Computational mining for terminator-like genes in soybean

By

HAMIDA MAHMOOD

B.Sc., University of Peshawar, Pakistan, 2007

A THESIS

submitted in partial fulfillment of the requirement for the degree

MASTER OF SCIENCE

Interdepartmental Genetics Program
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Co-Major Professor
Dr. Frank White

Approved by:

Co-Major Professor
Dr. Sanzhen Liu

Abstract

Plants and bacterial pathogens are in constant co-evolution to survive and sustain the next generation. Plants have two well-characterized levels of active defense -pathogens-associated molecular patterns (PAMPs)-triggered immunity (PTI) and effectors-triggered immunity (ETI). Some plants that are hosts for bacterial pathogens employing type three secretion system transcription activator-like (TAL) effectors have evolved a unique form of ETI, namely TAL effector-mediated ETI. TAL effectors induce expression of specific disease susceptibility (S) genes. Rice and pepper have evolved resistance genes termed terminator (T) genes, which have promoters that bind TAL effectors and, upon expression of the T gene, elicit a hypersensitive reaction (HR) and cell death. Only five T genes have been cloned, and the origin of most T genes is unknown. To determine the presence of candidate T genes in other plants species, a bioinformatics-based mining was designed. The basic approach utilized three structural features common to four terminator genes: a short trans-membrane domain, a secretion signal domain, and a length of <200 amino acid residues. Soybean was chosen as the test plant species, and 161 genes were retrieved that fulfilled the three parameters using R and Perl software programs. Further, functional annotation of candidate genes was conducted by comparisons to genes in public databases. Major classes of proteins found included unique and hypothetical, defense/stress/oxidative stress associated, DNA-binding, kinases, transferases, hydrolases, effector-related tRNA splicing, and F-box domain proteins. The potential T genes will serve as candidates for experimental validation and new resources for durable resistance strategies in crop species.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am elated to express my deepest gratitude and adoration to my major supervisor Dr. Frank White for his innovative ideas, inspiring guidance and constructive criticism. I am grateful to my co-supervisor Dr. Sanzhen Liu, for his mentorship on computational analysis and committee member Dr. Christopher Toomajian, for his technical guidance and generous support.

Words are too few to thank Dr. Tariq Mahmood, my husband, for his all-time availability and thought provoking discussion with me on the courses I took and the current project I completed. Many thanks to the past members of the White lab whose hard work laid the foundations for the present study. I thank all the scientists in the Department of Plant Pathology for their highly effective educational strategy and all members of Plant Pathology family for the helpful and cheerful atmosphere, particularly, the help extended by the office staff. I also express my appreciation to my fellow graduate students for the unstinted flow of help and encouragement throughout my course of study.

I am extremely thankful to my family, particularly, my father for his love and unflinching support toward my education.

Finally, a big thanks to K-State for the excellent facilities and the scholarship provided during my study.

Dedication

To my husband, Tariq; my kids, Ayesha and Hamza, and all those greats who serve humanity.

Chapter 1. Literature review

Plants have developed sophisticated forms of defense against pathogens, and an understanding of the defense mechanisms is crucial to managing healthy and sustainable crops. In addition to static structural and chemical barriers to infection, plants have an assortment of so-called active defense mechanisms. One important line of host defense is signaling and initiating responses through the interception of elicitors known as pathogen associated molecular patterns (PAMPs) or, in some cases, damage associated molecular patterns (DAMPs), in which case the molecules are derived from degraded host structures. PAMPs are commonly conserved molecules derived from pathogens that are recognized by a variety of so-called pathogen-associated pattern receptors (PPRs) (Chisholm *et al*., 2006). Elicitor interception initiates a concerted chain of signals for defense, ultimately comprised of chemical and biochemical responses that help the plant resist infection. This well-coordinated system results in the production of antimicrobial compounds, including reactive oxygen species (ROS) and increased callose production, the latter leading to the fortification of cell walls (Blazquez *et al*., 2006). ROS results in further downstream signaling events as part of the cascade of responses. Elicitor perception by host surface receptors has been termed PAMP-triggered immunity (PTI, Figure 1).

A major class of PPRs is the leucine rich repeat receptor-linked kinases (RLKs). Two of the best-known examples are FLS2 and EFR of the model plant *Arabidopsis*. Another well-known RLK is the product of the resistance gene *Xa21*, which provides resistance to the infection of rice by the bacterium *Xanthomonas oryzae* pv. *oryzae* (Xoo). FLS2 recognizes bacterial flagellin, while EFR recognizes the elongation factor Tu (EF-

Tu) protein of bacterial pathogen *Pseudomonas syringae* (Monoghan and Zipfel, 2012). Flagellin or EF-Tu perception, at the early stage of bacterial-plant interaction and in the absence of any suppression of the defense response, generally restricts bacterial invasion and contributes to resistance. The treatment of *Arabidopsis* with a 22 amino acid conserved core peptide derived from bacterial flagellin (flg22) induces numerous defense related genes and reduces the susceptibility of the plants to bacterial colonization. Plants mutant for *FLS2* (*fls2*) are compromised for flg22 recognition (Zipfel *et al*., 2004). At the same time, both wild type and *fls2* plants showed reduced susceptibility to *P syringae* pv *tomato*, when pretreated with crude extracts of different plant pathogenic bacteria, even with extracts that lacked flagellin, indicating elicitors other than flagellin are recognized (Zipfel *et al*., 2004).

FLS2 and EFR interact with an additional kinase BAK1, which is also a RLK co-receptor. The association of BAK1 with FLS2 or EFR is required for downstream signaling, including generation of ROS and activation of MAPKs and calcium dependent protein kinase (CDPKs) (Chinchilla *et al*., 2007; Sun *et al*., 2013). The ROS burst is generally catalyzed by NADPH oxidases to produce precursor superoxide (O^{-}_{2}), which is converted to H_2O_2 by peroxidases. NADPH oxidases are the members of respiratory burst oxidase homologs (RBOHs) family in plants (Torres and Dangl, 2005). In plants, RBOHs possess a core C-terminal region containing the transmembrane domain and a functional oxidase domain responsible for the ROS burst (Suzuki *et al*., 2011) in addition to an N-terminal domain that is regulated by CDPK (Marino *et al*., 2012; Suzuki *et al*., 2011), suggesting a catalyzing role for biochemical signals upstream of ROS burst (Figure 1).

ROS is an important intermediary for PTI-triggered immunity and is regulated by different kinases. ROS acts as signal for local and systemic secondary messengers, including gene expression and stomatal closure, as an antimicrobial compound, and as a cross linker of plant cell wall to inhibit penetration (Doke, 1983., Lamb and Dixon 1997, Suzuki *et al*., 2011). BIK1, another RLK protein, is reported to positively regulate the generation of ROS burst (Laluk *et al*., 2011; Lu *et al*., 2010). Another kinase BSK1 associates with FLS2 *in vivo* to produce ROS burst in response to flg22 in *Arabidosis* (Shi *et al*., 2013).

Bacterial secreted proteases are other elicitors that can activate defense responses. Some bacterial proteases trigger responses through the mitogen activated protein kinases (MAPKs) cascade and an as yet unknown receptor for activated C kinase1 (RACK1) (Cheng *et al*., 2015). Proteases activate heterotrimeric G protein complex signaling, which consists of α, β and γ subunits of G protein complex and is activated by RACK1. RACK1 functions as scaffold between G-complex upstream and MAPK signaling.

Bacteria, in turn, have evolved to cope with PTI plant defense responses. For example, some elicitor molecules of adapted pathogens have alterations that evade PPR recognition and, hence, host immune detection (Cai *et al*, 2011). Additionally, PTI can be overcome by bacteria that have adapted to the suppressive strategies (Jones and Dangle, 2006). One such strategy adopted by some invading bacteria is the secretion of effector molecules, which target pathogen recognition receptor (PRR) function and other components of the immune system to abrogate PAMP signaling within the host cells (Boller and He, 2009). Effectors are specialized arsenal of secreted proteins that interfere

with cellular processes for the benefit of bacterial multiplication, resulting in the onset of a disease. In the case of proteobacteria, many of the effectors are directly injected into the host cytoplasm through a type 3 secretion system (T3SS). T3SS effectors typically are indispensable as a suite of proteins whose proposed function is to suppress plant immune responses (Alfano and Colmer, 2004; Mudgett, 2005). T3SS effectors may also alter other aspects of host cell physiology, for example, the release of nutritional substrates from the host cells (Chen *et al*., 2010)**.**

Plants have also evolved another line of defense based on host cell perception of effectors, which has been termed effector-triggered immunity (ETI, Figure 1). ETI responses are typically more pronounced in comparison to PTI and associated with programmed cell death and a hypersensitive reaction (HR) at the site of infection (Chisholm *et al*., 2006; Mejia-Teniente *et al.*, 2015). Historically, the gene-for-gene interaction model postulated the presence of an *R* gene in the host corresponding to each *Avr* gene in the pathogen as major aspects of ETI (Flor, 1956). The interaction between R gene products and Avr proteins can be either direct or indirect. For example, the R protein PTO directly interacts with the Avr protein, AvrPto (Salmeron *et al*., 1994; Ntoukakis *et al*., 2013). AvrRps2, on the other hand, interacts with a protein known as RIN4, which exists in a complex with the R gene product RPS2 (Mackey *et al*., 2003). Many R proteins are members of the so-called NBS-LRR class, consisting of central nucleotide binding site (NBS) motif and C-terminal leucine rich repeats (LRRs). As noted for RPS2, recognition is not necessarily due to interactions of Avr-R pairs. A more encompassing model hypothesizes that NBS-LRR proteins act as guards of defense signaling complexes. Binding of the effector to the NBS-LRR component or

disturbance of the NBS-LRR-guarded component results in the triggering of the resistance response (Dodds and Rathjen, 2010). Bacterial pathogens have also evolved to defeat ETI by selection for modification of effector structure, acquisition of alternate effector genes, or suppression of the ETI mechanism (Jones and Dangle, 2006).

A unique class of T3SS effectors is transcription activator-like (TAL) effectors, which are found in some species of *Xanthomonas* and *Ralstonia*. TAL effectors function as transcription factors in host cells, inducing so-called host disease susceptibility (S) genes. Three of the characteristic features of eukaryotic transcription factors present in TAL effectors are a DNA binding domain, nuclear localization signals, and a transcription activation domain, respectively (Boch and Bonas, 2010). The DNA binding domain consists of different numbers of 33-35 amino acid repeating units arranged in tandem. Each repeat is almost identical in amino acid sequence, except for the amino acids 12^{th} and 13^{th}, which are polymorphic and referred to as repeat variable diresidues (RVD). The identity of the RVDs determines the specificity of the target genes in the host, where each repeat binds a single nucleotide in the promoter of the target gene. The specific sequence of nucleotides to which a TAL effector binds is known as the effector binding element (EBE) in the promoter of the target gene.

The Xoo– rice model reflects highly specialized example of evolutionary cycle between TAL effector and TAL effectors-associated susceptibility, and plants have evolved recessive and dominant forms of resistance in response to TAL effectors (Figure 2). One form of recessive resistance rice has evolved is changes in EBEs of the S genes, disrupting the binding of the cognate TAL effectors and preventing S gene

expression (White and Yang 2009). The R gene *xa13* is a recessive resistant allele of the dominant S gene *SWEET11*. *SWEET11* is induced by effector PthXo1 during bacterial blight disease of rice caused by Xoo (Yang *et al*., 2006). In recessive resistance, however, PthXo1 cannot bind to the EBE of *xa13*, and thus, cannot induce *OsSWEET11*. Similarly another TAL effector PthXo2 cannot induce the *S* gene *OsSWEET13* in rice cultivars Nipponbare and Kitake due to a single nucleotide change in the promoter, which, otherwise, is induced by PthXo2 in *indica* rice (Zhou *et al*., 2015). Recessive resistance, however, is prone to be defeated any time by emerging TAL effectors, and alternate TAL effector genes, which target alternate S genes are found in extant bacterial populations. For example, the strains harboring genes for TAL effectors, Avrxa7 or PthXo3, induce an alternate *S* gene, *SWEET 14* (Antony *et al*., 2010).

Some plants have adapted to TAL-dependent induction of host genes by acquisition through evolution of a form of ETI (Figure 1). Like the Avr–R interaction, the expression of a host gene, known as executor or terminator (T) gene, results in a HR and is dependent on binding of specific TAL effectors to the EBEs in the promoter of corresponding T genes. The central repeat region, C-terminus activation domain and nuclear localization signals of a TAL effector are indispensable for activation of T genes (Gu *et al*., 2005). Five of the T genes that have been cloned include *BS3* and *BS4C* from pepper, and *Xa27*, *Xa10* and *Xa23* from rice (Pierre *et al*., 2000; Strauss *et al*., 2012; Gu *et al*., 2004; Gu *et al*., 2008; Wang *et al*., 2014). While BS3 is a large protein with proposed catalytic function, the latter four proteins are relatively smaller in size and less than 200 amino acid residues in length. T proteins, with the exception of BS3, are

unique and not related to any other type of R proteins or known functional proteins in plants. The small T proteins, however, do possess transmembrane and signal domains as common features (Zhang *et al*., 2015). This project sought to develop a bioinformatics prediction approach to detect candidate T proteins in additional plant species.

Chapter 2. Introduction

Multiple TAL effectors are injected by Xoo into the plant cells via T3SS (Zhu *et al*., 2000). The effectors are localized to the host nucleus, guided by the nuclear localization signals, where the effectors bind to the EBEs and transcriptionally activate the cognate target genes (White and Yang, 2009). The binding code between a TAL effector and EBE (Boch *et al*., 2009; Moscou and Bogdanove, 2009) indicates that the RVD in each of the 33-35 amino acid central repeats of a TAL effector preferentially binds to a specific nucleotide of the EBE. This discovery led to the synthesis of custom designed TAL effectors for gene regulation and its derived TAL effector nucleases (TALENs) for genome editing (Li *et al*., 2011; Morbitzer *et al*., 2010; Weber *et al*., 2011).

The evolution of TAL effector-mediated dominant resistance is essentially based on EBE sequences located upstream of the T genes. The *Bs3* gene in pepper and the *Xa27* gene in rice are T genes that are transcriptionally activated by recruiting AvrBs3 and AvrXa27 TAL effectors to bind to the EBEs of the respective genes (Gu *et al*., 2005; Romer *et al*., 2007). *Bs3* and *Xa27* mediated resistance occurs only in the presence of TAL effectors AvrBs3 and AvrXa27, respectively. Another T gene, *Bs4C*, was identified in pepper through RNA-seq study, and is exclusively elicited by TAL effector AvrBS4 via AvrBs4 EBE in the *Bs4C* promoter (Strauss *et al*., 2012).

Since plant R genes generally mediate recognition of the effector proteins, microbial effectors can be used to screen germplasm collections to identify plant genotypes that contain cognate R genes (Vleeshouwer *et al.*, 2011). Tian *et al.* (2014) cloned and characterized another T gene, *Xa10,* from rice, which is activated by AvrXa10 from Xoo. The Xa10 protein occurs as a hexamer in the endoplasmic

reticulum (ER) and induces HR in rice, *Nicotiana benthamiana* and apoptosis in human HeLa cells via Ca^{++} depletion. Wang *et al* (2015) cloned a new T gene, *Xa23*, through map-based cloning method. Like *Xa27, Xa23* controls broad-based disease resistance to bacterial blight of rice. *Xa23* encodes a 113 amino acid protein that shares 50% sequence identity with XaA10. Xa23 triggers a strong HR in rice, tobacco, and tomato (Wang *et al*., 2015).

The exponential growth of sequence data necessitates the development of bioinformatics tools to manage and analyze the data. Traditional approach of identification of genes of interest is largely limited by the requirement of a priori knowledge about the physiological, biochemical or other functional aspects of the possible candidate genes (Zhu and Zhao, 2007). Digital candidate gene approach (DigiCGA), also known as *in silico* candidate gene approach, is a web resource-based candidate gene identification approach. DigiCGA is primarily based on gene ontology, where public web-based databases are objectively extracted, filtered, assembled and analyzed as per principles of gene ontology. The process involves complex statistical methods to computationally identify potential genes of specific interest and validate by actual association (Zhu and Zhao, 2007).

The key roles of bioinformatics for plant improvements include repository submission of sequences of interest and rational annotation of the sequences (Vassilev *et al*., 2005). T-genes and their products provide a window into proteins that trigger defense responses in plants and may provide a rich resource for the design of novel R gene strategies. Search for new T genes using bioinformatics tools is, therefore, needed to create a repository of T genes as a resource to be tested in novel strategies.

An attempt was made to establish a reference set of sequences to be used as a model for R genes that usually cluster in a genomic region with high number of homologs and pseudogenes (Gururani *et al*., 2012). Efforts are being made to develop efficient software applications, including the construction of plant resistance genes database (PRGdb), an open resource that hosts 16000 known and putative R genes belonging to 192 plant species (Sanseverino *et al*., 2012). The user-friendly interface consists of 73 manually curated R genes, 6308 putative R genes collected from NCBI and 10463 computationally predicted putative R genes.

The origin of most T genes is unknown. To create a pool of candidate T genes, the structural features of four T genes were used as a reference. These structural features were transmembrane (TM) domain, secretion signal peptide domain and < 200 amino acid residues size using computational analysis. Additional criteria of non-expressed genes or low expressed genes were also used to discover candidate T genes. Soybean was used as test plant species with the following objectives:

(1) Mining for candidate T proteins;

and

(2) Annotation for functional prediction of the retrieved candidate T proteins.

Two programming languages, Perl and R, were used as tools to achieve the first objective and public databases were used to accomplish the second.

Chapter 3. Materials and methods

Screen of potential soybean terminator genes

To mine for the candidate T genes in *Glycine max*, three criteria were chosen, which were based on the common features of majority of T genes including *Xa10, Xa23, Xa27* in rice and *Bs4C* in pepper. The three characteristic features are: possessing (1) TM domain (2) signal peptide cleavage domain and (3) protein length < 200 amino acids. The public database did not contain information about protein length. For this purpose, an in-house Perl code was used to select small proteins with <200 aa was designed. An additional criterion of no or low expressed genes was also set to search for candidate T proteins. Generally, signal domain consists of 5-30 amino acids peptide present at the N-terminus of the majority of newly synthesized proteins that are destined towards the secretory pathway into certain cell organelles, membranes, and extracellular environment (Blobel and Dobberstein, 1975). FASTA sequences of protein data for *G. max* were retrieved from the current release of public database *Gramene* FTP site (ftp://ftp.gramene.org/pub/gramene/CURRENT_RELEASE/data/fasta/glycine_max/pep/Glycine_max.V1.0.25.pep.all.fa.gz).

Homology of candidate genes

To understand the homology of candidate T genes in other plant species, orthologous datasets of 38 plant species were searched in EnsembleGenomes (ftp://ftp.ensemblgenomes.org/pub/release-26/plants/mysql/plants_mart_26/).

Functional annotation and expression data for candidate T genes during multiple developmental stages of soybean

Functional annotation of the candidate soybean proteins was done individually using different public databases (SoyBase.org and LegumeIP Database). These databases retrieved the information for functional annotation of the 161 soybean genes based on sequence relatedness of the 161 genes to the reference genes in the databases for either *Arabidopsis* or *Medicago* species. The resulting annotated genes from these databases were grouped into known, hypothetical and unique genes, respectively and termed collectively as candidate T genes. The expression for the 161 candidate T genes was retrieved from RNA-seq expression database (soybase.org/soyseq/) at different developmental stages of soybean.

Identification of disease and stress responsive candidate T genes

Since the homologs for the candidate T genes were present in different plant species searched, selected plant species were used to validate the approach of the present study. The 161 candidate T genes were searched in the published literature (Kawahara *et al.* 2012; Postnikova *et al*., 2012 and Salvo *et al*., 2014) with respect to expression profiles of rice, *Arabidopsis* and maize in response to biotic and abiotic stresses. The key words used to search for relevant publications were “expression”, “RNA-seq”, “microarray”, “disease resistance”, “resistance”, “defense”, “hypersensitive response”, “programmed cell death”, and “senescence” .

