

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

Expressão gênica associada à resistência da soja a *Piezodorus guildinii*

Ana Paula Mendes Silva

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

**Piracicaba
2014**

Ana Paula Mendes Silva
Bacharel em Ciências Biológicas

Expressão gênica associada a resistência da soja a *Piezodorus guildinii*

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

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

**Piracicaba
2014**

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

Silva, Ana Paula Mendes
Expressão gênica associada à resistência da soja a *Piezodorus guildinii* / Ana Paula
Mendes Silva. - - Piracicaba, 2014.
119 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz", 2014.
Bibliografia.

1. *Glycine max* 2. Transcriptoma 3. Resistência à insetos 4. RNA-Seq I. Título

CDD 633.34
S586e

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

A Deus pela oportunidade de viver.

Aos meu pais, meus exemplos de determinação, competência e sabedoria.

Aos meus parentes e amigos pelo apoio e energias positivas.

AGRADECIMENTOS

À Deus pelo dom da vida, por todas as graças a mim concedida e principalmente pela oportunidade de vencer mais essa etapa.

Aos meus pais, Lúcio e Maria Luiza, e meu irmão Rafael, por me apoiarem incondicionalmente, serem meus exemplos de superação e comprometimento e por me dedicarem todo o amor.

Ao Prof. Dr. José Baldin Pinheiro, pela orientação, ensinamentos, confiança e amizade durante esses dois anos de convivência.

A CAPES pelas bolsas de estudo concedidas para o Mestrado.

Ao Dr. Steven J. Clough por me receber em seu laboratório e me orientar nos momentos mais importantes das minhas análises da dissertação, contribuindo para a minha formação científica e principalmente pela amizade.

À Dra. Michelle da Fonseca Santos pelo apoio inicial através do fornecimento de artigos e material para o meu estudo, por disponibilizar seus dados de doutorado para a minha análise da dissertação, e principalmente por me ajudar no momento que eu mais precisava de orientação e suporte para o término das análises.

Ao Mestrando Lucas Mitsuo Taniguti pelas análises de bioinformática, pelos ensinamentos e suporte no desenvolvimento do trabalho e nas dúvidas até altas horas da madrugada, e pela amizade.

Ao Márcio Arruda por me ajudar nos momentos finais e cruciais das discussões sobre os resultados.

À Doutoranda Aline Rocha pela amizade e apoio.

A todos do Laboratório de Diversidade Genética e Melhoramento pelos momentos divididos durante esses dois anos de pós-graduação e pela ajuda em muitas etapas de concretização do meu projeto. Eu agradeço em especial ao Marcos, Carolline Pires, João Paulo, Clesivan, Alessandro, Felipe, Eleonora, Miklos, Carolina Grando, Fabiana, Fabiani e Mônica.

À Dra. Maria Imaculada Zucchi pela amizade.

Aos professores e técnicos do Departamento de Genética pelos ensinamentos e imenso suporte, em especial aos funcionários Marcio Araújo Silva e Cláudio Roberto Segatelli.

Aos meus orientadores e co-orientadores de graduação Everaldo de Gonçalves de Barros, Marcos Oliveira Pinto e Carla Quinhones por orientarem os meus primeiros passos na carreira acadêmica e científica.

Ao Marcos Araújo por toda a ajuda em todas as etapas do meu projeto, desde a parte prática de laboratório até a parte escrita, pelo carinho e incentivo que não me deixaram desanimar diante das dificuldades de fim de curso e pelo amor.

À minha amiga Vanessa Rizzi que me ajudou nos momentos de angústia e medo, me orientou e guiou meus passos na escrita e formatação da dissertação.

Às minhas irmãs de república Elisângela e Gilda pelo carinho e amizade nesse tempo de convivência.

Aos amigos do Yoga que tornaram meus dias mais prazerosos em Piracicaba.

A todos os meus familiares e amigos que mesmo distantes estão sempre torcendo por mim, em especial aos meus avós Dalísio, Luiza, Lúcio e Lourdes; meus tios e tias; aos primos Soninha, Brenda, Paul e Gabryel; e à família que escolhi e que guardo sempre do lado esquerdo do peito Prímula Viana, Leonardo Corrêa, Pedro Lacerda e Luiz Henrique.

EPÍGRAFE

“A natureza reservou para si tanta liberdade que não a podemos nunca penetrar completamente com o nosso saber e a nossa ciência.”

Johann Goethe

SUMÁRIO

RESUMO	11
ABSTRACT	13
LISTA DE FIGURAS	15
LISTA DE TABELAS	17
1 INTRODUÇÃO.....	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 Cultura e melhoramento de soja no Brasil.....	21
2.2 Complexo de percevejos.....	22
2.3 Variedades de soja	25
2.4 Estresses	26
2.4.1 Mecanismos de defesas em plantas contra herbivoria de insetos.....	27
2.4.1.1 Ordem <i>Hemiptera</i>	28
2.5 Novas tecnologias de sequenciamento	29
2.5.1 RNA-Seq	30
2.6 Bioinformática	33
2.6.1 TopHat e Cufflinks	34
2.6.2 STAR.....	37
2.6.3 iPATH.....	38
2.6.4 Blast2GO	39
2.6.5 Phytozome	39
3 OBJETIVOS.....	41
4 MATERIAL E MÉTODOS.....	43
4.1 Características do Experimento	43
4.2 Análises de Bioinformática	43
5 RESULTADOS	49
5.1 Resultados obtidos com o alinhador Tophat	49
5.2 Resultados obtidos com o alinhador <i>STAR</i>	74
6 DISCUSSÃO	99
7 CONCLUSÕES	105
REFERÊNCIAS	107

RESUMO

Expressão gênica associada à resistência da soja a *Piezodorus guildinii*

A soja é uma cultura de grande importância, movimentando aproximadamente 230 bilhões de dólares em todo o mundo, com produção anual estimada de 267,61 milhões de toneladas. Os percevejos sugadores de sementes são considerados uma das pragas de maior importância para a cultura da soja, sendo as espécies *Euschistus heros* (E.), *Piezodorus guildinii* (West.) e *Nezara viridula* (L.) as mais abundantes no Brasil. O ataque por percevejos causa diversos problemas como o atraso da maturação fisiológica, retenção foliar, perdas no rendimento e diminuição da qualidade e potencial germinativo das sementes. Esses insetos são responsáveis ainda pela transmissão de patógenos, e podem causar alterações na composição de óleos, proteínas e ácidos graxos das sementes. A obtenção de cultivares com resistência genética é indispensável, a fim de minimizar a necessidade de utilização de defensivos agrícolas. Dessa forma, este trabalho identificou genes possivelmente associados à resposta de resistência pela comparação da expressão gênica diferencial entre genótipos de soja resistente (IAC-100) e suscetível (CD-215), a partir da metodologia de sequenciamento de RNA (RNA-Seq). As vagens foram submetidas a dois tratamentos por 24 horas: infestação por percevejos *Piezodorus guildinii* (West.) e não infestação. Foi utilizado o alinhador STAR contra o genoma *hardmasked* do *phytozome* e o pacote *cufflinks*. Com os resultados das análises dos genes diferencialmente expressos e uma posterior filtragem foi possível identificar 128 genes candidatos, dentre estes: genes relacionados com defesa, metabolismo secundário de produção de terpenos e proteínas LRR de sinalização celular.

Palavras-chave: *Glycine max*; Transcriptoma; Resistência à insetos; RNA-Seq

ABSTRACT

Gene expression associated with soybean resistance to *Piezodorus guildinii*

Soybeans are an important crop with a world-wide annual production of 267.61 tons, worth 230 billion dollars. Stink bugs are a major soybean pest. The most abundant species in Brazil are *Euschistus heros* (E.), *Piezodorus guildinii* (West.) and *Nezara viridula* (L.). Stink-bug attack causes several problems, including delayed physiological maturity, leaf retention, yield losses, decreased seed quality, and decreased germination potential. The bugs are responsible for transmitting pathogens, and they can cause changes in the oil, protein, and fatty acids composition in the seed. Producing genetically resistant cultivars is critical in order to minimize the need for agricultural pesticides. Our work aims to identify genes associated with the stink-bug resistance response in soybean by comparing differential gene expression between resistant (IAC-100) and susceptible (CD-215) genotypes by analyzing RNA sequencing (RNA-Seq) data. Data was from soybean pods that were subjected to two treatments for 24 hours: infestation by the stink bug *Piezodorus guildinii* (West.), and non-infestation. *STAR* aligner was used against the hard-masked genome sequence from Phytozome, followed by analysis using the Cufflinks package. Analysis of the results of the differentially expressed genes and a subsequent filtering identified 128 candidate defense-associated genes, among them, defense-related such as LRR-RLKs and production of secondary metabolites like terpenes.

Keywords: *Glycine max*; Transcriptomic; Insect resistance; RNA-Seq

LISTA DE FIGURAS

Figura 1 - Diferentes níveis de infestação de percevejos durante o período de cultivo da soja	24
Figura 2 - Sequência de análises pelo <i>TopHat</i> e <i>Cufflinks</i> (TRAPNELL, 2012)	35
Figura 3 - Análises dos dados de RNA-Seq.....	46
Figura 4 - Vias metabólicas ativadas pelos transcritos expressos na comparação entre as amostras IAC-100 Infestada X IAC-100 Não Infestada.....	72
Figura 5 - Vias metabólicas ativadas pelos transcritos expressos na comparação entre as amostras CD-215 Infestada X CD-215 Não Infestada.....	73
Figura 6 - Mapa das vias metabólicas feito a partir dos transcritos diferencialmente expressos entre os tratamentos de Infestação e Não Infestação na variedade IAC-100.....	96
Figura 7 - Mapa das vias metabólicas feito a partir dos transcritos diferencialmente expressos entre os tratamentos de Infestação e Não Infestação na variedade CD-215.....	97

LISTA DE TABELAS

Tabela 1 - Números de pares de reads para cada amostra.....	44
Tabela 2 - Transcritos diferencialmente expressos, em ordem decrescente, resultantes da filtragem.....	50
Tabela 3 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador <i>TopHat</i>	55
Tabela 4 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador <i>TopHat</i>	59
Tabela 5 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador <i>TopHat</i>	63
Tabela 6 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador <i>TopHat</i>	67
Tabela 7 - Os 128 transcritos diferencialmente expressos somente na variedade IAC-100 infestada, em ordem decrescente – alinhador <i>STAR</i>	75
Tabela 8- Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador <i>STAR</i>	80
Tabela 9 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador <i>STAR</i>	84
Tabela 10 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador <i>STAR</i>	88
Tabela 11 - Ranking dos 100 primeiros genes que tiveram a mudança menos drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador <i>STAR</i>	92

1 INTRODUÇÃO

O Brasil é o segundo maior produtor de soja do mundo, com uma produção de 81,456 milhões de toneladas na safra 2012/13 (COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB, 2013). O cultivo da soja é realizado, principalmente, em grandes áreas agrícolas com plantio sucessivo, o que tem contribuído significativamente para o aumento na incidência de insetos-praga, causando danos crescentes à produção (MOURA, 1999). Dentro desse sistema intensivo de produção as características desejáveis que contribuem positivamente para aumentar a estabilidade e potencial de rendimento de cultivares de soja são: maior resistência às doenças limitantes, aos insetos-pragas e aos nematoides associados à cultura; boa resistência ao acamamento e a deiscência precoce; boa qualidade fisiológica da semente; adaptação a condições locais de ambiente e ideótipo de planta adequada ao sistema agrícola utilizado na região produtora (VERNETTI, 1983).

Os dois principais grupos de insetos-pragas causadores de prejuízos econômicos às lavouras de soja são constituídos pelos percevejos fitófagos da família Pentatomidae, *Euschistus heros* (F.), *Piezodorus guildinii* (West.) e *Nezara viridula* (L.), respectivamente conhecidas como, percevejo-marrom, percevejo pequeno e percevejo-verde, e pelas lagartas desfolhadoras *Helicoverpa armigera* e *Pseudoplusia (Chrysodeixis) includens* VENZON; RIPPOSATI; FERREIRA, 1999; HOFFMANN-CAMPO; MOSCARDI; CORRÊA-FERREIRA, 2000).

Tais percevejos alimentam-se diretamente nas vagens das plantas. Através da inserção de seus estiletes nos grãos de soja (CORRÊA-FERREIRA, 2005). Quando o aparelho bucal atinge as sementes, há a danificação dos tecidos, tornando os grãos murchos e enrugados, afetando, assim, a produção e a qualidade da soja (PANIZZI; SLANSKY-JÚNIOR, 1985). Nesse processo de infestação, o mecanismo de punctura facilita a contaminação dos grãos por patógenos que podem causar alterações na composição de óleos, proteínas e ácidos graxos das sementes, além de causar distúrbios fisiológicos, como retenção foliar, atraso da maturação fisiológica, perdas no rendimento e diminuição da qualidade e potencial germinativo das sementes (TODD; HERZOG, 1980).

O controle desses insetos-praga nas lavouras é feito com a aplicação sucessiva de defensivos agrícolas (QUINTANELA; FERREIRA; GUIMARÃES, 2006). Uma vez iniciado o processo de seleção natural dos insetos resistentes aos defensivos aplicados, aumenta-se a necessidade de aplicação de inseticidas nas lavouras e mesmo do desenvolvimento de defensivos cada vez mais agressivos ao meio ambiente (MORALES; SILVA, 2006; PANIZZI, 2006a; MOSCARDI; CORRÊA-FERREIRA; SOSA-GÓMEZ, 2009). A obtenção

de cultivares de soja com resistência genética a esses insetos torna-se indispensável a fim de minimizar a necessidade de utilização de defensivos agrícolas, desta forma reduzindo os custos da produção e limitando os danos ao meio ambiente (SOSA-GÓMEZ; CORSO; MORALES, 2001).

Existem cultivares de soja com resistência fisiológica ao complexo de percevejos como a IAC-100 (ROSSETTO, 1989), IAC-17, IAC-19 (VALLE; LOURENÇÃO, 2002) e IAC-24 (MIRANDA; BRAGA; LOURENÇÃO 2003; FUGI; LOURENÇÃO; PARRA, 2005). A combinação desse tipo de resistência com características econômicas desejáveis e elevado potencial produtivo, minimizaria os impactos ambientais pela redução do uso de inseticidas e pela redução da necessidade de expansão das fronteiras agrícolas da cultura da soja.

Tal combinação das características de resistência ao complexo de percevejos com qualidades agroeconômicas almejadas somente é viável através da utilização de metodologias biotecnológicas. Algumas destas metodologias são a análise dos genomas, transcriptomas e metabolomas por meio de sequenciamentos e comparação com bases de dados, a clonagem gênica e técnicas de cultura de tecidos (CORRÊA-FERREIRA; PANIZZI, 1999)

O entendimento dos mecanismos biológicos de regulação da expressão gênica, a partir do estudo dos transcritos expressos, da identificação dos genes e do estabelecimento dos processos metabólicos ativados ou desativados por eles, frente a determinadas condições torna possível uma futura manipulação biotecnológica desses genes, a fim de auxiliar os programas de melhoramento como o da soja (WAR, 2012).

A fim de entender melhor esses mecanismos, este trabalho objetivou estudar a expressão gênica da cultivar resistente IAC-100 e da cultivar susceptível CD-215 nas condições de infestação e não infestação com *Piezodorus guildinii*, bem como, investigar os genes responsáveis pela resistência fisiológica e estabelecer as vias metabólicas ativadas no processo de infestação.

A avaliação do transcriptoma de ambas as cultivares no presente estudo mostra quais genes são diferencialmente expressos quando submetidos à infestação e não obstante relaciona quais genes tem a sua expressão regulada negativamente ou positivamente.

2 REVISÃO BIBLIOGRÁFICA

2.1 Cultura e melhoramento de soja no Brasil

A soja (*Glycine max* L.) é uma das culturas mais importantes do mundo, sendo usualmente comercializada em grão, farelo e óleo bruto e refinado, constituindo grande fonte de proteína para toda a humanidade, além de apresentar grande potencial na fabricação de plásticos e biodiesel (SECRETARIA DE ESTADO DE AGRICULTURA E ABASTECIMENTO - SEAB, 2013). No mundo todo o agronegócio da soja movimenta aproximadamente 230 bilhões de dólares. Atualmente, são produzidas por ano, cerca de 267,61 milhões de toneladas deste grão (UNITED STATES DEPARTMENT OF AGRICULTURE - USDA, 2013). Nesse aspecto, o Brasil aparece como o segundo maior produtor, graças ao uso crescente e intensivo de tecnologias (ROESSING; SANCHES; MICHELLON, 2005), produzindo cerca de 66 milhões de toneladas (safra 2011/12). De acordo com a Companhia Nacional de Abastecimento - CONAB, a área cultivada com soja no Brasil na safra 2012/13 foi de 27,721 milhões de hectares, o que corresponde a um crescimento de 10,7%, em relação à safra anterior; e a produção foi de 81,456 milhões de toneladas, volume este superior em 22,7% (CONAB, 2013).

A cultura da soja, inicialmente cultivada apenas na Região Sul do Brasil, expandiu-se para o Cerrado e mais recentemente para a Região Norte e Nordeste do País. O uso de alta tecnologia de produção contribuiu para essa expansão da indústria da soja no país. Uma das consequências das grandes áreas cultivadas com soja está relacionada ao aumento de pragas que atacam essa oleaginosa, causando danos crescentes à produção (MOURA, 1999). Além disso, o processo de melhoramento genético visando à produtividade de grãos e à qualidade das plantas ou de seus derivados pode torná-las ainda mais vulneráveis a insetos-praga (PANIZZI; PANIZZI; BAYS, 1986; LUSTOSA, ZANUNCIO; LEITE, 1999). Nas diversas regiões produtoras do país, os percevejos sugadores das vagens, juntamente com as lagartas desfolhadoras, constituem os dois principais grupos de pragas que normalmente causam prejuízos econômicos às lavouras (VENZON; RIPPOSATI; FERREIRA, 1999).

Plantas transgênicas estão sendo desenvolvidas afim de amenizar os problemas ocasionados por essas pragas. A Soja Intacta RR2 PRO, aprovada em 2010 pela Comissão Técnica Nacional de Biossegurança – CTNBio, foi a primeira soja transgênica resistente a diversos tipos de lagartas como *Anticarsia gemmatalis*, lagarta falsa medideira (*Chrysodeixis includens* e *Rachiplusia nu*), lagarta das maçãs (*Heliothis virescens*) e broca das axilas ou broca dos ponteiros (*Crociosema aporema*), além de supressão às lagartas do tipo Elasmo

(*Elasmopalpus lignosellus*) e *Helicoverpa* (*H. zea* e *H. armigera*). A CTNBio é uma instância colegiada multidisciplinar, criada através da lei nº 11.105, de 24 de março de 2005, cuja finalidade é prestar apoio técnico consultivo e assessoramento ao Governo Federal na formulação, atualização e implementação da Política Nacional de Biossegurança relativa a Organismos Geneticamente Modificados (OGM), bem como no estabelecimento de normas técnicas de segurança e pareceres técnicos referentes à proteção da saúde humana, dos organismos vivos e do meio ambiente, para atividades que envolvam a construção, experimentação, cultivo, manipulação, transporte, comercialização, consumo, armazenamento, liberação e descarte de OGM e derivados (<http://www.ctnbio.gov.br/>).

Atualmente tem-se um total de 37 cultivares transgênicas de diferentes espécies como milho, feijão, algodão e soja liberados sendo comercializados. Em 1998 aconteceu a primeira aprovação de soja transgênica pela CTNBio, a Soja Roundup Ready, requerida pela Monsanto, que tinha a característica de ser tolerante ao Herbicida Glifosato. Atualmente existem mais quatro cultivares transgênicas comercializadas, sendo elas Soja Cultivance, Soja Liberty Link TM (PAT1), Soja Liberty Link TM (PAT2) e Soja Intacta RR2 PRO (http://www.ctnbio.gov.br/upd_blob/0001/1801.pdf).

2.2 Complexo de percevejos

O complexo de percevejos sugadores composto por *Euschistus heros*, *Nezara viridula* e *Piezodorus guildinii* causa danos significativos à produtividade dos grãos de soja (PANIZZI; SLANSKY JÚNIOR, 1985; CORRÊA-FERREIRA, 1999; PANIZZI; McPHERSON; JAMES, 2000). O percevejo-verde-pequeno, *P. guildinii* (West.) tem ampla distribuição ocorrendo tanto no Sul do Brasil como nas regiões produtoras de soja do Norte e Nordeste (CORRÊA-FERREIRA; PANIZZI, 1999). Embora as populações desse percevejo não atinjam níveis muito elevados, é a espécie que causa os maiores danos à qualidade das sementes de soja. Uma vez que, sua predação ocasiona uma retenção foliar (senescência anormal) das plantas de soja, sintoma conhecido popularmente como “soja-louca”.

O percevejo-verde, *N. viridula* (L.), ocorre em diferentes regiões produtoras, embora seja mais adaptado às regiões mais frias do Brasil (Região Sul). Essa espécie também é encontrada em outras culturas como a do feijão (PANIZZI; CORRÊA-FERREIRA, 1977). O percevejo-marrom *E. heros* (E.) é hoje o mais abundante no Brasil. Esta espécie é nativa da Região Neotropical e está bem adaptada às regiões mais quentes, sendo predominante do Norte do Paraná ao centro do Brasil no período dos meses de novembro a abril (PANIZZI; SLANSKY, 1985; FERNANDES; CORREIA; BORTOLI, 1990; CIVIDANES; PARRA,

1994). A cada ciclo da soja, o percevejo-marrom é capaz de produzir até três gerações. Durante o desenvolvimento biológico desse percevejo nos meses de maio até novembro ele passa por um estágio de quiescência, no qual não se alimenta. O percevejo permanece, assim, sob as folhas mortas e os restos das culturas, possibilitando a sua não predação pelos seus de inimigos naturais. Resultando numa alta taxa de percevejos sobreviventes até a safra seguinte (PANIZZI; NIVA, 1994; PANIZZI; VIVAN, 1997).

Todos esses percevejos são da ordem *Hemiptera* e pertencem à Família *Pentatomidae* e, durante seu desenvolvimento passam pelas fases de ovo, ninfa e adulto, completando o ciclo num período médio de 30 a 40 dias. A oviposição inicia aproximadamente no 15º dia da fase adulta, colocando, em média, 120 a 310 ovos por fêmea (CIVIDANES; PARRA, 1994a; CORRÊA-FERREIRA; PANIZZI, 1999).

Os percevejos se alimentam dos grãos da soja e, por isso, sua presença está relacionada diretamente à fase de desenvolvimento das vagens nas plantas. Dessa maneira, o período entre o começo da frutificação e o acúmulo máximo de matéria seca na semente é o de maior sensibilidade da soja ao ataque desses insetos sugadores (VICENTINI; JIMENEZ, 1977; PANIZZI; SMITH; PEREIRA, 1979; SCHUMANN; TODD, 1982; BRIER; ROGERS, 1991; GAMUNDI; ANDRIAN; BACIGALUPPO, 2003).

A partir do início do desenvolvimento das vagens, no estágio R3, os diferentes níveis de infestação de percevejos podem afetar negativamente o rendimento e a qualidade das sementes de soja (Figura 1). A magnitude dos danos é variável com a espécie de percevejo e também com a fase de desenvolvimento reprodutivo da soja, sendo, o período mais crítico o delimitado pelo final do desenvolvimento das vagens (R4) e o início do enchimento dos grãos (R6) (CORRÊA-FERREIRA; PANIZZI, 1999).

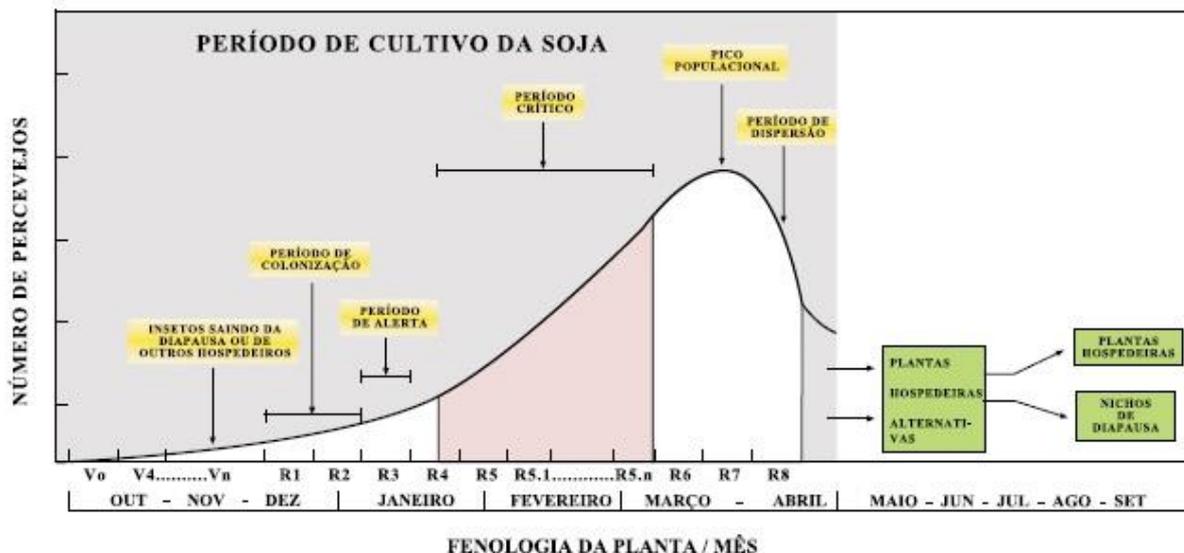


Figura 1 - Diferentes níveis de infestação de percevejos durante o período de cultivo da soja (CORRÊA-FERREIRA; PANIZZI, 1999)

Os percevejos sugadores de sementes alimentam-se inserindo os estiletes nas vagens e sugando os nutrientes dos grãos. Tanto as ninfas como os adultos obtêm seu alimento mediante a picada do tecido vegetal e posterior sucção dos conteúdos celulares. As puncturas nas sementes de soja, realizadas pelos percevejos, causam o aparecimento de manchas escuras e áreas esbranquiçadas (MINER, 1966). Os grãos quando atacados ficam menores, enrugados, murchos e com cor mais escura, podendo apresentar doenças como a mancha-fermento, causada pela levedura *Eremothecium corily Peglion*, que é transmitida durante a picada de prova e alimentação do percevejo. Além de reduções significativas na qualidade fisiológica (viabilidade e vigor), as sementes danificadas pelos percevejos sofrem alterações como redução de seus teores de óleo e aumento dos níveis de proteína (PANIZZI; SMITH; PEREIRA, 1979; VILLAS-BÔAS; GAZZONI; OLIVEIRA, 1990). Pode haver também retardamento da maturação da soja (retenção foliar/haste verde), dificultando a colheita (SOSA-GOMEZ e MOSCARDI, 1995).

As perdas no rendimento em função da presença destes insetos-praga nas lavouras podem variar de 80% a 100% (DEGRANDE; VIVAN, 2007), dependendo do nível de infestação. Além disso, há redução no poder germinativo e vigor das sementes (OLIVEIRA, 2010).

Na busca de amenizar os efeitos destes insetos-praga nas lavouras, extensivas aplicações de inseticidas têm sido realizadas. Estima-se que no Brasil sejam utilizados cerca de seis milhões de litros de inseticidas a cada safra, considerando apenas o controle de percevejos (ERENO, 2011). Todavia, a aplicação de inseticidas apresenta custo elevado e

nem sempre se mostra eficaz. Desde 1990 produtores de soja dos estados do Paraná e São Paulo relatam a não eficiência dos inseticidas do grupo dos organofosforados no controle de *E. heros* (SOSA-GÓMEZ; CORSO; MORALES, 2001). Um agravante é que as moléculas químicas provenientes deste grupo são as usualmente utilizadas no controle destes insetos. Além disso, o cenário brasileiro atual é de redução do número de ingredientes ativos disponíveis no mercado, proibição de alguns inseticidas organofosforados, falta de inovação e ausência de lançamento de novos inseticidas (GUEDES; ARNEMANN; STÜRMER, 2012).

Nos cultivos de soja que se destinam à produção de sementes, os prejuízos causados pelos percevejos devem ser mínimos. Dessa maneira, além do controle biológico e manejo da cultura, devem-se buscar alternativas para o controle dos percevejos (CORRÊA-FERREIRA; PANIZZI, 1999). Uma opção seria identificar plantas com resistência fisiológica ao complexo de percevejos e, conseguinte, obter variedades com características de interesse agrônomo associadas às de resistência.

2.3 Variedades de soja

As variedades de soja resistentes ao complexo de percevejos IAC-100 e IAC-17 (ROSSETTO, 1989; PINHEIRO, 1993; VEIGA; ROSSETTO; RAZERA, 1999) além das IAC-23 e IAC-24 (MIRANDA; BRAGA; MIRANDA, 2001; MIRANDA; LOURENÇÃO, 2002), são uma das poucas recomendadas para cultivo na região sudeste do Brasil, e atualmente pouco recomendadas.

A cultivar resistente IAC-100 é classificada dentro do grupo de maturidade (MG) 8 e foi desenvolvida pelo Instituto Agrônomo de Campinas em São Paulo, a partir de um cruzamento entre linhagens melhoradas de IAC-78-2318 e IAC-12 (ROSSETTO; GALLO; RAZERA, 1995). A linhagem IAC-78-2318 é uma fonte de genes de resistência múltipla contra-ataque de insetos de soja (LOURENÇÃO; MIRANDA, 1987).

A CD-215 é uma cultivar protegida desenvolvida pela Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda (COODETEC, 2013) e é proveniente do cruzamento entre OC88-127 e OC90-234. Essa variedade foi utilizada neste trabalho como o genitor susceptível ao percevejo, uma vez que não é fisiologicamente resistente fisiologicamente ao complexo de percevejo.

A utilização dessas duas variedades contrastantes (IAC-100 e CD-215) quanto à característica resistência ao complexo de percevejos é fundamental para se visualizar as diferenças dos mecanismos de respostas das cultivares frente à infestação. E, assim, tornar-se possível a identificação dos transcritos diferencialmente expressos entre as variedades,

possibilitando a apontamento dos genes da variedade IAC-100 que são os responsáveis pela resistência fisiológica.

2.4 Estresses

O estresse é todo e qualquer tipo de perturbação ambiental que condiciona uma modificação temporal de resposta fisiológica de um organismo vivo. Dessa forma, o metabolismo e desenvolvimento das plantas são constantemente influenciados por inúmeros estresses bióticos e abióticos. Como um dos resultados desses estresses temos a perda substancial na produtividade das culturas. Para contraporem-se aos estresses ambientais as plantas utilizam diferentes mecanismos incluindo a ativação de vias metabólicas de fitormônios de sinalização como resposta ao estresse, ultimamente sendo o principal tipo de resistência (BOSTOCK, 2005; CHEHAB; PEREA; GOPALAN, 2007).

Um componente essencial de todas as respostas aos estresses é a capacidade de modificação do transcriptoma (WALLEY; DEHESH, 2010). Quando uma planta é submetida a um estresse, ocorre uma sinalização que leva a modulação transcricional de uma série de genes, que resulta em uma alteração no nível de diversos metabólitos e proteínas, algumas das quais podem ser responsáveis por conferir um certo grau de proteção e tolerância (BHATNAGARMATHUR; VADEZ; SHARMA, 2008). A rapidez na percepção do sinal e as consequentes alterações transcricionais que levam a uma adaptação e eventual resistência ao estresse variam entre genótipos de uma mesma espécie (SREENIVASULU; SOROPY; KISHOR, 2007).

Na regulação gênica em nível transcricional há a ação de três elementos principais: remodelamento e modificação da cromatina, elementos cis-regulatórios que frequentemente são sítios de ligação como ‘*enhancers*’ e promotores, localizados antes ou após a região codante; e elementos trans-regulatórios, como fatores de transcrição. Além da transcrição, outros dois níveis de regulação também podem ocorrer, pós-transcricional e pós traducional. A modulação pós-transcricional pode ocorrer das seguintes maneiras: processamento de pré RNAm (*splicing*, poliadenilação); tráfego núcleo-citoplasmático de RNAm; *turnover* e estabilidade de RNAm; e tradução do RNAm. Em nível pós-traducional a planta pode realizar processos de fosforilação, sumoilação e ubiquitinação de proteínas (DUQUE; ALMEIDA; SILVA, 2013).

Embora grande parte dos esforços de caracterização de genes envolvam análises transcriptômicas, é relevante colocar que estas apenas refletem diferenças no acúmulo de RNAm, o que não significa necessariamente que o gene será convertido em um produto. Por

isso, além do perfil transcricional, faz-se importante a avaliação de outros produtos que estejam relacionados efetivamente ao processo de tradução, armazenamento e/ou degradação destes RNAm durante o tratamento de estresse (URANO; KURIHARA; SEKI, 2010).

Uma das estratégias biotecnológicas para enfrentar essa situação é o desenvolvimento de plantas geneticamente modificadas. Contudo, os mecanismos de defesas têm de ser bem estudados e elucidados para que possam ser produzidas culturas transgênicas transformadas com um ou vários genes que confirmam tolerância ao estresse de interesse (GATEHOUSE, 2002; FERRY, 2004).

