . . . Smyth, D. R. (1996). AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. *Plant Cell, 8*(2), 155-168.

Elowad, H. O., & Hall, A. E. (1987). Influences of early and late nitrogen fertilization on yield and nitrogen fixation of cowpea under well-watered and dry field conditions. *Field Crops Research, 15*(3), 229-244.

Eulgem, T., Rushton, P. J., Robatzek, S., & Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. [Review]. *Trends Plant Sci, 5*(5), 199-206.

Eulgem, T., & Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review]. *Curr Opin Plant Biol, 10*(4), 366-371. doi: 10.1016/j.pbi.2007.04.020

The Food and Agriculture Organization of the United Nations. (2011). FAOSTAT. Retrieved from faostat3.fao.org on June 22, 2013.

Format Converter v2.1.5. (2013) Retrieved from http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html on June 22, 2013.

Galeano, C., Fernandez, A., Gomez, M., & Blair, M. (2009). Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and

synteny analysis of common bean (Phaseolus vulgaris L.). *BMC Genomics, 10*(1), 629.

Gao, G., Zhong, Y., Guo, A., Zhu, Q., Tang, W., Zheng, W., . . . Luo, J. (2006). DRTF: a database of rice transcription factors. *Bioinformatics, 22*(10), 1286-1287. doi: 10.1093/bioinformatics/btl107

Gepts, P., Aragão, F. J., de Barros, E., Blair, M. W., Brondani, R., Broughton, W., . . . Lariguet, P. (2008). Genomics of Phaseolus beans, a major source of dietary protein and micronutrients in the tropics *Genomics of tropical crop plants* (pp. 113-143): Springer.

Gill, N., Findley, S., Walling, J. G., Hans, C., Ma, J., Doyle, J., . . . Jackson, S. A. (2009). Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiol, 151*(3), 1167-1174. doi: 10.1104/pp.109.137935

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., . . . Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Nucleic Acids Res, 40*(Database issue), D1178-1186. doi: 10.1093/nar/gkr944

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res, 38*(Database issue), D843-846. doi: 10.1093/nar/gkp798

Gu, Y. Q., Wildermuth, M. C., Chakravarthy, S., Loh, Y. T., Yang, C., He, X., . . . Martin, G. B. (2002). Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. *Plant Cell, 14*(4), 817-831.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol, 59*(3), 307-321. doi: 10.1093/sysbio/syq010

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., & Luo, J. (2005). DATF: a database of Arabidopsis transcription factors. *Bioinformatics, 21*(10), 2568-2569. doi: 10.1093/bioinformatics/bti334

Guo, A. Y., Chen, X., Gao, G., Zhang, H., Zhu, Q. H., Liu, X. C., . . . Luo, J. (2008). PlantTFDB: a comprehensive plant transcription factor database. *Nucleic Acids Res, 36*(Database issue), D966-969. doi: 10.1093/nar/gkm841

Gutterson, N., & Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr Opin Plant Biol, 7*(4), 465.

Hall, A., Ismail, A., Ehlers, J., Marfo, K., Cisse, N., Thiaw, S., & Close, T. (2002). Breeding cowpea for tolerance to temperature extremes and adaptation to drought. *Challenges and Opportunities for Enhancing Sustainable Cowpea Production*, 14-21.

Hall, A. E., Cisse, N., Thiaw, S., Elawad, H. O., Ehlers, J. D., Ismail, A. M., . . . Murdock, L. L. (2003). Development of cowpea cultivars and germplasm by the Bean/Cowpea CRSP. *Field Crops Research, 82*(2), 103-134.

Hall, A. E. (2004). Breeding for adaptation to drought and heat in cowpea. *European Journal of Agronomy, 21*(4), 447-454.

HarvEST Blast Server (http://www.harvest-blast.org/). Accessed on June 22, 2013.

Hiraoka, Y., & Sugimoto, Y. (2008). Molecular responses of sorghum to purple witchweed (Striga hermonthica) parasitism. *Weed Science, 56*(3), 356-363. doi: Doi 10.1614/Ws-07-136.1

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q., & Xiong, L. (2006). Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc Natl Acad Sci U S A, 103*(35), 12987-12992. doi: 10.1073/pnas.0604882103

Huang, K., Mellor, K. E., Paul, S. N., Lawson, M. J., Mackey, A. J., & Timko, M. P. (2012). Global changes in gene expression during compatible and incompatible interactions of cowpea (Vigna unguiculata L.) with the root parasitic angiosperm Striga gesnerioides. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *BMC Genomics, 13*(1), 402. doi: 10.1186/1471-2164-13-402

Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P., & Oinn, T. (2006). Taverna: a tool for building and running workflows of services. [Research Support, Non-U.S. Gov't]. *Nucleic Acids Res, 34*(Web Server issue), W729-732. doi: 10.1093/nar/gkl320

Ishiguro, S., & Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato. *Mol Gen Genet, 244*(6), 563-571.

Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O'Shea, C., & Skriver, K. (2010). The Arabidopsis thaliana NAC transcription factor family:

structure–function relationships and determinants of ANAC019 stress signalling. *Biochemical Journal, 426*(2), 183-196. doi: 10.1042/bj20091234

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS, 8(3), 275-282.

Kagaya, Y., Ohmiya, K., & Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res, 27*(2), 470-478. doi: 10.1093/nar/27.2.470

Kalavacharla, V., Liu, Z., Meyers, B. C., Thimmapuram, J., & Melmaiee, K. (2011). Identification and analysis of common bean (Phaseolus vulgaris L.) transcriptomes by massively parallel pyrosequencing. *BMC Plant Biol, 11*, 135. doi: 10.1186/1471-2229-11-135

Kemena, C., & Notredame, C. (2009). Upcoming challenges for multiple sequence alignment methods in the high-throughput era. [Review]. *Bioinformatics, 25*(19), 2455-2465. doi: 10.1093/bioinformatics/btp452

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., . . . Demura, T. (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev, 19*(16), 1855-1860.

Lai, Z., Li, Y., Wang, F., Cheng, Y., Fan, B., Yu, J.-Q., & Chen, Z. (2011). Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *The Plant Cell Online, 23*(10), 3824-3841.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D. M., Correa, L. G., .

. . Rensing, S. A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and correlation with complexity. [Comparative Study Research Support, Non-U.S. Gov't]. *Genome Biol Evol, 2*, 488-503. doi: 10.1093/gbe/evq032

Langyintuo, A., Lowenberg-DeBoer, J., Faye, M., Lambert, D., Ibro, G., Moussa, B., . . . Ntoukam, G. (2003). Cowpea supply and demand in West and Central Africa. *Field Crops Research, 82*(2), 215-231.

Lavin, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst Biol, 54*(4), 575-594.

Lee, S.-j., Park, J., Lee, M., Yu, J.-h., & Kim, S. (2010). Isolation and functional characterization of CE1 binding proteins. *BMC Plant Biol, 10*(1), 277.

Li, C. W., Su, R. C., Cheng, C. P., Sanjaya, You, S. J., Hsieh, T. H., . . . Chan, M. T. (2011). Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. *Plant Physiol, 156*(1), 213-227. doi: 10.1104/pp.111.174268

Li, H.-L., Zhang, L.-B., Guo, D., Li, C.-Z., & Peng, S.-Q. (2012). Identification and expression profiles of the WRKY transcription factor family in Ricinus communis. *Gene, 503*(2), 248-253. doi: http://dx.doi.org/10.1016/j.gene.2012.04.069

Luerssen, H., Kirik, V., Herrmann, P., & Misera, S. (1998). FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional

importance for the regulation of seed maturation in Arabidopsis thaliana. Plant J, 15(6), 755-764.

MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends in Genetics, 27*(4), 141-148. doi: http://dx.doi.org/10.1016/j.tig.2011.01.001

McClean, P. E., Lavin, M., Gepts, P., & Jackson, S. A. (2008). Phaseolus vulgaris: a diploid model for soybean *Genetics and Genomics of Soybean* (pp. 55-76): Springer.

McClean, P., Mamidi, S., McConnell, M., Chikara, S., & Lee, R. (2010). Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics, 11*(1), 184.

Missier, P., Soiland-Reyes, S., Owen, S., Tan, W., Nenadic, A., Dunlop, I., ... & Goble, C. (2010, January). Taverna, reloaded. In *Scientific and Statistical Database Management* (pp. 471-481). Springer Berlin Heidelberg.

Mitsuda, N., Seki, M., Shinozaki, K., & Ohme-Takagi, M. (2005). The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell, 17*(11), 2993-3006. doi: 10.1105/tpc.105.036004

Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K., & Tran, L. S. P. (2010). LegumeTFDB: an integrative database of Glycine max, Lotus japonicus and Medicago truncatula transcription factors. [Article]. *Bioinformatics, 26*(2), 290-291. doi: 10.1093/bioinformatics/btp645

Moore, R. C., & Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. Curr Opin Plant Biol, 8(2), 122-128. doi: http://dx.doi.org/10.1016/j.pbi.2004.12.001

Muchero, W., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., . . . Close, T. J. (2009). A consensus genetic map of cowpea [Vigna unguiculata (L) Walp.] and synteny based on EST-derived SNPs. *Proc Natl Acad Sci U S A, 106*(43), 18159-18164. doi: 10.1073/pnas.0905886106

Muchero, W., Ehlers, J. D., Close, T. J., & Roberts, P. A. (2011). Genic SNP markers and legume synteny reveal candidate genes underlying QTL for Macrophomina phaseolina resistance and maturity in cowpea [Vigna unguiculata (L) Walp.]. BMC Genomics, 12, 8. doi: 10.1186/1471-2164-12-8

Muller, T., & Vingron, M. (2000). Modeling amino acid replacement. J Comput Biol, 7(6), 761-776. doi: 10.1089/10665270050514918

Nag, R., Maity, M. K., & Dasgupta, M. (2005). Dual DNA binding property of ABA insensitive 3 like factors targeted to promoters responsive to ABA and auxin. Plant Mol Biol, 59(5), 821-838. doi: 10.1007/s11103-005-1387-z

Nakamura, S., Lynch, T. J., & Finkelstein, R. R. (2001). Physical interactions between ABA response loci of Arabidopsis. Plant J, 26(6), 627-635.

Nakano, T., Suzuki, K., Fujimura, T., & Shinshi, H. (2006). Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice. *Plant Physiol, 140*(2), 411-432. doi: 10.1104/pp.105.073783

Nielsen, S. S., Brandt, W. E., & Singh, B. B. (1993). Genetic-Variability for Nutritional Composition and Cooking Time of Improved Cowpea Lines. *Crop Science, 33*(3), 469-472. doi: 10.2135/cropsci1993.0011183X003300030010x

Nielsen, S., Ohler, T., & Mitchell, C. (1997). Cowpea leaves for human consumption: production, utilization, and nutrient composition. *Advances in cowpea research, 326*.

Notredame, C. (2007). Recent evolutions of multiple sequence alignment algorithms. [Research Support, Non-U.S. Gov't Review]. *PLoS Comput Biol, 3*(8), e123. doi: 10.1371/journal.pcbi.0030123

Notredame, C. (2008) T-Coffee: Cheat Sheet Tutorial and FAQ. Accessed from http://www.tcoffee.org/Documentation/t_coffee/t_coffee_tutorial.htm on June 22, 2013.

Nuruzzaman, M., Sharoni, A. M., Satoh, K., Moumeni, A., Venuprasad, R., Serraj, R., . . . Kikuchi, S. (2012). Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64. *Mol Genet Genomics, 287*(5), 389-410. doi: 10.1007/s00438-012-0686-8

Ohme-Takagi, M., & Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *The Plant Cell Online, 7*(2), 173-182.

Oinn, T., Greenwood, M., Addis, M., Alpdemir, M. N., Ferris, J., Glover, K., ... & Wroe, C. (2006). Taverna: lessons in creating a workflow environment for the life

sciences. *Concurrency and Computation: Practice and Experience*, *18*(10), 1067-1100.

Olsen, A. N., Ernst, H. A., Leggio, L. L., & Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci, 10*(2), 79-87. doi: 10.1016/j.tplants.2004.12.010

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., . . . Kikuchi, S. (2003). Comprehensive Analysis of NAC Family Genes in Oryza sativa and Arabidopsis thaliana. *DNA research, 10*(6), 239-247. doi: 10.1093/dnares/10.6.239

Oseni, T. O. (2010). Evaluation of sorghum-cowpea intercrop productivity in savanna agro-ecology using competition indices. *Journal of Agricultural Science, 2*(3), P229.

Pandey, S. P., & Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol, 150*(4), 1648-1655. doi: 10.1104/pp.109.138990

Pearson, W. R., Wood, T., Zhang, Z., & Miller, W. (1997). Comparison of DNA sequences with protein sequences. *Genomics, 46*(1), 24-36.

Perez-Rodriguez, P., Riano-Pachon, D. M., Correa, L. G., Rensing, S. A., Kersten, B., & Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res, 38*(Database issue), D822-827. doi: 10.1093/nar/gkp805

Pfam (2013). Accessed from http://pfam.sanger.ac.uk/ on June 22, 2013.

Phytozome (2013). Accessed from http://www.phytozome.net/ on June 22, 2013.

Pinheiro, G. L., Marques, C. S., Costa, M. D. B. L., Reis, P. A. B., Alves, M. S., Carvalho, C. M., . . . Fontes, E. P. B. (2009). Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene, 444*(1–2), 10-23. doi: http://dx.doi.org/10.1016/j.gene.2009.05.012

Pottorff, M., Ehlers, J. D., Fatokun, C., Roberts, P. A., & Close, T. J. (2012). Leaf morphology in Cowpea [Vigna unguiculata (L.) Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. *BMC Genomics, 13*, 234. doi: 10.1186/1471-2164-13-234

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. [Research Support, Non-U.S. Gov't]. *Nucleic Acids Res, 40*(Database issue), D290-301. doi: 10.1093/nar/gkr1065

Puranik, S., Sahu, P. P., Srivastava, P. S., & Prasad, M. (2012). NAC proteins: regulation and role in stress tolerance. [Research Support, Non-U.S. Gov't Review]. *Trends Plant Sci, 17*(6), 369-381. doi: 10.1016/j.tplants.2012.02.004

Qiu, D., Xiao, J., Ding, X., Xiong, M., Cai, M., Cao, Y., . . . Wang, S. (2007). OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. *Mol Plant Microbe Interact, 20*(5), 492-499. doi: 10.1094/mpmi-20-5-0492

Qiu, D., Xiao, J., Xie, W., Cheng, H., Li, X., & Wang, S. (2009). Exploring transcriptional signalling mediated by OsWRKY13, a potential regulator of multiple physiological processes in rice. *BMC Plant Biol, 9*, 74. doi: 10.1186/1471-

2229-9-74

Rambaut, A. FigTree. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/ on June 22, 2013.

Reiss, G., & Bailey, J. (1998). Striga gesnerioidesParasitising Cowpea: Development of Infection Structures and Mechanisms of Penetration. *Ann Bot, 81*(3), 431-440.

Riano-Pachon, D. M., Ruzicic, S., Dreyer, I., & Mueller-Roeber, B. (2007). PlnTFDB: an integrative plant transcription factor database. *BMC Bioinformatics, 8*, 42. doi: 10.1186/1471-2105-8-42

Richardt, S., Lang, D., Reski, R., Frank, W., & Rensing, S. A. (2007). PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. [Research Support, Non-U.S. Gov't]. *Plant Physiol, 143*(4), 1452-1466. doi: 10.1104/pp.107.095760

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., . . . Yu, G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science, 290*(5499), 2105-2110.

Riechmann, J. L., & Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. Review]. *Biol Chem, 379*(6), 633-646.

Roberts, P., Matthews, W., & Ehlers, J. (1996). New resistance to virulent root-knot nematodes linked to the Rk locus of cowpea. *Crop Science, 36*(4), 889-894.

Roberts, P., Ehlers, J., Hall, A., & Matthews, W. (1997). Characterization of new resistance to root-knot nematodes in cowpea. *Advances in cowpea research. Edited by BB Singh, DR Mohan Raj, KE Dashiell, and LEN Jackai. IITA Ibadan, Nigeria*, 207-214.

Rolland, F., Baena-Gonzalez, E., & Sheen, J. (2006). Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol, 57*, 675-709. doi: 10.1146/annurev.arplant.57.032905.105441

Romanel, E. A. C., Schrago, C. G., Couñago, R. M., Russo, C. A. M., & Alves-Ferreira, M. (2009). Evolution of the B3 DNA Binding Superfamily: New Insights into REM Family Gene Diversification. PLoS ONE, 4(6), e5791. doi: 10.1371/journal.pone.0005791

Rushton, P. J., Macdonald, H., Huttly, A. K., Lazarus, C. M., & Hooley, R. (1995). Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of alpha-Amy2 genes. [Research Support, Non-U.S. Gov't]. *Plant Mol Biol, 29*(4), 691-702.

Rushton, P. J., Bokowiec, M. T., Han, S., Zhang, H., Brannock, J. F., Chen, X., . . . Timko, M. P. (2008). Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. [Research Support, Non-U.S. Gov't]. *Plant Physiol, 147*(1), 280-295. doi: 10.1104/pp.107.114041

Rushton, P. J., Somssich, I. E., Ringler, P., & Shen, Q. J. (2010). WRKY transcription factors. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review]. *Trends Plant Sci, 15*(5), 247-258. doi: 10.1016/j.tplants.2010.02.006

Rushton, P. J. (2008). TOBFAC: Family AP2. Retrieved from http://compsysbio.achs.virginia.edu/tobfac/browse_family.pl?family=AP2 on June 22, 2013.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol, 4*(4), 406-425.

Salamov, A. A., & Solovyev, V. V. (2000). Ab initio gene finding in Drosophila genomic DNA. *Genome Research, 10*(4), 516-522.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., . . . Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature, 463*(7278), 178-183. doi: 10.1038/nature08670

Sharoni, A. M., Nuruzzaman, M., Satoh, K., Shimizu, T., Kondoh, H., Sasaya, T., . . . Kikuchi, S. (2011). Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. *Plant Cell Physiol, 52*(2), 344-360. doi: 10.1093/pcp/pcq196

Shin, R., Park, J. M., An, J.-M., & Paek, K.-H. (2002). Ectopic expression of Tsi1 in transgenic hot pepper plants enhances host resistance to viral, bacterial, and oomycete pathogens. *Molecular Plant-Microbe Interactions, 15*(10), 983-989.

Shiota, H., Satoh, R., Watabe, K., Harada, H., & Kamada, H. (1998). C-ABI3, the carrot homologue of the Arabidopsis ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of embryo-specific ABA-inducible genes. Plant Cell Physiol, 39(11), 1184-1193.

Shore, P., & Sharrocks, A. D. (1995). The MADS-Box Family of Transcription Factors. *European Journal of Biochemistry, 229*(1), 1-13. doi: 10.1111/j.1432-1033.1995.0001l.x

Singh, B. (2005). Cowpea [Vigna unguiculata (L.) Walp]. *Genetic resources, chromosome engineering, and crop improvement: Grain legumes, 1*, 117-162.

Singh, K., Foley, R. C., & Onate-Sanchez, L. (2002). Transcription factors in plant defense and stress responses. [Research Support, Non-U.S. Gov't Review]. *Curr Opin Plant Biol, 5*(5), 430-436. doi: 10.1016/s1369-5266(02)00289-3

Singh, B., Ajeigbe, H., Tarawali, S. A., Fernandez-Rivera, S., & Abubakar, M. (2003). Improving the production and utilization of cowpea as food and fodder. *Field Crops Research, 84*(1), 169-177.

Sleator, R. D. (2011). Phylogenetics. [Review]. *Arch Microbiol, 193*(4), 235-239. doi: 10.1007/s00203-011-0677-x

Smith, A.F.A., Hubley, R. & Green, P. (2010) RepeatMasker Open-3.0. Retrieved from http://www.repeatmasker.org on October 31, 2013.

Sonnhammer, E. L., Eddy, S. R., & Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins, 28*(3), 405-420.

Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res, 33*(suppl 2), W465-W467.

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res, 32*(suppl 2), W309-W312.

Sun, S., Yu, J.-P., Chen, F., Zhao, T.-J., Fang, X.-H., Li, Y.-Q., & Sui, S.-F. (2008). TINY, a dehydration-responsive element (DRE)-binding protein-like transcription factor connecting the DRE-and ethylene-responsive element-mediated signaling pathways in Arabidopsis. *Journal of Biological Chemistry, 283*(10), 6261-6271.

Takada, S., Hibara, K., Ishida, T., & Tasaka, M. (2001). The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. *Development, 128*(7), 1127-1135.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol, 28*(10), 2731-2739.

Tempel, S. (2012). Using and Understanding RepeatMasker. In Y. Bigot (Ed.), *Mobile Genetic Elements* (Vol. 859, pp. 29-51): Humana Press.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res, 22*(22), 4673-4680.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence

alignment aided by quality analysis tools. *Nucleic Acids Res, 25*(24), 4876-4882.

Timko, M. P., Ehlers, J. D., & Roberts, P. A. (2007). Cowpea *Pulses, sugar and tuber crops* (pp. 49-67): Springer.

Timko, M. P., Rushton, P. J., Laudeman, T. W., Bokowiec, M. T., Chipumuro, E., Cheung, F., . . . Chen, X. (2008). Sequencing and analysis of the gene-rich space of cowpea. *BMC Genomics, 9*, 103. doi: 10.1186/1471-2164-9-103

Timko, M. P., & Singh, B. (2008). Cowpea, a multifunctional legume. *Genomics of tropical crop plants*, 227-258.

Timko, M. P., Huang, K., & Lis, K. E. (2012). Host Resistance and Parasite Virulence in Striga–Host Plant Interactions: A Shifting Balance of Power. *Weed Science, 60*(2), 307-315. doi: 10.1614/ws-d-11-00039.1

Udvardi, M. K., Kakar, K., Wandrey, M., Montanari, O., Murray, J., Andriankaja, A., . . . Town, C. D. (2007). Legume transcription factors: global regulators of plant development and response to the environment. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review]. *Plant Physiol, 144*(2), 538-549. doi: 10.1104/pp.107.098061

Van Verk, M. C., Gatz, C., & Linthorst, H. J. M. (2009). Transcriptional Regulation of Plant Defense Responses. In L. C. VanLoon (Ed.), *Plant Innate Immunity* (Vol. 51, pp. 397-438). London: Academic Press Ltd-Elsevier Science Ltd.

Wanamaker, S., and T.J. Close. 2011. HarvEST. HarvEST:Cowpea version 1.27. Available at http://harvest.ucr.edu (verified 27 Sept. 2011). University of California, Riverside, CA.

Wang, Z., Libault, M., Joshi, T., Valliyodan, B., Nguyen, H. T., Xu, D., . . . Cheng, J. L. (2010). SoyDB: a knowledge database of soybean transcription factors. [Article]. *BMC Plant Biol, 10*. doi: 10.1186/1471-2229-10-14

Westwood, J. H., Roney, J. K., Khatibi, P. A., & Stromberg, V. K. (2009). RNA translocation between parasitic plants and their hosts. *Pest management science, 65*(5), 533-539. doi: 10.1002/ps.1727

Wu, C., Sun, S., Nimmakayala, P., Santos, F. A., Meksem, K., Springman, R., . . . Zhang, H.-B. (2004). A BAC- and BIBAC-Based Physical Map of the Soybean Genome. *Genome Research, 14*(2), 319-326. doi: 10.1101/gr.1405004

Xie, Q., Frugis, G., Colgan, D., & Chua, N. H. (2000). Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Genes Dev, 14*(23), 3024-3036.

Xie, Q., Sanz-Burgos, A. P., Guo, H., Garcia, J. A., & Gutierrez, C. (1999). GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. *Plant Mol Biol, 39*(4), 647-656.

Xu, Z. S., Chen, M., Li, L. C., & Ma, Y. Z. (2011). Functions and Application of the AP2/ERF Transcription Factor Family in Crop Improvement. [Review]. *Journal of Integrative Plant Biology, 53*(7), 570-585. doi: 10.1111/j.1744-7909.2011.01062.x

Yandell, M. D. RepeatRunner. Retrieved from http://www.yandell-lab.org/software/repeatrunner.html on October 31, 2013.

Yeh, R.-F., Lim, L. P. and Burge, C. B. (2001) Computational inference of homologous gene structures in the human genome. *Genome Res.* **11**: 803-816.

Yoshii, M., Yamazaki, M., Rakwal, R., Kishi-Kaboshi, M., Miyao, A., & Hirochika, H. (2010). The NAC transcription factor RIM1 of rice is a new regulator of jasmonate signaling. *Plant J, 61*(5), 804-815. doi: 10.1111/j.1365-313X.2009.04107.x

Zarei, A., Körbes, A. P., Younessi, P., Montiel, G., Champion, A., & Memelink, J. (2011). Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the PDF1. 2 promoter in Arabidopsis. *Plant Mol Biol, 75*(4-5), 321-331.

Zhang, G., Chen, M., Chen, X., Xu, Z., Guan, S., Li, L.-C., . . . Ma, Y. (2008). Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (Glycine max L.). *J Exp Bot, 59*(15), 4095-4107. doi: 10.1093/jxb/ern248

Zhang, H., Jin, J., Tang, L., Zhao, Y., Gu, X., Gao, G., & Luo, J. (2011). PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. [Research Support, Non-U.S. Gov't]. *Nucleic Acids Res, 39*(Database issue), D1114-1117. doi: 10.1093/nar/gkq1141

Zhao, C., Avci, U., Grant, E. H., Haigler, C. H., & Beers, E. P. (2008). XND1, a member of the NAC domain family in Arabidopsis thaliana, negatively regulates lignocellulose synthesis and programmed cell death in xylem. *Plant J, 53*(3), 425-436. doi: 10.1111/j.1365-313X.2007.03350.x

Zheng, X., Chen, B., Lu, G., & Han, B. (2009). Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. *Biochem Biophys Res*

Commun, 379(4), 985-989. doi: 10.1016/j.bbrc.2008.12.163

Zhu, Q.-H., Guo, A.-Y., Gao, G., Zhong, Y.-F., Xu, M., Huang, M., & Luo, J. (2007). DPTF: a database of poplar transcription factors. *Bioinformatics, 23*(10), 1307-1308.

Zhu, T., Nevo, E., Sun, D., & Peng, J. (2012). Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. [Research Support, Non-U.S. Gov't]. *Evolution, 66*(6), 1833-1848. doi: 10.1111/j.1558-5646.2011.01553.x

Zhuang, J., Deng, D.-X., Yao, Q.-H., Zhang, J., Xiong, F., Chen, J.-M., & Xiong, A.-S. (2010). Discovery, phylogeny and expression patterns of AP2-like genes in maize. *Plant Growth Regulation, 62*(1), 51-58. doi: 10.1007/s10725-010-9484-7

Chapter three

Analysis of AP2-EREBP, NAC and WRKY Expression in Cowpea During Interaction with *Striga gesnerioides*

Abstract

Cowpea, *Vigna unguiculata* (L.) Walp., is one of the most important food and forage legumes in semi-arid tropical regions. One of the most destructive causes of crop yield loss in cowpea is *Striga gesnerioides* (witch weed). Some cultivars of cowpea show resistance to some strains of *Striga*, and susceptibility to other strains; this makes cowpea an ideal subject of study for host-parasite interactions. Transcription factors, which regulate many regulatory pathways, including pathways involved with defense, may be an important aspect of these host-parasite interactions. In this study, three TF families related to defense, namely AP2-EREBP, NAC, and WRKY, were investigated for their potential roles in the interactions between cowpea and *Striga*. In the AP2-EREBP, NAC and WRKY families, it was found that some clades of the AP2-EREBP and NAC families only changed expression significantly when cowpea showed resistance to *Striga gesnerioides* race SG3, whereas several clades of the AP2-EREBP, NAC and WRKY families, some of which are associated with defense in other plants, only changed expression significantly when cowpea showed susceptibility to *Striga gesnerioides* race SG4z. Together, these results suggest that cowpea TF distribution and activity is influenced in part by the stresses in the environment such as *Striga*, and that *Striga* may be "shutting down" certain defense mechanisms in cowpea to gain access to nutrients.

Introduction

Cowpea is an important food and forage legume in semi-arid tropical regions such as sub-Saharan Africa (Ehlers & Hall, 1997; Timko & Singh, 2008). Of the 4.9 million metric tons of cowpea produced across 10.4 million hectares worldwide

(faostat3.fao.org), most is produced by subsistence farmers in western and central Africa (faostat3.fao.org; Singh, 2005; Timko et al., 2007; Langyintuo et al., 2003). In these regions, cowpea is an important source of carbohydrates and protein (Nielsen et al., 1993; Hall et al., 2003; Singh, 2005; Timko et al., 2007). In addition, cowpea can be used as a companion crop with cereals (Oseni, 2010; Dahmardeh et al., 2010). Moreover, cowpea can endure harsher conditions such as low quality soils and dry conditions (Hall et al., 2002; Hall et al., 2004).

However, cowpea growth and crop yields can be decimated by many stresses, from drought and salinity to fungal infection and nematodes (Timko et al., 2007; Singh 2005; Roberts et al. 1996, 1997; Das et al. 2010). Parasitic plants such as *Striga gesnerioides* (witch weeds) are particularly devastating (Timko et al., 2007), reducing crop yields significantly (Alonge et al., 2005a; Alonge et al., 2005b), potentially causing 100% yield loss (Cardwell & Lane, 1995). Despite its importance, cowpea has been under-studied compared to other crop plants (Timko & Singh, 2008). Therefore, it is important to study a legume such as cowpea, and to use bioinformatics to study how cowpea responds to the stresses it faces.

Cowpea is a diploid legume with a genome of 620 Mbp, smaller than other legumes (Arumuganathan & Earle, 1991; Timko et al., 2008). Yet for a long time, a comprehensive genomic understanding of cowpea did not exist. One major step towards this understanding was the sequencing and analysis of the gene-rich space of the cowpea genome (Timko et al., 2008). Other cowpea sequencing efforts include the sequencing of expressed sequence tags (ESTs) (Close et al., 2007; Close

et al., 2011; Muchero et al., 2009), and bacterial artificial chromosome (BAC)-end sequences (BESs) (Luo et al., 2003; DR Cook, UC Davis & NCBI).

The current genome assembly for cowpea (v0.03), which represents cowpea genotype IT97K-499-35 (Close et al., 2011), consists of a combination of assemblies from several sources: a SOAP de novo assembly with 67X coverage, 39 Gb of Illumina GAII paired-end sequences (70-130) using TrueSeq chemistry, 250,000 GSS sequences with an average length of 609 bases, 30,000 BAC-end sequences with an average length of 673 bases, and BLASTN hits for 97% of EST-derived consensus sequences from HarvEST cowpea assembly P12 from www.harvest-blast.org (Wanamaker & Close, 2011). This assembly is the most comprehensive to date, and thus presents an excellent opportunity to study the regulation of genetic expression. This type of study will be crucial, as regulating genetic expression is central to many biological processes. Thus, it is imperative to study transcription factors (TFs), which interact with specific DNA sequences (MacQuarrie et al., 2011), making them important in many aspects of plant survival, from developmental roles (Shore & Sharrocks, 1995; Kubo et al., 2005; Zhao et al., 2008) to defense (Singh et al., 2002). Thus TFs are important for crop yields (Rushton et al., 2008) and are an important subject of study (Richardt et al., 2007; Udvardi et al., 2007; Rushton et al., 2008). Given the many stresses that plants face, the study of TFs in defense pathways can prove to be particularly important; TFs are instrumental in several defense-signaling pathways, including the jasmonic acid (JA) pathway and the salicylic acid (SA) pathway (Qiu et al., 2007; Thaler et al., 2012). In *Arabidopsis thaliana*, for example, Schweizer et al. (2013) found 41 TFs across several families (including

bHLH, ERF, NAC and WRKY) that were induced upon insect herbivory. Some of these TFs were dependent on the JA signaling pathway and some of which were independent of the JA pathway (Schweizer et al., 2013). In another example, when Ryu et al. (2006) studied the interaction between rice and the fungal pathogen *Magnaporthe grisea*, they found that some WRKY TFs in rice are shown to respond to JA signaling, whereas others respond to SA signaling. Two WRKY TFs, OsWRKY30 and OsWRKY83, were found to respond to both JA and SA treatments (Ryu et al., 2006). Besides the JA and SA pathways, TFs in the abscisic acid (ABA) pathway can also be instrumental in defense responses. For example, NAC TFs can be induced by ABA signals, and in turn induce genes that activate a host of other genes, including MYB and AP2-EREBP TFs, which contribute to a defense response (Puranik et al., 2012).

It is thus probable that any TF involved in the defense of cowpea against parasitic weeds are involved in stress signaling pathways. It is also probable that different clades of certain TF families in cowpea are involved with different functions, and different types of responses to parasitic plants. Since different clades of some TF families, such as NAC (Zhu et al., 2012; Jensen et al., 2010), have tissue-specific expression and are specific to certain functions (e.g., the SNAC clade being involved with defense (Zhu et al., 2012; Xie et al., 1999)), it is likely that some TF families in cowpea have clades specifically dedicated to defense against parasitic weeds. Therefore, the data produced from a recent oligonucleotide probe-based microarray analysis of both compatible and incompatible interactions between cowpea and *Striga gesnerioides* (Huang et al., 2012) provide an opportunity to determine which

TFs in cowpea are involved with host response to *Striga* parasitism. However, the data are based on oligonucleotide probes and thus do not directly identify the specific sequences in the cowpea genome assembly sequenced by Close et al. (2011) that are involved with a host plant defense response. Moreover, the microarray data do not directly identify which clades of TF families are involved. Such information is essential in understanding the regulatory networks involved in responding to stresses. A prime example of this is the identification of members of certain clades of the *Arabidopsis* NAC family that are involved in defense against various stresses (Jensen et al., 2010).

In this study, three families commonly associated with defense, namely AP2-EREBP, NAC and WRKY (Singh et al., 2002; Olsen et al., 2005; Van Verk et al., 2009), are chosen for identification of clades (and their respective members) involved with host response to *Striga* parasitism.

Methods

Sources of Data

Cowpea Genome Assembly. In all bioinformatics analyses, the cowpea genome assembly used is the Cowpea genome v0.03 (Close et al., 2011), which is available for BLAST searches and sequence retrieval (www.harvest-blast.org, Pottorff et al., 2012).

Version 0.03 for cowpea genotype IT97K-499-35 (Close et al., 2011), was assembled from several sources: a SOAP de novo assembly with 67X coverage, 39 Gb of Illumina GAII paired-end sequences (70-130) using TrueSeq chemistry, 250,000 GSS sequences with an average length of 609 bases, 30,000 BAC-end sequences with an

average length of 673 bases, and BLASTN hits for 97% of EST-derived consensus sequences from assembly P12 from www.harvest-blast.org. From the high coverage of the SOAP de novo assembly to the combination of GSS, EST and BAC end sequences; the cowpea version v0.03 assembly is the most comprehensive cowpea genomic assembly to date (Close et al., 2011). The cowpea assembly was annotated and translated to protein using the MAKER annotation pipeline (Cantarel et al., 2008).

Two versions of the cowpea assembly were used, each of which was created using different methods of exon detection and translation to amino acid sequences (Stephen Turner, personal communication, October 29, 2013): one that was annotated via MAKER before being processed through AUGUSTUS (Stanke et al., 2004; Stanke & Morgenstern, 2005), and one in which the low-complexity regions were masked via RepeatMasker (http://www.repeatmasker.org/; Tempel et al., 2012) and RepeatRunner (http://www.yandell-lab.org/software/repeatrunner.html) before being annotated via MAKER.

Microarray data. In generating the microarray data, Huang et al. (2012) studied the interaction of cowpea cultivar B301 with two strains of *Striga gesnerioides*: SG3, to which cultivar B301 is immune; and SG4z, to which B301 is susceptible. In studying cowpea resistance, expression at the cowpea root was studied 6 days after infection and 13 days after infection. In studying cowpea susceptibility, expression at the cowpea root was studied 13 days after infection.

Huang et al. (2012) fabricated a 385,000-feature Nimblegen microarray (Nimblegen Inc., Madison, WI) based on 43,253 cowpea unigenes identified by Chen et al. (2007)

and Timko et al. (2008); 60-nt oligonucleotide probes based on these unigenes were used to do this. On the microarray, 6-8 probes represent each gene-coding region. Gene Ontology (GO) was used to annotate the sequences on the array, and qRT-PCR was used to validate transcript levels in tissue samples.

Identification and Classification of TFs

For each version of the assembly, TFs were found in the cowpea genome using a workflow created by Lawson & Mackey (2011) (Supplementary Figure S1) in Taverna (Missier et al. 2010; Hull et al. 2006; Oinn et al., 2006); this pipeline uses Hidden Markov models via HMMER 3 (Eddy, 2009; Eddy, 2011), as well as protein domain finding via PFAM (Punta et al., 2012; Sonnhamer et al., 1997; Bateman et al., 2000). TF classification was done using the classification rules described by Lang et al. (2010).

The TF repertoires for each version of the cowpea whole genome assembly were collapsed into a non-redundant set using the following process: Sequences with identical names were found in both versions of the assembly. Sequences with identical names were observed to have identical sequences. For instance, a sequence with a name like "scaffold8708-processed-0.1" in one set would be identical to the sequence "scaffold8708-0.1" in the other dataset. Redundant names were then eliminated in Excel, and custom Perl scripts were used to retrieve the sequence set based on the resulting non-redundant set of names.

Multiple Sequence Alignment of TF Sequences

Multiple sequence alignment was done using T-Coffee (Di Tommaso et al., 2011) algorithms Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena &

Notredame, 2009; Chang et al., 2012). For both methods, the options to combine the alignment with all other known alignment methods made available on the T-Coffee web server (Notredame, tcoffee.crg.cat) was chosen in order to create a more accurate alignment based on a consensus between multiple methods (http://www.tcoffee.org/Documentation/t_coffee/t_coffee_tutorial.htm). The alignments from both methods were combined using the T-Coffee Combine algorithm (Di Tommaso et al., 2011). MEGA 5.2.2 (Tamura et al., 2011) was used for manual curation and visualization of the alignment.

Phylogenetic Analysis of TF Families

Format Converter v2.0.5 at the Los Alamos National Laboratory page (http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html) was used to convert the alignments in FASTA format into PHYLIP interleaved format. The file was then manually adjusted so that the file was in the proper input format for ProtTest 3 (Darriba, et al. 2011) and for PhyML 3.0 (Guindon et al., 2010). Model testing for maximum likelihood phylogeny was done using ProtTest 3 (Darriba et al., 2011) in order to find the best fitting amino acid substitution model for the aligned sequences. Phylogenetic trees were made using PhyML 3.0 (Guindon et al., 2010), using approximate likelihood ratio test (aLRT) (Anasimova & Gascuel, 2006), and automatically selecting the better choice between nearest-neighbor interchange (NNI) and subtree pruning and regraphing (SPR), and automatically selecting the better choice between Chi-square or Shimodaira-Hasegawa (SH)-like branch supports (Guindon et al. 2010). The approximate likelihood ratio test (aLRT) compares the likelihoods of the two optimal arrangements around a given

branch on a tree (Anisimova & Gascuel, 2006; Guindon et al., 2010) by using a log ratio of the likelihood of the current arrangement on a branch to that of the most probable alternative (Guindon et al., 2010). This method was chosen because it has a comparable accuracy to bootstrap, but much faster (Anasimova & Gascuel, 2006). Phylogenetic trees were visualized using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

Deeper Analysis of AP2-EREBP, NAC and WRKY families. After alignment, the aligned set was manually curated in order to show the conserved domains for the AP2-EREBP, NAC and WRKY families. For the AP2-EREBP family, selected sequences from each clade of the *Arabidopsis* AP2-EREBP superfamily analyzed in Dietz et al. (2010) and from each clade of the rice AP2-EREBP superfamily analyzed in Sharoni et al. (2011) were used to determine the clades to which cowpea AP2-EREBP sequences belonged according to the grouping scheme by Nakano et al. (2006). To analyze the NAC family, sequences from each clade of the NAC family analyzed by Zhu et al. (2012) were used to classify cowpea NAC sequences into clades. To analyze the WRKY family, sequences from each clade of the WRKY family analyzed by Li et al. (2012) were used to classify the cowpea WRKY sequences into clades.

Comparison to Microarray Oligonucleotides

Microarray sequences used were based on the oligonucleotide data used in Huang et al. (2012), which measured changes in expression upon parasitism by *Striga gesnerioides* strains SG3 (compatible interaction) and SG4z (incompatible interaction). The AP2-EREBP, NAC and WRKY TF protein sequences from cowpea

were searched against the oligonucleotide sequence set from Huang et al. (2012) using FASTX version 36.3.5a (Pearson et al., 1997). After that, any oligonucleotide found was searched against the NCBI NR database. If an oligonucleotide was found to be a member of AP2-EREBP, NAC and WRKY, the sequence was kept for further analysis.

When examining the expression of a sequence, up- or down-regulation was considered significant if the microarray sequence was either up-regulated by at least 2.00 related to control or down-regulated to at most 0.50 related to control. P-values versus control needed to be less than 0.05 in order for an expression value to be considered significant.

Results

Analysis of Defense-related TF Families and Comparison to Oligonucleotide Sequences on Microarray

The phylogenetic trees for ERF, AP2, NAC and WRKY are shown in Figures 1, 3, 4, and 6 respectively, with a red bar next to clades containing sequences whose expression is changed upon *Striga* parasitism (whether during a compatible or incompatible reaction). The RAV tree is displayed in Supplementary Figure S2 since no RAV sequences were significantly up- or downregulated upon *Striga* parasitism.

AP2-EREBP.

ERF. An expression analysis based on the oligonucleotide microarray data in Huang et al. (2012) (see Materials and Methods; Figure 2) shows that in the ERF group, one sequence in the group III (DREB A4) clade was upregulated early in an incompatible interaction (against *Striga gesnerioides* SG3). Two sequences in the

group I clade, an ERF in the group VI (ERF B3) clade, and four sequences in group IX are only upregulated in compatible interactions (against *Striga gesnerioides* SG4z), and an ERF in the group VIII (ERF B4) clade showed up-regulation throughout the compatible reaction as well as in the incompatible reaction. One sequence in the group II clade and one sequence in the group IX clade showed up-regulation throughout the compatible reaction and in the late stages of the incompatible reaction.

AP2. Upon analysis of AP2 sequences based on the oligonucleotide microarray from Huang et al. (2012), it was found that the AP2 group had five sequences that were upregulated in late stages when cowpea showed resistance to *Striga gesnerioides* SG3, two sequences that were upregulated in both the late stages of resistance and throughout susceptibility, and three sequences that were down-regulated in both the susceptible and the resistant interaction with *Striga gesnerioides* (Figure 3).

RAV. Three RAV sequences were found in cowpea, and then aligned and made into a phylogenetic tree (Supplementary Figure S2). An analysis based on the oligonucleotides from Huang et al. (2012) shows that no RAV was significantly up- or downregulated upon *Striga* parasitism.

NAC. An analysis of NAC expression upon *Striga* parasitism based on the oligonucleotide data from Huang et al. (2012) (Figure 5) shows that in the NAC family, when cowpea showed resistance to *Striga*, two Group II sequences were downregulated late in resistance while two group IIIb and two group IIIc sequences were upregulated early in resistance and one sequence in group IIIc was

upregulated late in resistance. On the other hand, several other NAC TFs were upregulated only when cowpea showed susceptibility to *Striga*: one sequence from Group Ia, one from Group Ib, two from group IIIb, two from group Va(2), three from group Vb and two from VIc. NACs that were upregulated in both incompatible and compatible reactions to *Striga* were two NACs from group Ia, two from group Ib, two from Va(1), two from Vb and two from VIc. Two group Ic TFs were downregulated upon parasitism, regardless of whether the interaction was compatible or incompatible.

WRKY. In the cowpea WRKY family, no sequences were found that were specific to resistance (Figure 7). Instead, there were two Group I N-terminal domains and four Group IIe WRKY sequences that were upregulated only when cowpea showed susceptibility to parasitism by *Striga gesnerioides* strain SG4z. Two Group I C-terminal domains were expressed in both compatible and incompatible reactions to *Striga*.