Chapter 4. Results

With Perl programing code, using <200 aa in size as the first feature of set criteria, 19,775 out of 54,174 soybean genes were filtered from the Ensemble database. Data for proteins with signal peptide domain was also available from Ensemble, which, when used as the second filter, reduced the 54,174 soybean genes to 3845 candidate proteins. Using an R program for the identification of transmembrane (TM) domains, the 3845 protein set was further narrowed to 575 proteins containing both secretion signal and TM domains, respectively, but were not necessarily of <200 aa in size. By finding the common members of the 575 protein set and the original 19,775 proteins that were <200 aa in size, 161 proteins were obtained (Figure 3).

The candidate161 proteins were categorized based on predicted functions. The functional annotations of the 161 genes were searched in soybean database Soybase (Soybase.org). Seventy-five of the 161 candidate proteins were functionally predicted in the database based on the presence of related sequences and the respective functions, while 86 were functionally unknown (Figure 4). Of the unknown proteins, 44 were hypothetical proteins, defined here as having orthologs in either *Arabidopsis* or *Medicago* species, and 42 proteins were unique to soybean, defined here as having no detectible sequence relatedness, when 161 candidate T genes were scanned across 38 other plant species databases, including species of the genera *Arabidopsis* and *Medicago*.

The candidate T proteins were categorized into eighteen groups (Figure 5). The predominant classes were defense, stress and wound responsive, arabinogalactan, and membrane-bound, which consisted of 17, 13 and 15 proteins,

respectively. The next major class contained 4 proteins related to epidermal patterning factor and inflorescence deficient in abscission (IDA)-like functions, followed by 3 hydrolase and 3 CLE proteins. Serine-type endopeptidase inhibitor, Yos-like, and catalytic activity proteins were other categories possessing 2 proteins each. Single protein was in the classes of pectin lyase-like, AWPM-19, ATP dependent RNA helicase and bifunctional inhibitor/lipid transferase/seed storage 2S albumin superfamily proteins, respectively (Figure 5, Table 1).

By function, T genes are lethal to the host cells and tissues, and the function definition leads us to hypothesize that T genes should not be expressed or, if expressed, should be expressed at low levels during the normal plant development. To find out expression profile of 161 target genes, the expression database at Soyseq during normal development was searched, resulting in the identification of 94 out of 161 genes in the database, of which 9 were categorized as non-expressed, 36 were low expressed, 11 were moderately expressed, and 38 were highly expressed genes, respectively (Table 4). The reads per kilobase per million (RPKM) for these categories were 0, 1-5, 6-12 and 13-1061, respectively (Figure 6).

The non-expressed category included four functionally predicted, or known, genes and five unknown genes, the latter consisting of genes predicted to encode one unique and four hypothetical proteins. The predicted known proteins were GTP cyclohydrolase, cytochrome b 561, NOD 26-like intrinsic, and tetratricopeptide repeat candidate T proteins (Table 5). The low expressed candidate T proteins consisted of 11 known, 10 unique and 15 hypothetical proteins. The predicted known proteins were CLAVATA3, rhodanese, phytosulfokine 4 precursor, ethylene/abscission responsive

transcription factor, serine endopeptidase inhibitor (2 proteins), peroxidase, reductase, photosynthesis-associated YCF and major facilitator proteins (2 proteins). Taken together, non-expressed and low expressed proteins amounted to 45, with 15 known, 19 hypothetical and 11 unique proteins (Figure 7).

Sequence relatedness and structural conservation of genes across different species increase the confidence for functional prediction of the genes. The 161 candidate T genes were searched across 38 other plant species also including rice and maize in addition to *Arabidopsis* and *Medicago* species. The search resulted into either different number of orthologs found in each of rice, *Arabidopsis* and maize species or an ortholog present in one species but not in another, which gave different frequencies of soybean orthologs in each of these three plant species (Table 7)**.** Subsequently, related genes of rice, *Arabidopsis*, and maize were identified in studies of biotic stresses or non-normal plant development. The retrieved literature included studies of pathogenic fungal interaction with rice (Kawahara *et al*., 2012; Table 8), viral interaction with *Arabidopsis* (Postnikova *et al*., 2012; Table 9) and somatic embryogenesis, an unusual and non-normal development procedure (Salvo *et al*., 2014; Table 10). Of the 161 target genes with expression data from these experiments, 15 genes were found in rice, 8 genes were found in *Arabidopsi*s and 5 genes were found in maize. Two proteins, plantacyanin and rhodanese were common in rice, *Arabidopsis,* and maize. Rice shared 5 additional genes with *Arabidopsis* and 2 additional genes with maize. The rice proteins shared with maize were defense-related, including chitin-responsive ring/u box and dirigent candidate T proteins, respectively (Table 11, Figure 8).

Chapter 5. Discussion

Based on the functional domains, R genes have been grouped into different classes (Gurunani *et al*., 2012). T genes, however, represent a distinct class with regards to amino acid sequence homology, and it is impossible to place T genes in a separate class solely on the basis of sequence similarities. At the same time, T genes, with the exception of *Bs3*, do have structural similarities, including transmembrane domains, secretion signal domains, and consist of relatively short peptides. The present study is an effort to identify candidate T genes based on the structural similarities. Here, an arbitrary size limit of 200 aa was chosen for analysis. The analysis will benefit an understanding of relatively small predicted proteins for which there is little functional information available.

T genes provide a potential resource for broad and durable resistance, though identification of T genes is a challenging task. No T gene has been cloned from soybean. Another approach to look for the genes of interest is based on structural similarities (Zhu and Zhao, 2007). However, looking at the shortcomings of a single strategy, a combined and efficient strategy to mine the genes of interest with high degree of confidence is needed. In the present study, *in silico* computational approach, which combined a comparative genomics approach with a genome–wide expression approach, was used to mine the candidate T genes from soybean. The combined approach provided a structure informed approach to predict candidate T genes and enabled the systemic identification of genes that fit the criteria (Glazier *et al*., 2002). The approach extracts, filters, assembles and analyzes all possible information from public databases with respect to gene(s) of interest.

Expression profiles of genes help sets of genes responsible for a trait because of a direct relationship between a trait and gene(s) expression. RNA-seq, a powerful approach for transcriptome profiling, is being used to measure genes expression in both animals and plants at various developmental stages (Mortazavi *et al*., 2008; Severin *et al*., 2010). The expression is quantified in terms of the number of transcript copies, also known as reads. The number of reads is a direct measure of sequencing depth (Wang *et al*., 2009). The expression of T genes causes HR and cell death. It is logical that T gene should be not expressed, or, if expressed, expression should be lower than the threshold that is lethal for normal plant development. The genes with no or low level expression, here based on number of RNA-seq reads across all tissues during normal development of soybean, were also searched as candidate T genes, which resulted into different number of genes from predicted known, hypothetical and unique candidate T genes in the present study. Only unique genes with no/low expression criterion during normal development were considered potential candidates for T gene, here termed as type 1 candidate T genes. It is important to note that the no/low expression of candidate T genes was only under normal developmental conditions. When challenged with incompatible pathogens (Kawahara *et al*., 2012) or when grown in conditions other than normal (Salvo *et al*., 2014), the level of expression of defense/stress related genes in plants can be elevated. It is plausible that under additional conditions, a resistant/tolerant host would elevate the expression of defense/stress-related proteins, corroborating the approach of the present study to look for the no/low expressed genes during normal development conditions when narrowing down the search for candidate T genes.

Plants express sets of genes, including genes with no/low expression in spite of having important functions (Xiao *et al*., 2010). It is possible that no/low expressed genes are minimally required for normal functions but evolutionary preserved by the plants for some unusual functions, including HR to pathogens invasion, which obviously is a state of emergency. Certainly, there are genes or set of genes, which are expressed in a tissue-dependent manner. For example, *SWEET11*, which is an S gene and not a T gene, is expressed at a low level in most other tissues, but highly expressed in pollens during the normal development of rice (Yang *et al*., 2006). Genes with such expression patterns were not considered as candidate T genes in the present study. In the present study, the expression status of 94 out of 161 genes was available during normal development of the soybean. The 94 genes when overlapped with 42 unique candidate T genes resulted in 11 candidate T genes that were not only unique but also not/low expressed and, hereafter, will be referred to as type I candidate T genes. The 11 type I candidate T genes will serve as a subset of genes to be used for validation of T genes. The subset of 11 genes can further expand if the expression status of the remaining 67 genes is known, which include 18 more unique genes (Table 6, Figure 9). Tissue-specific expression analysis using real-time qPCR during normal soybean development can be the next step to know the expression status of 67 genes, or at least the 18 more unique genes. Though search for type I candidate T genes to pinpoint T genes was the priority in this study, the expression status of 15 functionally known but not/low expressed candidate T genes was important to know as to how these genes behave during pathogen infections or abiotic stress conditions. For example, rhodanese, a member of the 15 known genes category, which was not/low expressed

during normal growth conditions of soybean, increased in expression in response to fungal-rice, viral-*Arabidopsis* interactions, or during somatic embryogenesis, an additional growth condition other than normal, of maize (Figure 8). Thus *in silico* approach can serve a strong jumping off point to hunt down genes of interest and funnel these genes down according to priorities of objectives.

To validate that the type I candidate T genes are terminator genes, genes structure of these candidate T genes needs to be analyzed for the presence of complete open reading frame and proper regulatory elements, respectively. A localized expression of the genes, which fulfills the above criteria, is therefore needed to know whether the type I candidate T genes cause any HR. Designer TALe (dTALe), a customized transcription factor developed, is useful to induce virtually any gene in the genome. Based on the modular code assembly, 12^{th} and 13^{th} aa of each repeat in a dTALe tandemly bind to a single nucleotide of EBEs in the promoters of the candidate T genes (Boch *et al*., 2009; Moscou and Bogdanove, 2009), leading to predicted expression of type I candidate T genes and consequently, an anticipated HR phenotype. Thus, *in silico* approach followed by customized expression of the candidate genes present a promising recipe to discover some new T gene members from the abstract canvas of candidate T genes in the present study.

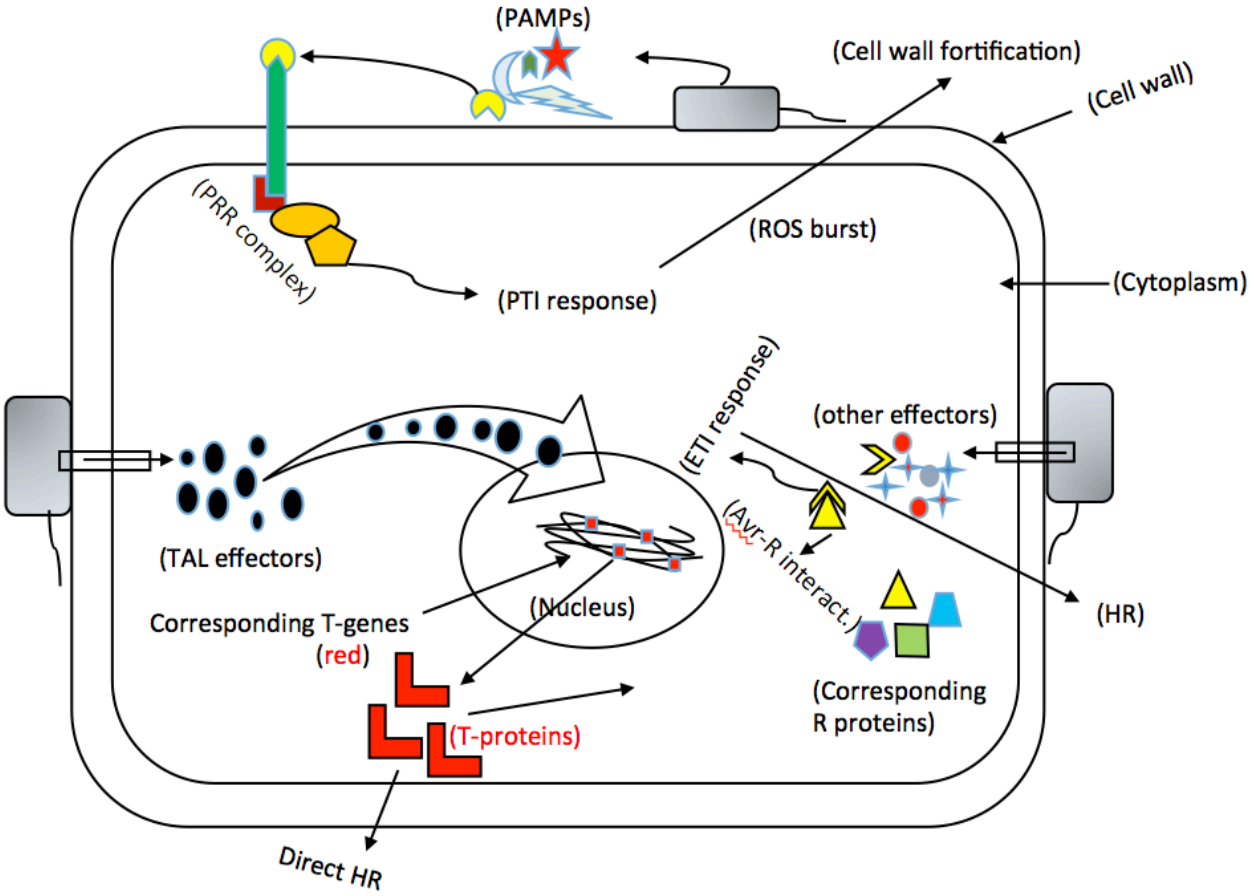

Figure 1. An overview of plant defense mechanisms against bacterial pathogens. The mechanisms include PAMPs-, ETI-, and TALe-mediated ETI-associated forms of defenses in plants. T, terminator; R, resistance; Avr, avirulence.

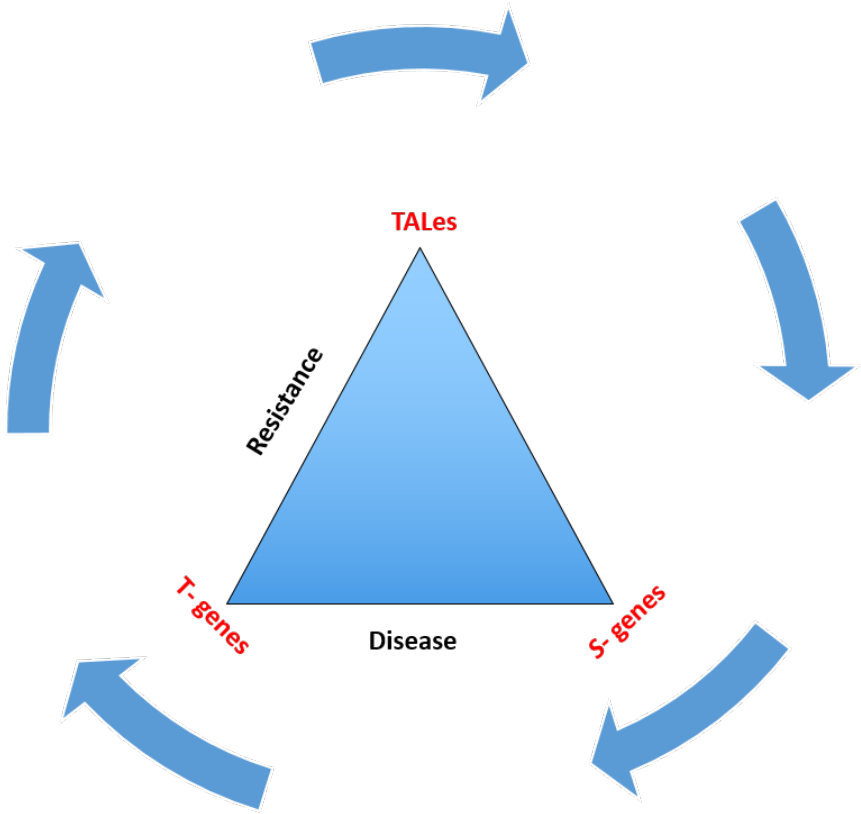

Figure 2. Evolutionary cycle between TALes, S-genes and T-genes. Disease condition exerts pressure on plants to select a resistance gene from gene pool, while host resistance drives bacteria to adopt new TALes for host susceptibility.

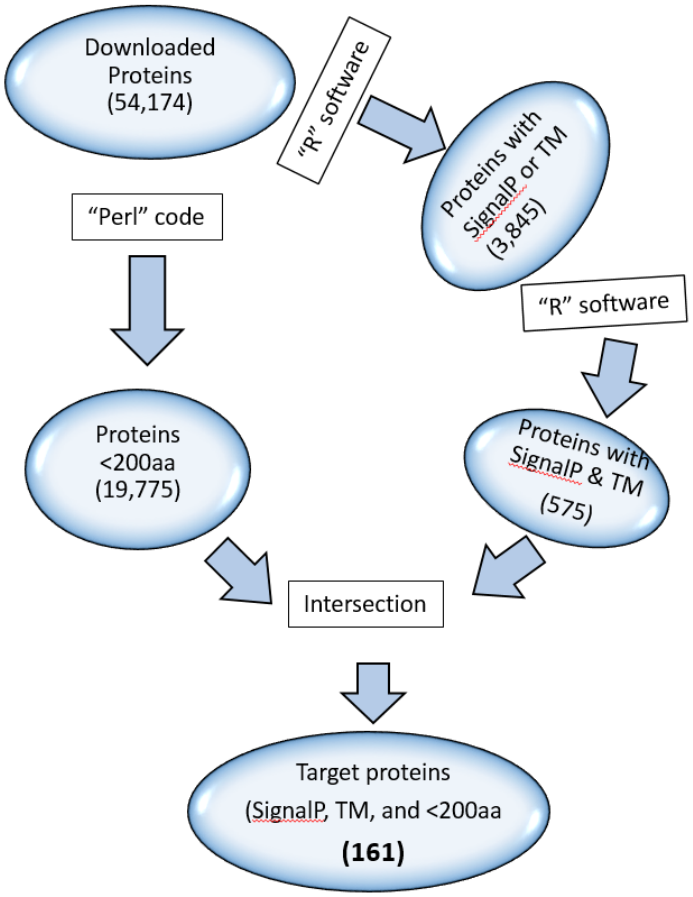

Figure 3. Filtering strategy for candidate T genes.