2.4.1 Mecanismos de defesas em plantas contra herbivoria de insetos

Plantas e insetos têm vivido juntos por mais de 350 milhões de anos. No processo de co-evolução, ambas desenvolveram estratégias para evitar sistemas de defesa uma da outra. Esta evolucionária corrida armamentista entre plantas e insetos tem resultado no desenvolvimento de um sistema de defesa em plantas que tem a habilidade de reconhecer moléculas estranhas ou sinais a partir de danos celulares, muito parecido com os animais, e ativa a resposta imune da planta contra os herbívoros (HOWE, 2008; VERHAGE, 2010; HARE, 2011). Para conter o processo de herbivoria, as plantas se especializaram em produzir estruturas morfológicas especializadas ou metabólitos secundários e proteínas que tenham alguma característica tóxica, repelente e/ou efeito não nutricional em herbívoros (USHA, 2010; WAR, 2011a, 2011b).

Defesas diretas contra esses herbívoros são mediadas por algumas características das plantas que afetam a biologia dos insetos herbívoros, assim como proteção mecânica na superfície de plantas (tricomos, espinhos e folhas mais grossas) ou produção de substâncias químicas tóxicas como terpenóides, alcalóides, fenóis e quinonas que podem retardar o desenvolvimento dos herbívoros ou mesmo matá-los (HANLEY, 2007). As defesas indiretas contra insetos são mediadas pela liberação de produtos voláteis que atraem especificamente inimigos naturais dos herbívoros e/ou através do fornecimento de alimentos (néctar floral) e habitação para melhorar a eficácia dos inimigos naturais (ARIMURA, 2009).

Entender a natureza da expressão gênica das características defensivas das plantas poderá ter uma enorme aplicação no desenvolvimento de plantas melhoradas com melhor proteção contra os herbívoros. Isto, por sua vez, reduzirá a necessidade do uso de prejudiciais pesticidas para o controle de insetos (VENTURA; PINHEIRO, 1999). O conhecimento da complexa interação química planta-herbívoro é necessária para otimizar a produção de novas cultivares (WAR, 2012).

Para se entender os mecanismos de defesa das plantas torna-se necessário também entender as características da relação planta-herbívoro em questão, bem como qual parte da planta é atacada, qual estado de fenológico se encontra a planta e a idade do herbívoro, intensidade e frequência do ataque (BEGON; MORTIMER; THOMPSON, 1996; CRAWLEY, 1997; STILING, 1999; BOEGE; MARQUIS, 2005).

2.4.1.1 Ordem Hemiptera

Os insetos da Ordem *Hemiptera* possuem as estruturas da boca altamente modificadas, como mandíbulas e maxilares dentro de um invólucro flexível e delgado conhecido como lábio. As maxilas formam dois canais de agulha (estiletos). O canal de alimentação é usado para ingerir fluidos vegetais, enquanto que o canal salivar é usado para entregar a saliva ao longo do percurso do estilete no local de alimentação (KALOSHIAN; WALLING, 2005). Apesar desta estrutura de alimentação única, existem diferenças substanciais entre os mecanismos de alimentação entre os hemípteros. Essas diferenças influenciam o contato com defesas pré-formadas e indução de vias de defesa que incluem: mecanismos de prova do estilete, a passagem do estilete entre ou dentro das células, a deposição ou ausência de saliva no estilete e composição química das salivas. Táticas de alimentação abrangem desde repetidas punções do estilete em alta frequência até uma simples punctura pontual e ejeção de enzimas celulares contendo solubilizantes. O conjunto de hemipteros, que inclui o fitófagos das ordens *Hemiptera* e *Thysanoptera* (tripes), provocam respostas contínuas das plantas hospedeiras que variam de feridas até indução de genes mais vulgarmente associados a infecções bacterianas e fúngicas (KALOSHIAN; WALLING, 2005).

Um dos hemípteros mais estudados atualmente é o aphídeo, *Aphis glycines* M., que causa danos à soja como nanismo, distorção foliar e reduz a formação de vagens (SUN; TIAN; WANG, 1990; HILL; LI; HARTMAN, 2004). Como um dos maiores grupos de insetos que se alimentam do floema, os aphídeos são um sério problema para muitas culturas. Estudos prévios sobre respostas das plantas aos aphídeos sugeriram que a regulação das vias de sinalização do ácido jasmônico, etileno, ácido salicílico foram pelo menos parcialmente ativadas pela alimentação do afídeo. A lipoxigenase, acredita-se que esteja envolvida na síntese de ácido jasmônico em plantas e que são induzidas por ferimentos (BELL; CREELMAN; MULLET, 1995; LEON; SANCHEZ-SERRANO, 1999). O gene codificador da proteína PR1 foi relatado como um bom marcador na via dependente de ácido salicílico (UKNES; MAUCH-MANI; MOYER, 1992; ROGERS; AUSUBEL, 1997) e foi induzida pela alimentação do aphídeo em tomate e *Arabidopsis* (MORAN; THOMPSON, 2001;

MARTINEZ DE ILARDUYA.; XIE; KALOSHIAN, 2003; De vos; Van OOSTEN; Van POECKE, 2005). Em trigo, a atividade das β -1,3-glucanase, peroxidase, e quitinase foram induzidas para altos níveis em cultivares ditas como resistentes do que nas susceptíveis (Van der WESTHUIZEN; QIAN; BOTHA, 1998a, 1998b).

Análises de microarranjos feitos por Li (2008) em soja identificaram 140 genes em resposta a 6 e 12 horas após infestação por afídeos na cultivar tida como resistente (Anexo A). Esses 140 genes identificados foram classificados dentro de 16 grupos funcionais baseados em anotação de diferentes bases de dados. Esses grupos incluem genes de desenvolvimento celular, parede celular, citoesqueleto, defesa, DNA/RNA, membrana, oxidação, metabolismo primário, proteína, metabolismo secundário, senescência, sinalização, estresse, heterogêneo (de várias classes protéicas), não conferem (anotações diferentes em diferentes bases de dados) e desconhecidos (LI, 2008).

2.5 Novas tecnologias de sequenciamento

As tecnologias de sequenciamento incluem um número de métodos que são agrupados amplamente como preparação da molécula recombinante de DNA (*template*), sequenciamento e imagem (bioluminescência), e análise de dados. Uma única combinação específica de protocolos distingue uma tecnologia das outras e determina o tipo de dados produzidos por cada plataforma (METZKER, 2010). A produção de um grande número de *reads* a baixo custo faz das plataformas de Sequenciamento de Nova Geração úteis para muitas aplicações como possibilitar a descoberta de regiões de interesse marcadas ou de genomas inteiros, montagens do tipo ‘*de novo*’ de genomas pequenos de eucariotos e bacterianos, catalogar o transcriptoma de células, tecidos e organismos (RNA-Seq) (WANG, 2009), definir o perfil de genomas completos de marcas epigenéticas e estrutura de cromatina utilizando outros métodos baseados em sequenciamento (*ChIP-seq*, *methyl-seq* e *DNase-seq*) (WOLD; MYERS, 2008), e classificar e/ou descobrir genes através de estudos metagenômicos (PETROSINO, 2009).

A utilização de técnicas de biologia molecular aliadas aos programas de melhoramento proporciona maior eficiência na identificação/seleção de genes de interesse. Avanços significativos têm sido obtidos através do desenvolvimento de técnicas moleculares capazes de isolar, analisar, sintetizar e sequenciar nucleotídeos e, a partir de estudos *in vitro* e *in vivo*, verificar a função dessas moléculas e quais os mecanismos que controlam sua expressão. O desenvolvimento destas técnicas está diretamente vinculado à criação de novas cultivares com características de interesse (VARSNHEY; HOISINGTON; TYAGI, 2006).

Algumas das técnicas utilizadas para análises da expressão gênica são a 454 Pirosequenciamento, Illumina Genome Analyser, Microarranjos e RT-qPCR (SCHUSTER, 2008; WANG; SUN; MACGRATH, 2008; MESTDAGH, 2009; METZKER, 2010).

A técnica conhecida como RNA-Seq sequencia amostras de RNA em um estado fisiológico específico, e é livre de muitas limitações das outras tecnologias anteriormente utilizadas, como a dependência de um conhecimento prévio do organismo, requerido por técnicas como microarranjos e PCR. Além disso, RNA-Seq promete resolver complexidades do transcriptoma inacessíveis anteriormente, como expressão alelo-específica e novos promotores e isoformas (WANG; SUN; MACGRATH, 2008).

2.5.1 RNA-Seq

A resposta fisiológica das plantas a estresses bióticos e abióticos envolve muitos genes e mecanismos bioquímicos e moleculares. A adaptação para esses estresses é alcançada através da regulação da expressão gênica nos níveis transcricional e pós-transcricional (BHATNA-GARMATHUR; VADEZ; SHARMA, 2008).

Para entendermos melhor as respostas da planta ao estresse, sucedidos experimentos de perfis de transcritos têm sido empregados para diferentes tipos de estresses bióticos e abióticos (De VOS; Van OOSTEN; Van POECKE, 2005; EULGEM, 2005; BOHNERT; QINGQIU; PINGHUA, 2006). Um tema comum que emergiu a partir desses experimentos foi que inicialmente as respostas transcricionais eram compostas de um conjunto básico de genes sensíveis a multi-estresses e com o progresso do tempo na indução dos estresses os genes, para cada caso, tornaram-se incrivelmente específicos (LOPEZ-MAURY; MARGUERAT; BAHLER, 2008).

A técnica do RNA-Seq é a mais recente e eficiente na identificação de transcritos em relação às outras técnicas; o número de publicações é crescente, e, de acordo com a base de dados *Web of Knowledge* (<http://wokinfo.com/>), até início de novembro de 2013 foram publicados 2212 artigos e 204 revisões utilizando RNA-Seq, sendo que 247 artigos e 24 revisões são com plantas.

O primeiro trabalho utilizando RNA-Seq foi desenvolvido por Nagalakshmi e colaboradores (2008) e investigou o perfil transcricional do genoma de levedura. Foi gerado um mapa de alta resolução do transcriptoma do genoma da levedura que demonstrou que 74,5% das sequências não repetitivas do genoma eram transcritas. Os resultados evidenciaram assim que o transcriptoma de levedura era mais complexo que o previamente conhecido (NAGALAKSHMI, 2008).

A técnica de RNASeq baseia-se no sequenciamento do RNA total de um organismo de interesse, ou seja, do seu transcriptoma. O transcriptoma é definido como sendo o conjunto completo de transcritos em uma célula, e suas quantidades, em um estágio específico do desenvolvimento ou condição fisiológica (WANG; GERSTEIN; SNYDER, 2009), incluindo, portanto RNA codificante (mRNA) e não codificante (rRNA, tRNA, RNA estrutural, RNA regulatório, e outros tipos de RNAs). Entender o transcriptoma é essencial para interpretar os elementos funcionais do genoma e revelar os constituintes moleculares das células e tecidos, e também para entender, por exemplo, o mecanismo de uma doença e seu desenvolvimento. Desde o início da biologia molecular, diversos estudos vêm sendo realizados com o objetivo de entender os mecanismos envolvidos no processo de transcrição nas células (SESHASAYEE; BERTONE; FRASER, 2006), uma vez que a alteração nos níveis de expressão está diretamente relacionada às modificações na fisiologia, metabolismo e consequentemente ao processo de adaptação celular (Van VLIET, 2010).

A aplicação de técnicas moleculares baseadas em microarranjos de DNA eram as mais eficientes ferramentas para a determinação de um amplo padrão de expressão gênica a nível celular (HINTON; HAUTEFORT; ERIKSSON, 2004). Entretanto, algumas limitações no uso desta tecnologia têm surgido (BLOOM; KHAN; KRUGLYAK, 2009), como por exemplo, a especificidade do arranjo para cada isolado, o *background* e a qualidade e densidade variáveis dos *spots*; fatores que têm dificultado a análise comparativa entre experimentos e, geralmente, levados a necessidade de métodos normalizadores complexos (HINTON; HAUTEFORT; ERIKSSON, 2004). Devido a isto, a aplicação de sequenciamento de cDNA surgiu como uma alternativa eficiente para solucionar estes problemas, bem como para suprir dados transcriptômicos independentemente da necessidade de uma sequência genômica de referência previamente descrita.

O surgimento do sequenciamento de nova geração e o desenvolvimento de protocolos específicos aplicados para a síntese e sequenciamento de cDNA em larga escala, como a tecnologia RNA-Seq, tornou possível a obtenção de uma alta resolução em perfis transcriptômicos (METZKER, 2010).

O sequenciamento na plataforma Solexa da Illumina, assim como o sequenciamento de Sanger, é realizado por síntese usando DNA polimerase e nucleotídeos terminadores marcados com diferentes fluoróforos (MARDIS, 2008; MOROZOVA; HIRST; MARRA, 2008). A inovação dessa plataforma consiste na clonagem *in vitro* dos fragmentos em uma plataforma sólida de vidro, processo também conhecido como PCR de fase sólida (FEDURCO; ROMIEU; WILLIAMS, 2006; TURCATTI; ROMIEU; FEDURCO, 2008).

A superfície de clonagem (*flow cells*) é dividida em oito *lanes* que podem ser utilizadas para o sequenciamento de até doze bibliotecas em cada *lane*. Em cada *lane*, adaptadores são fixados à superfície pela extremidade 5', deixando a extremidade 3' livre para servir na iniciação da reação de sequenciamento dos fragmentos imobilizados no suporte por hibridização (CARVALHO; DA SILVA, 2010).

Segundo o Protocolo de RNA-Seq (NAGALAKSHMI, 2010), os adaptadores são ligados em ambas as extremidades dos fragmentos de DNA da amostra, o que permite a fixação dos fragmentos ao suporte de sequenciamento por hibridização a um dos seus adaptadores fixados. Além disso *barcodes* são adicionados às sequências de cada biblioteca para a posterior distinção entre elas. No primeiro ciclo de amplificação, nucleotídeos não marcados são fornecidos para que haja a síntese da segunda fita do fragmento imobilizado no suporte. A alta densidade de adaptadores no suporte facilita a hibridização do adaptador livre dos fragmentos imobilizados à sua sequência complementar fixa perto do clone inicial durante o ciclo de anelamento. Após este ciclo, o fragmento forma uma estrutura em “ponte” na superfície de sequenciamento e a extensão ocorre, formando a fita complementar também em “ponte”. No ciclo de desnaturação, as fitas são separadas e linearizadas. Esses ciclos são repetidos até 35 vezes e assim as cerca de mil cópias geradas de cada fragmento nessa PCR de fase sólida permanecem próximas umas das outras, formando um *cluster* de sequenciamento. Etapas de desnaturação são necessárias para a separação dos *duplex* formados e, nos próximos ciclos de amplificação, nucleotídeos terminadores marcados são fornecidos para as reações de sequenciamento que ocorrem dentro de cada *cluster*. A alta densidade dos *clusters* de sequenciamento possibilita que o sinal de fluorescência gerado com a incorporação de cada um dos nucleotídeos terminadores tenha uma intensidade suficiente para garantir sua detecção exata. Até 50 milhões de *reads* podem ser produzidos por *lane*, correspondendo a uma representação satisfatória da biblioteca. Após a incorporação de cada nucleotídeo no fragmento em síntese, a leitura do sinal de fluorescência é realizada. Em seguida, ocorre uma etapa de lavagem para remoção dos reagentes excedentes e remoção do terminal 3' bloqueado e do fluoróforo do nucleotídeo incorporado no ciclo anterior para que a reação de sequenciamento prossiga. A leitura das bases é feita pela análise sequencial das imagens capturadas em cada ciclo de sequenciamento (SHENDURE; JI, 2008). Ao final do sequenciamento de alto rendimento (RNA-Seq) podem-se obter pequenas sequências a partir de uma extremidade (*single-end*) ou de ambas as extremidades (*pair-end*). Esses *reads* possuem tipicamente de 30 a 400bp, dependendo da tecnologia de sequenciamento utilizada.

Este novo enfoque no estudo de sequências expressas tem possibilitado a interpretação não somente a nível celular, mas também com uma visão global de transcritos em amostras de diferentes ambientes. A plataforma de sequenciamento de RNA (RNA-Seq) possibilita uma variedade de aplicações, incluindo análises da expressão relativa, de *splicing* alternativo, descoberta de novos transcritos e isoformas, edição do RNA, expressão alelo-específica e a exploração de transcriptomas de organismos não-modelos. A técnica permite analisar transcriptomas complexos, independente de conhecimento prévio do genoma, além de possibilitar o mapeamento e quantificação de transcriptomas, principalmente quando há um genoma de referência para proceder às comparações (TANG; BARBACIORU; WANG, 2009; FILICHKIN; PRIEST; GIVAN, 2010). Portanto, essa técnica permite a quantificação dos níveis de RNA, com o benefício adicional de que todo o transcriptoma é analisado sem qualquer conhecimento a priori das regiões transcritas (WILHELM; LANDRY, 2009).

Após o sequenciamento, as leituras resultantes são alinhadas com um genoma ou transcriptoma de referência, ou caso não exista um genoma de referência pode ser realizada uma montagem “*de novo*” (WANG; GERSTEIN; SNYDER, 2009). Por fim, é realizada a comparação do nível de expressão entre os diferentes genes, bem como em diferentes experimentos.

A técnica RNA-Seq deve ser complementada com normalização dos perfis de expressão por meio do PCR quantitativo em tempo real (q-PCR) a partir da utilização de picos de RNAs controles que possuem suas concentrações já conhecidas em comparação com as concentrações das amostras obtidas pelo sequenciamento (WORKMAN, 2002).

2.6 Bioinformática

A tradicional pesquisa biológica aborda tipicamente estudos de um gene ou alguns genes de uma só vez. Em contraste, a genômica de alto rendimento, proteômica e abordagens de bioinformática (como os dados gerados em microarranjos, proteômica, chip-on-chips, etc) estão emergindo como tecnologias alternativas que permitem aos pesquisadores medir simultaneamente as mudanças e regulação dos genes de genomas completos em determinadas condições biológicas (HUANG; LIU; ZHOU, 2010).

Atualmente, existem diversos *softwares* com o objetivo de alinhar leituras curtas oriundas de sequenciadores de segunda geração em sequências genômicas de interesse. Entre os mais utilizados estão o *BWA* e *Bowtie*. Leituras de experimentos de expressão gênica trazem uma característica que deve ser considerada: em uma leitura pode existir mais de um éxon, que no genoma encontram-se separados por regiões intergênicas. Com o objetivo de

contornar este problema surgiram os alinhadores *TopHat* (TRAPNELL, 2012) e o *STAR* (DOBIN, 2013). Ambos utilizam informações do arquivo de anotações e de sítios que são conservados em junções éxon/íntron para alinhar de forma mais adequada as leituras curtas aos genomas de interesse.

2.6.1 TopHat e Cufflinks

TopHat e *Cufflinks* são softwares livres (distribuídos com a licença GPL), que possibilitam alinhar, quantificar e inferir diferenças estatísticas em dados de RNA-Seq, e que consideram em sua normalização o tamanho de éxons (Fragmentos por *kilobase* de éxon por milhões de *reads* mapeados - FPKM). Juntos eles permitem aos biólogos identificarem novos genes e novas *splice* variantes de genes já conhecidos, assim como comparação entre a expressão do gene e do transcrito sob duas ou mais condições (TRAPNELL, 2012).

TopHat é uma programa que alinha sequências de leituras provenientes do sequenciamento do RNA (RNA-Seq) com um genoma a fim de identificar as junções éxon-éxon nas quais ocorre *splice*. O alinhamento é construído a partir do mapeamento ultra rápido das pequenas leituras com o genoma a partir do programa *Bowtie*. O *TopHat* funciona no **Linux** e no **OS X** (<http://tophat.cbcb.umd.edu/manual.shtml>). Esses alinhamentos são usados durante as análises decorrentes de várias maneiras. *Cufflinks* (TRAPNELL, 2012) (<http://cufflinks.cbcb.umd.edu/>) utiliza esse mapa contra o genoma para alinhar os *reads* dentro dos transcritos. *Cuffdiff*, uma parte do pacote do *Cufflinks*, analisa os *reads* alinhados aos respectivos genes e a partir de duas ou mais condições informa os genes e os transcritos que são diferencialmente expressos, utilizando rigorosas análises estatísticas. Essas ferramentas vêm conquistando grande aceitação e têm sido usadas em estudos de alta-resolução de transcriptomas (MIZUNO, 2010; GRAVELEY, 2011; LISTER, 2011; TWINE, 2011). A ferramenta *CummeRbund* oferece figuras e gráficos para auxiliar a interpretação dos dados de saída *Cuffdiff* (TRAPNELL, 2012). Na Figura 2 podemos observar todos os passos da análise pelo *TopHat* e *Cufflinks*.

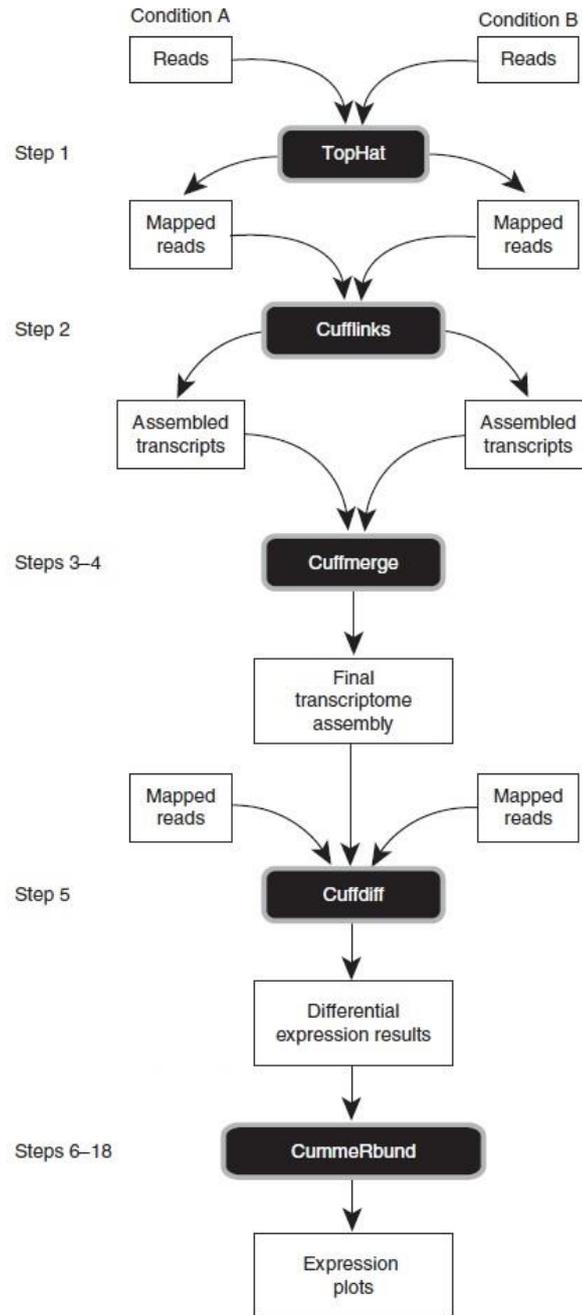


Figura 2 - Sequência de análises pelo *TopHat* e *Cufflinks* (TRAPNELL, 2012)

O *TopHat* e o *Cufflinks* foram projetado para trabalhar com leituras (*reads*) produzidas pelo *Illumina Genome Analyzer*, embora usuários tenham tido sucesso em utilizar o *TopHat* com *reads* provenientes de outras tecnologias (GOECKS; NEKRUTENKO; TAYLOR, 2010). Foi com o *TopHat* 1.1.0, que iniciou o apoio ao formato *Colorspace* da *Applied Biosystems*, *output* padrão das plataformas *Solid*. O software está otimizado para leituras de 75 *pb* ou mais. Uma das limitações do *TopHat* é que ele não é capaz de fazer uma leitura conjunta de *pair-end* e *single-end*. O protocolo do *TopHat* é apropriado para experimentos de

RNA-Seq em organismos com genomas já sequenciados. Para trabalhar com organismos que ainda não possuem o seus genomas sequenciados deve-se considerar a performance ‘*de novo*’ da montagem de transcriptomas usando uma das várias ferramentas assim como o *Trinity* (GRABHERR; HASS; YASSOUR, 2011), *Trans-Abyss* (ROBERTSON; SCHEIN; CHIU, 2010) ou *Oases* (<http://www.ebi.ac.uk/~zerbino/oases/>).

As junções *splice* são encontradas pelo *TopHat* sem a necessidade de uma anotação de referência. Primeiramente os *reads* do RNA-Seq são mapeados com o genoma, o *TopHat* identifica os potenciais éxons, uma vez que muitos *reads* do RNA-Seq vão se alinhar continuamente ao genoma. Usando essa informação inicial de mapeamento o *TopHat* constrói um conjunto de dados de possíveis junções *splice* e, em seguida, mapeia os *reads* contra essas junções para confirmá-las (TRAPNELL, 2012).

Plataformas de sequenciamento de leituras curtas têm produzido atualmente *reads* com 100 *pb* ou mais, mas muitos éxons são menores que isso então eles podem ser perdidos em um mapeamento inicial. O programa *TopHat* pode resolver esse problema principalmente através da divisão da entrada dos *reads* dentro de segmentos menores que são mapeados independentemente. Os segmentos alinhados são colocados juntos novamente em um passo final do programa que produz alinhamentos de uma extremidade à outra dos *reads*.

O *TopHat* gera a sua base de dados de possíveis junções *splice* a partir de duas fontes de evidências. A primeira e mais importante fonte de evidência de uma junção *splice* é quando dois segmentos de um mesmo *read* (para *reads* de pelo menos 45 *pb*) são mapeados até uma certa distância em uma mesma sequência genômica ou quando um segmento interno deixa de ser mapeado – novamente sugerindo que tais leituras estão abrangendo múltiplos éxons. Com essa abordagem, os íntrons "GT-AG", "GC-AG" e "AT-AC" serão encontrados desde o início. A segunda fonte são os pareamentos de “ilhas de cobertura”, que são regiões distintas de *reads* empilhados no início do mapeamento. Ilhas vizinhas frequentemente sofrem *splicing* juntas dentro do transcriptoma, assim o *TopHat* procura maneiras para juntar essas ilhas com um íntron. O uso dessa segunda opção (procura por cobertura) é sugerida apenas para *reads* curtos (menor que 45 *pb*). Esta última opção só irá relatar alinhamentos através dos íntrons "GT-AG" (TRAPNELL, 2012).

Recentemente uma nova versão do *TopHat* foi disponibilizada (v2.0.9) que trouxe sensíveis melhorias em relação a versão anterior, como a capacidade de mapear *reads* de regiões translocadas no genoma referência ou com *indels* (KIM, 2013).

Os programas *bowtie2* (<http://bowtie-bio.sourceforge.net/index.shtml>) e *samtools* (<http://samtools.sourceforge.net/>) são pacotes que devem ser instalados anteriormente ao

TopHat, eles são responsáveis pelo alinhamento inicial que será processado pelo *TopHat* e por realizar as manipulações nos arquivos de mapeamento de sequências (*SAM*), respectivamente. Algumas etapas do código do *TopHat* são escritas em *Python*, exigindo assim que exista também tal pacote instalado. Todos estes programas podem ser encontrados nos repositórios oficiais do *Ubuntu* (distribuição do *linux* mais utilizada atualmente), porém é recomendado verificar a versão dos programas instalados pois muitas vezes não são os mais recente.

As informações produzidas pelo processamento do *TopHat* são armazenadas em diferentes arquivos de saída, numa pasta especificada no momento de execução do programa ou no diretório padrão, caso não tenha sido especificado nenhum (diretório padrão: *tophat_out*) (TRAPNELL, 2012). Os arquivos output são descritos a seguir:

1. *accepted_hits.bam*. Dados do mapeamento dos *reads* contra o genoma referência. O formato *.bam* é o arquivo compactado do *SAM*. Entende-se *SAM* por formato de Sequência Alinhada/Mapeada (*Sequence Alignment/Map format*). É um texto de formato TAB-delimitado que consiste uma seção principal (*header*), que é opcional, e uma seção de alinhamento. Se presente a principal deve ser anterior aos alinhamentos. Linhas *header* começam com '@' enquanto que as linhas do alinhamento não. Cada linha de alinhamento tem 11 campos obrigatórios de informação essencial ao alinhamento, como posição de mapeamento e número variável de campos opcionais para flexibilidade ou informações específicas do alinhador (<http://samtools.sourceforge.net/>).

2. *junctions.bed*. Um UCSC BED relata o controle das junções pelo *TopHat*. Cada junção consiste em dois blocos BED conectados, onde cada bloco é tão longo quanto a projeção máxima de qualquer *read* que abrange a junção. A pontuação (*score*) é o número de alinhamentos que abrangem a junção.

3. *insertions.bed* e *deletions.bed*. UCSC BED relata o controle das inserções e deleções pelo *TopHat*.

2.6.2 STAR

O *Spliced Transcripts Alignment to a Reference* (STAR) é um alinhador que utiliza um algoritmo projetado especificamente para resolver muitos desafios do mapeamento dos dados do RNA-Seq, e utiliza uma nova estratégia para o alinhamento de transcritos *spliced*. A precisão do alinhador está na detecção do conjunto de junções *splice*. A alta velocidade e acurácia do mapeamento do STAR foram cruciais para analisar o grande conjunto de dados de transcriptoma da Enciclopédia de Elementos de DNA (Encyclopedia Of DNA Elements – ENCODE) com mais de 80 bilhões de *Illumina reads* (DJEBALI, 2012). O STAR tem uma

potencial acurácia para alinhar longos *reads* (várias kilobases) que está a emergir a terceira-geração de tecnologias de sequenciamento.

O STAR foi projetado para alinhar sequências não contíguas diretamente ao genoma referência. O seu algoritmo consiste de duas principais etapas: etapa de início da pesquisa e etapa de agrupamento/anotação. A principal idéia da fase de pesquisa do STAR é a busca sequencial pelo Prefixo Máximo Mapeável (MMP) (DOBIN, 2013). As junções *splice* são detectadas numa única passagem de alinhamento sem nenhum conhecimento *a priori* dos locais das junções *splice* ou suas propriedades, e sem uma passagem preliminar de alinhamento contíguo necessário para abordar os conjuntos de dados das junções. Na segunda fase do algoritmo, STAR constrói alinhamentos de toda a sequência do *read* agrupando todos os que foram alinhados ao genoma na primeira fase.

Primeiramente, os *reads* são agrupados pela proximidade de um conjunto selecionado como ‘âncora’. O procedimento ideal para a seleção âncora é através da limitação do número de locos genômicos para o alinhamento das âncoras. Embora a pesquisa MMP sequencial só encontre os *reads* que correspondem exatamente ao genoma, o procedimento de junção posterior é capaz de alinhar as leituras com um grande número de *mismatches*, *indels* e junções *splice*, condizentes com o comprimento de leitura estabelecido. Esta característica tem se tornado cada vez mais importante com a surgimento das tecnologias de sequenciamento de terceira geração (como Pacific Biosciences ou Ion Torrent) que produzem longos *reads* com elevadas taxas de erro (DOBIN, 2013).

2.6.3 iPATH

O *iPath* é uma ferramenta *online* de acesso aberto (<http://pathways.embl.de>) para visualizar e analisar as vias metabólicas. Um visualizador interativo que fornece simples navegação através de vários caminhos e permite fácil acesso aos produtos químicos e enzimas subjacentes. Mapas de vias metabólicas personalizados podem ser gerados e anotados utilizando diversas bases de dados externas. Por exemplo, por meio da fusão dos dados do genoma humano, com os de dois organismos comensais intestinais importantes, o *iPath* pode identificar a complementaridade das capacidades metabólicas hospedeiro-simbióticas (LETUNIC, 2008).

A publicação do Kyoto Enciclopédia de Genes e Genomas (*KEGG*) (<http://www.genome.jp/keeg/>) fornece uma visão global de mapa das vias metabólicas no formato de Gráficos e Vetores (KANEHISA; ARAKI; GOTO, 2007) o que representa um importante passo para a visualização em grande escala e interpretação de vários dados sobre

as atividades metabólicas. O mapa global foi construído manualmente usando 123 mapas clássicos do *KEGG* com uma média de 17 reações cada. O resultado é a visão geral de uma grande proporção de todas reações metabólicas conhecidas até à data, coletada a partir de várias sistemas biológicos (LETUNIC, 2008).

As linhas mostradas nos mapas disponibiliza detalhes como se relacionam as enzimas e os compostos envolvidos nas vias metabólicas de interesse. Outras informações podem ser acessadas através de *hiperlinks* a outros recursos *on-line*, como por exemplo, *KEGG*, *PubChem* (<http://pubchem.ncbi.nlm.nih.gov/>), *3DMET* (<http://www.3dmet.dna.affrc.go.jp/>) e *CHEBI* (DEGTYARENKO; MATOS; ENNIS, 2008).

2.6.4 Blast2GO

Blast2GO (CONESA; GÖTZ; GARCÍA-GÓMEZ, 2005) é uma ferramenta de bioinformática abrangente para a anotação funcional e análise de seqüências de genes ou proteínas. A ferramenta foi originalmente desenvolvido para fornecer uma interface entre a ontologia de um gene A do Gene Ontology e sua anotação (GÖTZ; GARCÍA-GÓMEZ; TEROL, 2008).

As recentes melhorias têm aumentado consideravelmente a funcionalidade da ferramenta de anotação como códigos da Enzima (CE), códigos de genes (*KEGG*), mapas e motivos InterPro (GÖTZ; GARCÍA-GÓMEZ; TEROL, 2008). Além disso, o aplicativo reúne uma grande variedade de ferramentas gráficas e analíticas para a manipulação de anotação e análise de dados.