Discussion

Since the AP2-EREBP, NAC and WRKY families are commonly associated with defense, and different clades in these families may perform different functions, the phylogenetic organization of these families, as well as the possibility that these families activate a defense response against *Striga gesnerioides,* was investigated. Until now, there has been a need not only for a comprehensive TF repertoire in cowpea, but also for the identification of TFs involved in the interaction between cowpea and *Striga gesnerioides*.

When the activity of AP2-EREBP, NAC and WRKY TFs upon *Striga* parasitism was investigated, it was found that while AP2-EREBP and NAC showed sequences whose expression only changed when cowpea showed resistance to *Striga*; several members of the AP2-EREBP, NAC and WRKY families only changed in expression (usually in the form of upregulation) when cowpea showed susceptibility to *Striga*. A complete table with these sequences is in Table 1.

Phylogenetic Analysis of Selected TF Families

In this study, gaining a more comprehensive understanding was not the only goal of investigating the AP2-EREBP superfamily in cowpea. Determining the function of each AP2-EREBP member was not the goal of this study. Rather, the AP2-EREBP family was studied with a focus on which members were up- or downregulated upon *Striga* parasitism. In the process, the AP2-EREBP family was separated into three smaller groups: the ERF, AP2 and RAV families. This separation allowed for easier manual curation of the multiple alignments and thus more satisfactory phylogenetic trees. The clades of each family whose members changed in activity upon interaction with *Striga* were also identified to gain a more detailed understanding of the mechanisms by which cowpea responds to *Striga*.

ERF.

Expression. ERFs are known for their roles in defense against pathogens (Singh et al., 2002; Gutterson & Reuber, 2004). ERFs have been shown to regulate pathogenesis-related (PR) gene expression by binding to DNA sequences such as the GCC box (GCCGCC) (Ohme-Takagi & Shinshi, 1995; Buttner & Singh, 1997; Zarei et

al., 2011). For this reason, this study investigates the potential roles of cowpea ERFs in interactions with *Striga gesnerioides*.

In this study, oligonucleotides from the microarray study by Huang et al. (2012) that were found to be AP2-EREBP sequences were examined for significant changes in expression upon *Striga* parasitism. Those oligonucleotides that showed significant changes in expression were searched against the cowpea genome in order to search for their closest homologs in the cowpea whole genome assembly. This analysis found eleven ERF sequences that showed a significant change in expression. First, two group I ERF sequences were found to be upregulated during a compatible interaction with *Striga*. RAP2.4, which is listed by Nakano et al. (2006) as being a member of group I, has been proposed by Lin et al. (2008) to regulate stress responses such as responses to drought and developmental processes such as flowering time and root growth, and may serve as a means of crosstalk between photo-response and ethylene (ETH) signaling pathways. Thus, it is possible that when cowpea shows susceptibility to *Striga gesnerioides*, that the parasitic plant is manipulating ethylene signaling, and that group I ERFs are affected.

Second, a group II ERF sequence was found to be upregulated throughout susceptibility and late in resistance. This most likely implies that group-II ERF sequences are part of a stress response that may be required for a general stress response. This argument is further supported when one considers that RAP2.1, shown by Nakano et al. (2006) to be a member of group II, is shown by Dong & Liu (2010) to be important in abiotic stress response, in that it negatively controls responses to cold and drought in an ABA-independent manner. Since the SA

signaling pathway negatively regulates these abiotic stress responses (Atkinson & Urwin, 2012), it is possible that upon *Striga* parasitic attack, that cowpea may employ Group II ERFs as part of the SA signaling pathway in an attempt to focus the host defenses on biotic stress response. The actual involvement of group II ERFs in such pathways, like other TFs in this study, may prove to be an invaluable avenue of future research.

A group III (DREB A-4 according to Chapter 2) ERF was found to be upregulated early in an incompatible interaction with *Striga*. This finding is consistent with the association of DREBs with defense responses (Sun et al., 2008; Xu et al., 2011). A group VI (ERF B-3 according to Chapter 2) ERF was only upregulated in susceptible interactions. This finding contrasts with studies that find that the activity of ERFs in the B-3 group is essential in defense responses against pathogens (Xu et al., 2011; Lee et al., 2010; Gutterson & Reuber, 2004; Gu et al., 2002; Berrocal-Lobo et al., 2002; Shin et al., 2002), because based on these studies, it could have been expected that an ERF B-3 would be upregulated when cowpea shows resistance to *Striga*. It can be suggested that *Striga* is somehow manipulating the transcriptional mechanisms in cowpea, thereby "disarming" the host plant. A group VIII (ERF-B4 according to Chapter 2) ERF exhibited upregulation in both the compatible and the incompatible interactions between cowpea and *Striga.* This could mean that this sequence is involved in a stress response other than *Striga* parasitism. In particular, it could be responding to parasitism as if it were a wounding stress (Huang et al., 2012; Hiraoka & Sugimoto, 2008). Hiraoka & Sugimoto (2008) found that varieties of sorghum that show resistance to *Striga hermonthica* showed upregulation of

genes involved with responses to both wounding and pathogens. The group-VIII ERF sequence in this study could thus be necessary, but not sufficient, for a successful defense response against parasitic plants. Lastly, four group IX sequences (three of which are concentrated in one sub-group) were only upregulated upon susceptibility while one group IX sequence was upregulated both upon susceptibility and late in resistance. Like the other ERFs that showed upregulation both upon susceptibility and upon resistance, the group IX ERF sequence shown to be upregulated upon both susceptibility and resistance may be part of a response to other stresses. The upregulation of four group IX ERFs only during susceptibility may indicate that the manipulation of crosstalk between JA and ETH pathways may be one of the ways in which *Striga* shuts down host defenses. The exact mechanism for such manipulation would be an interesting avenue of research.

AP2. The AP2 family is normally associated with developmental processes (Zhang et al., 2008; Elliott et al., 1996; Chuck et al., 1998; Boutilier et al., 2002; Reichmann & Meyerowitz, 1998). In this study, the identification of five AP2 sequences upregulated when responding to *Striga* may at first glance be seen as contrasting with previous reports suggesting that AP2 proteins are only associated with developmental processes. However, one must consider that developmental processes involve distribution of nutrients to various tissues. A prime example of this is sugar transporters (Doidy et al., 2012). Among the TF families that may be in the same metabolic network as sugar transporters are TFs in the AP2 family (Rolland et al., 2006; Cernac & Benning, 2004). Thus the upregulation of AP2 TFs

could be a way for the host plant to draw resources away from *Striga*, thus "starving" the parasitic weed. Besides these AP2s, two AP2s were upregulated and three AP2s downregulated late in resistance and during the susceptible interaction (Figure 3), suggesting that these TFs could be involved in a different type of stress response, or that these TFs are necessary but not sufficient for an effective response to *Striga* parasitism.

RAV. No RAVs were significantly upregulated or downregulated during parasitism. This contrasts with a study by Li et al. (2011) that found a tomato RAV sequence that was needed for defense against bacterial wilt; it was expected that if RAV sequences were involved with one biotic stress, that they may be involved in responding to other stresses. It is thus possible that RAV sequences could be part of the signaling pathways for responding to bacteria and not to parasitic plants. Whether or not this is true could be the subject of further avenues of research.

NAC. In this study, it was found that seven NACs changed in expression when cowpea showed resistance to *Striga* parasitism. Firstly, two group II NACs are downregulated late in resistance. Group II is also named ONAC4 in Zhu et al. (2012), because of its homology to ONAC4 in rice. Nuruzzaman et al. (2012) identifies an ONAC4 in rice as being orthologous to ANAC008, which according to an expression analysis by Jensen et al. (2010), belongs to a clade whose members are activated in roots and stems, and are upregulated when responding to temperature and abscisic acid (ABA). Because of the involvement of this TF in ABA signaling, it is possible that ANAC008, as well as others in its clade, are involved in defense against abiotic and biotic stresses. In the face of biotic stresses, the ABA signaling pathway can

repress abiotic stress responses and activate pathogen defense response genes (Atkinson & Urwin, 2012). It is thus possible that the group-II NACs downregulated in this study are associated with the abiotic stress responses that are repressed by ABA signaling in the presence of biotic stresses (Atkinson & Urwin, 2012).

Second, two group IIIb NACs were upregulated early in resistance. NTL1, NTL3 and NTL7, which according to Zhu et al. (2012) are members of group IIIb, are members of groups I-2 and IV-2 identified in Jensen et al. (2010). These groups, according to Jensen et al. (2010), respond to multiple abiotic stresses but for the most part, not to ABA signaling, and are mainly expressed in the roots and in old leaves. This presents the possibility that cowpea employs group-IIIb NACs as part of an ABA-independent component of host defense against *Striga* parasitism.

Two group IIIc NACs were upregulated early in resistance while one group IIIc NAC was upregulated late in resistance. Group IIIc is also named ANAC11 in Zhu et al. (2012), named after ANAC011, a NAC TF that is highly expressed in roots and heavily upregulated during salt stress (Jensen et al., 2010). According to Zhu et al. (2012), another member of the IIIc group is RIM1, a NAC TF that, according to Yoshii et al. (2010), both positively and negatively regulates early jasmonic acid (JA) signaling response genes. Considering that the activation of the JA signaling pathways has been shown to activate pathogen defense response genes (Atkinson & Urwin, 2012), it can be argued that the three upregulated group IIIc NACs in cowpea are JA signaling regulators that activate a defense response against *Striga*.

Not all NAC TFs in cowpea were shown to be associated only with resistance to *Striga*, though. In Group Ia, two sequences that changed in expression were

upregulated both in compatible and in the late phases of incompatible interactions with *Striga*, while one sequence was only upregulated when cowpea was susceptible. In Group Ib, two sequences were upregulated both in compatible and in early phases of incompatible interactions with *Striga*, and one sequence was only upregulated when cowpea was susceptible.

Group Ia sequences can have a variety of functions (Zhu et al., 2012; Puranik et al., 2012), and can be expressed in roots, old leaves and fruits according to an expression analysis of NACs by Jensen et al. (2010). These tend to regulate growth, development and tissue formation at several stages of plant life (Xie et al., 1999; Takada et al., 2001), and can inhibit viral reproduction when transiently overexpressed (Xie et al., 1999). In rice, a group-Ia NAC, ONAC045, can enhance resistance to abiotic stresses (Zheng et al., 2009). It is thus possible that two group-Ia sequences in cowpea that were upregulated during both host resistance and susceptibility are either part of a stress signaling pathway that responds to other types of stress, or are necessary, but not sufficient, for effective response against parasitism by *Striga*. It is possible that the group Ia sequence that was only upregulated during susceptibility could negatively regulate one or more steps in the SA signaling pathway, a pathway that is shown to be downregulated in sorghum varieties that are susceptible to witchweeds (Hiraoka & Sugimoto, 2008).

Group Ib sequences, also known as NAC1, are typically expressed in roots (Jensen et al., 2010; Xie et al., 2000), and are upregulated in response to ABA (Jensen et al., 2010). Moreover, they enhance the development of lateral roots via auxin-regulated pathways (Xie et al., 2000). This involvement of group Ib NACs in multiple signaling

pathways could explain why two group Ib cowpea NACs in this study are upregulated both during host resistance and during host susceptibility, whereas one group Ib sequence is only upregulated when cowpea is susceptible to witchweed.

Two cowpea group-Ic sequences in this study were shown to be downregulated in both the late phases of resistance and in susceptibility to parasitism. Group Ic sequences, which may also be known as VNDs (vascular-related NAC domain protein) (Pinheiro et al., 2009), are also usually expressed in roots although they sometimes are expressed in fruits and stems, according to data by Jensen et al. (2010). These sequences regulate the development of the vascular system, which may include the development of xylem tissue (Kubo et al., 2005) and thickening of secondary walls (Mitsuda et al., 2005). Since witch weeds attach to roots and make connections with the host vascular system (Bailey & Reiss, 1998), it is not unexpected that regardless of whether or not cowpea is resistant to witch weeds, that *Striga* would "hijack" the transcriptional mechanisms that regulate tissue development in roots such as the group-Ic NACs in this study.

Although two group IIIb NACs were upregulated only upon resistance to *Striga* activity as mentioned above, two other group IIIb NACs were only upregulated upon susceptibility to *Striga*. As mentioned above, group IIIb acts in a relatively ABA-independent manner (Jensen et al., 2010). Thus, ABA-independent pathways may be other targets of *Striga* manipulation.

The two cowpea group VIc NACs in this study that were only upregulated when cowpea showed susceptibility may also be targets of such hijacking of tissue

development, albeit under different circumstances, i.e., whereas group-Ic sequences (discussed above) are downregulated regardless of whether the host plant is resistant, group VIc sequence activity is only induced when cowpea is susceptible. Group VIc NACs are also known to respond to stress. In *Arabidopsis*, for example, ANAC104, which is mainly expressed in stems, is activated when responding to heat (Jensen et al., 2010). ANAC104 suppresses xylem vessel differentiation (Zhao et al., 2008), so it is not unexpected that *Striga* would force group VIc NACs to upregulate, thereby facilitating access to the host vascular system. Besides repression of xylem vessel differentiation, group VIc sequences can be activated as a result of ABA and SA signaling, as is the case in rice (Nuruzzaman et al., 2012). ABA and SA signaling pathways can both repress JA signaling, which can then activate pathogen defense genes (Atkinson & Urwin, 2012). Thus it is also possible that *Striga* is using the group VIc sequences to shut down defenses that would otherwise save the host plant from parasitism.

Such shutting down of defenses could also be a reason why three group Vb NACs are only upregulated during the compatible reaction with witch weeds. Group Vb NACs are known for their involvement in stress responses (Zhu et al., 2012); this has been shown in *Arabidopsis thaliana* (Jensen et al., 2010) and in soybean (Pinheiro et al., 2009). The involvement of these group-Vb sequences in defense is relatively complex. More specifically, group Vb NACs (known as Stress NACs in Pinhiero et al. (2009) and III-3 in Jensen et al. (2010)) are highly expressed in old leaves and roots, and are upregulated during salt stress (Jensen et al. (2010)). Moreover, in rice, group Vb sequences were shown to activate upon treatment with ABA (Nuruzzaman

et al., 2012), and have been shown to increase resistance to salt and drought stress (Hu et al., 2006). Furthermore, Nuruzzaman et al. (2012) show that in rice, Os03g60080, a group Vb sequence, is upregulated after SA treatment. Pinhiero et al. (2009) showed that in soybean, that GmNAC3, a stress NAC (which is equivalent to group Vb in Zhu et al. (2012)), was strongly induced upon JA, ABA, and salt treatments. It is possible that when those sequences are upregulated, they may be at least indirectly involved in repressing a pathogen defense response, thus deceiving the host plant into treating the *Striga* infection like an abiotic stress response such as wounding. This would be consistent with the studies done by Hiraoka & Sugimoto (2008), in which sorghum varieties that treated *Striga hermonthica* parasitism as if it were wounding because of activation of JA signaling and repression of SA signaling (making it possible that group Vb sequences can be activated by JA signaling without the need for SA signaling).

Two members of group Va(2) were only upregulated during susceptibility, which implies that these sequences, like members of several groups mentioned above, may be targets of *Striga* manipulation. Two members of Va(1) were upregulated both late in resistance and throughout susceptibility, which, like some other groups mentioned above, would make members of Va(1) part of a general response. Together, members of groups Va(1) and Va(2) are, according to Jensen et al. (2010), part of group III-2. Members of group III-2 are mostly expressed in old leaves, fruits and to a lesser extent inflorescence (Jensen et al. (2010)); these sequences respond to a variety of abiotic stresses, as well as ABA signaling. Thus, it is possible that members of groups Va(1) and Va(2) are part of an ABA-dependent pathway, but it is

also possible that these sequences are part of other pathways where abiotic stresses are regulated, such as the JA or SA pathways (Atkinson & Urwin, 2012). The exact involvement of groups Va(1) and Va(2) among other groups in the NAC family in host-parasite interactions would be an interesting avenue of future research.

In general, from the group II sequences that are downregulated only upon resistance to *Striga* to the group Vb sequences that are upregulated only upon susceptibility to *Striga,* the cowpea NAC TFs in this study are shown to exhibit a wide variety of responses to *Striga* parasitism.

WRKY.

Expression. When the expression of cowpea WRKY TFs was analyzed based on oligonucleotide microarray data from Huang et al. (2012), it was found that two group I C-terminal WRKY domain sequences are upregulated both in host susceptibility and late phases of host resistance.

The sequence-specific DNA binding of Group I WRKY is mainly done using the C-terminal domain (Chi et al., 2013; Ishiguro & Nakamura, 1994; de Pater et al., 1996). This activity can be used in defense responses. For example, two proteins, SIB1 and SIB2, which are activated by pathogens such as the necrotrophic fungus *Botrytis cinerea* (Lai et al., 2011), activate AtWRKY33 by binding to its C-terminal WRKY domain, and thus activate a defense response. Thus, it is possible that the upregulation of group-I WRKY C-terminal domains in both compatible and incompatible interactions with *Striga* could be due to regulatory pathways that could be, at least in part, independent of signaling pathways that respond to *Striga* parasitism. Like with other TFs that are upregulated in host resistance and

susceptibility both, the WRKY group-I C-terminal domains could be necessary, but not sufficient, for an effective response to witch weed parasitism.

The expression data also show that two group I WRKY N-terminal domain sequences were only upregulated in a susceptible reaction to parasitism and four group IIe WRKYs.

The N-terminus of Group I WRKY TFs AtWRKY25 and -33 are phosphorylated by a MAP-kinase that represses SA signaling (Andreasson et al., 2005). Hiraoka & Sugimoto (2008) show that a susceptible reaction to *Striga hermonthica* involves the repression of SA signaling. It is thus possible that when cowpea shows susceptibility to witch weeds, that group-I WRKY N-terminal domains are repressing SA signaling.

OsWRKY13, a WRKY identified by Pandey & Somssich (2009) as a group IIe WRKY, is shown to not only activate through both SA- and JA-dependent resistance pathways; but also to activate SA-dependent genes and suppress JA-dependent genes (Qiu et al., 2007). Moreover, OsWRKY13 regulates other TFs in rice, such as seven AP2 TFs, which are downregulated by the group IIe WRKY TF (Qiu et al., 2009). It is therefore possible that when cowpea shows susceptibility to *Striga gesnerioides* SG4z, the group IIe WRKY TFs in cowpea, when upregulated, are altering the expression of TFs from several families. For example, it is possible that AP2 TFs (that may not be represented by the microarray data in Huang et al. (2012)) are being downregulated by the Group IIe WRKYs in Figure 7c. Moreover, OsWRKY13 is shown to upregulate one NAC TF and downregulate three others (Qiu et al., 2009). Thus, the cowpea group IIe WRKYs in this study, which are

upregulated in the compatible interaction according to microarray data by Huang et al. (2012), are upregulating at least one of several NACs, in particular one of the NAC TFs shown in Figure 5. Furthermore, group IIe WRKYs regulate other WRKY TFs (Qiu et al., 2009; Berri et al., 2009). Berri et al. (2009) show that group IIe WRKYs indirectly up-regulate Group I WRKY TFs. It is thus possible that the group-IIe cowpea WRKY sequences in Figure 7c are indirectly up-regulating the N-terminal Group I WRKYs in Figure 7a.

Thus, though no cowpea WRKYs in this study seemed to confer resistance against witch weeds, there is potential for cowpea WRKYs to play a role in regulating defense-related genes, including other TFs (Qiu et al., 2009; Eulgem et al., 2000; Eulgem et al., 2007; Rushton et al., 2010).

Limitations

The sequences in the oligonucleotide-based microarray data used in Huang et al. (2012), which numbered almost 50,000, were based on the cowpea GSR assembly (Chen et al., 2007; Timko et al., 2008), and thus did not represent all of the sequences in the cowpea genome.

Conclusions

Overall, this study fulfills a long-standing need for the identification of genes in defense-related TF families that could possibly confer resistance to witchweed parasitism.

The use of microarray data from Huang et al. (2012) led to the identification of TFs (and their clades) in two defense-related TF families that were upregulated exclusively during host resistance to *Striga gesnerioides* SG3. It is quite possible that

some of them increase host resistance to parasitism by taking nutrients and water away from the host-parasite interface and back into the host plant.

Besides the identification of TFs up- and downregulated upon host resistance to *Striga gesnerioides* SG3, TFs have been identified that only change in expression when cowpea is susceptible to *Striga gesnerioides* variety SG4z; the identification of these TFs confirms the finding by Huang et al. (2012) that the parasitic plant facilitates entry by suppressing host functions, and changes the expression of genes involved in other host functions in order to gain access to nutrients and water from the host plant. One method for *Striga* to do this could involve the use of avirulence (Avr) proteins, effectors and virulence factors to avoid detection by host defenses and to suppress the host immune response (Timko et al., 2012). Another method used by *Striga* and other parasitic weeds could be to use small RNAs. Westwood et al. (2009) suggest that host plants transfer small RNAs to silence certain genes in parasitic weeds, and that it is also possible that parasitic weeds are also sending small RNAs to silence host resistance mechanisms. Both methods can be investigated as possible ways in which parasitic plants manipulate host genetic expression in an attempt to rob the host of nutrients and water.

The TFs identified in this study as being up- or downregulated when cowpea is resistant or susceptible to witchweed may be active in regulatory networks involved with host interaction with parasitic plants. The nature of the regulatory networks involved in cowpea defense against abiotic and biotic stresses warrants further study, because understanding these regulatory networks can eventually lead to

ways in which cowpea can be made more resilient in the face of parasitic plants, among other stresses and pests.

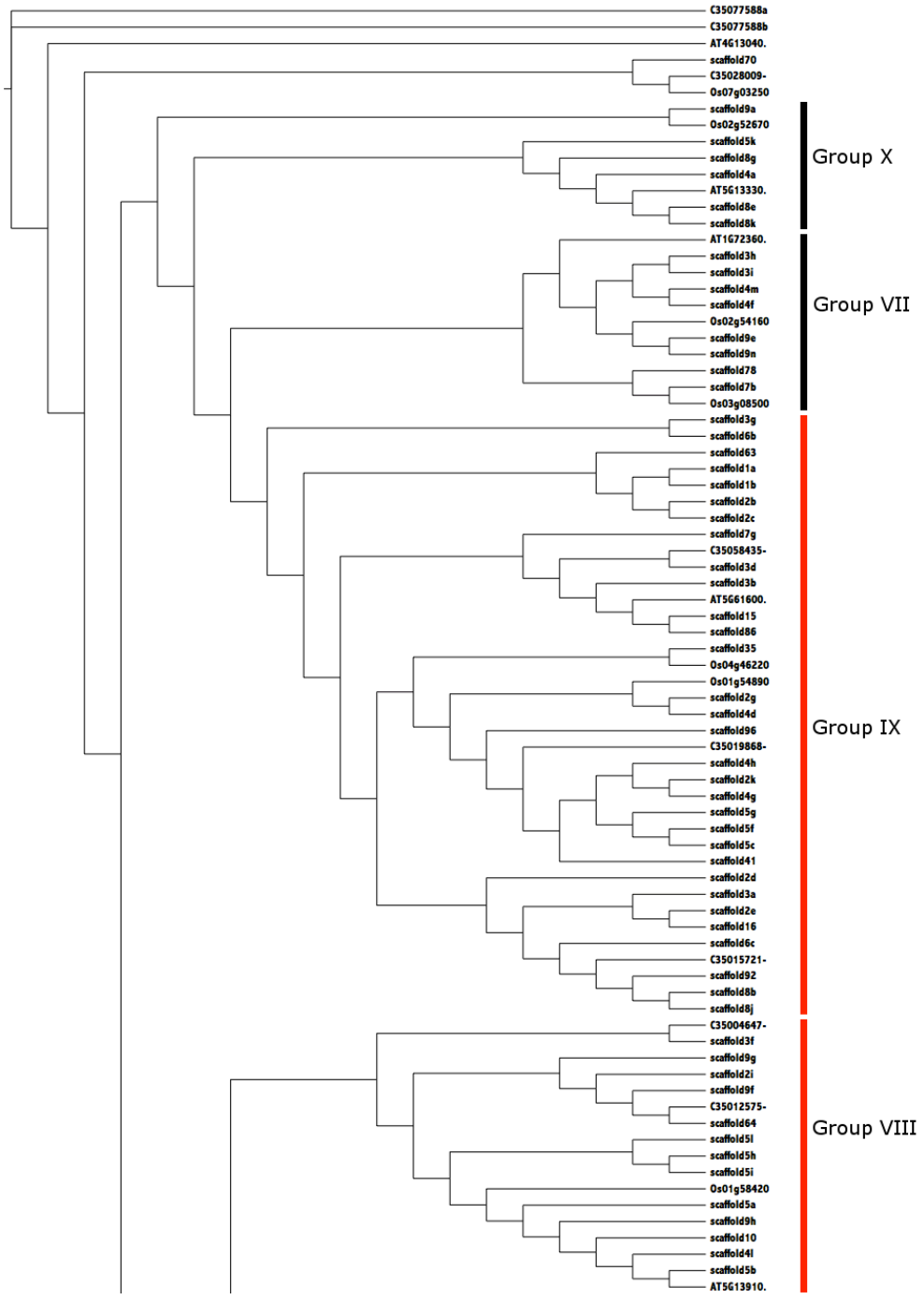


Figure 1. Cowpea ERF genes along with *Arabidopsis* and rice sequences. Shown is a phylogenetic tree of the ERF conserved domains constructed using the maximum

likelihood method via PhyML 3.0 (Guindon et al., 2010) with the LG amino acid substitution model (Le & Gascuel, 2008), automatically estimated invariable sites, and 5 gamma parameters. The tree was generated based on conserved cowpea ERF domains, as well as selected ERF domains from Arabidopsis and rice, whose selection is based on the ERF subgroup identification schemes of Dietz et al. (2010) and Sharoni et al. (2011), respectively. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S1.

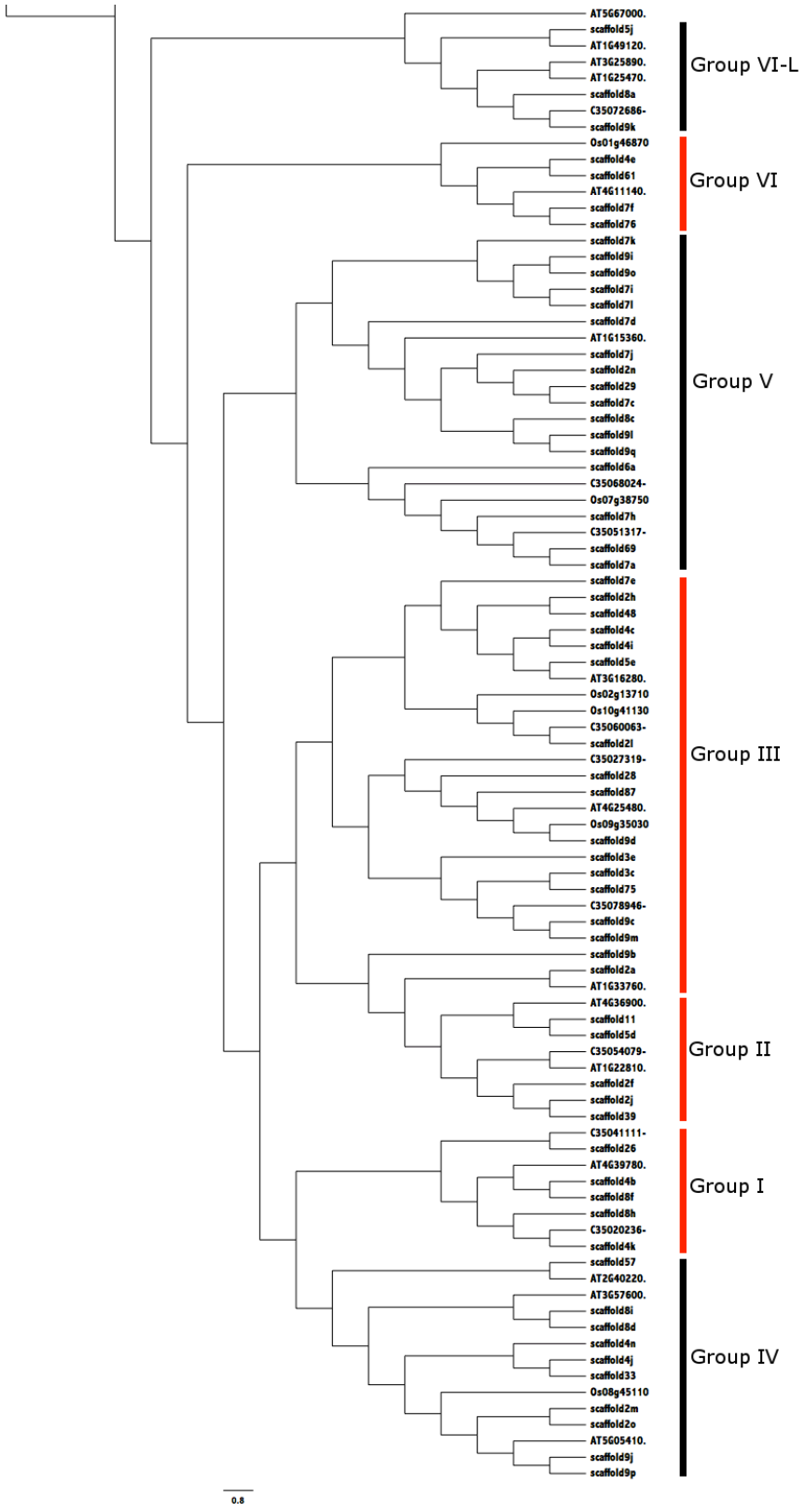
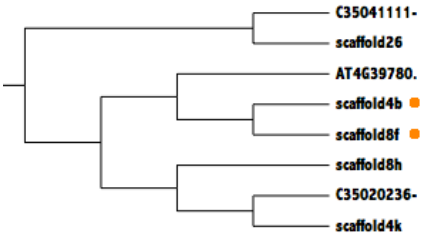
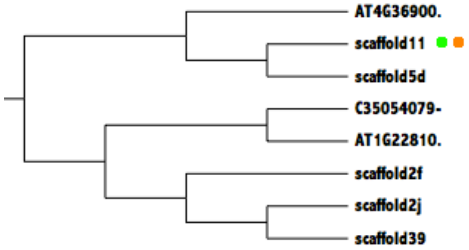


Figure 1, continued.

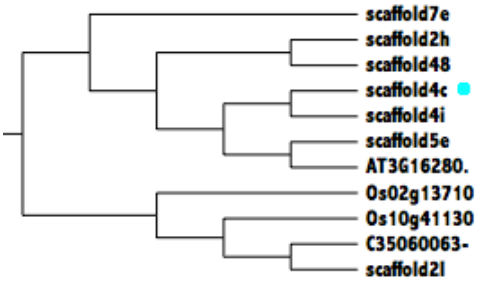
a) Group I



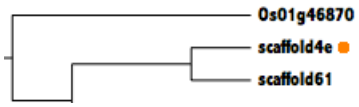
b) Group II



c) Group III



d) Group VI



e) Group VIII

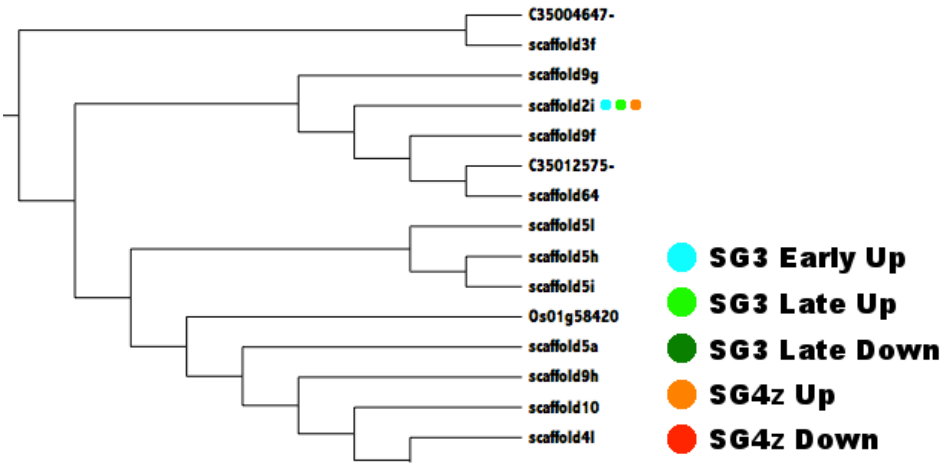


Figure 2. Clades of the cowpea ERF family involved with the cowpea host response to *Striga gesnerioides* strains SG3 (incompatible interaction) and/or SG4z (compatible) include the following: (a-c) Groups I-III, (d) Group VI, and (e-f) Groups VIII and IX.

f) Group IX

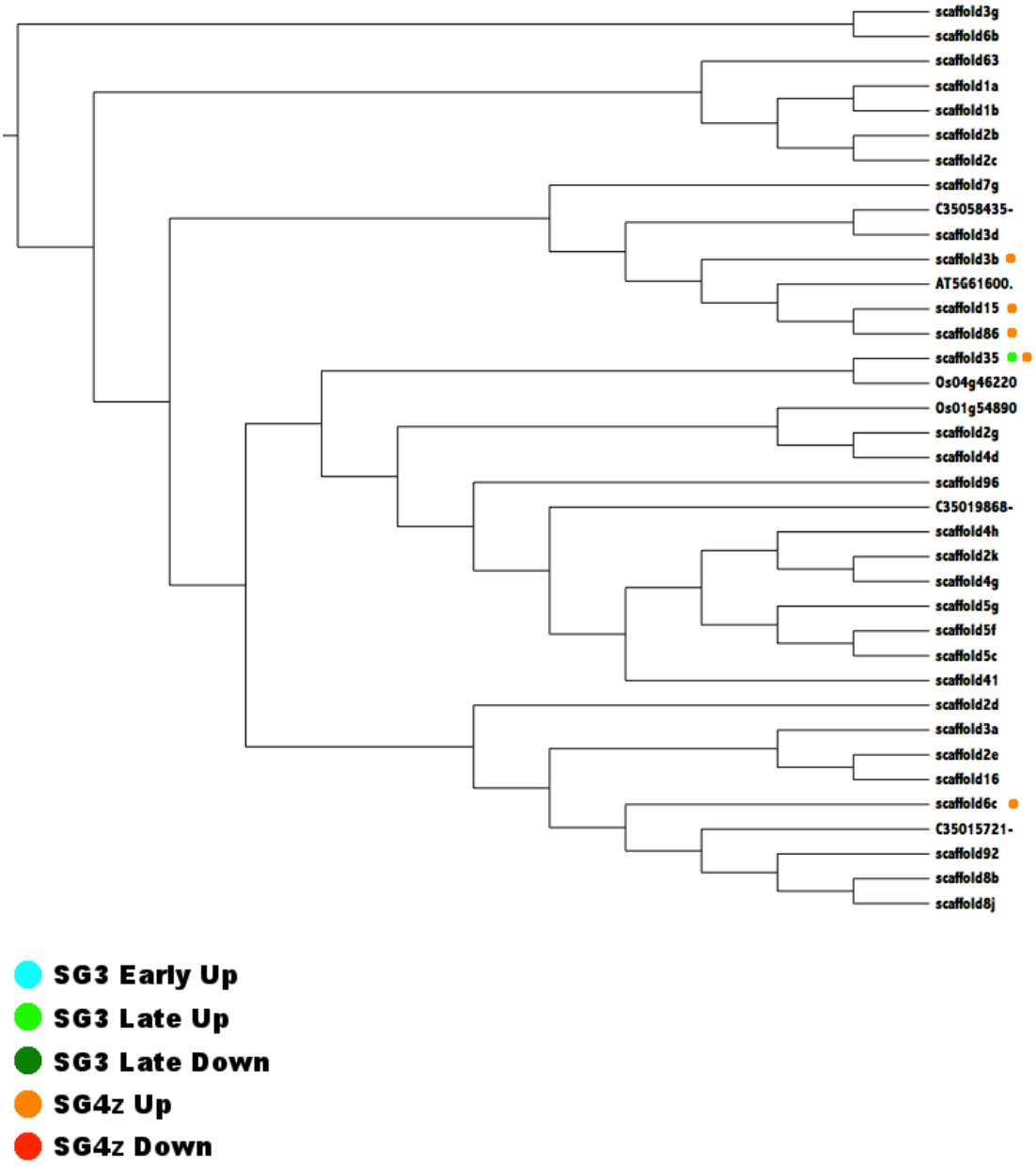


Figure 2, continued.

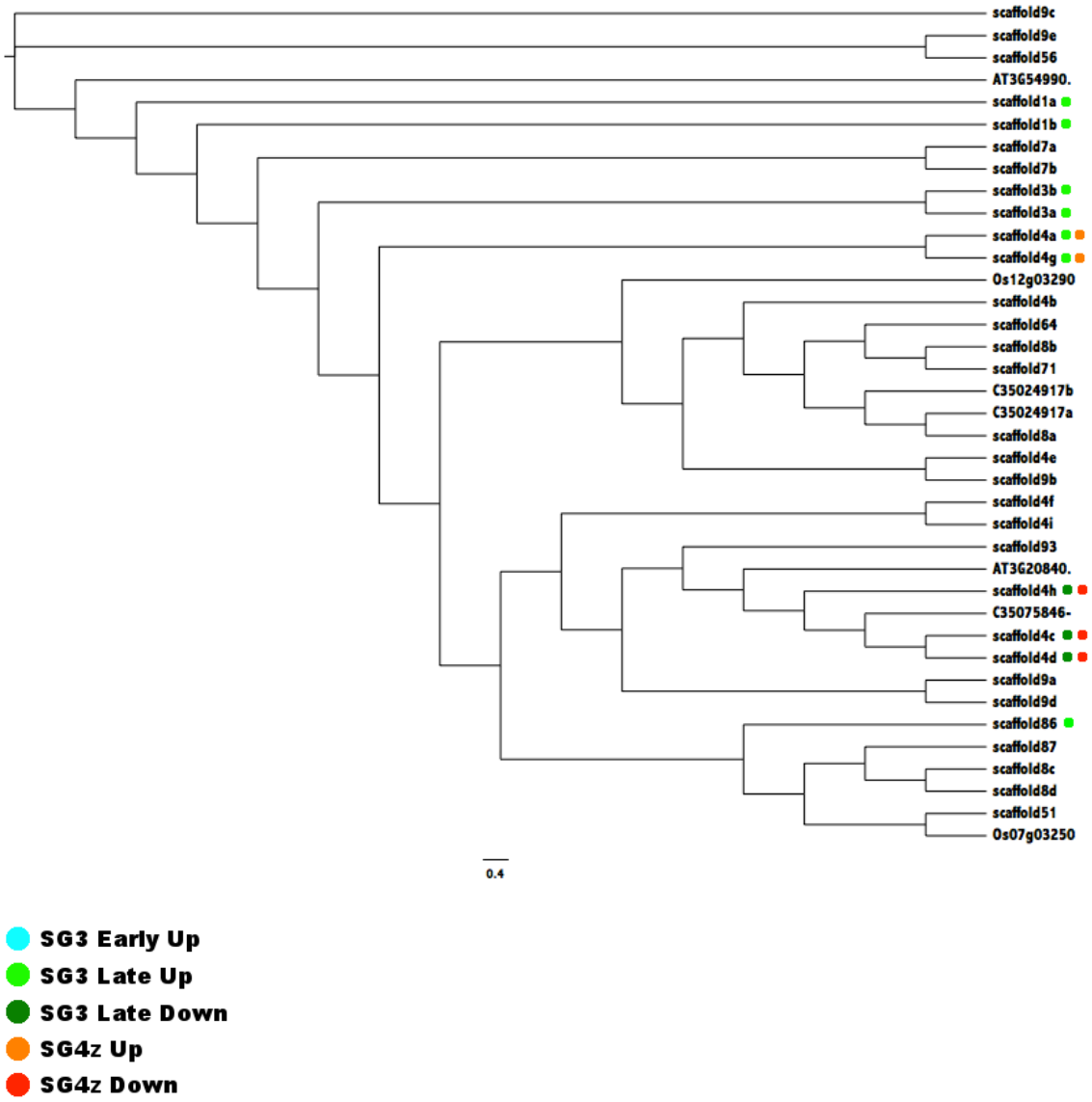


Figure 3. Cowpea AP2 genes. Shown is a phylogenetic tree of the AP2 conserved domains constructed using the maximum likelihood method via PhyML (Guindon et al., 2010) with the VT+G amino acid substitution model (Muller & Vingron, 2000), with an automatically estimated proportion of invariable sites, and 5 gamma parameters. The tree was calculated based on conserved AP2 domains, as well as representative ERF domains from Arabidopsis and rice, whose selection is based on the ERF subgroup identification schemes of Dietz et al. (2010) and Sharoni et al.

(2011), respectively. Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S2. On this tree are sequences whose expression is shown by the microarray data used in Huang et al. (2012) to change at least two-fold with $P < 0.05$ versus control.

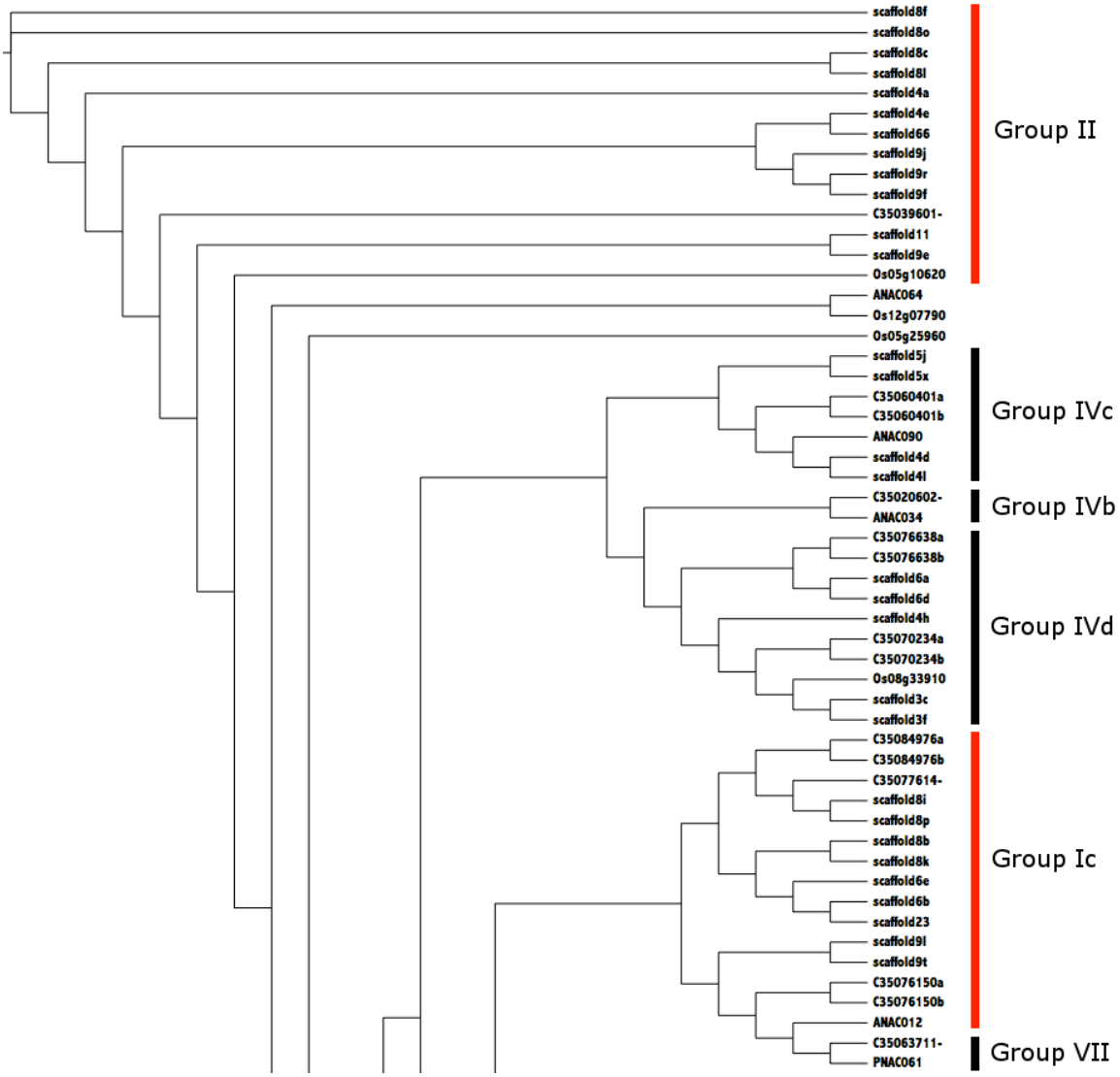


Figure 4. Cowpea NAC genes. Shown is a phylogenetic tree of the NAC conserved domains constructed using the maximum likelihood method with the JTT+G amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, and 5 gamma parameters. The tree was

calculated based on NAC domain sequences of cowpea, as well as the representative NAC domains used in the methodology of Zhu et al. (2012). Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S3.