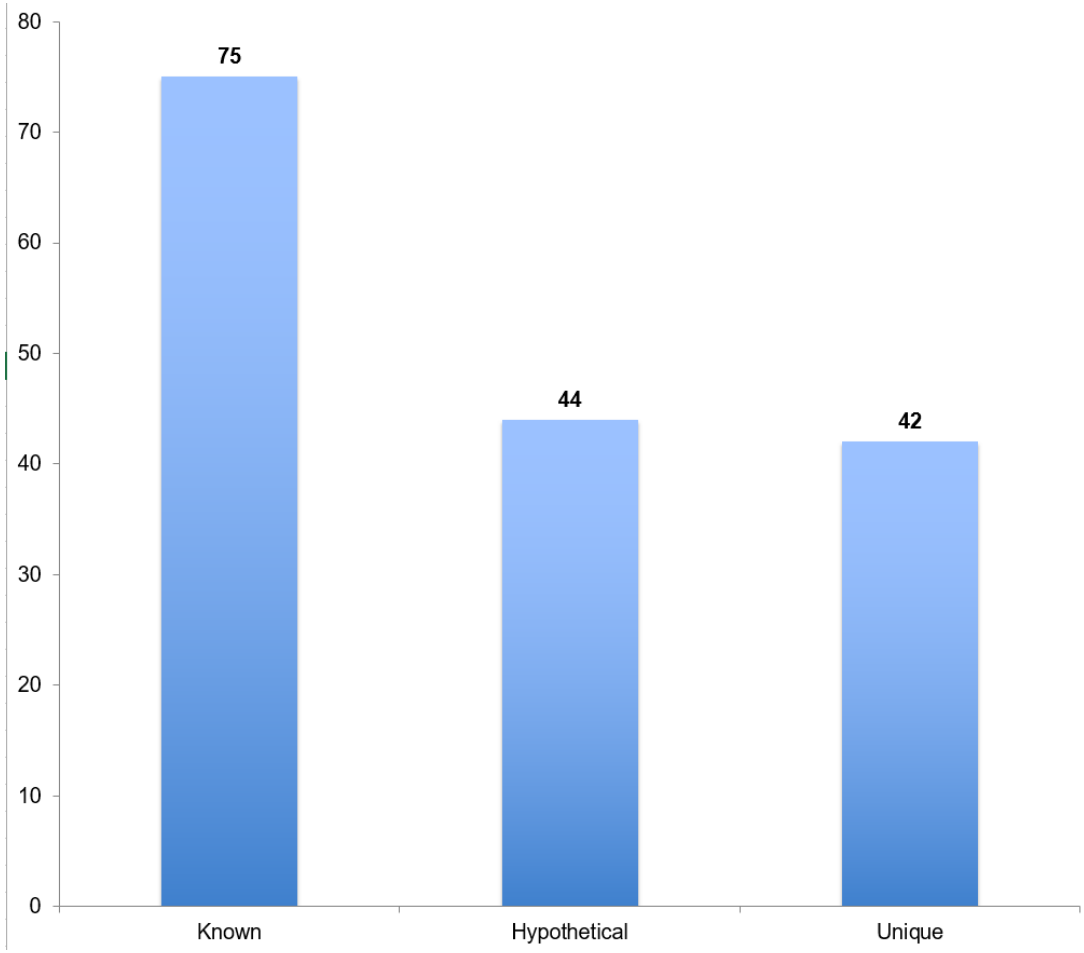

Figure 4. Sequence relatedness-based annotation of candidate T genes in soybean. Known, hypothetical and unique include genes with predicted functions, genes with no predicted functions but known orthologs in 38 other plant species searched, and genes unique to soybean with no predicted functions and no orthologs present in 38 other plant species respectively.

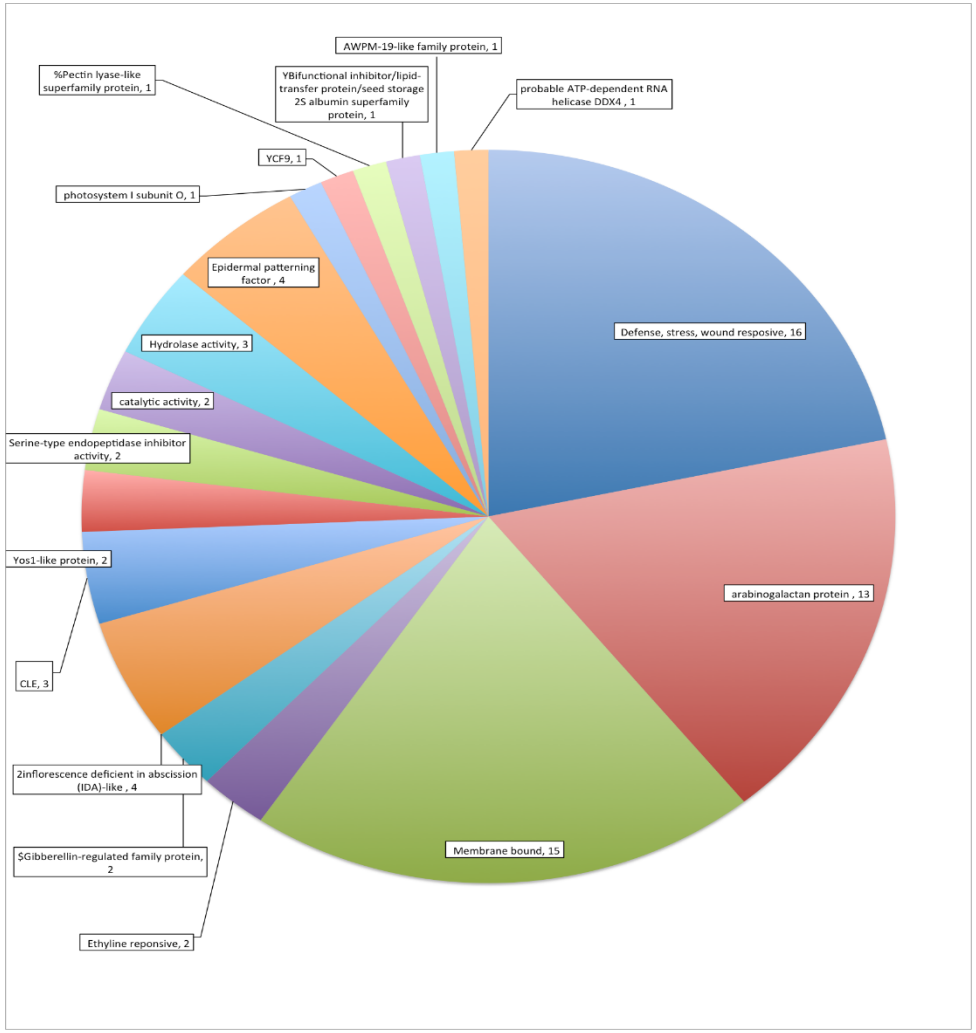

Figure 5. Categorization of functionally predicted candidate T genes. Genes related to defense/stress/wound constitute the largest category of the predicted genes.

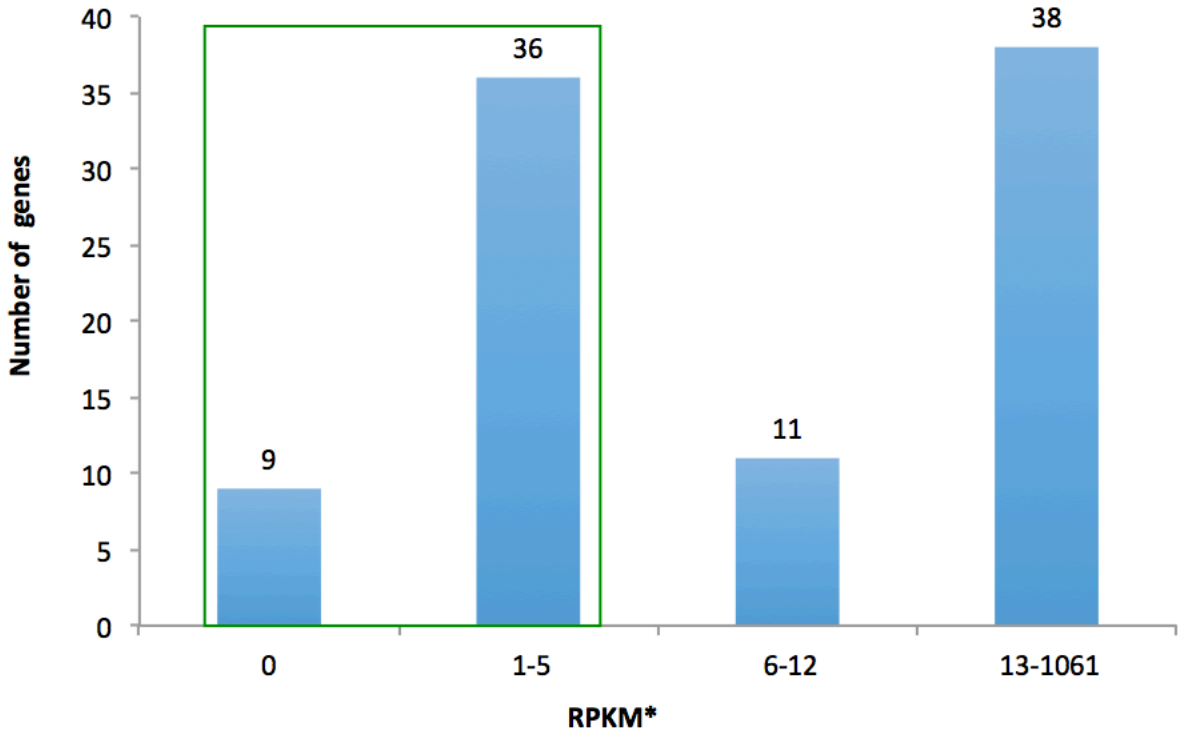

Figure 6. RNA-seq based categorization of 94 out of 161 candidate T genes during normal soybean development (Severin *et al*., 2010). Bars in the green boxed area represent non-expressed and low-expressed candidate T genes.
*****RPKM, reads per kilobase of transcripts per million of reads

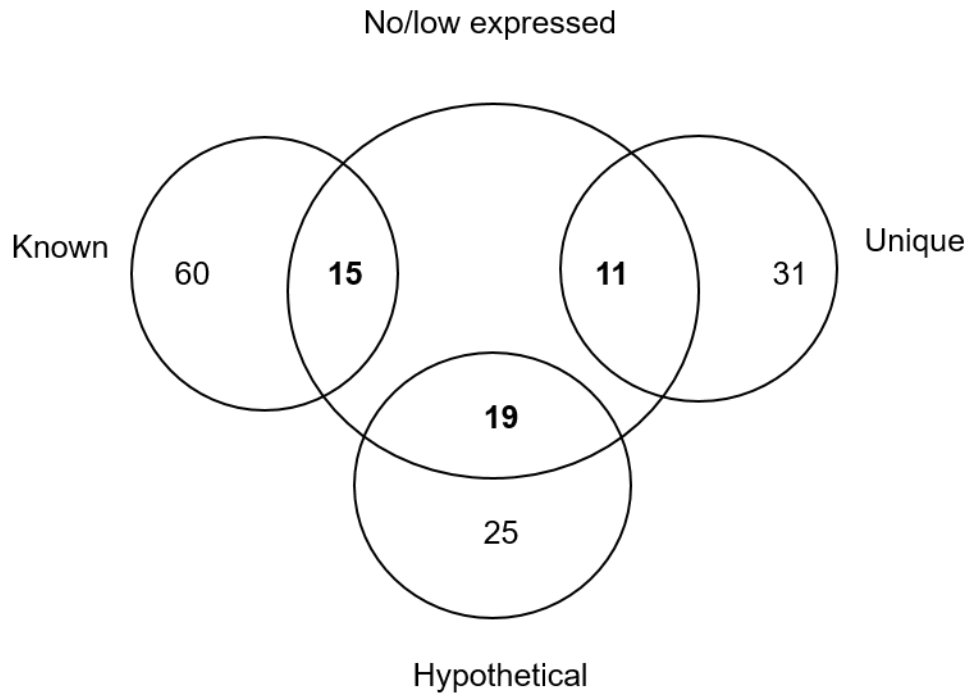

Figure 7 Overlap of expression and functional annotation of candidate T genes from soybean. The 11 unique and non/low expressed genes are termed here as type I candidate T genes.

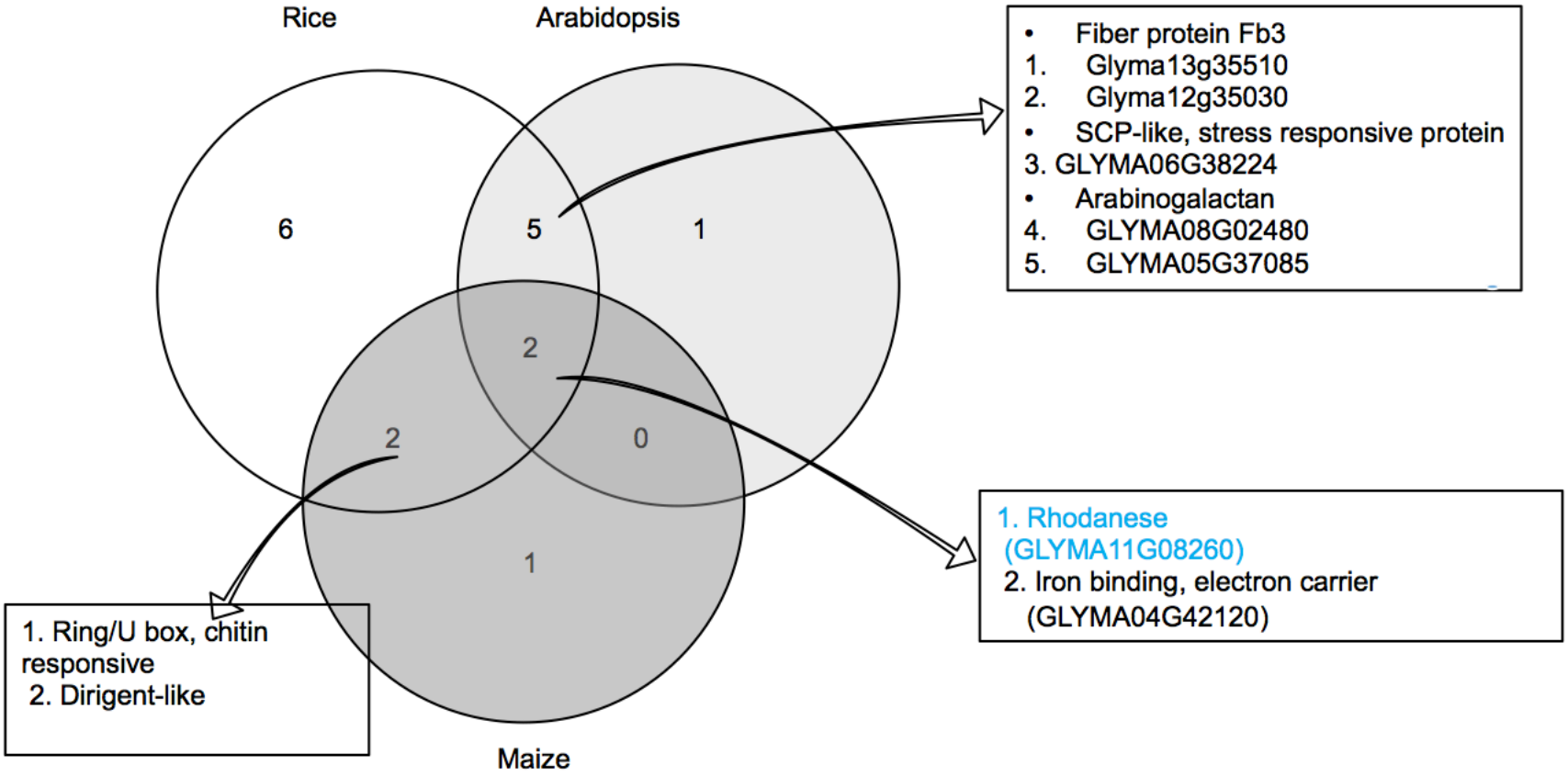

Figure 8. Literature based functional and expression analysis of soybean candidate T gene homologs in rice, *Arabidopsis* and maize. During fungal-rice, viral-*Arabidopsis* interactions, and additional than normal conditions for maize growth, the expression of rhodanese (blue font) elevated, opposite to no/low expression during normal growth conditions.

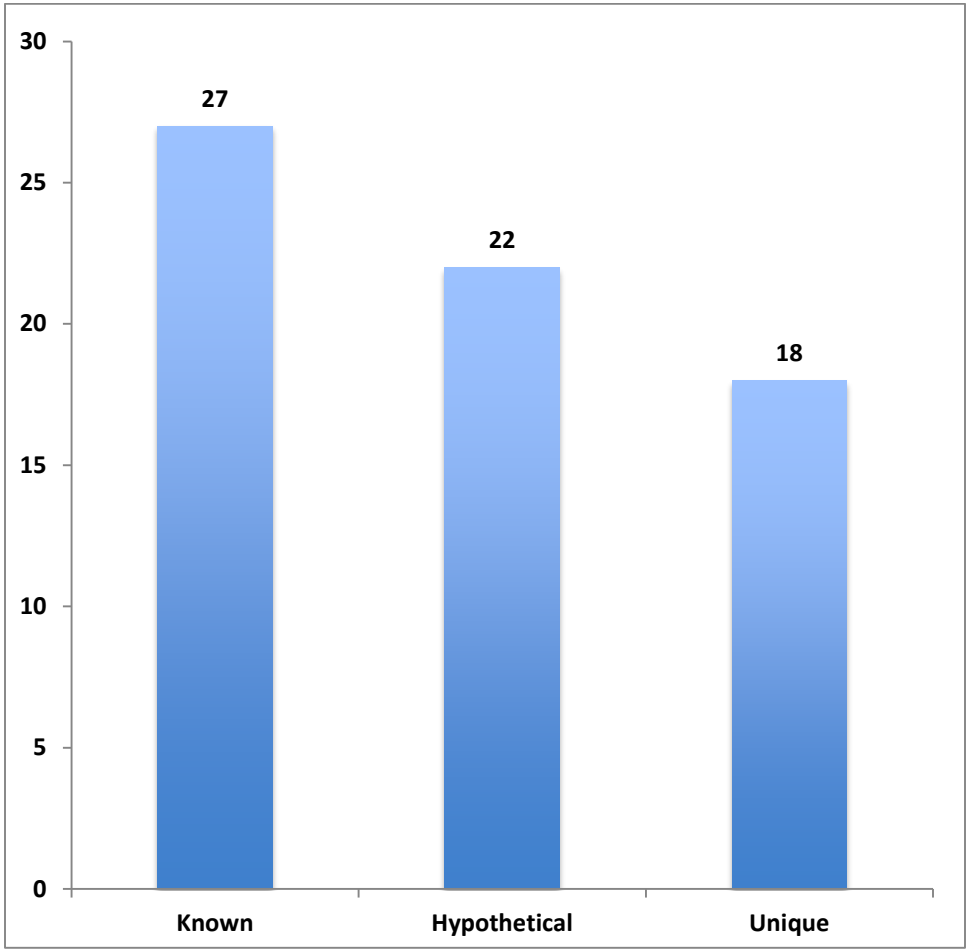

Figure 9. Categorization of 67 candidate T genes that were not present in RNA-seq expression database. Tissue-specific genes expression data was analyzed during normal growth of soybean (Severin *et al*., 2010).

Table 1. Predicted functions for known candidate T genes from soybean.