2.6.5 Phytozome

O portal *Phytozome* (<http://www.phytozome.net>), lançado em 2008, oferece um espaço de trabalho centralizado que permite aos usuários o acesso à famílias gênicas de plantas que já anotadas, navegar pelas suas histórias evolutivas, examinar os genes de plantas em seu contexto genômico, atribuir possíveis funções para seqüências não caracterizadas e fornecer acesso uniforme ao banco de dados genômico de Plantas, que consiste em genomas completos, gene e seqüências com seus alinhamentos relacionados (por exemplo gene homólogo), informações funcionais de genes e famílias de genes (GOODSTEIN; SHU; HOWSON, 2012).

A versão v7.0 de *Phytozome* contém dados de análises de 25 genomas de plantas, dos quais 18 já foram seqüenciados, montados e parcial ou completamente anotados no *Joint Genome Institute* - JGI (GOODSTEIN; SHU; HOWSON, 2012). Para genomas sem

anotações JGI, as seqüências do genoma já montadas e informações sobre determinado gene , transcrito e péptido são obtidas em formato *General Feature Format* - GFF ou em formato de seqüências nucleotídicas (FASTA) , e submetido a verificação de consistência .

3 OBJETIVOS

Investigar os transcritos diferencialmente expressos entre as cultivares IAC-100 e CD-215 após 24 horas de infestação com *Piezodorus guildinii* (West.), identificar os genes responsáveis pela resistência fisiológica à essa espécie e estabelecer as vias metabólicas ativadas no processo de infestação.

4 MATERIAL E MÉTODOS

4.1 Características do Experimento

Os dados analisados neste trabalho foram gerados na tese de Santos (2012). Nesta tese podem ser obtidos detalhes sobre a infestação, extração e sequenciamento de RNA. Os resultados do sequenciamento de RNA e as análises de bioinformática realizados correspondem ao tempo de 24 horas pós-infestação com *Piezodorus guildinii* nas cultivares IAC-100 e CD-215. Estes dados foram preliminarmente analisados por Santos (2012) e nesta dissertação buscou-se então a identificação de transcritos associados a reação de resistência a esta espécie de percevejo sugador das vagens.

4.2 Análises de Bioinformática

O Sequenciamento de RNA foi realizado apenas para as amostras provenientes das duas variedades de soja (IAC-100 e CD-215) submetidas ao tempo de 24 horas de infestação pelo percevejo *Piezodorus guildinii* (West.) e sem infestação. Foram feitas quatro réplicas biológicas para cada cultivar. O sequenciamento foi realizado na plataforma *Illumina HiSeq 2000* do tipo *pair-end*. Dessa forma, foram obtidos 32 arquivos *fastq* do sequenciamento de RNA proveniente das amostras de soja (Tabela 1).

Tabela 1 - Números de pares de reads para cada amostra

Amostra	Replicata Biológica	Pair-end	Sequência	Nº de reads	Variedade	Tratamento
1	1	Pair-end 1	1_R1.PF.fastq	16.199.010	IAC-100	24hrs, Não Infestado
		Pair-end 2	1_R2.PF.fastq	16.199.010		
2	2	Pair-end 1	2_R1.PF.fastq	12.516.152	IAC-100	24hrs, Não Infestado
		Pair-end 2	2_R2.PF.fastq	12.516.152		
3	3	Pair-end 1	3_R1.PF.fastq	15.819.440	IAC-100	24hrs, Não Infestado
		Pair-end 2	3_R2.PF.fastq	15.819.440		
4	4	Pair-end 1	4_R1.PF.fastq	23.305.130	IAC-100	24hrs, Não Infestado
		Pair-end 2	4_R2.PF.fastq	23.305.130		
5	1	Pair-end 1	5_R1.PF.fastq	14.193.004	IAC-100	24hrs, Infestado
		Pair-end 2	5_R2.PF.fastq	14.193.004		
6	2	Pair-end 1	6_R1.PF.fastq	14.929.043	IAC-100	24hrs, Infestado
		Pair-end 2	6_R2.PF.fastq	14.929.043		
7	3	Pair-end 1	7_R1.PF.fastq	16.638.714	IAC-100	24hrs, Infestado
		Pair-end 2	7_R2.PF.fastq	16.638.714		
8	4	Pair-end 1	8_R1.PF.fastq	15.068.147	IAC-100	24hrs, Infestado
		Pair-end 2	8_R2.PF.fastq	15.068.147		
9	1	Pair-end 1	9_R1.PF.fastq	14.895.220	CD-215	24hrs, Não Infestado
		Pair-end 2	9_R2.PF.fastq	14.895.220		
10	2	Pair-end 1	10_R1.PF.fastq	14.619.333	CD-215	24hrs, Não Infestado
		Pair-end 2	10_R2.PF.fastq	14.619.333		
11	3	Pair-end 1	11_R1.PF.fastq	12.899.899	CD-215	24hrs, Não Infestado
		Pair-end 2	11_R2.PF.fastq	12.899.899		
12	4	Pair-end 1	12_R1.PF.fastq	15.527.583	CD-215	24hrs, Não Infestado
		Pair-end 2	12_R2.PF.fastq	15.527.583		
13	1	Pair-end 1	13_R1.PF.fastq	15.903.430	CD-215	24hrs, Infestado
		Pair-end 2	13_R2.PF.fastq	15.903.430		
14	2	Pair-end 1	14_R1.PF.fastq	18.497.148	CD-215	24hrs, Infestado
		Pair-end 2	14_R2.PF.fastq	18.497.148		
15	3	Pair-end 1	15_R1.PF.fastq	13.035.656	CD-215	24hrs, Infestado
		Pair-end 2	15_R2.PF.fastq	13.035.656		
16	4	Pair-end 1	16_R1.PF.fastq	15.581.269	CD-215	24hrs, Infestado
		Pair-end 2	16_R2.PF.fastq	15.581.269		
Total de reads				499.256.356		

As bibliotecas do RNA-Seq foram preparadas com o *Kit* de preparação de amostras ‘*TruSeq RNaseq Sample Prep*’ da *Illumina*. Em seguida as bibliotecas foram reunidas em

concentração equimolar, quantificada por qPCR e sequenciadas em duas *lanes* por 100 ciclos cada extremidade dos fragmentos em um aparelho *HiSeq2000*.

O tamanho das *reads* sequenciadas foi de 100 nucleotídeos, sendo que o tamanho médio de fragmento do cDNA foi de 250 nucleotídeos.

A sequência dos adaptadores usados para fazer as bibliotecas *TruSeq* foram:

I) Sequência do adaptador no read_1 (pair-end 1):

AGATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGC
CGTCTTCTGCTTG

II) Sequência do adaptador no read_2 (pair-end 2):

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCC
GTATCAT

Os procedimentos descritos a seguir foram repetidos para cada uma das réplicas, de ambos os tratamentos.

As leituras *fastq* R1 e R2 obtidas a partir do sequenciamento foram verificadas quanto às suas qualidades, e para isso, utilizou-se o programa FASTQC para verificar visualmente a qualidade média das *reads* (qualidade do *Phred* de 20). Posteriormente as leituras foram alinhadas contra o genoma de *Glycine max* disponível no banco de dados *Phytozome*. Foi utilizado o *harmasked* que substitui sequências repetitivas identificadas pelo *RepeatMasker* por *Ns*. Os resultados deste relatório foram obtidos, separadamente, pelos alinhadores *TopHat* e *STAR*.

Os resultados obtidos juntamente com os arquivos de alinhamento (*.sam*) foram utilizados como *input* para o *Cufflinks* quantificar a expressão de cada gene, em FPKM. O programa *HTseq-count*, escrito em *Python*, foi utilizado para contar o número de *reads* que alinharam em cada gene presente na referência, para ser possível posterior análise com o pacote *DESeq*, paralelamente ao *Cufflinks*. O arquivo de anotações foi utilizado para auxiliar no processo de mapeamento pelo *pipeline Cuffdiff* (a partir de pacotes do *Cufflinks*) para obter os genes diferencialmente expressos (Figura 3). As anotações (Ids) dos transcritos diferencialmente expressos possuem os seguintes dados: quais são as proteínas produzidas por cada transcrito e arquivos de domínios *pfam*, termos do *Gene Ontology* (mostrando os produtos gênicos) e o Id de vias metabólicas do *Kyoto Encyclopedia of Genes and Genomes*

(KEGG). A base de dados UEKO foi utilizada juntamente com os programas *iPath* e *MapMan* para fazer a representação das vias metabólicas correspondentes aos transcritos diferencialmente expressos nas comparações entre as amostras IAC-100 Infestado x IAC-100 Não Infestado e CD-215 Infestado x CD-215 Não Infestado.

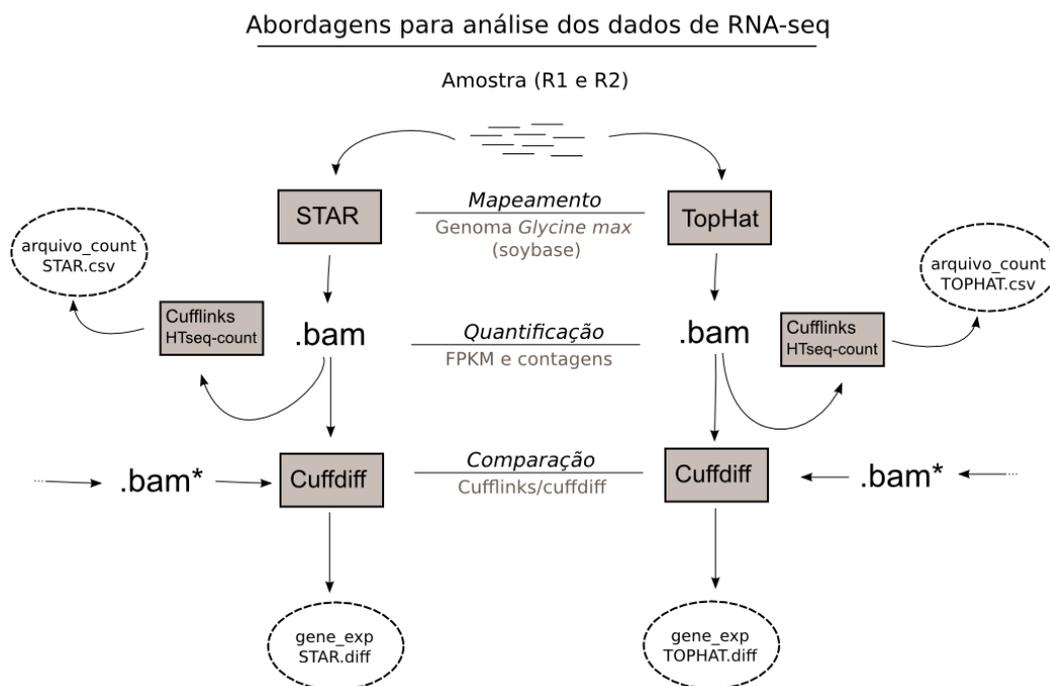


Figura 3 - Análises dos dados de RNA-Seq

Após o estudo inicial, foi feita uma filtragem dos transcritos afim de se identificar os expressos unicamente na situação IAC-100 Infestado. Para isso foi empregada a seguinte metodologia:

Após a obtenção dos resultados gerados pelo *Cuffdiff* foi realizado um refinamento da seguinte forma:

1º) A tabela gerada inicialmente dos transcritos em IAC-100 Infestado x CD-215 Infestado foi comparada com a tabela dos transcritos gerados pela comparação IAC-100 Não Infestado x CD-215 Não Infestado e os transcritos iguais em ambas as tabelas foram retirados. Sendo assim, ficaram apenas os transcritos expressos na situação de infestação.

2º) Os transcritos que ficaram da comparação anterior foram comparados com os presentes na tabela da comparação CD-215 Infestado x CD-215 Não Infestado. Sendo assim, os transcritos iguais em ambas as tabelas foram descartados, deixando apenas os transcritos que estariam presentes na variedade IAC-100 sob a condição Infestado. Uma vez que essa variedade é relatada como resistente ao complexo de percevejos.

Por fim, foi utilizado o *CummeRbund* para a criação de figuras e gráficos para publicação dos dados RNA-Seq.

5 RESULTADOS

5.1 Resultados obtidos com o alinhador Tophat

Primeiramente as amostras foram analisadas no software *TopHat2*, que utiliza o alinhador *Bowtie2* em seus processos iniciais. Depois, com o pacote *Cufflinks* foi realizada a quantificação dos transcritos por valores de *FPKM* e análise de expressão diferencial. Todo processo foi suprido pelo arquivo com anotações disponível no *Phytozome*.

As análises das comparações geraram os seguintes resultados:

1. IAC-100 Infestada x IAC-100 Não Infestada
2291 transcritos diferencialmente expressos
2. CD-215 Infestada x CD-215 Não Infestada
569 transcritos diferencialmente expressos
3. IAC-100 Infestada x CD-215 Infestada
613 transcritos diferencialmente expressos
4. IAC-100 Não Infestada x CD-215 Não Infestada
3858 transcritos diferencialmente expressos

Após esses resultados do *Cuffdiff* foi feito um refinamento da seguinte forma:

1º) A tabela dos transcritos em IAC-100 Infestada x CD-215 Infestada foi comparada com a dos em IAC-100 Não Infestada x CD-215 Não Infestada e os transcritos iguais foram retirados. Sendo assim, ficaram apenas os expressos na situação de infestação. Restando 174 transcritos de 613 iniciais.

2º) Os transcritos que restaram da comparação anterior foram confrontados com os presentes no contraste CD-215 Infestada x CD-215 Não Infestada. Sendo assim, foram descartados os transcritos que eram presentes tanto no resultado onde se obteve 174 transcritos quanto no contraste entre os dois tratamentos em CD (Inf x Não Inf). Restaram então apenas os transcritos que estariam presentes na variedade IAC-100 que é a de interesse, uma vez que esta é resistente ao complexo de percevejos. Logo, dos 613 transcritos iniciais restaram apenas 119 transcritos.

Na Tabela 2 encontram-se os transcritos resultantes da filtragem organizados de forma decrescente pelos seus valores de fold change que foram calculados a partir dos valores mostrados na tabela dos FPKM das amostras de IAC-100 Infestada e CD-215 Infestada pela eq.(1):

$$\log_2 \frac{FPKM_{IAC-100\ Infestada}}{FPKM_{CD-215\ Infestada}} \quad (1)$$

Tabela 2 - Transcritos diferencialmente expressos, em ordem decrescente, resultantes da filtragem

(continua)

Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma15g35221	0,1	1,9592	1,7977E+308	-
2	Glyma20g12720	0,1	0,89022	1,7977E+308	LRR and NB-ARC domains-containing disease resistance protein
3	Glyma20g12835	0,1	6,28945	1,7977E+308	-
4	Glyma15g39220	0,05564	1,43231	4,68603	Zinc finger C-x8-C-x5-C-x3-H type family protein
5	Glyma06g41380	0,04316	0,75093	4,12083	Disease resistance protein (TIR-NBS-LRR class)
6	Glyma14g11420	3,62952	36,0985	3,31409	17.6 kDa class II heat shock protein
7	Glyma08g11240	0,37478	3,61686	3,27061	FKBP-type peptidyl-prolyl cis-trans isomerase family protein
8	Glyma14g01700	1,75623	15,1081	3,10476	-
9	Glyma04g03470	0,40048	2,51164	2,64884	-
10	Glyma18g01030	0,78252	4,47987	2,51725	Translocase outer membrane 20-2
11	Glyma10g29300	0,37081	1,9081	2,36337	Actin-binding FH2 (formin homology 2) family protein
12	Glyma02g08500	0,63815	3,27817	2,36093	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
13	Glyma10g17561	0,25764	1,25428	2,28346	Calcium-dependent protein kinase 19
14	Glyma18g04490	1,5299	7,11367	2,21715	-
15	Glyma11g20860	0,49415	2,15679	2,12588	Cytokinin oxidase/dehydrogenase 6
16	Glyma13g33524	1,43526	5,8492	2,02693	-
17	Glyma19g44410	1,57871	4,5166	1,51649	Alpha/beta-Hydrolases superfamily protein
18	Glyma13g42061	2,58036	7,11037	1,46235	RNA binding
19	Glyma13g24270	1,30418	3,58853	1,46025	Cellulose synthase like G2
20	Glyma07g36530	19,3909	52,8726	1,44714	-
21	Glyma15g02840	0,89188	2,36921	1,40949	Indeterminate(ID)-domain 11
22	Glyma19g39480	5,9426	14,6679	1,30349	Glutaredoxin family protein
23	Glyma13g27510	7,06185	16,0678	1,18605	Protein of unknown function, DUF584
24	Glyma08g02261	3,9188	8,26395	1,07642	P-loop containing nucleoside triphosphate hydrolases superfamily protein
25	Glyma08g20330	3,9873	7,89509	0,985541	ATPase E1-E2 type family protein / haloacid dehalogenase-like hydrolase family protein
26	Glyma05g22470	1,68874	2,93403	0,79694	Arabinanase/levansucrase/invertase

Tabela 2 - Transcritos diferencialmente expressos, em ordem decrescente, resultantes da filtragem

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
27	Glyma08g09601	1,47941	2,56468	0,793756	Tetratricopeptide repeat (TPR)-like superfamily protein
28	Glyma06g46710	2,39705	4,1075	0,777	Zinc ion binding
29	Glyma10g38950	4,14724	5,30819	0,35607	-
30	Glyma01g24150	2,80397	1,80178	-0,638045	Protein kinase 1B
31	Glyma14g02680	4,23975	2,67232	-0,665886	Calmodulin-domain protein kinase 9
32	Glyma06g11140	2,27785	1,39675	-0,705597	Sulfate transporter 1
33	Glyma07g02400	15,7483	9,56018	-0,72009	Cyclin-dependent kinase B1
34	Glyma07g31910	6,01531	3,63276	-0,727574	Mitochondrial substrate carrier family protein
35	Glyma08g12630	4,33258	2,5919	-0,741213	Heat shock transcription factor A4A
36	Glyma10g18090	3,93675	2,31848	-0,763827	-
37	Glyma19g29031	4,24346	2,3598	-0,846578	-
38	Glyma08g26150	4,89812	2,57	-0,93046	Sec14p-like phosphatidylinositol transfer family protein
39	Glyma09g02900	15,1334	7,51321	-1,01023	6-phosphogluconate dehydrogenase family protein
40	Glyma07g31500	7,0061	3,34014	-1,0687	Aluminium induced protein with YGL and LRDR motifs
41	Glyma18g16130	3,53007	1,66169	-1,08704	IQ-domain 28
42	Glyma15g42140	20,8136	9,68114	-1,10428	ATP-citrate lyase A-1
43	Glyma07g06560	2,59909	1,16429	-1,15855	HhH-GPD base excision DNA repair family protein
44	Glyma03g27180	7,30868	2,89914	-1,33398	Squamosa promoter binding protein-like 8
45	Glyma10g33100	5,74615	2,08916	-1,45967	P-loop containing nucleoside triphosphate hydrolases superfamily protein
46	Glyma11g35240	4,81141	1,7378	-1,4692	SGNH hydrolase-type esterase superfamily protein
47	Glyma15g01960	3,74673	1,28607	-1,54267	HD-ZIP IV family of homeobox-leucine zipper protein with lipid-binding START domain
48	Glyma15g11910	25,2301	8,63571	-1,54676	Fibrillin
49	Glyma20g32870	1,66326	0,56555	-1,55629	Pleiotropic drug resistance 11
50	Glyma13g12041	60,9156	20,1851	-1,59352	-
51	Glyma13g12033	117,311	37,3012	-1,65304	-
52	Glyma13g11977	56,6388	17,9713	-1,6561	-

Tabela 2 - Transcritos diferencialmente expressos, em ordem decrescente, resultantes da filtragem

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
53	Glyma13g07530	1,88436	0,59418	-1,66511	Zinc-finger domain of monoamine-oxidase A repressor R1 protein
54	Glyma15g06770	65,4923	20,5938	-1,66911	Pathogenesis-related protein-1-like
55	Glyma13g11950	87,0692	27,1149	-1,68308	-
56	Glyma19g03730	11,0944	3,44154	-1,68871	HXXXXD-type acyl-transferase family protein
57	Glyma07g05740	21,0786	6,38524	-1,72297	Basic helix-loop-helix (bHLH) DNA-binding family protein
58	Glyma13g12070	55,5873	16,8273	-1,72395	-
59	Glyma13g21310	17,7675	5,37118	-1,72593	Sulfite exporter TauE/SafE family protein
60	Glyma15g14040	28,3434	8,56329	-1,72678	FAD-binding Berberine family protein
61	Glyma13g12090	79,2908	23,8427	-1,7336	-
62	Glyma12g10580	16,0694	4,80273	-1,74239	Aconitase 3
63	Glyma17g04780	10,9575	3,25916	-1,74935	Major facilitator superfamily protein
64	Glyma06g44250	2,16767	0,64084	-1,75811	NAC domain containing protein 73
65	Glyma03g14210	54,1529	15,8458	-1,77293	HXXXXD-type acyl-transferase family protein
66	Glyma08g13510	45,6107	13,3459	-1,77297	Plantacyanin
67	Glyma13g22890	13,8287	4,01756	-1,78328	Raffinose synthase family protein
68	Glyma11g18570	65,3415	18,6797	-1,80653	NAD(P)-binding Rossmann-fold superfamily protein
69	Glyma17g01330	21,3791	5,97925	-1,83817	Ethylene-forming enzyme
70	Glyma09g29371	67,5839	18,6822	-1,85501	-
71	Glyma10g29240	27,5535	7,35374	-1,90569	Translationally controlled tumor protein
72	Glyma07g33880	10,7232	2,83939	-1,91708	UDP-glycosyltransferase 73B4
73	Glyma06g17020	10,2942	2,71877	-1,9208	Clp ATPase
74	Glyma02g42180	21,1199	5,47849	-1,94675	HXXXXD-type acyl-transferase family protein
75	Glyma13g28970	8,44283	2,16657	-1,96231	Gibberellin 2-oxidase
76	Glyma03g33160	8,29584	2,10664	-1,97744	Eexocyst subunit exo70 family protein H4
77	Glyma10g44370	199,196	50,2039	-1,98832	-
78	Glyma09g37270	11,0956	2,60838	-2,08877	Transmembrane amino acid transporter family protein
79	Glyma16g01715	14,8131	3,40979	-2,11912	RING/U-box superfamily protein

Tabela 2 - Transcritos diferencialmente expressos, em ordem decrescente, resultantes da filtragem

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
80	Glyma02g26160	10,5401	2,34713	-2,16692	Lipoxygenase 2
81	Glyma17g07400	13,8232	3,06662	-2,17237	Deoxyxylulose-5-phosphate synthase
82	Glyma01g36590	7,63889	1,65407	-2,20735	Transmembrane amino acid transporter family protein
83	Glyma11g04130	23,0716	4,87748	-2,24191	TIFY domain/Divergent CCT motif family protein
84	Glyma09g32810	5,14022	1,07099	-2,26289	Auxin efflux carrier family protein
85	Glyma01g04760	13,7691	2,84423	-2,27533	Thioredoxin 2
86	Glyma06g15411	13,9159	2,80574	-2,31028	-
87	Glyma11g31330	9,44864	1,87333	-2,3345	Glutathione S-transferase TAU 19
88	Glyma01g26570	62,4646	12,1259	-2,36494	HSP20-like chaperones superfamily protein
89	Glyma14g36470	9,0345	1,72397	-2,38971	-
90	Glyma05g05880	24,6199	4,31431	-2,51263	Expansin-like B1
91	Glyma12g32381	8,82044	1,48906	-2,56645	Terpene synthase 02
92	Glyma08g44700	3,79448	0,63325	-2,58305	UDP-Glycosyltransferase superfamily protein
93	Glyma15g07710	7,59817	1,23868	-2,61685	-
94	Glyma12g32370	4,93734	0,77353	-2,6742	Terpene synthase 02
95	Glyma15g15200	4,60115	0,68579	-2,74616	Glycosyl hydrolase superfamily protein
96	Glyma02g06460	10,6095	1,46209	-2,85925	Major facilitator superfamily protein
97	Glyma12g25950	8,40378	1,15527	-2,86281	Flavodoxin-like quinone reductase 1
98	Glyma02g43580	7,29364	0,99215	-2,878	Ethylene-forming enzyme
99	Glyma18g03066	5,40566	0,71084	-2,92686	CRINKLY4 related 3
100	Glyma03g37610	6,70692	0,86597	-2,95326	C2H2 and C2HC zinc fingers superfamily protein
101	Glyma16g29490	0,88831	0,10698	-3,05379	Disease resistance family protein / LRR family protein
102	Glyma16g26940	7,81494	0,90627	-3,10823	Glutamate dehydrogenase 1
103	Glyma19g27130	19,5366	2,25588	-3,11442	NAD(P)-linked oxidoreductase superfamily protein
104	Glyma15g37520	3,18311	0,36087	-3,1409	UDP-glucosyl transferase 85A3
105	Glyma20g11600	8,75381	0,94993	-3,20402	Lipoxygenase 2
106	Glyma10g04190	3,12018	0,33278	-3,22901	Integrase-type DNA-binding superfamily protein

Tabela 2 - Transcritos diferencialmente expressos, em ordem decrescente, resultantes da filtragem

(conclusão)

N°	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
107	Glyma20g37351	3,12758	0,31296	-3,32102	-
108	Glyma01g37820	6,10839	0,53238	-3,52026	NmrA-like negative transcriptional regulator family protein
109	Glyma10g35640	11,9278	1,0187	-3,54954	Seven transmembrane MLO family protein
110	Glyma04g05510	2,23398	0,16771	-3,73561	Cytochrome P450, family 711, subfamily A, polypeptide 1
111	Glyma01g22880	4,96006	0,34077	-3,86348	Chalcone and stilbene synthase family protein
112	Glyma16g24831	19,7636	0,72814	-4,76249	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
113	Glyma06g45910	3,93182	0,13788	-4,83372	Peroxidase superfamily protein
114	Glyma17g15710	96,0235	2,99735	-5,00163	Expansin-like B1
115	Glyma16g25140	3,09996	0,03681	-6,39582	Disease resistance protein (TIR-NBS-LRR class), putative
116	Glyma18g47280	34,5412	0,24314	-7,15041	ACC synthase 1
117	Glyma13g12460	3,83999	0,1	-1,7977E+308	-
118	Glyma19g23446	2,13928	0,1	-1,7977E+308	GDSL-motif lipase 5
119	Glyma19g31811	2,05305	0,1	-1,7977E+308	-

Visando identificar o metabolismo completo de resposta à infestação, abrangendo todos os transcritos diferencialmente expressos na variedade IAC-100 que é resistente, foi feita a comparação entre as amostras de IAC-100 Infestada e IAC-100 Não Infestada. Como o resultado gerado foi de 2291 transcritos diferencialmente expressos, os mesmos foram classificados de forma decrescente de acordo com seus valores de fold change e, após essa classificação, realizou-se a análise dos 100 primeiros genes que obtiveram a mudança mais drástica em suas expressões (Tabela 3). Os 2291 transcritos também foram classificados de forma crescente de acordo com os valores de fold change para analisar os 100 genes que obtiveram as menores modificações em suas expressões, mostrando-se mais estáveis à presença do estresse de infestação (Tabela 4).

Tabela 3 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat* (continua)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma08g19470	0,1	1,45683	1,7977E+308	-
2	Glyma09g37310	0,1	3,39122	1,7977E+308	Mini zinc finger 2
3	Glyma20g35127	0,1	6,66356	1,7977E+308	-
4	Glyma19g43940	0,131734	27,1474	7,68705	Li-tolerant lipase 1
5	Glyma08g20170	0,195646	9,61851	5,61949	Homeobox protein 21
6	Glyma20g03661	0,330011	6,05694	4,198	Cell wall / vacuolar inhibitor of fructosidase 1
7	Glyma03g41330	0,231726	4,18857	4,17596	GDSL-like Lipase/Acylhydrolase superfamily protein ¹
8	Glyma20g03630	0,257451	4,30878	4,06491	Cell wall / vacuolar inhibitor of fructosidase 1
9	Glyma15g01370	2,1823	33,5247	3,9413	Protein of unknown function, DUF642
10	Glyma19g02810	1,26057	17,0477	3,75744	Expansin A15
11	Glyma04g00680	1,42273	17,4487	3,61638	Pollen Ole e 1 allergen and extensin family protein
12	Glyma03g15870	0,159501	1,93604	3,60147	Myb domain protein 118
13	Glyma09g03620	0,344907	3,62361	3,39315	Pectin lyase-like superfamily protein
14	Glyma03g36610	0,362323	3,79829	3,39	Peroxidase superfamily protein
15	Glyma02g03210	0,117096	1,22261	3,3842	Matrixin family protein
16	Glyma15g07600	0,206341	2,01294	3,2862	Serine carboxypeptidase-like 31
17	Glyma12g04960	2,01944	19,1124	3,24248	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin Superfamily protein
18	Glyma02g03250	13,483	121,807	3,17538	Matrixin family protein
19	Glyma12g29670	0,351029	3,07638	3,13157	FASCICLIN-like arabinogalactan-protein 11
20	Glyma13g35000	0,411284	3,5639	3,11525	Squamosa promoter-binding protein-like 12
21	Glyma15g34166	1,42317	11,9328	3,06776	-
22	Glyma07g16060	2,10984	17,021	3,01211	Laccase 17
23	Glyma12g07411	1,63804	12,1607	2,89218	FASCICLIN-like arabinogalactan-protein 11
24	Glyma18g13540	1,43111	10,5654	2,88414	GDSL-like Lipase/Acylhydrolase superfamily protein
25	Glyma18g04260	0,421046	3,08622	2,87379	Barwin-like endoglucanases superfamily protein
26	Glyma07g29990	9,75769	71,2247	2,86777	-
27	Glyma05g24330	1,68769	12,0594	2,83704	GDSL-like Lipase/Acylhydrolase superfamily protein
28	Glyma16g06640	0,857984	5,91425	2,78517	RmlC-like cupins superfamily protein

Tabela 3 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat*

(continuação)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
29	Glyma12g06480	5,01395	33,7985	2,75294	Laccase 2
30	Glyma16g04865	0,924349	6,16146	2,73676	HXXXD-type acyl-transferase family protein
31	Glyma08g19460	0,495309	3,28812	2,73086	Nodulin MtN21 /EamA-like transporter family protein
32	Glyma11g07845	1,62347	10,737	2,72543	-
33	Glyma10g34460	1,51751	10,0082	2,72141	Cytochrome P450, family 76, subfamily C, polypeptide 4
34	Glyma01g35800	0,411875	2,7105	2,71828	ATP-binding cassette 14
35	Glyma13g42770	0,864961	5,6731	2,71343	Pyridoxal phosphate phosphatase-related protein
36	Glyma11g04000	0,649471	4,20367	2,69431	HXXXD-type acyl-transferase family protein
37	Glyma05g04440	5,25655	33,5991	2,67623	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
38	Glyma18g42580	3,4935	22,0607	2,65873	Glycerol-3-phosphate acyltransferase 6
39	Glyma03g09140	0,433293	2,69661	2,63773	Transmembrane amino acid transporter family protein
40	Glyma15g14540	0,327823	2,00463	2,61235	Pectin lyase-like superfamily protein
41	Glyma16g26650	1,4235	8,50375	2,57866	HXXXD-type acyl-transferase family protein
42	Glyma19g05224	1,46862	8,7293	2,5714	HXXXD-type acyl-transferase family protein
43	Glyma13g26600	3,31347	19,5331	2,5595	Eukaryotic aspartyl protease family protein
44	Glyma18g42520	5,69117	33,283	2,54799	Laccase/Diphenol oxidase family protein
45	Glyma15g01500	0,448562	2,60754	2,53931	Gibberellin 3-oxidase 1
46	Glyma04g06230	0,620037	3,59794	2,53675	Glutamine dumper 4
47	Glyma06g12270	0,773466	4,41648	2,51349	Amino acid permease 3
48	Glyma07g01200	2,24996	12,7935	2,50744	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
49	Glyma11g14600	11,5684	65,0895	2,49224	Laccase 2
50	Glyma02g37990	2,59149	14,5801	2,49215	Matrix metalloproteinase
51	Glyma07g11580	0,38394	2,14172	2,47982	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
52	Glyma04g42520	0,2893	1,61289	2,47901	Amino acid permease 3
53	Glyma16g08300	0,262268	1,45285	2,46977	Long-chain fatty alcohol dehydrogenase family protein
54	Glyma19g44790	1,21253	6,71208	2,46874	Cytochrome P450, family 78, subfamily A, polypeptide 6
55	Glyma17g27520	3,24095	17,8652	2,46266	Integrase-type DNA-binding superfamily protein

