DISSECTING THE GENETIC BASIS OF SOYBEAN-AGROBACTERIUM COMPATIBILITY

by

TIMOTHY MICHAEL CHAPPELL

(Under the Direction of Wayne Parrott)

ABSTRACT

Soybean is one of the most widely grown crops in the world. The advent of genetic engineering in soybean has allowed genetic manipulation of soybean for a variety of modifications. Development of transgenic soybean plants is primarily through the use of particle bombardment to deliver DNA. This is because soybean is recalcitrant to the most commonly used method for genetic engineering, *Agrobacterium*-mediated transformation. Resistance to the microbe *Agrobacterium* by almost all soybean genotypes implicates the soybean immune system. A single line has been previously identified as extremely susceptible to *Agrobacterium*, the soybean accession Peking, identified as PI 548402. This accession is important for understanding how *Agrobacterium* susceptibility is controlled in most soybean varieties, thus opening the door to genetic engineering of more soybean varieties. Chapter one contains a brief history of soybeans in plant breeding and plant immunity, with a focus on previous work characterizing *Agrobacterium* and soybean interactions. Chapter two describes an identified candidate gene for susceptibility in soybean and subsequent testing with *Agrobacterium*, modified strains, and candidate gene knockouts. Chapter three details the

development of a mapping population derived from a Peking x Century inbred line and a resistant line, Jack or PI 540556, to determine the regions of the genome responsible for susceptibility, with further analysis of candidate genes identified within this region using RNA-Seq. Collectively, these chapters address finding the genomic regions in soybean that control susceptibility to *Agrobacterium*, using this knowledge to identify what changes could be implemented in the future to make transformation of all soybean genotypes possible. With these goals in mind, the objective of expanding *Agrobacterium* susceptibility to previously restricted soybean genotypes can be accomplished.

INDEX WORDS: *Agrobacterium*, Soybean, transformation, Host-microbe interactions, Recombinant Inbred Lines, CRISPR/cas9, Elongation factor thermo unstable, Elongation factor receptor

DISSECTING THE GENETIC BASIS OF SOYBEAN-*AGROBACTERIUM* COMPATIBILITY

by

TIMOTHY MICHAEL CHAPPELL

B.A., Vanderbilt University, 2014

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

DISSECTING THE GENETIC BASIS OF SOYBEAN-*AGROBACTERIUM*

COMPATIBILITY

by

TIMOTHY MICHAEL CHAPPELL

Major Professor:	Wayne Parrott
Committee:	Brian Kvitko
	Zenglu Li
	Peggy Ozias-Akins
	Chung-Jui Tsai

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2023

DEDICATION

In memory of Lane.

I will finish what we both started.

ACKNOWLEDGEMENTS

I would like to thank all who have joined me on this path and the organizations that allowed me the opportunity to pursue my PhD research. First, I want to thank my major advisor Wayne Parrott for giving a young kid with little plant experience a chance to study in this field, and teaching me about plant biotechnology, tissue culture, communication, and most importantly, how to tell a story.

I'm grateful for having great committee members who always asked the hardest questions. Thank you Dr. Kvitko for helping me learn about plant immunity and teaching me how to engineer bacteria. Thank you, Dr. Li, for your help with breeding populations and putting up with my attempts at genetic mapping. Thank you, Dr. Tsai, for somehow making all conversations fun and for asking poignant questions that helped me come to the proper conclusion. Thank you, Dr. Ozias-Akins, for not only being director of IPBGG for most of my time here, but for your continued support as I move forward.

For teaching me the methods necessary to succeed in this field, I must thank Dr. Pete LaFayette. Molecular cloning is a complicated subject and only a guru like you could teach me your ways and still want to drink a beer after work.

I'd like to thank my family and most importantly my parents, Teri & Mike. Without their financial support and help over the years, I would have never accomplished my dream that seemed so impossible 14 years ago.

I never thought it would be possible to go to graduate school in this field, and my undergraduate buddies along with my botany professor, Dr. Lisa Calfee, believed I could. Thank you all for giving me hope and the determination to chase a dream.

The Niki & Zack duo combo really helped me out in my mid-PhD crisis. Without your friendship, a home away from home, food to eat, cats to spay, or random acts of emergency assistance, I would have given up a long time ago.

To all my friends over the years who listened and understood my pains. To those early mornings, late nights, and sporadic breaks in time, I thank you all: Lane, Jacob, Gary, Evan, Brian, Mary, Jeff, Vincent, Hallie.

To Grace, for being the one who had to put up with me the most. I am so sorry to have bored you with work stuff and constant complaints, but I appreciate your willingness to always listen. Thank you so much for supporting me in every way through this process and making sure I retained the last vestiges of who I used to be.

To Helix, Elsie, Pinto, and Butter. My best buds. You defined my time here and hope I can give you what I have promised.

My last and least important acknowledgement is to all the soybean varieties known as Peking. The name alone grants great power and secrecy, one which I both hate and love.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction

Soybean [*Glycine max* (L.) Merr.] is one of the most cultivated crops in the world. As a top commodity crop in the United States in 2014, genetically engineered soybean was planted in 94% of total soybean hectarage, an increase of 77% from 1997 (Fernandez-Cornejo and Wechsler, 2016). This thirty-five billion dollar industry creates valuable goods including vegetable oil, animal feed, and non-food industrial products (Singh, 2010). Being a nitrogen-fixing crop, soybean increases nitrogen levels and benefits future crop yields in a rotation system (Crookston et al., 1991; Gepts et al., 2005). With the advent of genetic engineering and subsequent herbicide tolerant plants allowing a reduction in tilling, soybeans have helped reduce soil erosion while providing tangible benefits to farmers, consumers, and the world.

Due to soybean's susceptibility to many pests and pathogens and lack of genetic diversity for key traits, genetic engineering has been utilized to maintain and increase current production levels (Homrich et al., 2012). Almost all soybean in the United States is genetically modified for herbicide tolerance and oil quality (Bonny, 2008). Biotechnology has been vital to increasing production of soybean but the current technologies for developing genetically modified soybean are inefficient, fall under heavy regulations, and lack a method of introducing larger synthetic genetic constructs (Bawa and Anilakumar, 2013; Hamilton, 1997). Presently, microparticle bombardment or *Agrobacterium* are used to insert foreign DNA into soybean.

Soybean

Soybean was first domesticated in East Asia around 7000 BCE. Evidence supports the first cultivation occurring in China before radiating to nearby countries such as Korea and Japan (Lee 2011). In recent centuries, soybeans have spread worldwide and can be found throughout the Americas, Africa, Europe, and Asia. The major producers of soybean include China, India, Brazil, Argentina and the United States (Hymowitz, 1970).

Soybean was first brought to the United States in 1765 by Samuel Bowen, a sailor who obtained the seeds while visiting China. Bowen provided the first soybean seeds to his friend, Henry Yonge, who planted them near Savannah, Georgia on Skidaway Island. Until the twentieth century, soybeans were primarily grown for forage. In the early 1900's, William Morse and Charles Piper transformed the use of soybean in Western society by revealing its nutritional properties (Piper and Morse, 1923a).

The twentieth century ushered in the modern soybean era. After understanding the advantageous nutritional properties found in the seed, numerous other uses were developed. In addition to being grown for forage or oil, soybeans were now being used for many more industrial applications while becoming a common component in a wide range of food products. Some common consumables include soy sauce, tofu, or edamame (Hartman et al., 2011).

While many uses are well-known including soybean oil and animal feed, industrial applications are now commonplace. Henry Ford in 1931 sought to connect soybean cultivation and agriculture to the car industry. He succeeded in developing soybeans for a wide range of products, including plastics, textiles, and paint. Some more examples developed in recent years are adhesives, upholstery, crayons, and even

hydraulic fluids. What started as a forage crop in 1765 quickly became ubiquitous as agriculture and industry sought to utilize the tangible potential of the soybean (Piper and Morse, 1923a).

Soybean has many essential traits that are selectively bred for in the elite, commercially grown cultivars. These traits are important for maintaining and increasing the yield or production. Since becoming a major crop in the United States beginning in the twentieth century, soybean has been crucial for vegetable oil and meal production (Smith and Huyser, 1987; Wilcox, 2004). Incorporated traits affect yield but can be separated into different categories including agronomic value, disease resistance, or affecting geographical adaptation.

Because soybean yield or quality are the most essential traits, traits that impact or correlate with increased yield or quality are often maintained or incorporated into breeding projects. Typical quality traits such as oil or protein content are measured as they represent the most important products of processed soybean. Yield traits such as pods per plant, seed weight per plant, or 100-seed weight can be used to gauge yield. Other traits that may be measured include plant height, growth period, or branches (Chen et al., 2007).

In addition to quality or yield traits, many traits can be incorporated depending on the region where the soybean will be grown. As an example, geographical distribution affects flowering time and the maturity group that a soybean variety is assigned to. Natural variation regarding flowering time and plant maturity are measured among breeding populations to ensure that the plant is well-adapted for the specific environment where it will be grown (Valliyodan et al., 2016).

While many traits are important for soybean to be properly adapted to its environment in response to abiotic stresses, different regions of the United States also impact breeding in regard to biotic stresses. For example, in the southern United States, breeding focuses on yield but also on the incorporation of traits for pathogen resistance. In the southern United States, root-knot nematodes can affect yield upwards of 5% (Pham et al., 2013). Other economically damaging pathogens in the southern United States include soybean rust, insects, soybean stem canker, or frogeye leaf spot (Pantalone et al., 2017; Wrather and Koenning, 2009).

While soybean breeding will always emphasize yield with good quality traits, the objectives of funding agencies are important considerations. The United Soybean Board (USB) is a collection of farmer-directors who oversee investments resulting from a congressionally-mandated checkoff (Durham, 2003). The most important objectives of USB are capturing the full value of soybean meal and building preference of soybean oil. A third objective, sustainability, results in the need of more traits to fully promote sustainable agriculture. This results in some funds being used to develop traits for transgenic research in addition to traits affecting abiotic or pathogen response (Barnes, 2000; Pantalone et al., 2017). In recent years, especially in earlier maturity groups, soybean protein levels have been of concern. In order for soybeans to stay competitive with newer synthetics or additives in the meal business, the protein level needs to be maintained. Funding agencies are supporting research that can help increase protein levels and quality in order to help soybean farmers stay competitive in the global meal market (Durham, 2003).

The largest producers of soybean are the United States and Brazil (USDA-FAS, World Agricultural Production, 2017). Being the center of domestication for soybean, China maintains the largest and most diverse germplasm collection for soybean and related *Glycine* members. (Oliveira et al., 2010; Qiu et al., 2011). China's collection is in the National Crop Genebank at the Institute of Crop Science, Beijing, China. The breakdown of the curated 31,755 accessions represents 18,780 local landraces, 2370 local breeding lines, 1500 modern Chinese cultivars, 2156 cultivars from other countries, and 6,644 *G. soja* species. Additionally, China maintains three perennial *Glycine* spp. representing 125 accessions (Qiu et al., 2011).

The United States Department of Agriculture (USDA) Agricultural Research Service maintains the United States National Plant Germplasm System. The USDA Soybean Germplasm Collection is located at the University of Illinois containing around 22,000 accessions. These accessions represent 87 countries, with about 93% of the introductions being sourced from outside China. There are around 19,557 *Glycine max* accessions, 1181 *Glycine soja* accessions, and about 1038 representatives of 20 perennial *Glycine* species. Of note is that while China contributes more germplasm than it receives from the United States for many others species, the United States contributes more soybean accessions to China than it receives despite China being the center of diversity (Wang, 2012).

The overall total of accessions worldwide for *Glycine max* is at least 170,000 distributed throughout 70 countries (Nelson, 2009). Besides numerous countries maintaining collections, other entities, such as the World Vegetable Center (AVRDC), also maintain and distribute material. Located in Taiwan and maintaining 3,926 *Glycine*

spp., additional sources outside of national collections help ensure better pathways for moving around germplasm to broaden the gene pool (Tanksley and McCouch, 1997).

Unadapted germplasm has been of considerable importance for breeding desirable variation into cultivated soybean varieties. In the United States, only eight varieties were introduced prior to the United States Department of Agriculture collection beginning in 1898 (Piper and Morse, 1923b). These varieties are Ito San, Mammoth, Butterball, Buckshot, Kingston, Guelph, Eda and Ogemaw. These eight varieties were mostly collected or donated from China, Europe, or Japan (Piper and Morse, 1923a).

In the following few decades, more accessions and plant introductions became available. However, these new varieties were used primarily as donors for pathogen-related resistance. In 1972, it was reported that a total of 11 plant introductions contributed the majority of genetic diversity to the gene pool and that this situation has not changed much in recent years (Duvick, 1977; Johnson and Bernard, 1963). In 1991, a study found that only 62 plant introductions appeared in 221 released cultivars between 1949 and 1988 (Hymowitz and Bernard, 1991).

Of importance to the United States' soybean diversity, the contributions of ancestral lines often vary between the Northern and Southern United States. This is largely due to differences in maturity group and flowering, but also environmental factors. An example of these differences was noted in a study by Gizlice and Burton in 1994. A cultivar named Lincoln contributed genetically to 24.17% of Northern cultivars but only contributed 2.90% to Southern cultivars. A variety named S-100 is found only in 1.75% of Northern cultivars, whereas it is found in 21.31% of Southern cultivars (Gizlice et al., 1994). Thus, the pedigrees of Northern and Southern cultivars in the United States

are often the result of different parents due to regional differences that reflect maturity group.

This regional difference can be understood in terms of nematode resistance. In the North, PI 88788 contributed 0.38% while Peking contributed 0.09%. In the South, PI 88788 contributed 0.74% while Peking accounts for 1.14%. (Gizlice et al., 1994) This difference has often led to nematode resistance as being described from the Peking type, or the PI 88788 type. As resistance to nematodes break down, more unadapted germplasm may need to be evaluated and introgressed. While this serves as an example for breeding unadapted germplasm into breeding populations, many other examples exist in soybean (Concibido et al., 2003; Kuroda et al., 2009; Li, 2000; Li et al., 2010; Sebolt et al., 2000).

In 2002, Concibido described introgression of a quantitative trait locus (QTL) from soybean's wild progenitor, *Glycine soja*, into commercial soybean cultivars. Despite *G. soja* containing many undesirable and poor agronomic traits, the study succeeded in breeding in a locus for higher yield. While this project took many crosses and individuals with the plant introduction 407305, after a series of backcrosses to the elite parent, they found that this *G. soja* locus could increase yields by 9% in certain haplotype backgrounds (Concibido et al., 2003).

In 1995, Tanksley described an advanced backcross method that was used to breed valuable QTL alleles into elite lines. Instead of backcrossing a single time, Tanksley opted to backcross multiple times until selection to increase the elite background genotype in the breeding population. While a powerful tool, Tanksley notes that it requires one to two years to incorporate these QTLs from the wild populations.

However, many advantages are gained with backcrossing wild varieties into elite. This can be used to create more stable near isogenic lines (NILs). Later on in the breeding program, these NILs can be used to incorporate other alleles without requiring additional backcrossing to break up undesirable alleles and linkage drag (Tanksley and Nelson, 1996).

Soybean has been bred using many strategies, depending upon the desired goal or germplasm restrictions. The most used methods include single seed descent (SSD), marker assisted recurrent selection (MARS) or the bulk selection method. While many of these differ, they all rely upon the ability of soybean to self-fertilize and produce viable offspring with low chances of outcrossing. Although varied for potential applications, the following strategies are widely utilized in soybean (Brim, 1966; Hinchee et al., 1988; Wilcox and Cavins, 1995).

These methods above start with creating a hybrid. Depending upon the objective of the breeding program, a good parent is typically crossed with another good parent, whereas they both possess desirable alleles or combinations. By crossing desirable genotypes together, a breeder can create an F_1 hybrid that possesses the desired qualities from both parents. In SSD, an F_1 hybrid creates a large progeny of offspring, creating many combinations of genetic variation from the original parents. Typically, a single seed, but occasionally a pod containing a few seeds, is harvested from each F_2 plant. This ensures that ample genetic variation is captured from each original F_2 plant. This is carried onto later generations when homozygosity for each line nears 100%. Then, at the later generation, typically F_5 or F_6, all the offspring from each plant is planted in individual rows for selecting and finding the desired traits (Brim, 1966).

Marker-assisted recurrent selection is often incorporated into a breeding strategy to ensure that selected progeny or offspring contain a marker for a trait of interest. This allows a breeder to cull undesirable individuals early if they do not contain a known marker, while allowing recombination from the original hybrid to create new genotypes that may incorporate multiple traits of interest into a desired genotype (Concibido et al., 1996)

Depending upon the parents for a particular objective, backcrossing may be useful to bring in new alleles from wild germplasm. This has been used with a *G. soja* introduction, PI 468916, and a *G. max* cultivar, IAE2008. After an initial backcross, any standard breeding procedure, including MARS, bulk selection, or SSD, can be used with the initial backcross or a selected progeny backcrossed to the elite cultivar numerous times. Similar generations of selection or markers can be used to ensure that new desirable traits have been bred into the final population (Wang et al., 2004).

Common methods exist for breeding soybeans that don't rely on hybrid progeny, such as mutagenesis and transgenics. Mutagenesis is often used where, instead of introducing novel introgressions from other germplasm, radiation is used to delete or change the initial variety. This was originally discovered by Lewis Stadler and new methods have been utilized (Stadler and Sprague, 1936). Some current methods in soybean rely upon gamma irradiation to create drastic structural change, while others rely upon transposons to move around the genome, resulting in interruption of alleles or a change in expression (Campbell and Stupar, 2016; Kanizay et al., 2015).

While mutagenesis and breeding schemes rely upon modification or selection of more rare alleles, transgenics in soybean are often used to incorporate alleles from

outside the *Glycine* gene pool. An important example is transgenic soybean containing the *Bacillus thuringiensis* cry1Ac gene (Bt). Bt is used as an insecticide to reduce insect pressure on soybean plants, ultimately reducing necessary spraying of insecticides (Stewart Jr et al., 1996). Another incorporation of transgenics is the use of genome editing technology in soybeans. This was shown in a recent paper by Jacobs in 2015 where an endonuclease was used to knockout a green fluorescent protein in soybean. This work can be used to knockout any gene in soybean and allow functional characterization of the desired gene (Jacobs et al., 2015).

Agrobacterium

The first transgenic soybean was of the agronomically poor variety 'Peking,' and was created in 1988 using *Agrobacterium tumefaciens* (Hinchee et al., 1988). *A. tumefaciens* is a soil-dwelling bacterium that can transmit genes into plant tissue using the virulence and donor genes located on a tumor-inducing (Ti) plasmid. The Ti plasmid encodes virulence proteins that allow the transfer DNA (T-DNA) to be excised from the bacterium and shuttled through a type IV secretion system (T4SS) into the host plant. The major proteins for the tumorigenic ability of *A. tumefaciens* are within the *virulence* (*vir*) region consisting of six operons, *virABCDEG* (Stachel and Nester, 1986).

Four *A. tumefaciens vir* operons are required for virulence on the host plant. Mutations in any of the operons *virABDG* result in a complete loss of tumor formation whereas mutations in *virCE* greatly reduce virulence (Stachel and Nester, 1986). In some *A. tumefaciens* strains, *virF* may be required (Melchers et al., 1990). When *A. tumefaciens* senses plant cells either through recognition of a plant wound-induced phenolic compound, or exposure to acidity and sugars, the transmembrane *virA* protein

activates *virG* (Leroux et al., 1987). Being a two-component sensor system, transcriptionally-active *virG* then proceeds to activate the other proteins that are required for T-DNA transfer (Mantis and Winans, 1992).

The T-DNA of the Ti plasmid has twenty-five base-pair borders which are recognized by the endonuclease *virD2*, directing the protein to the T-DNA site to first cut, and then pilot the single-stranded T-DNA to the T4SS (Yadav et al., 1982; Yanofsky et al., 1986). The T4SS is a complex consisting of *virB* proteins and the remaining *virD* proteins (Cascales et al., 2013). The T4SS uses ATPase activity to deliver the *virD2*-guided T-DNA across the secretion channel into the host cell (Cascales et al., 2013).

The remaining two operons, *virC* and *virE*, are nonessential for virulence (Stachel and Nester, 1986). The *virC* operon encodes two proteins that help increase the number of T-DNAs being generated through binding of the overdrive sequence, while aiding in transport of the strand with *virD2* to the T4SS (Toro et al., 1988). The *virE* operon forms binding proteins, helping to protect the T-DNA during transfer into the host nucleus, while also preventing the binding of *virE1* to *virE2* (Deng et al., 1999; Gietl et al., 1987).

Through a complex sensing system, *A. tumefaciens* has evolved proteins that can respond to plant stimuli and allow the transfer of T-DNA into the host cell using the Ti plasmid's *virulence* region. *A. tumefaciens* hijacks the host machinery to transfer selfish genes, introducing additional genes that encode for the synthesis of opines, cytokinins, and auxins. With opines serving as a nutritional source for the bacterium, the plant hormones auxin and cytokinin induce crown galls, or tumors (Akiyoshi et al., 1984). After it was discovered that *A. tumefaciens* can transmit DNA to the host cell, deletions of the native T-DNA along with introduced binary vectors containing genes of interest, opened the field of plant engineering (Chilton et al., 1977; Hood et al., 1993).

Despite being fine-tuned and used for over 20 years, *Agrobacterium*'s major alternative is still the gene gun, or microprojectile bombardment (Hansen and Wright, 1999). The defining differences between the two transformation methods include variations in copy number insert, length of introduced DNA, and the host range of *A. tumefaciens* (Gelvin, 2008). Copy number inserts can be manipulated in biolistics to favor fewer insertions, but transformation efficiency may drop with a decrease in the amount of DNA, making the use of *A. tumefaciens* favorable when copy number concerns are present (Jackson et al., 2013). While both methods are able to integrate upwards of 150 kb of DNA, *A. tumefaciens* can introduce DNA with less chances of fragmentation and silencing (Hamilton et al., 1996; Mullen et al., 1998).

While *A. tumefaciens* and gene gun technology have been reported to generate high transformation efficiencies in many plant species, elite soybean cultivars still respond poorly to *A. tumefaciens* transformation. A key challenge to transformation is that the DNA must be delivered to cells able to differentiate into whole plants. In soybeans, this ability is limited to a few cells located in the apical and nodal meristems. The use of embryonic tissue leads to a much more efficient transformation, but the ability of *A. tumefaciens* to deliver DNA to these cells is very limited, as prior work suggests that soybean may recognize *A. tumefaciens* as a pathogen, with common defense responses resembling a hypersensitive response that results in death of the cells exposed to *Agrobacterium* (Olhoft et al., 2001). This response can also be visualized in soybean hypocotyls (Fig. 1.1). However, soybean's response to *Agrobacterium* remains unclear, and while common immune responses, such as reactive oxygen bursts and necrosis, are noticed in soybean cells when exposed to *Agrobacterium*, more evidence is needed to claim a hypersensitive response.

Several studies have attempted to pinpoint which genetic regions in *Agrobacterium* induce the host plant defense response. In tobacco, infiltrations of a control strain (ASE/pPZP211) and a Ti plasmid-less strain, A136, found that the control strain slightly increased induction of the pathogenesis-related 1 (*PR-1*) gene (Pruss et al., 2008). While the control strain was disarmed and thus, lacks the T-DNA region, this strain still possessed the *vir* region and the remaining genes encoded on the Ti plasmid, supporting that Ti plasmid genes play a minor role in eliciting host responses. Another study, in *Ageratum*, examined three genes previously identified relating to the plant defense response when infected with *Agrobacterium*. The induction of these genes, a peroxidase, a ribonuclease, and a pathogenesis-related (PR) protein, were measured using RT-PCR of *Ageratum* cells when exposed to different *Agrobacterium* strains. The strains used were A136, cured of the Ti plasmid, the disarmed or lacking T-DNA strain LBA4404, and the disarmed strain EHA105 but containing a binary plasmid with transfer capabilities. When compared to mock-inoculated cells, all strains exhibited a similar induction of the plant defense genes that were being measured. Another strain, similar to A136 but containing a mutation in a chromosomal virulence gene (*chvB*) required for proper attachment to the host cell, induced plant defense responses the most, suggesting that attachment of the bacterium to the host cell possibly dampens plant responses in an attachment-defective manner (Ditt et al., 2005). As plant host responses to bacteria vary widely and similar studies have not focused on soybean, it remains unclear which genetic component or components in *Agrobacterium* induce host responses in soybean.

Soybean and *Agrobacterium*

The bacterium *Agrobacterium tumefaciens* exhibits a wide host range, forming crown gall on thousands of plant species across hundreds of genera and families (De Cleene and De Ley, 1976). The earliest report on testing *Agrobacterium* with soybean was reported in 1936 and found a lack of disease symptoms. Although the author noted a few inoculations of old plants, the conclusion was that soybean is not susceptible to *Agrobacterium* (Lopatin, 1936).

Many pathogenic bacteria rely on accessible binding sites on the host plant cell surface to begin infection. When tissue cultured cells of carrot, tobacco, and soybean were tested, only soybean failed to produce tumors (Matthysse and Gurlitz, 1982). Up until this period, studies relied on using single strains of *Agrobacterium* or single genotypes of soybean.

In 1984, a group of twenty-four soybean cultivars were screened for susceptibility to *Agrobacterium* with the strain A348 (pTiA6). Three genotypes were judged to be highly susceptible- Peking, Jupiter, and Biloxi (Owens and Cress, 1985). A year later, another group infected twenty-seven genotypes of soybean from different maturity groups with six different strains of *Agrobacterium*. Interestingly, they reported the strain A281 as completely ineffective at causing crown gall disease on all genotypes tested. They concluded that the strain A208 was the optimal strain for infecting soybean genotypes, with Peking being highly susceptible (Byrne et al., 1987).

Contrary to the findings of Byrne et al., other studies reported that the strain A281 is hypervirulent on soybean and in particular, with the Peking variety (Hood et al., 1986; Hood et al., 1987; Owens and Smigocki, 1988). All of this previous work eventually led

to the *Agrobacterium*-mediated transformation of soybean in the Peking variety (Hinchee et al., 1988). The history of Peking and high susceptibility to *Agrobacterium* is well-documented and repeatable, making the Peking variety important to understanding the genetic basis of susceptibility to *Agrobacterium* in soybean (Fig. 1.2).

Host-plant Resistance to Pathogens

A great diversity of pathogens are constantly attacking plants. To defend against these pathogens, plants employ a two-branched innate immune system (Jones and Dangl, 2006). The first branch of this immunity uses plant cell-surface transmembrane pattern recognition receptors (PRRs) that recognized microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs) (Jones and Dangl, 2006). Two well-known examples of PRRs that detect bacteria in the environment are the *Arabidopsis* receptors FLS2 and EFR (Zipfel et al., 2006). These receptors and other PRRs bind to widely conserved, slowly evolving MAMPs, which are proteins from bacteria or fungi that cannot be lost without negative fitness consequences to the pathogen. After recognition of a PAMP by a PRR, signaling through plasma-membrane-associated co-receptor kinases and intracellular protein kinases occurs. This ligand-dependent association leads to downstream hallmarks of plant immunity to pathogens or PAMP-triggered immunity (PTI) (Ngou et al., 2021).

The second branch of plant immunity largely happens within the cell using intracellular nucleotide-binding, leucine-rich-repeat-containing receptors (NLRs) (Ngou et al., 2021). These receptors are encoded by polymorphic R genes and act as a sensor for detecting pathogen effectors have either a coiled-coil (CC) domain or a

Toll/interleukin-1 receptor/resistance protein (TIR) domain at their N terminus (Monteiro and Nishimura, 2018). Once a specific effector is directly or indirectly recognized by a NB-LRR protein, effector-triggered immunity (ETI) occurs. While PTI is typically considered to be a basal defense mechanism that aims to prevent colonization by a pathogen, ETI is a faster and stronger response that provides disease resistance and typically results in a hypersensitive response (HR) or cell death (Thomma et al., 2011).

The "zigzag" model was created to represent the current understanding on the plant immune system and is divided into four phases (Jones and Dangl, 2006). The first phase consists of PAMPs being recognized by plant PRRs and leads to PTI, thereby acting a basal defense response that protects against colonization. The second phase is considered as pathogens delivering effectors that increase the pathogen's virulence and can dampen PTI, leading to effector-triggered susceptibility (ETS). In phase three, plants can directly or indirectly recognize a bacterial effector with a NLR, resulting in effector-triggered immunity (ETI). In phase four, pathogens evolve to avoid ETI by acquiring new effectors, or by losing or mutating a recognized effector. While ETI and PTI share many similarities and represent a continuum, they are also sometimes viewed as the same response but varying in strength (Thomma et al., 2011).

The two best characterized MAMP proteins are flagellin (flg22) and elongation factor thermo unstable (EF-Tu), which are recognized by FLS2 and EFR respectively (Nicaise et al., 2009; Zipfel et al., 2006). These proteins vary widely between microbial species and serve as an identification system to the host plant. The EF-Tu protein from *A. tumefaciens* binds to EFR in *Arabidopsis*, and activates defense responses, whereas *efr* mutants exhibit enhanced susceptibility. Some shorter segments of the MAMP protein,

denoted as elongation factor (elf) 18 and 26, activate defensive responses when added to the plant cells, thus identifying the domain that is required for plant perception (Zipfel et al., 2006). A characterization of twenty elf18 sequences from phytopathogenic bacteria that infect *Arabidopsis* found two motifs that elicit much lower activity of the plant immune system, alluding to the ability of different species to evade detection from the host plant (Lacombe et al., 2010).

Identification of MAMP receptors allows disease resistances to be transferred between plant species, while allowing exploitation of plant pathology principles to increase *Agrobacterium*-mediated transformation (AMT) amenability in species that are still recalcitrant. In soybean, the lack of AMT is because of plant immune system induced by *A. tumefaciens.* While almost all soybean varieties are resistant to infection, one soybean variety in particular, Peking, is not resistant. Understanding what makes Peking compatible with *A. tumefaciens* can make it possible to transform all soybean varieties, and possibly other recalcitrant legumes.

Soybean and Rhizobium

The ability of *A. tumefaciens* to induce plant defense responses is hypothesized to result when the soybean immune system differentiates between pathogens and friendly bacteria. Such a differentiation is vital to the formation of nitrogen-fixing nodules with symbionts or to the initiation of plant defenses in the presence of pathogens.

Recently, all members of the *Agrobacterium* genus were reclassified into the *Rhizobium* genus, a taxon that also includes many other legume-specific symbionts involved in nitrogen fixation (Fig. 1.3) (Young et al., 2001). A key distinction of *A.*

tumefaciens, renamed to *Rhizobium radiobacter*, is the exploitation of host plants through the Ti plasmid, resulting in crown galls or tumorigenesis. With the reclassified species forming a monophyletic group within the *Rhizobium*, plants may have evolved the ability to perceive between the pathogenic *A. tumefaciens* and the symbiotic *Rhizobium* that facilitate nitrogen-fixation. Plants that form nitrogen-fixing nodules gain an evolutionary advantage, indicating why many members of the Fabaceae would evolve the ability to determine between symbiotic or pathogenic bacterial species within this monophyletic clade.

The rhizobia responsible for nodulating soybean include the slow-growing *Bradyrhizobium* and the fast-growing *Sinorhizobium* (Chen et al., 1988; Van Berkum and Eardly, 1998). While the relatedness of these rhizobia are often debated, recent reports have identified *Agrobacterium* strains that are capable of effective nodulation of soybean (Young, 1997; Young et al., 2001; Youseif et al., 2014). Amongst these groups of bacteria, many similarities exist between their plasmids. While *Agrobacterium* contains the tumor-inducing plasmid (pTi), certain rhizobia, such as *Sinorhizobium meliloti*, contain symbiotic plasmids, pSymA and pSymB, that are essential for nodulation (Barnett et al., 2001; Finan et al., 2001). Much in the same way that the *Agrobacterium* Ti plasmid encodes a functional system for delivering effectors and DNA using the Type IV secretion system (T4SS), similar extrachromosomal plasmids in *Sinorhizobium fredii* strains HH103, USDA257, and NGR234, encode a Type III secretion system (T3SS) to promote nodulation when infected, similar to *Bradyrhizobium* (Jiménez-Guerrero et al., 2022; Nester, 2015; Vinardell et al., 2015). The transfer of the pSym plasmids into *Agrobacterium* can result in nodules of variable size on a host plant, while transferring of

the Ti plasmid into nodulating species allows them to be capable of transforming plants (Kondorosi et al., 1982; Rudder et al., 2014).

In the context of forming a compatible interaction with soybean, the nodulating symbionts are extremely capable whereas *Agrobacterium* is very ineffective. There are many factors that control these processes, but understanding the differences between these species may reveal a method for making soybeans more amenable and susceptible to *Agrobacterium* (Fig. 1.4). While a few strains of symbiotic bacteria have been capable of transforming various species when a large plasmid encoding *Agrobacterium* machinery is introduced, this may not work for all plant species if the transfer machinery of *Agrobacterium* itself contains the proteins that are currently restricting transformation in recalcitrant species.

A major difference between these bacteria is the presence of a T3SS in the rhizobia. In certain *Bradyrhzobium* strains, the effector NopP is known to be secreted via the T3SS and interacts with a soybean nodulation gene, *Rj2*. When soybean genotypes harboring a specific allele of the *Rj2* gene are inoculated with the *Bradyrhizobium* strain USDA 122, the nopP effector is recognized and ETI occurs, preventing nodulation. If *nopP* is deleted, or certain T3SS structural components are deleted, the strain can nodulate effectively (Sugawara et al., 2018). While many potential effector sequences have been identified, not many have been characterized regarding their target or impact on nodulation. While *Agrobacterium* lacks a T3SS, introducing a T3SS plasmid from rhizobia may be able to increase transformation efficiencies by delivering effectors that dampen the soybean defense response. A similar strategy using the *Pseudomonas*

syringae T3SS in *Agrobacterium* resulted in significant increases in transformation efficiency in wheat, alfalfa, and switchgrass (Raman et al., 2022).

Current research has identified seven different genes that function in controlling nodule formation. The *rj1¸rj5¸and rj6* genes are all recessive, result in non-nodulation, and are nod factor receptors. Another gene, *rj7,* is also recessive but results in hypernodulation. This gene is known to be a LysM RLK and is believed to function in autoregulation of nodulation. These genes impact the ability of any rhizobia and are not genotype-specific (Hayashi et al., 2012).

The three dominant genes function in restricting nodulation and are strain specific. The gene *Rfg1* is allelic to *Rj2,* and is a R protein (TIR-NBS-LRR) that can restrict nodulation with specific *Sinorhizobium* and *Bradyrhizobium* strains (Sugawara et al., 2018). The last two genes, *Rj3* and *Rj4,* can restrict different species of *Bradyrhizobium* but their linkage group and function are still unknown (Hayashi et al., 2012). As most of the nodulation genes rely on effector recognition and ETI, it is unlikely that one of these genes could also be controlling susceptibility to *Agrobacterium* as *Agrobacterium* lacks a T3SS. It is interesting that among soybean lines, the ability to form effective nodules with *Sinorhizobium* is uncommon, and that these lines tend to be susceptible to *Agrobacterium* (Fig. 1.5) (Balatti and Pueppke, 1992).

As previous work has shown that moving pSym plasmids into *Agrobacterium* can result in nodule formation, it is important to note that besides a T3SS, the pSym plasmids typically contain nodulation genes that synthesize Nod factors or lipo-chitin oligosaccharides (LCOs) (Pueppke et al., 1998). Improving soybean susceptibility with soybean may be possible by using specific Nod factor genes from favorable symbionts as

a successful nodulation relies on specific Nod factor recognition for rhizobia to start the infection thread process (Buhian and Bensmihen, 2018).

Another aspect to consider when trying to understand what differences between *Agrobacterium* and the rhizobia allow effective symbiosis but poor susceptibility to *Agrobacterium* is the diversity of exopolysaccharides (EPS) and lipopolysaccharides (LPS). Between rhizobium species and strains, notable differences in the synthesis and decoration of the EPS and LPS occur. These sugars are involved in invasion and nodule development, suppression of plant defense response, and can protect the bacteria from plant antimicrobial compounds (Skorupska et al., 2006). In *Agrobacterium*, it was shown that LPS from virulent bacteria did not interfere with cell attachment, but LPS from avirulent strains inhibited cell attachment (Whatley et al., 1976). The importance of these LPSs in attachment was further supported by mutants with defective pathways that had reduced virulence (Pueppke and Benny, 1984). As bacterial cell walls are made up of LPSs and plant immunity often relies on recognition of external cell wall components, the large variation present within rhizobium is worth considering for fostering compatible interactions. While both EPS and LPS interactions with plants are still inconclusive, these polysaccharides could be involved in *Agrobacterium*-soybean interactions (De Castro et al., 2008).

In summary, the mechanisms by which legumes recognize members of the Rhizobium is much better understood than for *Agrobacterium*. As most plant species do not have a trade-off of recognizing whether the rhizobia are friendly, nitrogen-fixing symbionts or pathogenic, tumor-inducing pathogens, the ability to perceive these various Rhizobium appears to be limited to the legumes. According to the Zig-Zag model,

classical recognition of PAMPs lead to the first plant defense response, but it is well-known that the various symbionts of soybean leverage effectors to dampen plant immunity. While these symbionts possess functional T3SS and deliver effectors, there is currently no evidence to suggest that *Agrobacterium* either possesses a T3SS or delivers effectors to dampen plant immunity. It is well-known that *Agrobacterium* does deliver many proteins to the host cell, but the roles of these proteins are to aid in T-DNA transfer and have not been recognized as being an effector within plant immunity.

For soybeans, the immune response to *Agrobacterium* is known, but through which path it works is not. Understanding whether soybeans undergo PTI or ETI would aid in identifying the necessary genes to study to allow all soybeans to be susceptible to *Agrobacterium*. Another complicated factor is the presentation of nod factors and flavonoids with soybean and Rhizobium. The various symbionts are known to respond differently to soybean genotypes based on the specific soybean's flavonoid presentation. If flavonoids are disregarded, specific effectors and corresponding resistance genes in soybean can allow compatible or incompatible interactions through ETI. While ETI is commonly known for most nitrogen-fixing symbionts and their respective effectors, the mystery of *Agrobacterium* susceptibility remains unsolved. Rhizobium present many various flavors of EPS and LPS, which may be recognized by soybean to select for nitrogen-fixing symbionts over tumor-inducing pathogens.

A true understanding of all the factors that control soybean compatibility with Rhizobium, including *Agrobacterium*, will allow manipulation of either soybean genotypes or *Agrobacterium* species to facilitate a wide range of genotypes amenable to transformation.

References

1. Akiyoshi, D., Klee, H., Amasino, R., Nester, E., & Gordon, M. (1984). T-DNA of *Agrobacterium tumefaciens* encodes an enzyme of cytokinin biosynthesis. *Proceedings of the National Academy of Sciences, 81*(19), 5994-5998.
2. Balatti, P., & Pueppke, S. (1992). Identification of North American soybean lines that form nitrogen-fixing nodules with Rhizobium fredii USDA257. *Canadian Journal of Plant Science, 72*(1), 49-55.
3. Barnes, R. L. (2000). Why the American Soybean Association supports transgenic soybeans. *Pest Management Science, 56*(7), 580-583.
4. Barnett, M. J., Fisher, R. F., Jones, T., Komp, C., Abola, A. P., Barloy-Hubler, F., . . . Gouzy, J. (2001). Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid. *Proceedings of the National Academy of Sciences, 98*(17), 9883-9888.
5. Bawa, A., & Anilakumar, K. (2013). Genetically modified foods: safety, risks and public concerns—a review. *Journal of food science and technology, 50*(6), 1035-1046.
6. Bonny, S. (2008). Genetically modified glyphosate-tolerant soybean in the USA: adoption factors, impacts and prospects. A review. *Agronomy for Sustainable Development, 28*(1), 21-32.
7. Brim, C. A. (1966). A modified pedigree method of selection in soybeans. *Crop Science, 6*(2), 220-220.

8. Buhian, W. P., & Bensmihen, S. (2018). Mini-review: nod factor regulation of phytohormone signaling and homeostasis during rhizobia-legume symbiosis. *Frontiers in Plant Science, 9*, 1247.

9. Byrne, M. C., McDonnell, R. E., Wright, M. S., Carnes, M. G. J. P. C., Tissue, & Culture, O. (1987). Strain and cultivar specificity in the Agrobacterium-soybean interaction. *8*(1), 3-15.

10. Campbell, B. W., & Stupar, R. M. (2016). Soybean (Glycine max) Mutant and Germplasm Resources: Current Status and Future Prospects. *Current Protocols in Plant Biology*, 307-327.

11. Cascales, E., Atmakuri, K., Sarkar, M. K., & Christie, P. J. (2013). DNA substrate-induced activation of the Agrobacterium VirB/VirD4 type IV secretion system. *Journal of Bacteriology*, *195*(11), 2691-2704.

12. Chen, Q.-S., Zhang, Z.-C., Liu, C.-Y., Xin, D.-W., Qiu, H.-M., Shan, D.-P., . . . Hu, G.-H. (2007). QTL analysis of major agronomic traits in soybean. *Agricultural Sciences in China*, *6*(4), 399-405.

13. Chen, W., Yan, G., & Li, J. (1988). Numerical taxonomic study of fast-growing soybean rhizobia and a proposal that Rhizobium fredii be assigned to Sinorhizobium gen. nov. *International Journal of Systematic and Evolutionary Microbiology*, *38*(4), 392-397.

