

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

Identificação de QTLs em soja associados à resistência aos percevejos e a caracteres agronômicos utilizando a abordagem de mapeamento multivariado

Milene Möller

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

Piracicaba
2017

Milene Möller
Bióloga

Identificação de QTLs em soja associados à resistência aos percevejos e a caracteres agronômicos utilizando a abordagem de mapeamento multivariado
versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

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

Piracicaba
2017

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

Möller, Milene

Identificação de QTLs em soja associados à resistência aos percevejos e a caracteres agrônômicos utilizando a abordagem de mapeamento multivariado / Milene Möller. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2017.

131 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. *Glycine max* 2. Pentatomídeos fitófagos 3. Mapeamento de QTLs 4. Análise multivariada 5. Marcador molecular 6. Melhoramento de plantas I. Título

DEDICATÓRIA

Ao meu querido esposo **Maurício**,
pelo amor, companheirismo,
compreensão e incentivo,
dedico.

Aos meus pais, **Márcia e Gilson**,
pelo amor incondicional e
pelo exemplo de vida,
ofereço.

AGRADECIMENTOS

A Deus...

por permitir que eu superasse todos os obstáculos e por iluminar meu caminho para mais essa conquista;

À Família...

Ao meu querido esposo Maurício Terasawa, que de maneira inesperada completou a minha vida; Obrigada por todo o amor, companheirismo e me incentivar sempre a buscar novos desafios; Agradeço também pelas suas contribuições que foram fundamentais para a finalização dessa tese;

Aos meus pais, Gilson e Márcia, pelos princípios ensinados e por estarem sempre ao meu lado, me apoiando em todas as decisões; Obrigada por todo o incentivo e por me ajudarem a chegar onde estou;

Ao meu irmão Rodrigo e à minha cunhada Bianca, pelo apoio e carinho e por terem me presenteado com minhas sobrinhas Clara e Giovanna, duas princesas que alegram a minha vida;

À minha querida avó Ivone, pela sua alegria sempre contagiante; à tia Marcelina, pelo carinho; aos meus queridos tios, tias, primos e primas, por todo o incentivo e apoio; agradeço, em especial, à Tia Marli, por todas as orações e carinho sempre;

À minha segunda família, que Deus me presenteou, minha sogra Ana Cirte, meu sogro Francisco Terasawa, minha cunhada Daniela, meu cunhado Juliano e minha querida sobrinha Gabriela, uma princesa que desde o início encantou meu coração; Obrigada por estarem sempre presentes, pelo carinho e incentivo;

Ao meu avô Alvino Möller (*in memorian*), à minha bisavó Segunda (*in memorian*), ao meu avô Aristides (*in memorian*) e à minha avó Guilhermina Nair (*in memorian*), que estarão sempre no meu coração;

Aos Mestres...

Ao meu orientador, Prof. Dr. José Baldin Pinheiro, pela oportunidade, ensinamentos e confiança;

À Profa. Dra. Maria Imaculada Zucchi, pela co-orientação, amizade e ensinamentos;

A todos os membros da banca examinadora e suplentes, por disponibilizarem seus conhecimentos para essa defesa;

Ao Prof. Dr. Brian Diers, por me receber em seu laboratório durante o doutorado sanduiche;

Agradeço também...

Ao Prof. Dr. Antonio Augusto Franco Garcia e ao pós-doutorando Dr. João Ricardo Bachega Feijó Rosa, pela disponibilidade e prontidão em me ajudar e pelas valiosas contribuições nas análises;

Ao pós-doutorando Dr. Marcelo Mollinari, pela ajuda fundamental e ensinamentos;

À Dra. Michelle da Fonseca Santos, pela ajuda fundamental, e ao Dr. Steve Clough, pela amizade e ensinamentos;

Aos colegas de trabalho do Laboratório de Diversidade Genética e Melhoramento, que, de alguma forma, contribuíram para alguma etapa deste trabalho; Em especial à querida Diane, por sempre demonstrar sua alegria e otimismo;

À Escola Superior de Agricultura “Luiz de Queiroz” pela excelente qualidade de ensino e, em especial, ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, pela oportunidade em realizar o doutorado;

A todos os professores do Departamento de Genética da ESALQ - USP, pelos ensinamentos e conhecimentos compartilhados;

A todos os funcionários do Departamento de Genética, em especial ao Domingos de Sálvio Amaral, Márcio Araújo Silva, Ronaldo José Rabello, Cláudio Roberto Segatelli, José Monteiro, Berdan, Léia e Fernandinho, pela competência e compromisso com o trabalho; E aos funcionários da biblioteca central, em especial à Silvia e à Eliana, pela ajuda nessa etapa final;

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fulbright, pelo apoio financeiro concedido; E à FAPESP, pelo apoio financeiro concedido para a realização do projeto;

À FT Sementes, pela oportunidade de trabalho e crescimento profissional concedidos, e aos colegas da empresa, pelo apoio nessa etapa final;

A todos que de alguma forma contribuíram para o êxito deste trabalho, mesmo que aqui não mencionados, meus sinceros agradecimentos.

EPÍGRAFE

“O que prevemos raramente ocorre

O que menos esperamos

geralmente acontece”

Benjamin Disraeli

*“O cientista não é o homem que
fornece as verdadeiras respostas;
é quem faz as verdadeiras perguntas”*

Claude Lévi-Strauss

SUMÁRIO

RESUMO.....	9
ABSTRACT	10
1. INTRODUÇÃO	11
2. REVISÃO BIBLIOGRÁFICA.....	13
2.1. SOJA.....	13
2.2. PERCEVEJOS	15
2.3. MECANISMOS DE RESISTÊNCIA E CULTIVARES DE SOJA RESISTENTES A OS PERCEVEJOS	21
2.4. MARCADORES MOLECULARES EM SOJA	24
2.5. MAPEAMENTO DE QTLs.....	26
2.6. QTLs ASSOCIADOS À RESISTÊNCIA A INSETOS EM SOJA.....	28
2.7. ESTUDOS DE MAPEAMENTO DE QTLs PREVIAMENTE DESENVOLVIDOS	31
3. MATERIAL E MÉTODOS.....	33
3.1. MATERIAL VEGETAL.....	33
3.2. DADOS FENOTÍPICOS DA GERAÇÃO $F_{2:3}$	33
3.2.1. Análises Fenotípicas.....	34
3.3. EXTRAÇÃO DO DNA.....	35
3.4. DADOS GENOTÍPICOS DA GERAÇÃO F_2	35
3.4.1. Marcadores SNP	35
3.4.2. Marcadores SSR.....	36
3.4.3. Marcadores AFLP	37
3.4.4. Marcadores TRAP	37
3.5. MAPA GENÉTICO	37
3.6. MAPEAMENTO DE QTLs.....	38
4. RESULTADOS E DISCUSSÃO	43
4.1. ANÁLISES FENOTÍPICAS	43
4.2. ANÁLISES GENOTÍPICAS	49
4.3. MAPA GENÉTICO	49
4.4. MAPEAMENTO DE QTLs.....	53

4.4.1. Mapeamento MIM.....	54
4.4.2. Mapeamento MT-MIM.....	63
5. CONSIDERAÇÕES FINAIS	105
REFERÊNCIAS	107
APÊNDICES	127
ANEXOS	129

RESUMO

Identificação de QTLs em soja associados à resistência aos percevejos e a caracteres agronômicos utilizando a abordagem de mapeamento multivariado

A soja é a cultura agrícola brasileira que mais se expandiu nas últimas três décadas. Atualmente é uma *commodity* comercializada em larga escala na forma de grão, farelo e óleo. Por ser uma importante fonte de proteína, possui papel fundamental na indústria alimentícia, tanto humana quanto animal. O monocultivo em extensas áreas tem ocasionado o aumento da vulnerabilidade da cultura a patógenos e insetos-praga, com consequências relevantes na produção dos grãos. Dentre tais insetos-praga, os percevejos fitófagos representam um dos grupos de maior relevância para a cultura, pois podem comprometer o rendimento, a qualidade e a sanidade dos grãos. As perdas no rendimento da cultura devido ao ataque por percevejos são superiores a 30%, e o comprometimento no valor germinativo das sementes pode ser superior a 50%. Este trabalho teve como objetivo identificar QTLs (*Quantitative Trait Loci*), a partir de um conjunto de fenótipos associados à resistência aos percevejos e a caracteres agronômicos, utilizando a abordagem de mapeamento multivariado. Uma população de 228 indivíduos F₂, obtida a partir do cruzamento entre as cultivares IAC-100 e CD-215, foi utilizada para a obtenção dos dados genotípicos. Os caracteres agronômicos avaliados na geração F_{2:3} foram: número de dias para o florescimento (NDF), altura da planta no florescimento (APF), número de dias para a maturidade (NDM), altura da planta na maturidade (APM), acamamento (AC), valor agronômico (VA) e produtividade de grãos (PG). Oito caracteres associados à resistência aos percevejos foram avaliados: período de granação (PEG), retenção foliar (RF), número de vagens por planta (NVP), índice percentual de dano nas vagens (IPDV), número de sementes (NS), peso de cem sementes (PCS), peso de sementes boas (PSB) e peso de sementes manchadas (PSM). O mapa genético obtido, representando os 20 grupos de ligação (GLs) da soja, foi constituído por 417 marcadores SNP (*Single Nucleotide Polymorphism*), 61 SSR (*Simple Sequence Repeat*), 30 AFLP (*Amplified Fragment Length Polymorphism*) e 8 marcadores TRAP (*Target Region Amplification Polymorphism*). A cobertura do genoma da soja foi de 2.814,82 centiMorgans (cM), com um intervalo médio de 5,46 cM entre marcadores adjacentes. A identificação de QTLs (*Quantitative Trait Loci*) para os caracteres fenotípicos foi realizada utilizando-se o mapeamento de intervalos múltiplos univariado (MIM) e multivariado (MT-MIM), com estimativa dos efeitos principais dos QTLs. A abordagem MIM revelou um total de 60 QTLs, distribuídos por 13 GLs da soja, sendo 29 QTLs associados a caracteres de resistência aos percevejos e 31 QTLs relacionados a caracteres agronômicos. A percentagem da variação fenotípica explicada pelos QTLs (R^2) variou de 14,27% para AC a 65,45% para NDM. Na abordagem MT-MIM, foram selecionados nove conjuntos de variáveis, de acordo com a correlação entre as mesmas. Foram reveladas 20 regiões genômicas distintas, com uma alta tendência de concentração de QTLs em posições similares. No geral, para a maioria das características, os valores marginais de R^2 obtidos para os modelos MT-MIM foram superiores em relação aos modelos MIM, variando de 27,98% para APF a 65,30% para NDM. A abordagem MT-MIM permitiu a identificação de 13 novas posições genômicas, com efeito em pelo menos uma das variáveis analisadas, que não foram reveladas nos modelos MIM. Uma comparação com o banco de dados do Soybase forneceu evidências de que muitos QTLs, identificados nesta população em estudo, coincidem com QTLs descritos em outros *backgrounds* genéticos. No entanto, 56 QTLs identificados no presente estudo ainda não foram descritos na literatura. A maioria dos QTLs identificados explicam, individualmente, até 10% da variação fenotípica das características avaliadas. No entanto, QTLs presentes em oito novas regiões identificadas pela abordagem MT-MIM e oito novos QTLs identificados pela abordagem MIM contribuíram para explicar uma maior percentagem da variação dos fenótipos estudados. Esses QTLs devem ser melhor investigados considerando sua relevância para a seleção simultânea de características de interesse, permitindo uma maior eficiência de seleção e um maior ganho genético. Os resultados obtidos no presente estudo forneceram informações para auxiliar na melhor compreensão da arquitetura genética dos caracteres quantitativos analisados, bem como sobre a relação genética entre os mesmos.

Palavras-chave: *Glycine max*; Pentatomídeos fitófagos; Mapeamento de QTLs; Análise multivariada; Marcador molecular; Melhoramento de plantas

ABSTRACT

QTL identification in soybean related to stink bug resistance and agronomic traits using the multivariate multiple interval mapping approach

Soybean is the Brazilian crop with the most expansion along the past three decades. Currently, it is a commodity commercialized in large scale as grain, bran and oil. Because it is an important source of protein, it plays a fundamental role in the food industry, both human and animal. Soybean monoculture in large areas has increased crop vulnerability to pathogens and insect pests, with significant consequences on grain production. Among such pest insects, stink bugs are considered a major pest of soybean crop, feeding directly on seeds, reducing yield and seed quality. Losses in crop yield due to stink bugs attack are greater than 30% and seed germination compromising can be greater than 50%. This study aimed to identify QTL (*Quantitative Trait Loci*), for stink bug resistance traits and agronomic traits using the multivariate multiple interval mapping. An F₂ mapping population of 228 plants derived from a biparental cross between IAC-100 and CD-215 was used for genotyping. An F_{2:3} population was developed to evaluate eight stink bug resistant traits such as graining period, leaf retention, pod number per plant, percentage of pod damage, number of seeds, hundred seed weight, weight of healthy seeds and spotted seed weight. Other seven agronomic traits were evaluated such as number of days to flowering, plant height at flowering, number of days to maturity, plant height at maturity, lodging, agronomic value and grain yield. A total of 417 SNP (*Single Nucleotide Polymorphism*) markers, 61 SSR (*Simple Sequence Repeat*), 30 AFLP (*Amplified Fragment Length Polymorphism*) and 8 TRAP (*Target Region Amplification Polymorphism*) markers were mapped into 20 soybean linkage groups. The total map length was 2,814.82 cM with an average of 5.46 cM between markers. QTL mapping for those traits was performed using univariate (MIM) and multivariate (MT-MIM) multiple interval mapping, with main QTL effects estimates. MIM analysis identified a total of 60 QTL, through 13 soybean linkage groups, with 29 QTL related to stink bug resistant traits and 31 QTL related to agronomic traits. Phenotypic variation explained by QTL ranged from 14.27% for lodging to 65.45% for number of days to maturity. The traits were divided into nine groups for MT-MIM analysis considering their correlation coefficient. Twenty different genomic regions were identified showing a very high QTL clustering. For most of the traits phenotypic variation estimates for MT-MIM models were higher than MIM models, ranging from 27.98% to 65.30% for plant height at flowering and number of days to maturity, respectively. MT-MIM analysis showed 13 genomic regions controlling at least one of the evaluated traits which were not identified at MIM analysis. Comparison between identified QTL and QTL database at Soybase demonstrated that some QTL were similar to those described in different genetic background. However, 56 QTL detected in the present study were described for the first time in literature. Most of the QTL identified explain, individually, less than 10% of phenotypic variation. However, eight genomic regions identified with MT-MIM analysis and eight QTL identified with MIM analysis explain a great amount of phenotypic variation. These QTL should be investigated considering their importance for simultaneous selection for a high genetic gain. Results obtained in the present study provided information for a better understand of genetics architecture underlying quantitative traits studied and the genetic relation among them.

Keywords: *Glycine max*; Phytophagous pentatomids; QTL Mapping; Multivariate analysis; Molecular marker; Plant breeding

1. INTRODUÇÃO

A Soja [*Glycine max* (L.) Merrill] é um dos cultivos mais importantes e valiosos para a humanidade. No Brasil, é a cultura agrícola que mais se expandiu nas últimas três décadas, representando 49% da área plantada em grãos no país (CONAB 2017). As mudanças no cenário agrícola nas diversas regiões produtoras de soja do Brasil têm sido frequentes. A expansão do monocultivo da soja em extensas áreas tem ocasionado o aumento da vulnerabilidade da cultura a patógenos e insetos-praga, com consequências relevantes na produção dos grãos (Moura 1999; Boerma and Walker 2005). Dentre as pragas mais importantes, destacam-se os percevejos, que podem comprometer o rendimento, a qualidade e a sanidade dos grãos (Panizzi and Slansky Júnior 1985; Villas-Bôas et al. 1990; Brier and Rogers 1991; Panizzi et al. 2000; Corrêa-Ferreira and Azevedo 2002). Estudos demonstram que as perdas no rendimento da cultura devido ao ataque por percevejos são superiores a 30%, e o comprometimento no valor germinativo das sementes pode ser superior a 50% (Degrande and Vivan 2009; Soria et al. 2010). Estes insetos-praga têm uma significativa importância em termos de custo de produção e controle, além de ocasionarem redução na produtividade de grãos (Musser et al. 2011). Comparando-se as safras 2009/2010 e 2014/2015, houve um aumento significativo de 6,04% para 13,8% do custo total de produção associado às despesas com inseticidas, para o controle de percevejos na cultura da soja (Lantmann 2014). Além disso, há uma crescente dificuldade no controle destes insetos, decorrente da sua diversidade e densidade populacional elevada, tanto nas fases mais vulneráveis e críticas da soja, quanto nos demais períodos em que convivem com a cultura da soja (Guedes et al. 2012).

O manejo de percevejos tem evoluído em decorrência da crescente expansão da soja no Brasil, caracterizada pelo avanço e adaptação da cultura em novas fronteiras agrícolas. O cultivo em vários meses do ano também tem caracterizado as denominadas “pontes verdes hospedeiras”, que abrigam e multiplicam hospedeiros, sejam pragas ou plantas daninhas, durante a entressafra de outras regiões produtoras. Além disso, o impacto do plantio direto na preservação de algumas espécies e de seus inimigos naturais, a utilização desorientada e desenfreada de fungicidas e inseticidas e o aumento na utilização de cultivares indeterminadas têm contribuído para as constantes mudanças nas estratégias de controle.

No Brasil, três espécies de percevejos são mais comumente encontradas: o percevejo marrom (*Euschistus heros* Fabricius), o percevejo-verde-pequeno (*Piezodorus guildinii* Westwood) e o percevejo-verde (*Nezara viridula* Linnaeus). A variação populacional desses insetos pode ocorrer de acordo com as condições ambientais, com o período de semeadura e, até mesmo, com a cultivar de soja utilizada no cultivo (Magrini et al. 1996; Corrêa-Ferreira and Peres 2003). Entretanto, nas últimas safras, têm sido relatada a alta frequência e dispersão da espécie *E. heros*, na maioria das regiões produtoras de soja do Brasil (Corrêa-Ferreira et al. 2009; Panizzi et al. 2012; Sosa-Gómez and Omoto 2012).

As formas adultas e juvenis de percevejos alimentam-se diretamente dos grãos, ocasionando perdas significativas de rendimento, qualidade e potencial germinativo da semente, além de dificultarem a colheita mecanizada (Panizzi and Slansky Júnior 1985). Além disso, durante o processo de alimentação, podem transmitir patógenos que ocasionam a deterioração das sementes (Gazzoni 1998; Corrêa-Ferreira and Azevedo 2002). A aplicação de inseticidas tem sido o método mais utilizado para controle de percevejos na cultura da soja (Musser et al. 2011). Entretanto, a utilização excessiva e desenfreada de inseticidas pode ocasionar desequilíbrio ambiental e, conseqüentemente, eliminar as populações de inimigos naturais, bem como selecionar populações de insetos resistentes aos ingredientes ativos hoje utilizados (Byrne et al. 2003). Dessa forma, o investimento na característica de resistência genética a esses insetos deve ser valorizado, considerando os benefícios econômicos e ambientais, em virtude da redução da utilização de inseticidas, benefícios ecológicos, por reduzir o nível de agressão ao meio ambiente, e benefício direto aos agricultores.

O mapeamento de QTLs é uma técnica importante utilizada em programas de melhoramento genético para auxiliar melhoristas na identificação e seleção eficientes de plantas contendo os alelos desejáveis. Considerando o caráter poligênico das características envolvidas na resistência da soja aos percevejos, bem como da maioria das características de interesse agrônomo, a identificação de QTLs associados a essas características é de grande relevância visando a elucidação da base genética das mesmas. Diversos estudos foram publicados descrevendo QTLs para resistência a insetos em soja (Rector et al. 1998; Rector et al. 1999; Rector et al. 2000; Terry et al. 2000; Narvel et al. 2001; Hulburt et al. 2004; Walker et al. 2004; Boerma and Walker 2005; Guo et al. 2005; Komatsu et al. 2005; Guo et al. 2006; Zhu et al. 2006; Zhao et al. 2008; Zhu et al. 2008; Yesudas et al. 2010). Atualmente, somente os trabalhos desenvolvidos por Möller (2010) e Santos (2012), envolvendo o mapeamento de QTLs associados às características de resistência a percevejos, foram encontrados na literatura. De acordo com os resultados obtidos por Santos (2012), foi possível observar uma alta correlação genética entre as características associadas à resistência aos percevejos.

Nenhum trabalho foi encontrado na literatura utilizando a abordagem multivariada, em soja. A identificação de regiões genômicas controlando um conjunto de variáveis contribuirá para facilitar a seleção de genótipos elite, permitindo maiores ganhos de seleção para diversas características simultaneamente. Assim, a utilização de uma estratégia de mapeamento para múltiplas características simultaneamente, inédita para a espécie, pode fornecer informações para melhor compreender a arquitetura genética destes fenótipos, além de permitir um maior poder de detecção e estimativas mais precisas dos efeitos dos QTLs. Além disso, a identificação de QTLs afetando um conjunto de variáveis é desejável em um programa de melhoramento, por aumentar a eficiência de seleção e permitir maiores ganhos de seleção com a utilização da seleção simultânea de características.

Assim, o presente trabalho teve como objetivo identificar QTLs, a partir de um conjunto de fenótipos relacionados à resistência aos percevejos e a caracteres de importância agrônoma, em uma população $F_{2,3}$, de soja, utilizando a abordagem de mapeamento multivariado.

2. REVISÃO BIBLIOGRÁFICA

2.1. Soja

A soja cultivada comercialmente [*Glycine max* (L) Merrill] ($2n= 40$) é uma planta herbácea, autógama, cuja composição taxonômica se insere na divisão Magnoliophyta, classe Magnoliopsida, sub-classe Rosidae, ordem Fabales, família Fabaceae, subfamília Papilionoideae, tribo Phaseoleae, gênero *Glycine* (NCBI 2017).

O centro de origem da soja é o norte da China, que apresentava plantas selvagens com uma arquitetura que difere significativamente da soja cultivada atualmente. As plantas rasteiras se desenvolviam ao longo de rios e lagos chineses. O seu processo de domesticação ocorreu no século XI a.c., a partir de cruzamentos realizados por cientistas chineses, entre duas espécies selvagens (Sedyama 2009). De acordo com os registros históricos, a soja ficou restrita à China por muito tempo e só no século III d.C. foi difundida para a Coreia, Japão e países do sudeste da Ásia.

Estudos de diversidade genética realizados entre *G. max* e seu ancestral selvagem, *G. soja*, demonstram que a soja cultivada comercialmente apresenta uma estreita base genética, resultante do seu processo de domesticação (Hyten et al. 2006; Stupar 2010; Li et al. 2010b; Joshi et al. 2013; Hu et al. 2014; Zhang et al. 2016). Apesar das diferenças fenotípicas entre as espécies, Singh and Hymowitz (1988) demonstraram que é possível obter híbridos férteis, a partir do cruzamento entre as mesmas. Dessa forma, *G. soja* torna-se uma fonte promissora de genes e alelos, que podem ter sido perdidos durante o processo de domesticação, contribuindo para o processo de melhoramento da soja (Hu et al. 2014; Zhang et al. 2016; Lu et al. 2017).

A utilização da soja como alimento data de mais de 5.000 anos, quando o imperador chinês Shen-nung, considerado o “pai” da agricultura chinesa, iniciou o cultivo de grãos como alternativa ao abate de animais (APROSOJA, 2016a). Além de muito utilizada na alimentação para a produção de tofu, leite de soja, queijo, pão e óleo, a soja era utilizada como moeda, na troca por outras mercadorias. A sua introdução no Ocidente ocorreu no período das grandes navegações européias, entre o final do século XV e início do século XVI, sendo utilizada inicialmente como forrageira. No entanto, seu alto teor de óleo e proteína fez com que o grão fosse rapidamente valorizado, aumentando sua produção e tornando-se um item de comércio exterior de destaque no final da Primeira Guerra Mundial, em 1919. Foi então, em 1941, que a área ocupada para o cultivo de grãos superou a área para uso como forragem. A utilização da soja como forragem praticamente se extinguiu em meados dos anos 60, enquanto a produção de grãos crescia de forma exponencial (Embrapa 2003).

Atualmente, a soja é uma das culturas de maior importância econômica do mundo. Tornou-se uma *commodity* e é comercializada em larga escala na forma de grão, farelo e óleo. Sua utilização na indústria é bastante diversificada, sendo a matéria-prima em processos de produção nas indústrias de cosméticos, farmacêuticas, veterinárias, de vernizes de tintas e de plásticos, adesivos, adubos, entre outras. No entanto, é utilizada com maior destaque na indústria alimentícia, tanto humana quanto animal, devido ao seu alto teor de proteínas. No Brasil, 80% do farelo de soja é destinado à fabricação de ração animal, que também é composta por farelo de milho (APROSOJA 2016b). Já no caso da alimentação humana, o seu uso tem sido crescente e está associado aos benefícios propiciados à saúde como a redução de doenças cardiovasculares e derrame cerebral, além de ser considerada como uma ótima aliada ao antienvhecimento por ser rica em antioxidantes. Além disso, nos últimos anos, diversos produtos alimentícios à base de soja têm sido desenvolvidos, considerando o crescente número de casos de intolerância à proteína do leite. O alto teor de óleo presente no grão faz com que grande parte da produção seja destinada à fabricação de biocombustível, o combustível renovável que contribui para a redução da emissão de gases poluentes no meio ambiente. Os dois

principais biocombustíveis líquidos utilizados no Brasil são o etanol, obtido a partir da cana-de-açúcar, e o biodiesel, que é produzido a partir de óleos vegetais ou de gorduras animais e adicionado ao diesel de petróleo em proporções variáveis. A soja possui um papel relevante neste cenário, sendo responsável por mais de 82% da produção de biodiesel no Brasil (APROBIO 2016). O Brasil foi pioneiro no uso de biocombustíveis e ocupa uma posição de destaque na substituição do petróleo por fontes renováveis de energia. De acordo com a Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), cerca de 45% da energia e 18% dos combustíveis consumidos no Brasil já são renováveis, enquanto no restante do mundo 86% da energia ainda é dependente de fontes energéticas não renováveis (ANP 2016).

A produção mundial de soja na safra 2015-16 foi de 313 milhões de toneladas (USDA 2017). Os três maiores produtores são os Estados Unidos, o Brasil e a Argentina, que, juntos, representam 83% da produção mundial. Os Estados Unidos, maior produtor mundial de soja, atingiu produção de 107 milhões de toneladas na safra 2016, com estimativa de produção de 117 milhões de toneladas para a safra 2017 (USDA 2017). O Brasil, segundo maior produtor, produziu quase 95,5 milhões de toneladas na safra 2015-16, e a estimativa para a safra 2016-17 é de um aumento de 7,8 a 10% em relação à safra anterior, com produção variando de 103 a 105 milhões de toneladas (CONAB 2017). A área plantada com soja na safra brasileira de 2016-17, segundo a estimativa da Companhia Nacional de Abastecimento (CONAB), foi de 33,7 milhões de hectares, o que corresponde a um crescimento de 1,5% em relação à safra 2015-16.

A soja é a cultura agrícola brasileira que mais se expandiu nas últimas três décadas, representando, atualmente, 49% da área plantada em grãos do país (CONAB 2017). O Brasil é autossuficiente na produção de soja para o mercado interno e o excedente movimentou cerca de 28 bilhões de dólares por ano (SECEX 2017). A soja tem grande peso na balança comercial, contribuindo com quase 16% das exportações totais do país (SECEX 2017). A maior parte da soja brasileira exportada é adquirida pela China, que, em 2016, representou 59% das exportações do complexo soja do Brasil, seguida da União Européia com 19% (SECEX 2017). No entanto, considerando apenas a soja exportada como grão, farelo ou óleo, os maiores consumidores da soja brasileira exportada são Chi na (75%), União Européia (55 %) e Ásia (excluindo-se a China) (54 %), respectivamente. Atualmente, o Brasil é o segundo maior exportador do complexo soja, permanecendo atrás apenas dos Estados Unidos. De acordo com a Organização das Nações Unidas para a Alimentação e a Agricultura (FAO), em 2025, o Brasil deve alcançar o posto de maior exportador de soja, considerando a constante desvalorização do real, o que torna as exportações brasileiras mais competitivas (FAO 2017).

O primeiro registro da soja no Brasil é de 1882, na Bahia, quando o Professor Gustavo Dutra conduziu os primeiros estudos de avaliação de cultivares introduzidas dos Estados Unidos, na Escola de Agronomia da Bahia (Embrapa 2003). Inicialmente, não houve uma boa adaptação da soja às condições locais, pois, além das exigências hídricas e térmicas, a soja é sensível ao fotoperíodo, sendo considerada uma planta de dias curtos, o que inicialmente limitou a sua ampla difusão por todo o país (Sedyama 2009). Novas tentativas de cultivo foram realizadas, em 1891, no Instituto Agrônomo de Campinas, no estado de São Paulo, onde houve maior sucesso no seu desenvolvimento. No ano de 1900, a soja foi introduzida no Rio Grande do Sul (RS). Como os primeiros materiais de soja introduzidos no Brasil eram provenientes dos Estados Unidos, houve uma maior adaptação da planta devido à semelhança de latitude e das condições climáticas predominantes nas duas regiões. Foi então que, em 1914, registrou-se o primeiro plantio de soja no Brasil, no município de Santa Rosa, no RS. A soja consolidou-se como a principal cultura do agronegócio brasileiro somente em meados da década de 70, devido, principalmente, ao aumento da demanda mundial do produto e, também, ao incremento de novas tecnologias no processo de produção, proporcionando o aumento da produtividade nos cultivos (Missão 2006). No entanto, apesar das novas cultivares mais adaptadas, o cultivo da soja ainda se limitava à região sul do Brasil, onde prevalecia o cultivo do trigo no inverno e da soja no verão. Na época,

houve um grande avanço de produção de 1,5 milhões de toneladas, em 1970, para 15 milhões de toneladas, em 1979. Esse rápido aumento ocorreu tanto pelo aumento da área plantada, que passou de 1,3 para 8,8 milhões de hectares, mas também pelo expressivo incremento de produtividade (1.140 para 1.730 kg.ha⁻¹). Foi então que a FT-Cristalina, uma das primeiras cultivares lançadas pela FT PESQUISA & SEMENTES, constituiu-se em um dos principais pilares para tornar economicamente viável a produção de soja no Cerrado brasileiro, entre o final da década de 70 e o início da década de 90 (Terasawa 1982). Essa cultivar teve grande importância para os produtores e para a agricultura brasileira, sendo denominada, na época, de “Rainha do Cerrado”, pois chegou a constituir 95% da área cultivada com soja na região do Cerrado, entre o final da década de 70 e início da década de 80. Além disso, teve significativa contribuição para o incremento da produtividade brasileira, alcançando a produtividade inédita de 4.780 kg.ha⁻¹, no ano agrícola 1978/79, em experimentos na Fazenda Vereda, em Cristalina, conforme dados do Centro de Pesquisa Agropecuária dos Cerrados da Embrapa, órgão vinculado ao Ministério da Agricultura e Abastecimento (Gilioli et al. 1988). A soja foi, então, se expandindo para toda a região Centro-Oeste e, mais recentemente, para a Região Norte e Nordeste do Brasil. O uso de alta tecnologia de produção contribuiu para essa expansão da indústria da soja no país. Atualmente, o estado com maior produção é o Mato Grosso, com 28% da produção total do país, seguido do Paraná e Rio Grande do Sul, com 17,6 e 17%, respectivamente (CONAB 2017). A região conhecida como MATOPIBA, representada pelos Estados do Maranhão, Tocantins, Piauí e Bahia, é a que tem apresentado maior potencial de expansão da área cultivada com soja nos últimos anos.

No entanto, a expansão do monocultivo da soja em extensas áreas tem ocasionado o aumento da vulnerabilidade da cultura a patógenos e insetos-praga, com consequências relevantes na produção dos grãos (Moura 1999; Boerma and Walker 2005). A soja está sujeita ao ataque de diversas pragas e patógenos, de acordo com a região e época de cultivo. No entanto, em todas as 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 et al. 1999; Panizzi et al. 2012).

2.2. Percevejos

O aumento da área de produção de soja nos últimos anos foi acompanhado por um crescente incremento no número de pragas que atacam a cultura. Dentre tais pragas, que correspondem a diferentes grupos de insetos, os percevejos fitófagos representam um dos grupos de maior relevância para a cultura, pois podem comprometer o rendimento, a qualidade e a sanidade dos grãos (Panizzi and Slansky Júnior 1985; Villas-Bôas et al. 1990; Brier and Rogers 1991; Panizzi et al. 2000; Corrêa-Ferreira and Azevedo 2002).

Os percevejos são insetos sugadores e pertencem à família Pentatomidae. No Brasil, destaca-se um complexo de espécies, reconhecidos em ordem decrescente de importância econômica e de ocorrência nos últimos anos: o percevejo-marrom [*Euschistus heros* (F.)], o percevejo-verde-pequeno [*Piezodorus guildinii* (West.)], o percevejo-verde [*Nezara viridula* (L.)], e os percevejos-barriga-verde [*Dichelops* spp.]. Embora as espécies *Chinavia* sp. e *Edessa mediotubunda* tenham aumentado consideravelmente sua ocorrência nas lavouras de soja, ainda são consideradas de menor expressão (Panizzi and Slansky Júnior 1985; Corrêa-Ferreira et al. 2009; Panizzi et al. 2012; Sosa-Gómez and Omoto 2012).

A variação populacional e predominância das espécies de percevejos observadas em regiões e safras distintas estão diretamente relacionadas com a cultivar semeada, com o período de semeadura e com as condições climáticas locais (Cividanes and Parra 1994; Magrini et al. 1996). Além disso, os danos ocasionados são influenciados pela

densidade populacional das espécies, pela fase em que a cultura se encontra e pelo período (número de dias) em que os insetos permanecem se alimentando dos grãos (Cividanes and Parra 1994; Corrêa-Ferreira and Peres 2003).

A espécie *E. heros* tem apresentado maior dispersão e frequência em relação às demais espécies de percevejos na maioria das regiões produtoras de soja do Brasil (Corrêa-Ferreira et al. 2009; Panizzi et al. 2012; Sosa-Gómez and Omoto 2012). Além disso, *E. heros* também tem se destacado como uma praga de relevância econômica em regiões produtoras de soja da Argentina (Saluso et al. 2011). O rápido aumento da ocorrência dessa espécie de percevejo, no país, provavelmente está associado à uma melhor adaptação às regiões de temperaturas mais elevadas (Panizzi and Slansky Júnior 1985; Cividanes and Parra 1994). Registros da safra de soja de 2001/2002, em Balsas, no Maranhão (MA), indicaram que 89% dos pentatomídeos coletados a campo pertenciam à espécie *E. heros* (Panizzi 2002). Mais recentemente, na safra 2007/2008, em Londrina, no Paraná (PR), a espécie foi representada por 84% do total de percevejos coletados (Corrêa-Ferreira et al. 2010a).

O percevejo marrom é nativo da América Tropical e está bem adaptado às regiões mais quentes, sendo mais abundante do norte do Paraná ao norte e nordeste do Brasil. Essa espécie produz três gerações durante o ciclo da soja, sendo mais abundante nas lavouras no período de novembro a abril. Durante o verão, também pode ser observado se alimentando da erva-daninha conhecida como amendoim-bravo ou leiteiro (*Euphorbia heterophylla* L.) e, posteriormente, o inseto pode migrar para plantas hospedeiras como o carrapicho-de-carneiro (*Acanthospermum hispidum* DC), girassol (*Helianthus annuus* L.) e guandu (*Cajanus cajan* (L.) Millsp.). Nessa última espécie, o inseto completa a quarta geração antes de entrar em diapausa, permanecendo sem se alimentar sob restos de culturas e folhas mortas, o que garante sua sobrevivência até a próxima safra (Panizzi and Niva 1994; Panizzi and Vivan 1997; Corrêa-Ferreira and Panizzi 1999). O uso abusivo e errôneo de agrotóxicos no final do ciclo da soja tem ocasionado um aumento populacional de *E. heros* (Corrêa-Ferreira et al. 2010a), que tem migrado para lavouras de algodão (*Gossypium hirsutum* L.), no Mato Grosso do Sul (MS). Essa migração ocorre principalmente em áreas com algodão Bt (resistentes a lagartas), o que pode ser explicado pela redução na quantidade de agrotóxicos utilizados para controle das pragas do algodoeiro (Soria et al. 2010). Frequentemente, o percevejo-marrom também tem sido relatado na cultura do milho (*Zea mays* L.) causando injúrias de diferentes intensidades em plântulas e em espigas (Roza-Gomes 2010). No restante dos meses, de maio a outubro, essa espécie entra em estágio de dormência e permanece sob folhas mortas no solo, o que aumenta sua chance de sobrevivência e abundância até a safra seguinte (Panizzi and Niva 1994).

O percevejo-verde pequeno é uma espécie Neotropical, sendo observado desde o sul dos Estados Unidos até a Argentina (Panizzi et al. 2012). Esse percevejo também é a praga mais abundante e de maior importância econômica no Uruguai (Zerbino 2010). No Brasil, possui ampla distribuição geográfica, ocorrendo desde os estados produtores de soja do sul até as regiões de expansão recente do norte e nordeste do país (Hoffmann-Campo et al. 2000). É considerado a espécie que causa os maiores danos à qualidade dos grãos de soja em relação a outros pentatomídeos que atacam a cultura (Corrêa-Ferreira and Panizzi 1999; Hoffmann-Campo et al. 2000). Além disso, o comportamento de inserção e retirada dos estiletes de *P. guildinii* é o que causa maior lesão às paredes celulares em comparação com as outras espécies (Depieri and Panizzi 2011). Essa espécie completa três gerações em soja e uma quarta geração em crotalária (*Crotalaria lanceolata* E.Mey), feijão-guandu e outras espécies anileiras (*Indigofera hirsuta* L., *I. truxillensis* Kunth e *I. suffruticosa* Mill.). Em regiões mais frias do Rio Grande do Sul, também utiliza hospedeiros alternativos como ervilhaca (*Vicia sativa* L.), nabo forrageiro (*Raphanus sativus* L.) e tremoço (*Lupinus* spp.) (Silva et al. 2006).

O percevejo verde é originário do Norte da África, sendo encontrado nas regiões tropical e subtropical das Américas, na Austrália, na Ásia, na África e na Europa (Panizzi et al. 2000). No Brasil, sua ocorrência vem sofrendo

redução nos últimos anos, sendo geralmente observado nas regiões de baixa latitude (Kuss-Roggia 2009; Corrêa-Ferreira et al. 2010b; Kuss et al. 2012). É uma espécie polífaga, sendo encontrada em várias outras culturas, como feijão (*Phaseolus vulgaris* L.), arroz (*Oryza sativa* L.), algodão (*Gossypium hirsutum* L.), macadâmia (*Macadamia integrifolia* Maiden & Betche.) e a noqueira-pecã [*Carya illinoensis* (Wang.) K.Koch.] (Bundy and McPherson 2000; Hoffmann-Campo et al. 2000; Panizzi et al. 2000; Greene et al. 2001). O percevejo verde completa as três primeiras gerações na soja, a quarta geração em *C. lanceolata*, a quinta geração, durante o inverno, em nabo-bravo (*Raphanus raphanistrum* L.), mostarda (*Brassica rapa* L.) ou em feijão-guandu (*Cajanus cajan* L.) e, por fim, uma sexta geração em rubim (*Leonurus glaucescens* Bunge).

Outras espécies de percevejos, consideradas anteriormente como pragas secundárias, como é o caso do percevejo barriga-verde [*Dichelops melacanthus* (Dallas)], apresentaram um elevado crescimento populacional nos últimos anos (Panizzi 1997). Apesar de ser mais frequente na cultura do milho e do trigo, tem sido a espécie de percevejo com maior incidência na soja em algumas regiões produtoras do Paraná, nas últimas safras (Corrêa-Ferreira et al. 2009). Outras plantas hospedeiras dessa espécie são a trapoeraba (*Commelina benghalensis* L.), indigofera (*I. hirsuta*), crotalária (*Crotalaria pallida* Aiton) e capim-braquiária (*Brachiaria decumbens* Stapf) (Silva 2009). Após a colheita da soja, o percevejo barriga-verde permanece no solo sob restos culturais, alimentando-se de plantas de milho e trigo cultivadas em sistema de semeadura direta, o que contribui para sua sobrevivência por fornecer abrigo (palhada) e alimento (sementes maduras caídas no solo) (Chocorosqui 2001). Geralmente, essa espécie de percevejo era considerada secundária, pois o ataque desses insetos ocorria logo no início do desenvolvimento da cultura. As plantas recuperavam-se desse dano inicial e desenvolviam-se normalmente (Panizzi and Chocorosqui 1999). No entanto, nos últimos anos essa espécie de percevejo tem apresentado um aumento significativo em períodos reprodutivos da soja, atacando vagens com grãos verdes, ocasionando danos à cultura (Corrêa-Ferreira et al. 2009).

Outra espécie de percevejo que tem sido observada em soja e é uma praga de importância econômica na Argentina é a *Edessa mediotabunda* F. (Rizzo 1972). Esse percevejo é considerado praga de várias espécies de Solanaceae, incluindo tomate (*Solanum lycopersicum* L.) e batata (*Solanum tuberosum* L.), e de Fabaceae, como ervilha, soja e alfafa (Soria et al. 2010). Essa espécie alimenta-se preferencialmente da haste da soja, devido ao comprimento curto dos seus estiletos (Galileo and Heinrichs 1979; Panizzi and Machado-Neto 1992). No entanto, Husch et al. (2012) observaram, em ensaios de campo, que adultos de *E. mediotabunda* ocasionaram danos com redução significativa da qualidade das sementes de soja, principalmente quando as infestações ocorreram no período da maturação da soja (R7 a R9). Comparando-se sementes maduras atacadas por percevejos das espécies *E. mediotabunda* e *E. heros*, esses últimos ocasionaram danos mais profundos nas sementes, além das mesmas apresentarem tamanho reduzido (Silva et al. 2012a). Outras espécies de percevejos encontradas em soja no Brasil, porém em menor frequência, são *Chinavia* sp. e *Thyanta perditor* F. (Cividanes and Parra 1994; Corrêa-Ferreira and Panizzi 1999).

O ciclo de vida dos percevejos é o mesmo, independente da espécie, com variação interespecífica para o número de dias de cada fase. O ciclo inicia-se com a fase de ovo, seguida por cinco estádios (ínstares) de ninfas, e finaliza-se com a fase adulta (Corrêa-Ferreira and Panizzi 1999). A espécie *E. heros* é a que apresenta o menor período de desenvolvimento das ninfas e também a maior longevidade dos adultos, com média de 117 dias, em comparação com *P. guildinii*, com 54 dias, e *N. viridula*, com 53 dias. A fecundidade média das espécies é de 120 ovos/fêmea para *P. guildinii*, 150 para *N. viridula* e 170 para *E. heros*, com diminuição da postura à medida que as fêmeas envelhecem, parâmetro este influenciado pela dieta alimentar e temperatura.

Estudos demonstram que a intensidade dos danos ocasionados por percevejos depende da espécie, da densidade populacional, do estágio de desenvolvimento das plantas, da suscetibilidade dos genótipos de soja e da época

de semeadura (Panizzi et al. 1979; Schumann and Todd 1982; Villas-Bôas et al. 1990; Sosa-Gómez and Moscardi 1995; Corrêa-Ferreira and Azevedo 2002; Berlote et al. 2003; Corrêa-Ferreira 2010; Depieri and Panizzi 2011). No entanto, os danos ocasionados às sementes por insetos sugadores são semelhantes e irreversíveis (Gazzoni 1998; Gazzoni and Moscardi 1998; Berlote et al. 2003). Os danos são ocasionados devido ao comportamento alimentar desses pentatomídeos fitófagos, que inserem o estilete nas vagens e injetam a saliva que contém enzimas digestivas para sugarem o conteúdo liquefeito das mesmas (Panizzi et al. 2012; Silva et al. 2012a). O ataque das ninfas de 3º a 5º instares e de percevejos adultos durante a fase de formação de grãos pode resultar em aborto dos grãos ou até mesmo das vagens (Panizzi and Slansky Júnior 1985; Villas-Bôas et al. 1990; Brier and Rogers 1991; Gazzoni 1998; Panizzi et al. 2000; Corrêa-Ferreira and Azevedo 2002). Já no período de enchimento dos grãos, o ataque dos percevejos pode ocasionar enrugamento, deformações, aparecimento de manchas escuras e áreas esbranquiçadas, reduções no rendimento e na qualidade das sementes (viabilidade e vigor) (Silva and Ruedell 1983; Villas-Bôas et al. 1990; Gazzoni and Moscardi 1998). Em ataques mais intensos, além de ocorrerem perdas na germinação e vigor da semente, ocorrem alterações nos teores de óleo e proteínas, comprometendo a qualidade e o valor final dos grãos e sementes (Ramiro and Massariol 1977; Villas-Bôas et al. 1990; Gallo et al. 2002). Muitas vezes, é possível observar em áreas onde há infestação de percevejos a presença de plantas com hastes verdes e alta retenção foliar, o que dificulta a colheita mecânica (Lustosa et al. 1999). Tal fenômeno ocorre devido ao processo de alimentação dos percevejos que, ao injetarem toxinas e se alimentarem do conteúdo dos grãos, ocasionam a queda das vagens, o que impede a formação de hormônios de senescência das folhas e atrasa a maturidade da planta (Sosa-Gómez and Moscardi 1995).

Além dos danos diretos, esses insetos também são responsáveis pela transmissão de doenças durante a alimentação. A introdução do aparelho bucal, em forma de estilete, nas sementes facilita a entrada de organismos patogênicos, como é o caso da levedura *Eremothecium coryli* (antes denominada *Nematospora coryli* Peglion), causadora da mancha-fermento, que ocasiona a deterioração das sementes (Gazzoni 1998; Corrêa-Ferreira and Azevedo 2002). Os danos ocasionados por esses insetos-praga reduzem a remuneração do produtor e, também, aumentam o custo para o processamento da soja. Ao introduzir seu aparelho bucal no tegumento, o percevejo provoca uma ruptura dos tecidos que facilita a entrada de água e microrganismos que irão causar problemas futuramente na armazenagem desses grãos (Russin et al. 1988). Além disso, o grão atacado pelo percevejo apresenta uma maior incidência de *Fusarium* spp., bactérias e outros microrganismos que aceleram o processo de fermentação durante o beneficiamento e armazenamento dos grãos nos armazéns (Russin et al. 1988; Quirino 2012). Esse grão fermentado gera um custo maior à indústria, pois exige um tratamento para reduzir a acidez e clarear o óleo obtido do mesmo.

Uma preocupação constante entre os produtores de soja tem sido o custo adicional gerado pelo ataque por percevejos que não tem sido contabilizado. Quando o percevejo se alimenta do grão que está no estágio R6, o grão vai apresentar manchas ou puncturas características desse ataque. No entanto, quando o ataque dos percevejos ocorre no estágio de formação inicial do grão (R5.1 a R5.3), ocorre o abortamento, que é caracterizado como grão chocho, e as vagens apresentam-se planas, sem enchimento de grãos. Portanto, durante a colheita, essas vagens são descartadas na colhedora juntamente com a casca das vagens debulhadas e não são contabilizadas como uma perda de produção gerada pelos percevejos. Estudos demonstram que as perdas no rendimento da cultura devido ao ataque por percevejos são superiores a 30%, e o comprometimento no valor germinativo das sementes pode ser superior a 50% (Degrande and Vivian 2009; Soria et al. 2010). Um estudo comparativo da infestação das espécies de percevejos *P. guildinii*, *N. viridula*, *D. melacanthus* e *E. heros*, nos diferentes estádios de desenvolvimento dos grãos (R5 a R57), demonstrou que as plantas infestadas no estágio R5 são as que apresentam os danos com maior intensidade, enquanto que plantas infestadas na maturação fisiológica R7 apresentam sementes de qualidade semelhante às plantas testemunhas, sem

infestação por percevejos (Corrêa-Ferreira et al. 2009). Além disso, destaca-se que o local da lesão é mais relevante do que o número de picadas, pois uma picada sobre o hipocótilo inviabiliza a germinação, enquanto várias lesões nos cotilédones reduzem o vigor, a sanidade e a emergência, porém não prejudicam a germinação (Jensen and Newsom 1972; Trumper and Edelstein 2008).

Apesar da infestação de percevejos iniciar-se geralmente no final do período vegetativo (Vn) ou logo após a floração (R1 e R2), o período crítico da infestação e os maiores prejuízos à cultura são ocasionados durante o período inicial de desenvolvimento das vagens (R3) até o final do enchimento dos grãos (R6), quando os percevejos atingem o pico populacional máximo (Corrêa-Ferreira and Panizzi 1999). No período da maturação fisiológica da soja (R7), o nível populacional dos percevejos já tende a decrescer. Durante a fase de colheita dos grãos (R8), os percevejos remanescentes tendem a migrar para lavouras mais tardias e, posteriormente, para plantas hospedeiras alternativas, no caso do *N. viridula* e *P. guildinii*, e para os nichos de diapausa, no caso do *E. beros*. Portanto, o cultivo de entressafra, bem como as plantas hospedeiras próximas à lavoura contribuem para a sobrevivência e permanência das espécies de percevejos e influenciam a densidade de insetos colonizantes no início do próximo ciclo (Panizzi and Slansky Júnior 1985). O monitoramento adequado para identificar a presença de percevejos na lavoura é extremamente importante para que medidas adequadas sejam implantadas para o rápido controle populacional deste inseto-praga. O método adequado para a amostragem de insetos é conhecido como pano de batida, em que um pano branco é preso em duas varas, com um metro de comprimento, e estendido entre duas fileiras de soja (Embrapa 2011). As plantas compreendidas pela área do pano são então sacudidas vigorosamente sobre o mesmo para que ocorra a identificação e contagem das pragas ali presentes. Esse procedimento é realizado em diversos pontos aleatórios para realizar uma amostragem da área, preferencialmente nas primeiras horas da manhã ou no final da tarde, que é o período de menor atividade dos insetos, para permitir uma contagem mais eficaz. As amostragens devem ser iniciadas no período de formação das vagens até a maturação fisiológica das plantas, período que apresenta maior abundância de insetos na lavoura. Deve-se iniciar a amostragem pelas bordaduras, onde geralmente os percevejos iniciam o ataque. Para uma lavoura de até 10 hectares (ha), recomenda-se realizar, no mínimo, seis amostragens; para 30 ha, o mínimo aconselhado é de oito amostragens; até 100 ha, com um mínimo de 10 amostragens. No caso de propriedades maiores, recomenda-se a divisão em talhões de 100 ha (Hoffmann-Campo et al. 2000).

Até recentemente, o nível de controle recomendado era de quatro percevejos adultos ou ninfas com mais de 0,5 cm, observados na média das amostragens pelo pano de batida, e dois percevejos por amostragem no caso de campos de produção de sementes (Villas-Bôas et al. 1990). Considerando o aumento populacional desses insetos nas últimas safras, estabeleceu-se um novo nível de ação para os percevejos da soja, durante a Reunião da Comissão de Soja em 2007. Assim, deve-se iniciar o controle quando observada uma média igual ou superior a dois percevejos adultos ou ninfas com mais de 0,5 cm, por metro, e um percevejo para o caso de campos de produção de sementes (Embrapa 2011; Bueno et al. 2012). Atualmente, existem disponíveis cultivares com diferentes grupos de maturação, e as datas de plantio variam de acordo com a programação de cada agricultor. Portanto, deve-se redobrar a atenção nas áreas que possuem materiais mais tardios ou que são semeadas de forma mais tardia em relação às áreas vizinhas, pois os insetos migram rapidamente das áreas recém-colhidas em busca de alimento e podem causar danos irreversíveis.

Em comparação aos insetos desfolhadores, os percevejos são considerados uma praga silenciosa, pois os danos são contabilizados somente na colheita e pós-colheita, o que muitas vezes dificulta o seu controle mais rigoroso. O alto custo de inseticidas e o seu uso em conjunto com fungicidas, em uma mesma aplicação, também fazem com que o produtor reduza o número de aplicações e aumente o intervalo entre as mesmas, favorecendo a multiplicação das populações de percevejos. As despesas associadas ao controle químico dos percevejos são consideradas elevadas

no custo total de produção da cultura. Comparando-se as safras 2009/2010 e 2014/2015, houve um aumento significativo de 6,04% para 13,8% do custo total de produção associado às despesas com inseticidas, para o controle de percevejos na cultura da soja (Lantmann 2014).

A cada safra, os danos ocasionados por esses insetos têm se tornando mais frequentes, considerando o elevado número de espécimes presentes na lavoura, falta de um monitoramento adequado das pragas e, também, pelo surgimento de populações de insetos resistentes (Corrêa-Ferreira et al. 2009). Outros agravantes dessa situação envolvem a tecnologia de aplicação, em que o alvo muitas vezes não é atingido. Como o método de controle dos percevejos mais utilizado é o controle químico, pela pulverização de inseticidas, a escolha do produto e da tecnologia de aplicação devem ser criteriosas, uma vez que a eficiência biológica da pulverização é dependente desses dois fatores (Fiorin et al. 2011).

O elevado potencial reprodutivo de algumas espécies de percevejos, a falta de novas moléculas com diferentes modos de ação e o uso indiscriminado de inseticidas têm contribuído para o aumento do risco de seleção de populações de insetos resistentes, além de ocasionar desequilíbrio biológico dos inimigos naturais (Byrne et al. 2003; Sosa-Gómez and Omoto 2012). No Brasil, até o presente momento, o único caso de resistência a inseticidas detectado foi o do *E. heros* (Sosa-Gómez et al. 2001; Sosa-Gómez and Silva 2010). Um dos inseticidas mais utilizados na cultura da soja desde a década de 60, para controlar as populações de lagartas desfolhadoras e as três espécies mais importantes do complexo de percevejos, era o endossulfan, um organofosforado que, posteriormente, teve sua comercialização proibida. Os primeiros relatos de problemas de deficiência de controle do percevejos marrom, na soja, ocorreram em 1990 (Sosa-Gómez and Omoto 2012). O percevejo marrom também apresenta algumas características que o torna sujeito a uma maior pressão de seleção, em relação às outras espécies de percevejos. A ausência de um comportamento migratório e a limitada capacidade de vôo da espécie contribuem para um aumento da pressão de seleção nas populações desses insetos (Sosa-Gómez et al. 2012; Sosa-Gómez and Omoto 2012). Um estudo realizado com populações de *E. heros* e *N. viridula* indicaram que suas populações são estruturadas, com reduzido fluxo gênico entre as mesmas, mesmo em populações que se situam em localidades próximas, com aproximadamente 52 km de distância em linha reta (Sosa-Gómez et al. 2004; Sosa-Gómez et al. 2005). Geralmente, o percevejo marrom, juntamente com o percevejo verde pequeno apresentam até oito gerações por ano nas regiões mais quentes do Brasil, enquanto o percevejo verde da soja possui cinco gerações por ano, o que torna a pressão de seleção maior no primeiro caso (Cividanes and Parra 1994; Cividanes and Figueiredo 1997).