Tabela 3 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat*

(continuação)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
56	Glyma07g09500	0,364549	2,00009	2,45588	Flavin-binding monooxygenase family protein
57	Glyma16g24640	1,05879	5,70809	2,43059	Peroxidase superfamily protein
58	Glyma01g40460	1,14843	6,15153	2,42128	Xyloglucan endotransglucosylase/hydrolase 6
59	Glyma06g04831	0,34496	1,84733	2,42094	Family of unknown function (DUF566)
60	Glyma01g20740	1,72447	9,15459	2,40834	DNA glycosylase superfamily protein
61	Glyma09g38620	1,24035	6,55741	2,40238	ROP-interactive CRIB motif-containing protein 4
62	Glyma13g01970	1,21887	6,34266	2,37955	Phosphoglycerate mutase family protein
63	Glyma15g38271	1,85682	9,64675	2,37721	-
64	Glyma08g07380	15,2247	78,3028	2,36265	-
65	Glyma03g41061	0,335735	1,72609	2,36212	-
66	Glyma14g09110	1,04153	5,24918	2,33339	Cytochrome P450, family 707, subfamily A, polypeptide 4
67	Glyma12g06130	0,375012	1,88847	2,33221	Aldehyde dehydrogenase 3F1
68	Glyma19g07080	4,55139	22,8981	2,33085	GDSL-like Lipase/Acylhydrolase superfamily protein
69	Glyma02g14410	1,68789	8,46308	2,32596	-
70	Glyma03g38080	19,1023	95,1572	2,31657	Galactinol synthase 1
71	Glyma08g24900	0,338875	1,68228	2,31159	Chaperone DnaJ-domain superfamily protein
72	Glyma01g27900	0,749232	3,67191	2,29304	Glycerol-3-phosphate acyltransferase 6
73	Glyma09g24820	0,392885	1,91108	2,28221	MATE efflux family protein
74	Glyma15g38091	45,6457	220,284	2,27082	-
75	Glyma19g09290	1,62006	7,78674	2,26497	Cytochrome P450, family 96, subfamily A, polypeptide 1
76	Glyma20g27040	4,0742	19,4719	2,2568	Glycosyl hydrolase family 10 protein
77	Glyma06g03990	1,24977	5,92051	2,24406	-
78	Glyma18g40070	1,0081	4,71993	2,22712	Laccase 17
79	Glyma15g16590	0,634082	2,95734	2,22156	Purine permease 4
80	Glyma0022s00480	1,22861	5,71612	2,21801	DNA glycosylase superfamily protein
81	Glyma13g35970	1,834	8,47848	2,20881	BURP domain-containing protein
82	Glyma04g40600	0,401762	1,85294	2,2054	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
83	Glyma20g02210	8,99008	41,1787	2,19549	Polyketide cyclase/dehydrase and lipid transport superfamily protein

Tabela 3 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat*

(conclusão)

N°	GENE ID	IAC100Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
84	Glyma10g28020	14,384	65,8427	2,19456	Germin-like protein 10
85	Glyma07g17720	11,1411	50,9691	2,19373	Glycerol-3-phosphate acyltransferase 6
86	Glyma03g30180	8,61631	39,2401	2,18719	Peroxidase 2
87	Glyma05g00730	0,691408	3,12673	2,17705	Flavin-binding monooxygenase family protein
88	Glyma08g11560	1,41471	6,39669	2,17682	HXXXD-type acyl-transferase family protein
89	Glyma18g03540	1,00684	4,54339	2,17394	Concanavalin A-like lectin family protein
90	Glyma13g40960	8,23065	36,8348	2,16199	UDP-XYL synthase 5
91	Glyma13g19220	1,11115	4,93326	2,15048	GDSL-like Lipase/Acylhydrolase superfamily protein ¹
92	Glyma18g48530	4,24123	18,6891	2,13964	Subtilisin-like serine endopeptidase family protein
93	Glyma06g05240	0,504774	2,22016	2,13696	1-amino-cyclopropane-1-carboxylate synthase 8
94	Glyma11g02400	0,856388	3,72679	2,1216	Myb domain protein 103
95	Glyma19g07000	1,32273	5,75269	2,12072	GDSL-like Lipase/Acylhydrolase superfamily protein
96	Glyma18g33230	1,78446	7,76055	2,12067	-
97	Glyma07g09150	0,61578	2,66662	2,11453	Cytochrome P450, family 704, subfamily A, polypeptide 2
98	Glyma12g30420	0,849453	3,65334	2,10461	Serine carboxypeptidase S28 family protein
99	Glyma03g37420	0,645627	2,7678	2,09997	Glycosyl hydrolase 9B13
100	Glyma07g39340	2,02213	8,66403	2,09916	PA-domain containing subtilase family protein

Tabela 4 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat*

(continua)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma02g00890	7,31321	0,1	-1,7977E+308	Integrase-type DNA-binding superfamily protein
2	Glyma02g06140	9,71866	0,1	-1,7977E+308	-
3	Glyma03g35980	4,01428	0,1	-1,7977E+308	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
4	Glyma05g01780	1,93237	0,1	-1,7977E+308	-
5	Glyma05g08690	1,67531	0,1	-1,7977E+308	Myb domain protein 21
6	Glyma06g17625	1,44766	0,1	-1,7977E+308	Fatty acyl-ACP thioesterases B
7	Glyma06g45540	1,36004	0,1	-1,7977E+308	Myb domain protein 14
8	Glyma07g31061	7,91775	0,1	-1,7977E+308	-
9	Glyma08g26780	0,909737	0,1	-1,7977E+308	UDP-Glycosyltransferase superfamily protein
10	Glyma11g25660	3,47393	0,1	-1,7977E+308	Calmodulin-like 41
11	Glyma11g34000	2,4471	0,1	-1,7977E+308	Calcium-dependent lipid-binding (CaLB domain) family protein
12	Glyma14g35650	1,86387	0,1	-1,7977E+308	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
13	Glyma15g06000	1,14515	0,1	-1,7977E+308	UDP-glucosyl transferase 85A2
14	Glyma16g04360	1,35603	0,1	-1,7977E+308	Spermidine hydroxycinnamoyl transferase
15	Glyma16g32000	5,41153	0,1	-1,7977E+308	Cytochrome P450, family 71, subfamily A, polypeptide 26
16	Glyma16g33770	2,2705	0,1	-1,7977E+308	Kunitz family trypsin and protease inhibitor protein
17	Glyma17g01760	3,21144	0,1	-1,7977E+308	-
18	Glyma18g16780	1,47382	0,1	-1,7977E+308	Disease resistance protein (TIR-NBS class), putative
19	Glyma19g34650	2,10005	0,1	-1,7977E+308	Integrase-type DNA-binding superfamily protein
20	Glyma19g34696	2,10143	0,1	-1,7977E+308	Ethylene response factor 1
21	Glyma20g27480	0,760698	0,1	-1,7977E+308	Cysteine-rich RLK (RECEPTOR-like protein kinase) 25
22	Glyma20g27480	0,760698	0,1	-1,7977E+308	Cysteine-rich RLK (RECEPTOR-like protein kinase) 29
23	Glyma20g34810	4,89663	0,1	-1,7977E+308	Serine protease inhibitor, potato inhibitor I-type family protein
24	Glyma17g15710	96,0235	0,248019	-8,59679	Expansin-like B1
25	Glyma09g04191	25,119	0,085651	-8,19609	Glycosyl hydrolase superfamily protein
26	Glyma13g16950	25,4621	0,101762	-7,96701	Oxophytodienoate-reductase 3
27	Glyma15g05820	58,7334	0,305612	-7,58633	Peroxidase superfamily protein
28	Glyma04g35380	47,1006	0,303406	-7,27835	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein

Tabela 4 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat*

(continuação)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
29	Glyma20g36810	41,4882	0,277333	-7,22494	Pathogenesis-related 4
30	Glyma02g07940	28,7794	0,230664	-6,9631	Glutamate dehydrogenase 1
31	Glyma02g07940	28,7794	0,230664	-6,9631	Glutamate dehydrogenase 3
32	Glyma08g22530	55,6382	0,481869	-6,85129	-
33	Glyma05g25690	11,4296	0,105716	-6,75643	-
34	Glyma17g11490	13,3923	0,124803	-6,7456	Serine-rich protein-related
35	Glyma16g02320	20,4139	0,193172	-6,72352	Basic helix-loop-helix (bHLH) DNA-binding family protein
36	Glyma06g00880	18,1156	0,172988	-6,71041	SAUR-like auxin-responsive protein family
37	Glyma15g02850	29,8077	0,285397	-6,70657	Glycosyl hydrolases family 32 protein
38	Glyma17g34530	14,343	0,143381	-6,64436	Cytochrome P450, family 711, subfamily A, polypeptide 1
39	Glyma14g05355	154,479	1,59705	-6,59586	Ethylene-forming enzyme
40	Glyma02g06150	4,60727	0,0480449	-6,58338	-
41	Glyma07g16810	93,7072	0,991232	-6,56279	Glutathione S-transferase TAU 8
42	Glyma16g04740	68,3471	0,814062	-6,3916	NAC domain containing protein 47
43	Glyma05g05880	24,6199	0,304144	-6,33893	Expansin-like B1
44	Glyma06g19400	5,86226	0,0737306	-6,31305	Carbonic anhydrase 2
45	Glyma08g04670	34,0948	0,429855	-6,30956	Myb domain protein 62
46	Glyma14g22700	26,4824	0,334119	-6,30853	Regulator of chromosome condensation (RCC1) family protein
47	Glyma08g14900	15,0542	0,190254	-6,3061	Ferulic acid 5-hydroxylase 1
48	Glyma07g37240	907,179	11,5474	-6,29575	MLP-like protein 423
49	Glyma08g19180	4,60597	0,0586475	-6,29529	Peroxidase superfamily protein
50	Glyma16g27950	14,2906	0,18325	-6,28511	Integrase-type DNA-binding superfamily protein
51	Glyma15g05810	3,59225	0,0464729	-6,27236	Peroxidase superfamily protein
52	Glyma06g19490	31,9806	0,4195	-6,25238	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
53	Glyma14g08000	28,066	0,37595	-6,22214	Phloem protein 2-B10
54	Glyma10g33810	14,2976	0,19253	-6,21455	Integrase-type DNA-binding superfamily protein
55	Glyma14g05641	12,5129	0,171057	-6,1928	-
56	Glyma15g15590	79,308	1,10073	-6,17094	MLP-like protein 423

Tabela 4 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat*

(continuação)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
57	Glyma16g24831	19,7636	0,289969	-6,0908	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
58	Glyma15g18360	200,484	2,94595	-6,08861	Xyloglucan endotransglycosylase 6
59	Glyma07g37265	921,969	13,8862	-6,05299	-
60	Glyma08g05500	12,6867	0,197126	-6,00805	Ethylene-forming enzyme
61	Glyma10g32830	53,8422	0,842982	-5,99709	Serine protease inhibitor, potato inhibitor I-type family protein
62	Glyma15g01990	8,44056	0,133828	-5,97889	PLAC8 family protein
63	Glyma12g32381	8,82044	0,143694	-5,93978	Terpene synthase 02
64	Glyma14g05350	54,75	0,903608	-5,92102	Ethylene-forming enzyme
65	Glyma09g29340	9,77001	0,162321	-5,91144	Kunitz family trypsin and protease inhibitor protein
66	Glyma18g47280	34,5412	0,590575	-5,87005	ACC synthase 1
67	Glyma10g40320	25,6122	0,441656	-5,85776	-
68	Glyma03g26530	13,3749	0,233786	-5,8382	Ethylene-responsive element binding factor 13
69	Glyma02g00870	5,7822	0,101152	-5,83703	Ethylene response factor 1
70	Glyma17g03365	228,5	4,04033	-5,82158	MLP-like protein 423
71	Glyma06g26610	10,4553	0,188555	-5,79311	Blue-copper-binding protein
72	Glyma02g06460	10,6095	0,193579	-5,77629	Major facilitator superfamily protein
73	Glyma09g01900	49,3219	0,903354	-5,77079	-
74	Glyma18g45920	29,024	0,539461	-5,74958	BON association protein 2
75	Glyma07g16830	45,0977	0,843988	-5,73969	Glutathione S-transferase TAU 8
76	Glyma10g44170	23,478	0,449063	-5,70825	Homogentisate phytyltransferase 1
77	Glyma19g32700	66,0885	1,27032	-5,70113	Polygalacturonase inhibiting protein 1
78	Glyma10g32380	10,08	0,195999	-5,6845	Eukaryotic aspartyl protease family protein
79	Glyma13g18410	5,22553	0,101904	-5,6803	Ethylene response factor 1
80	Glyma02g39840	43,1738	0,845279	-5,67458	-
81	Glyma11g31330	9,44864	0,188989	-5,64373	Glutathione S-transferase TAU 19
82	Glyma08g26840	9,2773	0,185734	-5,6424	UDP-Glycosyltransferase superfamily protein
83	Glyma14g06710	34,0091	0,682153	-5,63968	HXXXD-type acyl-transferase family protein
84	Glyma04g00870	7,4658	0,150161	-5,63572	SAUR-like auxin-responsive protein family

Tabela 4 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *TopHat*

(conclusão)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
85	Glyma19g27130	19,5366	0,394415	-5,63032	NAD(P)-linked oxidoreductase superfamily protein
86	Glyma07g16840	22,6411	0,458306	-5,62649	Glutathione S-transferase TAU 8
87	Glyma08g44230	59,7011	1,21951	-5,61338	Nudix hydrolase homolog 2
88	Glyma11g06690	5,42905	0,111882	-5,60065	Cytochrome P450, family 71, subfamily B, polypeptide 34
89	Glyma03g33070	3,5641	0,0749724	-5,57103	C2H2 and C2HC zinc fingers superfamily protein
90	Glyma09g07070	60,0027	1,26466	-5,5682	Xyloglucan endotransglycosylase 6
91	Glyma12g09940	5,56914	0,119309	-5,54467	FAD/NAD(P)-binding oxidoreductase family protein
92	Glyma12g22880	59,2567	1,27362	-5,53997	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
93	Glyma03g38490	40,1271	0,862772	-5,53946	Glycosyl hydrolase superfamily protein
94	Glyma12g06310	33,3982	0,731025	-5,5137	Senescence-associated gene 13
95	Glyma10g33060	16,3356	0,360321	-5,5026	Ethylene response factor 1
96	Glyma11g04970	29,1018	0,642215	-5,50191	VQ motif-containing protein
97	Glyma18g51680	7,50344	0,166876	-5,49071	Integrase-type DNA-binding superfamily protein
98	Glyma07g02630	11,232	0,25062	-5,48597	WRKY DNA-binding protein 40
99	Glyma15g06780	122,792	2,76279	-5,47394	Basic pathogenesis-related protein 1
100	Glyma01g40320	4,74879	0,107348	-5,46719	VQ motif-containing protein

A comparação entre os resultados das Tabelas 2, 3 e 4 induz a alguns questionamentos como o fato desses transcritos apresentados nessas tabelas serem exclusivos à variedade IAC-100 ou também serem expressos na variedade CD-215 que é suscetível aos percevejos sugadores das vagens da soja. Dessa maneira, foi realizada a comparação entre as amostras CD-215 Infestada e CD-215 Não Infestada na qual foram obtidos 569 transcritos diferencialmente expressos. Para facilitar a análise desses transcritos resultantes, classificou-se os 100 genes que tiveram suas expressões reguladas de maneira mais acentuada pelo processo de infestação, ou seja, a partir dos maiores valores de fold change para os menores (Tabela 5). Os 100 transcritos que tiveram os menores valores de fold change, ou seja, que não foram acentuadamente influenciados pela infestação também foram classificados (Tabela 6), a fim de identificarmos os genes que não seriam os principais atuantes no mecanismo de resposta à infestação.

Tabela 5 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *TopHat* (continua)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma18g44680	4,31148	16,2071	1,91037	Lipid transfer protein 3
2	Glyma19g30460	2,17797	8,14987	1,90379	Alpha-L-arabinofuranosidase 1
3	Glyma02g37610	4,71322	16,8838	1,84085	Eukaryotic aspartyl protease family protein
4	Glyma03g38080	6,47224	21,0819	1,70367	Galactinol synthase 1
5	Glyma08g14670	5,50679	14,4637	1,39315	Myo-inositol-1-phosphate synthase 3
6	Glyma08g14670	5,50679	14,4637	1,39315	Myo-inositol-1-phosphate synthase 1
7	Glyma08g14670	5,50679	14,4637	1,39315	Myo-inositol-1-phosphate synthase 3
8	Glyma08g17010	7,48336	15,5528	1,05542	ATP-citrate lyase A-1
9	Glyma19g40300	2,04491	3,91763	0,937947	Beta-xylosidase 2
10	Glyma08g43460	7,11184	12,7847	0,846128	IAA carboxymethyltransferase 1
11	Glyma16g23240	2,00151	3,56838	0,834181	PLAC8 family protein
12	Glyma15g39820	4,35715	6,43773	0,563167	Interactor of constitutive active rops 1
13	Glyma11g10810	3,78313	3,25703	-0,216022	Mitogen-activated protein kinase kinase kinase 7
14	Glyma07g03680	49,0784	38,2654	-0,359047	Rubredoxin-like superfamily protein
15	Glyma10g35741	29,3852	22,5105	-0,384491	-
16	Glyma02g12970	16,6104	12,4079	-0,420824	Ureide permease 2
17	Glyma13g18805	4,91297	3,41463	-0,524867	Phosphate starvation response 1
18	Glyma11g05670	66,6862	44,8468	-0,572385	Sumo conjugation enzyme 1
19	Glyma01g42331	7,25645	4,68276	-0,631904	YTH family protein
20	Glyma18g49900	8,96573	5,63458	-0,670114	RING/U-box superfamily protein
21	Glyma16g30081	2,55965	1,55125	-0,722521	SKP1-like 21
22	Glyma02g47760	11,121	6,55982	-0,761558	-
23	Glyma11g25860	18,1538	10,451	-0,796625	-
24	Glyma16g04530	14,6568	8,11635	-0,852669	Adenosine-5'-phosphosulfate (APS) kinase 3
25	Glyma01g02110	29,3376	15,5913	-0,912008	Tetratricopeptide repeat (TPR)-like superfamily protein
26	Glyma17g13751	9,98206	5,28406	-0,917691	Protein kinase superfamily protein
27	Glyma02g41930	5,26745	2,7545	-0,935313	Mitochondrial substrate carrier family protein
28	Glyma07g03200	11,4322	5,93421	-0,945976	Leucine-rich repeat family protein

Tabela 5 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *TopHat*

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
29	Glyma12g30920	67,6786	33,3827	-1,0196	A20/AN1-like zinc finger family protein
30	Glyma02g04210	3,0118	1,48441	-1,02074	Cysteine-rich RLK (RECEPTOR-like protein kinase) 2
31	Glyma14g00380	10,9873	5,32427	-1,04518	Protein kinase superfamily protein
32	Glyma15g17080	2,45456	1,15111	-1,09243	K+ uptake permease 10
33	Glyma05g34870	19,7903	9,17819	-1,10851	PLANT CADMIUM RESISTANCE 2
34	Glyma09g08290	13,4214	6,21455	-1,11081	Jasmonate-zim-domain protein 1
35	Glyma07g37270	101,189	43,9503	-1,2031	MLP-like protein 423
36	Glyma10g31240	6,50859	2,81534	-1,20904	Phloem protein 2-B10
37	Glyma10g31240	6,50859	2,81534	-1,20904	F-box family protein
38	Glyma17g04350	4,67445	1,95666	-1,2564	Pleiotropic drug resistance 9
39	Glyma06g13550	10,9253	4,51107	-1,27613	ABL interactor-like protein 2
40	Glyma18g07090	7,55085	3,0901	-1,28899	RING/U-box superfamily protein
41	Glyma06g12020	29,7755	12,1636	-1,29156	thylakoidal ascorbate peroxidase
42	Glyma06g12020	29,7755	12,1636	-1,29156	Stromal ascorbate peroxidase
43	Glyma01g35480	3,75761	1,50362	-1,32137	Pectinacylesterase family protein
44	Glyma12g03680	2,11415	0,838518	-1,33416	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain
45	Glyma15g19840	19,9648	7,62554	-1,38855	Jasmonate-zim-domain protein 1
46	Glyma15g19840	19,9648	7,62554	-1,38855	Jasmonate-zim-domain protein 1
47	Glyma15g19840	19,9648	7,62554	-1,38855	Jasmonate-zim-domain protein 1
48	Glyma04g39350	9,33904	3,53054	-1,40338	Protein kinase family protein
49	Glyma04g39350	9,33904	3,53054	-1,40338	Protein kinase family protein
50	Glyma04g39350	9,33904	3,53054	-1,40338	Protein kinase family protein
51	Glyma08g43080	55,3549	20,4945	-1,43347	GDSL-like Lipase/Acylhydrolase superfamily protein
52	Glyma02g37710	18,0407	6,67021	-1,43545	Beta carbonic anhydrase 5
53	Glyma03g25651	9,3759	3,44458	-1,44463	-
54	Glyma18g43210	10,166	3,71328	-1,45299	Fatty acid desaturase 8
55	Glyma05g32756	12,4393	4,48044	-1,4732	Alpha/beta-Hydrolases superfamily protein
56	Glyma15g01470	5,22546	1,85156	-1,49681	Pleiotropic drug resistance 12

Tabela 5 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *TopHat*

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
57	Glyma15g40190	51,3613	18,1482	-1,50086	Glutathione S-transferase TAU 19
58	Glyma18g10930	18,2613	6,44195	-1,50322	Protein of unknown function, DUF584
59	Glyma17g12710	24,8383	8,75177	-1,50492	-
60	Glyma13g04670	11,5648	4,04511	-1,51548	Cytochrome P450, family 82, subfamily C, polypeptide 4
61	Glyma07g02420	13,3368	4,62825	-1,52687	-
62	Glyma10g42820	18,8966	6,52022	-1,53513	Arabidopsis phospholipase-like protein (PEARLI 4) family
63	Glyma13g32430	28,0691	9,68	-1,5359	Ribonuclease E inhibitor RraA/Dimethylmenaquinone methyltransferase
64	Glyma07g09220	7,4019	2,54457	-1,54047	Phosphorylase superfamily protein
65	Glyma14g00570	19,2823	6,62275	-1,54177	VQ motif-containing protein
66	Glyma19g42000	8,22433	2,81105	-1,54879	Alpha/beta-Hydrolases superfamily protein
67	Glyma01g29670	9,14351	3,10712	-1,55717	ALC-interacting protein 1
68	Glyma12g09830	20,1906	6,85428	-1,55861	Polynucleotidyl transferase, ribonuclease H-like superfamily protein
69	Glyma01g03150	11,1874	3,78278	-1,56436	Octicosapeptide/Phox/Bem1p family protein
70	Glyma13g43870	9,79686	3,29458	-1,57222	Pleiotropic drug resistance 12
71	Glyma02g48100	5,95208	1,98418	-1,58485	Protein kinase superfamily protein
72	Glyma16g28605	25,7611	8,58324	-1,5856	-
73	Glyma13g17180	15,6934	5,22317	-1,58716	Jasmonate-zim-domain protein 1
74	Glyma16g25550	5,61383	1,85391	-1,59841	C2H2-like zinc finger protein
75	Glyma17g34740	10,6862	3,52647	-1,59945	Protein of unknown function (DUF778)
76	Glyma08g26830	23,7932	7,83959	-1,6017	UDP-Glycosyltransferase superfamily protein
77	Glyma02g03280	15,4358	5,075	-1,60481	-
78	Glyma02g15370	63,8762	20,8948	-1,61214	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
79	Glyma06g19600	19,6956	6,42893	-1,61522	-
80	Glyma14g07360	14,5042	4,68669	-1,62983	Expansin A4
81	Glyma11g35800	91,1499	29,4476	-1,63009	Senescence associated gene 20
82	Glyma17g34420	34,3163	11,0588	-1,63371	-
83	Glyma05g15870	5,71691	1,83726	-1,63768	Calcium-binding EF-hand family protein

Tabela 5 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *TopHat*

(conclusão)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
85	Glyma05g28310	14,9144	4,78471	-1,6402	Xyloglucan endotransglucosylase/hydrolase 16
86	Glyma17g37580	9,32924	2,99072	-1,64126	Transcriptional factor B3 family protein / auxin-responsive factor AUX/IAA-related
87	Glyma17g35430	7,89706	2,52503	-1,64502	Salt tolerance zinc finger
88	Glyma01g41450	22,9912	7,31889	-1,65139	Alpha/beta-Hydrolases superfamily protein
89	Glyma15g10070	14,179	4,50763	-1,65331	Gibberellin 2-oxidase
90	Glyma17g16020	15,9063	5,03603	-1,65924	Chloroplast beta-amylase
91	Glyma10g00980	17,9277	5,63544	-1,66959	Ethylene response factor 1
92	Glyma08g25520	33,5696	10,5485	-1,67011	Kunitz trypsin inhibitor 1
93	Glyma06g04030	18,5553	5,81941	-1,67289	Protein of unknown function (DUF1645)
94	Glyma01g29930	49,7091	15,4192	-1,68878	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
95	Glyma02g41070	10,8104	3,35323	-1,68879	Eukaryotic aspartyl protease family protein
96	Glyma11g12980	16,7364	5,181	-1,69168	Hemoglobin 1
97	Glyma16g33840	18,2466	5,6347	-1,69522	YELLOW STRIPE like 7
98	Glyma06g35585	21,4587	6,62246	-1,69612	Tyrosine transaminase family protein
99	Glyma09g31810	19,9497	6,15023	-1,69765	Cytochrome P450, family 71, subfamily A, polypeptide 22
100	Glyma02g11610	11,5317	3,55447	-1,6979	UDP-glycosyltransferase 73B4

Tabela 6 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 - alinhador *TopHat*

(continua)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma09g04191	2,97136	0,1	-1,7977E+308	Glycosyl hydrolase superfamily protein
2	Glyma15g10630	2,64114	0,1	-1,7977E+308	Major facilitator superfamily protein
3	Glyma20g36810	7,53362	0,1	-1,7977E+308	Pathogenesis-related 4
4	Glyma15g15590	105,14	0,641737	-7,35611	MLP-like protein 423
5	Glyma15g05820	17,1253	0,181316	-6,56148	Peroxidase superfamily protein
6	Glyma13g16950	7,64908	0,10024	-6,25375	Oxophytodienoate-reductase 3
7	Glyma14g05650	19,5885	0,290435	-6,07565	Glycine-rich protein 3 short isoform
8	Glyma08g22530	17,2668	0,338655	-5,67204	-
9	Glyma11g27720	12,816	0,257644	-5,63642	Glutamine-dependent asparagine synthase 1
10	Glyma16g02320	5,85924	0,128662	-5,50906	Basic helix-loop-helix (bHLH) DNA-binding family protein
11	Glyma12g06310	5,8504	0,138036	-5,40542	Senescence-associated gene 13
12	Glyma17g11490	5,3221	0,128987	-5,3667	Serine-rich protein-related
13	Glyma14g05355	48,7439	1,20267	-5,34091	Ethylene-forming enzyme
14	Glyma07g16810	58,7336	1,51476	-5,27703	Glutathione S-transferase TAU 8
15	Glyma01g38870	1,59643	0,0416609	-5,26001	Cytochrome P450, family 82, subfamily C, polypeptide 4
16	Glyma04g35380	20,2422	0,540434	-5,2271	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
17	Glyma13g35320	80,1897	2,23466	-5,16529	Low-molecular-weight cysteine-rich 66
18	Glyma13g35320	80,1897	2,23466	-5,16529	Low-molecular-weight cysteine-rich 68
19	Glyma09g29340	4,13783	0,116421	-5,15146	Kunitz family trypsin and protease inhibitor protein
20	Glyma11g00580	4,47941	0,12781	-5,13123	Pyridoxal-5'-phosphate-dependent enzyme family protein
21	Glyma06g13110	22,0155	0,635967	-5,11343	Nodulin MtN3 family protein
22	Glyma06g19490	14,7156	0,444807	-5,04802	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
23	Glyma04g00870	5,80438	0,184548	-4,97508	SAUR-like auxin-responsive protein family
24	Glyma14g05630	6,66376	0,21659	-4,9433	Calcium-binding EF-hand family protein
25	Glyma17g37660	8,71657	0,287608	-4,92159	Putative lysine decarboxylase family protein
26	Glyma14g22700	7,10443	0,242492	-4,87271	Regulator of chromosome condensation (RCC1) family protein
27	Glyma15g18360	30,4431	1,05009	-4,85754	Xyloglucan endotransglycosylase 6
28	Glyma08g05500	3,3347	0,117272	-4,82962	Ethylene-forming enzyme

Tabela 6 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 - alinhador *TopHat*

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
29	Glyma07g35630	11,1427	0,394225	-4,82094	NAC-like, activated by AP3/PI
30	Glyma17g13730	9,12676	0,3249	-4,81204	Malate synthase
31	Glyma09g09430	25,72	0,943476	-4,76876	Cytidine/deoxycytidylate deaminase family protein
32	Glyma15g06780	63,037	2,38531	-4,72395	Basic pathogenesis-related protein 1
33	Glyma07g16830	28,1511	1,11614	-4,6566	Glutathione S-transferase TAU 8
34	Glyma12g22880	14,3034	0,580015	-4,62413	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
35	Glyma20g30835	4,27731	0,173932	-4,6201	Integrase-type DNA-binding superfamily protein
36	Glyma11g16120	3,2316	0,134842	-4,58291	Protein of unknown function (DUF793)
37	Glyma19g26400	7,48463	0,331444	-4,49709	WRKY DNA-binding protein 75
38	Glyma03g38490	3,05384	0,136519	-4,48344	Glycosyl hydrolase superfamily protein
39	Glyma02g43560	222,755	10,0413	-4,47143	Ethylene-forming enzyme
40	Glyma14g37980	23,5992	1,06738	-4,46659	-
41	Glyma16g04740	18,0644	0,817906	-4,46507	NAC domain containing protein 47
42	Glyma10g32830	24,9521	1,14614	-4,4443	Serine protease inhibitor, potato inhibitor I-type family protein
43	Glyma08g26840	3,85854	0,181757	-4,40797	UDP-Glycosyltransferase superfamily protein
44	Glyma12g13290	9,40248	0,44483	-4,40172	Protein phosphatase 2C family protein
45	Glyma04g41680	28,9869	1,37588	-4,39697	Nodulin MtN3 family protein
46	Glyma05g25690	4,42815	0,213648	-4,3734	-
47	Glyma07g02590	3,13857	0,151986	-4,3681	Acyl-CoA N-acyltransferases (NAT) superfamily protein
48	Glyma02g08835	4,40962	0,214111	-4,36422	Integrase-type DNA-binding superfamily protein
49	Glyma20g16920	15,5893	0,758083	-4,36206	Integrase-type DNA-binding superfamily protein
50	Glyma10g34760	6,71139	0,32649	-4,3615	AP2/B3 transcription factor family protein
51	Glyma07g08311	6,13789	0,299664	-4,35633	BON association protein 2
52	Glyma17g03365	122,24	6,16525	-4,30941	MLP-like protein 423
53	Glyma01g38590	4,85715	0,248319	-4,28984	Cytochrome P450, family 71, subfamily B, polypeptide 34
54	Glyma12g16750	3,06665	0,156972	-4,28809	GRAS family transcription factor
55	Glyma18g44320	11,8704	0,626606	-4,24367	Peroxidase superfamily protein
56	Glyma15g09560	10,0694	0,533955	-4,23711	GDSL-like Lipase/Acylhydrolase superfamily protein

Tabela 6 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 - alinhador *TopHat*

(continuação)					
Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
57	Glyma13g20170	4,37209	0,231866	-4,23696	Peroxidase superfamily protein
58	Glyma01g38390	11,0279	0,588027	-4,22914	Homeobox protein 40
59	Glyma08g04670	10,7649	0,575641	-4,22502	Myb domain protein 62
60	Glyma10g44170	20,8883	1,11776	-4,22402	Homogentisate phytyltransferase 1
61	Glyma02g07940	3,45639	0,186907	-4,20888	Glutamate dehydrogenase 3
62	Glyma02g07940	3,45639	0,186907	-4,20888	Glutamate dehydrogenase 1
63	Glyma11g33040	4,74921	0,270932	-4,13168	Oxidative stress 3
64	Glyma01g40360	6,34533	0,362057	-4,13141	Phytosulfokine 4 precursor
65	Glyma15g06010	194,732	11,2118	-4,1184	EXORDIUM like 2
66	Glyma17g34530	3,31171	0,192324	-4,10597	Cytochrome P450, family 711, subfamily A, polypeptide 1
67	Glyma14g09990	58,1681	3,47282	-4,06605	Phytosulfokine 4 precursor
68	Glyma14g08000	7,29265	0,438842	-4,05467	Phloem protein 2-B10
69	Glyma06g06290	3,67086	0,220906	-4,05461	Glutamine dumper 3
70	Glyma07g16840	11,9704	0,731857	-4,03177	Glutathione S-transferase TAU 8
71	Glyma08g14900	4,13222	0,254232	-4,0227	Ferulic acid 5-hydroxylase 1
72	Glyma14g05350	16,7048	1,02789	-4,02251	Ethylene-forming enzyme
73	Glyma05g21680	10,203	0,628365	-4,02125	Auxin-responsive GH3 family protein
74	Glyma07g18670	2,24662	0,138916	-4,01546	Ammonium transporter 2
75	Glyma18g02400	20,5754	1,2766	-4,01054	-
76	Glyma05g36100	66,7813	4,20367	-3,98972	Myo-inositol oxygenase 4
77	Glyma19g43360	8,78894	0,5535	-3,98903	NAD(P)-linked oxidoreductase superfamily protein
78	Glyma13g42530	3,03328	0,192403	-3,97868	Glycosyl hydrolases family 32 protein
79	Glyma14g16630	5,00029	0,320619	-3,96308	O-Glycosyl hydrolases family 17 protein
80	Glyma19g32700	24,8457	1,60683	-3,95071	Polygalacturonase inhibiting protein 1
81	Glyma16g27880	31,7499	2,09106	-3,92445	Peroxidase superfamily protein
82	Glyma18g53170	2,12299	0,140667	-3,91574	Class I glutamine amidotransferase-like superfamily protein
83	Glyma02g39840	23,7417	1,57411	-3,91481	-
84	Glyma11g37930	11,1617	0,755882	-3,88425	-

Tabela 6 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 - alinhador *TopHat*

(conclusão)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
85	Glyma07g36130	2,42411	0,164212	-3,88382	-
86	Glyma09g04510	85,8039	5,93711	-3,85321	MLP-like protein 423
87	Glyma18g45920	13,2367	0,917449	-3,85077	BON association protein 2
88	Glyma10g33060	6,2762	0,436986	-3,84423	Ethylene response factor 1
89	Glyma20g38000	1,63128	0,113693	-3,84279	Respiratory burst oxidase homolog B
90	Glyma20g20355	5,63116	0,399252	-3,81806	-
91	Glyma11g04970	7,72554	0,549485	-3,81349	VQ motif-containing protein
92	Glyma17g18230	3,97816	0,285945	-3,79829	-
93	Glyma08g23430	4,20863	0,307019	-3,77695	Acyl-CoA N-acyltransferases (NAT) superfamily protein
94	Glyma08g44230	16,8607	1,23025	-3,77665	Nudix hydrolase homolog
95	Glyma15g02850	8,6673	0,639111	-3,76144	Glycosyl hydrolases family 32 protein
96	Glyma16g05190	9,58214	0,730389	-3,71361	Integrase-type DNA-binding superfamily protein
97	Glyma16g27950	4,46176	0,343454	-3,69942	Integrase-type DNA-binding superfamily protein
98	Glyma15g13500	46,786	3,6133	-3,69469	Peroxidase superfamily protein
99	Glyma11g27480	1,64129	0,127942	-3,68126	Glutamine-dependent asparagine synthase 1
100	Glyma10g32380	1,81875	0,141987	-3,67912	Eukaryotic aspartyl protease family protein

A partir dos resultados obtidos pelo alinhador TopHat juntamente com o Cufflinks, foram usadas as sequências dos aminoácidos dos transcritos diferencialmente expressos provenientes das comparações IAC-100 Infestada X IAC-100 Não Infestada e CD-215 Infestada X CD-215 Não Infestada e feito o Blast com UEKO. Identificadores foram gerados e plotados nas vias metabólicas do iPATH. Dessa forma, foi elaborado um mapa metabólico para cada variedade, a fim de identificar as diferenças entre os mecanismos de respostas em relação ao tratamento de infestação. A seguir estão apresentadas as vias IAC-100 Infestada x IAC-100 Não Infestada (Figura 4) e CD-215 Infestada x CD-215 Não Infestada (Figura 5). As linhas azuis mostram as vias metabólicas correspondentes aos transcritos diferencialmente expressos provenientes das comparações supracitadas. Quanto mais espessa é a linha azul mostrada no mapa, maior é o número de transcritos responsáveis pela via. As vias que obtiveram o maior número de transcritos diferencialmente expressos foram as relacionadas

com a produção de terpenos, de ácido salicílico e ácido jasmônico, metabolismo energético, metabolismo de lipídios e metabolismo de co-fatores e vitaminas (Figura 4)..