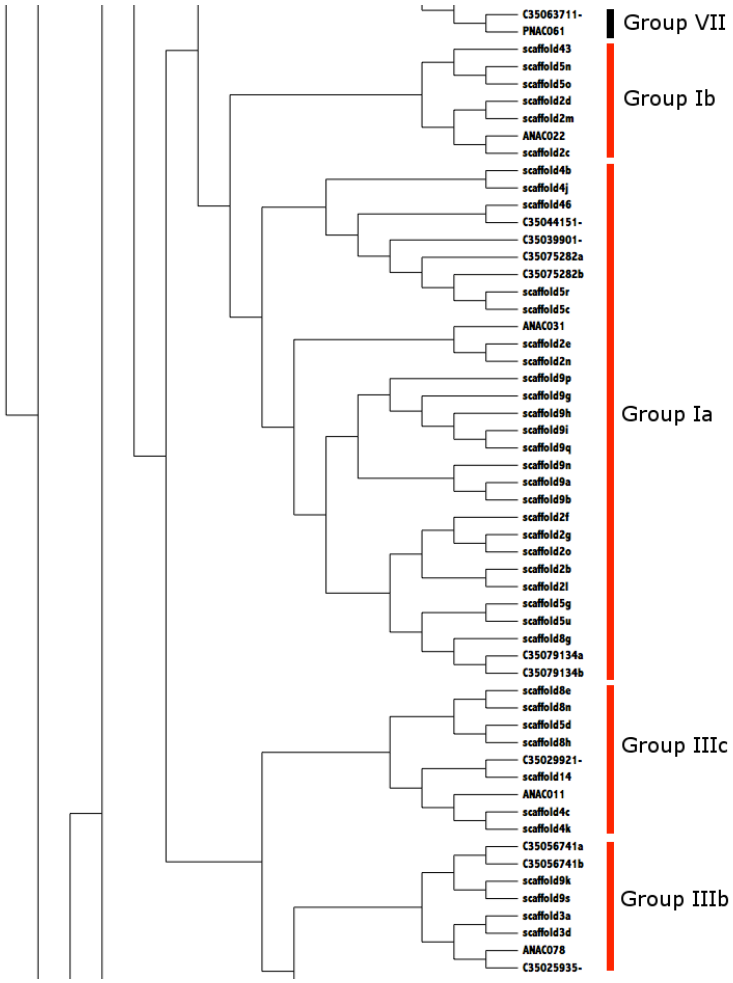


Figure 4, continued.

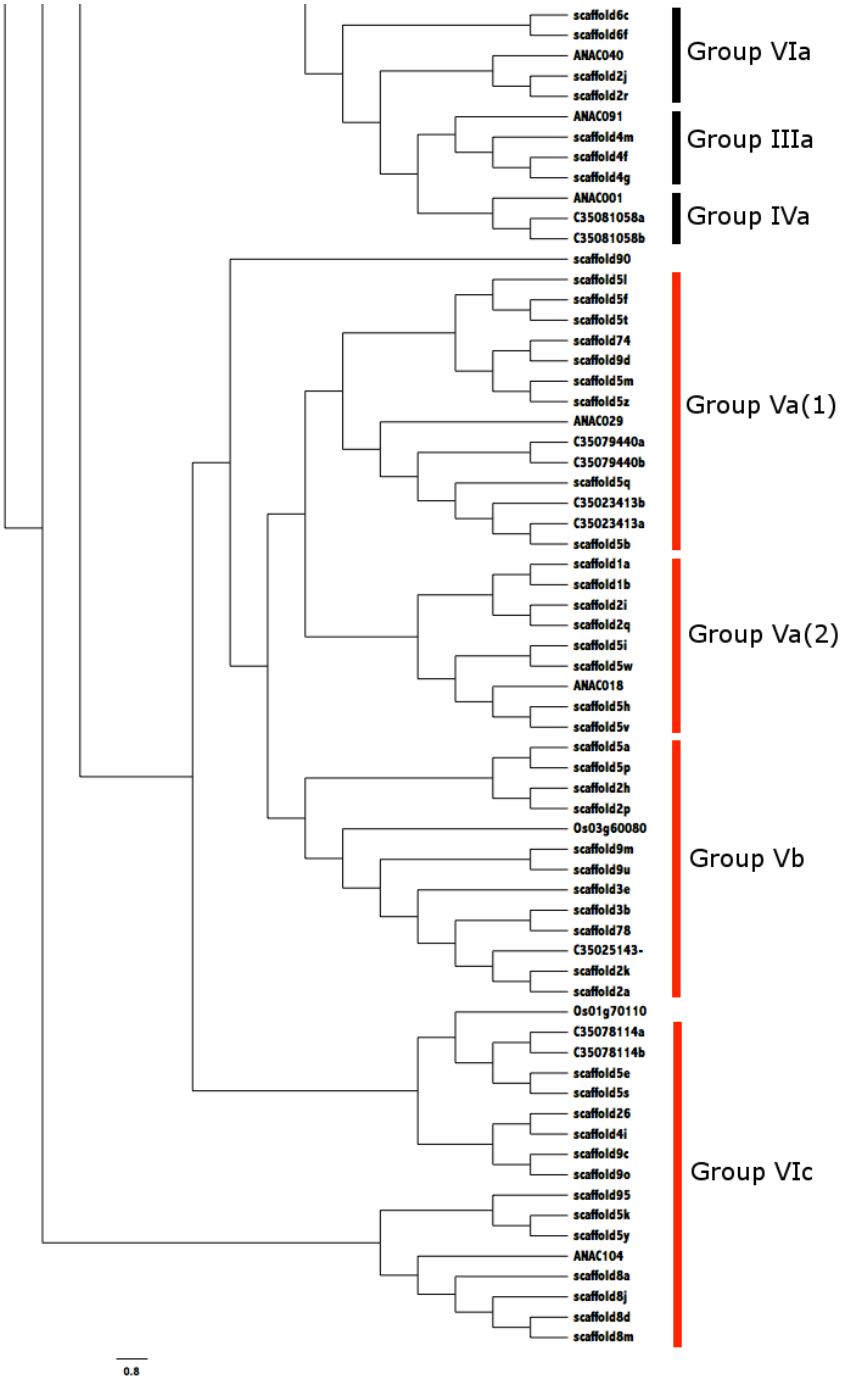


Figure 4, continued.

a) Group Ia

b) Group Ib

c) Group Ic

d) Group II

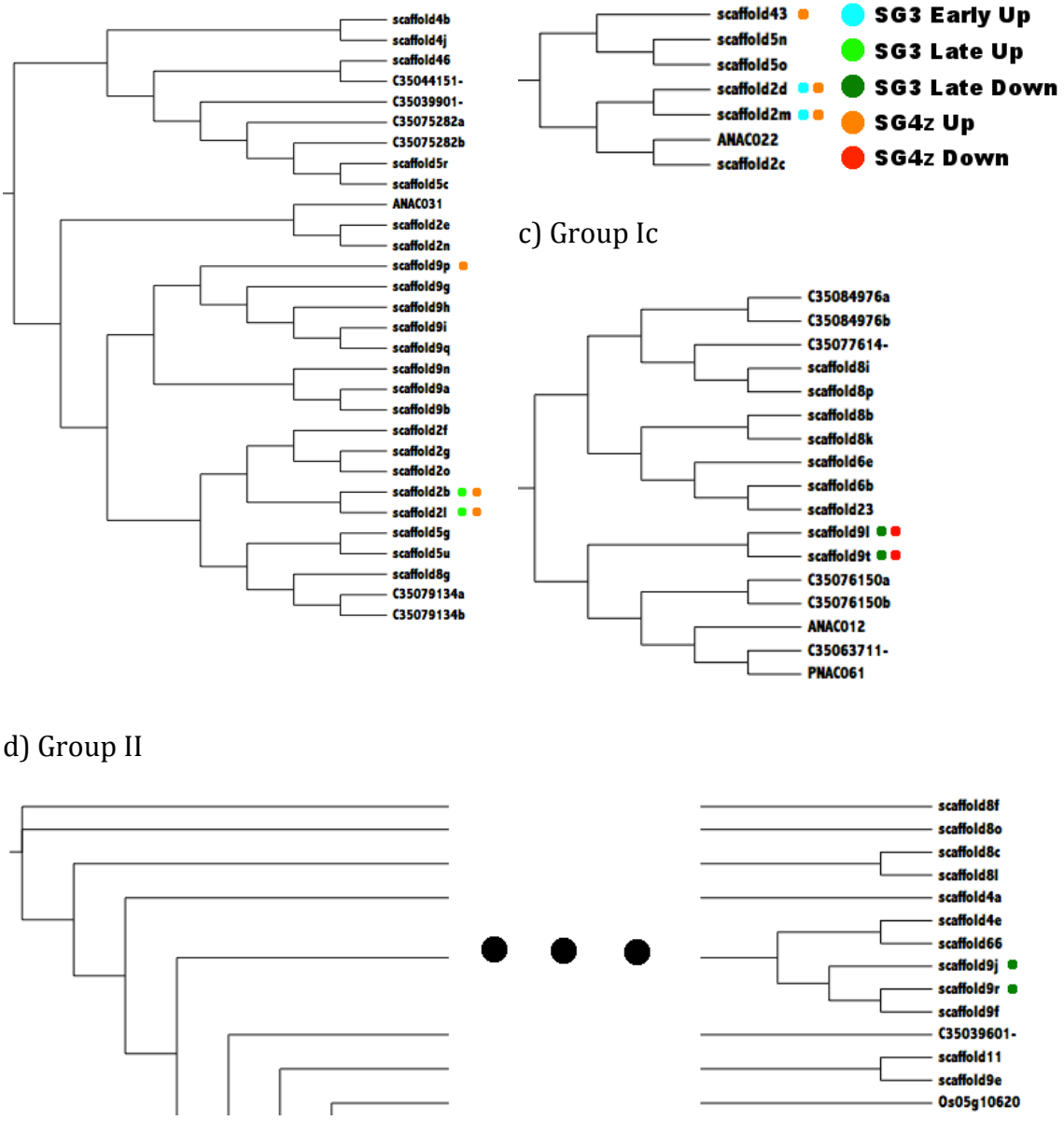
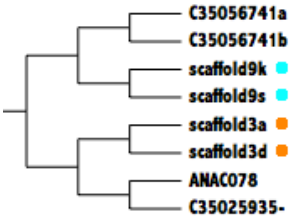


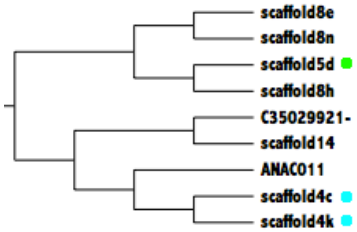
Figure 5. Clades of the cowpea NAC family involved with the cowpea host response to *Striga gesnerioides* strains SG3 (incompatible interaction) and/or SG4z (compatible) include the following groups: (a-c) Ia-c, (d) II, (e-f) IIIb-c, (g) Va(1), (h) Va(2), (i) Vb, and (j) VIc. Note: in (j), Os01g70110 belongs to the monocot-only VIb (Zhu et al., 2012). Since cowpea is a dicot and due to the expression of cowpea

sequences in this group, the group homologous to the VIb sequence is classified as being in group VIc.

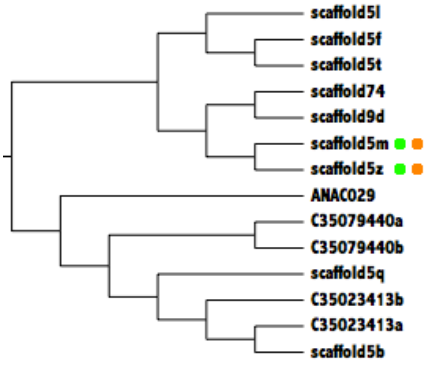
e) Group IIIb



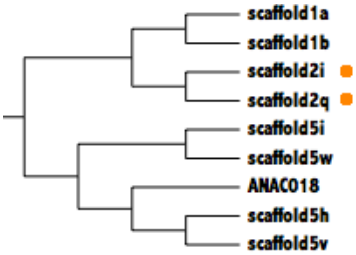
f) Group IIIc



g) Group Va(1)



h) Group Va(2)



i) Group Vb

j) Group VIc

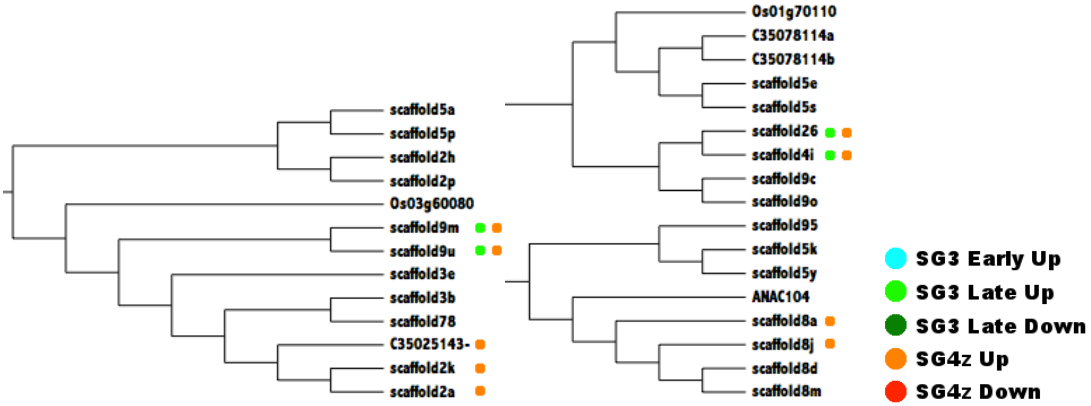


Figure 5, continued.

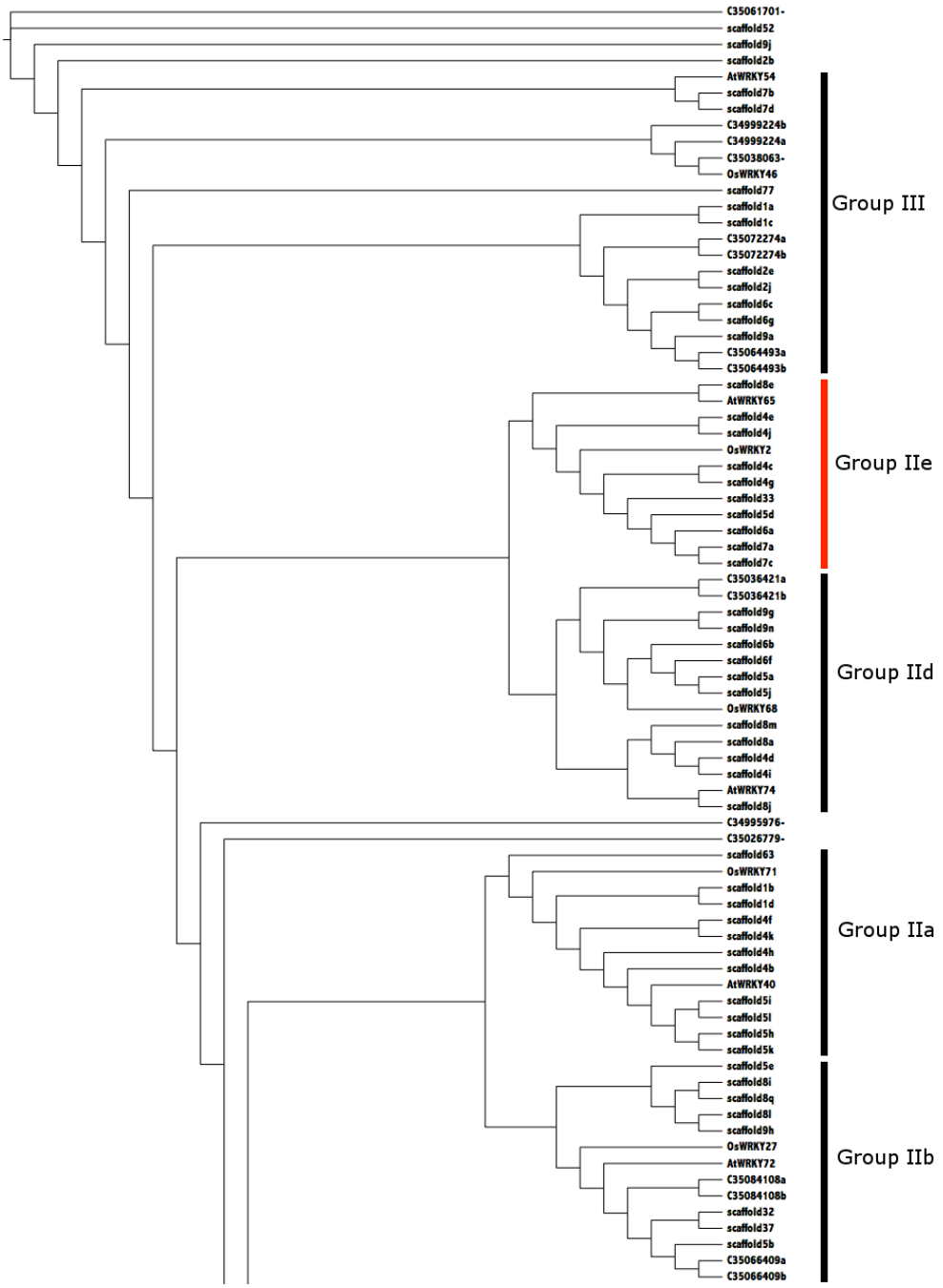


Figure 6. Cowpea WRKY genes. Shown is a phylogenetic tree of the WRKY conserved domains constructed using the maximum likelihood method via PhyML 3.0 (Guindon et al. (2010)) with the JTT+G amino acid substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable

sites, and 5 gamma parameters. The tree was generated from WRKY domain sequences of cowpea, as well as selected *Arabidopsis* and rice WRKY domains; this selection of *Arabidopsis* and rice WRKY sequences was based on the methodology of Li et al. (2012). Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S4.

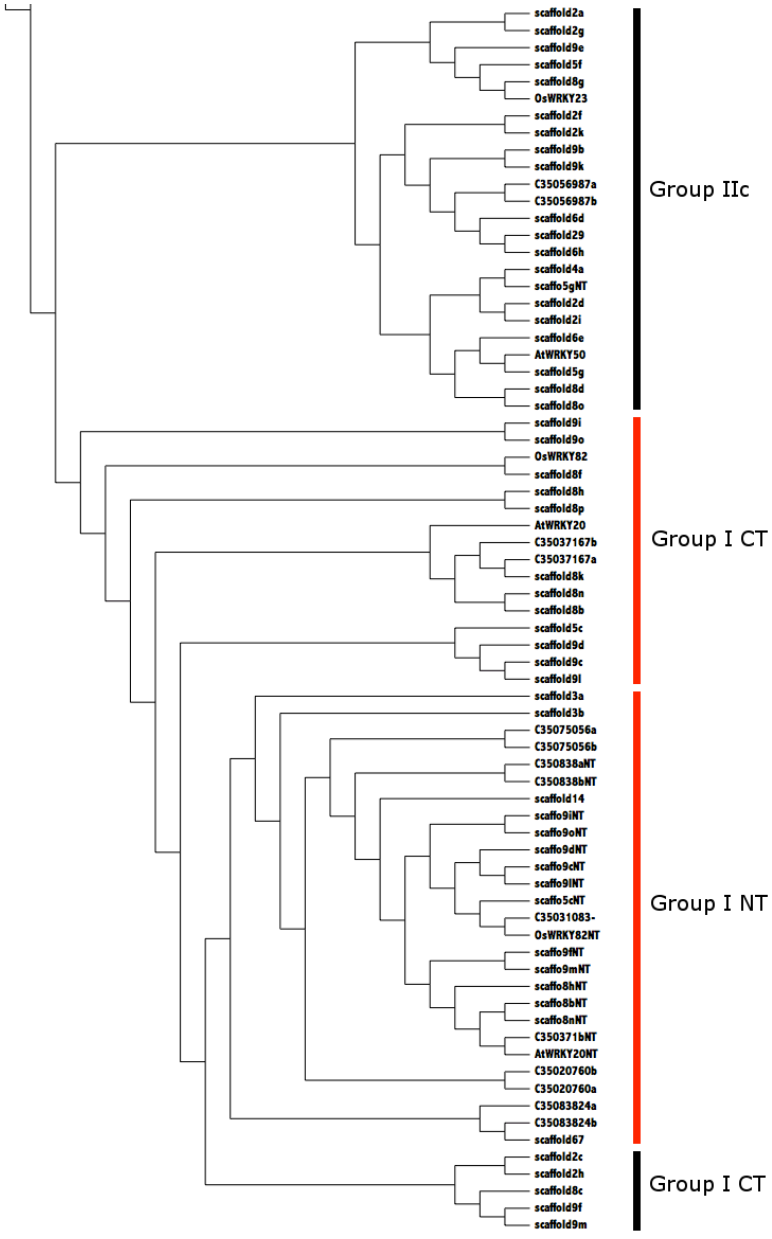
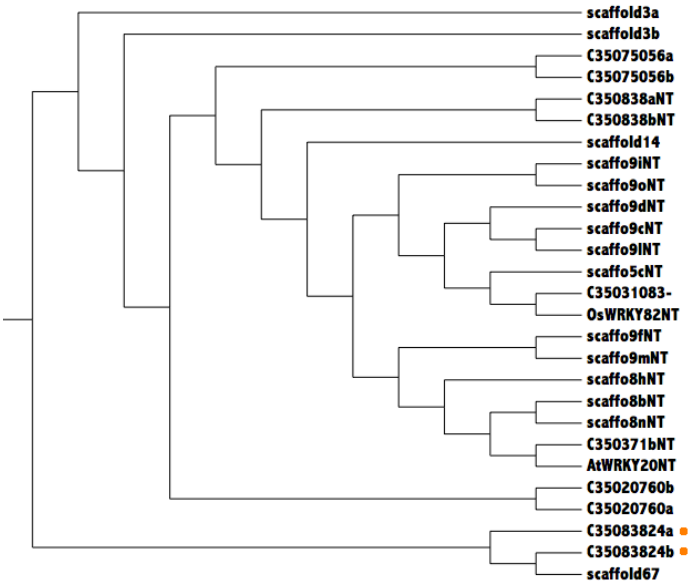
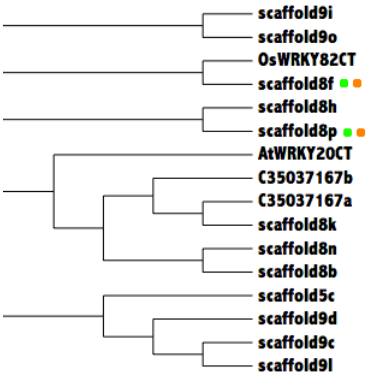


Figure 6, continued.

a) Group I NT



b) Group I CT



c) Group IIe

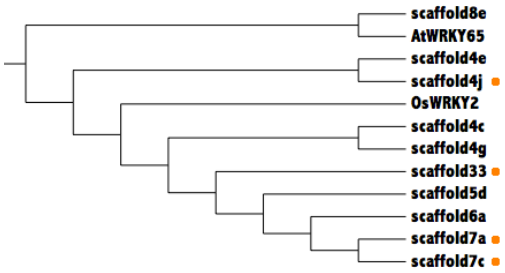
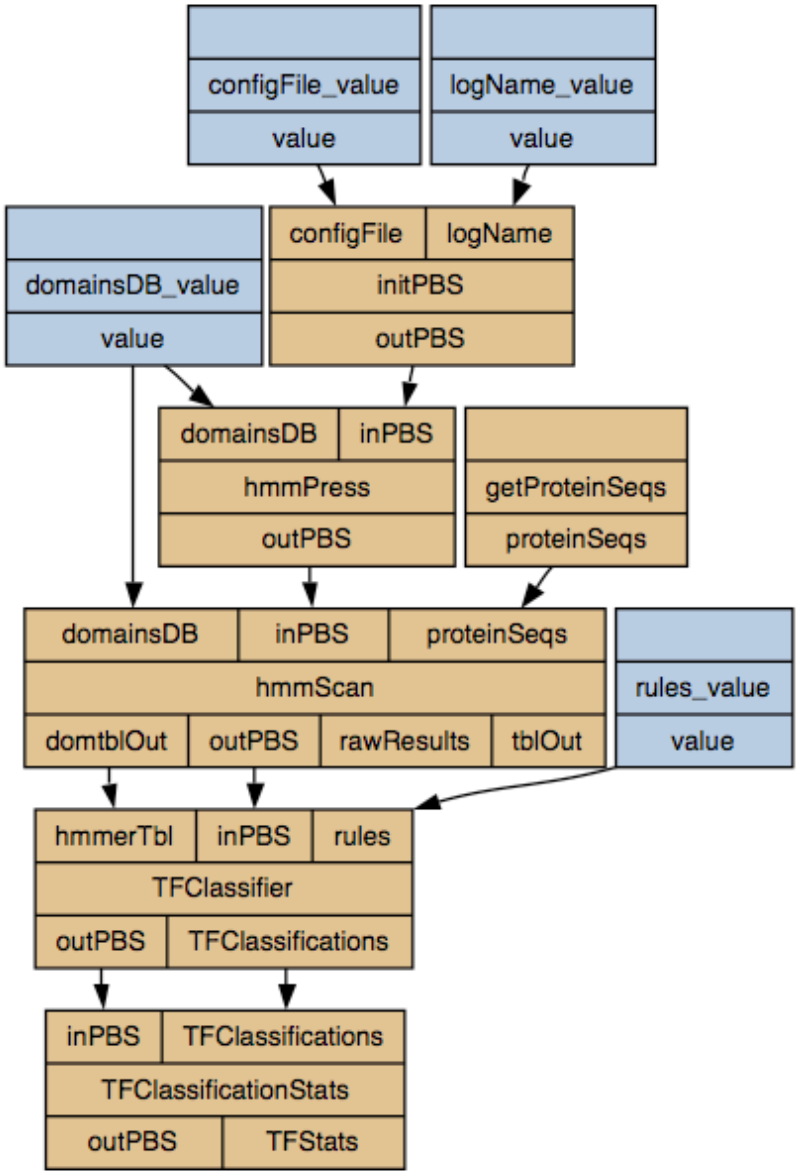


Figure 7. Clades of the cowpea WRKY family involved with the cowpea host response to *Striga gesnerioides* strains SG3 (incompatible inter-action) and/or SG4z (compatible) include the following groups: (a) I NT, (b) I CT, c) Group IIe.

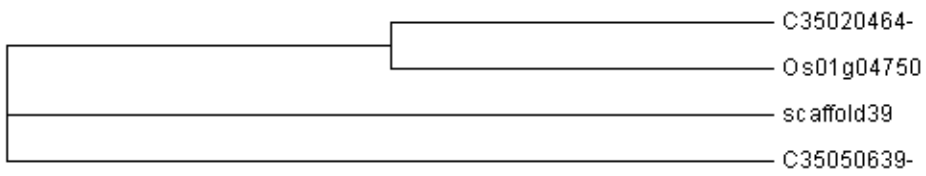
Table 1. Cowpea TFs from the AP2-EREBP, NAC and WRKY families whose expression is significantly changed upon parasitism by *Striga gesnerioides* strains SG3 and SG4z.

	SG3 Early Response		SG3 Late Response		SG4z	
	Up-regulated	Down-regulated	Up-regulated	Down-regulated	Up-regulated	Down-regulated
ERF						
Group I					scaffold40771-0.0 scaffold85129-0.0	
Group II			scaffold11245-0.1		scaffold11245-0.1	
Group III	scaffold42028-0.1					
ERF B3			scaffold42825-0.0			
Group VIII	scaffold24274-0.0		scaffold24274-0.0		scaffold24274-0.0	
Group IX			scaffold3552-0.0		scaffold15045-0.0 scaffold3016-0.0 scaffold3552-0.0 scaffold68868-0.0 scaffold86436-0.0	
AP2			scaffold10463-0.2 scaffold10463-augustus0.3 scaffold35030-0.1 scaffold35030-augustus0.3 scaffold40078-0.1 scaffold40078-augustus0.2 scaffold86592-0.1	scaffold40383-0.2 scaffold40383-0.3 scaffold40383-augustus0.4	scaffold40078-0.1 scaffold40078-augustus0.2	scaffold40383-0.2 scaffold40383-0.3 scaffold40383-augustus0.4
NAC						
Ia		scaffold21606-0.0 scaffold21606-augustus0.1			scaffold21606-0.0 scaffold21606-augustus0.1 scaffold93322-augustus0.2	
Ib	scaffold22594-0.2 scaffold22594-augustus0.3				scaffold22594-0.2 scaffold22594-augustus0.3 scaffold43420-0.0	
Ic				scaffold96551-0.0 scaffold96551-augustus0.1		scaffold96551-0.0 scaffold96551-augustus0.1
II				scaffold9386-0.1 scaffold9386-augustus0.6		
IIIb	scaffold94372-augustus0.4 scaffold94372-0.2				scaffold31854-0.1 scaffold31854-augustus0.4	
IIIc	scaffold42102-0.3 scaffold42102-augustus0.4		scaffold52810-0.1			
Va(1)			scaffold59215-0.1 scaffold59215-augustus0.4		scaffold59215-0.1 scaffold59215-augustus0.4	
Va(2)					scaffold27124-0.0 scaffold27124-augustus0.1	
Vb			scaffold9751-0.2 scaffold9751-augustus0.3		C35025143-0.1 scaffold20929-augustus0.4 scaffold20929-0.2	

					C35025143-0.1 scaffold9751-0.2 scaffold9751-augustus0.3	
VIc			scaffold2682-0.0 scaffold49445-0.2		scaffold2682-0.0 scaffold49445-0.2 scaffold82852-0.1 scaffold82852-augustus0.2	
WRKY						
I NT					C35083824-augustus0.3 C35083824-0.2	
I CT			scaffold87788-augustus0.3 scaffold87310-0.0		scaffold87788-augustus0.3 scaffold87310-0.0	
IIe					scaffold46551-augustus0.4 scaffold70584-augustus0.3 scaffold33087-0.0 scaffold70584-0.1	



Supplementary Figure S1. The TavernaPBS workflow used to classify TFs in cowpea, common bean and soybean. Figure adapted from Lawson & Mackey (2011).



Supplementary Figure S2. Phylogenetic tree of the RAV cowpea genes, as visualized in MEGA 5.2.2 (Tamura et al., 2011). The tree was calculated using PhyML 3.0 (Guindon et al., 2010) with the JTT+G substitution model (Jones, Taylor & Thornton, 1992), with an automatically estimated proportion of invariable sites, and 5 gamma parameters. Cowpea sequences are labeled as "C3…" or "scaffold…". Since the sequence names are in PHYLIP format, the full names can be found in Supplementary Table S5.

Supplementary Tables S1-S5 [Excel]. Full names of the sequences in Figures 1, 3, 4, 6 and Supplementary Figure S2.

References

Alonge, S. O., Lagoke, S. T. O., & Ajakaiye, C. O. (2005). Cowpea reactions to Striga gesnerioides I. effect on growth. Crop Protection, 24(6), 565-573. doi: http://dx.doi.org/10.1016/j.cropro.2004.08.014

Alonge, S. O., Lagoke, S. T. O., & Ajakaiye, C. O. (2005). Cowpea reactions to Striga gesnerioides II. Effect on grain yield and nutrient composition. Crop Protection, 24(6), 575-580. doi: http://dx.doi.org/10.1016/j.cropro.2004.08.015

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. [Comparative Study, Research Support, Non-U.S. Gov't]. Syst Biol, 55(4), 539-552. doi: 10.1080/10635150600755453

Andreasson, E., Jenkins, T., Brodersen, P., Thorgrimsen, S., Petersen, N. H., Zhu, S., . . . Mundy, J. (2005). The MAP kinase substrate MKS1 is a regulator of plant defense responses. *EMBO J, 24*(14), 2579-2589. doi: 10.1038/sj.emboj.7600737

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., . . . Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res, 34*(suppl 2), W604-W608.

Arumuganathan, K., & Earle, E. (1991). Nuclear DNA content of some important plant species. *Plant molecular biology reporter, 9*(3), 208-218.

Atkinson, N. J., & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. [Review]. *J Exp Bot, 63*(10), 3523-3543. doi: 10.1093/jxb/ers100

Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., & Sonnhammer, E. L. (2000). The Pfam protein families database. *Nucleic Acids Res, 28*(1), 263-266.

Berri, S., Abbruscato, P., Faivre-Rampant, O., Brasileiro, A. C., Fumasoni, I., Satoh, K., . . . Piffanelli, P. (2009). Characterization of WRKY co-regulatory networks in rice and Arabidopsis. *BMC Plant Biol, 9*, 120. doi: 10.1186/1471-2229-9-120

Berrocal-Lobo, M., Molina, A., & Solano, R. (2002). Constitutive expression of ETHYLENE-RESPONSE-FACTOR1 in Arabidopsis confers resistance to several necrotrophic fungi. *Plant J, 29*(1), 23-32.

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., . . . van Lookeren Campagne, M. M. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell, 14*(8), 1737-1749.

Buttner, M., & Singh, K. B. (1997). Arabidopsis thaliana ethylene-responsive element binding protein (AtEBP), an ethylene-inducible, GCC box DNA-binding protein interacts with an ocs element binding protein. *Proc Natl Acad Sci U S A, 94*(11), 5961-5966.

Byzova, M. V., Franken, J., Aarts, M. G., de Almeida-Engler, J., Engler, G., Mariani, C., . . . Angenent, G. C. (1999). Arabidopsis STERILE APETALA, a multifunctional

gene regulating inflorescence, flower, and ovule development. *Genes Dev, 13*(8), 1002-1014.

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., . . . Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res, 18*(1), 188-196. doi: 10.1101/gr.6743907

Cardwell, K. F., & Lane, J. A. (1995). Effect of soils, cropping system and host phenotype on incidence and severity of Striga gesnerioides on cowpea in West Africa. *Agriculture, Ecosystems & Environment, 53*(3), 253-262. doi: http://dx.doi.org/10.1016/0167-8809(94)00575-Y

Cernac, A., & Benning, C. (2004). WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *Plant J, 40*(4), 575-585. doi: 10.1111/j.1365-313X.2004.02235.x

Chang, J. M., Di Tommaso, P., Taly, J. F., & Notredame, C. (2012). Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. *BMC Bioinformatics, 13 Suppl 4*, S1. doi: 10.1186/1471-2105-13-s4-s1

Chen, X., Laudeman, T. W., Rushton, P. J., Spraggins, T. A., & Timko, M. P. (2007). CGKB: an annotation knowledge base for cowpea (Vigna unguiculata L.) methylation filtered genomic genespace sequences. *BMC Bioinformatics, 8*, 129. doi: 10.1186/1471-2105-8-129

Chi, Y., Yang, Y., Zhou, Y., Zhou, J., Fan, B., Yu, J.-Q., & Chen, Z. (2013). Protein–Protein Interactions in the Regulation of WRKY Transcription Factors. *Mol Plant, 6*(2), 287-300. doi: 10.1093/mp/sst026

Chuck, G., Meeley, R. B., & Hake, S. (1998). The control of maize spikelet meristem fate by the APETALA2-like gene indeterminate spikelet1. *Genes Dev, 12*(8), 1145-1154.

Close T. J., Luo M., Lonardi S., Alpert M., Duma D., Bozdag S., Farmer A., May G. D., Jackson S. A., Ma, Y., Gu, Y.Q., Wu, J.,You, F.M., Diop, N.N., Wanamaker S., Fenton R. D., Timko M. P., Roberts P. A. and Ehlers J. D. 2011. Genome Resources For Cowpea Genotype IT97K-499-35. Plant and Animal Genomes XIX Conference, Town and Country Convention Center, San Diego, CA.

Close, T., Wanamaker, S., Roose, M. L., & Lyon, M. (2007). HarvEST. In D. Edwards (Ed.), *Plant Bioinformatics* (Vol. 406, pp. 161-177): Humana Press.

Dahmardeh, M., Ghanbari, A., Syasar, B., & Ramrodi, M. (2009). Intercropping maize (Zea mays L.) and cow pea (Vigna unguiculata L.) as a whole-crop forage: Effects of planting ratio and harvest time on forage yield and quality. *Journal of Food, Agriculture and Environment, 7*(2), 505-509.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics, 27*(8), 1164-1165.

Das, S., Ehlers, J., Close, T., & Roberts, P. (2010). Transcriptional profiling of root-knot nematode induced feeding sites in cowpea (Vigna unguiculata L. Walp.) using a soybean genome array. *BMC Genomics, 11*(1), 480.

de Pater, S., Greco, V., Pham, K., Memelink, J., & Kijne, J. (1996). Characterization of a zinc-dependent transcriptional activator from Arabidopsis. *Nucleic Acids Res, 24*(23), 4624-4631.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., . . . Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res, 39*(suppl 2), W13-W17.

Dietz, K. J., Vogel, M. O., & Viehhauser, A. (2010). AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. *Protoplasma, 245*(1-4), 3-14. doi: 10.1007/s00709-010-0142-8

Doidy, J., Grace, E., Kuhn, C., Simon-Plas, F., Casieri, L., & Wipf, D. (2012). Sugar transporters in plants and in their interactions with fungi. [Review]. *Trends Plant Sci, 17*(7), 413-422. doi: 10.1016/j.tplants.2012.03.009

Dong, C.-J., & Liu, J.-Y. (2010). The Arabidopsis EAR-motif-containing protein RAP2.1 functions as an active transcriptional repressor to keep stress responses under tight control. *BMC Plant Biol, 10*(1), 47.

Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics, 14*(9), 755-763.

Eddy, S. R. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform, 23*(1), 205-211.

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput Biol, 7*(10), e1002195.

Ehlers, J., & Hall, A. (1997). Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Research, 53*(1), 187-204.

Elliott, R. C., Betzner, A. S., Huttner, E., Oakes, M. P., Tucker, W. Q., Gerentes, D., . . . Smyth, D. R. (1996). AINTEGUMENTA, an APETALA2-like gene of Arabidopsis

with pleiotropic roles in ovule development and floral organ growth. *Plant Cell, 8*(2), 155-168.

Elowad, H. O., & Hall, A. E. (1987). Influences of early and late nitrogen fertilization on yield and nitrogen fixation of cowpea under well-watered and dry field conditions. *Field Crops Research, 15*(3), 229-244.

Eulgem, T., Rushton, P. J., Robatzek, S., & Somssich, I. E. (2000). The WRKY superfamily of plant transcription factors. [Review]. *Trends Plant Sci, 5*(5), 199-206.

Eulgem, T., & Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review]. *Curr Opin Plant Biol, 10*(4), 366-371. doi: 10.1016/j.pbi.2007.04.020

The Food and Agriculture Organization of the United Nations. (2011). FAOSTAT. Retrieved from faostat3.fao.org on June 22, 2013.

Format Converter v2.1.5. (2013) Retrieved from http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html on June 22, 2013.

Galeano, C., Fernandez, A., Gomez, M., & Blair, M. (2009). Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (Phaseolus vulgaris L.). *BMC Genomics, 10*(1), 629.

Gao, G., Zhong, Y., Guo, A., Zhu, Q., Tang, W., Zheng, W., . . . Luo, J. (2006). DRTF: a database of rice transcription factors. *Bioinformatics, 22*(10), 1286-1287. doi: 10.1093/bioinformatics/btl107

Gepts, P., Aragão, F. J., de Barros, E., Blair, M. W., Brondani, R., Broughton, W., . . . Lariguet, P. (2008). Genomics of Phaseolus beans, a major source of dietary protein and micronutrients in the tropics *Genomics of tropical crop plants* (pp. 113-143): Springer.

Gill, N., Findley, S., Walling, J. G., Hans, C., Ma, J., Doyle, J., . . . Jackson, S. A. (2009). Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiol, 151*(3), 1167-1174. doi: 10.1104/pp.109.137935

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., . . . Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Nucleic Acids Res, 40*(Database issue), D1178-1186. doi: 10.1093/nar/gkr944

Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. (2010). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res, 38*(Database issue), D843-846. doi: 10.1093/nar/gkp798

Gu, Y. Q., Wildermuth, M. C., Chakravarthy, S., Loh, Y. T., Yang, C., He, X., . . . Martin, G. B. (2002). Tomato transcription factors pti4, pti5, and pti6 activate defense responses when expressed in Arabidopsis. *Plant Cell, 14*(4), 817-831.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood

phylogenies: assessing the performance of PhyML 3.0. *Syst Biol, 59*(3), 307-321. doi: 10.1093/sysbio/syq010

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., & Luo, J. (2005). DATF: a database of Arabidopsis transcription factors. *Bioinformatics, 21*(10), 2568-2569. doi: 10.1093/bioinformatics/bti334

Guo, A. Y., Chen, X., Gao, G., Zhang, H., Zhu, Q. H., Liu, X. C., . . . Luo, J. (2008). PlantTFDB: a comprehensive plant transcription factor database. *Nucleic Acids Res, 36*(Database issue), D966-969. doi: 10.1093/nar/gkm841

Gutterson, N., & Reuber, T. L. (2004). Regulation of disease resistance pathways by AP2/ERF transcription factors. *Curr Opin Plant Biol, 7*(4), 465.

Hall, A., Ismail, A., Ehlers, J., Marfo, K., Cisse, N., Thiaw, S., & Close, T. (2002). Breeding cowpea for tolerance to temperature extremes and adaptation to drought. *Challenges and Opportunities for Enhancing Sustainable Cowpea Production*, 14-21.

Hall, A. E., Cisse, N., Thiaw, S., Elawad, H. O., Ehlers, J. D., Ismail, A. M., . . . Murdock, L. L. (2003). Development of cowpea cultivars and germplasm by the Bean/Cowpea CRSP. *Field Crops Research, 82*(2), 103-134.

Hall, A. E. (2004). Breeding for adaptation to drought and heat in cowpea. *European Journal of Agronomy, 21*(4), 447-454.

Hiraoka, Y., & Sugimoto, Y. (2008). Molecular responses of sorghum to purple witchweed (Striga hermonthica) parasitism. *Weed Science, 56*(3), 356-363. doi: Doi 10.1614/Ws-07-136.1

Hu, H., Dai, M., Yao, J., Xiao, B., Li, X., Zhang, Q., & Xiong, L. (2006). Overexpressing a NAM, ATAF, and CUC (NAC) transcription factor enhances drought resistance and salt tolerance in rice. *Proc Natl Acad Sci U S A, 103*(35), 12987-12992. doi: 10.1073/pnas.0604882103

Huang, K., Mellor, K. E., Paul, S. N., Lawson, M. J., Mackey, A. J., & Timko, M. P. (2012). Global changes in gene expression during compatible and incompatible interactions of cowpea (Vigna unguiculata L.) with the root parasitic angiosperm Striga gesnerioides. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *BMC Genomics, 13*(1), 402. doi: 10.1186/1471-2164-13-402

Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P., & Oinn, T. (2006). Taverna: a tool for building and running workflows of services. [Research Support, Non-U.S. Gov't]. *Nucleic Acids Res, 34*(Web Server issue), W729-732. doi: 10.1093/nar/gkl320

Ishiguro, S., & Nakamura, K. (1994). Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato. *Mol Gen Genet, 244*(6), 563-571.

Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O'Shea, C., & Skriver, K. (2010). The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. *Biochemical Journal, 426*(2), 183-196. doi: 10.1042/bj20091234

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS, 8(3), 275-282.