SN[a]	Gene ID[b]	Gene ontology/Annotation	Cellular localization
1	GLYMA01G02310	Major facilitator superfamily protein |	
2	GLYMA01G11180	YCF9/ Photosystem II	
3	GLYMA01G22751	Inflorescence deficient in abscission (IDA)-like	
4	GLYMA01G31230	Catalytic activity/gamma interferon responsive lysosomal thiol (GILT) reductase family protein	
5	GLYMA02G14821	Inflorescence deficient in abscission (IDA)-like 1	
6	GLYMA02G26610	Defense response/disease resistance-responsive (dirigencandidate T protein) family protein	
7	GLYMA02G27131	Catalytic activity/Pectin lyase-like superfamily protein	Extracellular region
8	GLYMA02G44620	CLAVATA3/ESR-RELATED 22	
9	GLYMA02G45441	Cytochrome b561/ferric reductase transmembrane protein family	Chloroplast, integral to membrane
10	GLYMA03G14970	Yos1-like protein	
11	GLYMA03G28125	CLAVATA3/ESR-RELATED 22	
12	GLYMA04G05271	Arabinogalactan protein 15	
13	GLYMA04G11890	GTP cyclohydrolase I activity/GTP cyclohydrolase I	
14	GLYMA04G41720	Arabinogalactan protein 16	
15	GLYMA04G42120	Response to abscisic acid stimulus, copper ion binding/plantacyanin	
16	GLYMA04G42480	Acetyl-CoA metabolic process	
17	GLYMA05G07900	Receptor binding, apoplast/CLAVATA3/ESR-RELATED 13	
18	GLYMA05G13660	Hydrolase activity, apoplast/xyloglucan endotransglucosylase/hydrolase 15	
19	GLYMA05G32301	Inflorescence deficient in abscission (IDA)-like 2	Extracellular region
20	GLYMA05G26513	CLE31 protein	

[a] - SN indicates arbitrarily assigned sequence number

[b] - Protein ID in soybase and legumeIP databases

21	GLYMA05G37085	Arabinogalactan protein 14	
22	GLYMA06G13060	Arabinogalactan protein	
23	GLYMA06G40291	Arabinogalactan protein	
24	GLYMA06G46010	Transmembrane protein 97, predicted	
25	GLYMA06G43681	CLAVATA3/ESR-RELATED	
26	GLYMA07G09600	Cellular response to iron ion starvation/	
27	GLYMA08G02480	arabinogalactan protein 16	
28	GLYMA08G09950	Cytochrome b561/ferric reductase transmembrane protein family	integral membrane protein, mitochondrion/
29	GLYMA08G17190	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	integral membrane protein, mitochondrion
30	GLYMA08G29500	Regulation of defense response, response to arsenic-containing substance/NOD26-like major intrinsic protein 1	
31	GLYMA08G40260	Protein binding, response to gibberellin stimulus/Gibberellin-regulated family protein	
32	GLYMA09G07481	EPIDERMAL PATTERNING FACTOR-like protein	
33	GLYMA09G25983	CLE09 protein	
34	GLYMA09G28700	Serine-type endopeptidase inhibitor activity	extracellular region
35	GLYMA09G28730	Serine-type endopeptidase inhibitor activity	
36	GLYMA10G03050	Tetratricopeptide repeat (TPR)-like superfamily protein	
37	GLYMA10G06470	Membrane magnesium transporter	ER, mitochondrion
38	GLYMA10G12190	Cellular membrane fusion/Cornichon family protein	
39	Glyma10g12370	arabinogalactan protein 23	
40	GLYMA10G33941	RING/U-box superfamily protein	
41	GLYMA10G34960	Response to ethylene stimulus, abscission, transcription factor import into nucleus/Putative membrane lipoprotein	Extracellular region
42	GLYMA10G35010	AWPM-19-like family protein	Extracellular region
43	GLYMA10G37105	Inflorescence deficient in abscission (IDA)-like	Mitochondrion
44	GLYMA10G40190	Phytosulfokine 4 precursor, cell differentiation	Extracellular region, GA

45	GLYMA11G08260	Rhodanese/Cell cycle control phosphatase superfamily protein, systemic acquired resistance, salicylic acid mediated signaling pathway, Aging	Chloroplast
46	GLYMA11G19501	Transmembrane amino acid transporter family protein	PM
47	GLYMA12G09731	Photosystem I subunit O	Mitochondrion, Chloroplast
48	GLYMA12G10690	Transmembrane protein 97, predicted	
49	GLYMA12G32996	Epidermal patterning factor 1	Vacuole
50	GLYMA13G08400	Arabinogalactan protein 16	
51	GLYMA13G36831	CLE14 protein	
52	GLYMA13G37491	Epidermal patterning factor 1	
53	GLYMA14G03371	Cytochrome b561/ferric reductase transmembrane protein family	Chloroplast
54	GLYMA14G06841	Yos1-like protein	Mitochondrion
55	GLYMA14G09710	Glycosyl hydrolase family 35 protein	
56	GLYMA14G31650	arabinogalactan protein 16	PM
57	GLYMA15G32285	Seven transmembrane MLO family protein, calmodulin binding, defense response	Mitochondrion, PM
58	GLYMA15G36085	RING/U-box superfamily protein, defense response, response to chitin	PM
59	GLYMA15G41860	Peroxidase superfamily protein, heam binding	
60	GLYMA16G04471	Arabinogalactan protein 14	
61	GLYMA16G10175	Probable ATP-dependent RNA helicase DDX4	
62	GLYMA16G11475	NADH dehydrogenase subunit J, NADPH oxidoreductase activity	
63	GLYMA16G12146	Arabinogalactan protein 14	
64	GLYMA16G26965	Arabinogalactan protein 14	
65	GLYMA17G13100	CLAVATA3/ESR-RELATED 13	Extracellular region
66	GLYMA17G22121	Signal transduction histidine kinase, hybrid-type, ethylene sensor	
67	GLYMA18G17490	Gibberellin-regulated family protein	
68	GLYMA19G28574	ATP synthase subunit C family protein	
69	GLYMA19G28950	Arabinogalactan protein 14	

70	GLYMA20G05690	Major facilitator superfamily protein, transmembrane transporter activity	Chloroplast
71	GLYMA20G22351	Major facilitator superfamily protein	PM
72	GLYMA20G28830	EPIDERMAL PATTERNING FACTOR-like protein	
73	GLYMA20G29721	SNARE-like superfamily protein, TRANSPORT	
74	GLYMA20G32601	Inflorescence deficient in abscission (IDA)-like 1	
75	GLYMA20G33661	RING/U-box superfamily protein	

Table 2. Hypothetical candidate T genes from soybean.

SN	Gene ID	Function	Hypothetical	Function
1	GLYMA01G45370	Uncharacterized protein	AT4G14723	Unknown protein
2	GLYMA17G36320	Uncharacterized protein	MTR_1g020020	Unknown protein
3	GLYMA05G28100	Uncharacterized protein	AT5G08391	Protein of unknown function (DUF 3339)
4	GLYMA06G07900	Uncharacterized protein	AT4G31130	Protein of unknown function (DUF1218)
5	GLYMA06G45730	Uncharacterized protein	AT1G80133	Unknown protein
6	GLYMA07G06290	Uncharacterized protein	AT1G80134	Protein of unknown function (DUF 3339)
7	GLYMA08G06680	Uncharacterized protein	AT4G21310	Protein of unknown function (DUF1218)
8	GLYMA10G41700	Uncharacterized protein	AT1G68220	Protein of unknown function (DUF1218)
9	GLYMA11G36910	Uncharacterized protein	AT3G48660	Protein of unknown function (DUF 3339)
10	GLYMA12G11100	Uncharacterized protein	AT2G32280	Protein of unknown function (DUF1218)
11	GLYMA12G12730	Uncharacterized protein	AT1G03700	Uncharacterized protein family
12	GLYMA12G31130	Uncharacterized protein	AT5G09225	Protein of unknown function (DUF1218)
13	GLYMA12G32320	Uncharacterized protein	AT2G32280	Protein of unknown function (DUF1218)
14	GLYMA12G35030	Uncharacterized protein	AT4G27435	Protein of unknown function (DUF1218)
15	GLYMA13G25130	Uncharacterized protein	AT1G61065	Protein of unknown function (DUF1218)
16	GLYMA05G32010	Uncharacterized protein	MTR_8g092340	Unknown protein
17	GLYMA13G35510	Uncharacterized protein	AT4G27435	Protein of unknown function (DUF1218)
18	GLYMA13G38100	Uncharacterized protein	AT2G32280	Protein of unknown function (DUF1218)
19	GLYMA14G34190	Uncharacterized protein	AT3G27030	Unknown protein
20	GLYMA15G06910	Uncharacterized protein	AT4G21310	Protein of unknown function (DUF1218)
21	GLYMA15G06970	Uncharacterized protein	AT1G61667	Protein of unknown function
22	GLYMA16G02930	Uncharacterized protein	AT5G08391	Protein of unknown function (DUF 3339)
23	GLYMA20G25510	Uncharacterized protein	AT1G68220	Protein of unknown function (DUF 3339)
24	GLYMA02G03405	Putative uncharacterized	MTR_5g03565	Putative uncharacterized protein
25	GLYMA05G16420	Uncharacterized protein	AT3G50610	Unknown protein
26	GLYMA06G12301	Putative uncharacterized	AT1G77350	Unknown protein
27	GLYMA06G37230	Uncharacterized protein	AT1G53035	Unknown protein

28	GLYMA06G38224	Uncharacterized protein	AT4G27435	Protein of unknown function (DUF1218)
29	GLYMA08G08950	Uncharacterized protein	MTR_4g115720	Unknown protein
30	GLYMA08G16301	Uncharacterized protein	MTR_3g028340	Unknown protein
31	GLYMA08G25930	Uncharacterized protein	AT1G09645	Unknown protein
32	GLYMA10G42520	Uncharacterized protein	MTR_1g111990	Unknown protein
33	GLYMA12G35290	Uncharacterized protein	AT1G53035	Unknown protein
34	GLYMA13G35220	Uncharacterized protein	AT1G53035	Unknown protein
35	GLYMA13G39176	Uncharacterized protein	AT5G09225	Unknown protein
36	GLYMA14G08850	Uncharacterized protein	AT5G66815	Unknown protein
37	GLYMA15G08080	Uncharacterized protein	AT1G53035	Unknown protein
38	GLYMA15G42762	Uncharacterized protein	MTR_3g028270	Unknown protein
39	GLYMA07G31330	Uncharacterized protein	AT1G61065	Protein of unknown function
40	GLYMA07G30620	Uncharacterized protein	AT4G21310	Protein of unknown function (DUF1218)
41	GLYMA02G31140	Uncharacterized protein	MTR_1g052125	Unknown protein
42	GLYMA04G29430	Uncharacterized protein	MTR_3g015430	Unknown protein
43	GLYMA05G26021	Uncharacterized protein	MTR_4g115720	Unknown protein
44	GLYMA15G42774	Uncharacterized protein	MTR_3g028540	Unknown protein

Table 3. Unique candidate T genes from soybean.

S.N.	Gene ID	Function
1	GLYMA01G21521	Unknown
2	GLYMA02G37041	Unknown
3	GLYMA02G39725	Protein of unknown function (DUF1218)
4	GLYMA03G21909	Putative uncharacterized protein
5	GLYMA04G07830	Uncharacterized protein
6	GLYMA04G40810	Uncharacterized protein
7	GLYMA04G40823	Unknown
8	GLYMA04G40861	Putative uncharacterized protein
9	GLYMA06G23157	Unknown
10	GLYMA07G11035	Unknown
11	GLYMA07G16091	Uncharacterized protein
12	GLYMA08G07360	Uncharacterized protein
13	GLYMA08G46370	Unknown
14	GLYMA09G05645	Unknown
15	GLYMA09G05750	Uncharacterized protein
16	GLYMA09G06137	Unknown
17	GLYMA09G31094	Uncharacterized protein
18	GLYMA11G14061	Uncharacterized protein
19	GLYMA12G06030	Unknown
20	GLYMA12G07550	Unknown
21	GLYMA12G16173	Unknown
22	GLYMA13G01671	Unknown
23	GLYMA13G37370	Unknown
24	GLYMA14G12003	Unknown
25	GLYMA15G16860	Uncharacterized protein
26	GLYMA15G21980	Uncharacterized protein

27	GLYMA15G42725	Unknown
28	GLYMA15G42787	Uncharacterized protein
29	GLYMA16G03960	Unknown
30	GLYMA16G32110	Uncharacterized protein
31	GLYMA18G46745	Uncharacterized protein
32	GLYMA19G29690	Uncharacterized protein
33	GLYMA20G23630	Unknown
34	GLYMA20G26620	Unknown
35	GLYMA04G41941	Uncharacterized protein
36	GLYMA12G28680	Uncharacterized protein
37	GLYMA13G32350	Uncharacterized protein
38	GLYMA07G11031	Unknown
39	GLYMA08G16251	Uncharacterized protein
40	GLYMA12G36703	Unknown
41	GLYMA13G31270	Uncharacterized protein
42	GLYMA17G35370	Uncharacterized protein

Table 4. Normalized RNA-seq expression data for soybean development: Normalized reads/kilobase/million[a].

Gene ID	Young leaf	Flower	One cm pod	Pod shell 10DAF	Pod shell 14DAF	Seed 10D AF	Seed 14D AF	Seed 21D AF	Seed 25D AF	Seed 28D AF	Seed 35D AF	Seed 42D AF	Root	Nodule
Glyma01g02310	0	1	0	0	0	0	0	0	0	1	0	1	0	0
Glyma01g11180	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma01g31230	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Glyma01g45370	2	0	0	1	1	0	0	0	0	0	0	0	0	0
Glyma02g26610	12	4	15	8	14	28	93	55	6	2	1	0	6	36
Glyma02g31140	3	2	4	2	0	2	1	1	5	14	26	13	4	0
Glyma02g44620	1	0	1	1	0	1	4	3	2	2	2	1	0	0
Glyma03g14970	4	7	10	9	4	7	7	7	12	8	24	9	13	9
Glyma04g07830	0	1	1	0	1	1	1	1	0	1	0	0	1	3
Glyma04g11890	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma04g40810	59	105	100	98	271	16	10	4	6	2	2	2	201	60
Glyma04g41720	0	105	0	1	0	0	0	0	1	0	0	1	3	0
Glyma04g42120	64	11	69	66	63	0	6	8	3	0	3	0	36	11
Glyma04g42480	62	68	65	80	51	49	56	76	45	27	49	16	56	31
Glyma05g07900	0	0	0	0	0	0	0	0	0	0	0	0	12	0
Glyma05g13660	5	1	3	3	2	0	2	2	1	1	3	2	10	3
Glyma05g16420	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Glyma05g28100	0	0	2	0	0	0	0	0	0	0	0	0	5	0
Glyma05g32010	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma06g07900	3	6	3	2	3	3	4	2	1	2	2	2	2	1
Glyma06g13060	0	146	1	0	1	0	1	1	0	0	0	0	1	47
Glyma06g37230	0	1	0	1	1	0	0	0	0	0	0	0	5	1
Glyma06g45730	0	0	0	1	1	0	1	0	0	0	0	0	1	0
Glyma06g46010	4	6	7	8	6	1	5	3	1	1	1	1	3	1
Glyma07g06290	2	2	2	2	1	3	1	0	1	0	0	0	3	1

Glyma07g09600	18	10	3	16	67	8	9	3	8	2	2	1	11	15
Glyma07g30620	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Glyma07g31330	3	8	5	4	2	2	4	2	1	1	1	0	28	56
Glyma08g02480	165	33	325	249	163	61	172	143	195	115	146	156	53	1061
Glyma08g06680	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Glyma08g07360	0	1	4	10	27	0	6	8	0	3	4	6	31	10
Glyma08g08950	1	0	0	0	0	0	0	0	0	0	0	0	3	0
Glyma08g09950	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma08g17190	0	2	0	0	0	0	0	0	0	0	0	0	98	21
Glyma08g25930	13	9	9	12	13	4	6	5	6	2	8	3	13	11
Glyma08g29500	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma08g40260	0	0	0	0	0	3	18	22	256	442	994	801	0	0
Glyma08g46370	0	1	1	0	0	4	3	1	0	0	0	1	0	0
Glyma09g05750	0	1	0	0	0	0	0	0	0	0	0	0	1	0
Glyma09g28700	0	0	0	0	0	0	0	0	1	0	0	1	0	0
Glyma09g28730	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Glyma10g03050	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma10g06470	5	4	7	10	9	8	6	3	3	2	2	2	10	7
Glyma10g12190	13	12	6	7	4	3	4	4	2	1	2	2	4	3
Glyma10g12370	18	755	0	1	1	0	0	0	1	0	0	1	0	0
Glyma10g34960	0	1	1	0	0	0	0	0	0	0	0	0	0	0
Glyma10g35010	2	2	0	0	0	0	1	1	15	3	15	43	0	0
Glyma10g40190	3	5	3	2	1	0	0	0	0	0	0	0	1	0
Glyma10g41700	18	29	25	20	22	11	16	22	15	11	15	12	6	10
Glyma10g42520	17	14	20	13	8	1	3	7	5	4	4	0	5	2
Glyma11g08260	1	1	1	1	1	1	0	1	0	0	0	0	2	4
Glyma11g36910	0	73	7	4	10	200	41	11	6	10	8	5	6	11
Glyma12g06030	18	15	39	29	8	58	28	28	10	5	3	0	25	2
Glyma12g07550	1	1	3	0	1	0	1	1	0	0	1	0	0	1
Glyma12g10690	7	17	20	22	14	3	9	5	6	3	5	2	23	10

Glyma12g11100	1	1	2	4	3	2	2	1	1	0	0	0	1	2
Glyma12g12730	0	0	0	0	0	0	0	0	0	0	0	0	3	3
Glyma12g28680	4	10	4	6	4	4	2	3	3	2	6	4	8	13
Glyma12g31130	16	4	10	3	1	1	4	6	3	2	4	2	5	9
Glyma12g32320	1	1	0	0	0	0	0	1	0	0	0	0	1	0
Glyma12g35030	22	10	15	15	23	13	14	14	6	3	5	3	17	78
Glyma12g35290	1	7	5	3	2	2	1	2	1	1	1	1	1	1
Glyma13g08400	0	38	0	0	0	0	0	0	0	0	0	0	0	0
Glyma13g25130	2	0	0	0	0	0	0	0	0	0	0	0	0	191
Glyma13g31270	19	20	23	11	12	17	20	28	17	10	27	13	9	7
Glyma13g32350	2	2	2	3	1	1	0	4	1	2	3	0	1	0
Glyma13g35220	3	10	8	4	7	7	6	6	5	5	11	12	8	2
Glyma13g35510	41	18	48	55	66	12	12	14	13	13	14	10	63	9
Glyma13g37370	0	0	1	0	0	16	11	2	8	11	24	11	0	0
Glyma13g38100	2	0	1	1	0	0	0	1	0	0	0	0	1	1
Glyma14g08850	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma14g09710	2	0	0	0	0	3	8	10	9	5	7	3	0	0
Glyma14g31650	0	50	0	0	0	0	0	0	0	0	0	0	0	0
Glyma14g34190	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma15g06910	0	1	0	0	0	0	0	0	0	0	0	0	1	0
Glyma15g06970	0	1	1	1	0	0	1	3	1	1	1	0	1	0
Glyma15g08080	10	14	15	24	21	54	49	25	28	19	42	36	10	9
Glyma15g16860	0	0	1	0	0	1	0	1	1	0	0	0	1	0
Glyma15g21980	0	0	0	0	0	0	0	0	0	0	0	2	0	0
Glyma15g41860	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Glyma16g02930	17	1	3	7	8	9	10	6	2	3	5	4	24	7
Glyma16g03960	6	29	16	18	7	1	4	2	3	1	3	2	14	27
Glyma16g32110	4	4	3	6	1	2	2	2	1	1	2	1	3	2
Glyma17g13100	0	0	0	0	0	0	0	0	0	0	0	0	6	1
Glyma17g35370	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Glyma17g36320	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Glyma18g17490	0	2	5	9	4	4	7	15	63	83	216	221	0	0
Glyma19g28950	12	23	5	8	16	3	2	0	3	3	3	2	218	47
Glyma19g29690	2	13	0	3	4	0	1	1	4	1	6	2	5	1
Glyma20g05690	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Glyma20g23630	0	1	0	0	1	0	1	1	0	0	1	0	0	0
Glyma20g25510	50	29	21	19	16	13	17	15	10	7	11	6	19	9
Glyma20g26620	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Glyma20g28830	2	6	9	8	4	2	3	1	1	0	2	0	8	1

a- Data was obtained from soybase.org/soyseq/ (Severin *et al*., 2010).

Table 5. Not expressed and low expressed (not/low) candidate T genes in soybean[a].