Uma das medidas incorporadas ao manejo integrado de pragas (MIP), na década de 90, consiste no controle biológico dos percevejos da soja pelos parasitoides (Corrêa-Ferreira 1993). Os dois parasitoides mais utilizados são *Trissolcus basalís* e *Telenomus podisi*. (Corrêa-Ferreira 2003). Essas espécies se destacam pela eficiência, importância e abundância nas lavouras de soja. A espécie *T. podisi* parasita, preferencialmente, os ovos de *E. heros*, enquanto *T. basalís* tem preferência por parasitar ovos de *N. viridula* (Godoy et al. 2005). Ambos depositam seus ovos no interior dos ovos dos percevejos, matam o embrião e eliminam a praga no estágio inicial do seu desenvolvimento, antes mesmo de causar qualquer injúria às plantas (Bueno et al. 2012). Embora esses parasitoides ocorram naturalmente nas lavouras de soja, o constante uso de produtos seletivos na fase inicial da cultura e a falta de locais de refúgio durante a entressafra são fatores que prejudicam sua eficiência (Corrêa-Ferreira 2003). Portanto, recomenda-se a liberação de parasitoides na lavoura na época em que os percevejos iniciam a oviposição, para reduzir o número de percevejos-praga durante o período crítico da cultura. A recomendação é de 5.000 vespas por hectare, quando liberadas como adultos ou três cartelas de ovos parasitados, com aproximadamente 1.670 ovos/cartela. As fêmeas adultas, ao emergirem, são copuladas e vão depositar seus ovos em massas de ovos de percevejos ou em percevejos adultos, reiniciando o ciclo.

Além do parasitismo em ovos, também é comum, na cultura da soja, o parasitismo em adultos de percevejos, sendo que as moscas colocam seus ovos sob o corpo dos percevejos e as larvas se desenvolvem no interior do hospedeiro, matando-o quando emerge para empupar no solo (Corrêa-Ferreira 1984). A mosca *Trichopoda giacomelli* é um parasitoide importante em adultos de *N. viridula*, enquanto *Hexacladia smithii* (Encyrtidae) é o principal parasitoide em adultos de *E. heros* (Corrêa-Ferreira and Panizzi 1999; Bueno et al. 2012).

A falta de opções para a escolha de produtos com eficiência comprovada e com modos de ação diferentes e a contínua expansão da soja reforça a hipótese de que a forma mais eficaz para o controle desses insetos-praga seria a utilização de cultivares resistentes a percevejos (McPherson et al. 2007). Além disso, os benefícios do uso de plantas resistentes envolvem questões ambientais, econômicas e sociais. A redução no uso de inseticidas traz benefícios ambientais, minimiza os desequilíbrios biológicos, reduz o custo de produção, permitindo que um maior número de agricultores possa cultivar soja, e garante uma maior segurança ao produtor.

2.3. Mecanismos de resistência e cultivares de soja resistentes aos percevejos

Uma planta resistente a um determinado inseto-praga é aquela que não apresenta perdas no seu rendimento, ou seja, é menos danificada pelo inseto do que outra planta, em igualdade de condições (Painter 1951). A resistência das plantas é classificada por Painter (1951) em três categorias, sendo todas elas manifestadas na cultura da soja. A não preferência, que foi posteriormente denominada como antixenose (Kogan and Ortman 1978), é caracterizada pela menor utilização da planta pelo inseto, devido a fatores morfológicos ou químicos. A produção de compostos secundários pelas plantas ocasiona efeitos adversos sobre o comportamento do inseto, influenciando processos de alimentação, oviposição e abrigo. Na antibiose, a planta é utilizada como hospedeiro, porém ocorre a produção de toxinas ou compostos metabólicos secundários que causam efeitos adversos sobre a biologia do inseto resultando em alterações no crescimento, reprodução e sobrevivência do inseto. Os parâmetros mais comuns utilizados para identificação desse tipo de resistência de plantas a insetos são: a morte de indivíduos em estágio juvenil, aumento no período de desenvolvimento dos estádios iniciais de desenvolvimento do inseto, perda de peso ou redução do tamanho do inseto, redução da longevidade e fecundidade do inseto adulto e morte das pupas. No caso de plantas tolerantes, há a capacidade da planta suportar ou ainda recuperar-se dos danos sofridos por uma população de insetos, a qual normalmente causaria sérios prejuízos a um hospedeiro mais suscetível. Esse tipo de resistência é mais influenciado pelo ambiente, pois plantas mais vigorosas podem tolerar um maior ataque de pragas. No entanto, como não há influência da planta sobre a biologia ou comportamento do inseto, pode ocorrer o aumento populacional da praga em longo prazo, resultando em danos econômicos à lavoura. Portanto, no caso de plantas tolerantes, é necessária a utilização de estratégias complementares de manejo integrado de pragas para o controle das populações de insetos. O método mais eficaz para verificar a presença de plantas tolerantes é a comparação do dano e da produtividade de grãos da mesma planta utilizando-se um tratamento com e outro sem a infestação com insetos.

Nas últimas décadas, diversos programas de melhoramento têm utilizado a estratégia de identificação de genes de resistência e transferência de genes, por melhoramento genético, via cruzamentos, para a obtenção de variedades e cultivares elite mais produtivas. No entanto, devido ao grande número de pragas e doenças que as culturas estão sujeitas, a estratégia de piramidação de genes foi necessária para garantir um melhor desempenho dos materiais. O uso da seleção assistida por marcadores (SAM) tem sido empregado com sucesso no caso de genes de herança monogênica, como no caso da resistência a nematoide-de-cisto (*Heterodera glycines*) (Concibido et al. 2004; Silva et al. 2007). Porém, no caso de caracteres quantitativos, que normalmente são controlados por um maior número de genes, o uso da SAM não tem apresentado sucesso, tanto pela dificuldade na identificação de genes de maior efeito quanto

pela distância dos marcadores identificados em relação ao gene de interesse (Boerma and Walker 2005). Apesar da identificação de um grande número de QTLs após o advento dos marcadores moleculares, a utilização dos mesmos na SAM requer a validação dos mesmos em diferentes *backgrounds* genéticos e em ambientes distintos, além de um mapeamento fino para identificar marcadores mais próximos ao gene de interesse (Collard and Mackill 2008).

Pouco se conhece sobre os mecanismos e genes relacionados à resistência aos insetos praga, mas sabe-se que o mecanismo de controle/resposta ao ataque de percevejos apresenta natureza poligênica, o que justifica a dificuldade na obtenção de cultivares agronomicamente competitivas e resistentes a estes insetos (Souza and Toledo 1995; Godoi and Pinheiro 2009). A introdução de genes de resistência em cultivares elite tem sido realizada por meio de cruzamentos artificiais em soja, utilizando-se materiais descritos na literatura como resistentes a estes insetos-praga. Na década de 1960, três introduções de plantas (*Plant Introduction*) foram identificadas como resistentes ao ataque do besouro desfolhador *Epilachna varivestris* (Mulsant, 1850), sendo elas PI 171451 ('Kosamame'), PI 229358 ('Sodendaizu') e PI 227687 (Miyako White) (Van Duyn et al. 1971; Van Duyn et al. 1972). Posteriormente, estes mesmos genótipos foram identificados como resistentes a outros insetos, como os besouros *Cerotoma trifurcata* (Forster) e *Epicauta vittata* (F.), à lagarta *Helicoverpa zea* (Boddie) (Clark et al. 1972), à lagarta *Plathypena scabra* (F.), à lagarta falsa medideira *Trichoplusia ni* Hübner, e outras espécies de insetos praga da soja (Luedders and Dickerson 1977; Boerma and Walker 2005). Outras duas introduções de plantas descritas como resistentes ao ataque de insetos são PI 274454 (Rossetto 1989) e PI 171444 (Gilman et al. 1982). Desde então, muitos programas de melhoramento utilizaram essas introduções de plantas com o objetivo de transferir os genes de resistência das mesmas às cultivares elite, mas como pouco se conhece sobre os mecanismos desses genes e essas introduções de plantas apresentam baixo valor agrônomo, não houve sucesso (Lambert, L.; Tyler, 1999; Sosa-Gómez et al., 2012). Acredita-se que uma grande quantidade de genes de características agrônômicas não desejáveis encontra-se fortemente ligadas aos QTLs que conferem resistência a insetos, resultando em um arraste de genes indesejáveis associados aos genes de interesse (*linkage drag*) (Van Duyn et al. 1971; Boethel 1999; Boerma and Walker 2005). Além disso, a baixa herdabilidade das características de produtividade de grãos e resistência a insetos dificulta ainda mais a obtenção de linhagens resistentes (Sosa-Gómez et al. 2012). As primeiras cultivares americanas derivadas dessas introduções foram 'Crockett', 'Lyon', 'Lamar' e 'Shore'. No entanto, nenhuma delas apresentou o nível de resistência igual ao genitor doador e, além disso, as cultivares apresentaram baixa produtividade de grãos em relação às cultivares elite, o que não despertou interesse dos produtores (Boethel 1999).

No Brasil, os materiais derivados das introduções de plantas que foram utilizados como fonte de genitores resistentes são IAC 73-228 e IAC 78-2318. As cultivares desenvolvidas, a partir desses dois genitores, descritas na literatura como resistente aos percevejos são IAC-100, IAC-17, IAC-19, IAC-23 e IAC-24 (Rossetto et al. 1986; Pinheiro 1993; Carrao-Panizzi and Kitamura 1995; Veiga et al. 1999; Miranda et al. 2001; Miranda and Lourenção 2002; Miranda et al. 2003; Piubelli et al. 2003a; Campos et al. 2010; Silva et al. 2013). A resistência dessas cultivares provavelmente é herdada, em sua maioria, das introduções PI 274454 e PI 229358 (Rossetto et al. 1995; Veiga et al. 1999). Apesar destas cultivares já se encontrarem obsoletas em relação ao patamar de produtividade alcançado, elas continuam sendo utilizadas nos programas de melhoramento como fontes de genes de resistência a percevejos, pois também apresentam um maior número de caracteres agrônômicos desejáveis em relação às introduções de plantas, que geralmente são menos adaptadas (Rossetto et al. 1984; Panizzi and Slansky Júnior 1985; Rossetto et al. 1986; Lourenção et al. 1989; Rossetto 1989). No entanto, quando se comparam as PIs com as linhagens descendentes, as primeiras possuem resistência a insetos superior, o que confirma o grande número de genes envolvidos no mecanismo de

controle da resistência a insetos e a conseqüente dificuldade de sua transferência (Lourenção et al. 1985; Hoffmann-Campo 1995).

Apesar da baixa herdabilidade das características de produtividade de grãos e resistência a insetos dificultar o desenvolvimento de cultivares resistentes, algumas linhagens resistentes têm apresentado bons níveis de produtividade de grãos quando comparadas às cultivares comerciais (Sosa-Gómez et al. 2012). Dessa forma, programas de melhoramento de soja, visando a resistência a insetos, precisam desenvolver populações grandes para permitir o aparecimento das combinações gênicas para todos os caracteres de interesse. Como a herança para resistência a insetos é complexa (Souza and Toledo 1995), é preciso avaliar um modelo de seleção recorrente para obter linhagens com elevado nível de resistência e alto potencial produtivo (Sosa-Gómez et al. 2012). Outro cuidado importante deve ser estabelecido em relação à avaliação da infestação de insetos na área, pois o ataque não uniforme dos percevejos pode mascarar os resultados de testes de resistência realizados no campo experimental (Arias et al. 1999).

Dentre as cultivares apresentadas, a que mais tem se destacado como genótipo padrão de resistência aos insetos desfolhadores e sugadores é a cultivar IAC-100, pois apresenta características agrônômicas favoráveis e boa produtividade de grãos (Pinheiro et al. 2005). Fernandes et al. (1994) verificaram que a cultivar IAC-100 apresentou o menor índice de retenção foliar e a menor percentagem em peso de sementes totalmente deformadas, quando comparada com outras seis cultivares de soja (IAC8, IAC12, IAC17, Dourados, Emgopa 304 e Emgopa 309), submetidas ao ataque de percevejos. Souza (2013) avaliou cultivares e linhagens utilizadas no programa de melhoramento do Instituto Agrônômico de Campinas e também verificou que IAC-100 foi uma das cultivares menos danificadas pelos percevejos, embora não tenha sido uma das menos atrativas aos insetos. Esse mesmo autor verificou que a cultivar ocasiona o prolongamento do desenvolvimento ninfal, demonstrando a presença do mecanismo de resistência denominado antibiose. Outros estudos também relatam que a cultivar IAC-100 foi a cultivar menos atacada em condições de infestação com *P. guildinii* (Campos, 1996; Rodrigues, 1996). Silva et al. (2013) também verificou evidências do mecanismo antibiose, pois insetos de *P. guildinii*, alimentados com vagens da cultivar IAC-100, apresentaram alta mortalidade de ninfas (superior a 90%), além de um prolongamento do estágio ninfal. De acordo com Panizzi e Silva (2009), a mortalidade ninfal pode estar associada a fatores físicos da planta como a distância entre o grão e a vagem, tornando mais difícil o acesso ao alimento. A cultivar IAC-100 também foi utilizada com sucesso em programas de melhoramento de soja, nos Estados Unidos, como fonte de resistência para percevejos sugadores e insetos desfolhadores (McPherson et al. 2007; McPherson and Buss 2007).

Um dos mecanismos descritos para a cultivar IAC-100 consiste no mecanismo de tolerância, pois essa cultivar possui a capacidade de abortar vagens danificadas e substituí-las por novas, recuperando-se dos danos sofridos pelo ataque dos percevejos (Rossetto et al. 1995; Piubelli et al. 2005). Também apresenta um mecanismo de pseudo-resistência, denominado diluição de dano, pois apesar de possuir semente de tamanho pequeno, a quantidade de semente produzida por planta é geralmente maior em relação às outras cultivares, o que contribui para garantir uma boa produtividade final de grãos sem dano (Rossetto and Lara 1991). Além disso, Lara (1991) sugeriu um outro mecanismo de pseudo-resistência presente em IAC-100, denominado evasão, que também contribui para uma menor suscetibilidade da planta aos percevejos. A cultivar apresenta um menor período de granação, período em que se encontra mais suscetível ao ataque dos percevejos. Portanto, um menor período de exposição aos insetos resulta em menor quantidade de dano ocasionado nos grãos (Rossetto et al. 1995; Piubelli et al. 2005).

Estudos demonstram que os metabólitos flavonol rutina e isoflavonóide genistina, presentes na cultivar IAC-100, bem como nas introduções de plantas PI 274454 e PI 229358, afetam negativamente a biologia da lagarta *Anticarsia gemmatalis* (Piubelli 2004; Piubelli et al. 2005). Outros estudos demonstraram que o flavonol rutina também

provoca antibiose em *Trichoplusia ni* Hübner, 1803 (Hoffmann-Campo et al. 2001) e na lagarta-da-soja (*A. gemmatilis*) (Hoffmann-Campo et al. 2006). Estudos realizados com diferentes cultivares e introduções de plantas demonstraram que a cultivar IAC-100 e a PI 227697 possuem teores mais elevados de isoflavonas e provocam efeitos negativos mais acentuados (maior mortalidade e menor ganho de peso) em percevejos *N. viridula* quando alimentados com vagens desses genótipos (Piubelli et al. 2003b). Além disso, verificou-se um aumento significativo na concentração de isoflavonas daidzina e genistina, ambas precursoras das fitoalexinas gliccolinas em soja, em sementes da PI 227687 e PI 229358, após o ataque de percevejos *N. viridula*, ocasionando alterações negativas no comportamento dos mesmos (Piubelli et al. 2003a). Toledo (2005) também observou um pico no aumento das mesmas isoflavonas, no período de 48 horas após o ataque dos percevejos nos grãos danificados, além de uma indução mais rápida de gliccolinas em genótipos resistentes a insetos.

Com a finalidade de associar a preferência alimentar de *E. heros* às características morfológicas, Efrain (2013) comparou a quantidade e o tamanho de tricomas de diferentes cultivares e linhagens de soja. A alta densidade de tricomas observada na cultivar IAC-100 não inibiu a atratividade a campo dos percevejos. No entanto, o autor sugere que essa característica morfológica pode ter influenciado a resistência do tipo antibiose verificada na cultivar IAC-100, tendo contribuído para prolongamento ninfal. Embora a maior densidade de tricomas não tenha influenciado na alimentação de insetos adultos, provavelmente pode ter atuado como uma barreira física aos primeiros estádios iniciais (segundo e terceiro ínstars), dificultando a alimentação dos mesmos. Além disso, Hulburt et al. (2004) identificaram o QTL-E, associado à resistência a insetos, próximo ao loco *Pb*, que controla o tipo de pubescência na folha. Segundo os autores, linhagens que apresentaram tricomas com a extremidade do tipo afiada (*Pb*₋) demonstraram maior resistência ao ataque de insetos lepidópteros.

O estudo e elucidação dos mecanismos relacionados à resistência das plantas aos insetos-praga tem sido o objetivo de muitos pesquisadores. Dessa forma, a utilização de marcadores moleculares, juntamente com a genômica funcional, é fundamental para auxiliar no desenvolvimento de cultivares resistentes.

2.4. Marcadores moleculares em soja

Os primeiros mapas genéticos foram construídos com base em marcadores morfológicos e citológicos. Apesar da grande contribuição destes marcadores, sua disponibilidade se restringia basicamente às espécies modelo. Além disso, a disponibilidade dos mesmos é reduzida, impossibilitando uma eficiente cobertura dos genomas (Stuber et al. 1992).

O advento das técnicas de biologia molecular culminou no desenvolvimento de diversos marcadores moleculares disponíveis atualmente. Os marcadores moleculares representam variações na sequência do DNA, o que resultou em uma significativa contribuição para o processo de melhoramento de plantas, como a construção de mapas genéticos, associação entre genótipos e características fenotípicas e ainda a seleção de genótipos para caracteres de interesse (Schlötterer 2004; Borém et al. 2009).

Os marcadores moleculares podem ser classificados em dominantes ou co-dominantes, dependendo do tipo de polimorfismo que permitem identificar (Ferreira and Grattapaglia 1998; Borém et al. 2009). Desde a década de 80, diversos tipos de marcadores moleculares baseados em DNA foram desenvolvidos, com variações no tipo de tecnologia utilizada para detectar o polimorfismo, visando a redução de custos, a facilidade de uso, reprodutibilidade e a obtenção de dados em larga escala (Schlötterer 2004). Os marcadores RFLP (*Restriction Fragment Length Polymorphism*) (Botstein et al. 1980), RAPD (*Random Amplified Polymorphic DNA*) (Williams et al. 1990), SSR (Jacob et al. 1991) e AFLP (Vos et al. 1995) foram marcadores amplamente utilizados no melhoramento de plantas de diversas culturas (Cregan

et al. 1999a; Oliveira et al. 2001; Cholastova et al. 2011). Os primeiros mapas genéticos da soja, baseados em DNA, foram publicados na década de 90, sendo construídos inicialmente com marcadores RFLP, posteriormente também com marcadores AFLP, SSR e, recentemente, com a adição de milhares de marcadores SNP (*Single Nucleotide Polymorphism*), a partir das novas plataformas de sequenciamento e genotipagem (Keim et al. 1990; Mansur et al. 1993; Shoemaker and Specht 1995; Cregan et al. 1999b; Song et al. 2004; Wu et al. 2004; Choi et al. 2007; Hisano et al. 2007; Xia et al. 2007; Hwang et al. 2009; Hyten et al. 2010; Niu et al. 2013; Song et al. 2013). A integração de diferentes mapas da soja resultou na construção de um mapa consenso, denominado ConsensusMap 4.0, que está disponível no Banco de Dados da Soja (Soybase - <http://www.soybase.org>) (Grant et al. 2010). Atualmente, o mapa contém informações de seqüências e posição de aproximadamente 5.500 marcadores genéticos, distribuídos nos 20 GLs da soja (Hyten et al. 2010).

Os marcadores SSR ainda têm sido muito utilizados em soja para o mapeamento genético e identificação de QTLs de importância agrônômica e econômica envolvidos na produtividade de grãos e na resistência genética a pragas e doenças, características essas de herança complexa (Mian et al. 1999; Yuan et al. 2002; Hu et al., 2014; Qi et al., 2014; Jun et al., 2014; Kato et al., 2014). Marcadores SSR são loco-específicos, podem ser multialélicos e possuem caráter co-dominante. Os SSR consistem em pequenas seqüências (dois a seis pares de base) repetidas em tandem e amplamente distribuídas no genoma eucarioto (Akkaya et al. 1992). Cada SSR constitui um loco genético altamente polimórfico, cujas variações no número de repetições determinam diferentes alelos que podem ser separados por eletroforese para acessar o perfil de cada indivíduo. As seqüências de DNA que flanqueiam os microssatélites geralmente são únicas e bem conservadas, permitindo a construção de *primers* para a amplificação por meio da PCR (*Polymerase Chain Reaction*). A grande disponibilidade de marcadores SSR, distribuídos nos 20 GLs da soja, fazem com que a sua utilização seja frequente em estudos de mapeamento de QTLs. No entanto, o advento do sequenciamento de segunda geração, aliado ao desenvolvimento de técnicas de genotipagem em larga escala, contribuíram para tornar os marcadores SNP mais vantajosos em estudos de mapeamento.

Os marcadores SNP consistem em alterações de uma única base entre fragmentos homólogos de DNA, além de pequenas inserções ou deleções (INDELS), podendo ocorrer tanto em regiões codificadoras como em não codificadoras do genoma. Em regiões codificadoras, quando resultam em uma substituição de aminoácido na seqüência proteica, são denominados não sinônimos. Neste caso, a substituição pode ser conservativa ou não conservativa em função das características dos aminoácidos envolvidos na troca, podendo resultar em modificações estruturais e funcionais da proteína. Embora os SNP sinônimos não alterem a seqüência proteica, eles podem modificar a estrutura e a estabilidade do RNA mensageiro e, conseqüentemente, afetar a quantidade de proteína produzida. As mutações que ocorrem nos íntrons não podem afetar diretamente a estrutura da proteína, entretanto elas podem impedir a produção do RNA mensageiro, inibindo a união por *splicing* dos éxons. Algumas vezes, as alterações das bases nitrogenadas têm origem em erros de incorporação de bases durante a replicação do DNA, ou podem ser causadas por lesões no DNA por agentes ambientais. As mutações que ocorrem em células germinativas e são transmitidas às gerações seguintes podem se fixar na população. Quando a mutação está fixada em uma frequência mínima de 1% é considerada um SNP (Kwok and Gu 1999). O genoma da soja cultivada [*Glycine max* (L.) possui diversidade de nucleotídeos (θ) de 0.001, o que representa uma média de um SNP a cada 1000 pb (pares de bases) (Zhu et al. 2003; Choi et al. 2007). Já o genitor selvagem *Glycine soja* (Sieb and Zucc.) possui uma diversidade de nucleotídeos (θ) de 0.00235, com aproximadamente 1 SNP a cada 425 pb (Hyten et al. 2006). Essa redução da diversidade genética observada na soja cultivada é resultante do seu processo de domesticação (Stupar 2010; Li et al. 2010b; Joshi et al. 2013; Hu et al. 2014; Zhang et al. 2016).

Atualmente, diversos métodos de alto desempenho estão disponíveis para genotipar rapidamente um grande número SNP em um curto intervalo de tempo (Fan et al. 2006). O ensaio Illumina GoldenGate é um método de detecção de SNP em larga escala com capacidade de multiplex que pode genotipar de 384 a 1536 SNP em 192 amostras de DNA em um período de apenas três dias (Hyten et al. 2008; Hyten et al. 2009; Hyten et al. 2010). Recentemente, o Illumina Infinium foi desenvolvido, contendo 50.000 SNP distribuído pelos 20 GLs da soja (Song et al. 2013).

A abundância de marcadores SNP distribuídos pelos 20 GLs da soja e a constante inovação dos métodos de genotipagem em larga escala, permitem a obtenção de mapas genéticos com uma cobertura satisfatória do genoma para realizar estudos de mapeamento de QTLs, bem como o mapeamento fino dessas regiões. O uso de marcadores moleculares nos programas de melhoramento tem proporcionado formas mais eficientes de seleção de indivíduos superiores, pela SAM, no caso de genes de herança monogênica. Além de facilitar o rastreamento de genes de interesse, a SAM tem se mostrado uma importante ferramenta em programas de melhoramento por permitir a redução tanto do tempo quanto dos recursos necessários para a obtenção de cultivares melhoradas (Collard & Mackill, 2008; Sosa-Gómez et al., 2012). Nos programas de introgressão gênica, a SAM tem se mostrado eficaz para a seleção de indivíduos contendo o gene de interesse em populações de retrocruzamentos (Bouchez et al. 2002; Zhou et al. 2003; Toojinda et al. 2005; Mackill 2006). Outra vantagem dessa ferramenta é permitir a identificação de indivíduos resistentes a pragas e doenças sem que haja a infestação ou inoculação, proporcionando maior rapidez no programa de melhoramento. Além disso, tem sido relatado que a piramidação de genes de resistência utilizando a SAM dificulta a quebra da resistência pelo patógeno (Shanti et al. 2001; Singh et al. 2001). A piramidação de genes utilizando-se a SAM foi realizada com sucesso em arroz (Hittalmani et al. 2000; Datta et al. 2001; Arunakumari et al. 2016), trigo (Liu et al. 2000), aveia (Castro et al. 2003) e cevada (Riedel et al. 2011). No entanto, o grande desafio para o uso adequado da SAM em melhoramento é encontrar QTL cuja segregação explique uma porção significativa da variação fenotípica para justificar sua utilização (Sosa-Gómez et al. 2012). Além disso, o QTL precisa apresentar-se estável em diferentes ambientes e em populações com *background* genético distintos. No caso dos primeiros QTL identificados para resistência a insetos em soja, utilizando-se marcadores RFLP, devido às características intrínsecas dos marcadores, de baixo polimorfismo e elevada dificuldade técnica de obtenção dos mesmos, o mapeamento fino dessas regiões, utilizando-se os marcadores SNP, muito contribuiu para a seleção assistida desses QTL, tornando o processo muito mais eficiente e facilitando a introgressão desses QTL nas cultivares comerciais (Walker et al. 2010).

No caso de caracteres quantitativos, o uso da seleção genômica para a identificação de genótipos superiores tem sido frequente para diferentes espécies (Lorenz et al. 2012; Massman et al. 2013; Jarquín et al. 2014; Xavier et al. 2016). A grande disponibilidade de marcadores obtidos com a genotipagem em larga escala, aliada à redução do custo de genotipagem têm contribuído para a utilização da seleção genômica em programas de melhoramento. Basicamente, um conjunto de indivíduos é utilizado como população de treinamento, para a obtenção dos dados genotípicos e fenotípicos. A metodologia consiste em estimar o valor genético genômico (VGG) desses indivíduos para estimar o VGG de uma população de validação, que possui apenas os dados genotípicos (Meuwissen et al. 2001). Estudos tem demonstrado uma maior acurácia em modelos de seleção genômica com a utilização de QTLs, previamente identificados em estudos de mapeamento, como efeito fixo, juntamente com os demais marcadores como efeito aleatório (Bernardo 2014; Spindel et al. 2015; Arruda et al. 2016).

2.5. Mapeamento de QTLs

O advento dos marcadores moleculares aliado às técnicas de genotipagem em larga escala, permitem a obtenção de mapas de ligação saturados, que são de grande relevância para o mapeamento de QTLs (Ferreira and

Grattapaglia 1998). QTLs consistem em regiões cromossômicas que contêm genes (ou locos) que controlam caracteres poligênicos, cuja expressão fenotípica apresenta uma variação contínua (Thoday 1961). Muitas das características de interesse apresentam controle poligênico, sendo que cada loco geralmente possui um pequeno efeito no fenótipo da característica. A base fundamental do mapeamento genético consiste em identificar a associação preferencial entre marcadores moleculares ligados fisicamente ao gene de interesse (QTL), de tal forma que os dados do marcador fornecem informações sobre o mesmo. As associações entre genótipo e fenótipo, obtidas pelas informações da segregação dos marcadores moleculares e pelos valores de médias fenotípicas de cada indivíduo, respectivamente, permitem estimar o número, a posição, os efeitos, as interações dos alelos dos QTLs dentro dos locos (dominância) e entre locos (epistasia), efeitos pleiotrópicos e interações entre QTL e ambientes (Jiang and Zeng 1995; Ferreira and Grattapaglia 1998; Zeng 2001). Entretanto, a disponibilidade de um mapa genético fidedigno depende de vários fatores como o tipo de marcador utilizado, o tipo de população analisada e o tamanho da população. Um dos fatores de fundamental importância, para se obter dados consistentes em um trabalho de mapeamento, é o tamanho da amostra ou população a ser estudada. A resolução do mapa e a capacidade de se determinar a sequência de marcadores nele presentes estão diretamente relacionadas ao tamanho da amostra ou população. É fundamental que se estimem adequadamente distâncias entre os genes, estabelecendo-se o ordenamento correto e formando-se grupos de ligações que reflitam o número básico de cromossomos da espécie. A frequência de recombinação entre os locos é usada como referência para o cálculo de distância e ordenamento dos genes (ou marcadores) nos cromossomos (Schuster and Cruz 2004). De forma geral, os eventos de permutação ocorrem ao acaso ao longo do cromossomo, e a probabilidade de recombinação é maior para locos que se encontram a uma maior distância entre si do que para aqueles mais próximos. Entretanto, existem “*hot spots*”, ou seja, regiões preferenciais onde ocorre uma maior taxa de recombinação entre os genes. O entendimento da base genética dos caracteres quantitativos é fundamental em programas de melhoramento para que as estratégias de seleção sejam definidas, como no caso da ligação física, que pode ser transitória e modificada ao longo das recombinações, ou pleiotropia, que permanecerá em gerações futuras (Falconer and Mackay 1996).

O princípio básico que fundamenta as análises de mapeamento de QTL é a existência de desequilíbrio gamético de ligação decorrente da redução da frequência de recombinação entre genes situados em regiões próximas entre si ao longo de determinado cromossomo (Coelho 2000). Sem esse pré-requisito, os alelos dos locos marcadores e dos QTLs ocorrerão em combinações equivalentes à distribuição independente, fazendo com que os QTLs não sejam detectados. Este desequilíbrio gera efeitos quantitativos associados ao marcador, os quais podem ser detectados e estimados por meio de análises estatísticas adequadas. Diversas metodologias para o mapeamento de QTLs foram desenvolvidas, com sucessivo avanço tanto em relação ao número de informações obtidas como em relação à precisão das mesmas. O mapeamento por intervalos ou IM (*Interval Mapping*) (Lander and Botstein 1989) se baseia na frequência de recombinações de um par de marcadores adjacentes. A busca por um possível QTL ocorre de forma condicionada a certo intervalo. Apesar de ter sido muito utilizado, seu poder estatístico é relativamente baixo, o que é explicado, em parte, pela não independência dos diferentes intervalos quando existe mais de um QTL. Posteriormente, Zeng (1993) desenvolveu a abordagem de mapeamento por intervalos compostos ou CIM (*Composite Interval Mapping*), que combina o mapeamento por intervalos e regressão linear múltipla, mostrando-se mais eficiente. A vantagem dessa abordagem de mapeamento é que os marcadores são utilizados como co-fatores, e contribuem para reduzir a variação genética resultante dos demais QTLs não incluídos no modelo (Liu 1998). Dessa forma, o mapeamento CIM permite maior precisão na estimativa da localização do QTL, devido à inclusão de mais marcadores no modelo, o que contribui para a diminuição do resíduo. E, também, permite o controle de falsos positivos, pois os marcadores utilizados como co-fatores bloqueiam os efeitos de possíveis QTLs pertencentes às regiões adjacentes ao intervalo em estudo. Embora

apresente maior poder estatístico, a abordagem CIM considera um modelo com um único QTL, ou seja, o efeito aditivo de cada QTL é estimado considerando-se um único QTL por vez. Assim, são realizados múltiplos testes e o *threshold* é determinado por teste de permutação (Churchill and Doerge 1994). Considerando a existência de QTLs segregando conjuntamente para a população, Kao et al. (1999) propuseram a metodologia de mapeamento por intervalos múltiplos ou MIM (*Multiple Interval Mapping*). Esse método permite incluir o efeito de mais de um QTL no modelo, o que reduz a variância residual e aumenta o poder dos testes para identificação de QTLs e a precisão nas estimativas das posições dos mesmos. Além disso, o mapeamento MIM auxilia no maior entendimento da arquitetura genética dos caracteres quantitativos, permitindo estimar os efeitos de interação entre os QTLs (QQ) (Zeng 2001). Outra vantagem de destaque para a abordagem MIM é a possibilidade de obtenção da estimativa do valor genotípico dos indivíduos, os quais podem ser diretamente utilizados em programas de melhoramento genético (Zeng 2001).

As abordagens de mapeamento univariado têm apresentado sucesso na identificação de genes controlando características de interesse. No entanto, todas essas metodologias descritas permitem a identificação de genes que tenham efeito em uma única variável. Como, na prática, geralmente várias características são avaliadas em uma população de mapeamento de QTLs, Jiang and Zeng (1995) propuseram uma extensão do mapeamento por intervalo composto para o caso multivariado, denominado mapeamento por intervalo composto para múltiplas características ou MT-CIM (*Multiple Trait Composite Interval Mapping*). Posteriormente, Silva et al. (2012b) estenderam o MIM para a análise conjunta de fenótipos correlacionados, denominado mapeamento por intervalos múltiplos para múltiplas características ou MT-MIM (*Multiple Trait Multiple Interval Mapping*). Ambas as metodologias multivariadas de mapeamento de QTLs utilizam-se dos padrões de correlação representados na matriz de variâncias/covariâncias, o que resulta em um maior poder de detecção de regiões genômicas, com efeito sobre os fenótipos avaliados, que muitas vezes não são detectadas por análises univariadas. De acordo com Jiang and Zeng (1995) e Silva et al. (2012b), considerando a estrutura de correlação entre múltiplos fenótipos, pode-se obter um aumento no poder de detecção de QTLs, além de uma maior precisão da localização e das estimativas dos parâmetros genéticos em relação às análises de fenótipos individuais. Na abordagem multivariada, além do estudo dos efeitos de interações entre QTLs (QQ), é possível estimar a interação QTL x E (QE), o que auxilia na compreensão da arquitetura genética dos mesmos (Silva and Zeng 2010). A abordagem do mapeamento multivariado apresenta grande relevância no melhoramento de plantas, considerando que, para se obter materiais genéticos superiores, é necessário que os mesmos reúnam alelos favoráveis para diversas características (Walker et al. 2010). Portanto, a identificação de regiões genômicas controlando um conjunto de variáveis contribuirá para facilitar a seleção de genótipos elite, permitindo maiores ganhos de seleção para diversas características simultaneamente.

Nenhum trabalho foi encontrado na literatura utilizando a abordagem multivariada em soja. No entanto, estudos de mapeamento realizados em milho, para características fenotípicas associadas à absorção de fósforo (Gianotto 2015), e na área médica, para mapeamento de genes associados à hipertensão e obesidade (Duarte 2007), demonstram a superioridade da abordagem multivariada na identificação de QTLs, considerando variáveis correlacionadas, em relação à abordagem univariada. Duarte (2007) identificou um novo QTL, associado à diminuição da pressão sistólica, que não havia sido identificado na análise de traços individuais. Além disso, o estudo permitiu identificar as interações entre QTLs e em diferentes condições de tratamento, auxiliando na elucidação da base genética da pressão sistólica estudada em ratos.

2.6. QTLs associados à resistência a insetos em soja

Diversos estudos de QTLs associados à resistência da soja aos insetos desfolhadores foram publicados na literatura (Rector et al. 1998; Rector et al. 1999; Rector et al. 2000; Terry et al. 2000; Narvel et al. 2001; Hulburt et al. 2004; Walker et al. 2004; Boerma and Walker 2005; Guo et al. 2005; Komatsu et al. 2005; Guo et al. 2006; Zhu et al. 2006; Zhao et al. 2008; Zhu et al. 2008; Yesudas et al. 2010). No entanto, até o momento, apenas dois estudos envolvendo o mapeamento de QTLs em soja para resistência a percevejos foram descritos na literatura (Möller 2010; Santos 2012).

O QTL-M é considerado o principal QTL de resistência da soja a insetos coleópteros e lepidópteros, sendo inicialmente identificado em populações de mapeamento resultantes do cruzamento do genitor suscetível Cobb com a PI 229358 e a PI 171451, conferindo tanto mecanismos de resistência relacionados à antibiose ($R=22-28\%$) e à não-preferência ($R=37\%$) (Rector et al. 1998; Rector et al. 1999; Rector et al. 2000). Outros dois QTLs, considerados de menor efeito, foram descritos nessas mesmas populações, sendo o QTL-H, identificado nas três populações de mapeamento resultantes do cruzamento entre o genitor suscetível Cobb com as introduções de plantas PI 171451, PI227687 e PI 229358 (Rector et al. 1998; Rector et al. 1999; Rector et al. 2000), e o QTL-G, identificado na população do cruzamento de Cobb com a PI 229358 (Rector et al. 2000). O QTL-H confere mecanismos de resistência relacionados à não-preferência ($R=16\%$) e o QTL-G à antibiose ($R=19\%$). Posteriormente, o QTL-M foi confirmado em outros *backgrounds* genéticos como Jack x PI 229358 (Walker et al. 2002; Walker et al. 2004) e Benning x PI 229358 (Zhu et al. 2006), e também sua interação epistática com o QTL-G e QTL-H, o que confirma sua relevância como um QTL associado à resistência a insetos em soja. Apesar de QTL-G e QTL-H explicarem uma menor proporção fenotípica da resistência, linhagens que apresentaram o QTL-M e o QTL-H apresentaram menor desfolha pela lagarta-da-espiga em relação às linhagens somente com o QTL-H, enquanto linhagens com o QTL-M e o QTL-G apresentaram menor crescimento larval em relação às linhagens que apresentaram somente o QTL-G, demonstrando que deve haver alguma interação entre esses QTLs que intensifica a resistência aos insetos (Zhu et al. 2006; Zhu et al. 2008). Um mapeamento fino realizado na região do QTL-M possibilitou inferir que o mesmo está localizado em uma região de 0,52 cM entre os marcadores Sat_258 e Satt702 no GL M da soja, contribuindo para a seleção de genótipos de soja contendo esse QTL que confere alta resistência a insetos (Zhu et al. 2006). O QTL-H foi identificado em um intervalo de 0,5 cM, entre os marcadores Sat_122 e Satt541, e o QTL-G encontra-se entre os marcadores Satt472 e Satt191, em um intervalo de 4,4 cM (Narvel et al. 2001).

Posteriormente, um outro QTL de maior efeito foi identificado no GL E, conferindo mecanismos de antibiose e não-preferência, sendo denominado QTL-E (Hulburt et al. 2004). Esse QTL, inicialmente identificado em uma população proveniente do cruzamento entre PI 227687 x 'Cobb', foi mapeado próximo ao QTL denominado *Pb*, que controla o tipo de pubescência em soja. Os autores compararam a estrutura do tricoma e confirmaram que os genótipos que apresentaram tricomas com a extremidade do tipo afiada (*Pb*₋) demonstraram maior resistência ao ataque de insetos lepidópteros, em relação aos genótipos *pbpb*. Ortega et al. (2016) desenvolveram e avaliaram linhagens de Benning contendo diferentes combinações dos QTLs conferindo resistência a insetos. Os autores concluíram que linhagens contendo o QTL-M e QTL-E (Benning^{ME}) apresentaram melhor resposta ao ataque dos insetos, no caso do mecanismo antibiose. Já para o mecanismo de não-preferência, tanto Benning^{ME} quanto Benning^{MGHE} apresentaram as menores taxas de desfolha foliar.

Apesar dos programas de melhoramento de soja visando resistência a insetos utilizarem como fonte de resistência plantas com os QTLs descritos acima, muitas vezes a seleção baseada apenas em caracteres fenotípicos resulta na perda dos QTLs de menor efeito, sendo que a maioria das linhagens geralmente selecionadas apresenta apenas o QTL-M, de maior efeito, o que caracteriza uma descendência com uma menor resistência aos insetos em

relação ao genitor doador utilizado como fonte de resistência (Narvel et al. 2001). Além disso, como o QTL-M e QTL-E estão relacionados tanto ao modo de resistência de antibiose e não-preferência, o método escolhido para avaliar a resistência das linhagens interfere no modo de resistência que será selecionado, resultando na perda de alelos para antibiose ou não-preferência (Zhu et al. 2006). Isso torna mais evidente a importância da identificação de QTLs de interesse para o uso de marcadores moleculares na seleção assistida em programas de melhoramento.

Considerando a dificuldade na obtenção de cultivares resistentes a insetos via melhoramento convencional, a obtenção de plantas geneticamente modificadas contendo o gene *Bt* representa uma importante contribuição da engenharia genética para o melhoramento de plantas. As plantas modificadas são denominadas *Bt* porque recebem o gene da bactéria *Bacillus thuringiensis* Berliner, que produz as proteínas Cry, tóxicas para as larvas de lepidópteros, dípteros e coleópteros (Hongyu et al, 2000). Essas endotoxinas são digeridas pelas proteases liberadas no intestino dos insetos e liberam pequenos peptídeos (toxinas), capazes de associar-se às microvilosidades apicais das células do intestino do inseto, levando à formação de poros na membrana e, conseqüente, morte celular por choque osmótico (Sosa-Gómez et al. 2012). O uso de plantas transgênicas contendo o gene *Bt* iniciou-se na cultura do tomate, tabaco e algodão, em 1987, (Fischhoff et al. 1987; Umbeck et al. 1987; Vaeck et al. 1987), mas atualmente diversas culturas possuem cultivares ou híbridos contendo esse transgene, como o milho, algodão, batata, arroz, canola, alfafa e a soja. No Brasil, a primeira liberação de plantio comercial de plantas *Bt* ocorreu em 2005, com a liberação do algodão Bollgard™ 531 (CTNBio 2017). Já a soja *Bt* foi liberada para cultivo no Brasil em 2012, mas somente na safra 2013-2014 passou a ter significativa área de cultivo devido a barreiras de importação anteriormente impostas pela China. Desde o início da sua comercialização, a área de cultivo da soja *Bt* vem aumentando a cada safra e, hoje, representa quase 95% da área cultivada no Brasil (CONAB 2017). Apesar dos benefícios do uso da soja *Bt*, como a redução do uso de inseticidas, a soja *Bt* não é resistente a todos os insetos-praga. Além disso, a grande disseminação do uso da mesma tecnologia em diversas culturas tem causado preocupações como o aparecimento de populações de insetos resistentes. Não se pode deixar de destacar a importância do plantio de áreas de refúgio, com plantas que não possuem a tecnologia *Bt*, para garantir a existência de insetos suscetíveis à proteína *Bt*. Assim, quando a resistência for conferida por um alelo recessivo, o acasalamento cruzado entre as duas populações de insetos garantirá a existência de heterozigose, retardando a evolução da resistência na população (Sosa-Gómez et al. 2012). Dessa forma, outras fontes de resistência aos insetos-praga devem ser combinadas às estratégias já existentes, como a associação de QTLs relacionados à resistência a insetos-praga da soja às plantas de soja *Bt*. Segundo Walker et al. (2004), o QTL-M contribuiu para aumentar a eficácia do transgene *Bt cry1Ac* em soja, sendo que as linhagens contendo tanto o transgene quanto o QTL-M apresentaram menores danos em relação às linhagens contendo apenas o transgene. Recentemente, Warrington et al. (2008) realizaram a introgressão do QTL-M e do QTL-H na cultivar *Bt* Benning™, e não identificaram reduções na produtividade de grãos, como sugeriam estudos anteriores, devido ao arraste de genes indesejados juntamente com os QTLs de resistência aos insetos. Ortega et al. (2016) também avaliaram a piramidação dos QTLs de resistência a insetos, em soja, com o transgene *cry1Ac*. Os autores verificaram que as linhagens Benning^{ME+*cry1Ac*} demonstram maior resistência aos insetos em relação às linhagens Benning^{ME} e Benning^{*cry1Ac*}. Dessa forma, além de garantir uma maior proteção da planta contra insetos praga, a piramidação de QTLs que conferem resistência a insetos com o transgene *Bt* pode contribuir para a longevidade do uso dessa tecnologia, tendo em vista que diferentes mecanismos de resistência são utilizados (Boerma and Walker 2005; Zhu et al. 2006; Sosa-Gómez et al. 2012)

Até o presente momento, não há relatos de eventos transgênicos de soja e em nenhuma outra cultura visando o controle de percevejos. Nesta perspectiva, o desenvolvimento de cultivares de soja resistentes a estes insetos

tem sido um dos principais objetivos dos programas de melhoramento de soja. Assim, a identificação de QTLs de resistência à soja a percevejos é uma ferramenta importante para auxiliar no entendimento da base genética das características envolvidas no processo de resistência das plantas a esses insetos.

2.7. Estudos de mapeamento de QTLs previamente desenvolvidos

Para resistência a percevejos, o trabalho de mapeamento de QTLs desenvolvido por Möller (2010) foi o primeiro a ser relatado na literatura. Neste trabalho, uma população de 286 indivíduos F_2 , oriunda do cruzamento entre IAC-100 e CD-215, foi utilizada para o mapeamento. A cultivar IAC-100 foi desenvolvida pelo Instituto Agronômico de Campinas, do Estado de São Paulo, e apresenta resistência ao complexo de percevejos (Rossetto 1989; Pinheiro 1993; Carrao-Panizzi and Kitamura 1995), além de outras características agronômicas desejáveis como: resistência ao besourinho (*Diphaulaca viridipennis*), boa produtividade de grãos, baixa retenção foliar e poucas manchas nas sementes, mesmo na presença de uma grande infestação de percevejos (APASSUL 2017). A cultivar IAC-100 foi obtida a partir de cruzamentos entre as linhagens IAC78-2318 e IAC-12, sendo que a primeira apresenta múltiplas fontes de genes de resistência a insetos em soja (Rossetto et al. 1986; Lourenção and Miranda 1987). A cultivar CD-215 é uma cultivar protegida pela Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda, do Estado do Paraná. Essa cultivar é suscetível ao complexo de percevejos. Entretanto, apresenta resistência ao cancro da haste, resistência moderada ao oídio e tolerância à acidez do solo (COODETEC 2017). Essa cultivar foi obtida a partir do cruzamento entre OC88-127 e OC90-234.

Cruzamentos entre as cultivares IAC-100 e CD-215 foram realizados em casa de vegetação no Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz”, na Universidade de São Paulo (ESALQ–USP), entre novembro de 2007 e janeiro de 2008. A autofecundação natural da geração F_1 durante o inverno de 2008, em casa de vegetação, permitiu a obtenção da geração F_2 , utilizada na genotipagem dos marcadores moleculares para a construção do mapa genético. A população F_2 foi conduzida na Estação Experimental de Genética – Anhumas, do Departamento de Genética da ESALQ – USP, localizada no município de Piracicaba, São Paulo. Os caracteres agronômicos avaliados foram: número de dias para o florescimento (NDF), altura da planta no florescimento (APF), número de dias para a maturidade (NDM), altura da planta na maturidade (APM), acamamento (AC), valor agronômico (VA) e produtividade de grãos (PG). Dentre os caracteres associados à resistência aos insetos, foram avaliados: período de granação (PEG), retenção foliar (RF), índice percentual de dano nas vagens (IPDV), número de vagens por planta (NVP), número de sementes (NS), peso de sementes manchadas (PSM), peso de sementes boas (PSB) e peso de cem sementes (PCS).

A análise de ligação dos marcadores resultou em um mapa contendo 49 marcadores (35 AFLP e 14 TRAP), distribuídos por 12 GLs, com comprimento total de 696 cM. Möller (2010) utilizou a população F_2 para mapear QTLs utilizando a abordagem CIM e identificou oito QTLs para PSM, três QTLs para PSB, um QTL para PG e um QTL para VA. Considerando o fato da geração F_2 apresentar apenas uma planta por genótipo, Santos (2012) conduziu a geração $F_{2,3}$, na Estação Experimental do Departamento de Genética da ESALQ - USP, localizada no município de Anhumas, São Paulo. A avaliação das 15 características descritas acima foi realizada na geração $F_{2,3}$ para o mapeamento de QTLs. Foram selecionadas 15 sementes de cada planta F_2 , além dos genitores e de testemunhas (CD-208, IAC-17 e L.1.1.10), sendo que cada parcela era composta por 5 plantas, representando cada planta F_2 , com três repetições. A abordagem de mapeamento de QTLs, utilizada por Santos (2012), consistiu na estratégia de genotipagem seletiva. Uma seleção de 84 indivíduos foi realizada com base nas características: período de granação (PEG), peso de cem sementes (PCS), peso de sementes boas (PSB) e retenção foliar (RF). Nesse estudo, 337 SNP, 28 SSR, 13 TRAP e 41 AFLP

foram mapeados em 20 GLs e o comprimento total do mapa foi de 1.475 cM. O número de QTLs revelados para cada variável foi: quatro QTLs para as variáveis PEG, NDF, PCS, APM e PG; três QTLs para as variáveis APF, NDM, VA, NS e PSB; dois QTLs para as variáveis IPDV e NVP e um QTL para cada uma das variáveis RF e PSM. Os QTLs identificados localizam-se nos GLs C1, A1, C2, M, F, E, J, D2, L e I.

Esses dois estudos representaram uma etapa inicial deste projeto. Portanto, esse trabalho teve como objetivo aumentar o poder das estimativas do mapeamento de QTLs inicialmente realizado por Santos (2012), visando a identificação de QTLs de importância para o melhoramento da soja. Dessa forma, foi utilizada uma população de mapeamento com um maior número de indivíduos para a obtenção dos dados genotípicos. Além disso, foi feita a seleção de marcadores SSR no banco de dados do mapa da soja para reduzir os *gaps* observados no mapa inicial. A abordagem de mapeamento multivariado, utilizada neste trabalho, é inédita para a espécie. Os resultados obtidos neste estudo serão úteis na identificação de regiões promissoras para a seleção simultânea de características de interesse, além de auxiliarem no entendimento da base genética complexa das características avaliadas.

3. MATERIAL E MÉTODOS

3.1. Material vegetal

A população de 228 plantas F_2 utilizada neste trabalho, é proveniente da autofecundação de plantas F_1 , resultantes do cruzamento artificial, realizado em casa de vegetação, entre a cultivar IAC-100 (genitor resistente aos percevejos) e a cultivar CD-215 (genitor suscetível). Os cruzamentos e o avanço das plantas F_1 foram realizados em casa de vegetação do Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ/USP, em Piracicaba – SP, na safra 2007/2008 e inverno de 2008, respectivamente. A população F_2 foi conduzida por Möller (2010), na Estação Experimental de Genética – Anhumas, do Departamento de Genética da ESALQ – USP, localizada no município de Piracicaba, São Paulo, na safra 2008/2009.

Considerando o fato da geração F_2 apresentar apenas uma planta por genótipo, amostragens de 15 sementes de cada planta F_2 foram consideradas para a obtenção dos dados fenotípicos da geração $F_{2:3}$ para o mapeamento de QTLs. A população $F_{2:3}$ foi conduzida por Santos (2012), na safra 2009/2010, na Estação Experimental de Genética – Anhumas, do Departamento de Genética da ESALQ – USP, localizada no município de Piracicaba, São Paulo, utilizando-se o delineamento de blocos ao acaso.

3.2. Dados Fenotípicos da geração $F_{2:3}$

Os dados fenotípicos utilizados neste trabalho foram obtidos por Santos (2012), na avaliação da geração $F_{2:3}$, como já descrito. No total, foram avaliadas 15 características, sendo classificadas como sete características agronômicas e oito características diretamente relacionadas à resistência aos percevejos.

As sete características agronômicas avaliadas foram:

NDF: número de dias para o florescimento, contados da semeadura até a antese da primeira flor (R1) (Fehr and Caviness 1977);

APF: altura da planta (cm) no início do florescimento, medida desde a base da planta (no solo) até o ápice da haste principal;

NDM: número de dias para a maturidade, contados desde a semeadura até a data em que 95% das vagens apresentaram-se maduras (R8) (Fehr and Caviness 1977);

APM: altura da planta (cm) na maturidade, medida desde a base da planta (no solo) até o ápice da haste principal;

AC: acamamento, avaliado na maturidade por uma escala de notas visuais de 1 a 5, sendo a nota 1 correspondente à planta ereta e a 5 à planta acamada;

VA: valor agronômico, avaliado na maturidade através de uma escala de notas visuais de 1 a 5, sendo a nota 1 correspondente à planta sem nenhum valor agronômico e a nota 5 à planta com excelentes características agronômicas (grande número de vagens, altura superior a 60 cm, sem acamamento, ausência de hastes verdes e retenção foliar, sem debulha de vagens e sem sintomas de doenças);

PG: produtividade de grãos, avaliada pelo peso total de grãos (g) obtido por planta;

Já as oito características avaliadas, consideradas como diretamente associadas à resistência aos insetos, foram:

PEG: período de granação (dias), obtido pela diferença entre R7 (término do enchimento de grãos) e R5 (início do enchimento de grãos) (Fehr and Caviness 1977);

RF: retenção foliar, avaliada visualmente na maturidade por uma escala de notas visuais de 1 a 5, sendo a nota 1 a planta sem retenção foliar e 5 a planta com total retenção foliar (Godói et al. 2002);

NVP: número de vagens por planta, correspondendo ao somatório de todas as vagens presentes na planta (vagens planas, intermediárias e boas);

IPDV: índice percentual de dano nas vagens, correspondendo ao somatório da percentagem de vagens planas (sem enchimento de grãos) e metade da percentagem de vagens intermediárias (com falhas no enchimento de grãos) (Rossetto et al. 1986);

NS: número de sementes, contadas para cada planta individualmente após a colheita e beneficiamento dos grãos;

PSM: peso de sementes manchadas (g), para cada planta, sendo realizada a pesagem das sementes com manchas ou danos provocados pelos insetos e sementes com a presença da levedura *E. coryli* transmitida pelos mesmos;

PSB: peso de sementes boas (g), para cada planta, sendo realizada a pesagem das sementes sem aparentes danos provocados por percevejos;

PCS: peso de cem sementes (g), a partir de uma amostra ao acaso após padronização da umidade a 13%.