Kagaya, Y., Ohmiya, K., & Hattori, T. (1999). RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants. *Nucleic Acids Res, 27*(2), 470-478. doi: 10.1093/nar/27.2.470

Kalavacharla, V., Liu, Z., Meyers, B. C., Thimmapuram, J., & Melmaiee, K. (2011). Identification and analysis of common bean (Phaseolus vulgaris L.) transcriptomes by massively parallel pyrosequencing. *BMC Plant Biol, 11*, 135. doi: 10.1186/1471-2229-11-135

Kemena, C., & Notredame, C. (2009). Upcoming challenges for multiple sequence alignment methods in the high-throughput era. [Review]. *Bioinformatics, 25*(19), 2455-2465. doi: 10.1093/bioinformatics/btp452

Kubo, M., Udagawa, M., Nishikubo, N., Horiguchi, G., Yamaguchi, M., Ito, J., . . . Demura, T. (2005). Transcription switches for protoxylem and metaxylem vessel formation. *Genes Dev, 19*(16), 1855-1860.

Lai, Z., Li, Y., Wang, F., Cheng, Y., Fan, B., Yu, J.-Q., & Chen, Z. (2011). Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *The Plant Cell Online, 23*(10), 3824-3841.

Lang, D., Weiche, B., Timmerhaus, G., Richardt, S., Riano-Pachon, D. M., Correa, L. G., . . . Rensing, S. A. (2010). Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: a timeline of loss, gain, expansion, and

correlation with complexity. [Comparative Study Research Support, Non-U.S. Gov't]. *Genome Biol Evol, 2*, 488-503. doi: 10.1093/gbe/evq032

Langyintuo, A., Lowenberg-DeBoer, J., Faye, M., Lambert, D., Ibro, G., Moussa, B., . . . Ntoukam, G. (2003). Cowpea supply and demand in West and Central Africa. *Field Crops Research, 82*(2), 215-231.

Lavin, M., Herendeen, P. S., & Wojciechowski, M. F. (2005). Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Syst Biol, 54*(4), 575-594.

Lee, S.-j., Park, J., Lee, M., Yu, J.-h., & Kim, S. (2010). Isolation and functional characterization of CE1 binding proteins. *BMC Plant Biol, 10*(1), 277.

Li, C. W., Su, R. C., Cheng, C. P., Sanjaya, You, S. J., Hsieh, T. H., . . . Chan, M. T. (2011). Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. *Plant Physiol, 156*(1), 213-227. doi: 10.1104/pp.111.174268

Li, H.-L., Zhang, L.-B., Guo, D., Li, C.-Z., & Peng, S.-Q. (2012). Identification and expression profiles of the WRKY transcription factor family in Ricinus communis. *Gene, 503*(2), 248-253. doi: http://dx.doi.org/10.1016/j.gene.2012.04.069

Lin, R.-C., Park, H.-J., & Wang, H.-Y. (2008). Role of Arabidopsis RAP2.4 in Regulating Light- and Ethylene-Mediated Developmental Processes and Drought Stress Tolerance. *Molecular Plant, 1*(1), 42-57. doi: 10.1093/mp/ssm004

MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends in Genetics, 27*(4), 141-148. doi: http://dx.doi.org/10.1016/j.tig.2011.01.001

McClean, P. E., Lavin, M., Gepts, P., & Jackson, S. A. (2008). Phaseolus vulgaris: a diploid model for soybean *Genetics and Genomics of Soybean* (pp. 55-76): Springer.

McClean, P., Mamidi, S., McConnell, M., Chikara, S., & Lee, R. (2010). Synteny mapping between common bean and soybean reveals extensive blocks of shared loci. *BMC Genomics, 11*(1), 184.

Missier, P., Soiland-Reyes, S., Owen, S., Tan, W., Nenadic, A., Dunlop, I., ... & Goble, C. (2010, January). Taverna, reloaded. In *Scientific and Statistical Database Management* (pp. 471-481). Springer Berlin Heidelberg.

Mitsuda, N., Seki, M., Shinozaki, K., & Ohme-Takagi, M. (2005). The NAC transcription factors NST1 and NST2 of Arabidopsis regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell, 17*(11), 2993-3006. doi: 10.1105/tpc.105.036004

Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K., & Tran, L. S. P. (2010). LegumeTFDB: an integrative database of Glycine max, Lotus japonicus and Medicago truncatula transcription factors. [Article]. *Bioinformatics, 26*(2), 290-291. doi: 10.1093/bioinformatics/btp645

Muchero, W., Diop, N. N., Bhat, P. R., Fenton, R. D., Wanamaker, S., Pottorff, M., . . . Close, T. J. (2009). A consensus genetic map of cowpea [Vigna unguiculata (L)

Walp.] and synteny based on EST-derived SNPs. *Proc Natl Acad Sci U S A, 106*(43), 18159-18164. doi: 10.1073/pnas.0905886106

Muller, T., & Vingron, M. (2000). Modeling amino acid replacement. J Comput Biol, 7(6), 761-776. doi: 10.1089/10665270050514918

Nakano, T., Suzuki, K., Fujimura, T., & Shinshi, H. (2006). Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice. *Plant Physiol, 140*(2), 411-432. doi: 10.1104/pp.105.073783

Nielsen, S. S., Brandt, W. E., & Singh, B. B. (1993). Genetic-Variability for Nutritional Composition and Cooking Time of Improved Cowpea Lines. *Crop Science, 33*(3), 469-472. doi: 10.2135/cropsci1993.0011183X003300030010x

Nielsen, S., Ohler, T., & Mitchell, C. (1997). Cowpea leaves for human consumption: production, utilization, and nutrient composition. *Advances in cowpea research, 326*.

Notredame, C. (2007). Recent evolutions of multiple sequence alignment algorithms. [Research Support, Non-U.S. Gov't Review]. *PLoS Comput Biol, 3*(8), e123. doi: 10.1371/journal.pcbi.0030123

Notredame, C. (2008) T-Coffee: Cheat Sheet Tutorial and FAQ. Accessed from http://www.tcoffee.org/Documentation/t_coffee/t_coffee_tutorial.htm on June 22, 2013.

Nuruzzaman, M., Sharoni, A. M., Satoh, K., Moumeni, A., Venuprasad, R., Serraj, R., . . . Kikuchi, S. (2012). Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from

crossing Aday Selection (drought tolerant) and IR64. *Mol Genet Genomics, 287*(5), 389-410. doi: 10.1007/s00438-012-0686-8

Ohme-Takagi, M., & Shinshi, H. (1995). Ethylene-inducible DNA binding proteins that interact with an ethylene-responsive element. *The Plant Cell Online, 7*(2), 173-182.

Oinn, T., Greenwood, M., Addis, M., Alpdemir, M. N., Ferris, J., Glover, K., ... & Wroe, C. (2006). Taverna: lessons in creating a workflow environment for the life sciences. *Concurrency and Computation: Practice and Experience*, *18*(10), 1067-1100.

Olsen, A. N., Ernst, H. A., Leggio, L. L., & Skriver, K. (2005). NAC transcription factors: structurally distinct, functionally diverse. *Trends Plant Sci, 10*(2), 79-87. doi: 10.1016/j.tplants.2004.12.010

Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., . . . Kikuchi, S. (2003). Comprehensive Analysis of NAC Family Genes in Oryza sativa and Arabidopsis thaliana. *DNA research, 10*(6), 239-247. doi: 10.1093/dnares/10.6.239

Oseni, T. O. (2010). Evaluation of sorghum-cowpea intercrop productivity in savanna agro-ecology using competition indices. *Journal of Agricultural Science, 2*(3), P229.

Pandey, S. P., & Somssich, I. E. (2009). The role of WRKY transcription factors in plant immunity. *Plant Physiol, 150*(4), 1648-1655. doi: 10.1104/pp.109.138990

Pearson, W. R., Wood, T., Zhang, Z., & Miller, W. (1997). Comparison of DNA

sequences with protein sequences. *Genomics, 46*(1), 24-36.

Perez-Rodriguez, P., Riano-Pachon, D. M., Correa, L. G., Rensing, S. A., Kersten, B., & Mueller-Roeber, B. (2010). PlnTFDB: updated content and new features of the plant transcription factor database. *Nucleic Acids Res, 38*(Database issue), D822-827. doi: 10.1093/nar/gkp805

Pfam (2013). Accessed from http://pfam.sanger.ac.uk/ on June 22, 2013.

Phytozome (2013). Accessed from http://www.phytozome.net/ on June 22, 2013.

Pinheiro, G. L., Marques, C. S., Costa, M. D. B. L., Reis, P. A. B., Alves, M. S., Carvalho, C. M., . . . Fontes, E. P. B. (2009). Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. *Gene, 444*(1–2), 10-23. doi: http://dx.doi.org/10.1016/j.gene.2009.05.012

Pottorff, M., Ehlers, J. D., Fatokun, C., Roberts, P. A., & Close, T. J. (2012). Leaf morphology in Cowpea [Vigna unguiculata (L.) Walp]: QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. *BMC Genomics, 13*, 234. doi: 10.1186/1471-2164-13-234

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. [Research Support, Non-U.S. Gov't]. *Nucleic Acids Res, 40*(Database issue), D290-301. doi: 10.1093/nar/gkr1065

Puranik, S., Sahu, P. P., Srivastava, P. S., & Prasad, M. (2012). NAC proteins: regulation and role in stress tolerance. [Research Support, Non-U.S. Gov't Review]. *Trends Plant Sci, 17*(6), 369-381. doi: 10.1016/j.tplants.2012.02.004

Qiu, D., Xiao, J., Ding, X., Xiong, M., Cai, M., Cao, Y., . . . Wang, S. (2007). OsWRKY13 mediates rice disease resistance by regulating defense-related genes in salicylate- and jasmonate-dependent signaling. *Mol Plant Microbe Interact, 20*(5), 492-499. doi: 10.1094/mpmi-20-5-0492

Qiu, D., Xiao, J., Xie, W., Cheng, H., Li, X., & Wang, S. (2009). Exploring transcriptional signalling mediated by OsWRKY13, a potential regulator of multiple physiological processes in rice. *BMC Plant Biol, 9*, 74. doi: 10.1186/1471-2229-9-74

Rambaut, A. FigTree. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/ on June 22, 2013.

Reiss, G., & Bailey, J. (1998). Striga gesnerioidesParasitising Cowpea: Development of Infection Structures and Mechanisms of Penetration. *Ann Bot, 81*(3), 431-440.

Riano-Pachon, D. M., Ruzicic, S., Dreyer, I., & Mueller-Roeber, B. (2007). PlnTFDB: an integrative plant transcription factor database. *BMC Bioinformatics, 8*, 42. doi: 10.1186/1471-2105-8-42

Richardt, S., Lang, D., Reski, R., Frank, W., & Rensing, S. A. (2007). PlanTAPDB, a phylogeny-based resource of plant transcription-associated proteins. [Research Support, Non-U.S. Gov't]. *Plant Physiol, 143*(4), 1452-1466. doi: 10.1104/pp.107.095760

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., . . . Yu, G. (2000). Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science, 290*(5499), 2105-2110.

Riechmann, J. L., & Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. Review]. *Biol Chem, 379*(6), 633-646.

Roberts, P., Matthews, W., & Ehlers, J. (1996). New resistance to virulent root-knot nematodes linked to the Rk locus of cowpea. *Crop Science, 36*(4), 889-894.

Roberts, P., Ehlers, J., Hall, A., & Matthews, W. (1997). Characterization of new resistance to root-knot nematodes in cowpea. *Advances in cowpea research. Edited by BB Singh, DR Mohan Raj, KE Dashiell, and LEN Jackai. IITA Ibadan, Nigeria*, 207-214.

Rolland, F., Baena-Gonzalez, E., & Sheen, J. (2006). Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol, 57*, 675-709. doi: 10.1146/annurev.arplant.57.032905.105441

Rushton, P. J., Macdonald, H., Huttly, A. K., Lazarus, C. M., & Hooley, R. (1995). Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of alpha-Amy2 genes. [Research Support, Non-U.S. Gov't]. *Plant Mol Biol, 29*(4), 691-702.

Rushton, P. J., Bokowiec, M. T., Han, S., Zhang, H., Brannock, J. F., Chen, X., . . . Timko, M. P. (2008). Tobacco transcription factors: novel insights into transcriptional regulation in the Solanaceae. [Research Support, Non-U.S. Gov't]. *Plant Physiol, 147*(1), 280-295. doi: 10.1104/pp.107.114041

Rushton, P. J., Somssich, I. E., Ringler, P., & Shen, Q. J. (2010). WRKY transcription factors. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-

P.H.S. Review]. *Trends Plant Sci, 15*(5), 247-258. doi: 10.1016/j.tplants.2010.02.006

Rushton, P. J. (2008). TOBFAC: Family AP2. Retrieved from http://compsysbio.achs.virginia.edu/tobfac/browse_family.pl?family=AP2 on June 22, 2013.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol, 4*(4), 406-425.

Salamov, A. A., & Solovyev, V. V. (2000). Ab initio gene finding in Drosophila genomic DNA. *Genome Research, 10*(4), 516-522.

Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., . . . Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature, 463*(7278), 178-183. doi: 10.1038/nature08670

Sharoni, A. M., Nuruzzaman, M., Satoh, K., Shimizu, T., Kondoh, H., Sasaya, T., . . . Kikuchi, S. (2011). Gene structures, classification and expression models of the AP2/EREBP transcription factor family in rice. *Plant Cell Physiol, 52*(2), 344-360. doi: 10.1093/pcp/pcq196

Shin, R., Park, J. M., An, J.-M., & Paek, K.-H. (2002). Ectopic expression of Tsi1 in transgenic hot pepper plants enhances host resistance to viral, bacterial, and oomycete pathogens. *Molecular Plant-Microbe Interactions, 15*(10), 983-989.

Shore, P., & Sharrocks, A. D. (1995). The MADS-Box Family of Transcription Factors. *European Journal of Biochemistry, 229*(1), 1-13. doi: 10.1111/j.1432-1033.1995.0001l.x

Singh, B. (2005). Cowpea [Vigna unguiculata (L.) Walp]. *Genetic resources, chromosome engineering, and crop improvement: Grain legumes, 1*, 117-162.

Singh, K., Foley, R. C., & Onate-Sanchez, L. (2002). Transcription factors in plant defense and stress responses. [Research Support, Non-U.S. Gov't Review]. *Curr Opin Plant Biol, 5*(5), 430-436. doi: 10.1016/s1369-5266(02)00289-3

Singh, B., Ajeigbe, H., Tarawali, S. A., Fernandez-Rivera, S., & Abubakar, M. (2003). Improving the production and utilization of cowpea as food and fodder. *Field Crops Research, 84*(1), 169-177.

Sleator, R. D. (2011). Phylogenetics. [Review]. *Arch Microbiol, 193*(4), 235-239. doi: 10.1007/s00203-011-0677-x

Smith, A.F.A., Hubley, R. & Green, P. (2010) RepeatMasker Open-3.0. Retrieved from http://www.repeatmasker.org on October 31, 2013.

Sonnhammer, E. L., Eddy, S. R., & Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins, 28*(3), 405-420.

Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res, 33*(suppl 2), W465-W467.

Stanke, M., Steinkamp, R., Waack, S., & Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res, 32*(suppl 2), W309-W312.

Sun, S., Yu, J.-P., Chen, F., Zhao, T.-J., Fang, X.-H., Li, Y.-Q., & Sui, S.-F. (2008). TINY, a dehydration-responsive element (DRE)-binding protein-like transcription

factor connecting the DRE-and ethylene-responsive element-mediated signaling pathways in Arabidopsis. *Journal of Biological Chemistry, 283*(10), 6261-6271.

Takada, S., Hibara, K., Ishida, T., & Tasaka, M. (2001). The CUP-SHAPED COTYLEDON1 gene of Arabidopsis regulates shoot apical meristem formation. *Development, 128*(7), 1127-1135.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol, 28*(10), 2731-2739.

Tempel, S. (2012). Using and Understanding RepeatMasker. In Y. Bigot (Ed.), *Mobile Genetic Elements* (Vol. 859, pp. 29-51): Humana Press.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res, 22*(22), 4673-4680.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res, 25*(24), 4876-4882.

Timko, M. P., Ehlers, J. D., & Roberts, P. A. (2007). Cowpea *Pulses, sugar and tuber crops* (pp. 49-67): Springer.

Timko, M. P., Rushton, P. J., Laudeman, T. W., Bokowiec, M. T., Chipumuro, E., Cheung, F., . . . Chen, X. (2008). Sequencing and analysis of the gene-rich space of cowpea. *BMC Genomics, 9*, 103. doi: 10.1186/1471-2164-9-103

Timko, M. P., & Singh, B. (2008). Cowpea, a multifunctional legume. *Genomics of tropical crop plants*, 227-258.

Timko, M. P., Huang, K., & Lis, K. E. (2012). Host Resistance and Parasite Virulence in Striga–Host Plant Interactions: A Shifting Balance of Power. *Weed Science, 60*(2), 307-315. doi: 10.1614/ws-d-11-00039.1

Udvardi, M. K., Kakar, K., Wandrey, M., Montanari, O., Murray, J., Andriankaja, A., . . . Town, C. D. (2007). Legume transcription factors: global regulators of plant development and response to the environment. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review]. *Plant Physiol, 144*(2), 538-549. doi: 10.1104/pp.107.098061

Van Verk, M. C., Gatz, C., & Linthorst, H. J. M. (2009). Transcriptional Regulation of Plant Defense Responses. In L. C. VanLoon (Ed.), *Plant Innate Immunity* (Vol. 51, pp. 397-438). London: Academic Press Ltd-Elsevier Science Ltd.

Wanamaker, S. & Close, T. J. (2011) HarvEST Blast Server (http://www.harvest-blast.org/). Accessed on June 22, 2013.

Wang, Z., Libault, M., Joshi, T., Valliyodan, B., Nguyen, H. T., Xu, D., . . . Cheng, J. L. (2010). SoyDB: a knowledge database of soybean transcription factors. [Article]. *BMC Plant Biol, 10*. doi: 10.1186/1471-2229-10-14

Westwood, J. H., Roney, J. K., Khatibi, P. A., & Stromberg, V. K. (2009). RNA translocation between parasitic plants and their hosts. *Pest management*

science, 65(5), 533-539. doi: 10.1002/ps.1727

Wu, C., Sun, S., Nimmakayala, P., Santos, F. A., Meksem, K., Springman, R., . . . Zhang, H.-B. (2004). A BAC- and BIBAC-Based Physical Map of the Soybean Genome. *Genome Research, 14*(2), 319-326. doi: 10.1101/gr.1405004

Xie, Q., Frugis, G., Colgan, D., & Chua, N. H. (2000). Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Genes Dev, 14*(23), 3024-3036.

Xie, Q., Sanz-Burgos, A. P., Guo, H., Garcia, J. A., & Gutierrez, C. (1999). GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. *Plant Mol Biol, 39*(4), 647-656.

Xu, Z. S., Chen, M., Li, L. C., & Ma, Y. Z. (2011). Functions and Application of the AP2/ERF Transcription Factor Family in Crop Improvement. [Review]. *Journal of Integrative Plant Biology, 53*(7), 570-585. doi: 10.1111/j.1744-7909.2011.01062.x

Yandell, M. D. RepeatRunner. Retrieved from http://www.yandell-lab.org/software/repeatrunner.html on October 31, 2013.

Yeh, R.-F., Lim, L. P. and Burge, C. B. (2001) Computational inference of homologous gene structures in the human genome. *Genome Res.* **11**: 803-816.

Yoshii, M., Yamazaki, M., Rakwal, R., Kishi-Kaboshi, M., Miyao, A., & Hirochika, H. (2010). The NAC transcription factor RIM1 of rice is a new regulator of jasmonate signaling. *Plant J, 61*(5), 804-815. doi: 10.1111/j.1365-313X.2009.04107.x

Zarei, A., Körbes, A. P., Younessi, P., Montiel, G., Champion, A., & Memelink, J. (2011). Two GCC boxes and AP2/ERF-domain transcription factor ORA59 in jasmonate/ethylene-mediated activation of the PDF1. 2 promoter in Arabidopsis. *Plant Mol Biol, 75*(4-5), 321-331.

Zhang, G., Chen, M., Chen, X., Xu, Z., Guan, S., Li, L.-C., . . . Ma, Y. (2008). Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (Glycine max L.). *J Exp Bot, 59*(15), 4095-4107. doi: 10.1093/jxb/ern248

Zhang, H., Jin, J., Tang, L., Zhao, Y., Gu, X., Gao, G., & Luo, J. (2011). PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. [Research Support, Non-U.S. Gov't]. *Nucleic Acids Res, 39*(Database issue), D1114-1117. doi: 10.1093/nar/gkq1141

Zhao, C., Avci, U., Grant, E. H., Haigler, C. H., & Beers, E. P. (2008). XND1, a member of the NAC domain family in Arabidopsis thaliana, negatively regulates lignocellulose synthesis and programmed cell death in xylem. *Plant J, 53*(3), 425-436. doi: 10.1111/j.1365-313X.2007.03350.x

Zheng, X., Chen, B., Lu, G., & Han, B. (2009). Overexpression of a NAC transcription factor enhances rice drought and salt tolerance. *Biochem Biophys Res Commun, 379*(4), 985-989. doi: 10.1016/j.bbrc.2008.12.163

Zhu, Q.-H., Guo, A.-Y., Gao, G., Zhong, Y.-F., Xu, M., Huang, M., & Luo, J. (2007). DPTF: a database of poplar transcription factors. *Bioinformatics, 23*(10), 1307-1308.

Zhu, T., Nevo, E., Sun, D., & Peng, J. (2012). Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. [Research Support, Non-U.S. Gov't]. *Evolution, 66*(6), 1833-1848. doi: 10.1111/j.1558-5646.2011.01553.x

Zhuang, J., Deng, D.-X., Yao, Q.-H., Zhang, J., Xiong, F., Chen, J.-M., & Xiong, A.-S. (2010). Discovery, phylogeny and expression patterns of AP2-like genes in maize. *Plant Growth Regulation, 62*(1), 51-58. doi: 10.1007/s10725-010-9484-7

Chapter four

Genome-Wide Identification and Expression Analysis of SWEETs in Root-Parasitic Flowering Plants

Abstract

Root-parasitic plants in the *Orobanchaceae* family are among the most destructive causes of yield loss in a variety of crop plants. Parasitic plants attach to host plants using haustorial connections, and then drain the host of much needed nutrients. Among the most important of these nutrients are sugars. Recently, a type of sugar transporter known as SWEETs has been identified as having a role in phloem loading of sugars (Chen et al., 2012). Here, the SWEET repertoires in the genomes of three parasitic weeds, namely the holoparasitic *Orobanche aegyptiaca*, the obligate hemiparasite *Striga hermonthica*, and the facultative hemiparasite *Triphysaria versicolor*, are identified, and their phylogenetic relationships and expression profiles are analyzed. It was found that *Orobanche* SWEET expression is highest during growth and connection phases, whereas *Striga* and *Triphysaria* show the most expression of SWEETs during reproduction. *Triphysaria* SWEET expression is lowest during haustorial connection. Moreover, a phylogenetic study reveals that groups within the SWEET family of transporters are associated with different phases of the parasitic life cycle. This association was somewhat stronger in *Orobanche* and *Striga* than in *Triphysaria*. Together, these results suggest that SWEET expression in *Orobanche*, *Striga* and *Triphysaria* are typical for holoparasites, obligate and facultative hemiparasites, respectively. The possibility of SWEETs being in regulatory networks is discussed. This study thus fills a long-standing void in the literature regarding the molecular mechanisms of parasitism in the *Orobanchaceae*.

Introduction

Root-parasitic flowering plants of the *Orobanchaceae* family are among the most agronomically destructive biotic stresses. For example, *Striga gesnerioides* reduce cowpea crop yields significantly, with losses up to 100% being reported (Cardwell & Lane, 1995). Plants in the *Orobanche* species can cause similar losses in pea yield (Bernhard et al., 1998). Such losses amount to billions of US$ per annum (Spallek et al., 2013; Parker, 2009; Scholes & Press, 2008). Parasitic plants can attack many types of crop plants (Spallek et al., 2013), including wheat (Vasey et al., 2005; Mohamed et al., 2006), sorghum (Hiraoka & Sugimoto, 2008; Estep et al., 2011), rice (Swarbrick et al., 2008), maize (Matvienko et al., 2001), potato (Joel et al., 2007), and cowpea (Timko & Singh, 2008). Yet despite the damage these plants cause, *Orobanchaceae* remain relatively under-studied. It is therefore imperative that researchers study these parasitic plants.

Root-parasitic plant seeds can stay viable for many years (Bebawi et al., 1984). When these seeds germinate, they do so in response to chemical stimulants from host plants (Yoder, 2001; Sun et al., 2007). The parasitic plant then grows haustoria, which it uses to attach to the host root (Kuijt, 1969) and subsequently connect to host vascular tissue (Berner et al., 1997; Sun et al., 2007; Westwood et al., 2012). After that comes a phase in which the weed does the most damage to the host plant (Ejeta, 2007); the weed is able to grow, draining nutrients and water from its host, until they mature and are ready to go to seed (Berner et al., 1997; Sun et al., 2007).

Some of the earliest examples of genetic research on parasitic weeds were the use of random amplified polymorphic DNA (RAPD) analysis to understand the genetics of *Orobanche* plants, and the relationship between parasitic plants of the Orobanche genus (Katzir et al., 1996; Paran et al., 1997; Zeid et al., 1997). Since then, most research involving parasitic plants has revolved around how host plants respond to them (Kusumoto et al., 2007; Swarbrick et al., 2008) and phylogenetic relationships between *Orobanchaceae* plants (Schneeweiss et al., 2004; Weiss-Schneeweiss et al., 2006; Bennett & Mathews, 2006).

Yet few studies have involved the sequencing of these parasitic plants, a task essential in efficiently coming to an understanding of the mechanisms of root-parasitic action. Among the only studies of that kind was the sequencing of parasitic plant expressed sequence tags (ESTs) by Torres et al. (2005) and the sequencing of a full cDNA library and ESTs by Yoshida et al. (2010) for *Striga hermonthica*.

When ESTs alone are sequenced, however, very important aspects of parasitic plant activity may be left unseen. Whole genome assemblies, and resources that allow for analyses based on them, are required to more accurately and comprehensively study such activity. The Parasitic Plant Genome Project (Westwood et al., 2012) provides such resources, using data that is based on whole genome assemblies of several parasitic plants, including *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*; and providing a platform with which to perform searches of any of these genomes.

Despite this, there still remains a void in parasitic plant research. Several studies have been done on the genetic basis of host plant response to parasitic weeds, but

the genetic basis of parasitic plant activity has been seldom studied. The methods of nutrient acquisition by parasitic plants are especially important, and studying such methods can potentially lead to new ways of controlling parasitic plants and preventing them from infecting important crop plants. Since among the most fundamental nutrients are sugars, the study of sugar transporter activity in root-parasitic plants will be essential.

Sugar transporters are used in many processes in the growth and development of plants; they transport sugars from mesophyll cells in leaves to other locations throughout the plant, and have important roles in root elongation, as well as in interactions with symbiotic and pathogenic organisms (Bush, 1999; Williams, 2000; Buttner, 2010; Doidy et al., 2012).

There are three types of sugar transporters: monosaccharide transporters (MSTs) and sucrose transporters (SUTs) (Williams, 2000; Lalonde & Frommer, 2012), which transport unidirectionally (Doidy et al., 2012); and the most recently discovered SWEETs, which were discovered by Chen et al. (2010), and are known for sucrose efflux and for bidirectional transport of sugars (Chen et al., 2012; Doidy et al., 2012). Since SWEETs play important roles in phloem transport (Chen et al., 2012), it is probable that parasitic plants are using sugar transporters as part of a mechanism of draining the host plant of nutrients.

In this study, the SWEET repertoires of three parasitic weeds, *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*, are studied. Their expression throughout the life cycle of each plant studied is investigated, and a phylogenetic

analysis is performed to determine whether different clades of SWEETs are used in different phases of parasitic life cycles.

Methods

Sources of Data

All sequences are from the Parasitic Plant Genome Project (http://ppgp.huck.psu.edu/; Westwood et al., 2012). The parasitic plants investigated were *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*.

The parasitic plant sequences used in this study are based on tissue-specific and whole-plant transcriptomic data created by Westwood et al. (2012). This data was based on cDNA for sequencing made from mRNA as developed by Wickett et al. (2011), which consists of an expressed sequence tag (EST) assembly based on 1.25 million 454 FLX reads and 102.5 million Illumina GAIIx paired reads from different parasitic plants. For each species, the assembly had between 51,000 and 68,000 unigenes at least 200 bp in length with N_{50} values ranging from 663 to 952 (Wickett et al. (2011). PlantTribes (Wall et al., 2008) was used by Wickett et al. (2011) to assign each gene to its respective SuperTribe, Tribe and Orthogroup so as to investigate the overall classification of the transcriptomes studied. Westwood et al. (2012) used Gene Ontology (GO) (Ashburner et al., 2000) to annotate the sequences. Westwood et al. (2012) took two approaches to making the transcriptome assembly from the data produced by Wicket et al. (2011): first, the use of a CLC assembly cell (CLC bio, 10 Rogers St #101, Cambridge, MA 02142) to assemble several libraries in one step and to perform assemblies of one lane of Illumina data (Westwood et al.,

2012). The second approach involved the use of the NextGENe platform (SoftGenetics, LLC., 100 Oakwood Ave, Suite 350, State College, PA 16803) to make raw sequence reads by using a consolidation algorithm several times and then a maximum overlap assembly to assemble the most represented genes. These reads were mapped to assembled contigs, and the matched reads were deleted from the read pool (Westwood et al., 2012). Another consolidation step was done on the unmatched reads to obtain the next most represented genes. Highly represented genes were taken out of the read pool until finally, no contigs above 200 bp remained in the read pool (Westwood et al., 2012). Westwood et al. (2012) used several methods to measure assembly quality, including the capture and coverage of putatively conserved nuclear single copy genes (Duarte et al., 2010).

From this, Westwood et al. (2012) used the PlantTribes 2.0 ten-genome scaffold to identify 285 orthogroups that had one member from each of seven angiosperm genomes. In addition, Westwood et al. (2012) sorted each set of assembled unigenes into each of the 970 putative single-copy genes identified. Finally, Westwood et al. (2012) used GeneWise (GeneWise, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, U.K.) to translate the identified unigenes against reference proteins, align them to the reference gene, create a scaffold sequence, and for the reference gene report the ratio of the number of positions filled to gene length (i.e., coverage).

Westwood et al. (2012) measured the transcripts present at different stages: Stage 0, or germination; stage 1, or root before haustorial growth; stage 2, or root after haustorial growth; stage 3, or haustorial connection before vascular connection to

host; stage 4, or haustorial connection after vascular connection to host; stage 5, or stems and leaves before emergence (5.1) as well as pre-emergent roots (5.2); and finally stage 6, or emerged leaves and stems (and in the case of *Triphysaria versicolor*, roots) as well as reproductive structures.

Identification of SWEETs

Known Arabidopsis and Rice SWEET transporters were found on the NCBI Protein Database and were used as queries to search for potential SWEET transporters in the genomes of *Mimulus guttatus* and parasitic weeds *Orobanche aegyptiaca*, *Striga hermonthica,* and *Triphysaria versicolor.* The *Mimulus guttatus* genome v1.1, which was sequenced by the Willis Laboratory (www.mimulusevolution.org), was retrieved from Phytozome v9.0 (December 13, 2012) (www.phytozome.net; Goodstein et al. (2012)). The TFASTX method, part of the FASTA package (Lipman & Pearson, 1985; Pearson & Lipman, 1988), was used in searching for SWEET sequences. The potential SWEET amino acid sequences were retrieved using custom Perl scripts, and then searched against NCBI using the BLASTP program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Any sequence not identified as a SWEET was discarded. The list of SWEETs from the NCBI search was combined with a list of unigenes identified as SWEETs from a search against Interproscan (Zdobnov & Apweiler, 2001), and then collapsed into a non-redundant list of SWEET sequences. The sequences confirmed as SWEETs were kept for further analysis.

Analysis

Multiple sequence alignment was done using T-Coffee algorithms Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena & Notredame, 2009; Chang et al.,

2012). The alignments from both methods were combined using the T-Coffee Combine algorithm (Di Tommaso et al., 2011) so as to create a consensus between the structure-based alignments of Expresso with the profile-based alignments from PSI-Coffee. Format Converter v2.0.5 at Los Alamos National Laboratory site (http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html) was used to convert FASTA sequence files into PHYLIP interleaved format. The file was then manually adjusted so that the file was in the proper input format for ProtTest 3 (Darriba et al. 2011) and and PhyML 3.0 (Guindon et al., 2010). ProtTest 3 was used to find the best amino acid model for the aligned sequences. Phylogenetic trees were made using PhyML 3.0, using approximate likelihood ratio test (aLRT) (Anisimova & Gascuel, 2006), and automatically selecting the better choice between nearest-neighbor interchange (NNI) and subtree pruning and regraphing (SPR), and automatically selecting the better choice between Chi-square or Shimodaira-Hasegawa (SH)-like branch supports (Guindon et al. 2010). Trees were visualized using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

The expression values for the parasitic weeds upon infecting a host were based on read counts expressed as FPKM (fragments per kilobase of unigene length per million reads), read counts that were based on the results of read mapping for every unigene in the library for each species, after parasitic weeds were allowed to infect host plants; that is, after *Orobanche* infected *Arabidopsis thaliana* and tobacco, after *Striga* infected sorghum, and after *Triphysaria* infected *Medicago truncatula* (Westwood et al., 2012). The expression values for each SWEET sequence were

placed onto an Excel spreadsheet, and then converted into a heat map using Gitools 1.8.4 (Perez-Llamas & Lopez-Bigas, 2011).

The aLRT. The aLRT ensures the accuracy of the tree by comparing the likelihoods of the two optimal arrangements around a branch (Anisimova & Gascuel, 2006; Guindon et al., 2010); in doing so, the aLRT uses a log ratio of the likelihood of the current arrangement to the optimal alternative (Anisimova & Gascuel, 2006; Guindon et al., 2010).

Results

SWEET Repertoire in Parasitic Plants

In *Orobanche aegyptica*, eighteen SWEET transporters were identified, whereas in *Striga hermonthica*, thirty-two were found, and in *Triphysaria versicolor*, twenty-seven SWEET transporters were found. By comparison, a closely related non-parasitic plant organism, namely *Mimulus guttata*, has 38 SWEET transporters. By comparison, there are 21 known SWEETs in rice and 17 in *Arabidopsis thaliana* (Yuan & Wang, 2013).

Expression of SWEET Transporters Throughout the Parasitic Plant Life Cycle

Within each of the parasitic weeds studied, different SWEET transporters were expressed during different phases of the parasitic plant life cycle. In Figures 1-3, the SWEET activity throughout the life cycles of each parasitic plant studied is shown. The names of each stage are listed in Supplementary Table S1, and are described in detail in Westwood et al. (2012). The stages are divided as follows: Stages 0 through 2 were categorized as “initiation phases”, in which the seed germinates and the root and haustoria grow (Westwood et al., 2012). A SWEET transporter was

placed in this category if it was found to be active in stages 1 and/or 2 at minimum. Stages 3 through 4 were categorized as "connection phases", in which a parasitic plant first uses its haustoria to attach to the host plant, and then connect to the host plant vascular system (Westwood et al., 2012). A SWEET transporter was placed in this category if it was found to be active in stages 3 and/or 4. Stages 5.1 through 6.3 were categorized as "growth phases", in which the parasite grows, emerges and reproduces (Westwood et al., 2012). A SWEET transporter was placed in this category if it was found to be active in stages 5 through 6.1 at minimum. Finally, SWEETs that were expressed throughout a majority of life cycle stages was placed in a separate category. If a sequence is identified as being expressed in initiation, connection and growth, but not throughout life, this is due to that sequence having scattered expression. The results of this categorization are shown in Table 1.

Phylogenetic Analysis

To determine whether or not certain types of SWEETs were associated with different stages of the parasitic life cycle, a phylogenetic analysis of the SWEET repertoire for each plant studied was done. These trees are shown in Figures 4-6, and they include the categories of expression for each SWEET sequence. Their closest homologs in *Arabidopsis thaliana* and rice are shown in Supplementary Figures S1-S3. The closest homologs in *Arabidopsis* and rice were found in order to determine the nature of the organization of SWEETs in each parasitic plant. Supplementary Table S1 lists these homologues.

Discussion

In this study, the SWEET family of sugar transporters was identified in three agronomically destructive root-parasitic plants in the *Orobanchaceae* family, namely *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*. The SWEETs of each species were compared with respect to expression throughout the life cycle of each plant, as well as with respect to phylogenetic relationships between SWEETs involved with different phases of the parasitic life cycle.

SWEET Activity and Plant Activity Throughout the Life Cycle

Orobanche aegyptiaca, or Egyptian broomrape, is a holoparasite, completely dependent on its host, whereas *Striga hermonthica* is an obligate hemiparasite and *Triphysaria versicolor* is a facultative hemiparasite (Westwood et al., 2010).

In *Orobanche*, there were fewer SWEETs than in *Striga* and *Triphysaria*. The SWEETs in *Orobanche* were more active in the connection and growth phases than in *Striga* and *Triphysaria*. In *Striga* and *Triphysaria*, on the other hand, SWEETs were most active during growth and reproduction. These results are thus consistent with the finding that SWEETs efflux sucrose out of leaf parenchyma cells into the phloem (Chen et al., 2012), as well as the fact that *Orobanche* is a holoparasite and completely depends on the host rather than relying on photosynthesis (Westwood et al. 2010).

In addition, it was found that unlike in *Orobanche* and *Striga*, SWEET activity in *Triphysaria* was lowest during the connection phases, a finding consistent with the fact that *Triphysaria* is a facultative hemiparasite, and is thus less dependent on the host organism than holoparasites like *Orobanche*, and obligate hemiparasites like *Striga* (Irving & Cameron, 2009).

Phylogenetic Relationships Between SWEETs Involved with Different Stages of the Parasite Plant Life Cycle

In general, when certain SWEETs in the parasitic plants were more active in a certain phase of the life cycle, they were usually phylogenetically close. This was especially the case in *Orobanche* and *Striga*. In *Triphysaria*, on the other hand, there were fewer instances of SWEETs expressed in similar stages being in similar clades. Overall, a comparison with SWEETs from *Arabidopsis thaliana* and rice seems to confirm this finding. In *Orobanche*, sequences homologous with SWEET2 in *Arabidopsis* and rice were expressed throughout most stages of life, homologues to SWEET9 and SWEET10 were normally associated with growth phases, and homologues of SWEET11-15 were associated with both initiation and connection. In *Striga*, SWEET4 homologues were usually either associated with germination or reproduction; SWEET9 homologues were associated with reproduction; and SWEET10 homologoues were expressed throughout most stages of life. Homologues of SWEET15-17 were usually involved with both connection and growth phases, which is similar to the SWEET11-15 homologues in *Orobanche*. In *Triphysaria*, a comparison with *Arabidopsis* and rice shows some instances of *Triphysaria* homologues of different SWEETs being associated with different clades. Similar to *Orobanche*, SWEET2 homologues were expressed throughout most phases of life. SWEET5 homologues were involved with reproduction. Homologues of SWEET6 and SWEET7 were expressed in initiation, connection and growth but their expression patterns were too scattered to be SWEETs with expression throughout the life cycle. The only SWEET9 in *Triphysaria* was only involved in reproduction,

similar to *Orobanche* and *Striga*. Homologues of SWEET15 in *Triphysaria* were less likely to be expressed throughout life than in *Orobanche* or *Striga*. SWEET16 homologues were mostly inactive. Different SWEET1 homologues were expressed in different phases of life. This confirms the observation in this study that the association of homologues of different SWEETs with different stages of life is less prominent in *Triphysaria* than in *Orobanche* and *Striga*.

Host plants are known to use RNA interference (RNAi) to defend themselves against parasitic weeds (Westwood et al., 2009; Yoder et al., 2009); it is therefore theoretically possible that RNAi can be used to stop SWEET activity (among other mechanisms) during the connection phases, before parasitic plants can drain their hosts. The use of RNAi to stop such activity could therefore be a potentially useful avenue of research.

It was also observed that some SWEETs in *Arabidopsis* and rice had no homologues in parasitic weeds. For example, SWEET3 homologues were missing in *Orobanche* and *Striga*, while SWEET8 homologues were missing in *Orobanche* and *Triphysaria* and SWEET11 homologues were absent in *Striga* and *Triphysaria*. Together, these results suggest that the SWEET family in the *Orobanchaceae* family of parasitic plants is organized differently from other plants.

Conclusions

Together, these results suggest that it is possible that SWEETs in parasitic plants could be part of a regulatory network that determines when and how parasitic weeds undergo different stages of development. In particular, the finding that for the most part, different clades of SWEETs were associated with different phases of

parasitic plant development, suggests that certain TFs, which act in a sequence-specific manner (MacQuarrie et al., 2011), may be part of this regulatory network; this would be consistent with the studies reviewed by Rolland et al. (2006). Thus, the interactions between TFs and sugar transporters such as SWEETs could be an interesting future avenue of research in parasitic plants.

Moreover, it is probable that host organisms are responding with sugar transporters of their own (Doidy et al., 2012). However, SWEETs in particular may not be an integral part of this response. In rice, MtN3 (another name for SWEET) transporters were shown to respond to bacterial infection, but not parasitic plant infection (Narsai et al., 2013). And according to oligonucleotide-based microarray data by Huang et al. (2012), no oligonucleotides annotated as MtN3 in cowpea have been known to significantly change in expression when cowpea shows resistance or susceptibility to *Striga gesnerioides* (Timko Lab, unpublished data).

It is possible, however, that other types of sugar transporters (MSTs and SUTs) may be part of a host response to parasitic plants. According to microarray data by Huang et al. (2012), there are oligonucleotides annotated as sugar transporters whose expression in a compatible interaction with *Striga gesnerioides* is significantly different from their expression in an incompatible interaction (Timko Lab, unpublished data). In addition, it is probable that parasitic plants are also using MSTs and SUTs to drain sugars from their hosts, as MSTs and SUTs are known to express in roots (Doidy et al., 2012). It would therefore be beneficial to study MSTs and SUTs and their role in host-parasite interactions. The information gained from a study of MSTs and SUTs, together with information from studies on SWEETs such

as this one, would therefore provide a more comprehensive and precise view of the interactions between parasitic plants and their hosts.

A) B)

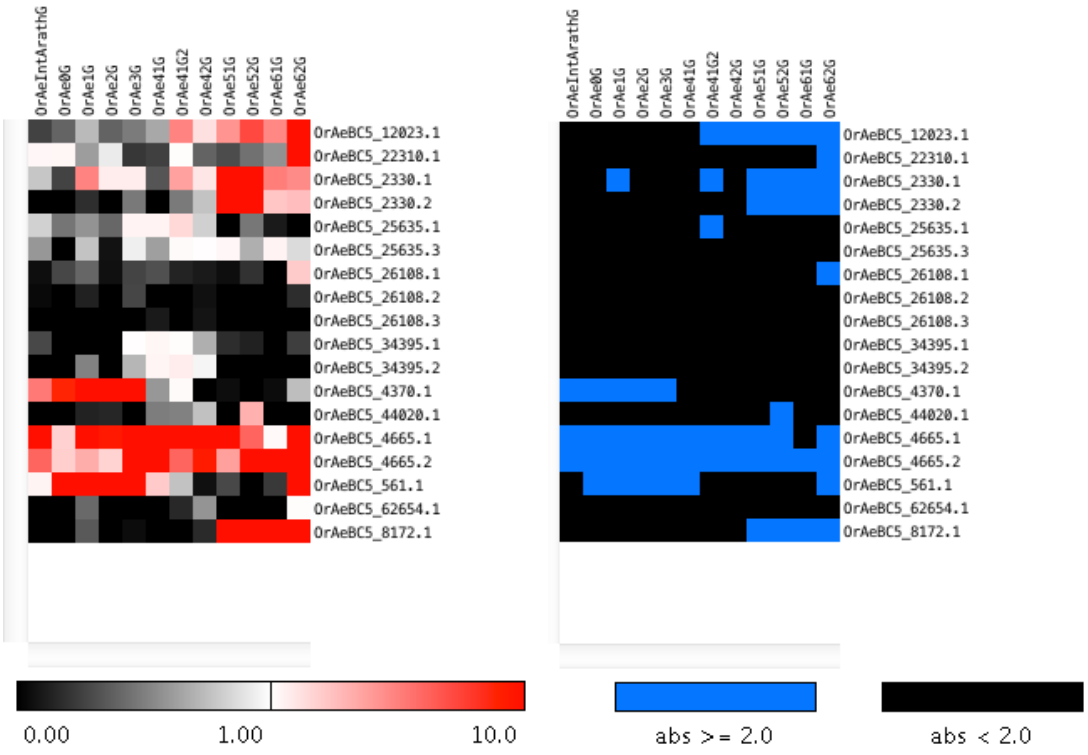


Figure 1. Expression levels of *Orobanche aegyptiaca* SWEET transporters during parasitism of *Arabidopsis thaliana*. (A) Expression values on a linear two-sided scale, with the highest values indicated by a red color (>= 10), neutral values (~1) indicated by white, and low values (<1) indicated by black. (B) Expression values based on significance. An expression value for a sequence at a certain stage was found to be significant if its value was 2.0 or above; here, such values are indicated by a blue color.