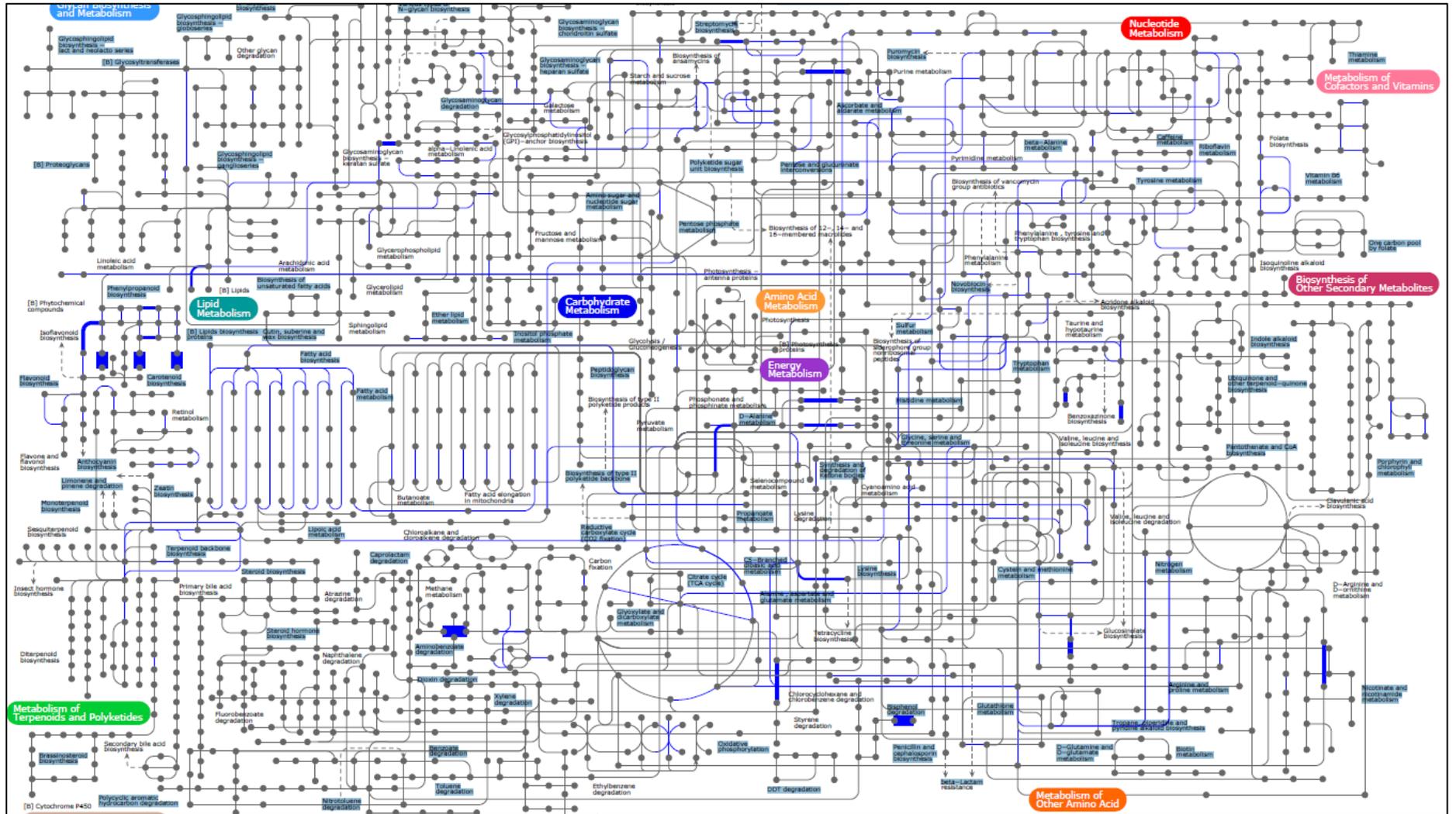


Figura 4 -Vias metabólicas ativadas pelos transcritos expressos na comparação entre as amostras IAC-100 Infestada X IAC-100 Não Infestada

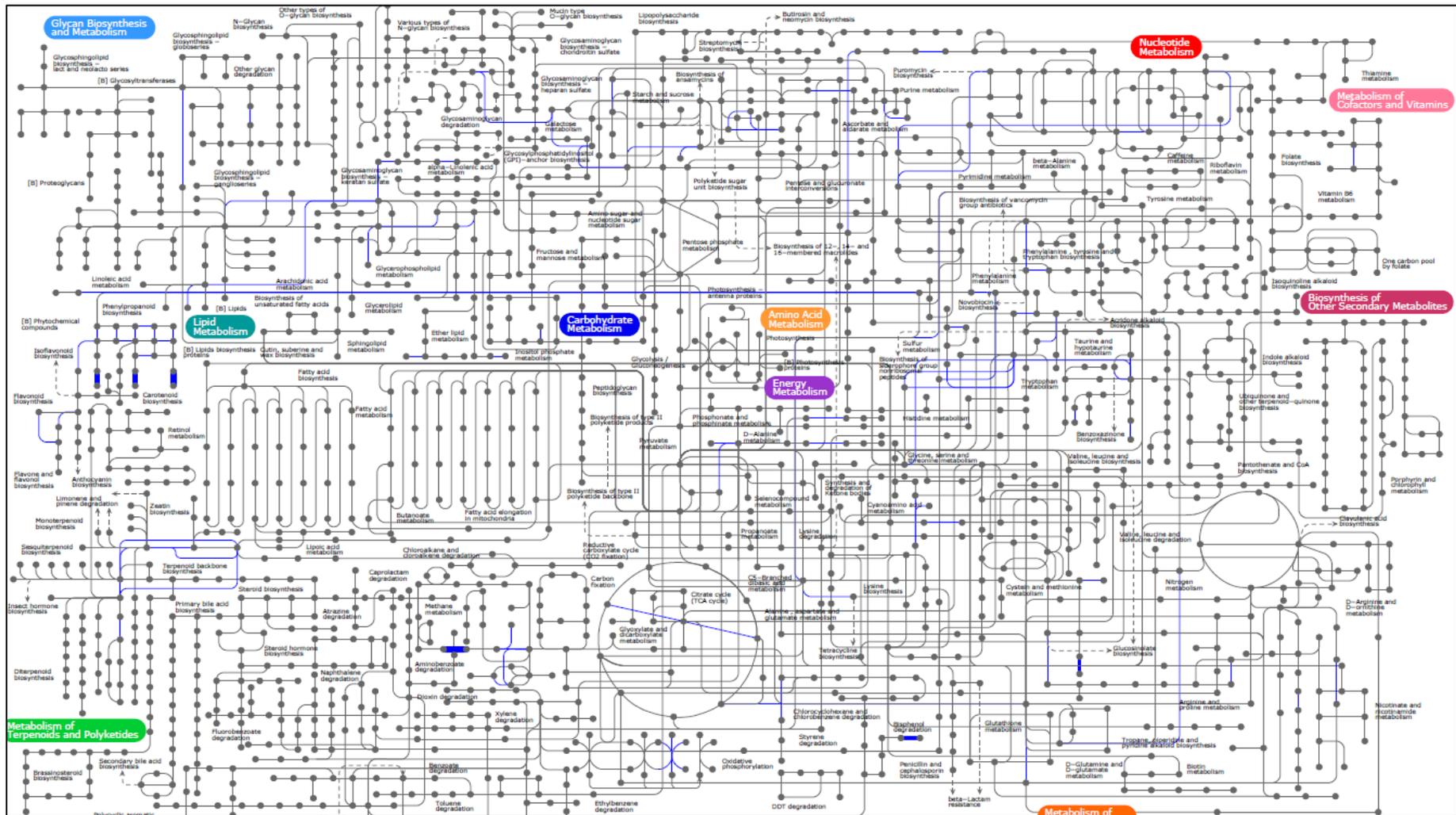


Figura 5 - Vias metabólicas ativadas pelos transcritos expressos na comparação entre as amostras CD-215 Infestada X CD-215 Não Infestada

5.2 Resultados obtidos com o alinhador *STAR*

Os resultados gerados pelo alinhador *STAR* foram de uma média de 90% de alinhamento único das *reads* em relação ao genoma referência, em contraposição à 84,6% de alinhamento único das *reads* realizado pelo alinhador *TopHat*.

As análises prévias das comparações entre as seguintes amostras foram:

- 1) IAC-100 Infestada x IAC-100 Não Infestada
2341 transcritos diferencialmente expressos
- 2) CD-215 Infestada x CD-215 Não Infestada
561 transcritos diferencialmente expressos
- 3) IAC-100 Infestada x CD-215 Infestada
619 transcritos diferencialmente expressos
- 4) IAC-100 Não Infestada x CD-215 Não Infestada
4043 transcritos diferencialmente expressos

Após esses resultados do *Cuffdiff*, foi feito um refinamento da seguinte forma:

1º) A tabela dos transcritos em IAC-100 Infestada x CD-215 Infestada foi comparada com a dos transcritos em IAC-100 Não Infestada x CD-215 Não Infestada e os transcritos iguais foram retirados. Sendo assim, ficaram apenas os transcritos expressos na situação de infestação.

2º) Os transcritos que restaram da comparação anterior foram comparados com os presentes na tabela dos CD-215 Infestada x CD-215 Não Infestada. Restaram somente os transcritos únicos da variedade IAC-100 sob a condição de infestação. Logo, dos 619 transcritos iniciais restaram apenas 128 transcritos.

Na tabela 7 encontram-se os transcritos resultantes da filtragem organizados de forma decrescente pelos seus valores de fold change que foram calculados a partir dos valores mostrados na tabela dos FPKM das amostras de IAC-100 Infestada e CD-215 Infestada pela eq.(1):

$$\log_2 \frac{FPKM_{IAC100Infestada}}{FPKM_{CD215Infestada}} \quad (1)$$

Tabela 7 - Os 128 transcritos diferencialmente expressos somente na variedade IAC-100 infestada, em ordem decrescente – alinhador *STAR*

(continua)

Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma19g31811	0,1	1,97948	1,7977E+308	-
2	Glyma18g47280	0,263918	34,5054	7,03059	ACC synthase 1
3	Glyma09g15135	0,099712	5,43359	5,768	BED zinc finger
4	Glyma16g25045	0,07105	3,05249	5,425	Disease resistance protein (TIR-NBS-LRR class), putative
5	Glyma17g15710	2,9702	93,1492	4,97091	Expansin-like B1
6	Glyma16g24831	0,767399	21,6171	4,81605	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
7	Glyma06g45910	0,140409	3,93433	4,80841	Peroxidase superfamily protein
8	Glyma19g01280	0,281768	7,21948	4,67931	Nodulin MtN3 family protein
9	Glyma16g25140	0,181276	3,33092	4,19966	Disease resistance protein (TIR-NBS-LRR class), putative
10	Glyma01g22880	0,288451	5,0504	4,13	Chalcone and stilbene synthase family protein
11	Glyma04g05510	0,1688	2,24949	3,73621	Cytochrome P450, family 711, subfamily A, polypeptide 1
12	Glyma10g35640	0,95973	10,5322	3,45603	Seven transmembrane MLO family protein
13	Glyma20g37351	0,315336	3,42863	3,44267	-
14	Glyma01g37820	0,506772	5,38012	3,40823	NmrA-like negative transcriptional regulator family protein
15	Glyma20g11600	0,946122	8,78408	3,21479	Lipoxygenase 2
16	Glyma16g26940	0,839969	7,41256	3,14156	Glutamate dehydrogenase 1
17	Glyma15g37520	0,377716	3,25861	3,10888	UDP-glucosyl transferase 85A3
18	Glyma09g04191	3,02905	25,3913	3,0674	Glycosyl hydrolase superfamily protein
19	Glyma16g29490	0,112706	0,908507	3,01093	Leucine-rich repeat transmembrane protein kinase
20	Glyma19g27130	1,69695	12,7272	2,9069	NAD(P)-linked oxidoreductase superfamily protein
21	Glyma03g37610	0,903044	6,7165	2,89484	C2H2 and C2HC zinc fingers superfamily protein
22	Glyma02g06460	1,48601	10,7788	2,85868	Major facilitator superfamily protein
23	Glyma02g43580	1,08172	7,82731	2,85519	Ethylene-forming enzyme
24	Glyma12g25950	1,21602	8,29455	2,76999	Quinone reductase family protein
25	Glyma15g15200	0,714011	4,85212	2,7646	Glycosyl hydrolase superfamily protein
26	Glyma12g32370	0,806963	5,11477	2,66409	Terpene synthase 02
27	Glyma15g07710	1,2829	8,04236	2,64821	-
28	Glyma16g25170	0,30402	1,90156	2,64494	disease resistance protein (TIR-NBS-LRR class), putative

Tabela 7 - Os 128 transcritos diferencialmente expressos somente na variedade IAC-100 infestada, em ordem decrescente – alinhador *STAR*

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
29	Glyma16g08990	0,608164	3,75189	2,62508	Non-specific phospholipase C4
30	Glyma08g44700	0,638571	3,89281	2,60789	UDP-Glycosyltransferase superfamily protein
31	Glyma12g32381	1,49712	8,86235	2,5655	Terpene synthase 02*
32	Glyma05g05880	4,20876	24,1382	2,51985	Expansin-like B1
33	Glyma15g40510	0,987737	5,5083	2,47941	NAC domain containing protein 1
34	Glyma14g36470	1,74261	9,2387	2,40644	-
35	Glyma01g26570	12,2355	63,5815	2,37753	HSP20-like chaperones superfamily protein
36	Glyma11g31330	1,85837	9,59461	2,36819	Glutathione S-transferase TAU 19
37	Glyma06g15411	2,50624	12,7578	2,34778	-
38	Glyma05g36050	0,419402	1,9868	2,24404	Phosphofruktokinase 3
39	Glyma09g32810	1,01343	4,78277	2,2386	Auxin efflux carrier family protein
40	Glyma01g36590	1,66398	7,68309	2,20705	Transmembrane amino acid transporter family protein
41	Glyma12g10620	1,0836	4,97186	2,19796	MATE efflux family protein
42	Glyma17g07400	3,0455	13,6931	2,1687	Deoxyxylulose-5-phosphate synthase
43	Glyma02g26160	2,4019	10,6123	2,14349	Lipoxygenase 2
44	Glyma13g25970	0,219863	0,957435	2,12257	LRR and NB-ARC domains-containing disease resistance protein
45	Glyma16g01715	3,45209	15,0283	2,12214	RING/U-box superfamily protein
46	Glyma09g37270	2,6281	11,3573	2,11153	Transmembrane amino acid transporter family protein
47	Glyma07g28940	64,1339	273,18	2,09069	BURP domain-containing protein
48	Glyma13g28970	2,11963	8,74761	2,04507	Gibberellin 2-oxidase
49	Glyma03g33160	2,11313	8,52468	2,01226	Exocyst subunit exo70 family protein H4
50	Glyma06g17020	2,6234	10,1627	1,95378	Clp ATPase
51	Glyma02g42180	5,60831	21,4762	1,9371	HXXXD-type acyl-transferase family protein
52	Glyma10g29240	6,88206	26,2769	1,93288	Translationally controlled tumor protein
53	Glyma07g33880	2,90572	10,9543	1,91454	UDP-glycosyltransferase 73B4
54	Glyma17g10450	2,78495	10,264	1,88187	Major facilitator superfamily protein
55	Glyma03g02370	1,57654	5,77924	1,87412	Calcium-dependent lipid-binding (CaLB domain) family protein
56	Glyma04g07390	1,26404	4,61264	1,86755	HCO3- transporter family

Tabela 7 - Os 128 transcritos diferencialmente expressos somente na variedade IAC-100 infestada, em ordem decrescente – alinhador *STAR*

(continuação)					
Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
57	Glyma09g29371	11,7789	42,8857	1,8643	-
58	Glyma17g01330	6,15595	21,6275	1,81281	Ethylene-forming enzyme
59	Glyma11g18570	18,8267	65,6819	1,80272	NAD(P)-binding Rossmann-fold superfamily protein
60	Glyma06g44250	0,629985	2,17976	1,79078	NAC domain containing protein 73
61	Glyma03g14210	16,1565	55,4994	1,78036	HXXXXD-type acyl-transferase family protein
62	Glyma13g22890	4,08689	14,0187	1,77828	Raffinose synthase family protein
63	Glyma12g10580	4,6119	15,7324	1,7703	Aconitase 3
64	Glyma17g04780	3,24472	10,9207	1,75089	Major facilitator superfamily protein
65	Glyma08g13510	13,3296	44,5362	1,74034	Plantacyanin
66	Glyma15g14040	8,77399	29,1117	1,73029	FAD-binding Berberine family protein
67	Glyma07g05740	6,68425	22,0361	1,72103	Basic helix-loop-helix (bHLH) DNA-binding family protein
68	Glyma19g03730	3,54588	11,5404	1,70248	HXXXXD-type acyl-transferase family protein
69	Glyma15g06770	21,0245	67,408	1,68085	Pathogenesis-related protein-1-like
70	Glyma13g21310	5,01634	16,051	1,67796	Sulfite exporter TauE/SafE family protein
71	Glyma13g12090	57,448	177,801	1,62993	-
72	Glyma13g04580	61,3019	186,113	1,60217	-
73	Glyma05g22770	0,411849	1,24704	1,59832	ACT domain repeat 1
74	Glyma11g35240	1,53287	4,63782	1,59721	SGNH hydrolase-type esterase superfamily protein
75	Glyma15g11910	8,78986	25,7564	1,55102	Fibrillin
76	Glyma13g40180	12,2743	35,4686	1,5309	Galactose oxidase/kelch repeat superfamily protein
77	Glyma20g32870	0,597521	1,65914	1,47337	Pleiotropic drug resistance 11
78	Glyma06g23920	0,755837	1,86928	1,30634	Stabilizer of iron transporter SufD / Polynucleotidyl transferase
79	Glyma09g09120	3,45395	8,42008	1,28559	-
80	Glyma02g15390	7,25533	17,6047	1,27884	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
81	Glyma03g27180	2,8866	6,79545	1,2352	Squamosa promoter binding protein-like 8
82	Glyma09g02900	6,64923	14,1835	1,09295	6-phosphogluconate dehydrogenase family protein
83	Glyma16g00291	1,05324	2,15265	1,03127	Protein of unknown function (DUF707)
84	Glyma02g38210	6,39086	11,727	0,875757	Phosphoglycerate mutase family protein

Tabela 7 - Os 128 transcritos diferencialmente expressos somente na variedade IAC-100 infestada, em ordem decrescente – alinhador *STAR*

(continuação)

Nº	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
85	Glyma13g00350	4,58658	8,31818	0,858848	Eukaryotic protein of unknown function (DUF872)
86	Glyma15g01830	2,2595	4,01944	0,83099	Mitochondrial substrate carrier family protein
87	Glyma04g43540	27,2789	47,588	0,802811	Methylenetetrahydrofolate reductase 2
88	Glyma19g39385	4,61625	8,04169	0,800778	Kinase interacting (KIP1-like) family protein
89	Glyma08g12630	2,50097	4,3405	0,795372	Heat shock transcription factor A4A
90	Glyma07g31910	3,2847	5,69123	0,792979	Mitochondrial substrate carrier family protein
91	Glyma04g13160	2,60676	4,41942	0,761598	-
92	Glyma19g37260	1,99477	3,28256	0,718597	Eukaryotic aspartyl protease family protein
93	Glyma18g46260	1,56976	2,43937	0,635967	EXS (ERD1/XPR1/SYG1) family protein
94	Glyma08g01320	9,81821	14,8049	0,592541	Ubiquitin family protein
95	Glyma01g37870	7,1801	10,0043	0,478543	4-phosphopantetheine adenyltransferase
96	Glyma09g08070	2,22253	3,08291	0,472085	Acyl-CoA N-acyltransferases (NAT) superfamily protein
97	Glyma02g07620	5,08384	6,43355	0,339698	Cysteine proteinases superfamily protein
98	Glyma06g04370	10,3829	12,055	0,215427	Remorin family protein
99	Glyma13g16790	4,01798	3,05629	-0,394688	Phospholipid:diacylglycerol acyltransferase
100	Glyma15g01400	8,00871	4,91975	-0,702983	-
101	Glyma05g22470	2,93886	1,74649	-0,750801	Arabinanase/levansucrase/invertase
102	Glyma08g09601	2,60004	1,50218	-0,791473	Tetratricopeptide repeat (TPR)-like superfamily protein
103	Glyma06g46710	4,36171	2,37673	-0,875915	zinc ion binding
104	Glyma18g51000	4,29683	2,17121	-0,984773	F-box family protein
105	Glyma04g12450	11,0683	5,52716	-1,00183	Sec14p-like phosphatidylinositol transfer family protein
106	Glyma14g14970	4,85295	2,28714	-1,08532	Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein
107	Glyma13g26540	4,90211	2,29103	-1,09741	Peptidase S24/S26A/S26B/S26C family protein
108	Glyma19g39480	14,6737	5,91413	-1,31099	Glutaredoxin family protein
109	Glyma06g08730	3,90208	1,50906	-1,3706	Alpha/beta-Hydrolases superfamily protein
110	Glyma15g02840	2,3551	0,908679	-1,37394	Indeterminate(ID)-domain 11
111	Glyma06g06581	2,90126	1,10633	-1,39089	-
112	Glyma13g24270	3,48517	1,30757	-1,41434	Cellulose synthase like G2

Tabela 7 - Os 128 transcritos diferencialmente expressos somente na variedade IAC-100 infestada, em ordem decrescente – alinhador *STAR*

(conclusão)					
N°	GENE ID	CD215Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
113	Glyma08g23050	10,634	3,83937	-1,46974	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
114	Glyma09g08221	2,71229	0,970083	-1,48333	Protein of unknown function (DUF300)
115	Glyma20g23120	2,56527	0,891702	-1,52448	F-box family protein
116	Glyma16g27510	8,2241	2,35485	-1,80423	-
117	Glyma02g04060	1,6133	0,404533	-1,99568	Domains rearranged methylase 1
118	Glyma13g33524	5,94169	1,41417	-2,07092	-
119	Glyma11g20860	2,23069	0,491244	-2,18298	Cytokinin oxidase/dehydrogenase 6
120	Glyma02g02441	2,9967	0,657114	-2,18916	ssDNA-binding transcriptional regulator
121	Glyma13g35970	8,19157	1,70038	-2,26828	BURP domain-containing protein
122	Glyma02g08500	3,29149	0,650447	-2,33924	Pyridoxal phosphate (PLP)-dependent transferases superfamily protein
123	Glyma10g29300	1,95472	0,370181	-2,40066	Actin-binding FH2 (formin homology 2) family protein
124	Glyma04g03470	2,67309	0,4217	-2,66422	-
125	Glyma08g11240	3,4069	0,386189	-3,14108	FKBP-type peptidyl-prolyl cis-trans isomerase family protein
126	Glyma14g01700	15,6782	1,6478	-3,25015	-
127	Glyma14g11420	36,7759	3,64491	-3,33481	17.6 kDa class II heat shock protein
128	Glyma20g12720	0,891408	0,1	-1,7977E+308	LRR and NB-ARC domains-containing disease resistance protein

A comparação entre as amostras de IAC-100 Infestada e IAC-100 Não Infestada foi realizada visando identificar todos os genes envolvidos com a resposta à infestação, mesmo que também estejam presentes na variedade CD-215. Como o resultado gerado foi de 2341 transcritos diferencialmente expressos, estes foram classificados de forma decrescente de acordo com seus valores de fold change e foi realizada análise dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão (Tabela 8). Os mesmo transcritos também foram classificados de forma crescente de acordo com seus valores de fold change para analisar os 100 genes que obtiveram a menor diferencial de expressão (Tabela 9).

Tabela 8 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR*

(continua)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma08g19470	0,1	1,36385	1,7977E+308	-
2	Glyma09g37310	0,1	3,18261	1,7977E+308	Mini zinc finger 2
3	Glyma18g28494	0,1	1,47384	1,7977E+308	TRICHOME BIREFRINGENCE-LIKE 43
4	Glyma20g35127	0,1	6,66611	1,7977E+308	-
5	Glyma19g43940	0,132812	9,9298	6,22431	Li-tolerant lipase 1
6	Glyma08g20170	0,197914	9,77745	5,62651	Homeobox protein 21
7	Glyma03g41330	0,260983	6,7971	4,70289	GDSL-like Lipase/Acylhydrolase superfamily protein
8	Glyma20g03630	0,272863	4,70732	4,10866	Cell wall / vacuolar inhibitor of fructosidase 1
9	Glyma15g01370	2,16462	33,6316	3,95763	Protein of unknown function, DUF642
10	Glyma20g03661	0,412568	6,20657	3,91109	Cell wall / vacuolar inhibitor of fructosidase 1
11	Glyma19g02810	1,2409	16,6388	3,74509	Expansin A15
12	Glyma03g15870	0,16067	1,93421	3,58958	Myb domain protein 118
13	Glyma04g00680	1,33647	16,0762	3,58842	Pollen Ole e 1 allergen and extensin family protein
14	Glyma03g36610	0,367014	3,67142	3,32243	Peroxidase superfamily protein
15	Glyma15g07600	0,195175	1,91212	3,29233	Serine carboxypeptidase-like 31
16	Glyma09g03620	0,362475	3,52881	3,28323	Pectin lyase-like superfamily protein
17	Glyma12g29670	0,279735	2,69265	3,26689	FASCICLIN-like arabinogalactan-protein 11
18	Glyma12g04960	2,04625	19,3182	3,2389	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
19	Glyma02g03250	13,6604	126,098	3,20648	Matrixin family protein
20	Glyma13g35000	0,398663	3,47917	3,1255	Squamosa promoter-binding protein-like 12
21	Glyma11g07845	1,19761	10,0956	3,0755	-
22	Glyma07g16060	2,16138	16,9438	2,97074	Laccase 17
23	Glyma02g03210	0,161015	1,25608	2,96366	Matrixin family protein
24	Glyma18g04260	0,426056	3,16491	2,89305	Barwin-like endoglucanases superfamily protein
25	Glyma12g07411	1,65763	12,2677	2,88768	FASCICLIN-like arabinogalactan-protein 11
26	Glyma18g13540	1,41157	10,3531	2,8747	GDSL-like Lipase/Acylhydrolase superfamily protein
27	Glyma07g29990	9,87463	71,8562	2,86331	-

Tabela 8 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR*
(continuação)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
28	Glyma05g24330	1,69832	11,7888	2,79524	GDSL-like Lipase/Acylhydrolase superfamily protein
29	Glyma16g04865	0,907889	6,26044	2,78568	HXXXD-type acyl-transferase family protein
30	Glyma16g06640	0,850735	5,82777	2,77616	RmlC-like cupins superfamily protein
31	Glyma08g19460	0,438972	2,93453	2,74093	Nodulin MtN21 /EamA-like transporter family protein
32	Glyma12g06480	5,06659	33,7642	2,73641	Laccase 2
33	Glyma01g35800	0,409938	2,72327	2,73186	ATP-binding cassette 14
34	Glyma10g34460	1,59195	10,093	2,66449	Cytochrome P450, family 76, subfamily C, polypeptide 4
35	Glyma11g04000	0,656823	4,15613	2,66166	HXXXD-type acyl-transferase family protein
36	Glyma05g04440	5,71216	36,0837	2,65924	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein
37	Glyma18g42580	3,54483	22,1194	2,64153	Glycerol-3-phosphate acyltransferase 6
38	Glyma13g42770	0,876058	5,46427	2,64093	Pyridoxal phosphate phosphatase-related protein
39	Glyma07g11580	0,332485	2,05478	2,62763	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
40	Glyma15g14540	0,319979	1,93444	2,59586	Pectin lyase-like superfamily protein
41	Glyma04g06230	0,628855	3,68601	2,55126	Glutamine dumper 4
42	Glyma16g26650	1,46066	8,55018	2,54933	HXXXD-type acyl-transferase family protein
43	Glyma03g09140	0,449038	2,604	2,53582	Transmembrane amino acid transporter family protein
44	Glyma18g42520	5,96116	34,435	2,53021	Laccase/Diphenol oxidase family protein
45	Glyma06g12270	0,766704	4,41619	2,52606	Amino acid permease 3
46	Glyma07g01200	2,36823	13,6394	2,5259	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
47	Glyma15g01500	0,464746	2,60162	2,4849	Gibberellin 3-oxidase 1 ²
48	Glyma04g42520	0,292686	1,63786	2,48438	Amino acid permease 3
49	Glyma02g37990	2,64072	14,7758	2,48423	Matrix metalloproteinase
50	Glyma19g05224	1,57192	8,75331	2,4773	HXXXD-type acyl-transferase family protein
51	Glyma11g14600	11,7225	65,0379	2,472	Laccase 2
52	Glyma16g08300	0,270314	1,49333	2,46582	Long-chain fatty alcohol dehydrogenase family protein
53	Glyma17g27520	3,04187	16,7195	2,45851	Integrase-type DNA-binding superfamily protein
54	Glyma07g09500	0,371623	2,02453	2,44568	Flavin-binding monooxygenase family protein

Tabela 8 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR*
(continuação)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
55	Glyma01g40460	1,07951	5,84628	2,43714	Xyloglucan endotransglucosylase/hydrolase 6
56	Glyma01g20740	1,37784	7,45798	2,43638	DNA glycosylase superfamily protein
57	Glyma13g26600	3,9099	21,0062	2,42562	Eukaryotic aspartyl protease family protein
58	Glyma19g44790	1,25206	6,68434	2,41649	Cytochrome P450, family 78, subfamily A, polypeptide 6
59	Glyma09g38620	1,19863	6,33228	2,40134	ROP-interactive CRIB motif-containing protein 4
60	Glyma16g24640	1,06886	5,60803	2,39142	Peroxidase superfamily protein
61	Glyma04g39240	0,485946	2,52781	2,37902	Plant glycogenin-like starch initiation protein 1
62	Glyma11g36620	51,1259	257,125	2,33034	S-adenosyl-L-homocysteine hydrolase
63	Glyma08g07380	15,675	78,7153	2,32818	-
64	Glyma14g09110	0,968486	4,85595	2,32595	Cytochrome P450, family 707, subfamily A, polypeptide 4
65	Glyma15g16590	0,597053	2,98392	2,32128	Purine permease 4
66	Glyma12g06130	0,369502	1,834	2,31134	Aldehyde dehydrogenase 3F1
67	Glyma08g14670	20,0274	99,3245	2,31017	Myo-inositol-1-phosphate synthase 3
68	Glyma08g14670	20,0274	99,3245	2,31017	Myo-inositol-1-phosphate synthase 1
69	Glyma08g14670	20,0274	99,3245	2,31017	Myo-inositol-1-phosphate synthase 3
70	Glyma19g07080	4,542	22,4109	2,3028	GDSL-like Lipase/Acylhydrolase superfamily protein ²
71	Glyma13g01970	1,17667	5,79305	2,29961	Phosphoglycerate mutase family protein
72	Glyma03g38080	19,4546	95,342	2,293	Galactinol synthase 1
73	Glyma19g09290	1,62986	7,90259	2,27758	Cytochrome P450, family 96, subfamily A, polypeptide 1
74	Glyma15g38091	46,6819	225,823	2,27426	-
75	Glyma15g38271	2,15221	10,3483	2,26551	-
76	Glyma09g24820	0,383887	1,79724	2,22703	MATE efflux family protein
77	Glyma13g35970	1,70038	7,91764	2,21921	BURP domain-containing protein
78	Glyma02g14410	1,99142	9,26756	2,21839	-
79	Glyma20g27040	4,01714	18,6746	2,21683	Glycosyl hydrolase family 10 protein
80	Glyma01g27900	0,785266	3,65045	2,21682	Glycerol-3-phosphate acyltransferase 6
81	Glyma06g03990	1,29043	5,96292	2,20816	-
82	Glyma18g40070	1,01467	4,66712	2,20153	Laccase 17