Apesar das características NVP e NS estarem relacionadas à produtividade da planta, ambas foram classificadas como características associadas à resistência aos insetos, pois refletem diretamente os danos ocasionados pelos percevejos. Conforme já descrito anteriormente, quando esses insetos se alimentam do grão no seu estágio de formação inicial (R5.1 a R5.3), ocorre o abortamento do mesmo, que é caracterizado como grão chocho, e as vagens apresentam-se planas, sem enchimento de grãos (Panizzi and Slansky Júnior 1985; Villas-Bôas et al. 1990; Brier and Rogers 1991; Gazzoni 1998; Panizzi et al. 2000; Corrêa-Ferreira and Azevedo 2002). A característica PCS também foi considerada como associada à resistência a percevejos, para essa população em estudo, pois estudos demonstram que genótipos resistentes a percevejos tendem a apresentar sementes menores, como é o caso da cultivar IAC-100 (Miranda et al. 1979; Lourenção et al. 1985; Panizzi et al. 1986).

3.2.1. Análises Fenotípicas

O experimento da geração $F_{2.3}$ utilizado para a avaliação dos caracteres fenotípicos foi conduzido por Santos (2012) em blocos ao acaso, com três repetições. No total, 228 plantas F_2 foram representadas por parcelas contendo cinco plantas da geração $F_{2.3}$, em um experimento contendo três repetições, totalizando 3.420 plantas $F_{2.3}$ avaliadas. As avaliações foram conduzidas para as cinco plantas de cada parcela, sendo, posteriormente, realizada a média da parcela experimental. Para garantir a infestação de percevejos nenhum controle químico para insetos foi utilizado durante a condução do experimento. Além disso, o experimento estava estrategicamente localizado para permitir a migração de insetos das áreas próximas.

Os dados fenotípicos obtidos como nota foram transformados utilizando-se \sqrt{x} , sendo as variáveis transformadas: acamamento (AC), valor agrônômico (VA) e retenção foliar. A variável índice percentual de dano nas vagens (IPDV), avaliada como percentagem, foi transformada utilizando-se $\sqrt{x/100}$. A análise de variância foi realizada por Santos (2012), utilizando-se o procedimento de modelos mistos no programa SAS (SAS INSTITUTE 2001). As variâncias genética e fenotípica foram estimadas considerando-se a média de progênies.

Os valores genotípicos e fenotípicos para cada característica foram obtidos pelo seguinte modelo estatístico:

$$Y_{ij} = \mu + B_i + G_j + \varepsilon_{ij} \quad (1)$$

em que,

Y_{ij} é a observação fenotípica do indivíduo j ($j = 1, 2, \dots, 228$) no bloco i ($i = 1, 2, \dots, r = 3$);

μ é o intercepto ou a média geral do experimento;

B_i é o efeito fixo do i -ésimo bloco;

G_j é o efeito aleatório do j -ésimo indivíduo;

ε_{ij} é o resíduo ou o erro associado ao Y_{ij}

O ajuste do modelo da Equação 1 foi realizado considerando resíduos independentes e identicamente distribuídos ($\varepsilon_{ij} \sim N(0, \hat{\sigma}_e^2)$). A variância genética das progênies ($\hat{\sigma}_g^2$) e a variância fenotípica foram estimadas com base na média de progênies ($\hat{\sigma}_p^2$), de acordo com as equações 2 e 3.

$$\hat{\sigma}_g^2 = \frac{QM_G - QM_E}{R} \quad (2)$$

$$\hat{\sigma}_g^2 = \frac{QM_G}{R}. \quad (3)$$

Sendo que, QM_G e QM_E correspondem aos quadrados médios das progênies $F_{2,3}$ e do resíduo, e R o número de repetições (Searle et al. 1992).

As herdabilidades no sentido amplo, ao nível de médias das parcelas, foram estimadas de acordo com a Equação 4.

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_p^2} \quad (4)$$

Foram calculadas as correlações de Pearson, entre os valores genéticos preditos das variáveis estudadas, no pacote *Performance Analytics* (Peterson et al. 2015), disponível no programa R (R Development Core Team 2017), para auxiliar na tomada de decisões do mapeamento de múltiplas características.

3.3. Extração do DNA

A extração do DNA de 228 plantas F_2 e dos genitores IAC-100 e CD-215 foi realizada a partir do material vegetal liofilizado armazenado em *freezer* -80°C para garantir uma boa qualidade e integridade do mesmo para a genotipagem dos marcadores. O DNA genômico total foi extraído de acordo com o método CTAB (*Cationic hexadecyl trimethyl ammonium bromide*) (Doyle and Doyle 1990) e a qualidade e concentração do DNA extraído foram avaliadas por eletroforese em gel de agarose 1% tratado com *SYBR Safe*, por comparação com o DNA-padrão do fago *lambda*.

3.4. Dados genotípicos da geração F_2

3.4.1. Marcadores SNP

O painel universal de ligação da soja (*Universal Soy Linkage Panel* - USLP 1.0) contendo 1.536 SNP, desenvolvido por Hyten et al. (2010), foi utilizado para a genotipagem dos 228 indivíduos F_2 e dos genitores IAC-100 (resistente) e CD-215 (suscetível), na plataforma Golden Gate (Illumina, San Diego, CA). As sequências utilizadas para o desenvolvimento de cada marcador SNP encontram-se disponíveis no Soybase (<http://www.soybase.org>) (Grant et al. 2010). A metodologia utilizada foi a mesma descrita em Fan et al. (2003) e Fan et al. (2006), e a genotipagem foi

realizada no laboratório do Dr. Perry Cregan, no Centro de Pesquisa em Agricultura da USDA, Beltsville, MD, US. O Illumina BeadStation 500G foi utilizado para determinar diferenças genotípicas referentes à plataforma Golden Gate.

O programa GenCall foi utilizado para verificar a intensidade da fluorescência Cy3 e Cy5, e assim determinar a presença de polimorfismo entre os genitores. Posteriormente, os locos polimórficos foram individualmente checados, no programa Genome Studio, para verificar a segregação da população F₂ e visualizar a separação da população em *clusters* de acordo com a emissão da fluorescência, evitando qualquer possível erro na leitura dos dados. A genotipagem seguiu o seguinte critério: “A” para indivíduos contendo o alelo do genitor resistente em homozigose, “B” para os indivíduos contendo o alelo do genitor susceptível em homozigose e “H” para os indivíduos heterozigotos para o marcador. Para os indivíduos que apresentaram falha na amplificação ou dúvida quanto à intensidade de fluorescência, o sinal “-“ foi utilizado.

Considerando que os dados das análises foram recebidos em três etapas diferentes (1ª genotipagem: 94 indivíduos, 2ª genotipagem: 94 indivíduos e 3ª genotipagem: 40 indivíduos), foi necessário realizar a análise separadamente, nesta etapa, para cada conjunto de indivíduos utilizados para a genotipagem. No entanto, o DNA dos genitores foi utilizado nas três etapas de genotipagem como DNA padrão para garantir a correta leitura dos dados. Uma rigorosa verificação dos dados foi realizada e, sempre que qualquer dúvida surgisse em relação aos genitores ou separação dos clusters, os dados de SNP eram imediatamente descartados para toda a população.

3.4.2. Marcadores SSR

O mapa previamente obtido por Santos (2012) apresentou alguns *gaps* em alguns GLs da soja. Para ampliar a cobertura do mapa e reduzir esses *gaps*, 300 locos SSR foram selecionados e sintetizados a partir das sequências disponíveis no SoyBase (<http://www.soybase.org>) (Grant et al. 2010). Para a seleção dos SSR, foram considerados os seguintes critérios: a sua localização nos GLs da soja, unidades de repetição constituídas por trinucleotídeos e valores de PIC (Polymorphism Information Content). Para cada par de *primer*, a sequência original do *primer* Forward foi modificada com a adição da cauda universal M13(-19) (5'- CACGACGTTGTAAAACGAC-3') na extremidade 5' (Schuelke 2000). A sequência do *primer* universal M13(-21) foi sintetizada com fluorescência 700 ou 800 nm para permitir a detecção da amplificação do loco pela emissão da fluorescência.

Uma genotipagem preliminar utilizando os genitores IAC-100 e CD-215 foi realizada para verificar a qualidade de amplificação de cada loco e para seleção dos locos polimórficos. As reações de amplificação foram realizadas em volume total de 15µl contendo 20ng de DNA genômico, 0,3µM do *primer* universal M13 *Forward* (IR-Dye 700 ou 800), 0,16µM do *primer Forward* com a cauda M13, 0,2µM do *primer Reverse*, 0,15mM de cada dNTP, 1X a solução tampão de PCR contendo 50mM de KCl e 10mM de Tris-HCl, pH 8,9, 2,0 mM de MgCl₂, 1X BSA e 1 unidade da enzima Taq DNA polimerase. O ciclo total de amplificação foi realizado em termociclador BioRad programado para iniciar com 5 minutos a 94°C, seguido por 10 ciclos de 1 minuto de denaturação a 94°C, 1 minuto de pareamento a 50°C, com decréscimo de 1°C por ciclo, 1 minuto de extensão a 72° C, 30 ciclos com 40 segundos de desnaturação a 94°C, 40 segundos de pareamento a 40°C e 1 minuto de extensão a 72° C. A extensão final foi de 10 minutos a 72°C. Os produtos da PCR foram separados por eletroforese no sistema Li-Cor® 4300 (Biosciences) e o tamanho dos alelos estimado por meio de marcador de peso molecular conhecido (Li-Cor IRDye 50-350pb).

Para a definição dos marcadores que seriam utilizados neste estudo, foi realizada uma rigorosa seleção da qualidade de amplificação dos locos, além de verificar o polimorfismo nos genitores. No total, 68 locos foram selecionados para a obtenção dos dados genotípicos da população F₂ e construção do mapa de ligação (APÊNDICE

A). A genotipagem seguiu o seguinte critério: “A” para indivíduos contendo o alelo do genitor resistente em homozigose, “B” para os indivíduos contendo o alelo do genitor susceptível em homozigose e “H” para os indivíduos heterozigotos para o marcador, assim como foi feito para os marcadores SNP.

3.4.3. Marcadores AFLP

Os marcadores AFLP utilizados neste estudo foram previamente obtidos por Möller (2010). O protocolo utilizado seguiu a metodologia proposta por Vos et al. (1995), com modificações. A partir da digestão do DNA de cada indivíduo da população F₂ e dos genitores, adaptadores específicos aos cortes das enzimas *EcoRI* e *MseI* foram ligados às extremidades dos fragmentos de DNA digerido. A pré-amplificação foi realizada com iniciadores para cada uma das enzimas, sem adição de nucleotídeos seletivos. O produto obtido foi verificado em gel de agarose 1% tratado com *SYBR Safe*. A amplificação seletiva foi realizada com iniciadores contendo três nucleotídeos seletivos na extremidade 3', sendo ‘E’ os *primers* correspondentes ao iniciador *EcoRI* e ‘M’ os *primers* correspondentes ao iniciador *MseI*. No total, foram utilizadas 14 combinações de *primers Eco*+3 e *Mse*+3 (ANEXO A), a partir de um *screening* realizado nos genitores. Os produtos de amplificação dos marcadores AFLP foram separados sob condições desnaturantes em gel de poliacrilamida 7%, uréia 7M, TBE 1X, por aproximadamente 4 horas a uma potência constante de 70W. Os fragmentos foram detectados por coloração em nitrato de prata, segundo protocolo descrito por Merrill et al. (1981), com as modificações propostas por Creste, Neto and Figueira (2001), e seus tamanhos estimados por meio de comparação com marcadores de pares de base. As marcas obtidas foram lidas como presença (1) e ausência (0), por se tratar de um marcador dominante.

3.4.4. Marcadores TRAP

Os marcadores TRAP utilizados neste trabalho foram previamente obtidos por Möller (2010). Foram utilizadas 11 combinações de *primers* fixo/arbitrário (ANEXO A). Os *primers* arbitrários, utilizados como *Reverse*, foram obtidos na literatura, sendo denominados Ga5, T03, T04 (Liu et al. 2005), Odd-26 (Zhi-Wei et al. 2007) e Arbi 2 (Alwala et al. 2006). As sequências dos *primers* fixos, utilizados como *Forward*, foram desenhadas por Möller (2010), com o auxílio do programa *Primer 3* (Rozen and Skaletsky 2000), a partir de sequências de genes da soja obtidas no NCBI (*National Center for Biotechnology Information*) (NCBI 2017), pela busca por palavras-chave associadas a genes envolvidos na rota metabólica de resistência a insetos. As sequências dos *primers* fixos e aleatórios utilizados encontram-se no ANEXO B.

Os produtos de amplificação dos marcadores TRAP foram separados sob condições desnaturantes em gel de poliacrilamida 7%, uréia 7M, TBE 1X, por aproximadamente 5 horas a uma potência constante de 70W. Os fragmentos foram detectados por coloração em nitrato de prata, segundo protocolo Merrill et al. (1981), com as modificações propostas por Creste et al. (2001), e seus tamanhos estimados por meio de comparação com marcadores de pares de bases. As marcas obtidas foram lidas como presença (1) e ausência (0), por se tratar de um marcador dominante.

3.5. Mapa genético

A ordem dos marcadores e a distância entre os mesmos, em cada GL, foi definida utilizando-se o pacote *OneMap* (Margarido et al. 2007). Os marcadores SSR e SNP utilizados neste trabalho encontram-se disponíveis no Soybase (<http://www.soybase.org>), e os mesmos já se encontram mapeados nos 20 GLs da soja (Grant et al. 2010). Portanto, para a construção do mapa na população em estudo, os marcadores SSR e SNP foram inicialmente designados para seus respectivos GLs, de acordo com o banco de dados.

Posteriormente, os marcadores foram ordenados dentro de cada GL, pois a ordem dos marcadores pode apresentar pequenas alterações em relação à ordem no mapa da soja (ConsensusMap 4.0). A função *order.seq* foi utilizada para automatizar a construção do mapa pelas funções *compare* e *try.seq*, que comparam todas as possíveis ordens dos marcadores e inclui o marcador no mapa de ligação sempre na posição de maior verossimilhança. Por fim, o comando *ripple* foi utilizado para verificar a possibilidade de modificar a ordem de algum marcador. Os *beatmaps* gerados foram analisados para um diagnóstico da posição dos marcadores em cada GL.

Após a obtenção do mapa com os marcadores co-dominantes, os marcadores dominantes, AFLP e TRAP, foram testados um a um para verificar se poderiam ser incluídos no mapa, utilizando-se a função *try.seq*. Após a inclusão de alguns marcadores dominantes no mapa de ligação, foi utilizado novamente o comando *ripple* para verificar a possibilidade de modificar a ordem de algum marcador.

Para a apresentação final do mapa, as frações de recombinação foram convertidas em distâncias genéticas e expressas em centiMorgans (cM), utilizando-se a função de mapeamento de Kosambi (1944). Para a visualização gráfica do mapa, foi utilizado o programa MapChart (Voorrips 2002).

3.6. Mapeamento de QTLs

O mapeamento de QTLs foi realizado utilizando-se o modelo de mapeamento de intervalos múltiplos univariado (MIM – *Multiple Interval Mapping*), considerando cada característica individualmente, e multivariado (MT-MIM – *Multiple Trait Multiple Interval Mapping*), considerando um conjunto de características simultaneamente. Na abordagem MIM (Kao and Zeng 1997; Kao et al. 1999; Zeng et al. 1999), o seguinte modelo de mapeamento foi empregado para cada uma das variáveis analisadas:

$$y_i = \mu + \sum_{r=1}^m \beta_{1r} x_{ir} + \sum_{r=1}^m \beta_{2r} z_{ir} + \varepsilon_i \quad (5)$$

em que,

y_i é o valor fenotípico do i -ésimo indivíduo ($i = 1, 2, \dots, 228$);

μ é o intercepto ou a média geral;

β_{1r} é o coeficiente da regressão que mede o efeito aditivo do r -ésimo QTL incluído no modelo ($r = 1, 2, \dots, m$);

x_{ir} é uma variável explanatória, definida de acordo com o modelo genético de Cockerham (Kao et al. 1999; Kao and Zeng 2002). Para cada indivíduo i da população $F_{2,3}$, a variável x_{ir} assume valor de 1, 0 ou -1 se o r -ésimo QTL em genótipo homocigoto dominante (QQ), heterocigoto (Qq) ou homocigoto recessivo (qq), respectivamente. As probabilidades dos genótipos do QTL foram calculadas de acordo com cada intervalo de marcadores no mapa genético de ligação, condicionalmente às informações dos marcadores flanqueadores;

β_{2r} é o coeficiente da regressão que mede o efeito de dominância do r -ésimo QTL incluído no modelo ($r = 1, 2, \dots, m$);

z_{ir} é uma variável explanatória, definida de acordo com o modelo genético de Cockerham (Kao et al. 1999; Kao and Zeng 2002). Para cada indivíduo i da população $F_{2,3}$, a variável z_{ir} assume valor de $1/2$ se o r -ésimo QTL tem genótipo heterozigoto (Qq) ou $-1/2$ caso contrário.

ε_i é o resíduo do i -ésimo indivíduo. Assume-se que os resíduos são independentes e identicamente distribuídos ($\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$).

Considerando o grande número de variáveis utilizadas neste trabalho, para reduzir a complexidade das análises do MT-MIM, foram selecionadas, inicialmente, variáveis que apresentaram valores de correlação igual ou superior a 0,6, significativas a 5% de probabilidade pelo teste t de *Student*. Após a análise dos resultados iniciais do MT-MIM e análise dos QTLs identificados no mapeamento univariado, alguns outros conjuntos de características foram selecionados para o mapeamento MT-MIM, considerando a correlação e a relevância de associação entre as mesmas. É importante ressaltar que outras opções de agrupamento entre as variáveis foram testadas, com, no máximo, cinco variáveis por grupo. Foram analisadas, no máximo, cinco variáveis por grupo para garantir um ajuste adequado no *OneQTL*. No entanto, os grupos definidos abaixo foram os que apresentaram o maior número de regiões genômicas reveladas pelo método MT-MIM. Considerando que o objetivo da análise MT-MIM consistiu em identificar o maior número de regiões genômicas controlando as características estudadas, foram definidos nove grupos de variáveis, para os quais foi realizado o mapeamento MT-MIM, sendo eles:

Multi-1: número de dias para o florescimento (NDF), número de dias para a maturidade (NDM), altura da planta na maturidade (APM), número de sementes (NS) e produtividade de grãos (PG);

Multi-2: altura da planta no florescimento (APF), altura da planta na maturidade (APM) e valor agrônômico (VA);

Multi-3: número de dias para a maturidade (NDM) e peso de sementes boas (PSB);

Multi-4: altura da planta na maturidade (APM) e número de vagens por planta (NVP);

Multi-5: período de granação (PEG); peso de sementes manchadas (PSM) e peso de cem sementes (PCS);

Multi-6: número de vagens por planta (NVP), número de sementes (NS), produtividade de grãos (PG) e valor agrônômico (VA);

Multi-7: número de vagens por planta (NVP), peso de sementes boas (PSB) e peso de sementes manchadas (PSM);

Multi-8: índice percentual de dano nas vagens (IPDV), número de sementes (NS), peso de sementes boas (PSB) e produtividade de grãos (PG);

Multi-9: período de granação (PEG), índice percentual de dano nas vagens (IPDV), peso de cem sementes (PCS) e produtividade de grãos (PG);

Na abordagem MT-MIM (Silva and Zeng 2010; Silva et al. 2012b), o seguinte modelo de mapeamento foi empregado para cada conjunto de características selecionadas:

$$y_{ti} = \mu_t + \sum_{r=1}^m \beta_{1tr} x_{ir} + \sum_{r=1}^m \beta_{2tr} z_{ir} + \varepsilon_{ti} \quad (6)$$

em que,

y_{ti} é o valor fenotípico do i -ésimo indivíduo ($i = 1, 2, \dots, 228$) para a t -ésima característica ($t = 1, 2, \dots, T$);

μ_t é o intercepto ou a média geral da t -ésima característica;

β_{1tr} é o coeficiente da regressão que mede o efeito aditivo do r -ésimo QTL incluído no modelo ($r = 1, 2, \dots, m$) para a t -ésima característica;

x_{ir} é uma variável explanatória, definida de acordo com o modelo genético de Cockerham (Kao et al. 1999; Kao and Zeng 2002). Para cada indivíduo i da população $F_{2,3}$, a variável x_{ir} assume valor de 1, 0 ou -1 se o r -ésimo QTL tem genótipo homocigoto dominante (QQ), heterocigoto (Qq) ou homocigoto recessivo (qq), respectivamente. As probabilidades dos genótipos do QTL foram calculadas de acordo com cada intervalo de marcadores no mapa genético de ligação, condicionalmente às informações dos marcadores flanqueadores;

β_{2r} é o coeficiente da regressão que mede o efeito de dominância do r -ésimo QTL incluído no modelo ($r = 1, 2, \dots, m$);

z_{ir} é uma variável explanatória, definida de acordo com o modelo genético de Cockerham (Kao et al. 1999; Kao and Zeng 2002). Para cada indivíduo i da população $F_{2,3}$, a variável z_{ir} assume valor de $1/2$ se o r -ésimo QTL tem genótipo heterocigoto (Qq) ou $-1/2$ caso contrário.

ε_{ti} é o resíduo do i -ésimo indivíduo na t -ésima característica. Assume-se que os resíduos são independentes e identicamente distribuídos ($\varepsilon_{it} \sim MNV(0, \Sigma_\varepsilon)$).

Todas as análises de QTL, incluindo a função *fdmain* e *backmain* e seus argumentos, foram realizadas utilizando-se o pacote *Oneqtl*, no programa R (R Development Core Team 2017). O pacote foi desenvolvido pelo Laboratório de Genética Estatística do Departamento de Genética da ESALQ e será disponibilizado, em breve, no Github e no CRAN.

Para o mapeamento univariado e multivariado foi realizada uma busca *forward* para os efeitos principais (efeitos aditivos e de dominância), utilizando-se a função *fdmain*, numa grade de posições genômicas de 1 cM (*sparse.grid* = 1) e uma janela de exclusão de busca por QTLs à direita e à esquerda de cada QTL presente no modelo de 5 cM (*msize* = 5). Os modelos das Equações 3 e 4 foram ajustados usando-se a aproximação de Haley-Knott (Haley and Knott 1992) e a significância para inclusão de QTLs no modelo a nível genômico (*genome-wide*) foi estabelecida em 10% (*sig.score* = 0.1) pela metodologia de reamostragem *score* (Zou et al. 2004; Silva and Zeng 2010; Laurie et al. 2014). As posições e os efeitos dos QTLs incluídos no modelo foram reestimadas com um passo de refinamento do modelo a cada busca por QTLs ao longo do genoma (*freq.refine* = 1). Além disso, uma etapa final foi realizada, após a inclusão de todos os QTLs no modelo, para refinar a posição dos mesmos, segundo a proposta original do método MIM (Kao and Zeng 1997; Kao et al. 1999; Zeng et al. 1999). Para a obtenção do modelo de mapeamento final, este procedimento foi repetido até que o próximo QTL candidato a entrar no modelo não excedesse o nível de significância pré-estabelecido.

Utilizando-se a abordagem descrita acima, para cada conjunto de variáveis analisadas, um modelo final foi selecionado e utilizado para estimar a magnitude dos efeitos principais para cada QTL, a proporção da variância fenotípica explicada pelos QTLs identificados, dada pela razão da soma das variâncias e covariâncias do efeito do QTL sobre a variância fenotípica vezes 100 (%VarCov), o coeficiente de determinação (R^2), dado pela soma das %VarCov de cada QTL, e o perfil de LOD ao longo dos cromossomos. Posteriormente, apenas os efeitos marginalmente significantes, para cada variável estudada individualmente, no mapeamento univariado, e para cada variável do conjunto analisado, foram considerados. Assim, o modelo da Equação 3, para o mapeamento univariado, e o modelo da Equação 4, para o mapeamento multivariado, obtidos pelo procedimento descrito acima, foram submetidos a um procedimento de eliminação *backward*. A função utilizada foi a *backmain*, com um valor de significância a nível genômico de 1% (*sig.score* = 0.01). Como todos os efeitos presentes no modelo já haviam sido previamente testados com proteção para

múltiplos testes na função *fvdm* e o nível de significância para permanência no modelo foi estridente, a eliminação *backward* não considerou nenhuma medida de controle de múltiplos testes (*multi.protes = FALSE*). Por fim, uma etapa de final foi realizada, para refinar a posição dos QTLs, segundo a proposta original do método MIM (Kao and Zeng 1997; Kao et al. 1999; Zeng et al. 1999).

A partir dos QTLs identificados no presente estudo foi feita uma comparação com o banco de dados de QTLs da soja, no Soybase, para verificar a presença de QTLs já descritos para essas variáveis. No caso da abordagem multivariada, a busca por QTLs foi realizada somente para as regiões genômicas que apresentaram efeitos significativos, para cada variável, considerando-se os resultados obtidos com a abordagem MT-MIM após a eliminação *backward*. Para padronizar as comparações, os marcadores flanqueando os QTLs identificados foram localizados no ConsensusMap4.0. Uma região de 20cM foi considerada para a comparação, para ambos os lados, a partir da posição dos marcadores. Como as populações utilizadas no mapeamento dos diferentes QTLs descritos no Soybase possuem diferentes *background* genéticos, e as posições dos marcadores pode sofrer alguma alteração em cada mapa genético utilizado no mapeamento de QTLs, optou-se por selecionar uma região mais abrangente para não superestimar o número de QTLs revelados nessa população em estudo, em relação aos QTLs já descritos na literatura. No entanto, como uma ampla região está sendo considerada, de no mínimo 40 cM, é evidente que não se pode afirmar, somente com base nessa comparação, que os QTLs identificados neste estudo coincidem com os QTLs descritos no Soybase.

4. RESULTADOS E DISCUSSÃO

4.1. Análises Fenotípicas

É importante ressaltar que houve infestação significativa no experimento da população $F_{2,3}$, conduzido na safra 2009/2010. Pode-se verificar na Figura 1 que houve um aumento significativo no número de percevejos observados, no período de enchimento de grãos (R5 a R7), chegando a atingir uma média de quase 19 insetos nas avaliações de pano de batida realizadas semanalmente por Santos (2012). No final do período R5, considerado o período crítico em que os grãos estão mais suscetíveis ao dano devido ao ataque dos percevejos, o número de insetos observados ficou acima do nível de dano econômico. De acordo com um estudo comparando-se a infestação nos estádios R5, R6 e R7 com as espécies de percevejos *P. guildinii*, *N. viridula*, *D. melacanthus* e *E. heros*, as plantas infestadas no estágio R5 são as que apresentam os danos com maior intensidade, enquanto que plantas infestadas na maturação fisiológica R7 apresentam sementes de qualidade semelhante às plantas testemunhas, sem infestação por percevejos (Corrêa-Ferreira et al. 2009).

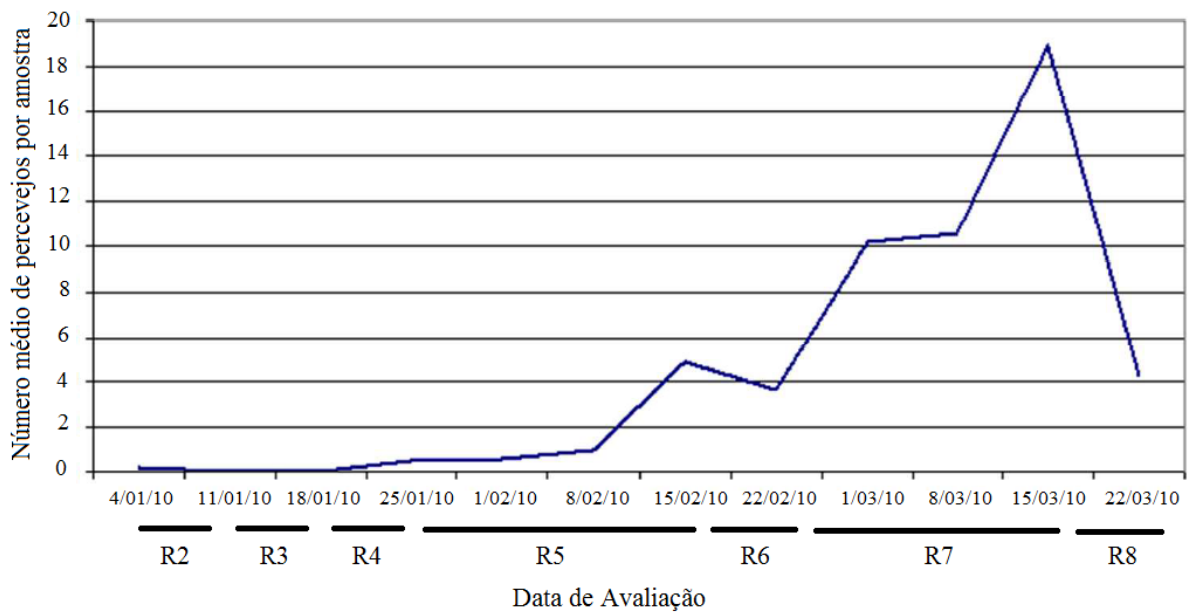


Figura 1. Número médio de percevejos por avaliação realizada semanalmente, entre os estádios R2 a R8, no campo experimental da população $F_{2,3}$, pelo método do pano de batida (dados obtidos por Santos (2012)).

A análise de variância realizada por Santos (2012) demonstrou que para as 15 variáveis estudadas todos os fenótipos avaliados apresentaram variação genotípica significativa na população $F_{2,3}$, exceto para acamamento (AC). O ANEXO C apresenta as estimativas dos valores de variância fenotípica, genética e residual obtidas para cada variável (Santos 2012). Na Tabela 1 pode-se verificar os valores da distribuição de cada variável na população $F_{2,3}$, a média da população, a média de cada genitor, a estimativa da herdabilidade (\hat{h}^2) e o coeficiente de variação (CV). Os valores estimados de herdabilidade variaram de 8,25% para acamamento (AC) a 90,02% para altura de planta na maturidade (APM). Os valores de herdabilidade estimados para período de granação (PEG) e índice percentual de dano nas vagens (IPDV) foram coerentes com os resultados obtidos por Godoi and Pinheiro (2009), que obtiveram estimativas de 74,7% e 49,6%, respectivamente. Para a variável número de sementes (NS), o valor de herdabilidade obtido neste trabalho, de 0,69, foi próximo ao valor obtido por Liang et al. (2010), de 0,79. No caso da variável produtividade de

grãos (PG), que é uma característica mais complexa, diferentes estimativas de herdabilidade foram descritas na literatura, variando de 0,47 (Yuan et al. 2002) a 0,90 (Hu et al. 2014). O valor obtido por Wang et al. (2014), de 0,65, foi próximo ao valor da estimativa de herdabilidade obtida neste trabalho, de 0,71.

Para a variável peso de cem sementes (PCS), Niu et al. (2013) relatam que características relacionadas ao tamanho da semente, que interferem diretamente na variável PCS, apresentam herdabilidade que variam entre 0,40 a 0,84. Portanto, a estimativa de herdabilidade observada para PCS de 0,75 está de acordo com a literatura. Já Hu et al. (2014) observaram valores mais elevados de herdabilidade para essa variável, de 0,91. Esses mesmos autores também observaram valores mais elevados de herdabilidade, em relação aos obtidos neste trabalho, para as variáveis NDF (0,98) e NDM (0,96), e um menor valor da estimativa de herdabilidade para a variável NVP (0,40).

Os coeficientes de variação estimados, no geral, apresentaram valores abaixo de 20%, o que indicam uma boa precisão ambiental (Pimentel-Gomes 2000). No entanto, as variáveis peso de sementes boas (PSB), produtividade de grãos (PG) e peso de sementes manchadas (PSM) apresentaram valores mais elevados de CV, sendo 24,47%, 28,97% e 62,08%, respectivamente. Esses resultados podem ter ocorrido devido à complexidade de tais características que estão diretamente relacionadas ao ataque dos insetos, além de sofrerem grande influência ambiental.

Tabela 1. Resumo dos dados obtidos por Santos (2012) apresentando a média obtida de cada variável para os genitores IAC-100 e CD-215, a média da população F_{2,3}, a distribuição dos valores para a população, as estimativas de herdabilidade genética (\hat{h}_G^2) e o coeficiente de variação (CV).

Caracteres	Média			Amplitude População	\hat{h}_G^2 (%)	CV(%)
	IAC-100	CD-215	População			
Resistência						
PEG (dias)	38,48	43,05	41,70	35,80 - 46,53	71,56 (64,25 - 77,17)	16,22
RF (nota)	1,43	1,60	1,39	1,00 - 2,17	20,60 (0,21 - 36,28)	12,08
NVP	148,42	89,27	122,67	77,93 - 261,83	69,72 (62,02 - 75,75)	18,52
IPDV (%)	16,61	34,22	26,90	16,23 - 36,46	53,14 (41,10 - 62,39)	12,06
NS	278,45	143,09	208,33	123,13 - 418,00	68,72 (60,69 - 74,90)	20,53
PSM (g)	2,64	2,23	3,40	1,17 - 9,86	40,33 (25,01 - 52,12)	62,08
PSB (g)	24,09	13,71	19,20	10,84 - 57,51	69,42 (61,57 - 75,46)	24,47
PCS (g)	9,87	11,55	11,73	8,26 - 14,89	75,00 (68,58 - 79,94)	10,68
Agronômicos						
NDF (dias)	52,14	43,06	45,42	42,60 - 50,40	83,41 (79,15 - 86,69)	2,82
APF (cm)	26,01	30,54	30,13	22,67 - 35,53	81,12 (76,27 - 84,85)	7,38
NDM (dias)	126,43	113,67	117,93	109,8 - 128,73	80,96 (76,07 - 84,72)	2,54
APM (cm)	34,91	32,03	33,78	24,37 - 49,00	90,02 (87,46 - 91,99)	7,42
AC (nota)	1,34	1,45	1,18	1,06 - 2,47	8,25 (-15,30 - 26,37)	16,22
VA (nota)	2,97	2,88	2,93	2,13 - 3,86	63,17 (53,71 - 70,44)	5,25
PG (g)	26,74	15,94	22,63	12,66 - 64,18	70,90 (63,43 - 76,65)	28,97

As correlações fenotípicas obtidas entre as variáveis analisadas podem ser visualizadas na Figura 2. Os histogramas e as densidades das distribuições dos valores das características avaliadas são apresentados na diagonal do correlograma. Pode-se verificar que a maioria das correlações apresentou significância a 1% ou 5% de probabilidade pelo Teste t. Foram observados valores de correlação quase próximos a 1, que indicam máxima correlação, e mesmo

sentido, entre as variáveis. Os dois valores de correlação mais altos observados foram entre peso de sementes boas (PSB) e produtividade de grãos (PG) ($r = 0,96^{***}$) e número de vagens por plantas (NVP) e número de sementes (NS) ($r = 0,94^{***}$). A alta correlação entre essas variáveis era esperada, uma vez que um maior peso de sementes boas (PSB) vai refletir diretamente em uma maior produtividade de grãos (PG). Da mesma forma, quanto maior o número de vagens por planta (NVP) maior o número de sementes (NS). Estudos realizados por Rocha et al. (2014), Pereira (2015) e Rocha (2015) também indicam alta correlação entre as variáveis PSB e PG. Outras variáveis não apresentaram nenhuma correlação significativa moderada ou forte, como é o caso da retenção foliar (RF). Já acamamento (AC), embora tenha apresentado uma correlação significativa a 1% e outras duas a 5% de probabilidade, todas apresentaram valores indicando fraca correlação. Essas duas características, RF e AC, foram as que apresentaram os menores valores de herdabilidade, 0,21 e 0,08, respectivamente, para a população em estudo.

Na Tabela 1, é possível verificar que a cultivar IAC-100 apresentou menor período de granação (PEG) em relação à cultivar CD-215. O dado obtido corrobora com os resultados apresentados por Rossetto et al. (1995) e Piubelli et al. (2005). O período de enchimento de grãos é o de maior suscetibilidade da cultura ao ataque dos percevejos. Portanto, um menor período de granação pode ser associado ao que se denomina evasão hospedeira, que é uma forma de pseudo-resistência, resultando em uma menor quantidade de dano nos grãos (Painter 1951; Rossetto and Lara 1991). Pelos resultados obtidos, também é possível observar que a cultivar IAC-100 obteve maior número de vagens por planta (NVP) e maior número de sementes (NS). Ambas as variáveis contribuíram para a maior produtividade de grãos (PG) da cultivar IAC-100 em relação à cultivar CD-215. Rossetto et al. (1995) e Piubelli et al. (2005) relataram que a cultivar IAC-100 apresenta o mecanismo de tolerância, pois possui a capacidade de abortar vagens danificadas e substituí-las por novas, recuperando-se dos danos sofridos pelo ataque dos percevejos. Esse mecanismo de pseudo-resistência é denominado diluição de dano, pois, apesar de possuir semente de tamanho pequeno (menor peso de cem sementes - PCS), a quantidade de semente produzida por planta é geralmente maior em relação às outras cultivares, o que contribui para garantir uma boa produtividade final de grãos sem dano. Estudos demonstram que genótipos resistentes a percevejos tendem a apresentar sementes menores, e a maior quantidade de sementes contribui para a diluição do dano ocasionado em algumas sementes (Miranda et al. 1979; Lourenção et al. 1985; Panizzi et al. 1986). Fernandes et al. (1994) e Pinheiro et al. (2005) também observaram a característica de alta produtividade de grãos da cultivar IAC-100 mesmo sob a infestação de percevejos. O resultado obtido de maior peso de sementes boas (PSB) da cultivar IAC-100 em relação à cultivar CD-215 demonstra que a primeira apresentou uma menor incidência da levedura *Ermothecium coryli*. Essa levedura é transmitida durante o processo de alimentação dos percevejos, pela inserção do aparelho bucal nas sementes (Corrêa-Ferreira and Azevedo 2002). De acordo com Rossetto et al. (1995), uma menor incidência de dano ocasionado por essa levedura pode sugerir resistência da cultivar aos percevejos. Em um estudo com populações provenientes de cruzamentos interespecíficos entre 11 genótipos de soja, também foi identificado que progênies resultantes do cruzamento entre CD-215 e IAC-100 apresentaram a maior percentagem de sementes sem incidência da levedura (87,67%) (Carlini-Garcia et al. 2011).

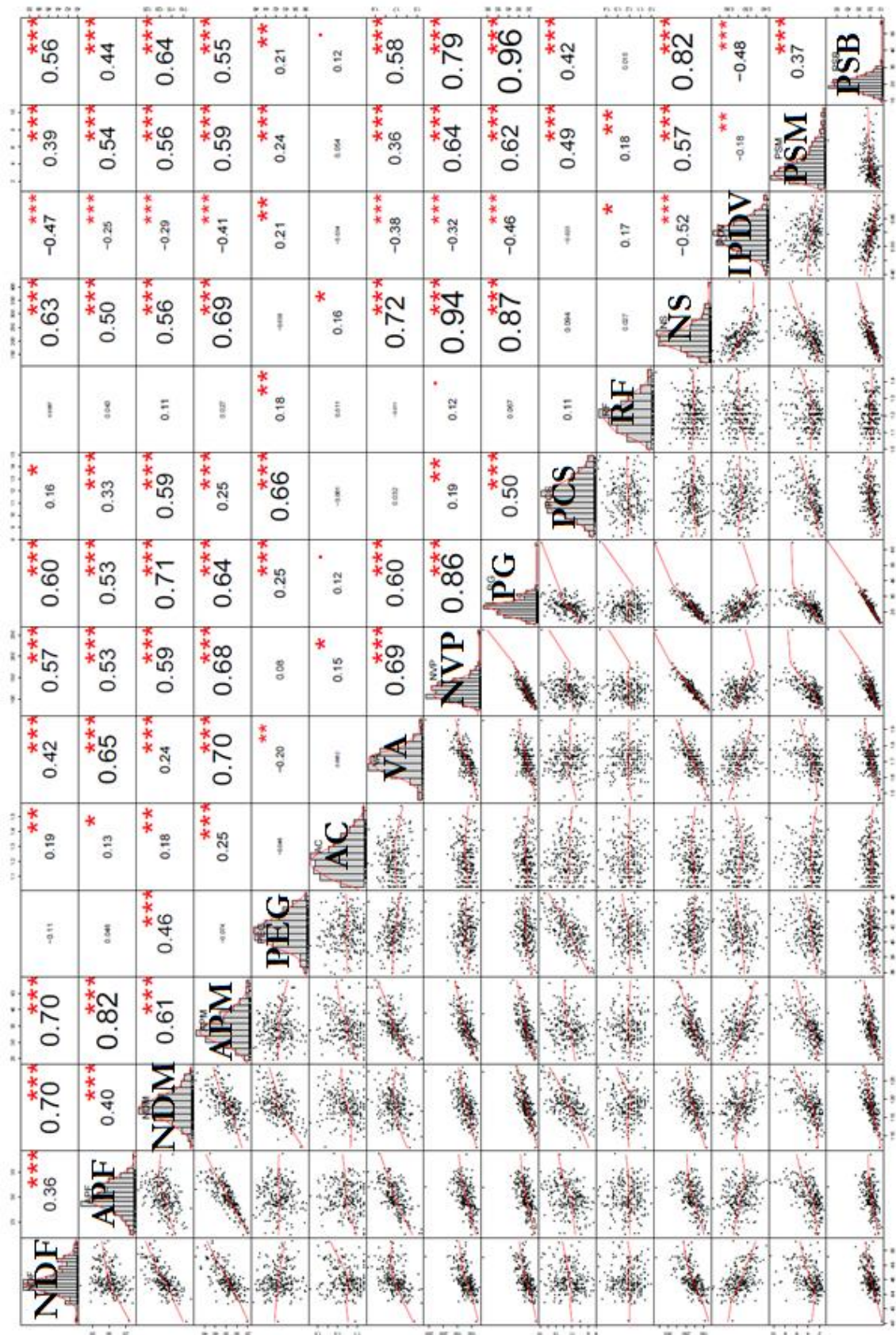


Figura 2. Correlação fenotípica entre as variáveis estudadas: NDF: número de dias para o florescimento; APF: altura da planta no florescimento; NDM: número de dias para a maturidade; APM: altura da planta na maturidade; PEG: período de granação; AC: acamamento; VA: valor agrônômico; NVP: número de vagens por planta; PG: produtividade de grãos; PCS: peso de cem sementes; RF: retenção foliar; NS: número de sementes; IPDV: índice percentual de dano nas vagens; PSM: peso de sementes manchadas; PSB: peso de sementes boas. Significância baseada em probabilidade pelo Teste t (* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$).

A correlação genética estimada entre as variáveis período de granação (PEG) e peso de sem sementes (PCS) foi positiva e moderada ($r_f = 0,66^{***}$) para essa população em estudo, indicando que um menor período de granação resultou em sementes menores. Como já mencionado, o período de granação representa o período de maior suscetibilidade dos grãos ao ataque dos percevejos. Portanto, é desejável que a cultivar apresente um menor PEG e, conseqüentemente, um menor período de exposição aos insetos. Moreira (2015), Oliveira (2015) e Pereira (2015) também relatam correlação positiva das variáveis PEG e PCS. Além disso, pode-se verificar correlação positiva e moderada entre as variáveis peso de cem sementes (PCS) e peso de sementes manchadas (PSM) ($r_f = 0,49^{**}$), sugerindo que, para essa população em estudo, os materiais que possuem sementes menores exibem uma menor quantidade de sementes danificadas pelos percevejos.

A variável produtividade de grãos (PG) também apresentou correlação positiva e forte com número de sementes (NS) ($r_f = 0,87^{***}$), peso de sementes boas (PSB) ($r_f = 0,96^{**}$) e número de vagens por planta (NVP) ($r_f = 0,86^{**}$). Considerando a variável número de vagens por planta (NVP), também foi possível observar correlação positiva e forte com as variáveis número de sementes (NS) ($r_f = 0,94^{**}$) e peso de sementes boas (PSB) ($r_f = 0,79^{**}$). A variável número de sementes (NS) apresentou correlação positiva e forte com peso de sementes boas (PSB) ($r_f = 0,82^{**}$). A correlação positiva e forte entre essas variáveis era esperada, uma vez que um maior número de vagens por plantas resulta em maior número de sementes e, conseqüentemente, maior produtividade de grãos e maior peso de sementes sem danos. Rocha et al. (2014) também observaram altos valores de correlação entre PSB e PG, e propuseram o uso dessa variável na seleção de genótipos resistentes a percevejos e com boa produtividade de grãos. A variável PSB está diretamente relacionada à produtividade de grãos (PG), uma vez que é mensurada após o beneficiamento das sementes, com a remoção dos grãos danificados pelos percevejos.

Na Tabela 1, é possível verificar que a cultivar IAC-100 apresentou menor altura de plantas no florescimento (APF) em relação à cultivar CD-215. No entanto, na maturidade, a média da altura de plantas da cultivar IAC-100 é maior do que a da cultivar CD-215. Ambos os genitores utilizados nos cruzamentos possuem hábito de crescimento determinado, que é caracterizado pelo fim do crescimento vegetativo a partir do início do florescimento (Verneti and Verneti Jr. 2009). Geralmente, no início do florescimento, os materiais de soja com hábito de crescimento determinado já apresentam 85 a 90% da sua altura final. Entretanto, essa diferença da altura de planta na maturidade e no florescimento é característico de cada cultivar. Para a população F₂₃ e para o genitor CD-215, é possível observar que houve pouca diferença na altura das plantas comparando-se o período do florescimento e da maturidade. A cultivar IAC-100 apresentou um maior número de dias para o florescimento (NDF) e também maior número de dias para a maturidade (NDM). O maior número de dias para a maturidade pode estar relacionado com o mecanismo de pseudo-resistência descrito para esta cultivar, denominado diluição de dano. A cultivar pode até sofrer algum dano pelo ataque dos percevejos, mas há tempo para se recuperar e produzir mais vagens, o que contribui para garantir uma boa produtividade final sem dano (Rossetto and Lara 1991).

A variável altura da planta na maturidade (APM) apresentou correlação positiva e forte com a variável altura da planta no florescimento (APF) ($r_f = 0,82^{**}$). Também apresentou correlação positiva e moderada com as variáveis número de dias para o florescimento (NDF) ($r_f = 0,70^{**}$), número de dias para a maturidade (NDM) ($r_f = 0,61^{**}$), número de vagens por planta (NVP) ($r_f = 0,68^{**}$), número de sementes (NS) ($r_f = 0,69^{**}$), produtividade de grãos (PG) ($r_f = 0,64^{**}$) e peso de sementes boas (PSB) ($r_f = 0,55^{***}$). A variável número de dias para o florescimento (NDF) apresentou correlação positiva e moderada com as variáveis número de sementes (NS) ($r_f = 0,63^{**}$), produtividade de grãos (PG) ($r_f = 0,60^{**}$) e peso de sementes boas (PSB) ($r_f = 0,56^{***}$). A variável número de dias para a maturidade (NDM) também apresentou correlação positiva e moderada com as variáveis peso de sementes boas (PSB) ($r_f = 0,64^{***}$)

e produtividade de grãos (PG) ($r = 0,71^{***}$). A correlação positiva entre essas variáveis era esperada, considerando que quanto maior é o período vegetativo e reprodutivo da cultivar, maior é a sua capacidade de crescimento, fatores estes que resultarão em uma maior quantidade de flores que serão posteriormente autofecundadas para a formação das vagens. Portanto, plantas mais altas tendem a apresentar um maior número de vagens e, conseqüentemente, maior o número de sementes que resultará em maior produtividade de grãos e maior peso de sementes boas, o que justifica a correlação obtida entre as variáveis. Sonah et al. (2015) também observaram correlação positiva e forte entre altura de planta e maturidade ($r = 0,77^{**}$). Os autores sugerem que plantas mais tardias apresentam um prolongamento do período vegetativo, resultando em plantas mais altas.

Para a variável valor agrônômico é atribuída uma nota visual para a parcela, na maturidade das plantas. A cultivar IAC-100 apresentou nota de valor agrônômico superior em relação à cultivar CD-215 (Tabela 1). Um dos critérios considerados na avaliação é o aspecto geral da parcela, sendo que o aspecto visual de produtividade de grãos contribui para uma maior nota de valor agrônômico. Pode-se verificar que, para a variável valor agrônômico (VA), foi observada correlação positiva e moderada com a variável produtividade de grãos (PG) ($r = 0,60^{**}$), bem como para as variáveis que influenciam em uma maior produtividade de grãos, como: altura da planta no florescimento (APF) ($r = 0,65^{**}$), altura da planta na maturidade (APM) ($r = 0,70^{**}$), número de vagens por planta (NVP) ($r = 0,69^{**}$) e número de sementes (NS) ($r = 0,72^{**}$).

A cultivar CD-215 apresentou retenção foliar maior que a cultivar IAC-100. Esse fenômeno pode ocorrer por diferentes razões, como deficiência nutricional e estresse hídrico, ou devido ao ataque de percevejos. A forma de alimentação dos percevejos ocasiona a má formação e, até mesmo, a queda das vagens, o que impede a formação de hormônios de senescência das folhas e atrasa a maturidade da planta (Sosa-Gómez and Moscardi 1995). Segundo Rossetto and Lara (1991) e Fernandes et al. (1994), a cultivar IAC-100 apresenta senescência normal, mesmo sob infestação de percevejos. O fenômeno de senescência normal é desejável, pois as folhas e hastes verdes dificultam o processo de colheita mecânica da lavoura (Souza and Toledo 1995). Apesar das estimativas de correlação obtidas para a variável retenção foliar terem apresentado valores baixos, três estimativas foram significativas e positivas, sendo elas: período de granação (PEG) ($r = 0,18^{**}$), índice percentual de dano nas vagens (IPDV) ($r = 0,17^*$) e peso de sementes manchadas (PSM) ($r = 0,18^{**}$). Apesar da baixa correlação, a associação entre essas variáveis é condizente com a literatura, pois materiais com menor período de granação apresentam menor retenção foliar e menor quantidade de grãos danificados, devido ao menor tempo de exposição aos insetos no período mais crítico para a cultura.

Os valores observados na Tabela 1 para a variável índice percentual de dano nas vagens (IPDV) sugere que a cultivar IAC-100 foi significativamente menos atacada pelos percevejos no estágio de formação inicial dos grãos (R5.1 a R5.3). Nesta fase, ao se alimentarem, os percevejos ocasionam o abortamento do grão, que é caracterizado como grão chocho, e as vagens apresentam-se planas. Pode-se verificar que a variável índice percentual de dano nas vagens (IPDV) apresentou correlação negativa e moderada com a variável número de sementes (NS) ($r = -0,52^{**}$). A correlação negativa entre essas variáveis está de acordo com o esperado, pois como a cultivar IAC-100 apresentou uma quantidade de sementes superior à cultivar CD-215, conseqüentemente possui um menor índice percentual de dano nas vagens. A variável também apresentou correlação negativa com a maioria das variáveis analisadas, exceto com as variáveis retenção foliar (RF) ($r = 0,17^*$) e período de granação (PEG) ($r = 0,21^{**}$). Apesar dos valores baixos de correlação obtidos com essas variáveis, ambos foram significativos e as três variáveis (IPDV, RF e PEG) são relevantes em relação à resistência aos percevejos. Como já mencionado, o ataque dos percevejos no estágio inicial de formação dos grãos pode ocasionar a queda das vagens, o que impede a formação de hormônios de senescência das folhas e atrasa a maturidade da planta (Sosa-Gómez and Moscardi 1995). Portanto, a correlação entre essas variáveis é positiva

e quanto maior o período de granação (PEG), maior será o índice percentual de dano nas vagens (IPDV) e, conseqüentemente, maior será o nível de retenção foliar (RF) ocasionado pelo ataque desses insetos.

Na Tabela 1, pode-se verificar que a cultivar IAC-100 obteve maior peso de sementes manchadas (PSM). Tal resultado deve ter ocorrido porque a cultivar IAC-100 apresentou número de vagens por planta (NVP) e número de sementes (NS) relativamente maiores do que a cultivar CD-215. Pode-se verificar que há uma correlação positiva e moderada entre a variável peso de sementes manchadas (PSM) e as variáveis número de vagens por plantas (NVP) ($r = 0,64^{***}$) e número de sementes (NS) ($r = 0,57^{***}$). Essa correlação positiva pode ser explicada pelo fato de que maior número de vagens e maior número de sementes resultam em maior peso de sementes manchadas do que um material com menor quantidade de sementes. Os danos geralmente ocorrem quando o percevejo se alimenta do grão que está no estádio R6, ocasionando manchas ou puncturas características desse ataque. Apesar de maior quantidade de sementes manchadas (PSM), a cultivar IAC-100 apresentou maior produtividade de grãos (PG) em relação à cultivar CD-215. Da mesma forma, apesar das progênies terem apresentado altos valores de sementes manchadas (PSM), os valores de produtividade de grãos (PG) obtidos foram superiores ao genitor resistente e mais produtivo, a cultivar IAC-100. A correlação positiva e moderada entre as variáveis produtividade de grãos (PG) e peso de sementes manchadas (PSM) ($r = 0,62^{***}$) confirma a hipótese de diluição de dano presente na cultivar IAC-100. Apesar desta cultivar apresentar maior quantidade de sementes danificadas em relação à CD-215, o tamanho reduzido da sua semente favorece a produção de uma quantidade superior de sementes por planta, o que garante uma alta produtividade final de grãos sem dano ocasionado pelos percevejos (Rossetto and Lara 1991).

4.2. Análises Genotípicas

Um total de 67 marcadores AFLP e 31 marcadores TRAP foram obtidos por Möller (2010). As combinações de *primers* utilizadas e o tamanho dos alelos amplificados podem ser visualizados no ANEXO A. Os marcadores SSR selecionados para a genotipagem da população F₂ encontram-se no APÊNDICE A, bem como o GL e sua posição, em centiMorgans (cM), descritos no mapa da soja (ConsensusMap 4.0). As seqüências dos *primers* SSR encontram-se disponíveis no Soybase (www.soybase.org) (Grant et al. 2010). A genotipagem utilizando o painel universal de ligação em soja (USLP 1.0 – *Universal Soybean Linkage Panel*) resultou em 442 SNP para a população F₂ utilizada neste estudo. O resultado obtido foi considerado satisfatório. O USLP 1.0 foi desenvolvido a partir da seleção de SNP com base em sua frequência de distribuição em três populações distintas de mapeamento, sendo elas ‘Minsoy’ x ‘Noir’, ‘Minsoy’ x ‘Archer’ (Mansur and Orf 1995; Mansur et al. 1996; Cregan et al. 1999a) e ‘Evans’ x ‘Peking’ (Concibido et al. 1997), além de um grupo de 96 acessos de materiais selvagens oriundos de diferentes países asiáticos (China, Coréia e Japão) e um grupo de 96 cultivares elite de soja dos Estados Unidos, lançadas entre 1990 e 2000 (Hyten et al. 2008). De acordo com Hyten et al. (2010), a previsão é de se obter aproximadamente 458 SNP em populações de mapeamento, utilizando-se ambos os genitores elite, 544 SNP em populações provenientes do cruzamento entre dois materiais selvagens e 590 SNP em populações resultantes do cruzamento entre um genitor elite e um material selvagem. Portanto, o número de marcadores SNP obtidos para essa população de mapeamento em estudo, obtida a partir do cruzamento de duas cultivares elite, encontra-se dentro do esperado. Jun et al. (2014) também utilizaram o USLP 1.0 em uma população de mapeamento de 93 RILs (*Recombinant Inbred Lines*), resultante do cruzamento entre um genitor elite e um material selvagem, e identificaram 516 SNP que foram utilizados para o mapeamento genético.