A) B)

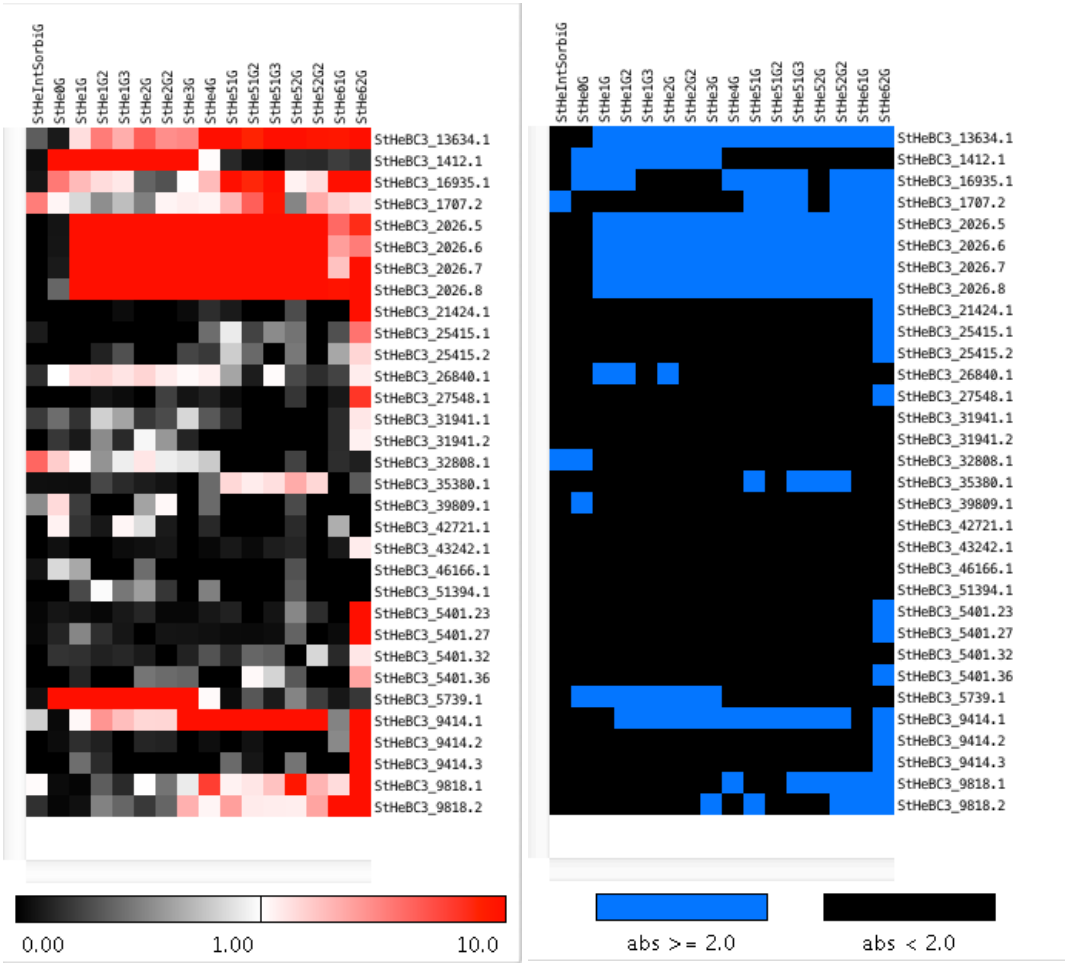


Figure 2. Expression levels of *Striga hermonthica* SWEET transporters during parasitism of *Arabidopsis thaliana*. (A) Expression values on a linear two-sided scale, with the highest values indicated by a red color (>= 10), neutral values (~1) indicated by white, and low values (<1) indicated by black. (B) Expression values based on significance. An expression value for a sequence at a certain stage was found to be significant if its value was 2.0 or above; here, such values are indicated by a blue color.

A) B)

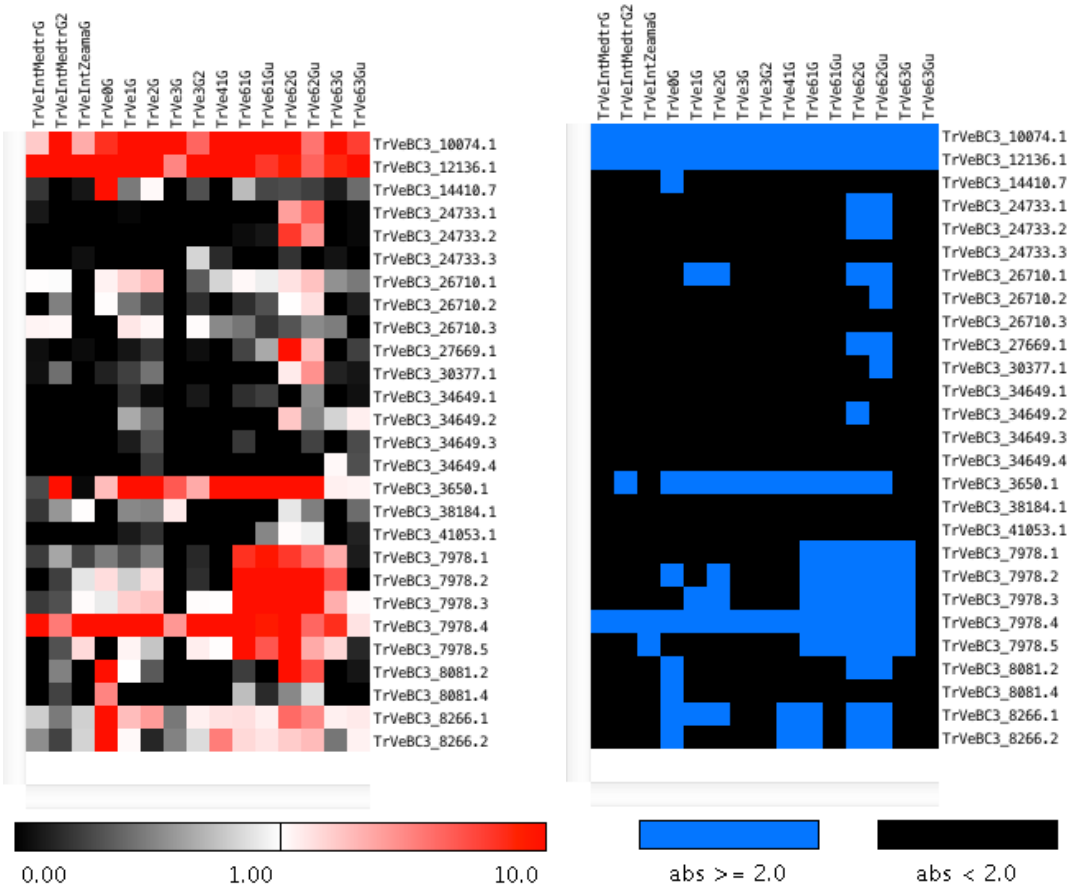


Figure 3. Expression levels of *Triphysaria versicolor* SWEET transporters during parasitism of *Arabidopsis thaliana*. (A) Expression values on a linear two-sided scale, with the highest values indicated by a red color (>= 10), neutral values (~1) indicated by white, and low values (<1) indicated by black. (B) Expression values based on significance. An expression value for a sequence at a certain stage was found to be significant if its value was 2.0 or above; here, such values are indicated by a blue color.

Species	# of SWEETs	Germination Without Other Initiation Phases (Stage 0, not stage 1 or 2)	Initiation (Stages 0-2)	Connection (Stages 3-4)	Growth (Stages 5-6)	Reproduction Without Growth Phases (Stage 6.2 without Stage 5)	Through-out
Orobanche aegyptiaca	18	0	3 / 18 = 16.7%	5 / 18 = 27.7%	5 / 18 = 27.7%	3 / 18 = 16.7%	2/ 18 = 11.1%
Striga hermonthica	32	2 / 32 = 6.3%	4 / 32 = 12.5%	5 / 32 = 15.6%	5 / 32 = 15.6%	9 / 32 = 28.1%	6 / 32 = 18.8%
Triphysaria versicolor	27	4 / 27 = 14.8%	4 / 27 = 14.8%	2/ 27 = 7.4%	6 / 27 = 22.2%	8 / 27 = 29.6%	4 / 27 = 14.8%

Table 1. Expression of SWEETs in *Orobanche*, *Striga*, and *Triphysaria* based on Figures 1B, 2B, and 3B. Note: some SWEETs are placed in multiple categories. Please note: some SWEETs were placed in multiple categories to more accurately describe their activity.

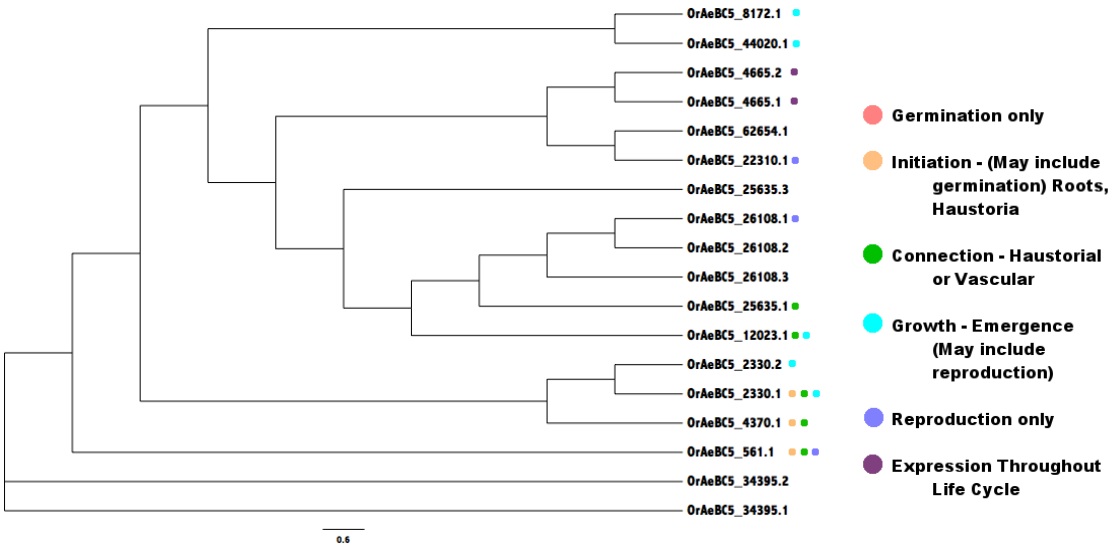


Figure 4. Phylogenetic tree of SWEET transporters from *Orobanche aegyptiaca*, with associated phases. This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the LG amino acid substitution model (Le & Gascuel, 2008), automatically estimated invariable sites, and 5 gamma parameters.

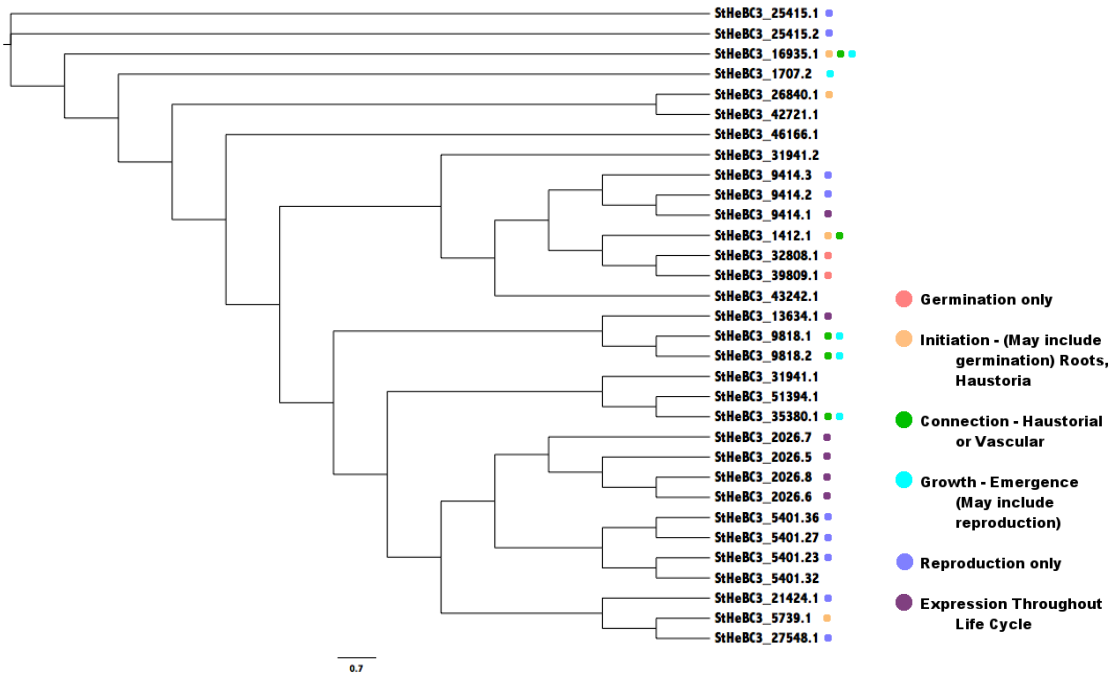


Figure 5. Phylogenetic tree of SWEET transporters from *Striga hermonthica*, with associated phases. This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the LG amino acid substitution model (Le & Gascuel, 2008), automatically estimated invariable sites, and 5 gamma parameters.

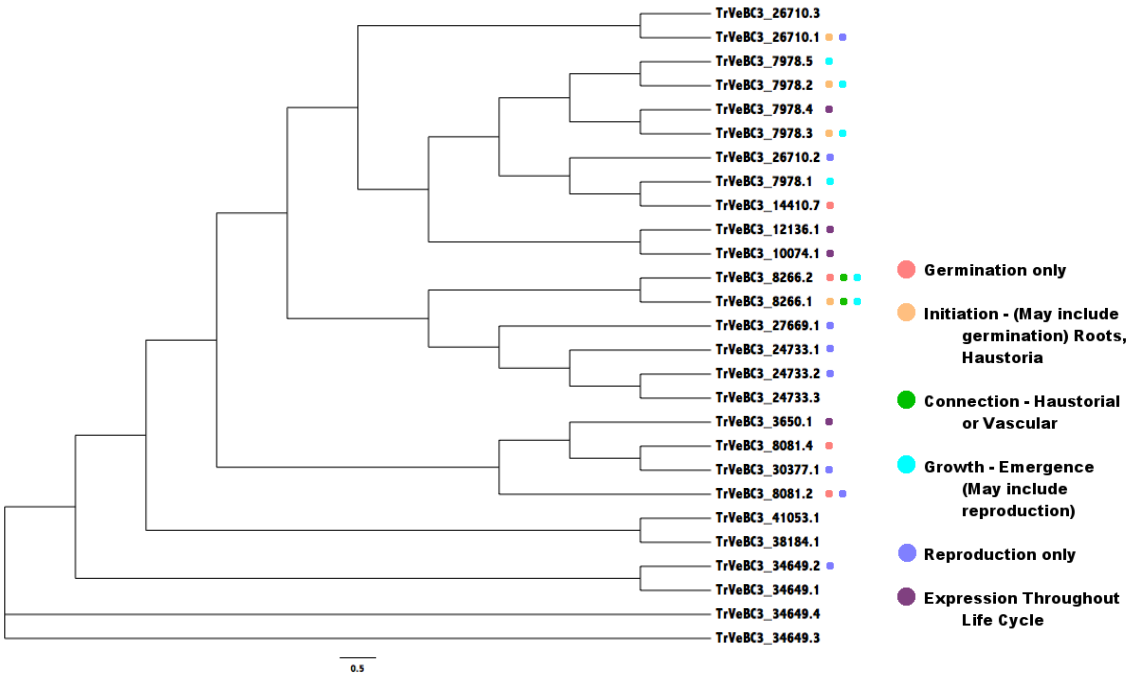
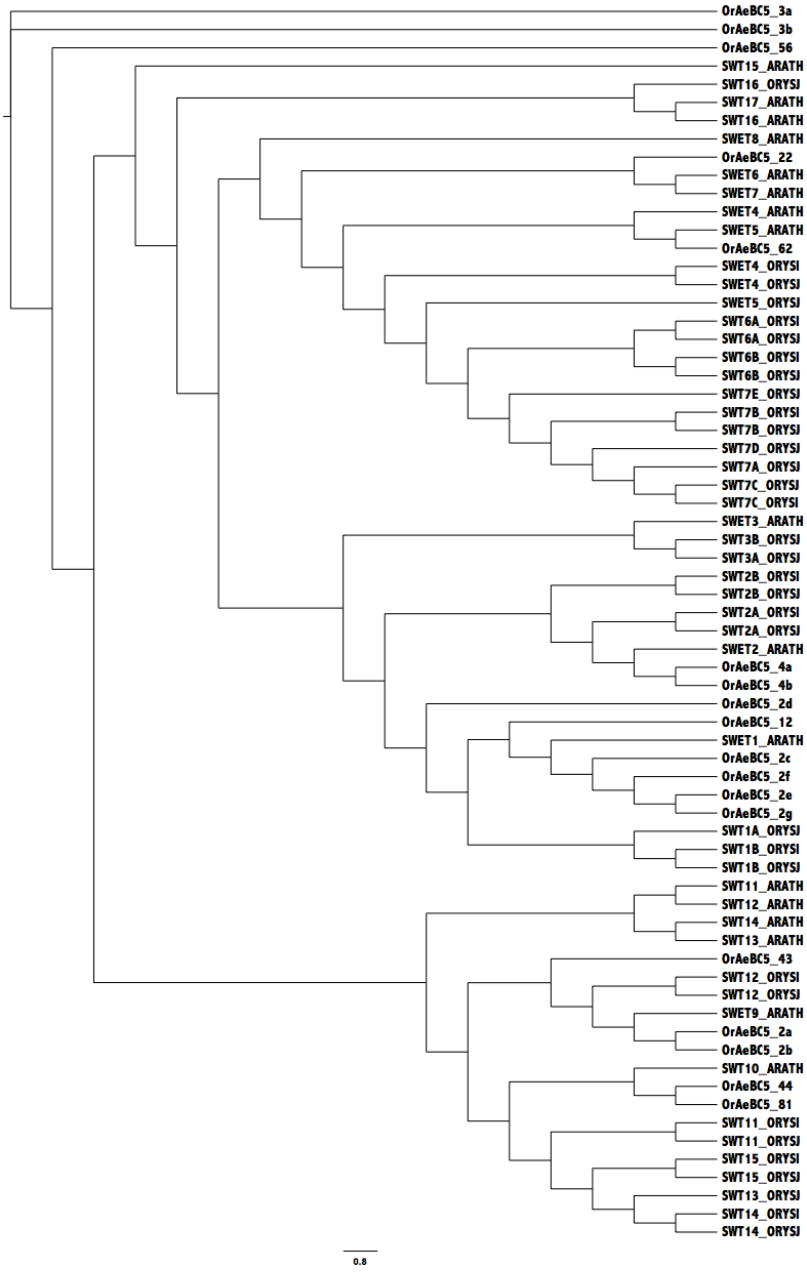


Figure 6. Phylogenetic tree of SWEET transporters from *Triphysaria versicolor*, with associated phases. This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the VT amino acid substitution model (Muller & Vingron, 2000), empirical frequencies, automatically estimated invariable sites, and 5 gamma parameters.

Supplementary Table S1. Arabidopsis and rice SWEETs from NCBI. The names are abbreviated in the trees (e.g., gi|122177696|sp|Q19VE6.1|SWT11_ORYSI is abbreviated to "SWT_ORYSI").

NCBI Accession
gi|122177696|sp|Q19VE6.1|SWT11_ORYSI
gi|75132597|sp|Q6YZF3.1|SWT11_ORYSJ
gi|75206789|sp|Q9SMM5.1|SWT11_ARATH
gi|75100713|sp|O82587.1|SWT12_ARATH
gi|75213043|sp|Q9SW25.1|SWT14_ARATH
gi|75170467|sp|Q9FGQ2.1|SWT13_ARATH
gi|322967576|sp|A2X5B4.1|SWT15_ORYSI
gi|75125443|sp|Q6K602.1|SWT15_ORYSJ
gi|322967574|sp|B8BKP4.1|SWT14_ORYSI
gi|122207452|sp|Q2R3P9.1|SWT14_ORYSJ
gi|122204154|sp|Q2QR07.1|SWT13_ORYSJ
gi|75180553|sp|Q9LUE3.1|SWT10_ARATH
gi|75173209|sp|Q9FY94.1|SWT15_ARATH
gi|322967558|sp|A2XGM7.1|SWT12_ORYSI
gi|122247024|sp|Q10LI8.1|SWT12_ORYSJ
gi|75216881|sp|Q9ZV02.1|SWET9_ARATH
gi|322967649|sp|Q84WN3.2|SWT17_ARATH
gi|75274282|sp|Q9LUR4.1|SWT16_ARATH
gi|122236833|sp|Q10LN5.1|SWT16_ORYSJ
gi|322967646|sp|A2WR31.2|SWT2A_ORYSI
gi|75105779|sp|Q5JJY5.1|SWT2A_ORYSJ
gi|75273203|sp|Q9LH79.1|SWET2_ARATH
gi|322967622|sp|B8A833.1|SWT2B_ORYSI
gi|75103724|sp|Q5N8J1.1|SWT2B_ORYSJ
gi|322967621|sp|B8AYH1.1|SWT1B_ORYSI

gi|75113860|sp|Q60EC2.1|SWT1B_ORYSJ
gi|75159095|sp|Q8RZQ8.1|SWT1A_ORYSJ
gi|75154590|sp|Q8L9J7.1|SWET1_ARATH
gi|322967645|sp|Q5NAZ9.2|SWT3B_ORYSJ
gi|322967644|sp|Q0DJY3.2|SWT3A_ORYSJ
gi|75127190|sp|Q6NQN5.1|SWET3_ARATH
gi|322967651|sp|Q9FM10.2|SWET5_ARATH
gi|75164203|sp|Q944M5.1|SWET4_ARATH
gi|322967626|sp|A2ZIM4.1|SWT7C_ORYSI
gi|122205774|sp|Q2QWX8.1|SWT7C_ORYSJ
gi|322967625|sp|A2YZ24.1|SWT7B_ORYSI
gi|322967642|sp|Q0J349.2|SWT7B_ORYSJ
gi|322967643|sp|Q0J361.2|SWT7A_ORYSJ
gi|322967624|sp|A2WSD3.1|SWT6B_ORYSI
gi|75161759|sp|Q8W0K2.1|SWT6B_ORYSJ
gi|322967623|sp|A2WSD8.1|SWT6A_ORYSI
gi|75157485|sp|Q8LR09.1|SWT6A_ORYSJ
gi|322967641|sp|B9G2E6.2|SWT7D_ORYSJ
gi|322967627|sp|A3BWJ9.1|SWT7E_ORYSJ
gi|75126698|sp|Q6L568.1|SWET5_ORYSJ
gi|322967140|sp|A2X3S3.1|SWET4_ORYSI
gi|75125196|sp|Q6K4V2.1|SWET4_ORYSJ
gi|75169746|sp|Q9C9M9.1|SWET6_ARATH
gi|75154973|sp|Q8LBF7.1|SWET7_ARATH
gi|75155877|sp|Q8LFH5.1|SWET8_ARATH

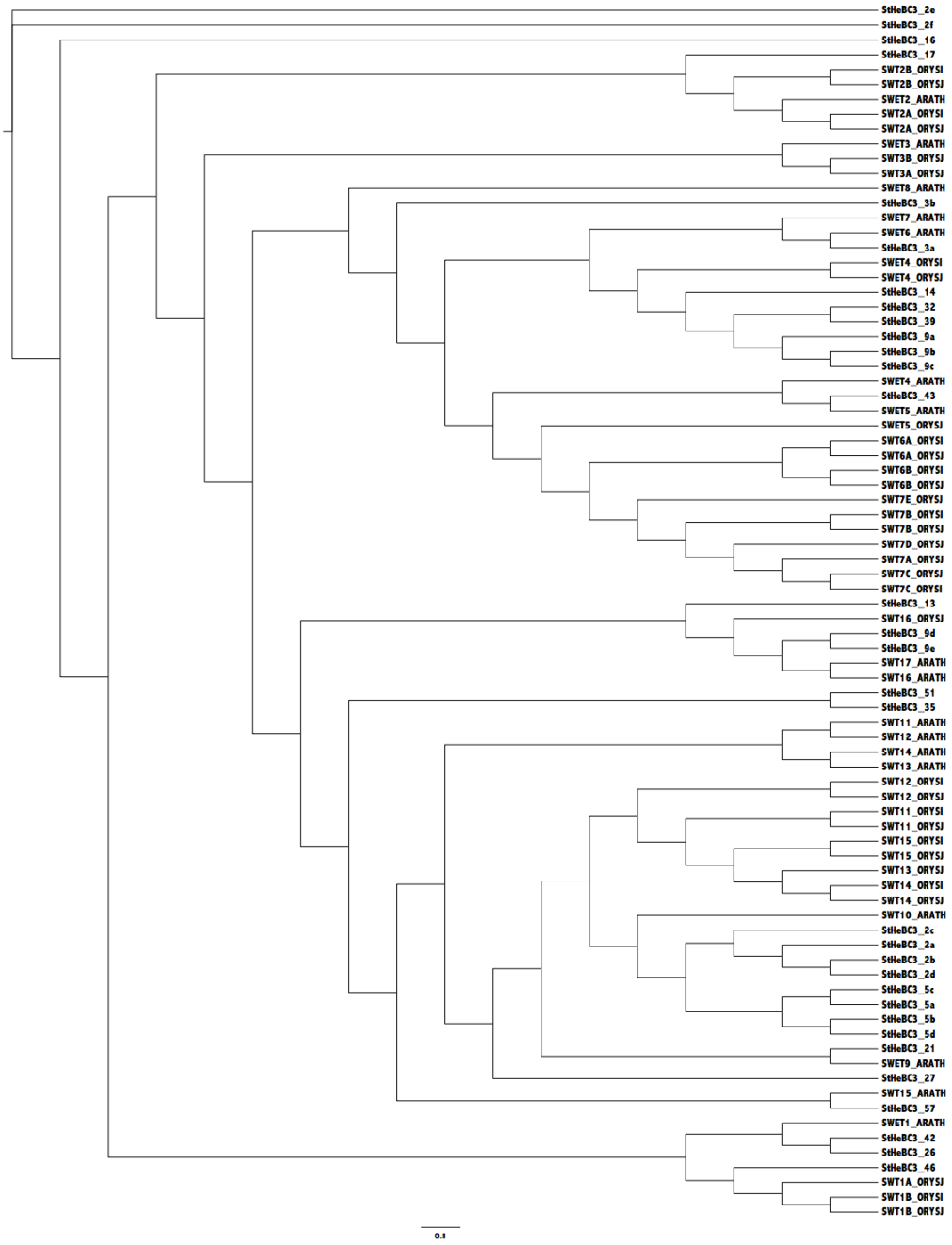


Supplementary Figure S1. Phylogenetic tree of SWEET transporters from *Orobanche aegyptiaca*, with associated *Arabidopsis* and rice SWEETs from NCBI. This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the JTT amino acid substitution model (Jones, Taylor & Thornton, 1992), automatically estimated invariable sites, and 5 gamma

parameters. The cowpea sequence names are abbreviated to PHYLIP format, so their full names are shown in Supplementary Table S1.

Full Name	Tree
OrAeBC5_12023.1	OrAeBC5_12
OrAeBC5_22310.1	OrAeBC5_22
OrAeBC5_2330.1	OrAeBC5_2a
OrAeBC5_2330.2	OrAeBC5_2b
OrAeBC5_25635.1	OrAeBC5_2c
OrAeBC5_25635.3	OrAeBC5_2d
OrAeBC5_26108.1	OrAeBC5_2e
OrAeBC5_26108.2	OrAeBC5_2f
OrAeBC5_26108.3	OrAeBC5_2g
OrAeBC5_34395.1	OrAeBC5_3a
OrAeBC5_34395.2	OrAeBC5_3b
OrAeBC5_4370.1	OrAeBC5_43
OrAeBC5_44020.1	OrAeBC5_44
OrAeBC5_4665.1	OrAeBC5_4a
OrAeBC5_4665.2	OrAeBC5_4b
OrAeBC5_561.1	OrAeBC5_56
OrAeBC5_62654.1	OrAeBC5_62
OrAeBC5_8172.1	OrAeBC5_81

Supplementary Table S2. *Orobanche aegyptiaca* SWEETs and their abbreviated names on the tree in Supplementary Figure S1.

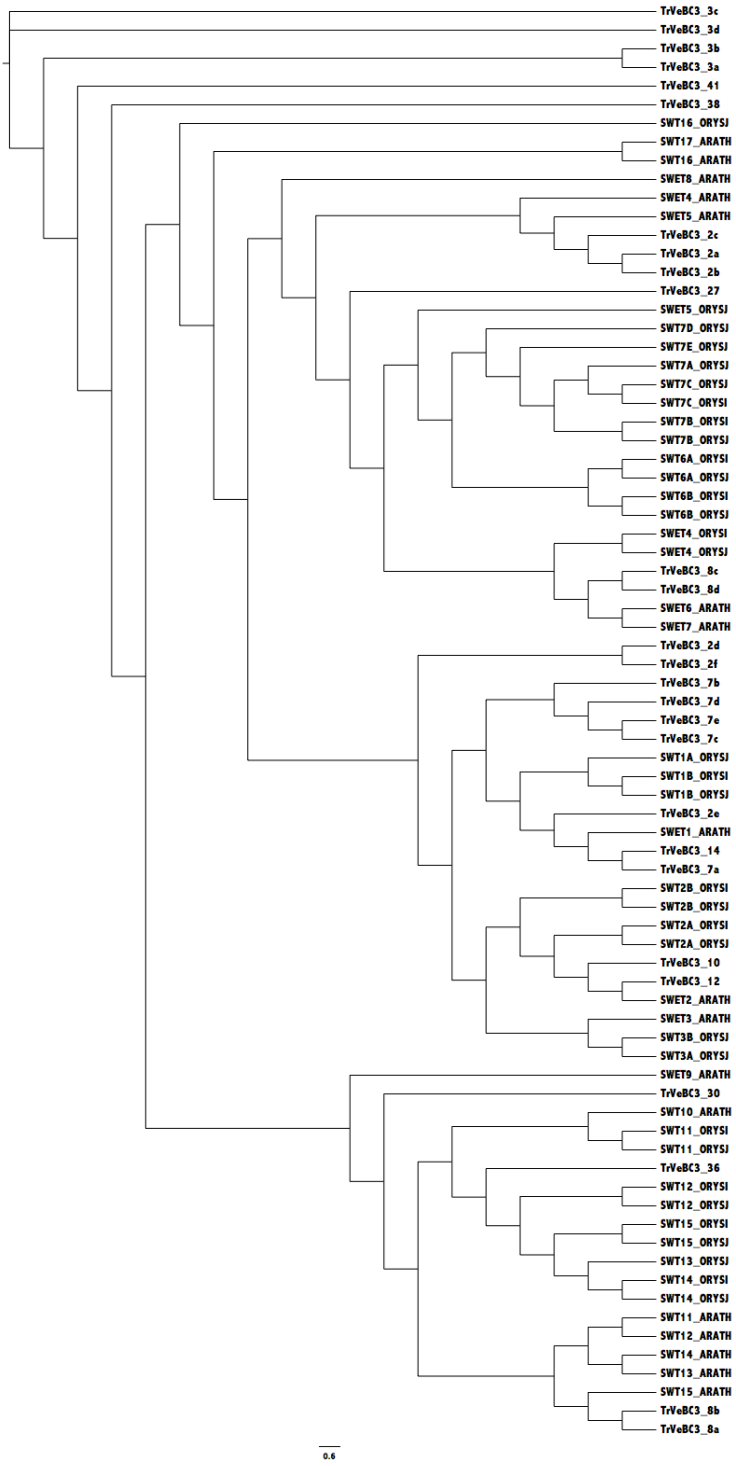


Supplementary Figure S2. Phylogenetic tree of SWEET transporters from *Striga hermonthica*, with associated *Arabidopsis* and rice SWEETs from NCBI. This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the LG amino acid substitution model (Le & Gascuel, 2008), automatically estimated invariable sites, and 5 gamma parameters. The full names

of the cowpea sequences, which are in PHYLIP format, are shown in Supplementary Table S3.

Full Name	Tree Name
StHeBC3_13634.1	StHeBC3_13
StHeBC3_1412.1	StHeBC3_14
StHeBC3_16935.1	StHeBC3_16
StHeBC3_1707.2	StHeBC3_17
StHeBC3_2026.5	StHeBC3_2a
StHeBC3_2026.6	StHeBC3_2b
StHeBC3_2026.7	StHeBC3_2c
StHeBC3_2026.8	StHeBC3_2d
StHeBC3_21424.1	StHeBC3_21
StHeBC3_25415.1	StHeBC3_2e
StHeBC3_25415.2	StHeBC3_2f
StHeBC3_26840.1	StHeBC3_26
StHeBC3_27548.1	StHeBC3_27
StHeBC3_31941.1	StHeBC3_3a
StHeBC3_31941.2	StHeBC3_3b
StHeBC3_32808.1	StHeBC3_32
StHeBC3_35380.1	StHeBC3_35
StHeBC3_39809.1	StHeBC3_39
StHeBC3_42721.1	StHeBC3_42
StHeBC3_43242.1	StHeBC3_43
StHeBC3_46166.1	StHeBC3_46
StHeBC3_51394.1	StHeBC3_51
StHeBC3_5401.23	StHeBC3_5a
StHeBC3_5401.27	StHeBC3_5b
StHeBC3_5401.32	StHeBC3_5c
StHeBC3_5401.36	StHeBC3_5d
StHeBC3_5739.1	StHeBC3_57
StHeBC3_9414.1	StHeBC3_9a
StHeBC3_9414.2	StHeBC3_9b
StHeBC3_9414.3	StHeBC3_9c
StHeBC3_9818.1	StHeBC3_9d
StHeBC3_9818.2	StHeBC3_9e

Supplementary Table S3. *Striga hermonthica* SWEETs, both the full names and the names on the tree in Supplementary Figure S2.



Supplementary Figure S3. Phylogenetic tree of SWEET transporters from *Triphysaria versicolor*, with associated *Arabidopsis* and rice SWEETs from NCBI. This tree was constructed using the maximum likelihood method via PhyML 3

(Guindon et al., 2010) with the VT amino acid substitution model (Muller & Vingron, 2000), automatically estimated invariable sites, and 5 gamma parameters. The full names of the cowpea sequences, which are in PHYLIP format, are shown in Supplementary Table S4.

Full Name	Tree Name
TrVeBC3_10074.1	TrVeBC3_10
TrVeBC3_12136.1	TrVeBC3_12
TrVeBC3_14410.7	TrVeBC3_14
TrVeBC3_24733.1	TrVeBC3_2a
TrVeBC3_24733.2	TrVeBC3_2b
TrVeBC3_24733.3	TrVeBC3_2c
TrVeBC3_26710.1	TrVeBC3_2d
TrVeBC3_26710.2	TrVeBC3_2e
TrVeBC3_26710.3	TrVeBC3_2f
TrVeBC3_27669.1	TrVeBC3_27
TrVeBC3_30377.1	TrVeBC3_30
TrVeBC3_34649.1	TrVeBC3_3a
TrVeBC3_34649.2	TrVeBC3_3b
TrVeBC3_34649.3	TrVeBC3_3c
TrVeBC3_34649.4	TrVeBC3_3d
TrVeBC3_3650.1	TrVeBC3_36
TrVeBC3_38184.1	TrVeBC3_38
TrVeBC3_41053.1	TrVeBC3_41
TrVeBC3_7978.1	TrVeBC3_7a
TrVeBC3_7978.2	TrVeBC3_7b
TrVeBC3_7978.3	TrVeBC3_7c
TrVeBC3_7978.4	TrVeBC3_7d
TrVeBC3_7978.5	TrVeBC3_7e
TrVeBC3_8081.2	TrVeBC3_8a
TrVeBC3_8081.4	TrVeBC3_8b
TrVeBC3_8266.1	TrVeBC3_8c
TrVeBC3_8266.2	TrVeBC3_8d

Supplementary Table S4. Full names of the *Triphysaria versicolor* SWEET sequences on the tree in Supplementary Figure S3.

References

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., . . . Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Research, 34*(suppl 2), W604-W608.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics, 25*(1), 25-29. doi: 10.1038/75556

Bebawi, F. F., Eplee, R. E., Harris, C. E., & Norris, R. S. (1984). Longevity of witchweed (Striga asiatica) seed. *Weed Science*, 494-497.

Bennett, J. R., & Mathews, S. (2006). Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. *Am J Bot, 93*(7), 1039-1051. doi: 10.3732/ajb.93.7.1039

Berner, D., Winslow, M., Awad, A., Cardwell, K., Raj, D. M., & Kim, S. (1997). Striga research methods. *International Institute for Tropical Agriculture, Ibadan*.

Bernhard, Jensen, & Andreasen. (1998). Prediction of yield loss caused by Orobanche spp. in carrot and pea crops based on the soil seedbank. *Weed Research, 38*(3), 191-197. doi: 10.1046/j.1365-3180.1998.00089.x

BLAST: Basic Local Alignment Search Tool. Accessed from http://blast.ncbi.nlm.nih.gov/Blast.cgi on July 20, 2013

Bush, D. R. (1999). Sugar transporters in plant biology. *Current Opinion in Plant Biology, 2*(3), 187-191. doi: http://dx.doi.org/10.1016/S1369-5266(99)80034-X

Buttner, M. (2010). The Arabidopsis sugar transporter (AtSTP) family: an update. [Review]. *Plant Biology, 12*, 35-41. doi: 10.1111/j.1438-8677.2010.00383.x

Cardwell, K. F., & Lane, J. A. (1995). Effect of soils, cropping system and host phenotype on incidence and severity of Striga gesnerioides on cowpea in West Africa. *Agriculture, Ecosystems & Environment, 53*(3), 253-262. doi: http://dx.doi.org/10.1016/0167-8809(94)00575-Y

Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., . . . Frommer, W. B. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Nature, 468*(7323), 527-532. doi: 10.1038/nature09606

Chang, J. M., Di Tommaso, P., Taly, J. F., & Notredame, C. (2012). Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. BMC Bioinformatics, 13 Suppl 4, S1. doi: 10.1186/1471-2105-13-s4-s1

Chen, L. Q., Qu, X. Q., Hou, B. H., Sosso, D., Osorio, S., Fernie, A. R., & Frommer, W. B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Science, 335*(6065), 207-211. doi: 10.1126/science.1213351

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics, 27*(8), 1164-1165.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., . . . Notredame, C. (2011). T-Coffee: a web server for the multiple sequence

alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research, 39*(suppl 2), W13-W17.

Doidy, J., Grace, E., Kuhn, C., Simon-Plas, F., Casieri, L., & Wipf, D. (2012). Sugar transporters in plants and in their interactions with fungi. [Review]. *Trends Plant Sci, 17*(7), 413-422. doi: 10.1016/j.tplants.2012.03.009

Duarte, J. M., Wall, P. K., Edger, P. P., Landherr, L. L., Ma, H., Pires, J. C., . . . dePamphilis, C. W. (2010). Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology, 10*, 61. doi: 10.1186/1471-2148-10-61

Ejeta, G. (2007). The Striga scourge in Africa: a growing pandemic. *Integrating New Technologies for Striga Control: Towards Ending the Witch-hunt*, 3-16.

Estep, M., Mourik, T., Muth, P., Guindo, D., Parzies, H., Koita, O., . . . Bennetzen, J. (2011). Genetic Diversity of a Parasitic Weed, Striga hermonthica, on Sorghum and Pearl Millet in Mali. *Tropical Plant Biology, 4*(2), 91-98. doi: 10.1007/s12042-011-9073-6

Format Converter v2.1.5. (2013) Retrieved from http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html on June 22, 2013.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., . . . Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Nucleic Acids Res, 40*(Database issue), D1178-1186. doi: 10.1093/nar/gkr944

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol, 59*(3), 307-321. doi: 10.1093/sysbio/syq010

Hiraoka, Y., & Sugimoto, Y. (2008). Molecular responses of sorghum to purple witchweed (Striga hermonthica) parasitism. *Weed Science, 56*(3), 356-363. doi: Doi 10.1614/Ws-07-136.1

Huang, K., Mellor, K. E., Paul, S. N., Lawson, M. J., Mackey, A. J., & Timko, M. P. (2012). Global changes in gene expression during compatible and incompatible interactions of cowpea (Vigna unguiculata L.) with the root parasitic angiosperm Striga gesnerioides. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *BMC Genomics, 13*(1), 402. doi: 10.1186/1471-2164-13-402

Irving, L. J., & Cameron, D. D. (2009). Chapter 3 You are What You Eat: Interactions Between Root Parasitic Plants and Their Hosts. In K. Jean-Claude & D. Michel (Eds.), *Advances in Botanical Research* (Vol. Volume 50, pp. 87-138): Academic Press.

Joel, D. M. (2007). Direct infection of potato tubers by the root parasite Orobanche aegyptiaca. [Article]. *Weed Research, 47*(4), 276-279. doi: 10.1111/j.1365-3180.2007.00570.x

Katzir, N., Portnoy, V., Tzuri, G., CastejonMunoz, M., & Joel, D. M. (1996). Use of random amplified polymorphic DNA (RAPD) markers in the study of the parasitic weed Orobanche. [Article]. *Theoretical and Applied Genetics, 93*(3),

367-372. doi: 10.1007/bf00223178

Kemena, C., & Notredame, C. (2009). Upcoming challenges for multiple sequence alignment methods in the high-throughput era. [Review]. *Bioinformatics, 25*(19), 2455-2465. doi: 10.1093/bioinformatics/btp452

Kuijt, J. (1969). The biology of parasitic flowering plants. University of California Press, Berkeley. *The biology of parasitic flowering plants. University of California Press, Berkeley.*

Kusumoto, D., Goldwasser, Y., Xie, X., Yoneyama, K., Takeuchi, Y., & Yoneyama, K. (2007). Resistance of red clover (Trifolium pratense) to the root parasitic plant Orobanche minor is activated by salicylate but not by jasmonate. [Article]. *Ann Bot, 100*(3), 537-544. doi: 10.1093/aob/mcm148

Lalonde, S., Wipf, D., & Frommer, W. B. (2004). Transport mechanisms for organic forms of carbon and nitrogen between source and sink. [Review]. *Annual Review of Plant Biology, 55*, 341-372. doi: 10.1146/annurev.arplant.55.031903.141758

Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science, 227*(4693), 1435-1441.

MacQuarrie, K. L., Fong, A. P., Morse, R. H., & Tapscott, S. J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends in Genetics, 27*(4), 141-148. doi: http://dx.doi.org/10.1016/j.tig.2011.01.001

Matvienko, M., Torres, M. J., & Yoder, J. I. (2001). Transcriptional responses in the hemiparasitic plant Triphysaria versicolor to host plant signals. *Plant Physiol, 127*(1), 272-282.

http://www.mimulusevolution.org/. [n.d.] Accessed on July 20, 2013.

Mohamed, K. I., Papes, M., Williams, R., Benz, B. W., & Peterson, A. T. (2006). Global invasive potential of 10 parasitic witchweeds and related Orobanchaceae. *Ambio, 35*(6), 281-288.

Narsai, R., Wang, C., Chen, J., Wu, J., Shou, H., & Whelan, J. (2013). Antagonistic, overlapping and distinct responses to biotic stress in rice (Oryza sativa) and interactions with abiotic stress. *BMC Genomics, 14*(1), 93.

