Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevejos

Michelle da Fonseca Santos

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba
2012

Michelle da Fonseca Santos
Engenheira Agrônoma

Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevejos

Orientador:
Prof. Dr**. JOSÉ BALDIN PINHEIRO**

Tese apresentada para obtenção do título de Doutor em Ciências. Área de concentração: Genética e Melhoramento de Plantas

Piracicaba
2012

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - ESALQ/USP

Santos, Michelle da Fonseca
Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevejos / Michelle da Fonseca Santos. - - Piracicaba, 2012.
119 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2012.

1. Expressão gênica 2. Correlação genética e ambiental 3. Herdabilidade 4. Mapeamento genético 5. Percevejo 6. Resistência genética vegetal 7. Soja I. Título

CDD 633.34
S237m

Dedico aos meus pais, **Everaldo e Neidinha**,

e ao meu irmão, **Vilker**, pelo amor incondicional

e apoio em todas as etapas de minha vida.

AGRADECIMENTOS

À Deus pelo dom da vida e pela oportunidade de alcançar mais esse objetivo.

Ao Prof. Dr. José Baldin Pinheiro, pela orientação, ensinamentos, confiança e amizade. O meu muito obrigada por tudo durante esses 5 anos de convivência.

Ao CNPq pelas bolsas de estudo concedidas para o Doutorado Direto e também para a formação complementar como aluna de Doutorado Sanduíche na Universidade de Illinois.

À FAPESP pelo apoio financeiro para a realização do projeto.

Ao Dr. Steven J. Clough por me receber em seu laboratório e contribuir com seus ensinamentos para a minha formação científica.

Ao Dr. Brian W. Diers e Dr. Ki-Seung Kim pela importante contribuição ao meu trabalho.

Aos meus pais, Everaldo e Neidinha, e meu irmão Vilker, por sempre me apoiarem e me dedicarem todo amor.

A todos do Laboratório de Diversidade Genética e Melhoramento por tantos excelentes momentos que dividimos durante esses anos de pós-graduação e pela ajuda em muitas etapas de concretização do meu projeto. Eu agradeço em especial a Fátima, Milene, Carlos, Fernanda, Amanda e Pedro pela amizade sincera e que foi tão essencial em tantos momentos. À Dra. Maria Imaculada Zucchi pela co-orientação e amizade.

Aos professores e técnicos do Departamento de Genética pelos ensinamentos e imenso suporte, em especial aos funcionários Marcio Araújo Silva, Sr. Irineu do Amaral, Sr. Natálio Vieira, Cláudio Roberto Segatelli e Sr. José Monteiro.

Aos meus orientadores e co-orientadores de graduação Arie Fitzgerald, Marcelo Mendonça, Renata Mann, Alessandra Boari e Genésio Tâmara por orientarem os meus primeiros passos na carreira acadêmica e científica.

À Sanzio por todo o carinho e incentivo que não me deixaram desanimar diante das dificuldades de fim de curso.

A todos os meus familiares e amigos que mesmo distantes estão sempre torcendo por mim, em especial aos meus avós Lizete e Joãozito; meus tios Fábio, Paulo, Maria, Dilma, Wilton e Rivaldo; aos primos Andray, Anderson, Carol e Bianca; e a essas pessoas muito especiais: Ana Cláudia, Ângela Oliveira, Anne Vasconcenlos, Bernarda Calla, Ísis Dantas, Lia Silvia, Lucimara Koga, Marcell Edward, Maria de Lurdes e Renata Océa,.

SUMMARY

RESUMO

Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevejos

O grupo de percevejos que mais frequentemente causa perdas econômicas à soja no Brasil é composto pelas espécies: *Nezara viridula, Piezodorus guildinii* e *Euchistus heros.* Assim, os objetivos desta pesquisa foram avaliar parâmetros genéticos e correlações entre as diferentes características de desenvolvimento e produção, mapear QTL associados à resistência da soja aos percevejos e determinar respostas de expressão gênica associadas à alimentação do inseto. Uma população $F_{2:3}$ foi desenvolvida através do cruzamento de IAC-100 (resistente) e CD-215 (suscetível) e avaliada em campo experimental. As características agronômicas avaliadas foram: número de dias para o florescimento (NDF), altura da planta no florescimento (APF), número de dias para a maturidade (NDM), altura da planta na maturidade (APM), acamamento (AC), valor agronômico (VA) e produtividade de grãos (PG). As características de resistência a percevejos avaliadas foram: período de enchimento de grãos (PEG), retenção foliar (RF), índice percentual de danos nas vagens (IPDV), número de vagens por planta (NVP), número de sementes (NS), peso de sementes manchadas (PSM), peso de sementes boas (PSB), e peso de cem sementes (PCS). Para se ter um total de 96 amostras, os dois pais foram genotipados juntamente com as 12 mais resistentes e 12 mais suscetíveis plantas F_2 para as características PEG, PCS e PSB, e as 11 mais resistentes e 11 mais suscetíveis para a característica RF. Para determinar o tempo de resposta de expressão gênica nas vagens à alimentação de percevejos, um estudo de microarranjo foi realizado com CD-215 avaliando a expressão relativa 5.5, 21, 24 e 41 horas após infestação em condições de casa-de-vegetação. Dentre as características de resistência, os maiores valores de herdabilidade foram observados para PCS e PEG. O caráter PCS apresentou correlação genética positiva e significativa com PSM e PEG. Neste estudo, 337 SNP, 28 SSR, 13 TRAP e 41 AFLP foram mapeados em 20 grupos de ligação. Quatorze QTL foram encontrados usando o modelo restrito de múltiplos QTL e análise de Kruskal-Wallis. A maioria dos QTL foi detectada para mais de uma característica e composta por genes com efeito menor. Na análise de microarranjo foi observada uma expressão diferencial clara para as amostras de 21, 24 e 41 horas com *P. guildinii.* Assim, para o experimento de campo as vagens foram infestadas com esta espécie e amostras de vagens foram coletadas 0, 8, 24 e 46 horas após infestação. Nesta etapa foi sequenciado somente o RNA mensageiro da amostra 24 horas. Na análise de RNA-seq realizada nas vagens sem nenhum tratamento, a cultivar resistente (IAC-100) apresentou 39,4% dos genes com maior expressão e 11,68% dos genes com menor expressão do que a cultivar suscetível (CD-215). Baseado nos resultados, a seleção indireta para PCS associado com PSB pode ser realizada com sucesso para a obtenção de genótipos resistentes a percevejo. Além disso, os resultados de mapeamento de QTL foram parcialmente consistentes com estudos anteriores para característica agronômicas, sugerindo que QTL reais foram mapeados.

Palavras-chave: Locos de características quantitativas; Sequenciamento de transcritos; Microarranjo; Resistência de planta; Herdabilidade; Correlação; *Glycine max*

ABSTRACT

QTL mapping and gene expression associated with soybean resistance to stink bug complex

The group of stink bugs most frequently causing economic losses in soybean in Brazil consists of the species: *Nezara viridula, Piezodorus guildinii,* and *Euchistus heros.* Therefore, the objectives of the current research were to evaluate genetic parameters and correlations among distinct development and yield traits, map QTL associated with soybean resistance to stink bugs, and determine plant gene expression profiles associated to insect feeding. An $F_{2:3}$ population was developed by crossing IAC-100 (resistant) and CD-215 (susceptible) and it was evaluated at an experimental field. The agronomic traits evaluated were number of days to flowering (NDF), plant height at flowering (PHF), number of days to maturity (NDM), plant height at maturity (PHM), lodging (L), agronomic value (AV), and grain yield (GY). The characteristics of stink bug resistance evaluated were grain filling period (GFP), leaf retention (LR), percentage index of pod damage (PIPD), number of pods per plant (NPP), number of seeds (NS), weight of spotted seeds (WSS), healthy seed weight (HSW), and weight of a hundred seeds (WHS). To have a total of 96 samples, the two parent lines were genotyped along with the 12 most resistant and 12 most susceptible F_2 plants for the traits GFP, WHS, and HSW, and the 11 most resistant and 11 most susceptible for the trait LR. In order to determine the timing of gene expression response in pods under stink bug feeding, a microarray study was carried out with the cultivar CD-215, evaluating relative transcription levels at 5.5, 21, 24, and 41 hours post-infestation under greenhouse conditions. Among the characteristics of resistance, the highest values of heritability were observed for WHS and GFP. The trait WHS exhibited positive and significant genotypic correlation with WSS and GFP. In this study, 337 SNP, 28 SSR, 13 TRAP, and 41 AFLP markers were mapped to 20 linkage groups. Fourteen QTL were found using the restricted multiple QTL model and Kruskal-Wallis analyses. The majority of the QTL was detected for more than one trait and consisted of genes with minor effects. A clear differential gene expression was observed in the microarray analysis for the samples at time points 21, 24, and 41 hours infested with *P. guildinii.* Thus, in field trials the pods were infested with this species and samples of pods were taken at 0, 8, 24, and 46 hours. In this study, only RNA from the 24 hour sample was sequenced. From RNA-seq analysis performed on pods without treatment, the resistant cultivar (IAC-100) showed 39.4% of genes with induced expression and 11.68% of genes with repressed expression in comparison to the susceptible cultivar (CD-215). Based on the results, indirect selection for WHS associated with HSW can be successfully employed for obtaining stink bug resistant genotypes. Moreover, mapping QTL results were partially consistent with previous studies for agronomic traits, suggesting that real QTL were mapped.

Keywords: Quantitative Trait Loci; RNA sequencing; Microarray; Plant resistance; Heritability; Correlation; Stink bug; *Glycine max*

1 INTRODUCTION

1.1 Soybean economical importance

Cultivated soybean [*Glycine max (L.*) Merr.] is a member of Fabaceae family, and is thought to have originated in Northern and Central China (PROBST; JUDD, 1973). Soybean is the most important legume crop in the world and a major agricultural commodity in Brazil, Argentina, and United States. Protein from soybean seed meal is the primary source of protein in livestock feed around the world, and also a major component of soymilk, soy flour, tofu, and many other retail human food products (MARTIN; WALDREN; STAMP, 2006).

Brazil has become an important soybean producer, especially during the last few years with the crop expansion into new areas in the Cerrado region. In 2010/2011, Brazil produced 72 million tons of grain in an area of 24 million hectares, with yields averaging 2.9 $Kg.ha^{-1}$ (CONAB, 2011). Soybean is now one of the main products of Brazilian exports. In 2010, Brazil exported 29 million tons of soy grain, 13.6 million tons of soybean meal and 1.56 million of crude and refined oil and others. Soybean exports represent approximately 24.69% of Brazilian agribusiness exports (around 8.70% of the total exports), according to statistics from Companhia Nacional do Abastecimento (CONAB, 2011). Soybean cultivation in Brazil is established in the South, Southeast, and more recently, in the Central West and Northeast regions. Almost all the soybean cultivation area in Brazil is distributed in the states of Mato Grosso, Paraná, Goiás, Rio Grande do Sul, Mato Grosso do Sul, Minas Gerais and Bahia (BROGIN, 2005).

Large areas of a single plant species, such as seen in cultivated soybean, contribute to increased damage by insect pests (MOURA, 1999). Furthermore, genetic breeding focusing higher grain yields without regarding pest resistance is also responsible for increasing soybean vulnerability to insect pests (LUSTOSA et al., 1999). Due to the increased threats of insect pests, it was estimated that the cost of pesticide use to protect soybean production in Brazil in 2010/2011 was approximately $200.00 per hectare (CONAB, 2011). In Brazil, the two most important types of insect pests causing economic damage to soybean crops are stink bugs and defoliating caterpillars (VENZON et al., 1999).

1.2 The stink bug complex in soybean

Stink bugs are considered to be the most important pests attacking soybean in Brazil. Of the complex of 25 stink bug species found in Brazil, the group of species that most frequently causes economical losses is called the "stink bug complex" and consists of three species: *Nezara viridula* (Linnaeus)*, Piezodorus guildinii* (Westwood), and *Euchistus heros* (Fabricius) (Heteroptera: Pentatomidae). Due to their piercing-sucking feeding habits, these insects cause damage mainly during pod formation, filling and maturation (GAZZONI, 1998; LOURENÇÃO et al., 2002).

N. viridula, commonly known as 'Southern green stink bug', is native to Ethiopia and has worldwide distribution. Despite being more adapted to the cooler areas of Southern Brazil (PANIZZI; CORRÊA-FERREIRA, 1997), *N. viridula* is expanding its distribution to more tropical regions. *P. guildinii*, commonly known as 'Red-banded stink bug', was first described in the island of St. Vincent (STONER, 1922) and has been frequently reported in Central and South America, as well as Southern United States. In Brazil, *P. guildinii* had seldom been found on soybean plants until the early 1970s. Subsequently, it has become more common, ranging from Rio Grande do Sul (32ºS latitude) to Piauí (5ºS latitude) (PANIZZI; SLANSKY JR., 1985). *E. heros* has a wide distribution in Brazilian soybean fields, being prevalent in Mato Grosso, Paraná, Goiás, and Mato Grossso do Sul (SOSA-GÓMEZ et al., 2004).

The stink bug populations are present in variable abundance and frequency among regions and years, with one species being more predominant than another depending on the weather conditions and plant cultivar (MAGRINI et al., 1996). Stink bug damage is caused mainly by older nymphs, from 3º a 5º instars, and adults that feed directly on soybean seeds, perforating the pods and extracting nutrients from the seed (McPHERSON; McPHERSON, 2000). In general, the presence of stink bugs in soybean crops is correlated with the presence of pods in the plant, as it is their preferred food source. The most sensitive period for stink bug damage is the period between the beginning of pod development and the point of maximum dry matter accumulation in the grain (PANIZZI et al., 2000).

Plants normally have a great capacity to recover from damages caused by defoliating insects, but stink bugs cause irreversible damages, as they injure the seeds directly. The stink bugs, while feeding on the soybean plant, can cause several types of damage to the beans such as punctures, stains (usually spots associated to the pathogenic yeast *Nematospora coryli,* Peglion), deformation and reduction of seed size, reduction of oil content, increase of the protein content, delayed maturation, uneven senescence (leaf retention), reduction of seed

germination, and decrease in production. Additionally, attacked plants are challenging to harvest mechanically (YEARGAN, 1977; PANIZZI; SLANSKY JR., 1985; GAZZONI; MOSCARDI, 1998).

The degree of damage caused by stink bugs was found to be dependent upon the developmental stage of the seed at the time of piercing. When young seeds are pierced, they become completely shriveled. However, when seeds are attacked after completing their development, the final dried seeds show only slight puncture marks (BLICKENSTAFF; HUGGANS, 1962).

1.3 Plant resistance mechanisms

Stink bug control is usually accomplished by pesticide spraying. However, the use of resistant cultivars would be a more economical and environmentally safe control. Resistant host plants have been described as a highly desirable control tactic with excellent potential to regulate insect pest populations in integrated management systems. This is primarily due to the compatibility of resistant host plants with other control methods (KOGAN, 1989).

Rossetto et al. (1995) listed traits that might enhance tolerance in soybeans against stink bug attacks, such as: shorter pod-filling period, higher seed-yield per plant, capacity to reject damaged immature pods and their substitution by new pods; along with normal senescence with leaf dropping at the maturation stage, and resistance to the pathogenic yeast *N. coryli* that can be transmitted by stink bugs.

Efforts to develop advanced breeding lines with multiple insect resistance and good agronomic qualities have been made (BOETHEL, 1999). Soybean breeding programs in Brazil have been directed towards pod-feeding pests, such as, stink bugs. A number of breeding lines have been developed in Brazil by crossing resistant plant introductions (PI) with commercial (adapted) varieties of soybean (PANIZZI; SLANKY JR., 1985; LOURENÇÃO; COSTA; MIRANDA, 1989; ROSSETO, 1989). Soybean varieties IAC-100, IAC-17 (ROSSETTO et al., 1989), IAC-23 and IAC-24 (MIRANDA; LOURENÇÃO, 2002), are reportedly more tolerant to stink bug feeding and adapted to Southeastern Brazil. The variety IAC-100 has been used in several studies as a standard genotype for enhanced resistance to defoliators and sucking insects. This cultivar has also been used in soybean breeding programs in the United States. McPherson, Buss, and Roberts (2007) and McPherson and Buss (2007) reported its great potential as a source of resistance to stink bugs.

Although improved cultivars with enhanced resistance can be successfully obtained, only a few studies have been carried out with the purpose of revealing the genetic mechanisms underlying these traits and the genetic parameters associated with them.

A classification of mechanisms involved in plant resistance to insect pest was proposed by Painter (1951). Non-preference, antibiosis, and tolerance are the three mechanisms influencing the ability of a plant to grow productively in the presence of an insect (PAINTER, 1951). The term non-preference refers to a behavioral response of the insect to a plant, whereas "antibiosis" and "tolerance" refer to plant characteristics. Two or more categories of these mechanisms can work together in one plant considered insect resistant (SMITH, 1989).

Non-preference describes resistance where the insect is either repelled from or not attracted to its normal host plant. Kogan and Ortman (1978) suggested the term "antixenosis" to describe the plant properties responsible for non-preference. The plant characters that influence non-preference include color, light reflection, type of pubescence, leaf angle, odor and taste.

Antibiosis describes insect resistance where feeding of the plant results in disruption of growth, development, or physiology of the insect. When toxins or inhibitors are present, or when the levels of essential nutrients decrease, antibiosis occurs. An example of soybean proteins found to cause antibiosis includes Kunitz trypsin inhibitor and a serine proteinase inhibitor (JOHNSTON; GATEHOUSE; ANSTEE, 1993). Non-choice experiments are often used to determine plant antibiosis resistance.

Tolerance differs from the other two mechanisms in that no observable adverse effect on the insect's life cycle is present. The mechanism consists in the ability of the plant to withstand or recover from the damage caused by insect populations, similar to those on susceptible cultivars (SMITH, 2005). The quantitative assessment of tolerance is often accomplished by evaluating the plant characteristics.

1.4 QTL mapping

Plant resistance to insects is usually a quantitatively inherited trait. Quantitative traits are genetically determined characteristics that can be measured as continuous values, are controlled by multiple genes and are strongly influenced by the environment. Mapping of quantitative trait loci (QTL) is a process that estimates the number of genes and individual gene effects, determines the approximate genome position of QTL on chromosomes, and

refines the genetic architecture for the quantitative trait. QTL mapping is based on the principle that genes and markers segregate via chromosome recombination (called crossing-over) during meiosis (i.e. sexual reproduction), thus allowing their analyses in the progeny (PATERSON, 1996).

Molecular genetic markers and QTL analysis offer plant breeders a more efficient approach to work with quantitative traits, whether specific genes or gene products are known or not. Detection of QTL is a demanding process, requiring the generation of a mapping population from a cross between two parents differing in insect resistance, and ‘‘genotyping’’ (scoring for a series of molecular markers) and ‘‘phenotyping’’ (scoring for the resistance traits of interest) each member of the mapping population (YENCHO; COHEN; BYRNE, 2000).

Several studies have been published with QTL identification associated with resistance to soybean insect defoliators (RECTOR et al., 2000; TERRY et al., 2000; BOERMA; WALKER, 2005; KOMATSU et al., 2005; ZHU et al., 2006) and aphids (ZHANG; GU; WANG, 2009). However, work on gene/QTL mapping associated with soybean resistance to stink bugs is absent from the literature.

A QTL mapping study in plants starts with the build-up of a mapping population. The parents selected for the mapping population will differ for one or more traits of interest. Generally, in self-pollinating species, mapping populations originate from parents that are both highly homozygous (inbred). F_2 populations, derived from F_1 hybrids and backcross (BC) populations, derived by crossing the F_1 hybrid to one of the parents, are the simplest types of mapping populations developed for self-pollinating species. Their main advantages are that they are easy to construct and require only a short time to produce (COLLARD et al., 2005).

For more precise results, QTL experiments require large numbers of individuals to be genotyped and phenotyped for the quantitative trait of interest. Since these are costly endeavors, investigators employ cost-saving strategies, such as; selective genotyping (LANDER; BOTSTEIN, 1989). This strategy consists in genotyping only the individuals belonging to the high and low tail-ends of a phenotypic distribution (the phenotypic extremes), after increasing the number of individuals phenotyped. The application of this strategy can substantially increase the power of QTL detection for the trait based on which the individuals are selected (DARVASI; SOLLER, 1992). The precision of the estimation of the QTL effect is increased in the same way, since precision and power are strongly related.

Data analyses involve the creation of linkage maps with markers sufficiently closely-spaced to facilitate QTL detection, followed by QTL mapping (ERICKSON et al. 2004). A variety of markers can be used to determine if variation at the molecular level is linked to variation in the quantitative trait. If this is the case, then the marker is linked to, or on the same linkage group, as a QTL that has allelic variants underlying the variation of the quantitative trait. Currently, Simple Sequenced Repeats (SSR) and Single Nucleotide Polymorphism (SNP) are the molecular markers commonly used for QTL mapping studies. However, dominant markers such as Amplified Fragment Length Polymorphism (AFLP) and Target Region Amplification Polymorphism (TRAP) are still important in these studies to contribute to the map saturation.

SSR DNA markers, also known as microsatellites or Short Tandem Repeats (STR), represent one of the most useful marker systems available to breeders. They are stretches of DNA consisting of tandem short-sequence motifs of 1 to 5 repeats spread throughout the genomes of most eukaryotic species, such as CTCTCTCTCT….. (FERREIRA; GRATTAPAGLIA, 1998). Some of the major core motifs of soybean SSR markers include ATT, AT, CTT, and CT (CHOI et al., 2007; SONG et al., 2004). The DNA sequences flanking SSRs are conserved, allowing the selection of PCR primers that will amplify the intervening SSR. SSR length polymorphism is detected when the PCR products from a particular locus differ in length, which is fairly common within a species. The uniqueness and value of microsatellites arises from their multiallelic nature, codominant transmission, ease of detection by PCR (Polymerase Chain Reaction), relative abundance, extensive genome coverage and requirement for only a small amount of starting DNA (POWELL; MACHRAY; PROVAN, 1996).

The first report of the polymorphism and heritability of SSR markers in plants was in soybean (AKKAYA; BHAGWAT; CREGAN, 1992). SSR markers exhibit high levels of length polymorphism in this crop plant. Maughan, Saghai Maroof, and Buss (1995) detected a total of 79 alleles at five SSR loci in a sample of 94 accessions of cultivated and wild soybean. The high level of SSR length polymorphism in soybean makes it likely that a particular locus will be polymorphic even in populations resulting from the hybridization of adapted genotypes. The second version of the soybean integrated genetic linkage map contains a total of 1,015 SSR loci (SONG et al., 2004).

SNPs can be used as molecular markers for a variety of applications in crop improvement, including; QTL discovery. A SNP marker corresponds to a single base change in a DNA sequence, with an usual alternative of two possible nucleotides at a given position.

Although in principle, at each position of a sequence stretch, any of the four possible nucleotide bases may be present, in practice, SNPs are usually biallelic. Mutation mechanisms result in either transitions: purine-purine (A↔G) or pyrimidine-pyrimidine (C↔T) exchanges, or transversions: purine-pyrimidine or pyrimidine-purine (A↔C, A↔T, G↔C, G↔T) exchanges (VIGNAL et al., 2002).

There is some evidence that the stability of SNPs and therefore, the relative fidelity of their inheritance, are higher than that of the other marker systems like SSRs and AFLPs. SNPs are biallelic as against the polyallelic nature of the once much sought-after SSRs. However, the extraordinary abundance of SNPs largely offsets the disadvantage of their being biallelic and makes them the most attractive molecular-marker system developed so far. According to recent estimates, one SNP occurs every 100–300 bp in any genome (GUPTA; ROY; PRASAD, 2001).

SNP markers are the most abundant variations in the genome. Due to their abundance and codominance, the use of SNPs as a marker system has the potential of providing highest map resolutions. SNP typing has progressed remarkably over the last several years, making genome-wide linkage analysis and molecular breeding rapid and efficient (NASU et al., 2002). SNPs were shown to be the most abundant source of DNA polymorphisms in soybean. It is estimated that there are 4 to 5 million SNPs in the soybean genome, based on the rate of 280 SNPs observed in 76 Kbp studied by Zhu et al. (2003).

Currently, several high-throughput methods are being designed to quickly genotype a large number of markers in the genome of different species. The Illumina Golden Gate Assay (Illumina, San Diego, CA) is a high-throughput SNP detection method capable of multiplexing from 96 to 1,536 SNPs in a single reaction over a 3-day period (FAN et al. 2003). Besides the high-throughput, it was observed that the GoldenGate assay is copy-number sensitive. This makes it possible to score heterozygous genotypes and/or bulks, and, therefore, the technique is effective for identifying markers associated with traits of interest by bulk segregant analysis or association analysis, as demonstrated by Hyten et al. (2009).

The AFLP technique was developed by Zabeau and Vos (1993) and it is based on PCR amplification. AFLPs are almost invariably dominant markers, but have the advantage of generating large numbers of polymorphic bands on polyacrylamide gels. The DNA is digested with one or two restriction enzymes, suitable adapters are ligated to the fragments and the ligated DNA fragments are selectively amplified with different primer combinations. With AFLP, molecular genetic polymorphisms are identified by the presence or absence of DNA fragments following restriction and amplification of genomic DNA. Despite the dominant

character of the marker, its use to construct genetic maps has been frequent due to the large amount of information generated in a single gel, allowing the rapid saturation of maps.

The TRAP marker technique was initially developed with sunflower (*Helianthus annuus* L.) DNA samples by Hu and Vick (2003). This simple and rapid PCR-based marker system uses EST information and a bioinformatics approach to generate polymorphic markers around targeted candidate-gene sequences. TRAPs are amplified by one fixed primer of about 18 nucleotides designed from a target gene of interest and a second primer of arbitrary sequence, containing an AT- or GC- rich motif that preferentially anneals to introns and exons, respectively.

The methods used for detecting QTLs are single-marker mapping, Interval Mapping (IM), Composite Interval Mapping (CIM) and Multiple Intervals Mapping (MIM), which allow statistical analyses of the associations between phenotype and genotype for the purpose of understanding and dissecting the regions of a genome affecting complex traits (DOERGE, 2002). According to Lynch and Walsh (1998), the association between markers and QTL can be assessed using one, or two, or more simultaneous tags. Single-marker analysis is the simplest method for detecting QTLs associated with single markers. This method investigates individual markers independently, and without reference to their position or order. Therefore, this method does not require a complete molecular marker linkage map. The statistical methods used for single-marker analysis include *t*-tests, analysis of variance (ANOVA) and linear regression. The disadvantage with this method is that the further a QTL is from a marker, the less likely its statistical detection is. This is due to the possibility of the occurrence of recombination between the marker and QTL (TANKSLEY, 1993).

The IM method, proposed by Lander and Botstein (1989), uses an estimated genetic map and analyzes the intervals between adjacent pairs of linked markers along chromosomes simultaneously, instead of analyzing single markers. Zeng (1994) presented the CIM method that combines interval mapping with linear regression and includes additional genetic markers in the statistical model, in addition to an adjacent pair of linked markers, for interval mapping. This method can remove the bias caused by QTLs that are linked to the position being tested. In CIM, cofactors are used to reduce residual variation by controlling the genetic background. MIM was proposed by Kao, Zeng, and Teasdale (1999) and uses multiple intervals markers simultaneously to fit multiple putative QTL directly in the model for QTL mapping. MIM is a method with improved precision and power for mapping. Numbers, locations, effects and epistasis effects between QTLs can be readily estimated and analyzed.

These QTL mapping methods discussed above share a common assumption; that the phenotype follows a normal distribution with equal variance in both parental. Under this assumption, the presence of a QTL can be tested by a simple parametric test. However, some phenotypes of interest are not normally distributed. In such cases, traditional QTL mapping methods cannot be directly applied. An alternative approach is to apply non-parametric methods to QTL mapping (KRUGLYAK; LANDER, 1995). 'Non-parametric' means that no assumptions are being made for the probability distribution (s) of the quantitative trait. The Kruskal-Wallis test (LEHMANN, 1975) can be regarded as the non-parametric equivalent of the one-way analysis of variance. The test ranks all individuals according to the quantitative trait, while it classifies them according to their marker genotype. Marker-QTL(s) associations are detected based on the level of differences in average rank between the marker and genotype classes (MOUSA et al., 2004).

1.5 High-throughput gene expression analysis

Plants are constantly challenged with biotic and abiotic stresses and have evolved a series of constitutive and inducible responses to adapt and survive in their environments. Induced defenses have received a lot of attention and, over the years, a large number of genes encoding defense-related genes have been identified. A vast majority of these defense-related genes are induced when the plant is attacked by diverse aggressors, like microbial pathogens, viruses, and insects (REYMOND, 2001).

Several technologies have been developed to deduce and quantify the transcriptome, including hybridization or sequence-based approaches. Many attribute the invention of DNA microarray technology, or cDNA 'gene chips', to researchers at Stanford University, as first described by Schena et al. (1995). This technology provides the important advantage of allowing parallel quantification of the expression levels of an almost unlimited number of genes from a given genome, in response to a given treatment.

Microarrays correspond to the immobilization of millions of single-strand copies of a given gene, as a spot of DNA on a solid support surface, with tens of thousands of such spots (genes) physically present on a single microarray. The array surface is then incubated with fluorescently labeled cDNA molecules to allow hybridization to complementary, spotted DNA. Out of this mixture, only the labeled molecules that represent the same gene as the immobilized DNA spots can form heteroduplexes, and these spots (spot of a single gene) will fluoresce at an intensity proportional to the amount of cDNA bound, which is therefore

proportional to the amount of mRNA corresponding to the gene in the tissue being studied. By measuring the amount of label that is bound to each array spot at the end of the hybridization reaction, a researcher can determine the relative transcript abundance level of each gene. Because each microarray consists in tens of thousands of spots (genes), RNA abundance levels for thousands of genes can be measured in a single experiment. By comparing abundance levels from several experiments, the investigator can correlate patterns of gene expression with particular tissues or experimental conditions (DEYHOLOS; GALBRAITH, 2001).

Plant defense responses are diverse and complex and, therefore, many researchers have tried to obtain a clearer idea of the physiology of these responses using microarrays to determine global gene expression patterns. For example, microarray studies have demonstrated that herbivory and mechanical damage elicit different gene expression responses, suggesting that plants respond to compounds released by herbivores, in addition to the simple wounding an insect may cause (KORTH, 2003). Microarrays are available for many plants, including soybean (VODKIN et al., 2004).

Although a lot of information can be gained from microarray analysis, these hybridization-based methods have several limitations, such as the requirement of a gene to be spotted in order to be analyzed (one must know and spot all the genes of an organism in order to study expression of all the genes), high background levels owing to cross-hybridization, and a limited dynamic range of detection due to both background and saturation of signals. Moreover, comparing expression levels across different experiments is often difficult and may require complicated normalization methods (WANG; GERSTEIN; SNYDER, 2009).

More recently, massively parallel RNA sequencing, known as ‘RNA-seq’ or whole-transcriptome shotgun sequencing, has been proposed as a method with important advantages over microarray-based methods (GRIFFITH et al., 2010). The RNA-seq method was originally developed to take advantage of the next-generation Illumina sequencing technology (SEVERIN et al., 2010). RNA-seq quantifies expression levels of thousands of genes by counting the number of transcripts encountered for each gene (HUDSON, 2008). The main steps in RNA-seq are: (i) isolate RNA from a sample, (ii) convert mRNA to cDNA fragments via reverse-transcription and fragmentation, (iii) use a high-throughput sequencer, such as those from Illumina (San Diego, CA, USA), Applied Biosystems (Foster City, CA, USA) and Roche 454 Life Science (Branford, CT, USA), to generate millions of reads from the cDNA fragments, (iv) map the reads to a reference genome or transcript set with a software alignment tool, and (v) count the reads mapped to each gene and use the counts to estimate

comparative expression levels (LI et al., 2010). The primary advantages of RNA-seq are its large dynamic range (spanning five orders of magnitude), low background noise, low sample amounts needed, and the ability to detect novel transcripts even in the absence of a sequenced genome (WANG; GERSTEIN; SNYDER, 2009).

The two principal differences between sequencing and microarray results are in the resolution of the output, and in the dynamic range of changes that can be observed (WILHELM; LANDRY, 2009). As RNA-seq is quantitative, it can be used to determine RNA expression levels more accurately than microarrays (WANG; GERSTEIN; SNYDER, 2009).

References

AKKAYA, M.S.; BHAGWAT, A.A.; CREGAN, P.B. Length polymorphisms of simple sequence repeat DNA in soybean. **Genetics**, Berlin, v. 132, p.1131-1139, 1992.

BLICKENSTAFF, C.C.; HUGGANS, J.L. **Soybean insects and related arthropods in Missouri**. Missouri: Agricultural Experiment Station, College of Agriculture, University of Missouri, 1962. 51p. (Research Bulletin, 803).

BOERMA, H.R.; WALKER, D.R. Discovery and utilization of QTLs for insect resistance in soybean. **Genetica**, Dordrecht, v. 123, p. 181-189, 2005.

BOETHEL, D.J. Assessment of soybean germplasm for multiple insect resistance. In: CLEMENT, S.L.; QUISENBERRY, S.S. **Global plant genetic resources for insect-resistant crops**. Boca Raton: CRC Press, 1999. p. 101-129.

BROGIN, R.L. **Mapeamento de genes de resistência à ferrugem e de QTLs envolvidos na resistência à septoriose em soja**. 2005. 93 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2005.

CHOI, I.Y.; HYTEN, D.L.; MATUKUMALLI, K.L.; SONG, Q.; CHAKY, M.J.; QUIGLEY, V.C.; CHASE, K.; LARK, G.K.; REITER, S.R.; YOON, M.S.; HWANG, E.Y.; YI, S.I.; YOUNG, N.D.; SHOEMAKER, R.C.; VAN, T.C.P.; SPECHT, J.E.; CREGAN, P.B. A soybean transcript map: gene distribution, haplotype an single-nucleotide polymorphism analysis. **Genetics**, Pittsburgh, v. 176, p. 685-696, 2007.

COLLARD, B.C.Y.; JAHUFER, M.Z.Z.; BROUWER, J.B.; PANG, E.C.K. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. **Euphytica**, Dordrecht, v. 142, p.169-196, 2005.

COMPANHIA NACIONAL DE ABASTECIMENTO – CONAB. Disponível em: http://www.conab.gov.br. Acesso em: 13 jun. 2011.

DARVASI, A.; SOLLER, M. Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. **Theoretical and Applied Genetics**, Berlin, v. 85, p. 353-359, 1992.

DEYHOLOS, M.K.; GALBRAITH, D.W. High-density microarrays for gene expression analysis. **Cytometry**, New York, v. 43, p. 229-238, 2001.

DOERGE, R.W. Mapping and analysis of quantitative trait loci in experimental populations. **Nature Reviews Genetics**, London, v. 3, p. 43-52, 2002.

ERICKSON, D.L.; FENSTER, C.B.; STENOIEN, H.; PRICE, D. Quantitative trait locus analyses and the study of evolutionary processes. **Molecular Ecology**, Oxford, v. 13, p. 2505-2522, 2004.

FAN, J.B.; OLIPHANT, A.; SHEN, R.; KERMANI, B.G.; GARCIA, F.; GUNDERSON, K.L.; HANSEN, M.; STEEMERS, F.; BUTLER, S.L.; DELOUKAS, P.; GALVER, L.; HUNT, S.; MCBRIDE, C.; BIBIKOVA, M.; RUBANO, T.; CHEN, J.; WICKHAM, E.; DOUCET, D.; CHANG, W.; CAMPBELL, D.; ZHANG, B.; KRUGLYAK, S.; BENTLEY, D.; HAAS, J.; RIGAULT, P.; ZHOU, L.; STUELPNAGEL, J.; CHEE, M.S. Highly parallel SNP genotyping. **Cold Spring Harbor Symposia on Quantitative Biology**, New York, v.68, p.69–78, 2003.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análises genéticas**. 3.ed. Brasília: EMBRAPA-CENARGEN, 1998. 220 p.

GAZZONI, D.L. Efeito de populações de percevejos na produtividade, qualidade da semente e características agronômicas da soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.33, p.1229-1237, 1998.

GAZZONI, D.L.; MOSCARDI, F. Effect of defoliation levels on recovery of leaf area, on yield and agronomic traits of soybeans. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 4, p. 411-424, 1998.

GRIFFITH, M.; GRIFFITH, O.L.; MWENIFUMBO, J.; GOYA, R.; MORRISSY, A.S.; MORIN, R.D.; CORBETT, R; TANG, M.J.; HOU, Y.C.; PUGH, T.J.; ROBERTSON, G.; CHITTARANJAN, S.; ALLY, A.; ASANO, J.K.; CHAN, S.Y.; LI, H.I.; MCDONALD, H.; TEAGUE, K.; ZHAO, Y.; ZENG, T.; DELANEY, A.; HIRST, M.; MORIN, G.B.; JONES, S.J.M.; TAI, I.T.; MARRA, M.A. Alternative expression analysis by RNA sequencing. **Nature methods**, New York, v.7, n. 10, p. 843-850, 2010.

GUPTA, P.K.; ROY, J.K.; PRASAD, M. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. **Current Science**, Bangalore, v. 80, n. 4, p. 524-535, 2001.

HU, J.; VICK, B.A. Target Region Amplification Polymorphism: A novel marker technique for plant genotyping. **Plant Molecular Biology Reporter**, New York, v. 21, p. 289-294, 2003.

HUDSON, M.E. Sequencing breakthroughs for genomic ecology and evolutionary biology. **Molecular Ecology Resources**, Oxford, v. 8, p. 3-17, 2008.

HYTEN, D.L.; SMITH, J.R.; FREDERICK, R.D.; TUCKER, M.L.; SONG, Q.; CREGAN, P.B. Bulked Segregant Analysis using the GoldenGate assay to locate the *Rpp3* locus that confers resistance to soybean rust in soybean. **Crop Science**, Madison, v. 49, p.265-271, 2009.

JOHNSTON, K.A.; GATEHOUSE, J.A.; ANSTEE, J.H. Effects of soybean protease inhibitors on the growth and development of larval *Helicoverpa armigera*. **Journal of Insect Physiology**, Oxford, v. 39, n. 8, p. 657-664, 1993.

KAO, C.H; ZENG, Z.B.; TEASDALE, R. Multiple interval mapping for quantative trait loci. **Genetics,** Pittsburgh, v. 152, p. 1023-1216, 1999.

KOGAN, M. Plant resistance in soybean insect control. In: CONFERÊNCIA MUNDIAL DE INVESTIGACIÓN EN SOJA, 4., 1989, Buenos Aires. **Actas**... Buenos Aires: Orientación Gráfica Editora SRL, 1989, v.3, p.1519-1525.

KOGAN, M.; ORTMAN, E.E. Antixenosis – a new term proposed to define Painter´s ´nonpreference´ modality of resistance. **Bulletin of the Entomological Society of America,** Lanhan, v. 24, p. 175-176, 1978.

KOMATSU, K.; OKUDA, S.; TAKAHASHI, M.; MATSUNAGA, R.; NAKAZAWA, Y. QTL Mapping of antibiosis resistance to common cutworm (*Spodoptera litura* Fabricius) in soybean. **Crop Science**, Madison, v. 45, p. 2044-2048, 2005.

KORTH, K.L. Profiling the response of plants to herbivorous insects. **Genome Biology**, London, v. 4, n. 7, article 221, 2003. Disponível em: <http://genomebiology.com/2003/4/7/221>. Acesso em: 08 out. 2011.

KRUGLYAK, L.; LANDER, E.S. A nonparametric approach for mapping quantitative trait loci. **Genetics**, Pittsburgh, v. 139, p. 1421-1428, 1995.

LANDER, E.S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Pittsburgh, v. 121, p. 185-199, 1989.

LEHMANN, E.L. **Nonparametrics:** Statistical methods based on ranks. San Francisco: Holden-Day, 1975. 480 p.

LI, B.; RUOTTI, V.; STEWART, R.M.; THOMSON, J.A.; DEWEY, C.N. RNA-Seq gene expression estimation with read mapping uncertainty. **Bioinformatics**, Oxford, v. 26, n. 4, p. 493-500, 2010.