4.3. Mapa genético

O mapa genético final foi constituído por 20 GLs que correspondem aos 20 cromossomos da soja (Figura 3). O mapa obtido foi comparado com o mapa da soja, disponível no Soybase. Observou-se alta colinearidade na posição dos marcadores. Na construção do mapa, alguns GLs foram subdivididos em subgrupos, sendo sua numeração sequencial do subgrupo identificada após a denominação do GL. Os GLs da soja subdivididos em dois subgrupos foram C1, C2, D1b, E, F, H e I. Por sua vez, os GLs B1, D1a e G foram subdivididos em três subgrupos e o GL B2 em quatro subgrupos. Os demais GLs não foram subdivididos (A1, A2, D2, J, K, L, M, N e O). Essa divisão foi realizada quando foram obtidas distâncias, maiores do que 30 cM, entre marcadores adjacentes. No entanto, algumas subdivisões não foram realizadas mesmo na presença desses *gaps*, quando se verificou que a distância obtida entre os marcadores se aproximava da distância apresentada entre os mesmos no mapa da soja (ConsensusMap 4.0). Portanto, a informação do genoma da soja foi utilizada como um critério biológico no caso dos *gaps* mantidos em alguns GLs. Jun et al. (2014) também obtiveram um mapa genético em uma população de RILs, utilizando esse painel USLP 1.0, com a distância entre SNP variando entre 5,0 a 30,8 cM. A presença de *gaps* na construção do mapa já era esperada, pois o chip de 1536 SNP não foi desenvolvido especificamente para a população em estudo, como já mencionado anteriormente. Além disso, 57% do genoma da soja é composto por regiões heterocromáticas repetitivas, localizadas próximas aos centrômeros dos cromossomos, as quais apresentam baixa taxa de recombinação (Schmutz et al. 2010).

O mapa genético final apresentou 516 marcadores, sendo que quase 85% dos marcadores utilizados para a genotipagem da população foram ancorados no mapa (Tabela 2). Apesar dos marcadores SSR e SNP utilizados possuírem sua posição conhecida nos GLs da soja, alguns tiveram que ser excluídos do mapa por estarem muito distantes dos outros marcadores. Como já mencionado, no caso de *gaps* maiores do que 30 cM, os GLs foram subdivididos. Os marcadores de cada subgrupo foram novamente ordenados utilizando-se a função *order.seq* e os *heatmaps* gerados pelo *OneMap* foram analisados. Muitas vezes, a análise do *heatmap* indicou que algum marcador posicionado na extremidade apresentava fração de recombinação alta e LOD score baixo. Dessa forma, esse marcador foi excluído do GL utilizando-se a função *drop.marker*. Os marcadores AFLP e TRAP são marcadores dominantes, sendo menos informativos do que marcadores co-dominantes, o que contribuiu para uma maior percentagem de marcadores não ancorados no mapa. A definição da posição e ordem desses marcadores dependem da quantidade de marcadores co-dominantes, presentes em cada GL, além de serem também influenciadas pela quantidade de dados perdidos para cada marcador. No entanto, a percentagem final de marcadores ancorados no mapa foi satisfatória, sendo que os mesmos apresentaram uma boa distribuição pelos 20 GLs da soja. Essa alta percentagem se deve ao elevado número de marcadores SSR ligados ao mapa e, principalmente aos marcadores SNP utilizados para a construção do mapa genético, abrangendo todos os 20 GLs da soja. Se considerarmos somente os marcadores SSR e SNP neste estudo, a percentagem de marcadores ancoradas no mapa sobe para quase 94%, o que é bastante elevada.

A cobertura do genoma da soja foi de 2.814,82 cM, com um intervalo médio de 5,46 cM entre os marcadores adjacentes (Figura 3). O mapa obtido foi considerado satisfatório por contemplar todos os GLs da soja. Além disso, a distribuição uniforme dos marcadores e a distância média obtida entre os marcadores foram importantes para o mapeamento de QTLs realizado neste trabalho. O mapa obtido no presente trabalho foi construído a partir da mesma população de mapeamento utilizada por Santos (2012), que adotou a estratégia de genotipagem seletiva, a partir da seleção de 84 indivíduos da população F₂, como já descrito anteriormente. O aumento da população de mapeamento para 228 indivíduos F₂ contribuiu para um maior número de marcadores SNP ancorados ao mapa genético (Tabela 2), em relação ao mapa genético construído por Santos (2012), que obteve o total de 337 marcadores SNP ancorados. Além disso, o mapa obtido por Santos (2012) auxiliou na identificação de regiões onde haviam *gaps* para a utilização de marcadores SSR para saturar esses GLs. Dessa forma, foi possível verificar que tanto o aumento da população de

mapeamento quanto a seleção de marcadores SSR em regiões específicas dos GLs contribuíram para a obtenção de um mapa mais saturado no presente estudo, além de permitirem uma maior cobertura do mapa genético em relação ao mapa inicial, que havia sido de 1.475 cM. Além disso, a inclusão de marcadores SNP e SSR em alguns GLs específicos foram relevantes no mapeamento de QTLs contribuindo para a identificação de um maior número de QTLs para as características avaliadas, como é o caso dos GLs C1-1, M, K e B1.

O mapa obtido neste trabalho é comparável a outros mapas publicados em trabalhos de mapeamento de QTLs em soja. Jun et al. (2014) utilizaram o mesmo painel de 1536 SNP descrito neste estudo e, a partir de 516 SNP genotipados em uma população de 93 RILs (*Recombinant Inbred Lines*), obtiveram uma cobertura de 2.005,3 cM nos 20 GLs da soja, com uma distância média de 5,0 cM entre marcadores. Kato et al. (2014) utilizaram duas populações de RILs, para a genotipagem com 304 e 170 marcadores SSR. O tamanho do mapa genético obtido para ambas foi similar ao obtido neste trabalho, sendo 2.609,1 cM e 2.671,1 cM, porém a distância média entre marcadores foi superior, sendo 13,8 cM e 15,7 cM, respectivamente. Por outro lado, Han et al. (2012) utilizaram 112 marcadores SSR que foram mapeados em 18 GLs da soja e também obtiveram um mapa com tamanho similar, com cobertura de 2.718 cM e distância média entre marcadores de 24,26 cM.

Tabela 2. Número de marcadores utilizados para a obtenção dos dados genotípicos da população F₂, total (e porcentagem) de marcadores ancorados e não ancorados no mapa genético.

Tipo de Marcador	Genotipagem	Ancorados (%)	Não ancorados (%)
SNP	442	417 (94,34)	25 (5,66)
SSR	68	61 (89,71)	7 (10,29)
AFLP	67	30 (44,78)	37 (55,22)
TRAP	31	8 (25,81)	23 (74,19)
Total	608	516(84,87)	92 (15,13)

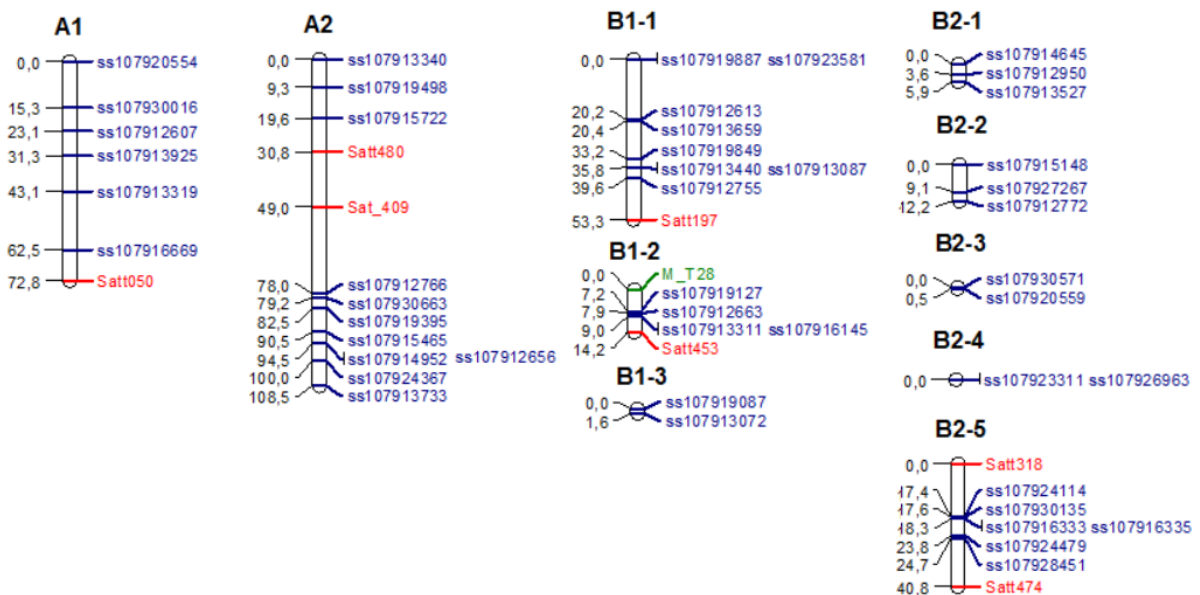


Figura 3. Mapa genético da população F₂, obtida a partir do cruzamento entre os genitores IAC-100 e CD-215, contendo os 20 GLs da soja. À direita, os nomes dos marcadores ancorados no mapa, sendo a cor azul correspondente aos marcadores SNP, a cor vermelha aos marcadores SSR, a cor rosa aos marcadores AFLP e a cor verde aos marcadores TRAP. À esquerda, as posições dos marcadores em cada grupo de ligação da soja, em centimorgans, utilizando-se a função de mapeamento de Kosambi.

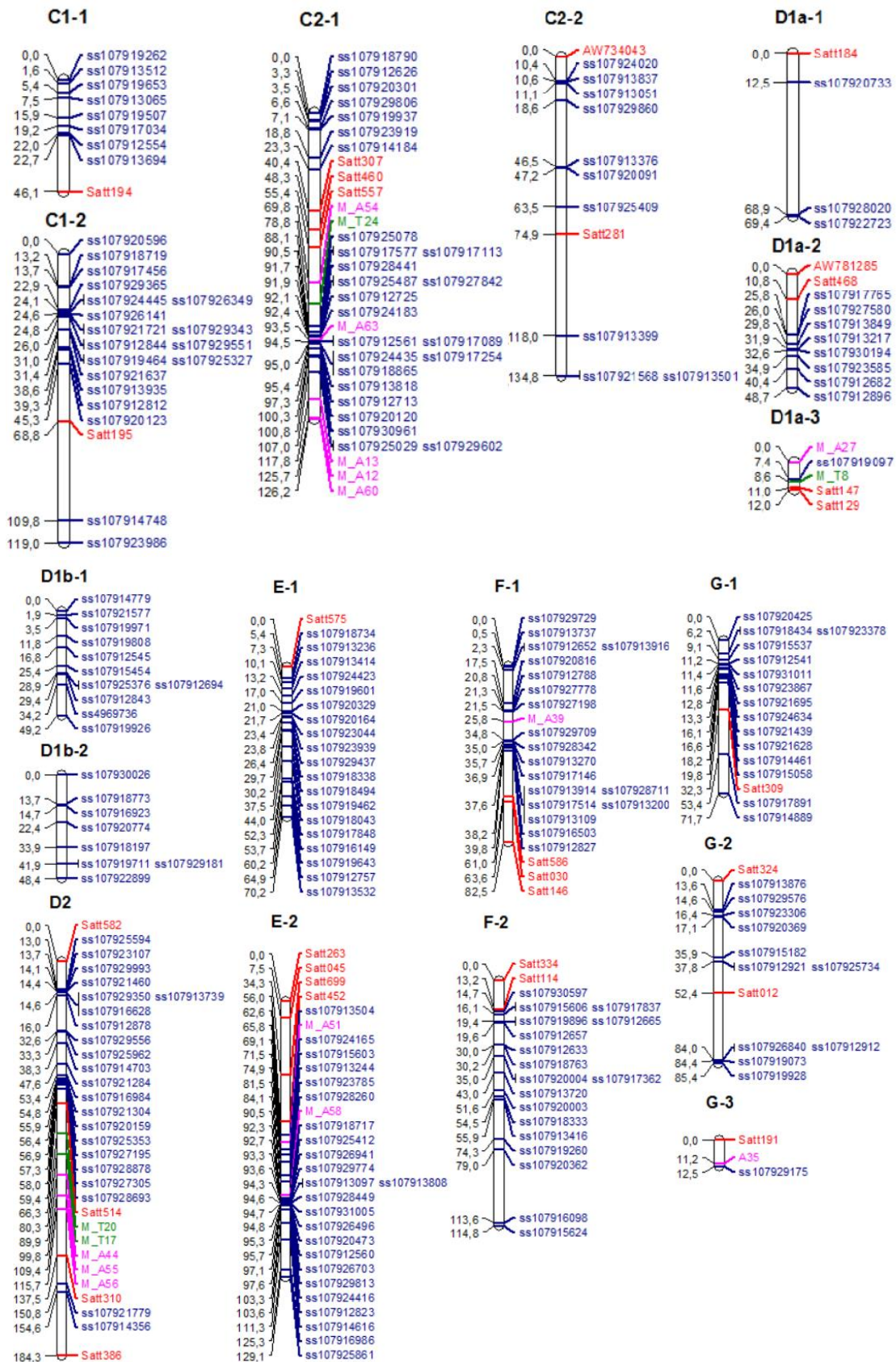


Figura 3 (continuação) Mapa genético da população F_2 , obtida a partir do cruzamento entre os genitores IAC-100 e CD-215, contendo os 20 GLs da soja. À direita, os nomes dos marcadores ancorados no mapa, sendo a cor azul correspondente aos marcadores SNP, a cor vermelha aos marcadores SSR, a cor rosa aos marcadores AFLP e a cor verde aos marcadores TRAP. À esquerda, as posições dos marcadores em cada grupo de ligação da soja, em centimorgans, utilizando-se a função de mapeamento de Kosambi.

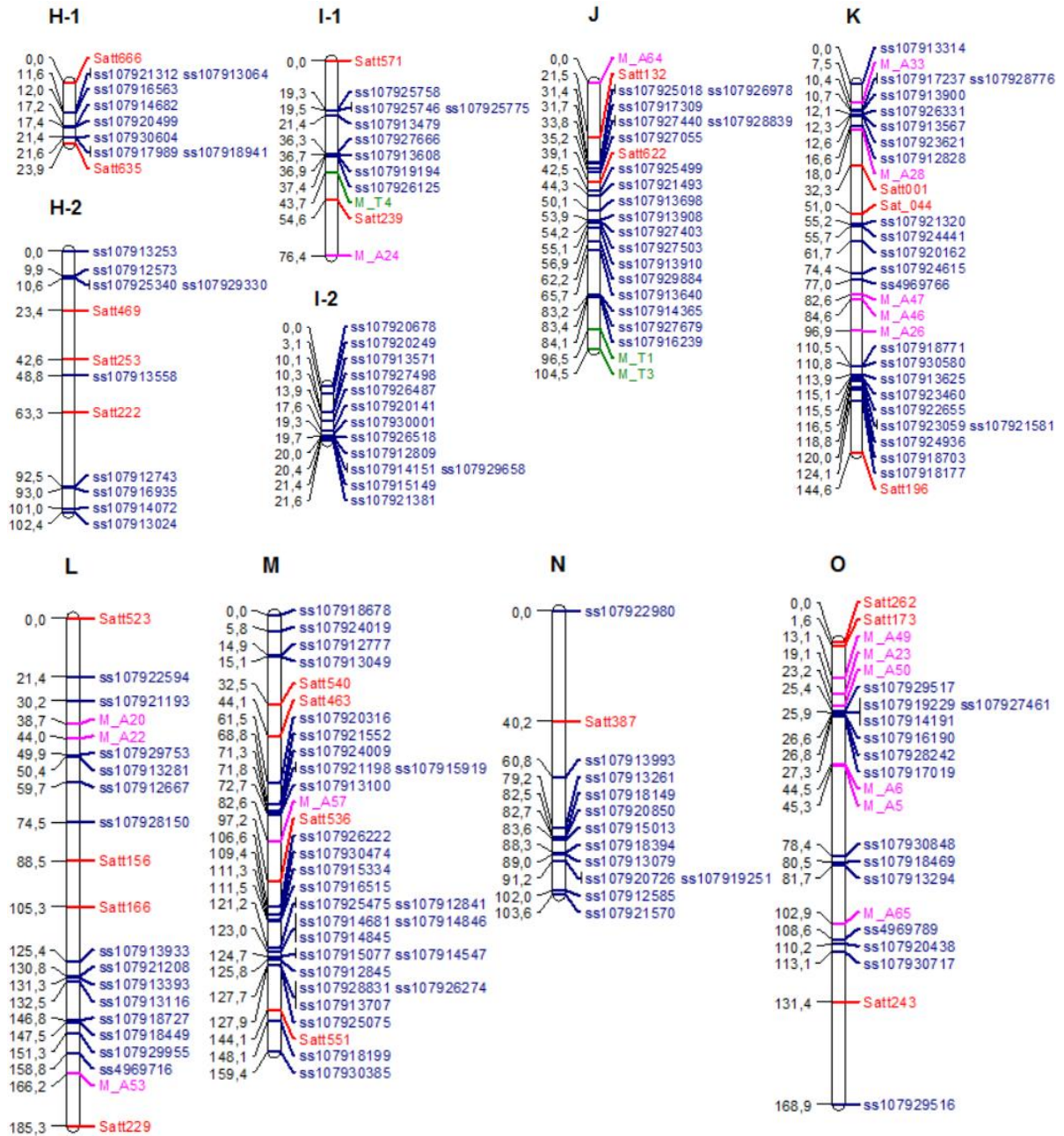


Figura 3 (continuação) Mapa genético da população F₂, obtida a partir do cruzamento entre os genitores IAC-100 e CD-215, contendo os 20 GLs da soja. À direita, os nomes dos marcadores ancorados no mapa, sendo a cor azul correspondente aos marcadores SNP, a cor vermelha aos marcadores SSR, a cor rosa aos marcadores AFLP e a cor verde aos marcadores TRAP. À esquerda, as posições dos marcadores em cada grupo de ligação da soja, em centimorgans, utilizando-se a função de mapeamento de Kosambi.

4.4. Mapeamento de QTLs

Os QTLs identificados neste trabalho receberam a seguinte notação: $F@GL_P$, em que 'F' representa o fenótipo no modelo de mapeamento univariado, sendo posteriormente substituído por "Multi" no mapeamento multivariado; 'GL' representa o grupo de ligação da soja onde o QTL se localiza; e 'P' representa a posição (em cM) do QTL, considerando o mapa de ligação construído para a população F_{2.3} em estudo. Tal atribuição foi anteriormente proposta por (Gianotto 2015).

4.4.1. Mapeamento MIM

O mapeamento de intervalos múltiplos (MIM) para os fenótipos individuais foi capaz de detectar um total de 60 QTLs, sendo 29 QTLs associados às características de resistência aos percevejos (Tabela 3) e 31 QTLs relacionados às características agronômicas (Tabela 4) avaliadas neste estudo (Figuras 6 e 7). Os QTLs identificados foram: dois QTLs afetando a variável acamamento (AC); três QTLs afetando cada uma das variáveis número de vagens por planta (NVP), peso de sementes boas (PSB) e valor agrônômico (VA); quatro QTLs afetando cada uma das variáveis altura da planta na maturidade (APM), índice percentual de dano nas vagens (IPDV) e produtividade de grãos (PG); cinco QTLs afetando as variáveis número de dias para o florescimento (NDF) e altura da planta no florescimento (APF); seis QTLs afetando as variáveis período de granação (PEG) e número de sementes (NS); sete QTLs afetando a variável peso de cem sementes (PCS); e oito QTLs afetando a variável número de dias para a maturidade (NDM). Nenhum QTL foi identificado para retenção foliar (RF) e peso de sementes manchadas (PSM). Para essas duas características, provavelmente a variação genética obtida para a população em estudo não foi suficiente para a identificação de QTLs. Além disso, as estimativas de herdabilidade genética obtidas para essas variáveis correspondem aos menores valores obtidos para a população em estudo (Tabela 1).

As Figuras 6 e 7 exibem os perfis de LOD *score* para cada variável analisada individualmente. Apenas para facilitar a visualização, considerando o grande número de características avaliadas, as mesmas foram separadas em características associadas à resistência aos percevejos (parte superior das Figuras 6 e 7) e características agronômicas (parte inferior das Figuras 6 e 7). Os GLs foram alinhados, considerando-se os QTLs identificados para as características de resistência e agronômicas, para facilitar a visualização e comparação dos QTLs identificados em um mesmo GL. Os 60 QTLs identificados encontram-se localizados em 13 diferentes GLs da soja, sendo eles: A1, B1, C1, C2, D2, E, F, G, I, J, L, M e N. Nas Figuras 6 e 7, os QTLs estão representados graficamente em 17 subgrupos, de acordo com as divisões dos GLs do mapa obtido para a população em estudo. Uma análise visual desses perfis pode fornecer evidências a favor da hipótese de pleiotropia, quando os pontos de máximo para as diferentes variáveis ocorrem na mesma posição, ou a favor da hipótese de QTLs ligados, quando os picos estão distantes entre si. Para os efeitos detectados nesses gráficos, há evidências de que os dois tipos de interações ocorrem nos subgrupos denominados A1, C1-1, C2-1, D2-2, E-1, I-1, J e M. Nesses subgrupos é possível visualizar dois ou mais pontos máximos em uma mesma posição (evidência de pleiotropia) e um outro pico em uma posição próxima (evidência de ligação). Já para o subgrupo L, a evidência é em favor somente da hipótese de ligação, pois os dois picos identificados para as duas características ocorreram em posições distintas. Para os demais subgrupos de ligação, B1-2, B1-3, C1-2, E1-2, F-1, G-2, I-2 e N, o pico obtido foi significativo apenas para uma variável. Estudos de mapeamento de QTLs, em soja, têm reportado uma alta frequência de QTLs identificados em posições similares, sugerindo a presença de pleiotropia. Zhang et al. (2015) realizaram mapeamento associativo em 309 acessos de germoplasma de soja e identificaram QTLs associados às características número de dias para o florescimento, número de dias para a maturidade, período entre o florescimento e a maturidade e altura da planta. Os autores também observaram, de forma semelhante ao resultado obtido neste trabalho, que a maioria dos QTLs identificados apresentou posições genômicas coincidentes, quando identificados no mesmo GL, sugerindo a evidência de pleiotropia. Hu et al. (2014) utilizaram 113 acessos de soja selvagem (*Glycine soja* Sieb. and Zucc), em um estudo de mapeamento associativo, e identificaram QTLs associados ao número de dias para o florescimento, período do florescimento à maturidade, peso de cem sementes, número de vagens por planta e produtividade de grãos. Da mesma forma, os autores observaram que alguns QTLs controlando diferentes características encontram-se em posições semelhantes ou muito próximas. Considerando os 18

SSR associados aos QTLs identificados, cinco encontram-se associados a duas ou mais características. Palomeque et al. (2009a) utilizaram uma população de RILs, avaliada em ambientes distintos, e identificaram uma mesma região contendo QTLs associados a diferentes características agrônômicas, como peso de cem sementes, número de vagens por planta, número de vagens por entrenó, altura da planta, número de dias para o florescimento, número de dias para a maturidade, e conteúdo de óleo e proteína. Wang et al. (2004) utilizaram uma população de retrocruzamento (BC₂F₄) para mapear QTLs, em soja, associados à produtividade de grãos. Os autores também observaram que a maioria das regiões que apresentaram QTL para produtividade de grãos também foi significativa para as variáveis altura de planta e acamamento. Bernard (1972) identificou o gene *Dt1* controlando a terminação da haste em soja. Considerando-se que plantas com hábito de crescimento determinado cessam o crescimento apical assim que a planta inicia o seu florescimento, Lee et al. (1996) e Zhang et al. (2004) investigaram QTLs associados às características altura da planta e maturidade, sugerindo que essas características podem apresentar uma base genética similar para alguns genes. Niu et al. (2013) utilizaram a abordagem do mapeamento associativo para identificar QTLs associados à morfologia da semente de soja. Os autores observaram uma alta frequência do mesmo marcador associado a QTLs que interferem no controle genético das diferentes características, sendo 25 marcadores associados a seis características diferentes e dois marcadores associados a quatro características distintas. Além disso, o trabalho também ressaltou que, apesar de terem sido identificados QTLs em todos os 20 GLs da soja, houve uma concentração de QTLs em 12 GLs. Portanto, os resultados obtidos no presente trabalho estão de acordo com os resultados descritos na literatura.

Todos os QTLs identificados pelo método mapeamento de intervalos múltiplos (MIM) para fenótipos individuais foram significativos a 10% de probabilidade, a partir da estatística *score*, e a maioria apresentou efeito aditivo sobre os fenótipos estudados. O QTL que explica a maior proporção da variação fenotípica foi o *PEG@I-1_39,00*. Para este QTL, o alelo da cultivar CD-215 contribuiu com o aumento de 1,364 desvios padrões da média da variável período de granação, o que correspondeu a 18,65% da variância deste fenótipo (Tabela 3). Considerando a resistência aos percevejos, como esse período é o mais crítico em relação ao dano ocasionado pelos percevejos, um material com menor PEG estaria menos exposto ao ataque desses insetos, resultando em maior produtividade de grãos. Portanto, o alelo proveniente da cultivar IAC-100, para esse QTL identificado, resulta em um fenótipo desejável. O banco de dados de QTLs do Soybase possui somente cinco QTLs descritos para a variável período de granação, sendo a maioria responsável por uma pequena percentagem da variação deste fenótipo. O QTL descrito no Soybase de maior magnitude está localizado no GL F e apresenta efeito aditivo de 1,1 desvios padrões da média da variável, explicando 12% da variância fenotípica de uma população de 295 indivíduos F₄ (Li et al. 2008).

Tabela 3. QTLs identificados usando a análise MIM para as características associadas à resistência da soja aos percevejos: **PEG**: período de granação; **NVP**: número de vagens por planta; **IPDV**: índice percentual de dano nas vagens; **NS**: número de sementes; **PSB**: peso de sementes boas e **PCS**: peso de cem sementes. QTLs destacados em negrito foram identificados somente na abordagem MIM. QTLs destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

QTL ¹	Marcadores Adjacentes ²		LOD ³	Efeito Aditivo ⁴ Dominância ⁵	% Var ⁶	R ² ⁷
PEG@A1_41,00	ss107913925	ss107913319	3,32	-0,644***	3,92	
			1,34	0,589*	1,51	
PEG@C1-1_22,68	ss107913694	ss107913694	6,30	0,902***	7,83	
				0,425		
PEG@D2-2_41,00	ss107914703	ss107921284	5,82	-0,851***	6,88	51,35
				-0,330		
PEG@E-1_16,00	ss107924423	ss107919601	3,59	-0,690***	4,20	
				-0,537		
PEG@I-1_40,00	ss107926125	T4	14,29	-1,364***	18,65	
				-0,158		
PEG@J_56,92	ss107913910	ss107913910	6,73	-0,894***	8,36	
				0,049		
NVP@C2-1_100,33	ss107912713	ss107920120	7,36	11,475***	11,40	
				0,353		
NVP@I-1_42,00	ss107926125	T4	7,32	11,718***	11,35	29,72
				3,502		
NVP@J_29,00	Satt132	ss107925018	4,63	9,605***	6,97	
				2,999		
IPDV@B1-2_28,00	ss107913659	ss107919849	3,54	0,006	3,54	
				-0,030***		
IPDV@I-1_43,67	ss107926125	T4	7,38	-0,027***	7,38	21,91
				0,011		
IPDV@J_54,18	ss107927403	ss107927403	3,80	-0,018***	3,80	
				0,006		
IPDV@M_72,73	ss107913100	ss107913100	4,66	0,021***	7,19	
				-0,004		
NS@C1-1_31,00	ss107913694	Satt194	3,16	15,074***	3,70	
				8,719		
NS@C2-1_100,00	ss107912713	ss107920120	10,44	23,024***	12,48	
				3,875		
NS@I-1_40,00	ss107926125	T4	10,73	23,602***	13,23	43,97
				-0,281		
NS@J_54,18	ss107927403	ss107927403	3,19	12,164***	3,71	
				3,192		
NS@L_132,00	ss107913393	ss107913116	4,72	14,904***	5,50	
				-3,149		
NS@M_108,00	ss107930474	ss107930474	4,69	-15,108***	5,35	
				6,233		
PSB@C1-1_28,00	ss107913694	Satt194	4,05	2,137**	6,82	
				0,387		
PSB@C2-1_91,00	ss107917113	ss107928441	3,66	1,780**	6,16	19,74
				-0,896		
PSB@M_103,00	Satt536	ss107926222	4,00	-1,934***	6,76	
				0,269		
PCS@A1_24,00	ss107912607	ss107913925	4,48	-0,457***	4,98	
				0,089		
PCS@B1-3_0,00	ss107919087	ss107919087	2,40	-0,154	3,20	
				0,563**		
PCS@D2-2_21,00	ss107912878	ss107929556	9,74	-0,732***	11,00	
				-0,001		
PCS@E-1_64,00	ss107919643	ss107912757	3,02	0,531***	6,51	42,00
				-0,078		
PCS@E-2_65,82	A51	A51	4,91	0,517***	6,47	
				0,304		
PCS@I-2_14,00	ss107926487	ss107920141	3,97	-0,426***	4,27	
				0,189		
PCS@M_107,00	ss107926222	ss107930474	2,94	-0,361***	3,05	
				0,466***	2,52	

¹Fenótipo@GL_posição(M); ²Marcadores adjacentes ao QTL identificado; ³LOD marginal com base na distribuição de $\chi^2=0,835$ (5%) ⁴Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁵Efeito de dominância; Significância dos efeitos baseada em p -valor estimado por reamostragem baseada na estatística *score* ($p < 0,05$; $**p < 0,01$; $***p < 0,001$); ⁶Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁷R² calculado com o modelo contendo apenas os efeitos marginais significativos.

Tabela 4. QTLs identificados usando a análise MIM para as características agrônômicas: **NDF**: número de dias para o florescimento; **APF**: altura da planta no florescimento; **NDM**: número de dias para a maturidade; **APM**: altura da planta na maturidade; **AC**: acamamento; **VA**: valor agrônômico; **PG**: produtividade de grãos. QTLs destacados em negrito foram identificados somente na abordagem MIM. QTLs destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

QTL ¹	Marcadores Adjacentes ²		LOD	Efeito Aditivo ³ Dominância ⁴	% Var ⁵	R ² ⁶
<i>NDF@C1-1_24,00</i>	ss107913694	Satt194	6,44	0,730*** -0,275	7,10	
<i>NDF@C2-1_100,00</i>	ss107912713	ss107920120	12,60	0,985*** 0,094	14,29	
<i>NDF@I-1_41,00</i>	ss107926125	T4	13,73	1,046*** -0,280	15,78	49,64
<i>NDF@J_52,00</i>	ss107913698	ss107913908	7,01	0,724*** -0,071	7,75	
<i>NDF@M_21,00</i>	ss107913049	ss107913049	4,52	-0,641*** -0,004	4,72	
<i>APF@D2-2_44,00</i>	ss107914703	ss107921284	5,15	-1,191*** 0,037	7,28	
<i>APF@E-1_70,21</i>	ss107913532	ss107913532	3,48	0,923*** -0,250	4,74	
<i>APF@I-1_37,40</i>	ss107926125	ss107926125	5,52 1,25	1,122*** 0,768*	7,74 1,66	37,24
<i>APF@J_54,18</i>	ss107927403	ss107927403	7,37	1,346*** 0,133	10,53	
<i>APF@N_5,00</i>	ss107922980	Satt387	3,81	-1,037*** -0,412	5,29	
<i>NDM@A1_39,00</i>	ss107913925	ss107913319	9,44	-1,681*** -0,169	7,87	
<i>NDM@C1-1_23,00</i>	ss107913694	Satt194	17,47	2,410*** 0,538	16,78	
<i>NDM@C1-2_21,00</i>	ss107917456	ss107929365	5,77	1,242*** -0,710	4,97	
<i>NDM@C2-1_90,48</i>	ss107917577	ss107917577	5,97	1,267*** -0,205	4,90	
<i>NDM@D2-2_39,00</i>	ss107914703	ss107921284	5,83	-1,246*** -0,508	4,76	65,45
<i>NDM@F-1_56,00</i>	ss107912827	Satt586	7,96	0,188 2,472***	6,60	
<i>NDM@I-1_67,00</i>	Satt239	A24	4,17	1,224*** -0,089	3,43	
<i>NDM@M_14,88</i>	ss107912777	ss107912777	17,47	-2,330*** 0,380	16,14	
<i>APM@C1-1_22,68</i>	ss107913694	ss107913694	2,80	1,286*** 0,837	3,41	
<i>APM@C2-1_100,79</i>	ss107930961	ss107930961	7,13	2,010*** 0,244	8,66	
<i>APM@I-1_39,00</i>	ss107926125	T4	13,65	2,835*** -0,405	17,91	42,51
<i>APM@J_55,00</i>	ss107927403	ss107927503	9,68	2,352*** -0,448	12,53	
<i>AC@G-2_68,00</i>	Satt012	ss107926840	3,28	0,020 -0,087***	5,94	
<i>AC@I-1_36,94</i>	ss107913479	ss107927666	3,02 1,79	0,038*** 0,043**	5,40 2,93	14,27
<i>VA@C2-1_98,00</i>	ss107912713	ss107920120	9,02	0,046*** 0,017	13,41	
<i>VA@I-1_40,00</i>	ss107926125	T4	9,51	0,047*** 0,009	14,20	34,53
<i>VA@J_54,28</i>	ss107913908	ss107927403	4,86	0,032*** 0,004	6,92	
<i>PG@C1-1_26,00</i>	ss107913694	Satt194	3,94	2,301** 0,486	6,18	
<i>PG@D2-2_23,00</i>	ss107912878	ss107929556	3,96	-2,243*** 1,030	6,18	
<i>PG@L_120,00</i>	Satt166	ss107913933	2,75 1,04	1,842** 1,781*	4,36 1,60	27,82
<i>PG@M_90,00</i>	A57	Satt536	5,91	-2,778*** 0,839	9,50	

¹Fenótipo@GL_posição(cM); ²Marcadores adjacentes ao QTL identificado; ³LOD marginal com base na distribuição de $\chi^2=0,835$ (5%) ⁴Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁵Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* ($p < 0,05$; ** $p < 0,01$; *** $p < 0,001$); ⁶Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁷R² calculado com o modelo contendo apenas os efeitos marginais significativos.

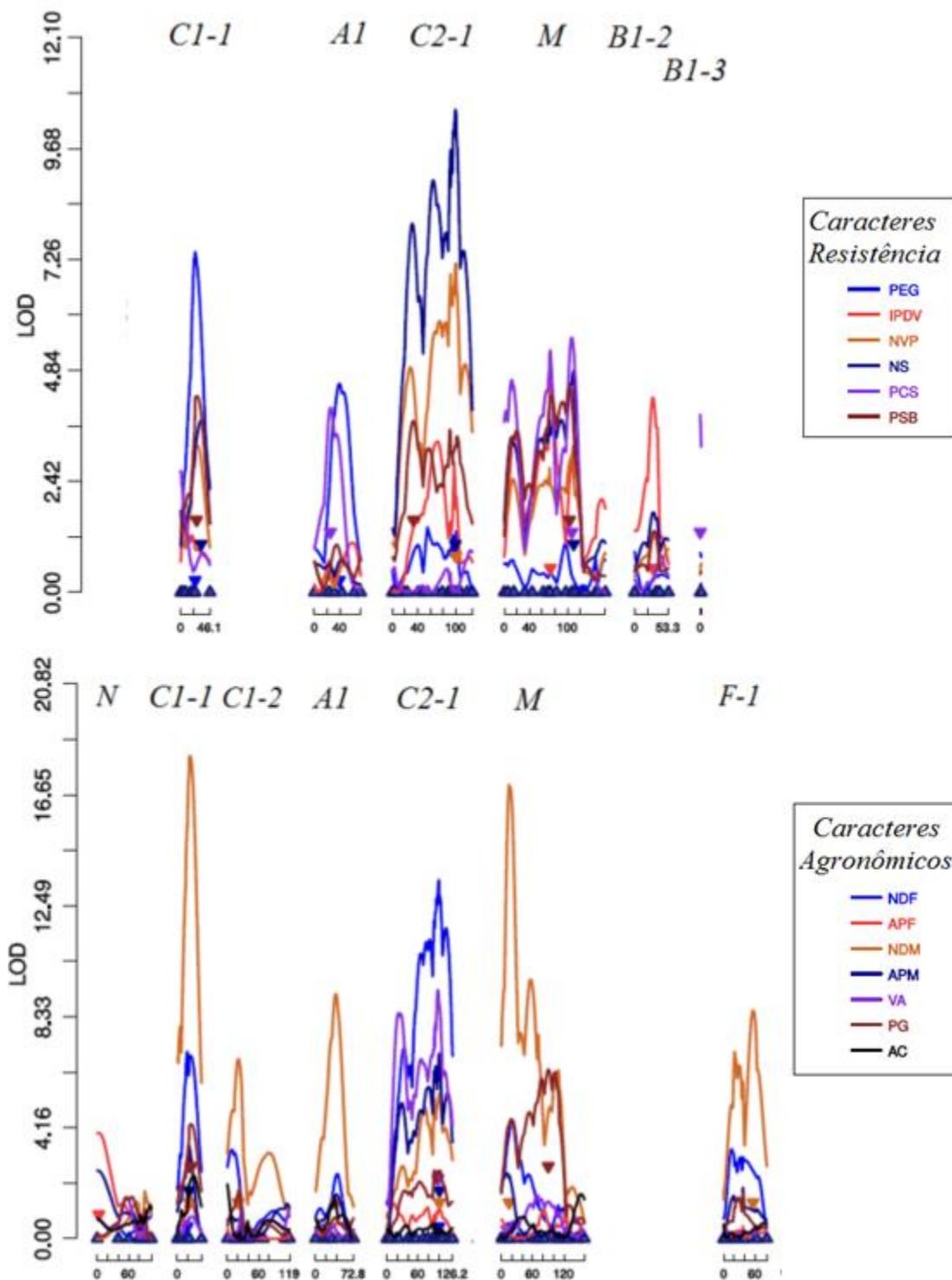


Figura 4. Perfis de LOD ao longo dos GLs N, C1-1, C1-2, A1, C2-1, M, B1-2, B1-3 e F-1, utilizando a abordagem univariada (MIM) para os fenótipos: caracteres de resistência a percevejos (gráfico superior): **PEG**: período de granação, **IPDV**: índice percentual de dano nas vagens; **NVP**: número de vagens por planta; **NS**: número de sementes; **PCS**: peso de cem sementes; **PSB**: peso de sementes boas; caracteres agronômicos (gráfico inferior): **NDF**: número de dias para o florescimento; **APF**: altura da planta no florescimento; **NDM**: número de dias para a maturidade; **APM**: altura da planta na maturidade; **VA**: valor agronômico; **PG**: produtividade de grãos e **AC**: acamamento.

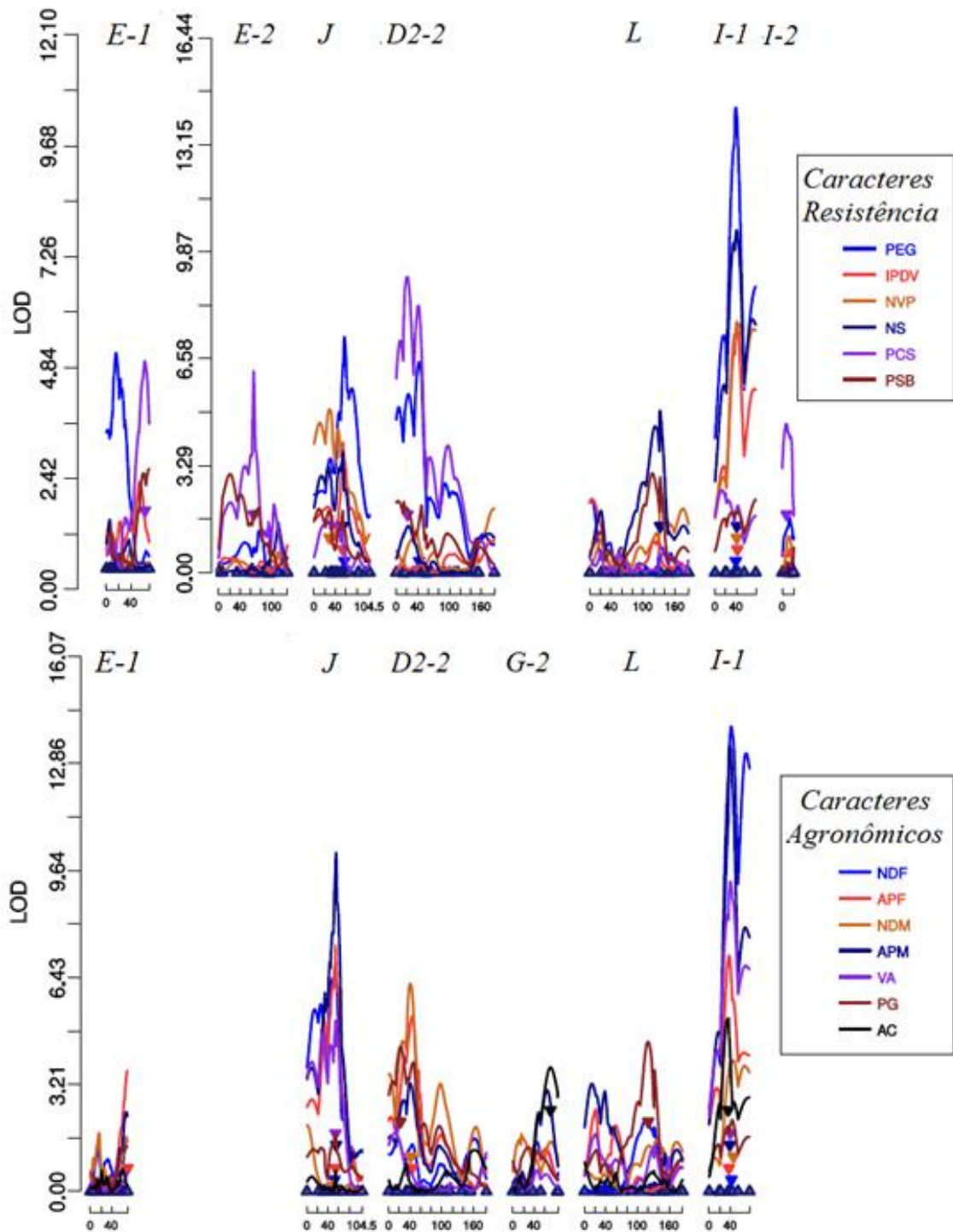


Figura 5. Perfis de LOD ao longo dos GLs E-1, E-2, J, D2-2, G2, L, I-1 e I-2 utilizando a abordagem univariada (MIM) para os fenótipos: caracteres de resistência a percevejos (gráfico superior): *PEG*: período de granação; *IPDV*: índice percentual de dano nas vagens; *NVP*: número de vagens por planta; *NS*: número de sementes; *PCS*: peso de cem sementes; *PSB*: peso de sementes boas; caracteres agronômicos (gráfico inferior): *NDF*: número de dias para o florescimento; *APF*: altura da planta no florescimento; *NDM*: número de dias para a maturidade; *APM*: altura da planta na maturidade; *VA*: valor agronômico; *PG*: produtividade de grãos e *AC*: acamamento.

Do total de 60 QTLs identificados nesta população de estudo, pelo método MIM para fenótipos individuais, cinco apresentaram tanto o efeito aditivo quanto o efeito de dominância significativos. No entanto, para esses QTLs, o efeito aditivo explicou uma maior porcentagem da variação genotípica da variável em estudo. Outros quatro QTLs

apresentaram somente o efeito de dominância significativo sobre os fenótipos estudados, afetando as variáveis número de dias para a maturidade ($NDM@F-1_{56,00}$), índice percentual de dano nas vagens ($IPDV@B1-2_{28,00}$), peso de cem sementes ($PCS@B1-3_{0,00}$) e acamamento ($AC@G-2_{68,00}$). Com exceção da variável AC, para as outras três variáveis foram identificados outros QTLs de maior magnitude apresentando efeito aditivo. Para o melhoramento da soja, espécie autógama, a identificação de QTLs que apresentam efeito aditivo é extremamente relevante, pois representam a substituição alélica no genótipo do QTL. Ou seja, o efeito aditivo representa a porção da variância genética que será transmitida para a próxima geração, sendo diretamente relacionado com o ganho de seleção para a característica de interesse. Em linhagens avançadas, após vários ciclos de autofecundação, a variância aditiva alcança, teoricamente, o valor máximo, com magnitude de duas vezes o valor observado na geração F_2 (Bernardo 2010). Por outro lado, o efeito de dominância representa a porção da variância genética que não é explicada pela variância aditiva, ou seja, corresponde a interação intraloco e entre os alelos, ou mesmo os desvios do modelo genético. O efeito de dominância é reduzido ao longo do avanço das gerações em populações autógamas, sendo desprezível em F_∞ . Dessa forma, a identificação de QTLs apresentando efeitos aditivos de maior magnitude refletem diretamente em uma maior eficiência de seleção para a característica de interesse.

A proporção da variância fenotípica explicada pelo mapeamento MIM para os fenótipos individuais variou de 14,27% para a variável acamamento (AC) a 65,45% para a variável número de dias para a maturidade (NDM) (Tabelas 3 e 4). Os valores obtidos de R^2 , nos modelos MIM, foram inferiores aos valores estimados de herdabilidade (h^2) para os caracteres fenotípicos avaliados (Tabela 1). Entretanto, esses resultados estão em consonância com estudos de mapeamento de QTLs encontrados para as características utilizadas neste trabalho. Por se tratarem de características quantitativas, os resultados obtidos corroboram com a hipótese de um grande número de genes controlando as mesmas, sendo que cada gene contribui para uma pequena porção do fenótipo. Além disso, o modelo considerado no presente estudo não incluiu os efeitos de epistasia entre QTLs, o que pode contribuir para aumentar os valores obtidos de R^2 . Outra questão relevante é o tamanho populacional utilizado, que pode não ter sido suficiente para identificar a variabilidade existente para cada característica analisada.

Os resultados demonstram que, para essa população de mapeamento em estudo, foram identificados tanto QTLs de pequeno efeito, bem como alguns QTLs de maior magnitude. No geral, a proporção da variância fenotípica explicada pelos QTLs foi considerada satisfatória. Dos 60 QTLs identificados, somente 18 QTLs explicam menos do que 5% da variância fenotípica das variáveis em estudo, 27 QTLs correspondem a 5 a 10%, oito QTLs explicam entre 10 a 15% e sete QTLs correspondem a 15 a 18% da variância fenotípica da característica. Considerando que as características avaliadas neste estudo possuem caráter quantitativo, que, normalmente, são controladas por muitos genes e sofrem grande influência ambiental, a identificação de genes de pequeno efeito já era esperada. Os resultados obtidos sugerem que as 13 características, para as quais foram identificados QTLs, são controladas por um conjunto de genes, cada um exercendo um pequeno efeito sobre o fenótipo. No entanto, a identificação de QTLs de pequeno efeito muitas vezes inviabiliza a sua utilização na seleção de genótipos superiores, pois sua detecção nem sempre se mantém em diferentes *backgrounds* genéticos e em diferentes ambientes (Walker et al. 2010). Portanto, a identificação de QTLs de maior magnitude é desejável no melhoramento de plantas, pois viabilizam a utilização dos mesmos para a seleção de genótipos superiores. Além disso, permitem uma maior eficiência na seleção, com ganhos genéticos superiores em programas de melhoramento (Bernardo 2010). Considerando as 13 variáveis utilizadas na abordagem de mapeamento MIM, foi possível identificar QTLs explicando entre 15 a 18% da variância do fenótipo para quatro destas variáveis, sendo dois QTLs dessa magnitude para número de dias para a maturidade (NDM), e um QTL para as variáveis número de dias para o florescimento (NDF), altura da planta na maturidade (APM) e período de granação

(PEG). Outras quatro variáveis apresentaram pelo menos um QTL de maior magnitude, explicando entre 10 a 15% da variância do fenótipo, sendo um QTL para as variáveis peso de cem sementes (PCS) e altura de planta no florescimento (APF), e dois QTLs para as variáveis número de vagens por planta (NVP), número de sementes (NS) e valor agrônômico (VA). Para as outras três variáveis, foram identificados dois QTLs explicando entre 5 a 10% da variação fenotípica das variáveis acamamento (AC), índice percentual de dano nas vagens (IPDV) e peso de sementes boas (PSB), e três QTLs dessa magnitude afetando a variável produtividade de grãos (PG). Todos os demais QTLs identificados para essas 13 variáveis são QTLs que, para essa população em estudo, explicaram menos do que 5% da variação fenotípica. No entanto, não se pode deixar de ressaltar a importância desses QTLs de menor magnitude, para o entendimento da base genética dessas características, pois cada um exerce um pequeno efeito no fenótipo, contribuindo para o valor final de R^2 obtido.

Zhang et al. (2004) realizaram o mapeamento de QTLs para características agrônômicas em uma população de RILs e identificaram quatro QTLs para PCS, nos GLs A2, B1 e D2 (2). Esses QTLs identificados explicam 37,5% da variação fenotípica da população. Portanto, o resultado obtido no presente estudo, para PCS ($R^2= 42,00\%$), encontra-se em consonância com os resultados relatados na literatura. Em um mapeamento associativo, realizado por Zhang et al. (2015), apesar de um grande número de QTLs ter sido identificado afetando cada característica avaliada, a percentagem total da variação fenotípica explicada pelos QTLs também foi menor do que a herdabilidade estimada. Para a variável número de dias para o florescimento, 27 QTLs foram identificados e explicam 77,2% da variação fenotípica ($b^2=95,6$). Para a variável número de dias para a maturidade, seis QTLs revelados explicam 52,5% da variação fenotípica ($b^2=95,6$), e outros 27 QTLs identificados para a variável altura da planta explicam 76,3% da variação fenotípica desta característica ($b^2=82,8$). Han et al. (2012) realizaram o mapeamento de QTLs para a característica peso de cem sementes, utilizando a abordagem CIM, em uma população de 125 RILs. Os autores identificaram a presença de 23 QTLs afetando a característica, e outros 17 QTLs condicionados pelo estágio de desenvolvimento dos grãos. A maioria dos QTLs apresentou R^2 menor do que 10%, o que, segundo os autores, sugere que a característica peso de cem sementes é controlada por muitos QTLs de pequeno efeito, o que dificulta o aumento no ganho de seleção baseado apenas em seleção fenotípica. Além disso, foram identificadas a presença de epistasia entre QTLs (QQ) e entre QTL x ambiente (QE). Segundo os autores, a maioria dos efeitos de interação entre QTLs estimados foi revelada em um curto período do estágio inicial de desenvolvimento dos grãos. Essa constatação da variação dos efeitos epistáticos, conforme o estágio de desenvolvimento dos grãos, limita a sua identificação. Essa pode ser uma das razões para a grande diferença de estimativas de interações epistáticas obtidas em diferentes estudos. Os resultados obtidos também indicaram que somente os QTLs identificados nos dois anos de avaliação apresentaram maiores valores de efeito aditivo e explicaram uma maior percentagem da variação fenotípica da característica avaliada. Estudos demonstram que o controle genético das características florescimento e maturidade, em soja, apresentam uma complexa base genética (Mao et al. 2017). Além disso, o período de florescimento é uma característica que interfere diretamente na produtividade da planta. Onze genes de maior magnitude já foram identificados, sendo denominados *E1*, *E2* (Bernard 1971), *E3* (Buzzell 1971), *E4* (Buzzell and Voldeng 1980), *E5* (McBlain and Bernard 1987), *E6* (Bonato and Vello 1999), *E7* (Cober and Voldeng 2001), *E8* (Cober et al. 2010), *E9* (Kong et al. 2014), *E10* (Samanfar et al. 2017) e *J* (Ray et al. 1995), sendo esse último associado ao maior período juvenil em soja. Os genes *E1*, *E2*, *E3*, *E4*, *E9* e *J* já foram clonados e caracterizados funcionalmente (Watanabe et al. 2009; Kong et al. 2010; Watanabe et al. 2011; Xia et al. 2012; Zhao et al. 2016; Yue et al. 2017). No entanto, além desses genes de maior magnitude, muitos outros genes de menor efeito encontram-se descritos no banco de dados de QTLs do Soybase (<http://www.soybase.org>). Estudos recentes demonstram a presença de genes, em soja, ortólogos a genes associados

ao florescimento em *Arabidopsis thaliana*, e sugerem que ambas as características, florescimento e maturidade, apresentam uma base genética complexa (Jung et al. 2012; Watanabe et al. 2012; Zhang et al. 2015). Considerando os resultados descritos acima, os valores de R^2 , estimados para a população em estudo, encontram-se em consonância com os valores estimados na literatura.

Os resultados apresentados para os fenótipos individuais foram obtidos após a utilização da função *backmain*, com a qual QTLs foram retirados a partir do modelo completo obtido via *forward*. No entanto, comparando-se esses resultados finais com os resultados da análise inicial realizada somente com a função *fdmain*, foi possível observar que a redução nos valores das estimativas dos efeitos dos QTLs e na porcentagem da variância explicada pelos mesmos foi pequena. Considerando-se o somatório da porcentagem da variação fenotípica explicada por todos os QTLs identificados para uma mesma característica, não houve redução maior do que três pontos percentuais após a utilização da função *backmain*, com exceção da variável denominada índice percentual de dano nas vagens (IPDV). Para essa variável, a redução no valor de R^2 foi de 8,64 pontos percentuais, devido à redução na porcentagem da variação fenotípica explicada por três QTLs afetando essa variável. No entanto, para a variável número de dias para a maturidade (NDM) e valor agrônômico (VA), houve um aumento no valor do efeito de alguns QTLs, após a utilização da função *backmain*, resultando em uma maior porcentagem da variação fenotípica explicada pelos QTLs identificados para essas características. A etapa final realizada também permitiu um refinamento da posição dos QTLs, com pequenos ajustes na posição dos QTLs revelados e, conseqüentemente, com modificações em alguns marcadores flanqueando os QTLs em questão.