SN	Gene ID	Annotation	Stage
1	Glyma04g11890	GTP cyclohydrolase I activity/GTP cyclohydrolase I	none
2	Glyma05g32010	Hypothetical	none
3	Glyma08g09950	Cytochrome b561/ferric reductase transmembrane protein family	none
4	Glyma08g29500	NOD26-like major intrinsic protein 1/ Regulation of defense response	none
5	Glyma10g03050	Tetratricopeptide repeat (TPR)-like superfamily protein	none
6	Glyma14g08850	Hypothetical	none
7	Glyma14g34190	Hypothetical	none
8	Glyma17g35370	Unique	none
9	Glyma17g36320	Hypothetical	none
10	Glyma01g11180	YCF9	Young leaf
11	Glyma05g16420	Hypothetical	Root
12	Glyma07g30620	Hypothetical	Root
13	Glyma08g06680	Hypothetical	Root
14	Glyma09g28700	Serine-type endopeptidase inhibitor activity	Seed
15	Glyma09g28730	Serine-type endopeptidase inhibitor activity	Seed
16	Glyma15g21980	Unique	Seed
17	Glyma15g41860	Peroxidase superfamily protein	Root
18	Glyma20g05690	Major facilitator superfamily protein	Pod
19	Glyma20g26620	Unique	Seed
20	Glyma05g28100	Hypothetical	Root
21	Glyma08g08950	Hypothetical	Young leaf, Root
22	Glyma09g05750	Unique	Flower, Root
23	Glyma10g34960	Response to ethylene stimulus, abscission, transcription factor import into nucleus/Putative membrane lipoprotein	Flower, Pod
24	Glyma12g12730	Hypothetical	Root, Nodule
25	Glyma15g06910	Hypothetical	Flower, Root
26	Glyma01g02310	Major facilitator superfamily protein	Flower, Seed

27	Glyma01g31230	gamma interferon responsive lysosomal thiol (GILT) reductase family protein	Young leaf, Flower, Pod
28	Glyma01g45370	Hypothetical	Young leaf, Flower, Pod shell
29	Glyma06g45730	Hypothetical	Pod, Seed
30	Glyma12g32320	Hypothetical	Young leaf, Flower, Root
31	Glyma06g37230	Hypothetical	Flower, Pod, Root, Nodule
32	Glyma15g16860	Unique	Flower, Pod, Seed
33	Glyma20g23630	Unique	Flower, Pod, Seed
34	Glyma08g46370	Unique	Flower, Pod, Seed
35	Glyma10g40190	phytosulfokine 4 precursor	Young leaf, Flower, Pod, Root
36	Glyma13g38100	Hypothetical	Young leaf, Pod, Seed, Root
37	Glyma12g07550	Unique	Young leaf, Flower, Pod, Seed, Nodule
38	Glyma04g07830	Unique	Young leaf, Flower, Pod, Seed, Root, Nodule
39	Glyma11g08260	Rhodanese/Cell cycle control phosphatase superfamily protein	Young leaf, Flower, Pod, Seed, Root, Nodule
40	Glyma15g06970	Hypothetical	Flower, Pod, Seed, Root
41	Glyma02g44620	CLAVATA3/ESR-RELATED 22	Young leaf, Pod, Seed,
42	Glyma07g06290	Hypothetical	Young leaf, Flower, Pod, Seed, Root, Nodule
43	Glyma12g11100	Hypothetical	Young leaf, Flower, Pod, Seed, Root, Nodule
44	Glyma13g32350	Unique	Young leaf, Flower, Pod, Seed, Root, Nodule
45	Glyma16g32110	Unique	Young leaf, Flower, Pod, Seed, Root, Nodule

a- The data is a sub-category derived from table 4, which was obtained from soybase.org/soyseq/ (Severin *et al*., 2010).

Table 6. Unique candidate T genes with no expression data available[a]

S.N.	Gene ID	Transcript ID	Function
1	GLYMA12G16173	GLYMA12G16173.1	unknown
3	GLYMA12G28680	GLYMA12G28680.1	Uncharacterized protein
7	GLYMA12G36703	GLYMA12G36703.1	unknown
3	GLYMA13G01671	GLYMA13G01671.1	unknown
10	GLYMA13G32350	GLYMA13G32350.1	unknown
4	GLYMA13G37370	GLYMA13G37370.1	unknown
5	GLYMA14G12003	GLYMA14G12003.1	unknown
6	GLYMA15G16860	GLYMA15G16860.1	Uncharacterized protein
7	GLYMA15G21980	GLYMA15G21980.1	Uncharacterized protein
8	GLYMA15G42725	GLYMA15G42725.1	unknown
11	GLYMA15G42787	GLYMA15G42787.1	Uncharacterized protein
12	GLYMA16G03960	GLYMA16G03960.1	unknown
13	GLYMA16G32110	GLYMA16G32110.1	Uncharacterized protein
14	GLYMA17G36320	GLYMA17G36320.1	unknown
15	GLYMA18G46745	GLYMA18G46745.1	Uncharacterized protein
16	GLYMA19G29690	GLYMA19G29690.1	Uncharacterized protein
17	GLYMA20G23630	GLYMA20G23630.1	unknown
18	GLYMA20G26620	GLYMA20G26620.1	unknown

a- The data is derived by overlapping sequence relatedness-based 42 unique genes with 67 genes for which no expression data was available in soybase.org/soyseq/.

Table 7. Soybean candidate T gene homologs in rice, *Arabidopsis* and maize[a].

Transcript	Rice	Arabidopsis thaliana	Maize
GLYMA01G11180	OS12G0504050	ATCG00300	GRMZM2G374263;GRMZM2G394732;GRMZM2G414660
GLYMA02G26610	OS11G0179500;OS11G0179400;OS11G0179000;OS11G0179700;OS11G0180000;OS11G0464600;OS11G0178800;OS12G0174700	NA	GRMZM2G159503;GRMZM2G320023;GRMZM2G178199;GRMZM2G059706;GRMZM2G132273
GLYMA02G45441	OS05G0565100	AT2G30890	GRMZM2G023133;GRMZM2G024448
GLYMA04G11890	OS04G0662700	AT3G07270	GRMZM2G062420;GRMZM2G106376
GLYMA04G4210	OS03G0850900;OS03G0709300;OS02G0731400;OS06G0266400	AT2G02850	GRMZM2G004160;GRMZM2G339943;GRMZM2G043300;GRMZM2G004106;GRMZM2G177934;GRMZM2G039381;GRMZM5G866053
GLYMA05G1360	OS06G0696500;OS06G0335975	NA	GRMZM2G026980
GLYMA05G28100	OS05G0103300;OS11G0496500	AT5G08391	GRMZM2G339866;GRMZM2G150691
GLYMA05G37085	OS06G0505700;OS05G0217000;OS01G0657000;OS02G0264800	AT5G53250	GRMZM2G020131;GRMZM2G050286;GRMZM2G002420
GLYMA06G07900	OS05G0121000	AT4G31130	GRMZM2G140998
GLYMA06G12301	OS05G0486200	AT1G77350	GRMZM2G131024
GLYMA06G37230	NA	AT1G53035;AT3G15358	NA
GLYMA06G38224	OS10G0495900;OS07G0462200;OS04G0678200	AT1G52910;AT3G15480	GRMZM2G084327;GRMZM2G138870;GRMZM5G872070;GRMZM2G147418;GRMZM2G396540
GLYMA06G45730	OS07G0545100	AT2G32280;AT1G05291	GRMZM2G064547;GRMZM2G358386
GLYMA06G46010	OS03G0646400;OS07G0458500;OS07G0458700	AT2G32380;AT1G05220;AT1G05210	GRMZM2G151651
GLYMA07G06290	OS11G0496500;OS05G0103300	NA	GRMZM2G150691;GRMZM2G339866
GLYMA07G30620	OS07G0545100	AT4G21310;AT1G11500	GRMZM2G358386;GRMZM2G064547
GLYMA07G31330	OS07G0462200;OS04G0678200;	AT1G61065	GRMZM2G147418;GRMZM2G084327

	OS10G0495900		;GRMZM2G138870;GRMZM2G396540;GRMZM5G872070
GLYMA08G02480	OS06G0505700;OS05G0217000;OS02G0264800;OS01G0657000	AT5G53250	GRMZM2G002420;GRMZM2G050286;GRMZM2G020131
GLYMA08G06680	OS07G0545100	AT1G11500;AT4G21310	GRMZM2G064547;GRMZM2G358386
GLYMA08G17190	NA	AT2G18370	NA
GLYMA08G29500	OS01G0112400	NA	GRMZM2G103214
GLYMA09G07481	NA	NA	GRMZM2G077219
GLYMA10G03050	OS03G0357500;OS07G0655800	AT3G63095	GRMZM2G448739;GRMZM2G309995;GRMZM2G063055
GLYMA10G06470	OS12G0566400;OS11G0568633	AT5G03345	GRMZM2G030408;GRMZM2G034511
GLYMA10G12190	OS12G0506400	AT3G12180	GRMZM2G124658;GRMZM2G018885
GLYMA10G33941	OS03G0788100;OS07G0159600	NA	GRMZM2G028188;AC206768.3_FG00RMZM2G135011;GRMZM2G029623;A12976.3_FG003
GLYMA10G41700	OS08G0261100	AT1G68220	GRMZM2G360615;GRMZM2G029184
GLYMA11G08260	OS12G0428000;OS02G0102300;OS04G0249600;OS02G0157600	AT2G17850;AT5G66170	GRMZM2G303374;AC204711.3_FG003;GRMZM2G131087
GLYMA11G19501	OS01G0857400	AT1G08230	GRMZM2G154958
GLYMA12G09731	OS04G0414700	AT1G08380	GRMZM2G001653
GLYMA12G10690	OS07G0458700;OS07G0458500;OS03G0646400	AT1G05210;AT2G32380;AT1G05220	GRMZM2G151651
GLYMA12G11100	OS07G0545100	AT2G32280;AT1G05291	GRMZM2G064547;GRMZM2G358386
GLYMA12G12730	OS01G0237000;OS05G0245300	AT1G03700;AT4G03540	GRMZM2G038780
GLYMA12G31130	OS12G0133600;OS11G0136701	AT5G09225	GRMZM2G064984;GRMZM2G171644
GLYMA12G32320	OS07G0545100	AT2G32280;AT1G05291	GRMZM2G064547;GRMZM2G358386
GLYMA12G32996	NA	AT2G20875	NA
GLYMA12G35030	OS10G0495900;OS07G0462200;OS04G0678200	AT1G52910;AT3G15480	GRMZM5G872070;GRMZM2G084327;GRMZM2G147418;GRMZM2G396540;GRMZM2G138870
GLYMA12G35290	NA	AT1G53035;AT3G15358	NA

GLYMA13G25130	OS10G0495900;OS07G0462200;OS04G0678200	AT1G61065	GRMZM5G872070;GRMZM2G147418;GRMZM2G396540;GRMZM2G084327;GRMZM2G138870
GLYMA13G32350	NA	AT1G61667	GRMZM2G173863;GRMZM2G144180
GLYMA13G35220	NA	AT3G15358;AT1G53035	NA
GLYMA13G35510	OS04G0678200;OS10G0495900;OS07G0462200	AT3G15480;AT1G52910	GRMZM2G396540;GRMZM5G872070;GRMZM2G138870;GRMZM2G147418;GRMZM2G084327
GLYMA13G37491	NA	AT2G20875	NA
GLYMA13G38100	OS07G0545100	AT1G05291;AT2G32280	GRMZM2G358386;GRMZM2G064547
GLYMA13G39176	OS11G0136701;OS12G0133600	AT5G09225	GRMZM2G064984;GRMZM2G171644
GLYMA14G03371	OS05G0565100	AT2G30890	GRMZM2G023133;GRMZM2G024448
GLYMA14G06841	OS03G0331900;OS09G0326850	NA	GRMZM2G140758;GRMZM2G106283
GLYMA14G09710	OS03G0206201	NA	GRMZM2G114232
GLYMA14G34190	NA	AT3G27030;AT5G40970	NA
GLYMA15G06910	OS07G0545100	AT1G11500;AT4G21310	GRMZM2G064547;GRMZM2G358386
GLYMA15G06970	NA	AT1G61667	GRMZM2G144180;GRMZM2G173863
GLYMA15G32285	OS05G0183566;OS05G0183300	NA	GRMZM2G325653;GRMZM2G000376
GLYMA15G36085	OS08G0484200;OS06G0218300;OS09G0468300;OS02G0759400;OS04G0450300;OS02G0572300;OS03G0188200;OS02G0572200	NA	GRMZM2G160966;GRMZM2G400784;GRMZM2G035704;GRMZM2G467187;GRMZM2G015753;GRMZM2G329195;GRMZM2G082653;GRMZM2G105460;GRMZM2G148706;GRMZM2G323013
GLYMA15G41860	OS07G0499500	AT1G30870	GRMZM2G010640
GLYMA16G02930	OS11G0496500;OS05G0103300	NA	GRMZM2G339866;GRMZM2G150691
GLYMA16G10175	NA	AT1G73020	NA
GLYMA16G11475	NA	ATCG00420	GRMZM2G070685;GRMZM2G003230;GRMZM2G394607;GRMZM2G175383;GRMZM2G404025
GLYMA17G22121	NA	AT3G04580	NA

GLYMA20G05690	OS03G0823200	NA	NA
GLYMA20G22351	OS07G0578200	AT3G60070;AT2G44280	GRMZM2G121458;GRMZM2G060762
GLYMA20G25510	OS08G0261100	AT1G68220	GRMZM2G029184;GRMZM2G360615
GLYMA20G28830	OS11G0581700;OS03G0161600	AT3G22820	GRMZM2G166445;GRMZM2G330384;GRMZM2G167812
GLYMA20G29721	NA	AT5G52270	NA
GLYMA20G33661	OS07G0159600;OS03G0788100	NA	AC212976.3_FG003;AC206768.3_FG005;GRMZM2G029623;GRMZM2G135011;GRMZM2G028188

a- Sets of soybean homologs for rice *Arabidopsis* and maize chosen from scanned 38 plant species with 161 candidate T genes with respect to sequence relatedness.

Table 8. Literature-based expression of soybean candidate T gene homologs at the initial stage of fungal infection of rice (Kawahara *et al*., 2012).

Rice gene ID	Soybean gene ID	Compatible[a]	Incompatible[b]
OS03G0709300	GLYMA04G42120	1.4	9
OS03G0850900	GLYMA04G42120	1.1	9.6
OS04G0450300	GLYMA15G36085	1.3	3.8
OS06G0218300	GLYMA15G36085	1.1	3.3
OS06G0505700	GLYMA08G02480	1	4.4
OS06G0505700	GLYMA05G37085	1	4.4
OS10G0495900	GLYMA13G35510	1.7	4.2
OS10G0495900	GLYMA06G38224	1.7	4.2
OS10G0495900	GLYMA13G25130	1.7	4.2
OS10G0495900	GLYMA07G31330	1.7	4.2
OS10G0495900	GLYMA12G35030	1.7	4.2
OS12G0174700	GLYMA02G26610	1.6	14.7
OS12G0506400	GLYMA10G12190	1.7	2.6
OS12G0566400	GLYMA10G06470	1.1	2
OS05G0486200	GLYMA06G12301	1.4	1.9
OS08G0484200	GLYMA15G36085	3.3	12.7
OS09G0326850	GLYMA14G06841	1.7	1.6
OS09G0468300	GLYMA15G36085	2.2	12.9
OS12G0428000	GLYMA11G08260	2.8	3

a- Fold change expression of the genes during compatible fungal-rice interaction relative to water control infiltration
b- Fold change expression of the genes during incompatible fungal-rice interaction relative to water control infiltration.

Table 9. Literature-based expression of candidate soybean T gene homologs during viral infection of *Arabidopsis* (Postnikova *et al*., 2012).

Arabidopsis ID	Soybean ID	Expressed in	Function
AT1G08380	GLYMA12G09731		Encodes subunit O of photosystem I.
AT1G30870	GLYMA15G41860	Root	oxidation reduction, response to oxidative stress
AT2G02850	GLYMA04G42120	transmitting tract of the pisti	Encodes plantacyanin
AT3G07270	GLYMA04G11890		GTP cyclohydrolase I
AT3G15480	GLYMA13G35510	22 plant structures during growth stages	Unknown
AT3G15480	GLYMA06G38224	22 plant structures during growth stages	Unknown
AT3G15480	GLYMA12G35030	22 plant structures during growth stages	Unknown
AT4G31130	GLYMA06G07900	22 plant structures during growth stages	Unknown
AT4G31130	GLYMA04G07830	25 plant structures during growth stages	Unknown
AT5G53250	GLYMA05G37085	25 plant structures during growth stages	Unknown
AT5G53250	GLYMA08G02480		Unknown
AT5G66170	GLYMA11G08260		Encodes a thiosulfate sulfurtransferase/rhodanese.

Table 10. Literature-based expression of candidate soybean T gene homologs during somatic embryogenesis of maize (Salvo *et al*., 2014).

Maize ID	Transcript	0hr	24hr	Expression
GRMZM2G004160	GLYMA04G42120.1	0.94	71.7	upregulated
GRMZM2G026980	GLYMA05G13660.1	0.28	23.51	upregulated
GRMZM2G035704	GLYMA15G36085.1	0.05	1.3	upregulated
GRMZM2G059706	GLYMA02G26610.1	0.56	12.66	upregulated
GRMZM2G082653	GLYMA15G36085.1	0.07	2.59	upregulated
GRMZM2G131087	GLYMA11G08260.1	8.56	294.2	upregulated
GRMZM2G400784	GLYMA15G36085.1	0.14	2.21	upregulated

Table 11. Evidence for the presence of candidate T genes overlapped between rice, *Arabidopsis* and maize during .biotic and abiotic stress conditions.

Names	Total	Gene ID
Arabidopsis, rice, maize	2[a]	GLYMA11G08260.1, GLYMA04G42120.1
Arabidopsis, rice	5	GLYMA13G35510.1, GLYMA12G35030.1, GLYMA06G38224.1, GLYMA08G02480.1, GLYMA05G37085.1
Maize, rice	2	GLYMA15G36085.1, GLYMA02G26610.1
Rice	6	GLYMA06G12301.1, GLYMA07G31330.3, GLYMA10G06470.1, GLYMA14G06841.1, GLYMA13G25130.2, GLYMA10G12190.3
Arabidopsis	5	GLYMA06G07900.1, GLYMA12G09731.1, GLYMA04G07830.4, GLYMA04G11890.1, GLYMA15G41860.1
Maize	1	GLYMA05G13660.1

a- Blue font, rhodanese is a candidate T gene, which is not/low expressed in normal growth conditions but elevated in expression by rice, *Arabidopsis* and maize during biotic or abiotic stress conditions as mentioned in discussion.

Table 12. Amino acid sequences of the candidate T proteins in soybean.