LOURENÇÃO, A.L.; COSTA, A.S.; MIRANDA, M.A.C. Sources of resistance to Insect pests and virus vectors in the soybean germplasm tested at Instituto Agronômico, SP, Brasil. In: WORLD SOYBEAN RESERCH CONFERENCE, 4., 1989, Buenos Aires. **Anais...** Buenos Aires: Gráfica SRL, 1989. v. 4, p. 1578-1581.

LOURENÇÃO, A.L.; BRAGA, N.R.; MIRANDA, M.A.C.; VALLE, G.E.; PEREIRA, J.C.V.N.A; RECO, P.C. Avaliação de danos de percevejos e de desfolhadores em genótipos de soja de ciclos precoce, semiprecoce e médio. **Neotropical Entomology**, Londrina, v. 31, n. 4, p. 623- 630, 2002.

LUSTOSA, P.R.; ZANUNCIO, J.C.; LEITE, G.L.D.; PICANÇO, M. Qualidade da semente e senescência de genótipos de soja sob dois níveis de infestação de percevejos (Pentatomidae). **Pesquisa Agropecuária Brasileira**, Brasília, v. 34, n. 8, p. 1347-1351, 1999.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Massachusetts: Sinauer Sunderland, 1998. 980 p.

MAGRINI, E.A.; SILVEIRA NETO, S.; BOTELHO, P.S.M; NEGRIM, S.G. Ocorrência de percevejos no município de Pirassununga-SP, associada com a fenologia da soja. **Revista de Agricultura,** Piracicaba, v.71, p.187-193, 1996.

MARTIN, J.H.; WALDREN, R.P.; STAMP, D.L. **Principles of field crop production**. 4th ed. Pearson Education, Upper Saddle River, NJ. 2006. 1030p.

MAUGHAN, P.J.; SAGHAI MAROOF, M.A.; BUSS, R.G. Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. **Genome**, Ottawa, v. 38, p. 715-723, 1995.

McPHERSON, J.E.; McPHERSON, R.M. **Stink bugs of economic importance in America north of Mexico.** Boca Raton, FL:CRC, 2000. 253p.

McPHERSON, R.M.; BUSS, G.R. Evaluating lepidopteran defoliation resistance in soybean breeding lines containing the stink bug (Hemiptera: Pentatomidae) resistance IAC-100 cultivar in their pedigrees. **Journal of Economic Entomology**, Lanham, v. 100, n. 3, p. 962-968, 2007.

McPHERSON, R.M.; BUSS, G.R.; ROBERTS, P.M. Assessing stink bug resistance in soybean breeding lines containing genes from germplasm IAC-100. **Journal of Economic Entomology**, Lanham, v. 100, n. 4, p. 1456-1463, 2007.

MIRANDA, M.A.C.; LOURENÇÃO, A.L. Melhoramento genético da soja para a resistência a insetos: uma realidade para aumentar a eficiência do controle integrado de pragas e viabilizar a soja orgânica. In: CONGRESSO BRASILEIRO DE SOJA E MERCOSOJA, 2., 2002, Londrina. **Anais**... Londrina: Embrapa-Soja, 2002. p. 52-60.

MOURA, M.F. **Avaliação de linhagens de soja (*Glycine max* (L.) Merrill) visando adaptação às condições de cerrado e resistência a insetos**. 1999. 156p. Dissertação (Mestrado em Agronomia) - Escola de Agronomia, Universidade Federal de Goiás, 1999.

MOUSA, M.A.A.; KANDEEL, N.M.; MOHAMED, M.F.; ABOUL NASR, M. H. Mapping QTL(s) for resistance to the beet cyst nematode (*Heterodera schachtii* Schm.) in radish (*Raphanus sativus* L.). **Assiut University Bulletin for Environmental Researches,** Assiut, v. 7, n. 2, p. 1-17, 2004.

NASU, S.; SUZUKI, J.; OHTA, R.; HASEGAWA, K.; YUI, R.; KITAZAWA, N.; MONNA, L.; MINOBE, Y. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. **DNA Research**, Oxford, v. 9, p. 163–171, 2002.

PAINTER, R.H. **Insect resistance in crop plants**. Lawrence: The University Press of Kansas, 1951. 520p.

PANIZZI, A.R.; CORRÊA-FERREIRA, B.S. Dynamics in the insect fauna adaptation to soybean in the tropics. **Trends in Entomology**, Trivandrum, v. 1, p. 71-88, 1997.

PANIZZI, A.R.; SLANSKY JÚNIOR, F. Review of phytophagous pentatomids (Hemiptera: Pentatomidae) associated with soybean in the Americas. **Florida Entomologist**, Lutz, v. 68, p. 184-214, 1985.

PANIZZI, A.R.; McPHERSON, J.E.; JAMES, D.G.; JAVAHERY, M.; McPHERSON, R.M. Stink bugs (Pentatomidae). In: SCHAEFER, C.W.; PANIZZI, A.R. (Ed.). **Heteroptera of economic importance**. Boca Raton, Florida: CRC, 2000. p. 432-434.

PATERSON, A.H. Making genetic maps. In: PATERSON, A.H. **Genome Mapping in Plants**, R.G. Landes , San Diego, California; Academic Press, Austin, Texas, 1996. p. 23-39.

POWELL, W.; MACHRAY, G.C.; PROVAN, J. Polymorphism revealed by simple sequence repeats. **Trends in Plant Science,** Oxford, v. 1, n. 7, p. 215-222, 1996.

PROBST, A.H.; JUDD, R.W. Origin, U.S. history and development, and world distribution. In: CALDWELL, B.E. (Ed.). **Soybeans:** improvement, production, and uses. Madison: American Society of Agronomy, 1973. p. 1-15.

RECTOR, B.G.; ALL, J.N.; PARROTT, W.A.; BOERMA, H.R. Quantitative trait loci for antibiosis resistance to corn earworm in soybean. **Crop Science**, Madison, v. 40, p. 233-238, 2000.

REYMOND, P. DNA microarrays and plant defense. **Plant Physiology and Biochemistry**, New Delhi, v. 39, p. 313-321, 2001.

ROSSETTO, C.J. Breeding for resistance to stink bugs. In: WORLD SOYBEAN RESEARCH CONFERENCE, 4., 1989, Buenos Ayres. **Proceedings**… Buenos Ayres: Associacion Argentina de la Soja Press, 1989. v. 4, p. 2046-2060.

ROSSETTO, C.J.; TISSELLI FILHO, O.; RAZERA, L.F.; GALLO, P.B.; PEDRO JÚNIOR, M.J.; CAMARGO, M.B.P.; IGUE, T.; TEIXEIRA, J.P.F. Integration of resistance cultivar and date of plant for cultivation of soybean with reduced used of insecticides. In: WORLD SOYBEAN RESEARCH CONFERENCE, 4., 1989, Buenos Ayres. **Proceedings…** Buenos Ayres: Association Argentina de la Soja Press, 1989. v. 4, p. 1582-1587.

ROSSETTO, C.J.; GALLO, P.B. , RAZERA, L.F.; BORTOLETTO, N.; IGUE, T.; MEDINA, P.F.; TISSELLI FILHO, O.; AQUILERA, V.; VEIGA, R.F.A.; PINHEIRO, J.B. Mechanisms of resistance to stink bug complex in the soybean cultivar 'IAC-100'. **Anais da Sociedade Entomológica do Brasil**, Londrina, v. 24, p. 517-522, 1995.

SCHENA M.; SHALON, D.; DAVIS, R.; BROWN, P. Quantitative monitoring of gene express, ion patterns with a complementary DNA microarray. **Science**, Washington, v. 270, p. 467-470, 1995.

SEVERIN, A.J.; WOODY, J.L.; BOLON, Y.T.; JOSEPH, B.; DIERS, B.W.; FARMER, A.D.; MUEHLBAUER, G.J.; NELSON, R.T.; GRANT, D.; SPECHT, J.E.; GRAHAM, M.A.; CANNON, S.B.; MAY, G.D.; VANCE, C.P.; SHOEMAKER, R.C. RNA-Seq atlas of *Glycine max*: A guide to the soybean transcriptome. **BMC Plant Biology**, London, v. 10, n. 160, p. 1-18, 2010. Disponível em: <http://www.biomedcentral.com/1471-2229/10/160>. Acesso em: 13 set.2011.

SMITH, C.M. **Plant resistance to insects:** A fundamental approach. New York: John Wiley , 1989. 294p.

SMITH, C.M. **Plant resistance to arthropods:** molecular and conventional approaches**.** The Netherlands: Springer. 2005. 432p.

SONG, Q.; MAREK, L.; SHOEMAKER, R.; LARK, K.; CONCIBIDO, V.; DELANNAY, X.; SPECHT, J.; CREGAN, P. A new integrated genetic linkage map of the soybean. **Theoretical and Applied Genetics**, Berlin, v. 109, p. 122-128, 2004.

SOSA-GÓMEZ, D.R.; DELPIN, K.E.; ALMEIDA, A.M.R.; HIROSE, E. Genetic differentiation among Brazilian populations of *Euschistus heros* (Fabricius) (Heteroptera: Pentatomidae) based on RAPD analysis. **Neotropical Entomology**, Londrina, v. 33, n. 2, p. 179-187, 2004.

STONER, D. Report on the Scutelleroidea collected by the Barbados-Antigua expedition from the University of Iowa in 1918. **University of Iowa Studies in Natural History**, Iowa, v. 10, p. 3-18, 1922.

TANKSLEY, S.D. Mapping polygenes. **Annual Review of Genetics**, Palo Alto, v. 27, p. 205-233, 1993.

TERRY, L.I.; CHASE, K.; JARVIK, T.; ORF, J.; MANSUR, L.; LARK, K.G. Soybean quantitative trait loci for resistance to insects. **Crop Science**, Madison, v. 40, p. 375-382, 2000.

VENZON, M; RIPPOSATI, J.G.; FERREIRA, J.A.M.; VIRÍSSIMO, J.H. Controle biológico de percevejos-da-soja no Triângulo Mineiro. **Ciência e Agrotecnologia**, Lavras, v. 23, n. 1, p. 70-78, 1999.

VIGNAL, A.; MILAN, D.; SANCRISTOBAL, M.; EGGEN, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, Paris, v. 34 p. 275-305, 2002.

VODKIN, L.O.; KHANNA, A.; SHEALY, R.; CLOUGH, S.J.; GONZALEZ, D.O.; PHILIP, R.; ZABALA, G.; THIBAUD-NISSEN, F.; SIDAROUS, M.; STRÖMVIK, M.V.; SHOOP, E.; SCHMIDT, C.; RETZEL, E.; ERPELDING, J.; SHOEMAKER, R.C.; RODRIGUEZ-HUETE, A.M.; POLACCO, J.C.; CORYELL, V., KEIM, P.; GONG, G., LIU, L; PARDINAS, J.; SCHWEITZER, P. Microarrays for global expression constructed with a low redundancy set of 27,500 sequenced cDNAs representing an array of developmental stages and physiological conditions of the soybean plant. **BMC Genomics**, London, v. 5, n. 1, p. 1-18, 2004. Disponível em: <http://www.biomedcentral.com/1471-2164/5/73>. Acesso em: 9 out. 2011.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, London, v. 10, n. 1, p. 57-63, 2009.

WILHELMA, B.T.; LANDRY, J.R. RNA-Seq: quantitative measurement of expression through massively parallel RNA-sequencing. **Methods**, San Diego, v. 48, p. 249–257, 2009.

YEARGAN, K.V. Effects of green stink bug damage on yield and quality of soybeans. **Journal of Economic Entomology**, Lanham, v. 70, n. 5, p. 619-622, 1977.

YENCHO, G.C.; COHEN, M.B.; BYRNE, P.F. Applications of tagging and mapping insect resistance loci in plantas. **Annual Review of Entomology**, Stanford, v. 45, p. 393-422, 2000.

ZABEAU, M.; VOS, P.**Selective restriction fragment amplification:** a general method for DNA fingerprinting. Munich: EUROPEAN PATENT OFFICE, 1993.
(Publication number 0 534 858 A1, bulletin 93/13) .

ZENG, J.W. Precision mapping of quantitative trait loci. **Genetics,** Pittsburgh, v. 136, p. 1457-1468, 1994.

ZHANG, G.; GU, C.; WANG, D. Molecular mapping of soybean aphid resistance genes in PI 567541B. **Theorical and Applied Genetics**, Berlin, v. 118, p. 473-482, 2009.

ZHU, S.; WALKER, D.R.; BOERMA, H.R.; ALL, J.N.; PARROTT, W.A. Fine mapping of a major insect resistance QTL in soybean and its interaction with minor resistance QTLs. **Crop Science**, Madison, v. 46, p. 1094-1099, 2006.

ZHU, Y.L.; SONG, Q.J.; HYTEN, D.L.; TASSELL, C.P. van; MATUKUMALLI, L.K. Single-nucleotide polymorphisms in soybean. **Genetics**, Pittsburgh, v. 163, p. 1123-1134, 2003.

2 HERITABILITY AND CORRELATIONS BETWEEN AGRONOMIC TRAITS AND TRAITS FOR STINK BUG RESISTANCE IN A $F_{2:3}$ SOYBEAN POPULATION

Abstract

In Brazil, the most important pest insects causing economic damage to soybean are stink bugs. Soybean genotypes with enhanced resistance to stink bugs have been obtained via indirect selection. The objective of the current research was to evaluate genetic parameters and correlations among distinct development and yield traits in an $F_{2:3}$ soybean population. The population was developed by crossing the resistant cultivar IAC-100 with the susceptible one CD-215. A total of 3,465 plants, including the parents and the $F_{2:3}$ generation, were sown in the field at the experimental area of ESALQ/USP, in Anhembi, São Paulo, in a randomized complete block design. The experimental units were plots consisting of five F_3 plants derived from the same F_2 plant. A total of 229 $F_{2:3}$ progenies and the parents were evaluated with three replicates. The experimental field was open to natural stink bug infestation via the absence of chemical insect control. The agronomic traits evaluated were: number of days to flowering (NDF), plant height at flowering (PHF), number of days to maturity (NDM), plant height at maturity (PHM), lodging (L), agronomic value (AV), and grain yield (GY). The characteristics of insect resistance evaluated were: grain filling period (GFP), leaf retention (LR), percentage index of pod damage (PIPD), number of pods per plant (NPP), number of seeds (NS), weight of spotted seeds (WSS), healthy seed weight (HSW), and weight of a hundred seeds (WHS). Results from the present research show that the natural infestation of stink bugs allows the discrimination of resistant genotypes, and that there is substantial genetic variability within progenies derived from the cross between IAC-100 and CD-215. Except for L, the studied traits showed significant differences based on the analysis of variance. The cultivar IAC-100 and its progenies presented superior performances to most of the traits evaluated. The highest value of heritability was observed for PHM (90.02), and the lowest values were for L (8.25) and LR (20.60). Among the characteristics of resistance, the highest values were for WHS (75.00) and GFP (71.56). The highest positive correlations occurred between the variables HSW and GY (r_G = 0.98), and NPP and NS (r_G = 0.96). The phenotypic and genotypic correlations between the WHS and the WSS were significant and positive, corresponding to 0.49 and 0.50, respectively. GFP also exhibited significant genotypic correlation (r_G = 0.66) with WHS. Thus, breeders can select genotypes for stink bug resistance indirectly employing the WHS associated with HSW.

Keywords: *Glycine max*; Insect resistance; Stink bugs; Heritability; Correlation

HERDABILIDADE E CORRELAÇÕES ENTRE CARACTERES AGRONÔMICOS E DE RESISTÊNCIA A PERCEVEJOS EM UMA POPULAÇÃO $F_{2:3}$ DE SOJA

Resumo

No Brasil, os insetos praga mais importantes que causam dano econômico à soja são os percevejos. Genótipos de soja com maior resistência aos percevejos têm sido obtidos via seleção indireta. O objetivo desta pesquisa foi avaliar parâmetros genéticos e correlações entre as diferentes características de desenvolvimento e produção em uma população $F_{2:3}$ de soja. Essa população foi desenvolvida pelo cruzamento da cultivar resistente IAC-100 com a cultivar suscetível CD-215. Um total de 3.465 plantas, incluindo os genitores e a geração $F_{2:3}$, foram semeadas no campo na área experimental da Escola Superior de Agricultura "Luiz de Queiroz"/Universidade de São Paulo, em Anhembi, São Paulo, em um delineamento em blocos casualizados. As unidades experimentais foram parcelas consistindo de cinco plantas F_3 derivadas da mesma planta F_2. Um total de 229 progênies $F_{2:3}$ e o genitores foram avaliados com três repetições. O campo experimental teve infestação natural de percevejos, ou seja, sem o controle químico de insetos. As características agronômicas avaliadas foram: número de dias para o florescimento (NDF), altura da planta no florescimento (APF), número de dias para a maturidade (NDM), altura da planta na maturidade (APM), acamamento (AC), valor agronômico (VA) e produtividade de grãos (PG). As características de resistência a insetos avaliadas foram: período de enchimento de grãos (PEG), retenção foliar (RF), índice percentual de danos nas vagens (IPDV), número de vagens por planta (NVP), número de sementes (NS), peso de sementes manchadas (PSM), peso de sementes boas (PSB) e peso de cem sementes (PCS). Nesta pesquisa, verificou-se que a infestação natural de percevejos permitiu a discriminação de genótipos resistentes e que existe considerável variabilidade genética nas progênies derivadas do cruzamento entre IAC-100 e CD-215. Exceto para AC, as características estudadas mostraram diferenças significativas com base na análise de variância. A cultivar IAC-100 e suas progênies apresentaram desempenhos superiores para a maioria das características avaliadas. O maior valor de herdabilidade foi observado para APM (90,02) e os menores valores foram para AC (8,25) e RF (20,60). Entre as características de resistência, os maiores valores foram para PCS (75,00) e PEG (71,56). Os maiores coeficientes de correlação positiva ocorreram entre as variáveis PSB e PG (r_G=0,98), e NVP e NS (0,96). As correlações fenotípicas e genotípicas entre PCS e PSM foram significativas e positivas, sendo de 0,49 e 0,50, respectivamente. O PEG também apresentou correlação genotípica significativa (r_G=0,66) com PCS. Assim, melhoristas podem selecionar genótipos para resistência a percevejos indiretamente através do PCS associado com PSB.

Palavras-chave: *Glycine max;* Resistência a insetos; Percevejos; Herdabilidade; Correlação

2.1 Introduction

Soybean (*Glycine max* (L.)Merrill) is the most important legume crop worldwide and a major agricultural commodity in Brazil. The plant is attacked by many insects during its life cycle. In Brazil, the most important pest insects causing economic damage to soybean are stink bugs. The group of stink bug species most frequently causing economic losses in the crop is called the "stink bug complex" and consists of three species: *Nezara viridula, Piezodorus guildinii,* and *Euchistus heros.* These insects cause damage mainly during soybean reproductive stages by piercing the pod and extracting nutrients from the seed with their piercing/sucking mouthparts (McPHERSON; McPHERSON, 2000).

Stink bug damage is associated with feeding punctures on seeds, as well as, with stains on the seeds caused by pathogen infections (mainly spots due to infection with the pathogenic yeast *Eremothecium coryli* (previously named *Nematospora coryli* Peglion). It is also responsible for the decrease of pod numbers, seed deformation and reduction of seed size, lower germination rates, reduction of oil content, increase in protein content, delayed maturation and decrease in production (PANIZZI; SLANSKY JR., 1985; GAZZONI; MOSCARDI, 1998). Stink bugs can be controlled by insecticides, however, insecticide usage has safety concerns and a high costs to growers. A preferred method of control is the use of resistant cultivars. Genotypes with enhanced resistance to stink bugs were obtained via indirect selection for shorter grain-filling periods, lower percentages of spotted seeds, lower hundred-seed weights, and lower indices of pod damage (GODÓI et al., 2002; PINHEIRO et al., 2005). According to Rossetto et al., (1995), these traits contribute to resistance mechanisms against stink bugs in soybean.

Soybean breeding programs in Brazil have developed cultivars with enhanced resistance to stink bugs. Some of those more resistant soybean varieties are IAC-100, IAC-17 (ROSSETTO, 1989), IAC-23 and IAC-24 (MIRANDA; LOURENÇÃO, 2002). It is, therefore, important to evaluate the heritability of agronomic and insect resistance traits, and to study correlations among these traits, in order to allow breeders to use the information in designing strategies and implementing selection. Heritability coefficient expresses the portion of the total variance attributed to the mean effect of genes, and it is an important predictive character using phenotypic value as a guide to genetic value (FALCONER; MACKAY, 1996). Correlation measures a co-relation, a joint property of two variables. When variables are jointly affected by the same external influences, the correlation may offer the most logical approach to analyze the data (STEEL; TORRIE, 1980). Correlation coefficients are very

useful in quantifying the size and direction of trait associations. Knowledge of significant correlations, either positive or negative, between agronomic and quality traits may be of direct or indirect use in improving traits in soybean (CICEK et al., 2006).

The objective of the current research was to evaluate genetic parameters and correlations among distinct development and yield traits in an $F_{2:3}$ soybean population.

2.2 Material and Methods

2.2.1 Plant material

An $F_{2:3}$ population was developed by crossing the cultivars IAC-100 and CD-215. The resistant cultivar IAC-100 is classified into the maturity group (MG) 8 and was developed by Instituto Agronômico de Campinas in São Paulo, from a cross between the breeding lines IAC 78-2318 and IAC-12 (ROSSETTO et al., 1995). The line IAC 78-2318 is a source of multiple-resistance genes against soybean-attacking insects (LOURENÇÃO; MIRANDA, 1987). CD-215, a protected cultivar developed by COODETEC (Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda), was used as stink bug susceptible parent. It is derived from a cross between OC88-127 and OC90-234. The crosses performed in order to obtain the F_1 generation were made artificially in the greenhouse of the Genetics Department at Escola Superior de Agricultura "Luiz de Queiroz", from the Universidade de São Paulo (ESALQ/USP), between November 2007 and January 2008. During the winter of 2008, F_1 seeds were sown in the greenhouse and the F_2 generation was obtained via natural selfing. In 2008/2009, F_2 seeds were sown in an experimental field at the experimental station of the Genetics Department of ESALQ/USP, located in Anhembi-SP. The $F_{2:3}$ generation was obtained via natural selfing and was used in the current study.

2.2.2 Field phenotyping

In order to ensure germination, $F_{2:3}$ generation and parents seeds were germinated in a growth chamber at 28 °C, under constant light (Figure 1). The seeds were treated with a fungicide to minimize fungal pathogen damage during germination and emergence. The seedlings were transplanted in plastic cups and grown to the V1 stage, and then transplanted to the field (Figure 2).

Figure 1 - Seed germination in growth chamber

Figure 2 - Seedlings growing in plastic cups

The experimental plot was transplanted in December 19th, 2009, in an experimental field at the ESALQ/USP, in Anhembi, São Paulo. A total of 3,465 plants, including the parents and the $F_{2:3}$ generation, was planted in the field in a randomized complete block design. The experimental units were plots, consisting of five F_3 plants derived from the same

F_2 plant. A total of 229 $F_{2:3}$ progenies and the parents were tested in an experiment with three replicates (5 plants per plot, 231 plots, 3 replicates from a total of 3,465 plants).

Trait measures were taken for five plants per plot. Experimental plots were 1.2 m long, 0.4 m apart, and the plants were spaced of 0.2 m within each plot. The distance between experimental rows was of 1.5 m. Two border rows were sown with cultivar BR-133 between experimental rows (at 0.5 m spacing), as well as, 2 or 3 border rows surrounding each experimental field, to encourage an uniform distribution of stink bug populations among the plots during the reproductive period, and to protect the experimental lines from being accidently stepped on during evaluations. The entire experimental field was approximately 70.8 m long and 29.5 m wide (Figures 3 and 4).

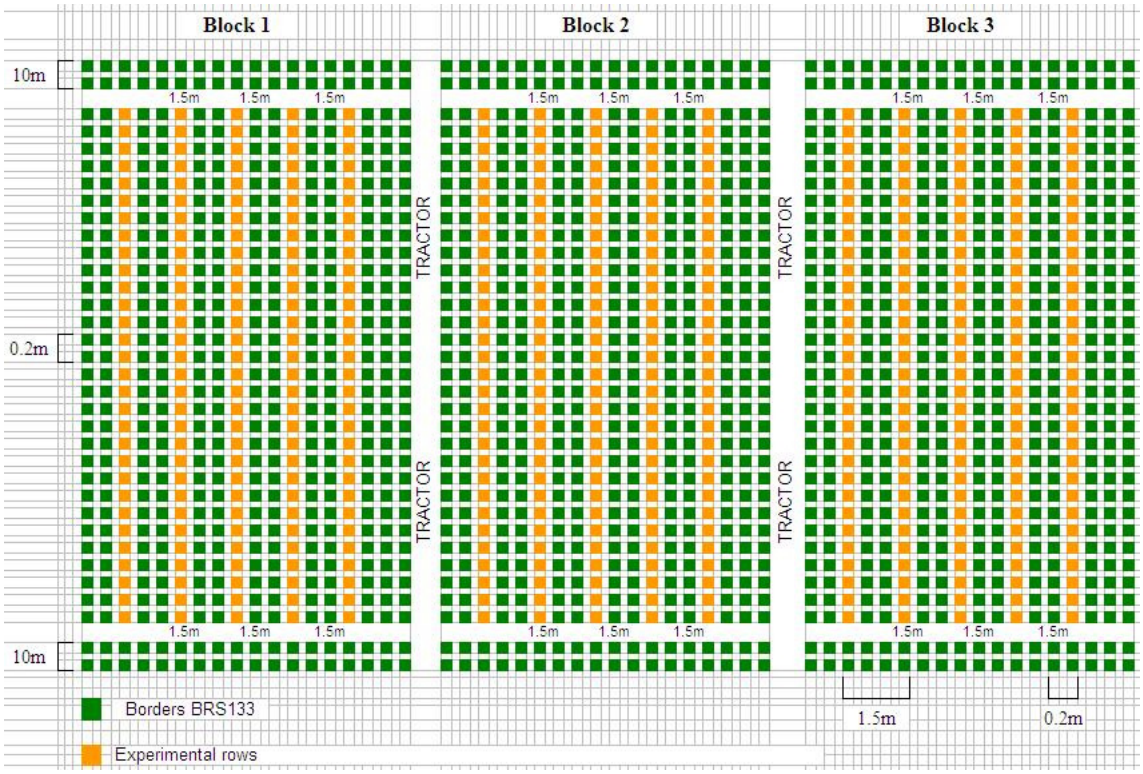


Figure 3 - Layout of the field plot for three blocks

Figure 4 - Field plot

The experimental field was open to natural stink bug infestation by the lack of chemical insect control. The experiment was planted late to increase natural infestation by insect migration at the final maturation or initial harvesting stages. Evaluation of stink bug infestation was carried out weekly, for 18 weeks from R2 to R8 stages (FEHR; CAVINESS, 1977), by the beating-tissue method with random sampling of the experimental area (GAZZONI, 1998). A total of 36 beatings were made per assessment, per day, totalizing 648 samples.

The agronomic traits evaluated were: a) number of days to flowering (NDF) - counted from sowing to first flower (R1 on the scale of FEHR; CAVINESS, 1977); b) plant height at flowering (PHF) - plant height in the beginning of the flowering, measured from the soil-level base of the plant to the apex of the main stem; c) number of days to maturity (NDM) -

counted from the sowing to the date at which the plant exhibit 95% of its pods ripe (R8 on the scale of FEHR; CAVINESS, 1977); d) plant height at maturity (PHM) - measured from the plant base to the apex of the main stem; e) lodging (L) - evaluated at plant maturity by a visual score ranging from 1 to 5, with 1 corresponding to erect plants and 5, to plants laying down; f) agronomic value (AV) - evaluated at maturity by a visual score from 1 to 5, with 1 corresponding to plants with no agronomic value, and 5 to plants with excellent agronomic characteristics; g) grain yield (GY) - total weight of seeds produced per plant, in grams.

The characteristics of insect resistance evaluated were: a) grain filling period (GFP) - obtained by the difference, in days, between reproductive stages R7 and R5 (FEHR; CAVINESS, 1977; PINHEIRO et al., 2005); b) leaf retention (LR) - evaluated in the field by a scale ranging from 1 to 5, where 1 equals normal senescence, and 5, green stems and leaves (non-harvestable) (GODÓI et al., 2002); c) percentage index of pod damage (PIPD) - obtained from an estimate of the percentage of good, intermediate or flat pods, followed by transformation using the formula PIPD = (0.5 x % intermediate pods) + % flat pods (ROSSETTO et al., 1986); d) number of pods per plant (NPP) - counted for each plant after harvest; e) number of seeds (NS) - counted for each individual plant after harvest and seed processing; f) weight of spotted seeds (WSS) - spotted seed weight in grams per plant (spotting caused by insects or colonization by yeast *Eremothecium coryli*); g) weight of a hundred seeds (WHS) - weight from a random sample of 100 seeds after standardization of humidity; h) healthy seed weight (HSW) - total weight of healthy seeds, without apparent stink bug damage, in grams, per plant. The last characteristic was evaluated after harvest and processing the seeds.

2.2.3 Variance and covariance analyses

The analysis of variance was performed for each trait using the GLM procedure of the SAS software (SAS INSTITUTE, 2001). The genetic variances of progenies ($\hat{\sigma}_G^2$) and the phenotypic variance based on means of progenies ($\hat{\sigma}_P^2$) estimated by equating the computed mean squares to their respective expectations, and solved as $\hat{\sigma}_G^2 = (QM_G - QM_E)/R$ and $\hat{\sigma}_P^2 = (QM_G)/R$ (SEARLE; CASELA; MACCULLOCK, 1992). Where, QM_G and QM_E are the mean squares of the $F_{2:3}$ progenies and the error; and R is the number of replicates. Broad-sense heritability coefficient on a progeny-mean basis was estimated as $\hat{h}_G^2 = \hat{\sigma}_G^2 / \hat{\sigma}_P^2$, and

95% confidence intervals were computed for the variance components and heritability estimates, as described in Burdick and Graybill (1992); Knapp, Stroup and Ross (1985) procedures, respectively. Covariance analysis was performed to estimate the genetic correlation coefficients between pairs of characters, following Vencovsky and Barriga (1992). Genetic and phenotypic correlations between characters x and y were computed as: $\hat{r}_{G(xy)} = \hat{C}ov_{G(xy)} \big/ \hat{\sigma}_{Gx}\hat{\sigma}_{Gy}$ and $\hat{r}_{\bar{P}(xy)} = \hat{C}ov_{\bar{P}(xy)} \big/ \hat{\sigma}_{\bar{P}x}\hat{\sigma}_{\bar{P}y}$, respectively; where $\hat{\sigma}_G$ and $\hat{\sigma}_{\bar{P}}$ are the square roots of the estimates of the genotypic and phenotypic variances. The significances of the genetic correlations were computed following Falconer and Mackay (1996), and the significance of phenotypic correlations was determined according to Steel and Torrie (1980).

2.3 Results and Discussion

2.3.1 Stink bug infestation

As illustrated in Figure 5, the increase of the stink bug population in the soybean growth area was greatest during the grain-filling period (R5-R7). The stink bugs may arrive during the vegetative period, increase in population gradually during the reproductive phase, with exponential and accelerated growth at the end of the crop cycle. However, the period between the beginning of fruiting and the point of maximum accumulation of grain dry matter, is the most sensitive period of soybean to sucking insect attacks (PANIZZI et al., 2000). In the end of the R7 period, the stink bug population increased exponentially. The population increase was due to the intense migration of adult insects from recently harvested crops, in search for better shelter, feeding and reproductive conditions. The population reached such a high level that it was necessary to use chemical control to ensure any harvest at the end of the crop cycle. The dramatic effect of the insecticide on stink bug populations at the R8 stage is obvious in Figure 5.

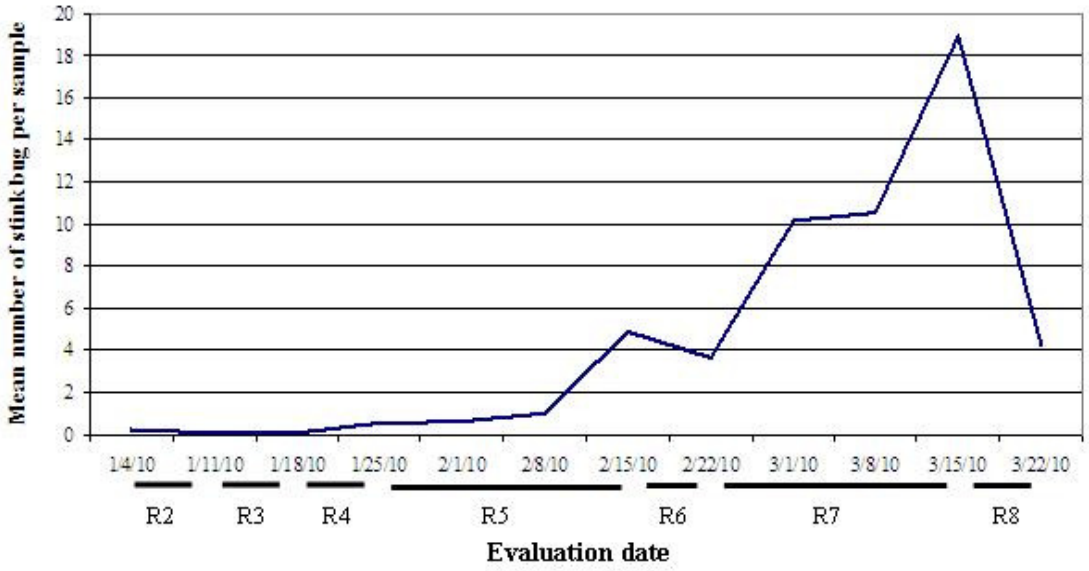


Figure 5 - Mean number of stink bugs per sample in the soybean field during the experiment, in the period comprise between maturity stages R2 and R8

2.3.2 Analysis of variance

The results of analyses of variance for the investigated traits are presented in Tables 1 and 2. Except for the trait lodging, the studied characteristics showed significant differences in the analysis of variance. The coefficient of variation (CV) is a parameter that indicates the data variability to the mean. Values of CV were low, indicating a homogeneous data set for the traits, except for healthy seed weight, which exhibit a high CV value. The observation can be explained by a combined effect, as the trait is directly related to stink bugs attack and distribution in the experimental field.

Cultivar IAC-100 exhibited a smaller mean value for grain filling period (38.48) in comparison to CD-215 (43.05). The means of the progenies ranged from, approximately, 35.80 to 46.53 days, with the average at 41.34 days. The grain filling is the most susceptible soybean developmental stage to stink bugs damage. When this stage is shorter, host evasion is more likely to occur (PAINTER, 1951), and the stink bugs damage is reduced.

The resistant parent IAC-100 displayed a higher mean of the characters number of pods per plant and number of seeds, in comparison to the susceptible cultivar CD-215. Its weight of a hundred seeds and percentage index of pod damage were lower. Progenies mean was intermediate between the parents for the number of pods per plant, number of seeds, and percentage index of pod damage. However, for weight of a hundred seeds, the progenies were

more similar to the susceptible parent CD-215. Miranda et al. (1979) and Lourenção, Rossetto, and Miranda (1985) observed that resistant genotypes produce small seeds and a larger number of pods, allowing a greater percentage of seeds to remain free from stink bugs attack. This pseudo-resistance mechanism is called 'damage dilution' (ROSSETTO; LARA, 1991) and explains the higher mean of grain yield by IAC-100. The mean of progenies for grain yield was closer to IAC-100 than to CD-215.

McPherson, Buss, and Roberts (2007) confirmed the resistance of IAC-100 in a field experiment. These authors evaluated sixty-five soybean breeding lines, containing IAC-100 in their pedigree, for the traits: percentage of seed damage, weight of a hundred seeds, and total yield. From these evaluations, four breeding lines, with either Hutcheson x IAC-100 or IAC-100 x ´V71-370` in their genealogy, produced the lowest stink bug-damaged seeds. Campos et al. (2010) also confirmed that IAC-100 carries pod and seed traits conferring resistance to stink bugs, in support of the findings by McPherson, Buss, and Roberts (2007).

The cultivar IAC-100 also displayed a higher mean for healthy seed weight (24.09g), in comparison to CD-215 (13.71g). The means of the progenies ranged from 9.53g to 57.51g, with an average of 19.13g, closer to IAC-100 than to CD-215. The result shows that IAC-100, and the majority of the progenies were less damaged by the yeast *E. coryli* (previously named *N. coryli*), which is transmitted by stink bugs. According Rossetto et al. (1995), resistance to *E. Coryli* is one of the resistance mechanisms to the stink bug complex. In a recent study, Carlini-Garcia et al. (2011) evaluated 25 soybean populations from intraspecific hibridization between 11 genotypes, and observed that progenies from crosses between CD-215 and IAC-100 had higher percentages of yeast-free seeds (87.67%).

The parent CD-215 was more affected by leaf retention than IAC-100. According to Rossetto and Lara (1991), IAC-100 exhibits normal leaf senescence at harvest, under stink bugs attack. The means of progenies ranged from 1.00 to 2.17, with an average of 1.46, closer to IAC-100. The leaf retention index is considered an important component of soybean resistance to stink bugs, since plants maturing and going through normal senescence processes under insect pressure still allow mechanical harvesting (SOUZA; TOLEDO, 1995).

Table 1 – Summary of the analysis of variance, coefficient of variation (CV), general mean, confidence interval (CI), and variation range (VR) of means of progenies of eight quantitative soybean traits[a]

Sources of variation	DF	Mean Square							
		NDF	PHF	NDM	PHM	GFP	L[b]	AV[b]	LR[b]
Block (B)	2	35.70	189.47	518.45	76.70	193.67	0.05	0.06	0.28
Treatment	230	10.09**	25.73**	47.15**	65.59**	14.07**	0.04	0.02**	0.03*
Residue	460	1.67	4.86	8.98	6.54	4.00	0.04	0.01	0.02
CV (%)		2.82	7.38	2.54	7.42	4.84	16.22	5.25	12.08
[c]Means		45.92	29.89	118.17	34.50	41.34	1.47	2.92	1.46
CI		44.24; 47.60	27.03; 32.75	114.28; 122.06	31.18; 37.82	38.74; 43.94	0.92; 2.16	2.53; 3.33	1.04; 1.96
VR		40.20; 51.27	22.40; 38.60	106.67; 128.73	24.37; 49.00	35.80; 46.53	1.06; 2.47	2.13; 3.87	1.00; 2.17
IAC-100		52.14	26.01	126.43	34.91	38.48	1.34	2.97	1.43
CI		50.46; 53.82	23.15; 28.87	122.54; 130.32	31.59; 38.23	35.88; 41.08	0.81; 2.00	2.58; 3.38	1.01; 1.92
CD-215		43.06	30.54	113.67	32.03	43.05	1.45	2.88	1.60
CI		41.38; 44.74	27.68; 33.41	109.78; 117.56	28.71; 35.36	40.45; 45.65	0.90; 2.13	2.50; 3.29	1.16; 2.11

[a]NDF = number of days to flowering, PHF = plant height at flowering (cm), NDM = number of days to maturity, PHM = plant height at maturity (cm), GFP = grain filling period (days), L = lodging, AV = agronomic value, and LR = leaf retention.
[b]data transformed in $\sqrt{x}$
[c]Mean of L, AV and LR are expressed in original data, without transformation.
* ,**, significant at 5% and 1% level of probability by F test.