Analisando-se os resultados obtidos neste trabalho, foi possível verificar que os modelos MIM utilizados permitiram revelar um maior número de QTLs, controlando as 13 variáveis analisadas, em relação ao mapeamento MQM (*Multiple QTL Mapping*) e o procedimento de Kruskal-Wallis (Lehmann 1975) utilizados por Santos (2012), no programa MapQTL® 4.0 (Van Ooijen et al. 2002). A abordagem MIM utilizada neste estudo foi capaz de revelar um total de 28 QTLs, que não haviam sido identificados anteriormente, para as 12 variáveis estudadas. Foram identificados seis QTLs para a variável NDM, em regiões distintas dos QTLs que já haviam sido identificados, quatro QTLs para PCS, três QTLs para NS, dois QTLs para as variáveis PEG, NDF, APF, IPDV, APM e PG, e um QTL para as variáveis NVP, VA e PSB. No caso dos QTLs que já haviam sido identificados por Santos (2012), observou-se neste estudo que a abordagem MIM permitiu uma maior precisão nas estimativas dos efeitos genéticos e na localização dos mesmos. Por fim, os QTLs que não foram revelados com o uso da abordagem MIM, mas que haviam sido previamente identificados por Santos (2012), apresentam um baixo valor de LOD. Como a abordagem MIM apresentou um *score* mais elevado, em relação ao *threshold* estabelecido na análise de mapeamento realizada anteriormente, os QTLs que apresentavam menor LOD não foram identificados no presente estudo. Além disso, deve-se considerar que o aumento da população de mapeamento utilizada neste trabalho, em relação à população utilizada por Santos (2012), que adotou a estratégia de genotipagem seletiva, certamente contribuiu para a identificação de um maior número de QTLs de menor efeito, além de reduzir a possibilidade do erro tipo I, ou seja, a identificação de falsos positivos.

Hwang (2010) identificou QTLs associados ao número de dias para a maturidade, altura da planta, peso de cem sementes e acamamento, comparando as abordagens CIM e MIM. O estudo concluiu que a estratégia de mapeamento MIM revelou um maior número de QTLs, os quais não haviam sido identificados pela estratégia CIM, para as três primeiras características, além de permitir um refinamento das posições dos QTLs identificados na abordagem CIM. O autor também destaca a relevância do uso da abordagem MIM, que permitiu identificar QTLs de pequeno efeito, considerando a importância dos mesmos no estudo de características quantitativas. Como a abordagem MIM se baseia em um modelo com múltiplos QTLs, conforme o número de QTLs incluídos no modelo, há uma

redução da variância residual, o que contribui para o maior poder de detecção de QTLs em relação ao mapeamento CIM, contribuindo também para a detecção de um maior número de QTLs de pequeno efeito. Por fim, o autor utilizou a abordagem MIM para estimar a epistasia presente entre os QTLs identificados para as mesmas características e a interação QTL x E, concluindo que nem todos os QTLs apresentavam o mesmo efeito nos dois diferentes anos de avaliação. Outros estudos recentes foram publicados, em soja, utilizando a abordagem MIM, para estimar o efeito de epistasia entre os QTLs controlando o conteúdo de óleo, bem como a interação desses QTLs com o ambiente (QTL x E) (Korir et al. 2011; Pathan et al. 2013; Qi et al. 2014a).

4.4.2. Mapeamento MT-MIM

Para o mapeamento de intervalos múltiplos multivariado (MT-MIM), foram selecionados nove conjuntos de variáveis. Como a complexidade das análises do modelo MT-MIM aumenta conforme o número de variáveis incluídas no modelo, as análises foram realizadas em grupos de, no máximo, cinco variáveis, sem incluir os efeitos de epistasia entre os QTLs. A ortogonalidade do modelo de Cockerham utilizado permite que os QTLs sejam mapeados mesmo sem incluir a epistasia no modelo, pois a estimação de um efeito (marginal ou epistático) não é afetada pela presença ou ausência de outros efeitos no modelo. Dessa forma, os efeitos estimados neste estudo correspondem aos efeitos principais dos QTLs. Do total de 15 variáveis utilizadas para a obtenção dos dados fenotípicos, apenas as variáveis acamamento (AC) e retenção foliar (RF) não foram utilizadas na abordagem multivariada. As duas variáveis não atenderam a um dos critérios utilizados para a definição dos conjuntos de características para a análise MT-MIM, que consistiu em obter valores de correlação superiores 0,6. O maior valor de r apresentado para a variável RF foi com as variáveis PSM e PEG ($r=0,18^{**}$), e a variável AC apresentou maior valor de r com a variável APM ($r=0,18^{**}$) (Figura 2). Considerando-se que a abordagem MT-MIM garante a obtenção de estimativas não viesadas pelo fato de utilizar a estrutura de correlação genética entre as variáveis para estimar os efeitos dos QTLs, recomenda-se não incluir características com baixa correlação no modelo. Além disso, a variável AC não apresentou valores significativos na análise de variância realizada por Santos (2012). As duas características foram avaliadas por notas, de 1 a 5. Pode-se verificar na Tabela 1, que a nota média para as variáveis AC e RF foi muito próxima entre os genitores. Além disso, a média da população F_{23} para essas variáveis foi abaixo das médias obtidas para ambos os genitores. Embora os valores da distribuição dessas variáveis demonstrem a presença de indivíduos da população F_{23} com valores bem acima dos valores médios apresentados pelos genitores, a análise dos dados contendo as notas de cada parcela, não apresentados neste trabalho, permitiu verificar que a maioria dos indivíduos da população F_{23} obteve notas entre 1 e 2 para ambas as variáveis. Isso fica evidente no histograma da Figura 2, com a concentração de indivíduos na região inferior do gráfico. Portanto, a população utilizada neste estudo provavelmente não apresentou variabilidade suficiente para essas características. Considerando-se as justificativas apresentadas acima, as variáveis RF e AC não foram incluídas no mapeamento MT-MIM. No entanto, como pode-se observar abaixo, outras variáveis foram utilizadas em mais de um grupo para a análise multivariada, considerando-se os valores de correlação obtidos com outras variáveis e a importância das mesmas para a resistência de cultivares aos percevejos. Como já relatado anteriormente, outras opções de agrupamento entre as variáveis foram realizadas, com no máximo cinco variáveis por grupo. No entanto, considerando o objetivo da análise MT-MIM, que consistiu em identificar o maior número de regiões genômicas controlando as características estudadas, os nove grupos abaixo foram os que atenderam ao objetivo sendo, portanto, selecionados:

Multi-1: NDF, NDM, APM, NS e PG;

Multi-2: APF, APM e VA;

Multi-3: NDM e PSB;

Multi-4: APM e NVP;

Multi-5: PEG, PSM e PCS;

Multi-6: NVP, NS, PG e VA;

Multi-7: NVP, PSB e PSM;

Multi-8: IPDV, NS, PSB e PG;

Multi-9: PEG, IPDV, PCS e PG.

O modelo de mapeamento multivariado (MT-MIM) foi capaz de detectar oito posições genômicas para o *Multi-1* (Tabela 5 e Figura 6), seis posições genômicas para o *Multi-2* (Tabela 6 e Figura 7), sete para o *Multi-3* (Tabela 7 e Figura 8), três para o *Multi-4* (Tabela 8 e Figura 9), seis para o *Multi-5* (Tabela 9 e Figura 10), 10 para o *Multi-6* (Tabela 10 e Figura 11), oito para o *Multi-7* (Tabela 11 e Figura 12), nove para o *Multi-8* (Tabela 12 e Figura 13) e 11 posições genômicas para o *Multi-9* (Tabela 13 e Figura 14), com efeito em pelo menos um dos fenótipos avaliados em cada grupo de variáveis. Pode-se verificar, na comparação dos perfis de LOD dos diferentes grupos de variáveis analisados, que há evidência em favor da hipótese de pleiotropia para algumas regiões, da mesma forma como foi observado no caso das análises MIM. Comparando-se as posições genômicas identificadas para o conjunto de variáveis *Multi-1* a *Multi-9*, pode-se verificar que há uma coincidência de posições, inclusive do marcador flanqueando o QTL identificado. Sendo assim, um total de 20 posições genômicas diferentes foram identificadas neste estudo, localizadas em 12 GLs, sendo GL A1, B1, C1 (2 regiões), C2, D2 (3 regiões), E (4 regiões), F (2 regiões), I (2 regiões), J, K, M e N. Apesar de não terem sido realizados testes, no *OneQTL*, para verificar a presença de pleiotropia ou ligação, os resultados sugerem que há uma tendência de concentração dos QTLs identificados nessas 20 diferentes posições genômicas identificadas.

Ao se avaliar somente os QTLs com efeitos marginais significativos, a proporção da variância fenotípica explicada pelo modelo MT-MIM, utilizando-se inicialmente a função *forward*, variou de 27,98% para altura de plantas no florescimento (APF) (*Multi-2*) (Tabela 8) a 65,30% para número de dias para a maturidade (NDM) (*Multi-3*) (Tabela 7), considerando-se o maior valor de R^2 , no caso de variáveis presentes em mais de um grupo. No geral, os modelos MT-MIM apresentaram valores de R^2 marginais superiores aos modelos MIM, com poucas exceções. As maiores diferenças nos valores de R^2 entre o modelo MIM e MT-MIM foram observadas para as variáveis IPDV, NVP, PG e PSB. A maioria dos QTLs para essas variáveis, que não foram identificados nos modelos de fenótipos individuais, mas foram revelados pela abordagem MT-MIM, consiste em QTLs de menor efeito, explicando uma pequena percentagem da variação do fenótipo. Os resultados demonstram o maior poder estatístico da abordagem multivariada em identificar QTLs de menor efeito em relação ao modelo MIM, no caso dessas variáveis. No entanto, da mesma forma como observado nos modelos MIM, os modelos MT-MIM também apresentaram valores de R^2 inferiores aos valores de herdabilidade das variáveis analisadas. Nenhum trabalho foi encontrado na literatura utilizando a abordagem de mapeamento multivariado (MT-MIM), em soja. Por se tratar do primeiro trabalho utilizando essa metodologia, não será possível comparar diretamente os valores de R^2 obtidos com valores de R^2 disponíveis na literatura científica. No entanto, estudos de mapeamento de QTLs utilizando a abordagem CIM, em soja, associados às variáveis analisadas neste trabalho, também relatam valores de R^2 dos QTLs em consonância com os resultados obtidos. Gianotto (2015) utilizou a abordagem multivariada para estudar características fenotípicas, em milho, associadas à absorção de fósforo e também verificou o seu maior poder estatístico em relação à abordagem univariada. Duarte (2007) também relata que a abordagem multivariada permitiu a identificação de QTLs, associados à hipertensão e obesidade, que não haviam sido revelados com a abordagem univariada.

Em todos os grupos de variáveis analisadas, *Multi-1* a *Multi-9*, foi possível detectar posições genômicas coincidentes com as análises MIM para os fenótipos individuais. No entanto, como regra geral, a abordagem MT-MIM forneceu estimativas de LOD superiores e uma maior precisão na estimativa da posição do QTL, quando comparados aos QTLs identificados nos modelos MIM. Além disso, a abordagem multivariada permitiu identificar 13 regiões, com efeito em pelo menos uma das variáveis analisadas, que não haviam sido identificadas nos modelos MIM de fenótipos individuais. Portanto, os resultados corroboram a tendência dos modelos MT-MIM apresentarem maior precisão e poder estatístico quando comparados aos modelos MIM. As regiões genômicas identificadas somente com o uso da abordagem MT-MIM encontram-se em negrito nas Tabelas 5 a 13. Para o grupo de variáveis do *Multi-1* (NDF, NDM, APM, NS e PG), uma nova região genômica foi identificada pela abordagem MT-MIM em relação à abordagem MIM, sendo o *Multi1@E-1_68,00*. Para as variáveis do grupo *Multi-2* (APF, APM e VA), uma nova região genômica foi identificada na abordagem MT-MIM (*Multi2@M_15,11*). Para as variáveis dos grupos *Multi-3* (NDM e PSB), *Multi-4* (APM e NVP) e *Multi-5* (PEG, PCS e PSM), a abordagem MT-MIM identificou as mesmas regiões já reveladas na abordagem MIM. Para o grupo de variáveis do *Multi-6* (NVP, NS, PG e VA), quatro regiões genômicas foram reveladas, sendo *Multi6@D2-2_47,63*, *Multi6@D2-2_154,60*, *Multi6@E-1_69,00* e *Multi6@K_27,00*, com efeito em pelo menos uma das características avaliadas. No caso do *Multi-7* (NVP, PSB e PSM), sete regiões genômicas foram reveladas, *Multi7@D2-2_23,00*, *Multi7@E-1_43,97*, *Multi7@E-1_67,00*, *Multi7@F-2_113,59*, *Multi7@I-1_72,00*, *Multi7@M_14,88* e *Multi7@N_59,00*. Para o *Multi-8* (IPDV, NS, PSB e PG), quatro regiões genômicas foram identificadas (*Multi8@D2-2_184,32*, *Multi8@E-1_67,00*, *Multi8@M_14,88* e *Multi8@N_88,29*), e outras cinco regiões genômicas foram reveladas para o *Multi-9* (*Multi9@C2-1_97,29*, *Multi9@D2-2_184,32*, *Multi9@F-2_41,00*, *Multi9@M_15,11* e *Multi9@N_88,29*), afetando pelo menos uma das variáveis analisadas (PEG, IPDV, PG e PCS). Os resultados obtidos evidenciam a superioridade do modelo MT-MIM em identificar QTLs em relação ao modelo MIM utilizado. Mesmo apresentando *score* mais elevado do que a abordagem MIM, o modelo é capaz de revelar regiões que controlam pelo menos um dos fenótipos analisados.

Tabela 5. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-1*: **NDF**: número de dias para o florescimento; **NDM**: número de dias para a maturidade; **APM**: altura da planta na maturidade; **NS**: número de sementes e **PG**: produtividade de grãos. Os QTLs destacados em negrito não foram identificados na abordagem MIM.

MT-MIM: <i>Multi-1</i>													
QTL ¹	Marcadores	Adjacentes ²	LOD	NDF		NDM		APM		NS		PG	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi1@C1-1_23,00</i>	ss107913694	Satt194	22,11	0,688*** -0,139	6,64	2,540*** 0,481	18,64	1,319*** 0,821*	3,52 0,71	13,777*** 5,446	4,26	1,565*** 0,587	3,30
<i>Multi1@C2-1_99,00</i>	ss107912713	ss107920120	20,94	1,028*** 0,184	15,46	1,258*** -0,259	4,77	2,052*** 0,457	8,91	23,126*** 5,047	12,51	2,282*** 0,593	7,32
<i>Multi1@D2-2_46,00</i>	ss107914703	ss107921284	15,59	-0,177 -0,342*	0,78	-1,153*** -0,740*	3,94 0,75	-1,343*** 0,608	3,75	-5,131 0,233	0,61 0,00	-2,029*** -0,413	5,70
<i>Multi1@E-1_68,00</i>	ss107912757	ss107913532	16,6	0,295* -0,087	1,24	1,109*** 0,215	3,61	1,006*** -0,305	2,08	-0,123 -1,075	0,00 0,01	1,769*** 0,03	4,29
<i>Multi1@F-1_36,85</i>	ss107917146	ss107917146	7,87	0,108 0,617***	2,69	0,056 1,945***	5,52	0,058 1,191*	1,45	0,943 12,464**	1,76	-0,307 2,522***	4,33
<i>Multi1@I-1_40,00</i>	ss107926125	T4	29,67	1,106*** -0,353*	18,17 0,81	1,240*** -0,423	4,71	3,116*** -0,307	20,84	26,929*** -0,093	17,22	2,160*** -0,182	6,66
<i>Multi1@J_54,18</i>	ss107927403	ss107927403	20,14	0,707*** -0,028	7,83	0,233 0,110	0,18 0,02	2,326*** -0,023	12,27	13,306*** 2,775	4,44	1,709*** 0,923	4,40
<i>Multi1@M_14,88</i>	ss107912777	ss107912777	11,36	-0,499*** 0,021	3,58	-1,988*** 0,272	11,75	-1,164*** 0,063	2,82	-11,189** -0,901	2,89	-2,226*** -0,014	6,87
R² ⁶				57,20		53,69		56,35		43,08		42,87	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

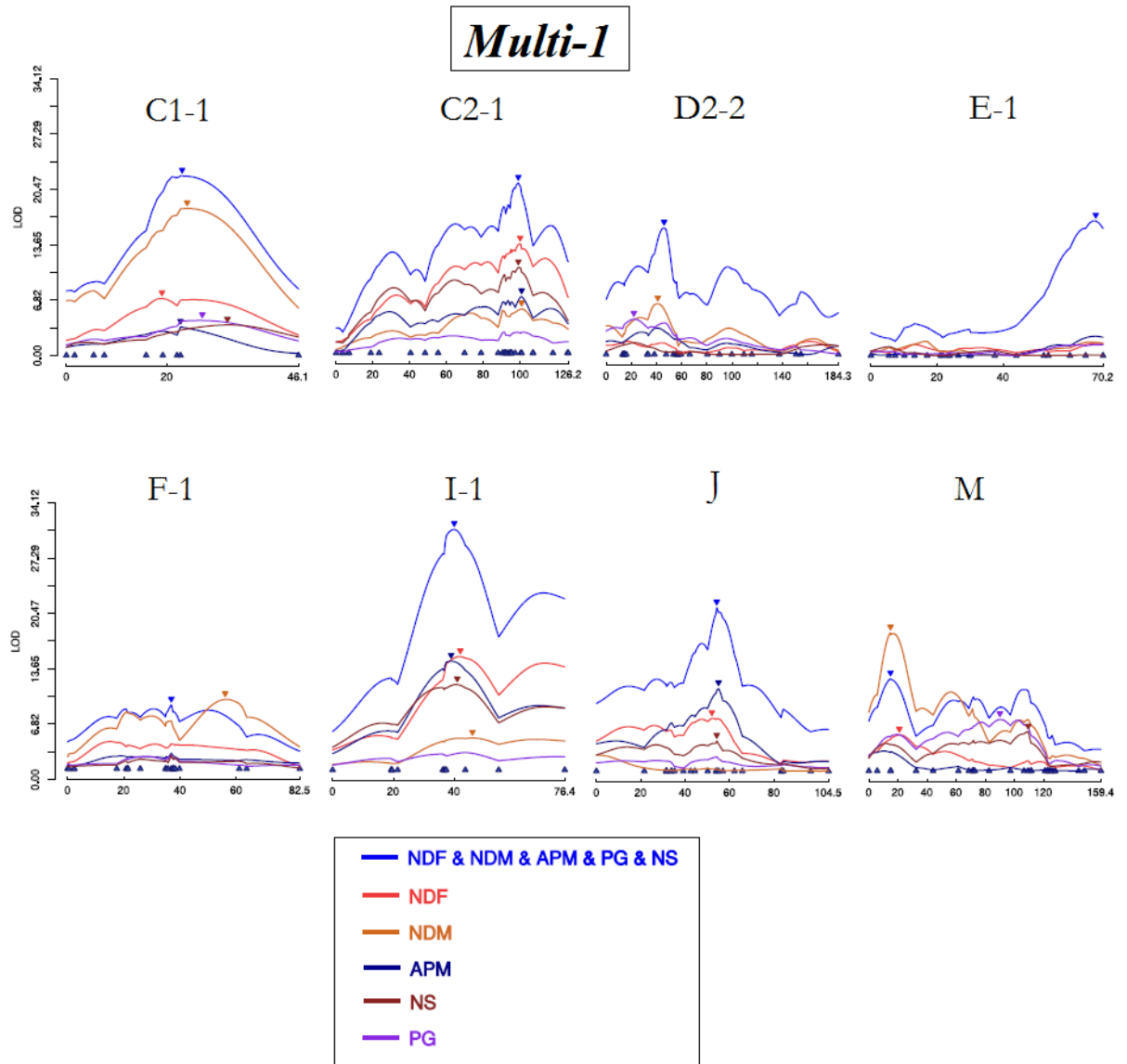


Figura 6. Representação gráfica do perfil de LOD obtido para as variáveis do grupo Multi-1 para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: NDF: número de dias para o florescimento; NDM: número de dias para a maturidade; APM: altura da planta na maturidade; NS: número de sementes e PG: produtividade de grãos. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 6. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-2*: **APF**: altura da planta no florescimento; **APM**: altura da planta na maturidade; **VA**: valor agronômico.

MT-MIM: <i>Multi-2</i>									
QTL ¹	Marcadores	Adjacentes ²	LOD	APF		APM		VA	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi2@C1-1_22,68</i>	ss107913694	ss107913694	7,67	0,443	1,96	1,403***	4,05	-0,001	0,00
				0,838*		0,848*		0,018	
<i>Multi2@C2-1_99,00</i>	ss107912713	ss107920120	17,17	0,470	1,20	2,005***	8,50	0,045***	13,10
				0,131		0,389		0,015	
<i>Multi2@D2-2_45,00</i>	ss107914703	ss107921284	7,45	-1,343***	9,38	-1,394***	3,95	-0,011	0,69
				0,157		0,586		0,009	
<i>Multi2@I-1_38,00</i>	ss107926125	T4	21,20	1,190***	8,45	2,855***	18,97	0,046***	14,73
				0,533		-0,191		0,004	
<i>Multi2@J_54,18</i>	ss107927403	ss107927403	12,31	1,187***	8,19	2,344***	12,46	0,034***	8,04
				0,310		-0,110		0,004	
<i>Multi2@M_15,11</i>	ss107913049	ss107913049	6,57	-0,205	0,22	-1,218***	3,06	-0,002	0,03
				0,014		-0,009		-0,013	
R² ⁶				27,76		48,71		35,24	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

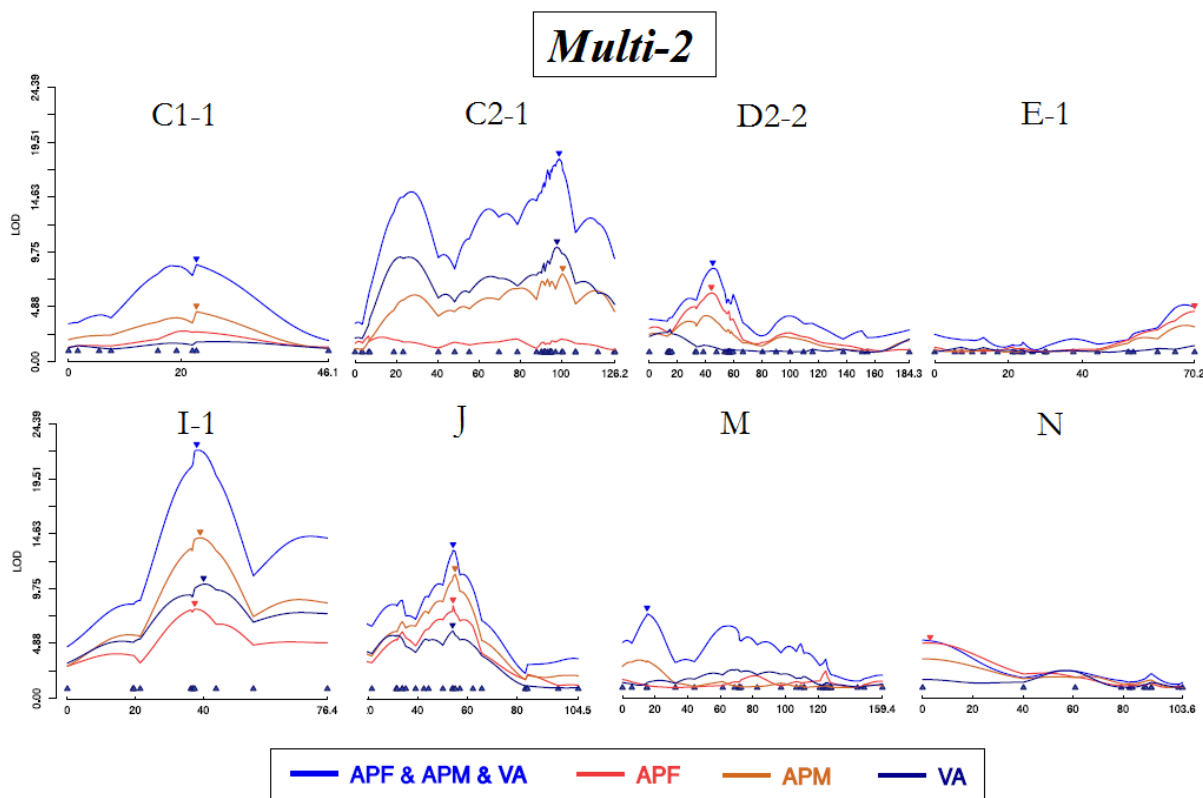


Figura 7. Representação gráfica do perfil de LOD obtido para as variáveis do grupo *Multi-2* para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: **APF**: altura da planta no florescimento; **APM**: altura da planta na maturidade; **VA**: valor agronômico. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 7. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-3*: **NDM**: número de dias para a maturidade; **PSB**: peso de sementes boas.

MT-MIM: <i>Multi-3</i>							
QTL ¹	Marcadores	Adjacentes ²	LOD	NDM		PSB	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi3@A1_40,00</i>	ss107913925	ss107913319	9,51	-1,704*** -0,063	8,20	-0,890 -0,448	1,45 0,16
<i>Multi3@C1-1_24,00</i>	ss107913694	Satt194	18,86	2,630*** 0,451	19,01	1,772*** 0,29	5,59
<i>Multi3@C1-2_24,10</i>	ss107924445	ss107924445	7,91	1,226*** -0883*	5,14 1,91	1,149 -1,565**	2,42
<i>Multi3@C2-1_97,29</i>	ss107912713	ss107912713	6,2	1,103*** 0,247	3,85	1,798** 0,689	6,63
<i>Multi3@D2-2_40,00</i>	ss107914703	ss107921284	6,31	-1,288*** -0,393	4,93	-1,462*** 0,131	2,75
<i>Multi3@F-1_56,00</i>	ss107912827	Satt586	7,49	0,313 2,366***	6,05	0,44 1,823**	2,33
<i>Multi3@M_14,88</i>	ss107912777	ss107912777	16,88	-2,335*** 0,281	16,21	-2,060*** -0,314	8,17
R² 6				40,89		27,89	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (*p* < 0,05; ***p* < 0,01; ****p* < 0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

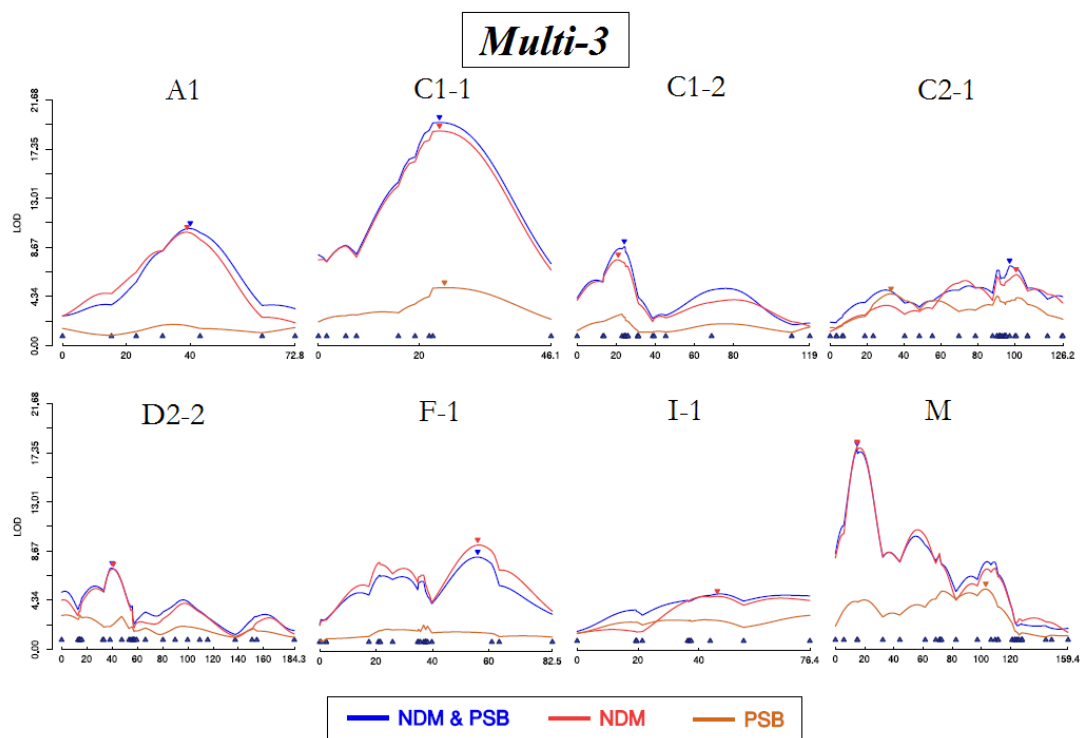


Figura 8. Representação gráfica do perfil de LOD obtido para as variáveis do grupo *Multi-3*, para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: **NDM**: número de dias para a maturidade; **PSB**: peso de sementes boas. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 8. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-4*: *APM*: altura da planta na maturidade; *NVP*: número de vagens por planta.

MT-MIM: <i>Multi-4</i>								
QTL ¹	Marcadores	Adjacentes ²	LOD	APM		NVP		
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	
<i>Multi4@C2-1_100,33</i>	ss107920120	ss107920120	8,97	2,065*** 0,127	9,14	11,510*** 0,517	11,48	
<i>Multi4@I-1_40,00</i>	ss107926125	T4	14,47	2,971*** -0,171	18,95	11,482*** 2,964	11,43	
<i>Multi4@J_55,00</i>	ss107927403	ss107927503	11,65	2,595*** -0,426	15,25	8,024*** 1,29	5,89	
R²⁶				28,09		22,91		

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* ($p < 0,05$; $**p < 0,01$; $***p < 0,001$); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

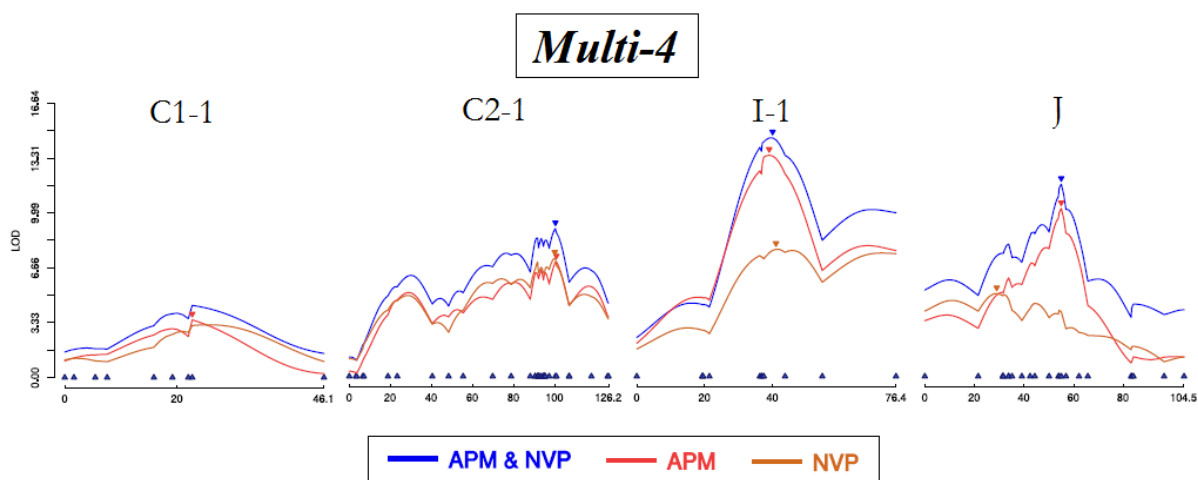


Figura 9. Representação gráfica do perfil de LOD obtido para as variáveis do grupo *Multi-4*, para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: *APM*: altura da planta na maturidade; *NVP*: número de vagens por planta. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 9. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-5*: **PEG**: período de granação; **PCS**: peso de cem sementes e **PSM**: peso de sementes manchadas.

MT-MIM: <i>Multi-5</i>									
QTL ¹	Marcadores	Adjacentes ²	LOD	PEG		PCS		PSM	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi5@C1-1_22,00</i>	ss107912554	ss107913694	9,93	0,893*** 0,411	7,73	0,064 0,088	0,09 0,09	0,144 0,205	0,32 0,34
<i>Multi5@D2-2_22,00</i>	ss107912878	ss107929556	9,09	-0,930*** -0,026	7,49	-0,746*** -0,048	11,29	-0,709*** 0,091	6,85
<i>Multi5@E-1_14,00</i>	ss107924423	ss107919601	6,07	-0,837*** -0,437	6,19	-0,491*** -0,238	5,00	-0,229 -0,201	0,73 0,30
<i>Multi5@E-1_68,00</i>	ss107912757	ss107913532	17,75	0,493** -0,144	2,34	0,931*** -0,091	19,54	0,272 -0,517*	1,90
<i>Multi5@I-1_40,00</i>	ss107926125	T4	18,25	-1,262*** -0,331	15,97	-0,333** -0,091	2,61	0,483** -0,096	3,69
<i>Multi5@J_54,00</i>	ss107913908	ss107927403	13,64	-0,761*** -0,058	6,10	0,181 0,039	0,81 0,02	0,399* -0,231	2,63
R² 6				45,82		38,44		15,07	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p* < 0,05; ***p* < 0,01; ****p* < 0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

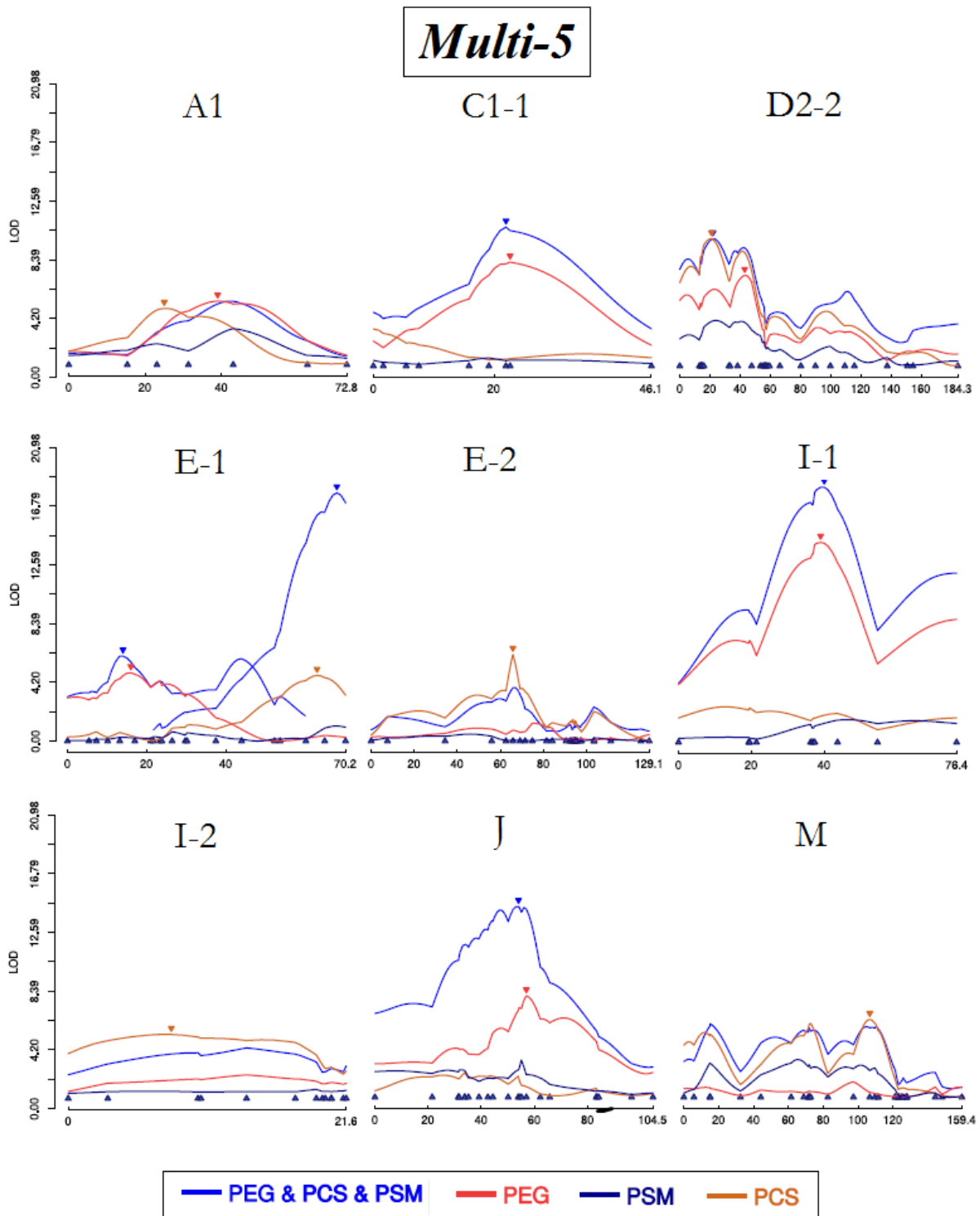


Figura 10. Representação gráfica do perfil de LOD obtido para as variáveis do grupo *Multi-5*, para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: **PEG**: período de granação; **PCS**: peso de cem sementes e **PSM**: peso de sementes manchadas. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 10. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-6*: **NVP**: número de vagens por planta; **NS**: número de sementes; **PG**: produtividade de grãos e **VA**: valor agrônômico. Os QTLs destacados em negrito não foram identificados na abordagem MIM.

MT-MIM: <i>Multi-6</i>											
QTL ¹	Marcadores	Adjacentes ²	LOD	NVP		NS		PG		VA	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi6@C1-1_29,00</i>	ss107913694	Satt194	11,54	9,740*** 4,623	5,98	19,207*** 5,391	6,37	2,424*** 0,457	6,09	-0,002 0,024*	1,20
<i>Multi6@C2-1_97,29</i>	ss107912713	ss107912713	13,52	10,557*** 2,917	9,99	21,426*** 7,332	11,27	2,044*** 1,148*	6,17	0,047*** 0,018*	14,40
<i>Multi6@D2-2_14,00</i>	ss107923107	ss107929993	7,84	-2,445 4,474	0,54	-5,654 9,997*	1,19	-1,388*** 1,402	2,87	-0,019** -0,005	2,34
<i>Multi6@D2-2_47,63</i>	ss107921284	ss107921284	9,77	-8,173*** -3,256	5,96	-12,549** -7,503	3,85	-2,647*** -1,389**	10,29	-0,003 0,006	0,07
<i>Multi6@D2-2_154,60</i>	ss107914356	ss107914356	8,29	8,633*** 0,693	7,47	16,713*** 2,556	7,67	2,282*** -0,279	8,60	0,007 0,005	0,32
<i>Multi6@E-1_69,00</i>	ss107912757	ss107913532	17,15	-0,119 -4,070	0,00	0,905 -6,409	0,02	1,761*** -1,148	4,30	0,009 -0,002	0,46
<i>Multi6@I-1_40,00</i>	ss107926125	T4	20,18	12,240*** 3,310	12,99	28,015*** 1,847	18,64	2,291*** 0,119	7,49	0,049*** 0,005	15,62
<i>Multi6@J_55,89</i>	ss107913908	ss107913908	7,18	5,471** 2,004	2,73	10,877*** 0,832	2,96	1,329** 0,656	2,65	0,034*** 0,006	8,02
<i>Multi6@K-1_27,00</i>	A28	Satt001	7,00	0,400 -4,946	0,01	0,098 -11,150*	1,15	0,730 -2,241***	2,80	0,005 0,001	0,14
<i>Multi6@M_75,00</i>	ss107913100	A57	13,49	-6,870*** 2,411	3,89	-13,854*** 2,561	4,33	-2,328*** 1,098	7,35	0,003 -0,025*	1,71
R² 6				49,01		57,43		60,98		44,33	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos

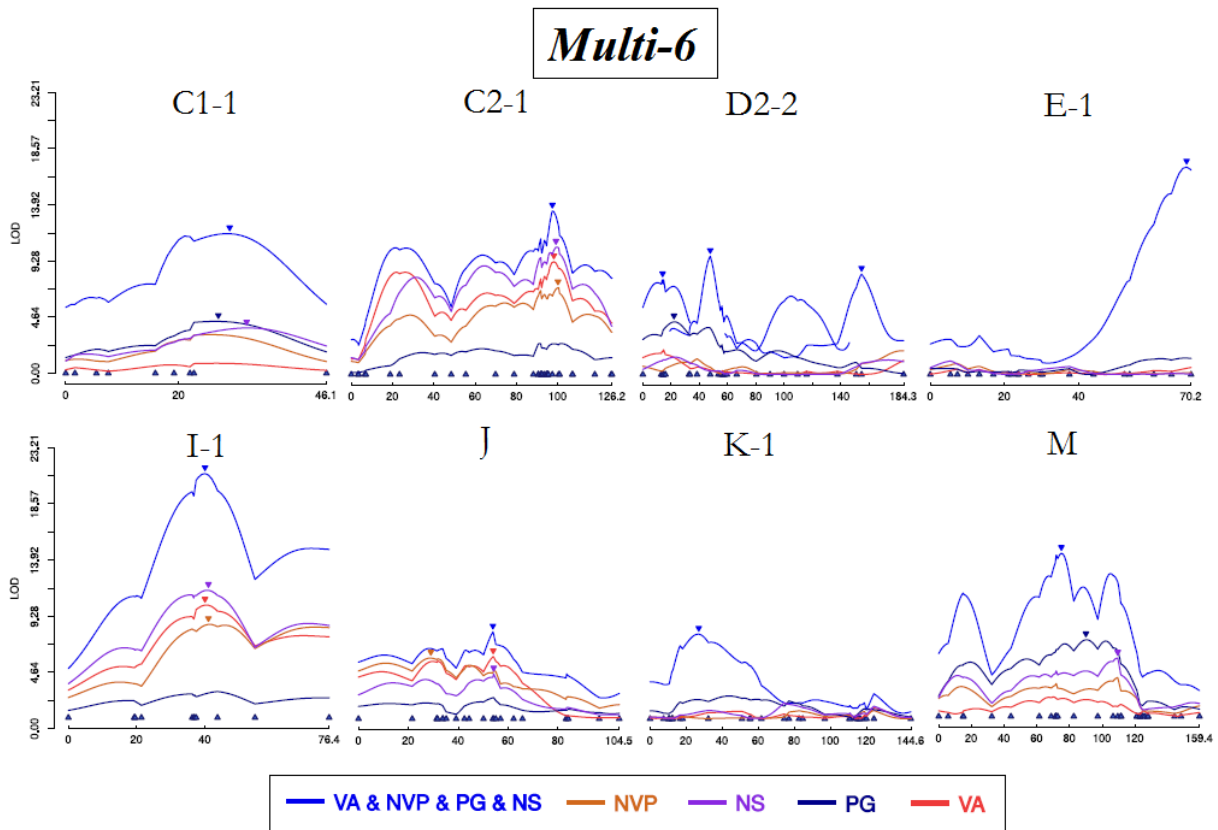


Figura 11. Representação gráfica do perfil de LOD obtido para as variáveis do grupo *Multi-6*, para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: **NVP**: número de vagens por planta; **NS**: número de sementes; **PG**: produtividade de grãos e **VA**: valor agrônômico. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 11. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-7*: **NVP**: número de vagens por planta, **PSB**: peso de sementes boas e **PSM**: peso de sementes manchadas. Os QTLs destacados em negrito não foram identificados na abordagem MIM.

MT-MIM: <i>Multi-7</i>									
QTL ¹	Marcadores	Adjacentes ²	LOD	NVP		PSB		PSM	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi7@C2-1_91,66</i>	ss107928441	ss107928441	12,23	12,930*** 4,514	14,34	2,361*** 0,705	10,94	0,318* 0,498**	1,58 2,03
<i>Multi7@D2-2_23,00</i>	ss107912878	ss107929556	8,02	-5,906** 0,434	2,60	-1,653*** 0,315	4,66	-0,708*** -0,07	6,79
<i>Multi7@E-1_43,97</i>	ss107918043	ss107918043	6,83	1,519 3,649	0,20 0,58	-0,573 -0,695	0,64 0,48	-0,12 0,473*	1,78
<i>Multi7@E-1_67,00</i>	ss107912757	ss107913532	19,33	-0,283 -3,591	0,01 0,51	2,011*** -0,083	7,70	0,359* -0,611**	1,96 2,66
<i>Multi7@F-2_113,59</i>	ss107916098	ss107916098	6,72	5,579** 3,205	2,99	0,825* -0,006	1,50	0,007 0,671**	3,12
<i>Multi7@I-1_72,00</i>	Satt239	A24	11,03	14,364*** 4,2912	13,60	1,812*** 0,093	4,95	0,557*** -0,109	3,72
<i>Multi7@M_14,88</i>	ss107912777	ss107912777	7,72	-6,411*** 1,822	3,46	-1,821*** -0,028	6,39	-0,515*** 0,406*	4,06 1,33
<i>Multi7@N_59,00</i>	Satt387	ss107913993	8,97	-3,281 10,371***	4,11	0,714 1,452**	1,84	-0,352* -0,482	1,83
R² 6				41,10		37,98		30,86	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* ($p < 0,05$; $**p < 0,01$; $***p < 0,001$); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

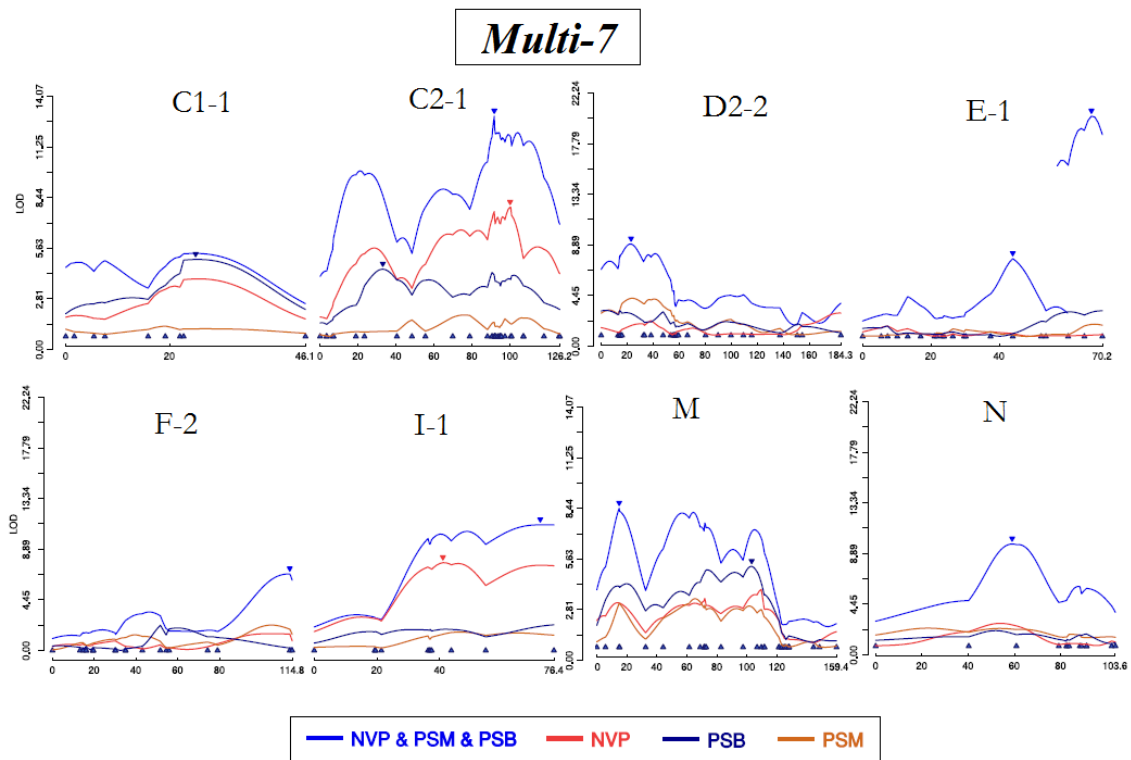


Figura 12. Representação gráfica do perfil de LOD obtido para as variáveis do grupo *Multi-7*, para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: **NVP**: número de vagens por planta, **PSB**: peso de sementes boas e **PSM**: peso de sementes manchadas. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 12. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-8*: **IPDV**: índice percentual de dano nas vagens; **NS**: número de sementes; **PSB**: peso de sementes boas e **PG**: produtividade de grãos. Os QTLs destacados em negrito não foram identificados na abordagem MIM.

MT-MIM: <i>Multi-8</i>											
QTL ¹	Marcadores	Adjacentes ²	LOD	IPDV		NS		PSB		PG	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi8@B1-2_30,00</i>	ss107913659	ss107919849	7,12	0,006 -0,028***	5,25	-5,904 17,501***	2,94	-0,646* 1,876***	0,84 2,82	-0,657 2,587***	3,86
<i>Multi8@C2-1_98,00</i>	ss107912713	ss107920120	13,56	-0,16*** -0,017**	4,17 2,38	23,935*** 7,630	13,67	2,093*** 0,649	8,72	2,465*** 1,017	8,71
<i>Multi8@D2-2_44,00</i>	ss107914703	ss107921284	11,53	0,003 -0,007	0,11 0,31	-5,981 -1,029	0,79 0,01	-1,503*** -0,39	4,17	-2,12*** -0,621	5,98
<i>Multi8@D2-2_184,32</i>	Satt386	Satt386	6,41	-0,008 -0,009	1,06 0,74	3,895 -14,795**	2,68	0,250 -0,711	0,13 0,52	0,342 -1,074	0,17 0,85
<i>Multi8@E-1_67,00</i>	ss107912757	ss107913532	17,10	-0,13** 0,001	2,65	-0,641 -6,064	0,01 0,39	1,408*** -0,422	3,77	1,623*** -0,881	3,61
<i>Multi8@I-1_40,00</i>	ss107926125	T4	22,20	-0,027*** 0,016**	11,98 1,84	27,045*** 1,790	17,37	1,674*** 0,217	5,55	2,107*** 0,052	6,34
<i>Multi8@J_55,00</i>	ss107927403	ss107927503	10,82	-0,020*** 0,008	6,91	15,984*** 2,860	6,40	1,449** 0,969*	4,39 0,93	1,969*** 0,701	5,84
<i>Multi8@M_14,88</i>	ss107912777	ss107912777	12,61	0,021*** 0,006	7,17	-9,826** -2,244	2,23	-1,749*** -0,577	5,89	-2,174*** -0,278	6,55
<i>Multi8@N_88,29</i>	ss107918394	ss107918394	9,45	-0,010* 0,026***	1,64 5,53	0,172 1,770	0,00 0,04	0,839* 0,252	1,52	0,517 0,248	0,41 0,04
R² ⁶				49,52		45,29		38,60		40,89	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p* < 0,05; ***p* < 0,01; ****p* < 0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

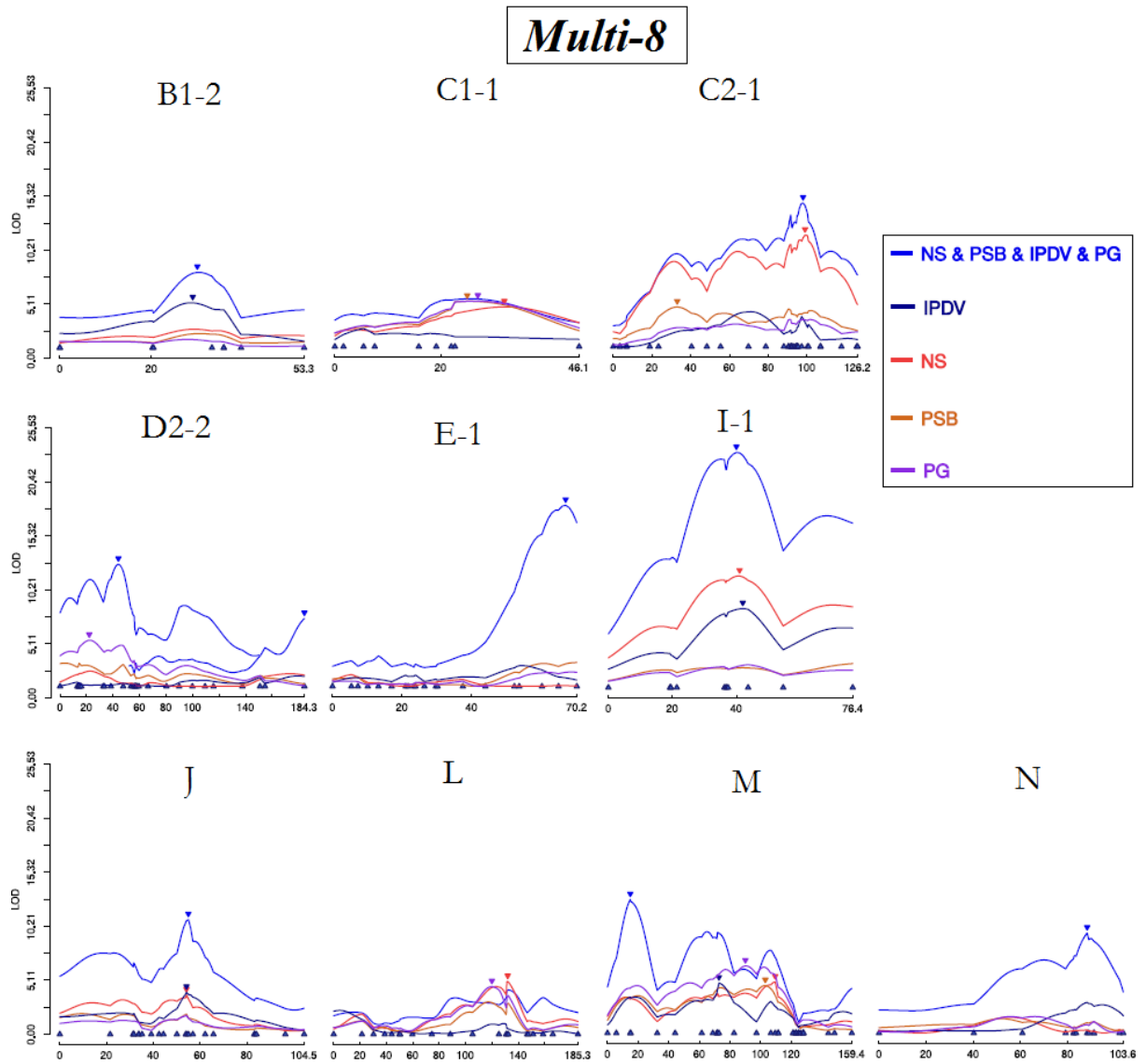


Figura 13. Representação gráfica do perfil de LOD obtido para as variáveis do grupo Multi-8, para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: IPDV: índice percentual de dano nas vagens; NS: número de sementes; PSB: peso de sementes boas e PG: produtividade de grãos. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Tabela 13. QTLs identificados usando a análise MT-MIM para as características do grupo *Multi-9*: **PG**: produtividade de grãos; **IPDV**: índice percentual de dano nas vagens; **PCS**: peso de cem sementes e **PEG**: período de granação. Os QTLs destacados em negrito não foram identificados na abordagem MIM.