Tabela 8 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR* (conclusão)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100Inf (FPKM)	log2.fold_change	DESCRIÇÃO
83	Glyma15g02450	0,31039	1,4272	2,20104	Leucine-rich repeat transmembrane protein kinase protein
84	Glyma10g28020	14,5778	66,4366	2,18821	Germin-like protein 10
85	Glyma13g19220	1,08624	4,93808	2,1846	GDSL-like Lipase/Acylhydrolase superfamily protein
86	Glyma05g00730	0,677094	3,07749	2,18433	Flavin-binding monooxygenase family protein
87	Glyma07g17720	11,3465	51,2967	2,17662	Glycerol-3-phosphate acyltransferase 6
88	Glyma03g30180	8,6665	39,0475	2,17171	Peroxidase 2
89	Glyma20g02210	9,05943	40,815	2,17161	Polyketide cyclase/dehydrase and lipid transport superfamily protein
90	Glyma0022s00480	1,37351	6,15985	2,16503	DNA glycosylase superfamily protein
91	Glyma08g11560	1,42541	6,33235	2,15137	HXXXD-type acyl-transferase family protein
92	Glyma19g07000	1,30914	5,76569	2,13888	GDSL-like Lipase/Acylhydrolase superfamily protein
93	Glyma18g48530	4,26727	18,7631	2,13652	Subtilisin-like serine endopeptidase family protein
94	Glyma18g03540	1,03837	4,53942	2,12819	Concanavalin A-like lectin family protein
95	Glyma11g02400	0,867031	3,7852	2,12621	Myb domain protein 103
96	Glyma13g40960	6,98446	30,4457	2,12402	UDP-XYL synthase 5
97	Glyma06g05240	0,51041	2,22167	2,12192	1-amino-cyclopropane-1-carboxylate synthase 8
98	Glyma07g39340	1,94505	8,44765	2,11874	PA-domain containing subtilase family protein
99	Glyma18g33230	1,64194	7,0908	2,11054	-
100	Glyma04g40600	0,427502	1,80705	2,07964	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein

Tabela 9 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR*

(continua)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma02g00890	7,94674	0,1	-1,7977E+308	Integrase-type DNA-binding superfamily protein
2	Glyma02g06140	9,47876	0,1	-1,7977E+308	-
3	Glyma03g35960	4,53436	0,1	-1,7977E+308	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
4	Glyma03g35980	4,22454	0,1	-1,7977E+308	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
5	Glyma05g01780	1,97643	0,1	-1,7977E+308	-
6	Glyma05g08690	1,59169	0,1	-1,7977E+308	Myb domain protein 21
7	Glyma06g17625	1,31512	0,1	-1,7977E+308	Fatty acyl-ACP thioesterases B
8	Glyma06g45540	1,40836	0,1	-1,7977E+308	Myb domain protein 14
9	Glyma07g31061	6,69495	0,1	-1,7977E+308	-
10	Glyma08g26780	0,910977	0,1	-1,7977E+308	UDP-Glycosyltransferase superfamily protein
11	Glyma11g25660	3,50765	0,1	-1,7977E+308	Calmodulin-like 41
12	Glyma11g34000	2,29205	0,1	-1,7977E+308	Calcium-dependent lipid-binding (CaLB domain) family protein
13	Glyma14g35650	1,86471	0,1	-1,7977E+308	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein
14	Glyma15g06000	1,17946	0,1	-1,7977E+308	UDP-glucosyl transferase 85A2
15	Glyma16g04360	1,39826	0,1	-1,7977E+308	Spermidine hydroxycinnamoyl transferase
16	Glyma16g32000	5,46506	0,1	-1,7977E+308	Cytochrome P450, family 71, subfamily A, polypeptide 26
17	Glyma16g33770	2,32552	0,1	-1,7977E+308	Kunitz family trypsin and protease inhibitor protein
18	Glyma17g01760	3,25391	0,1	-1,7977E+308	-
19	Glyma18g16780	1,44806	0,1	-1,7977E+308	Disease resistance protein (TIR-NBS class), putative
20	Glyma19g34650	2,08802	0,1	-1,7977E+308	Integrase-type DNA-binding superfamily protein
21	Glyma19g34696	2,16414	0,1	-1,7977E+308	Ethylene response factor 1
22	Glyma20g34810	5,54952	0,1	-1,7977E+308	Serine protease inhibitor, potato inhibitor I-type family protein
23	Glyma17g15710	93,1492	0,249068	-8,54686	Expansin-like B1
24	Glyma09g04191	25,3913	0,0916999	-8,1132	Glycosyl hydrolase superfamily protein
25	Glyma15g05820	58,768	0,306842	-7,58139	Peroxidase superfamily protein
26	Glyma13g16950	25,8021	0,146232	-7,46309	Oxophytodienoate-reductase 3
27	Glyma04g35380	47,1142	0,305325	-7,26967	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein

Tabela 9 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR*

(continuação)					
Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
28	Glyma20g36810	41,3616	0,27942	-7,20971	Pathogenesis-related 4
29	Glyma16g24831	21,6171	0,167001	-7,01617	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
30	Glyma16g02320	22,5219	0,192526	-6,87013	Basic helix-loop-helix (bHLH) DNA-binding family protein
31	Glyma08g22530	56,7207	0,48607	-6,86657	-
32	Glyma17g11490	13,9448	0,12522	-6,79911	Serine-rich protein-related
33	Glyma15g02850	29,7954	0,276483	-6,75175	Glycosyl hydrolases family 32 protein
34	Glyma17g34530	14,3572	0,134726	-6,7356	Cytochrome P450, family 711, subfamily A, polypeptide 1
35	Glyma06g00880	17,3699	0,175205	-6,63141	SAUR-like auxin-responsive protein family
36	Glyma02g07940	26,4902	0,269158	-6,62086	Glutamate dehydrogenase 1
37	Glyma02g07940	26,4902	0,269158	-6,62086	Glutamate dehydrogenase 3
38	Glyma02g06150	4,72083	0,048529	-6,60405	-
39	Glyma07g16810	104,404	1,1106	-6,5547	Glutathione S-transferase TAU 8
40	Glyma14g05355	185,079	2,11257	-6,453	Ethylene-forming enzyme
41	Glyma16g04740	68,5717	0,819281	-6,38711	NAC domain containing protein 47
42	Glyma08g14900	15,3069	0,183383	-6,38318	Ferulic acid 5-hydroxylase 1
43	Glyma14g22700	26,0218	0,321868	-6,33711	Regulator of chromosome condensation (RCC1) family protein
44	Glyma05g05880	24,1382	0,306	-6,30165	Expansin-like B1
45	Glyma06g19400	5,58908	0,0709195	-6,30028	Carbonic anhydrase 2
46	Glyma07g37240	930,485	11,8148	-6,29932	MLP-like protein 423
47	Glyma10g33810	14,6207	0,186064	-6,29607	Integrase-type DNA-binding superfamily protein
48	Glyma08g19180	4,65226	0,059242	-6,29517	Peroxidase superfamily protein
49	Glyma14g08000	28,2868	0,362974	-6,28412	Phloem protein 2-B10
50	Glyma08g04670	34,3463	0,445383	-6,26896	Myb domain protein 62
51	Glyma06g19490	31,9566	0,421567	-6,24421	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
52	Glyma05g25690	11,7477	0,15541	-6,24015	-
53	Glyma15g05810	3,51241	0,0465734	-6,23681	Peroxidase superfamily protein
54	Glyma15g15590	80,228	1,0982	-6,19089	MLP-like protein 423

Tabela 9 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR*

(continuação)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
55	Glyma15g18360	186,479	2,73896	-6,08925	Xyloglucan endotransglycosylase 6
56	Glyma10g44170	21,033	0,310295	-6,08287	Homogentisate phytyltransferase 1
57	Glyma16g27950	14,9952	0,224319	-6,06281	Integrase-type DNA-binding superfamily protein
58	Glyma10g32830	53,7039	0,805952	-6,05819	Serine protease inhibitor, potato inhibitor I-type family protein
59	Glyma07g37265	959,044	14,458	-6,05165	-
60	Glyma08g05500	12,5003	0,197246	-5,98581	Ethylene-forming enzyme
61	Glyma14g05350	57,3984	0,920428	-5,96256	Ethylene-forming enzyme
62	Glyma15g01990	8,35666	0,13434	-5,95897	PLAC8 family protein
63	Glyma09g29340	10,083	0,163554	-5,94602	Kunitz family trypsin and protease inhibitor protein
64	Glyma12g32381	8,86235	0,144658	-5,93697	Terpene synthase 02
65	Glyma18g47280	34,5054	0,565937	-5,93004	ACC synthase 1
66	Glyma07g16830	52,7771	0,869553	-5,92349	Glutathione S-transferase TAU 8
67	Glyma03g26530	13,8143	0,236083	-5,87072	Ethylene-responsive element binding factor 13
68	Glyma17g03365	229,528	3,95828	-5,85765	MLP-like protein 423
69	Glyma10g40320	26,3838	0,459282	-5,84413	-
70	Glyma12g09940	5,58788	0,0990072	-5,81862	FAD/NAD(P)-binding oxidoreductase family protein
71	Glyma06g26610	10,601	0,189976	-5,80224	Blue-copper-binding protein
72	Glyma19g27130	12,7272	0,234079	-5,76478	NAD(P)-linked oxidoreductase superfamily protein
73	Glyma09g01900	49,3451	0,90979	-5,76123	-
74	Glyma03g35990	9,1887	0,172135	-5,73825	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family
75	Glyma19g32700	67,2806	1,26282	-5,73547	Polygalacturonase inhibiting protein 1
76	Glyma13g18410	5,37767	0,102747	-5,70981	Ethylene response factor 1
77	Glyma10g32380	10,2471	0,196975	-5,70105	Eukaryotic aspartyl protease family protein
78	Glyma02g39840	44,3387	0,852411	-5,70087	-
79	Glyma14g06710	35,6414	0,691484	-5,68771	HXXXD-type acyl-transferase family protein
80	Glyma18g45920	28,0338	0,545776	-5,68272	BON association protein 2
81	Glyma07g16840	12,9642	0,254892	-5,66851	Glutathione S-transferase TAU 8

Tabela 9 - Ranking dos 100 primeiros genes que tiveram a menor mudança em sua expressão na comparação entre amostras infestadas e não infestadas da variedade IAC-100 – alinhador *STAR* (conclusão)

Nº	GENE ID	IAC100Inf (FPKM)	IAC100NInf (FPKM)	log2.fold_change	DESCRIÇÃO
82	Glyma08g26840	9,37259	0,186416	-5,65185	UDP-Glycosyltransferase superfamily protein
83	Glyma09g07070	59,6278	1,19056	-5,64627	Xyloglucan endotransglycosylase 6
84	Glyma04g00870	7,49879	0,15138	-5,63041	SAUR-like auxin-responsive protein family
85	Glyma02g06460	10,7788	0,217912	-5,62831	Major facilitator superfamily protein
86	Glyma11g06690	5,49978	0,112043	-5,61724	Cytochrome P450, family 71, subfamily B, polypeptide 34
87	Glyma08g44230	53,2287	1,09805	-5,59919	Nudix hydrolase homolog 2
88	Glyma02g00870	6,23464	0,128739	-5,59779	Ethylene response factor 1
89	Glyma03g38490	40,0137	0,837019	-5,57909	Glycosyl hydrolase superfamily protein
90	Glyma12g22880	59,6952	1,24957	-5,57811	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
91	Glyma11g00580	26,2443	0,579501	-5,50105	Pyridoxal-5'-phosphate-dependent enzyme family protein
92	Glyma01g40320	4,86837	0,107682	-5,49859	VQ motif-containing protein
93	Glyma15g06780	125,311	2,7829	-5,49278	Basic pathogenesis-related protein 1
94	Glyma12g06310	33,7962	0,759331	-5,47599	Senescence-associated gene 13
95	Glyma11g04970	31,4185	0,714526	-5,45848	VQ motif-containing protein
96	Glyma10g33060	16,4497	0,374978	-5,45511	Ethylene response factor 1
97	Glyma07g02630	11,031	0,252424	-5,44958	WRKY DNA-binding protein 40
98	Glyma18g51680	7,91461	0,185558	-5,41458	Integrase-type DNA-binding superfamily protein
99	Glyma19g43360	89,2248	2,12481	-5,39204	NAD(P)-linked oxidoreductase superfamily protein
100	Glyma02g09670	17,8234	0,425831	-5,38735	Basic helix-loop-helix (bHLH) DNA-binding family protein

A variedade CD-215 possui mecanismo de resposta diferente do apresentado pela variedade IAC-100. Logo, para a comparação dos transcritos expressos apresentados pela variedade CD-215, dita como não resistente, com os da resistente, foi feita a comparação entre as amostras CD-215 Infestada e CD-215 Não Infestada, obtendo-se 561 transcritos diferencialmente expressos. Para uma análise dos genes mais expressos nessa condição, foram classificados também os 100 genes que tiveram as mudanças mais drásticas em sua expressão (Tabela 10) e os 100 genes que tiveram a menor mudança (Tabela 11).

Tabela 10 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR*
(continua)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma19g30460	2,10504	7,88326	1,90495	Alpha-L-arabinofuranosidase 1
2	Glyma18g44680	4,32255	16,0943	1,8966	Lipid transfer protein 3
3	Glyma02g37610	5,31539	18,1449	1,77132	Eukaryotic aspartyl protease family protein
4	Glyma03g38080	6,62719	21,2155	1,67865	Galactinol synthase 1
5	Glyma08g14670	5,43516	14,2493	1,3905	Myo-inositol-1-phosphate synthase 1
6	Glyma08g14670	5,43516	14,2493	1,3905	Myo-inositol-1-phosphate synthase 3
7	Glyma08g17010	7,11388	14,9044	1,06703	ATP-citrate lyase A-1
8	Glyma19g04220	7,0799	13,4906	0,930157	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family
9	Glyma19g40300	2,05036	3,86418	0,914282	Beta-xylosidase 2
10	Glyma18g38620	3,63018	6,53605	0,848378	-
11	Glyma17g01640	3,12899	5,01851	0,681563	Ankyrin repeat family protein
12	Glyma11g12050	1,27176	1,64676	0,372803	Kinesin motor family protein
13	Glyma07g03680	43,2509	33,2164	-0,380835	Rubredoxin-like superfamily protein
14	Glyma01g42331	6,6664	4,13589	-0,68871	YTH family protein
15	Glyma18g49900	9,15628	5,66615	-0,692393	RING/U-box superfamily protein
16	Glyma02g47760	11,0404	6,35315	-0,797253	-
17	Glyma01g01860	3,15979	1,81381	-0,800805	Transcription factor jumonji (jmi) family protein / zinc finger (C5HC2 type) family protein
18	Glyma02g41930	5,03522	2,80831	-0,842355	Mitochondrial substrate carrier family protein
19	Glyma11g25860	18,0255	10,0317	-0,845475	-
20	Glyma17g13751	9,97449	5,30346	-0,91131	Protein kinase superfamily protein
21	Glyma06g04280	6,67239	3,45961	-0,947596	Photosystem I light harvesting complex gene 5
22	Glyma07g03200	10,3385	5,35352	-0,949471	Leucine-rich repeat family protein
23	Glyma04g01390	26,6634	13,7733	-0,952983	Adenine nucleotide alpha hydrolases-like superfamily protein
24	Glyma14g00380	10,3041	5,15192	-1,00003	Protein kinase superfamily protein
25	Glyma11g36010	7,23872	3,39137	-1,09387	bZIP transcription factor family protein
26	Glyma09g08290	13,4972	6,2403	-1,11297	Jasmonate-zim-domain protein 1
27	Glyma05g34870	19,5655	9,00368	-1,11972	PLANT CADMIUM RESISTANCE 2

Tabela 10 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR*
(continuação)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
28	Glyma07g37270	100,841	44,5486	-1,17863	MLP-like protein 423
29	Glyma06g13550	10,1671	4,38533	-1,21315	ABL interactor-like protein 2
30	Glyma12g32010	30,9807	13,3222	-1,21754	MATE efflux family protein
31	Glyma17g04350	4,52307	1,90377	-1,24845	Pleiotropic drug resistance 9
32	Glyma06g13420	1,85062	0,77451	-1,25665	Calcium-binding EF-hand family protein
33	Glyma04g41140	7,1736	2,95748	-1,27833	Oligopeptide transporter 1
34	Glyma20g28631	4,17775	1,72191	-1,27871	O-acetylserine (thiol) lyase (OAS-TL) isoform A1
35	Glyma04g09110	8,66975	3,52041	-1,30025	NADP-malic enzyme 3
36	Glyma01g35480	3,4644	1,35791	-1,35122	Pectinacetylsterase family protein
37	Glyma12g03680	2,06364	0,798902	-1,3691	Protein kinase protein with adenine nucleotide alpha hydrolases-like domain
38	Glyma18g10930	15,0479	5,78675	-1,37873	Protein of unknown function, DUF584
39	Glyma03g00830	11,0862	4,20618	-1,39818	MATE efflux family protein
40	Glyma16g29650	22,7707	8,58116	-1,40793	Heavy metal transport/detoxification superfamily protein
41	Glyma16g29650	22,7707	8,58116	-1,40793	Heavy metal transport/detoxification superfamily protein
42	Glyma02g37710	16,5181	6,06682	-1,44503	Beta carbonic anhydrase 5
43	Glyma03g25651	9,1211	3,32131	-1,45746	-
44	Glyma18g43210	9,83067	3,56947	-1,46158	Fatty acid desaturase 8
45	Glyma15g01470	5,5629	2,00032	-1,4756	Pleiotropic drug resistance 12
46	Glyma13g04670	11,2445	4,02554	-1,48197	Cytochrome P450, family 82, subfamily C, polypeptide 4
47	Glyma17g12710	25,2148	8,91929	-1,49927	-
48	Glyma05g32756	11,6446	4,11596	-1,50037	Alpha/beta-Hydrolases superfamily protein
49	Glyma17g03110	14,318	5,05678	-1,50154	2-oxoacid dehydrogenases acyltransferase family protein
50	Glyma15g40190	51,7686	18,161	-1,51123	Glutathione S-transferase TAU 19
51	Glyma19g42000	8,5662	3,00004	-1,51367	Alpha/beta-Hydrolases superfamily protein
52	Glyma14g00570	19,717	6,87904	-1,51916	VQ motif-containing protein
53	Glyma10g42820	19,0052	6,53521	-1,54009	Arabidopsis phospholipase-like protein (PEARLI 4) family
54	Glyma12g09830	20,4384	7,02241	-1,54125	Polynucleotidyl transferase, ribonuclease H-like superfamily protein

Tabela 10 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR*
(continuação)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
55	Glyma13g32430	28,2584	9,7056	-1,54179	Ribonuclease E inhibitor RraA/Dimethylmenaquinone methyltransferase
56	Glyma07g02420	13,3696	4,56917	-1,54895	-
57	Glyma20g21030	11,2101	3,82596	-1,55091	Ammonium transporter 1
58	Glyma13g43870	9,48868	3,23426	-1,55277	Pleiotropic drug resistance 12
59	Glyma01g03150	11,2891	3,8185	-1,56385	Octicosapeptide/Phox/Bem1p family protein
60	Glyma16g25550	5,6403	1,89169	-1,5761	C2H2-like zinc finger protein
61	Glyma01g29670	9,01988	2,97423	-1,60059	ALC-interacting protein 1
62	Glyma13g17180	15,2811	5,03188	-1,60258	Jasmonate-zim-domain protein 1
63	Glyma14g07360	13,6976	4,50886	-1,60309	Expansin A4
64	Glyma17g34420	37,9723	12,4877	-1,60444	-
65	Glyma02g03280	15,7848	5,19012	-1,6047	-
66	Glyma08g26830	23,8517	7,83906	-1,60534	UDP-Glycosyltransferase superfamily protein
67	Glyma06g19600	20,1928	6,63649	-1,60535	-
68	Glyma17g34740	10,7619	3,49195	-1,62383	Protein of unknown function (DUF778)
69	Glyma05g15870	5,78696	1,87753	-1,62397	Calcium-binding EF-hand family protein
70	Glyma05g28310	13,8424	4,48058	-1,62734	Xyloglucan endotransglucosylase/hydrolase 16
71	Glyma17g35430	8,04706	2,60172	-1,629	Salt tolerance zinc finger
72	Glyma02g15370	65,9929	21,3232	-1,62989	2-oxoglutarate (2OG) and Fe(II)- dependent oxygenase superfamily protein
73	Glyma18g53860	17,1777	5,5236	-1,63686	RmlC-like cupins superfamily protein
74	Glyma11g35800	93,1703	29,9007	-1,63969	Senescence associated gene 20
75	Glyma02g48100	5,26295	1,67704	-1,64996	Protein kinase superfamily protein
76	Glyma03g11580	60,8939	19,2623	-1,66052	NAD(P)-linked oxidoreductase superfamily protein
77	Glyma06g04030	18,8343	5,93951	-1,66495	Protein of unknown function (DUF1645)
78	Glyma11g12980	16,5381	5,20494	-1,66784	Hemoglobin 1
79	Glyma10g00980	18,4011	5,78749	-1,66878	Ethylene response factor 1
80	Glyma01g41450	23,4804	7,37575	-1,67059	Alpha/beta-Hydrolases superfamily protein
81	Glyma01g29930	43,1356	13,4831	-1,67773	2-oxoglutarate (2OG) and Fe(II)- dependent oxygenase superfamily protein

Tabela 10 - Ranking dos 100 primeiros genes que tiveram a mudança mais drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR* (conclusão)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
82	Glyma08g25520	34,2633	10,6957	-1,67964	Kunitz trypsin inhibitor 1
83	Glyma17g16020	16,0022	4,99308	-1,68027	Chloroplast beta-amylase
84	Glyma17g37580	9,24391	2,87989	-1,68249	Transcriptional factor B3 family protein / auxin-responsive factor AUX/IAA-related
85	Glyma01g26840	14,367	4,46085	-1,68737	PAR1 protein
86	Glyma06g35585	20,8838	6,46227	-1,69227	Tyrosine transaminase family protein
87	Glyma15g10070	13,975	4,32	-1,69375	Gibberellin 2-oxidase
88	Glyma16g33840	18,3937	5,66612	-1,69878	YELLOW STRIPE like 7
89	Glyma11g37450	11,625	3,56952	-1,70343	Heat shock protein 21
90	Glyma02g41070	11,3769	3,48561	-1,70663	Eukaryotic aspartyl protease family protein
91	Glyma13g39710	19,1595	5,85506	-1,7103	Xyloglucan endotransglucosylase/hydrolase 32
92	Glyma19g37108	45,5045	13,903	-1,71061	UDP-glucosyl transferase 73C2
93	Glyma02g11610	11,6663	3,55431	-1,71471	UDP-glycosyltransferase 73B4
94	Glyma11g20020	23,2477	7,05973	-1,7194	AMP-dependent synthetase and ligase family protein
95	Glyma09g31810	21,0776	6,37647	-1,72488	Cytochrome P450, family 71, subfamily A, polypeptide 22
96	Glyma20g35240	24,802	7,48823	-1,72776	Eukaryotic aspartyl protease family protein
97	Glyma18g16490	8,33784	2,4922	-1,74226	Nitrate transporter 1.7
98	Glyma09g24170	12,4301	3,69373	-1,75069	Heavy metal transport/detoxification superfamily protein
99	Glyma09g24170	12,4301	3,69373	-1,75069	Heavy metal transport/detoxification superfamily protein
100	Glyma01g36620	37,9406	11,2169	-1,75806	Rhodanese/Cell cycle control phosphatase superfamily protein

Tabela 11 - Ranking dos 100 primeiros genes que tiveram a mudança menos drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR*
(continua)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
1	Glyma20g36810	7,38518	0,1	-1,7977E+308	Pathogenesis-related 4
2	Glyma15g15590	106,573	0,705063	-7,23988	MLP-like protein 423
3	Glyma15g05820	17,1969	0,183179	-6,55275	Peroxidase superfamily protein
4	Glyma13g16950	7,50326	0,101224	-6,2119	Oxophytodienoate-reductase 3
5	Glyma14g05650	14,0791	0,20632	-6,09253	Glycine-rich protein 3 short isoform
6	Glyma16g02320	6,33732	0,116075	-5,77075	Basic helix-loop-helix (bHLH) DNA-binding family protein
7	Glyma08g22530	17,8166	0,341413	-5,70556	-
8	Glyma12g06310	5,72939	0,118429	-5,59628	Senescence-associated gene 13
9	Glyma17g11490	5,57012	0,129803	-5,42331	Serine-rich protein-related
10	Glyma11g00580	4,49006	0,111271	-5,33458	Pyridoxal-5'-phosphate-dependent enzyme family protein
11	Glyma14g05355	57,8895	1,44795	-5,32122	Ethylene-forming enzyme
12	Glyma01g38870	1,63493	0,0417025	-5,29295	Cytochrome P450, family 82, subfamily C, polypeptide 4
13	Glyma04g35380	20,4781	0,54542	-5,23057	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
14	Glyma11g27720	12,326	0,340538	-5,17775	Glutamine-dependent asparagine synthase 1
15	Glyma15g18360	28,5979	0,793085	-5,17229	Xyloglucan endotransglycosylase 6
16	Glyma09g29340	4,23957	0,117586	-5,17212	Kunitz family trypsin and protease inhibitor protein
17	Glyma13g35320	81,5417	2,26191	-5,17192	Low-molecular-weight cysteine-rich 68
18	Glyma13g35320	81,5417	2,26191	-5,17192	Low-molecular-weight cysteine-rich 66
19	Glyma13g35320	81,5417	2,26191	-5,17192	Low-molecular-weight cysteine-rich 68
20	Glyma07g16810	66,1886	1,86354	-5,15046	Glutathione S-transferase TAU 8
21	Glyma03g04920	113,292	3,51299	-5,01121	Lipid transfer protein 1
22	Glyma06g19490	14,3671	0,448879	-5,00029	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein
23	Glyma14g22700	6,97989	0,219531	-4,99071	Regulator of chromosome condensation (RCC1) family protein
24	Glyma06g13110	30,8492	0,982122	-4,97319	Nodulin MtN3 family protein
25	Glyma14g05630	6,76664	0,216926	-4,96316	Calcium-binding EF-hand family protein
26	Glyma07g35630	11,0831	0,40656	-4,76874	NAC-like, activated by AP3/PI
27	Glyma07g16830	31,8727	1,21201	-4,71684	Glutathione S-transferase TAU 8
28	Glyma15g06780	63,9426	2,44154	-4,71092	Basic pathogenesis-related protein 1

Tabela 11 - Ranking dos 100 primeiros genes que tiveram a mudança menos drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR*
(continuação)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
29	Glyma17g37660	8,11669	0,31629	-4,68157	Putative lysine decarboxylase family protein
30	Glyma17g13730	9,17911	0,36288	-4,66079	Malate synthase
31	Glyma19g26400	7,35294	0,290783	-4,66031	WRKY DNA-binding protein 75
32	Glyma20g30835	4,3683	0,175602	-4,63669	Integrase-type DNA-binding superfamily protein
33	Glyma11g16120	3,28297	0,13614	-4,59184	Protein of unknown function (DUF793)
34	Glyma04g00870	5,86562	0,247601	-4,56619	SAUR-like auxin-responsive protein family
35	Glyma12g22880	14,4643	0,633475	-4,51306	NAC (No Apical Meristem) domain transcriptional regulator superfamily protein
36	Glyma02g43560	229,935	10,1061	-4,50793	Ethylene-forming enzyme
37	Glyma02g08835	4,62793	0,203444	-4,50766	Integrase-type DNA-binding superfamily protein
38	Glyma08g05500	3,21103	0,142836	-4,4906	Ethylene-forming enzyme
39	Glyma16g04740	18,2223	0,81413	-4,4843	NAC domain containing protein 47
40	Glyma14g37980	24,1092	1,07836	-4,48267	-
41	Glyma03g38490	3,0643	0,13737	-4,47942	Glycosyl hydrolase superfamily protein
42	Glyma08g26840	3,92268	0,183235	-4,42008	UDP-Glycosyltransferase superfamily protein
43	Glyma10g32830	24,7103	1,15838	-4,41493	Serine protease inhibitor, potato inhibitor I-type family protein
44	Glyma20g16920	15,7704	0,740129	-4,4133	Integrase-type DNA-binding superfamily protein
45	Glyma07g08311	6,35919	0,301897	-4,39671	BON association protein 2
46	Glyma13g20170	4,38082	0,214715	-4,35071	Peroxidase superfamily protein
47	Glyma05g25690	4,60629	0,22878	-4,33157	-
48	Glyma12g16750	3,17118	0,157522	-4,3314	GRAS family transcription factor
49	Glyma07g02590	3,18988	0,160568	-4,31225	Acyl-CoA N-acyltransferases (NAT) superfamily protein
50	Glyma04g41680	18,0031	0,919634	-4,29104	Nodulin MtN3 family protein
51	Glyma01g38590	4,88641	0,250374	-4,28662	Cytochrome P450, family 71, subfamily B, polypeptide 34
52	Glyma10g34760	7,14793	0,369638	-4,27334	AP2/B3 transcription factor family protein
53	Glyma17g03365	122,286	6,33867	-4,26994	MLP-like protein 423
54	Glyma06g06290	3,72144	0,193188	-4,26778	Glutamine dumper 3
55	Glyma01g38390	10,7373	0,558284	-4,26548	Homeobox protein 40

Tabela 11 - Ranking dos 100 primeiros genes que tiveram a mudança menos drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR*
(continuação)

N°	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
56	Glyma17g34530	3,30849	0,17217	-4,26427	Cytochrome P450, family 711, subfamily A, polypeptide 1
57	Glyma18g44320	13,0362	0,688652	-4,2426	Peroxidase superfamily protein
58	Glyma14g05350	17,4392	0,921547	-4,24213	Ethylene-forming enzyme
59	Glyma08g04670	10,8921	0,580616	-4,22955	Myb domain protein 62
60	Glyma15g09560	9,93606	0,531724	-4,22392	GDSL-like Lipase/Acylhydrolase superfamily protein
61	Glyma12g13290	9,76526	0,523176	-4,22229	Protein phosphatase 2C family protein
62	Glyma10g44170	19,2339	1,03056	-4,22215	Homogentisate phytyltransferase 1
63	Glyma02g07940	3,10894	0,167466	-4,21448	Glutamate dehydrogenase 1
64	Glyma02g07940	3,10894	0,167466	-4,21448	Glutamate dehydrogenase 3
65	Glyma13g42530	2,9188	0,165032	-4,14455	Glycosyl hydrolases family 32 protein
66	Glyma11g33040	4,76635	0,272827	-4,12683	Oxidative stress 3
67	Glyma15g06010	199,477	11,499	-4,11664	EXORDIUM like 2
68	Glyma14g08000	7,21515	0,416633	-4,11418	Phloem protein 2-B10
69	Glyma01g40360	6,31338	0,364597	-4,11404	Phytosulfokine 4 precursor
70	Glyma14g09990	57,3034	3,36194	-4,09125	Phytosulfokine 4 precursor
71	Glyma05g36100	52,2139	3,09198	-4,07783	Myo-inositol oxygenase 4
72	Glyma05g21680	10,3708	0,633244	-4,03361	Auxin-responsive GH3 family protein
73	Glyma08g14900	4,17864	0,256086	-4,02834	Ferulic acid 5-hydroxylase 1
74	Glyma18g02400	21,107	1,29675	-4,02475	-
75	Glyma19g43360	8,68013	0,538031	-4,01196	NAD(P)-linked oxidoreductase superfamily protein
76	Glyma07g18670	2,22578	0,139663	-3,99429	Ammonium transporter 2
77	Glyma19g32700	25,2126	1,5978	-3,97998	Polygalacturonase inhibiting protein 1
78	Glyma10g33810	3,72263	0,239936	-3,9556	Integrase-type DNA-binding superfamily protein
79	Glyma16g27880	32,1885	2,09634	-3,9406	Peroxidase superfamily protein
80	Glyma18g53170	2,16459	0,141917	-3,93098	Class I glutamine amidotransferase-like superfamily protein
81	Glyma02g39840	24,4092	1,61511	-3,91772	-
82	Glyma14g16630	5,00206	0,33264	-3,91049	O-Glycosyl hydrolases family 17 protein

Tabela 11 - Ranking dos 100 primeiros genes que tiveram a mudança menos drástica em sua expressão na comparação entre amostras infestadas e não infestadas da variedade CD-215 – alinhador *STAR* (conclusão)

Nº	GENE ID	CD215Inf (FPKM)	CD215NInf (FPKM)	log2.fold_change	DESCRIÇÃO
83	Glyma08g44230	14,8337	0,991046	-3,90378	Nudix hydrolase homolog 2
84	Glyma07g36130	2,46577	0,165211	-3,89965	-
85	Glyma18g45920	13,0173	0,881752	-3,88391	BON association protein 2
86	Glyma15g02850	8,76034	0,603842	-3,85874	Glycosyl hydrolases family 32 protein
87	Glyma11g37930	11,8732	0,833902	-3,83169	-
88	Glyma09g04510	87,0801	6,15619	-3,82223	MLP-like protein 423
89	Glyma17g18230	4,04198	0,28764	-3,81273	-
90	Glyma20g20355	5,72267	0,407597	-3,81147	-
91	Glyma11g04970	8,51342	0,611411	-3,79952	VQ motif-containing protein
92	Glyma10g33060	6,32067	0,464154	-3,7674	Ethylene response factor 1
93	Glyma16g05190	9,74275	0,734548	-3,7294	Integrase-type DNA-binding superfamily protein
94	Glyma08g23430	4,27303	0,322688	-3,72705	Acyl-CoA N-acyltransferases (NAT) superfamily protein
95	Glyma15g13500	47,4573	3,61714	-3,71371	Peroxidase superfamily protein
96	Glyma14g05360	15,8147	1,22019	-3,69609	Ethylene-forming enzyme
97	Glyma10g32380	1,85254	0,142957	-3,69585	Eukaryotic aspartyl protease family protein
98	Glyma07g37265	368,7	29,0666	-3,66501	-
99	Glyma20g38000	1,59029	0,125932	-3,65858	Respiratory burst oxidase homolog B
100	Glyma16g27950	4,49492	0,36241	-3,6326	Integrase-type DNA-binding superfamily protein

Para uma melhor visualização dos mecanismos de respostas foram elaborados pelo programa iPath dois mapas metabólicos a partir dos transcritos diferencialmente expressos resultantes da comparação entre as amostras de IAC-100 Infestada x IAC-100 Não Infestada (Figura 6) e da comparação entre as amostras de CD-215 Infestada x CD-215 Não Infestada (Figura 7).