Table 2 - Summary of the analysis of variance, coefficient of variation (CV), general mean, confidence interval (CI), and variation range (VR) of means of progenies of seven soybean traits[a]

Sources of variation	DF	Mean Square						
		NPP	NS	GY	WHS	PIPD[b]	WSS	HSW
Block (B)	2	22825.12	178659.78	3081.21	49.52	0.18	98.89	2054.75
Treatment	230	1645.80**	6049.27**	99.58**	5.92**	0.01**	8.89**	71.64**
Residue	460	497.36	1892.29	28.97	1.48	0.00	5.31	21.91
CV (%)		18.52	20.53	23.58	10.68	12.06	62.08	24.47
[c]Means		120.44	211.85	22.83	11.39	26.24	3.71	19.13
CI		91.48; 149.39	155.37; 268.33	15.84; 29.82	9.81; 12.97	19.21; 33.95	0.72; 6.70	13.05; 25.21
VR		74.53; 261.83	120.60; 418.00	12.66; 64.18	8.24; 14.89	14.89; 39.92	1.17; 10.50	9.53; 57.51
IAC-100		148.42	278.45	26.74	9.87	16.61	2.64	24.09
CI		119.46; 177.37	221.97; 334.93	19.75; 33.73	9.97; 13.13	10.85; 23.32	0.00; 5.63	18.02; 30.17
CD-215		89.27	143.09	15.94	11.55	34.22	2.23	13.71
CI		60.31; 118.22	86.61; 199.57	8.95; 22.92	8.29; 11.45	26.48; 42.39	0.00; 5.22	7.63; 19.79

[a]NPP = number of pods per plant, NS = number of seeds, GY = grain yield (g/plant), WHS = weight of a hundred seeds (g), PIPD = percentage index of pods damage, WSS = weight of spotted seeds (g), and HSW = healthy seed weight (g).
[b]data transformed in arc sin $\sqrt{x}/100$
[c]Mean of L, AV and LR are expressed in original data, without transformation.
* ,**, significant at 5% and 1% level of probability by F test

2.3.3 Heritability

According to Brogin, Arias, and Toledo (2003), heritability values are considered small when they are smaller than 30%; intermediate, if between 30% and 60%; and high when higher than 60%. Traits with heritability estimates higher than 30% allow genetic gains by selection at initial endogamy generations; such as, the F_3 generation. For the population evaluated in the current study, heritability was generally observed to range from intermediate to high, showing that, for the most assessed traits, the variance was basically genetic in nature (Table 3). The heritability values were high for eleven traits (number of days to flowering, plant height at flowering, number of days to maturity, plant height at maturity, grain filling period, lodging, number of pods per plant, number of seeds, grain yield, weight of a hundred seeds, and healthy seed weight); intermediate, for two traits (percentage index of pod damage and weight of spotted seeds); and low, for other two traits (lodging and leaf retention). The highest value of heritability was observed for plant height at maturity (90.02), and the lowest values, for lodging (8.25) and leaf retention (20.60). Among the characteristics of resistance, the highest values were for weight of a hundred seeds (75.00) and grain filling period (71.56).

The estimated heritabilities for grain filling period (71.56) and percentage index of pod damage (53.14) were similar to the heritability coefficient in the broad sense found by Godói and Pinheiro (2009) (GFP=74.7% and PIPD=49.6%). According to these authors, the best strategy to obtain stink bug resistant genotypes is selection based on the percentage index of pod damage trait, at early generations (F_3 or F_4), followed by selection for grain filling period, leaf retention and weight of spotted seed, in following generations with higher endogamy levels.

Table 3 - Estimates of phenotypic ($\hat{\sigma}^2_P$), progenies ($\hat{\sigma}^2_G$) and environmental ($\hat{\sigma}^2_E$) variance, heritability ($\hat{h}^2_G$), and confidence interval of fifteen soybean traits[a]

Trait[a]	$\hat{\sigma}^2_P$	$\hat{\sigma}^2_G$	$\hat{\sigma}^2_E$	$\hat{h}^2_G(\%)$
NDF	3.36 (2.82; 4.07)	2.80 (2.27; 3.54)	1.67 (1.47; 1.91)	83.41 (79.15; 86.69)
PHF	8.58 (7.20; 10.39)	6.96 (5.61; 8.85)	4.86 (4.32; 5.61)	81.12 (76.27; 84.85)
NDM	15.71 (13.20; 19.03)	12.72 (10.26; 16.20)	8.98 (7.92; 10.26)	80.96 (76.07; 84.72)
PHM	21.86 (18.36; 26.48)	19.68 (16.21; 24.40)	6.54 (5.77; 7.48)	90.02 (87.46; 91.99)
GFP	4.69 (3.94; 5.68)	3.36 (2.63; 4.44)	4.00 (3.53; 4.57)	71.56 (64.25; 77.17)
L	0.014 (0.012; 0.017)	0.001 (0,000; 1.184)	0.039 (0.034; 0.044)	8.25 (-15.30; 26.37)
AV	0.007 (0.006; 0.009)	0.004 (0.003; 0.006)	0.008 (0.007; 0.009)	63.17 (53.71; 70.44)
LR	0.009 (0.007; 0.011)	0.002 (0.001; 0.008)	0.021 (0.019; 0.024)	20.60 (0.21; 36.28)
NPP	548.60 (460.67; 664.48)	382.81 (297.86; 510.31)	497.36 (438.86; 569.46)	69.78 (62.02; 75.75)
NS	2016.42 (1693.25; 2442.36)	1385.66 (1074.55; 1855.34)	1892.29 (1669.71; 2162.79)	68.72 (60.69; 74.90)
GY	33.19 (27.87; 40.20)	23.53 (18.39; 31.20)	28.97 (25.57; 33.12)	70.90 (63.43; 76.65)
WHS	1.97 (1.66; 2.39)	1.48 (1.17; 1.93)	1.48 (1.30; 1.69)	75.00 (68.58; 79.94)
PIPD	0.003 (0.002; 0.004)	0.001 (0.001; 0.002)	0.004 (0.004; 0.005)	53.14 (41.10; 62.39)
WSS	2.96 (2.49; 3.59)	1.19 (0.77; 2.09)	5.31 (4.68; 6.06)	40.33 (25.01; 52.12)
HSW	23.88 (20.05; 28.92)	16.58 (12.88; 22.13)	21.91 (19.33; 25.04)	69.42 (61.57; 75.46)

[a] NDF - number of days to flowering; PHF - plant height at flowering (cm); NDM - number of days to maturity; PHM - plant height at maturity (cm); GFP - grain filling period (days); L - lodging; AV - agronomic value; LR - leaf retention; NPP - number of pods per plant; NS - number of seeds; GY - grain yield (g/plant); WHS – weight of a hundred seeds (g); PIPD - percentage index of pods damage; WSS - weight of spotted seeds (g); HSW - healthy seed weight (g).
[1] Confidence interval $\alpha =$ 95% level of probability.

2.3.4 Correlation

Correlation coefficients were calculated for each pairwise combination of traits from $F_{2:3}$ means. The highest positive correlations occurred between the variables healthy seed weight and grain yield (r_G = 0.98), and number of pods per plant and number of seeds (r_G = 0.96).

The phenotypic and genotypic correlations between the weight of a hundred seeds and the weight of spotted seeds were significant and positive, corresponding to 0.49 and 0.50,

respectively. Thus, breeders are able to indirectly select lines for stink bug resistance via selection for the weight of a hundred seeds. A similar result was found by Lopes et al., 1997. This value (r_G = 0.50) indicates that the lines with smaller seed sizes were more resistant to stink bug infestation, probably due to the presence of a proportionally greater number of seeds with lower weight of spotted seeds, diluting the effect of the damage, as proposed by Rossetto and Lara (1991). Panizzi et al. (1986) also concluded that genotypes exhibiting the small-seed trait displayed, in general, lower weight of spotted seeds.

Significant genotypic correlation for grain filling period (r_G = 0.66) was solely observed for the weight of a hundred seeds, indicating that shorter grain filling periods result in lower weights of a hundred seeds. The occurrence of shorter grain filling period is desirable because it represents lower exposure of soybean plants to stink bug during its critical period of damage (R5-R7) (PINHEIRO, 1993). The shorter grain filling periods result in impaired seed development, and, thus, small seed size, one of the main characteristics of resistant cultivars to stink bug. A positive genotypic correlation was also found between number of days to maturity and the weight of a hundred seeds (r_G = 0.52), reinforcing discussed previously.

A positive genotypic correlation between healthy seed weight and the number of pods per plant (r_G = 0.83) was also observed, indicating that greater numbers of pods per plant correspond to greater numbers of seeds free of stink bug attack , consequently, leading to larger grain yields. This is showed by the high, positive correlation between healthy seed weight and grain yield (r_G = 0.98).

Results showed that the percentage index of pods damage is significantly and negatively correlated to the number of days to flowering, number of days to maturity, plant height at maturity, agronomic value, number of pods per plant, grain yield, and healthy seed weight. It was positively and significantly correlated only to number of seeds.

Significant genotypic correlations were found for number of days to flowering with plant height at flowering, number of days to maturity, and plant height at maturity. Similar results were found by Alliprandini and Vello (2004), indicating that populations with later flowering show greater height at flowering/maturity, and also had longer life cycles.

Table 4 - Estimate and significance of genotypic $(\hat{r}_G)$ (diagonal above) and phenotypic ($\hat{r}_P$) (diagonal below) correlation coefficients for fifteen soybean traits

Traits [a]	NDF	PHF	NDM	PHM	GFP	L	AV	NPP	GY	WHS	LR	NS	PIPD	WSS	HSW
NDF	..	0.41*	0.78**	0.76**	-0.17^{ns}	0.75**	0.58**	0.73**	0.73**	0.13^{ns}	-0.08^{ns}	0.82**	-0.71**	0.55**	0.72**
PHF	0.33**	..	0.45**	0.85**	0.03^{ns}	0.57*	0.76**	0.56**	0.58**	0.38*	0.07^{ns}	0.53**	-0.32^{ns}	0.81**	0.47**
NDM	0.71**	0.39**	..	0.66**	0.35^{ns}	0.66**	0.39*	0.70**	0.84**	0.52**	0.04^{ns}	0.71**	-0.52**	0.66**	0.82**
PHM	0.69**	0.82**	0.61**	..	-0.12^{ns}	0.87**	0.83**	0.77**	0.74**	0.25^{ns}	0.00^{ns}	0.80**	-0.58**	0.86**	0.64**
GFP	-0.13*	0.05^{ns}	0.42**	-0.08^{ns}	..	-0.28^{ns}	-0.27^{ns}	0.00^{ns}	0.20^{ns}	0.66**	0.43^{ns}	-0.13^{ns}	0.31^{ns}	0.24^{ns}	0.18^{ns}
L	0.18**	0.13*	0.17**	0.25**	-0.04^{ns}	..	0.67**	0.55*	0.50^{ns}	-0.10^{ns}	-0.60^{ns}	0.66**	-0.35^{ns}	0.37^{ns}	0.51^{ns}
AV	0.42**	0.65**	0.24**	0.70**	-0.20**	0.01^{ns}	..	0.80**	0.66**	-0.01^{ns}	0.22^{ns}	0.82**	-0.44*	0.58**	0.62**
NPP	0.58**	0.52**	0.60**	0.68**	0.06^{ns}	0.15*	0.69**	..	0.87**	0.08^{ns}	0.13^{ns}	0.96**	-0.43*	0.74**	0.83**
GY	0.60**	0.53**	0.71**	0.65**	0.23**	0.11^{ns}	0.60**	0.87**	..	0.47**	0.17^{ns}	0.85**	-0.51**	0.75**	0.98**
WHS	0.14*	0.33**	0.56**	0.24**	0.66**	-0.06^{ns}	0.03^{ns}	0.18**	0.49**	..	0.24^{ns}	-0.01^{ns}	0.00^{ns}	0.50**	0.40*
LR	0.00^{ns}	0.05^{ns}	0.10^{ns}	0.03^{ns}	0.18**	0.01^{ns}	-0.01^{ns}	0.11^{ns}	0.06^{ns}	0.11^{ns}	..	-0.01^{ns}	0.24^{ns}	0.11^{ns}	0.16^{ns}
NS	0.64**	0.49**	0.57**	0.69**	-0.05^{ns}	0.16*	0.72**	0.94**	0.87**	0.08^{ns}	0.02^{ns}	..	0.68**	0.83**	-0.29^{ns}
PIPD	-0.50**	-0.24**	-0.31**	-0.41**	0.23**	-0.03^{ns}	-0.39**	-0.34**	-0.47**	-0.01^{ns}	0.16*	0.54**	..	-0.29^{ns}	-0.53**
WSS	0.37**	0.54**	0.55**	0.59**	0.24**	0.06^{ns}	0.36**	0.64**	0.61**	0.49**	0.18**	0.57**	-0.18**	..	0.62**
HSW	0.57**	0.43**	0.64**	0.55**	0.19**	0.12^{ns}	0.58**	0.79**	0.96**	0.40**	0.01^{ns}	0.82**	-0.49**	0.37**	..

[a] NDF - number of days to flowering; PHF - plant height at flowering produção de grãos; NDM - number of days to maturity; PHM - plant height at maturity; GFP - grain filling period; L - lodging; AV - agronomic value; LR - leaf retention; NVP - number of pods per plant; NS - number of seeds; GY - grain yield; WHS - weight of a hundred seeds; PIPD - percent index of pods damage; WSS - weight of spotted seeds; HSW – healthy seed weight.
*;** Significance at the 5% and 1% level of probability by t test, respectively.

2.4 Conclusions

- The increase in stink bug population in the investigated soybean growth area was greatest during the grain filling period (R5-R7).
- Except for lodging, the studied traits show significant differences in the analysis of variance.
- The cultivar IAC-100 and its progenies displayed superior performances to the majority of traits evaluated.
- Heritability was generally observed to be intermediate to high, showing that, for the most assessed traits, the variance was basically genetic in nature. Among the characteristics of resistance, the highest values were found for weight of a hundred seeds and grain filling period, indicating that selection for these traits is likely to be effective for the current population.
- Weight of a hundred seeds presented significant and positive genotypic correlation with weight of spotted seeds and grain filling period.
- Based on the results from analyses of variance, heritability coefficient, and correlation among traits; indirect selection for weight of a hundred seeds associated with healthy seed weight can be successfully employed to obtain stink bug resistant genotypes.

References

ALLIPRANDINI, L.F.; VELLO, N.A. Heritability and correlations among traits in four-way soybean crosses. **Euphytica,** Dordrecht, v. 136, p. 81–91, 2004.

BROGIN, R.L.; ARIAS, C.A.A.; TOLEDO, J.F.F. Genetic control of soybean resistance to brown spot (*Septoria glycines*): First studies. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 3, n. 1, p. 35-44, 2003.

BURDICK; R.K.; GRAYBILL, F.A. **Confidence intervals on variance components**. New York: Marcel Dekker, 1992. v. 127, 211 p.

CAMPOS, M.; KNUTSON, A.; HEITHOLT, J.; CAMPOS, C. Resistance to seed feeding by southern green stink bug, *Nezara viridula* (Linnaeus), in soybean, *Glycine max* (L.) Merrill. **Southwestern Entomologist**, Weslaco, v. 35, n. 3, p. 233-239, 2010.

CARLINI-GARCIA, L.A.; LOURENÇÃO, A.L.; PINHEIRO, J.B.; SCARPELLINI, J.R.; SANTOS, M.daF.; MORAES, L.K.de. Avaliação de populações de soja quanto à produtividade e resistência a percevejos. **Pesquisa & Tecnologia**, Campinas, v. 8, n. 32, p. 1-3, 2011. Disponível em: <http://www.aptaregional.sp.gov.br/artigos>. Acesso em: 3 nov. 2011.

CICEK, M.S.; CHEN, P.; SAGHAI MAROOF, M.A.; BUSS, G.R. Interrelationships among agronomic and seed quality traits in an interspecific soybean recombinant inbred population. **Crop Science**, Madison, v.46, p.1253-1259, 2006.

FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th ed. London: Longman Scientific e Technical, 1996. 464p.

FEHR, W.R.; CAVINESS, C.E. **Stages of soybeans development**. Ames: Iowa State University of Science and Technology, 1977. 11p.

GAZZONI, D.L. Efeito de populações de percevejos na produtividade, qualidade da semente e características agronômicas da soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.33, p.1229-1237, 1998.

GAZZONI, D.L.; MOSCARDI, F. Effect of defoliation levels on recovery of leaf area, on yield and agronomic traits of soybeans. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 4, p. 411-424, 1998.

GODÓI, C.R.C.; PINHEIRO, J.B. Genetic parameters and selection strategies of soybean genotypes resistant to stink bug-complex. **Genetics and Molecular Biology**, Ribeirão Preto, v. 32, p. 328-336, 2009.

GODÓI, C.R.C.; PEREIRA, F.S.; UMENO, F.; AZARA, N.A., LIMA, L.P.M.S.; SILVA, R.P.; OLIVEIRA, A.B.; ARAUJO, I.M.; ZUCCHI, M.I.; PINHEIRO, J.B. Resistência a insetos em populações de soja com diferentes proporções gênicas de genitores resistentes. **Pesquisa Agropecuária Brasileira**, Brasília, v. 32, n.1, p.47-55, 2002.

KNAPP, S.J.; STROUP, W.W.; ROSS, W.M. Exact confidence intervals for heritability on a progeny mean basis. **Crop Science**, Madison, v. 25, n. 1, p. 192-194, 1985.

LOPES, E.C.A.; DESTRO, D.; MONTALVAN, R.; VENTURA, M.U.; GUERRA, E.P. Genetic gain and correlations among traits for stink bug resistance in soybeans. **Euphytica**, Dordrecht, v. 97, p. 161-166, 1997.

LOURENÇÃO, A.L.; MIRANDA, M.A.C. Resistência de soja a insetos: VIII. IAC 78-2318, Linhagem com resistência múltipla. **Bragantia**, Campinas, v. 46, n.1, p. 65-72, 1987.

LOURENCÃO, A.L.; ROSSETTO, C.J.; MIRANDA, M.A.C. Resistência de soja a insetos: IV. Comportamento de cultivares e linhagens em relação a *Hedilepta indicata* (Fabr.). **Bragantia**, Campinas, v. 44, n. 1, p. 149-157, 1985.

McPHERSON, R.M.; BUSS, G.R.; ROBERTS, P.M. Assessing stink bug resistance in soybean breeding lines containing genes from germplasm IAC-100. **Journal of Economic Entomology**, Lanham, v. 100, n. 4, p. 1456-1463, 2007.

McPHERSON, J.E.; McPHERSON, R.M. **Stink bugs of economic importance in America north of Mexico.** Boca Raton, FL. :CRC, 2000. 253p.

MIRANDA, M.A.C.; LOURENÇÃO, A.L. Melhoramento genético da soja para a resistência a insetos: uma realidade para aumentar a eficiência do controle integrado de pragas e viabilizar a soja orgânica. In: CONGRESSO BRASILEIRO DE SOJA E MERCOSOJA, 2., 2002, Londrina. **Anais...** Londrina: Embrapa-Soja, 2002. p. 52-60.

MOURA, M.F.; PINHEIRO, J.B. Selection of soybean genotypes resistant to insects adapted to the cerrados. **Crop Breeding and Applied Biotechnol**, Viçosa,v.2, p. 391-400, 2002.

MIRANDA, M.A.C.; ROSSETTO, C.J.; ROSSETTO, D.; BRAGA, N.R.; MASCARENHAS, H.A.A.; TEIXEIRA, J.P.F.; MASSARIOL, A. Resistência de soja a *Nezara viridula* e *Piezodorus guildinii* em condições de campo. **Bragantia**, Campinas, v.38, n.19 p.181-188, 1979.

PAINTER, R.H. **Insect resistance in crop plants**. Lawrence: The University Press of Kansas, 1951. 520 p.

PANIZZI, A.R.; SLANSKY JÚNIOR, F. Review of phytophagous pentatomids (Hemiptera: Pentatomidae) associated with soybean in the Americas. **Florida Entomologist**, Lutz, v. 68, p. 184-214, 1985.

PANIZZI, A.R.; PANIZZI, M.C.C.; BAYS, I.A.; ALMEIDA, L.A. Danos por percevejos em genótipos de soja com semente pequena. **Pesquisa Agropecuária Brasileira,** Brasília, v. 21, n. 6, p. 571-577, 1986.

PANIZZI, A.R.; McPHERSON, J. E.; JAMES, D.G.; JAVAHERY, M.; McPHERSON, R.M. Stink bugs (Pentatomidae). In: SCHAEFER, C.W.; PANIZZI, A.R. (Ed.). **Heteroptera of economic importance**. Boca Raton, Florida, USA: CRC, 2000. p.432-434.

PINHEIRO, J.B. **Dialelo parcial entre parentais de soja resistentes e suscetíveis a insetos**. 1993. 143 p. Dissertação Mestrado (Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1993.

PINHEIRO, J.B.; VELLO, N.A.; ROSSETTO, C.J.; ZUCCHI, M.I. Potential of soybean genotypes as insect resistance sources. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 5, p. 294-301, 2005.

ROSSETTO, C.J. Breeding for resistance to stink bugs, In: WORLD SOYBEAN RESEARCH CONFERENCE, 4., 1989, Buenos Ayres. **Proceedings…** Buenos Ayres: Asociación Argentina de la Soja Press, 1989. v. 4, p. 2046-2060.

ROSSETTO, C.J.; LARA, F.M. Diluição de dano, mecanismo da pseudo-resistência. In CONGRESSO BRASILEIRO DE ENTOMOLOGIA, 13, 1991, Recife, Pernambuco. **Anais…** Recife: SEB, 1991. p.535.

ROSSETTO, C.J.; IGUE, T.; MIRANDA, M.A.C.; LOURENÇÃO, A.L.R. Resistencia de soja a insetos: VI. Comportamento de genótipos em relação a percevejos. **Bragantia**, Campinas, v.45, n.2, p. 323-335, 1986.

ROSSETTO, C.J.; GALLO, P.B. , RAZERA, L.F.; BORTOLETTO, N.; IGUE, T.; MEDINA, P.F.; TISSELLI FILHO, O.; AQUILERA, V.; VEIGA, R.F.A.; PINHEIRO, J.B. Mechanisms of resistance to stink bug complex in the soybean cultivar 'IAC-100'. **Anais da Sociedade Entomológica do Brasil**, Londrina, v. 24, p. 517-522, 1995.

SAS INSTITUTE, **Statistical analysis software**: SAS 8.2. 2001. Cary-NC, EUA.

SEARLE, S.P.; CASELA, G.; MacCULLOCK. **Variance components**. New York: John Wiley 1992. 501 p.

SOUZA, R.F.; TOLEDO, J.F.F. Genetic analysis of soybean resistance to stink bug. **Brazilian Journal of Genetic**, Ribeirão Preto, v. 18, n. 4, p. 593-598, 1995.

STEEL, R.G.D.; TORRIE, J.H. **Principles and procedures of statistics**. New York: McGraw-Hill Book, 1980. 633 p.

VENCOVSKY, R.; BARRIGA, P. **Genética biométrica aplicada ao fitomelhoramento**. Ribeirão Preto: Sociedade Brasileira de Genética, 1992. 496 p.

3 QTL ANALYSIS OF STINK BUG RESISTANCE AND AGRONOMIC TRAITS IN SOYBEAN

Abstract

The group of stink bug species most frequently causing economic losses in soybean is called stink bug complex and consists of: *Nezara viridula, Piezodorus guildinii,* and *Euchistus heros.* These sucking insects damage the pods, affecting yield and quality of soybean seeds. Soybean breeding programs in Brazil are making an effort to develop cultivars resistant to the stink bug complex. However, the genetic mechanisms underlying the traits are not well-studied. Therefore, the objective of the current project was to map QTL associated to resistance to the stink bug complex, in a population developed from the cross IAC-100 (resistant) x CD-215 (susceptible), considering the strategy of selective genotyping. Segregants were grouped in extreme phenotypic classes for the traits grain filling period (GFP), weight of a hundred seeds (WHS), healthy seeds weight (HSW), and leaf retention (LR). The F_2 generation was genotyped with SSR, SNP, AFLP and TRAP markers, and the $F_{2:3}$ generation was also evaluated for agronomic traits [number of days to flowering (NDF), plant height at flowering (PHF), number of days to maturity (NDM), plant height at maturity (PHM), lodging (L), agronomic value (AV), and grain yield (GY)], as well as for others traits associated with stink bug resistance [percentage index of pod damage (PIPD), number of pods per plant (NPP), number of seeds (NS), and weight of spotted seeds (WSS)]. In the present study, 337 SNP, 28 SSR, 13 TRAP, and 41 AFLP markers were mapped to 20 linkage groups, and the total map length was 1,475 cM. The number of markers by linkage group ranged from 4 to 50, and the distance between two markers, from 0.1 cM to 32.4 cM. The map generated was used to conduct whole genome scans to identify stink bug resistance QTL, with the subset of 84 lines, since genotyping problems occurred in 10 of the 94 lines presented. Fourteen QTL were found using the restricted multiple QTL model and Kruskal-Wallis analyses. These QTL were localized on linkage groups (LGs) C1, A1, C2, M, F, E, J, D2, L, and I. The number of QTL per trait varied from one, for LR and WSS, to four, for GFP, NDF, WHS, PHM, and GY. The majority of the QTL were detected for more than one trait. A major QTL for WHS, with R^2 of 40.6%, was detected on Chr15-LGE. However, the most part of the loci underlying resistance to stink bug consisted of genes with minor effects. Many agronomic traits QTL mapped in this study were consistent with those found in others similar studies.

Keywords: QTL Mapping; Molecular marker; Stink bug; Soybean

ANÁLISE DE QTL DE RESISTÊNCIA A PERCEVEJOS E CARACTERÍSTICAS AGRONÔMICAS EM SOJA

Resumo

O grupo de espécies de percevejos que mais frequentemente causa perdas econômicas para a soja é chamado de complexo de percevejos e é composto por: *Nezara viridula*, *Piezodorus guildinii* e *Euschistus heros*. Estes insetos sugadores causam danos nas vagens, afetando o rendimento e a qualidade dos grãos de soja. Programas de melhoramento de soja no Brasil têm se esforçado para desenvolver cultivares resistentes ao complexo de percevejos. Entretanto, os mecanismos genéticos que controlam estas características ainda são pouco estudados. Assim, o objetivo deste projeto foi mapear QTL associados à resistência ao complexo de percevejos em uma população desenvolvida a partir do cruzamento IAC-100 (resistente) x CD-215 (suscetível) considerando a estratégia de genotipagem seletiva. Segregantes foram agrupados em classes fenotípicas extremas para os caracteres período de enchimento de grãos (PEG), peso de cem sementes (PCS), peso de sementes boas (PSB) e retenção foliar (RF). A geração F_2 foi genotipada com marcadores SSR, SNP, AFLP e TRAP e a geração $F_{2:3}$ foi avaliada também para as características agronômicas [número de dias para o florescimento (NDF), altura da planta no florescimento (APF), número de dias para maturidade (NDM), altura da planta na maturidade (APM), acamamento (AC), valor agronômico (VA) e produtividade de grãos (PG)], assim como para outras características associadas à resistência aos percevejos [índice percentual de dano nas vagens (IPDV), número de vagens por planta (NVP), número de sementes (NS) e peso de sementes manchadas (PSM)]. Nesse estudo, 337 SNP, 28 SSR, 13 TRAP e 41 AFLP foram mapeados em 20 grupos de ligação e o tamanho total do mapa gerado foi de 1.475 cM. O número de marcadores por grupo de ligação variou de 4 a 50 e a distância entre dois marcadores, de 0,1 cM a 32,4 cM. Este mapa foi usado para conduzir a varredura completa do genoma, visando identificar QTL de resistência a percevejo no subconjunto de 84 progênies, já que 10 das 94 apresentaram problemas de genotipagem. Quatorze QTL foram encontrados usando o modelo restrito de múltiplos QTL e análise de Kruskal-Wallis. Estes QTL foram localizados nos grupos de ligação C1, A1, C2, M, F, E, J, D2, L e I. O número de QTL por característica variou de 1 para RF e PSM a 4 para PEG, NDF, PCS, APM e PG. A maioria dos QTL foi detectada para mais de uma característica. Um QTL de efeito maior para PCS, com R^2 de 40,6%, foi detectado sobre Cr15-GLE. Entretanto, a maior parte dos locos controlando resistência a percevejos foi composta de genes com menor efeito. Muitos QTL associados a características agronômicas mapeados neste estudo foram consistentes com QTL encontrados por outros estudos similares.

Palavras-chave: Mapeamento de QTL; Marcador molecular; Percevejo; Soja

3.1 Introduction

Stink bugs are the major pest insect of soybean in Brazil, and three species are economically important: *Nezara viridula* (Linnaeus, 1758), *Piezodorus guildinii* (Westwood, 1837), and *Euschistus heros* (Fabricius, 1794) (Heteroptera: Pentatomidae). These insects feed primarily on pods, causing irreversible direct damage to developing seeds, such as, poor germination, low vigor, reduced oil content, and increased protein content (CORRÊA-FERREIRA; AZEVEDO, 2002). Along with the direct damage to soybean seeds, the transmission of microorganisms by these phytophagous pentatomids may also occur. A disease commonly associated with stink bug injury is the yeast spot disease, caused by *Eremothecium coryli* (previously named *Nematospora coryli* Peglion). In addition, plants attacked by stink bugs may have prolonged leaf retention, resulting in delayed maturation and poor harvest (BOETHEL et al., 2000).

Stink bug control is usually accomplished by pesticide spraying. However, a more economical and environmentally safe control approach would be the use of resistant cultivars. Host resistance to insect pest can occur by three means: antixenosis, antibiosis or tolerance (PAINTER, 1951). Antixenosis resistance affects insect behavior, and is expressed as the non-preference of the insect for certain plants. Antibiosis describes insect resistance, in which feeding on the plant results in disruption of the insect growth, development, or physiology. Tolerance is the ability of plants to withstand or recover from the damage caused by insect activities.

The genetics of insect resistance in soybean can be readily studied by the use of DNA markers in distinct plant materials. This genetic tool has proven to be useful in marker-assisted selection (MAS) for insect-resistant individuals in breeding populations, reducing the need of phenotype-based selection (BOERMA; WALKER, 2005). However, before carrying out MAS, the knowledge of the precise location of quantitative trait loci (QTL) controlling resistance is a prerequisite, since it is essential to know what markers should be used (PANTHEE et al., 2007). More recent studies of soybean resistance to insect pests have concentrated on mapping resistance QTL. Several studies have been published identifying QTL controlling resistance to soybean insect defoliators (RECTOR et al., 2000; TERRY et al., 2000; BOERMA; WALKER, 2005; KOMATSU et al., 2005; ZHU et al., 2006; YESUDAS; SHARMA; LIGHTFOOT, 2010) and aphids (ZHANG; GU; WANG, 2009). However, in the literature a single report on mapping genes/QTL associated with soybean resistance to stink bugs is available. Möller (2010) used an F_2 population, obtained from

crosses of IAC-100 and CD-215, probed with AFLP (Amplified Fragment Length Polymorphism), TRAP (Target Region Amplification Polymorphism), and SNP (Single Nucleotide Polymorphism) markers to identify loci associated with resistance to stink bugs, identifying 14 QTL by the composite interval mapping method. In the current research, the AFLP and TRAP markers identified by Möller (2010), in addition to SSR (Single Sequence Repeat) and novel SNP markers, will be used to fine map stink bug resistance QTL.

Various studies focusing stink bugs in soybean have been conducted in Brazil, and stink bug resistance cultivars, such as: IAC-100, IAC-17 (ROSSETTO, 1989), IAC-23 and IAC-24 (MIRANDA; LOURENCÃO, 2002), have been developed. Although improved cultivars with enhanced resistance can be successfully obtained, only a few studies have been carried out aiming to reveal the genetic mechanisms underlying these traits and the genetic parameters associated with them. Therefore, the objective of the current study was to map QTL associated to resistance to the stink bug complex, in a $F_{2:3}$ population developed from the cross IAC-100 (resistant) x CD-215 (susceptible), considering the strategy of selective genotyping. Segregants were grouped in extreme phenotypic classes for the traits grain filling period (GFP), weight of a hundred seeds (WHS), healthy seeds weight (HSW), and leaf retention (LR).

3.2 Material and Methods

3.2.1 Plant material

A total of 229 $F_{2:3}$ progenies were developed from a cross between the resistant cultivar IAC-100 and the stink bug-susceptible cultivar CD-215. All 229 plants were used for phenotyping. For genotyping, the $F_{2:3}$ phenotypic results (grain filling period, weight of a hundred seeds, healthy seed weight, and leaf retention) were used to choose extreme phenotypic classes for these traits. Thus, a selective genotyping was employed consisting of 12 resistant and 12 susceptible individuals for three of these four traits (GFP, WHS, and HSW), and 11 resistant and 11 susceptible for the trait LR. DNA from these 94 plants, along with DNA from the original parents, was used for genotyping, with AFLP, TRAP, SSR and SNP markers.

3.2.2 Field trial

The parents and $F_{2:3}$ progenies were evaluated for stink bug resistance during the 2009/2010 growing season at an experimental field at ESALQ/USP, in the city of Anhembi, São Paulo, Brazil. The experimental plot was arranged in a randomized complete block design (five F_2 offspring were planted together for each plot), consisting of 231 plots with three replicates (five plants per plot, 231 plots, three replicates for a total of 3,465 plants). Trait evaluation was done for all five plants per plot. Experimental plots were 1.2 m long, 0.4 m apart, and the plants were spaced at 0.2 m, within each plot. The distance between experimental rows was 1.5 m. Two border rows were sown with cultivar BR-133 between experimental rows (at 0.5 m spacing), as well as two or three border rows surrounding each experimental field to encourage an uniform distribution of stink bug populations among the plots during the reproductive period, and to protect the experimental lines from being accidently stepped on during evaluations. The entire experimental field was approximately 70.8 m in length and 29.5 m in width.

The experimental field was open to natural stink bug infestation via the absence of chemical insect control. In order to increase natural crop infestation by insect migration during the final maturation or initial harvesting stages, the experiment was planted in 11/30/2009, several weeks after the neighboring fields. In this manner, the insects would be more attracted to plants of this experiment as the neighboring plants would senesce earlier. Evaluation of stink bug infestation was carried out weekly, from R2 to R8 stages, by the beating-tissue method with random sampling of the experimental area (GAZZONI, 1998). The overall mean of stink bug populations during the observed period was 1.8 insects per meter of row. A maximum population equivalent to nine stink bugs per meter was reached during stage R7.

The agronomic traits evaluated were: a) number of days to flowering (NDF) - counted from sowing to anthesis of the first flower; b) plant height at flowering (PHF) - plant height in the beginning of the flowering, measured from the plant base (at the soil line) to the apex of the main stem; c) number of days to maturity (NDM) - counted from the sowing to the date when 95% of the pods were mature; d) plant height at maturity (PHM) - measured from the plant base to the apex of the main stem; e) lodging (L) - evaluated at plant maturity by a visual score from 1 to 5, with 1 corresponding to erect plant, and 5 to plants laying down; f) agronomic value (AV) - evaluated at maturity by a visual score from 1 to 5, with a 1

corresponding to plants with no agronomic value and a 5 to plants with excellent agronomic characteristics; g) grain yield (GY) - total weight of seeds produced in grams per plant.

The characteristics of insect resistance evaluated were: a) grain filling period (GFP) - obtained by calculating the difference in days between reproductive stages R7 and R5 (FEHR; CAVINESS, 1977; PINHEIRO et al., 2005); b) leaf retention (LR) - evaluated in the field based on a scale ranging from 1 to 5, where 1 was equal to normal senescence and 5, to stems and green leaves (harvesting impaired) (GODÓI et al., 2002); c) percentage index of pod damage (PIPD) - obtained from estimating the percentage of pods in good, intermediate or flat conditions, followed by transformation using the formula PIPD = (0.5 x % intermediate pods) + % flat pods (ROSSETTO et al., 1986); d) number of pods per plant (NPP) - counted for each plant after harvest; e) number of seeds (NS) - counted for each individual plant after the harvest and processing of seeds; f) weight of spotted seeds (WSS) - spotted seed weight (caused by insects or colonization by the pathogenic yeast *Eremothecium coryli*); g) healthy seed weight (HSW) - total weight of healthy seeds, without apparent stink bug damage, in grams per plant; h) weight of a hundred seeds (WHS) - weight from a random sample of 100 seeds after standardization of humidity.

3.2.3 DNA markers analysis

3.2.3.1 DNA extraction

Young leaflets were collected from each parent and individual F_2 plants and then transferred to the laboratory, where they were stored at -20ºC until the tissue could be freeze-dried in a lyophilizer, and subsequently ground to a fine powder. The DNA extraction was performed individually from parents and F_2 individuals by the CTAB (hexadecyltrimethyl ammonium bromide) method (DOYLE; DOYLE, 1990). The DNA was run on 1% (w/v) agarose gels stained with SYBR Safe™ to check DNA quality and to determine the approximate concentration by comparison to λ DNA standards of known concentration.

3.2.3.2 AFLP markers

Identification of AFLP markers was previously conducted (MÖLLER, 2010) and the markers were used in mapping in this Ph.D. project. The AFLP markers were identified essentially according to Vos et al. (1995). About 250 ηg of total genomic DNA was digested in a reaction with 2.5 U of each enzyme *Eco*RI (Invitrogen) and *Mse*I (Invitrogen) in 1X

buffer (50 mM Tris-HCl pH 8.0 and 10 mM $MgCl_2$). The mixture was incubated for 3 hours at 37°C and placed for 15 minutes at 70°C to inactivate the enzyme. The reaction products were visualized on 1% agarose gels stained with SYBR Safe™ for confirmation of the digestion. Subsequently, restriction fragments were ligated to double-strand adaptors *Eco*RI (5´-CTCGTAGACTGCGTACC and 5´-AATTGGTACGCAGTC) and *Mse*I (5´-GACGATGAGTCCTGAG and 5´-TACTCAGGACTCAT) using 0.5 U of T4 DNA ligase (Invitrogen) in a reaction with 1X buffer [10 mM Tris–HCl (pH 7.5), 50 mM KCl, 1 mM dithiothreitol (DTT), 50% (v/v) glycerol], 0.17 mM ATP, 0.83 µM *Mse*I adaptors, and 0.083 µM *Eco*RI adaptors. The mixture was incubated for 2 hours at 37°C, and then for 16 hours at 16°C.

Pre-amplification was performed with primers annealing to either the *Mse*I adaptor (GACTGCGTACCAATTC) or the *Eco*RI adaptor (GATGAGTCCTGAGTAA), with no selective nucleotides, in a total volume of 15 µL containing 2 µL of 6X diluted digestion–ligation DNA template, 1X PCR buffer (50 mM KCl, 10 mM Tris–HCl, pH 8.9), 0.2 mM of each dNTP, 2.0 mM $MgCl_2$, 3.33 µM *Eco*RI primer, 3.33 µM *Mse*I primer and 1U of Taq DNA polymerase (Fermentas). The pre-amplification was performed in a thermocycler Bio-RAD (MyCycler, Bio-RAD, USA) using the following program: 30 cycles of 94ºC for 30 seconds, 60ºC for 30 seconds and 72ºC for 1 minute. The PCR products were verified on a 1% agarose gel stained with SYBR Safe.

The selective amplification was performed with primers annealing to the restriction enzyme adaptors, added of three selective nucleotides at the 3' end. The amplification used 14 primer combinations, as determined by pre-screening conducted in the parents, where 'E' is the *Eco*RI primer and 'M' is the *Mse*I primer: E-ACC/M-CTG, E-ACC/M-CTC, E-ACT/M-CTG, E-ACT/M-CTC, E-ACT/M-CAT, E-AAC/M-CTT, E-AAC/M-CAT, E-AAC/M-CAG, E-AAG/M-CTG, E-AAG/M-CTC, E-AAG/M-CAA, E-AAG/M-CTA, E-AAG/M-CAG, and E-ACA/M-CAT. The selective PCRs were carried out in a total volume of 15 µl, containing 1.5 µl 10X diluted pre-amplification product, 1X PCR buffer, 0.2 mM of each dNTP, 2.0 mM $MgCl_2$, 1.33 µM of an *Eco*RI adaptor+3 primer, 1.33 µM of an *Mse*I adaptor+3 primer, and 1 U of Taq DNA polymerase (Fermentas). PCR was performed with a thermocycler employing the following touchdown program: 13 cycles of 94ºC for 30 seconds, 65ºC (-0.7ºC/cycle) for 30 seconds and 72ºC for 1 minute, followed by 23 cycles of 94ºC for 30 seconds, 56 ºC for 30 seconds, and 72ºC for 1 minute, with a final extension at 72ºC for 8 minutes. The PCR products were separated under denaturing conditions on 7% polyacrylamide gel. Bands were detected by silver staining following Merril et al. (1981) with modifications proposed by

Creste, Tulmann-Neto, and Figueira (2001) (Figure 1). The AFLP fragments were scored based on 0 (absence) and 1 (presence).