Paran, I., Gidoni, D., & Jacobsohn, R. (1997). Variation between and within broomrape (Orobanche) species revealed by RAPD markers. *Heredity (Edinb), 78*(Pt 1), 68-74. doi: 10.1038/sj.hdy.6881070

Parker, C. (2009). Observations on the current status of Orobanche and Striga problems worldwide. *Pest Manag Sci, 65*(5), 453-459. doi: 10.1002/ps.1713

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A, 85*(8), 2444-2448.

Perez-Llamas, C., & Lopez-Bigas, N. (2011). Gitools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE, 6*(5), e19541. doi: 10.1371/journal.pone.0019541

Phytozome (2013). Accessed from http://www.phytozome.net/ on June 22, 2013.

Rambaut, A. FigTree. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/ on July 20, 2013.

Schneeweiss, G. M., Colwell, A., Park, J.-M., Jang, C.-G., & Stuessy, T. F. (2004). Phylogeny of holoparasitic Orobanche (Orobanchaceae) inferred from nuclear ITS sequences. *Mol Phylogenet Evol, 30*(2), 465-478. doi:

http://dx.doi.org/10.1016/S1055-7903(03)00210-0

Scholes, J. D., & Press, M. C. (2008). Striga infestation of cereal crops – an unsolved problem in resource limited agriculture. *Curr Opin Plant Biol, 11*(2), 180-186. doi: http://dx.doi.org/10.1016/j.pbi.2008.02.004

Spallek, T., Mutuku, J. M., & Shirasu, K. (2013). The genus Striga: A witch profile. *Mol Plant Pathol*. doi: 10.1111/mpp.12058

Sun, Z., Matusova, R., & Bouwmeester, H. (2007). Germination of Striga and chemical signaling involved: a target for control methods.

Swarbrick, P. J., Huang, K., Liu, G., Slate, J., Press, M. C., & Scholes, J. D. (2008). Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant Striga hermonthica. *New Phytol, 179*(2), 515-529.

Timko, M. P., & Singh, B. (2008). Cowpea, a multifunctional legume. *Genomics of tropical crop plants*, 227-258.

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

Vasey, R. A., Scholes, J. D., & Press, M. C. (2005). Wheat (Triticum aestivum) Is Susceptible to the Parasitic Angiosperm Striga hermonthica, a Major Cereal Pathogen in Africa. *Phytopathology, 95*(11), 1294-1300. doi: 10.1094/phyto-95-1294

Wall, P. K., Leebens-Mack, J., Muller, K. F., Field, D., Altman, N. S., & dePamphilis, C. W. (2008). PlantTribes: a gene and gene family resource for comparative

genomics in plants. *Nucleic Acids Res, 36*(Database issue), D970-976. doi: 10.1093/nar/gkm972

Weiss-Schneeweiss, H., Greilhuber, J., & Schneeweiss, G. M. (2006). Genome size evolution in holoparasitic Orobanche (Orobanchaceae) and related genera. *American journal of botany, 93*(1), 148-156. doi: 10.3732/ajb.93.1.148

Westwood, J.H., dePamphilis, C.W., Timko, M.P., & Yoder, J. I. Parasitic Plant Genome Project. Accessed from http://ppgp.huck.psu.edu/ on July 20, 2013.

Westwood, J. H., Roney, J. K., Khatibi, P. A., & Stromberg, V. K. (2009). RNA translocation between parasitic plants and their hosts. *Pest management science, 65*(5), 533-539. doi: 10.1002/ps.1727

Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends Plant Sci, 15*(4), 227-235. doi: http://dx.doi.org/10.1016/j.tplants.2010.01.004

Westwood, J. H., dePamphilis, C. W., Das, M., Fernández-Aparicio, M., Honaas, L. A., Timko, M. P., . . . Yoder, J. I. (2012). The Parasitic Plant Genome Project: New Tools for Understanding the Biology of Orobanche and Striga. *Weed Science, 60*(2), 295-306. doi: 10.1614/ws-d-11-00113.1

Wickett, N. J., Honaas, L. A., Wafula, E. K., Das, M., Huang, K., Wu, B., . . . dePamphilis, C. W. (2011). Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. [Research Support, U.S. Gov't, Non-P.H.S.]. *Curr Biol, 21*(24), 2098-2104. doi: 10.1016/j.cub.2011.11.011

Williams, L. E., Lemoine, R., & Sauer, N. (2000). Sugar transporters in higher plants -

A diversity of roles and complex regulation. [Review]. *Trends Plant Sci, 5*(7), 283-290. doi: 10.1016/s1360-1385(00)01681-2

Yoder, J. I. (2001). Host-plant recognition by parasitic< i> Scrophulariaceae</i>. *Curr Opin Plant Biol, 4*(4), 359-365.

Yoder, J. I., Gunathilake, P., Wu, B., Tomilova, N., & Tomilov, A. A. (2009). Engineering host resistance against parasitic weeds with RNA interference. *Pest Manag Sci, 65*(5), 460-466. doi: 10.1002/ps.1696

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

Yuan, M., & Wang, S. (2013). Rice MtN3/Saliva/SWEET Family Genes and Their Homologs in Cellular Organisms. *Mol Plant, 6*(3), 665-674. doi: 10.1093/mp/sst035

Zdobnov, E. M., & Apweiler, R. (2001). InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics, 17*(9), 847-848.

Zeid, M., Madkour, M., Koraiem, Y., Nawar, A., Soliman, M., & Zaitoun, F. (1997). Molecular Studies on Orobanche. *Journal of Phytopathology, 145*(8-9), 351-355. doi: 10.1111/j.1439-0434.1997.tb00413.x

Chapter five

Genome-Wide Identification and Expression Analysis of Monosaccharide and Sucrose Transporters in Root-Parasitic Flowering Plants

Abstract

Root-parasitic flowering plants are extremely destructive to many crop plants, causing huge losses in crop yield. Parasitic weeds cause this destruction by attaching to host plants by using haustorial connections, followed by draining the host of nutrients. Sucrose and monosaccharides are among the most important of these nutrients. Thus, monosaccharide transporters (MSTs) and sucrose transporters (SUTs) are especially important. Here, the MSTs and SUTs are identified in three parasitic weeds, the holoparasitic *Orobanche aegyptiaca*, the obligate hemiparasite *Striga hermonthica*, and the facultative hemiparasite *Triphysaria versicolor*, and their expression profiles and phylogenetic relationships are investigated. It was found that MST and SUT expression in general increases with each phase of the parasitic plant life cycle. It was found that some clades of MSTs and SUTs were expressed throughout the parasitic plant life cycle, whereas other clades were expressed in specific phases. However, differences between the three parasitic plants were found. For example, in *Orobanche,* some clades of MSTs are inactive and the SUT4 clade is absent; and in *Triphysaria*, a sharp increase in SUT expression and a decrease in MST expression occur during haustorial connection. Overall, the MST and SUT phylogenetic relationships and expression profiles show that MST and SUT expression in *Orobanche*, *Striga* and *Triphysaria* are typical for holoparasites, obligate and facultative hemiparasites, respectively. The possible involvement of MSTs and SUTs in regulatory networks is discussed. This study thus fills a long-standing void in the literature regarding the molecular mechanisms of activity by root-parasitic flowering plants.

Introduction

Parasitic plants, including plants of the family *Orobanchaceae*, are among the most agronomically destructive constraints to crop growth in the world (Ejeta, 2007; Scholes & Press, 2008). They cause billions of US$ in losses (Spallek et al., 2013; Scholes & Press, 2008). Many crop plants, including sorghum (Hiraoka & Sugimoto, 2008; Estep et al., 2011), cowpea (Timko & Singh, 2008) and maize (Matvienko et al., 2001), can fall prey to these weeds (Spallek et al., 2013), sometimes causing 100% crop yield loss (Cardwell & Lane, 1995).

Orobanchaceae plant seeds, which stay viable for many years (Bebawi et al., 1984), can germinate in response to chemical stimulants (e.g., strigolactones) from host plants (Yoder, 2001; Sun et al., 2007), and then use haustoria to attach to the root of the host plant (Berner et al., 1997; Kuijt, 1969). The parasitic plant then connects to the host vascular tissue, and then grows, draining nutrients from the host plant (Berner et al., 1997); in this phase, most of the damage to the host plant is done (Ejeta, 2007). Finally, the parasitic plant flowers and goes to seed, completing the cycle (Berner et al., 1997; Sun et al., 2007).

Despite the threat that parasitic plants pose, researchers have paid relatively little attention to parasitic plants. Early examples of genetic research on parasitic plants of the *Orobanchaceae* family involved the use of random amplified polymorphic DNA (RAPD) analysis to understand the genetics of *Orobanche* plants, and the relationship between parasitic plants of the *Orobanche* genus (Katzir et al., 1996; Paran et al., 1997). Other research projects revolved around phylogenetic relationships between *Orobanchaceae* (Schneeweiss et al., 2004; Weiss-Schneeweiss

et al., 2006; Bennett & Mathews, 2006) and host plant response to the parasitic weeds (Kusumoto et al., 2007; Swarbrick et al., 2008; Huang et al., 2012). Yet for years, there were few attempts to sequence parasitic plant genomes, including the sequencing of expressed sequence tags (ESTs) by Torres et al. (2005) and the sequencing of a full cDNA library and ESTs of *Striga hermonthica* by Yoshida et al. (2010). The Parasitic Plant Genome Project (Westwood et al., 2012) provides the tools needed for a comprehensive and accurate understanding of parasitic plant activity, including the whole genome assemblies of several parasitic plants, such as *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*; one can search these genomes via BLAST and via Gene Ontology (GO) annotations (http://ppgp.huck.psu.edu/; Westwood et al., 2012).

One kind of study that can be gained from this type of resource is a study of how and when parasitic plants gain nutrients and water from their host. One prime example, which will prove essential, is the study of sugar transporters in parasitic plants. Sugar transporters are crucial in several processes, including growth and development of plants, transporting sugars from the leaves to other parts of the plant, elongating roots, and interactions with pathogenic organisms (Bush, 1999; Williams, 2000; Buttner, 2010; Doidy et al., 2012). There are three types: monosaccharide transporters (MSTs) and sucrose transporters (SUTs), which have roles in sugar influx (Williams, 2000; Lalonde & Frommer), and the recently discovered SWEETs (i.e., MtN3), which efflux sucrose out of cells (Chen et al., 2010). While SWEETs have a role in phloem loading (Chen et al., 2012), MSTs and SUTs have a variety of roles throughout the plant, including interactions with symbiotic

and pathogenic organisms (Doidy et al., 2012). MSTs and SUTs have been shown to play an important role in the interaction between host plants and fungi (Doidy et al., 2012; Lingner et al., 2011). It is probable that MSTs and SUTs play an important role in the interaction between host plant and parasitic plant as well.

In this study, the roles of MSTs and SUTs in parasitic plant activity against a host plant are investigated. The MST and SUT repertoires of three parasitic plants, *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*, are identified, and their expression patterns analyzed with the express purpose of determining how and when parasitic plants use MSTs and SUTs when attacking a host. Moreover, phylogenetic analyses are done in order to identify clades of MSTs and SUTs that are involved with different phases of parasitic plant growth.

Methods

Data Sources

All sequences are from the Parasitic Plant Genome Project (http://ppgp.huck.psu.edu/). The parasitic plants investigated were *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor*.

The PPGP data was based on tissue-specific and whole-plant transcriptomic data created by Westwood et al. (2012), which was based on cDNA for sequencing made from mRNA as developed by Wickett et al. (2011), which consists of an EST assembly based on 1.25 million 454 FLX reads and 102.5 million Illumina GAIIx paired reads from various parasitic plants. For each species, the assembly had between 51,000 and 68,000 unigenes at least 200 bp in length with N_{50} values ranging from 663 to 952 (Wickett et al. (2011). Wickett et al. (2011) investigated

the overall classification of the transcriptomes studied by using PlantTribes (Wall et al., 2008) to assign each gene to its respective SuperTribe, Tribe and Orthogroup. Westwood et al. (2012) used Gene Ontology (GO) (Ashburner et al., 2000) to annotate the sequences.

Westwood et al. (2012) took two approaches to making the transcriptome assembly from the data mentioned above: first, the use of a CLC assembly cell (CLC bio, 10 Rogers St #101, Cambridge, MA 02142) to assemble several libraries in one step and to perform assemblies of one lane of Illumina data (Westwood et al., 2012). The second approach involved the use of the NextGENe platform (SoftGenetics, LLC., 100 Oakwood Ave, Suite 350, State College, PA 16803) to make raw sequence reads by repeatedly using a consolidation algorithm, followed by a maximum overlap assembly to assemble the most represented genes, and then mapping the unmatched reads to assembled contigs, while deleting the matched reads from the read pool (Westwood et al., 2012). Another consolidation step was done on the unmatched reads to obtain the next most represented genes. Highly represented genes were removed from the read pool until all contigs in the read pool were at most 200 bp (Westwood et al., 2012). Westwood et al. (2012) used (among other methods) the capture and coverage of putatively conserved nuclear single copy genes to measure assembly quality (Duarte et al., 2010).

From this, Westwood et al. (2012) used the PlantTribes 2.0 (Wall et al., 2008) ten-genome scaffold to identify 285 orthogroups that had one member from each of seven angiosperm genomes. In addition, Westwood et al. (2012) sorted each set of assembled unigenes into each of the 970 putative single-copy genes identified.

Finally, Westwood et al. (2012) used GeneWise (GeneWise, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, U.K.) to translate the unigenes against reference proteins, align them to the reference gene, create a scaffold, and for the reference gene report the ratio of positions filled to gene length (i.e., coverage).

Westwood et al. (2012) measured the transcripts present at different stages: Stage 0, or seed germination; stage 1, or pre-haustorial root growth; stage 2, or post-haustorial root growth; stage 3, or haustorial connection to host before vascular connection; stage 4, or haustorial connection after vascular connection; stage 5, or pre-emergent shoots (5.1) as well as pre-emergent roots (5.2); and finally stage 6, or post-emergence leaves and stems (and in the case of *Triphysaria versicolor*, roots) and reproductive structures.

Identification of MSTs and SUTs

Known *Arabidopsis* and Rice MST and SUT sequences were found on the NCBI Protein Database and were used as queries to search for potential MST and SUT transporters in the genomes of parasitic weeds *Orobanche aegyptiaca*, *Striga hermonthica,* and *Triphysaria versicolor*. The TFASTX method, part of the FASTA package (Lipman & Pearson, 1985; Pearson & Lipman, 1988), was used in searching for MST and SUT sequences. An E-value cutoff of 10.0 was used in order to obtain more distantly related MSTs and SUTs. The potential MST and SUT amino acid sequences were retrieved using custom Perl scripts, and then searched against PFAM (Punta et al., 2012) with an E-value cutoff of 1.0. Any sequence found to have the domains DUF791, MFS_1, MFS_2 or Sugar_tr was kept for further analysis. The

resulting set of sequences was searched against NCBI using the BLASTP program with an E-value cutoff of 10.0 (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to ensure that all MSTs and SUTs were found. Any sequence not identified as an MST or SUT was discarded. The sequences confirmed as MST and SUTs were kept for further analysis.

Analysis

Multiple sequence alignment was done using T-Coffee algorithms Expresso (Armougom et al., 2006) and PSI-Coffee (Kemena & Notredame, 2009; Chang et al., 2012). The alignments from both methods were combined using the T-Coffee Combine algorithm (Di Tommaso et al., 2011) so as to create a consensus between the structure-based alignments of Expresso (Armougom et al., 2006) with the profile-based alignments from PSI-Coffee (Kemena & Notredame, 2009; Chang et al., 2012). Format Converter v2.0.5 at Los Alamos National Laboratory site (http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html) was used to convert FASTA sequence files into PHYLIP interleaved format. The file was then manually adjusted so that the file had the proper input format for ProtTest 3 (Darriba et al. 2011) and and PhyML 3.0 (Guindon et al., 2010). In order to do a proper maximum likelihood phylogenetic analysis, ProtTest 3 was used to find the best amino acid model for the alignment. Phylogenetic trees were made using PhyML 3.0, using approximate likelihood ratio test (aLRT) (Anisimova & Gascuel, 2006), and automatically selecting the better choice between nearest-neighbor interchange (NNI) and subtree pruning and regraphing (SPR), and automatically selecting the better choice between Chi-square or Shimodaira-Hasegawa (SH)-like

branch supports (Guindon et al. 2010). Trees were visualized using FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/), and expression values for the parasitic weeds upon infecting a host were based on read counts expressed as FPKM (fragments per kilobase of unigene length per million reads), which were based on the results of read mapping for every unigene in the library for each species, after parasitic weeds were allowed to attack host plants (Westwood et al., 2012). More specifically, *Orobanche* was allowed to infect *Arabidopsis thaliana* and tobacco, while *Striga* infected sorghum and *Triphysaria* infected *Medicago truncatula* (Westwood et al., 2012). The expression values for each MST and SUT sequence were placed onto an Excel spreadsheet; these values were placed in a heat map using Gitools 1.8.4 (Perez-Llamas & Lopez-Bigas, 2011).

The aLRT. The aLRT compares the two best arrangements around a branch on a phylogenetic tree by using a log ratio of the likelihood of an arrangement that exists on the tree to that of the optimal alternative (Anisimova & Gascuel, 2006; Guindon et al., 2010). The aLRT is robust to mild violations of substitution models, and is much faster than traditional bootstrap methods (Anisimova et al., 2006). ProtTest 3 (Darriba et al., 2011) is used before building a tree to choose the best model beforehand, thus ensuring the minimization of substitution model violations.

Results and Discussion

Orobanche aegyptiaca was found to have 141 MSTs and 8 SUTs, while *Striga hermonthica* was found to have 153 MSTs and 13 SUTs, and the *Triphysaria versicolor* transcriptome had 182 MSTs and 15 SUTs. *Orobanche aegyptiaca* is a holoparasite and *Striga* and *Triphysaria* are hemiparasites; since holoparasites have

lost anatomical features associated with photosynthesis (Bromham et al., 2013; Westwood et al., 2010), it is expected that MSTs and SUTs associated with photosynthesis would be lost in *Orobanche*. These data represent significant difference in repertoire size between the plants, which could be expected of different types of parasitism (such as *Orobanche* being a holoparasite as opposed to *Striga* and *Triphysaria* being hemiparasites) (Westwood et al., 2010). The differences in phylogenetic organization and expression profiles, as shown below, represent even more striking differences between the parasitic plants studied.

Phylogenetic Analysis of MSTs and SUTs

Lalonde & Frommer (2012) define six categories of MSTs: the yeast hexose transporters (yeast HXTs), which are not investigated in this study; Early Responsive to Dehydration 6-like (ERD6-like) clade (Kiyosue et al., 1998), which transports glucose out of the vacuole in response to stresses (Buttner, 2007); the GLT/SGB/GLUT1 clade, which includes plastidic glucose translocators (pGLTs) (Weber et al., 2000); the polyol transporter/vacuolar glucose transporter (PLT/VGT) clade, which is expressed in vacuoles and plays crucial roles in phloem loading and long-distance transport of polyols such as mannitol and sorbitol (Noiraud et al., 2001); the sugar transporter (STP) clade, which plays a variety of roles, including interaction with symbiotic and pathogenic fungi (Doidy et al., 2012), as well as pollen development and root development (Buttner, 2010); and the inositol transporter/tonoplast membrane transporter (INT/TMT) clade, whose members localize to the tonoplast and plasma membrane (Schneider et al., 2008;

Schneider et al., 2007) and are involved in vacuolar monosaccharide transport (Wormit et al., 2006; Doidy et al., 2012).

Kuhn & Grof (2010) define five groups of SUTs. Two of these groups, the SUT3 and SUT5, are monocot-specific and are not investigated here, as the plants studied are dicots.

The dicot-specific SUT1 clade, which can be found in companion cells and sieve elements (Doidy et al., 2012), is proposed to move from companion cells to sieve elements via the endoplasmic reticulum (ER) (Liesche et al., 2011; Kuhn & Grof, 2010). SUT1 proteins are shown to play roles in phloem loading (Burkle et al., 1998; Slewinski et al., 2010) and in interacting with symbiotic and pathogenic fungi (reviewed in Doidy et al., 2012).

Members of the SUT2 clade, which is found in both monocots and dicots (Kuhn & Grof, 2010), are expressed in sink cells and to a lesser extent, source leaves (Barker et al., 2000); like SUT1 clade members, SUT2 clade members can contribute to phloem loading and transport of sucrose into various sink cells (Sauer, 2007).

Finally, the SUT4 clade, also expressed in both monocots and dicots (Kuhn & Grof, 2010), is shown to localize in sieve elements (Weise et al., 2000) and in source leaves (Frost et al., 2012), and can be involved in several processes, from responses to dehydration and photosynthesis (Frost et al., 2012) and circadian rhythms (Chincinska et al., 2013), to nodule development (Flemetakis et al., 2003).

In this study, a phylogenetic analysis revealed that *Orobanche*, *Striga*, and *Triphysaria* have SUT families that are mostly similar. The only major difference between the three parasitic plants was the size of the SUT4 clade. *Triphysaria* has

the largest SUT4 clade while the SUT4 clade is absent in *Orobanche*. Since SUT4 is primarily expressed in source leaves (Weise et al., 2000) and since *Orobanche* is a holoparasite and thus lacks anatomical features associated with photosynthesis (Bromham et al., 2013; Westwood et al., 2010), it is expected that SUT4 homologues are absent in *Orobanche*.

In contrast, significant differences in MST repertoire were observed between the three parasitic angiosperms (Table 1). It was found that *Striga* has the smallest proportion of MSTs in the PLT/VGT clade, whereas *Orobanche* has the largest proportion. This could in part be due to the profile of polyols being host-dependent (Richter & Popp, 1992) (and each plant in this study infected a different host (see Materials & Methods)). *Triphysaria* had the greatest proportion of transporters in the GLT/SGB/GLUT1 clade; because pGLTs are in this clade (Weber et al., 2000), and since *Triphysaria* is a facultative hemiparasite (as opposed to an obligate hemiparasite like *Striga hermonthica* or a holoparasite like *Orobanche aegyptiaca*), it is expected that *Triphysaria* has a greater number of MSTs in the GLT/SGB/GLUT1 clade. In the STP clade, *Striga* had the greatest proportion of STPs and *Orobanche*, the lowest. Given that STPs function in root and pollen development (Buttner, 2010), the differences in sizes of the STP clade between the three parasitic plants in this study can thus be explained by when the STPs are expressed in each plant (see below). All three plants had a significantly lower proportion of sequences in the ERD6-like clade than in *Arabidopsis thaliana*. Since ERD6 transports glucose from vacuole to cytoplasm in non-parasitic plants (Buttner, 2007), it is expected that parasitic plants, getting their nutrients from host organisms, have less of a need for

ERD6-like MSTs. *Orobanche* had ten sequences that could not be categorized into any clade, because they were relatively close to the "root" of the phylogenetic tree in a polyphyletic family. Together, the sizes of the different MST clades in parasitic weeds alone may suggest that the types of MSTs found in *Orobanche* may be representative of a holoparasite. Moreover, the expression of different clades' members throughout different phases of the parasitic life cycle suggests that MST activity in *Orobanche* is to be expected for a holoparasitic plant.

Expression of MSTs and SUTs in Three Parasitic Plants

MSTs. Figures 1-3 and 7-9 show the activity of MSTs and SUTs, respectively throughout the life of the parasitic angiosperms studied. The names of each stage are described in detail in Westwood et al. (2012). Stages 0 through 2 were categorized as "initiation phases", representing germination and root and haustorial growth (Westwood et al., 2012). An MST or SUT was placed in this category if it was found to be active in stages 1 and/or 2 at minimum. Stages 3 through 4 were categorized as "connection phases", in which a parasitic plant uses its haustoria to attach to the host plant, and then connects to its vascular system (Westwood et al., 2012). An MST or SUT required activity stages 3 and/or 4 in order to be placed in this category. Stages 5.1 through 6.3 were categorized as "growth phases"; here, characterized as growth, emergence and reproduction (Westwood et al., 2012). Any transporter found to be active in stages 5 through 6.1 at minimum was placed in this category. Finally, MSTs and SUTs that were expressed throughout a majority of life cycle stages were placed in a separate category. The results of this categorization are shown in Table 1.

The heat maps for MSTs in parasitic plants in Figures 1 – 3 show that in general, all three parasitic plants gradually increase in MST expression throughout the life cycle. This is to be expected; since most damage to the host plant is done during growth phases (i.e., the later phases, after haustorial connection) (Ejeta, 2007), there are more sugars present for transport across more types of tissues in parasitic plants. In *Striga* and *Triphysaria*, this gradual increase culminates in a sharp increase in MST expression when reproducing. *Striga hermonthica* differs from the other plants studied in that it shows a dip in MST expression during early growth phases (Figure 2).

The MST expression in *Triphysaria*, on the other hand, is unique in that its MST expression levels are somewhat reduced during connection phases (Figure 3). *Triphysaria versicolor* is a facultative hemiparasite and can thus live independently of a host (Westwood et al., 2010); this may be expected because of the possibility that *Triphysaria* may have mechanisms for taking nutrients into the phloem from the soil that share similarities with its mechanisms for taking nutrients from a host. This argument can be strengthened when one considers the SUT activity in *Triphysaria* (discussed below).

ERD6-like. *Orobanche*, *Striga*, and *Triphysaria* have similar proportions of ERD6-like MSTs, yet the expression profiles of ERD6-like proteins in the three plants differ significantly (Figures 4-6). In *Orobanche*, ERD6-like proteins are mostly inactive during parasitism, although two clades (one larger and one smaller) are active throughout the stages of life. In *Striga*, one group of ERD6-like MSTs is expressed throughout life, while another is expressed mainly during growth. In

Triphysaria, most ERD6-like proteins are expressed throughout most stages of life. Since *Orobanche* is a holoparasite, getting all nutrients from its host, and since ERD6-like proteins transport sugars out of vacuoles (Buttner, 2007), it is expected that during parasitism, *Orobanche* does not rely as much on ERD6 proteins.

GLT/SGB/GLUT1. The GLT/SGB/GLUT1 clade is mainly expressed throughout most stages of life in *Orobanche* and *Striga* (Figures 4 and 5), whereas in *Triphysaria*, members of this clade are mainly expressed in germination and growth phases, and less often in connection phases (Figure 6). This could be attributed to *Triphysaria* being a facultative hemiparasite that can survive without a host (Westwood et al., 2010), hence the lesser need to connect to the host plant roots via haustoria.

PLT/VGT. The profile of polyols is host dependent (Richter & Popp, 1992). Consistent with this, polyol transporters and vacuolar glucose transporters are shown to be multifunctional in this study, with some groups in this clade being expressed in different phases of life (Figures 4-6). Moreover, each plant in this study expresses PLT/VGT-clade proteins differently.

In *Orobanche*, four groups are mainly expressed throughout life, while one group is mainly associated with initiation and to a smaller extent growth, and another group is mainly associated with connection and growth. Another clade has somewhat more scattered expression throughout life (Figure 4). In *Striga*, three groups are mainly expressed throughout life. One group within the PLT/VGT clade has a subgroup expressed in initiation phases and another subgroup expressed

throughout life (Figure 5). By contrast, in *Triphysaria*, all groups are mostly expressed throughout most stages of life (Figure 6).

PLTs and VGTs are known for long-distance transport and phloem loading (Noiraud et al., 2001). This may explain why many PLTs and VGTs in this study are expressed throughout most stages of life; expression that maintains high throughout life may require expression in multiple types of tissues.

Aluri & Buttner (2007) show that in *Arabidopsis*, VGT proteins are mainly expressed above ground. It is thus expected that in *Orobanche*, that there are VGTs that more specifically express in growth phases (and that there are VGTs expressed throughout phases including growth in *Striga* and *Triphysaria*).

STP. Like with the PLT/VGT clade, different members of the STP clade are expressed in different phases of life (Figures 4-6).

In *Orobanche*, there are two groups with scattered expression, one group associated with growth, one group expressed throughout life, one group expressed in initiation phases, and a group of two sequences involved with germination and growth. *Orobanche* was the only parasitic weed in this study to show groups of STPs that were uniquely involved with germination. In *Striga*, there are two groups expressed in reproduction, two groups expressed throughout life, and one group associated with growth and to a lesser extent, connection. In *Triphysaria*, there were two large groups associated with reproduction, two groups mainly associated with connection and growth, and two small groups expressed throughout life.

The expression of groups of STPs in growth phases is consistent with STPs in *Arabidopsis* being expressed during pollen development and root development (Buttner, 2010).

INT/TMT. The three plants studied expressed inositol transporters and tonoplastic monosaccharide transporters differently.

INT/TMTs in *Orobanche* are mostly expressed throughout life, and sometimes expressed in germination only. *Striga* INT/TMTs, on the other hand, are more scattered in their expression throughout life.

In *Triphysaria*, one group is involved throughout most stages of life, with somewhat more expression in connection and growth phases. Another group is expressed throughout life, with slightly more expression in initiation and growth phases. These findings are consistent with the finding that TMT1 overexpression leads to accelerated development and growth in *Arabidopsis* (Wingenter et al., 2010).

SUTs. SUT expression, like MST expression, tends to be strongest in the reproduction phases. Moreover, *Orobanche* SUT expression seems to mirror MST expression in that germination and reproduction are the most highly expressed stages of life.

Striga and *Triphysaria*, on the other hand, have SUT expression profiles that differ from their MST profiles. In *Striga*, SUT expression remains relatively constant until its sharp increase in expression upon reaching the reproduction stage. In *Triphysaria*, there is a sharp increase in SUT expression at the start of haustorial connection with the host. This is in contrast to the decrease in MST expression during that phase.

The most interesting differences between the parasitic plants, however, were in the expression of sequences in different clades of the MSTs and SUTs.

SUT4 expression differs in the three plants studied (Figures 7-9). There are no SUT4s expressed in *Orobanche,* which is, as stated before, expected due to *Orobanche* being a holoparasite and SUT4 being associated with photosynthesis (Weise et al., 2000). The SUT4 expressed in *Striga* is expressed throughout most stages of life; and in *Triphysaria*, SUT4 proteins are mostly expressed during haustorial connection (Figure 9). The finding that *Triphysaria* SUT4 proteins are mostly expressed during connection with the host contrasts with the finding that SUT4 is mainly expressed in leaves more than in sink cells (Weise et al., 2000). It is possible that SUT4 in *Triphysaria* is required for attachment to the host; thus, further investigation into SUT4 funciton in *Triphysaria* is warranted.

Whereas *Orobanche* and *Striga* SUT2 sequences are expressed throughout life, *Triphysaria* SUT2 sequences are expressed in connection and growth phases. Similarly, SUT1 expression is throughout the stages of life in *Orobanche* (although more scattered than SUT2). In *Striga*, four SUT1 sequences are expressed in the reproduction phases. In *Triphysaria*, some SUT1s are expressed during haustorial connection. In non-parasitic plants, SUT1 is associated with phloem loading (Barker et al., 2000; Kuhn et al., 2003; Slewinski et al., 2010) and SUT2 has been proposed to be a sucrose sensor (Barker et al., 2000). As discussed above, it is possible that since *Triphysaria* is a facultative hemiparisite, its mechanisms for taking nutrients into the phloem from the soil share similarities with its mechanisms for taking

nutrients from a host. To fully understand whether this is the case, though, studies on SUT2 regulation and regulation by SUT2 would have to be done.

Conclusions

In this study, the expression profiles of *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor* MSTs and SUTs, and the phylogenetic relationships between MSTs and SUTs were found. Clades of MSTs and SUTs were identified as being expressed in different phases of the parasitic plant life cycle; these clades differ, depending on the parasitic plant. These findings may suggest that different parasitic plants are regulating sugar transporters differently. This may partially be due to the differences in host preference; parasitic plants would have to change the expression of genes within their regulatory networks to most effectively parasitize a host. One important aspect of such regulation could come in the form of various families of transcription factors, which according to a review by Rolland et al. (2006) regulate various types of sugar transporters. This would be an exciting avenue of research to pursue.

A) B)

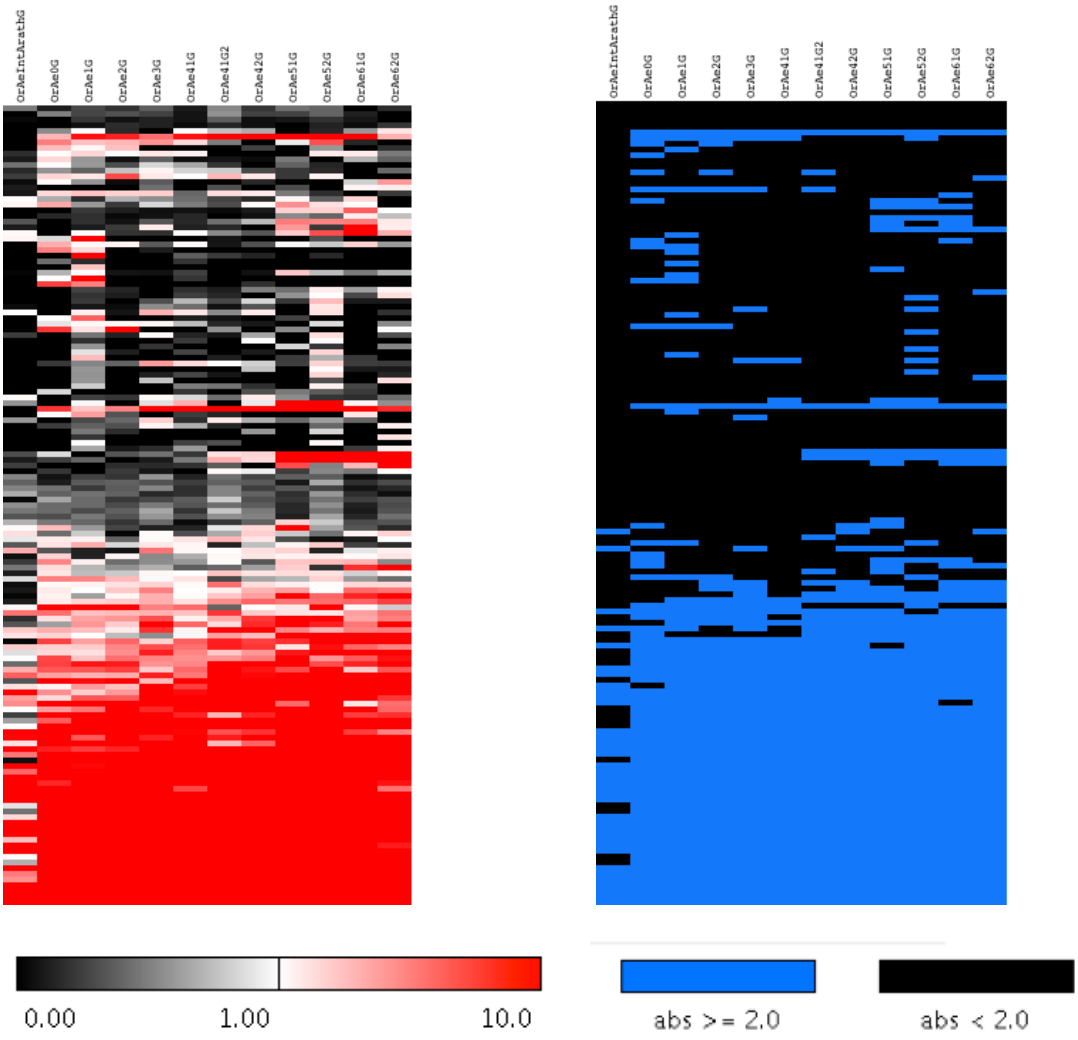


Figure 1. Heatmap showing expression of MSTs throughout the life cycle of *Orobanche aegyptiaca*. The rows were sorted in Gitools 1.8.4 (Perez-Llamas & Lopez-Bigas, 2011) in order of increasing expression levels, based on the multiplication aggregation method.

A) B)

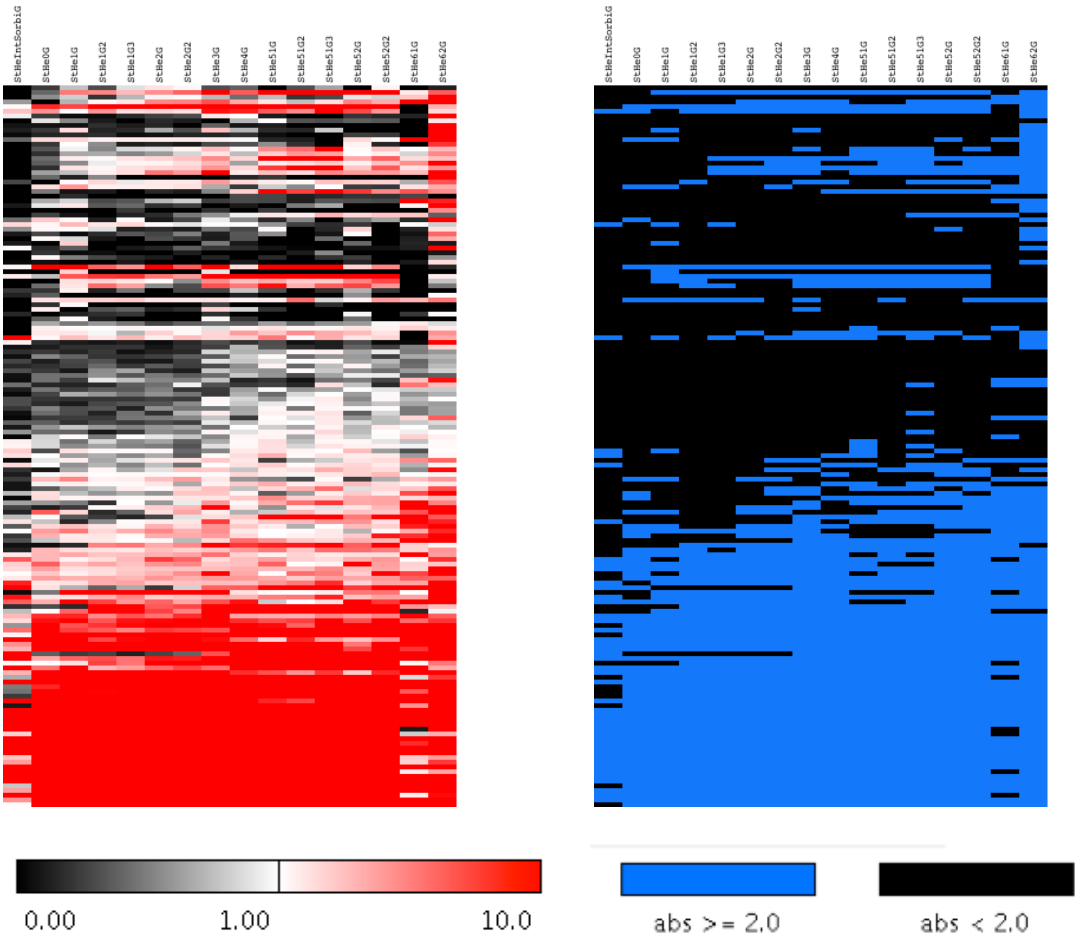


Figure 2. Heatmap showing expression of MSTs throughout the life cycle of *Striga hermonthica*. The rows were sorted in Gitools 1.8.4 (Perez-Llamas & Lopez-Bigas, 2011) in order of increasing expression levels, based on the multiplication aggregation method.

A) B)

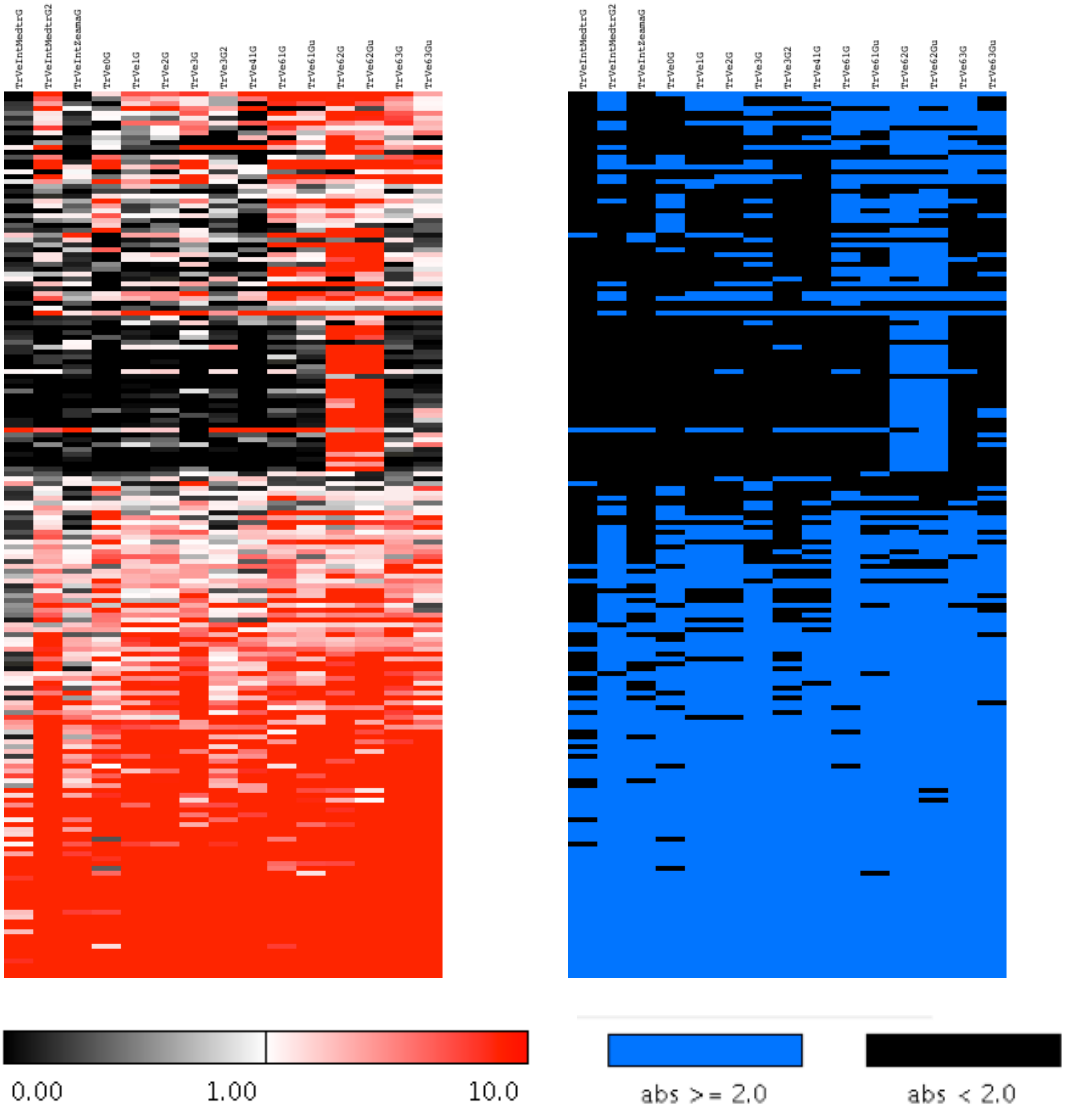


Figure 3. Heatmap showing expression of MSTs throughout the life cycle of *Triphysaria versicolor*. The rows were sorted in Gitools 1.8.4 (Perez-Llamas & Lopez-Bigas, 2011) in order of increasing expression levels, based on the multiplication aggregation method.