MT-MIM: <i>Multi-9</i>											
QTL ¹	Marcadores	Adjacentes ²	LOD	PG		IPDV		PCS		PEG	
				Aditivo ³ Domin. ⁴	% Var ⁵	Aditivo ³ Domin. ⁴	% Var ⁵	Aditivo ³ Domin. ⁴	% Var ⁵	Aditivo ³ Domin. ⁴	% Var ⁵
<i>Multi9@C1-1_22,68</i>	ss107913694	ss107913694	12,66	1,766**	4,22	-0,008*	0,93	0,16	0,58	1,004***	9,69
<i>Multi9@C2-1_97,29</i>	ss107912713	ss107912713	12,86	2,398***	8,49	-0,016***	4,12	0,046	0,05	-0,206	0,44
				1,014		-0,015**	1,99	0,169	0,35	-0,341	0,60
<i>Multi9@D2-2_18,00</i>	ss107912878	ss107929556	13,18	-2,385***	7,61	0,006	1,56	-0,744***	12,19	-0,850***	6,78
				1,099		-0,015*		0,112		-0,049	
<i>Multi9@D2-2_184,32</i>	Satt386	Satt386	6,46	0,101	0,01	-0,010*	1,54	-0,162	0,63	-0,327*	1,10
				-0,996		0,73		-0,015*		1,95	
<i>Multi9@E-1_67,00</i>	ss107912757	ss107913532	11,57	1,193*	1,95	-0,018***	4,92	0,594***	7,95	0,154	0,23
				-1,113		0,001		-0,147		0,018	
<i>Multi9@E-2_65,82</i>	A51	A51	6,58	0,899*	1,19	0,008	1,18	0,470***	5,34	0,433*	1,94
				0,468		-0,004		0,15		0,239	
<i>Multi9@F-2_41,00</i>	ss107917362	ss107913720	7,70	0,121	0,02	-0,018***	4,35	-0,239*	1,15	-0,451**	1,75
				-1,136		0,81		-0,004		-0,555***	
<i>Multi9@I-1_40,00</i>	ss107926125	T4	28,46	1,943***	5,39	-0,027***	11,38	-0,372***	3,24	-1,376***	18,99
				-0,043		0,011		-0,167		-0,507*	
<i>Multi9@J_53,00</i>	ss107913698	ss107913908	20,63	1,629**	3,84	-0,019***	5,84	0,141	0,47	-0,834***	7,08
				0,869		0,012*		1,11		0,140	
<i>Multi9@M_15,11</i>	ss107913049	ss107913049	12,45	-2,25***	6,95	0,022***	7,49	-0,549***	6,80	-0,355*	1,22
				0,258		0,002		0,151		0,299	
<i>Multi9@N_88,29</i>	ss107918394	ss107918394	9,53	0,487	0,37	-0,009**	1,42	0,184*	0,86	-0,164	0,29
				0,378		0,10		0,027***		5,94	
R² ⁶				39,64		56,89		40,70		50,73	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvio em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (*p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; ⁶R² calculado com o modelo contendo apenas os efeitos marginais significativos.

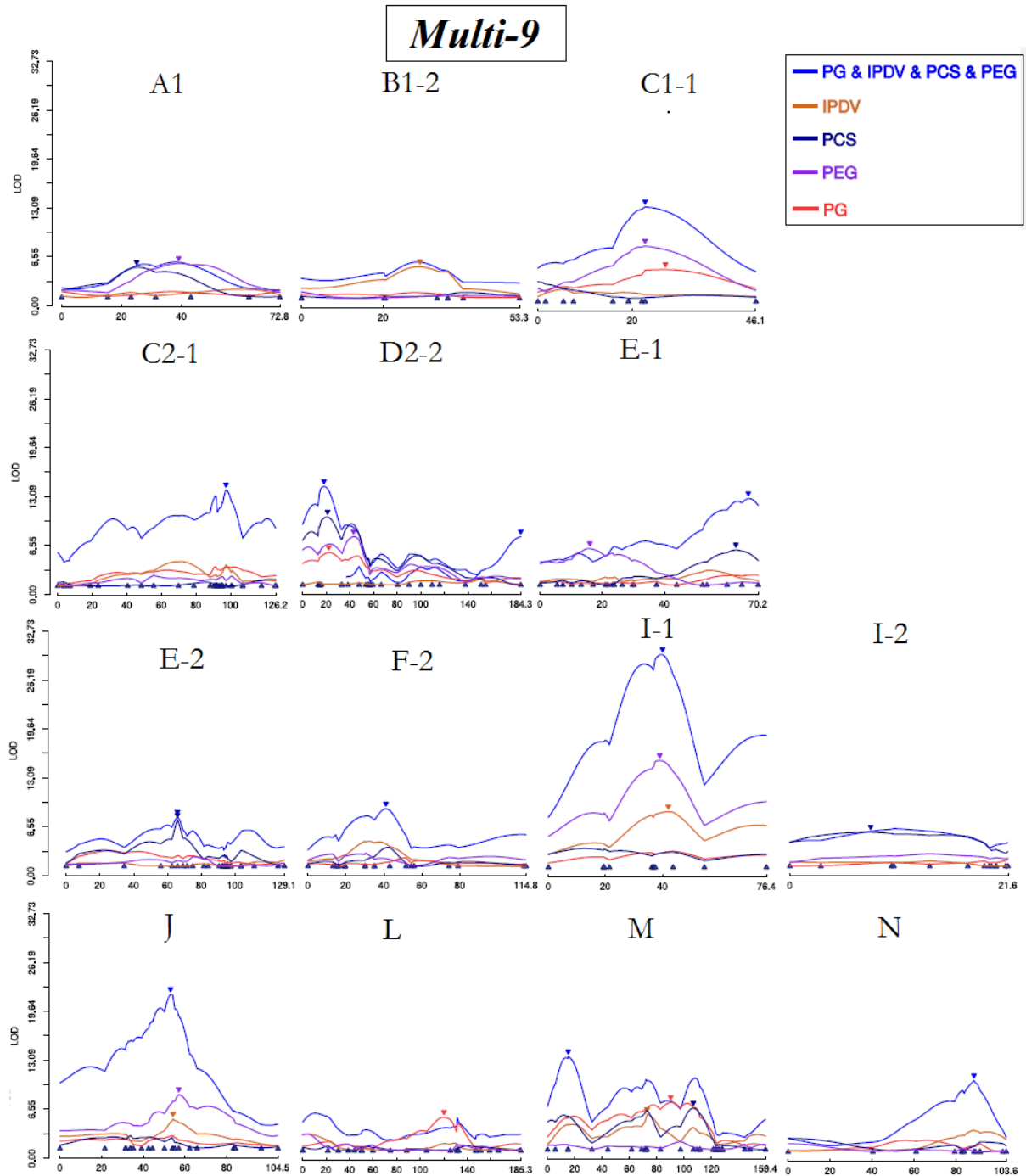


Figura 14. Representação gráfica do perfil de LOD obtido para as variáveis do grupo Multi-9, para o modelo de mapeamento univariado e multivariado. A cor azul clara representa o resultado obtido com o mapeamento multivariado. As demais cores estão indicadas na legenda e representam os resultados do mapeamento univariado para as características: IPDV: índice percentual de dano nas vagens; PCS: peso de cem sementes; PEG: período de granação e EPG: produtividade de grãos. O mapa genético estimado é representado no eixo x, em cM, pela posição dos marcadores genéticos ilustrada por triângulos de cor azul. Os triângulos invertidos indicam a posição dos QTLs incluídos no modelo final selecionado.

Como forma de refinar o modelo, a função *backmain* foi utilizada nas análises dos conjuntos *Multi-1* a *Multi-9*. Basicamente, o modelo completo, com todos os QTLs já inseridos, é contrastado com o modelo sem um QTL por vez e as razões de verossimilhanças são estimadas para determinar os QTLs que permanecerão no modelo. As Tabelas 14 a 22 apresentam os efeitos marginalmente significativos após o procedimento de eliminação *backward*, empregando um nível de significância de 1% ($p = 0,01$). Um nível de significância restritivo foi utilizado, para verificar os QTLs que

realmente apresentam efeito sobre as variáveis estudadas. Pode-se verificar que o modelo obtido após o procedimento da eliminação *backward* revelou as mesmas posições genômicas identificadas pelo procedimento *forward* (Tabelas 5 a 13), sendo oito posições genômicas para o *Multi-1* (Tabela 14), seis posições genômicas para o *Multi-2* (Tabela 15), sete para o *Multi-3* (Tabela 16), três para o *Multi-4* (Tabela 17), seis para o *Multi-5* (Tabela 18), 10 para o *Multi-6* (Tabela 19), oito para o *Multi-7* (Tabela 20), nove para o *Multi-8* (Tabela 21) e 11 posições genômicas para o *Multi-9* (Tabela 22). No entanto, para a maioria dos QTLs foi observada uma alteração na estimativa da posição dos mesmos. Essas diferenças podem ser explicadas pelo fato dos parâmetros do modelo terem sido re-estimados após a eliminação dos efeitos marginalmente não significativos, ocorrendo então um novo ajuste da posição dos QTLs. Pode-se verificar também que, para a maioria das variáveis analisadas, houve uma redução nos valores marginais de R^2 no modelo MT-MIM após a eliminação *backward*, cujos valores marginais de R^2 variaram de 14,03% para PSM a 59,84% para a variável NDM. A variável PSM, no *Multi-7*, foi a que apresentou a maior redução na percentagem da variação fenotípica explicada pelo QTL após a eliminação *backward*. Considerando somente os efeitos marginais significativos para essa variável, houve uma redução no valor do R^2 , que explicava 30,86% da variação fenotípica, no modelo MT-MIM *forward* (Tabela 11), para 14,03% no modelo MT-MIM *backward* (Tabela 20), o que representa uma redução de 54% no valor inicial do r^2 . Essa redução é explicada pelo fato do modelo *backward* não ter mantido cinco QTLs que foram identificados no modelo *forward*. Todos esses QTLs identificados exclusivamente pelo modelo MT-MIM *forward* apresentavam valor do LOD inferior aos demais. Além disso, verificou-se que a maioria desses QTLs apresentava significância de 10% ($p < 0,1$), demonstrando que o modelo *backward* é mais rigoroso em relação ao modelo *forward*.

Como já mencionado anteriormente, algumas características foram analisadas em mais de um conjunto de variáveis, considerando os dados obtidos de correlação e a importância das mesmas para a resistência aos percevejos. Foi possível observar uma grande diferença nos valores de R^2 marginais obtidos para uma mesma variável, em diferentes conjuntos das análises multivariadas. A maior diferença nos valores de R^2 marginais obtidos, em pontos percentuais, foi para a variável produtividade de grãos (PG). Além de ser uma característica bastante complexa, com alta influência ambiental, é controlada por muitos genes e possivelmente a maioria possui um pequeno efeito sobre a característica. No entanto, foi possível observar que, dos quatro conjuntos onde a variável estava presente (*Multi-1*, *Multi-6*, *Multi-8* e *Multi-9*), o maior valor de r^2 marginal obtido para PG foi no *Multi-6*, no qual também foram analisadas as variáveis número de vagens por plantas (NVP), número de sementes (NS) e valor agrônômico (VA) (Tabela 19). A correlação entre as variáveis NVP, NS e PG é positiva e forte (Figura 2), e uma planta com aspecto de ser mais produtiva, conseqüentemente receberá uma nota maior de valor agrônômico. O *Multi-6* identificou 10 QTLs com efeitos significativos para a variável PG, enquanto os outros conjuntos de variáveis identificaram somente oito QTLs com efeitos significativos, o que justifica o maior valor de R^2 para PG no *Multi-6*. Já a variável PSB foi analisada em três grupos diferentes. Pode-se verificar que também houve diferença nos valores de R^2 marginais obtidos para essa variável, sendo 28,97% para o *Multi-3*, 33,93% para o *Multi-8* e 40,73% para o *Multi-7*. Comparando-se os QTLs identificados em cada análise, pode-se concluir que o *Multi-7* foi o que mais identificou QTLs individuais para essa variável, contribuindo para o maior valor de R^2 marginal obtido para PSB. Nesse conjunto de variáveis, além de PSB, foram analisadas as variáveis número de vagens por planta (NVP) e peso de sementes manchadas (PSM). O maior valor de correlação observado para PSB foi com a variável NVP (Figura 2). Portanto, os resultados demonstram a importância da análise de correlação na escolha das características para a análise multivariada, visando a identificação de um maior número de QTLs explicando a variação fenotípica das mesmas.

Notavelmente, os modelos de mapeamento multivariado também foram capazes de explicar somente uma parte da variação fenotípica observada. Uma das razões dessa diferença observada entre os valores estimados de h^2 e

R^2 pode ser explicada pelo fato deste trabalho ter focado na busca de efeitos principais dos QTLs. Assim, os modelos de interação entre QTLs não foram incluídos no modelo de mapeamento final. Pode-se verificar em outros estudos, que muitos QTLs identificados em soja apresentam epistasia entre QTLs (QQ) e também interação QTL x E (QE) (Hwang 2010; Niu et al. 2013). Dessa forma, a inclusão dessas interações no modelo provavelmente contribuirá para reduzir a variância residual, fornecendo valores de R^2 superiores ao relatados neste trabalho. Como esse é o primeiro trabalho de mapeamento utilizando a abordagem MT-MIM em soja, não foi possível comparar os valores de R^2 estimados com resultados obtidos utilizando-se a mesma abordagem. No entanto, os estudos de mapeamento de QTLs que empregaram mapeamento CIM e um estudo que empregou a abordagem MIM, para algumas características aqui analisadas, revelaram QTLs com efeitos de magnitudes similares aos obtidos neste trabalho para o mapeamento de características individuais (Zhang et al. 2004; Hwang 2010; Han et al. 2012; Zhang et al. 2015). No entanto, os resultados obtidos com a abordagem MT-MIM se mostraram relevantes na identificação de novos QTLs para, no mínimo, uma das características analisadas conjuntamente, ou, em outros casos, para refinar os efeitos e as posições de QTLs já detectados nas análises individuais. Assim, os resultados obtidos evidenciam que as características estudadas apresentam arquitetura genética complexa, sendo que vários locos afetam tais características e cada um contribui com um pequeno efeito no fenótipo total.

Tabela 14. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-1*: **NDF**: número de dias para o florescimento; **NDM**: número de dias para a maturidade; **APM**: altura da planta na maturidade; **NS**: número de sementes e **PG**: produtividade de grãos. Os QTLs destacados em negrito não foram identificados na abordagem MIM. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-1</i>													
QTL ¹	Marcadores	Adjacentes ²	LOD	NDF		NDM		APM		NS		PG	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi1@C1-1_22,68</i>	ss107913694	Satt194	21,00	0,698*** -0,139	6,94	2,593*** 0,481	19,75	1,363*** 0,821	3,83	14,013*** 5,446	4,48	1,638*** 0,587	3,68
<i>Multi1@C2-1_99,00</i>	ss107912713	ss107920120	18,19	0,988*** 0,184	14,26	1,331*** -0,259	5,34	2,047*** 0,457	8,86	23,101*** 5,047	11,73	2,311*** 0,593	7,51
<i>Multi1@D2-2_46,00</i>	ss107914703	ss107921284	11,73	<i>sur</i>		-0,806*** -0,74	1,93	-1,013*** 0,608	2,14	<i>sur</i>		-1,364*** -0,413	2,58
<i>Multi1@E-1_68,00</i>	ss107912757	ss107913532	14,59	<i>sur</i>		0,780*** 0,215	1,79	0,789*** -0,305	1,28	<i>sur</i>		1,701*** 0,03	3,96
<i>Multi1@F-1_47,00</i>	ss107912827	Satt586	6,46	0,108 0,610**	1,72	0,056 2,079***	4,13	<i>sur</i>		<i>sur</i>		-0,395** 1,149***	0,23 0,59
<i>Multi1@I-1_40,00</i>	ss107926125	T4	27,46	1,060*** -0,353	16,67	1,058*** -0,423	3,43	2,935*** -0,307	18,49	24,948*** -0,093	14,79	1,825*** -0,182	4,75
<i>Multi1@J_54,18</i>	ss107927403	ss107927403	17,41	0,666*** -0,028	6,94	<i>sur</i>		2,115*** -0,023	10,14	11,603*** 2,775	3,38	1,461** 0,923	3,22
<i>Multi1@M_14,88</i>	ss107912777	ss107912777	12,09	-0,526*** 0,021	3,98	-2,616*** 0,272	13,30	-1,199*** 0,063	3,00	-11,502** -0,901	3,05	-2,285*** -0,014	7,24
R²				50,51		49,67		47,74		37,43		33,76	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (*p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 15. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-2*: **APF**: altura da planta no florescimento; **APM**: altura da planta na maturidade; **VA**: valor agrônomico. Os QTLs destacados em negrito não foram identificados na abordagem MIM. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-2</i>									
QTL ¹	Marcadores	Adjacentes ²	LOD	APF		APM		VA	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi2@C1-1_22,68</i>	ss107913694	ss107913694	5,16	<i>sur</i>		0,991*** 0,848	2,02	<i>sur</i>	
<i>Multi2@C2-1_99,00</i>	ss107912713	ss107920120	15,65	<i>sur</i>		1,479*** 0,389	4,63	0,038*** 0,015	9,02
<i>Multi2@D2-2_44,00</i>	ss107914703	ss107921284	6,14	-1,125*** 0,157	6,49	-1,146*** 0,586	2,62	<i>sur</i>	
<i>Multi2@I-1_38,00</i>	ss107926125	T4	21,60	1,134*** 0,671**	7,68 1,22	2,838*** -0,191	18,75	0,045*** 0,004	13,86
<i>Multi2@J_54,18</i>	ss107927403	ss107927403	11,33	1,233*** 0,310	8,84	2,357*** -0,110	12,59	0,032*** 0,004	7,07
<i>Multi2@M_15,11</i>	ss107913049	ss107913049	5,45	<i>sur</i>		-0,990*** -0,009	2,02	<i>sur</i>	
<i>R</i>²						24,23	42,63		29,95

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 16. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-3*: **NDM**: número de dias para a maturidade; **PSB**: peso de sementes boas. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-3</i>								
QTL ¹	Marcadores	Adjacentes ²	LOD	NDM		PSB		
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	
<i>Multi3@A1_42,00</i>	ss107913925	ss107913319	8,51	-1,442*** -0,063	6,18	<i>sur</i>		
<i>Multi3@C1-1_23,00</i>	ss107913694	Satt194	18,06	2,547*** 0,451	18,74	1,624** 0,29	4,94	
<i>Multi3@C1-2_21,00</i>	ss107917456	ss107929365	5,87	1,258*** -0,883	5,09	1,482** -1,565	4,58	
<i>Multi3@C2-1_90,48</i>	ss107917577	ss107917577	5,99	1,155*** 0,247	4,07	1,813** 0,689	6,49	
<i>Multi3@D2-2_41,00</i>	ss107914703	ss107921284	4,80	-1,164*** -0,393	3,92	-1,327*** 0,131	3,30	
<i>Multi3@F-1_55,00</i>	ss107912827	Satt586	6,62	0,313 2,389***	5,88	0,44 1,747**	2,04	
<i>Multi3@M_14,88</i>	ss107912777	ss107912777	16,11	-2,317*** 0,281	15,96	-1,990*** -0,314	7,62	
<i>R</i>²					59,84		28,97	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 17. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-4*: **APM**: altura da planta na maturidade; **NVP**: número de vagens por planta.

MT-MIM: <i>Multi-4</i>							
QTL ¹	Marcadores	Adjacentes ²	LOD	APM		NVP	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi4@C2-1_100,33</i>	ss107920120	ss107920120	9,22	2,039*** 0,127	8,92	11,721*** 0,517	11,90
<i>Multi4@I-1_40,00</i>	ss107926125	T4	14,47	3,008*** -0,171	19,42	11,479*** 2,964	11,43
<i>Multi4@J_55,00</i>	ss107927403	ss107927503	11,21	2,572*** -0,426	14,98	8,059** 1,29	5,94
R²				43,32		29,27	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* ($*p < 0,05$; $**p < 0,01$; $***p < 0,001$); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 18. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-5*: **PEG**: período de granação; **PCS**: peso de cem semetnes e **PSM**: peso de sementes manchadas. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-5</i>									
QTL ¹	Marcadores	Adjacentes ²	LOD	PEG		PCS		PSM	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi5@C1-1_22,68</i>	ss107913694	ss107913694	8,79	0,838*** 0,411	6,76	<i>sur</i>		<i>sur</i>	
<i>Multi5@D2-2_22,00</i>	ss107912878	ss107929556	8,73	-0,949*** -0,026	7,81	-0,749*** -0,048	11,37	-0,732*** 0,091	7,30
<i>Multi5@E-1_14,00</i>	ss107924423	ss107919601	4,51	-0,698*** -0,437	4,31	-0,396*** -0,238	3,25	<i>sur</i>	
<i>Multi5@E-1_67,00</i>	ss107912757	ss107913532	15,48	<i>sur</i>		0,707*** -0,091	11,28	<i>sur</i>	
<i>Multi5@I-1_40,00</i>	ss107926125	T4	18,85	-1,322*** -0,331	17,52	-0,354** -0,091	2,94	0,478** -0,096	3,60
<i>Multi5@J_54,18</i>	ss107927403	ss107927403	12,21	-0,967*** -0,058	9,91	<i>sur</i>		<i>sur</i>	
R²				46,31		28,84		10,90	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* ($*p < 0,05$; $**p < 0,01$; $***p < 0,001$); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 19. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-6*: **NVP**: número de vagens por planta; **NS**: número de sementes; **PG**: produtividade de grãos e **VA**: valor agrônomico. Os QTLs destacados em negrito não foram identificados na abordagem MIM. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-6</i>											
QTL ¹	Marcadores	Adjacentes ²	LOD	NVP		NS		PG		VA	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi6@C1-1_31,00</i>	ss107913694	Satt194	7,10	9,025*** 4,623	4,84	17,554*** 5,391	5,01	2,114*** 0,457	4,37	<i>sur</i>	
<i>Multi6@C2-1_99,00</i>	ss107912713	ss107920120	12,36	11,582*** 2,917	11,46	22,349*** 7,332	11,69	2,130*** 1.148	6,38	0,047*** 0,018	13,83
<i>Multi6@D2-2_14,13</i>	ss107929993	ss107929993	3,12	<i>sur</i>		<i>sur</i>		-0,774*** 1,402	0,90	<i>sur</i>	
<i>Multi6@D2-2_47,63</i>	ss107921284	ss107921284	6,69	-6,087*** -3,256	3,30	-8,877** -7,503	1,92	-2,200*** -1.389	7,10	<i>sur</i>	
<i>Multi6@D2-2_154,60</i>	ss107914356	ss107914356	4,78	6,440*** 0,693	4,16	12,248*** 2,556	4,12	1,704*** -0,279	4,79	<i>sur</i>	
<i>Multi6@E-1_68,00</i>	ss107912757	ss107913532	14,47	<i>sur</i>		<i>sur</i>		1,689*** -1,148	3,90	<i>sur</i>	
<i>Multi6@I-1_41,00</i>	ss107926125	T4	18,25	11,827*** 3,310	11,78	27,321*** 1,847	17,22	2,148*** 0,119	6,40	0,048*** 0,005	14,47
<i>Multi6@J_54,18</i>	ss107925409	ss107925409	5,49	5,714** 2,004	2,99	11,310*** 0,832	3,21	1,453** 0,656	3,18	0,033*** 0,006	7,33
<i>Multi6@K-1_27,00</i>	A28	Satt001	3,70	<i>sur</i>		<i>sur</i>		0,631*** -0,833**	0,58 0,39	<i>sur</i>	
<i>Multi6@M_73,00</i>	ss107913100	A57	11,95	-6,732*** 2,411	3,93	-14,033*** 2,561	4,68	-2,418*** 1,098	8,36	0,003 -0,027***	2,29
R² 6				42,46		47,85		46,35		37,92	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 20. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-7*: **NVP**: número de vagens por planta, **PSB**: peso de sementes boas e **PSM**: peso de sementes manchadas. Os QTLs destacados em negrito não foram identificados na abordagem MIM. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-7</i>										
QTL ¹	Marcadores	Adjacentes ²	LOD	NVP		PSB		PSM		
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	
<i>Multi7@C2-1_93,46</i>	A63	A63	10,09	10,545*** 4,514	9,16	1,999*** 0,705	7,52			<i>sur</i>
<i>Multi7@D2-2_24,00</i>	ss107912878	ss107929556	8,18	-5,941** 0,434	2,64	-1,682*** 0,315	4,83	-0,729*** -0,07		7,22
<i>Multi7@E-1_43,97</i>	ss107918043	ss107918043	5,05	1,519 3,181**	0,44	-1,050*** -0,695	2,17			<i>sur</i>
<i>Multi7@E-1_67,00</i>	ss107912757	ss107913532	17,59	<i>sur</i>		2,455*** -0,083	11,47			<i>sur</i>
<i>Multi7@F-2_114,00</i>	ss107916098	ss107915624	4,74	6,162*** 3,205	3,60	0,984** -0,006	2,10			<i>sur</i>
<i>Multi7@I-1_70,00</i>	Satt239	A24	9,70	13,371*** 3,631*	11,57 0,27	1,646*** 0,093	4,01	0,536** -0,109		3,38
<i>Multi7@M_17,00</i>	ss107913049	Satt540	5,84	-6,510** 1,822	3,23	-1,874*** -0,028	6,12	-0,497** 0,406		3,43
<i>Multi7@N_64,00</i>	ss107913993	ss107913261	6,76	-3,281 3,408**	0,40	1,198*** 1,452	2,51			<i>sur</i>
<i>R²</i>				31,31		40,73		14,03		

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 21. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-8*: **IPDV**: índice percentual de dano nas vagens; **NS**: número de sementes; **PSB**: peso de sementes boas e **PG**: produtividade de grãos. Os QTLs destacados em negrito não foram identificados na abordagem MIM. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-8</i>											
QTL ¹	Marcadores	Adjacentes ²	LOD	IPDV		NS		PSB		PG	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi8@B1-2_30,00</i>	ss107913659	ss107919849	6,15	0,006 -0,028***	5,11	-5,904 17,761***	2,39	-0,646 1,648**	2,18	-0,657 2,324***	3,12
<i>Multi8@C2-1_98,00</i>	ss107912713	ss107920120	12,97	-0,015*** -0,017**	3,68	23,508*** 7,630	13,19	2,059*** 0,649	8,44	2,410*** 1,017	8,33
<i>Multi8@D2-2_44,00</i>	ss107914703	ss107921284	10,71	<i>sur</i>		<i>sur</i>		-0,892*** -0,39	1,47	-1,399*** -0,621	2,60
<i>Multi8@D2-2_184,32</i>	Satt386	Satt386	3,87	<i>sur</i>		3,895 -8,714***	0,93	<i>sur</i>		<i>sur</i>	
<i>Multi8@E-1_66,00</i>	ss107912757	ss107913532	16,94	-0,14*** 0,001	3,10	<i>sur</i>		1,505*** -0,422	4,39	1,678*** -0,881	3,93
<i>Multi8@I-1_41,00</i>	ss107926125	T4	21,48	-0,027*** 0,019***	11,96	27,011*** 1,790	16,84	1,653** 0,217	5,26	2,075*** 0,052	5,97
<i>Multi8@J_55,00</i>	ss107927403	ss107927503	10,14	-0,020*** 0,008	7,16	16,180*** 2,860	6,56	1,467** 0,476***	4,50	2,000*** 0,701	6,02
<i>Multi8@M_15,00</i>	ss107912777	ss107913049	11,88	0,020*** 0,006	5,77	-9,928** -2,244	2,26	-1,748*** -0,577	5,84	-2,246*** -0,278	6,94
<i>Multi8@N_88,29</i>	ss107918394	ss107918394	8,43	-0,01 0,025***	5,19	<i>sur</i>		0,867*** 0,252	1,62	0,652*** 0,248	0,66
R²				46,87		42,17		33,93		37,57	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

Tabela 22. QTLs identificados após eliminação *backward* dos efeitos marginais não significativos do modelo MT-MIM para as características do grupo *Multi-8*: **IPDV**: índice percentual de dano nas vagens; **NS**: número de sementes; **PSB**: peso de sementes boas e **PG**: produtividade de grãos. Os QTLs destacados em negrito não foram identificados na abordagem MIM. QTLs com efeitos destacados em cinza correspondem a QTLs novos, ainda não descritos na literatura.

MT-MIM: <i>Multi-9</i>											
QTL ¹	Marcadores	Adjacentes ²	LOD	PG		IPDV		PCS		PEG	
				Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵	Aditivo ³ Domin. ⁴	%Var ⁵
<i>Multi9@C1-1_22,68</i>	ss107913694	ss107913694	9,09	1,072* 0,505	1,57	<i>sur</i>		<i>sur</i>		0,802*** 0,353	6,18
<i>Multi9@C2-1_100,00</i>	ss107912713	ss107920120	7,38	1,887** 1,014	5,04	<i>sur</i>		<i>sur</i>		-0,206 -0,463**	1,09
<i>Multi9@D2-2_20,00</i>	ss107912878	ss107929556	10,11	-2,129*** 1,099	5,76	<i>sur</i>		-0,795*** 0,112	13,20	-0,947*** -0,049	7,99
<i>Multi9@D2-2_184,00</i>	ss107914356	Satt386	2,88	0,101 -1,875***	2,52	<i>sur</i>		<i>sur</i>		<i>sur</i>	
<i>Multi9@E-1_64,00</i>	ss107919643	ss107912757	6,59	<i>sur</i>		<i>sur</i>		0,453*** -0,147	4,73	<i>sur</i>	
<i>Multi9@E-2_65,82</i>	A51	A51	4,20	0,851** 0,468	1,06	<i>sur</i>		0,353*** 0,239	3,02	<i>sur</i>	
<i>Multi9@F-2_42,00</i>	ss107917362	ss107913720	4,30	<i>sur</i>		-0,014** -0,004	2,66	-0,239 -0,334**	1,21	<i>sur</i>	
<i>Multi9@I-1_40,00</i>	ss107926125	T4	23,61	1,966** -0,043	5,52	-0,026*** 0,011	11,42	-0,357** -0,167	3,00	-1,342*** -0,507	18,05
<i>Multi9@J_54,18</i>	ss107927403	ss107927403	15,79	1,395** 0,869	2,93	-0,019*** 0,012	5,98	<i>sur</i>		-0,927*** 0,100	9,09
<i>Multi9@M_72,73</i>	ss107913100	ss107913100	7,18	-1,733*** 0,258	4,33	0,020*** 0,002	6,36	-0,290*** 0,151	2,00	<i>sur</i>	
<i>Multi9@N_88,29</i>	ss107918394	ss107918394	6,07	<i>sur</i>		-0,009 0,025***	5,08	0,209*** -0,045	1,12	<i>sur</i>	
R²				28,73		31,50		28,28		42,40	

¹*Multi@GL_posição(cM)*; ²Marcadores adjacentes ao QTL identificado; ³Efeito aditivo medido como desvios em relação à média (valores positivos indicam que a cultivar IAC-100 possui o alelo que aumenta a característica); ⁴Efeito de dominância; Significância dos efeitos baseada em *p*-valor estimado por reamostragem baseada na estatística *score* (**p*<0,05; ***p*<0,01; ****p*<0,001); ⁵Razão entre a variância do efeito do QTL sobre a variância fenotípica, vezes 100; *sur*: efeito não relacionado à regressão (*seemingly unrelated regression*).

As características avaliadas neste trabalho são frequentemente utilizadas em estudos de mapeamento de QTLs, em soja, a partir de diferentes *backgrounds* genéticos, utilizando-se, geralmente, a abordagem de mapeamento CIM ou mapeamento associativo. Alguns trabalhos recentes foram publicados envolvendo o mapeamento de características associadas ao florescimento, à maturidade e à altura de plantas (Zhang et al. 2010; Niu et al. 2013; Zuo et al. 2013; Takeshima et al. 2016). Características relacionadas à produtividade de grãos como número, tamanho e peso de sementes, número de vagens e peso de cem sementes também representam variáveis de grande interesse na identificação de QTLs, sendo frequentemente reportados estudos de mapeamento envolvendo essas características (Han et al. 2012; Niu et al. 2013; Yang et al. 2013; Hu et al. 2014; Kato et al. 2014; Rodrigues et al. 2016). O Soybase também possui um banco de QTLs, onde é possível verificar os QTLs identificados em cada GL, por diferentes instituições de pesquisa de soja. Assim, uma busca foi realizada para verificar a presença de QTLs afetando às variáveis analisadas, em posição genômica próxima à identificada no presente trabalho. É importante ressaltar que as informações depositadas no Soybase são provenientes do mapeamento de populações ou RILs (*Recombinant Inbred Lines*) com diferentes *backgrounds* genéticos. Além disso, os QTLs são mapeados com base em mapas genéticos obtidos a partir da genotipagem de populações de tamanhos distintos, utilizando-se diferentes marcadores moleculares. Outra questão relevante é a forma de condução dos experimentos que também é bastante distinta entre os trabalhos. Além disso, as estratégias utilizadas para o mapeamento de QTLs foram sendo modificadas e diferem entre os estudos, interferindo diretamente no poder de detecção dos QTLs, nos valores da percentagem da variação fenotípica explicada pelos mesmos e nas estimativas dos efeitos genéticos dos mesmos. No entanto, como já mencionado, o objetivo principal dessa busca constituiu-se em verificar a presença de QTLs já descritos para essas variáveis e não contrastar as estimativas obtidas. A presença de um QTL descrito em uma mesma região do QTL identificado neste estudo é uma evidência importante e sugere que o mesmo seja considerado para confirmação e validação em outras populações de mapeamento. Por outro lado, considerando a região que está sendo feita a busca por QTLs, o fato de não encontrar QTLs descritos no Soybase sugere que QTLs ainda não descritos na literatura podem ter sido identificados para essa população em estudo. Portanto, essa comparação com o banco de dados de QTLs do Soybase vai permitir identificar os QTLs que devem ser melhor investigados para confirmação em outras populações com diferentes *backgrounds* genéticos e em diferentes ambientes. Atualmente, o banco de dados da soja apresenta um grande número de QTLs descritos para a maioria das características analisadas no presente estudo. Inclusive, muitos dos QTLs descritos para uma mesma característica apresentam-se no mesmo GL, em uma posição genética muito próxima, ou até mesmo com sobreposição dos intervalos de confiança. Guo et al. (2006) reportaram que muitos QTLs descritos no Soybase se encontram repetidos no banco de dados. Considerando todas as questões abordadas acima, não é possível afirmar, com certeza, o número exato de QTLs já identificados para cada uma das características.

Como os QTLs identificados são referentes à essa população em estudo, é interessante verificar a presença de QTLs já descritos, na mesma região genômica, porém em outros *backgrounds* genéticos. No caso de regiões em que não há QTLs descritos na literatura, pode-se pensar na hipótese de que talvez os mesmos ainda não haviam sido identificados pelo fato dos estudos de mapeamento terem sido realizados utilizando-se somente abordagens de mapeamento que consideram uma única característica. Como foi constatado neste estudo, alguns QTLs só foram identificados pela abordagem multivariada, o que seria uma possível explicação, considerando a grande quantidade de estudos de mapeamento e de QTLs já descritos nos 20 GLs da soja.

A análise multivariada do grupo *Multi-1* revelou QTLs, associados às características NDF, NDM, APM, NS e PG, em oito GLs da soja, C1, C2, D2, E, F, I, J e M (Tabela 5). No GL C1, foram identificados QTLs, descritos no Soybase, próximos aos marcadores flanqueando o QTL revelado com a abordagem multivariada, afetando as variáveis

altura da planta, maturidade e produtividade de grãos (Palomeque et al. 2009a; Kim et al. 2012). O QTL associado à variável PG, identificado por Palomeque et al. (2009a), está associado ao mesmo marcador (Satt194) identificado flanqueando o QTL neste estudo, o que pode ser uma evidência de que o mesmo QTL descrito na literatura foi identificado nesta população F_{2,3}. Além disso, o autor relata que o QTL foi identificado em diferentes ambientes, o que é uma evidência a favor da estabilidade desse QTL. Para a variável NS, os QTLs descritos encontram-se em posições muito distantes (>50 cM) dos marcadores flanqueando o QTL identificado neste estudo. E para a variável NDF, nenhum QTL foi descrito neste GL. Portanto, há evidências de que, no presente trabalho, foram identificadas regiões genômicas, ainda não descritas na literatura, afetando as variáveis NDF e NS no GL C1. No GL C2, diversos QTLs foram descritos, próximos ao QTL identificado neste estudo, afetando as variáveis NDF, maturidade, altura de planta, número de sementes e produtividade de grãos (Keim et al. 1990; Mansur et al. 1996; Orf et al. 1999; Specht et al. 2001; Tasma et al. 2001; Kabelka et al. 2004; Wang et al. 2004; Zhang et al. 2004; Funatsuki et al. 2005; Reinprecht et al. 2006; Chen et al. 2007; Gai et al. 2007; Li et al. 2008; Du et al. 2009; Palomeque et al. 2009b; Liu et al. 2011; Rossi et al. 2013; Sun et al. 2013). Os QTLs descritos no Soybase para todas essas características mencionadas encontram-se muito próximos, inclusive com sobreposição dos intervalos de confiança. Portanto, os resultados obtidos sugerem que as regiões identificadas neste trabalho sejam melhor investigadas. No GL D2 foram identificados QTLs afetando maturidade, produtividade de grãos e altura de planta (Hwang 2010; Kazi et al. 2010; Kim et al. 2012; Qi et al. 2014a). QTLs afetando as variáveis maturidade e altura de planta também foram descritos por Hwang (2010), utilizando a abordagem MIM, no GL D2. Ambos foram revelados na mesma posição, associados ao marcador ss107917994, na posição 87,84 cM do mapa da soja (ConsensusMap 4.0). Já o QTL descrito neste estudo encontra-se próximo aos marcadores ss197914703 e ss107921284, nas posições 59,92 cM e 64,72 cM, respectivamente. Para as outras variáveis, maturidade, altura de planta e PG, foram identificados QTLs explicando essas características nas mesmas regiões identificadas neste estudo no GL E (Wang et al. 2004; Zhang et al. 2004; Panthee et al. 2007; Liu et al. 2011; Wang et al. 2011), no GL F (Specht et al. 2001; Reinprecht et al. 2006; Sun et al. 2006; Alcivar et al. 2007; Gai et al. 2007; Josie et al. 2007; Bachlava et al. 2009; Hwang 2010) e no GL J (Tasma et al. 2001; Sun et al. 2006; Li et al. 2008; Eskandari et al. 2013). Hwang (2010) também identificou um QTL afetando a variável maturidade no GL E, utilizando a abordagem MIM, próximo ao marcador ss107928260, na posição 71,42 cM no mapa da soja. Já o QTL identificado neste estudo, afetando NDM, no GL E, se encontra entre os marcadores ss107912757 e ss107913532, nas posições 54,94 cM e 57,24 cM, respectivamente. Também foram identificados QTLs para as variáveis NDF, maturidade, altura da planta e PG no GL I (Sebolt et al. 2000; Chapman et al. 2003; Chung et al. 2003; Nichols et al. 2006; Bachlava et al. 2009; Kuroda et al. 2013; Hu et al. 2014; Qi et al. 2014a). Os QTLs identificados no GL I, por Hu et al. (2014), afetam as variáveis NDF e NDM, e ambos estão associados ao marcador Satt614. Esse marcador se encontra localizado entre os marcadores que foram associados ao QTL identificado neste estudo, Satt194 e ss107913694, de acordo com o mapa da soja (ConsensusMap 4.0). Portanto, há evidências de que o QTL identificado no GL I, afetando as variáveis NDM e NDF, corresponde ao mesmo QTL identificado por Hu et al. (2014) nesse mesmo GL. Nos GLs I e J, nenhum QTL foi descrito para a variável NS. Portanto, os resultados sugerem que duas regiões ainda não descritas na literatura foram reveladas afetando essa variável, nos GLs I e J, considerando a população em estudo. No GL M, QTLs afetando as variáveis NDF, maturidade, altura de planta, NS e PG também foram identificados (Mansur et al. 1993; Lark et al. 1994; Lark et al. 1995; Mansur et al. 1996; Orf et al. 1999; Specht et al. 2001; Wang et al. 2004; Zhang et al. 2004; Hwang 2010). Também foi encontrado um QTL descrito por Hwang (2010), para a variável maturidade no GL M, na posição 18,67 cM, associado ao marcador Sat_316. O autor comparou as duas abordagens, CIM e MIM, para a mesma população estudada, e o QTL no GL M só foi revelado com o uso da abordagem MIM. O QTL identificado no

presente estudo, no mesmo GL, encontra-se na posição 31,48 cM, associado ao marcador ss107912777. Em resumo, a análise do *Multi-1* permitiu a identificação de regiões, ainda não descritas na literatura, afetando a variável NS, nos GLs C1, I e J, e afetando a variável NDF no GL C1.

A análise multivariada do grupo *Multi-2* revelou QTLs associados às características APF, APM e VA em seis GLs da soja (C1, C2, D2, I, J e M) (Tabela 6). Comparando-se cada GL em que foram identificados QTLs pela abordagem MT-MIM, foi possível verificar que todos apresentaram QTLs descritos no banco de dados de QTLs do Soybase para a variável altura da planta, sendo observada uma sobreposição de QTLs identificados em algumas posições. Foram identificados QTLs no Soybase associados à característica altura de planta no GL C1 (Lee et al. 1996; Kim et al. 2012), no GL C2 (Specht et al. 2001; Wang et al. 2004; Zhang et al. 2004; Reinprecht et al. 2006; Gai et al. 2007; Palomeque et al. 2009b; Liu et al. 2011; Rossi et al. 2013; Lee et al. 2015), no GL D2 (Kim et al. 2012), no GL I (Sebolt et al. 2000; Chapman et al. 2003), no GL J (Sun et al. 2006) e no GL M (Specht et al. 2001; Wang et al. 2004; Zhang et al. 2004; Pathan et al. 2013; Yao et al. 2015). Nenhum estudo de mapeamento de QTL para a variável valor agrônômico foi encontrado na literatura. No entanto, como já descrito neste trabalho, essa característica é bastante complexa e sua avaliação envolve diferentes variáveis, como altura de planta, número de sementes, número de vagens, produtividade de grãos, acamamento, retenção foliar e reação à pragas e doenças. Para comparação com os resultados obtidos no presente trabalho, uma busca por QTLs, associados à essas características, foi realizada na região próxima aos marcadores flanqueando os QTLs identificados. No GL C2, foram identificados QTLs afetando as características número de vagens por planta, número de sementes, quantidade de nódulos presentes na raiz e comprimento do internó da haste (Mansur et al. 1996; Zhang et al. 2004; Sun et al. 2006; Josie et al. 2007; Palomeque et al. 2009b; Hwang 2010; Liu et al. 2011). Já em relação à resistência a insetos, foram identificados QTLs conferindo resistência das plantas aos afídeos e à lagarta *Helicoverpa zea* (Rector et al. 1999; Terry et al. 2000; Zhang et al. 2013). No GL I, foram identificados QTLs associados ao número de vagens por planta, altura da planta, quantidade de nódulos presentes na raiz e comprimento do internó da haste (Sebolt et al. 2000; Chapman et al. 2003; Josie et al. 2007; Santos et al. 2013; Yao et al. 2015). Também foram identificados QTLs associados à resistência da soja ao nematoide de cisto *H. glycines* (Qiu et al. 1999; Winter et al. 2007), a *Phytophthora sojae* (Tucker et al. 2010; Pathan et al. 2013), a *Sclerotinia sclerotiorum* (Arabana et al. 2001; Huynh et al. 2010) e à podridão vermelha da raiz (*Fusarium solani*) (Iqbal et al. 2001; Swaminathan et al. 2016). No GL J, foram identificados QTLs, descritos no Soybase, associados ao número de vagens por planta, altura da planta e deiscência da vagem (Funatsuki et al. 2006; Sun et al. 2006; Kang et al. 2009; Yamada et al. 2009; Zhang et al. 2010). Os QTLs descritos nessa região associados às pragas e doenças afetam a resistência da soja ao nematoide de cisto *Heterodera glycines* (Glover et al. 2004; Guo et al. 2006; Chang et al. 2011; Ferreira et al. 2011; Jiao et al. 2015), a *Phytophthora sojae*, (Weng et al. 2007; Tucker et al. 2010), ao vírus do mosaico comum (VMCS) (Yan et al. 2015) e à podridão parda-da-haste (*Phialophora gregata*) (Lewers et al. 1999; Patzoldt et al. 2005a; Patzoldt et al. 2005b). Em resumo, a análise do *Multi-2*, permitiu a identificação de regiões genômicas, ainda não descritas na literatura, afetando a variável VA, nos GLs C2, I e J. A identificação de QTLs afetando outras variáveis que interferem diretamente na variável VA é uma evidência de que essa região pode apresentar um QTL para a variável valor agrônômico, como identificado neste estudo.

A análise multivariada do grupo *Multi-3* identificou sete regiões genômicas associadas às características NDM e PSB, distribuídas em seis GLs da soja: A1, C1 (2 regiões), C2, D2, F e M (Tabela 7). Para a variável NDM, foram identificados QTLs, associados à essa característica, descritos no banco de dados do Soybase, em todos os GLs citados acima, exceto para o GL F. Portanto, os resultados sugerem que o QTL identificado no presente estudo, afetando a variável NDM, no GL F, corresponde a um QTL ainda não descrito na literatura. Foram identificados

QTLs associados à maturidade no banco de dados do Soybase, no GL C2 (Wang et al. 2004; Rossi et al. 2013; Sun et al. 2013; Yao et al. 2015), um QTL no GL D2 (Kim et al. 2012), e QTLs no GL M (Specht et al. 2001; Wang et al. 2004; Zhang et al. 2004). No GL F, apesar de já terem sido identificadas regiões associadas à variável NDM, os QTLs descritos encontram-se à uma distância superior a 60 cM do marcador identificado flanqueando o marcador associado ao QTL deste estudo. Dessa forma, o QTL identificado no presente trabalho corresponde a um QTL que ainda não havia sido descrito na literatura. No GL C1, o presente estudo identificou QTLs afetando a variável NDM nos dois subgrupos do mapa de ligação obtido para a população em estudo, denominados C1-1 e C1-2. Comparando-se a posição dos marcadores flanqueando essas duas regiões identificadas neste trabalho, com suas respectivas posições no mapa genético da soja (ConsensusMap 4.0), foi possível verificar uma distância aproximada de 35 cM entre os QTLs identificados. Dessa forma, as duas regiões identificadas pela abordagem MT-MIM no presente estudo foram consideradas distintas. Para ambas as regiões há um QTL descrito no banco de dados de QTLs do Soybase, associado à variável NDM, tanto para a região identificada no subgrupo C1-1 deste estudo (Keim et al. 1990; Orf et al. 1999; Kim et al. 2012), quanto para a região revelada no subgrupo C1-2 (Orf et al. 1999; Zhang et al. 2004; Rossi et al. 2013; Sun et al. 2013). Não foi identificado no banco de dados de QTLs do Soybase nenhum QTL descrito para a variável peso de sementes boas (PSB). Portanto, pode-se sugerir que os QTLs identificados no presente estudo para PSB correspondem aos primeiros QTLs relatados na literatura para essa variável. Apenas para comparação com o banco de dados de QTLs do Soybase foi feita uma busca por QTLs associados à produtividade de grãos, que corresponde à variável que mais se aproxima da variável PSB no banco de dados do Soybase. Além disso, a correlação entre essas duas variáveis é positiva e foi o maior valor observado no presente estudo ($r=0,96^{***}$) (Figura 2). Os QTLs já descritos na literatura associados à produtividade de grãos foram identificados no GL C1 para a região identificada no subgrupo C1-1 deste trabalho (Palomeque et al. 2009a; Kim et al. 2012), e também para a região identificada no subgrupo C1-2 (Yuan et al. 2002; Palomeque et al. 2009a; Rossi et al. 2013), no GL C2 (Zhang et al. 2004; Reinprecht et al. 2006; Gai et al. 2007; Du et al. 2009; Palomeque et al. 2009a; Rossi et al. 2013; Yao et al. 2015), no GL D2 (Kazi et al. 2010; Qi et al. 2014a), no GL F (Reyna and Sneller 2001; Rossi et al. 2013) e no GL M (Wang et al. 2004; Zhang et al. 2004; Du et al. 2009; Wang et al. 2014). Dessa forma, a análise do *Multi-3* permitiu a identificação de novas regiões ainda não descritas na literatura, afetando a variável PSB nos GLs C1 (2), C2, D2, F e M, e uma região afetando a variável NDM no GL F.

A análise multivariada do grupo *Multi-4* identificou três regiões genômicas associadas às características APM e NVP, distribuídas em três GLs da soja: C2, I e J (Tabela 8). Para todos os GLs nos quais foram revelados QTLs associados a essas duas variáveis, no presente estudo, foram identificados QTLs já descritos no Soybase. No GL C2 os QTLs já descritos no Soybase encontram-se associados à altura de planta (Orf et al. 1999; Specht et al. 2001; Kabelka et al. 2004; Wang et al. 2004; Zhang et al. 2004; Reinprecht et al. 2006; Gai et al. 2007; Palomeque et al. 2009b; Liu et al. 2011; Rossi et al. 2013; Lee et al. 2015) e NVP (Sun et al. 2006; Palomeque et al. 2009b; Liu et al. 2011). No GL I, os QTLs encontrados afetam a variável altura de planta (Sebolt et al. 2000; Chapman et al. 2003) e para NVP somente um QTL foi encontrado descrito no Soybase na mesma região identificada no presente estudo (Li et al. 2010). Da mesma forma, no GL J foram encontrados QTLs já descritos para altura de planta (Funatsuki et al. 2006; Sun et al. 2006) e para NVP somente um QTL foi identificado (Zhang et al. 2010). Assim, na análise do *Multi-4*, todas as regiões identificadas neste estudo apresentam QTLs já descritos no Soybase, afetando as variáveis desse grupo.

A análise multivariada do grupo *Multi-5* identificou seis regiões genômicas associadas às características PEG, PCS e PSM, distribuídas por cinco GLs da soja: C1, E (2 QTLs), J, D2 e I (Tabela 9). Para a variável PEG, nenhum QTL foi encontrado descrito no Soybase, considerando os GLs em que foram identificados QTLs neste estudo. Não

foi identificado, no banco de dados de QTLs do Soybase, nenhum QTL descrito para a variável peso de sementes manchadas (PSM). Da mesma forma, nenhum outro estudo foi encontrado na literatura identificando QTLs para essa variável. Portanto, pode-se sugerir que os dois QTLs identificados no presente estudo para PSM correspondem aos primeiros QTLs relatados na literatura para essa variável. Para a variável PCS, foram identificados QTLs descritos no GL E, para as duas regiões identificadas no presente trabalho (GL E-1 e GL E-2), afetando essa variável (Lee et al. 2001; Liu et al. 2011). Para os demais GLs, também foram encontrados QTLs associados à PCS, no GL D2 (Teng et al. 2009; Han et al. 2012; Kuroda et al. 2013; Kato et al. 2014) e no GL I (Sebolt et al. 2000; Csanádi et al. 2001; Reinprecht et al. 2006; Liu et al. 2011; Han et al. 2012; Kato et al. 2014). O QTL identificado por Kato et al. (2014), afetando a variável PCS, no GL D2, foi revelado em duas populações de RILs, com *backgrounds* genéticos distintos. Além disso, o QTL se mostrou consistente em três ambientes diferentes, o que demonstra a estabilidade do mesmo. De acordo com o mapa da soja (ConsensusMap 4.0), o marcador Satt372, que se encontra ligado ao QTL identificado por Kato et al. (2014), encontra-se a 6 cM de distância do marcador SNP ss107912878, o marcador ligado ao QTL identificado neste estudo. Os resultados sugerem que talvez haja sobreposição no intervalo de confiança dos mesmos. Esse QTL do GL D2 foi o que apresentou o maior valor de R^2 para a variável PCS, na população $F_{2:3}$ deste estudo, sendo responsável por 13,20% da variação fenotípica observada. Dentre os QTLs identificados por Kato et al. (2014), para essa variável, o QTL do GL D2 também foi o que apresentou os maiores valores de R^2 , explicando entre 9,4 a 20,9% da variância fenotípica observada para a característica, dependendo da população e do local de avaliação. Considerando as evidências apresentadas acima, o QTL identificado no GL D2, neste trabalho, deve ser considerado em futuras investigações, como um importante QTL afetando a variável PCS. Em resumo, a análise do *Multi-5* permitiu a identificação de novas regiões ainda não descritas na literatura, afetando as variáveis PEG nos GLs C1, D2, E, I e J e para a variável PSM nos GLs D2 e I.