SN	Gene ID	Size (aa)	Seq (aa)
1	GLYMA01G02310.1	133	LLIIFCLIHMLNYVDRGAIASNGVNGSLATCTDSGICRGGSGIQVLSSAFMVGLLIASQYLLLLLLPII WKMVDGVLLCAGNVHFPKAGFVFFMMFLSNFKDELSVSPPQTHFFTESHYHYILIYYCIRSSKF
2	GLYMA01G11180.1	62	MTIAFQLTVFALIAISFILLISVPVVFASPEGWSNNKNVVFSGTSLWIGLVFLVGILNSLIS
3	GLYMA01G21521.1	120	MAKNEMKITGLMIMFMIVLGFSQANYNPSFVKIGSNRVYDGVYCQDKCSSACAKFLLLPGGVLAY VICLYACTAKCHDNPIDVAHDCITGCALKNSVDANIDARGLVVKDSFVQECRKN
4	GLYMA01G22751.1	94	MANYHSSKSLHLPCKTLSLVLLFVYLLLIGSCDAIRIGQTMKLNETREILKRKRNNQHGFPYKSMVF NFFPKGPVPPSGPSKRHNAVVDSTPNN
5	GLYMA01G31230.4	186	MKMRPFLRSFPRSLIFFFGCFLLLPLLLLLLVAPSSSGAPNEKVQHFRFVRCLERLTLEGRHNQWV NCFQMTGLGTSPIDCYTNGNGKAIDQKYAKETAQLNPPHRFVPWVVVNNQALQEDYQNFVTYIC RAYKGNVIPNACRSLSTRTYDSNEKVNSFLPVCYVDEARNLTLPLVITRLLSNESP
6	GLYMA01G45370.1	112	MEQTAKPNAFIVIIIIILIHIICQVHSRNMATSPSPAPASIAPETYARSLKPHEERKANPRLGSFPATCQ TKCNQCKPCVPVEVTIKTVAEEENEYYPIAWKCMCQSNIFSP
7	GLYMA02G03405.1	74	MASPLKFTVFSACVLVLFVVVTAQYGDGGNYGSDMPKDMPMGPAPHSNASLTYPAIITILFPFMLA FLAAKERI
8	GLYMA02G14821.1	95	MANYHSSKSLHLPCKTLSLVVLLFVYLLLIGSCDAIRMGQTMKLNERREILKGKHNNQHGFPYKSM VFNFFPKGSVPPSGPSKRHNAVVDSTPNN
9	GLYMA02G26610.1	196	MANPFPTFLPSSLIFFSFFFFLFSPLTAAKTHRFARTISPSSLGLDGEPEKLSHLHFFFHDVVSGQN QTAVRVAAAPATDKSPTLFGAVVMMDDPLTEQPEATSKVVGRAQGIYASASQSELGFLMAMNFA FTEGKYNGSSLAVLGRNTVASAVREMPVVGGSELFRFARGYAQAKTHSFSAVEAIVEYNVYVFHY
10	GLYMA02G27131.1	136	MKRSRSFSLLFYLYLLVIVFSTFSIVSVEARKTNTKKKLHKPHKGGHTRGSHSPCPILAPQGSTFDV LAFGAKGNGVSDDSEALLAAWNRACKVAGATVKIPAQLKFLMKHATLQGPCIPDLTLHVLFHFSLF YWF
11	GLYMA02G31140.1	68	MAPFGRFTAEILMAVMFVGLFCVTNIVAQDSEIAPTGQLEAGTGFALPVSKVIMCSSVLASLVAFM LQ
12	GLYMA02G37041.1	121	MAKNEMKIIGVVIMTMILFGFSQANYNPSFEKIGLNHVSDGIVCRLRCENKCKKFLIIPPGLIFYLICY HTCYANCHDMPIDAAHDCITSCGLTKPVDFNNDAPGLATDGVNSCLVKCYKK
13	GLYMA02G39725.1	71	MHVLYMFFFFFCRAWLMEYGCRESMGESVKLRRDSTMILMMFIAITLLFSMKQNIKILMVMMMRR RRRIPT

14	GLYMA02G44620.1	90	MHDLCLVESMRLMSLACFSILLLILFLFKISSSSAADGYAKSQSFRTEPVSNLGPFKGKSRQNTGE GGGEAFLGDEKRIIYTGPNPLHNR
15	GLYMA02G45441.1	173	MGDQQKLSAFLFHVSIVFVMFLLVSASQEHKKAKGGHSSKKDHNMKLAVLLATAGAIMSIKNFNN SFNNNHQRLGVALYGIIWLQVLVGIFRPQRGSKRRSLWFFAHWIMGTAVSLLGVLNVFIGLQAYQ EKTSKSITTWNILFTVQISLIVIFYLLQEKWVYIKNQGVIWDN
16	GLYMA03G14970.8	77	MGFWTLLEGFLLFANALAILNEDRFLARRGWTLAEMTGPQRNSLKGQVIGLIYACQFLRLPLILFNV ITIIVKLFSG
17	GLYMA03G21909.1	121	MSKNEMKLTGVVIMVMLVIGFAEADFNSAFVQIESNRVSSKPRLTCGPKCELSCLPYVLAFILYPIC VAACMKQCRQTPIDTIYSCLSGCDAVKSLPDNNEGRGRAAYVVDSCLQQCEDKK
18	GLYMA03G28125.1	100	MICVKRNNNQQKMSCVILFLLLLLLLLLSTTPCHAAARKARFDTLKGGSSSDDEFKFKPSNNFHG TGLQGNTNAQKDGDQVFGADKRKVYTGPNPLHNR
19	GLYMA04G05271.1	62	MASSQVVIMTAMAVLLAVVSAAHAADAPAPSPSSPASVISPSFAVGFLTAAVALVFGSSLRI
20	GLYMA04G07830.4	195	MAVTVKLMALTVSFFGLLSFILGVIAENKKPPAGTPVFGKDGVTCKFPADPTVALGYLSVIFLIASTV VGYLSLFYPYKGKTVPQGVLFKSMTFAVFFNVALFSTGLAATMLLWPTITEHLHLKRNVHLDLTYT CPTAKTGLFGGGAFLSLDSSLFWLVALLVADNAREDFLDEDKDDKLDVELPSLANHADMVI
21	GLYMA04G11890.1	128	MTMCQVLMLDAFSMCLLSSLFALLLSPCYCYRQKVKDIVQGALFPEAGLDNRVGHAGGVGGLVIV RDLDLLSYYGGAPSLNPTSSKVYDGPIKPVLEKHYLLRRSMKKCSNIDINSRLAPHTVIENNK
22	GLYMA04G29430.1	60	MMGYLKKVVINMILLFMLLLSMNLATARVPLWISEPLAHGRFLNADNPFGGNYHKRPPKP
23	GLYMA04G40810.1	71	TNIVPSSQTCILAMAKWISFAALFAILLLLISPELASAAQPSLILQLGRKLFQCGSAAVPVRGNTPSN GRS
24	GLYMA04G40823.1	91	MAFVGASKMMSSFGALLAILLLLMSSIMLATAAKPLGRKLLQQGYGGPVPDYGHRDPVPVRPTGR YRRPDPFPAYGHRDPVPVRPIHYGPP
25	GLYMA04G40861.1	66	MALLLFRASKLMSFAALLAILLLLMSSTMGSAAQPSFKLVRKLLQTKYPPAYPGYPGGGGYSPKTP
26	GLYMA04G41720.1	63	MAISSASFRVVAFLSLLLAALMTLASSQVTAPAPAPASDGTTIDQAVAYVLMLAALVLTYIMH
27	GLYMA04G41941.9	191	MGLLDIVCSSIFLFTLVEATSRSGIMWPLKIGVESTGTITLAVFSSSIFVLVALVLSTYLIFEHLAAYNQ PEFLSLLDSSAAFNCEVIRDCYEAFALYCFERYLIACLGGEDKTIQFMESMSLTESSTPLLKEAYAY GVVEHPFPVNCFLRDWYLGPDFYQSVKIGIVQYVCWILVNCAFVPLLFSLTGKE
28	GLYMA04G42120.1	132	MSKVAVIMREEAVHLYLLWSLVSLLCLLVLLERADAATYTVGGPGGWTFNTNAWPKGKRFRAGDI LIFNYDSTTHNVVAVDRSGYNSCKTPGGAKVFSSGKDQIKLARGQNYFICNYPGHCESGMKVAIN AV
29	GLYMA04G42480.3	125	MAGSGSSMLYSFLLFTVILSLQEMYRGKLASSELYTILGGFVSSLLFLVLLTFIGNFQETIGARTGW GAVIVAEAVALIAASTVHRVCITTCFLFSAALLYEVNKISGSALSTSDSRTKKQSGRA

30	GLYMA05G07900.1	118	MAMIIRFQHLLSFILWLFLFLVLFHGWFGSKSNNDKAIISNNQLQFSLSRNRRILAAGRGFDFTPFL RRHRHRHRHHHHHHQHHHHRSGVPESKETEIDPRYGVEKRLVPTGPNPLHH
31	GLYMA05G13660.1	197	MAPSSTHNNGFYVLMLVGIVVSAMVAICAASFYQDFDPTWVGDRAKIFNGGQLLSLSLDKVSGSG FKSKKEYLFGKIDMQLKLIAGNSAGTVTAYYRPRIDAAVAPVKKDSGGDIPCVGSLWWGGNVGDR MGASDGTVGGRNDLWVSVAFIARVAQYYSHQYSYTVINMPMLSLDWMMYVMCWLLVQQVILIPH PYT
32	GLYMA05G16420.2	156	MAKFQVLHKYFFIFLALVVCHGSLVAHGRKINVKPLNQQHYSLNTKTVANNNPYPSLPSLKTKVES PQYEEANKLGDSGADNTNAFRSTTPGGSTGVGHKIITSSEDNKMKTMVVVQSPDVEVFVTKGSK DDFKPTDPGHSPGVGHVYQNKIGQAN
33	GLYMA05G26021.1	65	MESCKIAILLTIYIMAMLVFAHCCVAENVAEDVAAIPPAPMESAGVHLCASAVFLATAFVVARFT
34	GLYMA05G26513.1	123	MIFIRHLFFCMAGDTASSFSTCMALFFSLLLLSHFTMAILEHNIFSPSTSNKEGVKKRNTKVVASVR EESEQNSKKDMALGETSKGSIHVPKREHSQHSQSPNRIFNASAHEVPSGPNPISNR
35	GLYMA05G28100.1	69	MSDWGPVFVSLVLFVLLTPGLLFQVPGRSRVVEFGNFQTSGAAILIHSLLYFALICVFLLAVRIHFYL G
36	GLYMA05G32010.1	85	MVVLNKRILLVAFFFLCFISIHARARVLKEKSNMVDSTSSDASTAHHSETHDHVFKPKEGKYNANE VFSMDYTPARRKPPIHNLN
37	GLYMA05G32301.1	77	MVLHRRPLILVIIWLFLFLFILGHCHGSRATTTVFKFKPKSPHHGHFSGFLPKRMPIPYSTPSRKHN DIGLRSWRSP
38	GLYMA05G37085.1	71	MAGGAPKQLFGFGVVALLATLILPLFMPAAVQAQSAAPAPAPTSDGSSLDQGIAYVLMLLALVLTYI IHSA
39	GLYMA06G07900.1	191	MAVTVKLMALIVSFFGLLSFILGVIAENKKPPAGMPVPVKDGVTCKFSADLTLALGYLSVIFLIASTV VGYLSLFYPYKGKAVPQRVLFKSTTFMVFFNVALFSTGLALTMLLWPTITEHLHLKRNVHHDLTYT CPTAKTGLLGGGAFLSLDSSLFWLVALMLADNAREDFLDEDKDVELPSHVNHADMVI
40	GLYMA06G12301.1	125	MAGSGSSMLYSFLLFTVILSLQEMYRGKLASSELYTILGGFVSSILFLVLLTFIGNLQETIGARTGWG AVIVAEAVALIAASTVHRVCITTCFLFSAALLYEVNKISGSALSTSDSRIKKQSGRA
41	GLYMA06G13060.2	63	MAVSSASLRVVAFLSLILAALMTVASSQVTAPAPAPTSDGTTIDQAVAYVLMLVALVLTYIMH
42	GLYMA06G23157.1	87	MLTYAFVVMLIFQFCGGYYADNTDSMFNSIKSYCCNLMKISEIFLCLFLLWRLISFKSSTENDLKIES PILFSIPFLLFSCFVYIIV
43	GLYMA06G37230.1	144	MVSSVFTLMCVLHSTIALTCGALMIFYSKEISVLGHGSETASKLQGTTPHDQLLIDTSDSFSGLLLFT IGFLLLMVAFVKDREFQSFFAKGCVMLHISMAVWRFFFERKLGDLAHEWPRHAVGDIALAISWVFF LVYTWREKYD
44	GLYMA06G38224.1	93	MASMFLLITVFVLDLIAFAFAVAAEQRRSTTRISTTIIVYDSDISTGLGVGAFLFLLASQDLIMVASRC FCCGKPLNPGGSRTLELVLFIICW

45	GLYMA06G40291.1	60	MDMKKVTCAVLIAAASVSAVVAATEVPAPAPGPSSGASASLPLIGSLVGASVLSFFALFH
46	GLYMA06G43681.1	93	MGNTSATLVPILALIMFSTFFMTLQARSLHGHNPLFAHKKVVDIQNFLHKSGIHLSKRVRIPFGDDL PLAPADRLAPGGPDPQHNVRAPPRKP
47	GLYMA06G45730.1	154	MLGLGAAVLLGLAHAITNLLGGCNCISSQQEFEKASSSRQLSTACLILTWVVLAIGLSMLVIGTMSN NRSDDSCGFSHHHFLSIGGICCFVHGLFCIAYYVSATASMD
48	GLYMA06G46010.1	168	MGGCVLKLVDAVLFLFFLLIAVVAPLIDAQTCLPLSYFPDILVQLKEHYTKDYGDYLVAEKPHFFVGL VWLELLFQWPLALLSLYAILTSKPWFNTTCLIYGVSVTTSMVAILSEMMNSKRASEKLLTIYASFLGL GVLALLRGLLTCSSKSSSALGKRSALARKKRA
49	GLYMA07G06290.1	70	MADWGPVFVSVVLFILLTPGLLIQIPGKGKMVEFGNFQTSGVSILVHSILYFALVCIFLMAIGVHMYT GS
50	GLYMA07G09600.1	63	MRTVNIFALPLLALFIMAISRLAQAQDLPLSPAPPPTSDGTAIDQGIAYILMVVALGITYMIH
51	GLYMA07G11031.1	80	MARTSLSSISIPFLILLIIAFTFFAQLAPVSADLQKRKFGPMSSPPPPPKFLLGPGPGRSGNDAPPPP R
52	GLYMA07G11035.1	73	MARTSLSSISIPFLILLIIAFTFFAQLAPVSADLKMRKLGPMPSPPPPPRNSPGPGPGRRKTARAAP PPMLIN
53	GLYMA07G16091.1	72	MARTLVSVISLIFFTFAVMVVVLFSIGVNSSQICPGNCENFSNCNAYCQAHYHKKGICISPPKRGYI CCCFS
54	GLYMA07G30620.1	173	MAKNCGFLVCILVLVLDIAAGILGIQAEIAQNKVKDLKVWALECRDPSYEAFKLGVAASIFLVFAHAI AHLLGRCICMQSKEECQGATANRVLAVTFLILSWILLGIAFSLLILGSSANSRSRESCGISNRRFLSI GGILCFIHGLFTIGYYVSVTATRREEKRQGNSFVGHT
55	GLYMA07G31330.4	183	MASSLLLAVVFVLDLVAFALAVAAEQRRNTASLNKDSAGRNYCQYDSDIATGLGVGSLFILVASQV IIMVVTRCLCCGRAMRPSGSRSWSICLFITSWVTFIIAASCLLAGSVRNAYHTKYRDLMGERAPSC QTLRKGVFGAGAAFIVLTGITSELYYVSFSKANNNGPPPYARDTGVRMANF
56	GLYMA08G02480.1	71	MAGAAPKQLFGFGVVAMLATLILALFMPAAVQAQSASPAPAPTSDGTSLDQGIAYVLMLLALVLTY IIHSP
57	GLYMA08G06680.1	173	MAKNYGFLVCIFVLVLDIVAGILGIQAEMAQNKVKDLKVWVLECRDPSYEAFKLGLAASICLVFAHAI AHLLGRCICMQSKEECQGATANGRLAVAFLILSWILLGIAFSLLIFGSLANSRSRESCGISNRRFLSI GGILCFIHGLFTIGYYVSVTATRREEKRQGNSIVGHT
58	GLYMA08G07360.1	98	MASHYSCIILAFVIAISFSSIDMALASRNLLQTTLPTLPQGNVPPLPVFPNLPLPPFSFPNIPIPKIPNIP NIPSNPIIPTIPTIPFLSPPPSTTTTP
59	GLYMA08G08950.1	66	MMECCKIAIVLTIYIMAMLVFAHCCVAENVAEDVAAIPPTPMESAGVHLCASAVFVVTAFVVAMFT

60	GLYMA08G09950.2	164	MMFFGFIFLIGLAILSHQTVPGSHETRKIIHLTLHFIAIILGIVGLCAVFKFYDMVNWEDVTSLHSWIGI GTFCLVGLQWLMGLGFMLQGSAQSRSTMAPWHVAAGRALFFMAICAALTGLMEKYTMIKLVTHH RESHLINFTGLAILLFGVFVDMSVGLPRYA
61	GLYMA08G16251.1	67	MEKKKVACVVLLVVTSISTAMAHAGHDKAEAPAIVPAPALAPKGGVAALGPFLGASLLSFLGYYLQ F
62	GLYMA08G16301.1	63	MASPARTFFAFILAVVALFFAAANAQDLSPAPSPDAGAAGSVSSSVAMIGASVVLSLMAILKH
63	GLYMA08G17190.1	139	MVKKRTRSKFIPASSSFIILFSATMKSSGALATLVLLILLLATTSEAAISCSDVIKDLRPCVSYLVSGS GQPPAACCSGAKALASAATTSEDKKAACNCIKSTSKSININSQLAQALPGNCGITLPVAISPNADCS KVG
64	GLYMA08G25930.1	105	MMVVRRSLQYPLLLALTSLHLQHLSGLDNNPSGAKNENKDVNHHASRSSVGLKIIIIFLGVVTVIAF CVFLFKLWQRKKREEQHARLLKLFEDDDELEVELGMRD
65	GLYMA08G29500.1	91	MVVFFWCLSLVQFHMLNLEVIKPITGASMNPARSLGPAIVHNEYKGIWIYLVSPTLGVVAGTWAYN FIRYTNKPVHEITKSASFLKGGEAK
66	GLYMA08G40260.1	116	MKNKRKTLLLLLLMAATLFCMPIVSYAVSNVNIQDHLTNISELVKGPNRRLLSFVDCGERCRVRCS LHSRPKICTRACGTCCMRCRCVPPGTYGNREMCGKCYTHMITHGNKPKCP
67	GLYMA08G46370.1	95	MAKTKALVVLSLFVMATLNGTATAERGFPGKDNDHGHHNDNDLLYEPQHFGVLGLWPFILLHPW LWLKNDKMAVPYYQGVPKVDDEGHGNSPTVP
68	GLYMA09G05645.1	72	MSLCLILLGICCVPVFVSDSYTFIIASYLCPVLYPCFLLKNPKLFTSQLQEPRMSAHPFSVKLFYSGV DNFS
69	GLYMA09G05750.1	75	MQMQLVAFYLLFILLPSFNSCGCRISSSRKYRFSLLSWIVPTGMFLFPSLVIVVPGEFFLICSSTDFY FLKPFGK
70	GLYMA09G06137.1	120	MTKNQIKTTALVIMLIIIMDFAQADYNALYVQNGPTSFPSKIDCYHKCELECLPLFITGAVYIGCVAAC FHDKCKKKPMDVVYNCISRCALTKSTEINNDDHDPATNVVDSCFEECQNEK
71	GLYMA09G07481.1	145	METRRIMCYNFLVASILLLSCLVCVKSRPLFLFDASTQQAFFKQSSTLDDDQTTRIQDVSKHVFDTT QGNEGKRILKAKANEYKEALRGLSRLGSTPPRCEHKCGGCIPCNPIQIPTNNDLLGAQYANYEPE GWKCKCGNSYYNP
72	GLYMA09G25983.1	73	MGVHKSRFLTTIFLLFILVILLHLSSCRHISWGLKEEIMEQRTRSRFPFSSFPHSYNSAEIKDNKAKIS SESV
73	GLYMA09G28700.1	118	MGLKNNMVVLKVCLVLLFLVGGTTSANLRLSKLGLLMKSDHHQHSNDDESSKPCCDQCACTKSN PPQCRCSDMRLNSCHSACKSCICALSYPAQCFCVDITDFCYEPCKPSEDDKENY
74	GLYMA09G28730.1	118	MGLKNNMVVLKVCLVLLFLVGGTTSANLRLSKLGLLMKSDHHQHSNDDESSKPCCDQCACTKSN PPQCRCSDMRLNSCHSACKSCICALSYPAQCFCVDITDFCYEPCKPSEDDKENY