As linhas azuis indicam as vias metabólicas que são comandadas pelos transcritos identificados. Quanto mais espessa a linha azul representada no mapa metabólico, significa que um grande número de transcritos recebeu o mesmo identificador do *UEKO*. Na Figura 6, pode-se ver com detalhe algumas das vias mais ativadas nas quais os transcritos estão

relacionados com o metabolismo de lipídios como ácido linoleico, o qual é diretamente relacionado com a via do ácido jasmônico, produção de isoflavonóides, carotenóides e terpenos. Todas essas vias supracitadas estão descritas na literatura como vias de resposta a defesa contra patógenos por parte das plantas e serão abordadas com mais detalhes no tópico de discussão.

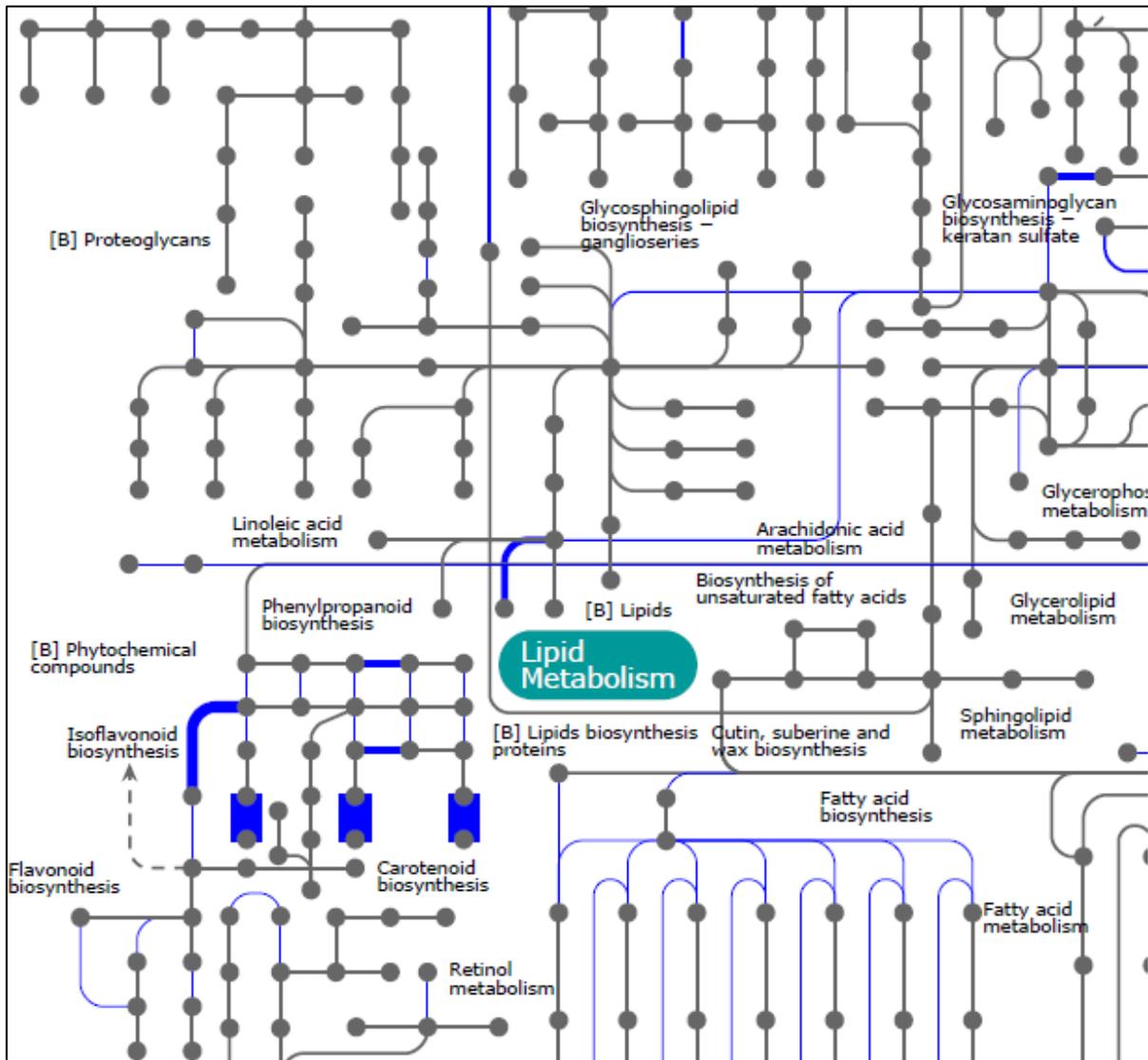


Figura 6 - Mapa das vias metabólicas feito a partir dos transcritos diferencialmente expressos entre os tratamentos de Infestação e Não Infestação na variedade IAC-100

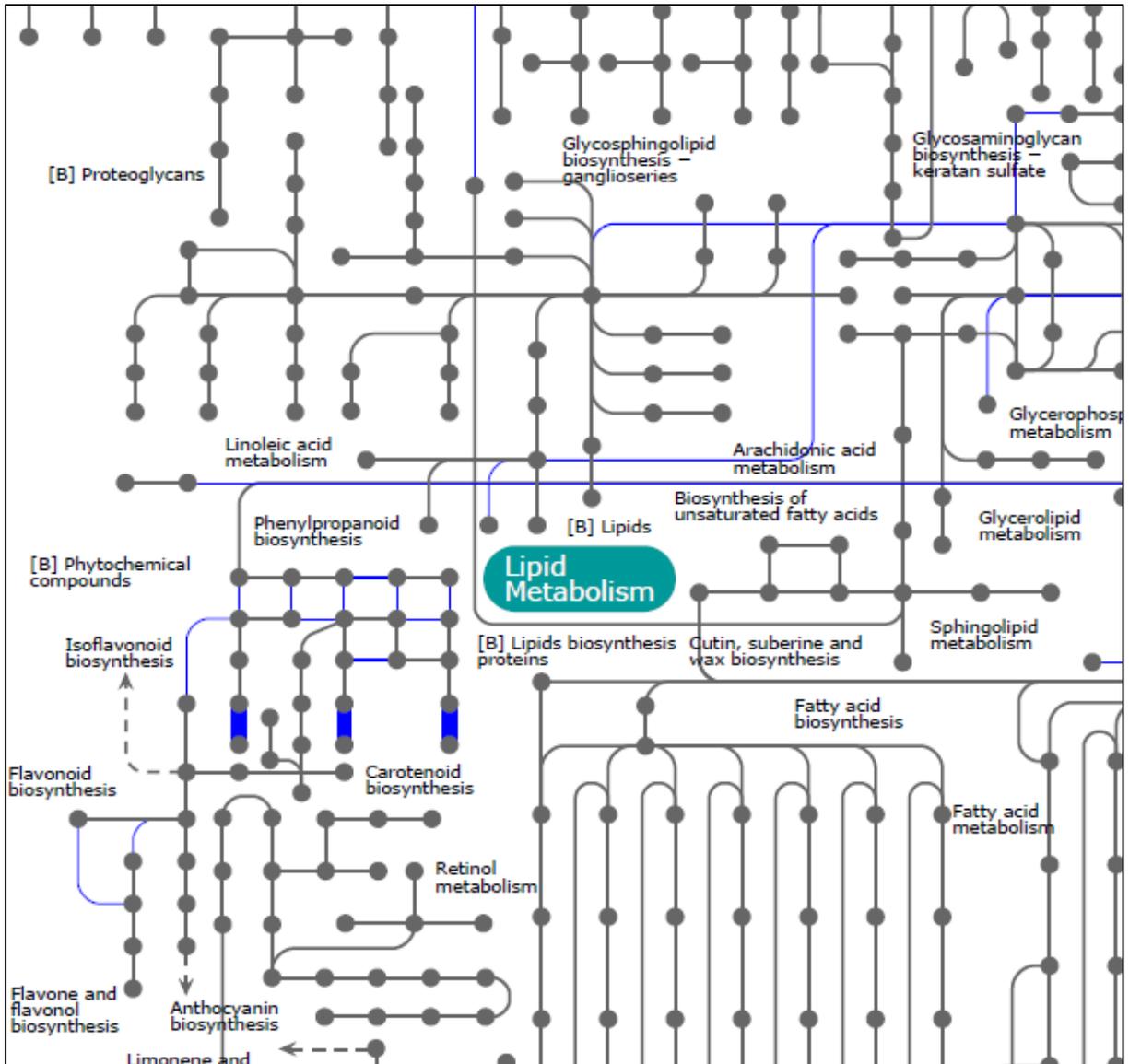


Figura 7 - Mapa das vias metabólicas feito a partir dos transcritos diferencialmente expressos entre os tratamentos de Infestação e Não Infestação na variedade CD-215

6 DISCUSSÃO

Como anteriormente abordados nos resultados, os dois alinhadores utilizados *TopHat* e *STAR* obtiveram resultados muito semelhantes, desde números de transcritos diferencialmente expressos em cada comparação das amostras, até nas listagens dos genes. Esses dois fatores foram muito importantes para garantia da confiabilidade das análises posteriormente feitas. A identificação de transcritos comuns aos relacionados à resistência a afídeo em soja, publicados no trabalho realizado por Li (2008), tais como *cytochrome P450*, *Protein kinase superfamily* e *Thioredoxin superfamily protein* mostra-nos que os mecanismos de defesa ao percevejo *P. guildinii* e ao afídeo são semelhantes em soja.

O nosso estudo propõe pelo alinhador *TopHat* que os 119 transcritos mostrados na Tabela 2 são os mais importantes no mecanismo de resistência ao percevejo *P. guildinii*, pois estão presentes unicamente na variedade IAC-100 sob condição de infestação. Dentre esses 119 existem 23 (19,33%) que ainda não possuem funções conhecidas. O mais interessante é que quatro desses transcritos com funções desconhecidas encontram-se entre os 10 primeiros listados no *ranking* de maiores *fold changes* (maiores modificações em sua expressão), sendo que o transcrito *Glyma15g35221* foi o que obteve a maior modificação de sua expressão. Dessa forma, dentro de um complexo metabólico que engloba todos os transcritos expressos pela variedade IAC-100 submetida à infestação (Tabela 3), esses são os transcritos mais representativos no fator resistência ao percevejo *P. guildinii*. De acordo com o alinhador *STAR* foram identificados 128 transcritos (Tabela 7) relacionados unicamente com o processo de infestação, pois estão presentes apenas na variedade resistente sob condição de infestação. Do total de 128 transcritos, 16 (12,5%) ainda não possuem suas funções conhecidas. Nos quais um deles é o transcrito que obteve o maior score de *fold change*, ou seja, teve a maior modificação em sua expressão entre todos os transcritos presentes unicamente na variedade IAC-100 e, portanto, o transcrito *Glyma19g31811* pode ser considerado o mais importante no processo de resistência ao percevejo *P. guildinii*. Descobrir as funções dos transcritos desconhecidos é o próximo passo para a identificação do processo metabólico de resposta, uma vez que esses encontram-se entre os que possuem as maiores modificações em suas expressões na variedade resistente infestada (Tabelas 2 e 7).

Comparando-se as Tabelas 2 e 7 contendo os transcritos únicos da IAC-100 sob infestação, provenientes dos resultados dos diferentes alinhadores, podemos notar que elas possuem 79 transcritos em comum. A diferença encontrada entre o número de transcritos diferencialmente expressos pelos alinhadores *TopHat* e *STAR* é devida aos diferentes algoritmos utilizados por esses programas.

Os genes de resistência a doenças (genes R) foram encontrados em todas as comparações das amostras (Tabelas 2 a 7) e estão relacionados com proteínas *TIR-NBS-LRR* e com os domínios *LRR* e *NB-ARC*.

As plantas evoluíram múltiplos mecanismos de defesa para restringir o crescimento de patógenos, e esses genes R são empregados mediante o reconhecimento da invasão de específicos patógenos ativando o reconhecimento dos fatores de avirulência (*avr*) correspondentes (FLOR, 1971). A caracterização dos genes R e elucidação e manipulação de vias de sinalização de resistência a doenças e pragas em plantas são o principal foco da investigação sobre a resistência no melhoramento de plantas. A maioria dos produtos desses genes R são estruturalmente caracterizados pela presença dos motivos repetidos ricos em leucina (LRR) (KOBÉ; DEISENHOFER, 1994). Um fator importante é a conservação estrutural desses domínios, dado que estes genes R especificam resistência a vírus, fungos, bactérias e patógenos como nematóides (BENT, 1996; MILLIGAN; BODEAU; YAGHOUBI, 1998; STASKAWICZ; AUSUBEL; BAKER, 1995).

Outros genes descritos como mais importantes no processo regulatório e de sinalização a mecanismos de estresses biológicos causados por artrópodes são os relacionados ao ácido jasmônico, ácido salicílico e/ou etileno (ARIMURA, 2009). Inúmeros transcritos relacionados com essas vias foram encontrados em ambas as análises feitas, tais como *ACC synthase 1*, *ethylene forming enzyme*, *terpene-synthase 0₂*, *gibberellin 2-oxidase*, *pleiotropic drug resistance 11*, *cytokinin oxidase/dehydrogenase 6*, *jasmonate-zim-domain protein 1* e *GDSL-like Lipase/Acylhydrolase superfamily protein* (Tabelas 2 e 7).

O ácido salicílico é uma importante molécula sinalizadora envolvida em respostas de resistência a doenças localmente e sistemicamente induzidas. Recentes avanços na nossa compreensão da defesa e sinalização da planta revelaram que as plantas utilizam uma rede de vias de transdução de sinais, algumas das quais são independentes de ácido salicílico. Estão surgindo evidências de que o ácido jasmônico e o etileno desempenham papéis fundamentais nessas vias de ácido salicílico independentes. A interrelação entre as vias dependentes de ácido salicílico e as independentes fornece grande potencial regulatório para ativar vários mecanismos de resistência em combinações variadas (CORNÉ, 1999).

O gás etileno (C₂H₄) é a única molécula de sinalização da planta que regula vários processos no desenvolvimento e resposta a estresse (ABELES; MORGAN; JUNIOR SALTVEIT, 1992). A enzima chave em dois dos passos de biossíntese do etileno é a *sintetase 1-aminociclopropano-1-carboxilato* (ACS), que catalisa a conversão de *S-adenosilmetionina* (AdoMet) em ácido *1-aminociclopropano-1-carboxilato* (ACC), o precursor do etileno. A

produção de etileno é aumentada por uma variedade de fatores externos incluindo ferimento, infecção viral, elicitores, tratamentos hormonais, injúrias por frio, seca e poluentes como íons (YANG; HOFFMAN, 1984; ABELES; MORGAN; JUNIOR SALTVEIT, 1992; BLEECKER; KENDE, 2000; THOMMA; PENNINCKX; BROEKAERT, 2001). Sobre estimulação, a atividade de ACS é induzida rapidamente e portanto ACS é considerado o passo limitante na biossíntese de etileno e o mais importante na indução do etileno (YANG; HOFFMAN, 1984; KENDE, 1993). Plantas sobre estresses de origem biótica e abiótica produzem níveis elevados de etileno, os quais são percebidos por receptores de etileno e desencadeiam respostas celulares. A fosforilação e de-fosforilação de proteínas resultaram na regulação da indução de etileno por estresses (SPANU; GROSSKOPF; FELIX, 1994; TATSUKI; MORI, 2001; WANG; LI; JUNIOR ECKER, 2002; CHAE; FAURE, KIEBER, 2003), com base em estudos utilizando proteínas cinases e inibidores de fosfatases. O uso de sistemas transgênicos para ganho de função de *SIPK*, uma *MAPK* (mitogen-activated protein kinase), resultou em aumento dramático da produção de etileno coincidindo com o aumento da atividade de ACS (KIM; LIU; THORNE, 2003). A cascata de *SIPK* mostrou-se relacionada à indução da biossíntese de etileno em resposta à infecção viral de *TMV* (*Tobacco mosaic virus*) em tabaco (KIM; LIU; THORNE, 2003). Desta forma a indução da expressão e por conseguinte da atividade de ACS estão relacionadas aos mecanismos de resposta das plantas frente aos estresses bióticos.

A partir da comparação entre as Tabelas 2 e 3, verificou-se que os transcritos expressos somente na variedade IAC-100 Infestada não foram classificados entre os 100 transcritos de modificações mais drásticas em suas expressões. Contudo, existem transcritos relacionados a outros locos, observados na Tabela 3, que realizam funções semelhantes aos da Tabela 2 como o *GDSL-like Lipase/Acylhydrolase superfamily protein* e o *gibberellin 3-oxidase 1*. Entre as Tabelas 2 e 4 (classificação dos 100 transcritos com os menores *fold change*) foram identificados 7 transcritos idênticos, ou seja, esses genes presentes unicamente na variedade IAC-100 Infestada não tiveram modificações acentuadas em suas expressões, não os classificando assim como os mais importantes para a resistência fisiológica ao percevejo.

A alteração da expressão do gene sob estresse, incluindo o ataque de insetos, leva a alterações qualitativas e quantitativas nas proteínas, que por sua vez desempenham um papel importante na transdução de sinal, e de defesa contra processos oxidativos (USHA RANI; JYOTHSNA, 2010; GULSEN; EICKHOFF; HENG-MOSS, 2010).

Quando analisamos o ranking dos 100 transcritos que obtiveram as maiores modificações de suas regulações na variedade IAC-100 pelo alinhador *STAR* (Tabela 8),

constata-se que apenas um transcrito contido nessa classificação também pertencia à lista dos 128 mais importantes para a expressão da resistência ao percevejo. Enquanto que pelo alinhador *TopHat*, nenhum dos transcritos da lista dos 119 estava presente no ranking dos 100 que obtiveram as maiores alterações (Tabela 3). Esses resultados nos mostram que a resposta relacionada à resistência ao percevejo não altera drasticamente o funcionamento desses genes, uma vez que essa modificação de expressão poderia levar a um gasto energético excessivo e uma possível morte da planta, por exemplo. Essa modificação mais branda pode ser confirmada quando fizemos a comparação das Tabelas (2 e 7) dos transcritos ditos como únicos com os Tabelas classificatórias dos 100 transcritos que tiveram a menor alteração em sua expressão na variedade IAC-100 (Tabelas 4 e 9). Pelas análises do *STAR* identificamos 8 transcritos dos 128 classificados entre os 100 de menor modificação em sua expressão (Tabela 9). Já pelo *TopHat* cinco transcritos dos 119 foram classificados entre os 100 de menor modificação no *fold change* (Tabela 4). Reforça-se assim, que apesar das diferenças entre os níveis de expressão desses transcritos sob os tratamentos de infestação e não infestação serem significativamente diferentes, eles não sofrem uma pressão acentuada nos graus de expressão.

Verificou-se também a presença de 56 transcritos exclusivos à variedade CD-215 Infestada. De acordo com a Tabela 5, apenas dois dos 56 transcritos encontram-se classificados entre os 100 transcritos que tiveram as modificações mais bruscas. Enquanto que 24 desses transcritos estão entre os 100 transcritos menores classificados de acordo com suas modificações nas expressões (Tabela 6). Esse resultado nos mostra que a variedade CD-215 também possui um mecanismo único de resistência, embora não eficaz contra o percevejo *P. guildinii*.

Outro estudo feito foi a partir da comparação entre os transcritos tidos como mais importantes na defesa ao percevejo, listados nas Tabelas 2 e 7, com os 140 transcritos identificados no trabalho de microarranjos sobre resistência afídeo em soja (LI, 2008). O objetivo foi o de verificar se havia algum transcrito em comum entre as duas análises. A Tabela com os 140 foi então comparada com as provenientes das seguintes amostras: IAC-100 Infestada x IAC-100 Não Infestada, CD-215 Infestada x CD-215 Não Infestada, IAC-100 Infestada x CD-215 Infestada e IAC-100 Não Infestada x CD-215 Não Infestada.

A partir da comparação com os dados provenientes do alinhador *TopHat* foram encontrados cinco transcritos em comum com as amostras de IAC-100 Infestada x CD-215 Infestada, sendo eles *Glyma12g11130*, *Glyma08g26150*, *Glyma02g03250*, *Glyma13g04710* e *Glyma03g38490*. A partir dos dados do *STAR* foram encontrados quatro transcritos em

comum com as amostras de IAC-100 Infestada x CD-215 Infestada, sendo eles *Glyma03g38490*, *Glyma13g04710*, *Glyma02g03250* e *Glyma12g11130*.

Os transcritos expressos somente nas amostras IAC-100 Não Infestada x CD-215 Não Infestada, evidenciou que vias metabólicas constitutivas são em parte semelhantes com as apresentadas pelo resultado dos transcritos elencados para resistência a afídeos, uma vez que foram obtidos 22 genes em comum pelo *TopHat* e 23 pelo *STAR*. Outros 25 transcritos foram expressos entre as condições de infestação e não-infestação na variedade IAC-100 pelo *TopHat*, porém apenas um deles (*Sec14p-like phosphatidylinositol transfer family protein*) também estava presente na tabela dos transcritos refinados (Tabela 2), julgados essenciais para o mecanismo de resistência ao percevejo *P. guildinii*. Pelo alinhador *STAR* foram identificados 27 transcritos idênticos aos presentes na comparação entre as amostras infestada e não infestada da variedade IAC-100, sendo que nenhum deles estava presente na Tabela 7 dos transcritos refinados, julgados essenciais para o mecanismo de resistência ao percevejo *P. guildinii*.

O trabalho realizado por Li (2008) utilizou a variedade de soja *Dowling* como resistente e *Williams 82* como susceptível para as comparações de infestação e não infestação por afídeo. Uma vez que as variedade utilizadas na pesquisa com afídeos foram diferentes das utilizadas nesse trabalho (IAC-100 e CD-215), não é possível uma otimização real de todos os transcritos compatíveis possíveis entre os dois processos diferentes de infestação ocasionados por *Aphis glycines* e pelo *P. guildinii*.

Os resultados obtidos condizem com uma resistência genética a *P. guildinii* em um modelo que envolve uma organização rápida dos sinais gerados pelo percevejo e os genes de resistência presentes na variedade IAC-100, uma vez que os dados são provenientes de amostras de 24 horas pós infestação e não infestação, mostrando as modificações significativas na expressão de inúmeros genes. Esse mecanismo de resposta rápido propicia um mecanismo de defesa que previne danos maiores aos grãos e à planta.

7 CONCLUSÕES

- Os dois alinhadores utilizados (*TopHat* e *STAR*) obtiveram resultados muito semelhantes, com 79 transcritos em comum (acima de 60% em cada uma das análises).
- Os mecanismos de defesa ao percevejo *Piezodorus guildinii* e ao afídeo *Aphis glycines* são semelhantes em soja.
- A resposta relacionada à resistência ao percevejo não altera drasticamente o funcionamento desses genes únicos à variedade resistente infestada (Tabelas 2 e 7) , uma vez que essa modificação brusca e rápida de expressão poderia levar a um gasto energético excessivo e uma possível morte da planta, por exemplo.
- A variedade CD-215 também possui um mecanismo único de resistência, embora não eficaz contra o percevejo *P. guildinii*.
- Uma vez que os transcritos que diferenciam a IAC-100 da CD-215 não foram os que mais responderam a infestação (com alteração da expressão), isso implica que estes transcritos provavelmente não condicionam uma resposta por si só e que no contexto do global de expressão gênica eles possivelmente influenciam a alteração da expressão de outros genes levando a resposta específica de defesa da planta (Tabelas 2 a 11).

REFERÊNCIAS

- ABELES, F.B.; MORGAN, P.W.; JUNIOR SALTVEIT, M.E. **Ethylene in plant biology**. 2nd ed. New York: Academic Press, 1992. 302 p.
- ARIMURA, G.I.; MATSUI, K.; TAKABAYASHI, J. Chemical and molecular ecology of herbivore-induced plant volatiles: proximate factors and their ultimate functions. **Plant Cell Physiology**, Otsu, v. 50, p. 911-923, 2009.
- BEGON, M.; MORTIMER, M.; THOMPSON, D.J. **Population ecology**: a unified study of animals and plants. 3rd ed. Oxford: Blackwell, 1996. 247p.
- BELL, E.; CREELMAN, R.A.; MULLET, J.E. A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in *Arabidopsis*. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 92, p. 8675-8679, 1995.
- BENT, A.F. Plant disease resistance genes: Functions meets structure. **Plant Cell**, Urbana-Champaign, v. 8, p. 1757-1771, 1996.
- BHATNAGAR-MATHUR, P.; VADEZ, V.; SHARMA, K.K. Transgenic approaches for abiotic stress tolerance in plants: retrospect and prospects. **Plant Cell Reports**, Patancheru, v. 27, p. 411-424, 2008.
- BLEECKER, A.B.; KENDE, H. Ethylene: a gaseous signal molecule in plants. **Annual Review of Cell Development Biology**, Madison, v. 16, p. 1-18, 2000.
- BLOOM, J.S.; KHAN, Z.; KRUGLYAK, L.; SINGH, M.; CAUDY, A.A. Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. **Biology Medical Central Genomics**, Princeton, v. 10, p. 221, 2009.
- BOEGE, K.; MARQUIS, R.J. Facing herbivory as you grow up: the ontogeny of resistance in plants. **Trends in Ecology and Evolution**, Saint Louis, v. 20, p. 441-448, 2005.
- BOHNERT, H.J.; QINGQIU, G.; PINGHUA, L.; MA, S. Unraveling abiotic stress tolerance mechanisms: getting genomics going. **Current Opinion of Plant Biology**, Urbana, v. 9, p. 180-188, 2006
- BORÉM, A.; MILACH, S.C.K. Melhoramento de plantas: o melhoramento de plantas na virada do milênio. **Biotecnologia Ciência & Desenvolvimento**, Viçosa, v. 7, p. 68-72, 1999.
- BOSTOCK, R.M. Signal crosstalk and induced resistance: straddling the line between cost and benefit. **Annual Review of Phytopathology**, Palo Alto, v. 43, p. 545-580, 2005
- BRIER, H.; ROGERS, D.J. Susceptibility of soybeans to damage by *Nezara viridula* (L.) (Hemiptera: *Pentatomidae*) and *Riptortus serripes* (F.) (Hemiptera: *Alydidae*) during three stages of pod development. **Journal of the Australian Entomological**, Kingaroy, v. 30, p. 123-128, 1991.

CARVALHO, M.C.C.G.; DA SILVA, D.C.G. Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. **Ciência Rural**, Santa Maria, v.40 , n. 3, p. 735-744, 2010.

CHAE, H.S.; FAURE, F.; KIEBER, J.J. The *eto1*, *eto2*, and *eto3* mutations and cytokinin treatment increase ethylene biosynthesis in *Arabidopsis* by increasing the stability of ACS protein. **Plant Cell**, Chapel Hill, v. 15, p. 545–559, 2003.

CHEHAB, E.W.; PEREA, J.V.; GOPALAN, B.; THEG, S.; DEHESH, K. Oxylin pathway in rice and *Arabidopsis*. **Journal of Integrative Plant Biology**, Davis, v. 49, p. 43–51, 2007.

CIVIDANES, F.J.; PARRA, J.R.P. Biologia em diferentes temperaturas e exigências térmicas de percevejos pragas da soja. III. *Piezodorus guildinii* (West., 1837) (Hepteroptera: Pentatomidae). **Científica**, Londrina, v. 22, p. 177-186, 1994.

COMPANHIA NACIONAL DE ABASTECIMENTO. **4º levantamento de grãos safra 2012/2013**. 2013. Disponível em: <http://www.conab.gov.br/OlalaCMS/uploads/arquivos/13_01_09_17_44_20_boletim_graos_janeiro_2013.pdf>. Acesso em: 06 fev. 2013.

CONESA, A.; GÖTZ, S.; GARCÍA-GÓMEZ, J.M.; TEROL, J.; TALÓN, M.; ROBLES, M. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, Valencia, v. 21, n. 18, p. 3674-3676, 2005.

COOPERATIVA CENTRAL DE AGROPECUÁRIA DE DESENVOLVIMENTO TECNOLÓGICO E ECONÔMICO LTDA. Disponível em: <<http://www.coodetec.com.br>>. Acesso em: 13 nov. 2013.

CORRÊA-FERREIRA, B.S. Suscetibilidade da soja a percevejos na fase anterior ao desenvolvimento das vagens. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 11, p. 1067-1072, 2005.

CORRÊA-FERREIRA, B.S.; PANIZZI, A.R. **Percevejo da soja e seu manejo**. Londrina: EMBRAPA, CNPSo, 1999. 45 p. (Circular Técnica, 24).

CRAWLEY, M.J. Plant-herbivores dynamics. In: _____. (Ed.). **Plant ecology**. 2nd ed. Oxford: Blackwell Science, 1997. p. 401-474.

DEGRANDE, P.E.; VIVAN, L.M. Pragas da soja. **Boletim de Pesquisa de Soja**, Rondonópolis, v. 11, p. 274, 2007.

DEGTYARENKO, K.; MATOS, P.; ENNIS, M.; HASTINGS, J.; ZBLINDEN, M.; McNAUGHT, A.; ALCÁNTARA, R.; DARSOW, M.; ASHBURNER, M. ChEBI: a database and ontology for chemical entities of biological interest. **Nucleic Acids Research**, Cambridge, v. 36, p. D344–D350, 2008.

De VOS, M.; Van OOSTEN, V.R.; Van POECKE, R.M.P.; Van PELT, J.A.; POZO, M.J.; MULLER, M.J.; BUCHALA, A.J.; METRAUX, J.P.; Van LOON, L.C.; DICKE, M.. Signal signature and transcriptome changes of *Arabidopsis* during pathogen and insect attack. **Molecular Plant-Microbe Interactions**, Utrechtv, v. 18, p. 923-937, 2005.

DJEBALI, S.; DAVIS, C.A.; MERKEL, A.; DOBIN, A.; LASSMANN, T.; MORTAZAVI, A.; TANZER, A.; LAGARDE, J.; LIN, W.; SCHLESINGER, F.; XUE, C.; MARINOV, G.K.; KHATUN, J.; WILLIAMS, B.A.; ZALESKI, C.; ROZOWSKY, J.; RODER, J.; KOKOCINSKI, F.; ABDELHAMID, R.F.; ALIOTO, T.; ANTOSHECHKIN, I.; BAER, M.T.; BAR, N.S.; BATUT, P.; BELL, K.; BELL, I.; CHAKRABORTTY, S.; CHEN, X.; CHRAST, J.; CURADO, J.; DERRIEN, T.; DRENKOW, J.; DUMAIS, E.; DUMAIS, J.; DUTTAGUPTA, R.; FALCONNET, E.; FASTUCA, M.; FEJES-TOTH, K.; FERREIRA, P.; FOISSAC, S.; FULLWOOD, M.J.; GAO, H.; GONZALEZ, D.; GORDON, A.; GUNAWARDENA, H.; HOWALD, C.; JHA, S.; JOHNSON, R.; KAPRANOV, P.; KING, B.; KINGSWOOD, C.; LUO, O.J.; PARK, E.; PERSAUD, K.; PREALL, J.B.; RIBECA, P.; RISK, B.; ROBYR, D.; SAMMETH, M.; SEE, L.; SHAHAB, A.; SKANCKE, J.; SUZUKI, A.M.; TAKAHASHI, H.; TILGNER, H.; TROUT, D.; WALTERS, N.; WANG, H.; WROBEL, J.; YU, Y.; RUAN, X.; HAVASHIZAKI, Y.; HARROW, J.; GERSTEIN, M.; HUBBARD, T.; REYMOND, A.; ANTONARAKIS, S.E.; HANNON, G.; GIDDINGS, M.C.; RUAN, Y.; WOLD, B.; CARNINCI, P.; GUIGÓ, R.; GINGERAS, T.R. Landscape of transcription in human cells. **Nature**, London, v. 489, p. 101-108, 2012.