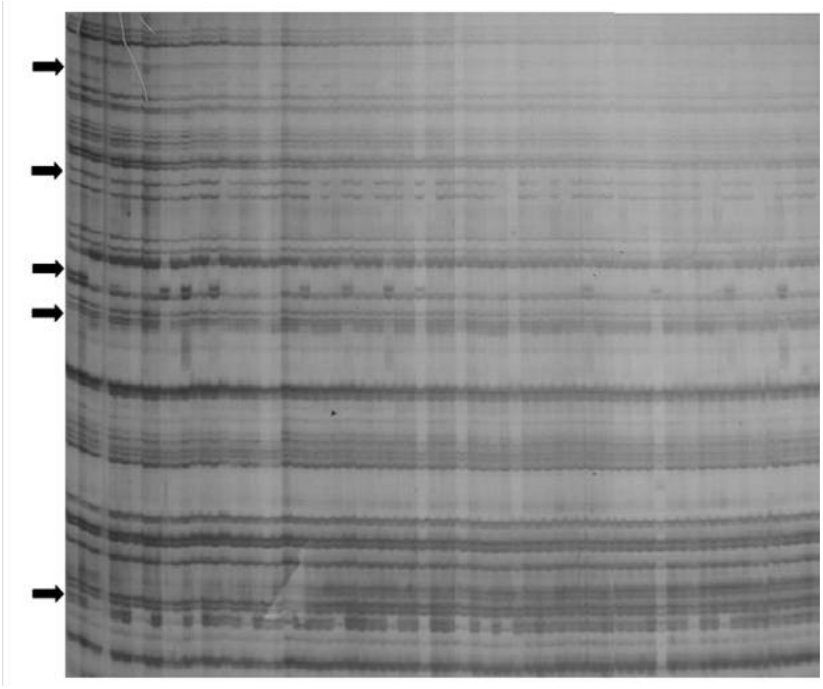


Figure 1 – AFLP polyacrylamide gel. Rows: polymorphic bands

3.2.3.3 TRAP markers

Identification of TRAP markers was conducted previously (MÖLLER, 2010) and the markers were used in mapping in this Ph.D. project. To generate TRAP markers for map construction, eight fixed primers (Table 1) related to genes involved in the metabolic pathways of plant resistance to insects were used in conjunction with five arbitrary primers obtained from the literature (Table 2).

Amplification reactions were carried out in a total volume of 15 µl, containing 100 ηg DNA, 1X PCR buffer, 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, 0.8 µM fixed primer, 0.8 µM arbitrary primer and 1.5 U of Taq DNA polymerase (Fermentas). A Bio-RAD thermocycler (MyCycler, Bio-RAD, EUA) was used with the following program: 94ºC for 2 minutes; 5 cycles of 94ºC for 45 seconds, 35ºC for 45 seconds and 72ºC for 1 minute; 35 cycles of 94ºC for 45 seconds, 50ºC for 45 seconds, 72ºC for 1 minute, with a final extension at 72ºC for 8 minutes. The PCR products were separated under denaturing conditions on 7% polyacrylamide gels. Bands were detected by silver staining following Merril et al. (1981)

with modifications proposed by Creste, Tulmann-Neto, and Figueira (2001). The TRAP fragments were scored based on 0 (absence) and 1 (presence).

Table 1- Fixed TRAP primers (MÖLLER, 2010)

Code	Sequences 5'-3'
KB10	GATGATTGGGTTCTGTGTCG
KB16	ATCCTCTTCCCAAACAAAGC
KB40	TCTGTCGTTTGTTCTCTTGATC
KB44	ACACTGAATGGCTGGTGAAG
KB53	CTCGTTTGGTTCCATTGTTGT
KB56	CACTGAATGGCTTCTGAAGG
KB65	TGTGCCACTTAGGTCACCAG
KB66	AAGCTGTAGCCCGTTCCA

Table 2 - Arbritary TRAP primers

Code	Sequences 5'-3'	Reference
Ga5	GGAACCAAACACATGAAGA	LIU et al. 2005
Odd-26	CTATCTCTCGGGACCAAAC	ZHI-WEI et al. 2007
Arbi 2	GACTGCGTACGAATTGAC	ALWALA et al. 2006
T04	CGTAGTGATCGAATTCTG	LIU et al. 2005
T03	CGTAGCGCGTCAATTATG	LIU et al. 2005

3.2.3.4 SSR markers

Identification of SSR markers was conducted previously (M. MÖLLER, unpublished data) and the markers were used in mapping in this Ph.D. project. A group of 43 SSR markers was selected based on the availability and parental polymorphism. PCR reactions were made in a total volume of 15 µl, using 20 ηg of template DNA, 0.3 µM of Forward primer (IR-Dye 700 or 800), 0.16 µM of Forward primer with M13 tail, 0.2 µM of Reverse primer, 0.15 mM each dNTP, 1X PCR buffer (50 mM KCl; 10 mM Tris–HCl, pH 8.9), 2.0 mM $MgCl_2$, 1X BSA and 1 U of Taq DNA polymerase.

PCR amplifications were performed on a BioRad thermocycler and the cycling conditions consisted of an initial step of 94ºC for 1 minute, followed by 10 cycles of denaturing of 94ºC for 1 minute, annealing of 50ºC for 1 minute, with a decrease of 1ºC per cycle, and extension at 72°C for 1 minute, and 30 cycles of denaturing of 94ºC for 40 seconds, annealing of 40ºC for 40 seconds, and extension at 72°C for 1 minute, followed by a final extension step at 72°C for 10 minute. Electrophoresis separation of the SSR markers was

made with a Li-Cor® 4300 DNA analyzer (LI-COR Biosciences) and the alleles size was estimated using a marker with known molecular weights (Li-Cor IRDye 50-350 bp) (Figure 2).

The SSR markers were scored co-dominantly as 'A' = homozygous for the marker allele from the resistant parent, 'H' = heterozygous for the marker, or 'B' = homozygous for the marker allele from the susceptible parent. The missing or ambiguous data were represented as a dash (-).

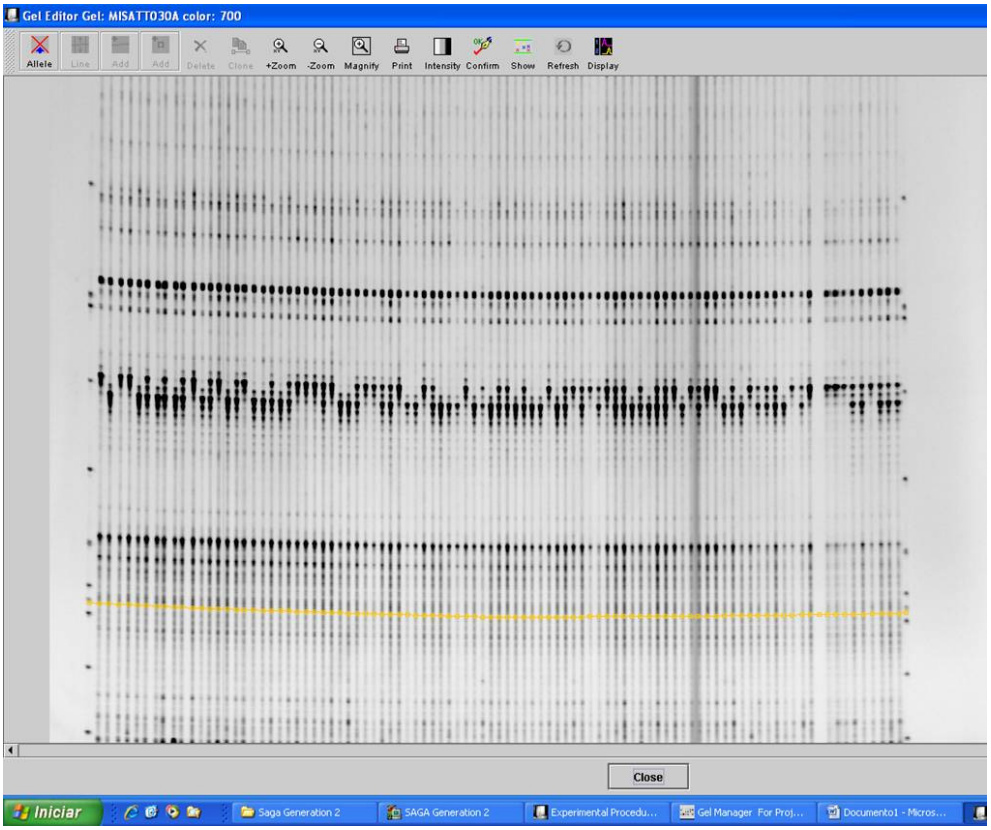


Figure 2 - Electrophoresis of the SSR marker SATT030 with a Li-Cor® 4300 DNA analyzer

3.2.3.5 SNP markers

The resistant (IAC-100) and susceptible (CD-215) parents were used, along with the most resistant and susceptible selected F_2 plants in a 1,536 SNP analysis. The GoldenGate assay was performed at the Cregan and Hyten Laboratories, at the USDA Agricultural Research Center (Beltsville, MD) as the manufacturer's protocol, according to the methodology and principles described in Fan et al. (2003) and Fan et al. (2006). The Illumina BeadStation 500G (Illumina Inc., San Diego, CA) was used to determine genotypic

differences subsequent to the GoldenGate assay. Based on the intensities detected from the two channels for the two respective alleles of each SNP, the automatic allele calling for each locus was done using GenCall software (Illumina Inc., San Diego, CA). A genotype that is homozygous for one or the other SNP alleles will display a signal in either the Cy3 or Cy5 channel, whereas a genotype that is heterozygous will display a signal in both channels (Figure 3). An "A", "B", and "H" genotype coding scheme was used, where 'A' = homozygous for the marker allele from the resistant parent, 'H' = heterozygous for the marker, and 'B' = homozygous for the marker allele from the susceptible parent. A dash (-) was used for ambiguous or missing genotype calls. The data were also manually checked. A SNP marker was considered to be associated to the stink bug resistance when it was identified as polymorphic between the two parents, and the data from the 47 resistant and 47 susceptible F_2 plants clustered tightly with the resistant (IAC-100) and susceptible (CD-215) parents, respectively, considering the traits grain filling period, leaf retention, weight of a hundred seeds, and healthy seeds weight.

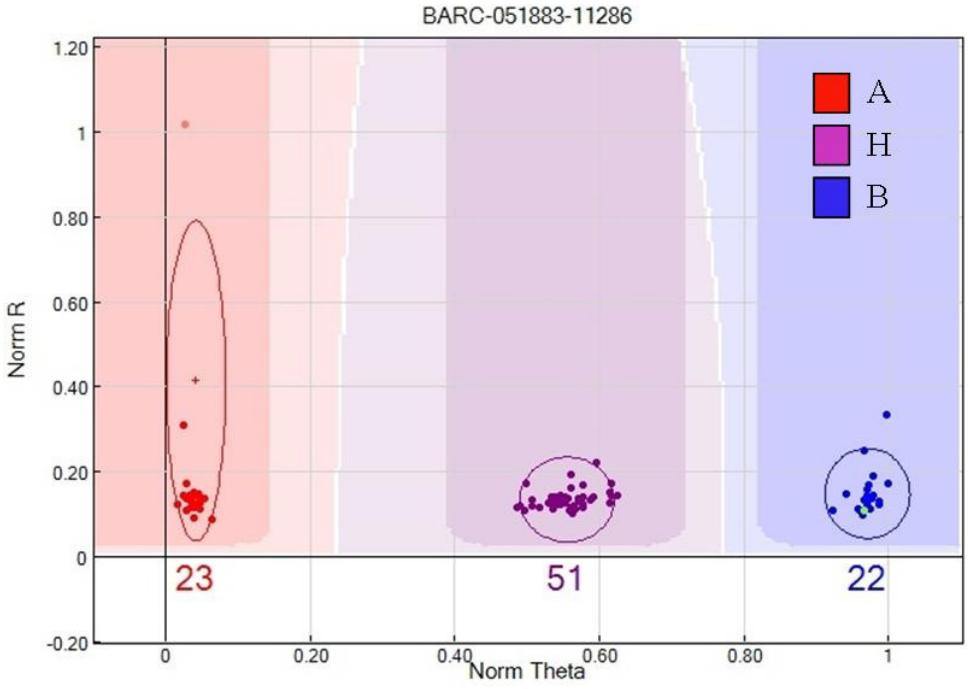


Figure 3 - Cluster separation using the average separation between the two homozygote clusters

3.2.4 Statistical analysis

Broad-sense heritability coefficient on a progeny-mean basis was estimated as $\hat{h}^2_{\bar{G}} = \hat{\sigma}^2_G / \hat{\sigma}^2_{\bar{P}}$, and 95% confidence intervals were computed for the heritability estimates following procedures by Knapp, Stroup, and Ross (1985). Molecular markers and phenotypic data were used to construct a genetic map with lines from IAC-100 x CD-215 population. Linkage analysis was performed with JoinMap® 3.0 (VAN OOIJEN; VOORRIPS, 2001) using the Kosambi mapping function. A logarithm (base 10) of the odds (LOD) score of ≥ 3.0 was used as a threshold to place markers into linkage groups. The maps were drawn using MapChart (VOORRIPS, 2002). Stink bug resistance QTL were mapped by the restricted multiple QTL model (MQM) and Kruskal-Wallis procedures in MapQTL® 4.0 (VAN OOIJEN et al., 2002) for traits with or without normal distribution, respectively. The restricted multiple-QTL model uses markers as cofactors in a multiple-QTL model. The QTL were determined by comparing the cofactor LOD with the LOD generated by a permutation test at 95% significance. Cofactor markers were selected by the automatic cofactor selection option ($\alpha = 0.010$) to identify a set of cofactors. For the Kruskal-Wallis procedure, a p-value ≤ 0.005 for each marker-trait association test was used as recommended by Van Ooijen and Maliepaard (1996).

The marker closest to the peak for each significant QTL was tested in a multivariate ANOVA with PROC GLM (SAS INSTITUTE, 2001). All two- and three-way interactions among significant markers were tested. The R^2 values from the multivariate models are estimates of the total phenotypic variance explained by the genes or QTL and their interactions. These R^2 values were used to estimate the portion of the genotypic variance explained by the QTL using the ratio R^2/h^2 (SCHÖN et al., 1994).

3.3 Results and Discussion

Significant differences among progenies were observed for all traits, except lodging. Therefore, for almost all traits the mean of progenies was intermediate between the resistant and susceptible parent, except for leaf retention, weight of spotted seeds, and weight of a hundred seeds. Heritability estimates varied from a low figure of 0.08 for lodging to a high, of 0.90, for plant height at maturity (Table 3).

Table 3 - Estimates of means, ranges, broad-sense heritabilities (h^2), portion of phenotypic variation explained in quantitative trait loci (QTL) models (R^2) with and without the interaction effect, and portion of genetic variance explained by the QTL and interactions

Trait[a]	Population range	Mean			h^{2b}	R^2 without interaction	R^2 with interaction	R^2/h^{2c}
		Population	IAC-100	CD-215				
Resistance								
GFP	35.80-46.53	41.70	38.48	43.05	0.71 (0.64, 0.77)[d]	0.41	NS[e]	0.57
LR	1.00-2.17	1.39	1.43	1.60	0.21 (0.00, 0.36)	0.13	-	0.62
PIPD	16.23-36.46	26.90	16.61	34.22	0.53 (0.41, 0.62)	0.30	NS	0.57
NPP	77.93-261.83	122.67	148.42	89.27	0.70 (0.62, 0.76)	0.28	NS	0.40
NS	123.13-418.00	208.33	278.45	143.09	0.69 (0.61, 0.75)	0.43	NS	0.62
WSS	1.17-9.86	3.40	2.64	2.23	0.40 (0.25, 0.52)	0.16	-	
WHS	8.26-14.89	11.73	9.87	11.55	0.75 (0.68, 0.80)	0.60	NS	0.80
HSW	10.84-57.51	19.20	24.09	13.71	0.69 (0.61, 0.75)	0.31	NS	0.45
L	1.06-2.47	1.18	1.34	1.45	0.08 (-0.15, 0.26)	-	-	-
Agronomic								
NDF	42.60-50.40	45.42	52.14	43.06	0.83 (0.79, 0.87)	0.59	NS	0.71
PHF	22.67-35.53	30.13	26.01	30.54	0.81 (0.76, 0.85)	0.40	NS	0.49
NDM	109.80-128.73	117.93	126.43	113.67	0.81 (0.76, 0.85)	0.39	NS	0.48
PHM	24.37-49.00	33.78	34.91	32.03	0.90 (0.87, 0.92)	0.48	NS	0.53
AV	2.13-3.86	2.93	2.97	2.88	0.63 (0.54, 0.70)	0.51	NS	0.81
GY	12.66-64.18	22.63	26.74	15.94	0.71 (0.63, 0.77)	0.36	0.52	0.73

[a] GFP - grain filling period (days); LR - leaf retention; PIPD - percent index of pods damage; NPP - number of pods per plant; NS - number of seeds; WSS - weight of spotted seeds (g); WHS – weight of a hundred seeds (g); HSW - healthy seed weight (g); NDF - number of days to flowering; PHF - plant height at flowering (cm); NDM - number of days to maturity; PHM - plant height at maturity (cm); AV - agronomic value; GY - grain yield (g/plant);

[b] h^2, broad-sense heritability.

[c] Portion of genetic variation explained with QTL model and interactions.

[d] Confidence interval at $\alpha = 95\%$ level of probability.

[e] NS – no significant interaction.

Frequency distributions of the means of each trait in the $F_{2:3}$ population were tested for normality by Shapiro and Wilk test statistic *W* (PROC Univariate, SAS, 2001). A probability of $P < 0.05$ was used to indicate absence of fitting. The population segregated quantitatively with normal distribution for grain filling period, number of days to flowering, plant height at flowering, number of days to maturity, and agronomic value. The population did not segregated with normal distribution for leaf retention, percent index of pods damage, number of pods per plant, number of seeds, weight of spotted seeds, weight of a hundred seeds, healthy seed weight, plant height at maturity or grain yield. The absence of normal distributions in some traits was most likely due to the use of the most susceptible and most resistant lines for analyses, which represented extreme scores.

In this study, 337 SNP, 28 SSR, 13 TRAP and 41 AFLP markers were mapped onto 20 linkage groups and the total map length was 1,475 cM (Figure 4). The Chr8-LGA2 was the shortest (34.4 cM), and Chr15-LGE, the longest (119.80 cM). The number of markers by LG ranged from 4 to 50, and the distance between two markers from 0.1 cM (Chr9-LGK: ss107923460 – ss107922655, Chr14-LGB2: ss107923311 – ss107926963, and Chr20-LGI: ss107925758 – ss107925746) to 32.4 cM (Chr5-LGA1: ss107913925 – ss107916669). The map was used to conduct a whole genome scan to identify stink bug resistance QTL with a subset of 84 lines, as 10 of the 94 lines had genotyping problems.

Significant QTL in the restricted MQM and Kruskal-Wallis analyses are reported in Tables 4 and 5, respectively. According to these results, fourteen QTL were detected in ten linkage groups. Among these, three QTL were detected on Chr4-LGC1. The QTL linked to the SNP marker ss107920596 (0.0 cM) was detected for number of days to flowering, number of seeds, healthy seeds weight, and grain yield; the QTL linked to ss107917456 (17.2 cM) was detected for number of days to maturity and explained 14.4% of the variation; and the last QTL linked to ss107914748 (73.2 cM) was detected for grain filling period and weight of a hundred seeds. For all traits, the allele with greater value is from the resistant parent IAC-100.

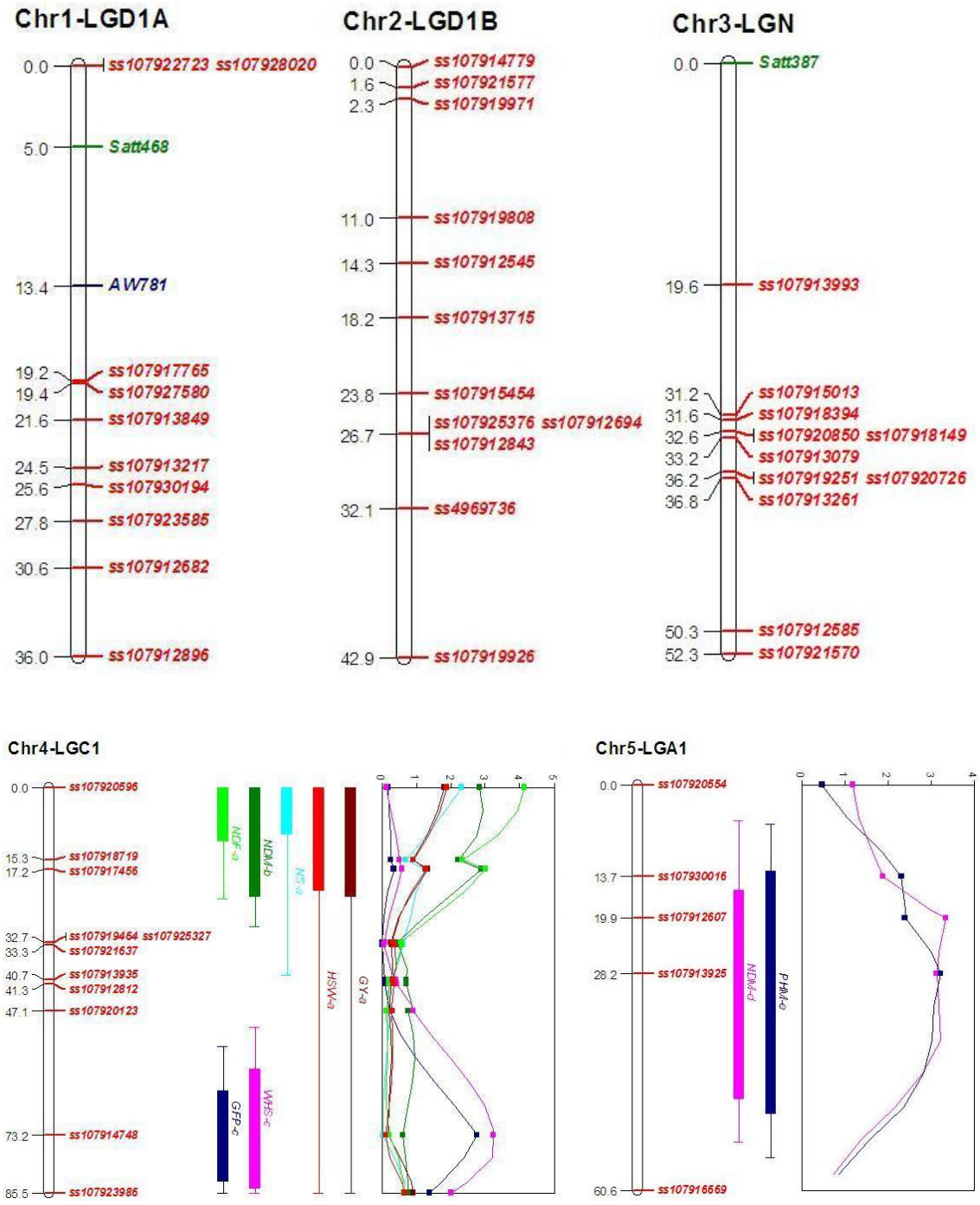


Figure 4 – Mapping of soybean QTL associated to stink bug resistance and agronomic traits using restricted multiple-QTL and Kruskal-Wallis analyses. The LG is represented on the left and the graphic representation of LOD values, on the right. Lettering corresponds to that in Tables 4 and 5

(Continues)

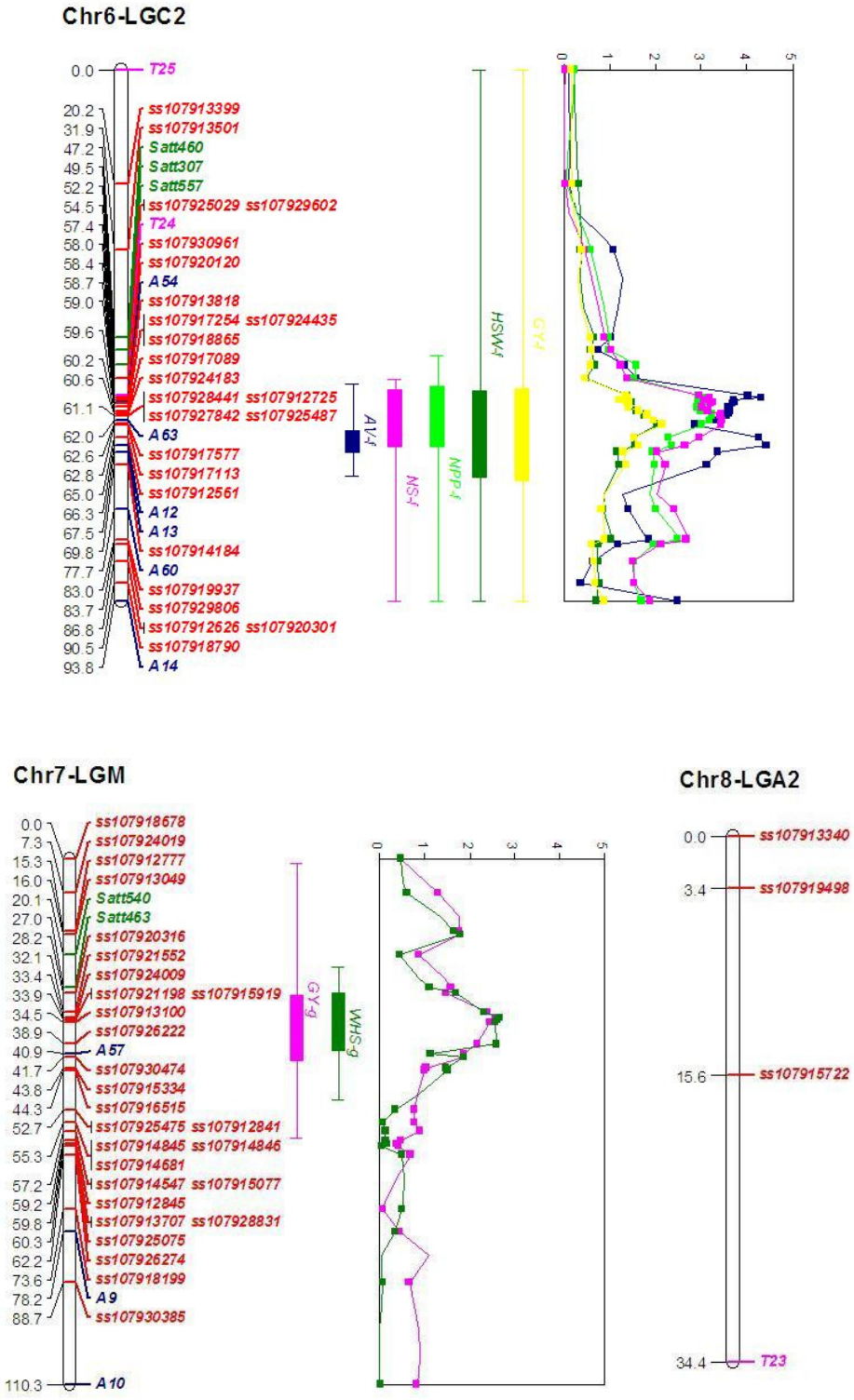


Figure 4 – Mapping of soybean QTL associated to stink bug resistance agronomic traits using restricted multiple-QTL and Kruskal-Wallis analyses. The LG is represented on the left and the graphic representation of LOD values, on the right. Lettering corresponds to that in Tables 4 and 5

(Continued)

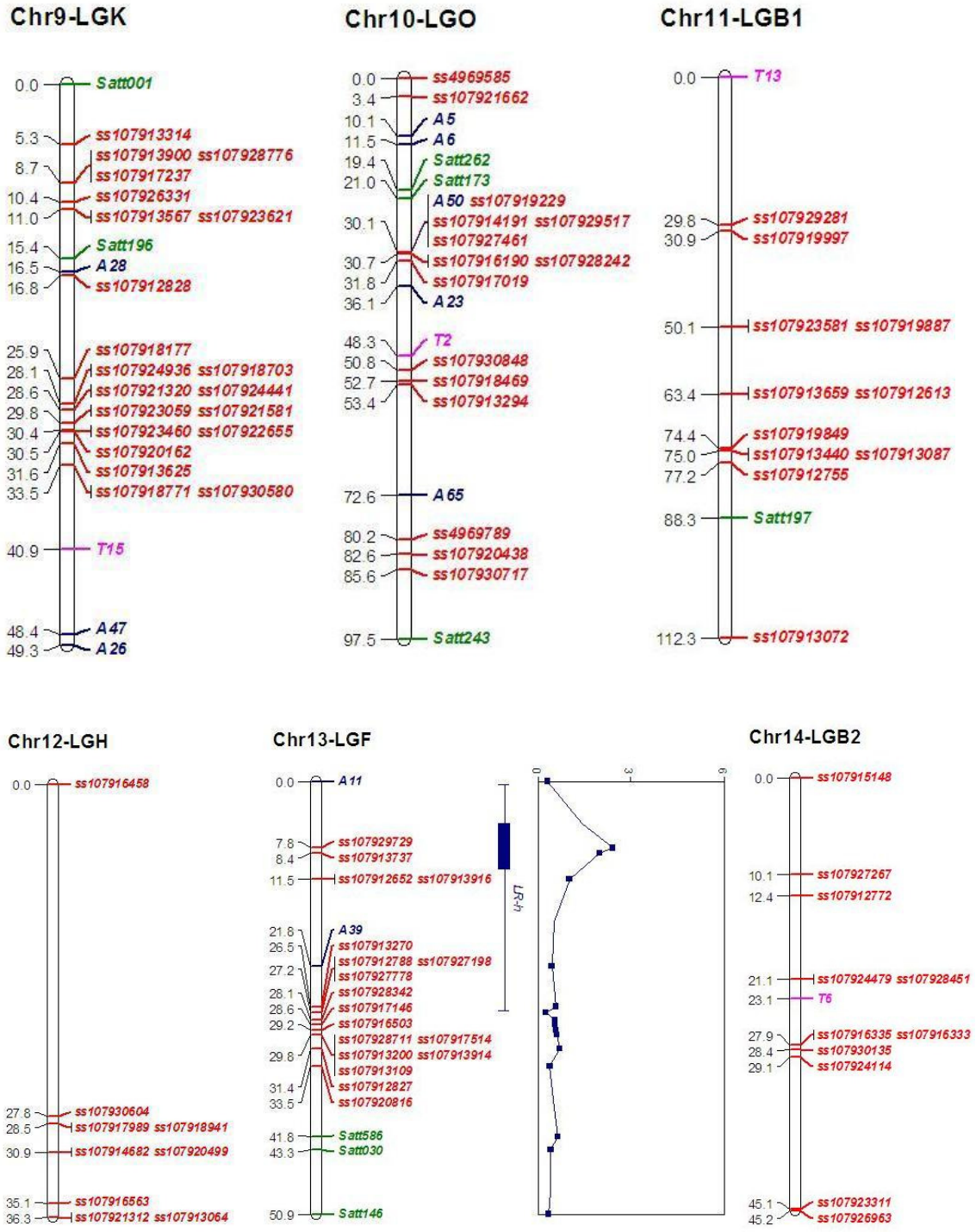


Figure 4 – Mapping of soybean QTL associated to stink bug resistance agronomic traits using restricted multiple-QTL and Kruskal-Wallis analyses. The LG is represented on the left and the graphic representation of LOD values, on the right. Lettering corresponds to that in Tables 4 and 5

(Continued)

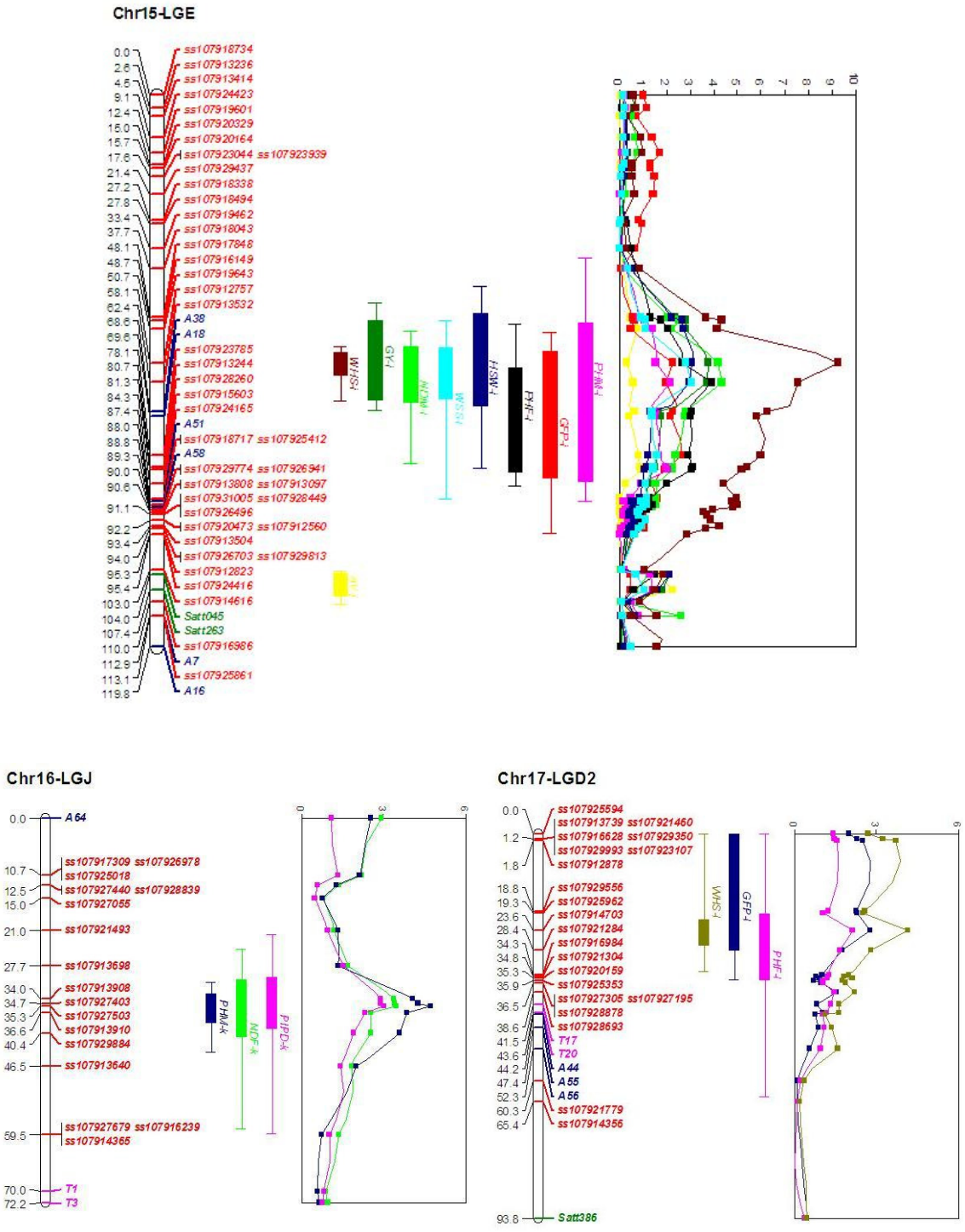


Figure 4 – Mapping of soybean QTL associated to stink bug resistance agronomic traits using restricted multiple-QTL and Kruskal-Wallis analyses. The LG is represented on the left and the graphic representation of LOD values, on the right. Lettering corresponds to that in Tables 4 and 5

(Continued)

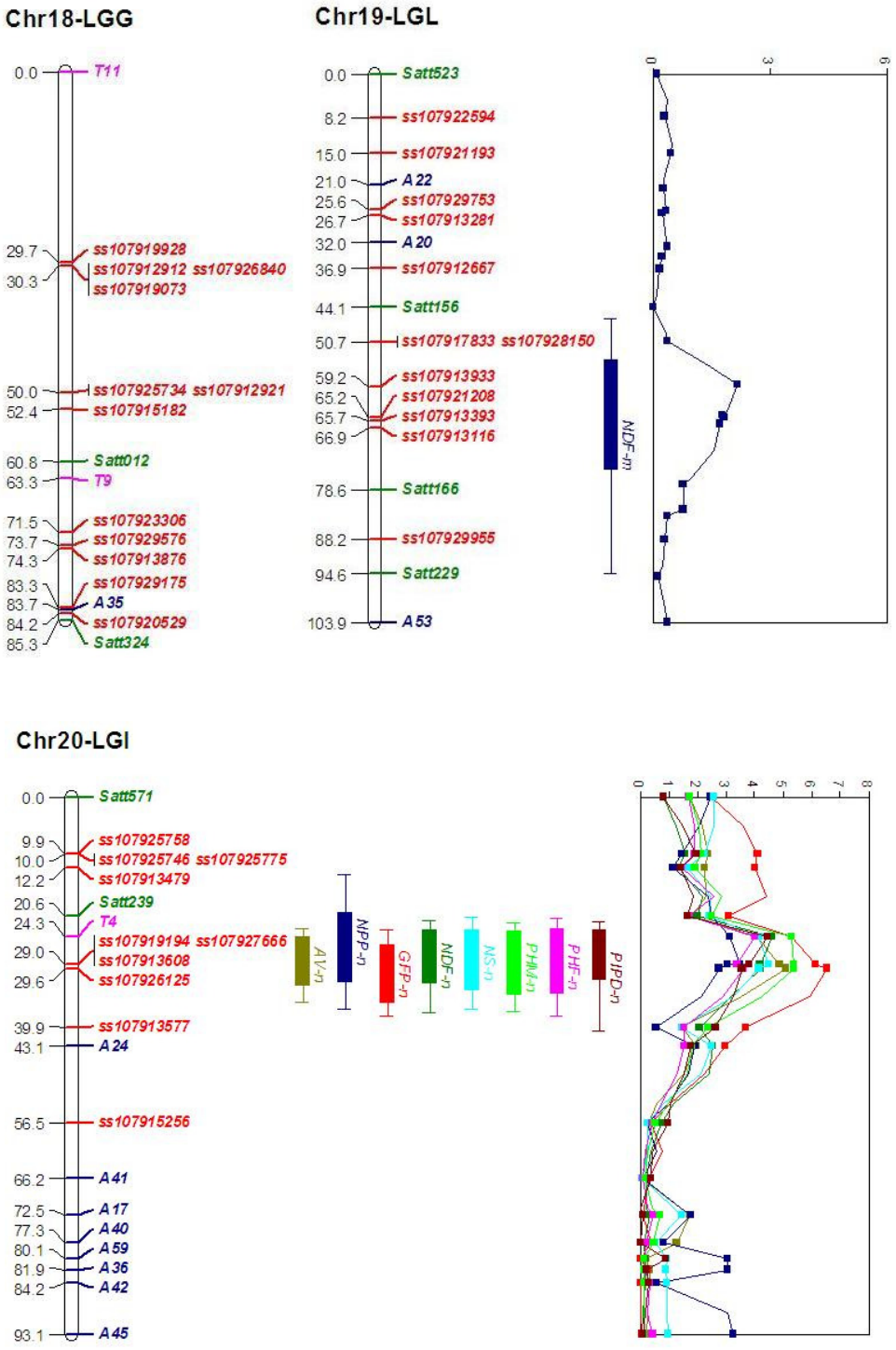


Figure 4 – Mapping of soybean QTL associated to stink bug resistance agronomic traits using restricted multiple-QTL and Kruskal-Wallis analyses. The LG is represented on the left and the graphic representation of LOD values, on the right. Lettering corresponds to that in Tables 4 and 5

(Conclusion)

Table 4 - Quantitative trait loci (QTL) for soybean resistance to stink bug complex based on restrict multiple QTL mapping

Traits	LG[a]	Map position (cm)	Marker nearest to the QTL peak	LOD[b]	R^{2c}	add[d]	QTL position[e]
Resistance traits							
GFP Grain Filling Period	4-C1	73.2	ss107914748	2.74	13.8	1.316	c
	15-E	78.1	ss107923785	2.62	13.1	1.467	i
	17-D2	23.6	ss107914703	2.76	13.8	-1.358	l
	20-I	29.6	ss107926125	6.53	29.5	-1.790	n
Agronomic traits							
NDF Number of Days to Flowering	4-C1	0.0	ss107920596	4.11	20.4	1.040	a
	16-J	35.3	ss107927503	3.46	17.4	-1.121	k
	19-L	66.9	ss107913116	3.14	15.7	1.074	m
	20-I	24.3	T4	4.62	22.5	1.090	n
PHF Plant Height at Flowering	15-E	62.4	ss107913532	3.20	13.6	1.710	i
	17-D2	23.6	ss107914703	2.41	11.3	-0.998	l
	20-I	24.3	T4	4.01	19.6	1.493	n
NDM Number of days to Maturity	4-C1	17.2	ss107917456	2.89	14.4	2.211	b
	5-A1	19.9	ss107912607	3.17	15.7	-2.423	d
	15-E	62.4	ss107913532	4.33	20.8	3.225	i
AV Agronomic Value	6-C2	58.0	ss107930961	5.05	21.8	0.171	f
	15-E	107.4	Satt263	2.78	14.9	-0.001	j
	20-I	29.6	ss107926125	5.62	24.2	0.100	n

[a] Linkage group
[b] Likelihood of odds (LOD) at QTL peak
[c] Percentage of phenotypic variation explained by a QTL
[d] Additive effect of an allele substitution equivalent to the difference between the means of the two homozygous classes divided by two.
[e] QTL position on map

Table 5 - Quantitative trait loci (QTL) for soybean resistance to stink bug complex based on Kruskal-Wallis analysis

Traits	LG[a]	Map position (cm)	Marker nearest to the QTL peak	K[b]	R^{2c}	Add[d]	QTL position[e]
Resistance traits							
LR Leaf Retention	13-F	7.8	ss107929729	11.145	12.5	0.0098	h
PIPD Percentage Index of Pod Damage	16-J	35.3 (34.0-36.6)	ss107927503	14.671	15.8	0.035	k
	20-I	29.0 (29.0-39.9)	ss107919194	15.768	18.5	-0.033	n
NPP Number of Pods per Plant	6-C2	61.1 (58.0-83.0)	ss1079128441	17.442	15.8	16.937	f
	20-I	29.0 (29.0-29.6)	ss107919194	13.501	15.1	12.84	n
NS Number of Seeds per Plant	4-C1	0.0	ss107920596	10.806	11.7	25.392	a
	6-C2	60.6 (58.0-83.0)	ss107924183	16.570	16.8	33.728	f
	20-I	29.0 (29.0-29.6)	ss107919194	16.844	21.4	32.376	n
WSS Weight of Spotted Seeds per Plant	15-E	62.4 (58.1-81.3)	ss107913532	17.278	15.4	1.083	i
WHS Weight of Hundred Seeds per Plant	4-C1	73.2	ss107914748	13.151	16.4	0.970	c
	7-M	38.9 (33.4-38.9)	ss107926222	12.031	13.6	-0.7114	g
	15-E	58.1 (48.1-95.4)	ss107912757	34.293	40.6	1.744	i
	17-D2	23.6 (1.8-28.4)	ss107914703	15.253	20.1	-1.042	l
HSW Healthy Seed Weight	4-C1	0.0	ss107920596	11.807	9.8	2.792	a
	6-C2	60.6 (60.2-62.8)	ss107924183	12.083	9.3	2.983	f
	15-E	58.1 (48.7-62.4)	ss107912757	12.629	15.4	3.917	i
Agronomic traits							
PHM	5-A1	28.2	ss107913925	11.408	15.0	-2.380	e
Plant Height at Maturity	15-E	62.4	ss107913532	11.017	11.0	2.821	i
	16-J	35.3 (34.0-35.3)	ss107927503	17.772	22.5	-3.901	k
	20-I	29.6 (29.0-29.6)	ss107926125	18.353	25.0	3.455	n
GY	4-C1	0.0	ss107920596	10.915	9.3	3.150	a
Grain Yield per Plant	6-C2	62.6	ss107917577	11.809	10.3	3.596	f
	7-M	33.9 (32.1-34.5)	ss107921198	11.53	12.9	-3.641	g
	15-E	58.1 (48.7-58.1)	ss107912757	15.795	18.4	4.990	i

[a] Linkage group
[b] Kruskal-Wallis coefficient (significance level of at least 0.5% of probability).
[c] Percentage of phenotypic variation explained by the QTL
[d] Additive effect of an allele substitution, equivalent to the difference between the means of the two homozygous classes divided by two.
[e] QTL position on map

In order to better understand the QTL for grain filling period and weight of a hundred seeds detected on Chr4-LGC1 at 73.2 cM, it is important to know that the most sensitive period for stink bug damage is between the beginning of pod development and the point of maximum dry matter accumulation in the grain (grain filling period). When the grain filling period is shorter, host evasion occurs and the pods are less damaged by stink bugs. As the resistant cultivar IAC-100 has a shorter grain filling period, its seeds are smaller, resulting in a smaller weight of a hundred seeds. According to Lourencão, Rossetto, and Miranda (1985), those resistant plants have small seeds and a large number of pods, allowing a greater number of seeds free of stink bug attack. This mechanism is called 'dilution of damage'. A possible explanation for the QTL for number of days to flowering, number of seeds, healthy seeds weight, and grain yield mapped on the Chr4-LGC1 at 0.0 cM is that an increase in the number of days to flowering will allow the plants an increased vegetative growth and greater grain yield. The increased number of seeds would allow more seeds to be free of stink bug attack, and increase the overall seed health, weight and yield.