14. Chilton, M.-D., Drummond, M. H., Merlo, D. J., Sciaky, D., Montoya, A. L., Gordon, M. P., & Nester, E. W. (1977). Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell*, *11*(2), 263-271.

15. Concibido, V., Denny, R., Lange, D., Orf, J., & Young, N. (1996). RFLP mapping and marker-assisted selection of soybean cyst nematode resistance in PI 209332. *Crop Science, 36*(6), 1643-1650.

16. Concibido, V., La Vallee, B., Mclaird, P., Pineda, N., Meyer, J., Hummel, L., . . . Delannay, X. (2003). Introgression of a quantitative trait locus for yield from Glycine soja into commercial soybean cultivars. *Theoretical and Applied Genetics, 106*(4), 575-582.

17. Crookston, R., Kurle, J., Copeland, P., Ford, J., & Lueschen, W. (1991). Rotational cropping sequence affects yield of corn and soybean. *Agronomy Journal, 83*(1), 108-113.

18. De Castro, C., Molinaro, A., Lanzetta, R., Silipo, A., & Parrilli, M. (2008). Lipopolysaccharide structures from Agrobacterium and Rhizobiaceae species. *Carbohydrate Research, 343*(12), 1924-1933.

19. De Cleene, M., & De Ley, J. (1976). The host range of crown gall. *The Botanical Review, 42*, 389-466.

20. Deng, W., Chen, L., Peng, W. T., Liang, X., Sekiguchi, S., Gordon, M. P., . . . Nester, E. W. (1999). VirE1 is a specific molecular chaperone for the exported single-stranded-DNA-binding protein VirE2 in Agrobacterium. *Molecular Microbiology, 31*(6), 1795-1807.

21. Ditt, R. F., Nester, E., & Comai, L. (2005). The plant cell defense and Agrobacterium tumefaciens. *FEMS Microbiology Letters, 247*(2), 207-213.

22. Durham, D. (2003). The United Soybean Board's better bean initiative: Building United States soybean competitiveness from the inside out.

23. Duvick, D. (1977). Major united states crops in 1976. *Annals of the New York Academy of Sciences, 287*(1), 86-96.

24. Fernandez-Cornejo, J., & Wechsler, S. J. (2016). USDA Economic Research Service-Recent Trends in GE Adoption.

25. Finan, T. M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorhölter, F. J., . . . Gouzy, J. (2001). The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endosymbiont Sinorhizobium meliloti. *Proceedings of the National Academy of Sciences, 98*(17), 9889-9894.

26. Gelvin, S. B. (2008). *Agrobacterium*-mediated DNA transfer, and then some. *Nature biotechnology, 26*(9), 998-1000.

27. Gepts, P., Beavis, W. D., Brummer, E. C., Shoemaker, R. C., Stalker, H. T., Weeden, N. F., & Young, N. D. (2005). Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. *Plant Physiology, 137*(4), 1228-1235.

28. Gietl, C., Koukolíková-Nicola, Z., & Hohn, B. (1987). Mobilization of T-DNA from Agrobacterium to plant cells involves a protein that binds single-stranded DNA. *Proceedings of the National Academy of Sciences, 84*(24), 9006-9010.

29. Gizlice, Z., Carter, T., & Burton, J. (1994). Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Science, 34*(5), 1143-1151.

30. Hamilton, C. M. (1997). A binary-BAC system for plant transformation with high-molecular-weight DNA. *Gene, 200*(1-2), 107-116. doi:10.1016/s0378-1119(97)00388-0

31. Hamilton, C. M., Frary, A., Lewis, C., & Tanksley, S. D. (1996). Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proceedings of the National Academy of Sciences, 93*(18), 9975-9979.

32. Hansen, G., & Wright, M. S. (1999). Recent advances in the transformation of plants. *Trends in plant science, 4*(6), 226-231. doi:10.1016/s1360-1385(99)01412-0

33. Hartman, G. L., West, E. D., & Herman, T. K. (2011). Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. *Food Security, 3*(1), 5-17.

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

35. Hinchee, M. A., Connor-Ward, D. V., Newell, C. A., McDonnell, R. E., Sato, S. J., Gasser, C. S., . . . Horsch, R. B. (1988). Production of transgenic soybean plants using *Agrobacterium*-mediated DNA transfer. *Bio/technology, 6*(8), 915-922.

36. Homrich, M. S., Wiebke-Strohm, B., Weber, R. L. M., & Bodanese-Zanettini, M. H. (2012). Soybean genetic transformation: A valuable tool for the functional study of genes and the production of agronomically improved plants. *Genetics and molecular biology, 35*(4), 998-1010.

37. Hood, E. E., Chilton, W. S., Chilton, M. D., & Fraley, R. T. (1986). Transfer DNA and opine synthetic loci in tumors incited by Agrobacterium tumefaciens A281 on soybean and alfalfa plants. *Journal of Bacteriology, 168*(3), 1283-1290.

38. Hood, E. E., Fraley, R. T., & Chilton, M. D. (1987). Virulence of Agrobacterium tumefaciens strain A281 on legumes. *Plant Physiology, 83*(3), 529-534. doi:10.1104/pp.83.3.529

39. Hood, E. E., Gelvin, S. B., Melchers, L. S., & Hoekema, A. (1993). New *Agrobacterium* helper plasmids for gene-transfer to plants. *Transgenic Research, 2*(4), 208-218. doi:10.1007/bf01977351

40. Hymowitz, T. (1970). On the domestication of the soybean. *Economic Botany, 24*(4), 408-421.

41. Hymowitz, T., & Bernard, R. (1991). Origin of the soybean and germplasm introduction and development in North America. *Use of Plant Introductions in Cultivar Development Part 1*(useofplantintro1), 147-164.

42. Jackson, M. A., Anderson, D. J., & Birch, R. G. (2013). Comparison of *Agrobacterium* and particle bombardment using whole plasmid or minimal cassette for production of high-expressing, low-copy transgenic plants. *Transgenic Research, 22*(1), 143-151.

43. Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., & Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC biotechnology, 15*(1), 1.

44. Jiménez-Guerrero, I., Medina, C., Vinardell, J. M., Ollero, F. J., & López-Baena, F. J. (2022). The Rhizobial Type 3 Secretion System: The Dr. Jekyll and Mr. Hyde in the Rhizobium–Legume Symbiosis. *International Journal of Molecular Sciences, 23*(19), 11089.

45. Johnson, H., & Bernard, R. (1963). Soybean genetics and breeding. The Soybean: Genetics, Breeding, Physiology, Nutrition and Management. Ed. by AG Norman. In: New York: Academic Press.

46. Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *Nature*, *444*(7117), 323-329.

47. Kanizay, L. B., Jacobs, T. B., Gillespie, K., Newsome, J. A., Spaid, B. N., & Parrott, W. A. (2015). HtStuf: High-throughput sequencing to locate unknown DNA junction fragments. *The Plant Genome, 8*(1).

48. Kondorosi, A., Kondorosi, E., Pankhurst, C., Broughton, W., & Banfalvi, Z. (1982). Mobilization of a Rhizobium meliloti megaplasmid carrying nodulation and nitrogen fixation genes into other rhizobia and Agrobacterium. *Molecular and General Genetics MGG, 188*, 433-439.

49. Kuroda, Y., Tomooka, N., Kaga, A., Wanigadeva, S., & Vaughan, D. A. (2009). Genetic diversity of wild soybean (Glycine soja Sieb. et Zucc.) and Japanese cultivated soybeans [G. max (L.) Merr.] based on microsatellite (SSR) analysis and the selection of a core collection. *Genetic Resources and Crop Evolution, 56*(8), 1045-1055.

50. Lacombe, S., Rougon-Cardoso, A., Sherwood, E., Peeters, N., Dahlbeck, D., Van Esse, H. P., . . . Staskawicz, B. (2010). Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. *Nature biotechnology, 28*(4), 365-369.

51. Leroux, B., Yanofsky, M., Winans, S., Ward, J., Ziegler, S., & Nester, E. (1987). Characterization of the virA locus of Agrobacterium tumefaciens: a

transcriptional regulator and host range determinant. *The EMBO journal*, *6*(4), 849.

52. Li, K. W. F. (2000). General Situation of Wild Soybean (G. soja) Germplasm Resources and Its Utilization of Introgression Into Cultivated Soybean in China [J]. *Review of China Agricultural Science and Technology*, *6*, 020.

53. Li, Y. H., Li, W., Zhang, C., Yang, L., Chang, R. Z., Gaut, B. S., & Qiu, L. J. (2010). Genetic diversity in domesticated soybean (Glycine max) and its wild progenitor (Glycine soja) for simple sequence repeat and single-nucleotide polymorphism loci. *New phytologist*, *188*(1), 242-253.

54. Lopatin, M. (1936). The susceptibility of plants to Bact. tumefaciens, the causative agent of the root-cancer of plants. *Mikrobiologia (Moskva)*, *5*, 716-724.

55. Mantis, N. J., & Winans, S. C. (1992). The Agrobacterium tumefaciens vir gene transcriptional activator virG is transcriptionally induced by acid pH and other stress stimuli. *Journal of Bacteriology*, *174*(4), 1189-1196.

56. Matthysse, A. G., & Gurlitz, R. (1982). Plant cell range for attachment of Agrobacterium tumefaciens to tissue culture cells. *Physiological Plant Pathology*, *21*(3), 381-387.

57. Melchers, L. S., Maroney, M. J., den Dulk-Ras, A., Thompson, D. V., van Vuuren, H. A., Schilperoort, R. A., & Hooykaas, P. J. (1990). Octopine and nopaline strains of Agrobacterium tumefaciens differ in virulence; molecular characterization of the virF locus. *Plant Molecular Biology*, *14*(2), 249-259.

58. Monteiro, F., & Nishimura, M. T. (2018). Structural, functional, and genomic diversity of plant NLR proteins: an evolved resource for rational engineering of plant immunity. *Annual review of phytopathology, 56*, 243-267.

59. Mullen, J., Adam, G., Blowers, A., & Earle, E. (1998). Biolistic transfer of large DNA fragments to tobacco cells using YACs retrofitted for plant transformation. *Molecular Breeding, 4*(5), 449-457.

60. Nelson, R. (2009). *Collection, conservation, and evaluation of soybean germplasm.* Paper presented at the The Abstract of Proceedings. Beijing: The Eight World Soybean Research Conference.

61. Nester, E. W. (2015). Agrobacterium: nature's genetic engineer. *Frontiers in Plant Science, 5*, 16. doi:10.3389/fpls.2014.00730

62. Ngou, B. P. M., Ahn, H.-K., Ding, P., & Jones, J. D. (2021). Mutual potentiation of plant immunity by cell-surface and intracellular receptors. *Nature, 592*(7852), 110-115.

63. Nicaise, V., Roux, M., & Zipfel, C. (2009). Recent advances in PAMP-triggered immunity against bacteria: pattern recognition receptors watch over and raise the alarm. *Plant Physiology, 150*(4), 1638-1647.

64. Olhoft, P., Lin, K., Galbraith, J., Nielsen, N., & Somers, D. (2001). The role of thiol compounds in increasing Agrobacterium-mediated transformation of soybean cotyledonary-node cells. *Plant Cell Reports, 20*(8), 731-737.

65. Oliveira, M. F., Nelson, R. L., Geraldi, I. O., Cruz, C. D., & de Toledo, J. F. F. (2010). Establishing a soybean germplasm core collection. *Field crops research, 119*(2), 277-289.

66. Owens, L. D., & Cress, D. E. (1985). Genotypic variability of soybean response to Agrobacterium strains harboring the Ti or Ri plasmids. *Plant Physiology, 77*(1), 87-94.

67. Owens, L. D., & Smigocki, A. C. (1988). Transformation of soybean cells using mixed strains of Agrobacterium tumefaciens and phenolic compounds. *Plant Physiology, 88*(3), 570-573.

68. Pantalone, V., Smallwood, C., & Fallen, B. (2017). Development of 'Ellis' Soybean with High Soymeal Protein, Resistance to Stem Canker, Southern Root Knot Nematode, and Frogeye Leaf Spot. *Journal of Plant Registrations*.

69. Pham, A.-T., McNally, K., Abdel-Haleem, H., Boerma, H. R., & Li, Z. (2013). Fine mapping and identification of candidate genes controlling the resistance to southern root-knot nematode in PI 96354. *Theoretical and Applied Genetics, 126*(7), 1825-1838.

70. Piper, C., & Morse, W. (1923a). The soybean. 329 pp. *New York: McCraw-Hill Book CO*.

71. Piper, C. V., & Morse, W. J. (1923b). *The soybean*: McGraw-Hill book Company, Incorporated.

72. Pruss, G. J., Nester, E. W., & Vance, V. (2008). Infiltration with Agrobacterium tumefaciens induces host defense and development-dependent responses in the infiltrated zone. *Molecular Plant-Microbe Interactions, 21*(12), 1528-1538.

73. Pueppke, S. G., & Benny, U. K. (1984). Adsorption of tumorigenic Agrobacterium tumefaciens cells to susceptible potato tuber tissues. *Canadian Journal of Microbiology, 30*(8), 1030-1037.

74. Pueppke, S. G., Bolanos-Vásquez, M. C., Werner, D., Bec-Ferté, M.-P., Promé, J.-C., & Krishnan, H. B. (1998). Release of flavonoids by the soybean cultivars McCall and Peking and their perception as signals by the nitrogen-fixing symbiont Sinorhizobium fredii. *Plant Physiology, 117*(2), 599-606.

75. Qiu, L.-J., Chen, P.-Y., Liu, Z.-X., Li, Y.-H., Guan, R.-X., Wang, L.-H., & Chang, R.-Z. (2011). The worldwide utilization of the Chinese soybean germplasm collection. *Plant Genetic Resources, 9*(1), 109-122.

76. Raman, V., Rojas, C. M., Vasudevan, B., Dunning, K., Kolape, J., Oh, S., . . . Pant, B. D. (2022). Agrobacterium expressing a type III secretion system delivers Pseudomonas effectors into plant cells to enhance transformation. *Nature communications, 13*(1), 2581.

77. Rudder, S., Doohan, F., Creevey, C. J., Wendt, T., & Mullins, E. (2014). Genome sequence of Ensifer adhaerens OV14 provides insights into its ability as a novel vector for the genetic transformation of plant genomes. *Bmc Genomics, 15*. doi:10.1186/1471-2164-15-268

78. Sebolt, A., Shoemaker, R., & Diers, B. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Science, 40*(5), 1438-1444.

79. Singh, G. (2010). *The soybean: botany, production and uses.*: CABI.

80. Skorupska, A., Janczarek, M., Marczak, M., Mazur, A., & Król, J. (2006). Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microbial cell factories, 5*, 1-19.

81. Smith, K. J., & Huyser, W. (1987). World distribution and significance of soybean. *Agronomy*.

82. Stachel, S. E., & Nester, E. W. (1986). The genetic and transcriptional organization of the vir region of the A6 Ti plasmid of Agrobacterium tumefaciens. *The EMBO journal, 5*(7), 1445.

83. Stadler, L., & Sprague, G. (1936). Genetic Effects of Ultra-Violet Radiation in Maize I. Unfiltered Radiation. *Proceedings of the National Academy of Sciences, 22*(10), 572-578.

84. Stewart Jr, C. N., Adang, M. J., All, J. N., Boerma, H. R., Cardineau, G., Tucker, D., & Parrott, W. A. (1996). Genetic transformation, recovery, and characterization of fertile soybean transgenic for a synthetic Bacillus thuringiensis cryIAc gene. *Plant Physiology, 112*(1), 121-129.

85. Sugawara, M., Takahashi, S., Umehara, Y., Iwano, H., Tsurumaru, H., Odake, H., . . . Yamakawa, T. (2018). Variation in bradyrhizobial NopP effector determines symbiotic incompatibility with Rj2-soybeans via effector-triggered immunity. *Nature communications, 9*(1), 3139.

86. Tanksley, S., & Nelson, J. (1996). Advanced backcross QTL analysis: a method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics, 92*(2), 191-203.

87. Tanksley, S. D., & McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science, 277*(5329), 1063-1066.

88. Thomma, B. P., Nürnberger, T., & Joosten, M. H. (2011). Of PAMPs and effectors: the blurred PTI-ETI dichotomy. *The plant cell, 23*(1), 4-15.

89. Toro, N., Datta, A., Yanofsky, M., & Nester, E. (1988). Role of the overdrive sequence in T-DNA border cleavage in Agrobacterium. *Proceedings of the National Academy of Sciences, 85*(22), 8558-8562.

90. Valliyodan, B., Qiu, D., Patil, G., Zeng, P., Huang, J., Dai, L., . . . Song, L. (2016). Landscape of genomic diversity and trait discovery in soybean. *Scientific reports, 6*.

91. Van Berkum, P., & Eardly, B. D. (1998). Molecular evolutionary systematics of the Rhizobiaceae. *The Rhizobiaceae: Molecular biology of model plant-associated bacteria*, 1-24.

92. Vinardell, J.-M., Acosta-Jurado, S., Zehner, S., Goettfert, M., Becker, A., Baena, I., . . . Jaenicke, S. (2015). The Sinorhizobium fredii HH103 genome: A comparative analysis with S. fredii strains differing in their symbiotic behavior with soybean. *Molecular Plant-Microbe Interactions, 28*(7), 811-824.

93. Wang, D., Graef, G., Procopiuk, A., & Diers, B. (2004). Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theoretical and Applied Genetics, 108*(3), 458-467.

94. Wang, F. Y. (2012). Flows of Crop Germplasm Resources into/out of China. *Journal of Plant Genetic Resources, 13*(3), 335-342.

95. Whatley, M., Bodwin, J., Lippincott, B., & Lippincott, J. (1976). Role of Agrobacterium cell envelope lipopolysaccharide in infection site attachment. *Infection and immunity, 13*(4), 1080-1083.

96. Wilcox, J. R. (2004). World distribution and trade of soybean. *Soybeans: Improvement, production, and uses*(soybeansimprove), 1-14.

97. Wilcox, J. R., & Cavins, J. F. (1995). Backcrossing high seed protein to a soybean cultivar. *Crop Science, 35*(4), 1036-1041.

98. Wrather, J., & Koenning, S. (2009). Effects of diseases on soybean yields in the United States 1996 to 2007. *Virus, 2*(2,144,000), 9.

99. Yadav, N. S., Vanderleyden, J., Bennett, D. R., Barnes, W. M., & Chilton, M.-D. (1982). Short direct repeats flank the T-DNA on a nopaline Ti plasmid. *Proceedings of the National Academy of Sciences, 79*(20), 6322-6326.

100. Yanofsky, M. F., Porter, S. G., Young, C., Albright, L. M., Gordon, M. P., & Nester, E. W. (1986). The virD operon of Agrobacterium tumefaciens encodes a site-specific endonuclease. *Cell, 47*(3), 471-477.

101. Young, J. (1997). Phylogeny and taxonomy of rhizobia. *Current Issues in Symbiotic Nitrogen Fixation*, 45-52.

102. Young, J., Kuykendall, L., Martinez-Romero, E., Kerr, A., & Sawada, H. (2001). A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of Agrobacterium Conn 1942 and Allorhizobium undicola de Lajudie et al. 1998 as new combinations: Rhizobium radiobacter, R. rhizogenes, R. rubi, R. undicola and R. vitis. *International Journal of Systematic and Evolutionary Microbiology, 51*(1), 89-103.

103. Youseif, S. H., Abd El-Megeed, F. H., Ageez, A., Mohamed, Z. K., Shamseldin, A., & Saleh, S. A. (2014). Phenotypic characteristics and genetic diversity of

rhizobia nodulating soybean in Egyptian soils. *European Journal of Soil Biology, 60*, 34-43.

104. Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D., Boller, T., & Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell, 125*(4), 749-760.

Fig. 1.1. Visual phenotype comparison between non-inoculated and inoculated soybean hypocotyls. The left image was mock inoculated, while the right image was inoculated with *Agrobacterium tumefaciens* A281.

Fig. 1.2. Comparison of typical resistant soybean genotype to the susceptible Peking. Jack is an embryogenic line but highly resistant to *Agrobacterium tumefaciens* A281. Peking is extremely susceptible to *Agrobacterium tumefaciens* A281.

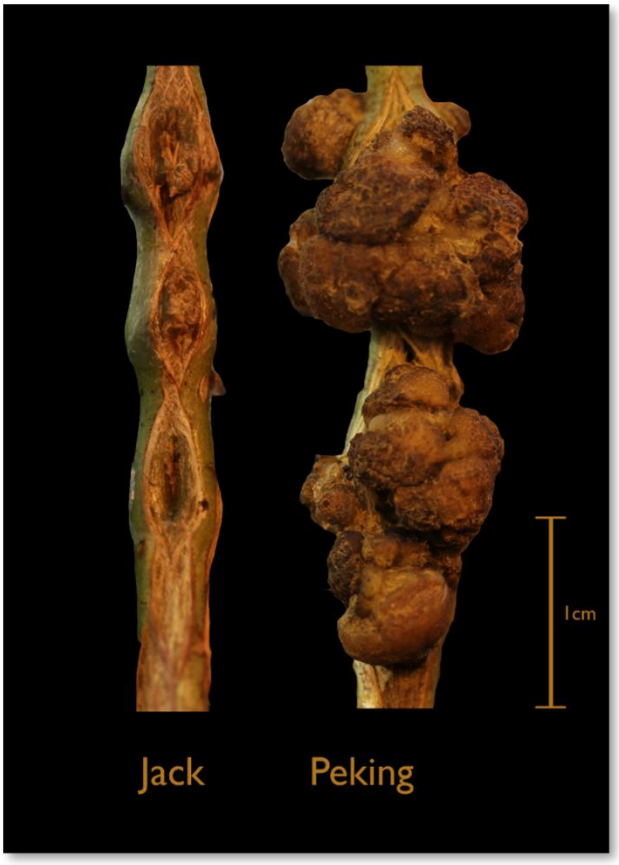


Fig. 1.3. Phylogeny constructed based on Rhizobiaceae species EF-Tu protein sequence using global alignment with the BLOSUM62 substitution matrix. The genetic distance model is Jukes-Cantor built with the Neighbor-Joining method with *Bradyrhizobium japonicum* USDA110 set as the outgroup. Numbers represent the number of amino acid substitutions per site. The rhizobium that form symbioses with soybean are closely related.

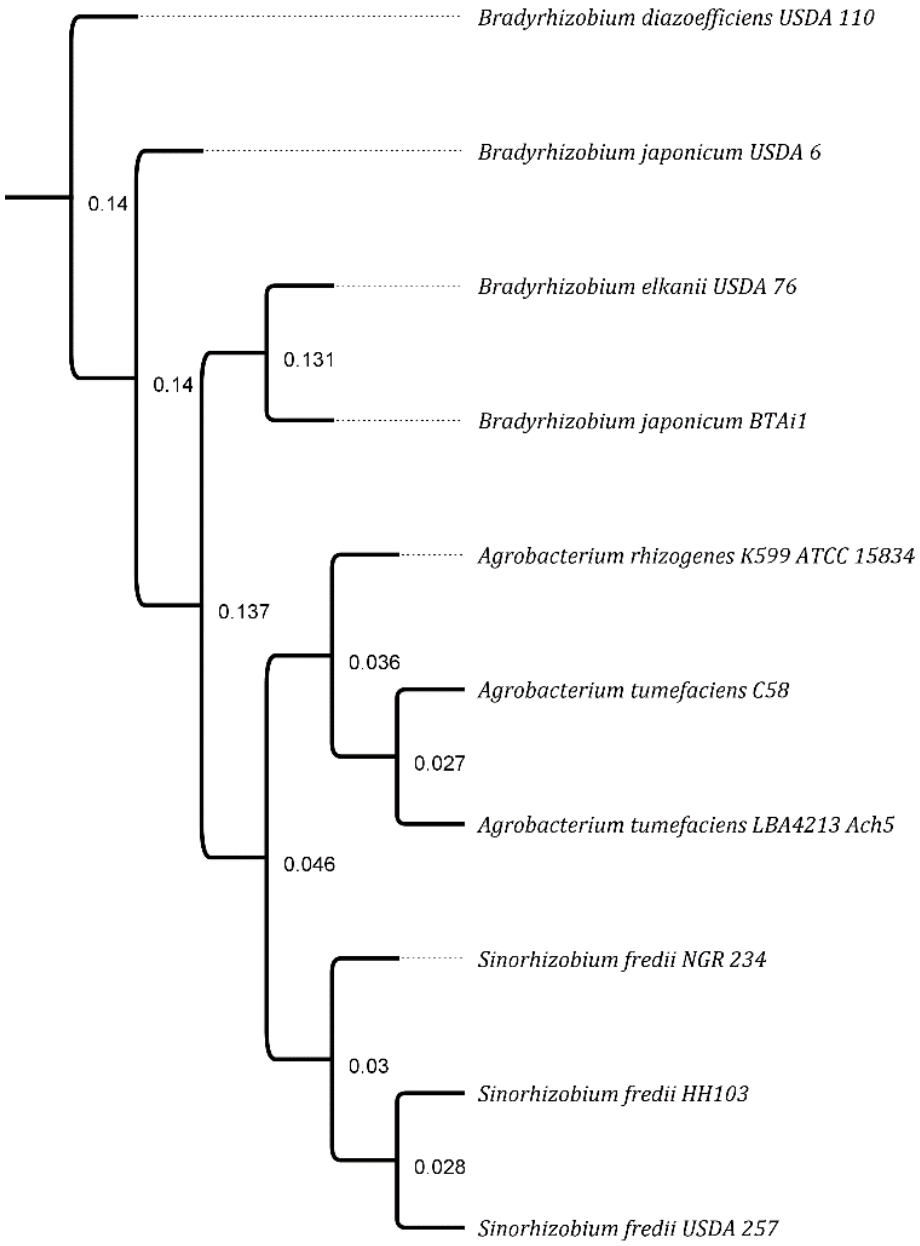


Fig. 1.4. Soybean genotypes are specific for certain species of nodulating bacteria. Bacteria highlighted in red form nitrogen-fixing relationships, while *Agrobacterium,* highlighted in blue, is a pathogen that produces crown gall disease. Many genotypes that are compatible with *Agrobacterium* strains are also compatible with *Sinorhizobium*.

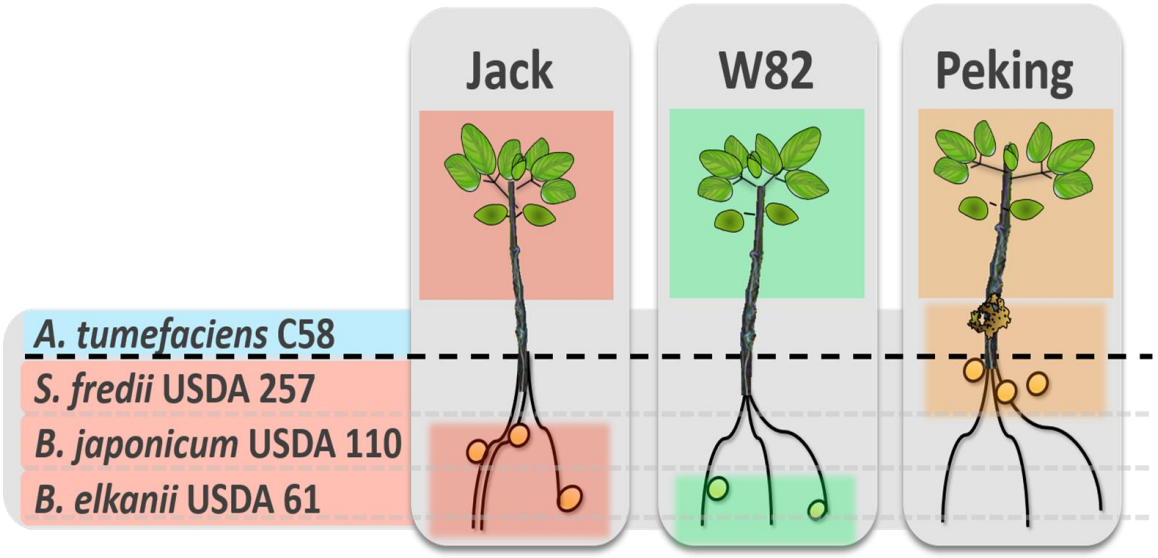
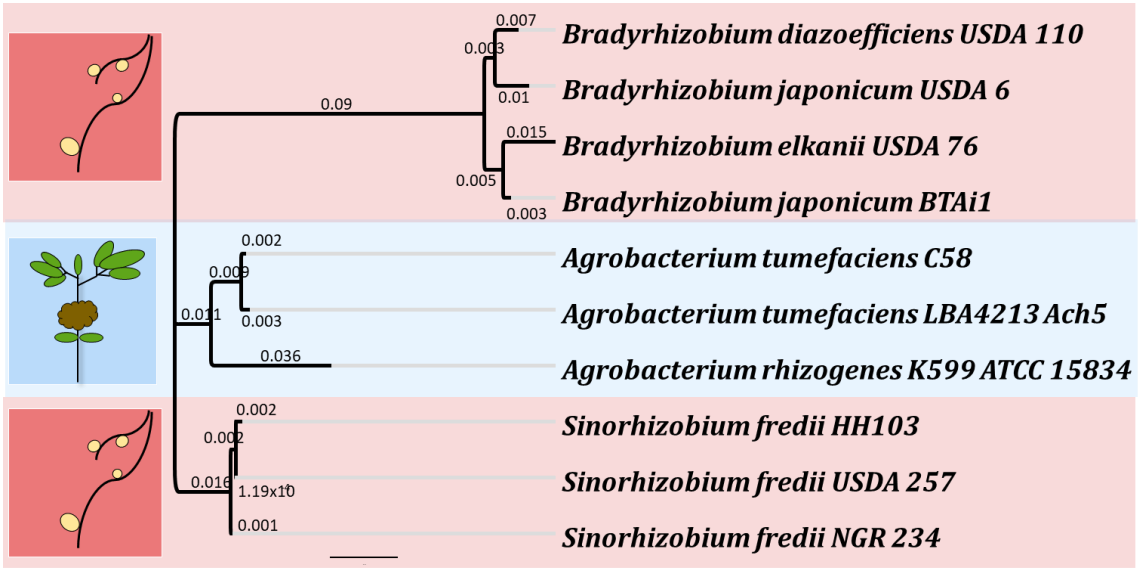


Fig. 1.5. *Agrobacterium* is closely related to both *Bradyrhizobium* and *Sinorhizobium*. Bacteria highlighted in red form nitrogen-fixing relationships, while *Agrobacterium*, highlighted in blue, is a pathogen that produces crown galls and hairy roots.



CHAPTER 2

ENHANCING AGROBACTERIUM-MEDIATED TRANSFORMATION IN SOYBEAN: INSIGHTS FROM COMPARATIVE GENOMICS AND EFR HOMOLOG ASSESSMENT [1]

[1]Timothy M. Chappell, Brian H. Kvitko, and Wayne A. Parrott. To be submitted to *Frontiers in Plant Science*.

Abstract

The ability of *Agrobacterium* to transform soybean [*Glycine max* (L.) Merr.] is a major limitation for genetic engineering. The *Agrobacterium tumefaciens* microbe-associated molecular pattern elongation factor Tu (EF-Tu) is known to be recognized by the arabidopsis pattern recognition receptor (PRR), EF-Tu receptor (EFR), resulting in plant defense responses that reduce transformation in arabidopsis. Comparative genomics identified an EFR homolog (GmEFR, *Glyma.09g216400*) that was absent across susceptible soybean genotypes. Engineered *Agrobacterium* strains were made using EF-Tu sequences from soybean symbionts to escape EFR perception. The engineered strains retained the ability to transform tobacco but were no different than wild-type strains for transforming soybean. CRISPR/Cas9 mutagenesis of GmEFR in a genotype resistant to *Agrobacterium* failed to increase the transformation efficiency with wild-type *Agrobacterium*. The results suggest that a homolog of Arabidopsis EFR in soybean (GmEFR) alone does not control the soybean host defense response to *Agrobacterium*. The strains constructed in this study may be of use in other crops, particularly the Brassicaceae family.

Introduction

Agrobacterium tumefaciens is a versatile bacterium widely used for genetic engineering due to its ability to transfer DNA into plant cells. In Arabidopsis, the EF-Tu receptor (AtEFR) plays a crucial role in recognizing the bacterial protein elongation factor thermo unstable (EF-Tu), an example of a microbe-associated molecular pattern (MAMP). AtEFR specifically recognizes a conserved region known as elf18, located at the N-terminal of EF-Tu from *Agrobacterium tumefaciens*. Notably, inactivation of AtEFR has been shown to enhance the efficiency of *Agrobacterium*-mediated transformation (AMT) (Zipfel et al., 2006)

While AtEFR has only been characterized in the Brassicaceae family, little is known about similar genes in other species. In soybean, several genes share similarities with AtEFR, with *Glyma.09g216400* being the most closely related soybean gene, sharing approximately 66.3% sequence identity with AtEFR, and hence termed GmEFR. Interestingly, GmEFR has been found to be absent in susceptible soybean genotypes, suggesting its potential role in modulating *Agrobacterium* resistance in recalcitrant genotypes.

The objective of this study was to investigate the hypothesis that the susceptibility of Peking could be attributed to the absence of GmEFR. To test this hypothesis, we employed CRISPR/Cas9 in the highly regenerable cultivar, Jack, to create GmEFR knockout plants. Additionally, novel *Agrobacterium* strains were engineered to encode different alleles of EF-Tu, replacing the wild-type EF-Tu with EF-Tu genes from nodulating symbionts compatible with soybean. The engineered strains would bypass the

need to breed or engineer susceptibility into the genotype of interest, expanding transformation and requiring less time to engineer soybean.

Results

Identification of GmEFR across genotypes

The soybean genotypes Jack (PI 540566) and Peking (PI 548402) exhibit a repeatable yet different host response upon inoculation with *Agrobacterium*, making these lines excellent genotypes to decipher the genetics underlying susceptibility (Figure 2.1). A large dataset of genes involved in plant immunity, such as WRKYs, MYBs, and RLKs, was compiled to search for any differences within these genes between Jack and Peking.

A short-read archive (SRA) dataset of Peking was used to identify variants between Jack and the reference genome, Williams82. When examining a homolog of Arabidopsis EF-Tu Receptor (AtEFR), annotated as *Glyma.09g216400*, short reads were found to only map to the 5’ UTR and 3’UTR region of the gene, indicating an absence of *Glyma.09g216400* (Figure 2.2). Because of the gene’s high similarity to the arabidopsis EFR gene, with a 66.3% similarity score based on homology to arabidopsis EFR, this gene was given a preliminary name, GmEFR. To confirm the deletion without relying on resequencing data and to avoid issues that arise with short reads, the Jack and Peking genotypes were *de novo* sequenced using 10x Chromium technology. The long-read assembly supported the prior SRA evidence, identifying a 4,421-bp deletion in Peking on Chr09 where GmEFR is positioned (Figure 2.3).

A multiplex PCR approach was designed to specifically detect the GmEFR deletion in the Jack and Peking genotypes. A single forward primer located in the 5'UTR region, combined with a reverse primer located within GmEFR, and another reverse primer located in the 3'UTR, successfully identified the presence or absence of GmEFR across the tested genotypes (Figure 2.4).

After confirming the multiplex PCR was able to distinguish between the alleles at the GmEFR position, additional soybean SRA datasets were analyzed for the Peking allele of GmEFR and the identified germplasm, along with previously characterized susceptible genotypes, were requested for subsequent screening. Many other genotypes have been documented regarding their response to *Agrobacterium*, but typically not often repeated in future studies. The multiplex PCR successfully identified genotypes harboring the GmEFR deletion found in Peking that were previously reported to be susceptible (Figure 2.5). To confirm if a correlation existed between absence of GmEFR and gall formation, the genotypes were inoculated with *Agrobacterium* to determine their susceptibility (Figure 2.6).

The following plant accessions were identified as having the GmEFR deletion, and susceptible to *Agrobacterium*: Peking (PI 438496A, PI 438496B, PI 438496C, PI 438497), Sable (PI 439498), Cloud (PI 548316), Sooty (PI 548415), Maple Arrow (PI 548593). Cloud was identified as having most of the gene intact, but Sanger sequencing revealed a small deletion that results in a premature termination (Figure 2.7). Another inbred line, called PxC, was bred at the University of Georgia from crossing Peking and Century and selected for susceptibility over many generations and also possessed the GmEFR deletion.

CRISPR/Cas9 of GmEFR

As all characterized susceptible genotypes possessed the Peking (deletion) allele, and all tested, resistant genotypes were confirmed to have the GmEFR gene present, we sought to create knockout mutants of GmEFR using CRISPR/Cas9 in the resistant Jack genotype. As Jack undergoes somatic embryogenesis and regeneration much better than other soybean genotypes, creating susceptibility to *Agrobacterium* would allow for greater transformation within the somatic embryogenesis system.

The first plasmid construct contained a single guide for EFR, EFRa (Table 2.1). This construct was biolistically delivered across ten attempts but had a poor efficiency, with most tissue dying during selection. This construct produced a total of four independent T_0 events, of which only two of the events were highly edited and resulted in complete knockouts (Figure 2.8). Events EFR_8_9 and EFR_10_4 had editing efficiencies of 47% and 75%, but knockout scores of 47% and 13%, respectively. The event EFR_8_9 presented as a monoallelic edit, whereas event EFR_10_4 displayed edits that did not result in frameshift mutations. Events EFR_8_2 and EFR_10_6 had editing efficiencies of 97-99% and knockout scores of 99-100%

To create more high-quality knockouts, a second construct was made containing two guides for EFR, EFRb and EFRc (Table 2.1). The new vector was introduced biolistically three times and produced twenty-five independent events with a wide range of editing efficiencies and knockout scores (Figure 2.9). The best high quality events were advanced to the T_2 generation and after screening for progeny that remained edited but had segregated out the cas9 backbone, were selected to be tested for susceptibility to *Agrobacterium*.

The selected GmEFR KO lines were tested along with Jack and Peking for susceptibility. However, the GmEFR-KO lines were not any more susceptible to *Agrobacterium* than wildtype Jack plants, indicating that the GmEFR gene alone is not response for conferring susceptibility to *Agrobacterium* (Figure 2.10). All Peking plants in the experiment were highly susceptible.

Engineering and testing of Agrobacterium EF-Tu mutants

While GmEFR did not condition the host response of soybean to *Agrobacterium*, we investigated whether EF-Tu is recognized by soybean through another receptor. To support this idea, the protein sequences of EF-Tu found in nodulating soybean symbionts, was compared to the EF-Tu sequence from *Agrobacterium*. The EF-Tu sequences to replace the *Agrobacterium* EF-Tu were selected from *Bradyrhizobium japonicum*, *Bradyrhizobium elkanii*, and *Agrobacterium rhizogenes.* The *Bradyrhizobium* species were selected as soybean varieties form symbiosis with these strains, with some soybean genotypes preferring *B. elkanii*, and others preferring *B. japonicum*. The EF-Tu from *Agrobacterium rhizogenes* was selected because unlike *Agrobacterium tumefaciens*, most soybeans are compatible with *A. rhizogenes* hairy root transformation.

The EF-Tu protein sequences among these strains of bacteria have several differences. While AtEFR from arabidopsis recognizes the first 18 amino acids of EF-Tu from *Agrobacterium*, sequence alignment identified a five amino acid insertion in the soybean symbionts at position 39 (Figure 2.11). This insertion is not found in the related soybean symbionts, the *Sinorhizobium*, but is present in *Pseudomonas syringae* pv. *glycinea*. Analysis of the insertion in Jmol presented the insertion as part of the external

structure of the protein, indicating it could be involved in determining host recognition of EF-Tu.

The *Agrobacterium* strain A281 was selected as the base strain for performing the swaps, or homologous recombinations, of the native EF-Tu sequences. The three engineered strains and A281 were first tested on *Nicotiana benthamiana* TW17 to confirm the strains were still functional. All four recombinant strains were able to transform and produce tumors on *N. benthamiana* (Figure 2.12).

The engineered strains and A281 were then tested on the soybean genotypes Jack and Peking. The strains were still able to form galls on Peking similar to the wild-type A281. The strains were no different than A281 when infecting Jack and no improvement of transformation was seen (Figure 2.13). Initially, the Peking genotype failed to show susceptibility with the engineered strain possessing *Bradyrhizobium japonicum* USDA110 EF-Tu, but a second experiment showed the strain was still functional. While GmEFR and the engineered strains indicate that *Glyma.09g216400*, is not an arabidopsis-like EFR for detecting *Agrobacterium* EF-Tu, these studies did uncover a few interesting observations.

During the course of these experiments and after ruling out the ability of the engineered strains to evade the host immune system, reactive oxygen species (ROS) assays were also conducted on a set of bioinformatically predicted peptides to try and identify other MAMPs that may be recognized by soybean. The peptide sequences were selected after filtering the core genome of *Agrobacterium* genes (Weisberg et al., 2020) and identified proteins that contained two to four sites with positive selection in a 25-bp window, similar to other methods that successfully predicted candidate MAMPs in

bacteria (Mott et al., 2016) (Table 2.2). To ensure an elicitor response could be measured when soybean was tested with these candidate peptides, three other flagellin-derived peptides (flg22) were included to serve as controls, *Pseudomonas syringae* pv. tomato flg22^{Psy}, *Ralstonia solanacearum* flg22^{Rso}, and *Pseudomonas aeruginosa* flg22^{Pae}, as soybean's response to these elicitors has been characterized (Wei et al., 2020). As this study was also focused on identifying MAMPs in *Agrobacterium*, the corresponding flg22 motif, flg22^{Atu} from *Agrobacterium*, was also tested.