75	GLYMA09G31094.1	78	MARTILSSISVPFLILLMIVFTLVAQITPISADVKMRRLGPLLSPPPPPDHADGPHPVGPHPVIPPPTDGPGRPPPIP
76	GLYMA10G03050.2	148	MKRVKIFILITLALTFVAFVPKIESQIMPPLIPFPPPSLHPLCLSQLALVSYACAMLPPTTTLPPSPLTPPPSPSSPGDDEGHGNDQSQGHHQTSQEENCCRWAREMDNQCVCEFLLLLPPFLTRPLHQYSISIGESCNVTYSCGGPI
77	GLYMA10G06470.1	106	MGLGFVVGFLGLLILFHAAYSTIQYKGLLKITEEEFSGPPLNVVIEVTLGLVFCMWAALTVPGKFLSIHPHSEENRIVSLPSNVDFMIFNHRYKVFPVEMDVKLKH
78	GLYMA10G12190.3	154	MGWNLLFWLAICFPSNIALLASTFYQVLILSDLESDYINPFDAASRINYFVLPEFVGQGALCALCLFTGHWFMFLLTVPVTCYHLRLYVKREHLIDVTEVFRVLNAEKKYRIAKLALYLTVLIVTIFRILAAGRVILYSSKFEELDIRLNPIGF
79	GLYMA10G12370.1	59	MDMKKVTCAVLIAAASMSAALAATEVPAPAPGPSSGASAAAAVGSLVGASVLSFFALFH
80	GLYMA10G33941.1	148	MNSSEETALLLVSFIVLTLIITLTAYVCSNNIPRTVPSVPGGATHSNNNHTTITVEPPEPRLDHTNVRSYPSLQFSKAKLCSSNSSSSSSSCSICLMDYKDCDSLKVLPACGHFFHVKCVDPWLRISLTCPVCRTPIALPDICKNLE
81	GLYMA10G34960.1	94	MANSHYSKTLHLPCKSFSMAILLVYLLLVGSCTAIRTGATMRLNEGSELLRRKQQQPRFPYKGLVFNFLPKGVPIPPSGPSKRHNSVVASTPKN
82	GLYMA10G35010.2	131	MFFLTFSILAAVLGIVSKLLGGNHMRTWRSDSLASAGATSMVAWAVTALAFGLACKQIHLGGHRGWRLRVVEAFIIILTFTQLLYLILIHAGLYSSRYGPGYHDTDYGHGHGVGGTTGDPMHKPATAGTRV
83	GLYMA10G37105.1	77	MDRRYLNLVLVLILWLIYFVGHSYGARHSHQVFKVQPKGEAFPPSFFGFLPKAMPIPPSGPSRKHNGIGLQSSNGEP
84	GLYMA10G40190.1	85	MKMSSKVTGATLCLAVLFLFLTFTYAGRLGPASSSITSIKTQHGVLEEEKLDVEETCDGIGEEECLMRRTLVAHTDYIYTQKHKP
85	GLYMA10G41700.1	195	MAVSVTILVLIIALHLIAFVFAVGAERRRSEAKVVPDEYDDQTFCHYTTDASTVYGLSAVALLLLSHTVLNGVTRCLCCGKGLVSGCSATSAVISFILSWISFLAAEACLLAGSARNAYHTKYRGYFVNHDLSCATLRKGVFAAGAALTLLSMLTAILYYWAHSKADTGFWEKHHNEGLGLATQHHHQGPDSDKA
86	GLYMA10G42520.1	129	MCFRVLLGVSLLILIYLINLAPAYPLHHTLLSEIVQNSKFEDKGILYGSRIGGAGHGGSSHGNGNGNSGSPETRGGGAALIPVYAAGAANKNHQHQTRHGAANCNLNKIRFSNLLMIILVYPLILSFLL
87	GLYMA11G08260.1	149	MAVAVAMLPRWSVFLLFLFVLCISGAKVVTIDVRAAKSLIQTGSIYLDVRTVEEFKKGHVYADNVLNIPYMLNTPKGKVKNGDFLKEVSSACNKEDHLVVGCQSGVRSLYATADLLSDGFKNAKDMGGGYVDWVKNKFPVNIPEAKEEL
88	GLYMA11G14061.1	72	MACISKSRAMLFFMVFFVGLLFAAGVSAQTSELPQAPAPAPTMDAGAGFVVTYSGAFVCSSLLLSLIALLCH

89	GLYMA11G19501.1	191	MLKGLCVCYLVLIVTFFSVSVSGYWAFGNESEGLILSNFVDNGKPLVPKWFIYMTNILIITQLSAVGV VYLQPTNEVLEQTFGDPKSPEFSKPNVIPRVISRSLATTISTTIAAMLPFFGDINSLIGAFGFIPLDFIL PMVFYNLTFKPSKRSPIFWLNVTIVVAFSALGAIAAVRQIVLDAKNYQLFANI
90	GLYMA11G36910.1	69	MADWGPVVIAVVLFVLLSPGLVFQLPGKSRVVEFGNMQTSAVSILVHTIIFFGLITIFLVAIGVHIYTG
91	GLYMA12G06030.1	72	MACISKSRAMFFFMAFFVGLLFAAGVSAQTSEFPQAPAPAPTMDAGAGFLVTYSGAFVCSSLLLS LIALLCH
92	GLYMA12G07550.1	76	MCFLCNCLCLLLLLPLLSTCKWGWNHTLNFLLLFNIYSSLFFSYLLKKSCLGILLNMTFLFFSVHPVP LDLTLLQI
93	GLYMA12G09731.2	128	MSMLCGMFCNLFAGFVKSPVSARNPLRKAVAMGNGRVTCFERNWLRADYSVIGFGLIGWLAPSS VPAINGKSLTGLFFESIGAELAHFPTPPALTSSFWLWLVTWHLGLFICLTFGQIGFKGRTEEYF
94	GLYMA12G10690.1	168	MGGCVLKLVDSVLFLFFLLIAVVAPLIDAQTCLPLSYFPDILVQLKEQYTNDYGDYLVAEKPHFFVG LVWLELLFQWPLALLSLYAMLTSKPWFNTTCLIYGVSVTTSMVAILSEMMNSKRASEKLLTIYAPFL GLGVLALLRGLLTCSSKSSSALGKRSALARKKRA
95	GLYMA12G11100.1	160	MARVAGIFLCLLILVMDVAAGILGFEAEIAQNKVKHLRLWIFECREPSHQAFMLGLGAAVLLGLAHA IANLLGGCNCICSQQEFEKASSNRQLSTACLILTWVVLAIGLSMLVIGTMSNNRSDGSCGFSHHHF LSIGGICCFVHGLFCIAYYVSATASMD
96	GLYMA12G12730.1	163	MAKTRRVCHLLLRFLAFAATLAAVIMMATSHETATIFTVSFEAKYTNSPAFKYFVIAYSVITVYGFLV LFLPAKSLLWQLVVALDLVFTMLVVSSFSASLAIAQVGKKGNSDAGWLPICDSVPKYCDQATRALI AGFIAMIIYIILLLHSIHTVIDPLLLRKS
97	GLYMA12G16173.1	166	MSIFCVKLFICMTVSSFSFLVFCIVPKENIEEIFLSLFLEGKTRHLYDVLICITCVACCLTVYYIPKINRK TNICYFLLHEFSMGATNLDAANLILCRLIRISYGICMSWSPKISVQDYCISFVWCNSYSWSSDHKNS FVVMDNYNSMTAWYVYLSSGICSLLLDV
98	GLYMA12G28680.1	103	MNSITRSGSLSIILVVFFSLHFILGFSDDSSPASKDATKTEPHAGRSTSAIVIIVLIVLVFFSLFSFVLFK LWRKKKREEQYARLLKLFEEDDELELELGLRD
99	GLYMA12G31130.1	58	MPHRTRPMTALLLFTALNTVLCATITPVYDFVCFHPYWERRRERRRQQREATIANSST
100	GLYMA12G32320.1	158	MDRGVGIFLCLLLVTMDISAGILGIEAEIAQNKVKHLRLWIFECKDPSHKAFMLGLAAAVLLALAHVI VNLVGGFNCLCSQPEADKASPNRQLSMASLILTWIVLAVGLSMLVIGTSSNNKSSGSCGFTHHHF LSTGGILCFVHALFSVVYYVSATAS
101	GLYMA12G32996.1	128	MRRNSLYIVAFVLIVLLCVPVVISARHINRSRSRHGHPDPVQENTEDGMNKVAKQNLHWEGRVTR KRNRGPDTLQIAGSRLPDCSHACRSCSPCRLVMVSFVCASLAEAESCPMAYKCMCHNKSYPVP
102	GLYMA12G35030.1	185	MASKLLLITVFVFDLIAFGLAVAAEQRRSTASIVPDNEKNYNYCVYNSDIATGYGVGAFLFLLVSQV LIMVASRCFCCGKPLNPGGSRACAVVLFIICWVFFIISEVCLLAGSVENAYHTKYRTIFGENPPSCE TVRKGVFAAGAAFIFFTAIVSEFYYINYSRARESFQPYAGGETGVGMGTYK

103	GLYMA12G35290.1	145	MGGSSCFFIICVLHSAIALTCGSLMVFYSKEIRVLGHGPKTASKLQGSTPHDQLLIQTSDSFSGLLL FTIGFLVFMVACVKDWEFQSFFAKGCVLLHISMAVWRFYFEGKLEDLAHDWPRHAVGDIALATSW LFFLVYMWREKYD
104	GLYMA12G36703.1	64	MAPTFKAFALFLLLALFSAASAQELSPASSPAPSPEAGAADSVPCSAIMIGASLVLSLMAVFKF
105	GLYMA13G01671.1	121	MTKKEMKVIRVVIMILIMFDFSQANYNPSSMQIGSNYGSDGRFACQIKCQKKCQKYLLLPPGVIFYII CYHSCFRKCDKMPIDVAQDCVTSCVLTKPIDVNIDVRGLTDDMVDSCIQECYK
106	GLYMA13G08400.1	61	MVVSKVVVFLGLILAALMSVACSQSAAPAPAPAPTSDGTSIDQAVAYVLMLVALVLTYIMH
107	GLYMA13G25130.2	183	MASSLLLAVVVVLDMVAFALAVAAKQRRNTASLNKDSAGRNYCQYDSDIATNLGVGSFFILVASQ VIIMVVTRCLCCGKAMRPSGSRSWAICLFITSWVTFFIAASCLLAGSVRNAYHTNYRDLMGERAPS CQTLRKGVFGAGAAFIIFKGITSDLYYVSFSKANNNGPPPYARDTGVRMGNL
108	GLYMA13G31270.4	148	MRSRMASSSSAFLVICILHSLIAVTCGGLMMFYMKEVYTFGHGVQAATKLLGSTPHDQLLIKTSDS FSGLLLVAIGFLLFMVSFVKDRDFQVFFAKGCTLLHLFMAMWRVYFERKVEDLAWDWLRQTVGD FLLALSWVFFLVYSWREN
109	GLYMA13G32350.1	168	MSMPPHRSLQAIFLVLFVVVLLPFSVSVAASYSSIHELLRSHGLPAGLFPEGVKSYNLDQRGRLEV NLDGPCMTKYETRVLFETVVRANLSFGQLKGLEGLSQEELFLWLPVKDIIVNDPSSGLILIDIGLAH KQLSLSLFEDPPVCRSQGLSLNIGGRKSIGFQDQR
110	GLYMA13G35220.1	145	MGGSSCFFIICVLHSAIALTCGSLMVFYSKELSVLGHGPKTASKLQGSTPHDQLLIQTSDSFSGLLL FTIGFLVFMVACVKDWEFQSFFAKGCVLLHISMAVWRFYFEGKLEDLAHDWPRHAVGDIALATSW LFFLVYMWREKYD
111	GLYMA13G35510.1	185	MASKLLLITVFVFDLIAFGLAVAAEQRRSTASVVTDNEKNYNYCVYNSDIATGYGVGAFLFLLVSQV LIMVASRCFCCGKPLNPGGSRACAVVLFIICWVFFIIAEVCLLAGSVENAYHTKYRTIFGENPPSCE TVRKGVFAAGAAFVFFTAIISEFYYINYSRARESFQPYAGGETGVGMGTYK
112	GLYMA13G36831.1	95	MANATMATRVSILIALIILSTFFMTLQASNLHGHPFIRENNIADSHHFLHKYLDDLSKHIHVQDADDA PHKDGNTHRLAPEGPDPHHNFATPPRN
113	GLYMA13G37370.1	82	MAYYSYYRHQSKKLIWFLMFVFLVFSLMETSLTEGREVYRLKDIGRSRSQVENGGQMKRLETLAE LLPRGPVPPSAPSPDIN
114	GLYMA13G37491.1	131	MRRNSLYIAAFVLVLLCVPIIISARHINRSRSLGHGGHPDPGEKNTKDGMNIVAKEKLHWEGRVVA RKRIGRGPDTLEIAGSRLPDCSHACGSCSPCRLVMVSFVCASLAEAESCPMAYKCMCHNKSYPV P
115	GLYMA13G38100.1	158	MERGVGIFLCLLLVTMDISAGILGIEAEIAQNKVKHLRLWIFECKDPSHKAFMLGLAAAVLLALAHVI VNLVGGFNCLCSQQEADKASPNRQLSMACLILTWVVLAVGLSMLVIGTSSNNKSNGSCGFTHHH FLSTGGILCFVHALFSVVYYVSATAS
116	GLYMA13G39176.1	58	MPHRTRPMTALLLFTALNAVLCATITPVYDFVCFHPYWERRRERHRQQREATIANSST

117	GLYMA14G03371.1	155	MGVQQKLSAFLFQLSMFLLVSASQEHKKAKGRHSSKKDHNIKKLAVLLATAGAIMSIKSFNNSFSN NHQRLGVALYCIIWLQVLVGIFRPQRGSKKRSLWFFARRVVGTAVSLLGVLNVFIGLQAYQEKTSK SITTWNILFTVQISLIVIFYLLQ
118	GLYMA14G06841.1	129	MINQQLKEKRCALANLGFCLLIGYLLLSSLIPPSTAPPPPPSTFRFCSFLREMGLWTLLEGFLLLAN ALAILNEDRFLTPRGWGLSDFSAGQTKSFKGQLIGLIYATQYLRFPLLLLNSVFIIVKLVSG
119	GLYMA14G08850.1	108	MANLKLVFTMSSILLVLVFFNGILPAMGRPLKKEHITTTYENSVKEMGTVEDNNILLWRRSIIENNAA NDGGVDKWIDDFRPMDPGHSPGAGHSSPTPKDATNGAPRP
120	GLYMA14G09710.1	124	MKKVSGFFVLLLVVGATLSFLNFLSPTCASWFSDLIATNCGDKATLIAVSRKLKESDSSTIKSRSNN KGDIGQVTLNDYNPIDPVPSSSKASINPGPIEHGTPLNPYIIPKPSPPNHPKPGDSN
121	GLYMA14G12003.1	93	MHLPWCLFCLFLSLCISGMHVCYIGEFYLGYCRRPKTSFLISVLKAQEEKRGSNTNIKMDEATREK TTYKRITCEEKSVTPNEEMRKRLQKIN
122	GLYMA14G31650.1	59	MVVSRVVVFLGLILATLVSLACSQSAAPAPAPTSDGTSIDQTVAYILMLVALVLTYIMH
123	GLYMA14G34190.1	69	MSDWAPVLIGVVLFVLLQPVLLFSFPGNGEQLEFGSMKTNDKAIFIHMLIFFALYYVLILAVKIHIYTS
124	GLYMA15G06910.1	175	MTKNYGFLICILVIVLDIVAGILGIEAEIAQNKEKHMWVWIFECRYPSYQAFKLGLAAATFLALAHVIA NLLGGCICVWSKEQYLSATANRKLAVAFLIFSWVVLAVALSMLMIGTLANSRSRKSCGMFSRHFLS IGGILCFIHGLFTVPYYVSATATKRDEKRPENASHTLDHL
125	GLYMA15G06970.1	163	MSMRHRSLLAIFLAFVGVFPFSVAASYSSIHELLRSHGLPAGLFPESVKSYNLDQSGRLEVNLDGP CMTKYETRVLFETVVRANLSFGQLKGLEGLSQEELFLWLPVKDIIVNDPSSGLILIDIGLAHKQLSLS LFEDPPVCRSQGLSLNIAGRKSIGFQDQR
126	GLYMA15G08080.1	148	MRRRMASSAFMVICILHSVIAMTCGALMMFYMKEVYTFGHGVQAATKLLGSTPHDQLLIKTSDSF SGLLLVAIGFLLFMVSFVKDRDFQVFFAKGCTLLHLFMAMWRVYFERKVEDLALDWLRQTVGDFL LALSWVFFLVYSWREKYD
127	GLYMA15G16860.1	59	MSVMILLLLLKSMPGTGIKNVGFIVSMGLDSKVNIFVVCVFFNMLWWVDFGFRTVVRIL
128	GLYMA15G21980.1	148	MVFQRSLTMAFNLVLLFCLVAVINNVVIPTSARQLMMTASPNEKEVMNIGTENRGHDEHDHQAQV LDDPTSKNNVEKKEERFDKNQLFPPFFPLPPFTLPFPRAAFPFPRFPLPGLPWPIGPAGASLFTYP TVDMPASPYLPKPTPLN
129	GLYMA15G32285.1	150	MYSSMFRKLFCSVLFSWLCFGGLAMAAGESSSSSRDLDQTPTWAVAAVCTVFILVSIALEKSLHK VWTWLGQNKKKALLEALEKVKAELMILGFISLLLTFDQSYIVRICIPEKLADNMLPCPYRYKEAKKA SVVKRNIVGNFCLMNVDI
130	GLYMA15G36085.1	163	MKFIVLFFLLTSSSSSSSSSYSYAQSPVAPAVAHVPSTRATLPMLLVIFLFALPFTALCSIFICYCSHE EQPQVLPEATRSTPRGVDPRVLATCPITSYSTVKMRTPQNPVVSLAEFDDANALRLLPKCSHVFH THCIDASLSSALSKKIKHRKKNYGMNRWQ

131	GLYMA15G41860.1	104	MRLHYLTLFLLLVPLELSIIYGLSTLGNVPKKSFKPLLPPEALLSIGHYHTTCPDTEGIISQKVAAWVK KDPTLAPAIIRLHFHDCAVRDHMDLVQKGAKGAKG
132	GLYMA15G42725.1	68	MARTSLSSISIPFLILLIFAFTFFAQLAPVSADPRMRKLGPGPSPPPPPGRNHGLSPFFPNAPPPPLI
133	GLYMA15G42762.1	66	MASQATSFKSFVIALIVALFFAAATAQDLSPAPAQGPDVGVAGSVSSSMAVIGASVVLSMFAIFKH
134	GLYMA15G42774.1	68	MASQAITFKSFATVIIVALFFAAAAASAQDLSPAPAPGPDAGAAGSVSSSVAVIGASVVLSMLAIFK H
135	GLYMA15G42787.1	67	MASQAVSFKSFAVALLVALFFAASASAQDLPPAAAPGPDAGAAGAVSSSVAMIGASVLLSMLAIFK H
136	GLYMA16G02930.1	70	MADWGPVFVSVVLFILLTPGLLIQIPGKGKMVEFGNFQTSGVSILVHSILYFALVCIFLMAIGVHMYT GS
137	GLYMA16G03960.1	131	MFVLVIAYLYIPASSALPCHPTWCGVCYSSSSSSKMQRNFMRLLMFFLCFSYVLSVSAIPATRTQN LKGEEEEDFSALPSLTRVDHALGNGEVVLIDMNEGFIERRVDLETQDYEGTGANKDHDPKSPGGP
138	GLYMA16G04471.1	59	MEASKMKFFLVLVIAMLAMVATGVSAAEAPAPGPSSDATTFFVPTALASLFVLAFGLLF
139	GLYMA16G10175.1	148	MILMFACAFPPAFAFAAVNNLMEIRTDALKLLVILRRPVPRAAATVGVWLNIFQFLILMSICTNCAILA WLYDEEGNWKIEPGLAAILIMEHVLLLTKFGFSRFFPEVIVLLPKSKCMFENTFKTLLVVRLFPYSYS FLYNVGTKHSP
140	GLYMA16G11475.1	126	MSTFIMPTLFLLLFIVFNIGDDCFGTECFLFFCICIVFILYVYVYKYLHSQCAYDVAPGGLLASVYHLT RLEYGIDQPEEVCIKIFVARRNPRIPSIFWVWESVDFQEKESYDMLGISYDNHSRQT
141	GLYMA16G12146.1	84	MAKYTAALLLLVLLVVAILNGSDASRRRLPEKDDLVYESQIFGFIPIFLHWKFCLLHPLLCLWRPKD HHVSAEIENLSIAPTNP
142	GLYMA16G26965.1	55	MEALKMKLFFVVMAMLIMAASAADSPAPSPTSDATTLFVPTAVASLVALAFGLLF
143	GLYMA16G32110.1	71	MVVFQHIFCILCHVPKLTQCLLEFTPTMLLHPRHPKMMDVAEYTSHNKFSRSSLHFCMLSFVYWVI IHLYT
144	GLYMA17G13100.1	114	MAMIIRFQHPLSFILWLFLFLVLFQYVWFGSKSNNDNNAVFTNNQLQFSLSRNRRILTTGRGFDFT PFLHHHHRGHHRRHHHRSRVPESKETEIDPRYGVEKRLVPTGPNPLHH
145	GLYMA17G22121.1	145	MERGLILLLFLLLVMVLSVYSNEVEYSHIHNVLVCQKVSDFFIAIAYFSIPLELLYFVSCSNVPFKLVF LQFIAFIVLCGLNHLLNAYTYYGRHSFQLFLSITIAKFLTALVSCATAISFPTLIPLLLKIKVRELFFWQ NVLELG
146	GLYMA17G35370.1	120	MASKLLLITVFVFYLIAFGLAVAAEQRKSIAKLESLIFYISFYELKFEPIISSKFSKPFTIDASVAGSLLT LEAQGPVQLVFFIIAEVCLLAGSVENSYHTKYMTIFGETIIIIREEIQE
147	GLYMA17G36320.1	108	MTNLKLVFTISSILLALVFINGISSVMGRPLKKEHIITTTYENSVKEMGTVEDNNILLWRRSIIENAAN DGGVDKWIDDFRPTDPGHSPGAGHSSPTPKDASNGAPRP