On Chr5-LGA1, two QTL were detected. The first one was detected for number of days to maturity, and is linked to ss107912607 (19.9 cM, R^2 = 15.7). The second one was detected for plant height at maturity, and is linked to ss107913925 (28.2 cM, R^2 = 15.0). These QTL exhibit additive effects of -2.42 and -2.38, showing that they received the higher allele from the susceptible cultivar CD-215.

One QTL was detected on Chr6-LGC2 for the traits: agronomic value, number of pods per plant, number of seeds, healthy seeds weight, and grain yield. Although the LOD peaks were not in the same position in the test for each trait, the peaks were within 4.6 cM, which is sufficiently close to indicate that the same genomic region is most likely controlling the traits. For all traits, the allele with the greater value is from IAC-100. A stink bug resistant soybean plant is expected to produce a higher number of pods per plant, and consequently, a higher number of seeds, resulting in higher weight of healthy seed, grain yield, and agronomic value.

The same proximity of peaks was also observed on Chr7-LGM, Chr15-LGE, and Chr20-LGI. On Chr7-LGM, QTL were detected for weight of a hundred seeds and grain yield within 5 cM of each other. On Chr15-LGE, QTL were detected for grain filling period, plant height at flowering, number of days to maturity, weight of spotted seeds, weight of a hundred seeds, healthy seeds weight, plant height at maturity, and grain yield, within a 20 cM region. On Chr20-LGI, QTL were detected for grain filling period, number of days to flowering, plant height at flowering, agronomic value, percent index of pods damage, number of pods per plant, number of seeds and plant height at maturity, within a 5.3 cM interval. The allele

contributing with greater values for the QTL on Chr7-LGM was from CD-215, whereas the allele of greater values for the QTL on Chr15-LGE was from IAC-100. For the QTL on Chr20-LGI, the additive effect varied from -1.79 to 32.38 among the traits.

A QTL for resistance to corn earworm (*Helicoverpa zea* Boddie) (RECTOR et al., 2000) and two QTL, linked to each other, for resistance to common cutworm (*Spodoptera litura* Fabricius) (KOMATSU et al., 2005) were found on LG M. These QTL and the QTL for weight of a hundred seeds found in this study were mapped at a similar position. A resistance gene to soybean aphid (*Aphis glycines* Matsumura), *Rag1,* was mapped between the SSR markers Satt540 and Satt463 on LGM (LI et al., 2007). Later, Kim et al. (2010) performed fine mapping on the gene in Dowling (PI 548663) and the region containing the *Rag1* locus was defined by SNP markers 46169.7 and 21A. In addition, a QTL for resistance to the soybean cyst nematode (*Heterodera glycines* Ichinohe) was also mapped on linkage group M (WEBB et al., 1995).

On Chr15-LGE, a QTL linked to Satt263 (107.4 cM, R^2 = 14.9%) was detected to be associated with agronomic value, and a single QTL was detected on Chr13-LGF for leaf retention, linked to ss107929729 (7.8 cM).

For Chr16-LGJ, a QTL was detected for the number of days to flowering, percent index of pods damage, and plant height at maturity, and it is linked to ss107927503 (35.3 cM). The additive effect for number the days to flowering was of -1.12, 0.03 for percent index of pods damage, and -3.90 for plant height at maturity. A QTL at the same region was found for resistance to corn earworm (RECTOR et al., 2000). In Chr17-LG2, a QTL linked to ss107914703 (23.6 cM) was detected for grain filling period, plant height at flowering and weight of a hundred seeds. For the three traits, the allele conferring a greater value is from the parent CD-215. On Chr19-LGL, one additional QTL was detected in relation to the number of days to flowering and it is linked to ss107913116 (66.9 cM). This QTL explained 15.7% of the phenotypic variation.

Markers linked to QTL and all possible two-way interactions were added to a multivariate SAS model (Table 3). A single significant interaction was revealed, corresponding to the interaction between ss107921198 and ss107912757 for grain yield. The QTL, along with the significant interaction, explained 52% of the total variation for grain yield. The proportion of the genetic variation explained by the QTL and the interactions ranged from 0.40 to 0.81, showing that for almost all traits the mapped QTL control most of the genetic variance for stink bug resistance (Table 3).

Nine QTL for seed filling period have been described on Soybase, but none at LG C1, E, D2 or I. Therefore, the four QTL detected in the present population are likely to be novel. A QTL for the number of days to flowering near Satt380 located on LG J has been previously detected in a population of 'PI317336' x 'Corsoy' (TASMA et al., 2001) and is located 7.7 cM from ss107927503, the marker used to map the QTL in the current population. Moreover, three QTL for number of days to flowering were mapped on LG L near Satt113 (ORF et al., 1999), A385_1/G173_1 (MANSUR et al., 1993), and Satt166 (ORF et al., 1999) in other populations. The first two QTL are 0.4 cM and 1.3 cM from the marker ss107913116, used to map the QTL in this study. The QTL near Satt166 maps between 65.50 and 67.50 cM. These QTL were detected in the same genome region, further indicating that this is an important region controlling the number of days to flowering in soybean.

For plant height at maturity, the QTL positions detected in the current population are very close to QTL previously detected on LG A1, E, J and I, near Satt382 (KABELKA et al., 2004), Satt321 (SPECHT et al., 2001), B166_1 (LEE et al., 1996), and Satt127 (SEBOLT; SHOEMAKER; DIERS, 2000), respectively. The QTL on LG A1 is 0.78 cM from ss107913925, the QTL on LG E is 6.83 cM from ss107913532, the one on LG J is 5.38 cM from ss107927503, and the one on LG I is 4.75 cM from ss107926125.

Among the QTL previously reported for the number of days to maturity and grain yield, only one of them is located in the same region as the QTL mapped in the current population originated from a cross between IAC-100 and CD-215. On LG A1, one QTL for the number of days to maturity near Satt382 (KABELKA et al., 2004) was detected in a population of 'BRS101' x 'LG82-8379', and it is, at least 5.52 cM, from ss107912607, which is the marker used to map the QTL in the current study. Moreover, a QTL for grain yield on LG M near Satt567/463 (WANG et al., 2004) was detected in a population of 'IA2008 x PI468916', and is located from 33.4 cM to 50.10 cM, whereas ss107921198 detected in the study is located in the range from 32.1 to 34.5 cM, with peak at 33.9 cM.

3.4 Conclusions

Little is known about the mechanisms of insect resistance in soybean. Therefore, in the current research, the QTL underlying fifteen traits were studied in a population of 84 $F_{2:3}$ lines developed from a cross between IAC-100 and CD-215. The parents greatly differ for the traits studied in the present work.

The results showed fourteen QTL underlying six agronomic and eight stink bug-resistance traits. The numbers of QTL detected for each trait varied from one for leaf retention and weight of spotted seeds to four for grain filling period, number of days to flowering, weight of a hundred seeds, plant height at maturity, and grain yield.

The majority of the QTL were detected for more than one trait. For example, two QTL affecting four agronomic traits, and four traits related to stink bug resistance were detected on Chr15-LGE and Chr20-LGI. A major QTL for weight of a hundred seeds, explaining 40.6% of the phenotypic variance, was detected on Chr15-LGE, and its peak is near the SNP marker ss107912757. However, the most part of the loci underlying resistance to stink bug consisted of genes of minor effects. The QTL for weight of a hundred seeds found in this study, on LG M, maps almost at the same position at regions underlying resistance to *Helicoverpa zea* and *Spodoptera litura.*

The results were partially consistent with previous studies for agronomic traits, suggesting that, although the mapping population size was small, real QTL were mapped.

References

BOERMA, H.R.; WALKER, D.R. Discovery and utilization of QTLs for insect resistance in soybean. **Genetica**, Dordrecht, v. 123, p. 181-189, 2005.

BOETHEL, D.J.; RUSSIN, J.S.; WIER, A.T.; LAYTON, M.B.; MINK, J.S., MICHAEL L. BOYD, M.L. Delayed maturity associated with southern green stink bug (Heteroptera: Pentatomidae) injury at various soybean phenological stages. **Journal of Economic Entomology**, Lanham, v. 93, n. 3, p. 707-712, 2000.

CRESTE, S.; TULMANN NETO, A.; FIGUEIRA, A. Detection of single sequence repeat polimorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, New York, v. 19. p. 299-306, 2001.

CORRÊA-FERREIRA, B.S.; AZEVEDO, J.de. Soybean seed damage by different species of stink bugs. **Agricultural and Forest Entomology**, Malden, v. 4, n. 2, p. 145-150, 2002.

DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA fresh tissue, **Focus**, Rockville, v. 12, p. 13 - 15, 1990.

FAN, J.B.; GUNDERSON, K.L.; BIBIKOVA, M.; YEAKLEY, J.M.; CHEN, J.; WICKHAM GARCIA, E.; LEBRUSKA, L.L.; LAURENT, M.; SHEN, R.; BARKER, D. Illumina universal bead arrays. **Methods in Enzymology**, New York, v.410, p.57-73, 2006.

FAN, J.B.; OLIPHANT, A.; SHEN, R.; KERMANI, B.G.; GARCIA, F.; GUNDERSON, K.L.; HANSEN, M.; STEEMERS, F.; BUTLER, S.L.; DELOUKAS, P.; GALVER, L.; HUNT, S.; MCBRIDE, C.; BIBIKOVA, M.; RUBANO, T.; CHEN, J.; WICKHAM, E.; DOUCET, D.; CHANG, W.; CAMPBELL, D.; ZHANG, B.; KRUGLYAK, S.; BENTLEY, D.; HAAS, J.; RIGAULT, P.; ZHOU, L.; STUELPNAGEL, J.; CHEE, M.S. Highly parallel SNP genotyping. **Cold Spring Harbor. Symposia on Quantitative Biology**, Cold Spring Harbor, v.68, p.69-78, 2003.

FEHR, W.R.; CAVINESS, C.E. **Stages of Soybeans Development.** Ames: Iowa State University of Science and Technology, 1977. 11p.

GAZZONI, D.L. Efeito de populações de percevejos na produtividade, qualidade da semente e características agronômicas da soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.33, p.1229-1237, 1998.

GODÓI, C.R.C.; PEREIRA, F.S.; UMENO, F.; AZARA, N.A., LIMA, L.P.M.S.; SILVA, R.P.; OLIVEIRA, A.B.; ARAUJO, I.M.; ZUCCHI, M.I.; PINHEIRO, J.B. Resistência a insetos em populações de soja com diferentes proporções gênicas de genitores resistentes. **Pesquisa Agropecuária Brasileira**, Brasília, v. 32, n.1, p.47-55, 2002.

KABELKA, E.A.; DIERS, B.W.; FEHR, W.R.; LEROY, A.R.; BAIANU, I.C.; YOU, T.; NELSON, R.L. Putative alleles for increased yield from soybean plant introductions. **Crop Science,** Madison, v. 44, n. 3, p. 784-791, 2004.

KIM, K.; BELLENDIR, S.; HUDSON, K.A.; HILL, C.B; HARTMAN, G. L.; HYTEN, D.L.; HUDSON, M.E.; DIERS, B.W. Fine mapping the soybean aphid resistance gene Rag1 in soybean. **Theoretical and Applied Genetics**, Berlin, v. 120, p. 1063-1071, 2010.

KNAPP, S.J.; STROUP, W.W.; ROSS, W.M. Exact confidence intervals for heritability on a progeny mean basis. **Crop Science**, Madison, v. 25, n. 1, p. 192-194, 1985.

KOMATSU, K.; OKUDA, S.; TAKAHASHI, M.; MATSUNAGA, R.; NAKAZAWA, Y. QTL mapping of antibiosis resistance to common cutworm (*Spodoptera litura* Fabricius) in soybean. **Crop Science**, Madison, v. 45, n. 5, p. 2044-2048, 2005.

LEE, S.H.; BAILEY, M.A.; MIAN, M.A.R.; CARTER JÚNIOR., T.E.; ASHLEY, D.A.; HUSSEY, R.S.; PARROTT, W.A.; BOERMA, H.R. Molecular markers associated with soybean plant height, lodging, and maturity across locations. **Crop Science**, Madison v. 36, n. 3, p. 728-735, 1996.

LI, Y.; HILL, C.B.; CARLSON, S.R.; DIERS, B.W.; HARTMAN, G.L. Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. **Molecular Breeding**, Dordrecht, v.19, n. 1, p.25–34, 2007.

LOURENCÃO, A.L.; ROSSETTO, C.J.; MIRANDA, M.A.C. Resistência de soja a insetos: IV. Comportamento de cultivares e linhagens em relação à *Hedilepta indicata* (Fabr.). **Bragantia**, Campinas, v. 44, n. 1, p. 149-157, 1985.

MANSUR, L.M.; LARK , K.G.; KROSS, H.; OLIVEIRA, A. Interval mapping of quantitative trait loci for reproductive, morphological, and seed traits of soybean *(Glycine max* L.). **Theoretical and Applied Genetics**, Berlin, v. 86, n. 8, p. 907-913, 1993.

MERRIL, C.R.; GOLDMAN, D.; SEDMAN, S.A.; EBERT, M.H. Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. **Science**, Washington, v. 211, p. 1437-1438, 1981.

MIRANDA, M.A.C.; LOURENÇÃO, A.L. Melhoramento genético da soja para a resistência a insetos: uma realidade para aumentar a eficiência do controle integrado de pragas e viabilizar a soja orgânica. In: CONGRESSO BRASILEIRO DE SOJA E MERCOSOJA, 2., 2002, Londrina. **Anais...** Londrina: Embrapa-Soja, 2002. p. 52-60.

MÖLLER, M. **Mapeamento de locos de resistência quantitativa da soja ao complexo de percevejos**. 2010. 86 p. Dissertação (Mestrado em Genética e Melhoramento de Plantas) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2010.

ORF, J.H.; CHASE, K.; JARVIK, T.; MANSUR, L.M.; CREGAN, P.B.; ADLER, F.R.; LARK, K.G. Genetics of soybean agronomic traits: I. Comparison of three related recombinant inbred populations. **Crop Science,** Madison, v. 39, n. 6, p.1642–1651, 1999.

PAINTER, R.H. **Insect resistance in crop plants.** Lawrence: The University Press of Kansas, 1951. 520p.

PANTHEE, D.R.; PANTALONE, V.R.; SAXTON, A.M.; WEST, D.R.; SAMS, C.E. Quantitative trait loci for agronomic traits in soybean. **Plant Breeding,** Malden, v. 126, n. 1, p. 51-57, 2007.

PINHEIRO, J.B.; VELLO, N.A.; ROSSETTO, C.J.; ZUCCHI, M.I. Potential of soybean genotypes as insect resistance sources. **Crop Breeding and Applied Biotechnology**,Viçosa, v. 5, p. 294-301, 2005.

RECTOR, B.G.; ALL, J.N.; PARROTT, W.A.; BOERMA, H.R. Quantitative trait loci for antibiosis resistance to corn earworm in soybean. **Crop Science**, Madison, v. 40, p. 233-238, 2000.

ROSSETTO, C.J. Breeding for resistance to stink bugs, In: WORLD SOYBEAN RESEARCH CONFERENCE, 4., 1989, Buenos Ayres. **Proceedings…** Buenos Ayres: Associacion Argentina de la Soja Press, 1989. v. 4, p. 2046-2060.

ROSSETTO, C.J.; IGUE, T.; MIRANDA, M.A.C.; LOURENÇÃO, A.L. R Resistencia de soja a insetos: VI. Comportamento de genótipos em relação a percevejos. **Bragantia**, Campinas, v.45, n.2, p. 323-335, 1986.

SAS INSTITUTE, **Statistical analysis software**: SAS 8.2. 2001. Cary-NC, EUA.

SCHÖN, C.C.; MELCHINGER, A.E.; BOPPENMAIER, J.; BRUNKLAUS- JUNG, E., HERRMANN, R.G.; SEITZER, J.F. RFLP mapping in maize: Quantitative trait loci affecting testcross performance of elite European flint lines. **Crop Science**, Madison v. 34, n. 2, p. 378–389, 1994.

SEBOLT, A.M.; SHOEMAKER, R.C.; DIERS, B.W. Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. **Crop Science**, Madison, v. 40, n. 5, p. 1438-1444, 2000.

SPECHT, J.E.; CHASE, K.; MACRANDER, M.; GRAEF, G.L.; CHUNG, J.; MARKWELL, J. P.; GERMANN, M.; ORF, J.H.; LARK, K.G. Soybean response to water: A QTL analysis of drought tolerance. **Crop Science**, Madison, v. 41, n. 2, p. 493-509, 2001.

TASMA, I.M.; LORENZEN, L.L.; GREEN, D.E.; SHOEMAKER, R.C. Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. **Molecular Breeding**, Dordrecht, v. 8, n. 1, p.25–35, 2001.

TERRY, L.I.; CHASE, K.; JARVIK, T.; ORF, J.; MANSUR, L.; LARK, K.G. Soybean quantitative trait loci for resistance to insects. **Crop Science**, Madison, v. 40, p. 375-382, 2000.

VAN OOIJEN; MALIEPAARD, 1996. MapQTL version 3.0: Software for the calculation of QTL positions on genetic maps. Plant Genome IV Abstracts, Web site: http://www.intl-pag.org.

VAN OOIJEN, J.W.; VOORRIPS. R.E**. JoinMap 3.0, software for the calculation of genetic linkage maps.** Plant Research International, Wageningen, The Netherlands, 2001.

VAN OOIJEN, J.W.; VOORIPS, R.E.; JANSEN, T.C.; MALIEPAARD, C. **MapQTL 4.0, software for the calculation of QTL positions on genetic maps.** Plant Research International, Wageninigen, 2002. CD-ROM.

VOORRIPS, R.E. MapChart: software for the graphical presentation of linkage maps and QTLs. **The Journal of Heredity**, Oxford, v. 93, n. 1, 2002. Disponível em: http://jhered.oxfordjournals.org/cgi/reprint/93/1/77. Access in: July 17,2011.

VOS, P.; HOGERS, R.; BLEEKER, M.; REIJANS, M.; LEE, T. van der; HORNES, M.; FRIJTERS, A.; POT, J.; PELEMAN, J.; KUIPER, M.; ZABEAU, M. AFLP: a new technique for DNA fingerprinting. **Nucleic Acids Research**, Oxford, v. 23, p. 4407-4414, 1995.

WANG, D.; GRAEF, G.L.; PROCOPIUK, A.M.; DIERS, B.W. Identification of putative QTL that underlie yield in interspecific soybean backcross populations. **Theoretical and Applied Genetics**, Berlin, v. 108, n. 3, p.458-467, 2004.

WEBB, D.M.; BALTAZAR, B.M.; RAO-ARELLI, A.P.; SCHUPP, J.; CLAYTON, K.; KEIM, P.; BEAVIS, W.D. Genetic mapping of soybean cyst nematode race-3 resistance loci in the soybean PI 437.654. **Theoretical and Applied Genetics**, Berlin, v.91, n. 4, p. 574-581, 1995.

YESUDAS, C.R.; SHARMA, H.; LIGHTFOOT, D.A. Identification of QTL in soybean underlying resistance to herbivory by Japanese beetles (*Popillia japonica*, Newman). **Theoretical and Applied Genetics**, Berlin v. 121, n. 2, p. 353-362, 2010.

ZHANG, G.; GU, C.; WANG, D. Molecular mapping of soybean aphid resistance genes in PI 567541B. **Theoretical and Applied Genetics**, Berlin, v. 118, n. 3, p. 473–482, 2009.

ZHU, S.; WALKER, D.R.; BOERMA, H.R.; ALL, J.N.; PARROTT, W.A. Fine mapping of a major insect resistance QTL in soybean and its interaction with minor resistance QTLs. **Crop Science**, Madison, v. 46, p. 1094-1099, 2006.

4 EXPRESSION PROFILE OF SOYBEAN RESISTANCE RESPONSE TO STINK BUG

Abstract

The genome-wide expression approach is especially promising for cases where defense-related genes are difficult to map precisely, such as when resistance is governed by many genes and each gene has a minor contribution to the defense response. Therefore, the present study takes advantage of modern genomic techniques to perform a transcript profile analysis of stink bug resistant and susceptible soybean cultivars to evaluate the plant defense response to the pest insects. In the present work, a pilot experiment was developed to determine the timing of the most significant gene expression changes in soybean in response to stink bug feeding using microarray analysis. Subsequently, a field trial of insect feeding was carried out to identify genes associated with the resistance responses by comparing the differential expression between resistant and susceptible soybean cultivars with RNA-seq. Additionally, gene expression from resistant and susceptible plants in the absence of stink bugs was examined to identify genes exhibiting constitutive differential regulation. For the pilot experiment, pods of the highly susceptible soybean cultivar CD-215 were infested with insects of the species *Nezara viridula* and *Euschistus heros* in the greenhouse. Pods were collected at time points 5.5, 21, 24, and 41 hours post-infestation. Pods used for time points 5.5 and 21 hours were infested with insects previously starved for 24 hours. Microarrays were used to characterize gene expression changes at the four time points with both species of insects. A clear differential expression was observed for the samples at time points 21, 24, and 41 hours infested with *P. guildinii*. For the field experiment, pods were infested with *P. guildinii* and samples were taken at time points 0, 8, 24, and 46 hours. However, due to limited financial resources, sequencing was only carried out for RNA obtained from the 24 hour time point, since it generated the best results in the pilot experiment along with the 41 hour sample. As the field experiment was carried out at the end of September in 2011, samples were not ready to be sent for RNA-seq analysis until the beginning of November. Thus, RNA-seq results will be available after the submission of the thesis. The resistant cultivar showed 39.4% of the genes with higher expression levels (fold change > 2.0 up) and 11.68%, with lower expression (fold change > 2.0 down) in comparison to the susceptible cultivar in the RNA-seq analysis performed on pods without treatment.

Keywords: cDNA microarray; RNA-seq; Gene expression; Plant resistance; Stink bug; Soybean

4 PERFIL DE EXPRESSÃO DA RESPOSTA DE RESISTÊNCIA DA SOJA A PERCEVEJOS

Resumo

A abordagem de expressão gênica de genomas completos é especialmente promissora para casos onde genes relacionados à defesa são difíceis de mapear precisamente, tal como quando a resistência é governada por muitos genes e cada gene apresenta uma pequena contribuição para a resposta de defesa. Assim, o presente estudo tira proveito de técnicas modernas de genômica para fazer uma análise do perfil transcricional de cultivares de soja resistente e suscetível a percevejo, visando avaliar a resposta de defesa da soja a estes insetos-praga. Neste trabalho, um experimento piloto foi desenvolvido para determinar o tempo em que ocorrem as maiores mudanças na expressão gênica em soja em resposta à alimentação de percevejo, usando análises de microarranjo. Em seguida, foi realizado um estudo de alimentação do inseto no campo para identificar genes associados à resposta de resistência pela comparação da expressão gênica diferencial entre materiais de soja resistente e suscetível, empregando o sequenciamento de RNA (RNA-seq). Adicionalmente, foi investigada a expressão gênica de plantas resistente e suscetível na ausência dos percevejos para identificar genes com diferenças constitutivas de expressão. Para o experimento piloto, vagens da cultivar de soja suscetível ao percevejo, CD-215, foram infestadas com insetos das espécies *Nezara viridula* e *Euschistus heros* em casa de vegetação. Foram coletadas vagens às 5.5, 21, 24 e 41 horas após infestação. Vagens usadas para 5.5 e 21 horas foram infestadas com insetos que estavam previamente sem alimentação por 24 horas. Microarranjos foram usados para caracterizar as mudanças na expressão gênica nos quatro tempos de infestação e com ambas as espécies de insetos. Foi observada uma clara expressão diferencial para as amostras infestadas às 21, 24 e 41 horas com *P. guildinii*. Para o experimento de campo, vagens foram infestadas com *P. guildinii* e amostras coletadas 0, 8, 24 e 46 horas após infestação. Entretanto, devido aos recursos financeiros para o projeto serem limitados, somente RNA da amostra coletada após 24 horas foi sequenciado, já que 24 e 41 horas foram os tempos de infestação que geraram os melhores resultados no experimento-piloto. Como o experimento de campo foi realizado no fim de setembro de 2011, as amostras não estavam prontas para serem enviadas para a análise de RNA-seq até o início de novembro. Assim, os resultados de RNA-seq estarão disponíveis após a submissão da tese. A cultivar resistente apresentou 39,4% dos genes com expressão induzida (duas vezes maior) e 11,68% de genes com expressão reprimida (duas vezes menor) em comparação com a cultivar suscetível na análise de RNA-seq realizada em vagens na ausência de infestação.

Palavras-chave: Microarranjo de cDNA; Sequenciamento de RNA;Expressão gênica;Resistência vegetal; Percevejo; Soja

4.1 Introduction

Recent advances in sequencing technologies have greatly enhanced the understanding of the structural and functional aspects of plant genomes (VARSHNEY; GRANER; SORRELLS, 2005). Genomics is the study of genomes, which includes the study of all coding and non-coding regions of a genome. An important element of genomics is the determination of the expression profile of all genes in response to environmental and developmental factors. These great advances in genomics are providing valuable information to assist researchers in modern breeding in the pursuit of improved crop performance.

The advent of lost-cost, whole-genome sequencing has led to the sequencing of the complete genome of hundreds of organisms. In order to take advantage of the large and rapidly increasing sequence information, new technologies are constantly being developed. Several tools and analyses now exist to conduct expression profiling to aid the investigation of how an organism responds to environmental changes, or in response to a specific stimulus. For example, one of the first methods to examine genome-wide gene expression was the closed-format hybridization technology called 'microarrays', which consists of thousands of oligonucleotides or cDNA corresponding to all known genes of an organism (SCHENA et al. 1995).

Microarrays are typically used to directly compare samples from two different sources (control and treatment) to determine gene expression ratios. In two-color microarray technology, two RNA samples are labeled separately with distinct fluorescent tags [for example, cyanine 3 and cyanine 5 (Cy3, Cy5)], hybridized to a single microarray and scanned to generate fluorescent images from the two fluorescent channels. A two-color graphical overlay can then be used to visualize genes that are activated or repressed, and specific softwares are used to express numerical data of the fluorescence levels, which are proportional to mRNA levels (STEARS; MARTINSKY; SCHENA, 2003).

Although very informative, microarray technology is limited by the requirement of having the known sequence of a gene to be able to quantify the mRNA level of that gene. On the other hand, the recent development of next-generation massively parallel sequencing of RNA, called “RNA-seq” (RNA sequencing), allows quantitative determination of RNA levels, comparable to microarrays, with the added benefit that the entire transcriptome is surveyed without any a priori knowledge of the transcribed regions (WILHELM; LANDRY, 2009). In general, in an RNA-seq assay a population of RNA is converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule is then sequenced

in a high-throughput manner to obtain short sequences. The sequencing reads are typically 30-400bp, depending on the DNA sequencing technology. Following sequencing, the resulting reads are aligned to a reference genome or transcript, or assembled *de novo* without the genomic sequence to produce a genome-scale transcription structure and/or level of expression for each gene (WANG; GERSTEIN; SNYDER, 2009).

The use of high-throughput gene expression technologies can bring insights into how an organism responds to abiotic and biotic stresses, as hundreds or even thousands of genes may respond, representing many metabolic reactions. For example, plants have been shown to express many thousands of genes during interactions with pests, whereas under non-stressed conditions, these same genes might not be expressed. Several studies have provided clues as to how plants are defending themselves against insect feeding. Li et al. (2008) identified 140 genes in soybean showing specific responses to aphids (*Aphis glycines* Matsumura), including genes related to cell wall, defense, DNA/RNA, secondary metabolism, signaling, and other processes. Casteel et al. (2008) showed a decrease in the levels of defense-associated genes in soybean grown under elevated CO_2, and that the plants were more susceptible to Japanese beetle (*Popillia japonica*) herbivory. Kempema et al. (2007) identified that distinct defense pathways were triggered by phloem feeding and chewing insects in *Arabidopsis thaliana*.

The genome expression approach is especially promising for cases where defense-related genes are difficult to map precisely, such as, when resistance is governed by many genes (QTLs, or quantitative trait loci) and each gene has a minor contribution to the defense response (CLOUGH; VODKIN, 2004). For example, Calla et al. (2009) identified several defense-related genes, such as PR5, located within a disease resistance QTL and also significantly differentially expressed.

The present study takes advantage of modern genomic techniques to perform a transcript profile analysis of soybean cultivars that are either resistant or susceptible to stink bugs in order to evaluate the soybean defense response to these pest insects. Stink bugs are considered the most important pest insect attacking soybean (*Glycine max*) in Brazil and are also becoming a serious problem in the United States. The group of stink bug species most frequently causing economic losses in Brazil is called the “stink bug complex” and is composed of three species: *Nezara viridula* (Linnaeus)*, Piezodorus guildinii* (Westwood) and *Euchistus heros* (Fabricius) (Heteroptera: Pentatomidae). The most susceptible period of the soybean life cycle to stink bug damage is the pod-filling period, when the stink bugs extract nutrients from seeds using their piercing and sucking mouth. The impact of stink bug feeding may be significant losses in soybean yield, quality, and germination potential. Moreover, stink

bugs are also associated with the transmission of the yeast *Eremothecium coryli* (previously named *Nematospora coryli* Peglion), and other diseases (CORRÊA-FERREIRA; AZEVEDO, 2002).

In this present work, a pilot experiment was developed to determine how early one can detect gene expression changes in soybean in response to stink bug feeding using microarray analysis. Subsequently, an insect feeding study in the field was carried out to identify genes associated with resistance responses by comparing the differential gene expression between resistant and susceptible soybean employing RNA-seq. Additionally, we examined gene expression patterns in resistant and susceptible plants in the absence of stink bugs to identify genes with constitutive expression differences between the cultivars used.

4.2 Material and Methods

4.2.1 Plant material

Soybean cultivars IAC-100 and CD-215 were used in the present study. IAC-100 is partially resistant to stink bugs, is classified in maturity group 8, and was developed by Instituto Agronômico de Campinas in São Paulo, Brazil. It is derived from the cross between IAC 78-2318 and IAC-12 (ROSSETTO et al., 1995). The line IAC 78-2318 is a source of multiple-resistance genes against soybean-attacking insects (LOURENÇÃO; MIRANDA, 1987). The cultivar CD-215, a protected plant developed by COODETEC (Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda), from the state of Paraná, Brazil, was used as the stink bug susceptible parent and is derived from a cross between the OC88-127 and OC90-234.

4.2.2 Pilot experiment

The stink bug susceptible soybean cultivar CD-215 was grown during the second semester of 2010 in a stink-bug-free greenhouse at the University of Illinois, Urbana campus, to early pod fill (stage R5). Insects of the species *Nezara viridula* and *Piezodorus guildinii* were obtained from USDA-ARS entomologist Dr. Clint Allen, in Stoneville, Mississippi, and allowed to feed on pods within attached small bags surrounding an individual pod (Figure 1). The manipulations for both insect species were similar.

Pods were collected at time points 5.5, 21, 24, and 41 hours post-infestation. Pods used from time points 5.5 and 21 hours were infested with insects previously starved for 24 hours. An empty bag was wrapped around a pod at a similar growth stage, on a separate plant, to serve as controls for each insect-infested sample from the same time point. Biological replicates were not done, because the goal of the pilot experiment was simply to determine how early after feeding one could detect changes in gene expression using the soybean oligo microarrays. Harvested pods were placed immediately in liquid nitrogen after removal from the plants to rapidly stop gene expression activity. Subsequently, samples were stored at -80ºC until RNA extraction.

Figure 1 - A nylon bag wrapped around a single pod containing stink bugs

4.2.3 Field trial

The field study was conducted at the USDA-ARS experiment station in Stoneville, Mississippi, in the USA in collaboration with entomologist Dr. Clint Allen. Seeds of IAC-100 and CD-215 were planted in four separate 6-foot by 24-foot plots, each covered with a 6-foot high cage covered with a nylon mesh to restrict the entry of herbivores into the plots (Figure 2). However, it was difficult to completely eliminate tiny insects, such as whiteflies, during plant development.

Figure 2 - Cages covering field plots in Stoneville, Mississippi, USA

Plantings were on June 2nd, 20th, and 30th 2011 to ensure that plants of distinct genotypes and maturity groups were at the same developmental stage at the time of sampling. One of the cages was blown over by strong winds resulting in many lodged plants. However, this fact did not derail the experiment as plants recovered and produced healthy seed pods, and treated and control plants were obtained from within the same plot. Infestations were all done on the same day, when a sufficient number of plants of both genotypes had seed pods at the R5 stage. On that day, September 22nd, 2011, the cages were removed and individual pods were covered with small nylon bags with or without two insects, similar to our pilot experiment. Infestations were performed at each of the four plots, each plot representing one experimental block consisting of three control and three infested pods per treatment. The treatments were 0, 8, 24, and 46 hours post-infestation. Adult stink bugs of the species *Piezodorus guildinii* used in the study were obtained from soybean fields in southeast Texas and fed with green beans for two to three days prior to the day of infestation. The experiment schedule is described in Table 1. Pods were harvested at each time point and placed immediately in a plastic pouch and into a dry ice/ethanol bath, and then on dry ice to remain frozen until transferred to -80ºC, several hours later. During the approximate 10-hour return trip to Urbana, samples were kept deep frozen within a large ice chest of dry ice, and then, stored again at -80ºC in Urbana, Illinois until RNA extraction.

Table 1 - Summary of the field experiment			
	Date		
Hours	**09/22/2011**	**09/23/2011**	**09/24/2011**
8a.m.	Insects placed in vials without food (2 insects/vial)		
9 a.m.			
10 a.m.	Set up T8h		
11 a.m.			Collected T46h
12 p.m.	Set up T24h	Collected T24h	Verified stink bugs still alive
1p.m.	Set up T46h	Verified stink bugs still alive	
2p.m.	Collected T0h		
6p.m.	Collected T8h		
7p.m.	Verified stink bugs still alive		

4.2.4 RNA extraction

For RNA extraction, pods frozen in liquid nitrogen were ground to a powder and total RNA isolated by treatment with Fruit-mate (Takara, Japan) followed by extraction with TRIzol reagent (Invitrogen, Carlsbad, CA) in tubes containing Phase Lock Gel-Heavy (Brinkmann Instruments, Inc., Westbury, NY). Subsequently, the supernatants were transferred to new centrifuge tubes and extracted with chloroform. Following chloroform extraction, RNA in the aqueous phase was precipitated with isopropyl alcohol, washed with ethanol, and suspended in water. RNA samples were checked using a BioAnalyzer 2100 (Agilent Technologies, Palo Alto, CA) to verify RNA integrity, and the concentrations were determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE).