The previously characterized peptides, flg22^{Psy} and flg22^{Rso} were recognized by the soybean genotype Jack, with the strongest response to flg22^{Psy}. Interestingly, the soybean genotype Peking did respond to flg22^{Psy}, but typically exhibited a much-diminished response when presented with flg22^{Rso} when compared to Jack (Figure 2.14). For both tested genotypes, the addition of flg22^{Atu} did not elicit a ROS response, similar to the DMSO control (Figure 2.14). While flg22^{Atu} has been shown to be recognized in *Vitis riparia*, the evidence presented here suggests that the two soybean genotypes tested do not recognize flg22^{Atu} (Fürst et al., 2020). Peking's lower ROS activity upon recognition of flg22^{Rso} remains undetermined.

After confirmation that ROS generation can be measured using known elicitors in soybean, the candidate *Agrobacterium* peptides were tested, along with either flg22^{Pae} or flg22^{Psy} (Figure 2.15). All twenty of the *Agrobacterium* predicted peptides failed to elicit an immune response in Jack or Peking. The response of the soybean genotypes to seven of the candidate peptides, including positive control peptides, flg22^{Pae}, or flg22^{Psy}, are presented in Figure 2.15.

Discussion

The findings of this research identify a soybean homolog of AtEFR that is absent in Peking and other susceptible soybean varieties. However, a resistant genotype engineered with CRISPR/Cas9 to knockout GmEFR displayed similar behavior to wild-type, resistant plants, suggesting that GmEFR alone may not solely control susceptibility to *Agrobacterium* in soybean. Additionally, the novel *Agrobacterium* strains expressing EF-Tu proteins from compatible soybean symbionts retained their ability to infect tobacco but did not exhibit an increased capacity to infect resistant soybean genotypes.

The ability of genotypes possessing the GmEFR deletion to be susceptible remains interesting. In all but one case, these genotypes are older varieties or landraces that were initially collected in China. These genotypes are distant from elite cultivars planted in the United States today, and they tend to prefer different rhizobium for nodulation than elite germplasm. While GmEFR alone was ruled out, there may be a haplotype in the region surrounding GmEFR that is shared in these genotypes, possibly explaining their ability to be susceptible.

This work advances our understanding of the molecular basis of soybean-*Agrobacterium* interactions, shedding light on the role of GmEFR and its lack of an impact on soybean susceptibility. These findings have implications for crop biotechnology, as they provide valuable insights into potential mechanisms regulating *Agrobacterium*-mediated transformation in soybean and identify methods for reliably characterizing soybean's response to *Agrobacterium*. The knowledge gained from this study may pave the way for targeted strategies to enhance transformation efficiency in soybean.

While the engineered strains created in this work did not affect outcomes with soybean, the strains may be of use for transforming members of the Brassicaceae, as many species are known to recognize elf18. These species, such as *Arabidopsis thaliana,* and important food crops like *Brassica napa*, *Brassica napus* or *Brassica oleracea,* may not recognize the symbiont EF-Tu sequence, resulting in greater transformation efficiency.

Materials and Methods

Plant Materials

The soybeans were planted in Fafard 3B potting mix (Sun Gro Horticulture Inc., Vancouver, Canada) in a 32-cell tray. The trays were placed in a growth chamber at 25º C and grown under a 23h photoperiod and a light intensity of 100 µmol m^{-2} s^{-1} for seven days with humidity domes (Hummert International, Earth City, MO). After nine days (VC growth stage), each genotype was thinned down to 3 to 4 plants and inoculated with *Agrobacterium* strain A281. After infection, plants were returned to the growth chamber maintaining the 23-h photoperiod at a temperature of 23º C and light intensity of 200 µmol m^{-2} s^{-1} for two days. The inoculated seedlings were then transplanted into 15.24 cm nursery pots containing a soil mixture of 3B, sterile farm soil, and sand. The plants were grown in the greenhouse with supplemental lighting to maintain 18/6-hour light/dark and keeping the temperature constant between 23.9º C – 26.7º C. After 60 days in the greenhouse, or around R1/R2 for Jack and Peking, the plants were phenotyped.

Marker design and genotyping

A multiplex PCR was used to screen for the Jack allele, the Peking (from PxC parent) allele, or both alleles across tested genotypes. Due to this deletion being present or absent in any individual, three primers were used. Two primers, TC1463_F and JHR452_R, detect the wild-type allele by binding to the 5' UTR and 452 bp downstream of the start codon, respectively. The third primer, TCEFR_R, binds downstream of the 3'UTR. In the susceptible genotypes, or having the deletion, this primer pairs with TC1463_F but the amplicon of 329 bp is significantly smaller than the resistant genotypes, or wild-type allele, of 4760 bp. The JHR452_R primer detects the wild-type allele while producing a smaller amplicon than the possible 4760 bp amplicon, allowing PCR to rapidly detect the genotype for *Glyma.09g216400*.

The PCR mix contains two premixed ready-to-use solutions, GoTaq Long and GoTaq Green (Promega Corporation, Wisconsin, United States). Instead of the protocol-referenced final concentration of 1X, the two master mixes were prepared to 0.5X final concentration, with the sum of the two being 1X. The forward primer TC1463_F was used at a final concentration of 15pmol, with the reverse primers JHR452_R at 8.75 pmol and TCEFR_R at 6.25 pmol. PCR conditions were followed per manufacturer's instructions unless noted. The primer annealing temperature was set to 60° C with an extension time of 30 seconds. For each sample, 1µL of a 10 ng μL^{-1} of DNA was added to 10µL of master mix.

Phenotyping

Plants were evaluated after sixty days in the greenhouse. The genotypes Jack and Peking were typically in the R1/R2 growth stage, with some variation between the genotypes based on maturity group. Images of each plant's entry or exit wounds, along with a ruler for calibration, were taken with a camera. The images were loaded into ImageJ, and the ruler used to calculate distance per pixel using the Measure function. Each gall was outlined by hand and size recorded (Schneider et al., 2012). Measurements of gall area were collected and analyzed using Microsoft Excel and RStudio (Racine, 2012).

Inoculation

The *Agrobacterium tumefaciens* strain A281 (C58 containing the armed pTiBo542) was obtained from Eugene Nester and was streaked from a glycerol stocks on yeast-mannitol medium containing 50 mg mL^{-1} rifampicin (Chem-Impex International, Wood Dale, Illinois) and grown for three days in a 28° C incubator to produce the inoculum used for infection (Jin et al., 1993; Vincent, 1970a). Sterile spatulas were used to spread *Agrobacterium* inoculum onto Giant Push Pins (Officemate, Edison, NJ). Each hypocotyl was punctured 2 cm above the soil completely through the stem. After puncturing, additional *Agrobacterium* inoculum was rubbed into each side of the puncture wound using a sterile spatula.

Experimental design for GmEFR knockout comparison

A randomized complete block design (RCBD) was used to compare the differences in susceptibility to *Agrobacterium*. In each of the three replicates, eight T_2 seeds from four independent events which had segregated out the cas9 and vector backbone but remained edited for the GmEFR were planted and inoculated. Eight Jack and Peking plants were also planted and inoculated. The T_2 plants were sequenced to confirm they possessed the GmEFR knockout. The border plants consisted of alternating Jack and Peking plants.

Bioinformatic prediction of candidate peptides

Assemblies of *Agrobacterium tumefaciens* biovar 1 were retrieved from NCBI BioProject PRJNA607555 (Weisberg et al., 2020), and compared, identifying 3475 core genes. Genes were clustered into ortholog groups. For each single-copy core gene, codon alignments and phylogenies were generated. For each gene, an overall dN/dS was calculated (HyPhy SLAC) and individual sites under positive selection were identified (HyPhy FUBAR) (Murrell et al., 2013; Pond and Frost, 2005). Sliding windows of 25 or 90 sites were analyzed for regions with multiple sites under positive selection. After visual inspection of the predicted regions, 4 mg of each 25-mer peptide was ordered with a minimum of 95% purity and standard TFA removal. Peptides were dissolved according to individual solubility tests performed on each peptide (GenScript, Piscataway, NJ).

Measurement of ROS generation

The ROS bursts were measured using a luminol and peroxidase-based assay as previously reported (Sang and Macho, 2017). Seeds of the soybean genotypes Jack (PI 540556) and Peking (548402) were planted in 15.24 cm nursery pots containing a mixture of Fafard 3B potting mix (Sun Gro Horticulture Inc., Vancouver, Canada), sterile farm soil, and sand. The plants were grown in the greenhouse with supplemental lighting to maintain a long day cycle (18/6-hour light/dark), with temperature kept constant between 23º C - 27º C. When the plants were about 4- to 5-week-old, approximately V2 growth stage, the ROS assays were conducted using leaf discs. All peptides were tested at a concentration of 100µM and dissolved in DMSO, along with a DMSO mock control (100mM), using twenty-four leaf discs from each genotype. The ROS burst in response to the peptides was reported as ROS response over time and as total accumulated relative luminescence units (RLU).

Vector construction

The CRISPR/cas9 vectors were made using NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, MA, US). Construction was based on p201H as described with minor modifications (Jacobs et al., 2015). The target sites in the resistant soybean Jack were amplified and sequenced to ensure proper target selection. The GmEFR targets or gRNAs were ordered as single-stranded oligomers (primers) from Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO). The selected targets for the vectors are listed in Table 1.

Soybean embryogenesis and transformation

The transgenic soybean lines were created from Jack somatic embryos as previously described with minor modifications (Trick, 1997). Somatic embryos were induced from immature cotyledons in a basal Murashige and Skoog medium supplemented with 40 mg L^{-1} 2,4-D for six weeks, and four additional weeks for maintenance with 20 mg L^{-1} 2,4-D. The embryos were transformed using microprojectile bombardment. Each shot used 10 ng of plasmid DNA with 0.6- μm gold particles and shot at a pressure of 1100 psi. After one week of recovery, the embryos were transferred to FNL media supplemented with 20 mg L^{-1} hygromycin (Samoylov et al., 1998). After 8 weeks, embryos were selected and transferred into soybean histodifferentiation and maturation (SHaM) medium (Schmidt et al., 2005). After five weeks in SHaM medium, the embryos were desiccated for a week before undergoing germination.

CRISPR/Cas9 analysis

A young trifoliolate leaf or clump of somatic embryos from putative events was collected from each event. The samples were ground in a Geno/Grinder (SPEX SamplePrep LLC, Metuchen, NJ) before undergoing DNA extraction using the Promega Wizard DNA extraction kit (Promega Corporation, Madison, WI). A master mix of Apex Taq RED polymerase (Apex Bioresearch Products, Houston, TX) and two primers, TC1463_F and JHR452_R, were used to amplify from the 5' UTR to 452 bp after the start codon, capturing the region of *Glyma.09g216400* targeted by the CRISPR/Cas9 vectors. After confirmation of amplification with gel electrophoresis, column purified PCR products were sent for Sanger sequencing (Azenta Life Sciences, Chelmsford, MA).

Sequencing results were compared to unedited control samples using the ICE analysis tool to calculate editing efficiency, knockout-score, and type of indel created (Conant et al., 2022).

Construction of Agrobacterium tumefaciens A281 mutants

Agrobacterium tumefaciens strain A281 was obtained from Eugene Nester (Jin et al., 1987). The strain was streaked from freeze stocks on medium known as Luria-Bertani (LB) containing 50 mg mL^{-1} rifampicin (Bertani, 1952; Chem-Impex International, Wood Dale, Illinois, United States) and grown for two days in a 28 °C incubator (Vincent, 1970b). The *tufA* sequence of *Agrobacterium* A281 was replaced with tuf sequences from *Bradyrhizobium japonicum USDA110*, *Bradyrhizobium elkanii*, or *Agrobacterium rhizogenes* K599. DNA fragments to replace genes were synthesized or amplified from bacterial stocks and sequenced. Compatible *Agrobacterium* flanks (~ 1000 bp) for recombination were amplified and fused using NEBuilder (New England Biolabs, MA). Complete donor constructs were transformed into DH10 α for sequencing prior to being transformed into *E. coli* suitable for conjugation. Competent *E. coli* RHO5, requiring diaminopimelic acid to survive, were provided by Brian Kvitko at the University of Georgia (Kvitko et al., 2012). Donor *E. coli* and *Agrobacterium* were placed on Immobilon-NY+ filters and after one day, moved to DAP-free LB medium containing 10% sucrose before being plated on LB medium without (for replacement of *tufA*) or with the antibiotic spectinomycin (for replacement of *tufB*) to recover double-crossover mutants with the spectinomycin resistance gene replacing the original sequence of interest followed by PCR and sequencing confirmation (Kvitko and Collmer, 2011).

References

1. Conant, D., Hsiau, T., Rossi, N., Oki, J., Maures, T., Waite, K., . . . Holden, K. (2022). Inference of CRISPR edits from Sanger trace data. *The CRISPR Journal, 5*(1), 123-130.
2. Fürst, U., Zeng, Y., Albert, M., Witte, A. K., Fliegmann, J., & Felix, G. (2020). Perception of Agrobacterium tumefaciens flagellin by FLS2XL confers resistance to crown gall disease. *Nature plants, 6*(1), 22-27.
3. Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., & Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC biotechnology, 15*(1), 1.
4. Jin, S., Komari, T., Gordon, M., & Nester, E. J. J. o. B. (1987). Genes responsible for the supervirulence phenotype of Agrobacterium tumefaciens A281. *169*(10), 4417-4425.
5. Jin, S., Song, Y. n., Pan, S. Q., & Nester, E. W. (1993). Characterization of a virG mutation that confers constitutive virulence gene expression in Agrobacterium. *Molecular Microbiology, 7*(4), 555-562.
6. Kvitko, B. H., Bruckbauer, S., Prucha, J., McMillan, I., Breland, E. J., Lehman, S., . . . Schweizer, H. P. (2012). A simple method for construction of pir+ Enterobacterial hosts for maintenance of R6K replicon plasmids. *BMC research notes, 5*, 1-7.

7. Kvitko, B. H., & Collmer, A. (2011). Construction of *Pseudomonas syringae* pv. tomato DC3000 mutant and polymutant strains. *Plant Immunity: Methods and Protocols*, 109-128.

8. Mott, G. A., Thakur, S., Smakowska, E., Wang, P. W., Belkhadir, Y., Desveaux, D., & Guttman, D. S. (2016). Genomic screens identify a new phytobacterial microbe-associated molecular pattern and the cognate Arabidopsis receptor-like kinase that mediates its immune elicitation. *Genome biology, 17*, 1-15.

9. Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular biology and evolution, 30*(5), 1196-1205.

10. Pond, S. L., & Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution, 22*(5), 1208-1222.

11. Racine, J. S. (2012). RStudio: a platform-independent IDE for R and Sweave. In: JSTOR.

12. Samoylov, V., Tucker, D., Thibaud-Nissen, F., & Parrott, W. (1998). A liquid-medium-based protocol for rapid regeneration from embryogenic soybean cultures. *Plant Cell Reports, 18*, 49-54.

13. Sang, Y., & Macho, A. P. (2017). Analysis of PAMP-triggered ROS burst in plant immunity. *Plant pattern recognition receptors: methods and protocols*, 143-153.

14. Schmidt, M., Tucker, D., Cahoon, E., & Parrott, W. (2005). Towards normalization of soybean somatic embryo maturation. *Plant Cell Reports, 24*, 383-391.

15. Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods, 9*(7), 671-675.

16. Trick, H. N. J. P. T. C. B. (1997). Recent advances in soybean transformation. *3*, 9-26.

17. Vincent, J. M. (1970a). A manual for the practical study of the root-nodule bacteria. *A manual for the practical study of the root-nodule bacteria.*

18. Vincent, J. M. J. A. m. f. t. p. s. o. t. r.-n. b. (1970b). A manual for the practical study of the root-nodule bacteria.

19. Wei, Y., Balaceanu, A., Rufian, J. S., Segonzac, C., Zhao, A., Morcillo, R. J., & Macho, A. P. (2020). An immune receptor complex evolved in soybean to perceive a polymorphic bacterial flagellin. *Nature communications, 11*(1), 3763.

20. Weisberg, A. J., Davis, E. W., Tabima, J., Belcher, M. S., Miller, M., Kuo, C.-H., . . . Chang, J. H. (2020). Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science, 368*(6495), eaba5256.

21. Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D., Boller, T., & Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell, 125*(4), 749-760.

Figure 2.1. Typical reactions of the resistant Jack genotype and susceptible Peking genotype to infection with *Agrobacterium tumefaciens* A281.

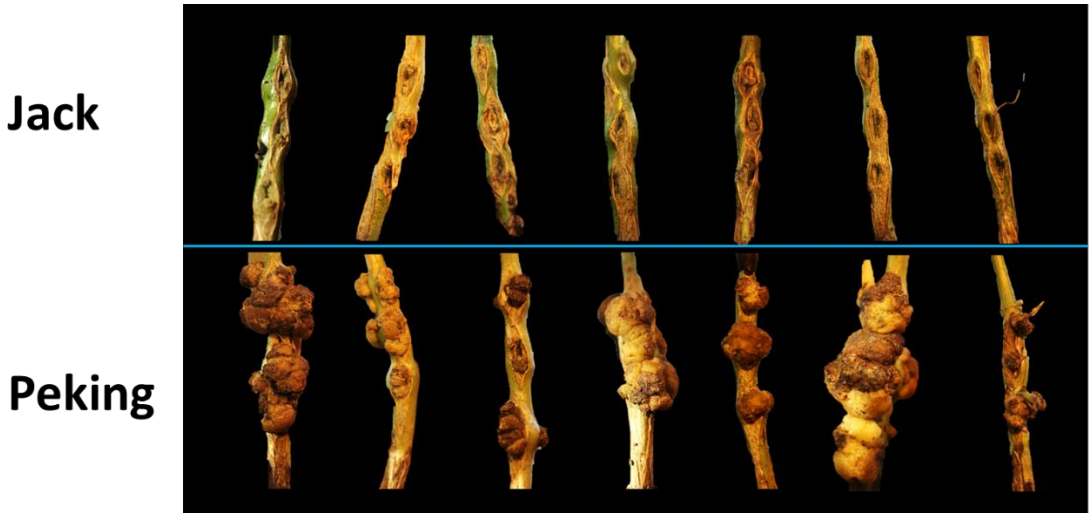


Figure 2.2. Short reads aligned to a soybean homolog (GmEFR, *Glyma.09g216400*) of the Arabidopsis Elongation Factor Receptor. The black rectangles represent reads aligning to the model. The absence of the reads for susceptible or genotypes is evident by the lack of reads aligning.

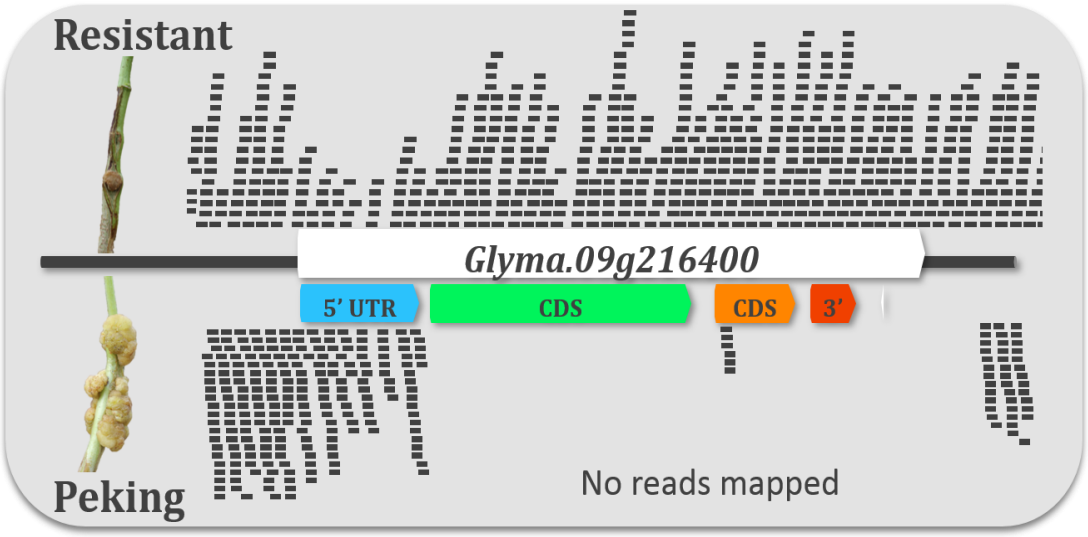


Figure 2.3. Alignment from a 10X assembly of *Glyma.09g216400* on Chr09 in the susceptible Peking variety, top, to the corresponding region in the reference genome Williams 82. The deletion (orange rectangle) was confirmed to be 4,421 bp in the gene model (blue rectangle). Circles represent SNPs or indels between the two genotypes.

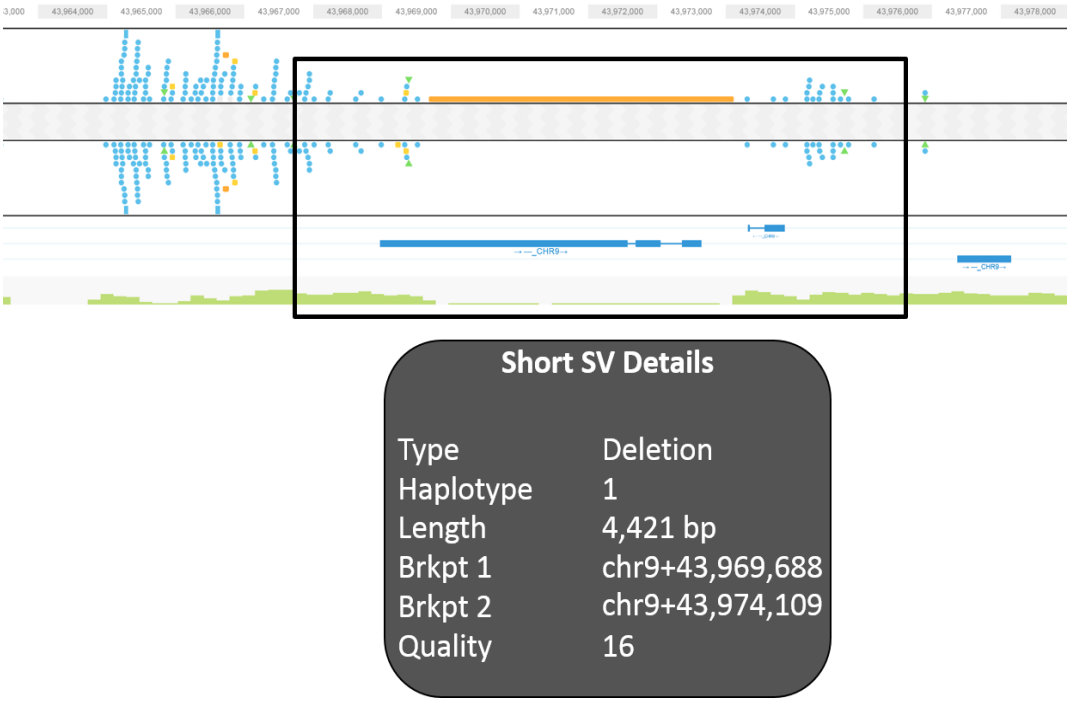


Figure 2.4. A PCR screen for presence or absence of the soybean homolog of arabidopsis Elongation Factor Receptor (GmEFR). In the left image, a primer pair amplifying from the promoter to the CDS will indicate if the gene is present. In the right image, a primer pair amplifying from the promoter to the 3' UTR will produce a small amplicon if the gene is deleted.

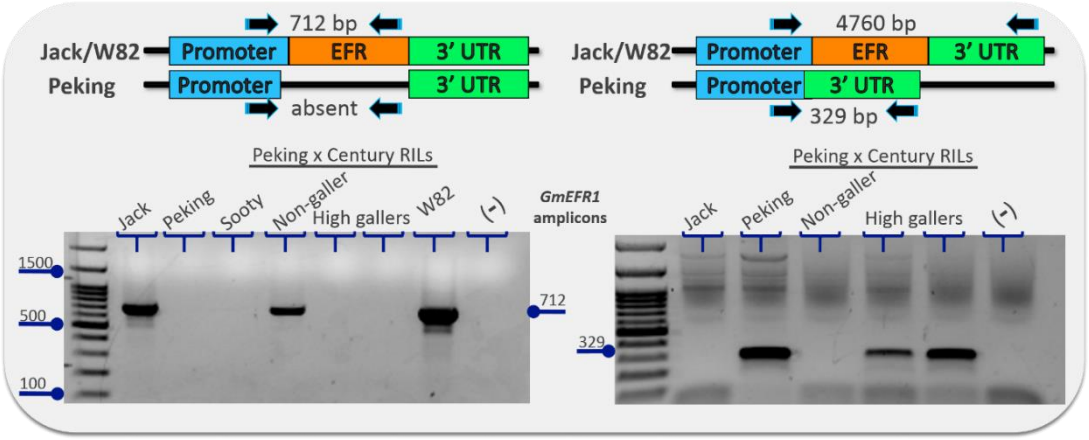


Figure 2.5. Multiplex PCR for determining presence or absence of the soybean EFR gene across diverse genotypes (*Glyma.09g216400*).

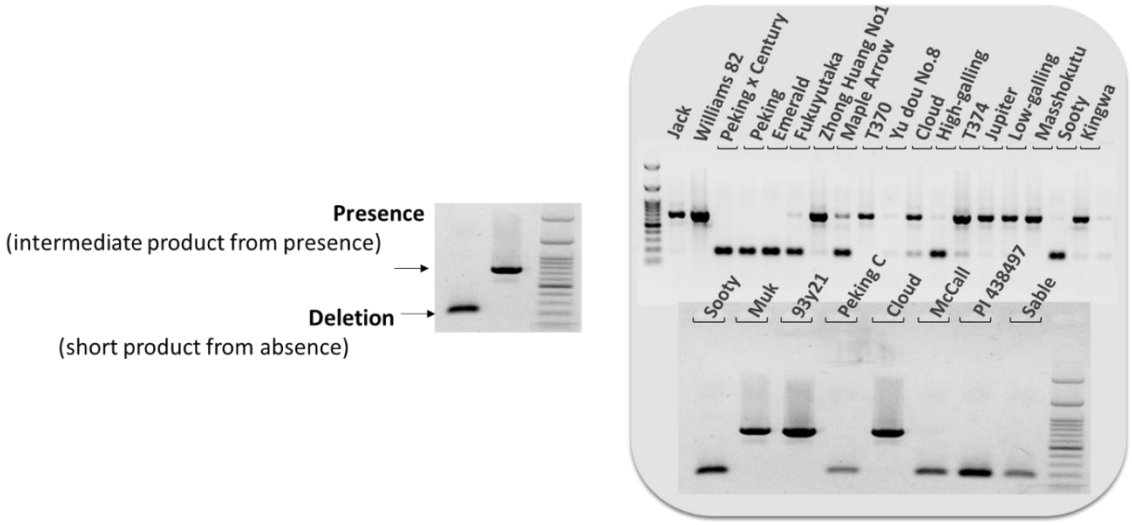


Figure 2.6. Genotypes lacking GmEFR tend to be susceptible to *Agrobacterium*. Corresponding names to the accessions are Sable (PI 438498), Peking (PI 438497), Peking (PI 438496A), Peking (PI 438496B), Peking (PI 438496C), Maple Arrow (PI 548593) and Cloud (PI 548316). The accession Cloud is marked with an asterisk as the deletion is different from the other lines.

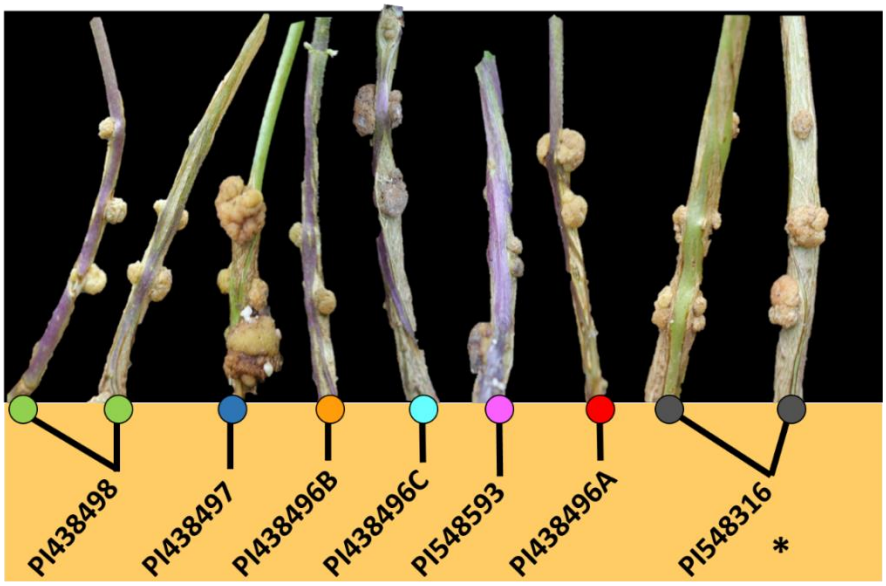


Figure 2.7. Sequence analysis of the Cloud genotype with a small deletion within GmEFR, resulting in premature termination.

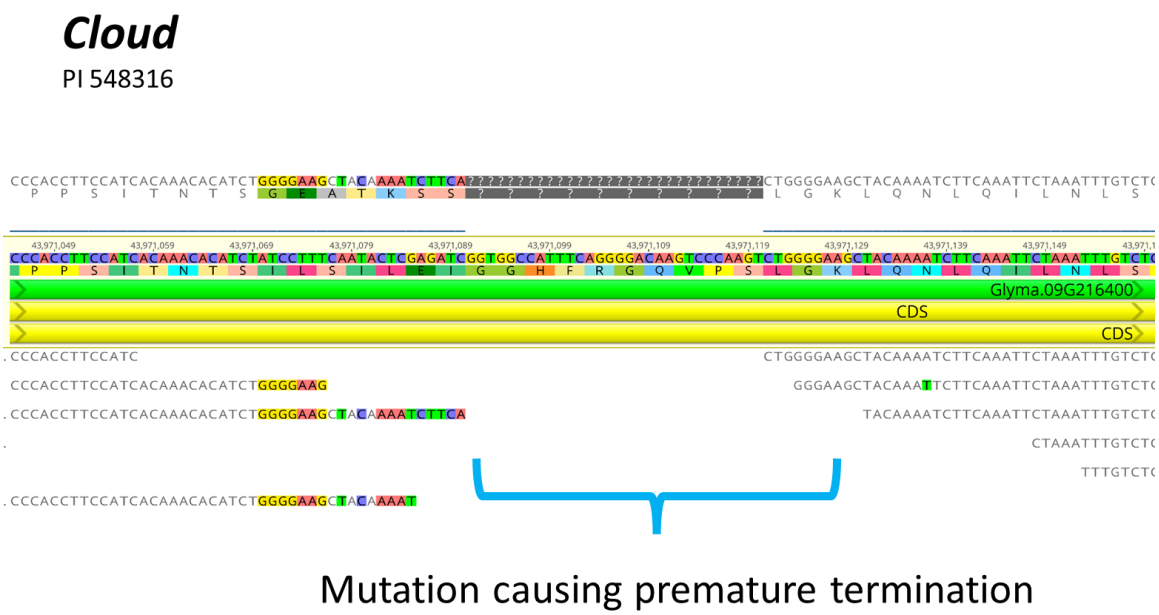


Table 2.1. Target sequences for CRISPR/cas9 editing of GmEFR. The PAM is highlighted in red.

Guide	Target Gene	Sequence (5' - 3')	Location (Chr09)
EFRa	Glyma.09g216400	GAGATATGAATCCCTTCAACTGG	43,970,470-43,970,451
EFRb	Glyma.09g216400	GTCGTCTTAAAAGCCTGAAATGG	43,970,840-43,970,859
EFRc	Glyma.09g216400	GAATTGTTACCTAGATTGTTCGG	43,971,182-43,971,163

Figure 2.8. Example edits in GmEFR using CRISPR/Cas9 in four independent events. Event EFR_8_9 had a -2 bp indel in 47% of the amplicons, with 53% of the amplicons being wildtype. Event EFR_10_4 presented with only 59% of amplicons being edited, but with a -3 bp indel. Events EFR_8_2 and EFR_10_6 presented with at least 97% editing efficiency, with the edits being -11 bp, - 8 bp, or -7 bp, ensuring the reading frame would be disrupted in the first exon, making these knockout events.

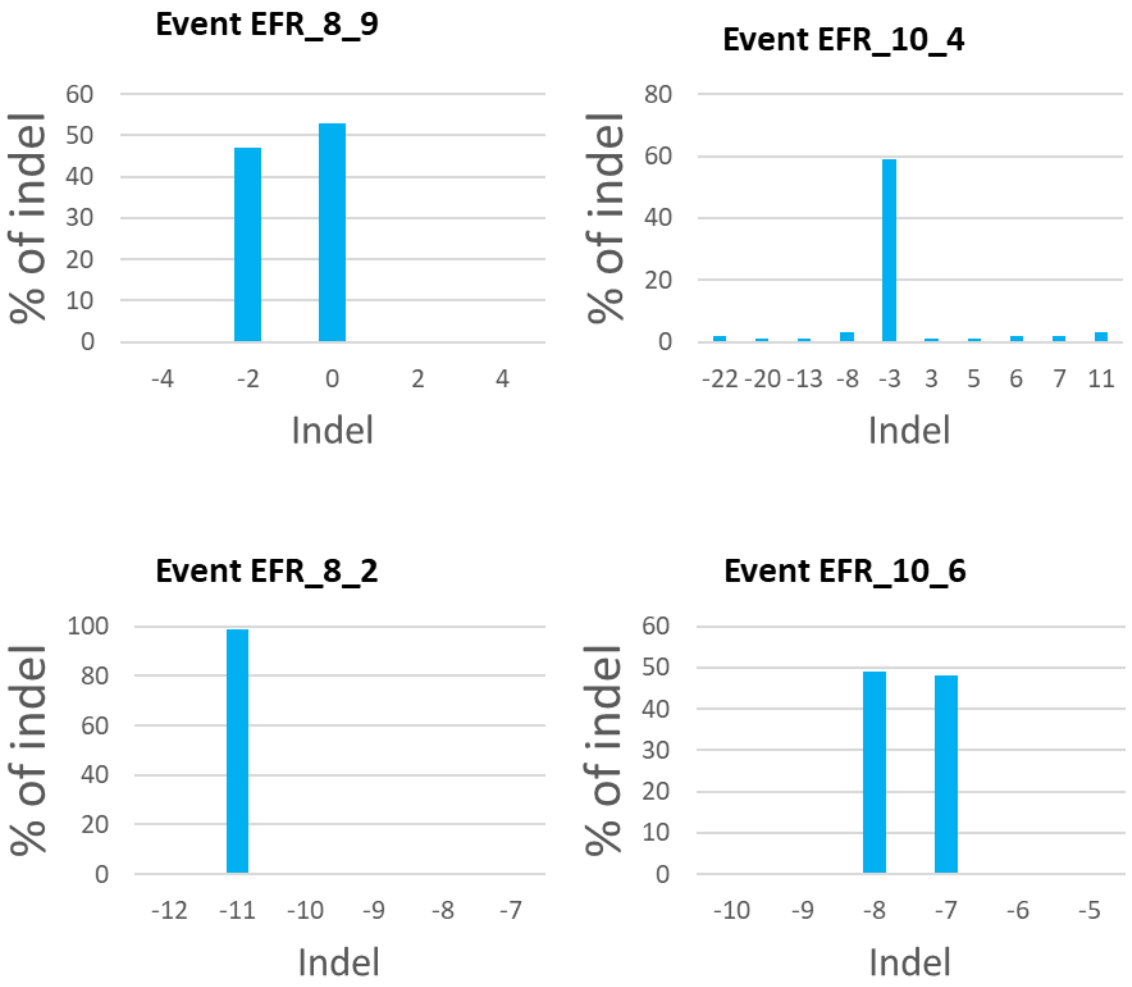


Figure 2.9. CRISPR/Cas9 of selected events exhibiting high knockout efficiency. Leaf samples were collected for DNA extraction and PCR of the target region. Amplicons were sequenced and compared to unedited samples using ICE Analysis (Synthego, Redwood City, CA).

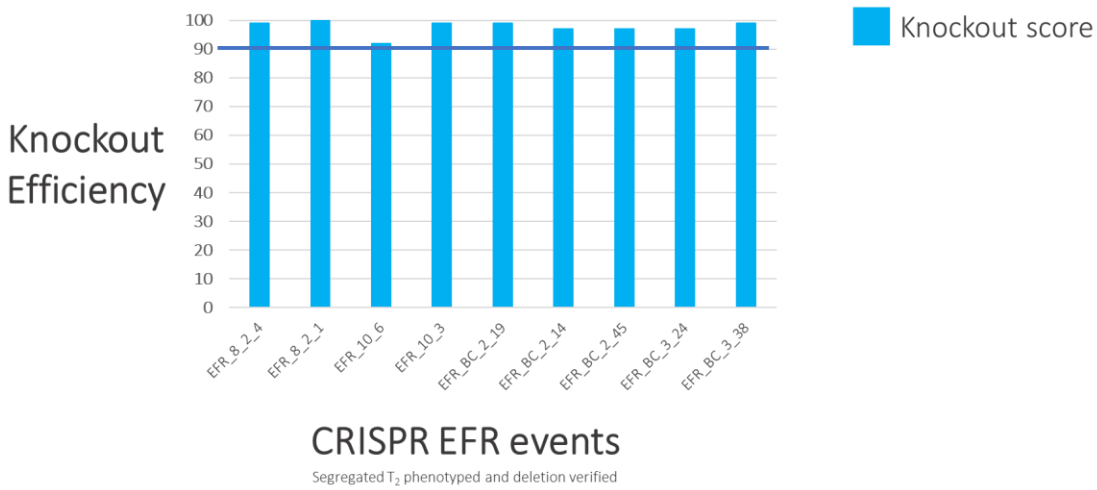


Figure 2.10. Inoculation of four different GmEFR CRISPR mutant lines with high knockout scores and check lines with *Agrobacterium tumefaciens* A281. Eight T_2 of each event were checked for knockout score and tested.

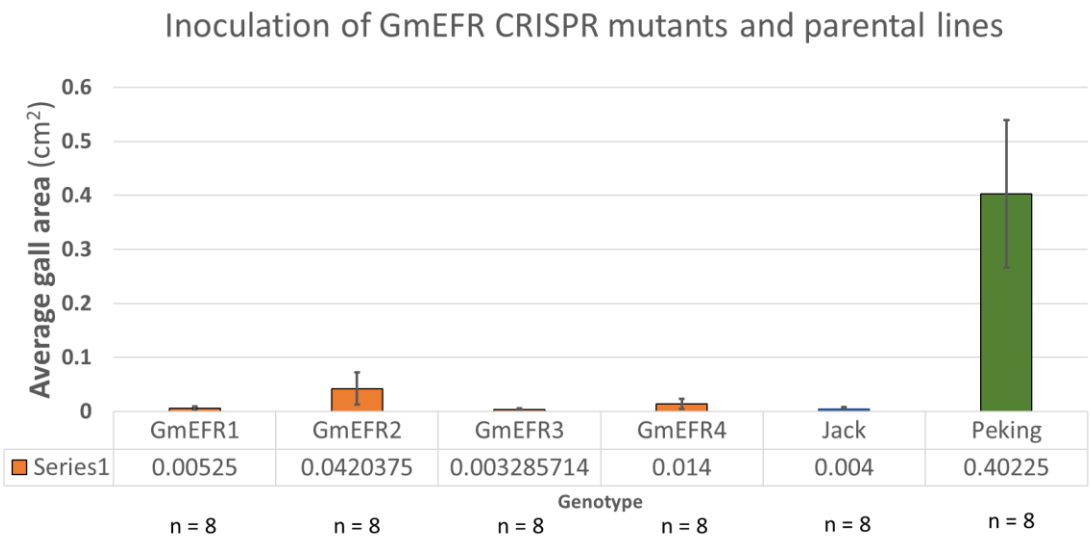


Figure 2.11. Alignment of various EF-Tu sequences from *Agrobacterium* and related rhizobia.

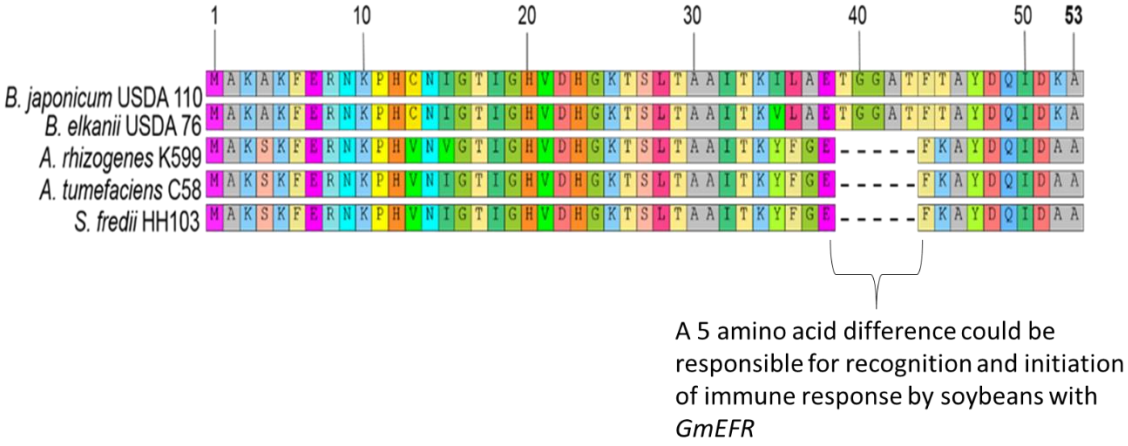


Figure 2.12. The engineered EF-Tu strains maintained the ability to infect tobacco. The labels under each image represent which bacterial EF-Tu sequence replaced the native *Agrobacterium* EF-Tu.