148	GLYMA18G17490.2	115	MEKKRKTLLLLLLMAATLFCMPIVSYAVSSVNIQGHLTHSELVKGPNRRLLPFVDCGARCRVRCSL HSRPKICSRACGTCCFRCRCVPPGTYGNREMCGKCYTDMITHGNKPKCP
149	GLYMA18G46745.1	80	MLSGIFLHGLLLPAPLTISEAIKEVLEATWKVQETIASFTLLHAFGLSPLLFYLGFLVSIQLTYLCISQL FIFLVSLQTS
150	GLYMA19G28574.1	71	MNLIIFAAFVIAAGLTVGLASIGPGVEGIARQPKTKGKIRGTLLLSLVFMEALTIYGLVVALTLLFANPF V
151	GLYMA19G28950.1	59	MEASKMKFFLVLVVSVLAMAATGVSAAEAPAPGPSSDATTLFVPTAFASLFVLAFGFLF
152	GLYMA19G29690.1	95	MSFKSPLQLLIIILLVFSFVVSSAALQTTRRLLPNKEKLSTQITSDKGVEELRNGEEMLDMAEEFMV EGRIDLESNDYPGTGANNRHDPKTPGGP
153	GLYMA20G05690.1	136	YSFVCSWFLFLALGGLLFGYDIGATSGATISLQSPELSGISWFNLSAIQLGLVVSGSLYGALLGSLV AFARADFLGRKKQLITATLLYKVLDRRLQKAWARTTKEDPRILMNLRGILSSLCSRDFFFLYFHVFL EW
154	GLYMA20G22351.1	126	MVVNLSLYAVALIVFSVINGKTHDDVENQYCWIAYLSIFIGCCFVGVFHLATKEPRLKVDVHGMVH ARISWDYWFKRILYYHVGPVYVLTRLVLNVSQAYLAFFVINDLQMAQSAKALVIYKFCPF
155	GLYMA20G23630.1	69	MFVLSSSSSSSSSLLLLLLLLFQIEDGFLELADPCRIVILCPVHCSMCVAYDISFLVFIGIFSLRLS
156	GLYMA20G25510.1	196	MAVSVTILVLIIALHLIAFVFAIGAERRRSEAKVVPDEYDDRTFCVYTTDASSVYGLAAVALLLLSHTV LNGVTRCLCCGKGLVSGCSATCAVFSFILSWISFLAAEACLLAGSARNAYHTKYRGYFVKHDLSC ATLRKGVFAAGAALTLLSMLTAILYYWAHSKADTGFWEKHRNEGLGLATQHHHHQGPDSDKA
157	GLYMA20G26620.1	63	MKGYSLAVAFNAIFLVIFLLLLLSYSNNVESTRTLKDQSSSPAFIGLIINRAYSGPSHRGAGH
158	GLYMA20G28830.1	125	MGVTRRRRHHHHFHWLQIFTTFTFLFFSSTSAITTLTPHPSGGAVKQLNREEKQNGSTVVPVNRV VVEEKRFGGPGSSPPSCRSKCGWCSPCFPVHVPVQPGLIIRLEYYPEAWRCKCGNKLFMP
159	GLYMA20G29721.1	188	MALAFFGFLFLSSFLSCIAINFDYDYTLLVLSYLVENGVVFIVLCESTYPRKLAFHYLQDIQKEFEKF DKTLIGKITRPYSFVKFDGIIANISRQYIDTRTQANLSKLNANRKQDLDIATEDIYKILERKRNSETMR RLPVTPQPESTIWCSPQLEVIALKWTPIMIIVITSMALLWASLALTDDFIV
160	GLYMA20G32601.1	94	MANSHYSKTLHLPCKAFSLAILLAFLLLVGSCTATRTGATMRLNEGSELLRPKQQKPRFPYKGLVF NFFPKGVPIPPSGPSKRHNSLVASTPQN
161	GLYMA20G33661.1	146	MNSSEETVLLLVSFIVLTLIVTLTAYVCSNNIPRTIPSVPGGATHSNNNHTIITVETPEPRLDHTSVRS YPSLQFSKAKLCSSNSNSSSSSCSICLMDYKECDSLRVLPACAHFFHVKCVDPWLRINLTCPVCR TPIALPDICKNL

References

Alfano, J.R., and Colmer, A. (2004). Type III secretion system effector proteins double agents in bacterial disease and plant defense. *Phytopathol.*,**42**; 385–414.

Antony, G., Zhou, J., Huang, S., Li, T., Liu, B., White, F., and Yang, B. (2010). Rice xa13 recessive resistance to bacterial blight is defeated by induction of the disease susceptibility gene Os-11N3. *The Plant Cell*, **22**; 3864-3876.

Blobel, G. and Dobberstein, B. (1975). Transfer of proteins across membranes. I. presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane- bound ribosomes of murine myeloma. *The J.Cell Biol,* **67**; 835-851.

Boch, J., and Bonas, U. (2010). *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. *Phytopathol*., **48**; 419.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**; 1509-1512.

Boller, T., and He, S. Y. (2009). Innate immunity in plants: an arms race between pattern recognition receptors in plants and effectors in microbial pathogens. *Science* **324**; 742-744.

Cai, R., Lewis, J., Yan, S., Liu, H., Clarke, C. R., Campanile, F., Almeida, N.F., Studholme, D.J., Lindeberg, M., Schneider, D. and Zaccardelli, M. (2011). The plant pathogen *Pseudomonas syringae* pv. tomato is genetically monomorphic and under strong selection to evade tomato immunity. *PLoS Pathog*en, **7**; e1002130.

Chen, L. Q., Hou, B. H., Lalonde, S., Takanaga, H., Hartung, M. L., Qu, X. Q., Guo, W.J., Kim, J.G., Underwood, W., Chaudhuri, B. and Chermak, D. (2010). Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*, **468**;527-532.

Cheng, Z., Li, J., Niu, Y., Zhang, X., Woody, O. Z., Xiong, Y., Djonović, S., Millet, Y., Bush, J., McConkey, B.J. Ausubel, F. M. (2015). Pathogen- secreted proteases activate a novel plant immune pathway. *Nature,* **521**; 213-216.

Chinchilla, D., Zipfel, C., Robatzek, S., Kemmerling, B., Thorsten Nürnberger, Jonathan, D. G. J., Felix, G. Boller, T. (2007). A flagellin- induced complex of the receptor FLS2 and BAK1 initiates plant defence. *Nature,* **448**; 497.

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006). Host-microbe interactions: shaping the evolution of the plant immune response. *Cell*, **124**; 803-814.

Christian, M., Cermak, T., Doyle, E. L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. & Voytas, D. F. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**; 757-761.

Chung, J., Staswick, P. E., Graef, G. L., Wysong, D. S., and Specht, J. E. (1998). Inheritance of a disease lesion mimic mutant in soybean. *The Journal of heredity (USA)*.

Dodds, P. N., and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**; 539-548.

Doke, N. (1983). Generation of superoxide anion by potato tuber protoplasts during the hypersensitive response to hyphal wall components of Phytophthora infestans and specific inhibition of the reaction by suppressors of hypersensitivity. *Physiol. Plant Pathol.* **23**; 359-367.

Flor, H. H. (1956). Mutations in flax rust induced by ultraviolet radiation. *Science*, **124**; 888-889.

Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science*, **298**; 2345-2349.

Gu, K., Tian, D., Yang, F. W. U. L., Wu, L., Sreekala, C., Wang, D., Wang, G.L. and Yin, Z. (2004). High-resolution genetic mapping of Xa27 (t), a new bacterial blight resistance gene in rice, Oryza sativa L. *Theo. App Genet.*,**108**, 800-807.

Gu, K., Yang, B., Tian, D., Wu, L., Wang, D., Sreekala, C., Yang, F., Chu, Z., Wang, G.L., White, F.F. and Yin, Z., (2005). R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature*, **435**; 1122-1125.

Gu, K., Sangha, J.S., Li, Y. and Yin, Z., (2008). High-resolution genetic mapping of bacterial blight resistance gene Xa10. *Theo. App. Genet.*, ***116***, pp.155-163.

Gururani, M. A., Venkatesh, J., Upadhyaya, C. P., Nookaraju, A., Pandey, S. K., & Park, S. W. (2012). Plant disease resistance genes: current status and future directions. *Physiol. Mol. Plant Pathol.*, **78**; 51-65.

Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *Nature*, **444**; 323-329.

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., & Minami, E. (2012). Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PloS One*, ***7***; e49423.

Lacombe, S., Rougon-Cardoso, A., Sherwood, E., Peeters, N., Dahlbeck, D., Van Esse, H.P., Smoker, M., Rallapalli, G., Thomma, B.P., Staskawicz, B. and Jones, J.D. (2010). Interfamily transfer of a plant pattern- recognition receptor confers broad- spectrum bacterial resistance. *Nat. Biotech.,* **28**; 365-369.

Laluk, K., Luo, H., Chai, M., Dhawan, R., Lai, Z., and Mengiste, T. (2011). Biochemical and genetic requirements for function of the immune response regulator BOTRYTIS-INDUCED KINASE1 in plant growth, ethylene signaling, and PAMP-triggered immunity in Arabidopsis. *The Plant Cell*, **23**; 2831-2849.

Lamb, C., and Dixon, R. A. (1997). The oxidative burst in plant disease resistance. *Ann. Rev. Plant. Biol*., **48**; 251-275.

Li, T., Huang, S., Jiang, W., Wright, D., Spalding, M., Weeks, D., and Yang, B. (2011). TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res*., **39**; 359-372.

Li, T., Huang, S., Zhao, X., Wright, D. A., Carpenter, S., Spalding, M. H., Weeks, D.P. and Yang, B. (2011). Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*., gkr188.

Lu, D., Wu, S., Gao, X., Zhang, Y., Shan, L., & He, P. (2010). A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proc. Natl. Acad.Sci.,* **107**; 496-501.

Mackey, D., Belkhadir, Y., Alonso, J. M., Ecker, J. R., and Dangl, J. L. (2003). Arabidopsis RIN4 is a target of the type III virulence effector AvrRpt2 and modulates RPS2- mediated resistance. *Cell,* **112**; 379-389.

Marino, D., Dunand, C., Puppo, A., and Pauly, N. (2012). A burst of plant NADPH oxidases. *Trend. Plant Sc*i., **17**, 9-15.

Mejía-Teniente, L., Torres-Pacheco, I., González-Chavira, M. M., Ocampo-Velazquez, R. V., Herrera-Ruiz, G., Chapa-Oliver, A. M., and Guevara-González, R. G. (2015). Use of elicitors as an approach for sustainable agriculture. *African J. Biotech***., 9**, 9155-9162.

Monaghan, J., and Zipfel, C. (2012). Plant pattern recognition receptor complexes at the plasma membrane. *Curr. Opin. Plant. Biol*., **15**, 349-357.

Morbitzer, R., Römer, P., Boch, J., & Lahaye, T. (2010). Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *Proc. Natl. Acad. Sci.*, **107;** 21617-21622.

Mortazavi, A., Brian, A. W., Mccue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Meth., **5***, 621-628.

Moscou, M. J., and Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science,* **326**; 1501-1501.

Mudgett, M.B., (2005). New insights to the function of phytopathogenic bacterial type III effectors in plants. *Plant Biol.*, **56;** 509-531.

Ntoukakis, V., Balmuth, A. L., Mucyn, T. S., Gutierrez, J. R., Jones, A. M., & Rathjen, J. P. (2013). The tomato Prf complex is a molecular trap for bacterial effectors based on Pto transphosphorylation. *PLoS Pathog*, **9**; e1003123.

Pierre, M., Noël, L., Lahaye, T., Ballvora, A., Veuskens, J., Ganal, M., and Bonas, U. (2000). High-resolution genetic mapping of the pepper resistancelocus Bs3 governing recognition of the *Xanthomonas*campestris pv vesicatora AvrBs3 protein. *Theor. App. Genet*., **101:** 255-263.

Postnikova, O. A., & Nemchinov, L. G. (2012). Comparative analysis of microarray data in Arabidopsis transcriptome during compatible interactions with plant viruses. *Virology J.,* **9*:*** 1.

Römer, P., Hahn, S., Jordan, T., Strauß, T., Bonas, U. and Lahaye, T., 2007. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science*, **318**; 645-648.

Salmeron, J. M., Barker, S. J., Carland, F. M., Mehta, A. Y., and Staskawicz, B. J. (1994). Tomato mutants altered in bacterial disease resistance provide evidence for a new locus controlling pathogen recognition. *The Plant Cell*, **6:** 511-520.

Salvo, S. A., Hirsch, C. N., Buell, C. R., Kaeppler, S. M., & Kaeppler, H. F. (2014). Whole transcriptome profiling of maize during early somatic embryogenesis reveals altered expression of stress factors and embryogenesis-related genes. *PloS One*, **9:** e111407.

Sanseverino, W., Hermoso, A., D'Alessandro, R., Vlasova, A., Andolfo, G., Frusciante, L., Lowy, E., Roma, G. & Ercolano, M. R. (2012). PRGdb 2.0: towards a community-based database model for the analysis of R-genes in plants. *Nucl. Acids Res.,* gks1183.

Severin, A. J., Woody, J. L., Bolon, Y., Joseph, B., Diers, B. W., Farmer, A. D., . . . Shoemaker, R. C. (2010). RNA- seq atlas of *glycine max* : A guide to the soybean transcriptome. *BMC Plant Biol.,* **10:** 160-160.

Shi, H., Ye, T., Chen, F., Cheng, Z., Wang, Y., Yang, P., Zhang, Y. and Chan, Z. (2013). Manipulation of arginase expression modulates abiotic stress tolerance in *Arabidopsis*: effect on arginine metabolism and ROS accumulation. *J. Exp. Bot.,* **64:** 1367-1379.

Strauss, T., van Poecke, R. M. P., Strauss A., Römer P., Minsavage G. V., Singh S., *et al*. (2012). RNA- seq pinpoints a *Xanthomonas* TAL-effector activated

resistance gene in a large-crop genome. *Proc. Natl. Acad. Sci.,* USA*,* **109:** 19480-19485.

Sun, Y., Li, L., Macho, A. P., Han, Z., Hu, Z., Zipfel, C., Zhou, J.M. Chai, J. (2013). *Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex*. *Science*, **342:** 624-628.

Suzuki, N., Miller, G., Morales, J., Shulaev, V., Torres, M. A., & Mittler, R. (2011). Respiratory burst oxidases: the engines of ROS signaling. *Curr. Opin. Plant Biol.,* **14:** 691-699.

Tian, D., Wang, J., Zeng, X., Gu, K., Qiu, C., Yang, X., Zhou, Z., Goh, M., Luo, Y., Murata-Hori, M. and White, F.F. (2014). The rice TAL effector–dependent resistance protein XA10 triggers cell death and calcium depletion in the endoplasmic reticulum. *The Plant Cell*, ***26:***, 497-515.

Torres, M. A., and Dangl, J. L. (2005). Functions of the respiratory burst oxidase in biotic interactions, abiotic stress and development. *Curr. Opin. Plant Biol*., **8:** 397-403.

Vassilev, D., Leunissen, J., Atanassov, A., Nenov, A. and Dimov, G., 2005. Application of bioinformatics in plant breeding. *Biotech. Biotechnol. Equip.*, **19**, 139-152.

Vleeshouwers, V. G., Raffaele, S., Vossen, J. H., Champouret, N., Oliva, R., Segretin, M. E., Rietman, H., Cano, L.M., Lokossou, A., Kessel, G. & Pel, M. A. (2011). Understanding and exploiting late blight resistance in the age of effectors. *Ann. Rev. Phytopathol.*, **49:** 507-531.

Wang, C., Fan, Y., Zheng, C., Qin, T., Zhang, X., & Zhao, K. (2014). High-resolution genetic mapping of rice bacterial blight resistance gene Xa23. *Mol. Genet. Genomics*, **289:** 745-753.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA- seq: A revolutionary tool for transcriptomics. *Nat. Rev., Genet*., **10:** 57-63.

Weber, E., Gruetzner, R., Werner, S., Engler, C., & Marillonnet, S. (2011). Assembly of designer TAL effectors by Golden Gate cloning. *PloS one*, ***6:*** e19722.

White, F. F., and Yang, B. (2009). Host and pathogen factors controlling the rice-*Xanthomonas oryzae* interaction. *Plant Physiol*., **150:** 1677-1686.

Xiao, Y., Redman, J. C., Monaghan, E. L., Zhuang, J., Underwood, B. A., Moskal, W. A., Wang, W., Wu, H.C. and Town, C. D. (2010). High throughput generation of promoter reporter (GFP) transgenic lines of low expressing genes in Arabidopsis and analysis of their expression patterns. *Plant Methods, **6*:** 18-18.

Yang, B., Sugio, A., & White, F. F. (2006)**.** Os8N3 is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sc.,* USA *103* 10503-10508.

Zhang, J., Yin, Z., and White, F. (2015). TAL effectors and the executor R genes. *Front. Plant Sci., 6*.

Zhao, Z., Chen, C., Hu, C.Q., Ren, C.H., Zhao, J.J., Zhang, L.P., Jiang, X., Luo, P. and Wang, Q.B. (2010). The type III secretion system of *Vibrio alginolyticus* induces rapid apoptosis, cell rounding and osmotic lysis of fish cells. *Microbiology*, *156*:.2864-2872.

Zhou, J., Peng, Z., Long, J., Sosso, D., Liu, B., Eom, J.S., Huang, S., Liu, S., Vera Cruz, C., Frommer, W.B. and White, F.F.. (2015). Gene targeting by the TAL effector PthXo2 reveals cryptic resistance gene for bacterial blight of rice. *The Plant J*, *82*(4), 632-643.

Zhu, M., and Zhao, S. (2007). Candidate gene identification approach: progress and challenges. *Int. J. Biol. Sci. **3***: 420-427.

Zhu, W., Magbanua, M.M., White, F.F. (2000) Identification of two novel hrp-associated genes in the hrp gene cluster of *Xanthomonas oryzae* pv. *oryzae*. *J Bacteriol. **182*****: 1**844-1853.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D. G., Boller, T., and Felix, G. (2006). Perception of the bacterial PAMP EF- tu by the receptor EFR restricts agrobacterium- mediated transformation. *Cell,* **125:** 749-760.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E. J., Jonathan, D. G. J., Felix,G., and Boller, T. (2004). Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature*, **428:** 764-767.