DOBIN, A.; DAVIS, C.A.; SCHLESINGER, F.; DRENKOW, J.; ZALESKI, C.; JHA, S.; BATUT, P.; CHAISSON, M.; GINGERAS, T.R. STAR: ultrafast universal RNA-seq aligner. **Bioinformatics**, New York, v. 29, n. 1, p. 15-21, 2013.

DUDAREVA, N.; NEGRE, F.; NAGEGOWDA, D.A.; ORLOVA, I. Plant volatiles: recent advances and future perspectives. **Critical Reviews in Plant Science**, West Lafayette, v. 25, p. 417-40, 2006.

DUQUE, A.S.; ALMEIDA, A.M.; SILVA, A.B.; SILVA, J.M.; FARINHA, A.P.; SANTOS, D.; FEVEREIRO, P.; ARAÚJO, S.S. Abiotic stress responses in plants: unraveling the complexity of genes and networks to survive. In: **Abiotic stress: plant responses and applications in agriculture**. Rijeka, 2013. chap. 3.

ERENO, D. **Armadilhas biológicas: percevejo-marrom da soja pode ser monitorado com feromônio sintético desenvolvido pela Embrapa**, 2011. Disponível em: <<http://revistapesquisa.fapesp.br/?art=4560&bd=1&pg=1&lg>>. Acesso em: 05 fev. 2013.

EULGEM, T. Regulation of the *Arabidopsis* defense transcriptome. **Trends Plant Science**, Riverside, v. 10, p. 71-80, 2005.

FEDURCO, M.; ROMIEU, A.; WILLIAMS, S.; LAWRENCE, I. & TURCATT, G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. **Nucleic Acids Research**, Lausanne, v. 34, n. 3, e. 22, p. 1-33, 2006.

FERNANDES, O.A.; CORREIA, A.C.B.; BORTOLI, S.A. **Manejo integrado de pragas e nematóides**. Jaboticabal: FUNEP, 1990. v. 2, 352 p.

FERRY, N.; EDWARDS, M.G.; GATEHOUSE, J.A. & GATEHOUSE, A.M.R. Plant-insect interactions: molecular approaches to insect resistance. **Current Opinion on Biotechnology**, Newcastle, v. 15, p. 1-7, 2004.

FILICHKIN, S.A.; PRIEST, H.D.; GIVAN, S.A.; SHEN, R.; BRYANT, D.W.; FOX, S.E.; WONG, W.K.; MOCKLER, T.C. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. **Genome Research**, Corvallis, v. 20, p. 45-58, 2010.

FLOR, H. Current status of the gene-for-gene concept. **Annual Review of Phytopathology**, Palo Alto, v. 9, p. 275-296, 1971.

FUGI, C.G.Q.; LOURENÇÃO, A.L.; PARRA, J.R.P. Biology of *Anticarsia gemmatalis* on soybean genotypes with different degrees of resistance to insects. **Scientia Agricola**, Piracicaba, v. 62, p. 31-35, 2005.

GAMUNDI, J.C.; ANDRIAN, M.; BACIGALUPPO, D.; LAGO, M.; LENZI, L.; RANDAZZO, P.; BODRERO, M. Incidencia del complejo de chinches en el cultivo de soja con diferentes espaciamentos entre líneas . Soja: para mejorar la producción de soja. [S.l.]: **INTA - Estación Experimental Agropecuaria Oliveros**, Oliveros, v. 24, p. 79-86, 2003.

GATEHOUSE, J.A. Plant resistance towards insect herbivores: a dynamic interaction. **New Phytology**, Durham, v. 156, p. 145-169, 2002.

GOECKS, J.; NEKRUTENKO, A.; TAYLOR, J. Galaxy team galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. **Genome Biology**, Atlanta, v. 11, p. R86, 2010.

GOODSTEIN, M.D.; SHU, S.; HOWSON, R.; NEUPANE, R.; HAYES, R.D.; FAZO, J.; MITROS, T.; DIRKS, W.; HELLSTEN, U.; PUTNAM, N.; ROKHSAR, D.S. Phytozome: a comparative platform for green plant genomics. **Nucleic Acids Research**, San Francisco, v. 40, p. D1178-D1186, 2012.

GÖTZ, S.; GARCÍA-GÓMEZ, J.M.; TEROL, J.; WILLIAMS, T.D.; NAGARAJ, S.H.; NUEDA, M.J.; ROBLES, M.; TALON, M.; DOPAZO, J.; CONESA, A. High-throughput functional annotation and data mining with the blast2go suite. **Nucleic Acids Research**, San Francisco, v. 36, n. 10, p. 3420-3435, 2008.

GRABHERR, M.G.; HASS, B.J.; YASSOUR, M.; LEVIN, J.Z.; THOMPSON, D.A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACOEN, N.; GNIRKE, A.; RHIND, N.; di PALMA, F.; BIRREN, B.W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Full-length transcriptome assembly from RNA-seq data without a reference genome. **Nature Biotechnology**, Cambridge, v. 29, p. 644-652, 2011.

GRAVELEY, B.R.; BROOKS, A.N.; CARLSON, J.W.; DUFF, M.O.; LANDOLIN, J.M.; YANG, L.; ARTIERI, C.G.; van BAREN, J.; BOLEY, N.; BOOTH, B.W.; BROWN, J.B.; CHERBAS, L.; DAVIS, C.A.; DOBIN, A.; LI, R.; LIN, W.; MALONE, J.H.; MATTIUZZO, N.R.; MILLER, D.; STURGILL, D.; TUCH, B.B.; ZALESKI, C.; ZHANG, D.; BLANCHETTE, M.; DUDOIT, S.; EADS, B.; GREEN, R.E.; HAMMONDS, A.; JIANG, L.; KAPRANOV, P.; LANGTON, L.; PERRIMON, N.; SANDLER, J.E.; WAN, K.H.; WILLINGHAM, A.; ZHANG, Y.; ZOU, Y.; ANDREWS, J.; BICKEL, P.J.; BRENNER, S.E.; BRENT, M.R.; CHERBAS, P.; GINGERAS, T.R.; HOSKINS, R.A.; KAUFMAN, T.C.; OLIVER, B.; CELNIKER, S.E. The developmental transcriptome of *Drosophila melanogaster*. **Nature**, London, v. 471, p. 473-479, 2011.

GUEDES, J.V.C.; ARNEMANN, J.A.; STÜRMER, G.R.; MELO, A.A.; BIGOLIN, M.; PERINI, C.R.; SARI, B.G. Percevejos da soja: novos cenários, novo manejo. **Revista Plantio Direto**, Santa Maria, v. 12, n. 1, p. 24-30, 2012.

GULSEN, O.; EICKHOFF, T.; HENG-MOSS, T.; SHEARMAN, R.; BAXENDALE, F.; SARATH, G.; LEE, D. Characterization of peroxidase changes in resistant and susceptible warmseason turfgrasses challenged by *Blissus occiduus*. **Arthropod-Plant Interact**, Kayseri, v. 4, p. 45-55, 2010.

HANLEY, M.E.; LAMONT, B.B.; FAIRBANKS, M.M.; RAFFERTY, C.M. Plant structural traits and their role in antiherbivore defense. **Perspectives Plant Ecology Evolution Systematics**, Plymouth, v. 8, p. 157-178, 2007.

HARE, J.D. Ecological role of volatiles produced by plants in response to damage by herbivorous insects. **Annual Review of Entomology**, Riverside, v. 56, p. 161-180, 2011.

HILL, C.B.; LI, Y.; HARTMAN, G.L. Resistance to the soybean aphid in soybean germplasm. **Crop Science**, Urbana, v. 44, p. 98-106, 2004.

HINTON, J.C.; HAUTEFORT, I.; ERIKSSON, S.; YHOMPSON, A. & RHEN, M. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. **Current opinion in Microbiology**, London, v. 7, p. 277-282, 2004.

HOFFMANN-CAMPO, C.B.; MOSCARDI, F.; CORRÊA-FERREIRA, B.S.; OLIVEIRA, L.J.; SOSA-GOMEZ, D.R.; PANIZZI, A.R.; CORSO, I.C.; GAZZONI, D.L.; OLIVEIRA, E.B. **Pragas da soja no Brasil e seu manejo integrado**. Londrina: Embrapa Soja, 2000. 70 p. (Circular Técnica, 30).

HOWE, G.A.; JANDER, G. Plant immunity to insect herbivores. **Annual Review of Plant Biology**, East Lansing, v. 59, p. 41-66, 2008.

HUANG, H.; LIU, C.; ZHOU, X.J. Bayesian approach to transforming public gene expression repositories into disease diagnosis databases. **Proceedings of the National Academy of Sciences of the USA**, Washington, v. 107, n. 15, p. 6823-6828, 2010.

KALOSHIAN, I.; WALLING, L.L. Hemipterans as plant pathogens. **Annual Review of Plant Biology**, Riverside, v. 43, p. 491-521, 2005.

KANEHISA, M.; ARAKI, M.; GOTO, S.; HATTORI, M.; HIRAKAWA, M.; ITOH, M.; KATAYAMA, T.; KAWASHIMA, S.; OKUDA, S.; TOKIMATSU, T.; YAMANISHI, Y. KEGG for linking genomes to life and the environment. **Nucleic Acids Research**, Kyoto, v. 36, p. D480-D484, 2007.

KENDE, H. Ethylene biosynthesis. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 44, p. 283-307, 1993.

KIM, C.Y.; LIU, Y.; THORNE, E.T.; YANG, H.; FUKUSHIGE, H.; GASSMANN, W.; HILDEBRAND, D.; SHARP, R.E.; ZHANG, S. Activation of a stress-responsive mitogen-activated protein kinase cascade induces the biosynthesis of ethylene in plants. **The Plant Cell**, Columbia, v. 15, p. 2707-2718, 2003.

KIM, D.; PERTEA, G.; TRAPNELL, C.; PIMENTEL, H.; KELLEY, R.; SALZBERG, S.L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. **Genome Biology**, Maryland, v. 14, p. R36, 2013

KOBE, B.; DEISENHOFER, J. The leucine-rich repeat: a versatile binding motif. **Trends Biochemical Science** Dallas, v. 19, p. 415-421, 1994.

LEON, J.; SANCHEZ-SERRANO, J.J. Molecular biology of jasmonic acid biosynthesis in plants. **Plant Physiology and Biochemistry**, Madrid, v. 37, p. 373-380, 1999.

LETUNIC, I.; YAMADA, T.; KANEHISA, M.; BORK, P. iPath: interactive exploration of biochemical pathways and networks. **Trends in Biochemical Sciences**, Heidelberg, v. 33, n. 3, p. 101-103, 2008

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**, Champaign, v. 179, p. 185-195, 2008.

LISTER, R.; PELIZZOLA, M.; KIDA, Y.S.; HAWKINS, R.D.; NERY, J.R.; HON, G.; ANTOSIEWICZ-BOURGET, J.; O'MALLEY, R.; CASTANON, R.; KLUGMAN, S.; DOWNES, M.; YU, R.; STEWART, R.; REN, B.; THOMSON, J.A.; EVANS, R.M.; ECKER, J. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. **Nature**, London, v. 470, p. 68-73, 2011.

LOPEZ-MAURY, L.; MARGUERAT, S.; BAHLER, J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. **Nature Review Genetics**, London, v. 9, p. 583-593, 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, p. 65-72, 1987.

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, p. 1347-1351, 1999.

MARDIS, E.R. Next-generation DNA sequencing methods. **Annual Review of Genomics and Human Genetics**, St. Louis, v. 9, p. 387-402, 2008.

MARTINEZ DE ILARDUYA, O.; XIE, Q.G.; KALOSHIAN, I. Aphid-induced defense responses in *Mi-1* mediated compatible and incompatible tomato interactions. **Molecular Plant-Microbe Interactions**, Riverside, v. 16, p. 699-708, 2003.

MESTDAGH, P.; Van VLIERBERGHE, P.; DE WEER, A.; MUTH, D.; WESTERMANN, F.; SPELEMAN, F.; VANDESOMPELE, J. A novel and universal method for microRNA RT-qPCR data normalization. **Genome Biology**, Ghent, v. 10, R64, 2009.

METZKER, M.L. Sequencing technologies: the next generation. **Nature Review of Genetics**, Houston, v. 11, p. 31-46, 2010.

MILLIGAN, S.B.; BODEAU, J.; YAGHOUBI, J.; KALOSHIAN, I.; ZABEL, P.; WILLIAMSON, V.M. The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. **Plant Cell**, Davis, v. 10, p. 1307-1319, 1998.

MINER, F.D. **Biology and control of stink bugs on soybeans**. Fayetteville: Arkansas Experiment Station, 1966. 40 p. (Bulletin, 708).

MIRANDA, M.A.C.; LOURENÇÃO, A.L. Melhoramento genético da soja para 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, Foz do Iguaçu. **Anais...** Londrina: Embrapa Soja, 2002. p. 52-60.

MIRANDA, M.A.C.; BRAGA, N.R.; LOURENÇÃO, A.L.; MIRANDA, F.T.S.; UNÊDA, S.H.; ITO, M.F. Descrição, produtividade e estabilidade da cultivar de soja IAC-24, resistente a insetos. **Bragantia**, Campinas, v. 62, p. 29-37, 2003.

MIRANDA, M.A.C.; BRAGA, N.R.; MIRANDA, F.T.S.; UNEDA, S.H.; LOURENÇÃO, A.L.; ITO, M.F. IAC-23 e IAC-24: cultivares de soja resistentes a insetos para o estado de São Paulo. In CONGRESSO BRASILEIRO DE MELHORAMENTO DE PLANTAS, 1., 2001, Goiania. **Resumos...** Goiania: SBMP, 2001.1 CD-ROM.

MIZUNO, H.; KAWAHARA, Y.; SAKAI, H.; KANAMORI, H.; WAKIMOTO, H.; YAMAGATA, H.; OONO, Y.; WU, J.; IKAWA, H.; ITOH, T.; MATSUMOTO, T. Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L.). **Bio Med Central Genomics**, Tsukuba, v. 11, p. 683, 2010.

MORALES, L.; SILVA, M.T.B. Desafios do MIPsoja na região sul do Brasil e o plantio direto. In: CONGRESSO BRASILEIRO DE SOJA, 4., 2006, Londrina. **Anais...** Londrina: Embrapa Soja, 2006. p. 134-139.

MORAN, P.J.; THOMPSON, G.A. Molecular responses to aphid feeding in arabidopsis in relation to plant defense pathways. **Plant Physiology**, Tucson, v. 125, p. 1074-1085, 2001.

MOROZOVA, O.; HIRST, M.; MARRA, M.A. Applications of new sequencing technologies for transcriptome analysis. **Annual Review of Genomics and Human Genetics**, Vancouver, v. 10, p. 135-151, 2009.

MOSCARDI, F.; CORRÊA-FERREIRA, B.S.; SOSA-GÓMEZ, D.R.; CORSO, I.C.; BUENO, A.F.; HOFFMANN-CAMPO, C.B.; PANIZZI, A.R. Diagnóstico da situação atual do manejo de pragas na cultura da soja no Brasil. In: CONGRESSO BRASILEIRO DE SOJA, 5., 2009, Londrina. **Anais...** Londrina: Embrapa Soja, 2009. 1 CD-ROM.

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

MUSSER, F.R.; CATCHOT, A.L.; GIBSON, B.K.; KNIGHTEN, K.S. Economic injury levels for southern green stink bugs (*Hemiptera: Pentatomidae*) in R7 growth stage soybeans. **Crop Protection**, Mississippi, v.30, p.63-69, 2011.

NAGALAKSHMI, U.; WAERN, K.; SNYDER, M. RNA-Seq: a method for comprehensive transcriptome analysis. **Current Protocol in Molecular Biology**, New Haven, cap. 4, n. 4.11, p. 11-13, 2010.

NAGALAKSHMI, U.; WANG, Z.; WAERN, K.; SHOU, C.; RAHA, D.; GERSTEIN, M.; SNYDER, M. The transcriptional landscape of the yeast genome defined by RNA sequencing. **Science**, New Haven, v. 320, n. 5881, p. 1344-1349, 2008.

OLIVEIRA, J.R. Ovos de *Euschistus heros* parasitados por *Telenomus podisi*. **Informativo do Manejo Ecológico de Pragas**, Brasília, v. 63, p. 738-739, 2010.

PANIZZI, A.R.; NIVA, C.C. Overwintering strategy of the brown stink bug in northern Paraná. **Pesquisa Agropecuária Brasileira**, Brasília, v. 29, p. 509-511, 1994.

PANIZZI, A.R.; VIVAN, L.M. Seasonal abundance of the neotropical brown stink bug, *Euschistus heros*, in overwintering sites, and the breaking of dormancy. **Entomologia Experimentalis et Applicata**, Londrina, v. 82, p. 213-217, 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.; PANIZZI, M.C.C.; BAYS, I.A. ; ALMEIDA, L.A. Danos por percevejo em genótipos de soja com semente pequena. **Pesquisa Agropecuária Brasileira**, Brasília, v. 21, n. 6, p. 571-577, 1986.

PANIZZI, A.R.; SMITH, J.G.; PEREIRA, L.A.G. ; YAMASHITA, J. Efeito dos danos de *Piezodorus guildinii* (Westwood, 1837) no rendimento e qualidade da soja. In: SEMINÁRIO NACIONAL DE PESQUISA DA SOJA, 1., 1978, Londrina. **Anais...** Londrina: EMBRAPA, CNPSo, 1979. v. 2, p. 59-78.

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: CRC Press, 2000. p. 432-434.

PANIZZI, A.R.; CORRÊA, B.S.; GAZZONI, D.L.; OLIVEIRA, E.B.; NEWMAN, G.G.; TURNIPSEED, S.G. **Insetos da soja no Brasil**. Londrina: EMBRAPA, CNPSo, 1977. 20 p. (Boletim Técnico, 1).

PETROSINO, J.F.; HIGHLANDER, S.; LUNA, R.A.; GIBBS, R.A.; VERSALOVIC, J. Metagenomic pyrosequencing and microbial identification. **Clinical Chemical**, Houston, v. 55, p. 856-866, 2009.

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

- QUINTELA, E.D.; FERREIRA, S.B.; GUIMARÃES, W.F.F.; OLIVEIRA, L.F.C.; OLIVEIRA, A.C.; CZEPAK, C. Desafios do MIP em soja em grandes propriedades no Brasil Central. In: CONGRESSO BRASILEIRO DE SOJA, 4., 2006, Londrina. **Anais...** Londrina: Embrapa Soja, 2006. p. 127-133.
- ROBERTSON, G.; SCHEIN, J.; CHIU, R.; CORBETT, R.; FIELD, M.; JACKMAN, S.D.; MUNGALL, K.; LEE, S.; OKADA, H.M.; QIAN, J.Q.; GRIFFITH, M.; RAYMOND, A.; THIESSEN, N.; CEZARD, T.; BUTTERFIELD, Y.S.; NEWSOME, R.; CHAN, S.K.; SHE, R.; VARHOL, R.; KAMOH, B.; PRABHU, A.L.; TAM, A.; ZHAO, Y.; MOORE, R.A.; HIRST, M.; MARRA, M.A.; JONES, S.J.; HOODLESS, P.A.; BIROL, I. *De novo* assembly and analysis of RNA-seq data. **Nature Methods**, Vancouver, v. 7, p. 909–912, 2010.
- ROESSING, A.C.; SANCHES, A.C.; MICHELLON, E. As perspectivas de expansão da soja. In: CONGRESSO DA SOBER, 43., 2005, Ribeirão Preto. **Anais...** São Paulo: SOBER, 2005. p. 1-20.
- ROGERS, E.E.; AUSUBEL, F.M. *Arabidopsis* enhanced disease susceptibility mutants exhibit enhanced susceptibility to several bacterial pathogens and alterations in PR-1 gene expression. **Plant Cell**, Boston, v. 9, p. 305-316, 1997.
- ROSSETTO, C.J. Breeding for resistance to stink bugs. In: CONFERENCIA MUNDIAL DE INVESTIGACIÓN EN SOJA, 4., 1989, Buenos Aires. **Actas...** Buenos Aires: Orientación Gráfica Editora SRL, 1989. v. 4, p. 2046-2060.
- ROSSETTO, C.J.; GALLO, P.B.; RAZERA, L.F.; BORTOLETTO, N.; IGHE, T.; MEDINA, P.F.; TISSELLI-FILHO, O.; AQUILERA, V.; VEIGA, R.F.A.; PINHEIRO, J.B. Mecanismos de resistência ao complexo de percevjos na cultivar de soja IAC-100. **Anais da Sociedade Entomológica do Brasil**, Londrina, v. 24, n. 3, p. 517-522, 1995.
- SANTOS, M.F. **Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevjos**. 2012. 120 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2012.
- SCHUMANN, F.W.; TODD, J.W. Population dynamics of the southern green stink bug (Heteroptera: Pentatomidae) damage to soybean seed. **Journal of Economic Entomology**, Riverside, v. 80, p. 215-220, 1982.
- SCHUSTER, S.C.; Next-generation sequencing transforms today's biology. **Nature Methods**, Pennsylvania, v. 5, p. 16-18, 2008.
- SECRETARIA DE ESTADO DE AGRICULTURA E ABASTECIMENTO. **Soja: análise da conjuntura agropecuária**. 2012. Disponível em: <http://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=9&cad=rja&sqi=2&ved=0CHUQFjAI&url=http%3A%2F%2Fwww.agricultura.pr.gov.br%2Farquivos%2FFile%2Fderal%2FPrognosticos%2Fsoja_2012_13.pdf&ei=uEqEUuuhAeP42AX1qIC4BQ&usq=AFQjCNEXRvFplfuP-hfi3rcSdqMd5TXk8g&sig2=lskRAopthZsQDT6A6u01ew&bvm=bv.56343320,d.b2I>. Acesso em: 03 jan. 2014.

- SESHASAYEE, A.S.N.; BERTONE, P.; FRASER, G.M.; LUSCOMBE, N.M. Transcriptional regulatory networks in bacteria: from input signals to output responses. **Current Opinion in Microbiology**, Cambridge, v. 9, n. 5, p. 511–519, 2006.
- SHENDURE, J.; JI, H. Next-generation DNA sequencing. **Nature Biotechnology**, Seattle, v. 26, n. 10, p. 1135-1145, 2008.
- SOSA-GÓMEZ, D.R.; MOSCARDI, F. Retenção foliar diferencial em soja provocada por percevejos (Heteroptera: Pentatomidae). **Anais da Sociedade Entomológica do Brasil**, Londrina, v. 24, n. 2, p. 401-404, 1995.
- SOSA-GÓMEZ, D.R.; CORSO, I.C.; MORALES, L. Insecticide resistance to endosulfan, monocrotophos and metamidophos in the neotropical brown stink bug, *Euschistus heros* (F.). **Neotropical Entomology**, Londrina, v. 30, p. 317-332, 2001.
- SPANU, P.; GROSSKOPF, D.G.; FELIX, G.; BOLLER, T. The apparent turnover of 1-aminocyclopropane-1-carboxylate synthase in tomato cells is regulated by protein phosphorylation and dephosphorylation. **Plant Physiology**, Basel, v. 106, p. 529–535, 1994.
- SREENIVASULU, N.; SOROPY, S.K.; KISHOR, P.B.K. Deciphering the regulatory mechanisms of abiotic stress tolerance in plants by genomic approaches. **Gene**, Gatersleben, v. 388, p. 1-13, 2007.
- STASKAWICZ, B.J.; AUSUBEL, F.M.; BAKER, B.J.; ELLIS, J.G.; JONES, J.D.G. Molecular genetics of plant disease resistance. **Science**, Berkeley, v. 268, p. 661-667, 1995.
- STILING, P. **Ecology: theories and applications**. 3rd ed. Upper Saddle River: Prentice-Hall, 1999.
- SUN, Z.; TIAN, P.Z.; WANG, J. Study on the uses of aphid-resistant character in wild soybean. I. Aphid-resistance performance of f2 generation from crosses between cultivated and wild soybeans. **Soybean Genetics Newsletter**, Gongzhuling, v. 17, p. 43-48, 1990.
- TANG, F.; BARBACIORU, C.; WANG, Y.; NORDMAN, E.; LEE, C.; XU, N.; WANG, X.; BODEAU, J.; TUCH, B.B.; SIDDIQUI, A.; LAO, K.; SURAN, M.A. mRNA-Seq whole-transcriptome analysis of a single cell. **Nature Methods**, Cambridge, v. 6, p. 377-382, 2009.
- TATSUKI, M.; MORI, H. Phosphorylation of tomato 1-aminocyclopropane-1-carboxylic acid synthase, LE-ACS2, at the C-terminal region. **Journal Biological Chemistry**, Nagoya, v. 276, p. 28051–28057, 2001.
- THOMMA, B.P.H.J.; PENNINCKX, I.A.M.A.; BROEKAERT, W.F.; CAMMUE, B.P.A. The complexity of disease signaling in *Arabidopsis*. **Current Opinion in Immunology**, Heverlee-Leuven, v. 13, p. 63–68, 2001.
- TODD, J.W.; HERZOG, D.C. Sampling phytophagous Pentatomidae on soybean. In: KOGAN, M.; HERZOG, D.C. (Ed.). **Sampling methods in soybean entomology**. New York: Springer-Verlag, 1980. p. 438-478.

TOPHAT a spliced read mapper for RNA-Seq. Disponível em: <<http://tophat.cbcb.umd.edu/manual.shtml>>. Acesso em: 01 nov.2013.

TRAPNELL, C.; ROBERTS, A.; GOFF, L.; PERTEA, G.; KIM, D.; KELLEY, D.R.; PIMENTEL, H.; SALZBERG, S.L.; RINN, J.L.; PACTER, L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, Cambridge, v. 7, n. 3, p. 562-578, 2012.

TURCATTI, G.; ROMIEU, A.; FEDURCO, M.; TAIRI, A.P. A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. **Nucleic Acids Research**, Coinsins, v. 36, n. 25, p. e25, 2008.

TWINE, N.A.; JANITZ, K.; WILKINS, M.R.; JANITZ, M. Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease. **Plos One**, Sydney, v. 6, p. ed16266, 2011.

UKNES, S.; MAUCH-MANI, B.; MOYER, M.; POTTER, S.; WILLIAMS, S.; DINCHER, S.; CHANDLER, D.; SLUSARENKO, A.; WARD, E.; RYALS, J. Acquired resistance in *Arabidopsis*. **Plant Cell**, North Carolina, v. 4, p. 645-656, 1992.

UNITED STATES DEPARTMENT OF AGRICULTURE. **World agricultural supply and demand estimates**. Disponível em: <<http://www.usda.gov/oce/commodity/wasde/latest.pdf>>. Acesso em: 04 fev. 2013.

URANO, K.; KURIHARA, Y.; SEKI, M.; SHINOZAKI, K. Omics analyses of regulatory networks in plant abiotic stress responses. **Current Opinion in Plant Biology**, Tsukuba, v. 13, p. 132-138, 2010.

USADEL, B.; POREE, F.; NAGEL, A.; LOHSE, M.; CZEDIK-EYSENBERG, A.; STITT, M. A guide to using mapman to visualize and compare omics data in plants: a case study in the crop species, maize. **Plant, Cell and Environment**, Potsdam-Golm, v. 32, n. 9, p. 1211-1229, 2009.

USHA RANI, P.; JYOTHSNA, Y. Biochemical and enzymatic changes in rice as a mechanism of defense. **Acta Physiology Plant**, Bhaganagar, v. 32, p. 695-701, 2010.

VALLE, G.E.; LOURENÇÃO, A.L. Resistência de genótipos de soja a *Bemisia tabaco* (Genn.) biótipo b (Hemiptera: Aleyrodidae). **Neotropical Entomology**, Campinas, v. 31, p. 285-295, 2002.

Van der WESTHUIZEN, A.J.; QIAN, X.M.; BOTHA, A.M. Beta-1,3-glucanases in wheat and resistance to the Russian wheat aphid. **Physiologia Plantarum**, Rondebosch, v. 103, p. 125-131, 1998a.

_____. Differential induction of apoplastic peroxidase and chitinase activities in susceptible and resistant wheat cultivars by Russian wheat aphid infestation. **Plant Cell Reports**, Rondebosch, v. 18, p. 132-137, 1998b.

Van VLIET, A.H.M. Next generation sequencing of microbial transcriptomes: challenges and opportunities. **FEMS Microbiology Letters**, Norwich, v. 302, p. 1-7, 2010.

VARSHNEY, R.K.; HOISINGTON, D.A.; TYAGI, A.K. Advances in cereal genomics and applications in crop breeding. **Trends in Biotechnology**, Patancheru, v. 24, p. 490–499, 2006.

VEIGA, R.F.A.; ROSSETTO, C.J.; RAZERA, L.F.; GALLO, P.B.; BORTOLETO, N.; MEDINA, P.F.; TISSELLI FILHO, O.; CIONE, J. **Caracterização morfológica e agrônômica do cultivar de soja IAC-100**. São Paulo: Instituto Agronômico de Campinas, 1999. 23 p. (IAC. Boletim Técnico, 177).

VENTURA, M.U.; PINHEIRO, J.B. Resistência a insetos, In: DESTRO, D.; MONTALVÁN, R. (Org.). **Melhoramento genético de plantas**. Londrina: Ed. UEL, 1999. p. 467-516.

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, p. 70-78, 1999.

VERHAGE, A.; VAN WESS, S.C.M.; PIETERSE, C.M.J. Plant immunity: it's the hormones talking, but what do they say? **Plant Physiology**, Utrecht, v. 154, p. 536-540, 2010.

VERNETTI, F.J. **Soja: genética e melhoramento**. Campinas: Fundação Cargill, 1983. v. 2.

VICENTINI, R.; JIMENEZ, H.A. **El vanejo de los frutos em soja**. Oliveros: INTA, 1977. 47 p. (INTA Serie Técnica).

VILLAS-BÔAS, G.L.; GAZZONI, D.L.; OLIVEIRA, M.C.N.; COSTA, N.P.; ROESSING, A.C.; FRANÇA-NETO, J.B.; HENNING, A. Efeito de diferentes populações de percevejos sobre o rendimento e seus componentes, características agrônômicas e qualidade de semente de soja. **Boletim de Pesquisa Embrapa, CNPSo**, Londrina, v. 1, p. 43, 1990.

WALLEY, J.W.; DEHESH, K. Molecular mechanisms regulating rapid stress signalling networks in *Arabidopsis*. **Journal of Integrative Plant Biology**, San Diego, v. 52, n. 4, p. 354-359, 2010.

WANG, K.L; LI, C.; JUNIOR ECKER, H. Ethylene biosynthesis and signaling networks. **Plant Cell**, La Jolla, v. 14, p. S131–S151, 2002.

WANG, X.; SUN, Q.; MACGRATH, S.D.; MARDIS, E.R.; SOLOWAY, P.D.; CLARK, A.G. Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. **Plos One**, New York, v. 3, p. e3839, 2008.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Review of Genetics**, New Haven, v. 10, n.1, p. 57-63, 2009.

WAR, A.R.; PAULRAJ, M.G.; WAR, M.Y. Ignacimuthu S. Jasmonic acid- mediated induced resistance in groundnut (*Arachis hypogaea* L.) against *Helicoverpa armigera* (Hubner) (Lepidoptera: Noctuidae). **Journal Plant Growth Regulation**, Chennai, v. 30, p. 512-523, 2011a.

_____. Ignacimuthu S. Herbivore- and elicitor-induced resistance in groundnut to Asian armyworm, *Spodoptera litura* (Fab.) (Lepidoptera: Noctuidae). **Plant Signal & Behavior**, Chennai, v. 6, p. 1769-1777, 2011b.

WAR, A.R.; PAULRJ, M.G.; AHMAD, T; BUHROO, A.A.; HUSSAIN, B.; SHARMA, H.C. Mechanisms of plant defense against insect herbivores. **Plant Signaling & Behavior**, Patancheru, v. 7, p. 1306-1320, 2012.

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.

WOLD, B.; MYERS, R.M. Sequence census methods for functional genomics. **Nature Methods**, Pasadena, v. 5, p. 19-21, 2008.

WORKMAN, C.; JENSEN, L.J.; JARMER, H.; BERKA, R.; GAUTIER, L.; NIELSER, H.B.; SAXILD, H.H.; NIELSEN, C.; BRUNAK, S.; KNUDSEN, S. A new non-linear normalization method for reducing variability in DNA microarray experiments. **Genome Biology**, Basel, v. 3,p. 48, 2002.

YANG, S.F.; HOFFMAN, N.E. Ethylene biosynthesis and its regulation in higher plants. **Annual Review of Plant Physiology**, Davis, v. 35, p. 155–189, 1984.