Figure 2.13. The engineered strains did not gain the ability to transform the resistant genotype Jack. The labels under each image represent which bacterial EF-Tu sequence replaced the native *Agrobacterium* EF-Tu.

A281 wild-type | *B. elkanii* EF-Tu | *B. japonicum* EF-Tu | *A. rhizogenes* EF-Tu

Table 2.2. Bioinformatically selected *Agrobacterium* peptide sequences tested for recognition in soybean. The twenty peptide sequences were selected after filtering the core genome of *Agrobacterium* genes and identifying peptides that possessed sites undergoing positive selection in a 25-bp window. The *Pseudomonas syringae* pv. tomato (Psy) flg22 peptide, *Ralstonia solanacearum* (Rso) flg22 peptide, and *Pseudomonas aeruginosa* (Pae) peptide served as controls for soybean recognition of flagellin as measured through reactive oxygen species (ROS).

***Agrobacterium* Gene**	**Selected Peptide Sequence**
Atu0479_09	VGNSRHEQMLKAFFPKARVEGFDGY
Atu0579	VVFSNGPITTKSDETTADSNNSAGS
Atu0579_09	AVNSAAANPWANAGASSSDKNAASL
Atu0686_09	MIWVKFPVKDAAPALSGVSVVIWTT
Atu0686_09_2	AKLTFKFVRDVKSEELAAITCAHPL
Atu0816	MSAAAANTPSSSATILPFAEHSKVA
Atu0891	GLRLVSPVQIDFNQAILVAVIGLVV
Atu0920_09	ELLDKQLASGDKPADVADIAADAEK
Atu1159	GYKTRYHPQEHLTPRGWEIYSPKEE
Atu1364	KAAKTADEAEILVAEADKSDEDKPRR
Atu2324	DREGPGNNLEEYFFDEPFMIAEPSL
Atu2330	HTPVIPSGFRLQEHVSIAVPDFRKF
Atu2330_09	KPHLRGIAERHERDAAENWRSGERR
Atu3041_09	MKMPGLSSRLAMLALGTAMALPLVP
Atu3514	NGATAIEYGLIAGIISAALIAGLGN
Atu3514_09	IAGLGNISSGINAVFQFIVDAFPKG
Atu3725	LGEFVLASENETVRYHEKTQFRRAG
Atu3725_09	KGPAFGETVTSGVWTESGLGAAFAR
Atu4020_09_2	LWNTISILVCLAIAGIISFANSKTL
Atu4478	MCQERDYVNRTERFSSKTFFNLSRY
Psy flg22	TRLSSGLKINSAKDDAAGLQIA
Rso flg22	QRLSTGLRVNSAQDDSAAYAAS
Pae flg22	QRLSTGSRINSAKDDAAGLQIA
Atu flg22	SRVSSGLRVKSASDNAAYWSIA

Figure 2.14. Reactive oxygen species (ROS) response by soybean genotypes to flg22 peptides. The average relative luminescence units (RLU) from 1 to 64 minutes or total accumulated RLU are shown. The experiment used 24 leaf discs per genotype and peptide combination. Error bars represent the standard error of the mean. The soybean Jack is presented in the top panel, Peking in the bottom panel.

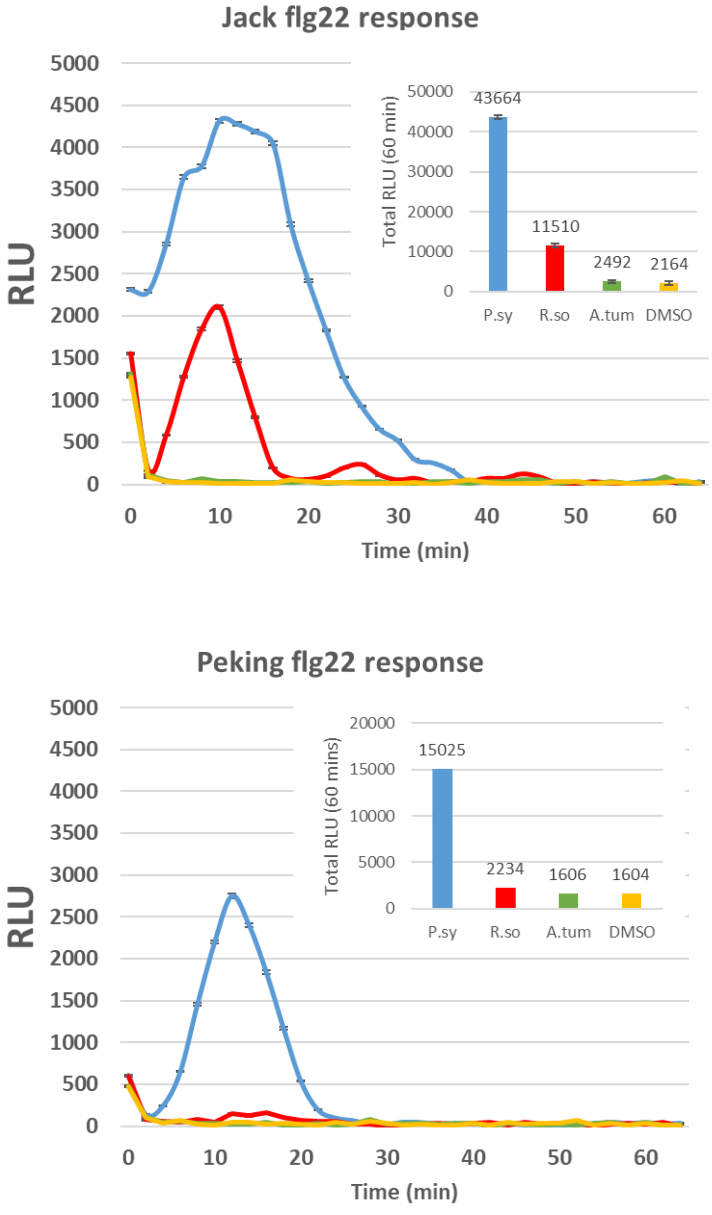
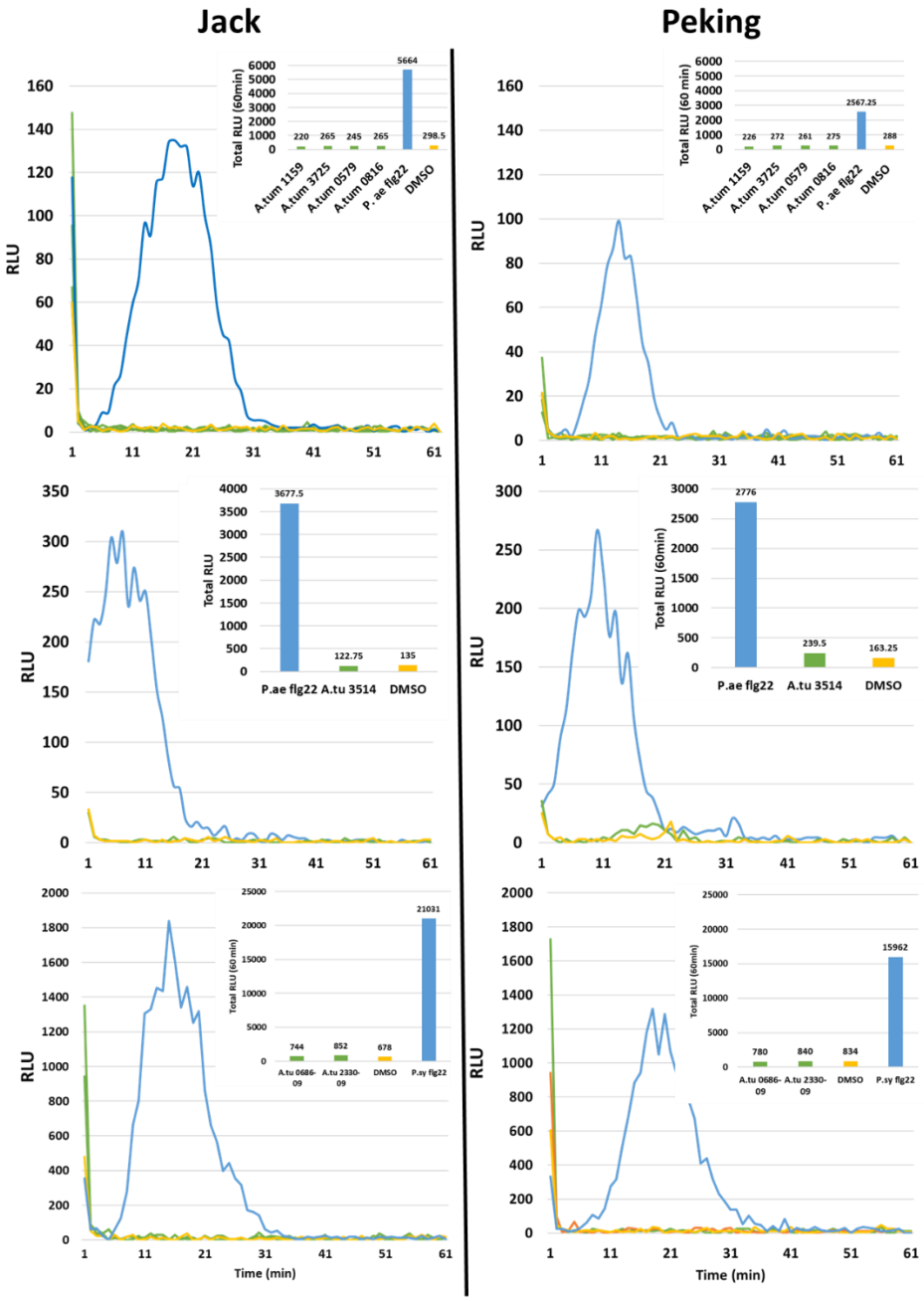


Figure 2.15. Reactive oxygen species (ROS) response by soybean genotypes to predicted candidate *Agrobacterium* peptides. The average relative luminescence units (RLU) from 1 to 64 minutes or total accumulated RLU are shown. The experiment used 24 leaf discs per genotype and peptide combination. The soybean Jack is presented in the left panel, Peking in the right panel.



CHAPTER 3

UNRAVELING THE GENETIC BASIS OF SOYBEAN-AGROBACTERIUM INTERACTIONS: QTL ANALYSIS AND RNA-SEQ PROFILING[2]

[2]Timothy M. Chappell, Hallie Wright, Liza Zaytseva, Qijan Song, Zenglu Li, and Wayne A. Parrott. To be submitted to *Frontiers in Plant Science*.

Abstract

Agrobacterium-mediated transformation is widely used for genetic engineering in soybean [*Glycine max* (L.) Merr.], but the underlying host-microbe interactions remain poorly understood. We developed a mapping population from the resistant parent Jack (PI 540556) and an inbred F_5 line, PxC, derived from Peking crossed with Century and selected for susceptibility. The F_5 lines exhibited varying susceptibility to *Agrobacterium*, with some lines outperforming the susceptible Peking. A major QTL on Chr06 was identified for gall size and number, explaining 18.88% and 27.93% of their variation, respectively. This QTL region harbors numerous host defense genes and displays significant genetic variation between Jack and Peking genomes. To refine the selection of candidates genes from the regions identified in the QTL, we utilized an RNA-Seq approach. We compared the transcriptomic differences in Jack and Peking at four timepoints post-inoculation. Gene expression in the two genotypes showed substantial variation in their hypocotyls across timepoints, independent of *Agrobacterium* inoculation. Time and genotype were the main factors influencing gene expression, with infection having a minor effect. The previously identified QTL for *Agrobacterium* susceptibility contained multiple differentially expressed candidate genes, including peroxidases, transcription factors, leucine-rich repeat protein kinases, and isoflavone pathway genes. These findings shed light on the expression differences in response to *Agrobacterium* infection between the two genotypes, providing valuable insights into the potential pathways that modulate soybean-*Agrobacterium* interactions. The identified region will aid fine-mapping, candidate gene testing, and breeding Peking-type susceptibility into elite cultivars.

Introduction

Genetically engineered soybeans are essential for modern agriculture, achieved through biolistic bombardment or *Agrobacterium*-mediated transformation (AMT). However, the recalcitrance of most soybean genotypes to tissue culture and regeneration hinders biolistic transformation, and the resistance of most genotypes to *Agrobacterium* creates bottlenecks in soybean breeding and the adoption of new technologies like CRISPR/Cas9. Understanding *Agrobacterium*-resistance in soybean is crucial for expanding genetic engineering tools to diverse genotypes, facilitating direct trait delivery into elite cultivars, and speeding up trait introgression.

The first transgenic soybean was created through *Agrobacterium*-mediated transformation (AMT) of the soybean genotype Peking (Hinchee et al., 1988). Susceptibility of Peking to *Agrobacterium* has been recorded extensively, yet no other genotype has been reported across studies to be susceptible (Byrne et al., 1987; Owens and Cress, 1985). Later studies attempted to understand the inheritance of susceptibility, revealing a complex nature, where different combinations of parents suggested some genotypes harbor more genes than other genotypes that are involved with controlling susceptibility (Bailey et al., 1994; Mauro et al., 1995).

More recent studies have focused on mitigating the host defense response through varied methods or molecules. A technology called sonication-assisted *Agrobacterium*-mediated transformation (SAAT) was described in 1997 as a method of creating wounds in embryos (Trick and Finer, 1998). These wounds were hypothesized to allow deeper penetration of *Agrobacterium* into the tissues and possibly into meristematic tissues, allowing the engineered cell to survive selection and undergo regeneration into a whole

plant. While this new technology boasted dramatic fold increases in transient expression and was necessary to obtain stable transformants of soybean, transformation efficiencies remain low.

Wounding methods, including SAAT, have not solved the limitations with the stable transformation of soybean. When soybean and *Agrobacterium* are co-cultivated, host tissue can quickly become necrotic or undergo a hypersensitive response (Olhoft et al., 2001; Van Breusegem and Dat, 2006). In recent years, medium modifications that alleviate the plant response have shown moderate utility. Examples include antioxidants that arrest the buildup of reactive oxygen species (ROS) which unchecked, trigger the programmed cell death (PCD) of the host cell (Dan, 2008; Dan et al., 2009). While the recognition mechanism of *Agrobacterium* by soybean remains unknown, parallels between *Agrobacterium*, and related members of the Rhizobiaceae, *Bradyrhizobium* and *Sinorhizobium*, suggest that *Agrobacterium* and soybean interactions may act through analogous pathways.

As current methods seek to alleviate the response of soybean to *Agrobacterium*, we aimed to genetically map and identify quantitative trait loci underlying Peking's extreme susceptibility to *Agrobacterium*, and to further identify differentially expressed genes across genotypes using RNA-Seq. By using a F_5 mapping population (Figure 3.1) from the resistant parent Jack (PI 540556) and an inbred F_5 line, PxC, derived from Peking, we seek to identify QTL, that when combined with RNA-Seq data across genotypes and timepoints, can pinpoint candidate genes within a narrow genetic region that are involved with susceptibility to *Agrobacterium*. Furthermore, by inoculating both

genotypes at multiple timepoints, typical of the different immune responses, we sought to understand soybean's innate immune reactions to the bacterium.

Results

Susceptibility within population

The ability of the recombinant inbred lines to be susceptible to *Agrobacterium* was evident within this Jack x PxC population (Figure 3.2, Figure 3.3). The inbred population produced an overall mean across the replicates that was intermediate of the parental lines but were not normally distributed (Figure 3.4). ANOVA analysis of the group means across replicates had a *p value* of 0.00178, indicating a significant difference among the genotypic groups. To determine which genotypic groups significantly differed, a Tukey multiple comparisons of means was employed. Interestingly, the Peking and JxPC groups were not significantly different, but the groups JxPC/Jack ($p = 0.009$) and Peking/Jack ($p = .0016640$) were. In general, the population behaved similarly to Peking, but was significantly more susceptible than Jack. This finding implicates that Peking-type susceptibility to *Agrobacterium* can be bred into other soybean varieties through selection, as evident by the population not being significantly different than the susceptible genotype Peking and having a much improved frequency distribution in terms of gall area when compared across the parents (Figure 3.5).

While the normality issue of the data did not fit the assumptions for performing an ANOVA within all individual recombinant inbred lines and parentals, an ANOVA was still performed (Table 3.1). According to the ANOVA and the Scott-Knott test, between twenty to thirty RILs were significantly more susceptible to *Agrobacterium* than the other

RILs (Table 3.2). However, many of the JxPC RILs behaved similar to Jack, similar to Peking, or better than Peking (Fig. 3.6).

As these data were not amenable to all tested transformations and remained non-parametric, a Kruskal-Wallis test was performed, which supported significant differences within the population ($p = 3.872$e-05). To cluster the RILs into groups based on similar means, the Scott-Knott test was used with 5% level of probability (Table 3.3). Twenty-nine RILs formed a group of the lines most susceptible to *Agrobacterium*. A plot of every line and respective average gall size in comparison to Peking shows many lines with a greater average gall size than Peking but display much more variability across the replications (Figure 3.6). While variability was higher with the recombinant inbred lines than the inbred parents, a few lines warrant being used in future studies. The lines 143, 139, and 142 had the lowest variation among the highest galling lines. Lines 130 and 118 have great galling potential but may still be segregating for factors that are involved in susceptibility. Nonetheless, these lines may be useful for breeding hypersusceptibility into other populations or to perform comparative analyses with lines that performed poorly or similar to Jack.

QTL region

After construction of the genetic map (Figure 3.7, Figure 3.8), composite interval mapping (CIM) was performed. A significant QTL was identified on Chr06 for gall size and gall number that accounted for 18.88% and 27.93% of the phenotypic variation in these traits, respectively. These results imply that these traits have a complex nature due to the polygenic effects and their environmental conditions. Many lines galled similar to

or better than Peking, and as the deletion of GmEFR was fixed within this mapping population, we are unable to interpret the effect this deletion had on the phenotypes overall, but variation for susceptibility to *Agrobacterium* still existed within this population.

The major QTL on chromosome 6 for gall size peaked at 32.7 cM with a LOD score of 7.3 (Figure 3.9, Figure 3.10, Figure 3.12, Table 3.4). Two markers represented this position, Gm06_11659627 and Gm06_11786580, spanning 126 kb between them and covering 17 gene models. A group of four markers 3 cM upstream had a LOD score of 6.702. These linked markers covered a physical distance of 836 kb. Similarly, another group of four markers were located 3.4 cM downstream, covering a physical distance of 342 kb. This QTL in total spans 6.4 cM based on these markers.

The major QTL on chromosome 6 for gall number was mapped between Gm_10823424 and Gm06_11786580 with a LOD score peak of 6.6 over 3 cM (Figure 3.9, Figure 3.11, Figure 3.13, Table 3.4). These markers span a physical distance of 963 kb.

While only a single, nearly identical QTL was identified for both gall size and gall number, each phenotype also identified three minor QTL (LOD score > 2.0). The three chromosomes containing the minor QTL for gall size were Chr02, Chr09, and Chr10 (Fig. 3.9). The most significant physical marker on Chr02 was Gm02_41400852_G_A with a LOD score of 1.33. For Chr09, the physical marker identified was Gm09_33894091_A_C with a LOD score of 2.77. On Chr10, the physical marker Gm10_3465857_C_T had a LOD score of 2.23.

The minor QTL for gall number were different than the minor QTL identified for gall size (Figure 3.9). The gall number minor QTL were found on Chr07, Chr11, and Chr20. The most significant physical marker on Chr07 was Gm07_5097877_A_C with a LOD score of 2.27. For Chr11, the physical marker identified was Gm11_1704799_C_T with a LOD score of 2.89. On Chr20, the physical marker Gm20_39698421_G_A had a LOD score of 2.21. These results suggest that other genetic regions may impact either gall size and gall number when infected with *Agrobacterium tumefaciens,* but these QTL are relatively minor and may have no effect.

Heritability was calculated using $H^2 = V_g/V_p$, where H is the heritability estimate, V_g was the variation in genotype, and V_p was the variation in phenotype. The heritability (H^2) of gall size and gall number were 31.2% and 25%, respectively. While the heritability of both traits is an acceptable value, it is evident that environment is a large factor.

Candidate genes for susceptibility to Agrobacterium

Around 169 gene models are annotated on the Williams82 v2.1 reference genome across the QTL region defined by the physical markers Gm_06_10823424 and Gm_06_12129311 was thoroughly investigated to identify genes controlling soybean's host defense response to *Agrobacterium* (Table 3.5). Leucine-rich repeat receptor-like protein kinases, pathogenesis-related proteins, WRKYs, MYBs, and peroxidases were among the genes found within this region. Notably, the gene *Glyma.06g131500,* a Dof-type zinc finger transcription factor involved in phytohormone signaling, exhibited an additional base pair insertion in Peking, resulting in premature termination.

In Peking, *Glyma.06g134700*, a leucine-rich repeat receptor-like protein kinase, displayed four SNPs, with two resulting in amino acid changes, making it of particular interest due to its known role in host defense. Additionally, the heat shock protein *Glyma.06g134900* had a single base pair insertion leading to premature termination, potentially affecting plant immunity and abiotic stress responses.

A rearrangement was observed in Peking between the gene models *Glyma.06g136500* and *Glyma.06g137600*, involving a duplication of three genes and a single gene translocation from Chr04. The impact of this rearrangement on neighboring genes remains to be fully understood. Furthermore, *Glyma.06g137800* had a large deletion in Peking, rendering it nonfunctional. This gene is relevant to plant immunity and can induce defense responses.

In comparison to Williams82, Peking lacked several genes for ribosomal proteins in the QTL region. Additionally, *Glyma.06g141100*, a leucine-rich repeat receptor-like protein kinase, had a single base pair insertion leading to premature termination, suggesting potential involvement in plant immunity and host defense.

Evidence of a TNT 1-94 transposon was found in Peking between the genes *Glyma.06g145000* and *Glyma.06g145100*, with potential effects on gene expression in nearby nodulins, a superoxide dismutase, and a R-R type MYB protein.

A gene encoding a protein of unknown function (DUF260) exhibited differences from Williams82 and Jack, possibly influencing xylem cell differentiation. Genes *Glyma.06g146400*, *Glyma.06g146500*, and *Glyma.06g147500* were located within the highest LOD score region and encoded a leucine-rich repeat TIR domain protein, a SKP-

1 interacting partner, and a WRKY13. Although they did not exhibit presence/absence variation, several promoter changes might impact expression. An interesting novel gene without homology to any characterized soybean genes was identified, resembling an FBD-associated F-box protein or a cysteine-rich receptor-like protein kinase 20. Its potential involvement in susceptibility to *Agrobacterium* warrants further investigation, pending supporting expression data.

Interestingly, many genes within this identified QTL have been previously identified as either a flavonol synthase gene, flavonol 3-O-methyltransferase genes, or a chalcone-flavonone isomerase gene (Knizia et al., 2021). These genes were identified as contributing to the amount of daidzein or genistein present within soybean seeds. While more transcriptomic work will help determine if these genes are also involved with the compatibility of *Agrobacterium* and soybean, prior research has identified the importance of isoflavone levels in mediating symbiotic interactions in soybean. The isoflavone content of Peking has been studied before, and literature supports that these nod-gene-inducing flavonoids help to characterize the cultivar-specific interactions during symbiosis (Pueppke et al., 1998).

RNA-Seq analysis of major QTL on Chr06 for Agrobacterium susceptibility

As this study identified genetic regions associated with susceptibility in the mapping population, we aimed to further characterize the QTL region with RNA-Seq data to identify potential candidate genes that are involved with the *Agrobacterium*-soybean interaction. An RNA-Seq study was designed using Jack and Peking across four timepoints and two treatments, mock inoculation and inoculation with *Agrobacterium*

(Figure 3.14, Table 3.6). While forty-eight libraries were initially attempted, one failed library preparation and one was removed due to being an outlier. The remaining libraries produced six G of raw data, with over twenty million reads per library (Table 3.7). Principal component analysis and heat map clustering of the samples showed the samples clustering together in distinct groups for the experiment, while indicating that genotype and time were the most important causes for variation within the experiment (Figure 3.15, Figure 3.16).

The scope of the RNA-Seq experiment captured many genome-wide transcriptomic differences between the genotypes, timepoints, and treatments (Figure 3.17, Table 3.8). For the purposes of understanding differences within the QTL identified on Chr06, analysis focused on the 1.3 mb region to identify candidate genes.
At the twenty-minute timepoint, six transcripts were significant between Jack-*Agrobacterium* and Peking-*Agrobacterium* (Table 2.9). The under-expressed genes in Jack are annotated as a transcription factor PHOX2, and a chalcone-flavanone isomerase family protein. The overexpressed genes in Jack are annotated as an integral component of the membrane, and as a cleavage site for pathogenic type III effector avirulence factor, *Avr*. Another gene that was identified lacks annotations and has no corresponding gene in arabidopsis.

At the eight-hour time point, five transcripts were significant between Jack-*Agrobacterium* and Peking-*Agrobacterium* (Table 3.10). The only overexpressed gene in Jack is a cytochrome B5 isoform A. The under-expressed genes in Jack are an oxidoreductase protein, a ribosomal S26e protein, and a membrane-associated kinase regulator, *Glyma.06g147200*.

At the twenty-four-hour timepoint, three transcripts levels were significantly different between Jack-*Agrobacterium* and Peking-*Agrobacterium* (Table 3.11). The transcript that has a corresponding gene ID does not have any annotation or corresponding gene in Arabidopsis, and the remaining transcripts do not have a corresponding soybean gene ID. These transcripts have low base means, lacking uniformity within replicates, and would be filtered with more stringent parameters. These genes could be considered as playing a role in soybean compatibility if the other identified transcripts are disregarded.

At the seventy-two-hour timepoint, six transcripts were significant between Jack-*Agrobacterium* and Peking-*Agrobacterium* (Table 3.12). Two transcripts were under-expressed in Jack, but only one is annotated as a Yippee family zinc-binding protein. Another transcript was overexpressed in Jack, a Cytochrome B5 isoform A.

As Jack and Peking are different in their genetic background and transcriptomes, a contrast was also calculated comparing Jack-Mock to Jack-*Agrobacterium*. These specific contrasts address the question of what is being differentially expressed between Jack-Mock and Jack-*Agrobacterium* across these timepoints within the previously identified QTL region.

At the eight-hour timepoint, five genes were identified that were all under-expressed in Jack-Mock when compared to Jack-*Agrobacterium* (Table 3.13). A peroxidase is overexpressed in Jack-*Agrobacterium*, a hallmark of plant defense. A chalcone-flavanone isomerase family protein, alginate lyase protein, and a transducin/WD40-repeat protein are also overexpressed in Jack-*Agrobacterium*.

Transducins or WD40-repeat proteins have been shown to be involved in many processes, including plant cell wall formation, and regulating plant immunity.

At the twenty-four-hour timepoint, fourteen transcripts were identified within this region (Table 3.14). The only overexpressed genes in Jack-Mock were a Dof-type ring finger DNA-binding family protein and a nucleotide binding protein. Two transcripts belonging to proteins annotated as integral components of the membrane were overexpressed in Jack-*Agrobacterium*. Similarly overexpressed were a leucine-rich repeat protein kinase, a nodulin, a different transducing/WD-repeat protein, a calcium-dependent lipid-binding protein, and the same peroxidase identified at the eight-hour-timepoint. Leucine-rich repeat protein kinases play large roles in plant immunity, as well as calcium, an important signaling molecule. Only a single transcript was identified at the seventy-two-hour timepoint, annotated as an integral component of the membrane (Table 3.15).

In general, contrasts between Jack-Mock and Jack-*Agrobacterium* had a greater number of significant differentially expressed genes in this region than contrasts between Peking-Mock and Peking-*Agrobacterium*. When compared at twenty minutes, Peking-Mock had only a single transcript identified as being under-expressed, a major facilitator protein (Table 3.16). At the eight-hour time point, the Peking-*Agrobacterium* group increased expression of a mitochondrial import inner membrane translocase protein and a cytochrome B5 isoform E protein (Table 3.17). The Peking-*Agrobacterium* group also increased expression of genes seen in the Jack-Mock and Jack-*Agrobacterium* group, a peroxidase, and a chalcone-flavanone isomerase family protein. The fold change of the chalcone-flavanone isomerase family protein was similar across genotypes when

infected, but Peking had much higher expression of the peroxidase than Jack. Compared to Jack at twenty-four hours and significant changes in possible plant defense expression, Peking-Mock had two transcripts upregulated, a transmembrane protein and a photosystem subunit (Table 3.18). At seventy-two hours, Peking-*Agrobacterium* had two significant overexpressed genes, the chalcone-flavanone isomerase family protein, and the peroxidase. These genes were also identified at earlier timepoints, in both Jack and Peking (Table 3.19).

Galling ability of inbred lines

The results shown here identify a subset of a recombinant inbred line population that surpasses the previously most susceptible line Peking. Based on CRISPR/Cas9 mutagenesis and microbiology work, the gene representing GmEFR does not appear to control susceptibility or recognize bacterial EF-Tu, but selection of the Peking allele (deletion) during development of this population suggests that a nearby gene to GmEFR may be involved with controlling susceptibility to *Agrobacterium*. Other susceptible accessions in the germplasm collection also possess the Peking allele, indicating a shared haplotype in this region that may be harboring an important variant.

The GmEFR gene, *Glyma.09g216400*, is within a haplotype found in Peking and many landraces possessing identical SNPs, but elite varieties have the opposite SNP calls, indicating this region has been selected against during domestication. This region contains many plant defense response genes, including leucine-rich repeat receptor-like protein kinases (*Glyma.09g* 215700, Glyma.*09g216100, Glyma.09g217500*), nodulin-like protein (*Glyma.09g216300*), and cytoplasmic protein kinase domain-containing proteins

(*Glyma.09g217200)*. The region also contains glycosyl hydrolases, carbon catabolite repressor proteins, and a pectinacetylesterase.

Haplotype analysis of the recombinant inbred lines and mapping suggested the RILs inherited the block of DNA containing the GmEFR deletion with recombination events in the flanking regions, as would be expected. The immediately adjacent SNPs to the Peking allele of GmEFR was present in all RILs when examined in GenomeStudio. Additionally, the RILs were tested across many generations using PCR and confirmed to have the deletion. As these SNPs are monomorphic within the mapping population, no QTL could be mapped as it is fixed in all lines. Using the RNA-Seq dataset, two noticeable differences were identified in the regions flanking GmEFR on Chr09 between Jack and Peking during infection.

Two genes were identified using transcriptomic data that differed in a 100 kb region flanking GmEFR on Chr09. The two genes, *Glyma.09g216100* and *Glyma.09g216300,* had minor differences in Peking when compared to Jack. While *Glyma.09g216100,* a leucine-rich repeat receptor-like protein kinase, was expressed in Jack, it was not expressed in Peking across timepoints or treatment. The second gene, *Glyma.09g216300,* a nodulin-like protein, was expressed in Peking and Jack, but unlike Jack, the first exon was never expressed. The deletion of GmEFR did not result in novel transcripts in this region.

Plant immunity differences

While the RNA-Seq study was aimed to understand differentially expressed genes within a QTL region on Chr06, a few differences across the transcriptome between Jack

and Peking were noted. Most importantly at the eight-hour timepoint in this study, classical signs of plant immunity and subsequent changes in expression were present. W

While GO enrichment identifies thiazole and thiamine biosynthesis families as being the most enriched in Jack, enrichment also identified seven TIR (toll interleukin 1 receptor) domain containing proteins, overlapping with nine NB-ARC proteins and 13 leucine-rich repeat domain superfamily proteins (Figure 3.18, Table 3.20).

To further the support for these immunity genes as playing a role in Jack's resistance to *Agrobacterium*, two other hallmarks of plant immunity signaling are also overexpressed in Jack. The first group are GDSL lipase/esterase-like proteins. A lipase mutant in arabidopsis was found to exhibit enhanced auxin responses and to be more susceptible to bacteria than wild-type plants (Lee et al., 2009). A similar study in pepper found that GDSL-type lipases modulate disease susceptibility to *Xanthomonas* and *Pseudomonas* (Hong et al., 2008). The current literature suggests that these lipases play a role in bacterial resistance through negative regulation of auxin signaling, as well as regulating glucose and oxidative stress signaling.

A second group of proteins enriched in Jack belonged to the heat shock protein (DnaJ) superfamily. Heat shock proteins (Hsp) are considered critical for plant defense responses and are involved with modulating the structure of R proteins through acting as a chaperone as well as being implicated in resistance gene regulation (Elmore et al., 2011). In rice, Hsp90 was found to be required for transporting the rice chitin receptor to properly defend against the rice blast fungus (Chen et al., 2010). Similar to Hsp90, Hsp70 was found to be a target of a *P. syringae* effector, HopI1, to interfere and promote infection in arabidopsis (Jelenska et al., 2010). In tobacco, the soybean Hsp40 caused a

hypersensitive response (HR) cell death, and when silenced in soybean, enhanced susceptibility to soybean mosaic virus. The response was found to be depended on MAP kinase pathways.(Liu and Whitham, 2013)

The data suggests that at the eight-hour timepoint, Jack is undergoing a potent immune response to *Agrobacterium.* A plausible model can be constructed from the RNA-Seq data, starting with TIR containing R proteins that work through downstream mediators of lipases and MAPKs, eventually leading to cell death.

Discussion

This study identified a QTL on Chr06 that is implicated in controlling *Agrobacterium*-soybean interactions within the Peking and Jack genotypes. The genes within these regions were candidates of interest for analyzing their gene expression profiles using RNA-Seq. As many recombinant inbred lines surpassed that of the parent Peking in terms of galling ability, the effect of the marker used for selection of the GmEFR allele from Peking may have had an impact on susceptibility, but this warrants more study as many lines were similar to Jack regarding susceptibility. While GmEFR was ruled out for controlling *Agrobacterium* susceptibility, numerous other genes are located near GmEFR. Interestingly, another leucine-rich repeat kinase is three gene models upstream of GmEFR, and transcriptome data shows this gene is not expressed in Peking.

A previous study has attempted to define the genetic control of *Agrobacterium* susceptibility in soybean (Bailey et al., 1994). While this study relied on different parents than the present study, it concluded that one gene or two genes control susceptibility to

Agrobacterium, depending on the parental genotypes. The data presented here may support a gene from Peking on chromosome 6, identified as a significant QTL, that is important for susceptibility. The use of the GmEFR on Chr09 during population development suggests a nearby gene is also present within the population and is linked to the Peking allele of GmEFR, as breeding of susceptibility into the RILs was evident through using the GmEFR marker. Another study suggested two to three genes are involved in the quantitative inheritance of susceptibility in soybean (Mauro et al., 1995). It may be possible that within the population presented here that a majority of the RILs possess two of the genes important for susceptibility, and that the hypothesized third gene is segregating and may explain while some RILs outperformed Peking. It is clear that other genes and environment are important, as evident by at least three different minor QTL regions that were also identified for gall size and gall number.

Soybean and *Agrobacterium* interactions have remained elusive when trying to understand the role of the plant immune system and the factors that influence susceptibility or resistance. This study identified a QTL on Chr06 which contained 169 candidate gene models. Through RNA-Seq analysis, twenty genes were identified as being differentially expressed between the resistant and susceptible genotypes when infected. The twenty identified genes are excellent candidate genes and targets for increasing susceptibility to *Agrobacterium* in soybean. Additionally, the use of the GmEFR marker during population development suggests a nearby gene involved in susceptibility, and transcriptomic data supports two adjacent genes to GmEFR as candidate genes.

Materials and Methods

Population development

A biparental cross between the maternal recipient Jack (PI 540556), a resistant but embryogenic cultivar, and the paternal donor PxC, an inbred susceptible variety developed in the Parrott Lab, was made in a greenhouse in Athens, Georgia, during the Fall 2016 (Bailey et al., 1994). In Spring 2017, a PCR for detecting the presence/absence of the homolog of arabidopsis Elongation Factor Receptor, *Glyma.09g216400* (GmEFR), a candidate gene for controlling *Agrobacterium* susceptibility in soybean. The primers were used to confirm the F_1 seed and planted in a greenhouse in Athens, Georgia. Seed from the F_2 generation were screened to be homozygous for the absence of GmEFR and were planted in the greenhouse in Athens, Georgia. By fixing the population for the GmEFR deletion, the expectation is that the population would still be segregating for other factors affecting susceptibility. Two generations of single seed descent method were used to advance the population to the F_4 generation at the USDA-ARS Winter Nursery, Puerto Rice. A total of 150 F_5 lines were planted in a greenhouse located in Athens, Georgia for the experiments.

Marker design and genotyping

A multiplex PCR was used to screen for the Jack allele, the Peking (from PxC parent) allele, or both alleles during the development of the population. Due to this deletion being present or absent in any individual, three primers were used. Two primers, TC1463_F and JHR452_R, detect the wild-type allele by binding to the 5’ UTR and 452 bp downstream of the start codon, respectively. The third primer, TCEFR_R, binds

downstream of the 3'UTR. In the susceptible genotypes, or having the deletion, this primer pairs with TC1463_F but the amplicon of 329 bp is significantly smaller than the resistant genotypes, or wild-type allele, of 4760 bp. The JHR452_R primer detects the wild-type allele while producing a smaller amplicon than the possible 4760 bp amplicon, allowing PCR to rapidly detect the genotype for *Glyma.09g216400*.

The PCR mix contains two premixed ready-to-use solutions, GoTaq Long and GoTaq Green (Promega Corporation, Wisconsin, United States). Instead of the protocol-referenced final concentration of 1X, the two master mixes were prepared to 0.5X final concentration, with the sum of the two being 1X. The forward primer TC1463_F was used at a final concentration of 15pmol, with the reverse primers JHR452_R at 8.75 pmol and TCEFR_R at 6.25 pmol. PCR conditions were followed per manufacturer's instructions unless noted. The primer annealing temperature was set to 60° C with an extension time of 30 seconds. For each sample, 1µL of a 10 ng μL^{-1} of DNA was added to 10µL of master mix.

Phenotyping

For mapping, eight seeds from each RIL were planted in Fafard 3B potting mix (Sun Gro Horticulture Inc., Vancouver, Canada) across two cells in a 32-cell tray. The trays were placed in a growth chamber at 25º C and grown under a 23-h photoperiod (to prevent premature flowering induction) with a light intensity of 100 µmol m^{-2} s^{-1} for seven days with humidity domes. After nine days (VC growth stage), each genotype was thinned down to three plants and inoculated with *Agrobacterium* inoculum scraped from a yeast-mannitol plate. After infection, plants were returned to the growth chamber for

two days, maintaining the 23-h photoperiod but the temperature was lowered to 23º C, and the light intensity increased to 200 µmol m^{-2} s^{-1}. The inoculated seedlings were then transplanted into 15.24 cm nursery pots containing a soil mixture of 3B, sterile farm soil, and sand. The plants were grown in the greenhouse with supplemental lighting to maintain 18/6-hour light/dark and keeping the temperature constant between 23º C - 27º C. After 60 days in the greenhouse, the plants were phenotyped.

For RNA-Seq, seeds of soybean genotypes Jack (PI 540566) and Peking (PI 548402) were planted in Fafard 3B potting mix (Sun Gro Horticulture Inc., Vancouver, Canada) across two cells in a 32-cell tray. The trays were placed in a growth chamber at 25º C and grown under a 23h photoperiod and a light intensity of 100 µmol m^{-2} s^{-1} for seven days with humidity domes. After nine days (VC growth stage), each genotype was thinned down to 3 to 4 plants and inoculated with *Agrobacterium*. After infection, plants were returned to the growth chamber maintaining the 23-h photoperiod at a temperature of 23 ºC and light intensity of 200 µmol m^{-2} s^{-1} until the collection timepoint.

Statistical design

For mapping, a randomized complete block design (RCBD) with three replications was performed. Each replicate contained nine plants of each parent, the embryogenic Jack and *Agrobacterium*-susceptible Peking, and 150 recombinant inbred lines, with each line represented once per replication. Alternating border plants of Jack and Peking surrounded each block.

For RNA-Seq, a total of 48 libraries were attempted spanning across two genotypes, two treatments, four timepoints. Each combination of treatment, genotype,

and timepoint had 3 biological replicates. The sample Jack 24 hour mock failed library preparation three times and thus was not included in the final study. Another library, Peking 24 hour inoculated (Sample 41) was discarded from the study as an outlier after performing a principal component analysis of all libraries.