A análise multivariada do grupo *Multi-6* permitiu a identificação de 10 regiões genômicas, com efeito em pelo menos uma das características avaliadas NVP, NS, PG e VA, distribuídas por oito GLs da soja: C1, C2, D2 (3 QTLs), E, I, J, K e M) (Tabela 10). Para a variável NS, nenhum QTL foi encontrado descrito no banco de dados de QTLs do Soybase nos GLs C1, D2 (considerando os dois QTLs identificados), I e J. Portanto, a população em estudo permitiu a identificação de seis QTLs, afetando a variável NS, que não haviam sido descritos na literatura. Para os GLs C2 e M, foi identificada uma região em cada GL, afetando a variável NS (Mansur et al. 1996). Para a variável NVP, somente para o GL M não foi encontrado nenhum QTL descrito no Soybase, na região próxima ao marcador SNP (ss107913100) flanqueando o QTL identificado neste trabalho. Foram encontrados QTLs, descritos no Soybase, associados às variáveis NVP e PG no GL C1 (Palomeque et al. 2009a; Kim et al. 2012), no GL C2 (Zhang et al. 2004; Reinprecht et al. 2006; Sun et al. 2006; Gai et al. 2007; Du et al. 2009; Palomeque et al. 2009a; Liu et al. 2011; Rossi et al. 2013; Yao et al. 2015), no GL D2 (Kabelka et al. 2004; Kazi et al. 2010; Hu et al. 2013; Kuroda et al. 2013), no GL I (Sebolt et al. 2000; Chung et al. 2003; Nichols et al. 2006; Reinprecht et al. 2006; Wang et al. 2014) e no GL J (Zhang et al. 2010; Eskandari et al. 2013; Wang et al. 2014; Fox et al. 2015). Para PG, também foram identificados QTLs descritos no GL E (Wang et al. 2004; Reinprecht et al. 2006; Zhang et al. 2010; Rossi et al. 2013; Wang et al. 2014), no GL K (Yuan et al. 2002; Kabelka et al. 2004; Wang et al. 2004; Vieira et al. 2006; Eskandari et al. 2013; Rossi et al. 2013) e no GL M (Wang et al. 2004; Zhang et al. 2004; Du et al. 2009; Wang et al. 2014). Essa região identificada no GL M encontra-se localizada a aproximadamente 7 cM do principal gene que confere resistência a insetos coleópteros e lepidópteros (Rector et al. 1998; Rector et al. 1999; Rector et al. 2000), sugerindo ser uma importante região responsável pela resistência a insetos. O marcador Satt389, associado ao QTL identificado por Hu et al. (2014), no GL D2, afetando a variável NVP, encontra-se a menos de 5 cM de distância do marcador associado ao QTL identificado

neste estudo, ss107921284. Os resultados sugerem que o mesmo QTL identificado por Hu et al. (2014), para a variável NVP, tenha sido revelado nesta população em estudo. A população em estudo permitiu a identificação de quatro novos QTLs para a variável VA. Para comparação com os resultados obtidos no presente trabalho, uma busca por QTLs afetando características que podem estar associadas à variável valor agrônômico foi realizada na região próxima aos marcadores flanqueando os QTLs identificados. No GL C2, foram identificados QTLs, na mesma região identificada neste estudo, afetando altura da planta (Orf et al. 1999; Specht et al. 2001; Kabelka et al. 2004; Wang et al. 2004; Zhang et al. 2004; Reinprecht et al. 2006; Gai et al. 2007; Palomeque et al. 2009a; Liu et al. 2011; Rossi et al. 2013; Lee et al. 2015), acamamento (Mansur et al. 1993; Mansur et al. 1996; Specht et al. 2001; Zhang et al. 2004; Rossi et al. 2013), número de nódulos na raiz (Wu et al. 2004; Gai et al. 2007), comprimento do internó da haste (Alcivar et al. 2007) e QTLs conferindo resistência das plantas aos afídeos e à lagarta *H. zea* (Rector et al. 1999; Terry et al. 2000; Zhang et al. 2013). No GL I, foram encontrados QTLs descritos afetando altura da planta (Sebolt et al. 2000; Chapman et al. 2003), número e massa de nódulos na raiz (Santos et al. 2013; Yao et al. 2015) e comprimento do interno na haste (Josie et al. 2007). Já em relação à resistência a insetos, foram identificados QTLs conferindo resistência das plantas ao nematoide de cisto *H. glycines* (Qiu et al. 1999; Winter et al. 2007), a *Phytophthora sojae*, (Lee et al. 2013), a *Sclerotinia sclerotiorum* (Arabana et al. 2001; Huynh et al. 2010) e à podridão vermelha da raiz (*Fusarium solani*) (Iqbal et al. 2001; Swaminathan et al. 2016). No GL J, foram identificados QTLs, descritos no Soybase, associados ao número de vagens por planta, altura da planta e deiscência da vagem (Funatsuki et al. 2006; Sun et al. 2006; Kang et al. 2009; Yamada et al. 2009; Zhang et al. 2010). Os QTLs descritos nessa região associados à pragas e doenças afetam a resistência da soja ao nematoide de cisto *H. glycines* (Glover et al. 2004; Guo et al. 2006; Chang et al. 2011; Ferreira et al. 2011; Jiao et al. 2015), a *Phytophthora sojae*, (Weng et al. 2007; Tucker et al. 2010), ao vírus do mosaico comum (VMCS) (Yan et al. 2015) e à podridão parda-da-haste (*Phialophora gregata*) (Lewers et al. 1999; Patzoldt et al. 2005a; Patzoldt et al. 2005b). No GL M, foram identificados QTLs, descritos no Soybase, associados à altura da planta (Wang et al. 2004; Pathan et al. 2013; Yao et al. 2015). Os QTLs descritos nessa região, associados à pragas e doenças, afetam a resistência da soja a afídeos (Zhang et al. 2009), a *Phytophthora sojae* (Lee et al. 2013), à lagarta *H. zea* (Narvel et al. 2001) e a *Sclerotinia sclerotiorum* (Kim and Diers 2000). Resumindo, a análise do *Multi-6*, permitiu a identificação de novas regiões ainda não descritas na literatura, afetando a variável NS nos GLs C1, D2 (2 QTLs), I e J, um QTL afetando a variável NVP no GL M e quatro QTLs afetando a variável VA nos GLs C2, I, J e M.

A análise multivariada do grupo *Multi-7* permitiu a identificação de oito regiões genômicas, com efeito em pelo menos uma das características avaliadas NVP, PSB e PSM, distribuídas por sete GLs da soja: C2, D2, E (2 regiões), F, I, M e N) (Tabela 11). Para a variável NVP, nenhum QTL foi encontrado, descrito no banco de dados de QTLs do Soybase, nos GLs E, F, M e N. Portanto, a população em estudo permitiu a identificação de quatro novas regiões genômicas, afetando NVP, que não haviam sido descritas para essa variável. Para as outras regiões identificadas nesta análise, foram encontrados QTLs descritos no Soybase, afetando a variável NVP, no GL C2 (Sun et al. 2006; Palomeque et al. 2009b; Liu et al. 2011), no GL D2 (Li et al. 2010a) e no GL I (Li et al. 2010a). Como já mencionado anteriormente, não foi identificado no banco de dados de QTLs do Soybase nenhum QTL descrito para as variáveis PSM e PSB. Portanto, pode-se sugerir que os QTLs identificados no presente estudo, afetando as variáveis PSM e PSB, correspondem aos primeiros QTLs relatados na literatura. Para comparação com o banco de dados de QTLs do Soybase, foi feita uma busca por QTLs associados à produtividade de grãos, que corresponde à variável que mais se aproxima da variável PSB no banco de dados do Soybase. Os QTLs já descritos na literatura associados à produtividade de grãos, foram identificados no GL C2 (Zhang et al. 2004; Reinprecht et al. 2006; Gai et al. 2007; Du et al. 2009; Palomeque et al. 2009a; Rossi et al. 2013; Yao et al. 2015), no GL D2 (Kazi et al. 2010; Qi et al. 2014a), no GL E

(Wang et al. 2004; Reinprecht et al. 2006; Rossi et al. 2013; Wang et al. 2014), no GL F (Specht et al. 2001), no GL I (Sebolt et al. 2000; Nichols et al. 2006; Reinprecht et al. 2006; Wang et al. 2014), GL J (Eskandari et al. 2013; Wang et al. 2014; Fox et al. 2015), no GL M (Wang et al. 2004; Zhang et al. 2004; Du et al. 2009; Wang et al. 2014) e no GL N (Kabelka et al. 2004). Em resumo, a análise do *Multi-7* permitiu a identificação de QTLs, ainda não descritos na literatura, afetando a variável PSM nos GLs D2, I e M, a variável PSB nos GLs C2, D2, E (2), F, I, M e N, e a variável NVP nos GLs E, F, M e N.

A análise multivariada do grupo *Multi-8* permitiu a identificação de nove regiões genômicas, com efeito em pelo menos uma das características avaliadas IPDV, NS, PSB e PG, distribuídas por oito GLs da soja: B1, C2, D2 (2 regiões), E, I, J, M e N (Tabela 12). Não foi identificado no banco de dados de QTLs do Soybase nenhum QTL descrito para a variável índice percentual de dano nas vagens (IPDV). Da mesma forma, nenhum outro estudo foi encontrado na literatura descrevendo QTLs para essa variável. Da mesma forma, nenhum QTL para a variável PSB foi encontrado no Soybase e na literatura. Portanto, pode-se sugerir que os QTLs identificados no presente estudo, para as variáveis IPDV e PSB, correspondem aos primeiros QTLs relatados na literatura. Para a variável NS, nenhum QTL foi encontrado descrito no banco de dados de QTLs do Soybase nos GLs D2, I, J e M. Portanto, a população em estudo permitiu a identificação de quatro QTLs, afetando a variável NS, que não haviam sido descritos anteriormente. No GL B1, foi identificado um QTL descrito para PG (Du et al 2009) e dois QTLs associados ao NS (Lian et al 2010). No GL C2, foram encontrados QTLs descritos para PG (Zhang et al. 2004; Reinprecht et al. 2006; Gai et al. 2007; Du et al. 2009; Palomeque et al. 2009a; Rossi et al. 2013; Yao et al. 2015) e um QTL associado ao NS (Mansur et al. 1996). Foram encontrados QTLs descritos associados à variável PG no GL D2 (Kabelka et al. 2004; Kazi et al. 2010; Wang et al. 2014), no GL E (Wang et al. 2004; Reinprecht et al. 2006; Rossi et al. 2013; Wang et al. 2014), no GL I (Sebolt et al. 2000; Chung et al. 2003; Nichols et al. 2006; Reinprecht et al. 2006; Wang et al. 2014), no GL J (Li et al. 2008; Eskandari et al. 2013; Wang et al. 2014; Fox et al. 2015), no GL M (Wang et al. 2004; Zhang et al. 2004) e no GL N (Kabelka et al. 2004). Para comparação com o banco de dados de QTLs do Soybase, foi feita uma busca por QTLs associados à produtividade de grãos, que corresponde à variável que mais se aproxima da variável PSB no banco de dados do Soybase. Os QTLs já descritos na literatura, associados à produtividade de grãos, foram identificados no GL B1 (Du et al. 2009), GL C2 (Zhang et al. 2004; Reinprecht et al. 2006; Gai et al. 2007; Du et al. 2009; Palomeque et al. 2009a; Rossi et al. 2013; Yao et al. 2015), GL D2 (Kazi et al. 2010; Qi et al. 2014a), GL E (Wang et al. 2004; Reinprecht et al. 2006; Rossi et al. 2013; Wang et al. 2014), GL I (Sebolt et al. 2000; Nichols et al. 2006; Reinprecht et al. 2006; Wang et al. 2014), GL J (Eskandari et al. 2013; Wang et al. 2014; Fox et al. 2015), GL M (Wang et al. 2004; Zhang et al. 2004; Du et al. 2009; Wang et al. 2014) e GL N (Kabelka et al. 2004). Resumindo, a análise do *Multi-8* permitiu a identificação de QTLs, ainda não descritos na literatura, afetando a variável PSB e IPDV nos GLs B1, C2, E, I, J, M e N, um QTL para PSB no GL D2, e quatro QTLs para a variável NS nos GLs D2, I, J e M.

A análise multivariada do grupo *Multi-9* permitiu a identificação de 11 regiões genômicas, com efeito em pelo menos uma das características avaliadas PEG, IPDV, PG e PCS, distribuídas por nove GLs da soja: C1, C2, D2 (2 regiões), E (2 regiões), F, I, J, M e N (Tabela 13). Para a variável PEG, nenhum QTL foi encontrado descrito no banco de dados de QTLs do Soybase nos GLs C1, D2, I e J. Portanto, a população em estudo permitiu a identificação de quatro QTLs, que ainda não haviam sido descritos para a variável PEG, nos GLs C1, D2, I e J. Considerando ainda a variável PEG, um QTL foi encontrado descrito no Soybase no GL C2 (Mansur et al. 1993). Para IPDV, nenhum QTL foi encontrado descrito no Soybase. Portanto, essa população em estudo permitiu revelar cinco QTLs afetando a variável IPDV. Para a variável PG, foram encontrados QTLs, descritos no Soybase, no GL C1 (Palomeque et al. 2009a; Kim et al. 2012), no GL C2 (Reinprecht et al. 2006; Gai et al. 2007; Du et al. 2009; Palomeque et al. 2009a;

Rossi et al. 2013; Yao et al. 2015), no GL D2 para as duas regiões identificadas neste trabalho (Kabelka et al. 2004; Kazi et al. 2010; Wang et al. 2014), no GL E (Wang et al. 2004; Reinprecht et al. 2006; Rossi et al. 2013; Wang et al. 2014), no GL I (Sebolt et al. 2000; Chung et al. 2003; Nichols et al. 2006; Reinprecht et al. 2006; Wang et al. 2014), no GL J (Eskandari et al. 2013; Wang et al. 2014; Fox et al. 2015) e no GL M (Wang et al. 2004; Zhang et al. 2004; Du et al. 2009; Wang et al. 2014). Para PCS, foram identificados QTLs já descritos no GL D2 (Teng et al. 2009; Han et al. 2012; Kuroda et al. 2013; Kato et al. 2014), no GL E (Liu et al. 2011), no GL F (Lee et al. 2001; Hyten et al. 2004; Teng et al. 2009; Eskandari et al. 2013; Rossi et al. 2013; Yan et al. 2015), no GL I (Sebolt et al. 2000; Csanádi et al. 2001; Nichols et al. 2006; Reinprecht et al. 2006; Liu et al. 2011; Kato et al. 2014), no GL M (Orf et al. 1999; Han et al. 2012) e no GL N (Chen et al. 2007; Kim et al. 2010). A região identificada no GL M, com QTLs significativos para IPDV, PG e PCS, se encontra localizada à aproximadamente 7 cM do principal gene que confere resistência a insetos coleópteros e lepidópteros (Rector et al. 1998; Rector et al. 1999; Rector et al. 2000), sugerindo ser uma importante região responsável pela resistência a insetos. No GL E, foram identificadas duas posições genômicas, neste trabalho, afetando as variáveis PG e PCS no grupo *Multi-9*, sendo uma em cada subgrupo obtido no mapa da população em estudo (GL E-1 e GL E-2). O marcador flanqueando o QTL identificado no GL E-2 é um marcador AFLP, e não possui sua posição definida no mapa da soja (ConsensusMap 4.0), como os marcadores SNP e SSR. Portanto, não foi possível definir a sua posição correspondente no mapa da soja para realizar a busca por QTLs já descritos. No entanto, os dois marcadores SNP mais próximos (ss107913504 e ss107924165), com distância de aproximadamente 3 cM do marcador AFLP, de acordo com o mapa obtido para a população em estudo, foram utilizados para definir a região de busca por QTLs no Soybase. Foi possível verificar que o marcador ss107924165, utilizado para definir a região de busca por QTLs, encontra-se a menos de 0,2 cM do marcador flanqueando o QTL identificado por Kato et al. (2014), afetando a variável PCS, no GL E. Esse QTL foi revelado em duas populações de RILs, com *backgrounds* distintos, em dois locais de avaliação, explicando entre 4,8 a 13,2 % da variância fenotípica observada, dependendo da população em estudo e do local de avaliação. Os resultados obtidos no presente estudo são consistentes com os resultados obtidos por Kato et al. (2014), sendo que o QTL identificado nesse GL E, para a população F₂₃ em estudo, foi responsável por 5,34% da variância fenotípica observada. Hu et al. (2014), utilizando mapeamento associativo, também identificou um QTL associado à variável PCS, nessa mesma região genômica do GL E. O marcador identificado associado à variável PCS encontra-se à aproximadamente 1,5 cM do marcador ss107924165, que está localizado próximo ao marcador AFLP associado ao QTL identificado neste estudo. Os resultados sugerem que o QTL identificado neste estudo corresponde ao mesmo QTL identificado por Hu et al. (2014) e Kato et al. (2014). O QTL identificado no GL N, afetando a variável PCS, encontra-se próximo ao QTL identificado neste mesmo GL por Hu et al. (2014). Os marcadores flanqueando o QTL identificado neste trabalho se encontram nas posições 49,69 cM (ss107913993) e 60,67 cM (ss107918394) do mapa da soja (ConsensusMap 4.0). O marcador Satt521, associado ao QTL identificado por Hu et al. (2014), encontra-se dentro desse intervalo, na posição 52,41 cM do ConsensusMap 4.0. Os resultados sugerem que o QTL identificado para PCS, no GL N, corresponde ao mesmo QTL descrito por Hu et al. (2014). Resumindo, a análise do *Multi-9* permitiu a identificação de QTLs ainda não descritos na literatura, afetando a variável PEG nos GLs C1, D2, I e J, e afetando a variável IPDV nos GLs F, I, J, M e N.

Cinco regiões genômicas identificadas neste estudo apresentaram marcadores dominantes, AFLP ou TRAP, flanqueando o QTL identificado no presente estudo. No entanto, nas regiões genômicas dos subgrupos K-1 (posição 27 cM), I-1 (posição 40 e 70 cM) e GL M (posição 73 cM), um dos marcadores flanqueadores corresponde a um marcador SNP ou SSR, que possuem posição conhecida no mapa da soja (ConsensusMap 4.0). Já no caso do subgrupo E-2 (posição 65,82 cM), o QTL identificado foi associado somente a um marcador dominante AFLP (A51). Tal fato

demonstra a vantagem do uso de marcadores já descritos no mapa da soja, facilitando a comparação com o banco de dados da soja. Além disso, os marcadores dominantes são menos informativos em relação aos marcadores co-dominantes, na construção de mapas genéticos, não permitindo distinguir o genótipo heterozigoto do homozigoto dominante.

A busca por QTLs no banco de dados do Soybase evidenciou que muitos dos QTLs identificados neste estudo se encontram em regiões genômicas onde há a presença de QTLs já descritos para as características analisadas. Tal resultado possui uma grande relevância, sugerindo que, possivelmente, QTLs já descritos também tenham sido revelados nessa população de mapeamento em estudo. Muitas regiões reveladas apresentaram até o mesmo marcador descrito como ligado ao QTL, o que aumenta essa evidência. No entanto, para nove variáveis analisadas não foram identificados QTLs descritos para determinadas regiões identificadas neste estudo. Essa constatação é uma evidência importante que alguns dos QTLs identificados neste estudo consistem em regiões ainda não descritas na literatura, sendo fundamentais para auxiliar no entendimento da arquitetura genética dessas características. Como a maioria das regiões identificadas pela abordagem MT-MIM, para os nove grupos analisados, concentra-se em algumas posições genômicas, foi feita uma comparação no caso das variáveis presentes em mais de um grupo, para definir o número de QTLs revelados nessa população que ainda não foram descritos na literatura. No total, a população em estudo permitiu revelar 51 QTLs que ainda não foram descritos na literatura, sendo que 30 deles só foram revelados utilizando-se a abordagem multivariada. Esse resultado demonstra a relevância do uso da abordagem multivariada na identificação de QTLs no caso de variáveis correlacionadas, permitindo a identificação de novos QTLs que irão auxiliar na elucidação da base genética das mesmas. Como não foram encontrados QTLs, descritos no Soybase, para as variáveis VA, IPDV, PSM e PSB, todas as regiões identificadas neste estudo, pela abordagem multivariada, foram consideradas como QTLs ainda não descritos na literatura. Dessa forma, foram identificados, nessa população em estudo, 16 QTLs ainda não descritos na literatura, para a variável PSB, localizados nos GLs B1, C1 (2 QTLs), C2, D2 (2 QTLs), E (2 QTLs), F (2 QTLs), I (2 QTLs), J, M e N (2 QTLs). Do total de 16 QTLs revelados para PSB, 13 QTLs não haviam sido identificados na análise de fenótipos individuais, sendo revelados somente pela abordagem MT-MIM (GLs B1, C1 (1), D2 (2), E (2), F (2), I (2), J e N (2)). Esse resultado demonstra a relevância da utilização da abordagem multivariada na identificação de QTLs para a variável PSB. Essa população em estudo permitiu identificar nove QTLs ainda não descritos na literatura, afetando a variável IPDV, localizados nos GLs B1, C2, E, F, I, J, M (2 QTLs) e N. Deste total de nove QTLs revelados para IPDV, cinco deles só foram revelados utilizando a abordagem MT-MIM (GLs C2, E, F, M (1) e N. Para a variável VA, foram identificados, neste trabalho, quatro QTLs ainda não descritos na literatura, localizados nos GLs C2, I, J e M, sendo que o QTL do GL M só foi revelado utilizando a abordagem MT-MIM. Neste estudo, quatro QTLs, ainda não descritos na literatura, foram revelados afetando a variável PSM, localizados nos GLs D2, I (2 QTLs) e M, sendo que nenhum desses QTLs havia sido revelado pela abordagem MIM. Para as demais características que apresentam QTLs descritos no Soybase, foram revelados, neste estudo, seis novos QTLs, para a variável NS, localizados nos GLs C1, D2 (2 QTLs), I, J e M, sendo que os dois QTLs identificados no GL D2 só foram revelados com o uso da abordagem MT-MIM. Para a variável PEG foram revelados cinco QTLs, ainda não descritos na literatura, afetando essa característica, localizados nos GLs C1, D2, E, I e J. Esses QTLs afetando PEG foram revelados em ambas abordagens, tanto na MIM quanto na MT-MIM. Os cinco novos QTLs identificados para a variável NVP localizam-se nos GLs E, F, M (2 QTLs) e N, sendo que todos eles só foram revelados pela abordagem MT-MIM. Um novo QTL foi revelado, nessa população em estudo, para a variável NDF, no GL C1, e NDM, no GL F, sendo que ambos foram identificados pelas duas abordagens, tanto MIM quanto MT-MIM.

A maioria dos 51 novos QTLs identificados neste estudo explica até 10% da variação do fenótipo da característica, para essa população em estudo. No entanto, alguns QTLs de maior magnitude também foram revelados, como é o caso dos QTLs *Multi2@C2-1_99,00* e *Multi2@I-1_38,00*, que explicam 13,10% e 14,73%, respectivamente, da variação fenotípica da variável VA (Tabela 15). Já o QTL *Multi5@I-1_40,00* foi responsável por 15,97% da variação fenotípica da variável PEG (Tabela 18) e o QTL *Multi6@I-1_40,00* explicou 18,64% da variável NS (Tabela 19). Para a variável IPDV, foi revelado o QTL *Multi8@I-1_40,00* explicando 11,98% da variação fenotípica dessa variável (Tabela 21). Como já abordado, a identificação de QTLs de maior magnitude afetando as características estudadas são relevantes para o melhoramento da soja, pois devido ao caráter quantitativo das mesmas, a maioria dos QTLs que controlam essas características apresenta pequeno efeito na manifestação fenotípica. Portanto, os QTLs de maior magnitude identificados podem ser úteis na seleção de indivíduos superiores, contribuindo para um maior ganho de seleção.

No geral, o perfil de LOD do modelo MT-MIM foi consistentemente superior, no sentido de apresentar maiores valores, em relação aos modelos MIM individuais para as variáveis analisadas em cada grupo. No entanto, algumas exceções foram verificadas, sendo que, em alguns GLs, a presença de QTLs somente foi revelada com a abordagem MIM. No total, 15 QTLs foram revelados somente com a abordagem MIM, sendo um QTL para cada uma das variáveis APM (*APM@C1-1_22,68* - Figura 9), NDM (*NDM@I-1_67,00* - Figura 8), PEG (*PEG@A1_41,00* - Figura 10 e Figura 14), PSB (*PSB@C1-1_28,00* - Figura 12 e Figura 13) e IPDV (*IPDV@B1-2_28,00* - Figura 14); dois QTLs para cada uma das variáveis APF (*APF@N_5,00* e *APF@E-1_70,21* - Figura 7), NS (*NS@C1-1_31,00* e *NS@L_132,00* - Figura 13) e PG (*PG@C1-1_27,00* - Figura 13 e *PG@L_120,00* - Figura 13 e Figura 14), e quatro QTLs revelados para PCS (*PCS@A1_24,00* e *PCS@I-2_14,00* - Figura 10 e Figura 14, *PCS@E-2_65,82* e *PCS@M_107,00* - Figura 10). De acordo com uma simulação realizada por Silva et al. (2012b), os modelos MT-MIM podem apresentar menor poder de detecção em regiões com efeito restrito a apenas um subconjunto das características avaliadas. Essa pode ser a razão dos QTLs mencionados acima não terem sido detectados pelo modelo MT-MIM, sendo considerados específicos para o fenótipo da variável para a qual o QTL foi identificado. Os QTLs revelados somente pela abordagem MIM explicaram de 3,05% (*PCS@M_107,00*) a 6,82% (*PSB@C1-1_28,00*) da variação fenotípica dessas variáveis.

Dessa forma, a estratégia adotada neste trabalho, de aliar a abordagem multivariada com a abordagem univariada, foi satisfatória e permitiu a identificação de diferentes regiões contendo QTLs para as variáveis analisadas. Para verificar se os QTLs revelados somente pela abordagem MIM se localizam em região onde já foram descritos QTLs para essas variáveis, uma busca no banco de dados do Soybase foi realizada, seguindo os mesmos critérios descritos anteriormente. No GL A1, na região genômica onde foi identificado um QTL para a variável PEG, neste estudo, não foram encontrados QTLs descritos no Soybase para essa variável. Portanto, há evidências de que um QTL afetando a variável PEG foi identificado, para essa população em estudo, no GL A1. Nesse mesmo GL, foi identificada uma região, no Soybase, contendo QTL para a variável PCS, próxima à região identificada neste estudo (Kato et al. 2014). Como já relatado anteriormente, não foi encontrado nenhum QTL com a descrição da variável IPDV, no Soybase, o que indica que, para essa população em estudo, um QTL no GL B1 foi revelado. No GL C1 uma nova região foi revelada, neste estudo, afetando a variável NS, uma vez que nenhum QTL foi encontrado, no Soybase, associado à essa variável nessa região identificada nesse estudo. Nesse mesmo GL, foi identificada, no Soybase, uma região afetando APM (Kim et al. 2012), e duas regiões afetando PG (Palomeque et al. 2009a; Kim et al. 2012). Como não há um QTL com a descrição de PSB no Soybase, os resultados evidenciam que uma região genômica ainda não descrita na literatura foi identificada no GL C1, afetando essa variável. No GL E, foram identificadas, no Soybase,

duas regiões afetando a variável PCS (Teng et al. 2009; Han et al. 2012) e três regiões associadas à altura da planta (Sun et al. 2006; Chen et al. 2007; Liu et al. 2011). No GL I, foram identificadas duas regiões no Soybase, associadas à variável NDM (Sebolt et al. 2000), e uma região associada à variável PCS (Hyten et al. 2004). O QTL *PCS@I-2_14,00*, revelado pela abordagem MIM, encontra-se localizado próximo ao QTL identificado por Hwang (2010), que também utilizou a abordagem MIM para identificar QTLs associados a variável PCS. O QTL identificado neste estudo encontra-se entre os marcadores ss107926487 e ss1079201441, que se localizam nas posições 105,11 e 105,67 cM, respectivamente, no mapa da soja (CompositeMap 4.0). Já o QTL identificado por Hwang (2010) se encontra na posição 95 cM, próximo ao marcado Satt581. Para o GL L, duas regiões foram identificadas próximas ao QTL identificado neste estudo, afetando a variável PG (Orf et al. 1999; Panthee et al. 2007). E no GL N, próximo à região onde foi identificado o QTL para a variável APF, neste estudo, foram encontradas três regiões, no Soybase, com QTLs já descritos para a variável altura de plantas (Lark et al. 1995; Specht et al. 2001; Kabelka et al. 2004). No caso da variável RF, nenhum QTL foi identificado, neste estudo, utilizando-se a abordagem MIM e essa variável não foi utilizada na abordagem MT-MIM, como já explicado anteriormente. Já para a variável AC, o método MIM permitiu a identificação de duas regiões indicando a presença de QTLs (*AC@G-2_68,00* e *AC@I-2_36,94*) (Tabela 4). A verificação no banco de QTLs descritos no Soybase permitiu identificar uma região, no GL G, associada à variável AC (Kim et al. 2012). No entanto, no GL I, o QTL descrito no Soybase, afetando essa variável, se encontra muito distante da região identificada neste estudo (>50 cM). Portanto, há evidência de que uma região ainda não descrita, afetando a variável AC, foi revelada neste estudo, no GL I. Portanto, os resultados sugerem que, dos 17 QTLs identificados neste estudo, somente pelo emprego da abordagem MIM, cinco ainda não foram descritos na literatura, sendo um QTL para cada uma das variáveis PEG (GL A1), IPDV (GL B1), NS (GL C1), AC (GL I) e PSB (GL C1).

A abordagem multivariada revelou QTLs que afetam um conjunto de variáveis analisadas nos nove grupos selecionados (*Multi-1* a *Multi-9*). Verificando-se os resultados, foi possível identificar uma alta coincidência de QTLs afetando diferentes conjuntos de características nas mesmas posições genômicas. Dessa forma, foi realizada uma comparação entre as mesmas posições genômicas, reveladas pelos diferentes grupos analisados, para identificar as regiões que apresentam o alelo proveniente do mesmo genitor favorecendo o fenótipo de cada QTL. Como o marcador identificado está flanqueando o QTL, assume-se que ambos estão segregando conjuntamente. Portanto, tanto o marcador quanto o QTL apresentam alelos provenientes do mesmo genitor. Como não foram estimados os efeitos de interação, não é possível afirmar se há a presença de pleiotropia ou de ligação em cada posição genômica. No entanto, do ponto de vista do melhoramento, o desejável é que o alelo proveniente do mesmo genitor contribua para favorecer o fenótipo de todas as variáveis, seja aumentando ou diminuindo os fenótipos em questão. No caso de QTLs ligados, se o alelo responsável pelo fenótipo de cada QTL de interesse é proveniente de genitores diferentes, seria necessário identificar os indivíduos que apresentam *crossing over* entre as regiões dos QTLs, permitindo assim que todos os QTLs apresentem o fenótipo desejável. A necessidade de um evento de *crossing over* entre QTLs próximos interfere diretamente no tamanho da população a ser avaliada, para aumentar a chance de identificação de indivíduos contendo os alelos de interesse para cada um dos QTLs. A informação do tipo de interação e do alelo responsável pelo efeito do QTL é imprescindível em um programa de melhoramento, para a definição da estratégia de condução e seleção das populações. Para as características avaliadas neste estudo, o fenótipo desejável é uma planta que apresente maior NDF, NDM, APF, APM, NVP, NS, VA, PG e PSB, e um menor PEG, IPDV, PCS e PSM. Uma planta com maior período vegetativo e reprodutivo apresenta maior capacidade de crescimento, o que resultará em uma maior quantidade de flores que serão posteriormente autofecundadas para a formação das vagens. Portanto, um maior número de vagens e de sementes resultará em uma maior produtividade de grãos. As características de altura da planta e aspecto visual de

produtividade de grãos são fatores que afetam diretamente a nota de VA. Por outro lado, uma planta com menor PEG apresenta sementes menores, com menor PCS, e fica menos exposta ao período crítico de ataque de percevejos, o que resulta em um menor IPDV e um menor PSM.

Para a região identificada no GL C2-1, o alelo proveniente do genitor IAC-100 contribui para favorecer todas as variáveis afetadas por essa região. Os grupos de variáveis que apresentaram QTL significativo nessa região foram *Multi-1*, *Multi-2*, *Multi-3*, *Multi-4*, *Multi-6*, *Multi-7*, *Multi-8* e *Multi-9* (Tabela 14, Tabela 15, Tabela 16, Tabela 17, Tabela 19, Tabela 20, Tabela 21 e Tabela 22). Nessa região genômica, o alelo proveniente do genitor IAC-100, para os QTLs identificados nos grupos mencionados, permite a obtenção de um genótipo com maior NDF, NDM, APM, NVP, NS, VA, PG e PSB, e menor PEG e IPDV.

Outras regiões identificadas também apresentaram o alelo do genitor IAC-100 contribuindo com todas as variáveis. No GL J, para os grupos *Multi-1*, *Multi-2*, *Multi-4*, *Multi-5*, *Multi-6*, *Multi-8* e *Multi-9* (Tabela 14, Tabela 15, Tabela 17, Tabela 18, Tabela 19, Tabela 21 e Tabela 22), o alelo IAC-100 permite a obtenção de um genótipo com maior NDF, APF, APM, NVP, NS, VA, PSB e PG e menor PEG e IPDV. No caso do QTL identificado no GL F-2 (posição 114 cM), nos grupos *Multi-7* e *Multi-9* (Tabela 20 e Tabela 22), genótipo com maior NVP, PSB e menor PCS e IPDV pode ser obtido com a presença do alelo proveniente do genitor IAC-100. Para o QTL *Multi-3@C1-2_21*, é possível obter planta com maior NDM e PSB (Tabela 16), para o QTL *Multi-5@E-1_67,00*, menor PEG e menor PCS (Tabela 18), para o QTL *Multi6@K-1_27,00*, maior PG (Tabela 19) e, para o QTL *Multi-7@E-1_43,97*, maior NVP e maior PSB (Tabela 20).

No caso do alelo proveniente do genitor CD-215, a região identificada no GL D2-2 (posição 46 cM) permite a obtenção de genótipo com maior APF, NDM, APM, NVP, NS, PSB e PG. Os grupos de variáveis que apresentaram QTL significativo nessa região foram *Multi-1*, *Multi-2*, *Multi-3*, *Multi-6* e *Multi-8* (Tabela 14, Tabela 15, Tabela 16, Tabela 19 e Tabela 21). Para o QTL *Multi-3@A1_42,00*, o alelo proveniente do genitor CD-215 permite a obtenção de planta com maior NDM (Tabela 16), e para o QTL *Multi-8@B1-2_30,00*, planta com maior NS, PSB e PG, e um menor IPDV (Tabela 21).

Analisando-se esses resultados apresentados acima, foi possível identificar que para os QTLs revelados na análise multivariada, o alelo proveniente da cultivar IAC-100 contribui em maior proporção para favorecer o fenótipo das variáveis, seja para aumentar ou reduzir o fenótipo em questão. Para as outras regiões que não foram discutidas, nem sempre o alelo do mesmo genitor favorece todas as variáveis. Como o mapeamento de QTLs foi realizado em uma única população, para que os QTLs identificados nessa população possam ser utilizados em programas de melhoramento, é necessário que as regiões genômicas de maior interesse sejam melhor investigadas. Além de confirmar a presença e a posição dos QTLs em outros *backgrounds* genéticos, é necessário validá-los em diferentes ambientes, para verificar sua representatividade no germoplasma da soja e estabilidade.

Como já discutido anteriormente, a identificação de QTLs de maior magnitude é relevante, pois resulta em uma maior eficiência de seleção para a característica de interesse. Do total de 20 regiões genômicas identificadas pela abordagem MT-MIM, oito se destacaram por apresentarem QTLs de maior magnitude para a maioria das variáveis analisadas nos diferentes grupos, sendo GLs C1-1, C2-1, D2-2 (posição 20 cM), E-1 (posição 67 cM), I-1 (posição 40 cM), J, M (posição 14 cM) e N. Como já discutido acima, para as regiões C2-1 e J, o alelo proveniente do genitor IAC-100 contribui para favorecer todas as variáveis que são afetadas pelos QTLs identificados nos diferentes grupos. Essas regiões devem ser melhor investigadas, pois contribuem para explicar a maior percentagem da variação fenotípica para a maioria das características analisadas nos diferentes grupos (*Multi-1* a *Multi-9*). Além de um maior valor de R^2 , essas regiões apresentam grande interesse no caso de seleção simultânea de características, pelo fato de afetarem diferentes

variáveis. A seleção simultânea é desejável em programas de melhoramento, pois permite uma maior eficiência na seleção, com ganhos genéticos superiores em programas de melhoramento (Bernardo 2010). A região genômica identificada no GL C1-1 contribuiu para explicar os maiores valores da variação fenotípica das características avaliadas nos grupos *Multi-1*, *Multi-3*, *Multi-4*, *Multi-5*, *Multi-6* e *Multi-9*. A região genômica identificada no GL C2-1 contribuiu para explicar os maiores valores da variação fenotípica das características avaliadas nos grupos *Multi-1*, *Multi-2*, *Multi-3*, *Multi-4*, *Multi-6*, *Multi-7* e *Multi-9*. A região genômica identificada no GL D2-2 (posição 20 cM) contribuiu para explicar os maiores valores da variação fenotípica das características avaliadas nos grupos *Multi-7* e *Multi-9*. E a região genômica identificada no GL E1 (posição 67 cM) contribuiu para explicar os maiores valores da variação fenotípica das características avaliadas nos grupos *Multi-5* e *Multi-7*. As regiões genômicas identificadas no GL I-1 (posição 40 cM) e no GL J contribuem para explicar os maiores valores da variação fenotípica das características avaliadas nos grupos *Multi-1*, *Multi-2*, *Multi-4*, *Multi-5*, *Multi-6*, *Multi-8* e *Multi-9*. A região genômica identificada no GL M (posição 14 cM) contribuiu para explicar os maiores valores da variação fenotípica das características avaliadas nos grupos *Multi-1*, *Multi-3*, *Multi-7* e *Multi-8*. Além dessas regiões genômicas identificadas explicando uma porcentagem significativa das características avaliadas em pelo menos dois dos grupos analisados, a região identificada no GL N, identificada somente na abordagem multivariada, no *Multi-8* e *Multi-9*, explica 5,19% da variação fenotípica da variável IPDV, sendo considerada uma região que também deve ser melhor investigada. Portanto, as oito regiões destacadas acima, identificadas na abordagem multivariada, representam as regiões genômicas que explicam a maior parte da variação fenotípica das características analisadas na população em estudo.

Para os QTLs identificados somente na análise de fenótipos individuais, e que não foram identificados na análise multivariada, os que explicaram uma alta porcentagem da variação fenotípica das variáveis analisadas e também devem ser melhor investigados para confirmação em outras populações e validação são: para a variável APF, o QTL *APF@N_3,00* explica 5,29% da variação fenotípica dessa variável (Tabela 4), para a variável NDM, os QTLs *NDM@A-1_39,00* e *NDM@F-1_56,00* explicam 7,87% e 6,60%, respectivamente, da variação fenotípica dessa variável (Tabela 4). Como a variável AC não foi utilizada na análise MT-MIM, o QTL *AC@I-1_36,00*, identificado na análise MIM, deveria ser melhor investigado, considerando que explica 5,40% da variação fenotípica dessa variável (Tabela 4). Já o QTL *NS@C2-1_99,00* explica 12,48% da variação fenotípica dessa variável e o QTL *PCS@E-2_65,82* explica 6,47% da variação fenotípica desta variável (Tabela 3). Considerando a importância de um fenótipo apresentar maior produtividade, os QTLs *PG@L_120,00* e *PG@M_90,00* deveriam ser melhor investigados por explicarem 4,36% e 9,50%, respectivamente, da variação fenotípica dessa variável (Tabela 4). Assim, as oito regiões destacadas acima, identificadas na abordagem univariada, representam as regiões genômicas que explicam a maior parte da variação fenotípica de características analisadas individualmente nessa população em estudo. Considerando que essas regiões não foram identificadas na abordagem multivariada, e a alta porcentagem de variação fenotípica explicada pelas mesmas, seria interessante uma investigação dessas regiões.

O emprego da abordagem multivariada é inédito na espécie e os resultados da análise de mapeamento de QTLs nessa população em estudo estão em consonância com o observado na literatura, de que essas características são controladas por muitos genes, sendo a maioria de pequeno efeito (Zhang et al. 2004; Hwang 2010; Han et al. 2012; Zhang et al. 2015). A maioria dos QTLs identificados nessa população em estudo, utilizando tanto a abordagem univariada quanto a abordagem multivariada, apresentou pequenos efeitos sobre a variância das características. No entanto, alguns QTLs de maior magnitude foram identificados, explicando uma maior porcentagem da variação fenotípica das características avaliadas. Os resultados obtidos pelo mapeamento MT-MIM, aliado ao mapeamento MIM, permitiram a identificação de 16 regiões, que explicam uma alta porcentagem da variação fenotípica da população

em estudo. Portanto, considerando a importância dessas regiões, as mesmas devem ser melhor investigadas e validadas para confirmação das informações verificadas nessa população em estudo.

A comparação dos resultados obtidos com QTLs descritos na literatura sugere que, muitos dos QTLs identificados na população utilizada neste trabalho, coincidem com regiões onde há QTLs descritos para as mesmas características, em outras populações de mapeamento. Apesar da abordagem de mapeamento diferir entre os estudos, e nenhum trabalho ter sido encontrado utilizando a abordagem MT-MIM, em soja, a identificação de QTLs afetando a mesma característica, em posições próximas do genoma, em diferentes populações, pode ser uma evidência de que um QTL já descrito na literatura também foi revelado na população em estudo. Algumas das comparações realizadas permitiram identificar o mesmo marcador molecular próximo ao QTL, o que aumenta essa evidência. Além disso, alguns estudos utilizando o mapeamento associativo, para algumas das características analisadas, também forneceram uma evidência maior para a confirmação desses QTLs. O mapeamento associativo, ao contrário das análises de ligação, utiliza indivíduos não aparentados, ou seja, populações não estruturadas, que apresentam menor desequilíbrio de ligação (DL) entre alelos. Além disso, o mapeamento associativo tem se mostrado um método de alta resolução para mapear locos de características quantitativas com base em DL (Yu and Buckler 2006). Considerando que a soja cultivada (*G. max* L. Merrill) apresenta uma base genética estreita, ou seja, possui baixa diversidade genética, e o processo de domesticação da soja selvagem (*G. soja* Sieb. and Zucc.) resultou em um alto desequilíbrio de ligação na espécie, a validação desses QTLs em estudos de mapeamento associativo é uma estratégia importante e que vem sendo utilizada (Hu et al. 2014; Zhang et al. 2015). Sonah et al. (2015) analisaram os 20 GLs da soja e verificaram que cada cromossomo apresenta um DL, variando de 250 kb a 2.5 Mb. Além disso, Hwang et al. (2014) identificaram valores de DL de 360 kb e 9600 kb para as regiões heterocromáticas e eucromáticas, respectivamente.

Novas regiões genômicas, ainda não descritas na literatura, também foram reveladas neste estudo, afetando as variáveis estudadas. Apesar do banco de dados de QTLs do Soybase apresentar uma grande quantidade de QTLs já descritos, para diferentes características, a descoberta de novos QTLs é de grande importância, tendo em vista o caráter quantitativo das características estudadas. Além disso, considerando os diferentes *backgrounds* genéticos utilizados em programas de melhoramento de soja, e a interação QTL x ambiente, nem sempre os mesmos QTLs serão identificados em populações distintas de mapeamento. A identificação de QTLs, em uma mesma posição genômica, afetando mais de uma característica, como observado neste trabalho, tem sido reportada em diferentes estudos de mapeamento (Palomeque et al. 2009a; Palomeque et al. 2009b; Niu et al. 2013; Hu et al. 2014).

A alta concentração de QTLs descritos no Soybase, em posições similares, controlando diferentes características, demonstra a relevância da utilização da abordagem de mapeamento de múltiplas características, utilizada neste trabalho. A identificação de regiões genômicas afetando um conjunto de características, é desejável em programas de melhoramento, tornando mais eficaz o processo de seleção de genótipos superiores. Dessa forma, os resultados obtidos na abordagem MT-MIM evidenciam que essas regiões devem ser melhor investigadas para auxiliarem na elucidação da base genéticas das mesmas. Além de fornecer maior precisão, pelo ajuste de múltiplos QTLs em um mesmo modelo, o mapeamento de intervalos múltiplos permite a obtenção de estimativas dos valores genotípicos dos indivíduos, os quais podem ser diretamente utilizados em programas de melhoramento genético (Zeng 2001). A identificação de regiões genômicas controlando um conjunto de variáveis é desejável em um programa de melhoramento e poderá contribuir para a seleção simultânea desse conjunto de características pelo uso da SAM. Após validação dessas regiões genômicas que se mostraram mais interessantes, deve-se identificar marcadores mais próximos aos QTLs de interesse, pelo mapeamento fino, para evitar que um *crossing over* ocorra entre o marcador e o QTL.

A soja é uma espécie que apresenta muitos genes duplicados em seu genoma. Tais características dificultam a elucidação dos mecanismos moleculares envolvidos no controle genético das características. No entanto, a construção do mapa da soja integrando informações de marcadores provenientes de cinco diferentes populações (ConsensusMap 4.0), aliada a grande disponibilidade de marcadores SNP distribuídos pelos 20 GLs da soja, propiciam grandes benefícios para a elucidação da base genética dos caracteres de importância para a cultura. Como as sequências genômicas dos marcadores SSR e SNP utilizados neste estudo encontram-se disponíveis no Soybase, uma próxima etapa importante será realizar o alinhamento das mesmas com dados do transcriptoma da soja para identificar genes candidatos. A comparação dessas sequências com genoma de espécies modelo, como *Arabidopsis thaliana*, também permitirá identificar a presença de genes ortólogos. Watanabe et al. (2012) utilizaram 46,367 anotações de genes da soja para comparação com genes já identificados em *Arabidopsis thaliana* responsáveis pelo florescimento. O estudo permitiu identificar 333 genes em soja ortólogos a 92 genes em *Arabidopsis*, revelando que muitos genes responsáveis pelo florescimento em soja apresentam múltiplas cópias, como já havia sido relatado na literatura (Schmutz et al. 2010). Os autores relatam que a presença de epistasia, interação com o ambiente e o grande número de genes duplicados no genoma são fatores que dificultam a elucidação dos mecanismos moleculares das características. Dessa forma, é importante que as informações obtidas em mapeamento de QTLs sejam posteriormente validadas em diferentes populações. A estratégia de mapeamento fino e a identificação de genes candidatos a partir das informações disponíveis em espécies modelo, como *Arabidopsis*, são ferramentas essenciais na elucidação dos mecanismos moleculares envolvidos na determinação das características de importância para o melhoramento da soja.

Estudos tem demonstrado que a utilização de QTLs, como efeito fixo, em modelos de seleção genômica contribui para uma maior acurácia do modelo (Bernardo 2014; Spindel et al. 2015; Arruda et al. 2016). Dessa forma, os QTLs identificados neste trabalho podem ser importantes para futuros estudos de seleção genômica, contribuindo para a identificação de genótipos superiores em programas de melhoramento de soja.

5. CONSIDERAÇÕES FINAIS

A utilização da abordagem multivariada no melhoramento de plantas é recente e tem se mostrado uma ferramenta interessante para o mapeamento de características complexas que apresentam correlação. Os resultados obtidos neste estudo permitem inferir que as abordagens MIM univariada e multivariada são complementares na busca por QTLs ao longo do genoma, com o objetivo de dissecar a arquitetura dos caracteres quantitativos. A abordagem MT-MIM apresentou maior poder de detecção de QTLs, além de uma maior precisão das estimativas dos parâmetros genéticos e na localização dos QTLs, em relação às análises de fenótipos individuais. Já a abordagem MIM permitiu revelar alguns QTLs particulares para cada variável analisada. Neste trabalho, a utilização da abordagem multivariada, aliada à abordagem univariada, foi satisfatória, permitindo a identificação de diferentes posições genômicas controlando as características avaliadas.

A análise univariada permitiu identificar 60 QTLs, sendo 29 QTLs associados às características de resistência aos percevejos e 31 QTLs relacionados às características agrônômicas avaliadas neste estudo. Os QTLs identificados encontram-se distribuídos por 13 GLs da soja, sendo GLs A1, B1, C1, C2, D2, E, F, G, I, J, L, M e N. A utilização de uma população com maior número de indivíduos, aliada à abordagem MIM, permitiram identificar um maior número de QTLs, em relação ao mapeamento realizado por Santos (2012). Foram revelados 28 QTLs que não haviam sido identificados anteriormente, para as variáveis NDM (6), PCS (4), NS (3), PEG (2), NDF (2), APF (2), IPDV (2), APM (2), PG (2), NVP (1), VA (1) e PSB (1). Esses QTLs contribuirão para um melhor entendimento da arquitetura genéticas dessas variáveis.

A abordagem multivariada permitiu identificar 20 posições genômicas, afetando os diferentes grupos de variáveis analisadas (*Multi-1 a Multi-9*), localizadas em 12 GLs da soja, sendo GLs A1, B1, C1 (2 regiões), C2, D2 (3 regiões), E (4 regiões), F (2 regiões), I (2 regiões), J, K, M e N. Apesar de algumas coincidências terem sido observadas nas regiões genômicas identificadas pelas duas abordagens, a abordagem MT-MIM forneceu estimativas de LOD superiores e uma maior precisão nas estimativas das posições dos QTLs, quando comparados aos QTLs identificados nos modelos MIM. Além disso, 13 regiões, com efeito em pelo menos uma das variáveis analisadas, foram reveladas somente pela abordagem MT-MIM. Portanto, os resultados corroboram a tendência dos modelos MT-MIM apresentarem maior precisão e poder estatístico quando comparados aos modelos MIM.

Uma comparação com o banco de dados de QTLs descritos no Soybase permitiu constatar que muitos dos QTLs revelados na população em estudo coincidiram com QTLs já descritos na literatura. Esse resultado pode ser uma evidência de que QTLs já descritos na literatura também tenham sido revelados nessa população. No entanto, considerando as posições genômicas reveladas nesse estudo, pela abordagem multivariada, e as características afetadas pelas mesmas, foram identificados 51 QTLs, que ainda não foram descritos na literatura, afetando nove características avaliadas. Desse total, 30 QTLs foram revelados somente com a utilização da abordagem multivariada, o que demonstra sua importância na identificação de QTLs para essa população em estudo. Os QTLs ainda não descritos na literatura foram identificados afetando as variáveis NDF (1), NDM (1), PEG (5), NVP (5), IPDV (9), NS (6), VA (4), PSB (16) e PSM (4). Considerando os QTLs que só foram revelados pela análise univariada, a comparação com o Soybase permitiu identificar cinco QTLs que ainda não foram descritos na literatura, sendo um QTL afetando cada uma das variáveis PEG, IPDV, NS, PSB e AV. Por se tratarem de características quantitativas controladas por muitos genes geralmente de pequeno efeito, a identificação desses 56 QTLs, nessa população em estudo, ainda não descritos na literatura, contribuirão para o entendimento da arquitetura genética dessas características. Como esses QTLs foram revelados em uma única população de mapeamento, avaliada em um único ambiente, é importante que suas posições

e efeitos sejam confirmados e os QTLs validados em outras populações com diferentes *backgrounds* genéticos, para verificar a representatividade dos mesmos no germoplasma da soja, e em outros ambientes, para verificar a estabilidade dos mesmos.

A maioria das regiões genômicas reveladas pela análise multivariada explica até 10% da variação fenotípica das características avaliadas. No entanto, oito regiões se destacaram por explicarem a maior parte da variação fenotípica das características, para os diferentes grupos analisados, sendo GLs C1-1, C2-1, D2-2 (posição 20 cM), E-1 (posição 67 cM), I-1 (posição 40 cM), J, M (posição 14 cM) e N. Para as regiões identificadas pela abordagem univariada, também foram identificados oito QTLs que explicam uma alta percentagem da variação fenotípica das variáveis APF (1), NDM (2), AC (1), NS (1), PCS (1) e PG (2). Portanto, esses 16 QTLs destacados foram os que explicaram maiores percentagens da variação dos fenótipos estudados. Assim, os mesmos poderão contribuir na seleção genótipos superiores, visando acelerar o processo de seleção e contribuir para um maior ganho genético. Já os QTLs identificados na abordagem MT-MIM são de grande relevância para a seleção simultânea de características de interesse, tornando mais eficaz o processo de seleção. Sendo assim, além da confirmação e validação desses QTLs, é necessário realizar um mapeamento fino nessas regiões de interesse, para identificar marcadores mais próximos aos QTLs, para que os mesmos possam ser utilizados na SAM.

Além disso, após o mapeamento fino, a comparação das sequências dos marcadores identificados próximos aos QTLs, com o banco de dados de transcriptoma da soja, bem como com o banco de dados de espécies modelo, como *Arabidopsis thaliana*, permitirão a identificação de genes candidatos, sendo ferramentas essenciais na elucidação dos mecanismos moleculares envolvidos na determinação das características de importância para o melhoramento da soja.

Por se tratarem de características quantitativas, a maioria dos QTLs identificados neste trabalho apresenta um pequeno efeito. A utilização desses QTLs para futuros estudos de seleção genômica pode ser importante para programas de melhoramento, contribuindo para a identificação de genótipos superiores.