4.2.5 Microarray analysis

RNA from the pilot study was analyzed using soybean oligonucleotide microarrays from Dr. Lila Vodkin (University of Illinois). RNA labeling was conducted essentially according to Zou et al. (2005), with some modifications. Each purified total RNA (RNeasy kit, Qiagen, Valencia, CA) was reverse-transcribed into cDNA incorporating aminoallyl-dUTP (aa-dUTP; Sigma, St. Louis, MO). The reverse transcription reaction was performed using 30 µg of total RNA concentrated to an 11 µl volume under vacuum centrifugation. A sample of 11 µl RNA and 2 µl oligo $d(T)_{25}$ (3 µg/µl) was mixed and heated at 70°C for 10 minutes and immediately chilled on ice for 10 minutes. Following centrifugation for 30 s at

13,000 × *g*, 29.5 µl of the reverse transcription master mix (2.5 mM each of dATP, dCTP, and dGTP; 1.5 mM dTTP; 1.0 mM aminoallyl-dUTP [all nucleotides from Sigma]; 0.1 M dithiothreitol [DTT] [Invitrogen]; 1000 units of Superscript III reverse transcriptase [Invitrogen]; and 5× FS reaction buffer [Invitrogen]) were added to a total volume of 42.5 µl. After a 3-hour incubation at 50°C, RNA was hydrolyzed by addition of 13 µl 1M NaOH and 13 µl 0.5 M EDTA (pH8.0) and incubated for 15 minutes at 65°C. Samples were neutralized with 13 µl 1 M HCl. Newly synthesized cDNA was purified using QiaQuick PCR purification columns (Qiagen, Valencia, CA, U.S.A.) with modified Tris-free buffers (50 ml of wash buffer: 10 ml of nuclease-free H_2O, 40 ml of ethanol; 50 ml of elution buffer: 0.19 ml of 1M K_2HO_4, 0.01 ml of 1 M KH_2PO_4, 49.80 ml of nuclease-free H_2O). Successful reverse transcription was verified by measuring light absorption at 260 nm with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies) and observing a cDNA yield corresponding to at least 2% total RNA. Samples were dried by vacuum centrifugation, followed by suspension in 4.5 µl of 0.1 M sodium carbonate buffer (1 M $NaHCO_3$ made fresh monthly and adjusted to pH 9.0 with concentrated HCl). For the addition of 4.5 µl of Cy3 or Cy5 monofunctional dye [#PA23001 and #PA25001 (Amersham Pharmacia)], each tube of dye was suspended with 73 µl of anhydrous dimethyl sulfoxide [(Aldrich, Milwaukee, WI, U.S.A.), pipetted in single-use aliquots and stored at –80°C], and added to each sample. Labeling reactions were incubated for 1 h at room temperature in the dark. Unincorporated dyes from Cy3 and Cy5 reactions were removed using QiaQuick PCR purification columns (Qiagen). The column eluates were combined and the fluorescent labeling was verified using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies). Light absorption at 260, 550, and 650 nm was measured to determine the concentration of cDNA, incorporation of Cy3 and Cy5, respectively.

Hybridization, washing, scanning, image analysis, and data acquisition were performed as previously described (ZOU et al., 2005). The purified labeled probe (concentrated to 40 µl), 4 µl Poly $d(A)_{50}$ (25 µg/µl) was added and the solution was placed into just-boiled (95 to 99°C) water for 3 minutes. After centrifugation at 13,000 × *g* for 30 seconds, 44 µl of 2× hybridization solution (50% formamide, 10× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 0.2% sodium dodecyl sulfate, preheated to 42°C) was added. The solution was then pipetted onto one slide containing half the oligonucleotide set, and subsequently, the slide containing the remaining half was gently placed on top. This 'sandwich' hybridization was incubated in a sealed chamber within a 42°C water bath for three

days, and then, washed (HEGDE et al., 2000) and scanned for fluorescence emission using a Perkin-Elmer (Foster City, CA, U.S.A.) ScanArray Express Microarray Scanner.

Images were analyzed with the software GenePix version 5.0 (Axon Instruments, Union City, CA, U.S.A.) to generate fluorescent intensity and background values for each spot in both Cy3 and Cy5 channels. Single spots or areas of the array with obvious blemishes were flagged and excluded from subsequent analyses. Images in the TIFF format and GenePix result data files were uploaded into GeneTraffic (Iobion, La Jolla, CA, U.S.A.) for data storage, visualization and analysis. The response variable studied was the log2 of the normalized, background-subtracted, fluorescence intensity. Hierarchical clustering was applied to both axes using the similarity metrics method based on Pearson correlation with the flavor correlation (centered), as implemented in the program Cluster (M. Eisen; http://www.eisenlab.org/eisen/?page_id=42) (EISEN et al., 1998). The results were analyzed with TreeView (M. Eisen; http://www.eisenlab.org/eisen/?page_id=42) (EISEN et al., 1998).

4.2.6 RNA sequencing

Purified (RNeasy kit, Qiagen, Valencia, CA) total RNA from pods (see methods in 4.2.1.4) were treated by DNAseI in a reaction containing 28 μg RNA, 10 x buffer, and DNaseI, and incubated at 37 ºC for 20 minutes. The products were purified using RNeasy MinElute Cleanup Kit (50) (Qiagen, Hilden, Germany). For the investigation of constitutively differentially expressed genes in the absence of insect treatment, RNA from pods of greenhouse-grown plants was used and RNA from three replicates were mixed for each cultivar and submitted to RNA-seq analysis. For the field study, RNA was extracted from individual pods, with three pods being collected per treatment per time point per plot. The three replicates of each treatment at each time point were pooled prior to submitting to RNA-seq analysis.

The high-throughput RNA sequencing (RNA-seq) was performed at the Keck Center of the University of Illinois, using the Illumina HiSeq2000 sequencer. For sequencing, the cDNA from different samples were marked with tag sequences to differentiate one sample from the other by sorting with the added tag. These RNA-seq libraries were prepared with Illumina's 'TruSeq RNA-seq Sample Prep kit'. The libraries were quantified by qPCR and pooled in equimolar concentrations of two samples. It was possible to sequence two samples per lane using a TruSeq SBS sequencing kit version 5, with a sequence target of 100 nucleotides per sequence, and about 40-50 million sequence reads per sample. Analyses were

made using the pipeline version 1.8. The sequence reads were aligned to the soybean reference genome (http://www.phytozome.net/soybean) using the software Tophat, which uses the ultra high-throughput short-read aligner Bowtie. The aligned RNA-seq reads were run through a second software called Cufflinks that assembles the alignments into a parsimonious set of transcripts. Subsequently, Cufflinks software estimates the relative abundances of the transcripts based on how many reads support each one, taking into account biases in the library preparation protocols. The end-result was a table of *Glycine max* (Glyma) IDs and frequencies of how often each gene was represented by transcripts in the samples.

4.3 Results and Discussion

4.3.1 Pilot experiment

Microarrays were used to characterize gene expression changes for four time points (5.5, 21, 24 and 41 hours after infestation) with both species of insects (*P. guildinii* and *N. viridula*). Schematically, pods with feeding stink bugs were compared to those without insects at each time point (Figure 3). Two sets of soybean cDNA microarray slides, each with 19,000 gene oligonucleotides, were used totaling 38,000 screened genes.

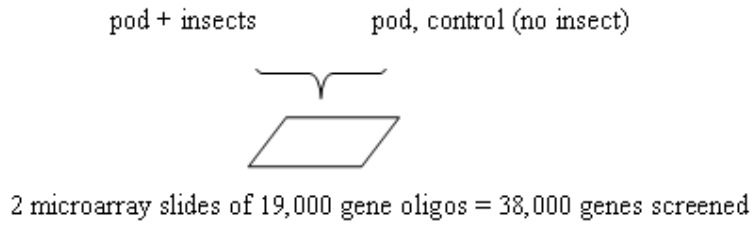


Figure 3 - Schematic representation of the microarray analyses to determine gene expression changes for each time point investigated

For each time point, fluorescent cDNA probes, labeled with Cy5 dye, were prepared from each mRNA sample extracted from pods infested with insects. Reference cDNA probes, labeled with Cy3, were prepared from mRNA obtained from pods without feeding insect. Each Cy5-labeled experimental cDNA probe was combined with Cy3-labeled reference probe and the mixture was hybridized to the microarray. The fluorescence ratio was quantified for each gene and reflected the relative abundance of transcripts corresponding to the given gene in each experimental sample in comparison to the reference mRNA. Data was uploaded to an

in-house microarray analysis database, GeneTraffic, and the differentially expressed genes were selected based on the fold-change ratio (> 3x).

Figure 4 provides an overview of the variation in gene expression throughout the samples. A hierarchical clustering algorithm was used to group genes based on the similarity of expression patterns among the samples. The data are shown in a matrix format, with each row representing the hybridization results for a single cDNA element of the array, and each column representing the measured expression levels for each gene in a single treatment comparison. To visualize the results, the expression level of each gene was color-coded, with black representing no change in expression; red, upregulation due to feeding; and green representing downregulation due to feeding. The color intensities reflect the magnitude of the change.

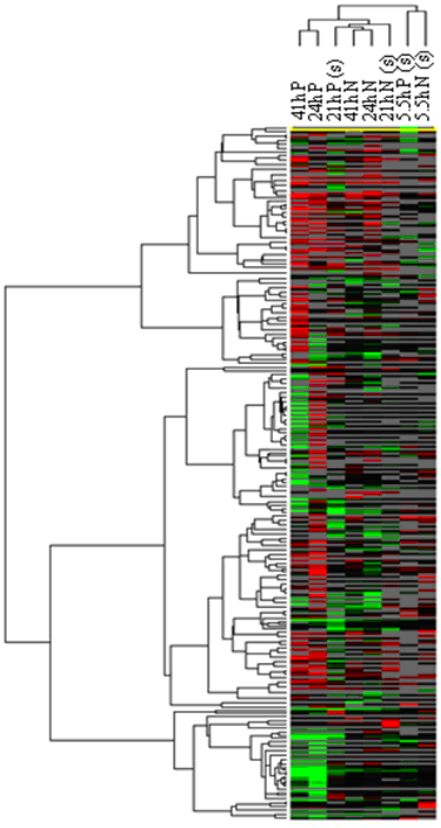
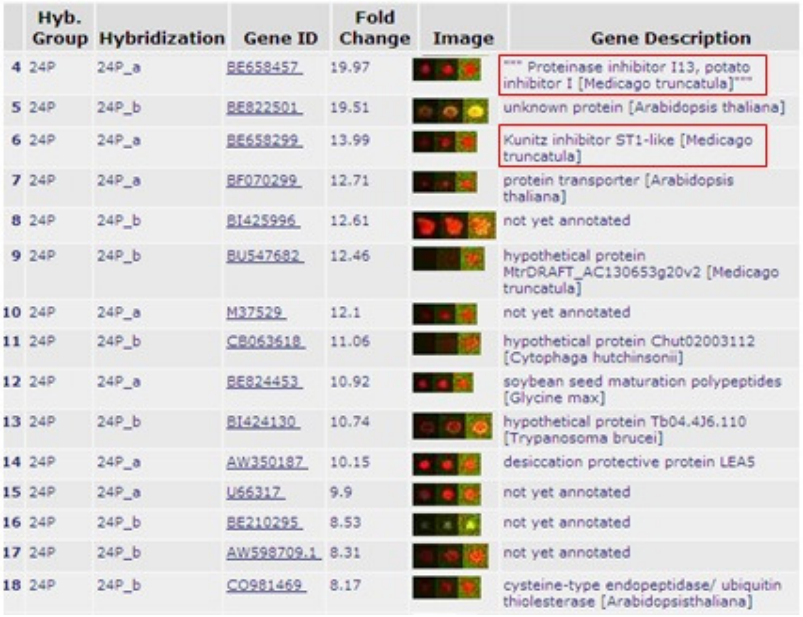


Figure 4 - Hierarchical clustering of the relative gene expression levels of soybean pods across four time points (5.5, 21, 24 and 41 hours after infestation) and two species of insects (*Nezara viridula* and *Euchistus heros*). Rows represent cDNA clones and columns, the treatments. Treatments: 41hP (41 hours, *P. guildinii*), 24hP (24 hours, *P. guildinii*), 21hP(s) (21 hours, starved *P. guildinii*), 5.5hP(s) (5.5 hours, starved *P. guildinii*), 41hN (41 hours, *N. viridula*), 24hN (24 hours, *N. viridula*), 21hN(s) (21 hours, starved, *N. viridula*), 5.5hN(s) (5.5 hours, starved, *N. viridula*)

Gene expression results from GeneTraffic allowed the observation that some of the most strongly induced genes for the different samples were genes typically associated with insect defense, such as proteinase inhibitors and lipoxygenases (Figures 5 and 6). However, looking at gene expression throughout the different samples, a lot of variation was observed (Figure 4); and since only one replication was performed, conclusions on soybean gene expression responses to stink bug feeding will be made only from the RNA-seq analysis of the field study. The pilot experiment served mainly to identify when detection of differential gene expression could be made. Results from the clustering analyses of the pilot experiment data from all treatments indicated that there was a clear differential expression among the samples from time points 21, 24, and 41 hours with *P. guildinii*. Focusing on consistently differentially expressed genes, it was noted that some typical defense-responsive genes were induced (Figure 7). Taking together these results, and the observation that the differential expression was increasing over time, we concluded that gene expression changes detected in the pilot-experiment were related to soybean response to insect feeding, and that differential gene expression can be detected within 24 hours after application of the insects.

In the pilot experiment, pods infested with *P. guildinii* induced greater changes in gene expression (Figure 4), and it was noticed that *P. guildinii* was the more active of the two species tested. Therefore, we used *P. guildinii* for the field study and collected samples at time points 0, 8, 24 and 46 hours. Due to limited financial resources for the project, only RNA from the 24 hour samples was submitted for RNA-seq analysis. The RNA extracted from samples taken at 0, 8 and 48 hours were stored at -80ºC and will be used in future studies.



	Hyb. Group	Hybridization	Gene ID	Fold Change	Image	Gene Description
4	24P	24P_a	BE658457	19.97		""" Proteinase inhibitor I13, potato inhibitor I [Medicago truncatula]"""
5	24P	24P_b	BE822501	19.51		unknown protein [Arabidopsis thaliana]
6	24P	24P_a	BE658299	13.99		Kunitz inhibitor ST1-like [Medicago truncatula]
7	24P	24P_a	BF070299	12.71		protein transporter [Arabidopsis thaliana]
8	24P	24P_b	BI425996	12.61		not yet annotated
9	24P	24P_b	BU547682	12.46		hypothetical protein MtrDRAFT_AC130653g20v2 [Medicago truncatula]
10	24P	24P_a	M37529	12.1		not yet annotated
11	24P	24P_b	CB063618	11.06		hypothetical protein Chut02003112 [Cytophaga hutchinsonii]
12	24P	24P_a	BE824453	10.92		soybean seed maturation polypeptides [Glycine max]
13	24P	24P_b	BI424130	10.74		hypothetical protein Tb04.4J6.110 [Trypanosoma brucei]
14	24P	24P_a	AW350187	10.15		desiccation protective protein LEA5
15	24P	24P_a	U66317	9.9		not yet annotated
16	24P	24P_b	BE210295	8.53		not yet annotated
17	24P	24P_b	AW598709.1	8.31		not yet annotated
18	24P	24P_b	CO981469	8.17		cysteine-type endopeptidase/ ubiquitin thiolesterase [Arabidopsisthaliana]

Figure 5 - Output from GeneTraffic software showing gene expression changes in pods 24 hours post-infestation with *Piezodorus guildinii*

	Hyb. Group	Hybridization	Gene ID	Fold Change	Image	Gene Description
1	24N	24N_a	CA819146	11.23		not yet annotated
2	24N	24N_a	M37529	10.04		not yet annotated
3	24N	24N_a	BI969433	9.01		unknown protein [Arabidopsis thaliana]
4	24N	24N_a	BU546523	5.92		unnamed protein product [Glycine max]
5	24N	24N_a	AW755768	5.22		not yet annotated
6	24N	24N_a	AW351243	4.94		FAS2 [Glycine max]
7	24N	24N_a	AI899977	4.57		not yet annotated
8	24N	24N_a	BI967838	4.36		glutathione S-transferase GST 10 [Glycine max]
9	24N	24N_a	BE822993	4.19		Mcp20 [Matricaria chamomilla]
10	24N	24N_a	BE820488	3.93		Mcp20 [Matricaria chamomilla]
11	24N	24N_a	AI939193	3.85		not yet annotated
12	24N	24N_a	AJ401230	3.83		not yet annotated
13	24N	24N_a	AW348586	3.83		unknown protein [Arabidopsis thaliana]
14	24N	24N_a	AW349864	3.66		GGH_SOYBN Gamma-glutamyl hydrolase precursor (Gamma-Glu-X carboxypeptidase)(Conjugase) (GH)
15	24N	24N_a	BU551268	3.64		PspA [Streptococcus pneumoniae]
16	24N	24N_a	AW351113	3.6		orf [Medicago sativa]
17	24N	24N_a	CA799194	3.49		not yet annotated
18	24N	24N_a	X58684	3.45		not yet annotated
19	24N	24N_a	CO984765	3.34		unnamed protein product [Glycine max]
20	24N	24N_a	CO978720	3.3		lipoxygenase [Phaseolus vulgaris]
21	24N	24N_a	AW570476	3.27		unknown protein [Arabidopsis thaliana]
22	24N	24N_a	BE658940	3.23		expansin [Pyrus communis]
23	24N	24N_a	BU550216	3.14		transcription factor EIL2 [Vigna radiata]

Figure 6 - Output from GeneTraffic software showing gene expression changes in pods 24 hours post-infestation with *Nezara viridula*

Figure 7 - Genes exhibiting the most significant differential expression levels in pods 24 and 41 hours post-infestation with *Piezodorus guildinii*. Star (*) represents defense-response genes

4.3.2 Constitutively differentially expressed genes between IAC-100 and CD-215 under greenhouse conditions

RNA-seq analysis was performed on RNA extracted from pods of plants grown in the absence of the insects in order to compare possible constitutive gene expression differences between the resistant (IAC-100) and susceptible (CD-215) cultivars. In the comparison, the resistant cultivar showed 39.4% of genes with higher expression levels (fold change > 2.0 up) and 11.68% of genes with lower expression (fold change > 2.0 down) in contrast with the levels observed in the susceptible cultivar.

Genome location of SNP [identified by USDA-ARS at the Beltsville Area Research Center, (BARC)] and SSR markers for stink bug resistance QTL identified in the current study were obtained from "Soybase and the Soybean Breeder's Toolbox" (http://soybase.org/gb2/gbrowse/gmax1.01/). Genes located within 500 kb downstream and 500 kb upstream the markers were investigated. Gene annotations were obtained from Phytozome (http://www.phytozome.net/soybean) and NCBI (http://www.ncbi.nlm.nih.gov/). The lg2 IAC/CD values for these genes ranged from -8.39 to 10.29.

Genes in the vicinity of the identified QTL are listed in Tables 2-15 and were classified into functional groups: cell development / division, defense, DNA / RNA, energy, lipids, membrane, miscellaneous, no hits, oxidation, oxidative state, primary metabolism, protein, secondary metabolism, signaling, stress, and unknown. Gray highlighted genes correspond to putative defense-related genes. Among these genes, the lg2 IAC/CD values

ranged from -5.95 (protease inhibitor/seed storage protein) to 9.75 (pathogenesis-related protein Bet v I family).

Table 2 - Genes in the vicinity of the QTL identified on chromosome 4 at position 0.0cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
04g07540.1	-3.401	actin	Cell development / division	Cytoskeleton
04g07500.1	-0.335	paramyosin-related	Cell development / division	Cytoskeleton
04g08210.1	0.443	avirulence induced gene (AIG)	Defense	Elicitor induced
04g08130.1	-0.125	RNA binding protein	DNA / RNA	Binding
04g08540.1	-1.591	RNA-binding protein La	DNA / RNA	Binding-RNA
04g07460.1	0.442	LUG (LEUNIG); protein binding / proteinheterodimerization/ transcription repressor	DNA / RNA	Gene regulation
04g07580.1	0.658	transcription elongation factor S-II	DNA / RNA	Gene regulation
04g08060.1	-1.203	WRKY DNA -binding domain / transcriptation factors	DNA / RNA	Gene regulation
04g08290.1	0.017	basic region leucine zipper	DNA / RNA	Gene regulation
04g08320.1	2.752	NAC/NAM (no apical meristem) protein	DNA / RNA	Gene regulation
04g08400.1	0.302	bZIP protein	DNA / RNA	Gene regulation
04g08550.1	2.492	MYB-like DNA-binding domain	DNA / RNA	Gene regulation
04g08570.1	0.365	RNA polymerase I-associated factor - PAF67	DNA / RNA	Synthesis-RNA
04g07700.1	1.065	SWIM zinc finger	DNA / RNA	Zinc ion binding
04g07720.1	1.060	MYM-type Zinc finger with FCS sequence motif	DNA / RNA	Zinc ion binding
04g08220.2	0.578	oleosin	Lipids	Storage lipids
04g07660.1	-0.142	syntaxin	Membrane	Membrane protein
04g07680.1	2.381	syntaxin	Membrane	Membrane protein
04g08480.1	1.105	beta2-adaptin appendage	Membrane	Trafficking
04g07950.1	-0.882	plasma membrane H+-transporting ATPase	Membrane	Transport
04g08190.1	-0.252	cation transport protein chac-related	Membrane	Transport
04g08600.1	5.326	VHS and GAT domain protein	Membrane	Transport
04g07860.1	0.511	predicted integral membrane metal-binding protein (DUF2296)	Membrane	
04g07820.1	1.996	X8 domain	Miscellaneous	Carbohydrate binding
04g07570.1	-3.456	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
04g07710.1	8.498	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
04g07980.1	6.542	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
04g08330.1	-0.366	SAM methyltransferase	Miscellaneous	
04g07810.1	-1.252	no hits	No hits	
04g07480.1	4.207	2-oxoglutarate-dependent dioxygenase / FE(II)	Oxidation	Multiple annotations
04g07520.1	-0.297	iron/ascorbate family oxidoreductases	Oxidation	
04g08560.1	-0.989	aspartate aminotransferase	Primary metabolism	Amino acid
04g08070.1	-1.625	sarcosine oxidase	Primary metabolism	Amino acid metabolism
04g07790.1	2.053	aspartate/glutamate/uridylate kinase familyprotein	Protein	Amino acid biosynthesis
04g07850.3	2.791	ubiquitin-specific protease 26	protein	Degradation
04g08200.1	-0.171	proteasome subunit alpha/beta	Protein	Degradation
04g08280.1	-0.209	26S proteasome regulatory subunit	Protein	Degradation
04g08510.1	0.155	signal peptide peptidase	Protein	Degradation
04g08620.1	1.446	acylamino acid releasing enzyme	Protein	Degradation
04g08000.1	1.108	kelch-related proteins	Protein	Protein binding
04g07940.1	0.348	protein translation initiation factor	Protein	Synthesis
04g07450.1	-0.365	flavin-containing monooxygenase	Secondary metabolism	
04g08250.1	3.003	SAUR family protein	Signalling	Hormone
04g07560.1	-1.355	serine/threonine protein phosphatase 2A	Signalling	Phosphatase
04g08470.1	-0.263	GTP binding protein	Signalling	G protein
04g08040.1	-0.634	tyrosine phopatase	Signalling	Phosphatase
04g08410.1	-0.515	RD-22 like protein	Stress	Dehydration
04g08150.1	-0.166	unknown	Unknown	
04g08180.1	8.913	unknown	Unknown	
04g08230.1	-0.665	unknown	Unknown	
04g08240.1	-2.444	unknown	Unknown	
04g08310.1	0.293	unknown	Unknown	
04g08390.1	0.838	unknown	Unknown	
04g08580.1	1.132	unknown	Unknown	
04g08680.1	0.982	unknown	Unknown	
04g08700.1	0.089	unknown	Unknown	
04g08710.1	-0.747	unknown	Unknown	
04g07640.1	9.626	unknown	Unknown	
04g07650.1	0.343	unknown	Unknown	
04g07830.1	2.247	unknown	Unknown	
04g07920.1	2.198	unknown	Unknown	
04g07930.1	0.403	unknown	Unknown	
04g07970.1	-3.508	unknown	Unknown	
04g08020.1	0.252	unknown	Unknown	
04g08050.1	-1.018	unknown	Unknown	
04g08100.1	1.048	cellullose synthase	Unknown	

Table 3 - Genes in the vicinity of the QTL identified on chromosome 4 at position 17.2cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
04g09840.1	1.651	cyclin d2	Cell development / division	Cell division
04g09680.1	2.925	rhamnogalacturonate lyase	Cell development / division	Cell wall
04g09350.1	-1.191	alpha tubulin	Cell development / division	Cytoskeleton
04g10080.1	8.820	kinesin heavy chain	Cell development / division	Cytoskeleton
04g09930.1	0.242	leucine rich repeat	Defense	R gene
04g10230.1	-1.268	double-stranded RNA binding	DNA / RNA	Binding-RNA
04g09710.1	1.675	DNA binding protein	DNA / RNA	DNA binding
04g10190.1	3.542	MYB-related transcription factor	DNA / RNA	Gene regulation
04g10330.1	-0.141	GATA zinc finger	DNA / RNA	Gene regulation
04g10340.1	3.254	GATA zinc finger	DNA / RNA	Gene regulation
04g09480.1	2.489	mRNA cleavage and polyadenylation factor II complex	DNA / RNA	Processing-RNA
04g09640.1	-1.366	pentatricopeptide repeat-containing protein	DNA / RNA	Processing-RNA
04g10140.3	8.820	RPB10; DNA-directed RNA polymerase II subunit L	DNA / RNA	Synthesis-RNA
04g09920.1	0.120	GRAM domain containing protein	Membrane	Membrane-associated
04g09310.1	0.021	amino acid transrpoter	Membrane	Transport
04g09460.1	-1.534	sucrose transporter	Membrane	Transport
04g09630.1	1.649	OPT oligopeptide transporter protein	Membrane	Transport
04g09770.1	1.790	mitochondrial oxoglutarate/malate carrier proteins	Membrane	Transport
04g09830.1	-0.042	divalent cation tolerance-related protein	Membrane	Transport
04g10060.2	-3.500	mechanosensitive ion channel	Membrane	Transport
04g09690.1	-0.462	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
04g10260.1	-0.131	tetratricopeptide repeat	Miscellaneous	TPR containig protein
04g10200.1	-0.267	NADPH:cytochrome P450 reductase	Oxidative state	Oxidoreductase
04g09340.1	-0.608	xylulose kinase	Primary metabolism	Carbohydrate
04g09510.1	-0.552	phosphoenolpyruvate carboxykinase (ATP) activity	Primary metabolism	Carbohydrate
04g09590.1	-0.680	phosphoglycerate mutase	Primary metabolism	Carbohydrate
04g09960.1	1.420	sentrin-specific cysteine protease (Ulp1 family)	Primary metabolism	Degradation
04g09740.1	-1.694	aspartyl protease	Protein	Degradation
04g10030.1	-0.244	DNAJ/HSP40 (heat shock protein binding)	Protein	Folding
04g09470.1	-0.568	beta-1,3-glucuronosyltransferase	Protein	modification
04g09730.1	-2.584	ubiquitin carboxyl-terminal hydrolase	Protein	Modification
04g10240.1	0.768	rab geranylgeranyl transferase like protein	Protein	Prenylation
04g09520.1	-0.521	gutamyl-tRNA reductase	Protein	Synthesis
04g09820.1	-1.008	14-3-3-like protein B	Signalling	14-3-3
04g09550.1	-0.542	aux/IAA protein	Signalling	Hormone
04g10090.1	-0.080	phosphatidylinositol 3-kinase	Signalling	Inositol
04g09370.1	3.983	serine/threonine/tyrosine kinase	Signalling	Kinase
04g09610.1	-0.609	serine/threonine protein kinase	Signalling	Kinase
04g09900.1	-1.363	serine/threonine protein kinase	Signalling	Kinase
04g10320.1	0.019	WD-containing protein and/or Gprotein subunit	Signalling	
04g09600.1	0.011	unknown	Unknown	Expressed
04g09330.1	2.398	unknown	Unknown	Predicted
04g09490.1	2.130	unknown	Unknown	
04g09570.1	5.446	unknown	Unknown	
04g09700.1	0.031	unknown	Unknown	
04g09750.1	3.378	unknown	Unknown	
04g09790.1	-0.073	unknown	Unknown	
04g09880.1	-2.273	unknown	Unknown	
04g09970.1	-0.953	unknown	Unknown	
04g10040.1	-1.008	unknown	Unknown	
04g10050.1	3.737	unknown	Unknown	
04g10170.1	1.157	unknown	Unknown	
04g10300.1	5.523	unknown	Unknown	

Table 4 - Genes in the vicinity of the QTL identified on chromosome 4 at position 73.2cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
04g40000.1	-0.126	expansin	Cell development / division	Cell wall
04g40630.1	0.638	harpin-induced protein 1 (Hin1)	Defense	Elicitor induced
04g40760.1	-0.132	RNA binding protein	DNA / RNA	Binding
04g40880.1	-0.100	WRC domain	DNA / RNA	Binding
04g40010.1	4.366	b-Zip DNA binding protein	DNA / RNA	Gene regulation
04g40130.1	3.551	WRKY DNA -binding domain	DNA / RNA	Gene regulation
04g40150.1	-0.158	transcription elongation factor s-II	DNA / RNA	Gene regulation
04g40450.1	1.750	NAC/NAM (no apical meristem) protein	DNA / RNA	Gene regulation
04g40610.1	-0.277	WD-40 containing protein	DNA / RNA	Gene regulation
04g40730.1	5.147	auxin-independent growth promoter protein	DNA / RNA	Gene regulation
04g40960.1	1.775	homeobox protein	DNA / RNA	Gene regulation
04g39980.1	-0.009	ribonuclease	DNA / RNA	Nuclease-RNA
04g41200.1	4.691	WD-40 repeat family protein	DNA / RNA	Processing-RNA
04g41220.1	-0.238	acyl-CoA-binding protein	Lipids	Lipid metabolism
04g41020.1	-0.344	OPT Oligopeptide transporter protein	Membrane	Transport
04g41140.1	2.303	OPT Oligopeptide transporter protein	Membrane	Integral Membrane Transporter
04g39880.1	-0.090	xanthine/uracil permease	Membrane	Transport
04g39960.1	-0.244	KUP system potassium uptake protein	Membrane	Transport
04g40420.1	0.486	band 7 family	Membrane	Transport
04g40500.1	3.567	copper chaperone	Membrane	Transport
04g40680.1	-0.717	phosphatidylcholine transfer	Membrane	Transport
04g40240.1	-0.379	zinc-binding dehydrogenase / alcohol dehydrogenase related	Miscellaneous	Alcohol dehydrogenase
04g41040.1	0.134	AAA-ATPase	Miscellaneous	ATP
04g41060.1	3.561	ATP binding protein	Miscellaneous	ATP
04g40140.1	0.185	ubiquinol-cytochrome c reductase complex 6.7 kDaprotein	Miscellaneous	Cytochrome related
04g40790.1	1.434	chloroplast small heat shock protein	Miscellaneous	Heat shock protein
04g41030.1	0.457	zinc-finger protein	Miscellaneous	Multiple functions
04g40020.1	0.171	zinc finger family, RING domain	Miscellaneous	Multiple functions
04g41160.1	-0.255	yippee putative zinc-binding protein	Miscellaneous	Multiple functions
04g40220.1	-0.307	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
04g39910.1	0.274	pentatricopeptide repeat-containing protein	Miscellaneous	PPR containing
04g40310.1	5.110	tetratricopeptide repeat / chaperone binding protein	Miscellaneous	
04g40360.1	-0.418	SAM domain (sterile alpha motif)	Miscellaneous	
04g40540.1	-0.275	nascent polypeptide associated complex alphasubunit	Miscellaneous	
04g39860.1	8.973	anionic peroxidase	Oxidative state	Peroxidase
04g39930.1	-0.271	manganese superoxide dismutases	Oxidative state	SOD
04g40990.1	-0.562	6-phosphogluconolactonase - like protein	Primary metabolism	Carbohydrate
04g40090.1	0.129	tRNA synthetases	Primary metabolism	Nucleotide metabolism
04g40970.1	2.887	aldo/keto reductase	Primary metabolism	Oxidoreductase activity
04g40980.1	0.448	ubiquitin/ribosomal protein	Protein	Degradation
04g41100.1	1.672	f-box domain	Protein	Degradation
04g39940.1	-0.333	40s ribosomal protein S26	Protein	Synthesis
04g40200.1	-0.765	60s ribosomal protein L10	Protein	Synthesis
04g40430.1	0.149	60S ribosomal protein L9	Protein	Synthesis
04g40470.1	0.273	60S ribosomal protein L9	Protein	Synthesis
04g40650.1	0.030	chloroplast ribosomal protein	Protein	Synthesis
04g40720.1	0.468	60s acidic ribosomal protein P1	Protein	Synthesis
04g40030.1	-0.168	chalcone-flavanone isomerase	Secondary metabolism	Flavonoid biosynthesis
04g40580.1	-0.496	caffeic acid3-O-methyltransferase	Secondary metabolism	Phenylpropanoid
04g40930.1	-0.622	SAUR family protein	Signalling	Hormone
04g40870.1	4.981	serine-threonine protein kinase, plant-type	Signalling	Kinase
04g41000.1	1.115	Ypt/Rab-specific GTPase-activating protein	Signalling	GTP related
04g40440.1	3.298	IAA-amino acid hydrolase	Signalling	Hormone
04g40080.1	-0.644	serine/threonine/tyrosine kinase	Signalling	Kinase
04g40920.1	2.392	CPDK-related protein kinase (calcium/calmodulin-dependent protein kinase)	Signalling	Kinase
04g40180.1	1.215	serine/threonine protein kinase / leucine-rich repeat transmembrane protein kinase	Signalling	Membrane associated/receptor
04g40100.1	2.548	response regulator	Signalling	Two-component system
04g41080.1	1.416	unknown	Unknown	Predicted/rhodanese-like
04g39970.1	8.235	unknown	Unknown	
04g39990.1	-0.805	unknown	Unknown	
04g40050.1	-0.998	unknown	Unknown	
04g40060.1	-0.233	unknown	Unknown	
04g40110.1	-0.358	unknown	Unknown	
04g40230.1	4.850	unknown	Unknown	
04g40270.1	2.951	unknown	Unknown	
04g40290.1	0.483	unknown	Unknown	
04g40300.1	0.485	unknown	Unknown	
04g40320.1	4.777	unknown	Unknown	
04g40350.1	-0.448	unknown	Unknown	
04g40390.1	2.529	unknown	Unknown	
04g40400.1	-0.941	unknown	Unknown	
04g40400.2	-2.076	unknown	Unknown	
04g40520.1	4.448	unknown	Unknown	
04g40620.1	1.691	unknown	Unknown	
04g40660.1	4.045	unknown, (plastid transcriptionally active15, PTAC15)	Unknown	
04g40670.1	2.386	unknown	Unknown	
04g40700.1	4.158	unknown	Unknown	
04g40780.1	-2.771	unknown	Unknown	
04g40890.1	0.006	unknown	Unknown	
04g40900.2	1.964	unknown	Unknown	
04g41070.1	3.291	unknown	Unknown	
04g41090.1	0.288	unknown	Unknown	
04g41120.1	0.612	unknown	Unknown	

Table 5 - Genes in the vicinity of the QTL identified on chromosome 5 at position 19.9cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
05g01800.1	1.168	glucose inhibited division protein	Cell development / division	Cell division
05g01870.1	5.612	LIM domain	Cell development / division	Cytoskeleton
05g01580.1	3.226	dynein light chain type 1	Cell development / division	Cytoskeleton
05g02370.1	2.685	leucine rich repeat	Defense	R gene
05g01420.1	-0.511	leucine-rich repeat family protein / proteinkinase family protein	Defense	R gene
05g02340.1	6.507	U3 small nucleolar ribonucleoprotein (snoRNP) subunit - Mpp10p	DNA / RNA	Binding-RNA
05g01150.1	-0.181	helix-loop-helix DNA-binding domain	DNA / RNA	Gene regulation
05g01320.1	7.049	nuclear protein ZAP	DNA / RNA	Gene regulation
05g01390.1	1.066	homeobox associated leucine zipper	DNA / RNA	Gene regulation
05g01590.1	-0.169	helix-loop-helix DNA-binding domain	DNA / RNA	Gene regulation
05g01640.1	-1.168	MYB-like DNA-binding domain	DNA / RNA	Gene regulation
05g01760.1	2.776	transcription factor E2F	DNA / RNA	Gene regulation
05g01910.1	-0.745	transcription elongation factor S-II protein N terminal	DNA / RNA	Gene regulation
05g01120.1	-0.265	nucleosome assembly protein (NAP)	DNA / RNA	Nuclear organization
05g01170.1	0.441	transducin /WD-40 repeat protein family	DNA / RNA	Processing-RNA
05g02240.1	-1.157	utp13 specific WD40 associated domain	DNA / RNA	Processing-RNA
05g02000.1	6.662	RNA dependent RNA polymerase	DNA / RNA	RNA
05g01230.1	4.981	ABC transporter	Energy	ATP
05g02290.1	-2.005	GDSL-like lipase/acylhydrolase	Lipids	Utilization/Redox/catabolism
05g01430.1	9.074	OPT oligopeptide transporter protein	Membrane	Transport
05g01350.1	-1.003	porin/voltage-dependent anion-selective channel protein	Membrane	Anion transport
05g01340.1	1.697	synaptotagmin	Membrane	Membrane proteins
05g01160.1	0.803	HVA22 like Proteins	Membrane	Transport
05g01380.1	8.552	nitrate transporter NTL1	Membrane	Transport
05g01440.1	2.025	nitrate transporter	Membrane	Transport
05g01450.1	8.235	nitrate transporter	Membrane	Transport
05g01460.1	1.957	cation-transporting ATPase	Membrane	Transport
05g01540.1	-2.403	AAA-type ATPase-like protein	Miscellaneous	ATP
05g01680.2	-2.277	HEAT repeat	Miscellaneous	Intracellular transport
05g02220.1	0.201	dof domain, zinc finger	Miscellaneous	Multiple functions
05g01690.1	0.291	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
05g02130.1	3.667	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
05g02260.1	-0.525	nitrogen fixation protein Nifu	Miscellaneous	Nodulation
05g01480.1	-0.034	pentatricopeptide (PPR) repeat-containing protein	Miscellaneous	PPR containing
05g01650.1	0.954	pentatricopeptide (PPR) repeat-containing protein	Miscellaneous	PPR containing
05g01820.1	0.832	miscellaneous	Miscellaneous	
05g01600.1	0.884	no hits	No hits	
05g01770.1	0.279	aldehyde dehydrogenase	Oxidative state	Oxidoreductase
05g01830.1	8.665	mannosyl-oligosaccharide 1,2-alpha-mannosidase	Primary metabolism	Carbohydrate
05g01570.1	1.959	galactosyltransferase	Primary metabolism	Catalytic activity
05g02420.1	-0.712	retinaldehyde binding protein-related	Primary metabolism	Catalytic activity
05g01930.1	-0.048	serine protease	Protein	Degradation
05g01980.1	6.241	ubiquitin-conjugating Enzime E2	Protein	Degradation
05g01330.1	3.641	small heat shock protein (HSP20)	Protein	Folding
05g01270.1	-0.689	ubiquitin-conjugating enzyme	Protein	Modification
05g01490.1	0.803	ubiquitin carboxyl-terminal hydrolase	Protein	Modification
05g01180.1	-0.265	40S ribosomal protein S9	Protein	Synthesis
05g01560.1	-0.730	"60S ribosome subunit biogenesis protein NIP7,putative [Ricinus communis]	Protein	Synthesis
05g01860.1	-0.605	calmodulin binding protein-like	Protein	Synthesis
05g02100.1	0.036	ribosomal protein S4	Protein	Synthesis
05g02190.1	0.273	1-deoxy-D-xylulose 5-phosphate reductoisomerase C-terminal	Secondary metabolism	Terpene/tocopherol biosynthesis
05g01240.1	-0.288	calmodulin	Signalling	Calcium
05g01470.1	3.938	Ca2+/calmodulin-dependent protein kinase, EF-hand protein superfamily	Signalling	Calcium
05g01920.1	9.019	ras-related small GTPase, Rho type	Signalling	GTP binding protein
05g01790.1	2.066	transducin family protein / WD-40 repeat familyprotein	Signalling	GTP related
05g02080.1	-0.117	protein kinase domain	Signalling	Kinase
05g02150.1	-3.389	serine/threonine/tyrosine kinase	Signalling	Kinase
05g01970.1	0.315	RING finger containing	Stress	Hypoxia induced protein conserved region
05g01250.1	8.721	unknown	Unknown	
05g01310.1	-0.135	unknown	Unknown	
05g01510.1	-1.551	unknown	Unknown	
05g01530.1	0.635	unknown	Unknown	
05g01630.1	1.127	unknown	Unknown	
05g01660.1	-1.225	unknown	Unknown	
05g01670.1	0.294	unknown	Unknown	
05g01740.1	0.390	unknown	Unknown	
05g01810.1	0.364	unknown	Unknown	
05g01880.1	-3.679	unknown	Unknown	
05g02010.1	-0.945	unknown	Unknown	
05g02030.1	6.874	unknown	Unknown	
05g02140.1	5.190	unknown	Unknown	
05g02210.1	1.450	unknown	Unknown	
05g02320.1	-0.646	unknown	Unknown	
05g02330.1	8.614	unknown	Unknown	
05g02350.1	1.836	unknown	Unknown	
05g02380.1	8.913	unknown	Unknown	