Inoculation

For mapping, *Agrobacterium tumefaciens* strain A281 (C58 containing the armed pTiBo542) was obtained from Eugene Nester and streaked from a glycerol stock on yeast-mannitol medium containing 50 mg mL^{-1} rifampicin (Chem-Impex International, Wood Dale, Illinois) and grown for three days in a 28° C incubator to produce the inoculum used for infection (Jin et al., 1993; Vincent, 1970). Sterile spatulas were used to spread *Agrobacterium* inoculum onto Giant Push Pins (Officemate, Edison, NJ). Each hypocotyl was punctured 2 cm above the soil completely through the stem. After puncturing, additional *Agrobacterium* inoculum was rubbed into each side of the puncture wound using a sterile spatula.

For RNA-Seq analysis, A281 was streaked from a glycerol stock and sterile spatulas were used to collect freshly grown two-day old *Agrobacterium* from yeast-mannitol plates and the inoculum spread onto Giant Push Pins (Officemate, Edison, NJ). The plants were stabbed completely through the hypocotyl, about 1 cm above the soil, and the pushpin pulled upward, creating a wound of about 1 cm. After puncturing, additional *Agrobacterium* inoculum was rubbed into each side of the puncture wound using a sterile spatula. The mock plants were treated similarly, but without

Agrobacterium. Stem sections containing the inoculation site were frozen in liquid nitrogen at their respective timepoints and placed in the -80 for later processing.

Phenotyping

Plants were evaluated after sixty days in the greenhouse. Most genotypes were approaching R1 growth stage, with some variation between genotypes based on maturity group. Images of each plant's entry or exit wounds, along with a ruler for calibration, were taken with a camera. The images were loaded into ImageJ, and the ruler used to calculate distance per pixel using the Measure function. Each gall was outlined by hand and size recorded (Schneider et al., 2012). For each plant, the number of galls was visually counted and recorded. Measurements of gall area and gall number were collected and analyzed using Microsoft Excel and RStudio (Racine, 2012).

Genotyping

A young trifoliolate leaf was collected from each genotype within each replication and lyophilized. The samples were ground in a Geno/Grinder (SPEX SamplePrep LLC, Metuchen, New Jersey) before undergoing DNA extraction using the Promega Wizard DNA extraction kit. Each extraction was resuspended in sterile water before quantification. After DNA quality was checked by gel electrophoresis, each individual line across the three replicates were pooled to ensure equimolar concentrations between the individuals. The samples were sent to the USDA in Beltsville, Maryland for genotyping with the Soybean BARCSoySNP6K Infinium chip (Song et al., 2020).

BLUP calculation

To capture the effects of having recombinant inbred lines repeated over three replications, the gall size and gall number phenotypic data were analyzed in JMP Pro 16.0.0 (SAS Institute, 2021). Using JMP, best linear unbiased predictions. (BLUPs) of the recombinant inbred lines and their measured phenotypes within replications were fit using a standard least squares model treating both genotype and replication as a random variable. The phenotypes were analyzed in R to calculate heritability.

Mapping analysis

After initial filtering in GenomeStudio based on visual inspection of SNP calling cluster quality and subsequent removal of monomorphic markers, the remaining 2424 markers were reformatted and analyzed in R using RStudio. The package r/qtl was used to create a genetic linkage map using the Kosambi mapping function (Broman et al., 2003). After checking each marker for segregation distortion using a Chi-squared threshold of significance and dropping markers with identical genotype information, 1706 markers remained. For each chromosome, plots of the estimated recombination fractions and LOD scores were visually inspected. Problematic markers, or markers with a large positive LOD score with a corresponding large change in genetic length, were excluded, leaving 1676 markers for mapping. After construction of the genetic maps, Composite Interval Mapping and Simple Interval Mapping were performed using the R package ‘qtl’. Genotypes were simulated using the command sim.geno with 1-cM step size across 32 simulation replicates using the Kosambi mapping function with an error

probability of 0.0001. After genotype simulations and calculating genotype probabilities, the mapping functions were performed for gall size and gall number.

RNA extraction

The stems were ground using mortar and pestle with liquid nitrogen or ground in a Geno/Grinder with liquid nitrogen pre-chilled blocks (SPEX SamplePrep LLC, Metuchen, New Jersey). The samples underwent a CTAB extraction to help remove polysaccharides found in soybean stems before performing the RNA extraction (Murray and Thompson, 1980). The total RNA of each tissue sample was extracted with Tri-Reagent according to the manufacturer's instructions (Thermo Fisher Scientific).

While the typical extraction procedure was sufficient for most mock-inoculated stem sections, an issue became apparent when extracting from *Agrobacterium*-infected stems, especially with the Peking genotype. These samples typically precipitated a gelatinous mass when the Trizol supernatant was added to the ethanol. As this prevented proper centrifugation through columns and tended to reduce yield and quality, additional steps including 24:1 chloroform:isoamyl alcohol and 25:24:1 phenol:chloroform:isoamyl alcohol wash were included to improve quality and yield.

The total extractions were each transferred to a Zymo-Spin IC Column in a collection tube and purified according to the manufacturer's protocol to remove genomic DNA and ensure high quality. All RNA extractions were analyzed on an Agilent 2100 (Agilent Technologies, Santa Clara, CA) and had RNA Integrity Number values greater than 7.0.

References

1. Bailey, M. A., Boerma, H. R., & Parrott, W. A. (1994). Inheritance of *Agrobacterium tumefaciens*-induced tumorigenesis of soybean. *Crop Science, 34*(2), 514-519.
2. Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics, 19*(7), 889-890.
3. Byrne, M. C., McDonnell, R. E., Wright, M. S., Carnes, M. G. J. P. C., Tissue, & Culture, O. (1987). Strain and cultivar specificity in the Agrobacterium-soybean interaction. *8*(1), 3-15.
4. Chen, L., Hamada, S., Fujiwara, M., Zhu, T., Thao, N. P., Wong, H. L., . . . Shibuya, N. (2010). The Hop/Sti1-Hsp90 chaperone complex facilitates the maturation and transport of a PAMP receptor in rice innate immunity. *Cell Host & Microbe, 7*(3), 185-196.
5. Dan, Y. (2008). Biological functions of antioxidants in plant transformation. *In Vitro Cellular & Developmental Biology-Plant, 44*(3), 149-161. doi:10.1007/s11627-008-9110-9
6. Dan, Y., Armstrong, C. L., Dong, J., Feng, X., Fry, J. E., Keithly, G. E., . . . Duncan, D. R. (2009). Lipoic acid-an unique plant transformation enhancer. *In Vitro Cellular & Developmental Biology-Plant, 45*(6), 630-638. doi:10.1007/s11627-009-9227-5
7. Elmore, J. M., Lin, Z.-J. D., & Coaker, G. (2011). Plant NB-LRR signaling: upstreams and downstreams. *Current opinion in plant biology, 14*(4), 365-371.

8. Hinchee, M. A., Connor-Ward, D. V., Newell, C. A., McDonnell, R. E., Sato, S. J., Gasser, C. S., . . . Horsch, R. B. (1988). Production of transgenic soybean plants using *Agrobacterium*-mediated DNA transfer. *Bio/technology, 6*(8), 915-922.

9. Hong, J. K., Choi, H. W., Hwang, I. S., Kim, D. S., Kim, N. H., Choi, D. S., . . . Hwang, B. K. (2008). Function of a novel GDSL-type pepper lipase gene, CaGLIP1, in disease susceptibility and abiotic stress tolerance. *Planta, 227*, 539-558.

10. Jelenska, J., Van Hal, J. A., & Greenberg, J. T. (2010). Pseudomonas syringae hijacks plant stress chaperone machinery for virulence. *Proceedings of the National Academy of Sciences, 107*(29), 13177-13182.

11. Jin, S., Song, Y. n., Pan, S. Q., & Nester, E. W. (1993). Characterization of a virG mutation that confers constitutive virulence gene expression in Agrobacterium. *Molecular Microbiology, 7*(4), 555-562.

12. Knizia, D., Yuan, J., Bellaloui, N., Vuong, T., Usovsky, M., Song, Q., . . . Lakhssassi, N. (2021). The soybean high density ‘forrest’by ‘williams 82’snp-based genetic linkage map identifies QTL and candidate genes for seed isoflavone content. *Plants, 10*(10), 2029.

13. Lee, D. S., Kim, B. K., Kwon, S. J., Jin, H. C., & Park, O. K. (2009). Arabidopsis GDSL lipase 2 plays a role in pathogen defense via negative regulation of auxin signaling. *Biochemical and biophysical research communications, 379*(4), 1038-1042.

14. Liu, J. Z., & Whitham, S. A. (2013). Overexpression of a soybean nuclear localized type–III DnaJ domain-containing HSP40 reveals its roles in cell death and disease resistance. *The plant journal, 74*(1), 110-121.

15. Mauro, A., Pfeiffer, T., & Collins, G. (1995). Inheritance of soybean susceptibility to Agrobacterium tumefaciens and its relationship to transformation. *Crop Science, 35*(4), 1152-1156.

16. Murray, M., & Thompson, W. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research, 8*(19), 4321-4326.

17. Olhoft, P., Lin, K., Galbraith, J., Nielsen, N., & Somers, D. J. P. C. R. (2001). The role of thiol compounds in increasing Agrobacterium-mediated transformation of soybean cotyledonary-node cells. *20*(8), 731-737.

18. Owens, L. D., & Cress, D. E. (1985). Genotypic variability of soybean response to Agrobacterium strains harboring the Ti or Ri plasmids. *Plant Physiology, 77*(1), 87-94.

19. Pueppke, S. G., Bolanos-Vásquez, M. C., Werner, D., Bec-Ferté, M.-P., Promé, J.-C., & Krishnan, H. B. (1998). Release of flavonoids by the soybean cultivars McCall and Peking and their perception as signals by the nitrogen-fixing symbiont Sinorhizobium fredii. *Plant Physiology, 117*(2), 599-606.

20. Racine, J. S. (2012). RStudio: a platform-independent IDE for R and Sweave. In: JSTOR.

21. Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods, 9*(7), 671-675.

22. Song, Q., Yan, L., Quigley, C., Fickus, E., Wei, H., Chen, L., . . . Hyten, D. (2020). Soybean BARCSoySNP6K: An assay for soybean genetics and breeding research. *The plant journal, 104*(3), 800-811.

23. Trick, H., & Finer, J. (1998). Sonication-assisted Agrobacterium-mediated transformation of soybean [Glycine max (L.) Merrill] embryogenic suspension culture tissue. *Plant Cell Reports, 17*(6-7), 482-488.

24. Van Breusegem, F., & Dat, J. F. J. P. p. (2006). Reactive oxygen species in plant cell death. *141*(2), 384-390.

25. Vincent, J. M. (1970). A manual for the practical study of the root-nodule bacteria. *A manual for the practical study of the root-nodule bacteria.*

Fig. 3.1. Population development of the Jack x PxC population. Plants were selected to be homozygous for a deletion from the PxC parent, also found in Peking, in the F_2 generation before undergoing single-seed descent.

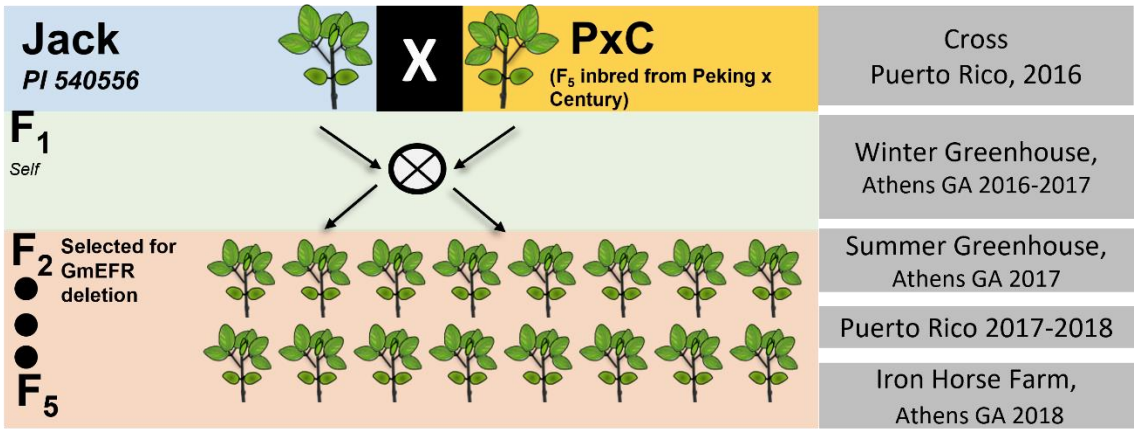


Fig. 3.2. Early testing of the F_4 generation of the recombinant inbred lines displayed susceptible phenotypes.

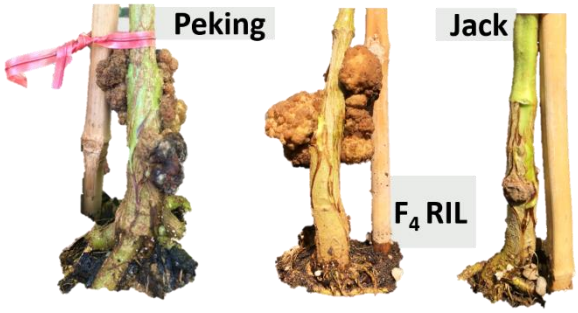


Fig. 3.3. Early testing of the F_4 generation of recombinant inbred lines exhibited varying degrees of susceptibility to *Agrobacterium*.

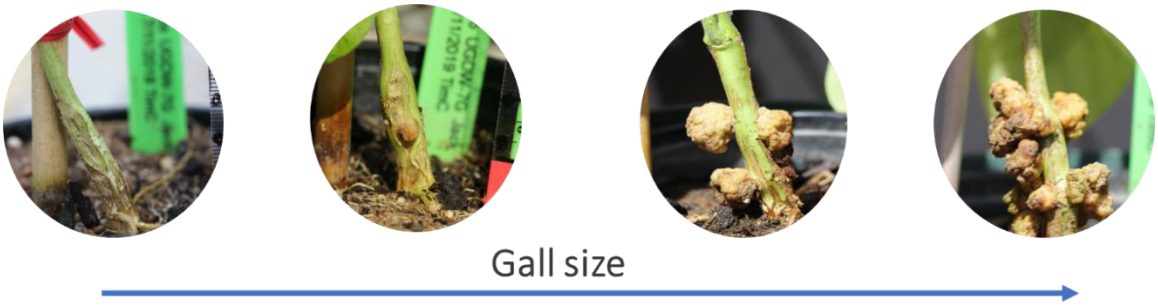


Fig. 3.4. Average gall size across three replicates of the parental genotypes Jack and Peking (n=9) and recombinant inbred lines (n=150) as groups. Error bars represent the standard error of the mean (SEM). As a group, the RILs performed intermediate of Jack and Peking.

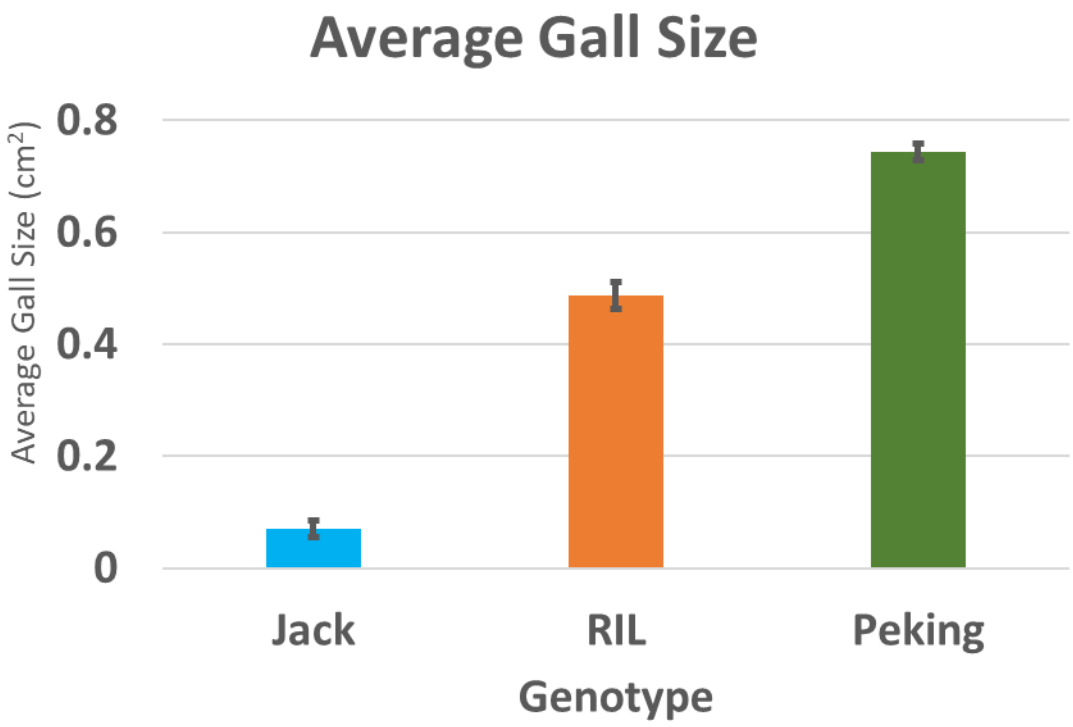


Table 3.1. ANOVA of the recombinant inbred lines for gall size across three replications.

ANOVA

	Df	*SS*	*MS*	*F*	*P-value*
Line	149	113.7	0.763	1.698	6.12E-05
Residuals	300	134.8	0.4493		
Total	449	248.5			

Table 3.2. Mean gall size and grouping of the recombinant bred lines using ANOVA from the RCBD across three replications. The data presented with a non-normal distribution and transformations including Box-Cox were tested.

Genotype	Mean Gall Size	Group
JxPC_143	2.154	a
JxPC_118	1.989	ab
JxPC_137	1.893	abc
JxPC_39	1.885	abc
JxPC_191	1.850	abcd
JxPC_130	1.842	abcd
JxPC_18	1.700	abcde
JxPC_6	1.649	abcdef
JxPC_101	1.595	abcdefg
JxPC_132	1.520	abcdefgh
JxPC_23	1.495	abcdefghi
JxPC_33	1.479	abcdefghij
JxPC_194	1.478	abcdefghij
JxPC_65	1.340	abcdefghijk
JxPC_38	1.314	abcdefghijkl
JxPC_142	1.298	abcdefghijklm
JxPC_97	1.283	abcdefghijklmn
JxPC_3	1.221	abcdefghijklmno
JxPC_179	1.164	abcdefghijklmnop
JxPC_45	1.141	abcdefghijklmnopq
JxPC_98	1.130	abcdefghijklmnopqr
JxPC_150	1.086	abcdefghijklmnopqrs
JxPC_123	1.014	bcdefghijklmnopqrst
JxPC_24	0.992	bcdefghijklmnopqrst
JxPC_1	0.989	bcdefghijklmnopqrst
JxPC_111	0.956	bcdefghijklmnopqrst
JxPC_19	0.927	bcdefghijklmnopqrst
JxPC_15	0.867	cdefghijklmnopqrst
JxPC_102	0.867	cdefghijklmnopqrst
JxPC_193	0.783	defghijklmnopqrst
JxPC_141	0.778	defghijklmnopqrst
JxPC_7	0.777	defghijklmnopqrst
JxPC_122	0.747	efghijklmnopqrst
JxPC_4	0.746	efghijklmnopqrst
JxPC_91	0.690	efghijklmnopqrst
JxPC_149	0.682	efghijklmnopqrst

JxPC_17	0.678	efghijklmnopqrst
JxPC_41	0.627	efghijklmnopqrst
JxPC_76	0.626	efghijklmnopqrst
JxPC_173	0.625	efghijklmnopqrst
JxPC_112	0.619	fghijklmnopqrst
JxPC_213	0.588	fghijklmnopqrst
JxPC_99	0.576	fghijklmnopqrst
JxPC_13	0.525	ghijklmnopqrst
JxPC_40	0.521	ghijklmnopqrst
JxPC_55	0.514	hijklmnopqrst
JxPC_89	0.504	hijklmnopqrst
JxPC_25	0.484	hijklmnopqrst
JxPC_11	0.472	hijklmnopqrst
JxPC_61	0.464	hijklmnopqrst
JxPC_86	0.461	hijklmnopqrst
JxPC_144	0.460	hijklmnopqrst
JxPC_68	0.431	ijklmnopqrst
JxPC_66	0.417	jklmnopqrst
JxPC_131	0.417	jklmnopqrst
JxPC_156	0.415	jklmnopqrst
JxPC_116	0.412	jklmnopqrst
JxPC_77	0.404	jklmnopqrst
JxPC_115	0.391	klmnopqrst
JxPC_119	0.388	klmnopqrst
JxPC_206	0.372	klmnopqrst
JxPC_95	0.372	klmnopqrst
JxPC_31	0.370	klmnopqrst
JxPC_129	0.369	klmnopqrst
JxPC_134	0.367	klmnopqrst
JxPC_50	0.365	klmnopqrst
JxPC_2	0.365	klmnopqrst
JxPC_90	0.357	klmnopqrst
JxPC_103	0.356	klmnopqrst
JxPC_114	0.350	klmnopqrst
JxPC_53	0.350	klmnopqrst
JxPC_69	0.345	klmnopqrst
JxPC_74	0.337	klmnopqrst
JxPC_177	0.334	klmnopqrst
JxPC_36	0.325	klmnopqrst
JxPC_204	0.321	klmnopqrst
JxPC_165	0.316	klmnopqrst
JxPC_184	0.313	klmnopqrst
JxPC_46	0.313	klmnopqrst

JxPC_108	0.307	klmnopqrst
JxPC_151	0.296	klmnopqrst
JxPC_16	0.277	klmnopqrst
JxPC_67	0.272	klmnopqrst
JxPC_140	0.263	lmnopqrst
JxPC_56	0.254	lmnopqrst
JxPC_106	0.243	lmnopqrst
JxPC_138	0.241	lmnopqrst
JxPC_135	0.233	mnopqrst
JxPC_62	0.226	mnopqrst
JxPC_70	0.220	nopqrst
JxPC_124	0.219	nopqrst
JxPC_96	0.212	nopqrst
JxPC_32	0.204	opqrst
JxPC_28	0.194	opqrst
JxPC_44	0.193	opqrst
JxPC_113	0.185	opqrst
JxPC_139	0.184	opqrst
JxPC_85	0.182	opqrst
JxPC_164	0.182	opqrst
JxPC_136	0.180	opqrst
JxPC_59	0.179	opqrst
JxPC_120	0.176	opqrst
JxPC_60	0.171	opqrst
JxPC_47	0.170	opqrst
JxPC_186	0.165	opqrst
JxPC_71	0.164	opqrst
JxPC_64	0.162	opqrst
JxPC_104	0.156	opqrst
JxPC_176	0.141	pqrst
JxPC_154	0.140	pqrst
JxPC_73	0.139	pqrst
JxPC_87	0.136	pqrst
JxPC_180	0.135	pqrst
JxPC_29	0.134	pqrst
JxPC_52	0.131	pqrst
JxPC_117	0.127	pqrst
JxPC_214	0.125	pqrst
JxPC_133	0.125	pqrst
JxPC_161	0.123	pqrst
JxPC_107	0.119	pqrst
JxPC_57	0.113	pqrst
JxPC_105	0.110	pqrst

JxPC_10	0.108	pqrst
JxPC_128	0.096	pqrst
JxPC_125	0.084	qrst
JxPC_49	0.084	qrst
JxPC_100	0.074	qrst
JxPC_48	0.069	qrst
JxPC_127	0.067	qrst
JxPC_30	0.066	qrst
JxPC_34	0.056	rst
JxPC_83	0.051	st
JxPC_42	0.048	st
JxPC_121	0.044	st
JxPC_147	0.044	st
JxPC_109	0.042	st
JxPC_54	0.036	st
JxPC_110	0.029	st
JxPC_94	0.025	st
JxPC_175	0.023	st
JxPC_78	0.019	st
JxPC_58	0.015	st
JxPC_51	0.007	t
JxPC_20	0.003	t
JxPC_126	0.000	t
JxPC_146	0.000	t
JxPC_174	0.000	t
JxPC_22	0.000	t
JxPC_88	0.000	t
JxPC_93	0.000	t

Table 3.3. Mean gall size and grouping of the recombinant bred lines according to the Scott-Knott test from the RCBD across three replications.

Genotype	Mean Gall Size	Group
JxPC_143	2.15	a
JxPC_118	1.99	a
JxPC_137	1.89	a
JxPC_39	1.88	a
JxPC_191	1.85	a
JxPC_130	1.84	a
JxPC_18	1.7	a
JxPC_6	1.65	a
JxPC_101	1.6	a
JxPC_132	1.52	a
JxPC_23	1.49	a
JxPC_33	1.48	a
JxPC_194	1.48	a
JxPC_65	1.34	a
JxPC_38	1.31	a
JxPC_142	1.3	a
JxPC_97	1.28	a
JxPC_3	1.22	a
JxPC_179	1.16	a
JxPC_45	1.14	a
JxPC_98	1.13	a
JxPC_150	1.09	a
JxPC_123	1.01	a
JxPC_24	0.99	a
JxPC_1	0.99	a
JxPC_111	0.96	a
JxPC_19	0.93	a
JxPC_15	0.87	a
JxPC_102	0.87	a
JxPC_193	0.78	b
JxPC_141	0.78	b
JxPC_7	0.78	b
JxPC_122	0.75	b
JxPC_4	0.75	b
JxPC_91	0.69	b
JxPC_149	0.68	b
JxPC_17	0.68	b

JxPC_41	0.63	b
JxPC_76	0.63	b
JxPC_173	0.62	b
JxPC_112	0.62	b
JxPC_213	0.59	b
JxPC_99	0.58	b
JxPC_13	0.52	b
JxPC_40	0.52	b
JxPC_55	0.51	b
JxPC_89	0.5	b
JxPC_25	0.48	b
JxPC_11	0.47	b
JxPC_61	0.46	b
JxPC_86	0.46	b
JxPC_144	0.46	b
JxPC_68	0.43	b
JxPC_66	0.42	b
JxPC_131	0.42	b
JxPC_156	0.41	b
JxPC_116	0.41	b
JxPC_77	0.4	b
JxPC_115	0.39	b
JxPC_119	0.39	b
JxPC_206	0.37	b
JxPC_95	0.37	b
JxPC_31	0.37	b
JxPC_129	0.37	b
JxPC_134	0.37	b
JxPC_50	0.37	b
JxPC_2	0.36	b
JxPC_90	0.36	b
JxPC_103	0.36	b
JxPC_114	0.35	b
JxPC_53	0.35	b
JxPC_69	0.34	b
JxPC_74	0.34	b
JxPC_177	0.33	b
JxPC_36	0.32	b
JxPC_204	0.32	b
JxPC_165	0.32	b
JxPC_184	0.31	b

JxPC_46	0.31	b
JxPC_108	0.31	b
JxPC_151	0.3	b
JxPC_16	0.28	b
JxPC_67	0.27	b
JxPC_140	0.26	b
JxPC_56	0.25	b
JxPC_106	0.24	b
JxPC_138	0.24	b
JxPC_135	0.23	b
JxPC_62	0.23	b
JxPC_70	0.22	b
JxPC_124	0.22	b
JxPC_96	0.21	b
JxPC_32	0.2	b
JxPC_28	0.19	b
JxPC_44	0.19	b
JxPC_113	0.19	b
JxPC_139	0.18	b
JxPC_85	0.18	b
JxPC_164	0.18	b
JxPC_136	0.18	b
JxPC_59	0.18	b
JxPC_120	0.18	b
JxPC_60	0.17	b
JxPC_47	0.17	b
JxPC_186	0.16	b
JxPC_71	0.16	b
JxPC_64	0.16	b
JxPC_104	0.16	b
JxPC_176	0.14	b
JxPC_154	0.14	b
JxPC_73	0.14	b
JxPC_87	0.14	b
JxPC_180	0.14	b
JxPC_29	0.13	b
JxPC_52	0.13	b
JxPC_117	0.13	b
JxPC_214	0.13	b
JxPC_133	0.12	b
JxPC_161	0.12	b

JxPC_107	0.12	b
JxPC_57	0.11	b
JxPC_105	0.11	b
JxPC_10	0.11	b
JxPC_128	0.1	b
JxPC_125	0.08	b
JxPC_49	0.08	b
JxPC_100	0.07	b
JxPC_48	0.07	b
JxPC_127	0.07	b
JxPC_30	0.07	b
JxPC_34	0.06	b
JxPC_83	0.05	b
JxPC_42	0.05	b
JxPC_121	0.04	b
JxPC_147	0.04	b
JxPC_109	0.04	b
JxPC_54	0.04	b
JxPC_110	0.03	b
JxPC_94	0.03	b
JxPC_175	0.02	b
JxPC_78	0.02	b
JxPC_58	0.02	b
JxPC_51	0.01	b
JxPC_20	0	b
JxPC_22	0	b
JxPC_174	0	b
JxPC_146	0	b
JxPC_88	0	b
JxPC_93	0	b
JxPC_126	0	b

Fig. 3.5. Frequency distribution of average gall size among parental lines and recombinant inbred line groups across replicates.

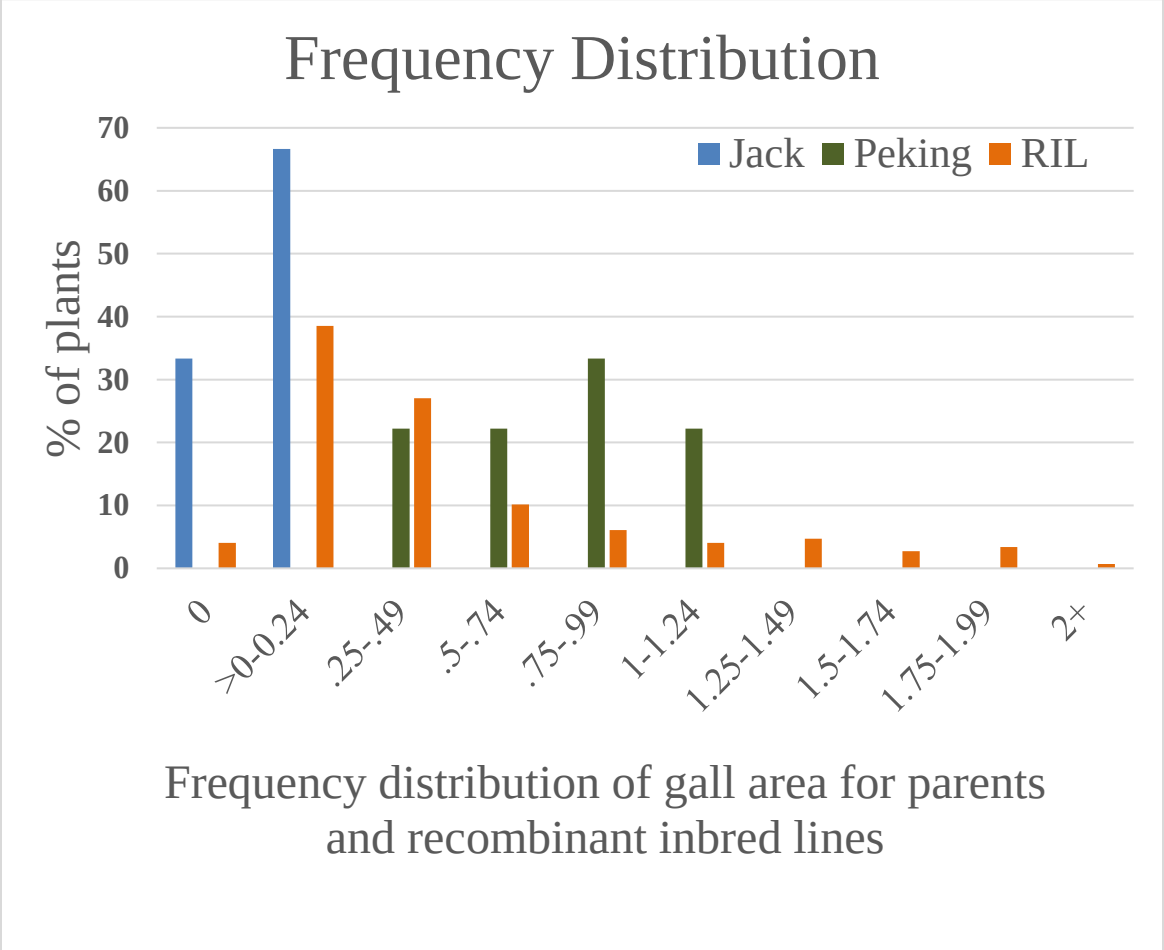


Fig. 3.6. Average gall size of each recombinant inbred line across replications. Jack is shown in blue between the individuals 127 and 100. Error bars represent the standard error of the mean (SEM). Due to the high number of RILs (n = 150), every other line is displayed on the x-axis.

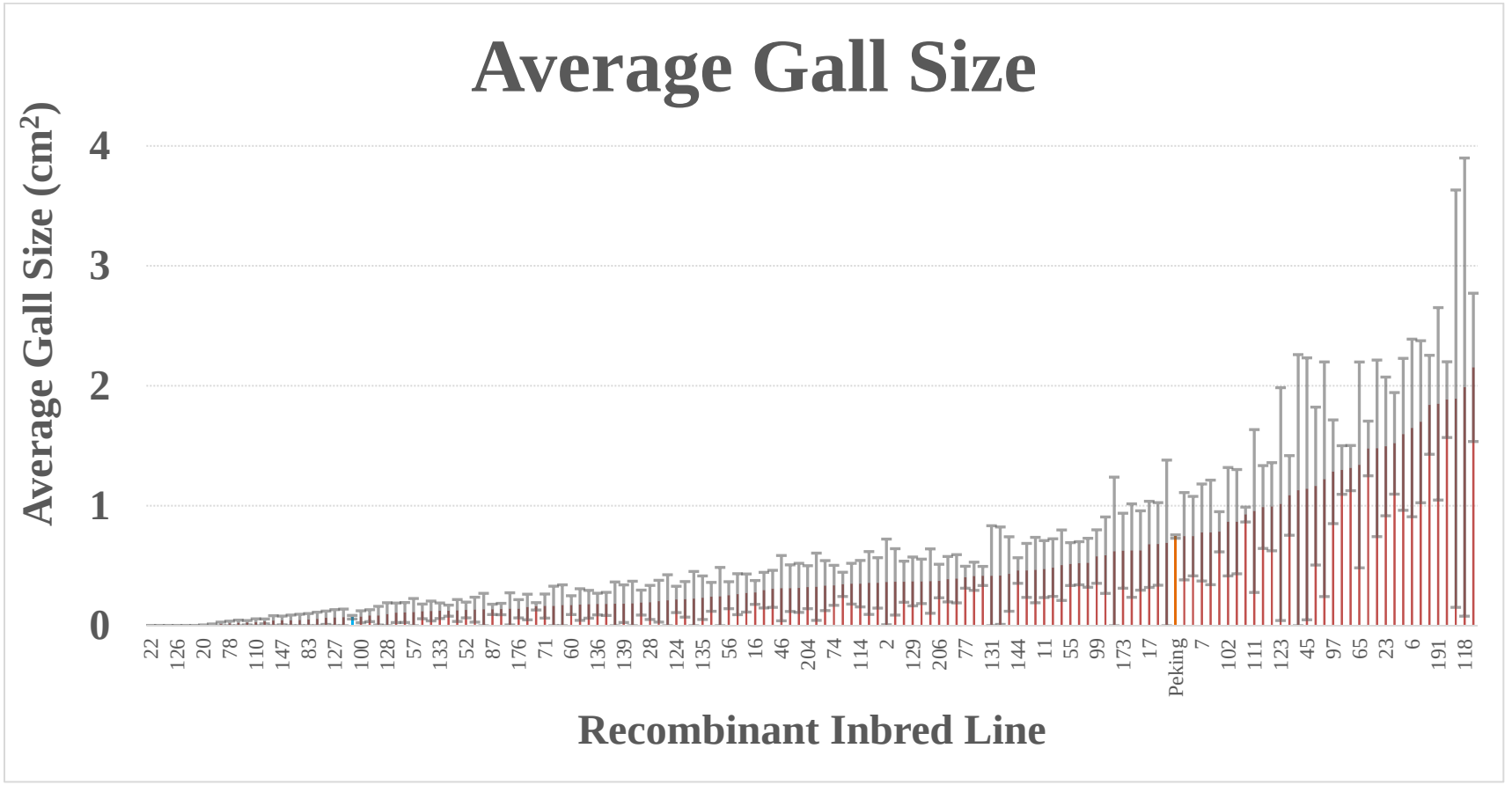


Fig. 3.7. Genetic map of the Jack x PxC population.

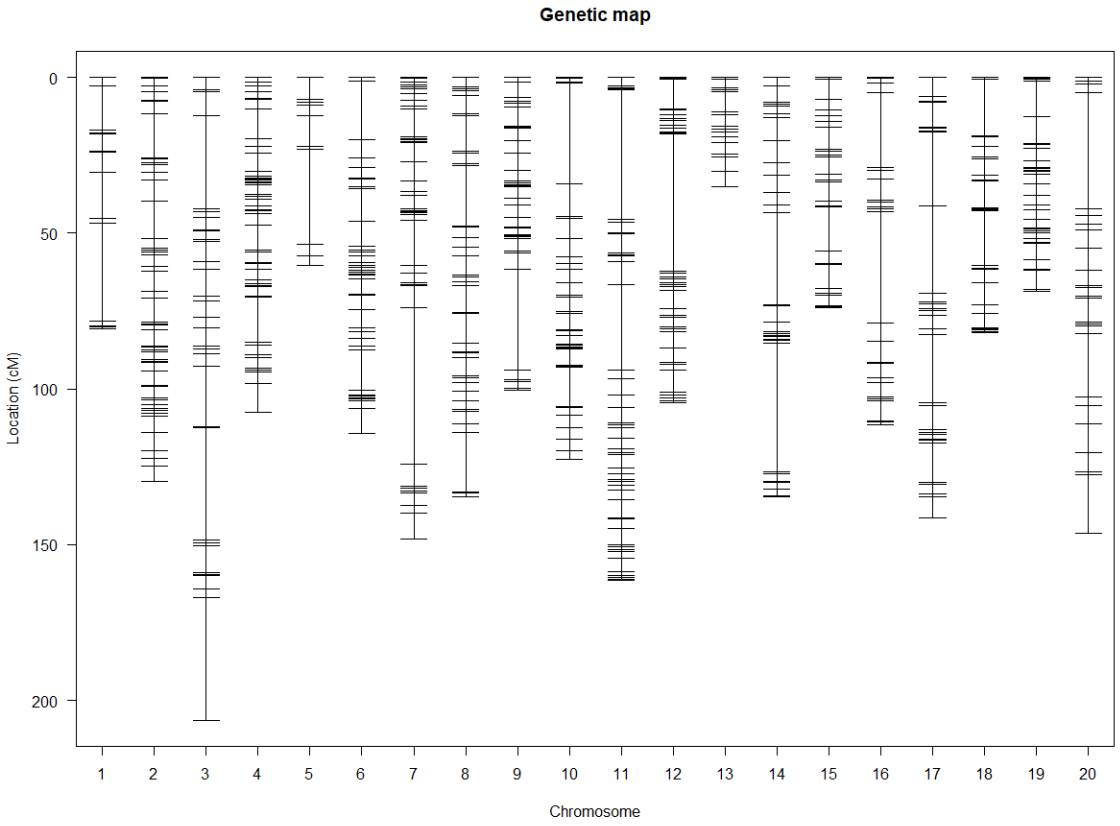


Fig. 3.8. Final pairwise recombination and LOD score plot across all 20 chromosomes in the Jack x PxC recombinant inbred population.

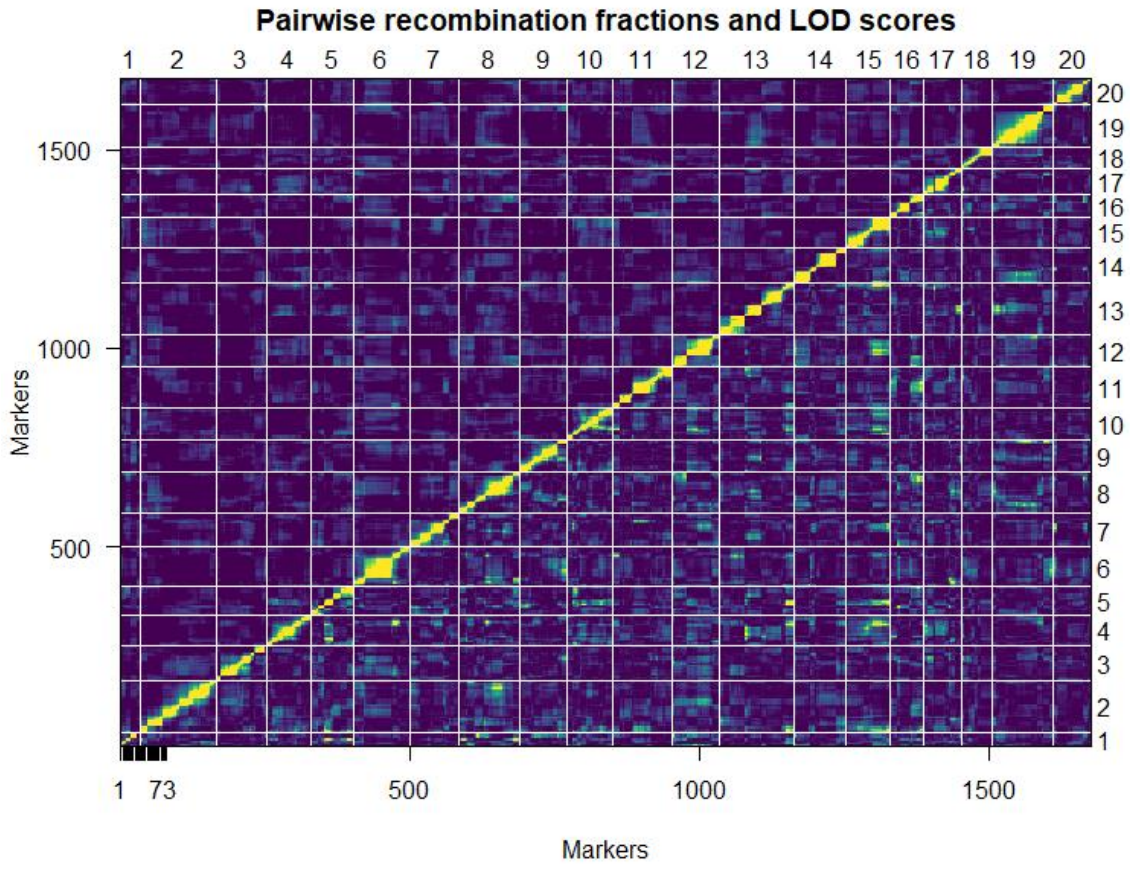


Fig. 3.9. QTL plot for gall size and gall number across all chromosomes. Three minor QTL (LOD > 2.0) for gall size were located on Chr02, Chr09, and Chr10. Three minor QTL for gall number were located on Chr07, Chr11, and Chr20.