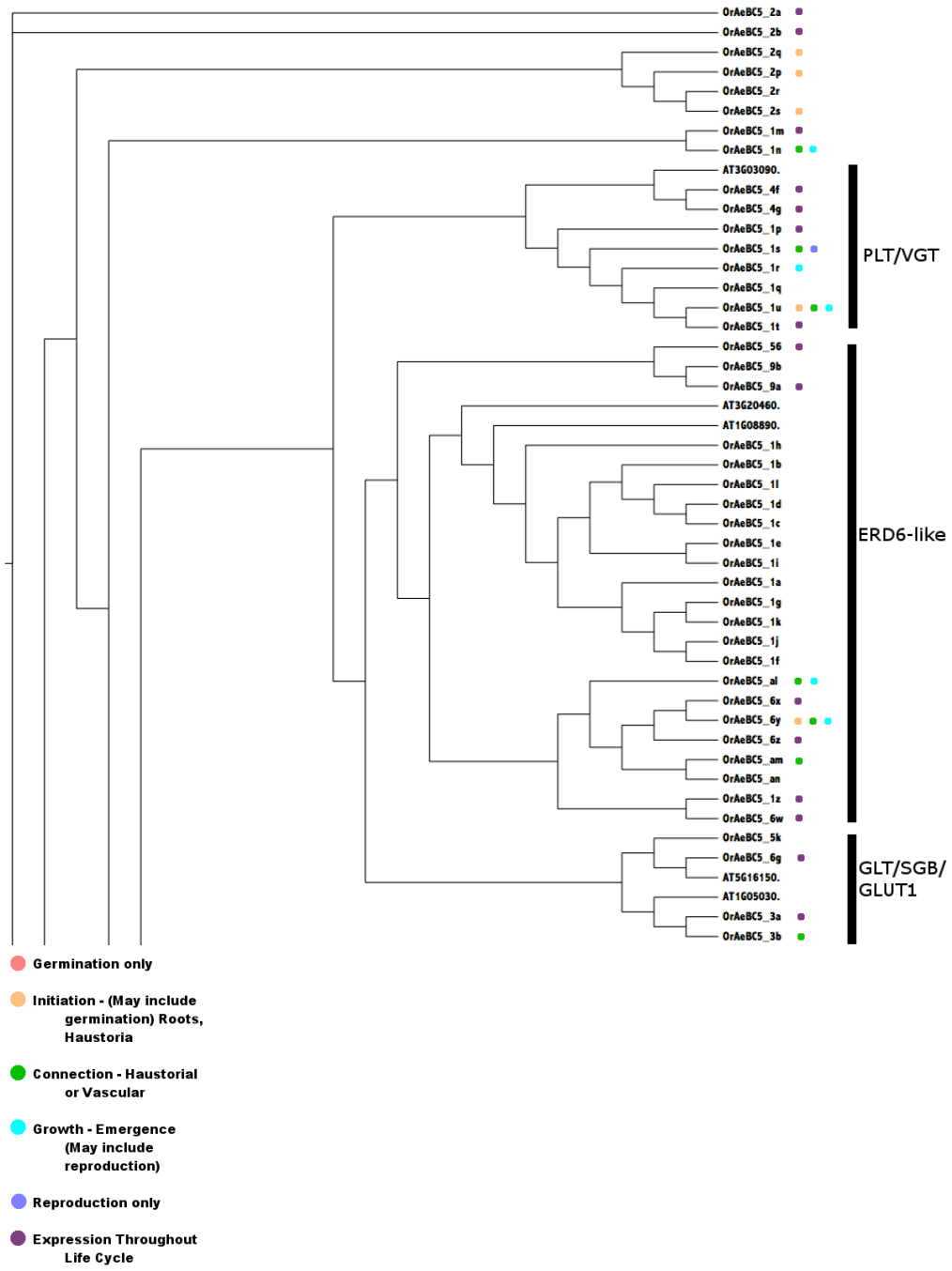


Figure 4. Phylogenetic tree showing the MSTs of *Orobanche aegyptiaca*, along with selected sequences from *Arabidopsis thaliana* from each clade of the MST superfamily from Lalonde & Frommer (2012) (Supplementary Table S7). This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al.,

2010) with the LG amino acid substitution model (Le & Gascuel, 2008), automatically estimated invariable sites and empirical frequencies, and 5 gamma parameters. Since the *Orobanche* sequence names are in PHYLIP format, the full names of these sequences are shown in Supplementary Table S1. Asterisks represent scattered expression in a phase.

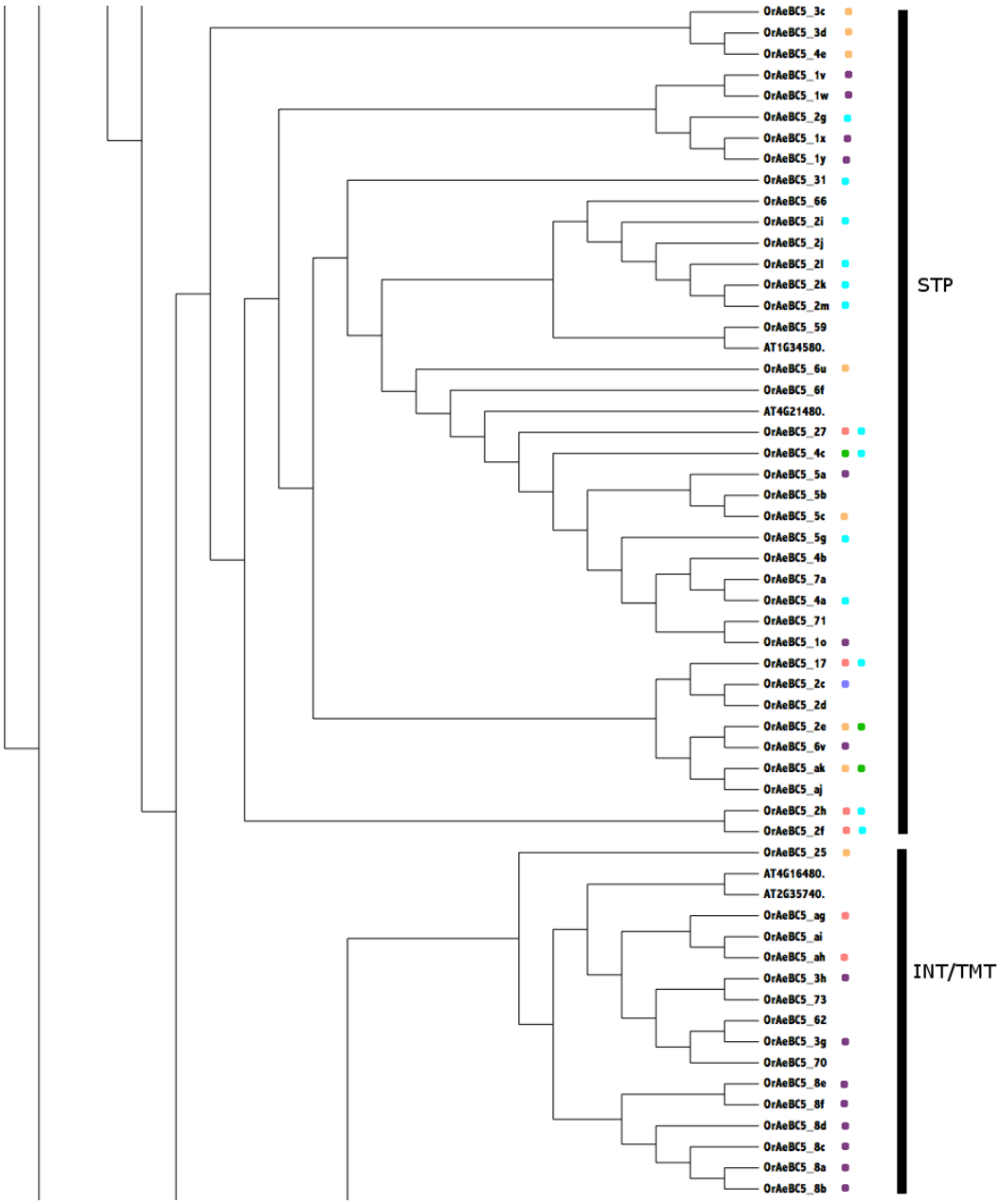


Figure 4, cont'd.

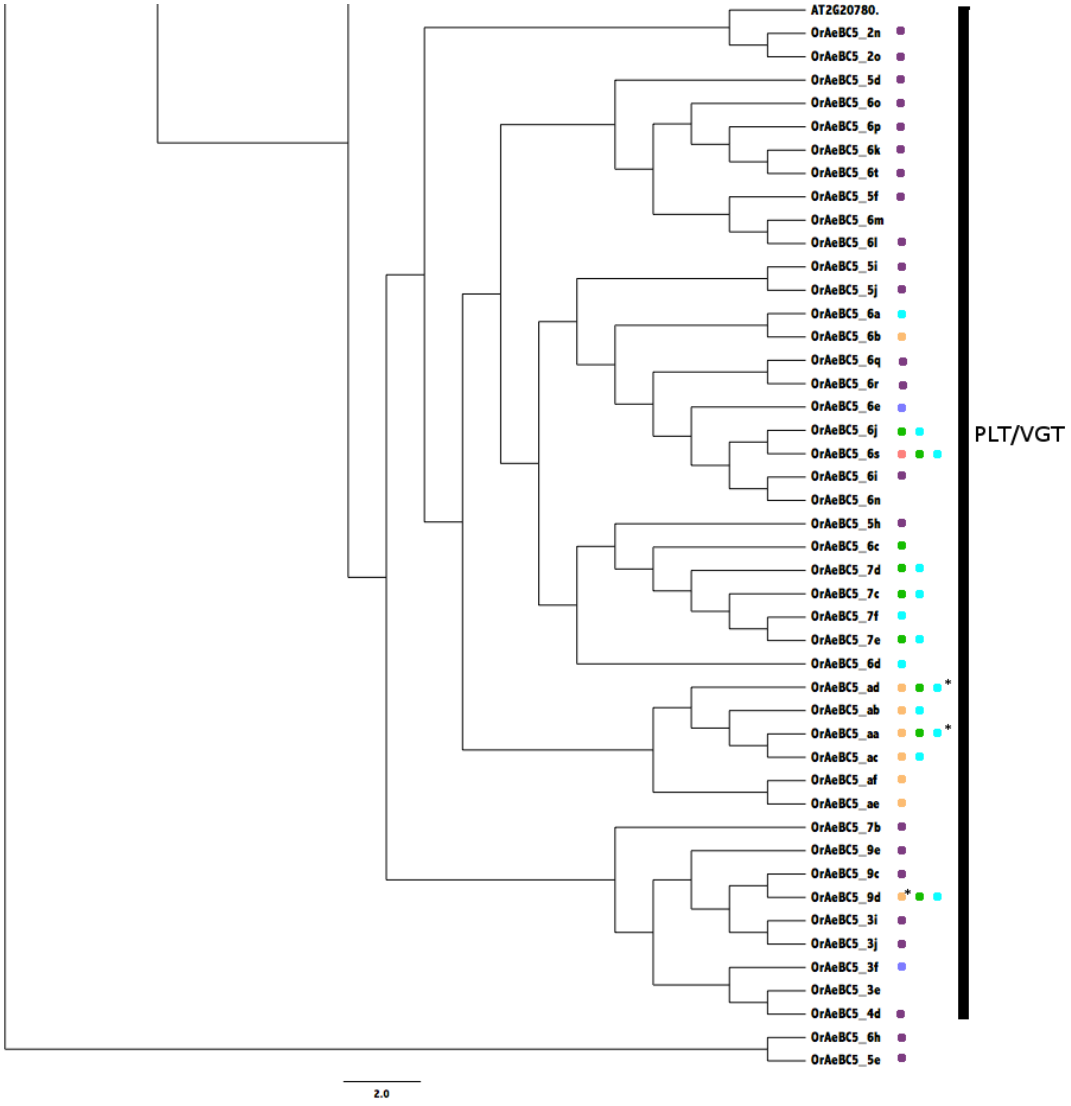


Figure 4, cont'd.

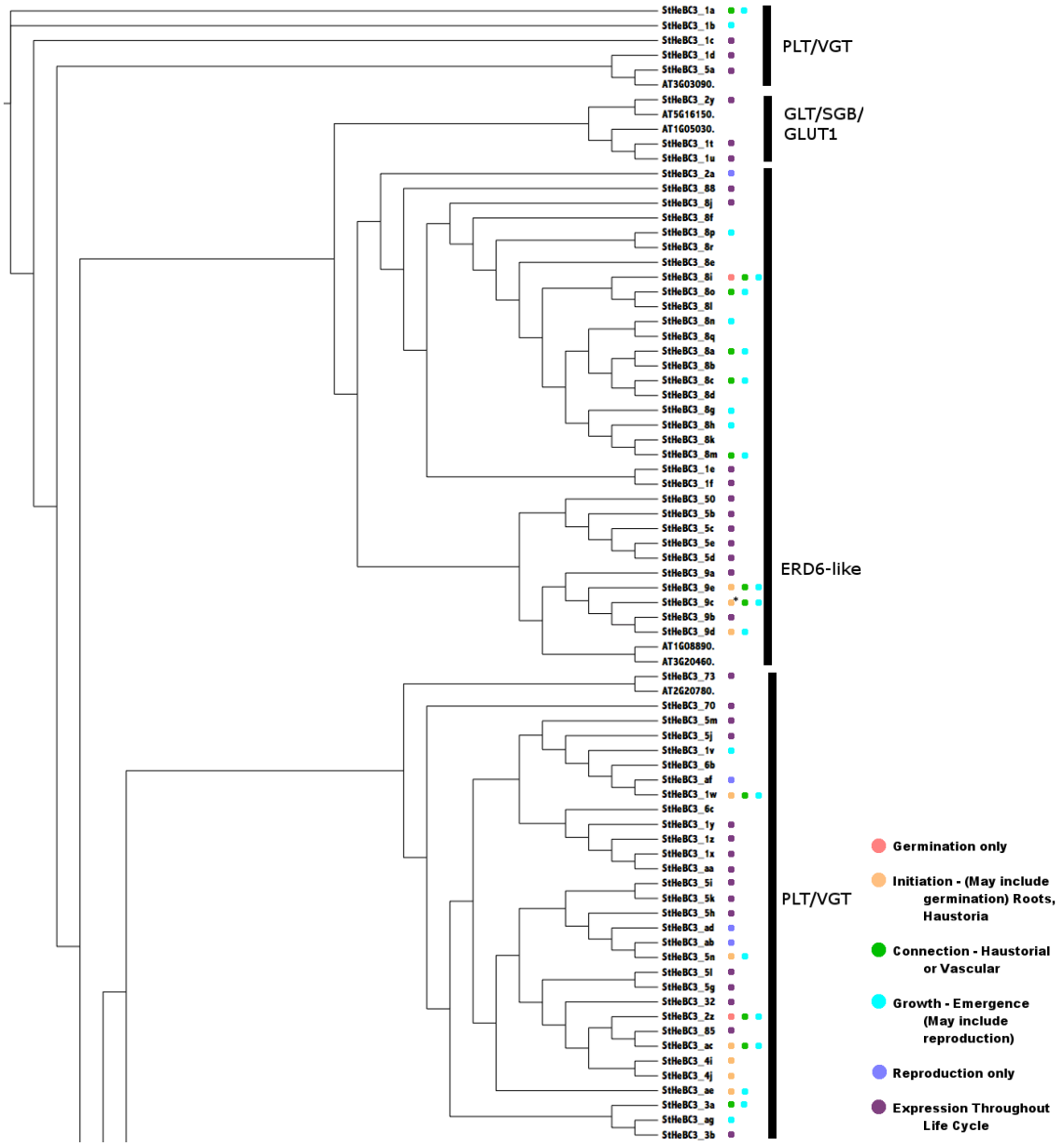


Figure 5. Phylogenetic tree showing the MSTs of *Striga hermonthica*, along with selected sequences from *Arabidopsis thaliana* from each clade of the MST superfamily from Lalonde & Frommer (2012) (Supplementary Table S7). This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the VT amino acid substitution model (Muller & Vingron, 2000), automatically estimated invariable sites and empirical frequencies, and 5 gamma

parameters. Since the *Striga* sequence names are in PHYLIP format, the full names of these sequences are shown in Supplementary Table S2. Asterisks represent scattered expression.

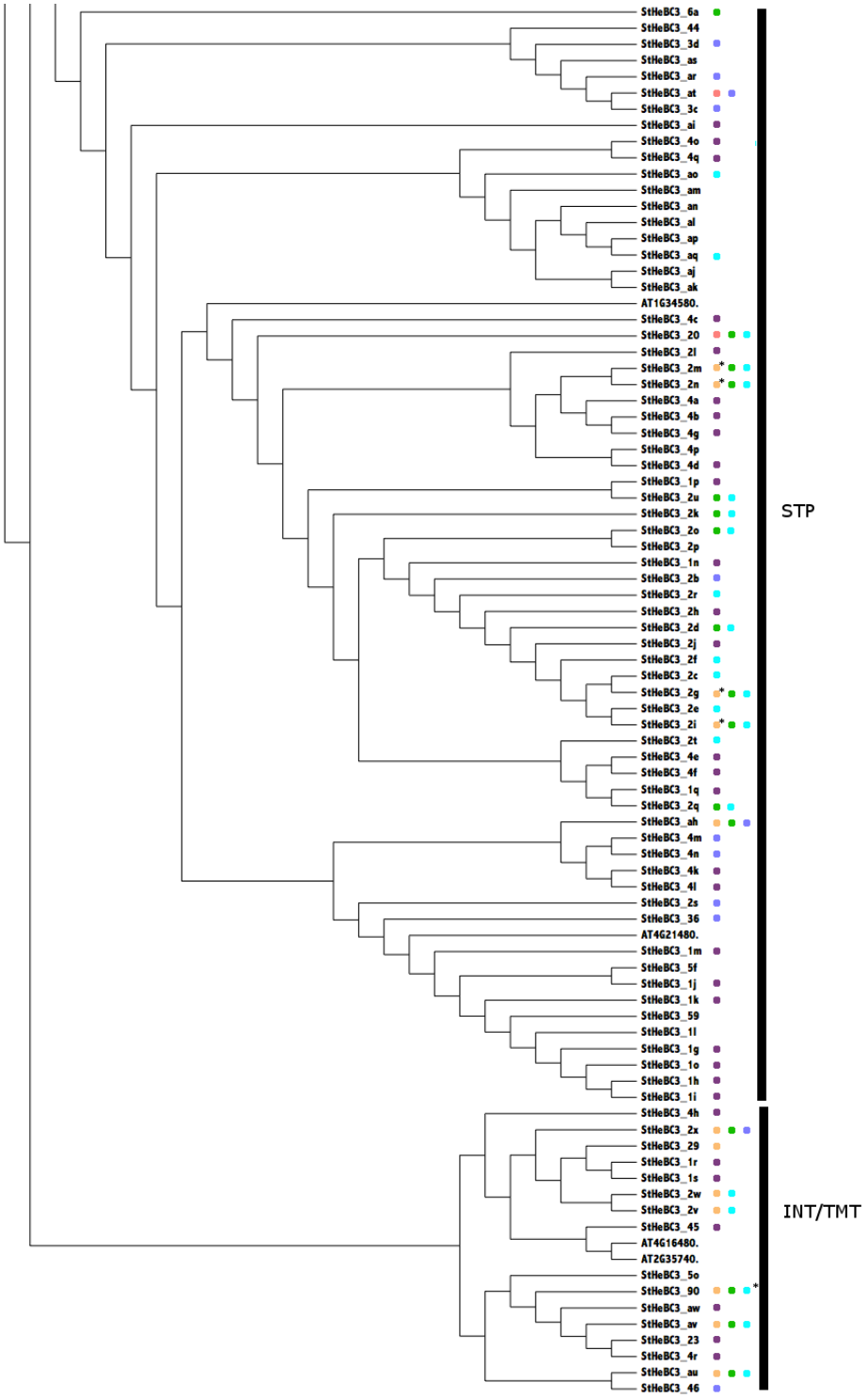


Figure 5, cont'd.

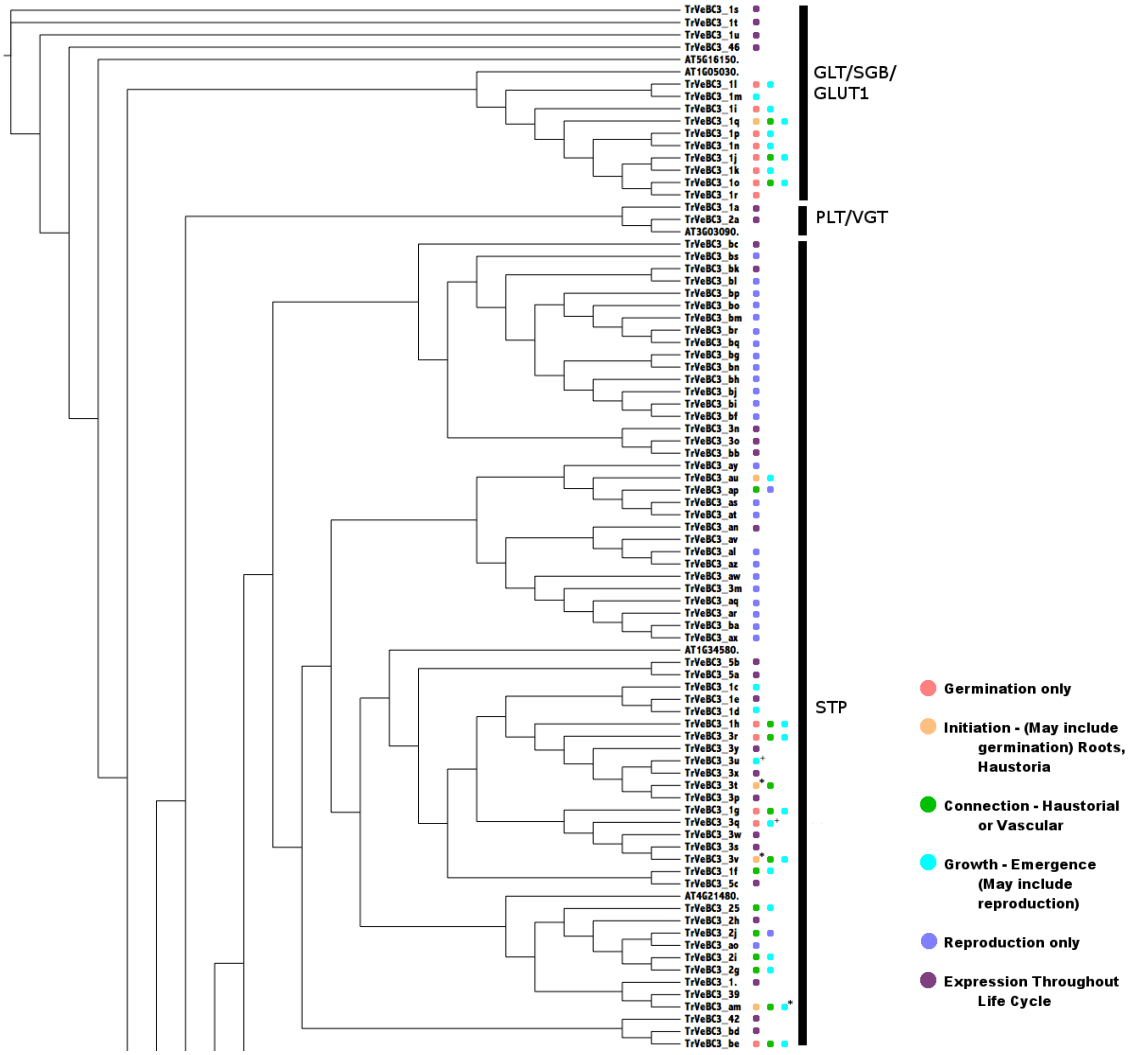


Figure 6. Phylogenetic tree showing the MSTs of *Triphysaria versicolor*, along with selected sequences from *Arabidopsis thaliana* from each clade of the MST superfamily from Lalonde & Frommer (2012) (Supplementary Table S7). This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the JTT amino acid substitution model (Jones et al., 1992), automatically estimated invariable sites and empirical frequencies, and 5 gamma parameters. Since the *Triphysaria* sequence names are in PHYLIP format, the full names of these

sequences are shown in Supplementary Table S3. Asterisks represent scattered expression and the plus sign (+) represents some growth stages without reproduction.

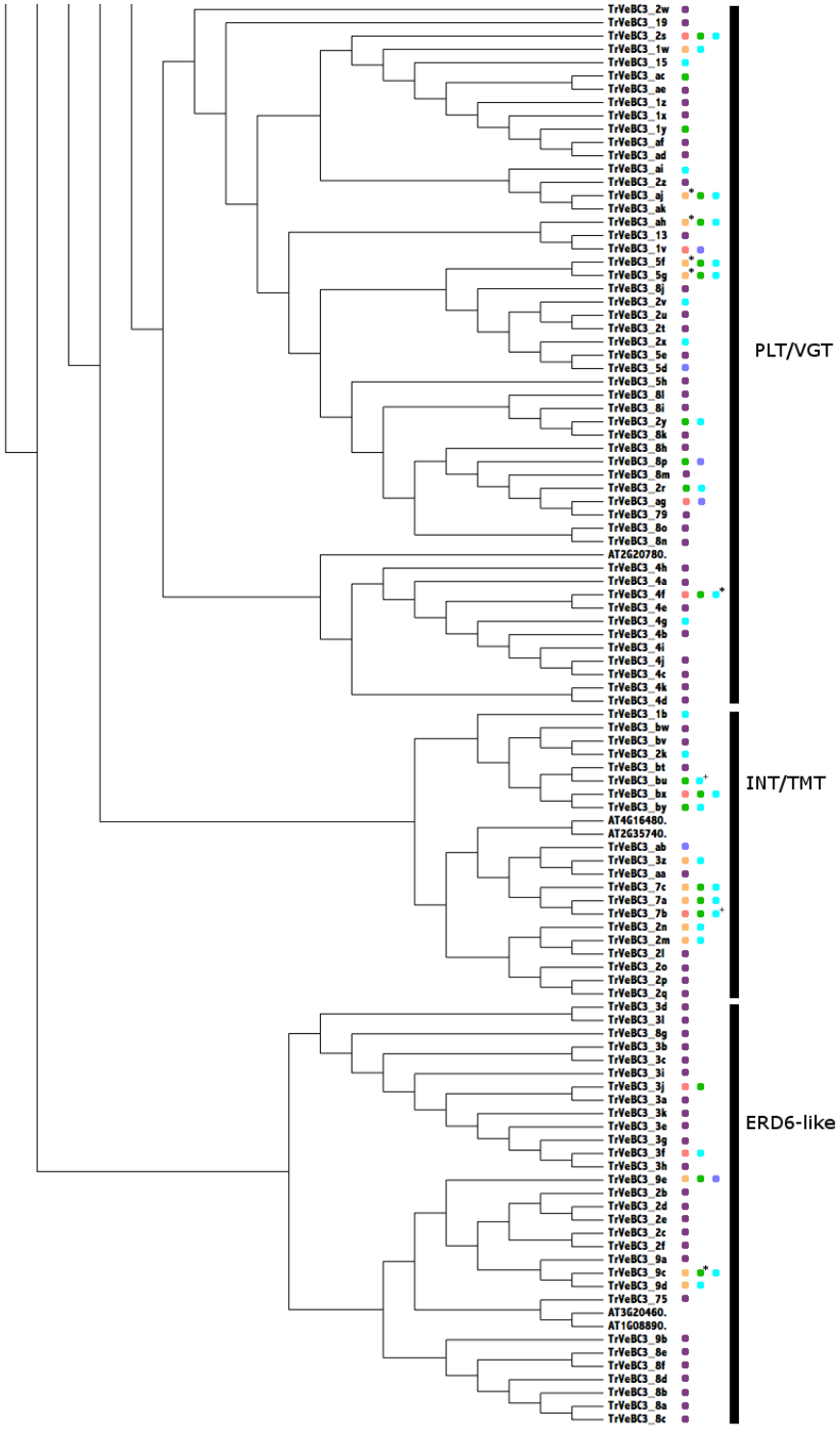


Figure 6, cont'd.

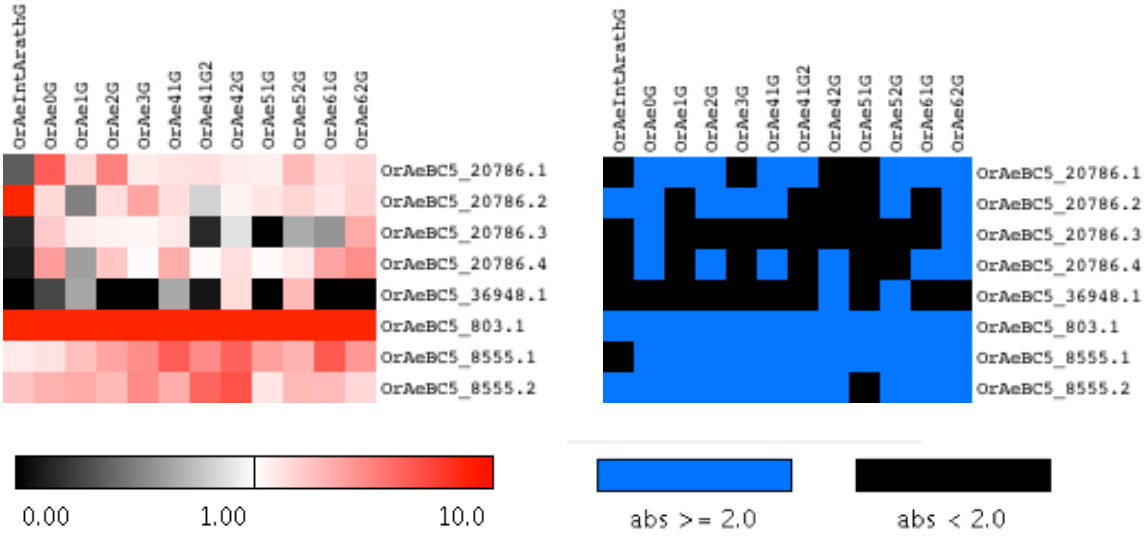


Figure 7. Heatmap showing expression of SUTs throughout the life cycle of *Orobanche aegyptiaca*.

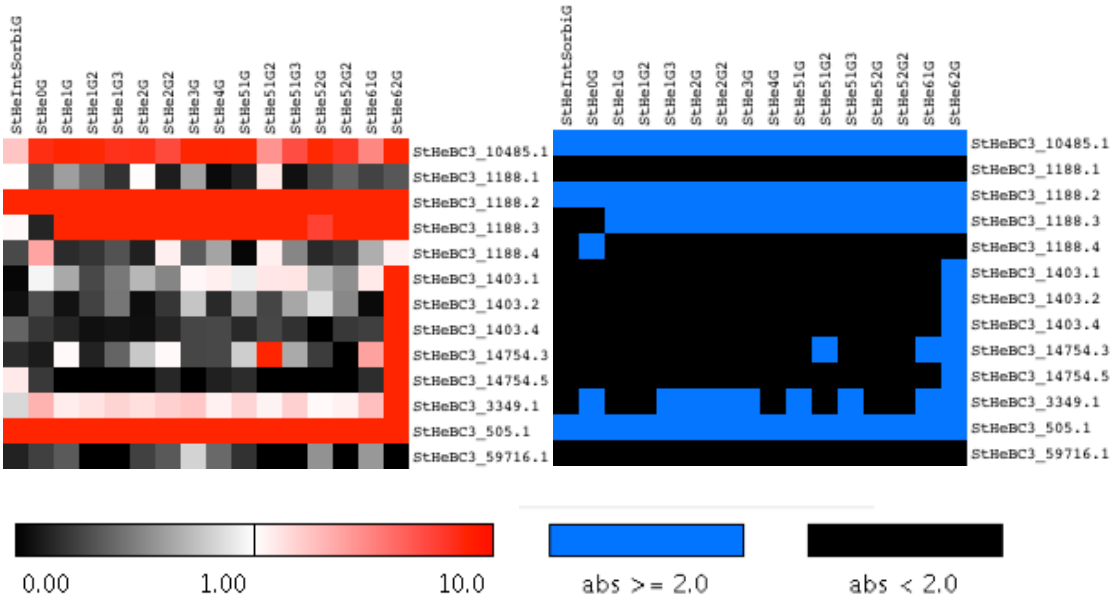


Figure 8. Heatmap showing expression of SUTs throughout the life cycle of *Striga hermonthica*.

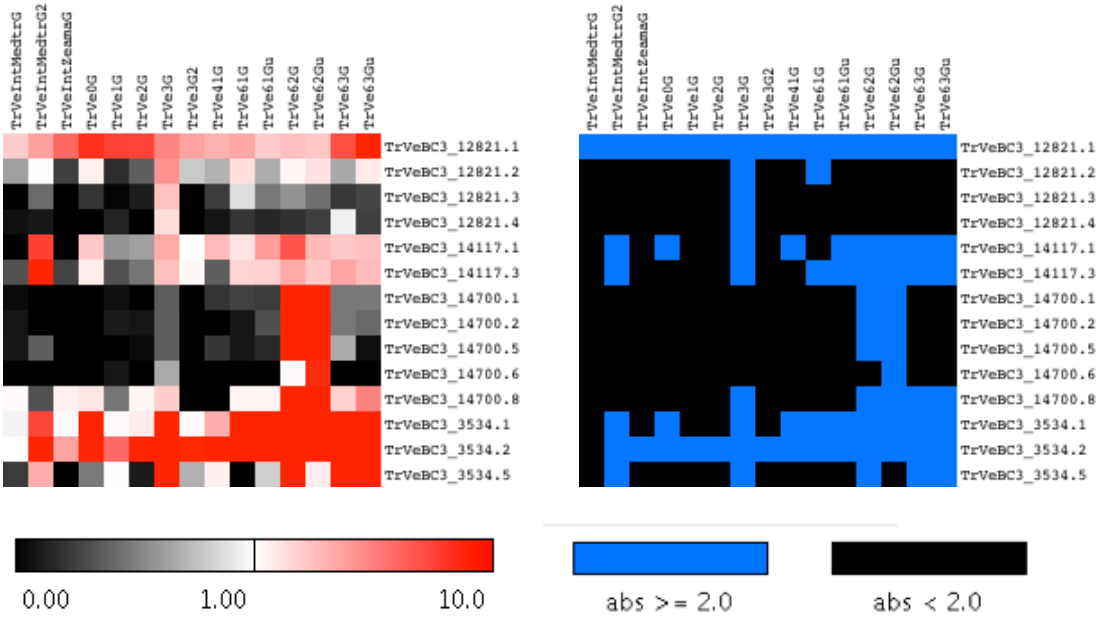


Figure 9. Heatmap showing expression of SUTs throughout the life cycle of *Triphysaria versicolor*. The rows were sorted in Gitools 1.8.4 (Perez-Llamas & Lopez-Bigas, 2011) in order of increasing expression levels, based on the multiplication aggregation method.

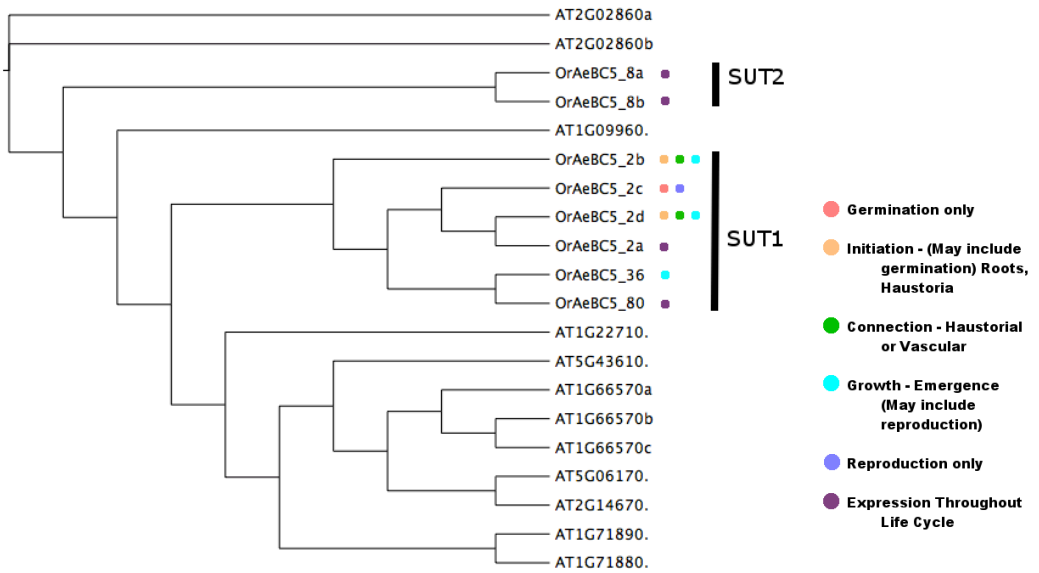


Figure 10. Phylogenetic tree showing the SUTs of *Orobanche aegyptiaca,* with associated phases. This tree was constructed using the maximum likelihood method

via PhyML 3 (Guindon et al., 2010) with the WAG amino acid substitution model (Whelan & Goldman, 2001), automatically estimated invariable sites and empirical frequencies, and 5 gamma parameters. The Arabidopsis SUTs are from Lalonde & Frommer (2012) (Supplementary Table S8). The full names of the *Orobanche* sequences are provided in Supplementary Table S4.

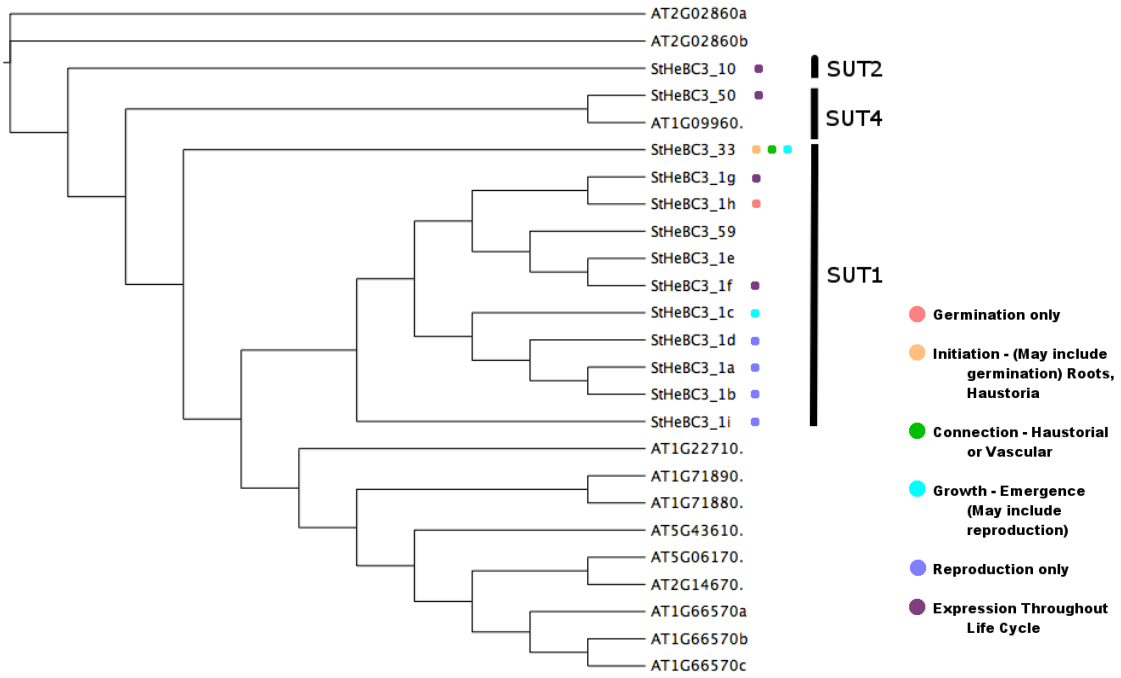


Figure 11. Phylogenetic tree showing the SUTs of *Striga hermonthica*, with associated phases. This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the WAG amino acid substitution model (Whelan & Goldman, 2001), automatically estimated invariable sites, and 5 gamma parameters. The Arabidopsis SUTs are from Lalonde & Frommer (2012) (Supplementary Table S8). The full names of the *Striga* sequences are provided in Supplementary Table S5.

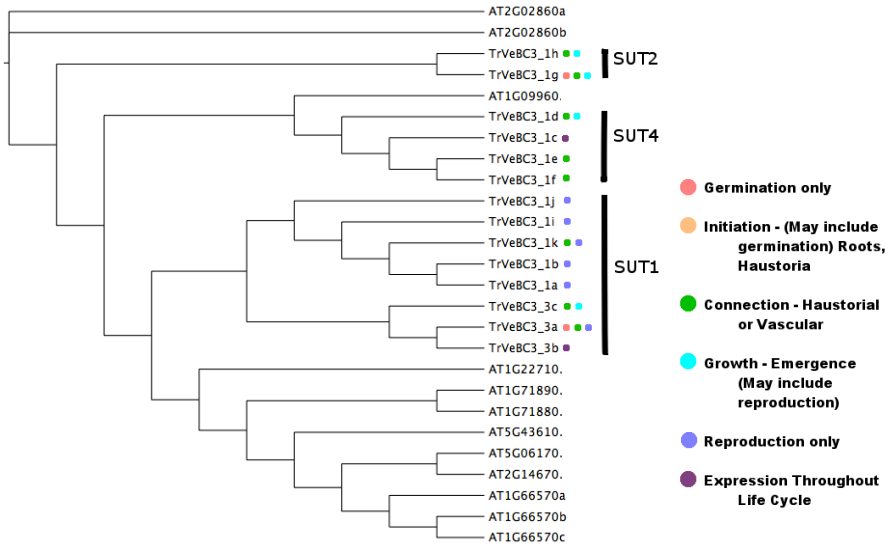


Figure 12. Phylogenetic tree showing the SUTs of *Triphysaria versicolor*, with associated phases. This tree was constructed using the maximum likelihood method via PhyML 3 (Guindon et al., 2010) with the WAG amino acid substitution model (Whelan & Goldman, 2001), automatically estimated invariable sites, and 5 gamma parameters. The Arabidopsis SUTs are from Lalonde & Frommer (2012) (Supplementary Table S8). The full names of the *Triphysaria* sequences are provided in Supplementary Table S6.

Supplementary Tables S1-S3 [Excel file]. Full names of *Orobanche, Striga*, and *Triphysaria* MST sequences.

Full Name	Tree Name
OrAeBC5_20786.1	OrAeBC5_2a
OrAeBC5_20786.2	OrAeBC5_2b
OrAeBC5_36948.1	OrAeBC5_36
OrAeBC5_8555.1	OrAeBC5_8a
OrAeBC5_8555.2	OrAeBC5_8b
OrAeBC5_20786.3	OrAeBC5_2c
OrAeBC5_20786.4	OrAeBC5_2d
OrAeBC5_803.1	OrAeBC5_80

Supplementary Table S4. Full names of *Orobanche* SUT sequences in Figure 10.

Full Name	Tree Name
StHeBC3_1403.1	StHeBC3_1a
StHeBC3_1403.2	StHeBC3_1b
StHeBC3_14754.3	StHeBC3_1c
StHeBC3_14754.5	StHeBC3_1d
StHeBC3_10485.1	StHeBC3_10
StHeBC3_1188.1	StHeBC3_1e
StHeBC3_1188.3	StHeBC3_1f
StHeBC3_505.1	StHeBC3_50
StHeBC3_1188.2	StHeBC3_1g
StHeBC3_1188.4	StHeBC3_1h
StHeBC3_1403.4	StHeBC3_1i
StHeBC3_3349.1	StHeBC3_33
StHeBC3_59716.1	StHeBC3_59

Supplementary Table S5. Full names of *Striga* SUT sequences in Figure 11.

Full Name	Tree Name
TrVeBC3_14700.5	TrVeBC3_1a
TrVeBC3_14700.6	TrVeBC3_1b
TrVeBC3_12821.1	TrVeBC3_1c
TrVeBC3_12821.2	TrVeBC3_1d
TrVeBC3_12821.3	TrVeBC3_1e
TrVeBC3_12821.4	TrVeBC3_1f
TrVeBC3_14117.1	TrVeBC3_1g
TrVeBC3_14117.3	TrVeBC3_1h
TrVeBC3_14700.2	TrVeBC3_1i
TrVeBC3_14700.1	TrVeBC3_1j
TrVeBC3_14700.8	TrVeBC3_1k
TrVeBC3_3534.1	TrVeBC3_3a
TrVeBC3_3534.2	TrVeBC3_3b
TrVeBC3_3534.5	TrVeBC3_3c

Supplementary Table S6. Full names of *Orobanche* SUT sequences in Figure 12.

Supplementary Tables S7-S8 [Excel file]. Full names of Arabidopsis MST (Supplementary Table S7) and SUT sequences (Supplementary Table S8).