Table 6 - Genes in the vicinity of the QTL identified on chromosome 5 at position 28.2cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
05g02370.1	2.685	leucine rich repeat	Defense	R gene
05g02620.1	8.377	leucine rich repeat	Defense	R gene
05g02930.1	-0.449	RPM1-interacting protein 4 (RIN4) family protein	Defense	R gene related
05g02240.1	-1.157	utp13 specific WD40 associated domain	DNA / RNA	Processing-RNA
05g02340.1	6.507	U3 small nucleolar ribonucleoprotein (snoRNP) subunit - Mpp10p	DNA / RNA	Binding-RNA
05g02540.1	-3.310	histone deacetylase	DNA / RNA	Nuclear organization
05g02550.1	-1.871	myb-like DNA-binding domain	DNA / RNA	Gene regulation
05g02590.1	0.106	ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
05g02800.1	0.336	RNA-binding protein	DNA / RNA	Binding
05g02880.1	-1.236	BTB/POZ domain	DNA / RNA	Gene regulation
05g02900.1	1.421	replication factor C	DNA / RNA	Synthesis
05g03070.1	10.001	bromodomain containign protein	DNA / RNA	Gene regulation
05g03150.1	-0.144	nucleolar Protein NOP56	DNA / RNA	Nuclear organization
05g03280.1	-0.054	SAP domain	DNA / RNA	Binding
05g03290.1	0.258	BRG-1 associated factor 60 (BAF60)	DNA / RNA	Binding
05g02290.1	-2.005	GDSL-like lipase/acylhydrolase	Lipids	Utilization/Redox/Catabolism
05g02950.1	0.180	GDSL-motif lipase/hydrolase family protein	Lipids	Utilization/Redox/Catabolism
05g02790.1	2.014	amino acid transporters	Membrane	Transport
05g02990.2	1.239	ABC1 family	Membrane	Transport
05g03260.1	2.663	Vps51/Vps67	Membrane	Trafficking
05g02220.1	0.201	dof domain, zinc finger	Miscellaneous	Multiple functions
05g02260.1	-0.525	nitrogen fixation protein Nifu	Miscellaneous	Nodulation
05g02730.1	6.017	cytochrome P450	Miscellaneous	
05g02760.1	4.083	cytochrome P450	Miscellaneous	
05g02850.1	-0.400	WD40 repeat	Miscellaneous	WD-40 containign protein
05g02910.1	1.260	MA3 domain	Miscellaneous	Several possible functions
05g02920.1	-0.504	zinc finger (CCCH-type) family protein	Miscellaneous	Multiple functions
05g02420.1	-0.712	retinaldehyde binding protein-related	Primary metabolism	Catalytic activity
05g02510.1	1.812	raffinose synthase	Primary metabolism	Carbohydrate
05g02710.1	-0.503	NADA; quinolinate synthase	Primary metabolism	Metabolism of Cofactors and Vitamins
05g02810.1	-0.272	protein-L-isoaspartate O-methyltransferase	Primary metabolism	Transferase activity
05g02890.1	-0.808	pectate lyase	Primary metabolism	
05g03090.1	3.029	isocitrate lyase/malate synthase	Primary metabolism	Catalytic activity
05g03100.1	-0.219	isocitrate lyase/malate synthase	Primary metabolism	Catalytic activity
05g03190.1	-0.235	argininosuccinate synthase	Primary metabolism	Amino acid
05g03210.1	-1.847	argininosuccinate synthase	Primary metabolism	Amino acid
05g02100.1	0.036	ribosomal protein S4	Protein	Synthesis
05g02130.1	3.667	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
05g02570.1	0.213	60s ribosomal protein L27	Protein	Synthesis
05g02630.1	-0.661	ubiquitin-associated (UBA)/TS-N domain	Protein	Modification
05g02670.1	-0.744	mitochondrial translation elongation factor Tu	Protein	Synthesis
05g02940.1	0.239	26S proteasome regulatory complex	Protein	Degradation
05g02970.1	1.872	ribosomal protein S21	Protein	Synthesis
05g02980.1	0.224	26S proteasome regulatory complex	Protein	Degradation
05g02190.1	0.273	1-deoxy-D-xylulose 5-phosphate reductoisomerase C-terminal	Secondary metabolism	Terpene
05g02520.1	-0.476	tryptophan aminotransferase related 2/alliinase-like	Secondary metabolism	
05g02680.1	-1.193	protein phosphatase 2c	Signalling	Phosphatase
05g02080.1	-0.117	protein kinase domain	Signalling	Kinase
05g02150.1	-3.389	protein kinase domain	Signalling	Kinase
05g02470.1	1.501	serine/threonine/tyrosine kinase	Signalling	Kinase
05g02560.1	5.091	armadillo/beta-catenin-like repeat	Signalling	
05g02610.1	0.066	serine/threonine/tyrosine kinase	Signalling	Kinase
05g03010.1	0.009	nucleoside diphosphate kinase	Signalling	Kinase
05g03300.1	-0.170	AMP-activated protein kinase, gamma subunit	Signalling	Kinase
05g02010.1	-0.945	unknown	Unknown	
05g02030.1	6.874	unknown	Unknown	
05g02140.1	5.190	unknown	Unknown	
05g02210.1	1.450	unknown	Unknown	
05g02320.1	-0.646	unknown	Unknown	
05g02330.1	8.614	unknown	Unknown	
05g02350.1	1.836	unknown	Unknown	
05g02380.1	8.913	unknown	Unknown	
05g02460.1	-0.387	unknown	Unknown	
05g02480.1	-0.287	unknown	Unknown	
05g02500.1	0.820	unknown	Unknown	
05g02530.1	3.774	unknown	Unknown	
05g02640.1	-0.879	unknown	Unknown	
05g02690.1	-1.812	unknown	Unknown	
05g02820.1	-0.624	unknown	Unknown	
05g02840.1	4.114	unknown	Unknown	
05g03050.1	-0.657	unknown	Unknown	
05g03080.1	1.695	unknown	Unknown	

Table 7 - Genes in the vicinity of the QTL identified on chromosome 6 at position 58-62.6cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
06g32850.1	2.676	anaphase-promoting complex (APC), Cdc16 subunit	Cell development / division	Cell division
06g30860.1	0.513	cellulose synthase	Cell development / division	Cell wall
06g22390.1	2.749	kinesin	Cell development / division	Cytoskeleton
06g36650.1	3.585	dynamin	Cell development / division	Cytoskeleton
06g39740.1	0.896	myosin heavy chain	Cell development / division	Cytoskeleton
06g38600.1	4.289	amine oxidase	Defense	Apoptosis
06g38540.1	2.215	pathogenesis-related protein Bet v I family	Defense	Response to biotic stimulus
06g33940.1	-1.694	RNA-binding protein	DNA / RNA	Binding
06g25310.1	0.147	methyl-CpG binding domain	DNA / RNA	DNA binding
06g30350.1	0.322	methyl-CpG binding domain	DNA / RNA	DNA binding
06g30590.1	1.003	methyl-CpG binding domain	DNA / RNA	DNA binding
06g30620.3	-2.036	UV excision repeair protein RAD23	DNA / RNA	DNA binding
06g36140.1	1.758	SBP domain	DNA / RNA	DNA binding
06g21780.1	-0.008	MYB-like DNA-binding domain	DNA / RNA	Gene regulation
06g22240.1	0.297	TCP family transcription factor	DNA / RNA	Gene regulation
06g22650.1	-1.337	MADS BOX Protein	DNA / RNA	Gene regulation
06g24730.1	2.948	MYB-like DNA-binding domain	DNA / RNA	Gene regulation
06g27440.2	-0.041	WRKY DNA -binding domain	DNA / RNA	Gene regulation
06g34330.1	-0.563	TCP family transcription factor	DNA / RNA	Gene regulation
06g35330.1	-0.658	helix-loop-helix DNA-binding domain	DNA / RNA	Gene regulation
06g35560.1	-0.519	homeobox domain	DNA / RNA	Gene regulation
06g35710.1	-0.667	AP2 domain	DNA / RNA	Gene regulation
06g38410.1	0.896	no apical meristem (NAM) protein	DNA / RNA	Gene regulation
06g38440.1	-0.657	no apical meristem (NAM) protein	DNA / RNA	Gene regulation
06g39810.1	-1.192	B-box zinc finger	DNA / RNA	Gene regulation
06g21530.1	8.961	ATP-dependent helicase SMARCA	DNA / RNA	Helicase
06g21830.1	0.834	mRNA splicing factor ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
06g22750.1	0.264	dead box ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
06g23290.1	-0.475	dead box ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
06g21970.1	1.926	helicase SKI2W	DNA / RNA	Helicase activity
06g27640.1	-0.954	methyladenine glycosylase	DNA / RNA	Modifying-DNA
06g34850.1	6.429	BAH domain	DNA / RNA	Modifying-RNA
06g23240.2	0.024	nuclear factor Y, subunit B13	DNA / RNA	Nuclear organization
06g24500.1	0.244	histone h2a	DNA / RNA	Nuclear organization
06g35090.1	1.434	3'-5' exonuclease ERI1-related	DNA / RNA	Nucleic acids synthesis
06g22000.1	-0.188	U5 small nuclear ribonucleoprotein 200 KDA Helicase	DNA / RNA	Processing-RNA
06g23620.1	3.543	pentatricopeptide repeat-containing protein	DNA / RNA	Processing-RNA
06g24890.1	0.471	DNA-dependent RNA polymerase	DNA / RNA	Processing-RNA
06g29700.1	-0.151	pentatricopeptide repeat-containing protein	DNA / RNA	Processing-RNA
06g32720.1	3.500	pentatricopeptide repeat-containing protein	DNA / RNA	Processing-RNA
06g38110.1	1.118	pentatricopeptide repeat-containing protein	DNA / RNA	Processing-RNA
06g35650.1	2.027	reverse transcriptase, DNA polymerase, RNA dependent	DNA / RNA	Reverse transcriptase
06g23920.1	6.466	translation initiation factor 2C	DNA / RNA	RNA binding
06g23940.1	0.600	GRAS family transcription factor	DNA / RNA	RNA binding
06g36590.1	1.847	plastocyanin-like domain	Energy	Photosynthesis
06g23560.1	1.682	acyl-ACP thioesterase	Lipids	Fatty Acids
06g24640.1	-1.335	lysophospholipase	Lipids	Lysophospholipase
06g23340.1	0.363	oleosin	Lipids	Storage lipids
06g39700.1	1.066	cytochrome C biogenesis protein transmembrane region	Membrane	Cytochrome biogenesis
06g26810.1	2.728	reticulon	Membrane	Integral membrane protein
06g33130.1	-3.527	GRAM domain	Membrane	Membrane associated
06g22160.1	9.475	exo70 exocyst complex subunit	Membrane	Transport
06g28090.1	3.959	endomembrane protein 70	Membrane	Transport
06g29640.1	-0.426	amino acid transporter protein	Membrane	Transport
06g38400.1	4.054	ABC-2 type transporter	Membrane	Transport
06g39620.1	-0.310	organic solute transporter-related	Membrane	Transport
06g38120.1	1.946	choline transporter-like protein	Membrane	Transporter
06g21910.1	1.278	AAA-ATPase	Miscellaneous	ATP
06g23220.1	9.520	haloacid dehalogenase-like hydrolase	Miscellaneous	ATP
06g21920.1	1.687	cytochrome P450	Miscellaneous	Cytochrome
06g37250.1	-0.011	cytochrome C oxidase copper chaperone	Miscellaneous	Cytochrome
06g25620.1	0.899	serine esterase	Miscellaneous	Hydrolase
06g36640.1	1.182	VQ motif	Miscellaneous	Many possible functions
06g21730.1	2.212	zinc finger, C2H2 type	Miscellaneous	Multiple functions
06g22800.1	0.291	dof domain, zinc finger	Miscellaneous	Multiple functions
06g34960.1	0.312	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
06g22850.1	6.839	pentatricopeptide repeat-containing protein	Miscellaneous	PPR containing
06g35950.1	0.901	tetratricopeptide repeat	Miscellaneous	Protein-protein interaction
06g29710.1	1.943	core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	Miscellaneous	Realted to human inmune-syste
06g33850.1	0.209	thioredoxin	Miscellaneous	Redox
06g33880.1	-0.973	galactosyltransferase	Miscellaneous	Transferase
06g22440.1	-0.162	reversibly glycosylated polypeptide 3	Miscellaneous	
06g26370.1	1.396	phloem protein 2-A9	Miscellaneous	
06g35630.1	0.872	aminotransferase class I and II	Miscellaneous	
06g36210.1	-0.123	cytochrome P450	Miscellaneous	
06g36290.1	3.259	3-hydroxyacyl-CoA dehydrogenase, C-terminal domain	Miscellaneous	
06g36430.1	-0.008	nuclear pore complex protein (NUP)	Miscellaneous	
06g39970.1	-2.803	1,4-benzoquinone reductase-like; Trp repressor binding protein-like/protoplast-secreted protein	Oxidative state	Reduction

(Continues)

Table 7 - Genes in the vicinity of the QTL identified on chromosome 6 at position 58-62.6cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
06g22010.1	3.192	X8 domain	Primary metabolism	Carbohydrate binding
06g34890.1	1.190	homogentisate 1,2-dioxygenase	Primary metabolism	Amino acid
06g34940.1	0.018	homogentisate 1,2-dioxygenase	Primary metabolism	Amino acid
06g22820.1	0.761	UDP-glucuronosyl and UDP-glucosyl transferase	Primary metabolism	Carbohydrate
06g23080.1	0.704	enolase	Primary metabolism	Carbohydrate
06g33380.1	-0.940	phosphoenolpyruvate carboxylase	Primary metabolism	Carbohydrate
06g35110.1	-0.060	UDP-glucuronosyl and UDP-glucosyl transferase	Primary metabolism	Carbohydrate
06g36520.1	8.420	UDP-glucuronosyl and UDP-glucosyl transferase	Primary metabolism	Carbohydrate
06g40390.1	-1.568	UDP-glucuronosyl and UDP-glucosyl transferase	Primary metabolism	Carbohydrate
06g36190.1	0.626	serine/threonine protein kinase WNK (with no lysine)-related	Primary metabolism	Kinase activity
06g36070.1	3.614	methyltransferase activity	Primary metabolism	Methyltransferase activity
06g37960.1	1.578	cytidylyltransferase	Primary metabolism	Nucleotide-sugars
06g34190.1	0.375	lactate/malate dehydrogenase	Primary metabolism	Oxidoreductase activity
06g37970.1	0.097	thioredoxin	Primary metabolism	Oxidoreductase activity
06g38160.1	-0.697	short chain dehydrogenase	Primary metabolism	Oxidoreductase activity
06g33080.1	-0.391	polynucleotide 3'-phosphatase.	Primary metabolism	Phosphatase activity
06g36620.1	-0.251	enolase-phosphatase E-1	Primary metabolism	Phosphatase activity
06g38480.1	0.069	methyltransferase	Primary metabolism	SAM
06g22730.1	1.303	glycosyl transferase family 8	Primary metabolism	Transferase activity
06g24600.1	-0.886	protein arginine N-methyltransferase	Primary metabolism	Transferase activity
06g24780.1	0.300	MUTT/NUDIX Hydrolase	Primary metabolism	Transferase activity
06g28540.1	4.423	1-acylglycerol-3-phosphate O-acyltransferase.	Primary metabolism	Transferase activity
06g23320.1	0.992	translation initiation factor 2B	Primary metabolism	
06g24470.1	8.373	heme-binding protein-related	Primary metabolism	
06g24870.1	0.516	gpi-anchor transamidase	Primary metabolism	
06g23950.1	1.402	f-box domain	Protein	Degradation
06g27880.1	-0.297	26S proteasome regulatory subunit	Protein	Degradation
06g39710.1	0.069	proteasome subunit alpha/beta	Protein	Degradation
06g24830.1	-0.330	DNAJ heat shock	Protein	Folding
06g38470.1	3.923	peptidyl-prolyl cis-trans isomerase TLP38,chloroplast	Protein	Folding
06g22810.1	1.646	O-fucosyltransferase family protein	Protein	Modification
06g39770.1	-1.283	UBA/TS-N domain	Protein	Modification
06g22320.1	3.334	kelch repeat domain	Protein	Protein binding
06g39720.1	6.935	leucine rich repeat	Protein	Protein binding
06g22900.1	-8.392	60S ribosomal protein L21	Protein	Synthesis
06g30000.1	3.634	eukaryotic translation initiation factor 3	Protein	Synthesis
06g30880.1	-0.881	eukaryotic translation initiation factor 4 gamma	Protein	Synthesis
06g35690.1	0.346	ribosomal protein S20	Protein	Synthesis
06g36460.1	-0.218	ribosomal protein L6	Protein	Synthesis
06g38250.1	8.721	ribosomal protein S12	Protein	Synthesis
06g30020.1	-0.503	secy/sec61-alpha family member	Protein	Transport
06g23060.1	0.424	strictosidine synthase	Secondary metabolism	Alkaloids
06g24480.1	-0.461	beta-ketoacyl-CoA synthase [Helianthus annuus]	Secondary metabolism	
06g39910.1	-0.023	calmodulin	Signalling	Calcium
06g22360.1	0.334	transducin/WD40 repeat-like superfamily protein	Signalling	GTP related
06g22840.1	6.256	transducin/WD40 repeat-like superfamily protein	Signalling	GTP related
06g38170.1	3.671	transducin/WD40 repeat-like superfamily protein	Signalling	GTP related
06g39690.1	1.352	ARF GTPase activator activity	Signalling	GTP-related
06g25880.1	0.613	inositol monophosphatase	Signalling	Inositol
06g38490.1	5.484	inositol monophosphatase	Signalling	Inositol
06g21790.1	-3.293	leucine rich repeat	Signalling	Kinase
06g23590.1	5.557	serine/threonine protein kinase	Signalling	Kinase
06g27230.1	-0.329	serine/threonine protein kinase	Signalling	Kinase
06g33920.1	3.154	serine/threonine protein kinase	Signalling	Kinase
06g36230.1	4.684	serine/threonine protein kinase	Signalling	Kinase
06g39760.1	1.146	diacylglycerol kinase	Signalling	Kinase
06g39930.1	4.083	serine/threonine protein kinase	Signalling	Kinase
06g40240.1	3.614	serine/threonine protein kinase	Signalling	Kinase
06g40370.1	3.325	serine/threonine protein kinase	Signalling	Kinase
06g40400.1	1.189	serine/threonine protein kinase	Signalling	Kinase
06g40520.1	8.333	serine/threonine protein kinase	Signalling	Kinase
06g40560.1	8.721	serine/threonine protein kinase	Signalling	Kinase
06g40610.1	2.708	serine/threonine protein kinase	Signalling	Kinase
06g22060.1	1.741	phosphoserine phosphatase	Signalling	Phosphatase
06g36150.1	0.392	protein phosphatase 2c	Signalling	Phosphatase
06g39800.1	-0.446	universal stress protein family	Stress	Response to stress
06g21800.1	8.235	unknown	Unknown	
06g21850.1	-0.005	unknown	Unknown	
06g21860.1	3.017	unknown	Unknown	
06g21890.1	4.415	unknown	Unknown	
06g22130.1	4.634	unknown	Unknown	
06g22220.1	0.560	unknown	Unknown	
06g22250.1	1.744	unknown	Unknown	
06g22260.1	-2.207	unknown	Unknown	
06g22270.1	3.509	unknown	Unknown	
06g22290.1	1.328	unknown	Unknown	
06g23070.1	0.732	unknown	Unknown	
06g23250.1	0.340	unknown	Unknown	
06g23380.1	0.955	unknown	Unknown	
06g23580.1	4.505	unknown	Unknown	
06g28560.1	1.441	unknown	Unknown	
06g29540.1	0.285	unknown	Unknown	
06g29560.1	-0.356	unknown	Unknown	
06g29570.1	-0.607	unknown	Unknown	
06g29610.1	0.554	unknown	Unknown	
06g29980.1	-0.194	unknown	Unknown	
06g30630.1	0.154	unknown	Unknown	
06g32700.1	1.100	unknown	Unknown	
06g32860.1	-0.561	unknown	Unknown	
06g32870.1	-1.945	unknown	Unknown	
06g33970.2	-2.805	unknown	Unknown	
06g33980.1	1.806	unknown	Unknown	
06g34250.1	1.532	unknown	Unknown	
06g34340.1	0.383	unknown	Unknown	
06g34970.1	-0.492	unknown	Unknown	
06g35550.1	-1.603	unknown	Unknown	
06g35550.2	-0.068	unknown	Unknown	
06g35940.1	2.868	unknown	Unknown	
06g36170.1	3.446	unknown	Unknown	
06g36310.1	9.021	unknown	Unknown	
06g36610.1	-1.620	unknown	Unknown	
06g37230.1	-0.120	unknown	Unknown	
06g37260.1	-0.904	unknown	Unknown	
06g38130.1	0.682	unknown	Unknown	
06g39640.1	0.157	unknown	Unknown	

(Conclusion)

Table 8 - Genes in the vicinity of the QTL identified on chromosome 7 at position 33.9-38.9cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
07g13200.1	-0.435	cyclin	Cell development / division	Cell division
07g13230.1	-0.018	actin	Cell development / division	Cytoskeleton
07g10790.1	-4.153	kinesin motor domain	Cell development / division	Cytoskeleton
07g11160.1	-0.064	harpin-induced protein 1 (Hin1)	Defense	Elicitor induced
07g11700.1	3.453	harpin-induced protein 1 (Hin1)	Defense	Elicitor induced
07g11720.1	3.795	harpin-induced protein 1 (Hin1)	Defense	Elicitor induced
07g11950.1	0.231	B-cell receptor-associated 31-like	Defense	Pro-apoptopic
07g13210.1	-1.238	RNA-binding protein	DNA / RNA	Binding
07g13900.1	0.687	hyaluronan / mRNA binding family	DNA / RNA	Binding
07g11140.1	3.935	poly(A) polymerase	DNA / RNA	Binding-RNA
07g11510.1	-0.554	Zn-finger in ran binding protein and others	DNA / RNA	Binding-RNA
07g11560.1	-0.574	nucleolar protein NOP56	DNA / RNA	Binding-RNA
07g13260.1	-2.379	methyl-CpG binding domain	DNA / RNA	DNA binding
07g11840.1	1.175	nuclear transcription factor, X-box binding 1 (NFX1)	DNA / RNA	Gene regulation
07g12070.1	-4.052	MYB-like DNA-binding domain	DNA / RNA	Gene regulation
07g12140.1	-0.940	zinc finger homeodomain protein SZF-HD	DNA / RNA	Gene regulation
07g13410.1	-0.805	helix-loop-helix DNA-binding domain	DNA / RNA	Gene regulation
07g13420.1	1.646	helix-loop-helix DNA-binding domain	DNA / RNA	Gene regulation
07g13500.1	0.668	helix-loop-helix DNA-binding domain	DNA / RNA	Gene regulation
07g14270.1	0.782	homeobox protein	DNA / RNA	Gene regulation
07g14330.1	0.487	mitochondrial transcription terminationfactor-related / mTERF-related [Arabidopsis	DNA / RNA	Gene regulation
07g11880.1	-2.404	dead box ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
07g11980.1	2.568	Ssu72-like protein, RNA Polymerase	DNA / RNA	Synthesis
07g14100.1	1.745	ATPase proton transporting	Energy	ATPase
07g13840.1	1.118	plastocyanin-like domain	Energy	Photosynthesis
07g13470.1	-0.562	ATP Synthase Delta	Energy	ATP
07g11650.1	0.134	SURF1 family	Energy	Electron transfer
07g12630.1	4.236	cytochrome b	Energy	Electron transfer
07g13790.1	0.154	patatin	Lipids	Degradation/utilization
07g12750.1	3.489	fatty acid hydroxylase superfamily	Lipids	Fatty Acids
07g10820.1	4.266	nonspecific lipid-transfer protein	Membrane	Biogenesis
07g10900.1	8.913	reticulon	Membrane	Integral membrane protein
07g13040.1	0.689	vacuolar sorting protein 39 domain 2	Membrane	Trafficking
07g13120.1	-0.355	vacuolar membrane protein	Membrane	Trafficking
07g11240.1	1.045	multidrug resistance pump	Membrane	Transport
07g11250.1	3.287	multidrug resistance pump	Membrane	Transport
07g11550.1	-2.535	membrane transport protein	Membrane	Transport
07g12180.1	-4.064	multidrug resistance pump	Membrane	Transport
07g12680.1	0.882	ABC transporter	Membrane	Transport
07g12280.1	8.820	autophagocytosis associated protein	Miscellaneous	Autophagocytosis protein
07g12080.1	2.181	embryo sac development arrest 7	Miscellaneous	Development
07g13350.1	2.900	alpha/beta hydrolase fold	Miscellaneous	Hydrolase
07g13070.1	-2.560	ferredoxin	Miscellaneous	Iron related
07g12090.1	1.531	inositol polyphosphate 5-phosphatase or nuclease	Miscellaneous	Multiple annotations
07g12170.1	2.392	zinc finger, C2H2 type	Miscellaneous	Multiple functions
07g10930.1	1.374	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
07g12450.1	-0.777	nodulin-like	Miscellaneous	Nodulin
07g13330.1	4.922	cytochrome P450	Miscellaneous	P450
07g11410.1	3.701	pentatricopeptide repeat-containing protein	Miscellaneous	PPR
07g11930.1	4.409	pentatricopeptide repeat-containing protein	Miscellaneous	PPR
07g13170.1	0.236	pentatricopeptide repeat-containing protein	Miscellaneous	PPR containing
07g12850.1	2.586	dynamin GTPase effector domain	Miscellaneous	Vacuolar sorting
07g13760.1	5.226	gag-pol polyprotein	Miscellaneous	Viral
07g13140.1	-0.113	no hits	No hits	
07g11910.1	2.025	no hits	No hits	
07g11520.1	-2.994	glutathione S-transferase	Oxidative state	Antioxidant
07g11530.1	5.879	FAD NADPH dehydrogenase/oxidoreductase	Oxidative state	Oxidoreductase

(Continues)

Table 8 - Genes in the vicinity of the QTL identified on chromosome 7 at position 33.9-38.9cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
07g12190.1	-1.322	hexokinase	Primary metabolism	Carbohydrate
07g13130.1	1.223	UDP-glucuronosyl and UDP-glucosyl transferase	Primary metabolism	Carbohydrate
07g13560.1	-4.172	UDP-glucuronosyl and UDP-glucosyl transferase	Primary metabolism	Carbohydrate
07g11310.1	-0.390	glycosyl hydrolase	Primary metabolism	Carbohydrate metabolism
07g12130.1	4.218	dihydroxyacetone kinase family protein	Primary metabolism	Carbohydrate metabolism
07g11610.1	5.132	fructose-bisphosphate aldolase class-II	Primary metabolism	Carbohydrates
07g11810.1	-0.941	glutamine synthetase	Primary metabolism	Nitrogen metabolism
07g12660.1	-3.737	ribose-phosphate pyrophosphokinase	Primary metabolism	Nucleic acid
07g12840.1	-0.460	adenylate kinase	Primary metabolism	Nucleic acid
07g13710.1	0.286	nucleoside diphosphate kinase	Primary metabolism	Nucleic acid
07g11180.1	-1.022	oxidoreductase, 2OG-Fe(II) oxygenase familyprotein	Primary metabolism	Oxidoreductase activity
07g13100.1	1.683	2OG-Fe(II) oxygenase familyprotein	Primary metabolism	Oxidoreductase activity
07g11580.1	8.373	methyltransferase	Primary metabolism	SAM
07g12780.1	3.801	sterol desaturase	Primary metabolism	Sterols
07g13150.1	9.557	glycosyltransferase	Primary metabolism	Transferase
07g13320.1	-0.556	aminotransferase class-III	Primary metabolism	Transferase
07g11890.1	-0.040	dihydroorotate dehydrogenase	Primary metabolism	
07g10940.1	0.852	lysosomal acid lipase-related	Protein	Degradation
07g11690.1	0.439	molecular chaperone Hsp40/DnaJ family protein	Protein	Folding
07g12980.1	-0.189	chaperonin 10 Kd subunit	Protein	Folding
07g14250.1	5.096	peptidylprolyl isomerase, FKBP-type	Protein	Folding
07g14120.1	4.998	protein farnesyltransferase alpha subunit	Protein	Modification
07g12340.1	5.068	glutamyl-tRNA synthetase	Protein	Synthesis
07g13380.1	0.100	eukaryotic translation initiation factor 4 gamma	Protein	Synthesis
07g11590.1	1.216	vacuolar protein sorting 26	Protein	Transport
07g14460.1	-0.124	obtusifoliol14-alpha-demethylase	Secondary metabolism	Sterols
07g11390.1	3.228	calmodulin	Signalling	Calcium
07g12030.1	-0.741	annexin	Signalling	Calcium
07g11970.1	5.050	G-protein alpha subunit	Signalling	GTP binding protein
07g11420.1	6.151	RAS-related GTPASE	Signalling	GTP-related
07g10490.1	3.216	serine/threonine protein kinase	Signalling	Kinase
07g10570.1	0.625	serine/threonine protein kinase	Signalling	Kinase
07g10690.1	0.734	serine/threonine protein kinase	Signalling	Kinase
07g10760.1	8.820	serine/threonine protein kinase	Signalling	Kinase
07g11280.1	7.202	CDC2-related Kinase	Signalling	Kinase
07g11430.1	8.290	serine/threonine protein kinase	Signalling	Kinase
07g11670.1	-0.597	protein kinase superfamily protein	Signalling	kinase
07g11680.1	0.589	receptor-like kinase 1	Signalling	kinase
07g13390.1	3.685	serine/threonine protein kinase	Signalling	Kinase
07g13440.1	-0.988	serine/threonine protein kinase	Signalling	Kinase
07g13960.1	0.144	serine/threonine protein kinase	Signalling	Kinase
07g12950.1	4.997	calcineurin-like phosphoesterase	Signalling	Phosphatase
07g10950.1	-0.780	octicosapeptide/phox/bem1p family protein	Signalling	
07g12600.1	-0.091	XYPPX repeat	Unknown	XYPPX repeat
07g10710.1	5.127	unknown	Unknown	
07g10780.1	3.913	unknown	Unknown	
07g10800.1	2.566	unknown	Unknown	
07g11300.1	-0.303	unknown	Unknown	
07g11630.1	1.757	unknown	Unknown	
07g11800.1	0.875	unknown	Unknown	
07g12360.1	0.461	unknown	Unknown	
07g12380.1	-0.936	unknown	Unknown	
07g12530.1	-2.890	unknown	Unknown	
07g12570.1	5.910	unknown	Unknown	
07g12810.1	0.433	unknown	Unknown	
07g12890.1	-2.142	unknown	Unknown	
07g14350.1	-1.351	unknown	Unknown	
07g14410.1	1.223	unknown	Unknown	
07g14440.1	8.391	unknown	Unknown	
07g14450.1	4.423	unknown	Unknown	

(Conclusion)

Table 9 - Genes in the vicinity of the QTL identified on chromosome 13 at position 7.8cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
13g04230.1	5.126	leucine rich repeat	Defense	Apopitosis
13g03770.1	8.847	leucine rich repeat	Defense	Apoptosis / PCD
13g03760.1	0.187	RRM-containing RNA-binding protein-like protein	DNA / RNA	Binding-RNA
13g03840.1	2.258	RNA recognition motif (RRM)-containing protein	DNA / RNA	Binding-RNA
13g03660.1	-0.217	transcription factor EIL2	DNA / RNA	Gene regulation
13g03700.1	-0.262	transcription factor EIL2	DNA / RNA	Gene regulation
13g03740.1	8.633	transcription factor bZIP34	DNA / RNA	Gene regulation
13g04030.1	8.404	MYB-like DNA-binding domain	DNA / RNA	Gene regulation
13g03920.1	-0.901	vacuolar protein sorting-associated protein 26	Membrane	Trafficking
13g03940.1	-0.920	peroxisomal membrane protein MPV17 and related proteins	Membrane	Transport
13g04330.1	0.506	zinc finger, C3HC4 type (RING finger)	Miscellaneous	Multiple functions
13g04020.1	6.724	RINT-1 / TIP-1 family	Miscellaneous	Several possible functions
13g03900.1	8.235	gag-pol polyprotein	Miscellaneous	Viral
13g03650.1	4.833	ascorbate oxidase	Oxidative state	Antioxidant
13g04150.1	-0.390	nudix hydrolase homolog 3	Primary metabolism	Transferase activity
13g04190.1	-0.405	nudix hydrolase homolog 3	Primary metabolism	Transferase activity
13g03810.1	0.098	protease inhibitor/seed storage/LTP family	Protein	Degradation
13g04250.1	0.148	cystatin	Protein	Degradation
13g04350.1	1.719	small heat shock protein (HSP20)	Protein	Folding
13g04140.1	-0.538	ubiquitin-associated (UBA)/TS-N domain-containing protein	Protein	Modification
13g04120.1	1.028	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	Secondary metabolism	Terpene
13g03820.1	0.206	squalene/phytoene synthase	Secondary metabolism	Terpenoids
13g03910.1	0.157	calmodulin	Signalling	Calcium
13g03990.1	4.530	serine/threonine protein kinase	Signalling	Kinase
13g04000.1	0.969	unknown	Unknown	
13g04420.1	0.230	unknown	Unknown	

Table 10 - Genes in the vicinity of the QTL identified on chromosome 15 at position 107.4cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
15g26790.1	-1.378	leucine rich repeat	Defense	R gene
15g26820.1	3.521	gag-pol polyprotein	Miscellaneous	Viral
15g26330.1	0.646	serine/threonine protein kinase	Signalling	Kinase
15g26590.1	-0.474	temperature-induced lipocalin	Stress	Temperature incuded membrane protein

Table 11 - Genes in the vicinity of the QTL identified on chromosome 15 at position 62.4cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
15g14390.1	-0.333	CDC2-related kinase	Cell development / division	Cell division
15g16650.1	0.297	fasciclin-like arabinogalactan protein	Cell development / division	Cell wall
15g16900.1	-1.185	cellulose synthase	Cell development / division	Cell wall
15g18360.1	0.162	xyloglucan endo-transglycosylase (XET) C-terminus	Cell development / division	Cell wall
15g14420.1	4.664	hydroxyproline-rich glycoprotein family protein	Cell development / division	Cell wall
15g16880.1	8.235	pectin lyase-like superfamily protein	Cell development / division	Cell wall
15g17640.1	0.235	villin	Cell development / division	Cytoskeleton
15g13970.1	-0.622	beta tubulin	Cell development / division	Cytoskeleton
15g15900.1	-0.068	kisenin heavy chain	Cell development / division	Cytoskeleton
15g15910.1	-1.075	translationally controlled tumour protein	Cell development / division	Cytoskeleton
15g16050.1	-0.559	dynein light chain	Cell development / division	Cytoskeleton
15g16060.1	-1.144	dynein light chain	Cell development / division	Cytoskeleton
15g13720.1	3.537	myosin heavy chain-related	Cell development / division	Cytoskeleton
15g17690.1	3.095	metaxin	Defense	Apopitosis
15g15360.1	-1.078	chitinase class I	Defense	Fungal cell wall degradation
15g13740.1	1.927	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
15g13750.1	0.981	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
15g13760.1	-5.955	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
15g13770.1	-3.305	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
15g18290.1	2.831	leucine rich repeat	Defense	R gene
15g16310.1	9.526	disease resistance protein (TIR-NBS-LRR class) family	Defense	Resistance protein
15g17310.1	8.830	disease resistance protein (TIR-NBS-LRR class) family	Defense	Resistance protein
15g15610.1	2.518	pathogenesis-related protein Bet v I family	Defense	Response to biotic stimulus
15g15660.1	9.753	pathogenesis-related protein Bet v I family	Defense	Response to biotic stimulus
15g15680.1	0.046	pathogenesis-related protein Bet v I family	Defense	Response to biotic stimulus
15g18010.1	0.255	KH domain	DNA / RNA	Binding
15g18050.1	-0.087	KH domain	DNA / RNA	Binding
15g16570.1	-0.307	DNA replication licensing factor, MCM6 component	DNA / RNA	DNA binding
15g13990.1	1.235	ethylene-responsive nuclear protein	DNA / RNA	Gene regulation
15g14860.1	1.966	WRKY DNA -binding domain	DNA / RNA	Gene regulation
15g14890.1	-0.639	sequence-specific DNA binding transcription factors	DNA / RNA	Gene regulation
15g15520.1	1.768	type-b response regulator	DNA / RNA	Gene regulation
15g15950.1	-0.789	transducin/WD40 repeat-like superfamily protein	DNA / RNA	Gene regulation
15g16260.1	-0.952	AP2 domain	DNA / RNA	Gene regulation
15g16390.1	4.825	mitochondrial transcription termination factor, mTERF	DNA / RNA	Gene regulation
15g16400.1	2.316	mitochondrial transcription termination factor, mTERF	DNA / RNA	Gene regulation
15g16410.1	2.350	mitochondrial transcription termination factor, mTERF	DNA / RNA	Gene regulation
15g16410.2	8.235	mitochondrial transcription termination factor, mTERF	DNA / RNA	Gene regulation
15g16430.1	-0.148	mitochondrial transcription termination factor, mTERF	DNA / RNA	Gene regulation
15g16730.1	1.412	transducin/WD40 repeat-like superfamily protein	DNA / RNA	Gene regulation
15g17040.1	4.105	PLATZ transcription factor	DNA / RNA	Gene regulation
15g17550.1	-3.039	sequence-specific DNA binding transcription factors;transcription regulators	DNA / RNA	Gene regulation
15g18250.1	2.146	WRKY DNA -binding domain	DNA / RNA	Gene regulation
15g18320.1	6.052	homeobox protein	DNA / RNA	Gene regulation
15g18380.1	-0.629	GATA zinc finger	DNA / RNA	Gene regulation
15g18450.1	-0.180	transducin family protein / WD-40 repeat family protein	DNA / RNA	Gene regulation
15g14240.1	2.774	toprim domain	DNA / RNA	Helicase / Topoisomerase
15g14470.1	0.544	dead box ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
15g16830.1	-1.256	SET domain	DNA / RNA	Nuclear organization
15g17030.1	9.328	SET domain	DNA / RNA	Nuclear organization
15g15110.1	8.756	GRAS family transcription factor	DNA / RNA	RNA binding
15g14490.1	-1.313	YT521-B-like family	DNA / RNA	Splicesome
15g15250.1	-0.207	DNA-directed RNA Polymerase	DNA / RNA	Synthesis-RNA
15g15450.1	0.655	SWIM zinc finger	DNA / RNA	Zinc ion binding
15g15450.2	3.361	SWIM zinc finger	DNA / RNA	Zinc ion binding
15g14930.1	6.519	GDSL-like lipase/acylhydrolase	Lipids	Utilization/Redox/catabolism
15g16330.1	0.304	lipase (class 3)	Lipids	Utilization/Redox/catabolism
15g16360.1	1.204	synaptotagmin	Membrane	Membrane proteins
15g16560.1	0.185	DREPP plasma membrane polypeptide	Membrane	Membrane-associated
15g15810.1	8.498	peripheral-type benzodiazepine receptor	Membrane	Receptor
15g16810.1	0.200	vacuolar sorting receptor-like protein	Membrane	Trafficking
15g14290.1	0.020	BT1 family	Membrane	Transport
15g15480.1	0.011	importin alpha	Membrane	Transport
15g15760.1	-0.584	synaptobrevin	Membrane	Transport
15g16370.1	-0.077	mitochondrial carrier protein	Membrane	Transport
15g18180.1	0.286	calcium transporting ATPase	Membrane	Transport
15g18230.1	1.020	triose-phosphate Transporter family	Membrane	Transport
15g16580.1	-0.300	purine transporter	Membrane	Transporter
15g17530.1	0.068	haloacid dehalogenase-like hydrolase	Miscellaneous	ATP
15g15640.1	-0.103	melanoma-associated antigen (MAGE Antigen)	Miscellaneous	Hum an immune system
15g15310.1	2.083	disulphide oxidoreductase	Miscellaneous	Hydrogenase/oxidoreductase
15g16500.1	8.963	CW-type zinc finger	Miscellaneous	Multiple functions
15g13870.1	-3.871	MtN19 protein	Miscellaneous	Nodulation
15g13880.1	-4.209	MtN19 protein	Miscellaneous	Nodulation
15g16030.1	-0.104	MtN3/saliva family	Miscellaneous	Nodulation
15g17580.1	6.141	nitrogen fixation protein nifU	Miscellaneous	Nodulation
15g14330.1	-1.048	cytochrome P450	Miscellaneous	P450
15g13930.1	1.273	pentatricopeptide repeat-containg protein	Miscellaneous	Pentatricopeptide repeat-containg protein
15g16840.1	1.948	pentatricopeptide repeat-containing protein	Miscellaneous	Pentatricopeptide repeat-containing
15g17500.1	-0.211	pentatricopeptide repeat-containing protein	Miscellaneous	Pentatricopeptide repeat-containing
15g17780.1	8.798	pentatricopeptide repeat-containing protein	Miscellaneous	Pentatricopeptide repeat-containing
15g17240.1	0.122	ankyrin repeat	Miscellaneous	Protein-protein interactions
15g16960.1	3.064	methyltransferase domain	Miscellaneous	SAM
15g17820.1	8.406	gag-pol polyprotein	Miscellaneous	Viral
15g14260.1	-0.887	cytomatrix protein-related	Miscellaneous	
15g14970.1	0.114	huntingtin interacting protein HYPK	Miscellaneous	
15g15220.1	8.899	WD40/YVTN repeat-like-containing domain;Bromodomain	Miscellaneous	
15g17000.1	-0.019	no hits	No hits	