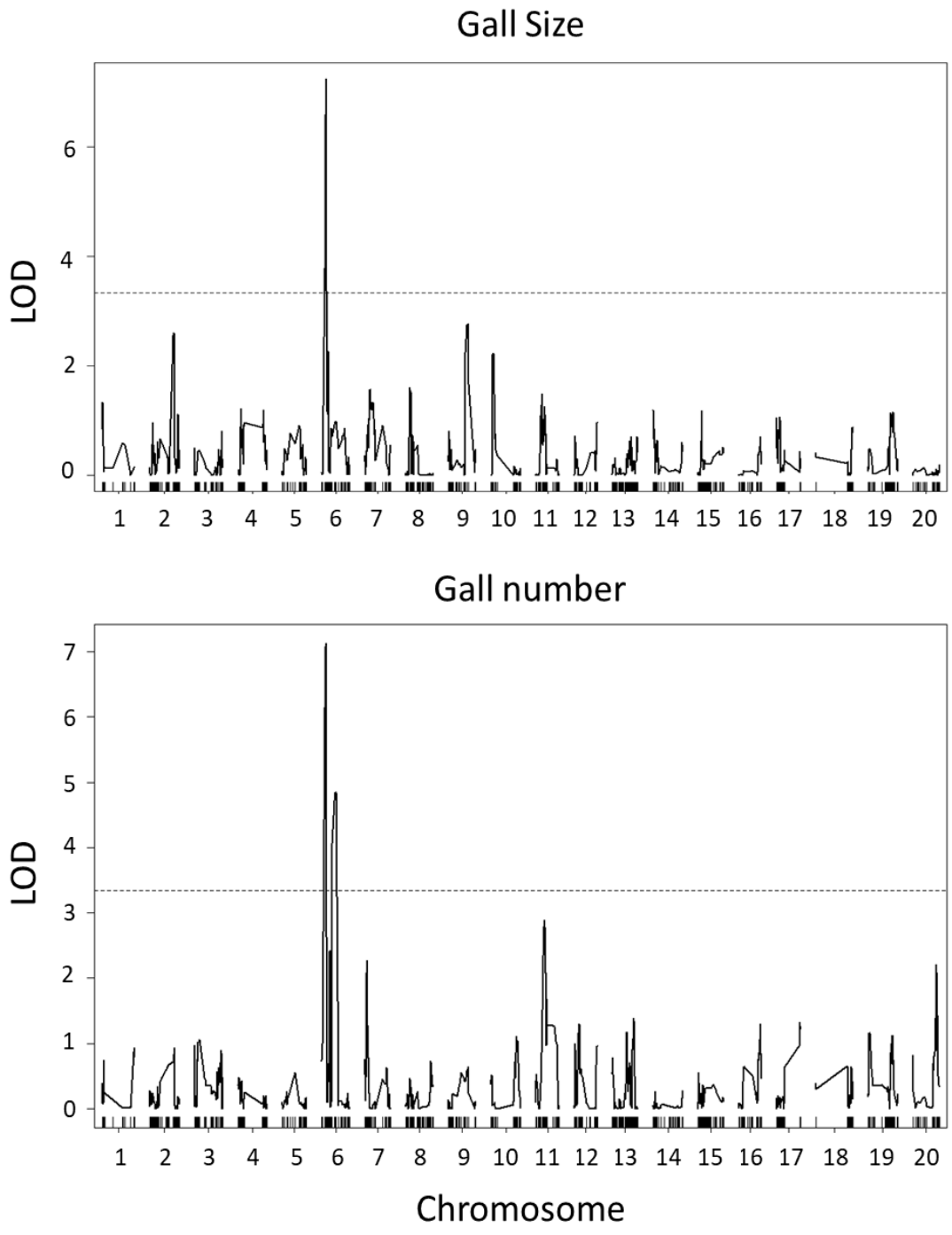


Table 3.4. Identified QTL for gall size and gall number within the Jack x PxC population using CIM.

CIM	*Chr*	*SNP*	*Position* (cM)	*LOD*
Gall Size				
	6	Gm06_10823424_C_T	29.739	6.702
	6	Gm06_10891060_T_C	29.739	6.702
	6	Gm06_10993554_A_G	29.739	6.702
	6	Gm06_11243298_G_T	29.739	6.702
	6	Gm06_11659627_A_G	32.746	7.277
	6	Gm06_11786580_T_G	32.746	7.277
	6	Gm06_11844498_A_G	36.148	5.124
	6	Gm06_11948808_G_A	36.148	5.124
	6	Gm06_12028624_A_G	36.148	5.124
	6	Gm06_12129311_G_A	36.148	5.124
Gall Number				
	6	Gm06_10823424_C_T	29.7385	6.573
	6	Gm06_10891060_T_C	29.7385	6.573
	6	Gm06_10993554_A_G	29.7385	6.573
	6	Gm06_11243298_G_T	29.7385	6.573
	6	Gm06_11659627_A_G	32.7455	5.557
	6	Gm06_11786580_T_G	32.7455	5.557

Fig. 3.10. QTL underlying gall size on Chr06 within the Jack x PxC population using composite interval mapping.

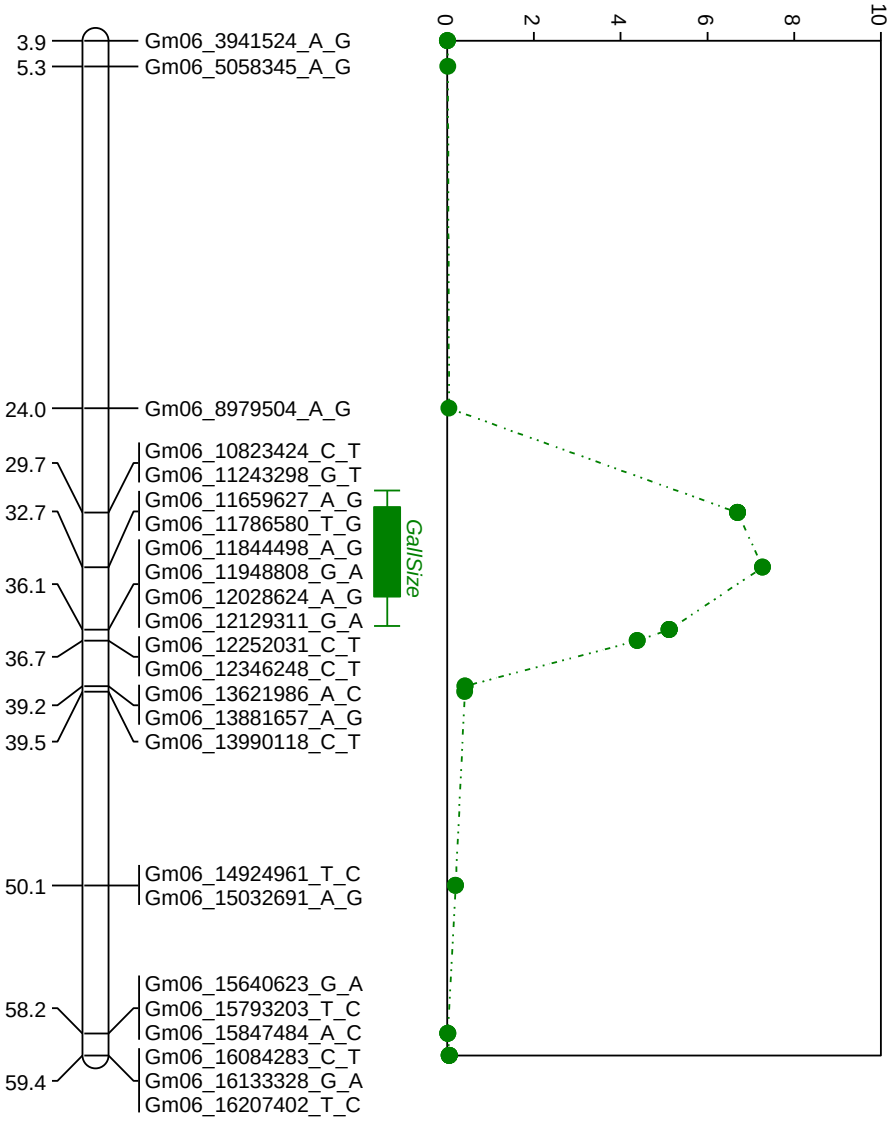


Fig. 3.11. QTL underlying gall number on Chr06 within the Jack x PxC population using composite interval mapping.

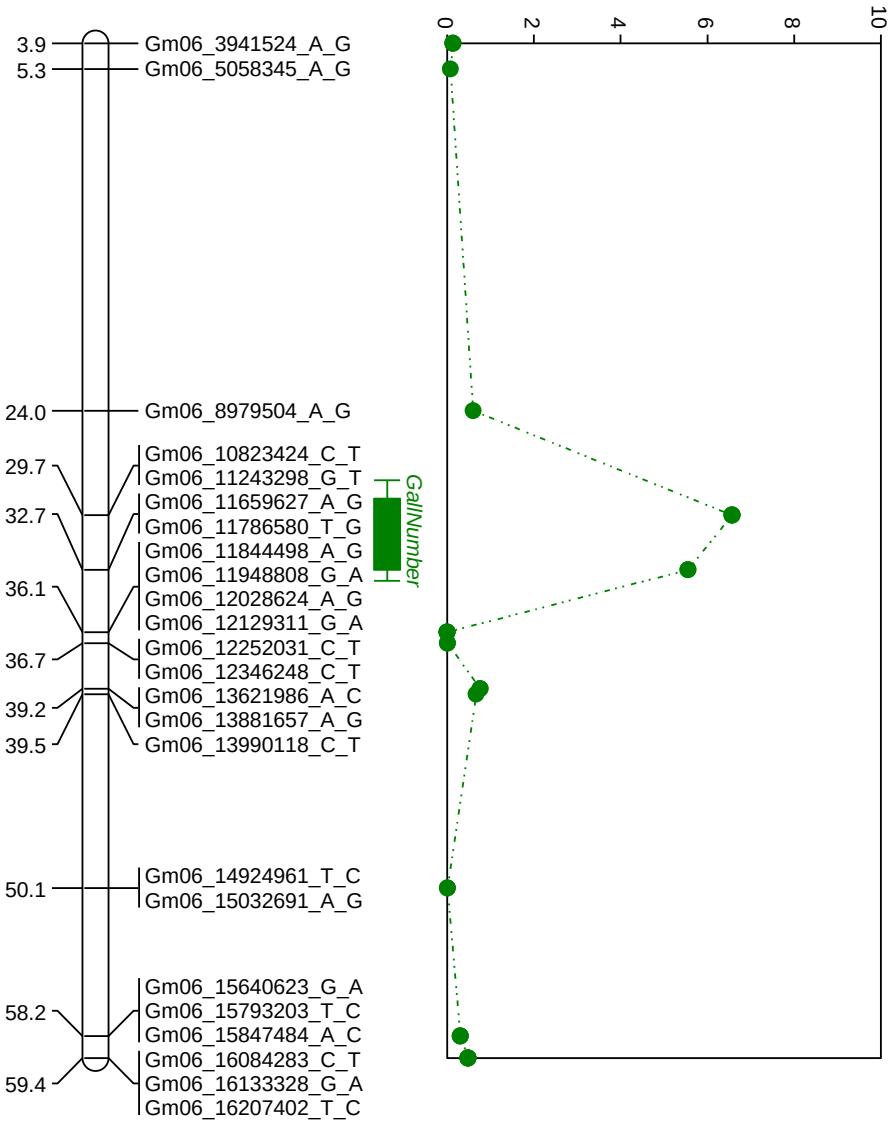


Fig. 3.12. Genotypic variant associated with controlling average gall size on Chr06.

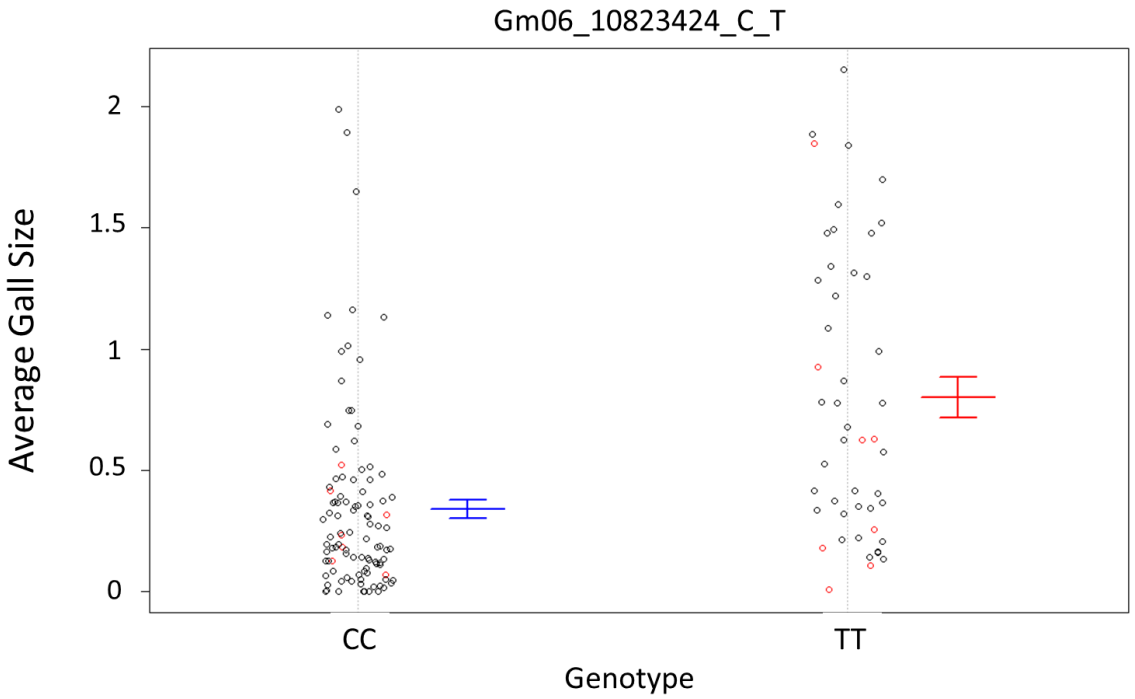


Fig. 3.13. Genotypic variant associated with controlling average gall number on Chr06.

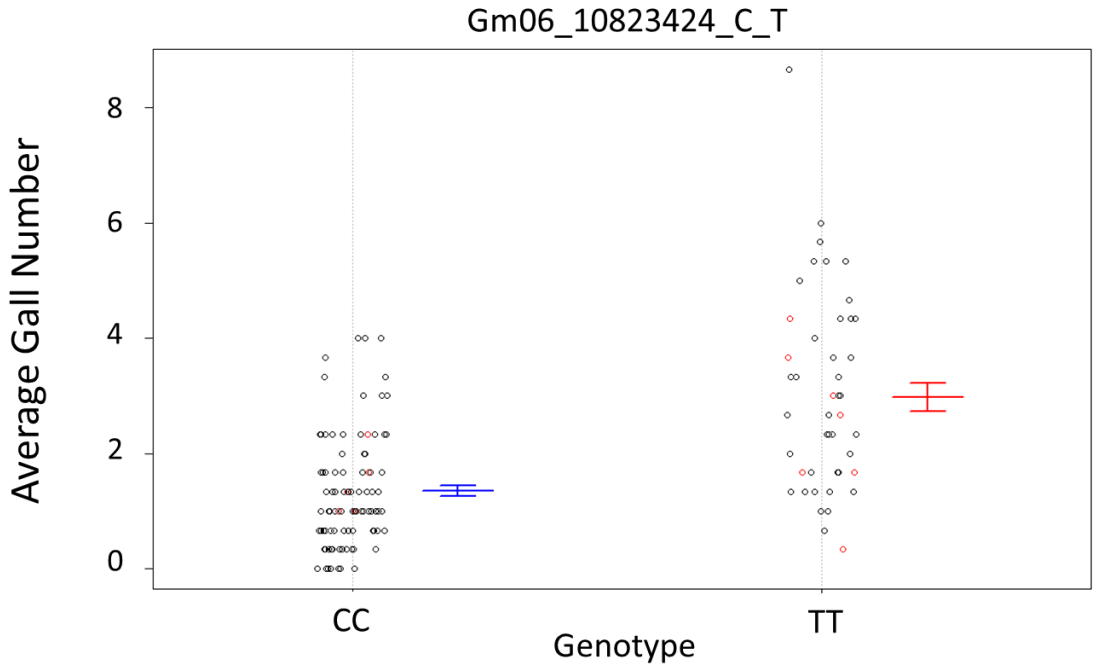


Table. 3.5. Soybean candidate genes within the QTL for gall size and gall number.

Gene	Annotation
Glyma.06g131400	WD repeat domain
Glyma.06g131500	Dof-type zinc finger
Glyma.06g131600	Yippee-type zinc-binding protein
Glyma.06g131700	RNA recognition binding
Glyma.06g131800	Oligopeptide transporter
Glyma.06g132100	ARM repeat family
Glyma.06g133100	Protein of unknown function (DUF 659)
Glyma.06g133500	6-Phosphogluconolactonase
Glyma.06g133800	Transcription factor homeobox 7
Glyma.06g133900	Plasminogen
Glyma.06g134000	Auxin responsive protein
Glyma.06g134200	Serine-threonine protein kinase CDPK
Glyma.06g134000	SSXT protein
Glyma.06g134600	Growth-regulating factor 7
Glyma.06g134700	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.06g134900	HSP20
Glyma.06g135000	
Glyma.06g136200	Phosphatidylcholine transfer protein
Glyma.06g136300	
Glyma.06g136700	Late embryogenesis abundant protein
Glyma.06g136800	RING-H2 zing finger protein
Glyma.06g136900	WD repeat containing protein
Glyma.06g137000	Oxidoreductase family protein
Glyma.06g137600	Hypothetical protein
Glyma.06g137700	Nascent polypeptide related
Glyma.06g137800	Nascent polypeptide related
Glyma.06g137900	Peroxidase
Glyma.06g138100	NAC transcription factor-like 9
Glyma.06g138200	Carbon catabolite repressor protein 4
Glyma.06g138400	
Glyma.06g138600	
Glyma.06g138700	
Glyma.06g138800	Major facilitator superfamily protein
Glyma.06g139000	Thaumatin family
Glyma.06g139100	Double-stranded DNA binding
Glyma.06g139300	Response to low sulfur 2
Glyma.06g139800	
Glyma.06g139900	Transducin/WD40 repeat-like protein
Glyma.06g141100	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.06g141200	Methyl esterase

Glyma.06g141400	Protein of unknown function (DUF3133)
Glyma.06g141600	
Glyma.06g142000	WRKY70
Glyma.06g142100	WRKY55
Glyma.06g142500	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.06g142700	U-Box domain protein
Glyma.06g143200	bZIP52
Glyma.06g143300	Expansin
Glyma.06g143400	Mini-chromosome maintenance replisome factor
Glyma.06g143600	R-R type MYB protein
Glyma.06g144500	Superoxide dismutase
Glyma.06g144600	nodulin transporter family protein
Glyma.06g144700	nodulin transporter family protein
Glyma.06g144800	
Glyma.06g144800	Pentatricopeptide repeat protein
Glyma.06g145100	
Glyma.06g145300	Peroxidase
Glyma.06g145400	Protein of unknown function (DUF260)
Glyma.06g144500	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.06g145800	
Glyma.06g145900	Ring finger and SRY domain-containing
Glyma.06g146400	Leucine-rich repeat containing protein TIR domain
Glyma.06g146500	SKP1-interacting partner
Glyma.06g146800	Cleavage site for pathogenic type III effector avirulence factor Avr
Glyma.06g147000	Programmed cell death protein 2-related
Glyma.06g147100	WRKY51
Glyma.06g147200	
Glyma.06g147500	WRKY13
Glyma.06g147600	BRI1

Fig. 3.14. Design of the RNA-Seq study.

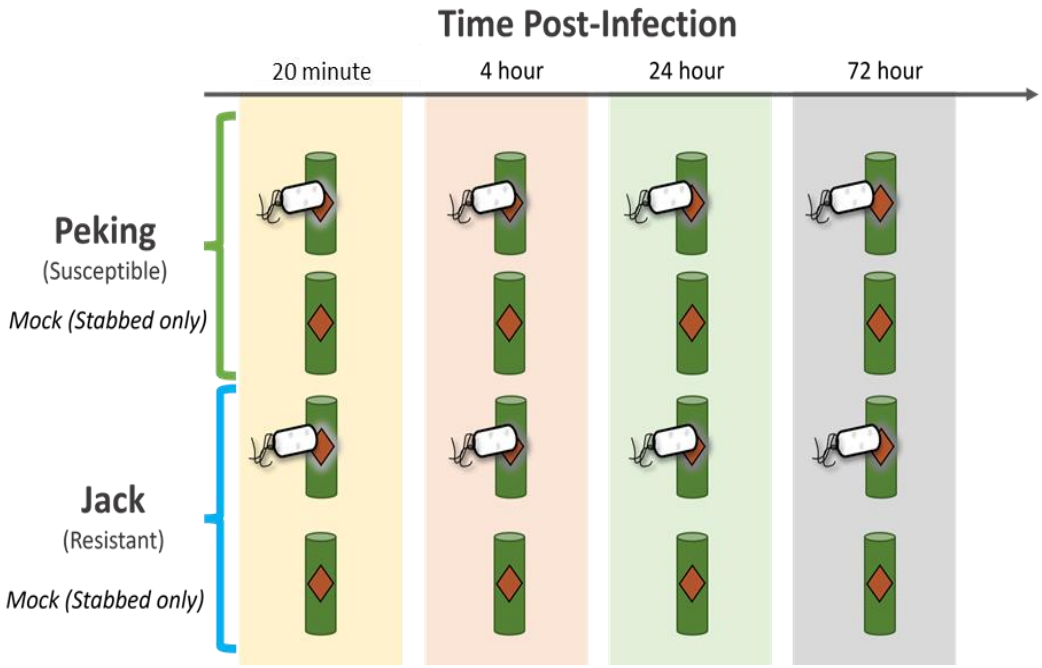


Table 3.6. Design of the RNA-Seq study. Two libraries were excluded, denoted by an asterisk.

Host Accession	Treatment	Time Point	Reps
Peking	A281	20min	3
Peking	Mock	20min	3
Peking	A281	8hpi	3
Peking	Mock	8hpi	3
Peking	A281	24hpi	2*
Peking	Mock	24hpi	3
Peking	A281	72hpi	3
Peking	Mock	72hpi	3
Jack	A281	20min	3
Jack	Mock	20min	3
Jack	A281	8hpi	3
Jack	Mock	8hpi	3
Jack	A281	24hpi	3
Jack	Mock	24hpi	2*
Jack	A281	72hpi	3
Jack	Mock	72hpi	3
		Total	**46**

Table 3.7. Statistics of the of the RNA-Seq data. Total raw data of 363.7 G. ID references the specific identifier used for barcoding and sequencing each library.

Sample	ID	Raw reads	Raw data	Effective(%)	Error(%)	Q20(%)	Q30(%)	GC(%)
J-20M-1	HA	47650676	7.1	99.55	0.03	97.9	93.7	45
J-20M-2	HB	42052216	6.3	99.7	0.03	98	94	45
J-20M-3	HC	41062198	6.2	99.74	0.03	97.8	93.5	45
J-20A-4	HD	56739592	8.5	99.81	0.03	97.8	93.6	45
J-20A-5	HE	55579776	8.3	99.73	0.03	98	94	45
J-20A-6	HF	53178464	8	99.22	0.03	98	93.9	45
J-8M-7	HG	47827176	7.2	99.62	0.03	97.8	93.6	46
J-8M-8	HH	56983700	8.5	99.58	0.03	97.9	93.8	45
J-8M-9	HI	44538184	6.7	99.65	0.03	97.8	93.6	45
J-8A-10	HJ	60776646	9.1	98.99	0.03	97.9	94	47
J-8A-11	HK	54708758	8.2	99.64	0.03	97.8	93.6	45
J-8A-12	HL	65727778	9.9	99.5	0.03	97.9	93.9	46
J-24M-13	JB	49695244	7.5	99.49	0.02	98.2	94.5	44
J-24M-14	JC	54228572	8.1	99.11	0.02	98.2	94.5	43
J-24A-15	JD	66597734	10	99.58	0.03	98	93.9	46
J-24A-16	JE	40583594	6.1	99.61	0.03	97.8	93.6	45
J-24A-17	JF	71828274	10.8	99.66	0.03	98	93.9	45
J-72M-18	JG	52192478	7.8	99.36	0.02	98.2	94.3	44
J-72M-19	JH	51833372	7.8	99.43	0.03	98	94	45
J-72M-20	JI	50556938	7.6	99.68	0.03	97.9	93.7	45
J-72A-21	HM	50467626	7.6	99.58	0.03	97.7	93.4	45
J-72A-22	HN	59994882	9	99.46	0.03	98	94.1	45
J-72A-23	JJ	53427066	8	99.7	0.03	97.9	93.7	44
P-20M-24	HO	46544600	7	99.51	0.03	98	93.9	45
P-20M-25	HP	55971450	8.4	99.49	0.03	97.8	93.5	46
P-20M-26	HQ	52392364	7.9	99.51	0.03	98	94	45
P-20A-27	PA	45505680	6.8	99.74	0.03	97.8	93.5	45
P-20A-28	HR	43536880	6.5	99.67	0.03	98.1	94.1	46
P-20A-29	PB	46859526	7	99.73	0.03	97.8	93.5	45
P-8M-30	PC	44878298	6.7	99.46	0.02	98.1	94.2	45
P-8M-31	PD	45829306	6.9	99.67	0.02	98.2	94.5	46
P-8M-32	PE	45986320	6.9	99.58	0.03	97.9	93.7	45

P-8A-33	PF	50400794	7.6	99.63	0.02	98.1	94.2	46
P-8A-34	PG	57221422	8.6	99.59	0.03	98	94	46
P-8A-35	PH	49625084	7.4	99.47	0.03	98	94.1	45
P-24M-36	PI	48144186	7.2	99.62	0.03	98	94	45
P-24M-37	PJ	57278324	8.6	99.68	0.02	98.1	94.2	45
P-24M-38	PK	45091306	6.8	99.63	0.02	98.1	94.2	44
P-24A-39	PL	55621480	8.3	99.6	0.03	97.9	93.9	47
P-24A-40	PM	45813948	6.9	99.69	0.03	97.9	93.7	45
P-24A-41	PN	44276524	6.6	99.51	0.02	98.1	94.3	46
P-72M-42	PO	50705024	7.6	99.54	0.03	97.9	93.6	45
P-72M-43	PP	71842838	10.8	99.66	0.02	98.1	94.1	45
P-72M-44	PQ	45848126	6.9	99.69	0.03	97.8	93.6	45
P-72A-45	PR	41807330	6.3	99.49	0.02	98.2	94.4	44
P-72A-46	PS	49839594	7.5	99.61	0.03	98	94	44
P-72A-47	PT	55214788	8.3	99.46	0.02	98	94.2	46

Fig. 3.15. Principal component analysis of the RNA-Seq data. The PCA was created using the model ~time + plant + treatment + plant*treatment.

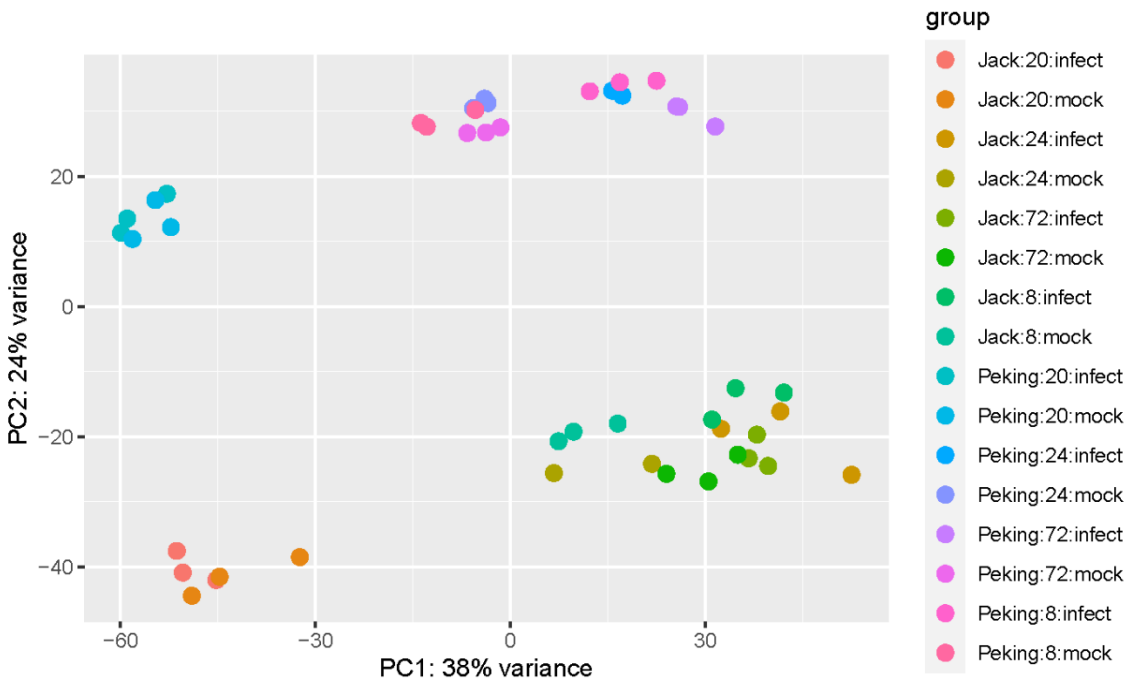


Fig. 3.16. Sample distance matrix of the RNA-Seq data used in the analysis. Euclidean distances were used to cluster samples based on gene expression similarities to create the dendrogram. A darker scale color indicates the samples share higher gene expression similarity.

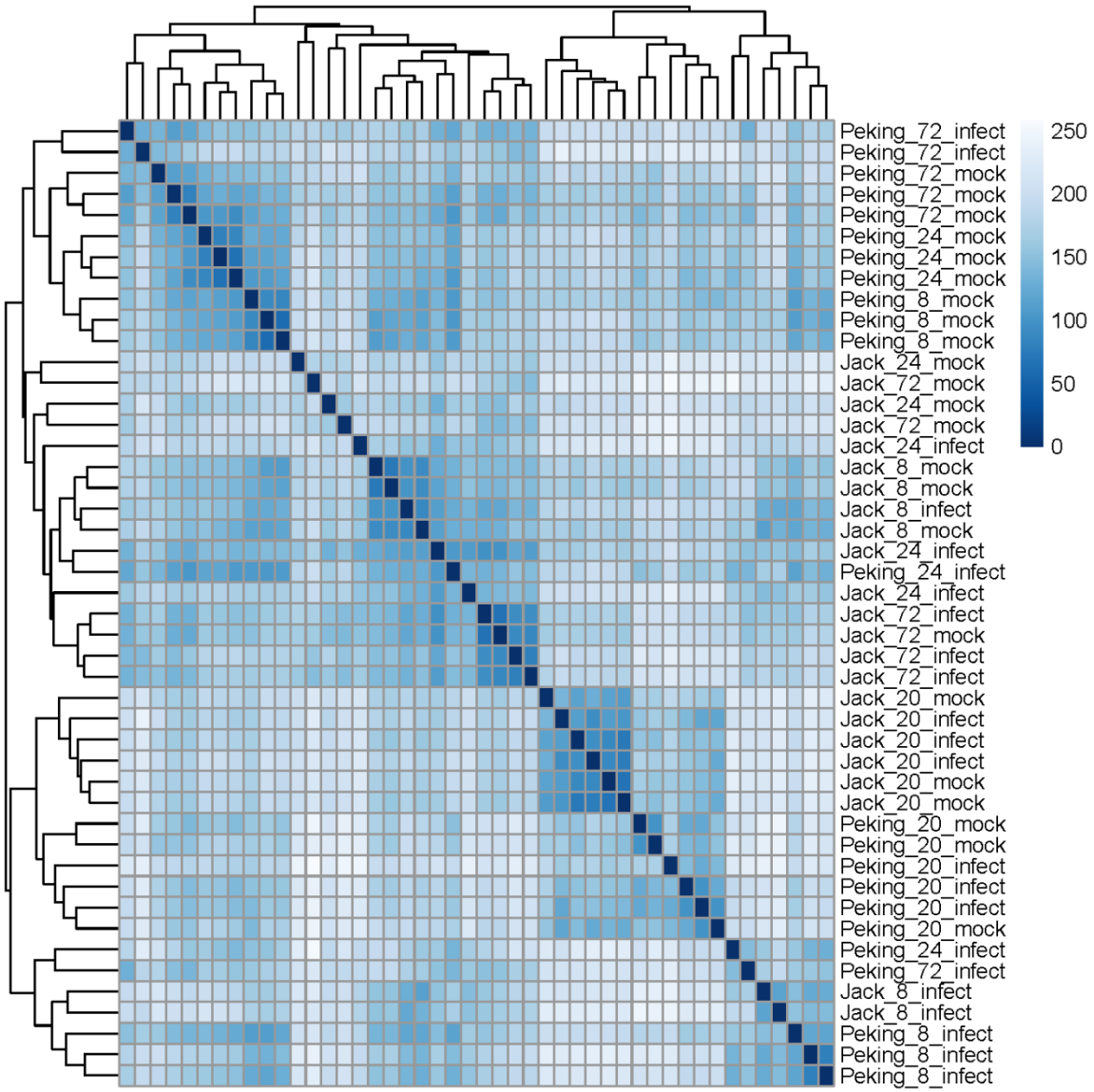


Table 3.8. Significant DEGs among treatment contrasts using a broad grouping method, without timepoints. Default DESeq2 cutoffs were used ($p < 0.1$).

Group	Contrast	Upregulation	Downregulation
Jack-*Agrobacterium*	Jack-Mock	48	115
Jack-*Agrobacterium*	Peking-*Agrobacterium*	1660	1912
Jack-Mock	Peking-Mock	2189	1958
Peking-*Agrobacterium*	Peking-Mock	1576	773

Fig. 3.17. Venn diagram comparing DEGs between Jack-Mock (JA), Jack-*Agrobacterium* (JA), Peking-Mock (PM), and Peking-*Agrobacterium* (PA) across the four timepoints using an adjusted *p*-value < 0.01 for cutoff.

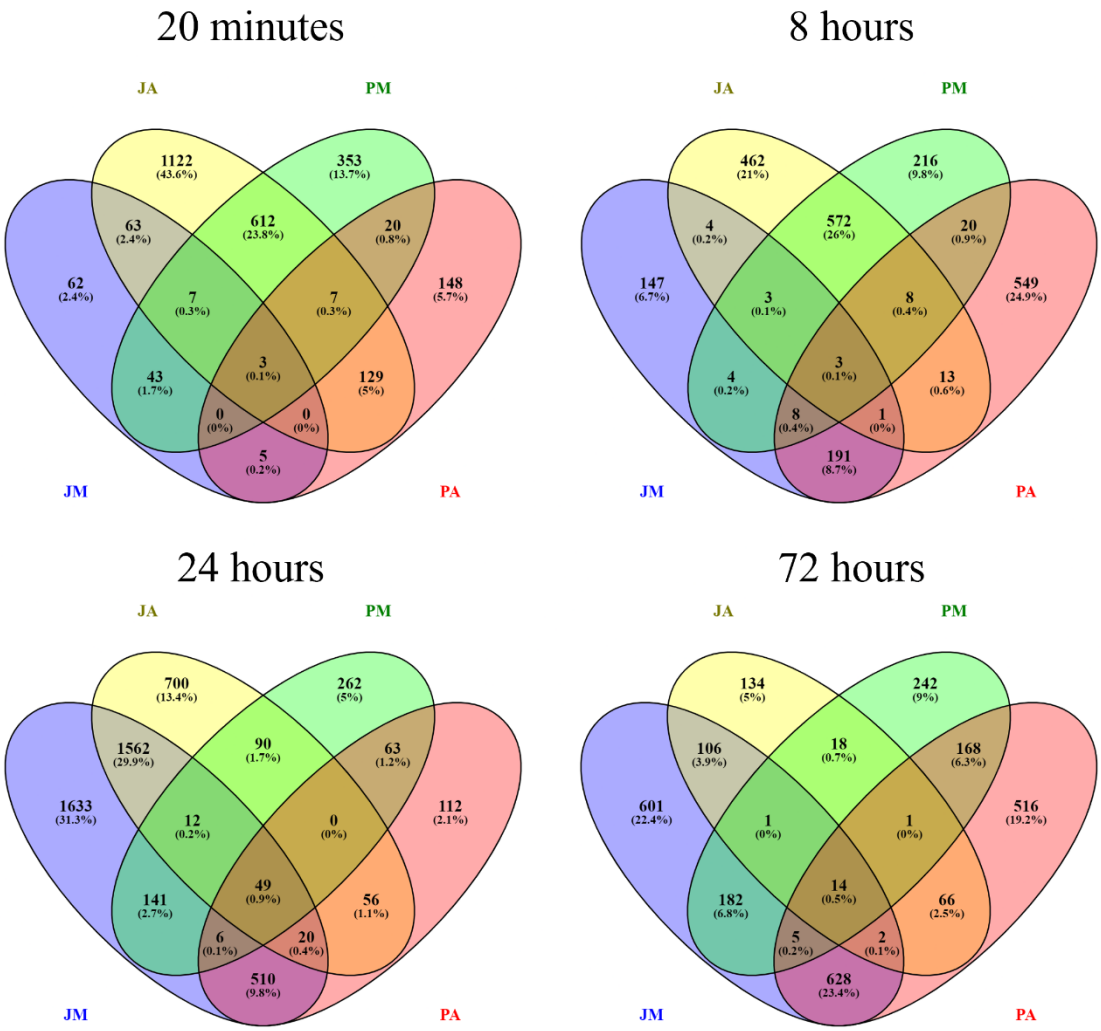


Table 3.9. DEGs within Chr06 QTL region between Jack and Peking 20 minutes after inoculation with *Agrobacterium*.

DEGs in the Chr06 QTL Region

Group	*Contrast*
Jack-*Agrobacterium*	Peking-*Agrobacterium*

20 minutes

Gene Model	baseMean	logFC	Adj P-value	Annotation
Glyma.06G133800	917	-3.5	4E-07	Transcription factor PHOX2/ARIX, contains HOX domain
Glyma.06G143000	31788	-2.2	4E-04	Chalcone-flavanone isomerase family protein
Glyma.06G144000	1394	2.4	7E-03	No annotation or TAIR
Glyma.06G146000	1649	1.0	2E-05	Integral component of membrane
Glyma.06G146800	2042	1.0	4E-04	Cleavage site for pathogenic type III effector avirulence factor Avr
MSTRG.13974	69	-9.9	3E-07	

Table 3.10. DEGs within Chr06 QTL region between Jack and Peking 8 hours after inoculation with *Agrobacterium*.

Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G131900	83	23.0	8E-09	Cytochrome B5 isoform A
Glyma.06G140800	89	-8.7	3E-04	Metallo-hydrolase/oxidoreductase superfamily protein
Glyma.06G144400	8729	-0.7	6E-03	Ribosomal protein S26e family protein
Glyma.06G147200	18	-8.5	9E-03	Membrane-associated kinase regulator
MSTRG.13974	69	-10.3	8E-08	

Table 3.11. DEGs within Chr06 QTL region between Jack and Peking 24 hours after inoculation with *Agrobacterium*.

24 hours				
Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G132200	37	18.4	6E-10	No annotation or TAIR
MSTRG.13176	8	21.5	1E-05	
MSTRG.13974	69	-10.2	1E-06	

Table 3.12. DEGs within Chr06 QTL region between Jack and Peking 72 hours after inoculation with *Agrobacterium*.

Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G131600	1225	-1.0	5E-03	Yippee family putative zinc-binding protein
Glyma.06G131900	83	20.0	1E-06	Cytochrome B5 isoform A
Glyma.06G132200	37	-22.0	3E-19	No annotation or TAIR
Glyma.06G137000	5639	2.1	4E-03	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
MSTRG.13885	16	-20.3	4E-05	
MSTRG.13974	69	-8.4	4E-05	

Table 3.13. DEGs within Chr06 QTL region between mock-inoculated Jack and *Agrobacterium*-inoculated Jack after 8 hours.

Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G133300	18494	-1.7	3.3E-04	Cytochrome B5 isoform E
Glyma.06G136900	2395	-2.1	4.4E-04	Transducin/WD40 repeat-like superfamily protein
Glyma.06G143000	31788	-2.7	2.6E-05	Chalcone-flavanone isomerase family protein
Glyma.06G145300	9148	-2.5	9.3E-03	Peroxidase superfamily protein
Glyma.06G149100	6386	-3.3	3.3E-04	Alginate lyase

Table 3.14. DEGs within Chr06 QTL region between mock-inoculated Jack and *Agrobacterium*-inoculated Jack after 24 hours.

Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G131500	1814.14	1.90	4.98E-05	Dof-type zinc finger DNA-binding family protein
Glyma.06G131700	1531.50	0.90	5.09E-03	Nucleotide binding,nucleic acid binding
Glyma.06G132200	37.03	-18.64	1.95E-10	Integral component of membrane
Glyma.06G133900	38.02	-8.10	3.18E-03	Integral component of membrane
Glyma.06G134700	52.97	-8.20	8.39E-03	Leucine-rich repeat protein kinase family protein
Glyma.06G137600	65.95	-9.56	8.67E-04	No annotation or TAIR
Glyma.06G138300	115.15	-8.22	7.26E-04	Nodulin MtN21 /EamA-like transporter family protein
Glyma.06G138800	54.26	-8.12	4.81E-03	Major facilitator superfamily protein
Glyma.06G139900	3716.82	-1.25	1.02E-06	Transducin/WD40 repeat-like superfamily protein
Glyma.06G140100	36.11	-7.39	8.26E-03	Calcium-dependent lipid-binding (CaLB domain) family protein
Glyma.06G145300	9147.78	-2.62	6.16E-03	Peroxidase superfamily protein
MSTRG.13835	92.93	-7.85	1.10E-03	
MSTRG.13885	15.72	23.40	1.19E-05	
MSTRG.13905	22.75	-14.49	8.86E-04	
MSTRG.13953	47.29	-8.42	9.24E-04	

Table 3.15. DEGs within Chr06 QTL region between mock-inoculated Jack and *Agrobacterium*-inoculated Jack after 72 hours.

Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G132200	37.03	19.55	6.67E-14	Integral component of membrane
MSTRG.13885	15.72	22.99	7.93E-06	

Table 3.16. DEGs within Chr06 QTL region between mock-inoculated Peking and *Agrobacterium*-inoculated Peking after 20 minutes.

Gene Model	baseMean	log2FoldChange	Adjusted P-value
Glyma.06G132500	31656.78	-1.68	8.74E-03

Table 3.17. DEGs within Chr06 QTL region between mock-inoculated Peking and *Agrobacterium*-inoculated Peking after 8 hours.

8 hours				
Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G132300	2228.85	1.33	6.85E-03	MAG2 interacting protein
Glyma.06G133300	18494.31	-1.96	2.10E-05	Cytochrome B5 isoform E
Glyma.06G139200	4959.62	-0.78	5.56E-03	Mitochondrial import inner membrane translocase subunit
Glyma.06G139800	3138.69	1.13	1.65E-03	GPI-anchored adhesin-like protein
Glyma.06G141900	3764.71	-1.14	4.54E-03	Transmembrane protein
Glyma.06G143000	31787.65	-2.87	2.83E-06	Chalcone-flavanone isomerase family protein
Glyma.06G143500	3244.67	1.01	9.89E-03	Ribonuclease II family protein
Glyma.06G145300	9147.78	-4.08	2.70E-07	Peroxidase superfamily protein

Table 3.18. DEGs within Chr06 QTL region between mock-inoculated Peking and *Agrobacterium*-inoculated Peking after 24 hours.

Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G132200	37.03	20.69	2.88E-12	Transmembrane protein
Glyma.06G150300	16710.54	1.78	8.31E-03	Photosystem I reaction center subunit PSI-N

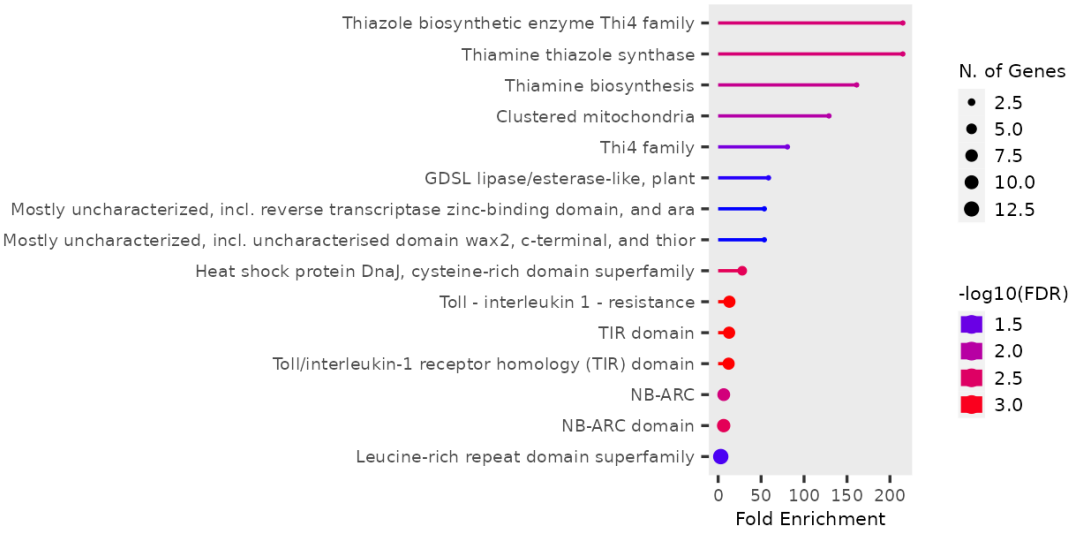
Table 3.19. DEGs within Chr06 QTL region between mock-inoculated Peking and *Agrobacterium*-inoculated Peking after 72 hours.

Gene Model	baseMean	log2FoldChange	Adjusted P-value	
Glyma.06G131900	82.52	23.30	1.35E-08	Cytochrome B5 isoform A
Glyma.06G143000	31787.65	-2.16	1.40E-03	Chalcone-flavanone isomerase family protein
Glyma.06G145300	9147.78	-2.63	4.60E-03	Peroxidase superfamily protein

Table 3.20. Enriched defense-related genes in Jack during infection with *Agrobacterium*. The gene models with an asterisk are not expressed or have little expression in Peking.

Gene Model	Annotation	Chr	Start	Stop
Glyma.01g032900*	TIR-NBS-LRR	1	3455319	3459556
Glyma.01g112200*	TIR-NBS-LRR	1	38052284	38062064
Glyma.14g024400*	TIR-NBS-LRR	14	1746860	1749862
Glyma.14g024500*	TIR-NBS-LRR	14	1751063	1753753
Glyma.16g127900*	TIR-NBS-LRR	16	28004128	28012396
Glyma.16g210600*	TIR-NBS-LRR	16	36968641	36973853
Glyma.16g213700*	TIR-NBS-LRR	16	37160318	37164316
Glyma.06g060300	Heat shock protein DnaJ	6	4573317	4575454
Glyma.08g213400*	Heat shock protein DnaJ	8	17225312	17230465
Glyma.08g239000*	Heat shock protein DnaJ	8	20403057	20403985
Glyma.12g190100	Heat shock protein DnaJ	12	35168889	35171299
Glyma.10g251500	Thiazole/thiamine synthase family	10	47969725	47971337
Glyma.20g142000	Thiazole/thiamine synthase family	20	38055037	38057101
Glyma.14g214100	GDSL lipase/esterase-like plant	14	47899920	47902239
Glyma.16g144900	GDSL lipase/esterase-like plant	16	30549575	30553059
Glyma.08g119200*	NB-ARC domain	8	9192002	9196512
Glyma.08g317400*	NB-ARC domain	8	43698290	43703231
Glyma.13g190000	NB-ARC domain	13	30355157	30359208
Glyma.13g193300*	NB-ARC domain	13	30662869	30665830
Glyma.06g319700	Leucine-rich repeat domain superfamily	6	50856253	50858320
Glyma.08g195400	Leucine-rich repeat domain superfamily	8	15751282	15754886
Glyma.09g107600*	Leucine-rich repeat domain superfamily	9	20356992	20359486
Glyma.12g231800*	Leucine-rich repeat domain superfamily	12	39169878	39182758
Glyma.18g204300	Leucine-rich repeat domain superfamily	18	48725617	48729076
Glyma.18g286800	Leucine-rich repeat domain superfamily	18	56696744	56698723

Fig. 3.18. GO enrichment of upregulated genes in Jack-*Agrobacterium* when compared to Peking-*Agrobacterium* at eight hours post-infection.



CHAPTER 4

CONCLUSIONS

Understanding the interaction of soybean and *Agrobacterium* is a complex process. Before the work presented here, knowledge of which soybean accessions were susceptible to *Agrobacterium* was sparse. While a few studies tested susceptibility across different groups of accessions and support Peking as the most susceptible genotype in soybean, these studies failed to test a diverse panel of germplasm. Of the lines tested in this work, Peking and similarly named accessions were the most susceptible. Other lines tested based on previous work that were confirmed to be susceptible include PI 417139, Thorne, and Jupiter. Later studies on *Agrobacterium*-mediated transformation of soybean would include compounds to dampen the soybean immune system, but up until now, how the soybean immune system responds to *Agrobacterium* was largely ignored. Knowledge of which genes in soybean initiate the immune response and through which pathways this immune system works can lead to higher transformation efficiencies of soybean with *Agrobacterium*,

Agrobacterium is a pathogen that prefers to evade recognition by a host plant, while also transforming DNA into the host to help meet the bacterium's nutritional needs. As most bacteria do not transmit DNA into their host nucleus, *Agrobacterium* in a way, is also very similar to a virus. *Agrobacterium* shares similarities with both pathogens and viruses, making the interacting proteins and host response to *Agrobacterium* harder to elucidate. This interaction is further complicated as *Agrobacterium* shares many

characteristics with the nitrogen-fixing symbionts of soybean, and soybeans have evolved to perceive and recognize Rhizobiaceae members.

In Chapter 2, a potential candidate gene in soybean for recognizing *Agrobacterium* was tested. This gene, GmEFR (*Glycine max* Elongation Factor Receptor), is an ortholog of the arabidopsis EFR gene. In arabidopsis, EFR was found to recognize a bacterial protein, Elongation Factor Thermo-unstable, and initiate plant defenses. In the case of arabidopsis, EFR was capable of recognizing EF-Tu from several bacterial species, including *Agrobacterium*. Through perception of *Agrobacterium* EF-Tu and activation of the immune system, arabidopsis becomes chlorotic and undergoes cell death, thereby limiting transformation. In EFR mutant lines, the plant fails to recognize *Agrobacterium* EF-Tu and higher transformation is observed.

Genome sequencing of the susceptible soybean line revealed that GmEFR was absent in this accession. Based on the arabidopsis literature, GmEFR became the best candidate gene for improving transformation in other soybean accessions. Through traditional breeding methods and CRISPR/cas9 mutagenesis of GmEFR in resistant accessions with subsequent experiments, GmEFR was shown to not behave similar to the arabidopsis EFR. The knockout plants did not become susceptible, and while the use of the deletion marker in the breeding population did improve susceptibility, the cause of this increased susceptibility is not the lack of GmEFR.

While the soybean homolog of EFR does not play a role in defending against *Agrobacterium*, it is possible that soybeans possess a different receptor that does recognize *Agrobacterium* EF-Tu. To test this question, recombinant strains of *Agrobacterium* were made. Instead of the wild-type *Agrobacterium* EF-Tu sequence,

these strains had EF-Tu sequences from different Rhizobiaceae that normally nodulate and are compatible with soybeans. The strains were tested with tobacco and were still functional, but when testing with a resistant soybean genotype, no change in transformation ability was found.

Despite a large amount of work in Chapter 2, very little was learned about the underlying genetics in soybean that modulate susceptibility or resistance to *Agrobacterium*. While the original hypothesis of a single gene was tested, the hypothesis was adjusted to include broader approaches to allow a more in-depth search for regions that underlie soybean's response to *Agrobacterium*.

In Chapter 3, a mapping approach was employed to identify QTL that explain the variation in response to *Agrobacterium*. Through using recombinant inbred lines derived from a susceptible parent, Peking, crossed to a resistant parent, Jack, new lines were created that vary widely in their susceptibility to *Agrobacterium*. This work identified a novel area on chromosome 6 underlying susceptibility within the recombinant inbred population. While this QTL region was not identified by the threshold used in the GWAS panel, a SNP approximately 4 Mb away was significant for gall size (*p = 0.000001I)* in the GWAS population. Several genes have mutations, insertions, and premature stop codons introduced within the strongest peak of the QTL identified on chromosome six.

Another finding in Chapter 3 was that several recombinant lines were more susceptible than the susceptible parent, Peking, indicating there is transgressive segregation. The exact reason for this is unknown, but there are a few possibilities. Earlier work in arabidopsis supported the presence of a receptor, EFR, that recognizes *Agrobacterium* EF-Tu, triggering plant defense responses, thereby reducing

transformation efficiency. After a soybean homolog of this gene was found to be absent in the susceptible Peking genotype, the population was screened to contain the absence of *Glyma.09g216400* found in Peking. Through selection of the GmEFR absence using a marker during the development of this population, a gene near to GmEFR that confers susceptibility was also selected for. While all the lines tested lack the GmEFR gene, several genotypes repeatedly fail to be susceptible to *Agrobacterium,* further supporting that GmEFR is not the gene. Interestingly, the region surrounding GmEFR contains many immunity genes, providing support that this inherited section of the genome in the recombinant inbred lines may harbor a gene involved in susceptibility to *Agrobacterium*. A gene, *Glyma.09g216100,* three gene models upstream of GmEFR, annotated as a leucine-rich repeat family protein, was found to have little to no expression in Peking when compared to Jack when infected with *Agrobacterium*. An adjacent gene, *Glyma.09g216300,* annotated as a nodulin-like protein, also exhibits different splicing behavior in Peking when compared to Jack.

Continuing Chapter 3, an RNA-Seq dataset was generated from mock-inoculated and *Agrobacterium*-inoculated stems from the soybean genotypes Jack and Peking. The data generated is useful not only for determining expression differences between Jack and Peking, but also for examining differences in the identified QTL region in the mapping population. The results support that resistant soybeans undergo a plant defense response to *Agrobacterium*, with initial responses driven by TIR-NBS-LRR proteins, with downstream proteins including peroxidases and heat shock proteins that culminate in a hypersensitive response. The RNA-Seq data identified around twenty significant DEGs within the major QTL on Chr06 that controls susceptibility identified in a biparental Jack

by PxC population. Candidate genes within this region are numerous, but they include a peroxidase, a leucine-rich repeat protein, and isoflavone-related genes. While the soybean ortholog of the arabidopsis EFR that restricts *Agrobacterium* transformation was ruled out using CRISPR/cas9, the selection of the Peking allele of GmEFR resulted in increased susceptibility within a recombinant inbred population, indicating a nearby gene is involved with susceptibility. The RNA-Seq data identified two adjacent candidate genes to GmEFR on Chr09.

The appendix contains a study of diverse germplasm to examine the frequency of *Agrobacterium* susceptibility and to identify markers associated with susceptibility. As only a handful of soybean accessions have been repeatedly tested for susceptibility to *Agrobacterium*, a large panel of accessions were tested, including Jack and Peking. The information learned from this study identified areas underlying susceptibility to *Agrobacterium* in a diverse population that differs from the region identified by genetic mapping. While this study suffered from low statistical power because of the number of genotypes included and given that rare alleles are hard to detect within GWAS studies, the significant markers could be associated with genetic variants that do impact susceptibility to *Agrobacterium.* Many other candidate genes are located near the significant markers identified by the GWAS, especially on chromosomes 8, 10, 13, 17, and 19. The gene models near these markers were analyzed using the RNA-Seq dataset to determine if they play a role in susceptibility to *Agrobacterium,* but these variants may only be present in certain genotypes. The marker on Chr13, ss71561891, is located adjacent to two FASCIN genes, annotated as cellulases or glucan 1,3-betaglucosidases. The markers on Chr19 from 28 mb to 30 mb are neighboring a bucentaur-related protein,

but this region also contains a LRR island, with five different leucine-rich repeat proteins containing NB-ARC domains. These alleles may be distinctly different from those in Peking controlling susceptibility and could lead to creating more transformable soybean varieties when combined with Peking-type susceptibility.

While the RNA-Seq study identified different genes as being the most significantly differentially expressed than the genes within regions identified in earlier chapters, twenty significant genes were still located within the previously identified QTL and nearby associated markers. The results support that Jack and Peking are very different in terms of expression, complicating the finding of differentially expressed genes. Nonetheless, numerous important genes were identified in Jack from eight hours to seventy-two hours post-infection. Furthermore, genes within the major QTL mapped to Chr06, and nearby a previous marker that correlates with susceptibility, were also identified. Before the work presented here, the common belief was that soybean undergoes a hypersensitive response to *Agrobacterium*, but no data supported this claim besides death of the soybean tissue. Using this data, the immune system of soybean is definitively involved with resistance to *Agrobacterium*. The genes and their pathways can now be further studied and manipulated to facilitate greater transformation efficiencies across soybean genotypes.

APPENDIX A

GENOME-WIDE ASSOCIATION STUDY OF SOYBEANS ACCESSIONS TO IDENTIFY GENOMIC REGIONS UNDERLYING RESPONSE TO AGROBACTERIUM INFECTION

Abstract

Susceptibility to *Agrobacterium* enables genetic engineering in plants for applications such as herbicide or pest resistance, CRISPR/cas9 mutagenesis, or for adding beneficial traits. In soybean [*Glycine max* (L.) Merr.], susceptibility to *Agrobacterium* is found in the Peking variety, PI 548402 but the mechanism controlling susceptibility remains unknown. A collection of 106 diverse genotypes were evaluated in greenhouse assays and characterized for susceptibility to *Agrobacterium* through measurements of gall size and gall number. Using 50K single nucleotide polymorphism (SNP) data, a genome-wide association study (GWAS) was conducted to identify SNPs associated with susceptibility to *Agrobacterium*. Across all four replications, results were consistent and susceptible genotypes were identified. The Peking genotype was the most susceptible genotype. A total of 31 SNPs ($P \leq 0.0000001$) that showed a significant association with gall size and gall number for the average across all replications were identified by GWAS. The largest locus was located on Chr19 and spanned by ten markers. The remaining most significant SNPs tagged smaller loci and were located on Chr08, Chr10, and Chr13. Numerous candidate genes were identified that are implicated in host defense to bacteria. Further work is necessary to confirm the associated markers and evaluate their impact on susceptibility to *Agrobacterium*.

Introduction

While progress towards achieving *Agrobacterium*-mediated transformation has progressed for many species in recent years, soybean [*Glycine max* (L.) Merr.] remains difficult to efficiently transform with *Agrobacterium*. Despite many decades of active research, a single genotype, Peking or PI548402, remains as the lone genotype that is highly susceptible to *Agrobacterium* transformation. Additional soybean genotypes with desirable agronomic characteristics but possessing susceptibility to *Agrobacterium* would benefit soybean transformation and breeding.

Investigations into the genetic architecture controlling *Agrobacterium* susceptibility in soybeans have been limited. To date, no QTL have been mapped in soybean for this trait. An earlier study on the inheritance of susceptibility in soybean concluded that selection in early generations should be possible, but that due to the quantitative mode of inheritance and lack of distinct phenotypic classes, susceptibility to *Agrobacterium* is controlled by multiple host genes, with one parental cross supporting a single gene model, and another cross supporting a two gene model (Bailey et al., 1994). Both of these crosses used Peking as parental line.

This study examined one hundred six diverse soybean genotypes for their response to *Agrobacterium*-infection (Fig. Af.1). The genotypes were selected based on a few criteria. Many of the genotypes have been previously tested for their response to *Agrobacterium* and serve as check lines while providing population diversity. Several genotypes are known nodulation mutants to determine if these genes that condition symbiosis may also be controlling *Agrobacterium* susceptibility. The remaining lines in the study were included if Peking was in their pedigree, to increase genetic diversity at

poorly represented sites, or to incorporate accessions that have contributed to many varieties in North American soybean programs. Of particular interest for this study was the evaluation of soybean accessions with the name Peking. Within the USDA collection, there exists seven varieties named Peking or were once known also as Peking.

Results

Identification of susceptible lines

The soybean varieties named “Peking” outperformed most other genotypes tested (Fig. Af.2). Phenotypic and statistical analysis of gall size, gall number, and similar grouping were conducted (Table At.2, Table At.3, Table At.4, Table At.5, Table At.6). The most common Peking accession, PI 548402, produced the largest total gall sizes per plant, with an average size of 0.9593 cm^2. This accession also produced the highest number of galls per plant, with an average of eight galls. The next best accession was also Peking, PI 548402S, which descended from the original Peking variety, PI 17852B, but was maintained in Stoneville, Mississippi (Bernard et al., 1987). These two genotypes have been commonly tested for susceptibility to *Agrobacterium* but only represent two out of the many Peking accessions. Along with the Peking derived accession PI 548205, these three genotypes were significantly different and the most susceptible lines when analyzed using the Kruskal-Wallis test with Scott-Knott rankings.

Five other Peking varieties were also tested in this experiment: PI 297543, PI 438496A, PI 438496B, PI 438496C, and PI 438497. The PI 438498, named Sable, was also considered to be a Peking variety and tested (Morse and Cartter, 1927). The accession PI 297543 presents with a round, yellow seed coat, unlike the typical black,

flatter seed typical of Peking varieties and was not susceptible to *Agrobacterium*. The remaining Peking varieties did prove to be susceptible to *Agrobacterium*, with PI 438497 and PI 438496B outperforming the others.

Several accessions with better agronomic characteristics performed similar to the Peking varieties. As we were interested in lines that are commonly used in *Agrobacterium*-mediated transformation of soybean, the PI 564718, or Thorne, was included due to it being commonly used in the cot node system (Widholm et al., 2010). Thorne showed a high degree of susceptibility but tended to produce fewer galls overall when compared to Peking, exhibiting characteristics that coincide with its use in *Agrobacterium*-mediated transformation. Two other improved varieties, Maple Arrow (PI 548593) and Hardee (PI 548666) showed susceptibility to *Agrobacterium* and can be considered as a donor for incorporating *Agrobacterium* susceptibility into other soybean lines.

Two varieties collected in Japan show promising susceptibility. While PI 417138 initially showed excellent tumorigenic response in the first replication, the line often maintained a viny habit and performed poorly in the blocks within the greenhouse. The other line, PI 417139, fared better under these growing conditions and performed similar to Peking.

In previous experiments, two landrace accessions, Cloud (PI 548316) and Sooty (PI 548415), often exhibited varying degrees of susceptibility. During these experiments, Sooty performed similar to Peking varieties while Cloud showed minimal susceptibility. Sooty was selected from Cloud in 1907 because of a segregating difference in seed coat bloom but is more susceptible than its parent.

Using a mixed linear model, heritability for both traits was found to be quite high. For gall size, heritability (h^2) was 75.1%, and for gall number, heritability (h^2) was 59.6%.

Susceptibility-associated SNPs

This study was attempted with 120 soybean lines, but only 106 lines germinated reliably across replications. This decrease in accessions hurts the statistics of this study, leaving it underpowered, but this small number of missing accessions would not have resulted in enough power. For future iterations of this work, at least two hundred lines should be considered in the future.

For this study, a generalized linear model (GLM) was used because of the model's ability to allow for response variables that have error distributions that may not be normal. For helping reduce noise due to kinship and non-informative pseudo-SNPs, a total of 100,000 permutations was run. These results were then compared to a mixed-linear model (MLM), as these methods are useful for controlling population structure and relatedness, to help draw final conclusions.

The GLM was tested with variations of five to seven principal components, and while results were similar across the models, seven principal components were chosen. The GLM was conducted with 100,000 permutations and identified significant markers for gall size (Fig. Af.3, Table At.7). For gall size, markers above a significance threshold of 5 or $-\log_{10}(0.00000001)$ (with $p = 0.00000001$) were identified. Single markers were identified on chromosomes 8, 10, and 13. For chromosome 19, two clusters of markers were identified (Fig Af.4, Table At.3). As the GLM model was underpowered and

exhibited relatedness among the lines, the model included 100,000 permutations. Based on permutation *p* values, only two SNPs were above the significance threshold of 1.3 or *p* = 0.05 (Fig. Af.5)

Compared to the gall size trait, gall number resulted in fewer significant markers. A single SNP on chromosome 19 was above the threshold of significance (Fig. Af.6, Table At.7). Interestingly, this SNP was not identified with gall size, but is within 100 kb of SNPs identified for gall size.

To compare results with the GLM, a mixed linear model was used for gall size and gall number. With a threshold of significance of 5 or *p* = 0.0001, a single SNP on chromosome 13 was identified for gall size and gall number (Fig.Af.8, Fig. Af.9, Fig. Af.10). Many other SNPs were significant with *p* = 0.001 (Table At.9, Table At.10).

Investigation of associated regions

The genes surrounding the associated markers identified for gall size and gall number were identified and included for further analysis if the annotations support a possible role in plant immunity and host defense. As whole genome sequence data for all the lines included in this study with high susceptibility are not available, further work examining these regions in more depth may identify changes in the areas surrounding the identified markers.

The region on Chr04 did not have many genes that are implicated in plant immunity. A single gene, *Glyma.04g107900,* is a universal stress protein that could be elevated during infection and regulate other immunity processes.

The regions identified on Chr07 have two strong candidates, a calmodulin binding transcriptional activator and a MYB protein, MYB 111. Similarly, Chr10 regions were found to have additional MYB proteins, thioredoxins, AUX/IAA proteins, and leucine rich repeat receptor-like protein kinases. MYB proteins are known transcription factors for being involved in a variety of processes, especially plant immunity. The thioredoxin proteins have been implicated in nodulation processes, and given *Agrobacterium*'s relation to the symbionts, may be of significance.

The soybean Chr13 contains many leucine-rich repeat receptor-like protein kinases. In total seven were identified near associated markers. Other DNA modification proteins, sulfate transporters, calcium-related proteins, and a MAPKK were near markers identified in the GWAS.

The strongest associations were found across Chr19. The specific regions of Chr19 associated with gall size and number have numerous genes implicated in plant immunity. These include typical leucine-rich repeat receptor-like protein kinases, numerous MYBs, and several transcription factors. As this region was the most significantly associated marker with the phenotype, expression and sequencing data will be important to determining their role in controlling susceptibility.

Evaluation of the study

While this study found significant results for SNPs underlying susceptibility and identified numerous other genotypes besides Peking that exhibit susceptibility, a few modifications should be considered for future work.

The current data presented a challenge in conducting the analysis. Most importantly, the study is underpowered and should include more accessions with broad diversity to increase power. As the data violated the assumption of normality, standard data transformations were attempted such as logarithm (*log(x)*, reciprocal (*1/x)*, square root ($x^{0.5}$) and $(x + 0.5)^{0.5}$ but failed to correct for normality when analyzed using the Shapiro-Wilk test. A Box-Cox transformation was attempted, and the lambda calculated, but distribution remained non-normal. It is important to note that a normal distribution was not expected due to the nature of the trait.

To draw conclusions, two approaches were used. While not typical due to violating the assumption of normality, an ANOVA was performed to identify genotype groupings (Table At.3). Because of the normality, the Kruskal-Wallis test was used to compare the plant accessions. The Kruskal-Wallis test found a significant difference ($p = 3.872e-5$) indicating that gall size differs among the 106 genotypes. To allow for ranking of the accessions, the Scott-Knott test (significance level = 0.5) was used (Table At.4).

Overall, susceptibility to *Agrobacterium* is a rare trait in soybeans and this research supports this conclusion. While little was known about Peking varieties besides PI 548402, it is evident that most other varieties with the name Peking are also highly susceptible. Besides Peking varieties, several other improved varieties and landraces from Japan were identified as also being susceptible. This research identifies other soybean genotypes that could be of use in identifying the genetic basis underlying susceptibility to *Agrobacterium* not found in the Peking varieties.

As other studies have found Peking to be the most susceptible genotype to *Agrobacterium*, this research continues to support that claim (Bailey et al., 1994; Byrne et

al., 1987; Owens and Cress, 1985). Susceptible lines identified in other studies include PI 417138, Jupiter, Biloxi, and Manchu. While PI 417138 did not perform well across replicates in the present study, the line did perform well in a single replicate. A related line, PI 417139, did perform well across replicates. Jupiter was moderately susceptible across all three replicates, similar to previously reported results. Seeds of Biloxi and Manchu were not tested in this panel. An important finding of this work is that many Peking accessions are highly susceptible, and the genotype commonly used in cotyledonary node transformation, Thorne, is also highly susceptible.

Discussion

Although rare, soybean accessions other than the varieties of the name Peking exhibit susceptibility to *Agrobacterium*. While originally believed to be controlled by one or two major genes, the results here point to the trait being of a more polygenic nature. It is possible that the non-Peking germplasm found to be susceptible in this study possess different mechanisms controlling susceptibility than those in Peking. Genetic mapping in a Peking-derived population identified a major QTL on Chr06, but the nearest significant marker to the QTL identified in this population is located 5 Mb away. Given that susceptibility to *Agrobacterium* can be selected for, crossing lines with the best Peking-type susceptibility to lines identified in this GWAS may allow for greater susceptibility. It is interesting to note that gall size and susceptibility may be unrelated, and the process determining gall size may rely more upon DNA integration, cell wall crosslinking, or other similar molecular differences between soybean accessions. However, current

literature supports the involvement of the soybean immune system to control compatibility with *Agrobacterium*.

Materials and Methods

Plant genotypes

The following genotypes were used in this study. Peking, the *Agrobacterium*-susceptible variety, derived from seeds obtained from Roger Boerma at the University of Georgia in 1988. This variety is most similar to the United States Department of Agriculture (USDA) Germplasm Resources Information Network (GRIN) plant accession PI 548402. Jack, PI 540556, is an embryogenic variety obtained from Wayne Parrott, University of Georgia. The remaining lines were requested from the USDA GRIN database and listed in Table At.1.

Planting

Eight seeds from each accession were planted in Fafard 3B potting mix (Sun Gro Horticulture Inc., Vancouver, Canada) across two cells in a 32-cell tray. The trays were placed in a growth chamber at 25º C and grown under a 23-h photoperiod (to prevent premature flowering induction) with a light intensity of 100 µmol m^{-2} s^{-1} for seven days with humidity domes. After nine days (VC growth stage), each genotype was thinned down to three plants and inoculated with *Agrobacterium* inoculum scraped from a yeast-mannitol plate. After infection, plants were returned to the growth chamber for two days, maintaining the 23-h photoperiod but the temperature was lowered to 23º C, and the light intensity increased to 200 µmol m^{-2} s^{-1}. The inoculated seedlings were then transplanted

into 15.24 cm nursery pots containing a soil mixture of 3B, sterile farm soil, and sand. The plants were grown in the greenhouse with supplemental lighting to maintain 18/6-hour light/dark and keeping the temperature constant between 23º C - 27º C. After 60 days in the greenhouse, the plants were phenotyped.

Statistical Design

The experimental design was a randomized complete block design (RCBD) with four replications containing a single plant of each of the 106 genotypes. Each block had border rows of alternating Jack (PI 540566) and Peking (PI 548402).

Inoculation

Agrobacterium tumefaciens strain A281 (C58 containing the armed pTiBo542) was obtained from Eugene Nester and streaked from a glycerol stock on yeast-mannitol medium containing 50 mg mL^{-1} rifampicin (Chem-Impex International, Wood Dale, Illinois) and grown for three days in a 28° C incubator to produce the inoculum used for infection (Jin et al., 1993; Vincent, 1970). Sterile spatulas were used to spread *Agrobacterium* inoculum onto Giant Push Pins (Officemate, Edison, NJ). Each hypocotyl was punctured 2 cm above the soil completely through the stem. After puncturing, additional *Agrobacterium* inoculum was rubbed into each side of the puncture wound using a sterile spatula.

Phenotyping

Plants were evaluated after sixty days in the greenhouse. Most genotypes were approaching R1 growth stage, with some variation between genotypes based on maturity group. Images of each plant's entry or exit wounds, along with a ruler for calibration, were taken with a camera. The images were loaded into ImageJ, and the ruler used to calculate distance per pixel using the Measure function. Each gall was outlined by hand and size recorded (Schneider et al., 2012). For each plant, the number of galls was visually counted and recorded. Measurements of gall area and gall number were collected and analyzed using Microsoft Excel and RStudio (Racine, 2012).

Genome-wide association analysis

GWAS was performed with Tassel 5.0 using both GLM and MLM models (Bradbury et al., 2007) . A total of seven principal components were used in the models. Soybean 50K SNP data was downloaded from SoyBase and formatted for Tassel with Excel or RStudio (Grant et al., 2009; Song et al., 2013). The Newick tree was constructed in Tassel using 50K SNP data.

References

1. Bailey, M. A., Boerma, H. R., & Parrott, W. A. (1994). Inheritance of *Agrobacterium tumefaciens*-induced tumorigenesis of soybean. *Crop Science, 34*(2), 514-519.
2. Bernard, R., Juvik, G., & Nelson, R. (1987). USDA Soybean germplasm collection inventory. Vol. 1. *Intl. Agric. Publ. INTSOY Ser, 30*.
3. Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. J. B. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *23*(19), 2633-2635.
4. Byrne, M. C., McDonnell, R. E., Wright, M. S., Carnes, M. G. J. P. C., Tissue, & Culture, O. (1987). Strain and cultivar specificity in the Agrobacterium-soybean interaction. *8*(1), 3-15.
5. Grant, D., Nelson, R. T., Cannon, S. B., & Shoemaker, R. C. J. N. a. r. (2009). SoyBase, the USDA-ARS soybean genetics and genomics database. *38*(suppl_1), D843-D846.
6. Jin, S., Song, Y. n., Pan, S. Q., & Nester, E. W. (1993). Characterization of a virG mutation that confers constitutive virulence gene expression in Agrobacterium. *Molecular Microbiology, 7*(4), 555-562.
7. Morse, W., & Cartter, J. (1927). Soy beans; culture and varieties. US Dept. Agr. *Farmers Bull, 1520*.

8. Owens, L. D., & Cress, D. E. (1985). Genotypic variability of soybean response to Agrobacterium strains harboring the Ti or Ri plasmids. *Plant Physiology, 77*(1), 87-94.

9. Racine, J. S. (2012). RStudio: a platform-independent IDE for R and Sweave. In: JSTOR.

10. Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods, 9*(7), 671-675.

11. Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., & Cregan, P. B. J. P. o. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *8*(1), e54985.

12. Vincent, J. M. (1970). A manual for the practical study of the root-nodule bacteria. *A manual for the practical study of the root-nodule bacteria.*

13. Widholm, J. M., Finer, J. J., Vodkin, L. O., Trick, H. N., LaFayette, P., Li, J., & Parrott, W. (2010). Soybean. *Genetic Modification of Plants: Agriculture, Horticulture and Forestry*, 473-498.

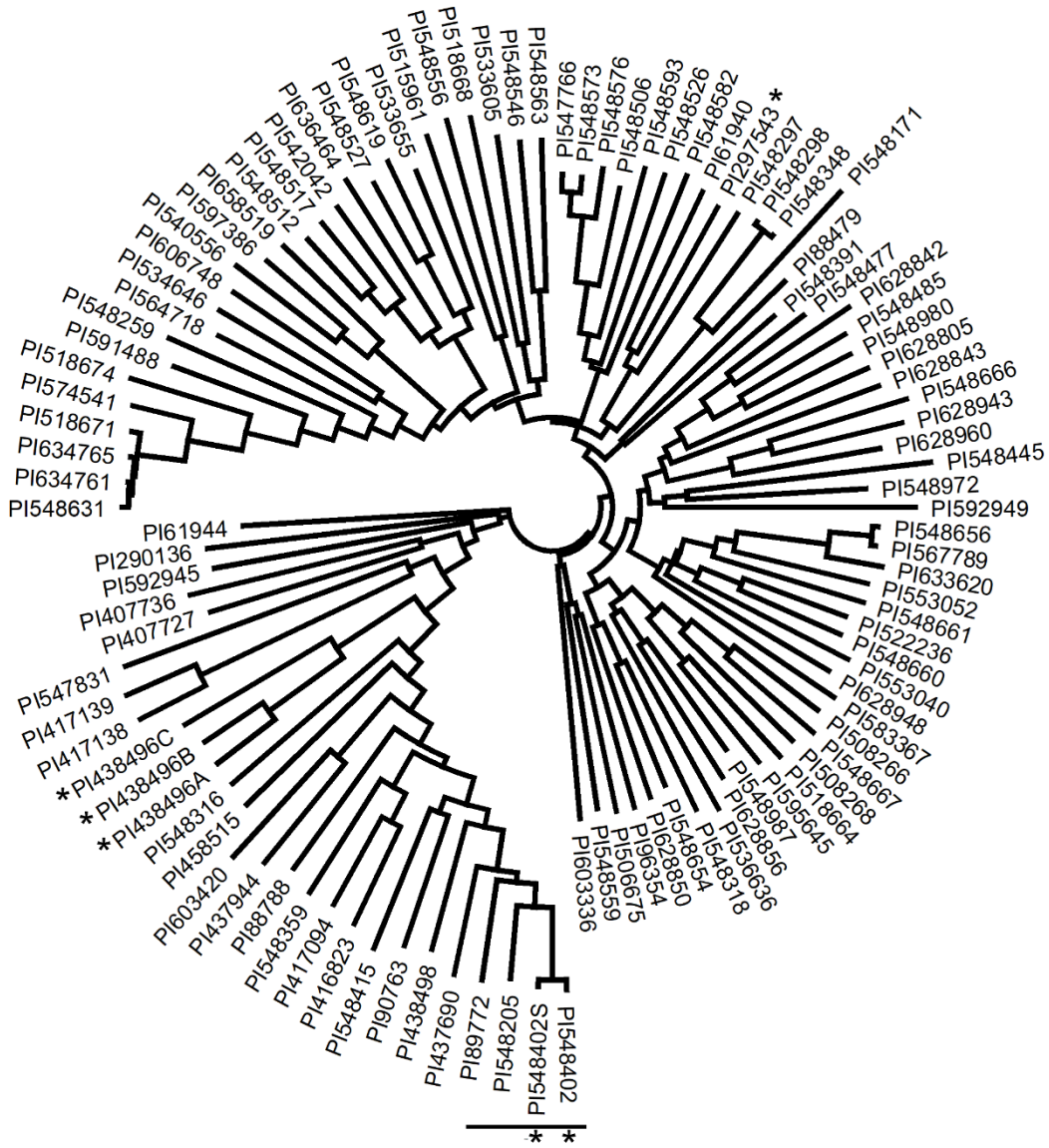
Table At.1. Plant accessions and names used in the genome-wide association study.

Accession	Name
PI 61940	
PI 61944	
PI 88479	Kungchuling Improved No. 77
PI 88788	5913
PI 89772	
PI 90763	7570
PI 96354	649
PI 290136	Noir 1
PI 297543	Peking
PI 407727	
PI 407736	
PI 416823	Cha Masshokutou
PI 417094	Kuro Masshokutou (Kou 205)
PI 417138	Masshokutou
PI 417139	Masshokutu (Kou 503)
PI 437679	Nan-Cou
PI 437690	Pin-din-guan
PI 437944	VIR 569
PI 438496A	Peking
PI 438496B	Peking
PI 438496C	Peking
PI 438497	Peking
PI 438498	Sable
PI 458515	Tie Zhugan
PI 506675	Fukuyutaka
PI 518668	TN4-86
PI 518671	Williams 82
PI 518674	Fayette
PI 522236	Thomas
PI 533605	Cordell
PI 533655	Burlison
PI 536636	Ripley
PI 540556	Jack
PI 542042	Kato
PI 547766	L64-2139
PI 547831	L83-0215
PI 548171	T134
PI 548205	T221

PI 548259	Century 84
PI 548297	A.K.
PI 548298	AK (Harrow)
PI 548316	Cloud
PI 548359	Kingwa
PI 548391	Mukden
PI 548402	Peking
PI 548402S	Peking
PI 548415	Sooty
PI 548445	CNS
PI 548477	Ogden
PI 548485	Roanoke
PI 548506	Amsoy
PI 548512	Century
PI 548517	Bonus
PI 548526	Hardin
PI 548527	Calland
PI 548546	Custer
PI 548556	Elf
PI 548559	Emerald
PI 548563	Franklin
PI 548573	Harosoy
PI 548576	Harwood
PI 548582	McCall
PI 548593	Maple Arrow
PI 548619	Sparks
PI 548631	Williams
PI 548654	Hill
PI 548655	Forrest
PI 548656	Lee
PI 548660	Bragg
PI 548661	Semmes
PI 548666	Hardee
PI 548667	Essex
PI 548972	Jupiter
PI 548974	Bedford
PI 548976	Dyer
PI 548980	Hood
PI 548987	Dare
PI 548988	Pickett
PI 553039	Davis

PI 553040	Jeff
PI 553052	Narow
PI 564718	Thorne
PI 567789	Bossier
PI 574541	LN89-5612
PI 591488	L91-8060
PI 592945	Zhong Huang No.1
PI 592949	Yu Dou No.8
PI 595645	Benning
PI 597386	Dwight
PI 603336	Qing Pi Si Li Huang
PI 603420	Hei dou
PI 606748	Rend
PI 628805	Savana
PI 628842	IAC-1
PI 628843	IAC-4
PI 628850	IAC-100
PI 628856	IAC-5
PI 628943	IAC-2
PI 628948	IAC-17
PI 628960	UFV-9
PI 633620	NC55
PI 634761	T370
PI 634765	T374
PI 636464	LDX01-1-65
PI 656647	DS4-SCN05
PI 658519	LD00-2817F

Fig. Af.1. Newick tree displaying phylogeny of the 106 soybean accessions used in the GWAS study. The Tassel program generated the plot using the 50K SNP data of the accessions used in the study. Accessions named Peking are marked with an asterisk.