References

Aluri, S., & Büttner, M. (2007). Identification and functional expression of the Arabidopsis thaliana vacuolar glucose transporter 1 and its role in seed germination and flowering. *Proceedings of the National Academy of Sciences, 104*(7), 2537-2542. doi: 10.1073/pnas.0610278104

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. [Comparative Study Research Support, Non-U.S. Gov't]. *Syst Biol, 55*(4), 539-552. doi: 10.1080/10635150600755453

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., . . . Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Research, 34*(suppl 2), W604-W608.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics, 25*(1), 25-29. doi: 10.1038/75556

Barker, L., Kuhn, C., Weise, A., Schulz, A., Gebhardt, C., Hirner, B., . . . Frommer, W. B. (2000). SUT2, a putative sucrose sensor in sieve elements. *Plant Cell, 12*(7), 1153-1164.

Bebawi, F. F., Eplee, R. E., Harris, C. E., & Norris, R. S. (1984). Longevity of witchweed (Striga asiatica) seed. *Weed Science*, 494-497.

Bennett, J. R., & Mathews, S. (2006). Phylogeny of the parasitic plant family Orobanchaceae inferred from phytochrome A. *Am J Bot, 93*(7), 1039-1051. doi: 10.3732/ajb.93.7.1039

Berner, D., Winslow, M., Awad, A., Cardwell, K., Raj, D. M., & Kim, S. (1997). Striga research methods. *International Institute for Tropical Agriculture, Ibadan*.

BLAST: Basic Local Alignment Search Tool. Accessed from http://blast.ncbi.nlm.nih.gov/Blast.cgi on July 20, 2013.

Bromham, L., Cowman, P., & Lanfear, R. (2013). Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evolutionary Biology, 13*(1), 126.

Burkle, L., Hibberd, J. M., Quick, W. P., Kuhn, C., Hirner, B., & Frommer, W. B. (1998). The H+-sucrose cotransporter NtSUT1 is essential for sugar export from tobacco leaves. *Plant Physiol, 118*(1), 59-68.

Bush, D. R. (1999). Sugar transporters in plant biology. *Current Opinion in Plant Biology, 2*(3), 187-191. doi: http://dx.doi.org/10.1016/S1369-5266(99)80034-X

Buttner, M. (2007). The monosaccharide transporter(-like) gene family in Arabidopsis. *FEBS Lett, 581*(12), 2318-2324. doi: 10.1016/j.febslet.2007.03.016

Buttner, M. (2010). The Arabidopsis sugar transporter (AtSTP) family: an update. [Review]. *Plant Biology, 12*, 35-41. doi: 10.1111/j.1438-8677.2010.00383.x

Cardwell, K. F., & Lane, J. A. (1995). Effect of soils, cropping system and host phenotype on incidence and severity of Striga gesnerioides on cowpea in

West Africa. *Agriculture, Ecosystems & Environment, 53*(3), 253-262. doi: http://dx.doi.org/10.1016/0167-8809(94)00575-Y

Chang, J. M., Di Tommaso, P., Taly, J. F., & Notredame, C. (2012). Accurate multiple sequence alignment of transmembrane proteins with PSI-Coffee. BMC Bioinformatics, 13 Suppl 4, S1. doi: 10.1186/1471-2105-13-s4-s1

Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., . . . Frommer, W. B. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Nature, 468*(7323), 527-532. doi: 10.1038/nature09606

Chen, L. Q., Qu, X. Q., Hou, B. H., Sosso, D., Osorio, S., Fernie, A. R., & Frommer, W. B. (2012). Sucrose efflux mediated by SWEET proteins as a key step for phloem transport. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Science, 335*(6065), 207-211. doi: 10.1126/science.1213351

Chincinska, I., Gier, K., Krugel, U., Liesche, J., He, H., Grimm, B., . . . Kuhn, C. (2013). Photoperiodic regulation of the sucrose transporter StSUT4 affects the expression of circadian-regulated genes and ethylene production. *Front Plant Sci, 4*, 26. doi: 10.3389/fpls.2013.00026

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics, 27*(8), 1164-1165.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., . . . Notredame, C. (2011). T-Coffee: a web server for the multiple sequence

alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res, 39*(suppl 2), W13-W17.

Doidy, J., Grace, E., Kuhn, C., Simon-Plas, F., Casieri, L., & Wipf, D. (2012). Sugar transporters in plants and in their interactions with fungi. [Review]. *Trends Plant Sci, 17*(7), 413-422. doi: 10.1016/j.tplants.2012.03.009

Duarte, J. M., Wall, P. K., Edger, P. P., Landherr, L. L., Ma, H., Pires, J. C., . . . dePamphilis, C. W. (2010). Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. *BMC Evol Biol, 10,* 61. doi: 10.1186/1471-2148-10-61

Ejeta, G. (2007). The Striga scourge in Africa: a growing pandemic. *Integrating New Technologies for Striga Control: Towards Ending the Witch-hunt*, 3-16.

Estep, M., Mourik, T., Muth, P., Guindo, D., Parzies, H., Koita, O., . . . Bennetzen, J. (2011). Genetic Diversity of a Parasitic Weed, Striga hermonthica, on Sorghum and Pearl Millet in Mali. *Tropical Plant Biology, 4*(2), 91-98. doi: 10.1007/s12042-011-9073-6

Flemetakis, E., Dimou, M., Cotzur, D., Efrose, R. C., Aivalakis, G., Colebatch, G., . . . Katinakis, P. (2003). A sucrose transporter, LjSUT4, is up-regulated during Lotus japonicus nodule development. *J Exp Bot, 54*(388), 1789-1791. doi: 10.1093/jxb/erg179

Format Converter v2.1.5. (2013) Retrieved from http://hcv.lanl.gov/content/sequence/FORMAT_CONVERSION/form.html on June 22, 2013.

Frost, C. J., Nyamdari, B., Tsai, C. J., & Harding, S. A. (2012). The tonoplast-localized sucrose transporter in Populus (PtaSUT4) regulates whole-plant water relations, responses to water stress, and photosynthesis. *PLoS ONE, 7*(8), e44467. doi: 10.1371/journal.pone.0044467

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol, 59*(3), 307-321. doi: 10.1093/sysbio/syq010

Hiraoka, Y., & Sugimoto, Y. (2008). Molecular responses of sorghum to purple witchweed (Striga hermonthica) parasitism. *Weed Science, 56*(3), 356-363. doi: Doi 10.1614/Ws-07-136.1

Huang, K., Mellor, K. E., Paul, S. N., Lawson, M. J., Mackey, A. J., & Timko, M. P. (2012). Global changes in gene expression during compatible and incompatible interactions of cowpea (Vigna unguiculata L.) with the root parasitic angiosperm Striga gesnerioides. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *BMC Genomics, 13*(1), 402. doi: 10.1186/1471-2164-13-402

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Computer applications in the biosciences: CABIOS, 8(3), 275-282.

Katzir, N., Portnoy, V., Tzuri, G., CastejonMunoz, M., & Joel, D. M. (1996). Use of random amplified polymorphic DNA (RAPD) markers in the study of the parasitic weed Orobanche. [Article]. *Theoretical and Applied Genetics, 93*(3),

367-372. doi: 10.1007/bf00223178

Kemena, C., & Notredame, C. (2009). Upcoming challenges for multiple sequence alignment methods in the high-throughput era. [Review]. *Bioinformatics, 25*(19), 2455-2465. doi: 10.1093/bioinformatics/btp452

Kiyosue, T., Abe, H., Yamaguchi-Shinozaki, K., & Shinozaki, K. (1998). ERD6, a cDNA clone for an early dehydration-induced gene of Arabidopsis, encodes a putative sugar transporter. *Biochim Biophys Acta, 1370*(2), 187-191.

Kuhn, C., & Grof, C. P. L. (2010). Sucrose transporters of higher plants. [Review]. *Curr Opin Plant Biol, 13*(3), 287-298. doi: 10.1016/j.pbi.2010.02.001

Kuhn, C., Hajirezaei, M. R., Fernie, A. R., Roessner-Tunali, U., Czechowski, T., Hirner, B., & Frommer, W. B. (2003). The sucrose transporter StSUT1 localizes to sieve elements in potato tuber phloem and influences tuber physiology and development. *Plant Physiol, 131*(1), 102-113. doi: 10.1104/pp.011676

Kuijt, J. (1969). The biology of parasitic flowering plants. University of California Press, Berkeley. *The biology of parasitic flowering plants. University of California Press, Berkeley.*

Kusumoto, D., Goldwasser, Y., Xie, X., Yoneyama, K., Takeuchi, Y., & Yoneyama, K. (2007). Resistance of red clover (Trifolium pratense) to the root parasitic plant Orobanche minor is activated by salicylate but not by jasmonate. [Article]. *Ann Bot, 100*(3), 537-544. doi: 10.1093/aob/mcm148

Lalonde, S., & Frommer, W. B. (2012). SUT Sucrose and MST Monosaccharide Transporter Inventory of the Selaginella Genome. *Frontiers in plant science, 3*, 24. doi: 10.3389/fpls.2012.00024

Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol Biol Evol, 25*(7), 1307-1320. doi: 10.1093/molbev/msn067

Liesche, J., Krugel, U., He, H., Chincinska, I., Hackel, A., & Kuhn, C. (2011). Sucrose transporter regulation at the transcriptional, post-transcriptional and post-translational level. *J Plant Physiol, 168*(12), 1426-1433. doi: 10.1016/j.jplph.2011.02.005

Lipman, D. J., & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science, 227*(4693), 1435-1441

Matvienko, M., Torres, M. J., & Yoder, J. I. (2001). Transcriptional responses in the hemiparasitic plant Triphysaria versicolor to host plant signals. *Plant Physiol, 127*(1), 272-282.

Muller, T., & Vingron, M. (2000). Modeling amino acid replacement. *J Comput Biol, 7*(6), 761-776. doi: 10.1089/10665270050514918

Noiraud, N., Maurousset, L., & Lemoine, R. (2001). Transport of polyols in higher plants. *Plant Physiology and Biochemistry, 39*(9), 717-728. doi: Doi 10.1016/S0981-9428(01)01292-X

Paran, I., Gidoni, D., & Jacobsohn, R. (1997). Variation between and within broomrape (Orobanche) species revealed by RAPD markers. *Heredity (Edinb), 78*(Pt 1), 68-74. doi: 10.1038/sj.hdy.6881070

Pearson, W. R., & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A, 85*(8), 2444-2448.

Perez-Llamas, C., & Lopez-Bigas, N. (2011). Gitools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS ONE, 6*(5), e19541. doi:

10.1371/journal.pone.0019541

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. [Research Support, Non-U.S. Gov't]. *Nucleic Acids Res, 40*(Database issue), D290-301. doi: 10.1093/nar/gkr1065

Rambaut, A. FigTree. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/ on July 20, 2013.

Richter, A., & Popp, M. (1992). The physiological importance of accumulation of cyclitols in Viscum album L.*. *New Phytologist, 121*(3), 431-438. doi: 10.1111/j.1469-8137.1992.tb02943.x

Rolland, F., Baena-Gonzalez, E., & Sheen, J. (2006). Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol, 57*, 675-709. doi: 10.1146/annurev.arplant.57.032905.105441

Sauer, N. (2007). Molecular physiology of higher plant sucrose transporters. [Review]. *FEBS Lett, 581*(12), 2309-2317. doi: 10.1016/j.febslet.2007.03.048

Schneeweiss, G. M., Colwell, A., Park, J.-M., Jang, C.-G., & Stuessy, T. F. (2004). Phylogeny of holoparasitic Orobanche (Orobanchaceae) inferred from nuclear ITS sequences. *Mol Phylogenet Evol, 30*(2), 465-478. doi: http://dx.doi.org/10.1016/S1055-7903(03)00210-0

Schneider, S., Beyhl, D., Hedrich, R., & Sauer, N. (2008). Functional and physiological characterization of Arabidopsis INOSITOL TRANSPORTER1, a novel tonoplast-localized transporter for myo-inositol. *The Plant Cell Online, 20*(4), 1073-1087.

Schneider, S., Schneidereit, A., Udvardi, P., Hammes, U., Gramann, M., Dietrich, P., & Sauer, N. (2007). Arabidopsis INOSITOL TRANSPORTER2 mediates H+ symport of different inositol epimers and derivatives across the plasma membrane. *Plant Physiol, 145*(4), 1395-1407.

Scholes, J. D., & Press, M. C. (2008). Striga infestation of cereal crops – an unsolved problem in resource limited agriculture. *Curr Opin Plant Biol, 11*(2), 180-186. doi: http://dx.doi.org/10.1016/j.pbi.2008.02.004

Slewinski, T. L., Garg, A., Johal, G. S., & Braun, D. M. (2010). Maize SUT1 functions in phloem loading. *Plant Signal Behav, 5*(6), 687-690.

Spallek, T., Mutuku, J. M., & Shirasu, K. (2013). The genus Striga: A witch profile. *Mol Plant Pathol*. doi: 10.1111/mpp.12058

Sun, Z., Matusova, R., & Bouwmeester, H. (2007). Germination of Striga and chemical signaling involved: a target for control methods.

Swarbrick, P. J., Huang, K., Liu, G., Slate, J., Press, M. C., & Scholes, J. D. (2008). Global patterns of gene expression in rice cultivars undergoing a susceptible or resistant interaction with the parasitic plant Striga hermonthica. *New Phytol, 179*(2), 515-529.

Timko, M. P., & Singh, B. (2008). Cowpea, a multifunctional legume. *Genomics of tropical crop plants*, 227-258.

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

Wall, P. K., Leebens-Mack, J., Muller, K. F., Field, D., Altman, N. S., & dePamphilis, C. W.

(2008). PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic Acids Research, 36*(Database issue), D970-976. doi: 10.1093/nar/gkm972

Weber, A., Servaites, J. C., Geiger, D. R., Kofler, H., Hille, D., Gröner, F., . . . Flügge, U.-I. (2000). Identification, purification, and molecular cloning of a putative plastidic glucose translocator. *The Plant Cell Online, 12*(5), 787-801

Weise, A., Barker, L., Kuhn, C., Lalonde, S., Buschmann, H., Frommer, W. B., & Ward, J. M. (2000). A new subfamily of sucrose transporters, SUT4, with low affinity/high capacity localized in enucleate sieve elements of plants. *Plant Cell, 12*(8), 1345-1355.

Weiss-Schneeweiss, H., Greilhuber, J., & Schneeweiss, G. M. (2006). Genome size evolution in holoparasitic Orobanche (Orobanchaceae) and related genera. *American journal of botany, 93*(1), 148-156. doi: 10.3732/ajb.93.1.148

Westwood, J. H., dePamphilis, C. W., Das, M., Fernández-Aparicio, M., Honaas, L. A., Timko, M. P., . . . Yoder, J. I. (2012). The Parasitic Plant Genome Project: New Tools for Understanding the Biology of Orobanche and Striga. *Weed Science, 60*(2), 295-306. doi: 10.1614/ws-d-11-00113.1

Westwood, J.H., dePamphilis, C.W., Timko, M.P., & Yoder, J. I. Parasitic Plant Genome Project. Accessed from http://ppgp.huck.psu.edu/ on July 20, 2013.

Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends Plant Sci, 15*(4), 227-235. doi: http://dx.doi.org/10.1016/j.tplants.2010.01.004

Whelan, S., & Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol, 18*(5), 691-699.

Wickett, N. J., Honaas, L. A., Wafula, E. K., Das, M., Huang, K., Wu, B., . . . dePamphilis, C. W. (2011). Transcriptomes of the parasitic plant family Orobanchaceae reveal surprising conservation of chlorophyll synthesis. [Research Support, U.S. Gov't, Non-P.H.S.]. *Current Biology, 21*(24), 2098-2104. doi: 10.1016/j.cub.2011.11.011

Williams, L. E., Lemoine, R., & Sauer, N. (2000). Sugar transporters in higher plants - A diversity of roles and complex regulation. [Review]. *Trends Plant Sci, 5*(7), 283-290. doi: 10.1016/s1360-1385(00)01681-2

Wingenter, K., Schulz, A., Wormit, A., Wic, S., Trentmann, O., Hoermiller, I. I., . . . Neuhaus, H. E. (2010). Increased Activity of the Vacuolar Monosaccharide Transporter TMT1 Alters Cellular Sugar Partitioning, Sugar Signaling, and Seed Yield in Arabidopsis. *Plant Physiol, 154*(2), 665-677. doi: 10.1104/pp.110.162040

Wormit, A., Trentmann, O., Feifer, I., Lohr, C., Tjaden, J., Meyer, S., . . . Neuhaus, H. E. (2006). Molecular identification and physiological characterization of a novel monosaccharide transporter from Arabidopsis involved in vacuolar sugar transport. [Research Support, Non-U.S. Gov't]. *Plant Cell, 18*(12), 3476-3490. doi: 10.1105/tpc.106.047290

Yoder, J. I. (2001). Host-plant recognition by parasitic< i> Scrophulariaceae</i>. *Curr Opin Plant Biol, 4*(4), 359-365.

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

Chapter six

General discussion

The study in chapter two was done to determine the number of transcription factors (TFs) in each cowpea TF family, as well as to determine the phylogenetic organization of the cowpea AP2-EREBP, NAC and WRKY families, and to compare several TF families in cowpea to their counterparts in *Phaseolus vulgaris* (common bean). In chapter two it was found that many TF families in cowpea are significantly different in size (as a percentage of TF repertoire) from common bean and soybean. It was also found that cowpea TF families are mostly similar to their counterparts in common bean, with one exception being the ABI3-VP1 family, which is heavily involved with the abscisic acid (ABA) pathway (Rolland et al., 2006; Nakamura et al., 2001; Shiota et al., 1998). The ABA pathway and its response to stress is complex; it can activate abiotic stress responses while repressing the activation of defense genes via the salycilic acid (SA), jasmonic acid (JA) and ethylene (ETH) signaling (de Torres Zabala et al., 2009; Mauch-Mani & Mauch, 2005; Nahar et al., 2012). However, instances of ABA signaling being a positive regulator of biotic stress response have been shown as well (reviewed in Bari & Jones, 2009; Atkinson & Urwin, 2012).

It is therefore possible that the differences between cowpea and common bean with regards to the ABI3-VP1 family, and thus differences in ABA signaling, could at least partially explain potential differences between the two legumes in terms of response to biotic stresses. In common bean, down-regulation of ABA responses contributed to an immune response to the pathogenic fungus *Colletotrichum lindemuthianum* (Oblessuc et al., 2012).

ABA signaling in cowpea may be more complex. In chapter three, it was shown that two group II NACs were downregulated when cowpea showed resistance to *Striga*, and as discussed in chapter three, group II NACs may be shut down by ABA signaling (Nuruzzaman et al., 2012; Atkinson & Urwin, 2012; Jensen et al., 2010). On the other hand, cowpea group Vb NACs are only upregulated during cowpea susceptibility. As discussed in chapter three, these NAC TFs may be induced by ABA signaling (Nuruzzaman et al., 2012; Pinheiro et al., 2009). Since ABA signaling can regulate JA, SA, and ETH signaling (Mauch-Mani & Mauch, 2005), and since, as discussed in chapter three, ERFs, NACs and WRKYs can be involved in JA and SA signaling pathways (Schweizer et al., 2013; Ryu et al., 2006), it is possible that during cowpea susceptibility, *Striga gesnerioides* is sending out signals that may in part be using ABA signals to manipulate other signaling pathways and "shut down" JA- and SA-signaling defenses. Therefore, it is possible that ABA signaling activity in cowpea may be dependent on other signals being sent, depending on the strain of *Striga gesnerioides*.

In chapter two, it was found that common bean and cowpea have different proportions of two subtypes of ABI3-VP1, namely ABI3 and REM (Romanel et al., 2009), and that common bean may have a higher proportion of the REM type of ABI3-VP1. Perhaps the differences in types of ABI3-VP1 between the two legumes could at least partially explain why ABA signaling may have a different effect in common bean than it does in cowpea.

The interactions between the ABA pathway and other signaling pathways, including the JA, SA and ETH pathways, will therefore be a productive avenue of research in the study of cowpea interactions with parasitic weeds.

TFs and Sugar Transporters

The studies in chapters three through five were done to determine the possibility of TF involvement, as well as sugar transporter involvement, in a host-parasite interaction; together, these studies were used to determine the possibility that TFs and sugar transporters were interacting with each other. First, the genome-wide analysis of TFs in cowpea, together with the phylogenetic analysis of selected TF families and an expression analysis based on microarray data from Huang et al. (2012), determined that in the interaction between cowpea and parasitic angiosperm *Striga gesnerioides* (witch weed), that while several cowpea TFs were only upregulated when cowpea shows susceptibility, there were cowpea TFs whose expression only changed when cowpea showed resistance. Secondly, the genome-wide expression and phylogenetic analysis of SWEET sugar transporters in parasitic angiosperms *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor* established that while parasitic angiosperms tend to express SWEETs more strongly when reproducing, different clades of the SWEET family in the parasitic plants express more strongly in different phases of life, including connection of the haustorium to a host plant. Thirdly, the genome-wide expression and phylogenetic analysis of monosaccharide transporters (MSTs) and sucrose transporters (SUTs) in *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor* showed that in general, MST expression increases as parasitic plants grow and mature, culminating

in high expression at the reproductive stage. In addition, it was found that in general, SUT expression was strongest in reproduction.

Together, these results show that parasitic plants are using sugar transporters when parasitizing a host, much like bacterial or fungal pathogens (Chen et al., 2010; Doidy et al., 2012), and that the host responds by employing TFs. The microarray data used in Huang et al. (2012) show that when cowpea shows resistance, expression profiles of a few sugar transporters in cowpea significantly differ from their expression profiles when cowpea shows resistance (unpublished data). When this is taken into account, along with the analysis of TFs in chapter three, it can be deduced that it may be possible for interaction between TFs and sugar transporters in the host upon resistance to parasitic weeds.

This deduction can be further understood when findings of the studies of parasitic plant sugar transporters are discussed with respect to each other, and when the findings from the study of cowpea TFs is discussed with respect to the findings from the studies of parasitic plant sugar transporters.

Sugar Transporters in Parasitic Angiosperms

The studies of sugar transporters in parasitic angiosperms reveal differences between *Orobanche aegyptiaca*, *Striga hermonthica*, and *Triphysaria versicolor.* These studies also reveal that with few exceptions, MSTs, SUTs and SWEETs show similar expression profiles.

Comparison of Parasitic Angiosperms

Egyptian broomrape. *Orobanche aegyptiaca* (syn. *Phelipanche aegyptiaca*), or Egyptian broomrape, is a holoparasite and thus lacks anatomical features

dedicated to photosynthesis (Bromham et al., 2013; Westwood et al., 2010). The studies of chapters three and four suggest that unlike in *Striga hermonthica* and *Triphysaria versicolor*, *Orobanche* SWEET expression is high during connection to the host and growth. The expression during connection and growth contrasts with MST and SUT expression profiles, in which expression in *Orobanche* slightly and gradually increases throughout life. However, in *Orobanche,* all other aspects of SWEET expression are similar to the MST and SUT expression. Together with the finding that clades of MSTs and SUTs sometimes associated with photosynthesis are not as large (or even absent) in *Orobanche*, it can be argued that sugar transporter activity in *Orobanche* may be typical for a holoparasitic plant.

Purple witchweed. *Striga hermonthica*, or purple witchweed, is an obligate hemiparasite; although it requires a host, it is still capable of photosynthesis, unlike *Orobanche* (Westwood et al., 2010). The studies of chapters three and four suggest that *Striga* SWEET expression sharply increases when reproducing. This sharp expression when reproducing is also shown in the *Striga* MST and SUT expression profiles. This is similar to the expression profiles of *Triphysaria*, discussed below; it can be argued that the sharp increase in expression upon reproduction, which is shown in chapters three and four to be shared by *Striga* and *Triphysaria,* are unique to hemiparasites because since hemiparasites are capable of photosynthesis (Westwood et al., 2010), sugars being transported to the seeds not only come from the parasitic interaction with a host, but from photosynthesis as well. Together with the findings that in *Striga*, some clades of MSTs and SUTs are expressed during connection and growth phases; and that in SUTs, unlike *Triphysaria*, there is no

sharp increase in expression during haustorial connection; it can be argued that sugar transporter activity in *Striga hermonthica* may be typical for an obligate hemiparasite.

Yellowbeak owl's clover. *Triphysaria versicolor*, or yellowbeak owl's clover, is a facultative hemiparasite; although capable of both parasitism and photosynthesis like *Striga*, *Triphysaria* does not require a host (Westwood et al., 2010). The studies of chapters three and four suggest that *Triphysaria* SWEET expression is similar to *Triphysaria* MST expression in that there is a decrease in expression upon connection to a host.

Though SWEET expression in *Triphysaria* during different phases of life are not quite as specific to certain clades as in *Striga* or *Orobanche*, MST and SUT expression through different phases of life in *Triphysaria* is attributed to certain clades. MST expression in *Triphysaria* is such that some clades were not as expressed during connection to the host. On the other hand, although *Triphysaria* shows a "spike" in SUT expression upon reproduction, there is another sharp increase in SUT expression during haustorial connection to the host. This may be attributed to *Triphysaria* being a facultative hemiparasite; as discussed in chapter five, the SUTs that are expressed when *Triphysaria* gathers nutrients from the ground could be the same SUTs that are expressed during haustorial connection to a host. Together, these findings suggest that sugar transporter activity in *Triphysaria* may be typical for a facultative hemiparasite.

Thus, it has been established that sugar transport is an essential component of parasitic angiosperm activity. The discovery of mechanisms regulating sugar

transporters is thus an excellent future avenue of research, as exploiting such mechanisms to shut down sugar transport in parasitic weeds may help control parasitic plant populations; this would be especially effective in holoparasites like *Orobanche aegyptiaca*.

When a host plant interacts with bacterial and fungal pathogens, a competition between host and pathogen for sugars ensues (Chen et al., 2010; Doidy et al., 2012). As stated above, an analysis of microarray data by Huang et al. (2012) shows that in cowpea there exist sugar transporters whose expression during resistance to parasitism is different from expression when cowpea shows susceptibility (unpublished data). Given the sugar transporter expression profiles of parasitic weeds studied, it can be argued that host and parasitic plant engage in a "tug-of-war" for sugars, similar to host interactions with fungal parasites as discussed by Doidy et al. (2012).

Glucose and sucrose are part of a signaling network that includes a variety of TFs (Rolland et al., 2006). The study of TFs in cowpea in chapter three is therefore essential to understanding the possibility that in cowpea, sugar transporters and TFs could be in the same regulatory networks.

Cowpea TFs and Sugar Transport

In chapter three it was found that in general, some clades of TF families only change expression when cowpea showed resistance to *Striga gesnerioides* race SG3, whereas other clades of those same families only change in expression when cowpea shows susceptibility to *Striga gesnerioides* race SG4z. It is possible that the changes in expression in the TF families studied, namely AP2-EREBP, NAC and

WRKY, could be involved with sugar transport and therefore may play a role in the tug-of-war for sugars with parasitic plants.

AP2-EREBP. In chapter three, it was shown that one Group III ERF and four AP2 TFs are upregulated only when cowpea shows resistance to *Striga*. On the other hand, one group VI ERF was upregulated only when cowpea showed susceptibility.

ABI4 is a group IV ERF (Nakano et al., 2006) and is involved with sugar signaling (Rolland et al., 2006; Hu et al., 2012). No group IV ERFs in cowpea were found to be either significantly upregulated or downregulated. Therefore, it is possible that sugar signaling in response to *Striga* parasitism could be involved with different TFs.

Group III ERFs can be involved with gibberellin signaling, which is involved with growth and modulation of JA and SA signaling (Magome et al., 2004; Navarro et al., 2008; Pieterse et al., 2012), and VI ERFs are involved with defense (Zhou et al., 1997; Gu et al., 2000). As discussed in chapter three, AP2 TFs, which are involved in development (Zhang et al., 2008; Elliott et al., 1996; Chuck et al., 1998; Boutilier et al., 2002; Reichmann & Meyerowitz, 1998), could be drawing nutrients away from parasitic plants. Whether these AP2-EREBP TFs are indirectly involved with sugar transport can potentially be a subject for future avenues of research.

NAC. In chapter three it was shown that in one clade of NAC sequences, which is defined as Group Ia by Zhu et al. (2012), two NACs are upregulated in both resistance and in susceptibility to *Striga,* and one is only upregulated during susceptibility.

Matallana-Ramirez et al. (2013) show that in *Arabidopsis thaliana*, a NAC TF, namely ORE1, which binds to the promoter of SWEET15, a transporter that is induced upon infection by *Pseudomonas syringae* (Chen et al., 2010). According to an expression analysis by Jensen et al. (2010), ORE1 is mainly expressed in stems and responds to abscisic acid (ABA) signaling. Moreover, ORE1 is in the same clade as ANAC031, which, according to Zhu et al., (2012) is a group Ia NAC.

Therefore, at least one of two scenarios are possible: *Striga gesnerioides* is using group-Ia NACs to "hijack" sugar transport mechanisms, or group-Ia NACs are a necessary, but not a sufficient, component of a response to parasitism. Either of these scenarios is possible because an analysis based on the microarray data of Huang et al. (2012) shows that when cowpea shows resistance, two sequences annotated as SWEETs act differently than when cowpea shows susceptibility (unpublished data). However, their expression is not significantly different from control; this may imply that cowpea may use group Ia NACs for different processes, possibly including other sugar transport mechanisms. It is also possible that group Ia NACs interact with SWEET sequences that are not annotated as SWEETs or are not represented in the Huang et al. (2012) microarray data. The possible interaction between NAC TFs and sugar transporters is but one type of interaction that would be an excellent subject of future study.

WRKY. In chapter three it was shown that while the N-terminal regions of two Group I WRKYs were only upregulated when cowpea showed susceptibility to parasitism, the C-terminal regions of two Group I WRKYs were upregulated both when cowpea showed resistance and when cowpea showed susceptibility.

SUSIBA2, a group I WRKY, has been found to contribute to sugar transport activity by binding to sugar responsive elements (SUREs) and by activating sugar signaling transcriptionally (Sun et al., 2003; Sun et al., 2005). Therefore, just as with the group-Ia NACs described above, group-I WRKY TFs in cowpea might be manipulated by *Striga* to hijack sugar transport, or group-I C-terminal WRKY domains are a necessary, but not sufficient, component of an effective stress response pathway. Together, the results of the analysis of TFs suggest that in cowpea, the TFs from the AP2-EREBP, NAC, and WRKY families that may be involved in sugar signaling pathways may be manipulated by parasitic plants, as well as in sugar signaling pathways that are necessary but not sufficient for an effective response to parasitic plants. As for the TFs that were only upregulated when cowpea showed resistance, it is still possible that they may be at least indirectly downstream targets of sugar transporters (or vice versa); these TFs may, after all, may be examples of TFs regulated by sugar signaling (Rolland et al., 2006). It is also possible that members of other TF families are directly interacting with sugar transporters when cowpea shows resistance. Understanding the interactions between TFs and other genes such as sugar transporters in a host-parasite interaction could be part of an interactome study, which would be an interesting future avenue of research.

Conclusions

Overall, the studies done in chapters two and three suggest that some of the changes in TF activity in cowpea upon different types of interactions with *Striga gesnerioides* may in part be explained by changes in the activity of the JA, SA, and ETH pathways, which may in part be mediated by the ABA pathway. In addition, the studies from

chapters three through five, along with microarray data from Huang et al. (2012), host and parasitic plant engage in a competition for sugars and that it is possible that TFs are interacting with the sugar transporters involved. However, finding evidence for such interactions, as well as finding further evidence of interactions between the signaling pathways in cowpea upon *Striga* parasitism, whether direct or indirect, would require the pursuit of several avenues of research:

- An updated oligonucleotide expression array for cowpea during resistant and susceptible interactions is warranted, since the array used in Huang et al. (2012) is based on the cowpea gene-space read (GSR) assembly sequenced by Timko et al. (2008), and not the latest cowpea assembly on HarvEST (Close et al., 2007), which is also the most comprehensive cowpea assembly to date. An updated array may thus help provide a more comprehensive understanding of how genetic expression changes when cowpea shows resistance to witch weeds, as opposed to when cowpea shows susceptibility.
- Genome-wide analyses of TFs in parasitic plants would be highly informative. If it is possible that TFs interact with sugar transporters in the host organism, then it is also possible that interactions between TFs and sugar transporters exist in parasitic weeds as well.
- Studying RNA interference (RNAi) mechanisms in the interaction between cowpea and *Striga gesnerioides* will be a necessary step in understanding the differences between genome-wide expression in cowpea resistance to *Striga* and expression during cowpea susceptibility. Both host and parasitic plants

are known to employ RNAi mechanisms (Westwood et al., 2009; Runo et al., 2011; Leblanc et al., 2012).

- A genome-wide interactome study in cowpea may provide the most direct evidence of interactions between genes during defense against parasitism. Such a study could involve (among other technologies) protein microarrays, which can be used to reconstruct interaction networks such as host-pathogen interactions (Uzoma & Zhu, 2013; Braun et al., 2013). Understanding such networks will allow researchers to make more informed decisions on how to create cultivars of crops that are not only resistant to parasitic plants and other pathogens, but also safer for consumers (Runo et al., 2011).

Combinations of findings from these avenues of research will not only increase our understanding of host-parasite interactions, but will allow researchers to use this understanding to make crop plants more resilient in the face of parasitic plants. The methods employed in some of these avenues of research may also be applied to plant response to other stresses (Braun et al., 2013).

References

Atkinson, N. J., & Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. [Review]. J Exp Bot, 63(10), 3523-3543. doi: 10.1093/jxb/ers100

Bari, R., & Jones, J. G. (2009). Role of plant hormones in plant defence responses. Plant Mol Biol, 69(4), 473-488. doi: 10.1007/s11103-008-9435-0

Boutilier, K., Offringa, R., Sharma, V. K., Kieft, H., Ouellet, T., Zhang, L., . . . van Lookeren Campagne, M. M. (2002). Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell, 14*(8), 1737-1749.

Braun, P., Aubourg, S., Van Leene, J., De Jaeger, G., & Lurin, C. (2013). Plant protein interactomes. *Annu Rev Plant Biol, 64*, 161-187. doi: 10.1146/annurev-arplant-050312-120140

Bromham, L., Cowman, P., & Lanfear, R. (2013). Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evolutionary Biology, 13*(1), 126.

Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., . . . Frommer, W. B. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Nature, 468*(7323), 527-532. doi: 10.1038/nature09606

Chuck, G., Meeley, R. B., & Hake, S. (1998). The control of maize spikelet meristem fate by the APETALA2-like gene indeterminate spikelet1. *Genes Dev, 12*(8),

1145-1154.

Close, T., Wanamaker, S., Roose, M. L., & Lyon, M. (2007). HarvEST. In D. Edwards (Ed.), *Plant Bioinformatics* (Vol. 406, pp. 161-177): Humana Press.

De Torres Zabala, M., Bennett, M. H., Truman, W. H., & Grant, M. R. (2009). Antagonism between salicylic and abscisic acid reflects early host–pathogen conflict and moulds plant defence responses. The Plant Journal, 59(3), 375-386. doi: 10.1111/j.1365-313X.2009.03875.x

Doidy, J., Grace, E., Kuhn, C., Simon-Plas, F., Casieri, L., & Wipf, D. (2012). Sugar transporters in plants and in their interactions with fungi. [Review]. *Trends Plant Sci, 17*(7), 413-422. doi: 10.1016/j.tplants.2012.03.009

Elliott, R. C., Betzner, A. S., Huttner, E., Oakes, M. P., Tucker, W. Q., Gerentes, D., . . . Smyth, D. R. (1996). AINTEGUMENTA, an APETALA2-like gene of Arabidopsis with pleiotropic roles in ovule development and floral organ growth. *Plant Cell, 8*(2), 155-168.

Gu, Y.-Q., Yang, C., Thara, V. K., Zhou, J., & Martin, G. B. (2000). Pti4 is induced by ethylene and salicylic acid, and its product is phosphorylated by the Pto kinase. *The Plant Cell Online, 12*(5), 771-785.

Hu, Y.-F., Li, Y.-P., Zhang, J., Liu, H., Tian, M., & Huang, Y. (2012). Binding of ABI4 to a CACCG motif mediates the ABA-induced expression of the ZmSSI gene in maize (Zea mays L.) endosperm. *J Exp Bot, 63*(16), 5979-5989. doi: 10.1093/jxb/ers246

Huang, K., Mellor, K. E., Paul, S. N., Lawson, M. J., Mackey, A. J., & Timko, M. P. (2012). Global changes in gene expression during compatible and incompatible

interactions of cowpea (Vigna unguiculata L.) with the root parasitic angiosperm Striga gesnerioides. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *BMC Genomics, 13*(1), 402. doi: 10.1186/1471-2164-13-402

Jensen, M. K., Kjaersgaard, T., Nielsen, M. M., Galberg, P., Petersen, K., O'Shea, C., & Skriver, K. (2010). The Arabidopsis thaliana NAC transcription factor family: structure–function relationships and determinants of ANAC019 stress signalling. *Biochemical Journal, 426*(2), 183-196. doi: 10.1042/bj20091234

Leblanc, M., Kim, G., & Westwood, J. H. (2012). RNA trafficking in parasitic plant systems. *Front Plant Sci, 3*, 203. doi: 10.3389/fpls.2012.00203

Magome, H., Yamaguchi, S., Hanada, A., Kamiya, Y., & Oda, K. (2004). dwarf and delayed-flowering 1, a novel Arabidopsis mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. *Plant J, 37*(5), 720-729.

Matallana-Ramirez, L. P., Rauf, M., Farage-Barhom, S., Dortay, H., Xue, G.-P., Dröge-Laser, W., . . . Mueller-Roeber, B. (2013). NAC Transcription Factor ORE1 and Senescence-Induced BIFUNCTIONAL NUCLEASE1 (BFN1) Constitute a Regulatory Cascade in Arabidopsis. *Mol Plant*. doi: 10.1093/mp/sst012

Mauch-Mani, B., & Mauch, F. (2005). The role of abscisic acid in plant–pathogen interactions. Curr Opin Plant Biol, 8(4), 409-414.

Nahar, K., Kyndt, T., Nzogela, Y. B., & Gheysen, G. (2012). Abscisic acid interacts antagonistically with classical defense pathways in rice–migratory nematode

interaction. New Phytologist, 196(3), 901-913. doi: 10.1111/j.1469-8137.2012.04310.x

Nakamura, S., Lynch, T. J., & Finkelstein, R. R. (2001). Physical interactions between ABA response loci of Arabidopsis. Plant J, 26(6), 627-635.

Nakano, T., Suzuki, K., Fujimura, T., & Shinshi, H. (2006). Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice. *Plant Physiol, 140*(2), 411-432. doi: 10.1104/pp.105.073783

Navarro, L., Bari, R., Achard, P., Lison, P., Nemri, A., Harberd, N. P., & Jones, J. D. (2008). DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. *Curr Biol, 18*(9), 650-655. doi: 10.1016/j.cub.2008.03.060

Nuruzzaman, M., Sharoni, A. M., Satoh, K., Moumeni, A., Venuprasad, R., Serraj, R., . . . Kikuchi, S. (2012). Comprehensive gene expression analysis of the NAC gene family under normal growth conditions, hormone treatment, and drought stress conditions in rice using near-isogenic lines (NILs) generated from crossing Aday Selection (drought tolerant) and IR64. Mol Genet Genomics, 287(5), 389-410. doi: 10.1007/s00438-012-0686-8

Oblessuc, P. R., Borges, A., Chowdhury, B., Caldas, D. G., Tsai, S. M., Camargo, L. E., & Melotto, M. (2012). Dissecting Phaseolus vulgaris innate immune system against Colletotrichum lindemuthianum infection. PLoS ONE, 7(8), e43161. doi: 10.1371/journal.pone.0043161

Pieterse, C. M., Van der Does, D., Zamioudis, C., Leon-Reyes, A., & Van Wees, S. C. (2012). Hormonal modulation of plant immunity. *Annu Rev Cell Dev Biol, 28*, 489-521. doi: 10.1146/annurev-cellbio-092910-154055

Pinheiro, G. L., Marques, C. S., Costa, M. D. B. L., Reis, P. A. B., Alves, M. S., Carvalho, C. M., . . . Fontes, E. P. B. (2009). Complete inventory of soybean NAC transcription factors: Sequence conservation and expression analysis uncover their distinct roles in stress response. Gene, 444(1–2), 10-23. doi: http://dx.doi.org/10.1016/j.gene.2009.05.012

Riechmann, J. L., & Meyerowitz, E. M. (1998). The AP2/EREBP family of plant transcription factors. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S. Review]. *Biol Chem, 379*(6), 633-646.

Rolland, F., Baena-Gonzalez, E., & Sheen, J. (2006). Sugar sensing and signaling in plants: conserved and novel mechanisms. *Annu Rev Plant Biol, 57*, 675-709. doi: 10.1146/annurev.arplant.57.032905.105441

Romanel, E. A. C., Schrago, C. G., Couñago, R. M., Russo, C. A. M., & Alves-Ferreira, M. (2009). Evolution of the B3 DNA Binding Superfamily: New Insights into REM Family Gene Diversification. PLoS ONE, 4(6), e5791. doi: 10.1371/journal.pone.0005791

Runo, S., Alakonya, A., Machuka, J., & Sinha, N. (2011). RNA interference as a resistance mechanism against crop parasites in Africa: a 'Trojan horse' approach. *Pest management science, 67*(2), 129-136. doi: 10.1002/ps.2052

Ryu, H. S., Han, M., Lee, S. K., Cho, J. I., Ryoo, N., Heu, S., . . . Jeon, J. S. (2006). A

comprehensive expression analysis of the WRKY gene superfamily in rice plants during defense response. Plant Cell Rep, 25(8), 836-847. doi: 10.1007/s00299-006-0138-1

Schweizer, F., Bodenhausen, N., Lassueur, S., Masclaux, F. G., & Reymond, P. (2013). Differential Contribution of Transcription Factors to Arabidopsis thaliana Defense Against Spodoptera littoralis. Front Plant Sci, 4, 13. doi: 10.3389/fpls.2013.00013

Shiota, H., Satoh, R., Watabe, K., Harada, H., & Kamada, H. (1998). C-ABI3, the carrot homologue of the Arabidopsis ABI3, is expressed during both zygotic and somatic embryogenesis and functions in the regulation of embryo-specific ABA-inducible genes. Plant Cell Physiol, 39(11), 1184-1193.

Sun, C., Palmqvist, S., Olsson, H., Boren, M., Ahlandsberg, S., & Jansson, C. (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. [Research Support, Non-U.S. Gov't]. *Plant Cell, 15*(9), 2076-2092.

Sun, C., Höglund, A.-S., Olsson, H., Mangelsen, E., & Jansson, C. (2005). Antisense oligodeoxynucleotide inhibition as a potent strategy in plant biology: identification of SUSIBA2 as a transcriptional activator in plant sugar signalling. *The Plant Journal, 44*(1), 128-138. doi: 10.1111/j.1365-313X.2005.02515.x

Timko, M. P., Rushton, P. J., Laudeman, T. W., Bokowiec, M. T., Chipumuro, E., Cheung, F., . . . Chen, X. (2008). Sequencing and analysis of the gene-rich space of cowpea. *BMC Genomics, 9*, 103. doi: 10.1186/1471-2164-9-103

Uzoma, I., & Zhu, H. (2013). Interactome mapping: using protein microarray technology to reconstruct diverse protein networks. *Genomics Proteomics Bioinformatics, 11*(1), 18-28. doi: 10.1016/j.gpb.2012.12.005

Westwood, J. H., Roney, J. K., Khatibi, P. A., & Stromberg, V. K. (2009). RNA translocation between parasitic plants and their hosts. *Pest management science, 65*(5), 533-539. doi: 10.1002/ps.1727

Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends Plant Sci, 15*(4), 227-235. doi: http://dx.doi.org/10.1016/j.tplants.2010.01.004

Zhang, G., Chen, M., Chen, X., Xu, Z., Guan, S., Li, L.-C., . . . Ma, Y. (2008). Phylogeny, gene structures, and expression patterns of the ERF gene family in soybean (Glycine max L.). *J Exp Bot, 59*(15), 4095-4107. doi: 10.1093/jxb/ern248

Zhou, J., Tang, X., & Martin, G. B. (1997). The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. *The EMBO Journal, 16*(11), 3207-3218.

Zhu, T., Nevo, E., Sun, D., & Peng, J. (2012). Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. [Research Support, Non-U.S. Gov't]. *Evolution, 66*(6), 1833-1848. doi: 10.1111/j.1558-5646.2011.01553.x