(Continues)

Table 11 - Genes in the vicinity of the QTL identified on chromosome 15 at position 62.4cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
15g18440.1	0,681	amine oxidase	Oxidative state	Oxidase
15g15070.1	1,315	aldehyde dehydrogenase family	Oxidative state	Oxidoreductase
15g15830.1	0,570	arginine N-methyltransferase	Primary metabolism	Amino acid
15g14140.1	0,272	alpha-glucosidase 1	Primary metabolism	Carbohydrate
15g14150.1	2,763	alpha-glucosidase 1	Primary metabolism	Carbohydrate
15g14430.1	3,342	NAD Dependent Epimerase/Dehydratase	Primary metabolism	Carbohydrate
15g15820.1	-0,631	alpha-1,6-xylosyltransferase	Primary metabolism	Carbohydrate
15g18220.1	-0,573	oligosaccharyltransferase complex/magnesium transporter family protein	Primary metabolism	Carbohydrate
15g18430.1	2,797	beta-galactosidase	Primary metabolism	Carbohydrate
15g14020.1	4,383	berberine	Primary metabolism	Carbohydrate / Lipid
15g14040.1	0,239	berberine	Primary metabolism	Carbohydrate / Lipid
15g14170.1	5,546	berberine	Primary metabolism	Carbohydrate / Lipid
15g15370.1	-0,464	glycosyl hydrolase	Primary metabolism	Carbohydrate metabolism
15g14460.1	3,971	ribose 5-phosphate isomerase	Primary metabolism	Carbohydrate/PPP pathway
15g15010.1	0,296	ATP-citrate synthase	Primary metabolism	Citric acid cycle
15g15020.1	0,263	ATP-citrate synthase	Primary metabolism	Citric acid cycle
15g17330.1	-0,060	histidine triad (HIT) protein	Primary metabolism	Hydrolase activity
15g16950.1	1,289	cholinephosphate cytidylyltransferase	Primary metabolism	Nucleotidyltransferase activity
15g16490.1	4,241	2OG-Fe(II) oxygenase super family	Primary metabolism	Oxidoreductase activity
15g13860.1	4,105	3-hydroxyisobutyrate dehydrogenase	Primary metabolism	Pentose phosphate pathway
15g14690.1	0,650	f-box domain	Protein	Degradation
15g17830.1	0,364	subtilase	Protein	Degradation
15g18110.1	-4,335	ATP-Dependent CLP Protease	Protein	Degradation
15g14700.1	-0,198	DNAJ/HSP40 (heat shock protein binding)	Protein	Folding
15g14870.1	-0,191	prefoldin subunit	Protein	Folding
15g15930.2	-0,229	DNAJ/HSP40 (heat shock protein binding)	Protein	Folding
15g15100.1	0,638	ubiquitin-conjugating enzyme	Protein	Modification
15g18190.1	-0,416	O-fucosyltransferase family protein	Protein	Modification
15g13850.1	9,151	KRR1 family protein	Protein	Synthesis
15g14280.1	0,218	ribosomal protein L7/L12	Protein	Synthesis
15g14550.1	-0,540	30S ribosomal protein S1, chloroplast precursor	Protein	Synthesis
15g14710.1	0,476	ribosomal protein S6	Protein	Synthesis
15g15320.1	0,607	60S ribosomal protein L29	Protein	Synthesis
15g15800.1	-0,389	40S ribosomal protein S21	Protein	Synthesis
15g15840.1	-0,300	ARM repeat superfamily protein	Protein	synthesis
15g17010.1	-0,452	60S ribosomal protein L27	Protein	Synthesis
15g18500.1	0,268	translation initiation factor SUI1	Protein	Synthesis
15g15350.1	1,247	calmodulin-binding protein	Signalling	Calcium
15g15380.1	-3,745	cAMP-regulated phosphoprotein 19-related protein	Signalling	cAMP
15g15330.1	0,266	GTP-binding protein sar1	Signalling	GTP binding protein
15g17870.1	3,789	microtubule-associated protein	Signalling	GTP binding protein
15g14760.1	0,399	rab5-interacting family protein	Signalling	GTP related
15g13830.1	-0,447	GTP-binding protein	Signalling	GTP related
15g14220.1	-5,446	sec14p-like phosphatidylinositol transfer family protein	Signalling	Inositol
15g16510.1	1,409	5'-AMP-activated protein kinase, gamma subunit	Signalling	Kinase
15g13840.1	5,532	serine/threonine protein kinase	Signalling	Kinase
15g16670.1	0,188	serine/threonine protein kinase	Signalling	Kinase
15g17360.1	-0,140	serine/threonine protein kinase	Signalling	Kinase
15g18470.1	4,776	serine/threonine protein kinase	Signalling	Kinase
15g14980.1	-0,553	phytochrome B2	Signalling	Light
15g16270.1	2,920	phospholipase D	Signalling	Phospholipase
15g15990.1	-1,178	late embryogenesis abundant protein	Stress	Drought
15g15870.1	-0,708	multidrug resistance-associated protein 4	Stress	
15g14130.1	5,124	uncharacterized ACR, COG1565	Unknown	
15g14530.1	0,032	unknown	Unknown	
15g14660.1	-0,259	unknown	Unknown	
15g14720.1	-1,888	unknown	Unknown	
15g14790.1	-2,242	unknown	Unknown	
15g15050.1	4,877	unknown	Unknown	
15g15130.1	-0,271	unknown	Unknown	
15g15140.1	-1,470	unknown	Unknown	
15g15270.1	1,267	unknown	Unknown	
15g15390.1	-1,274	unknown	Unknown	
15g15470.1	5,169	unknown	Unknown	
15g15500.1	0,271	unknown	Unknown	
15g15730.1	-0,092	unknown	Unknown	
15g15920.1	8,369	unknown	Unknown	
15g16000.1	1,554	unknown	Unknown	
15g16100.1	-2,291	unknown	Unknown	
15g16130.1	1,857	unknown	Unknown	
15g16180.1	2,206	unknown	Unknown	
15g16190.1	-0,944	unknown	Unknown	
15g16320.1	-0,200	unknown	Unknown	
15g16630.1	0,393	unknown	Unknown	
15g16970.1	-0,228	pigment defective 320	Unknown	
15g16980.1	-0,596	unknown	Unknown	
15g17300.1	7,234	unknown	Unknown	
15g17340.1	-0,015	unknown	Unknown	
15g17600.1	0,180	unknown	Unknown	
15g17650.1	9,001	unknown	Unknown	
15g17710.1	8,775	unknown	Unknown	
15g17920.1	-0,510	unknown	Unknown	
15g18120.1	4,816	unknown	Unknown	
15g18120.2	3,543	unknown	Unknown	
15g18150.1	-0,477	unknown	Unknown	
15g18280.1	2,551	unknown	Unknown	
15g18300.1	-0,141	unknown	Unknown	
15g18370.3	1,232	unknown	Unknown	

(Conclusion)

Table 12 - Genes in the vicinity of the QTL identified on chromosome 16 at position 35.3cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
16g26790.1	-0.701	dynein light chain type 1	Cell development / division	Cytoskeleton
16g26860.1	3.610	glucan endo-1,3-beta-glucosidase	Defense	Fungal cell wall degradation
16g26780.1	0.263	exosome complex exonuclease	Defense	R gene
16g27560.1	4.826	leucine rich repeat	Defense	R gene
16g27670.1	8.512	poly-adenylate binding protein, unique domain	DNA / RNA	Binding-RNA
16g26740.1	8.425	NAC/NAM (no apical meristem) protein	DNA / RNA	Gene regulation
16g26810.1	8.643	NAC/NAM (no apical meristem) protein	DNA / RNA	Gene regulation
16g26850.1	5.063	transcription regulator NOT2/NOT3/NOT5 familyprotein	DNA / RNA	Gene regulation
16g26870.1	-0.354	GATA zinc finger	DNA / RNA	Gene regulation
16g27070.1	2.974	tuftelin interacting protein 11	DNA / RNA	Gene regulation
16g27170.1	0.053	GATA zinc finger	DNA / RNA	Gene regulation
16g27280.1	0.178	zinc finger, C2H2 type	DNA / RNA	Gene regulation
16g26580.1	-0.029	dead box ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
16g27680.1	2.750	dead box ATP-dependent RNA helicase	DNA / RNA	Helicase / Topoisomerase
16g26890.1	2.798	gemin 2	DNA / RNA	RNA binding
16g27270.1	1.484	DNA Polymerase I	DNA / RNA	Synthesis-DNA
16g26710.1	0.665	syntaxin	Membrane	Membrane protein
16g27050.1	-0.200	syntaxin	Membrane	Membrane protein
16g26720.1	-3.095	aquaporin (major intrinsic protein family)	Membrane	Transport
16g27130.1	-0.014	aquaporin (major intrinsic protein family)	Membrane	Transport
16g27320.1	3.444	sucrose transporter	Membrane	Transport
16g27370.1	2.863	multidrug resistance pump	Membrane	Transport
16g27500.1	9.060	ABC1 family	Membrane	Transport
16g26660.1	-1.134	flowering locus T-like protein	Miscellaneous	Multiple functions
16g27460.1	0.323	nodulin-like	Miscellaneous	Nodulin
16g27480.1	3.125	multicopper oxidase	Miscellaneous	Oxidoreductase
16g26970.1	-0.582	no hits	No hits	
16g27020.2	-0.680	Cu2+/Zn2+ superoxide dismutase SOD1	Oxidative state	Redox
16g27210.1	0.076	glutathione reductase	Oxidative state	Redox
16g27100.1	1.441	tryptophan/tyrosine permease	Primary metabolism	Amino acid transporters
16g26930.1	0.336	galactose-1-phosphate uridyltransferase	Primary metabolism	Carbohydrate
16g27440.1	0.278	UDP-glucuronosyl and UDP-glucosyl transferase	Primary metabolism	Carbohydrate
16g27180.1	-2.595	N-acetyltransferase	Primary metabolism	N-acetyltransferase activity
16g26940.1	8.954	glutamate dehydrogenase 1	Primary metabolism	Nitrogen metabolism
16g27090.1	0.752	alpha-1,4-fucosyltransferase	Primary metabolism	Transferase activity
16g26920.1	0.005	chaperonin complex component, TCP-1 theta subunit (CCT8)	Protein	Folding
16g27360.1	8.235	ubiquitin carboxyl-terminal hydrolase, family 1	Protein	Modification
16g26600.1	0.220	peptide chain release factor 1 (eRF1)	Protein	Synthesis
16g27420.2	-0.967	regulator of Vps4 activity in the MVB pathway protein	Protein	Trafficking
16g26950.1	-1.031	protein-L-isoaspartate methyl transferase	Protein	
16g26650.1	9.730	anthranilate N-benzoyltransferase protein	Secondary metabolism	Phytoalexin Biosynthesis
16g27630.1	5.606	tryptophan aminotransferase related 2// Alliinase-like	Secondary metabolism	
16g27760.1	-0.291	inositol polyphosphate 5-phosphatase	Signalling	Inositol
16g27250.1	-2.037	serine/threonine protein kinase	Signalling	Kinase
16g27260.1	6.548	serine/threonine protein kinase	Signalling	Kinase
16g27380.1	-1.366	serine/threonine protein kinase	Signalling	Kinase
16g26560.1	8.614	unknown	Unknown	
16g26570.1	0.592	unknown	Unknown	
16g26590.1	-1.753	unknown	Unknown	
16g26990.1	-1.323	unknown	Unknown	
16g27190.1	-0.191	unknown	Unknown	
16g27300.1	0.445	unknown	Unknown	
16g27400.1	-0.527	unknown	Unknown	
16g27410.1	8.498	unknown	Unknown	
16g27590.1	-0.048	unknown	Unknown	

Table 13 - Genes in the vicinity of the QTL identified on chromosome 17 at position 23.6cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
17g15600.1	1.859	bax inhibitor	Defense	Apoptosis / PCD
17g14620.1	1.175	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
17g14840.1	-2.417	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
17g14850.1	2.291	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
17g14890.1	-0.362	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
17g14900.1	0.949	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
17g14910.1	3.274	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
17g14930.1	2.585	protease inhibitor/seed storage/LTP family	Defense	Protease inhibitor
17g15140.1	-0.608	zinc finger, C2H2 type	DNA / RNA	Gene regulation
17g15310.1	2.029	AP2 domain	DNA / RNA	Gene regulation
17g15330.1	-0.806	MYB transcription factor	DNA / RNA	Gene regulation
17g15380.1	0.288	homeobox protein	DNA / RNA	Gene regulation
17g15460.1	-2.342	AP2 domain	DNA / RNA	Gene regulation
17g15480.1	-2.498	AP2 domain	DNA / RNA	Gene regulation
17g15610.1	0.756	GATA zinc finger	DNA / RNA	Gene regulation
17g15320.1	8.302	acyl-CoA oxidase	Lipids	Fatty Acids
17g15530.1	-0.578	triacylglycerol lipase	Lipids	Utilization/Redox/catabolism
17g15590.1	1.523	vesicle-associated membrane protein	Membrane	Trafficking
17g14830.1	2.610	nitrate transporter	Membrane	Transport
17g15260.1	1.582	acetylglucosaminyltransferase EXT1/exostosin 1	Miscellaneous	Glycosyltransferase
17g14730.1	-0.418	multicopper oxidase	Miscellaneous	Oxidoreductase
17g14950.1	2.292	L-lactate dehydrogenase	Miscellaneous	Oxidoreductase
17g15170.1	-0.306	growth regulator like protein	Miscellaneous	Plant development
17g14650.1	1.022	elongation factor G C-terminus	Miscellaneous	
17g15090.1	-0.241	asparaginyl-tRNA synthetase	Primary metabolism	Amino acid metabolism
17g14750.1	4.704	beta-fructofuranosidase	Primary metabolism	Carbohydrate
17g14760.1	-1.672	transaldolase	Primary metabolism	Carbohydrate
17g15060.1	5.974	glycogenin	Primary metabolism	Carbohydrate
17g15080.1	-5.154	peroxisomal farnesylated protein	Primary metabolism	Lipid
17g15240.1	-2.606	retinoblastoma	Primary metabolism	Nucleic acid binding
17g15240.2	-1.880	retinoblastoma	Primary metabolism	Nucleic acid binding
17g15190.1	1.452	formyltetrahydrofolate deformylase	Primary metabolism	Nucleic acid metabolism
17g15020.1	1.018	aspartyl protease	Protein	Degradation
17g15410.1	2.459	PPPDE putative peptidase domain	Protein	Degradation
17g15720.1	0.460	ATP-dependent CLP protease	Protein	Degradation
17g15230.1	0.218	60s ribosomal protein L15	Protein	Synthesis
17g15630.1	8.820	40S ribosomal protein S24	Protein	Synthesis
17g15660.1	8.235	40S ribosomal protein S24	Protein	Synthesis
17g14680.1	0.916	miscellaneous	protein	Trafficking
17g14820.1	4.384	oxysterol-binding protein	Secondary metabolism	Terpene
17g14810.1	1.373	phosphatidic acid phosphatase-related /PAP2-related	Signalling	Phosphatase
17g15400.1	0.136	dual specificity phosphatase	Signalling	Phosphatase
17g14660.1	0.292	tetratricopeptide repeat / chaperone binding protein	Stress	Stress-inducible protein, putative
17g14800.1	-1.452	unknown	Unknown	
17g14940.1	-0.079	unknown	Unknown	
17g14980.1	1.649	unknown	Unknown	
17g14990.1	-0.262	unknown	Unknown	
17g15040.1	-0.479	unknown	Unknown	
17g15160.1	-1.053	unknown	Unknown	
17g15200.1	-1.541	phosphoinositide binding	Unknown	
17g15290.1	1.425	unknown	Unknown	
17g15300.1	3.186	unknown	Unknown	
17g15370.2	3.115	unknown	Unknown	
17g15510.1	5.332	unknown	Unknown	

Table 14 - Genes in the vicinity of the QTL identified on chromosome 19 at position 66.9cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
19g37430.1	1,763	kinase	Cell development / division	Cdc2-related protein kinase CRK2
19g37770.1	1,751	kinase	Cell development / division	Cdc2-related protein kinase CRK2
19g37250.1	4,993	pelota	Cell development / division	Cell division
19g37820.1	-2,240	peptidoglycan-binding LysM domain-containing protein	Cell development / division	Cell wall
19g38150.1	9,075	kinesin	Cell development / division	Cytoskeleton
19g37940.1	0,494	tubulin/FtsZ family protein	Cell development / division	Plant development
19g38010.1	-0,226	SOUL heme-binding family protein	Defense	*probably bacterial effector binding domain
19g37540.1	0,402	histone deacetylase 2C	DNA / RNA	Chromatin structure
19g37740.1	2,333	WRC	DNA / RNA	DNA binding
19g37410.1	1,996	transcription factor GT-2	DNA / RNA	Gene regulation
19g37580.1	0,185	heat shock transcription factor	DNA / RNA	Gene regulation
19g37660.1	0,958	transcription factor GT-2	DNA / RNA	Gene regulation
19g37050.1	1,509	transducin family protein / WD-40 repeat family protein	DNA / RNA	Gene regulation
19g37190.1	-0,234	PHD-finger	DNA / RNA	Gene regulation
19g37800.1	5,905	basic-leucine zipper (bZIP) transcription factor family protein	DNA / RNA	Gene regulation
19g37910.1	1,298	basic-leucine zipper (bZIP) transcription factor family protein	DNA / RNA	Gene regulation
19g38130.1	-0,335	small nuclear ribonucleoprotein	DNA / RNA	Processing-RNA
19g37760.1	10,291	ATP-binding cassette transporter	Energy	ATP
19g37810.1	1,543	GDSL-like lipase/acylhydrolase superfamily protein	Lipids	Degradation
19g37280.1	-1,279	polyglutamine binding protein	Miscellaneous	Cactus-binding C-terminus of cactin protein
19g38180.1	6,100	alpha/beta hydrolase fold	Miscellaneous	Catalytic function
19g37090.1	0,954	ARM/HEAT repeat	Miscellaneous	Intracellular transport
19g37920.1	0,663	f-box family protein	Miscellaneous	Lectin-like
19g37600.1	1,603	zinc finger, C2H2 type	Miscellaneous	Multiple functions
19g36880.1	4,452	zinc finger protein	Miscellaneous	Multiple functions
19g37840.1	-1,371	zinc finger protein	Miscellaneous	Multiple functions
19g36900.1	-0,075	rhodanese-like domain	Miscellaneous	Possible chimeric
19g37310.4	2,449	ubiquitin-protein ligase	Miscellaneous	Possible Chimeric
19g37780.1	-0,416	60s acidic ribosomal protein	Miscellaneous	Possible chimeric
19g37970.1	0,240	ubiquinol-cytochrome C reductase	Miscellaneous	Possible Chimeric
19g37320.1	1,972	PPR repeat	Miscellaneous	PPR domain
19g37490.1	4,292	PPR repeat	Miscellaneous	PPR domain
19g37620.1	4,412	PAN domain	Miscellaneous	Protein-protein or protein-carbohydrate interactions
19g37530.1	0,265	zinc finger C-x8-C-x5-C-x3-H type family protein	Miscellaneous	Several possible functions
19g37960.1	3,104	WD repeat containing protein	Miscellaneous	Several possible functions
19g36850.1	-0,075	no hits	No hits	
19g38040.1	-0,153	glutaredoxin	Oxidative state	Antioxidant
19g37500.2	-0,539	peroxiredoxin	Oxidative state	Oxidase
19g37420.1	0,171	pyruvate kinase	Primary metabolism	Carbohydrate
19g37520.1	-0,680	enrolase	Primary metabolism	Carbohydrate
19g38160.1	-2,681	beta-fructofuranosidase	Primary metabolism	Carbohydrate
19g36860.1	-0,086	molybdopterin	Primary metabolism	Cofactor
19g37930.1	4,694	iron-sulphur cluster biosynthesis	Primary metabolism	Nitrogen fixation
19g37990.1	1,136	glucose-6-phosphate acetyltransferase 1	Primary metabolism	
19g37260.1	-0,465	aspartyl protease	Protein	Degradation
19g38170.1	-7,767	ubiquitin	Protein	Degradation
19g37100.1	3,906	glucosyl/glucuronosyl transferases	Protein	Synthesis
19g37120.1	3,944	glucosyl/glucuronosyl transferases	Protein	Synthesis
19g37130.1	-0,325	glucosyl/glucuronosyl transferases	Protein	Synthesis
19g37140.1	3,847	glucosyl/glucuronosyl transferases	Protein	Synthesis
19g37550.1	4,579	threonyl-tRNA synthetase, putative / threonine--tRNA ligase, putative	Protein	Synthesis
19g37880.1	1,548	golgi membrane protein YIP1	Protein	Traffiking
19g38110.1	-1,488	signal recognition particle 54 kDa subunit precursor	Protein	Transport
19g37790.1	-1,560	IQ calmodulin-binding motif family protein	Signalling	Calcium
19g37020.1	0,000	RAB GTPase homolog C2A	Signalling	GTP-related
19g38120.1	1,191	GTPase-activator protein	Signalling	GTP-related
19g37850.1	8,235	acid phosphatase related	Signalling	PA
19g37860.1	0,966	acid phosphatase related	Signalling	PA
19g36950.1	0,186	protein phosphatase 2A	Signalling	Phosphatase
19g36960.2	0,083	unknown	Unknown	
19g37450.1	4,898	unknown	Unknown	
19g37460.1	4,351	unknown	Unknown	
19g37610.1	0,370	unknown	Unknown	
19g37630.1	-1,047	unknown	Unknown	
19g37950.1	2,142	unknown	Unknown	
19g38020.1	-2,510	unknown	Unknown	
19g38060.1	0,707	unknown	Unknown	

Table 15 - Genes in the vicinity of the QTL identified on chromosome 20 and position 24.3-29.6cM

GlymaID	lg2 IAC/CD	Gene call	Category	Sub category
20g19640.1	8,710	kinase	Cell development / division	Cdc2-related protein kinase CRK2
20g17500.1	0,443	exo70 exocyst complex subunit	Cell development / division	Exocytosis
20g19210.1	0,854	inosine triphosphate pyrophosphatase	DNA / RNA	Synthesis
20g17040.1	-0,236	GRAM domain family protein	Membrane	Membrane associated
20g18870.1	-1,135	ABC1 family	Membrane	Transport
20g18420.2	-2,593	f-box domain	Miscellaneous	Lectin-like
20g19200.1	1,601	pectate lyase	Miscellaneous	Nodulation
20g17400.1	8,286	NPH3 family	Miscellaneous	Phototropism
20g18890.2	-1,714	ankyrin repeat	Miscellaneous	Protein-protein interactions
20g18900.1	9,394	ankyrin repeat	Miscellaneous	Protein-protein interactions
20g18010.1	-0,140	PPR repeat	Miscellaneous	Several possible functions
20g18550.1	4,723	p-loop containing nucleoside triphosphate hydrolases superfamily protein	Miscellaneous	
20g19000.1	0,475	aldo/keto reductase family	Oxidation	Redox
20g17990.1	1,079	ureD urease accessory protein	Primary metabolism	Nitrogen
20g17440.1	0,544	uricase	Primary metabolism	Nitrogen fixation
20g17280.1	6,004	core-2/I-branching beta-1,6-N-acetylglucosaminyltransferase family protein	Primary metabolism	
20g17090.1	3,579	ubiquitin	Protein	Degradation
20g17560.1	5,392	ATP-dependent CLP protease	Protein	Degradation
20g17960.1	0,640	GTP-binding protein-related	Signalling	G protein
20g18290.1	-1,539	Uncharacterized conserved protein, AMMECR1	Unknown	
20g19620.1	0,841	unknown	Unknown	

Plants possess a vast arsenal of defenses to protect them against insect damage (WALLING, 2000). Defenses include physical and chemical approaches, some induced upon insect attack, others constitutively present. In this study, several genes that might be associated with defense based on their annotations, were constitutively differentially expressed between the resistant and susceptible cultivars within QTL associated with resistance. Some of these genes will be discussed below.

Chemical defenses are well known in plants. One such chemical class consists of the terpenes, which have been associated with pathogen and insect resistance (GERSHENZON; DUDAREVA, 2007). A protein family that might play a role in terpene metabolism, are the oxysterol-binding proteins. A gene encoding for an oxysterol-binding proteins, was identified within a QTL on chromosome 17. On chromosome 16, there is a QTL containing a gene encoding an alliinase-like protein, an enzyme that synthesizes alliin, a sulfoxide known to be toxic to many insects as well as to microbes (DUGRAVOT et al., 2005). Transcripts for both of these putative defense related genes were present at levels approximately 20-40 fold (log2 4.4 and 5.6) in the resistant genotype versus the susceptible.

Constitutive defenses also include physical barriers like the cell wall that are very important for protection against pathogen and insect pests. In this study, several genes related to the cell-wall category were differently expressed at least 2 fold between parents. These differentially expressed genes included ones encoding a rhamnogalacturonate lyase, a hydroxyproline-rich glycoprotein family protein, and a pectin lyase-like superfamily protein.

RNA sequencing of the two parents identified 54 genes that were classified as DNA or RNA related that had a strong (log2$\geq$2 or $\leq$-2) differential expression between the two

genotypes, with 47 of these genes showing a higher expression in IAC-100 than in CD-215. Among these upregulated genes, one was a MYB-related transcription factor and two were homologs to WRKY transcription factors. Genes involved in production of jasmonic acid, a well known inducer of defenses against insects (FARMER; RYAN, 1990) may be regulated by WRKY type transcription factors, and a gene encoding a putative WRKY was found to be expressed about 4 fold (log2 2.1) higher in the resistant genotype on chromosome 15. In addition to identifying these genes that are related to DNA and RNA, the study also identified 42 genes related to signal transduction, any of which could be having an effect on defense levels.

Of the differentially expressed genes that were located within the QTLs, 16 of them had functional annotations suggesting that they are related to defense and also showed a higher abundance (fold change > 2) in resistant IAC-100 than in CD-215. These genes were annotated as being homologous to leucine rich repeat (LRR) proteins, metaxin, amine oxidase, harpin-induced protein 1 (Hin1), protease inhibitor/seed storage/LTP family, and pathogenesis-related protein Bet v I family. LRR domains are often found in resistance (R) genes and are believed to trigger apoptosis or a hypersensitive-type resistance reaction in response to microbe-specific stimuli known as effector proteins. The majority of R genes encode a nucleotide binding site (NBS) in addition to the LRR domains. LRRs are multiple, serial repeats of a motif of approximately 24 amino acids in length. LRRs contain leucines or other hydrophobic residues at regular intervals and can also contain regularly spaced prolines and asparagines (BENT, 1996). The first cloned insect resistance gene belongs to the NBS-LRR class of R gene. The *Mi-1* gene is from wild tomato (*Lycopersicon peruvianum*) and confers resistance to some isolates of the potato aphid (*Macrosiphum euphorbiae*) and silverleaf whitefly (*Bemisia tabaci*) (ROSSI et al., 1998; NOMBELA; WILLIAMSON; MUNIZ, 2003). The soybean aphid resistance gene, *Rag1*, may also be an LRR type R gene as the region that *Rag1* is fine mapped to, is full of LRR proteins (KIM et al, 2010). Supporting that *Rag1* might be an LRR-type R gene was the study by Li et al. (2008) where the authors examined gene expression from aphid (*A. glycines*) resistant (Dowling) and susceptible (Williams 82) soybean cultivars without insect treatment to identify genes with constitutive expression differences. At the start of the experiment (T0), 68 genes showed higher expression in resistant (R) than in susceptible (S) cultivar. However, only five genes showed constitutively higher expression in R than in S throughout the three time points (0, 6, and 12 hours post-application). Among these genes, in addition to a MYB factor homolog, was a homolog of the soybean NBS-LRR type resistance gene *RPG1-b*.

Another class of proteins typically associated with insect resistance in plants are the protease inhibitors (RYAN, 1990), which are antidigestive proteins that influence herbivore performance by inhibiting insect digestive enzymes (KESSLER; BALDWIN, 2002). Six protease inhibitor/seed storage/LT family genes were differentially expressed between the two genotypes. Three of these six were downregulated in the resistant genotype, while three were upregulated. The upregulated protease inhibitors were located within the same QTL on chromosome 17.

A total of 190 genes were categorized as having unknown function, making it the largest category. Genes were placed in this category if they failed to match any sequences from the Phytozome and NCBI or matched to sequences whose functions have not yet been characterized. Some of these genes were expressed to a much higher level in the resistant IAC-100 than in CD-215. Another 111 genes were classified as miscellaneous, as they had either multiple possible annotations, or encoded for proteins with multiple functions making them difficult to categorize precisely. Moreover, six genes did not have significant homology with any record on the databases, so they were classified as no hits.

4.3.3 Differentially expressed genes between IAC-100 and CD-215 infested with stink bug under field conditions

RNA from the field experiment was collected, purified, DNase treated, and sent for sequencing in November 2011. However, due to a back-log of sequencing requests, the sequencing center was not able to begin processing the samples until January 2012. Therefore, results from these analyses will not be available to include in this thesis, but instead will be incorporated into a manuscript describing this study.

4.4 Conclusions

- Gene expression responses can be detected in soybean within 24 hours of insect feeding.
- *Piezodorus guildinni* induced a stronger gene expression response than did *Nezara viridula.*
- Candidate defense related genes were found that were differentially expressed and located within 500 kb upstream or downstream of known QTL markers.

- Genes related to terpenes, alliin, R genes, protease inhibitors, MYBs, and WRKYs are the best candidate defense-related genes associated with the mapped QTL for resistance.
- RNA-seq expression data from the field experiment could not be processed in time to be included in this thesis, and will be analyzed at a later date.

References

BENT, A.F. Plant disease resistance genes: Function meets structure. **The Plant Cell**, Rockville, v. 8, p. 1757-1771, 1996.

CALLA, B.; VUONG, T.; RADWAN, O.; HARTMAN, G. L.; CLOUGH, S. J. Gene expression profiling soybean stem tissue early response to *Sclerotinia sclerotiorum* and *in silico* mapping in relation to resistance markers. **The Plant Genome**, Madison, v. 2, n. 2, p. 149-166, 2009.

CASTEEL, C.L.; O'NEILL, B.F.; ZAVALA, J.A.; BILGIN, D.D.; BERENBAUM, M.R.; DeLUCIA, E.H. Transcriptional profiling reveals elevated CO_2 and elevated O_3 alter resistance of soybean (*Glycine max*) to Japanese beetles (*Popillia japonica*). **Plant, Cell and Environment**, Oxford, v. 31, n. 4, p.419-434, 2008.

CLOUGH, S.J.; VODKIN, L.O. Soybean microarrays: a genomic tool for crop improvement. In: WILSON, R. F.; BRUMMER, C. **Genomics for legume crops**. Urbana: American Oil Chemists Society Press, 2004. p. 267-282

CORRÊA-FERREIRA, B.S.; AZEVEDO, J. de. Soybean seed damage by different species of stink bugs. **Agricultural and Forest Entomology**, Malden, v. 4, n. 2, p. 145-150, 2002.

DUGRAVOT, S.; MONDY, N.; MANDON, N.; THIBOUT, E. Increased sulfur precursors and volatiles production by the leek *Allium porrum* in response to specialist insect attack. **Journal of Chemical Ecology**, New York, v. 31, n. 6, 2005.

EISEN, M.B.; SPELLMAN, P.T.; BROWN, P.O.; BOSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences**. Washington, v. 95, p. 14863-14868, 1998.

FARMER, E.E.; RYAN, C.A. Interplant communication: Airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. **Proceedings of the National Academy of Sciences**, Washington, v. 87, p. 7713-7716, 1990.

GERSHENZON, J.; DUDAREVA, N. The function of terpene natural products in the natural world. **Nature Chemical Biology**, Cambridge, v. 3, n. 7, p. 408-414, 2007.

HEGDE, P.; QI, R.; ABERNATHY, K.; GAY, C.; DHARAP, S.; GASPARD, R.; HUGHES, J.E.; SNESRUD, E.; LEE, N.; QUACKENBUSH, J. A concise guide to cDNA microarray analysis. **BioTechniques**, Natick, v. 29, p. 548-562, 2000.

KEMPEMA, L.A.; CUI, X.; HOLZER, F.M.; WALLING, L.L. *Arabidopsis* transcriptome changes in response to phloem-feeding silverleaf whitefly nymphs. Similarities and distinctions in responses to aphids. **Plant Physiology**, Waterbury, v. 143, n. 1, p. 849-865, 2007.

KESSLER, A.; BALDWIN, I. T. Plant responses to insect herbivory: the emerging molecular analysis. **Annual Review of Plant Biology**, Palo Alto, v. 53, p. 299-328, 2002.

KIM, K.S.; BELLENDIR, S.; HUDSON, K.A.; CURTIS B. HILL, C.B.; GLEN L. HARTMAN, G.L.; DAVID L. HYTEN, D.L.; HUDSON, M.E.; · BRIAN W. DIERS, B.W. Fine mapping the soybean aphid resistance gene *Rag1* in soybean. **Theoretical and Applied Gene**tics, Berlin, v. 120, p.1063–1071, 2010.

LI, Y.; ZOU, J.; LI, M.; BILGIN, D.D.; VODKIN, L.O.; HARTMAN, G.L.; CLOUGH, S.J. Soybean defense responses to the soybean aphid. **New Phytologist**, Lancaster, v.179, n. 1, p. 185-195, 2008.

LOURENÇÃO, A.L.; MIRANDA, M.A.C. Resistência de soja a insetos: VIII. IAC 78-2318, Linhagem com resistência múltipla. **Bragantia**, Campinas, v. 46, n.1, p. 65-72, 1987.

NOMBELA, G.; WILLIAMSON, V.M.; MUNIZ, M. The root-knot nematode resistance gene *Mi-1.2* of tomato is responsible for resistance against the whitefly *Bemisia tabaci*. **Molecular Plant-Microbe Interaction**, Saint Paul, v. 16, p. 645-649, 2003.

VARSHNEY, R.K.; GRANER, A.; SORRELLS, M.E. Genomics-assisted breeding for crop improvement. **Trends in Plant Science,** London, v.10, n. 12, p. 621-630, 2005.

ROSSETTO, C.J.; GALLO, P.B. , RAZERA, L.F.; BORTOLETTO, N.; IGUE, T.; MEDINA, P. F.; TISSELLI FILHO, O.; AQUILERA, V.; VEIGA, R.F.A.; PINHEIRO, J. B. Mechanisms of resistance to stink bug complex in the soybean cultivar 'IAC-100'. **Anais da Sociedade Entomológica do Brasil**, Londrina, v. 24, p. 517-522, 1995.

ROSSI, M.; GOGGIN, F.L.; MILLIGAN, S.B.; KALOSHIAN, I.; ULLMAN, D.E.; WILLIAMSON, V.M. The nematode resistance gene *Mi* of tomato confers resistance against the potato aphid. **Proceedings of the National Academy of Sciences**, Washington, v. 95, p. 9750-9754, 1998.

RYAN, C.A. Protease inhibitors in plants: genes for improving defenses against insects and pathogens. **Annual Review of Phytopathology**, v. 28, p. 425-49, 1990.

SCHENA M.; SHALON, D.; DAVIS, R.W.; BROWN, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science**, Washington, v. 270, p. 467–470, 1995.

SMITH, C.M.; BOYKO, E.V. The molecular bases of plant resistance and defense responses to aphid feeding: current status, **Entomologia Experimentalis et Applicata**, Dordrecht, v. 122, p.1-16, 2007.

STEARS, R.L.; MARTINSKY, T.; SCHENA, M. Trends in microarray analysis. **Nature Medicine**, New York, v. 9, n. 1, p. 140-145, 2003.

WALLING, L.L. The myriad plant responses to herbivores. **Journal of Plant Growth Regulation**, New York, v. 19, p. 195-216, 2000.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, London ,v. 10, n. 1, p.57-63, 2009.

WILHELM, B.T.; LANDRY, J.R. RNA-Seq - quantitative measurement of expression through massively parallel RNA-sequencing. **Methods**, San Diego, v. 48, p. 249–257, 2009.

ZOU, J.J.; RODRIGUEZ-ZAS, S.; ALDEA, M.; LI, M.; ZHU, J.; GONZALEZ, D. O.; VODKIN, L.O.; DELUCIA, E.; CLOUGH, S.J. Expression profiling soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-specific downregulation of photosynthesis. **Molecular Plant–Microbe Interactions,** Saint Paul, v. 18, n. 11, p. 1161–1174, 2005.