.

Fig. Af.2. Average Gall Size of the accessions used in the GWAS. Due to the number of accessions presented on the x-axis, half of the labels are omitted. Peking is on the right, shown in green. Jack, PI 540556, is in blue. For checking other accessions not labeled, refer to the tables.

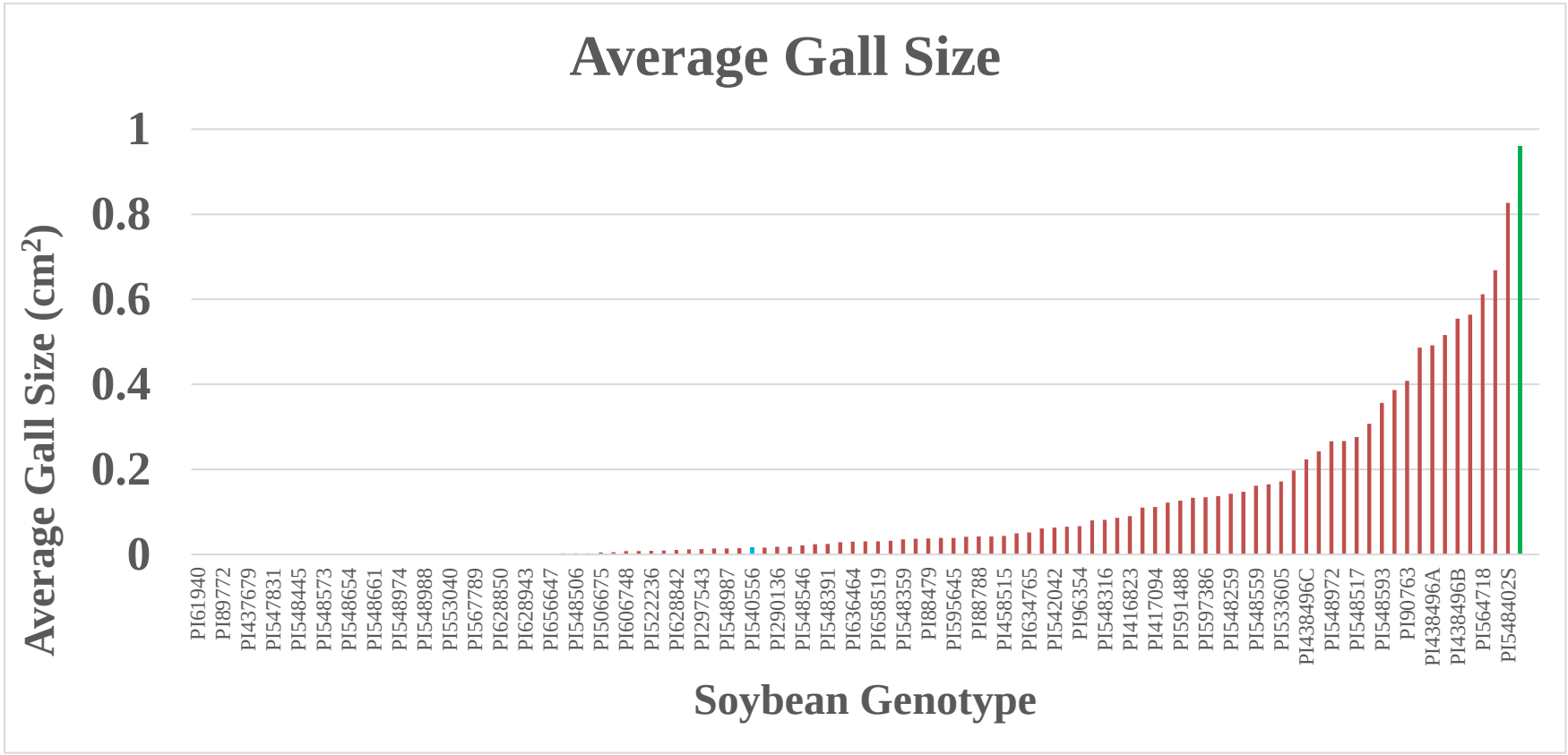


Table At.2. ANOVA Statistics of accessions used in the GWAS for average gall size.

ANOVA

	Df	*SS*	*MS*	*F*	*P-value*
Line	105	13.05	0.12426	2.254	7.17E-08
Residuals	270	14.89	0.05514		
Total	375	27.94			

Table At.3. Phenotype analysis of gall size. Four replications were used for analysis, with some replications missing data for certain genotypes.

Trait	Effect	df	SS	MS	F	*P* value	Broad sense heritability (H^2)
Gall Size	Genotype	149	36.06	0.24201	2.2876	<.0001	0.3124524
	Replication	2	0.604	0.30183	2.853	0.05943	
	Residuals	267	28.247	0.10579			
	Total	418	64.911	0.64963			

Table At.4. Phenotype analysis of gall number. Four replications were used for analysis, with some replications missing data for certain genotypes.

Trait	Effect	df	SS	MS	F	*P* value	Broad sense heritability (H^2)
Gall Number	Genotype	149	722.97	4.8522	2.0458	<.0001	0.2499739
	Replication	2	46.11	23.0553	9.7207	<.0001	
	Residuals	292	28.247	0.10579			
	Total	443	797.327	28.01329			

Table At.5. ANOVA group means of accessions used in the GWAS for average gall size.

Genotype	Mean Gall Size	Group
PI 548402	0.959	a
PI 548402S	0.827	ab
PI 548205	0.668	abc
PI 564718	0.612	abcd
PI 438497	0.564	bcde
PI 438496B	0.555	bcde
PI 548415	0.516	bcdef
PI 438496A	0.492	bcdefg
PI 417139	0.487	bcdefgh
PI 90763	0.408	cdefghi
PI 548666	0.387	cdefghij
PI 548593	0.357	cdefghijk
PI 437690	0.307	defghijkl
PI 548517	0.276	defghijkl
PI 437944	0.267	defghijkl
PI 548972	0.266	defghijkl
PI 438498	0.243	efghijkl
PI 438496C	0.223	efghijkl
PI 592945	0.198	fghijkl
PI 533605	0.172	ghijkl
PI 518674	0.165	ghijkl
PI 548559	0.161	ghijkl
PI 548976	0.148	ghijkl
PI 548259	0.143	hijkl
PI 548171	0.138	hijkl
PI 597386	0.135	hijkl
PI 592949	0.133	ijkl
PI 591488	0.127	ijkl
PI 634761	0.122	ijkl
PI 417094	0.112	ijkl
PI 548619	0.111	ijkl
PI 416823	0.090	ijkl
PI 548631	0.086	ijkl
PI 548316	0.082	ijkl
PI 548563	0.080	jkl
PI 96354	0.067	jkl
PI 574541	0.065	jkl
PI 542042	0.063	jkl

PI 603336	0.061	jkl
PI 634765	0.052	kl
PI 548297	0.050	kl
PI 458515	0.044	kl
PI 603420	0.043	kl
PI 88788	0.042	kl
PI 548512	0.042	kl
PI 533655	0.039	kl
PI 595645	0.039	kl
PI 88479	0.038	kl
PI 548582	0.037	kl
PI 548359	0.035	kl
PI 417138	0.032	kl
PI 658519	0.031	kl
PI 548485	0.031	kl
PI 636464	0.030	kl
PI 548556	0.029	l
PI 548391	0.025	l
PI 628960	0.024	l
PI 548546	0.022	l
PI 290136	0.018	l
PI 548527	0.018	l
PI 407736	0.016	l
PI 540556	0.016	l
PI 628948	0.015	l
PI 548987	0.014	l
PI 518671	0.014	l
PI 297543	0.013	l
PI 536636	0.012	l
PI 628842	0.011	l
PI 518668	0.009	l
PI 522236	0.009	l
PI 606748	0.008	l
PI 628843	0.008	l
PI 548526	0.006	l
PI 506675	0.005	l
PI 548506	0.003	l
PI 548656	0.003	l
PI 548660	0.003	l
PI 407727	0	l
PI 437679	0	l

PI 547766	0	I
PI 547831	0	I
PI 548298	0	I
PI 548445	0	I
PI 548477	0	I
PI 548573	0	I
PI 548576	0	I
PI 548654	0	I
PI 548655	0	I
PI 548661	0	I
PI 548667	0	I
PI 548974	0	I
PI 548980	0	I
PI 548988	0	I
PI 553039	0	I
PI 553040	0	I
PI 553052	0	I
PI 567789	0	I
PI 61940	0	I
PI 61944	0	I
PI 628805	0	I
PI 628850	0	I
PI 628856	0	I
PI 628943	0	I
PI 633620	0	I
PI 656647	0	I
PI 89772	0	I

Table At.6. Scott-Knott groupings of accessions used in the GWAS for gall size.

Genotype	Maturity Group	Origin	Seed Coat Color	Mean Gall Size	Group
PI 548402	IV	China	Black	0.96	a
PI 548402S	V	China	Black	0.83	a
PI 548205		United States	Black	0.67	a
PI 564718	III	United States	Yellow	0.61	b
PI 438497	III	United States	Black	0.56	b
PI 438496B	III	United States	Black	0.55	b
PI 548415	IV	China	Black	0.52	b
PI 417139	I	Japan	Brown	0.49	b
PI 438496A	II	United States	Black	0.49	b
PI 90763	IV	China	Black	0.41	b
PI 548666	VIII	United States	Yellow	0.39	b
PI 548593	00	Canada	Yellow	0.36	b
PI 437690	III	China	Black	0.31	b
PI 548517	IV	United States	Yellow	0.28	c
PI 437944	II	China	Black	0.27	c
PI 548972		United States	Yellow	0.27	c
PI 438498	IV	United States	Black	0.24	c
PI 438496C	IV	United States	Black	0.22	c
PI 592945	III	China	Yellow	0.2	c
PI 518674	III	United States	Yellow	0.17	c
PI 533605	V	United States	Light green	0.17	c
PI 548559	IV	United States	Green	0.16	c
PI 548976	V	United States	Light green	0.15	c
PI 548171		United States	Yellow	0.14	c

PI 548259		United States	Gray Green	0.14	c
PI 591488		United States	Yellow	0.13	c
PI 592949	IV	China	Yellow	0.13	c
PI 597386	II	United States	Yellow	0.13	c
PI 634761		United States	Yellow	0.12	c
PI 417094	III	China	Black	0.11	c
PI 548619	IV	United States	Yellow	0.11	c
PI 416823	III	China	Green/Brown	0.09	c
PI 548631	III	United States	Yellow	0.09	c
PI 548316	III	China	Black	0.08	c
PI 548563	IV	United States	Yellow	0.08	c
PI 96354	VI	North Korea	Yellow	0.07	c
PI 542042	I	United States	Yellow	0.06	c
PI 574541	III	United States	Yellow	0.06	c
PI 603336	II	China	Green	0.06	c
PI 548297	IV	China	Yellow	0.05	c
PI 634765		United States	Yellow	0.05	c
PI 458515	IV	China	Black	0.04	c
PI 533655	II	United States	Yellow	0.04	c
PI 548359	IV	China	Black	0.04	c
PI 548512	II	United States	Yellow	0.04	c
PI 548582	00	United States	Yellow	0.04	c
PI 595645	VII	United States	Yellow	0.04	c
PI 603420	II	China	Black	0.04	c
PI 88479	II	China	Yellow	0.04	c
PI 88788	III	China	Black	0.04	c

PI 417138	II	Japan	Black	0.03	c
PI 548485	VII	China	Yellow	0.03	c
PI 548556	III	United States	Yellow	0.03	c
PI 636464		United States	Yellow	0.03	c
PI 658519		United States		0.03	c
PI 290136	0	France	Black	0.02	c
PI 407736	IV	China	Yellow	0.02	c
PI 540556	II	United States	Yellow	0.02	c
PI 548391	II	China	Yellow	0.02	c
PI 548527	III	United States	Yellow	0.02	c
PI 548546	IV	United States	Yellow	0.02	c
PI 628960		Brazil		0.02	c
PI 297543	II	China	Yellow	0.01	c
PI 518668	IV	United States	Yellow	0.01	c
PI 518671	III	United States	Yellow	0.01	c
PI 522236	VII	United States	Yellow	0.01	c
PI 536636	IV	United States	Yellow	0.01	c
PI 548526	I	United States	Yellow	0.01	c
PI 548987	V	United States	Yellow	0.01	c
PI 606748	IV	United States	Yellow	0.01	c
PI 628842	VIII	Brazil	Yellow	0.01	c
PI 628843	VIII	Brazil	Yellow	0.01	c
PI 628948	VIII	Brazil	Yellow	0.01	c
PI 407727	IV	China	Yellow	0	c
PI 437679	IV	China	Black	0	c
PI 506675	VI	Japan	Yellow	0	c

PI 547766		United States	Yellow	0	c
PI 547831		United States	Black	0	c
PI 548298	III	China	Yellow	0	c
PI 548445	VII	China	Yellow	0	c
PI 548477	VI	United States	Light green	0	c
PI 548506	II	United States	Yellow	0	c
PI 548573	II	Canada	Yellow	0	c
PI 548576	II	Canada	Yellow	0	c
PI 548654	V	United States	Yellow	0	c
PI 548655	V	United States	Light green	0	c
PI 548656	VI	United States	Yellow	0	c
PI 548660	VII	United States	Yellow	0	c
PI 548661	VII	United States	Yellow	0	c
PI 548667	V	United States	Yellow	0	c
PI 548974	V	United States	Yellow	0	c
PI 548980	VI	United States	Yellow	0	c
PI 548988	VI	United States	Yellow	0	c
PI 553039	VI	United States	Yellow	0	c
PI 553040	VI	United States	Yellow	0	c
PI 553052	V	United States	Yellow	0	c
PI 567789	VIII	United States	Light green	0	c
PI 61940	III	China	Yellow	0	c
PI 61944	IV	CN	Yellow	0	c
PI 628805	VIII	Brazil	Yellow	0	c
PI 628850		Brazil	Yellow	0	c
PI 628856	V	Brazil	Yellow	0	c

PI 628943		Brazil		0	c
PI 633620		US	Black	0	c
PI 656647		United States	Yellow	0	c
PI 89772	IV	China	Black	0	c

Figure Af.3. Manhattan plot of average gall size using a generalized linear model with permutations. The red line represents the threshold of significance (*p = 0.00000001*), the blue line represents a general level of significance (*p = 0.00001*).

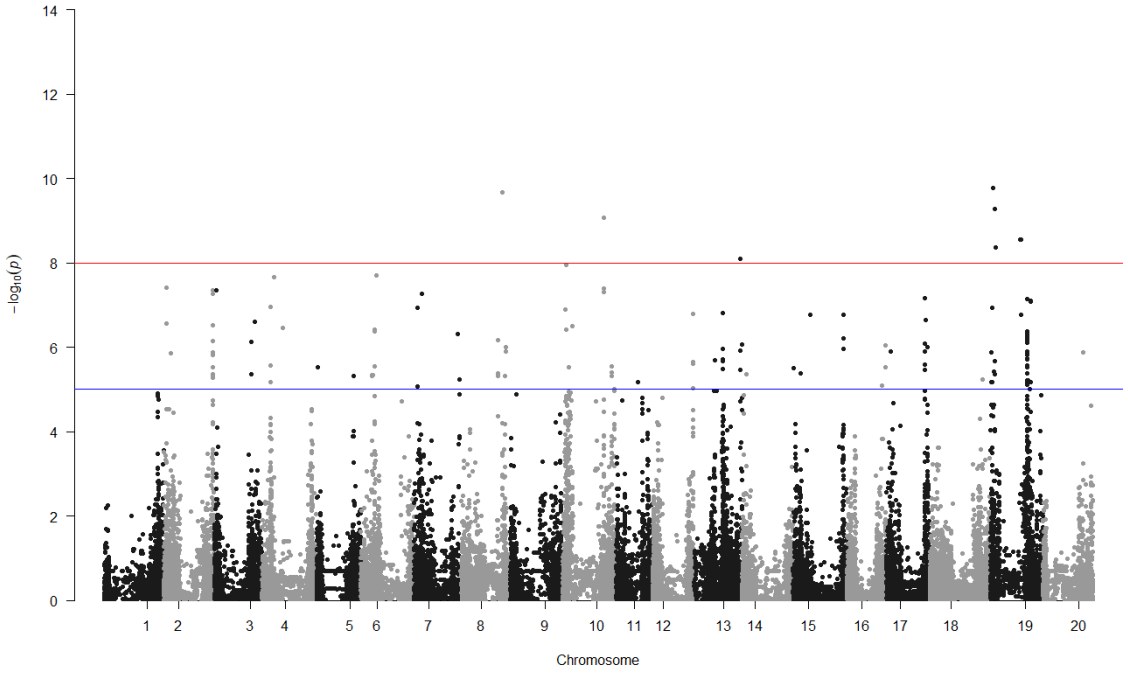


Figure Af.4. Manhattan plot average gall size using a generalized linear model with permutations of chromosome 19. The red line represents the threshold of significance (*p = 0.00000001*), the blue line represents a general level of significance (*p = 0.00001*).

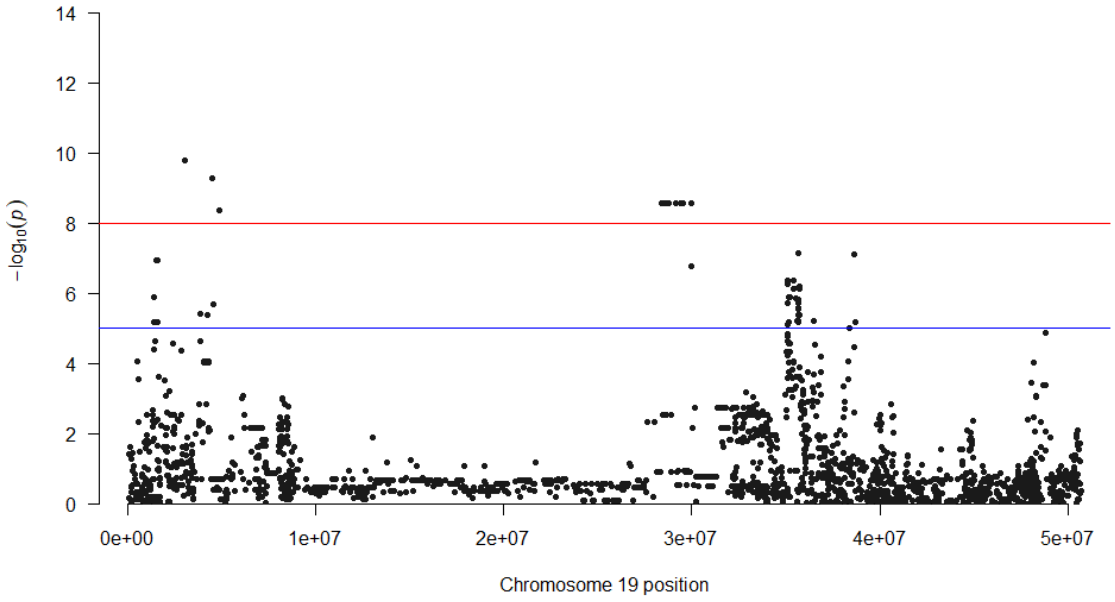


Fig. Af.5. Manhattan plot of average gall size using a generalized linear model with -log 10 permutation test (100,000) p-values for each SNP. The red line indicates the threshold of significance with a permutation test p-value of 0.05.

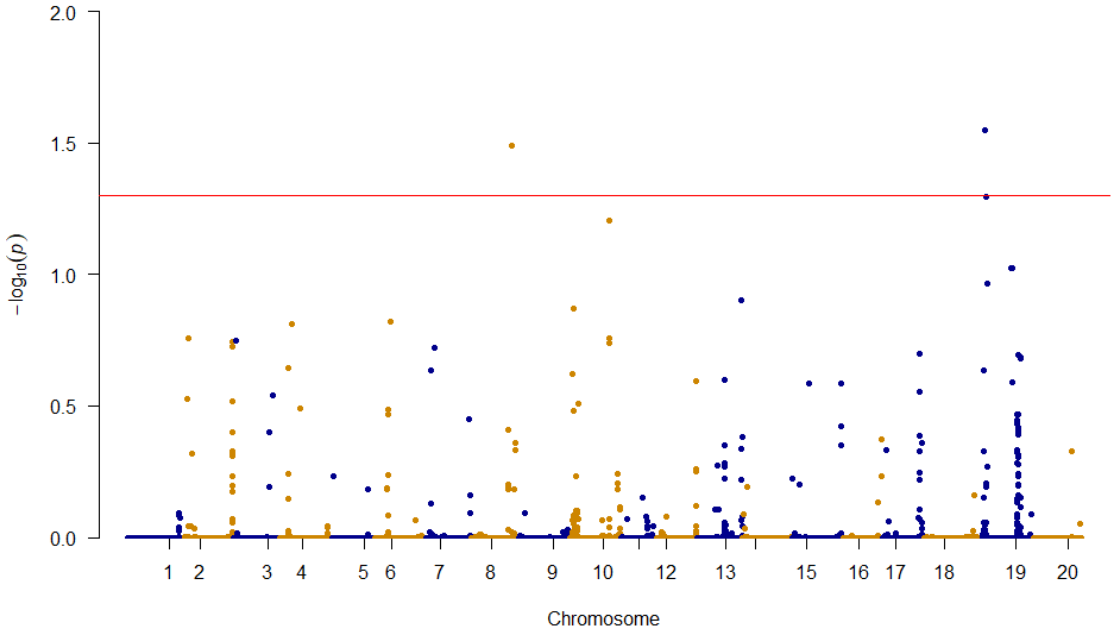


Fig. Af.6. Manhattan plot of average gall number using a generalized linear model with permutation values. The red line represents the threshold of significance (*p = 0.00000001*), the blue line represents a general level of significance (*p = 0.00001*).

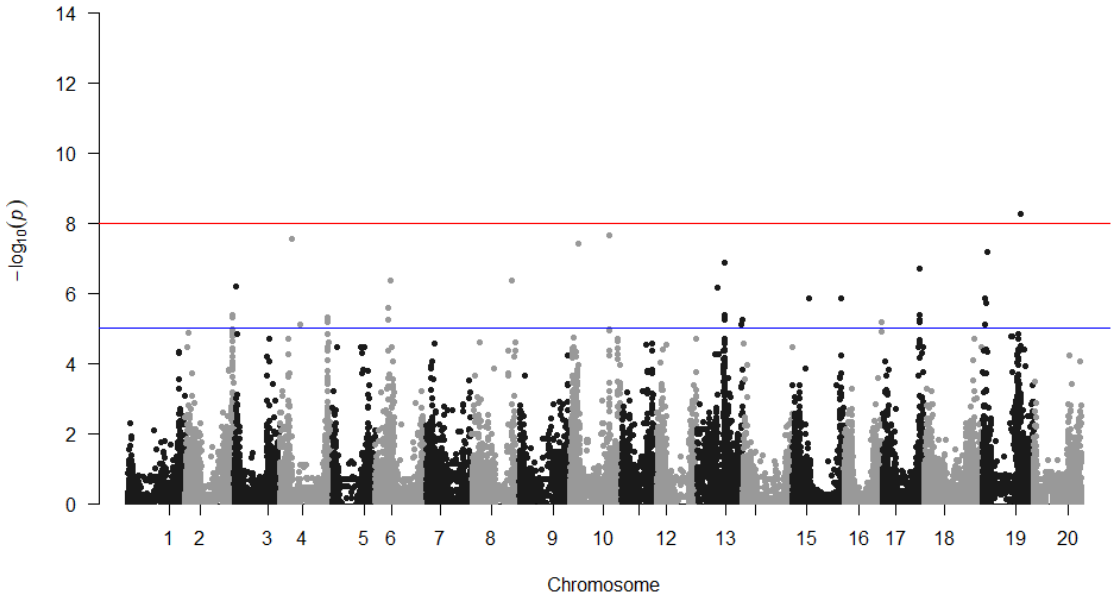


Fig. Af.7. Manhattan plot of average gall number using a generalized linear model with -log 10 permutation test (100,000) p-values for each SNP. The red line indicates the threshold of significance with a permutation test p-value of 0.05.

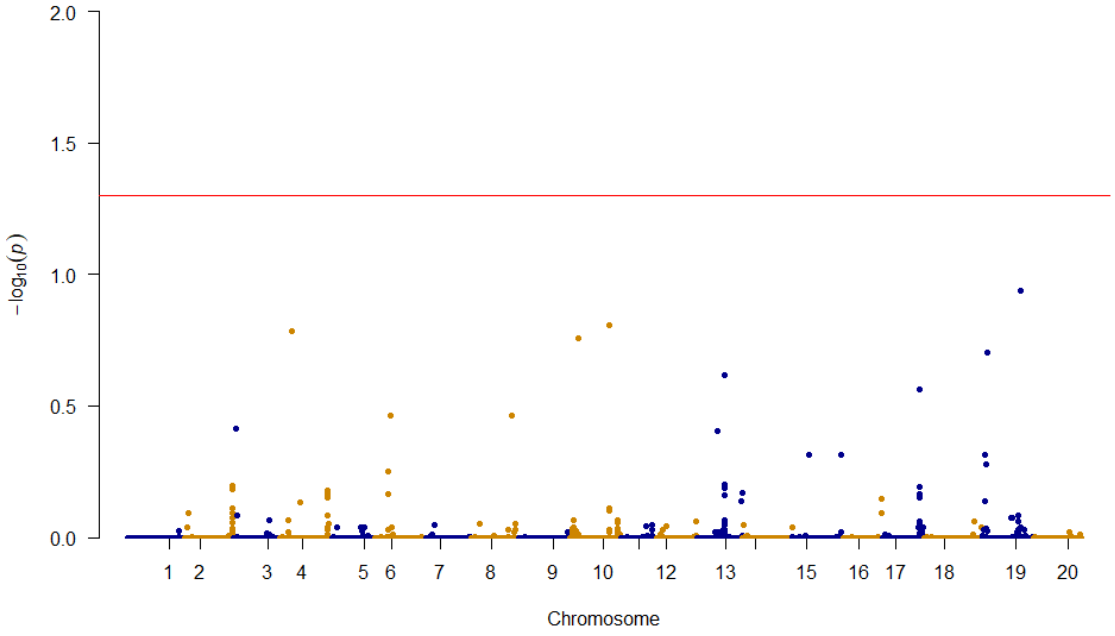


Fig. Af.8. Manhattan plot of average gall size using a mixed linear model. The red line represents the threshold of significance (*p = 0.0001*), the blue line represents a general level of significance (*p = 0.01*).

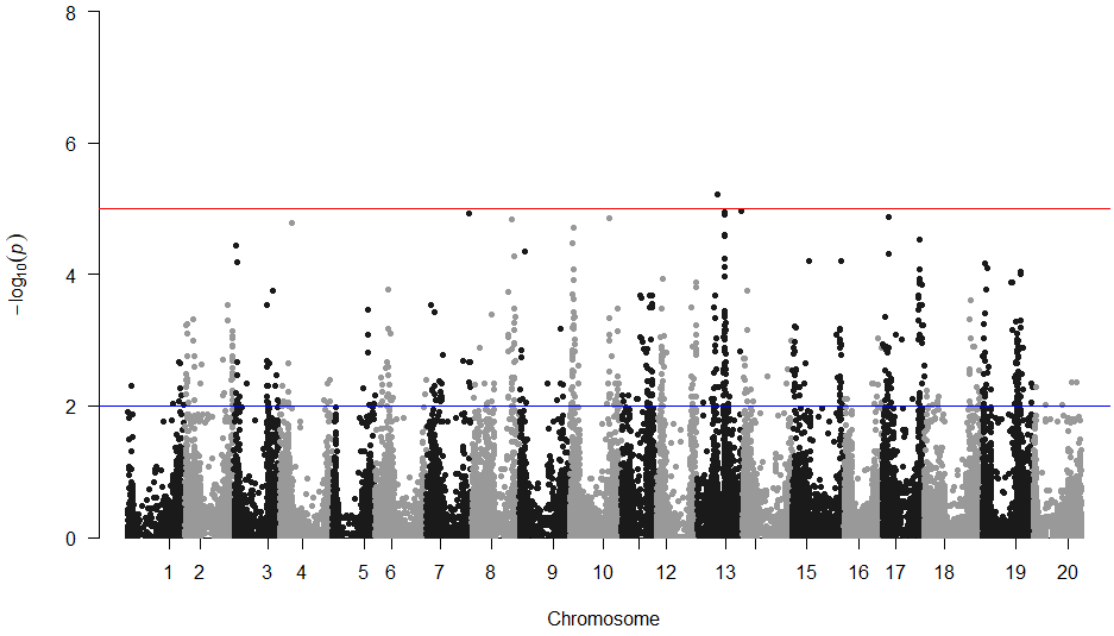


Fig. Af.9. Manhattan plot average gall size using a mixed linear model of chromosome 13. The red line represents the threshold of significance (*p = 0.0001*), the blue line represents a general level of significance (*p = 0.01*).

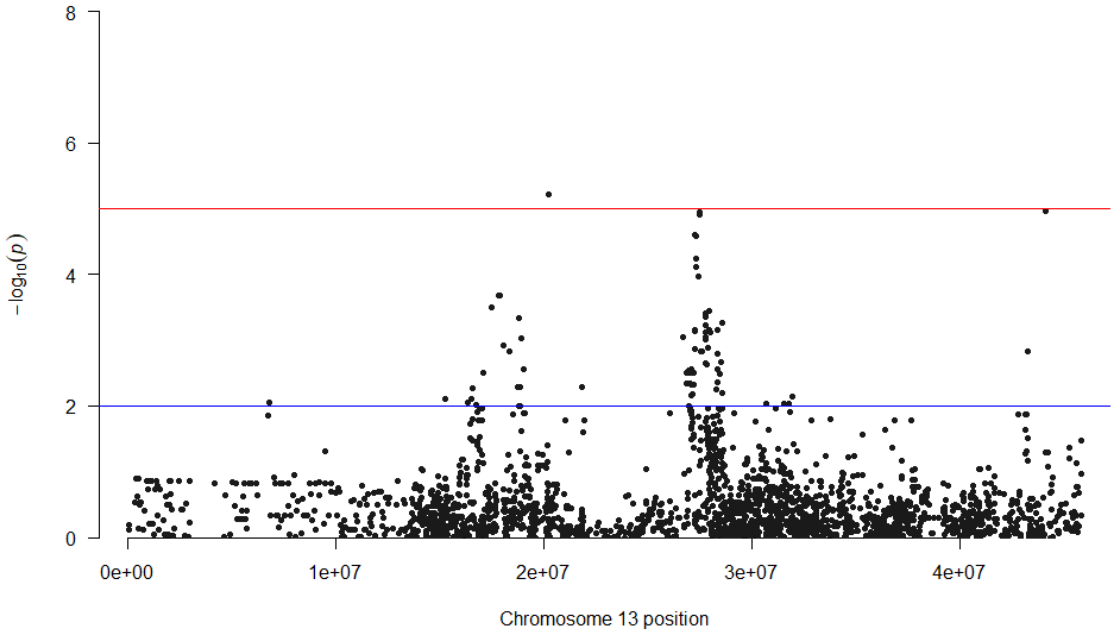


Fig. Af.10. Manhattan plot of average gall number using a mixed linear model. The red line represents the threshold of significance (*p = 0.0001*), the blue line represents a general level of significance (*p = 0.01*).

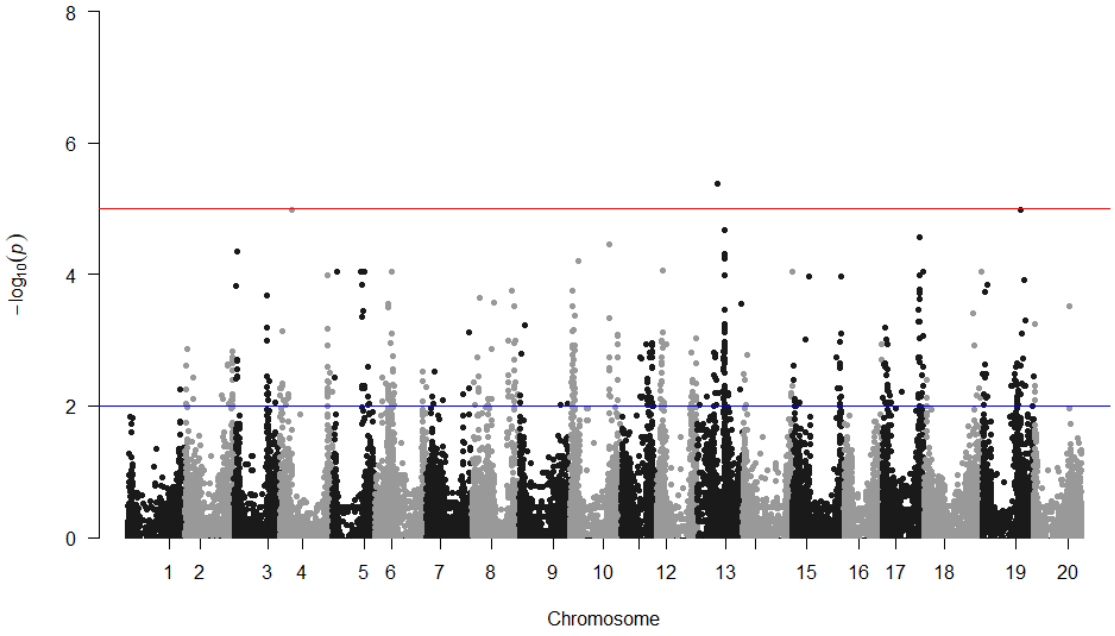


Table At.7. Significant markers identified for average gall size using a generalized linear model with permutations.

Average Gall Size

Marker	*Chr*	*Position (bp)*	*F Value*	*P Value*	*Permutation P Value*
ss715633667	19	3027614	52.16475	1.66E-10	0.02837
ss715601728	8	40516663	51.41076	2.11E-10	0.03231
ss715635337	19	4460496	48.60204	5.28E-10	0.05077
ss715606806	10	39608897	47.13863	8.58E-10	0.06225
ss715633612	19	28364090	43.6497	2.79E-09	0.09487
ss715633614	19	28562399	43.6497	2.79E-09	0.09487
ss715633618	19	28784237	43.6497	2.79E-09	0.09487
ss715633629	19	29142283	43.6497	2.79E-09	0.09487
ss715633640	19	29430536	43.6497	2.79E-09	0.09487
ss715633647	19	29541876	43.6497	2.79E-09	0.09487
ss715633660	19	29977321	43.6497	2.79E-09	0.09487
ss715635871	19	4857303	42.49361	4.30E-09	0.10784
ss715616515	13	44102683	23.31657	7.80E-09	0.1246

Table At.8. Significant markers identified for average gall number using a generalized linear model with permutations.

Average Gall Number

Marker	*Chr*	*Position (bp)*	*F Value*	*P Value*	*Permutation P Value*
ss715634741	19	38654143	23.77668	5.55E-09	0.11539

Table At.9. Significant markers identified for average gall size using a mixed linear model.

Average Gall Size

Marker	*Chr*	*Position (bp)*	*F Value*	*P Value*
ss715613891	13	20226704	1.40E+01	6.0136E-06
ss715616515	13	44102683	13.10984	1.06E-05
ss715614457	13	27466801	21.73436	1.11E-05
ss715598090	7	42230209	21.52655	1.19E-05
ss715614455	13	27443516	21.49536	1.21E-05
ss715628077	17	5980729	21.23533	1.35E-05
ss715606806	10	39608897	21.12858	1.41E-05
ss715601728	8	40516663	21.09951	1.43E-05
ss715584735	4	11448308	20.74752	1.66E-05
ss715606238	10	3253282	20.38979	1.95E-05
ss715614423	13	27227443	19.75943	2.52E-05
ss715614431	13	27251455	19.75943	2.52E-05
ss715614442	13	27308262	19.75414	2.56E-05
ss715627113	17	36714072	19.36239	2.99E-05
ss715627129	17	36739648	19.36239	2.99E-05
ss715627132	17	36744319	19.36239	2.99E-05

Table At.10. Significant markers identified for average gall number using a mixed linear model.

Average Gall Number

Marker	*Chr*	*Position (bp)*	*F Value*	*P Value*
ss715613891	13	20226704	14.48872	4.09E-06
ss715634741	19	38654143	13.11784	1.03E-05
ss715584735	4	11448308	21.85171	1.04E-05
ss715614457	13	27466801	20.18879	2.12E-05
ss715627113	17	36714072	19.59543	2.71E-05
ss715627129	17	36739648	19.59543	2.71E-05
ss715627132	17	36744319	19.59543	2.71E-05

Table At.11. Soybean candidate genes located nearby associated markers for gall size and number.

Gene	Annotation
Glyma.04g107900	Universal stress protein family
Glyma.07g242000	Calmodulin binding transcriptional activator
Glyma.07g242600	MYB-like DNA-binding protein (MYB111)
Glyma.10g037000	MYB-like DNA-binding protein (MYB14)
Glyma.10g161200	Transcription factor GT-2 related
Glyma.10g161600	Thioredoxin-like protein
Glyma.10g161900	Calcium binding protein
Glyma.10g162000	NAD dependent epimerase
Glyma.10g162100	CAMP-response element binding protein-related
Glyma.10g162200	mttA/Hcf106 family
Glyma.10g162300	HLH DNA-binding
Glyma.10g162400	AUX/IAA family
Glyma.10g162600	HEVEIN-like protein
Glyma.10g162700	HEVEIN-like protein
Glyma.10g163200	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g087200	Sulfate transporter
Glyma.13g087400	FASCIN Cellulase
Glyma.13g087500	FASCIN Cellulase
Glyma.13g087800	Phosphoglycerate mutase family protein
Glyma.13g088400	TPR repeat containing protein
Glyma.13g089600	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g090000	Serine/threonine protein kinase RIO
Glyma.13g156400	DNA Replication Licensing factor
Glyma.13g156500	Helicase-related protein
Glyma.13g154900	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g159600	EF-hand Calcium-binding domain protein
Glyma.13g159000	AUX/IAA transcriptional regular family protein
Glyma.13g158900	Beta-1,4-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase
Glyma.13g158800	WUS-interacting protein 2
Glyma.13g158700	Zinc finger domain-containing protein
Glyma.13g159100	RabGAP/TBC domain containing protein
Glyma.13g349600	Beta-galactosidase related
Glyma.13g349800	MAPKK
Glyma.13g350000	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g351100	Phospholipase D

Glyma.13g351500	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g352700	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g352800	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g352900	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g353000	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.13g353100	Zing finger FYVE domain protein
Glyma.13g353400	Phosphofructokinase
Glyma.17g076100	Chitinase
Glyma.17g076200	Plant calmodulin-binding protein-related
Glyma.17g076600	Synaptosomal associated protein
Glyma.17g217500	Chromatin remodeling factor
Glyma.17g217700	Autoinhibited Cas2+ ATPase
Glyma.17g217800	
Glyma.17g217900	
Glyma.17g218000	ARM repeat superfamily protein
Glyma.19g023500	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.19g023700	Receptor-like kinase in flowers 1
Glyma.19g024100	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.19g024200	Rare lipoprotein A (RlpA)-like double-psi beta-barrel
Glyma.19g024700	MYB-like DNA-binding protein MYB (MYB16)
Glyma.19g024800	Protein of unknown function (DUF1997)
Glyma.19g024900	Apoptosis-inducing factor 2
Glyma.19g025000	MYB-like DNA-binding protein MYB (MYB133)
Glyma.19g026100	Kinase associated domain 1
Glyma.19g033100	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.19g034000	Oxidoreductase
Glyma.19g034800	Phopholipase D
Glyma.19g035800	Glucosyl/Glucuronosyl transferases
Glyma.19g035900	BED zinc finger hAT family dimerisation domain
Glyma.19g036000	Glucosyl/Glucuronosyl transferases
Glyma.19g036100	Glucosyl/Glucuronosyl transferases
Glyma.19g079500	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.19g081200	Leucine-Rich Repeat Receptor-like Protein Kinase
Glyma.19g126800	bZIP transcription factor (BZIP17)
Glyma.19g127000	MYB-like DNA-binding protein
Glyma.19g217200	Cyclin family protein
Glyma.19g127700	Tubulin beta
Glyma.19g217800	Transcription elongation factor family protein