REFERÊNCIAS

- Akkaya MS, Bhagwat AA, Cregan PB (1992) Length polymorphisms of simple sequence repeat DNA in soybean. *Genetics* 132:1131–1139.
- Alcivar A, Jacobsen J, Rainho J, et al (2007) Genetic analysis of soybean plant height, hypocotyl and internode lengths. *J Agric Food Environ Sci* 1:1–20.
- Alwala S, Kimbeng C, Gravois K, Bischoff K (2006) Trap, a new tool for sugarcane breeding: comparison with AFLP and coefficient of parentage. *J Am Soc Sugar Cane Technol* 26:62–86.
- ANP (2016) Agência nacional de petróleo, gás natural e biocombustíveis. <http://www.anp.gov.br/wwwanp/biocombustiveis>. Accessed 27 Nov 2016
- APASSUL (2017) Associação dos produtores e comerciantes de sementes e mudas do Rio Grande do Sul. http://www.apassul.com.br/index.php?menu=cultivares_mostra&acao=mostrar&chave=291. Accessed 27 Nov 2016
- APROBIO (2016) Associação dos produtores de biodiesel do Brasil. <http://aprobio.com.br/2015/11/24/soja-e-a-materia-prima-de-82-do-biodiesel-produzido-no-brasil/>. Accessed 27 Nov 2016
- APROSOJA (2016a) Associação dos produtores de soja - A história da soja. In: A história da soja. <http://aprosojabrasil.com.br/2014/sobre-a-soja/a-historia-da-soja/>. Accessed 27 Nov 2016
- APROSOJA (2016b) Associação dos produtores de soja - Uso da soja. In: Uso da soja. <http://aprosojabrasil.com.br/2014/sobre-a-soja/uso-da-soja/>. Accessed 27 Nov 2016
- Arahana V, Graef G, Specht J, et al (2001) Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in soybean. *Crop Sci* 41:180–188.
- Arias CAA, Brogin RL, Toledo JFF, et al (1999) Caracterização da reação de genótipos de soja à mancha parda em casa-de-vegetação. In: Congresso Brasileiro de Soja. Embrapa Soja, Londrina, p 468
- Arruda MP, Lipka AE, Brown PJ, et al (2016) Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol Breed*. doi: 10.1007/s11032-016-0508-5
- Arunakumari K, Durgarani C V., Satturu V, et al (2016) Marker-assisted pyramiding of genes conferring resistance against bacterial blight and blast diseases into indian rice variety MTU1010. *Rice Sci* 23:306–316. doi: 10.1016/j.rsci.2016.04.005
- Bachlava E, Dewey R, Burton J, Cardinal A (2009) Mapping and comparison of quantitative trait loci for oleic acid seed content in two segregating soybean populations. *Crop Sci* 49:433–442.
- Berlote L, Ramiro Z, Faria A, Marino C (2003) Danos causados por percevejos (Hemiptera: Pentatomidae) em cinco cultivares de soja (*Glycine max* (L.) Merrill, 1917) no município de Araçatuba, SP. *Arq do Inst Biológico* 70:169–175.
- Bernard R (1972) Two genes affecting stem termination in soybeans. *Crop Sci* 12:235–239.
- Bernard R (1971) Two genes for time of flowering in soybeans. *Crop Sci* 11:242–244.
- Bernardo R (2010) Breeding for quantitative traits in plants. Stemma Press, Woodbury
- Bernardo R (2014) Genomewide selection when major genes are known. *Crop Sci* 54:68–75. doi: doi:10.2135/cropsci2013.05.0315
- Boerma HR, Walker DR (2005) Discovery and utilization of QTLs for insect resistance in soybean. *Genetica* 123:181–189. doi: 10.1007/s10709-004-2741-9

- Boethel DJ (1999) Assessment of soybean germplasm for multiple insect resistance. In: Global plant genetic resources for insect resistant crops. CRC, Boca Raton - Florida, pp 101–129
- Bonato ER, Vello NA (1999) E6, a dominant gene conditioning early flowering and maturity in soybeans. *Genet Mol Biol* 22:229–232.
- Borém A, Vieira MLC, Colli W (2009) Glossário de biotecnologia. Visconde do Rio Branco, Suprema
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32:314–331.
- Bouchez A, Hospital F, Causse M, et al (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics* 162:1945–1959.
- Brier H, Rogers J (1991) Susceptibility of soybean to damage by *Nezara viridula* (L.) (Hemiptera: Pentatomidae) and *Riptortus serripes* (F.) (Hemiptera: Alydidae) during three stage of pod development. *J Aust Entomol Soc* 30:123–128.
- Bueno A de F, Panizzi RR, Corrêa-Ferreira BS, et al (2012) Histórico e evolução do manejo integrado de pragas da soja no Brasil. In: Soja: manejo integrado de insetos e outros artrópodes-praga. Embrapa, Brasília - DF, pp 37–74
- Bundy CS, McPherson RM (2000) Morphological examination of stink bug (Heteroptera: Pentatomidae) eggs on cotton and soybeans, with a key to genera. In: *Annals of the Entomological Society of America*. pp 616–624
- Buzzell R (1971) Inheritance of a soybean flowering response to fluorescent-daylength conditions. *Can J Genet Cytol* 13:703–707.
- Buzzell R, Voldeng H (1980) Inheritance of insensitivity to long day length. *Soybean Genet* 7:26–29.
- Byrne FJ, Castle S, Prabhaker N, Toscano NC (2003) Biochemical study of resistance to imidacloprid in B biotype *Bemisia tabaci* from Guatemala. *Pest Manag Sci* 59:347–352.
- Campos M, Knutson A, Heitholt J, Campos C (2010) Resistance to seed feeding by Southern green stink bug, *Nezara viridula* (Linnaeus), in soybean, *Glycine max* (L.) Merrill. *Southwest Entomol* 35:233–239.
- Campos OR (1996) Influência de duas variedades de soja *Glycine max* (L.) Merrill sobre insetos-praga e predadores entomófagos. Dissertation, Universidade Estadual Paulista - Ilha Solteira
- Carlini-Garcia LA, Lourenção AL, Pinheiro JB, et al (2011) Avaliação de populações de soja quanto à produtividade e resistência a percevejos. *Pesqui e Tecnol* 8:1–3.
- Carrao-Panizzi MC, Kitamura K (1995) Isoflavone content in Brazilian soybean cultivars. *Breed Sci* 45:295–300.
- Castro A, Chen X, Hyes P, Johnston M (2003) Pyramiding QTL alleles determining resistance to barley stripe rust: effects on resistance at the seedling stage. *Crop Sci* 43:651–659.
- Chang W, Dong L, Wang Z, et al (2011) QTL underlying resistance to two HG types of *Heterodera glycines* found in soybean cultivar “L-10”. *BMC Genomics* 12:233. doi: 10.1186/1471-2164-12-233
- Chapman A, Pantalone V, Ustun A, et al (2003) Quantitative trait loci for agronomic and seed quality traits in an F2 and F4:6 soybean population. *Euphytica* 129:387–393.
- Chen Q, Zhang Z, Liu C, et al (2007) QTL analysis of major agronomic traits in soybean. *Agric Sci China* 6:399–405.
- Chocorosqui VR (2001) Bioecologia de *Dichelops* (*Diceraeus*) *melacanthus* (Dallas, 1851) (Heteroptera: Pentatomidae), danos e controle em soja, milho e trigo no Norte do Paraná. Thesis, Universidade Federal do Paraná
- Choi IY, Hyten DL, Matukumalli LK, et al (2007) A soybean transcript map: Gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics* 176:685–696. doi: 10.1534/genetics.107.070821

- Cholastova T, Soldanova M, Pokorný R (2011) Random amplified polymorphic DNA (RAPD) and simple sequence repeat (SSR) marker efficacy for maize hybrid identification. *African J Biotechnol* 10:4794–4801. doi: 10.5897/AJB10.1344
- Chrispim TP, Ramos JM (2007) *Revista científica eletrônica de engenharia florestal - issn 1678-3867 p.*
- Chung J, Babka HL, Graef GL, et al (2003) The seed protein, oil, and yield QTL on soybean linkage group I. *Crop Sci* 43:1053–1067. doi: 10.2135/cropsci2003.1053
- Churchill G, Doerge R (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971.
- Cividanes FJ, Figueiredo JG (1997) Previsão de ocorrência de picos populacionais de percevejos pragas da soja em condições de campo. *An da Soc Entomológica do Bras* 26:517–525.
- Cividanes FJ, Parra JRP (1994) Zoneamento ecológico de *Nezara viridula* (L.), *Piezodorus guildinii* (West.) e *Euschistus beros* (Fabr.) (Heteroptera: Pentatomidae) em quatro estados produtores de soja no Brasil. *An da Soc Entomológica do Bras* 23:219–226.
- Clark WJ, Harris FA, Maxwell FG, Hartwig EE (1972) Resistance to certain soybean cultivars to bean leaf beetles, striped blister beetles, and bollworm. *J Econ Entomol* 65:1669–1672.
- Cober E, Molnar S, Charette M, Voldeng H (2010) A new locus for early maturity in soybean. *Crop Sci* 50:524–527.
- Cober E, Voldeng H (2001) Low R: FR light quality delays flowering of E7E7 soybean lines. *Crop Sci* 41:1823–1826. doi: 10.2135/cropsci2001.1823
- Coelho A (2000) Considerações gerais sobre a análise de QTL's. In: *Análise de QTL no melhoramento de plantas*. Funape, Goiânia, pp 1–36
- Collard BCY, Mackill DJ (2008) Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Phil Trans R Soc B* 363:557–572. doi: 10.1098/rstb.2007.2170
- CONAB (2017) Companhia Nacional de Abastecimento. <http://www.conab.gov.br/>. Accessed 27 Jan 2017
- Concibido V, Lange D, Denny R, et al (1997) Genome mapping of soybean cyst nematode resistance genes in “Peking”, PI 90763, and PI 88788 using DNA markers. *Crop Sci* 258–264.
- Concibido VC, Diers BW, Arelli PR (2004) A decade of QTL mapping for cyst nematode resistance in soybean. *Crop Sci* 44:1121–1131. doi: 10.2135/cropsci2004.1121
- COODETEC (2017) Cooperativa central agropecuária de desenvolvimento tecnológico e econômico Ltda. In: *Cultiv. CD 215*. http://www.coodetec.com.br/php/detalhes_cultivar.php?id=29. Accessed 5 Jan 2017
- Corrêa-Ferreira BS (2010) Ocorrência e dano de percevejos em cultivares de soja de crescimento determinado e indeterminado. In: *Anais da 31 Reunião de Pesquisa de Soja da Região Central do Brasil*. Embrapa Soja, Brasília, pp 78–80
- Corrêa-Ferreira BS (1993) Utilização do parasitoide de ovos *Trissolcus basalisi* (Wollaston) no controle de percevejos da soja. *Circ. Técnica* 11 Embrapa 40p.
- Corrêa-Ferreira BS (2003) Soja orgânica alternativas para o manejo dos insetos-pragas. Embrapa Soja, Londrina
- Corrêa-Ferreira BS (1984) Incidência do parasitoide *Eutrichopodopsis nitens* (Blanchard, 1966) em populações do percevejo verde *Nezara viridula* (Linnaeus, 1758). *An da Soc Entomológica do Bras* 13:321–330.
- Corrêa-Ferreira BS, Alexandre TM, Pellizzaro EC, et al (2010a) Práticas de manejo de pragas utilizadas na soja e seu impacto sobre a cultura. *78 Circ. Técnica* Embrapa 16p.
- Corrêa-Ferreira BS, Azevedo J (2002) Soybean seed damage by different species of stink bugs. *Agric For Entomol* 4:145–150.

- Corrêa-Ferreira BS, Krzyzanowski FC, Minami CA (2009) Percevejos e a qualidade da semente de soja – série sementes. *Circ Técnica* 67 Embrapa 1–16. doi: 10.1590/2317-1545v36n3919
- Corrêa-Ferreira BS, Lima D, Krzyzanowski FC (2010b) Ocorrência e dano de percevejos em cultivares de soja de crescimento determinado e indeterminado. In: *Anais da 31 Reunião de Pesquisa de Soja da Região Central do Brasil*. Embrapa Soja, Londrina, pp 78–80
- Corrêa-Ferreira BS, Panizzi AR (1999) Percevejos da soja e seu manejo. Embrapa-CNPSO - *Circ. Técnica* 24
- Corrêa-Ferreira BS, Peres WAA (2003) Comportamento da população dos percevejos-pragas e a fenologia da soja. In: *Soja orgânica: alternativa para o manejo dos insetos-pragas*, 1st edn. Embrapa Soja, Londrina, pp 27–32
- Cregan PB, Jarvik T, Bush AL, et al (1999a) An integrated genetic linkage map of the soybean genome. *Crop Sci* 39:1464–1490. doi: 10.2135/cropsci1999.3951464x
- Cregan PB, Mudge J, Fickus EW, et al (1999b) Targeted isolation of simple sequence repeat markers through the use of bacterial artificial chromosomes. *Theor Appl Genet* 98:919–928. doi: 10.1007/s001220051151
- Creste S, Neto AT, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Report* 19:299–306. doi: 10.1007/BF02772828
- Csanádi G, Vollmann J, Stift G, Lelley T (2001) Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor Appl Genet* 103:912–919.
- CTNBio (2017) Comissão Técnica Nacional de Biossegurança. <http://ctnbio.mcti.gov.br/>. Accessed 15 Jan 2017
- Da Silva MF, Schuster I, Da Silva JFV, et al (2007) Validation of microsatellite markers for assisted selection of soybean resistance to cyst nematode races 3 and 14. *Pesqui Agropecu Bras* 42:1143–1150. doi: 10.1590/S0100-204X2007000800011
- Datta K, Tu J, Oliva N, et al (2001) Enhanced resistance to sheath blight by constitutive expression of infection-related rice chitinase in transgenic elite Indica rice cultivars. *Plant Sci* 160:405–414.
- Degrande PE, Vivan LM (2009) Boletim de pesquisa de soja. Fundação MT, Rondonópolis
- Depieri RA, Panizzi AR (2011) Duration of feeding and superficial and in-depth damage to soybean seed by selected species of stink bugs (Heteroptera: Pentatomidae). *Neotrop Entomol* 40:197–203.
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA fresh tissue. *Focus (Madison)* 12:13–15.
- Du W, Wang M, Fu S, Yu D (2009) Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. *J Genet Genomics* 36:721–731.
- Duarte N (2007) Análise multivariada no mapeamento genético de traços quantitativos. Dissertation, Universidade de São Paulo - USP
- Embrapa (2003) Tecnologias de produção de soja - Região central do Brasil 2003. Embrapa Soja, Londrina
- Embrapa (2011) Tecnologias de produção de soja – Região central do Brasil 2012 e 2013. Embrapa Soja, Londrina
- Eskandari M, Cober E, Rajcan I (2013) Genetic control of soybean seed oil: II. QTL and genes that increase oil concentration without decreasing protein or with increased seed yield. *Theor Appl Genet* 126:1677–1687.
- Falconer D, Mackay T (1996) Introduction to quantitative genetics. Edinburgh, Longman
- Fan J-B, Oliphant A, Shen R, et al (2003) Highly parallel SNP genotyping. *Cold Spring Harb Symp Quant Biol* 68:69–78. doi: 10.1101/sqb.2003.68.69
- Fan J, Gunderson K, Bibikova M, et al (2006) Illumina universal bead arrays. *Methods Enzymol* 410:57–73.
- FAO (2017) Organização das nações unidas para a alimentação e a agricultura. <http://www.fao.org/brasil/pt/>. Accessed 27 Jan 2017

- Fehr WR, Caviness CE (1977) Stages of soybeans development. Iowa State University of Science and Technology, Ames
- Fernandes F, Athayde M, Lara F (1994) Comportamento de cultivares de soja no campo em relação ao ataque de percevejos. *Pesqui Agropecu Bras* 29:363–368.
- Ferreira ME, Grattapaglia D (1998) Introdução ao uso de marcadores moleculares em análises genéticas, 3.ed. Embrapa-CENARGEN, Brasília - DF
- Ferreira MF da S, Cervigni GDL, Ferreira A, et al (2011) QTLs for resistance to soybean cyst nematode, races 3, 9, and 14 in cultivar Hartwig. *Pesqui Agropecu Bras* 46:420–428. doi: 10.1590/S0100-204X2011000400012
- Fiorin RA, Stürmer GR, Guedes JVC, et al (2011) Métodos de aplicação e inseticidas no controle de percevejos na cultura da soja. *Semin Agrar* 32:139–146.
- Fischhoff D., Bowdish KS, Perlak FJ, et al (1987) Insect tolerant transgenic tomato plants. *Nat Biotechnol* 5:807–813.
- Fox C, Cary T, Nelson R, Diers B (2015) Confirmation of a seed yield QTL in soybean. *Crop Sci* 55:992–998.
- Funatsuki H, Ishimoto M, Tsuji H, et al (2006) Simple sequence repeat markers linked to a major QTL controlling pod shattering in soybean. *Plant Breed* 125:195–197.
- Funatsuki H, Kawaguchi K, Matsuba S, et al (2005) Mapping of QTL associated with chilling tolerance during reproductive growth in soybean. *Theor Appl Genet* 111:851–861.
- Gai J, Wang Y, Wu X, Chen S (2007) A comparative study on segregation analysis and QTL mapping of quantitative traits in plants-with a case in soybean. *Front Agric China* 1:1–7.
- Galileo MHM, Heinrichs EA (1979) Danos causados a soja em diferentes níveis e épocas de infestação durante o crescimento. *Pesqui Agropecu Bras* 14:279–282.
- Gallo D, Nakano O, Silveira Neto S, et al (2002) Entomologia agrícola. Fealq, Piracicaba
- Gazzoni DL (1998) Efeito de populações de percevejos na produtividade, qualidade da semente e características agrônômicas da soja. *Pesqui Agropecu Bras* 33:1229–1237.
- Gazzoni DL, Moscardi F (1998) Effect of defoliation levels on recovery of leaf area, on yield and agronomic traits of soybeans. *Pesqui Agropecu Bras* 33:411–424.
- Gianotto A (2015) Desenvolvimento de modelos de causalidade com informações de QTLs para estudo do relacionamento de caracteres fenotípicos relativos à absorção de fósforo em milho. Thesis, Universidade de São Paulo
- Gilioli JL, Terasawa F, Trecenti R, et al (1988) Descrição e uso de cultivares de soja para a região dos cerrados. FT Pesquisa e Sementes, Brasília
- Gilman DF, McPherson RM, Newsom LD, Herzog DC (1982) Resistance of soybean to the Southern green stink bug. *Crop Sci* 22:573–576.
- Glover K, Wang D, Arelli P, et al (2004) Near isogenic lines confirm a soybean cyst nematode resistance gene from PI 88788 on linkage group J. *Crop Sci* 44:936–941.
- Godói C, Pereira F, Umeno F, et al (2002) Resistência a insetos em populações de soja com diferentes proporções gênicas de genitores resistentes. *Pesqui Agropecu Bras* 32:47–55.
- Godoi CRC, Pinheiro JB (2009) Genetic parameters and selection strategies of soybean genotypes resistant to stink bug complex. *Genet Mol Biol* 32:328–336.
- Godoy KB, Galli JC, Ávila CJ, Corrêa-Ferreira BS (2005) Parasitismo em ovos de percevejos da soja *Euschistus heros* (Fabricius) e *Piezodorus guildinii* (Westwood) (Hemiptera: Pentatomidae) em São Gabriel do Oeste, MS. *Ciência Rural* 35:455–458.

- Grant NRTD, Cannon SB, Shoemaker RC (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acids Res* 38:843–846. doi: 10.1093/nar/gkp798
- Greene JK, Turnipseed SG, Sullivan MJ, May OL (2001) Treatment threshold for stink bug (Hemiptera: Pentatomidae) in cotton. *J Econ Entomol* 94:403–409.
- Guedes JVC, Arnemann JA, Stürmer GR, et al (2012) Percevejos da soja: novos cenários, novo manejo. *Rev. Plantio Direto* 28–34.
- Guo B, Sleper DA, Lu P, et al (2006) QTLs associated with resistance to soybean cyst nematode in soybean: Meta-analysis of QTL locations. *Crop Sci* 46:595–602. doi: 10.2135/cropsci2005.04-0036-2
- Guo B, Sleper D, PR A, et al (2005) Identification of QTLs associated with resistance to soybean cyst nematode races 2, 3 and 5 in soybean PI 90763. *Theor Appl Genet* 111:965–971.
- Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity (Edinb)* 69:315–324.
- Han Y, Xie D, Teng W, et al (2012) QTL underlying developmental behaviour of 100-seed weight of soybean. *Plant Breed* 131:600–606. doi: 10.1111/j.1439-0523.2012.01987.x
- Hisano H, Sato S, Isobe S, et al (2007) Characterization of the soybean genome using EST-derived microsatellite markers. *DNA Res* 14:271–281.
- Hittalmani S, Parco T, Mew T, Zeigler R (2000) Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in rice. *Theor Appl Genet* 100:1121–1128.
- Hoek J, Fehr W, Shoemaker R (2003) Molecular marker analysis of seed size in soybean.
- Hoffmann-Campo CB (1995) Role of the flavonoids in the natural resistance of soybean to *Heliothis virescens* (F.) and *Trichoplusia ni* (hübner). University of Reading
- Hoffmann-Campo CB, Harborne JB, Miccaffery AR (2001) Pré- ingestive and post-ingestive effects of soya bean extracts and rutin on *Trichoplusia ni* growth. *Entomol Exp Appl* 98:181–194.
- Hoffmann-Campo CB, Moscardi F, Corrêa-Ferreira BS, et al (2000) Pragmas da soja no Brasil e seu manejo integrado. *Circ Técnica* No 30 70p.
- Hoffmann-Campo CB, Ramos Neto JA, Oliveira MCN d., Oliveira LJ (2006) Detrimental effect of rutin on a main soybean defoliator pest, *Anticarsia gemmatalis*. *Pesqui Agropecu Bras* 41:1453–1459.
- Hu Z, Zhang D, Zhang G, et al (2014) Association mapping of yield-related traits and SSR markers in wild soybean (*Glycine soja* Sieb. and Zucc.). *Breed Sci* 63:441–9. doi: 10.1270/jsbbs.63.441
- Hu Z, Zhang H, Kan G, et al (2013) Determination of the genetic architecture of seed size and shape via linkage and association analysis in soybean (*Glycine max* L. Merr.). *Genetica* 141:247–254.
- Hulburt DJ, Boerma H, All JN (2004) Effect of pubescence tip on soybean resistance to lepidopteran insects. *J Econ Entomol* 97:621–627.
- Husch PG, Oliveira MCN, Sosa-Gómez DR (2012) Danos ocasionados por *Edessa meditabunda* (F.) e *Piezodorus guildinii* (West.) (Hemiptera: Pentatomidae) em sementes de soja. In: *Anais do 6 Congresso Brasileiro de Soja*. Embrapa Soja, Londrina, p 31
- Huynh T, Bastien M, Iqura E, et al (2010) Identification of QTLs associated with partial resistance to white mold in soybean using field-based inoculation. *Crop Sci* 50:969–979.
- Hwang S (2010) Linkage & selective mapping and QTL analysis in a Harosoy x Clark population of 300 RILs. University of Nebraska - Lincoln

- Hwang S, Ray J, Cregan P, et al (2014) Genetics and mapping of quantitative traits for nodule number, weight, and size in soybean (*Glycine max* L. [Merr.]). *Euphytica* 195:419–434.
- Hwang TY, Sayama T, Takahashi M, et al (2009) High-density integrated linkage map based on SSR markers in soybean. *DNA Res* 16:213–225. doi: 10.1093/dnares/dsp010
- Hyten D, Pantalone V, Sams C, et al (2004) Seed quality QTL in a prominent soybean population. *Theor Appl Genet* 109:552–561.
- Hyten D, Smith J, Frederick R, et al (2009) Bulk segregant analysis using the GoldenGate assay to locate the *Rpp3* locus that confers resistance to soybean rust in soybean. *Crop Sci* 49:265–271.
- Hyten D, Song Q, Zhu Y, et al (2006) Impacts of genetic bottlenecks on soybean genome diversity. *Proc Natl Acad Sci* 103:16666–16671.
- Hyten DL, Choi IY, Song Q, et al (2010) A high density integrated genetic linkage map of soybean and the development of a 1536 universal soy linkage panel for quantitative trait locus mapping. *Crop Sci* 50:960–968. doi: 10.2135/cropsci2009.06.0360
- Hyten DL, Song Q, Choi I-Y, et al (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theor Appl Genet* 116:945–952.
- Iqbal M, Meksem K, Njiti V, et al (2001) Microsatellite markers identify three additional quantitative trait loci for resistance to soybean sudden-death syndrome (SDS) in Essex x Forrest RILs. *Theor Appl Genet* 102:187–192.
- Jacob H, Lindpainter K, Lincoln S, et al (1991) Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. *Cell* 167:213–224.
- Jarquín D, Kocak K, Posadas L, et al (2014) Genotyping by sequencing for genomic prediction in a soybean breeding population. 1–10.
- Jensen RL, Newsom LD (1972) Effects of stink bug damaged soybean seed on germination, emergence, and yield. *J Econ Entomol* 65:261–264.
- Jiang C, Zeng Z (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics* 140:1111–1127.
- Jiao Y, Vuong T, Liu Y, et al (2015) Identification and evaluation of quantitative trait loci underlying resistance to multiple HG types of soybean cyst nematode in soybean PI 437655. *Theor Appl Genet* 128:15–23.
- Joshi T, Valliyodan B, Wu J, et al (2013) Genomic differences between cultivated soybean, *G. max* and its wild relative *G. soja*. *BMC Genomics* 14:S5. doi: 10.1186/1471-2164-14-S1-S5
- Josie J, Alcivar A, Rainho J, Kasse M (2007) Genomic regions containing QTL for plant height, internode length, and flower color in soybean [*Glycine max* (L.) Merr.]. *BiosOne J* 78:119–126.
- Jun T, Freewalt K, Michel A, Mian R (2014) Identification of novel QTL for leaf traits in soybean. *Plant Breed* 133:61–66. doi: 10.1111/pbr.12107
- Jung C, Wong C, Singh M, Bhalla P (2012) Comparative genomic analysis of soybean flowering genes. *PLoS One*. doi: <http://dx.doi.org/10.1371/journal.pone.0038250>
- Kabelka EA, Diers BW, Fehr WR, et al (2004) Putative alleles for increased yield from soybean plant introductions. *Crop Sci* 44:784–791.
- Kang S-T, Kwak M, Kim H-K, et al (2009) Population-specific QTLs and their different epistatic interactions for pod dehiscence in soybean [*Glycine max* (L.) Merr.]. *Euphytica* 166:15–24.
- Kao C, Zeng Z-B (1997) General formulas for obtaining the MLEs and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* 53:653–665.

- Kao C, Zeng Z-B (2002) Modeling epistasis of quantitative trait loci using Cockerham's model. *Genetics* 160:1243–1261.
- Kao C, Zeng Z, Teasdale R (1999) Multiple interval map-ping for quantitative trait loci. *Genetics* 152:1203–1216.
- Kato S, Sayama T, Fujii K, et al (2014) A major and stable QTL associated with seed weight in soybean across multiple environments and genetic backgrounds. *Theor Appl Genet* 127:1365–1374. doi: 10.1007/s00122-014-2304-0
- Kazi S, Shultz J, Afzal J, et al (2010) Iso-lines and inbred-lines confirmed loci that underlie resistance from cultivar “Hartwig” to three soybean cyst nematode populations. *Theor Appl Genet* 120:633–644. doi: 10.1007/s00122-009-1181-4
- Keim P, Diers BW, Olson TC, Shoemaker RC (1990) RFLP mapping in soybean: Association between marker loci and variation in quantitative traits. *Genetics* 126:735–742.
- Kim H, Diers B (2000) Inheritance of partial resistance to *Sclerotinia* stem rot in soybean. *Crop Sci* 40:55–61.
- Kim K-S, Diers BW, Hyten DL, et al (2012) Identification of positive yield QTL alleles from exotic soybean germplasm in two backcross populations. *Theor Appl Genet* 125:1353–1369.
- Kim KS, Chirumamilla A, Hill CB, et al (2014) Identification and molecular mapping of two soybean aphid resistance genes in soybean PI 587732. *Theor Appl Genet* 127:1251–1259. doi: 10.1007/s00122-014-2296-9
- Kim M, Hyten DL, Bent a F, Diers BW (2010) Fine mapping of the SCN resistance locus *rhg1-b* from PI 88788. *Plant Genome* 3:81–89. doi: 10.3835/plantgenome2010.02.0001
- Kogan M, Ortman EE (1978) Antixenosis - a new term proposed to replace Painter's “Nonpreference” modality of resistance. *Bull Entomol Soc Am* 24:175–176.
- Komatsu K, Okuda S, Takahashi M, et al (2005) QTL mapping of antibiosis resistance to common cutworm (*Spodoptera litura* Fabricius) in soybean. *Crop Sci* 45:2044–2048. doi: 10.2135/cropsci2004.0744
- Kong F, Liu B, Xia Z, et al (2010) Two coordinately regulated homologs of *FLOWERING LOCUS T* are involved in the control of photoperiodic flowering in soybean. *Plant Physiol* 154:1220–1231. doi: 10.1104/pp.110.160796
- Kong F, Nan H, Cao D, et al (2014) A new dominant gene *E9* conditions early flowering and maturity in soybean. *Crop Sci* 54:2529–2535. doi: 10.2135/cropsci2014.03.0228
- Korir P, Qi B, Wang Y, et al (2011) A study on relative importance of additive, epistasis and unmapped QTL for Aluminium tolerance at seedling stage in soybean. *Plant Breed* 130:551–562.
- Kosambi DD (1944) The estimation of map distance from recombination values. *Eugenics* 12:172–175.
- Kuroda Y, Kaga A, Norihiko T, et al (2013) QTL affecting fitness of hybrids between wild and cultivated soybeans in experimental fields. *Ecol Evol* 3:2150–2168.
- Kuss-Roggia RCR (2009) Distribuição espacial e temporal de percevejos da soja e comportamento de *Piezodorus guildinii* (Westwood, 1837) (Hemiptera: Pentatomidae) na soja (*Glycine max* (L.) Merrill) ao longo do dia. Thesis, Universidade Federal de Santa Maria
- Kuss CC, Toaldo VDB, Berghetti J, et al (2012) Percentagem de espécies de percevejos pentatomídeos ao longo do ciclo da soja no Norte do Paraná. In: *Anais da 7 Jornada Acadêmica da Embrapa Soja*. Embrapa Soja, Londrina, pp 30–34
- Kwok P, Gu Z (1999) Single nucleotide polymorphism libraries: why and how are we building them? *Mol Med Today* 5:538:543.
- Lambert L, Tyler J (1999) Appraisal of insect-resistant soybeans. In: *Economic, Environmental, and Social Benefits of Insect Resistance in Field Crops*. Entomological Society of America, Lanham, pp 131–148

- Lander E, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185–199.
- Lantmann A (2014) Gastos com defensivos. <http://www.projetosojabrasil.com.br/artigo-gastos-com-defensivos/>. Accessed 20 Dec 2016
- Lara FM (1991) Princípios de resistência de plantas a insetos. Ícone, São Paulo
- Lark K, Orf J, Mansur L (1994) Epistatic expression of quantitative trait loci (QTL) in soybean [*Glycine max* (L.) Merr.] determined by QTL association with RFLP alleles. *Theor Appl Genet* 88:486–489.
- Lark KG, Chase K, Adler F, et al (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Pnas* 92:4656–4660. doi: 10.1073/pnas.92.10.4656
- Laurie C, Wang S, Carlini-Garcia LA, Zeng Z-B (2014) Mapping epistatic quantitative trait loci. *BMC Genet* 15:112.
- Lee J, Yoo M, Jung J, et al (2015) Detection of novel QTLs for foxglove aphid resistance in soybean. *Theor Appl Genet* 128:1481–1488.
- Lee S, Mian M, McHale L, et al (2013) Novel quantitative trait loci for partial resistance to *Phytophthora sojae* in soybean PI 398841. *Theor Appl Genet* 126:1121–1132.
- Lee SH, Bailey MA, Mian MAR, et al (1996) Molecular markers associated with soybean plant height, lodging, and maturity across locations. *Crop Sci* 36:728–735.
- Lee SH, Park KY, Lee HS, et al (2001) Genetic mapping of QTLs conditioning soybean sprout yield and quality. *Theor Appl Genet* 103:702–709.
- Lehmann E (1975) *Nonparametrics: Statistical methods based on ranks*. Holden-Day, San Francisco
- Leite DC, Pinheiro JB, Campos JB, et al (2016) QTL mapping of soybean oil content for marker-assisted selection in plant breeding program. *Genet Mol Res* 15:1–11. doi: 10.4238/gmr.15017685
- Lewers K, Crane E, Bronson C, et al (1999) Detection of linked QTL for soybean brown stem rot resistance in “BSR 101” as expressed in a growth chamber environment. *Mol Breed* 5:33–42.
- Li D, Sun M, Han Y, et al (2010a) Identification of QTL underlying soluble pigment content in soybean stems related to resistance to soybean white mold (*Sclerotinia sclerotiorum*). *Euphytica* 172:49–57.
- Li W, Zheng D, Van K, Lee S (2008) QTL mapping for major agronomic traits across two years in soybean (*Glycine max* L. Merr.). *J Crop Sci Biotech* 11:171–190.
- Li Y, Li W, Zhang C, et al (2010b) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *New Phytol* 188:242–253.
- Liang Q, Cheng X, Mei M, et al (2010) QTL analysis of root traits as related to phosphorus efficiency in soybean. *Ann Bot* 106:223–234.
- Liu H (1998) *Statistical genomics: linkage, mapping and QTL analysis*. CRC Press, Boca Raton
- Liu J, Liu D, Tao W, et al (2000) Molecular marker-facilitated pyramiding of different genes for powdery mildew resistance in wheat. *Plant Breed* 119:21–24.
- Liu W, Kim M, Kang Y, et al (2011) QTL identification of flowering time at three different latitudes reveals homeologous genomic regions that control flowering in soybean. *Theor Appl Genet* 4:545–553.
- Liu Z, Anderson J, Hu J, et al (2005) A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci. *Theor Appl Genet* 111:782–794.
- Lorenz AJ, Smith KP, Jannink J-L (2012) Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. *Crop Sci* 52:1609–1621.

- Lourenção A, Miranda M (1987) Resistência de soja a insetos: VIII. IAC 78-2318, linhagem com resistência múltipla. *Bragantia* 46:65–72.
- Lourenção A, Rossetto C, Miranda M de (1985) Resistência de soja a insetos. III. Seleção de linhagens resistentes a percevejos. *Bragantia* 44:77–86.
- Lourenção AL, Costa AS, Miranda MAC de (1989) Resistência de campo ao vírus da queima-do-broto em genótipos de soja resistentes a insetos. *Bragantia* 48:209–214.
- Lu X, Xiong Q, Cheng T, et al (2017) A PP2C-1 allele underlying a quantitative trait locus enhances soybean 100-seed weight. *Mol Plant* 10:670–684. doi: 10.1016/j.molp.2017.03.006
- Luedders VD, Dickerson WA (1977) Resistance of selected soybean genotypes and segregating populations to cabbage looper feeding. *Crop Sci* 17:395–396.
- Lustosa PR, Zanuncio JC, Leite GLD, Picanço M (1999) Qualidade da semente e senescência de genótipos de soja sob dois níveis de infestação de percevejos (Pentatomidae). *Pesqui Agropecu Bras* 34:1347–1351.
- Ma Y, Wang DG, Li HC, et al (2011) Fine mapping of the R SC14Q locus for resistance to soybean mosaic virus in soybean. *Euphytica* 181:127–135. doi: 10.1007/s10681-011-0457-3
- Mackill D (2006) Breeding for resistance to abiotic stresses in rice: the value of quantitative trait loci. In: *Plant breeding: the Arnel R. Hallauer International Symposium*. Blackwell Publication, Ames, pp 201–212
- Magrini EA, Silveira Neto S, Botelho PSM, Negrim SG (1996) Ocorrência de percevejos no município de Pirassununga-SP, associada com a fenologia da soja. *Rev Agric* 71:187–193.
- Mansur L, Orf J (1995) Agronomic performance of soybean recombinant inbreds in northern USA and Chile. *Crop Sci* 35:422–425.
- Mansur L, Orf J, Chase K, et al (1996) Genetic mapping of agronomic traits using recombinant inbred lines of soybean. *Crop Sci* 36:1327–1336.
- Mansur L, Orf J, Lark K (1993) Determining the linkage of quantitative trait loci to RFLP markers using extreme phenotypes of recombinant inbreds of soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 86:914–918.
- Mao T, Li J, Wen Z, et al (2017) Association mapping of loci controlling genetic and environmental interaction of soybean flowering time under various photo-thermal conditions. *BMC Genomics* 18:415–431. doi: 10.1186/s12864-017-3778-3
- Margarido GRA, Souza AP, Garcia AAF (2007) OneMap: Software for genetic mapping in outcrossing species. *Hereditas* 144:78–79. doi: 10.1111/j.2007.0018-0661.02000.x
- Massman JM, Jung HJG, Bernardo R (2013) Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Sci* 53:58–66. doi: 10.2135/cropsci2012.02.0112
- McBlain B, Bernard R (1987) A new gene affecting the time of flowering and maturity in soybean. *J Hered* 78:160–162.
- McPherson RM, Buss GR (2007) Evaluating lepidopteran defoliation resistance in soybean breeding lines containing the stink bug (Hemiptera: Pentatomidae) resistance IAC-100 cultivar in their pedigrees. *J Econ Entomol* 100:962–968.
- McPherson RM, Buss GR, Roberts PM (2007) Assessing stink bug resistance in soybean breeding lines containing genes from germplasm IAC-100 assessing stink bug resistance in soybean breeding lines containing genes from germplasm IAC-100. *J Econ Entomol* 100:1456–1463.

- Merrill C, Goldman D, Sedman S, Ebert M (1981) Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science* (80-) 211:1437–1438.
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Mian M, Bailey M, Tamulonis J, et al (1996) Molecular markers associated with seed weight in two soybean populations. *Theor Appl Genet* 93:1011–1016.
- Mian MAR, Wang T, Phillips DV, et al (1999) Molecular mapping of the *Rcs3* gene for resistance to frogeye leaf spot in soybean. *Crop Sci* 39:1687–1691.
- Miranda MAC, Braga NR, Miranda FTS, et al (2001) IAC-23 e IAC-24: cultivares de soja resistentes a insetos para o estado de São Paulo. In: *Anais do Congresso Brasileiro de Melhoramento de Plantas*. Sociedade Brasileira de Melhoramento de Plantas, Goiânia, p 12
- Miranda MAC, Lourenção AL (2002) Melhoramento genético da soja para a resistência a insetos: uma realidade para aumentar a eficiência do controle integrado de pragas e viabilizar a soja orgânica. In: *2 Congresso Brasileiro de Soja e Mercosoja*. Embrapa Soja, Londrina, Paraná, pp 52–60
- Miranda MAC, Rossetto CJ, Rossetto D, et al (1979) Resistência de soja a *Nezara viridula* e *Piezodorus guildinii* em condições de campo. *Bragantia* 38:181–188.
- Miranda MAC de, Braga NR, Lourenção AL, et al (2003) Descrição, produtividade e estabilidade da cultivar de soja IAC-24, resistente a insetos. *Bragantia* 62:29–37.
- Missão MR (2006) Soja: origem, classificação, utilização e uma visão abrangente do mercado. *Rev Ciências Empres* 3:7–15.
- Möller M (2010) Mapeamento de locos de resistência quantitativa da soja ao complexo de percevejos. Dissertation, Universidade de São Paulo
- Moreira F (2015) Resistência ao complexo de percevejos em linhagens elite de soja. Dissertation, Universidade de São Paulo
- Moura MF (1999) Avaliação de linhagens de soja (*Glycine max* (L.) Merrill) visando adaptação às condições de cerrado e resistência a insetos. Dissertation, Universidade Federal de Goiás
- Musser FR, Catchot AL, Gibson BK, Knighten KS (2011) Economic injury levels for Southern green stink bugs (Hemiptera: Pentatomidae) in R7 growth stage soybeans. *Crop Prot* 30:63–69. doi: 10.1016/j.cropro.2010.09.011
- Narvel JM, Walker DR, Rector BG, et al (2001) A retrospective DNA marker assessment of the development of insect resistant soybean. *Crop Sci* 41:1931–1939. doi: 10.2135/cropsci2001.1931
- NCBI (2017) National Center for Biotechnology Information, Taxonomy. <https://www.ncbi.nlm.nih.gov/>. Accessed 5 Apr 2017
- Nichols D, Glover K, Carlson S, et al (2006) Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Sci* 46:834–839.
- Niu Y, Xu Y, Liu X-F, et al (2013) Association mapping for seed size and shape traits in soybean cultivars. *Mol Breed* 31:785–794. doi: 10.1007/s11032-012-9833-5
- Oldach KH, Peck DM, Nair RM, et al (2014) Genetic analysis of tolerance to the root lesion nematode *Pratylenchus neglectus* in the legume *Medicago littoralis*. *BMC Plant Biol* 14:100. doi: 10.1186/1471-2229-14-100
- Oliveira Kc (2015) Seleção e desenvolvimento de genótipos de soja resistentes ao complexo de percevejos. Thesis, Universidade de São Paulo

- Oliveira RP de, Cristofani M, Machado MA (2001) Marcadores RAPD para mapeamento genético e seleção de híbridos de citrus. *Rev Bras Frutic* 23:477–481.
- OrfJ, Chase K, Jarvik T, et al (1999) Genetics of soybean agronomic traits I. Comparison of three related recombinant inbred populations. *Crop Sci* 39:1642–1651.
- Ortega MA, All JN, Boerma HR, Parrott WA (2016) Pyramids of QTLs enhance host-plant resistance and Bt-mediated resistance to leaf-chewing insects in soybean. *Theor Appl Genet* 129:703–715. doi: 10.1007/s00122-015-2658-y
- Painter RH (1951) *Insect resistance in crop plants*, 1st edn. Macmillan and Co., New York
- Palomeque L, Li-Jun L, Li W, et al (2009a) QTL in mega-environments: I. Universal and specific seed yield QTL detected in a population derived from a cross of high-yielding adapted x high-yielding exotic soybean lines. *Theor Appl Genet* 119:417–427.
- Palomeque L, Li-Jun L, Li W, et al (2009b) QTL in mega-environments: II. Agronomic trait QTL co-localized with seed yield QTL detected in a population derived from a cross of high-yielding adapted x high-yielding exotic soybean lines. *Theor Appl Genet* 119:429–436.
- Panizzi AR (2002) Stink bugs on soybean in northeastern Brazil and a new record on the Southern green stink bug, *Nezara viridula* (L.) (Heteroptera: Pentatomidae). *Neotrop Entomol* 31:331–332.
- Panizzi AR (1997) Wild hosts of pentatomids: ecological significance and role in their pest status on crops. *Annu Rev Entomol* 42:99–122.
- Panizzi AR, Bueno AF, Silva FAC da (2012) Insetos que atacam vagens. In: *Soja: manejo integrado de insetos e outros artrópodes-praga*, 1st edn. Embrapa, Brasília, pp 335–420
- Panizzi AR, Chocorosqui VR (1999) Pragas: eles vieram com tudo! *Cultiv Gd Cult* 1:8–10.
- Panizzi AR, Machado-Neto E (1992) Development of nymphs and feeding habits of nymphal and adult *Edessa meditabunda* (Heteroptera: Pentatomidae) on soybean and sunflower. In: *Annals of the Entomological Society of America*. pp 477–481
- Panizzi AR, Niva CC (1994) Overwintering strategy of the brown stink bug in northern Paraná. *Pesqui Agropecu Bras* 29:509–511.
- Panizzi AR, Panizzi MCC, Bays IA, Almeida LA (1986) Danos por percevejos em genótipos de soja com semente pequena. *Pesqui Agropecu Bras* 21:571–577.
- Panizzi AR, Parra JRP, Santos CH, Carvalho DR (2000) Rearing the Southern green stink bug using artificial dry diet and artificial plant. *Pesqui Agropecu Bras* 35:1709–1715.
- Panizzi AR, Silva FAC (2009) Insetos sugadores de sementes (Heteroptera). In: *Bioecologia e nutrição de insetos. Base para o manejo integrado de pragas*. Embrapa, Brasília,
- Panizzi AR, Slansky Júnior F (1985) Review of phytophagous pentatomids (Hemiptera: Pentatomidae) associated with soybean in the Americas. *Florida Entomol* 68:184–214.
- Panizzi AR, Smith JC, Pereira LAG, Yamashita J (1979) Efeito de danos de *Piezodorus guildinii* (Westwood, 1837) no rendimento e qualidade da soja. In: *Anais do 1 Seminário Nacional de Pesquisa de Soja*. Embrapa Soja, Londrina, Paraná, pp 59–78
- Panizzi AR, Vivan LM (1997) Seasonal abundance of the neotropical brown stink bug, *Euschistus heros* in overwintering sites and the breaking of dormancy. *Entomol Exp Appl* 82:213–217.
- Panthee D, Pantalone V, Saxton A, et al (2007) Quantitative trait loci for agronomic traits in soybean. *Plant Breed* 126:51–57.

- Pathan SM, Vuong T, Clark K, et al (2013) Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Sci* 53:765–774. doi: 10.2135/cropsci2012.03.0153
- Patzoldt M, Carlson S, Diers B (2005a) Characterization of resistance to brown stem rot of soybean in five accessions from central China. *Crop Sci* 45:1092–1095.
- Patzoldt M, Grau C, Stephens P, et al (2005b) Localization of a quantitative trait locus providing brown stem rot resistance in the soybean cultivar Bell. *Crop Sci* 45:1241–1248.
- Pereira FB (2015) Estratégias de seleção para resistência a percevejos e alta produtividade em populações segregantes de soja Piracicaba. Thesis, Universidade de São Paulo
- Peterson BG, Carl P, Boudt K, et al (2015) Econometric tools for performance and risk analysis - PerformanceAnalytics.
- Pimentel-Gomes F (2000) Curso de estatística experimental. Degaspari, Piracicaba
- Pinheiro JB (1993) Dialelo parcial entre parentais de soja resistentes e suscetíveis a insetos. Dissertation, Universidade de São Paulo
- Pinheiro JB, Vello NA, Rossetto CJ, Zucchi MI (2005) Potential of soybean genotypes as insect resistance sources. *Crop Breed Appl Biotechnol* 5:294–301.
- Piubelli G, Hoffmann-Campo C, Moscardi F, et al (2005) Are chemical compounds important for soybean resistance to *Anticarsia gemmatalis*. *J Chem Ecol* 31:1515–1531.
- Piubelli GC (2004) Bioatividade de genótipos de soja resistentes a *A. gemmatalis* hübner (Lepidoptera: Noctuidae) e interações de suas substâncias químicas com inimigos naturais. Thesis, Universidade Federal do Paraná
- Piubelli GC, Campo CBH, Arruda IC, et al (2003a) Flavonoid increase in soybean genotypes as response of *Nezara viridula* injury and its effect on insect- feeding preference. *J Chem Ecol* 29:1223–1233.
- Piubelli GC, Campo CBH, Arruda IC, Lara FM (2003b) Nymphal development, lipid content, growth and weight gain of *Nezara viridula* (L.) (Heteroptera: Pentatomidae) fed on soybean genotypes. *Neotrop Entomol* 32:127–132.
- Qi Z, Han X, Hou M, et al (2014a) QTL analysis of soybean oil content under 17 environments. *Can J Plant Sci* 94:245–261. doi: 10.4141/cjps2013-219
- Qi Z, Hou M, Han X, et al (2014b) Identification of quantitative trait loci (QTLs) for seed protein concentration in soybean and analysis for additive effects and epistatic effects of QTLs under multiple environments. *Plant Breed* 133:499–507. doi: 10.1111/pbr.12179
- Qiu B, Arelli P, Sleper D (1999) RFLP markers associated with soybean cyst nematode resistance and seed composition in a “Peking” x “Essex” population. *Theor Appl Genet* 98:356–364.
- Quirino RJ (2012) Consequências na armazenagem da soja devido ao ataque de percevejos nas lavouras de soja. In: Palestra Congr. Bras. Soja. <http://www.cbsoja.com.br/anais/palestras/JoseRonaldoQuirino.pdf>. Accessed 20 Dec 2016
- R Development Core Team (2017) R: A language and environment for statistical computing.
- Ramiro ZA, Massariol AA (1977) Manejo de insetos na cultura da soja. In: A soja no Brasil central. Fundação Cargil, Campinas, pp 141–155
- Ray J, Hinson K, Mankono E, Malo F (1995) Genetic control of a long-juvenile trait in soybean. *Crop Sci* 35:1001–1006.
- Rector B, All J, Parrott W, Boerma H (1999) Quantitative Trait Loci for Antixenosis Resistance to Corn Earworm in Soybean. *Crop Sci* 39:531–538.

- Rector BG, All JN, Parrott WA, Boerma HR (2000) Quantitative trait loci for antibiosis resistance to corn earworm in soybean. *Crop Sci* 40:233–238.
- Rector BG, All JN, Parrott WA, Boerma HR (1998) Identification of molecular markers linked to quantitative trait loci for soybean resistance to corn earworm. *Theor Appl Genet* 96:786–790.
- Reinprecht Y, Poysa V, Yu K, et al (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase - free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome* 49:1510–1527.
- Reyna N, Sneller C (2001) Evaluation of marker-assisted introgression of yield QTL alleles into adapted soybean. *Crop Sci* 41:1317–1321.
- Riedel C, Habekuß A, Schliephake E, et al (2011) Pyramiding of Ryd2 and Ryd3 conferring tolerance to a German isolate of Barley yellow dwarf virus-PAV (BYDV-PAV-ASL-1) leads to quantitative resistance against this isolate. *Theor Appl Genet* 123:69–76. doi: 10.1007/s00122-011-1567-y
- Rizzo HFE (1972) Insectos y otros animales enemigos de la soja [*Glycine max* (L.) Merrill] en la Argentina. *Fitotec Latinoam* 8:44–49.
- Rocha F da (2015) Seleção de genótipos de soja para resistência ao complexo de percevejos. Thesis, Universidade de São Paulo
- Rocha F da, Bermudez F, Ferreira MC, et al (2014) Effective selection criteria for assessing the resistance of stink bugs complex in soybean. *Crop Breed Appl Biotechnol* 14:174–179. doi: 10.1590/1984-70332014v14n3a26
- Rodrigues CJ (1996) Influência de duas cultivares de soja *Glycine max* (L.) Merrill sobre insetos-praga e seus inimigos naturais. Dissertation, Universidade Estadual Paulista - Ilha Solteira
- Rodrigues JI da S, De Miranda FD, Piovesan ND, et al (2016) QTL mapping for yield components and agronomic traits in a brazilian soybean population. *Crop Breed Appl Biotechnol* 16:265–273. doi: 10.1590/1984-70332016v16n4a41
- Rossetto CJ (1989) Breeding for resistance to stink bugs. In: *Anais da 4 Conferencia Mundial de Investigacion en Soja*. Bueno Aires - Argentina, pp 2046–2060
- Rossetto CJ, Gallo PB, Razera LF, et al (1995) Mechanisms of resistance to stink bug complex in the soybean cultivar “iAC-100.” *An da Soc Entomológica do Bras* 24:517–522.
- Rossetto CJ, Igue T, Miranda MAC, Lourenção AL (1986) Resistência de soja a insetos: VI. Comportamento de genótipos em relação a percevejos. *Bragantia* 45:323–335.
- Rossetto CJ, Lara FM (1991) Diluição de dano, mecanismo da pseudo-resistência. In: *13 Congresso Brasileiro de Entomologia*. Sociedade Brasileira de Entomologia, Recife, p 535
- Rossetto CJ, Lourenção AL, Miranda MAC, Igue T (1984) Resistência de soja a insetos: ii. Teste de livre escolha entre a linhagem iAC73-228 e o cultivar Paraná, infestados por *Nezara viridula* (L.) em telado. *Bragantia* 43:141–153.
- Rossi M, Orf J, Liu L, et al (2013) Genetic basis of soybean adaptation to North American vs. Asian mega-environments in two independent populations from Canadian × Chinese crosses. *Theor Appl Genet* 127:7.
- Roza-Gomes MF (2010) Avaliação de danos de quatro espécies de percevejos (Heteroptera: Pentatomidae) em trigo, soja e milho. Thesis, Universidade de Passo Fundo
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386.
- Russin JS, Orr DB, Layton MB, Boethel DJ (1988) Incidence of microorganisms in soybean seeds damaged by stink bug feeding. *Phytopathology* 78:306.

- Saluso A, Xavier L, Panizzi AR (2011) An invasive pentatomid pest in Argentina: neotropical brown stink bug, *Euschistus heros* (F.) (Hemiptera: Pentatomidae). *Neotrop Entomol* 40:704–705.
- Samanfar B, Molnar S, Charette M, et al (2017) Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor Appl Genet* 130:377–390.
- Santana FA, Freire M, Kellen J, et al (2014) Marker-assisted selection strategies for developing resistant soybean plants to cyst nematode. *Crop Breed Appl Biotechnol* 14:180–186.
- Santos M da F (2012) Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevejos. Thesis, Universidade de São Paulo
- Santos M, Geraldi I, Garcia A, et al (2013) Mapping of QTLs associated with biological nitrogen fixation traits in soybean. *Hereditas* 150:17–25.
- SAS INSTITUTE (2001) Statistical Analysis software.
- Schlötterer C (2004) The evolution of molecular markers—just a matter of fashion? *Nat Rev Genet* 5:63–69. doi: 10.1038/nrg1249
- Schmutz J, Cannon SB, Schlueter J, et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–83. doi: 10.1038/nature08670
- Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments A poor man 's approach to genotyping for research and high-throughput diagnostics . *Prism* 18:1–2. doi: 10.1038/72708
- Schumann FW, Todd JW (1982) Population dynamics of the Southern green stink bug (Heteroptera: Pentatomidae) in relation to soybean phenology. *J Econ Entomol* 75:748–753.
- Schuster I, Cruz C (2004) Estatística genômica aplicada a populações derivadas de cruzamentos controlados. Viçosa, MG
- Searle SP, Casela G, MacCulloch CE (1992) Variance components, 1st edn. Sons, John Wiley &, New York
- Sebolt A, Shoemaker R, Diers B (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci* 40:1438–1444.
- SECEX (2017) Ministério da indústria, comércio exterior e serviços. <http://www.mdic.gov.br/>. Accessed 27 Jan 2017
- Sedyama T (2009) Tecnologias de produção e usos da soja. Mecenas, Londrina
- Shanti M, George M, Cruz C, et al (2001) Identification of resistance genes effective against rice bacterial blight pathogen in eastern India. *Plant Dis* 85:506–512.
- Shoemaker R, Specht J (1995) Integration of the soybean molecular and classical genetic linkage groups. *Crop Sci* 35:436–4446.
- Silva FAC (2009) Fatores físicos e químicos mediadores da oviposição de pentatomídeos fitófagos em substratos artificiais. Thesis, Universidade Federal do Paraná
- Silva FAC, Silva JJ, Depieri RA, Panizzi AR (2012a) Feeding activity, salivary amylase activity, and superficial damage to soybean seed by adult *Edessa meditabunda* (F.) and *Euschistus heros* (F.) (Hemiptera: Pentatomidae). *Neotrop Entomol* 41:386–390. doi: doi:10.1007/s13744-012-0061-9
- Silva JPGF, Baldin ELL, Souza ES, et al (2013) Characterization of antibiosis to the redbanded stink bug *Piezodorus guildinii* (Hemiptera: Pentatomidae) in soybean entries. *J Pest Sci* (2004) 86:649–657. doi: 10.1007/s10340-013-0527-5
- Silva L, Wang S, Zeng Z-B (2012b) Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. *BMC Genet* 13:67.

- Silva L, Zeng Z-B (2010) Current progress on statistical methods for mapping quantitative trait loci from inbred line crosses. *J Biopharm Stat* 20:454–481.
- Silva MF, Schuster I, Silva JFV, et al (2007) No Validation of microsatellite markers for assisted selection of soybean resistance to cyst nematode races 3 and 14. *Pesqui Agropecu Bras* 42:1143–1150.
- Silva MTB, Ferreira BS., Sosa-Gómez DR (2006) Controle de percevejos em soja. In: *Tecnologia de aplicação de defensivos agrícolas*, 1st edn. Aldeia Norte, Passo Fundo, pp 109–122
- Silva MTB da, Ruedell J (1983) Ocorrência de percevejos fitófagos da família pentatomidae em soja (*Glycine max* (L.) Merrill). *Trigo e Soja* 65:4–6.
- Singh RJ, Hymowitz T (1988) The genomic relationship between *Glycine max* (L.) Merr. and *G. soja* Sieb. and Zucc. as revealed by pachytene chromosome analysis. *Theor Appl Genet* 76:705–711.
- Singh S, SIDHU J, Huang N, et al (2001) Pyramiding three bacterial blight resistance genes (xa5, xa13 and Xa21) using marker-assisted selection into Indica rice cultivar PR106. *Theor Appl Genet* 102:1011–1015.
- Sonah H, Donoughue LO, Cober E, Rajcan I (2015) Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol J* 13:211–221. doi: 10.1111/pbi.12249
- Song Q, Hyten DL, Jia G, et al (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8:1–12. doi: 10.1371/journal.pone.0054985
- Song QJ, Marek LF, Shoemaker RC, et al (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109:122–128. doi: 10.1007/s00122-004-1602-3
- Soria MF, Degrande PE, Panizzi AR (2010) Algodoeiro invadido. *Rev Cultiv* 131:18–20.
- Sosa-Gómez DR, Corson IC, Morales L (2001) Insecticide resistance to endosulfan, monocrotophos and metamidophos in the neotropical Brown stink bug, *Euschistus berus* (F.). *Neotrop Entomol* 30:317–320.
- Sosa-Gómez DR, de Carvalho M da C da CG, Marcelino-Guimarães FC, Hoffmann-Campo CB (2012) A biotecnologia, o melhoramento e o manejo de pragas da soja. In: *Soja - Manejo integrado de insetos e outros artrópodes-praga*. Embrapa, Brasília, pp 725–788
- Sosa-Gómez DR, Delpin KE, Almeida A, Hirose E (2004) Genetic differentiation among Brazilian populations of *Euschistus berus* (Fabricius) (Heteroptera: Pentatomidae) based on RAPD analysis. *Neotrop Entomol* 33:179–187.
- Sosa-Gómez DR, Moscardi F (1995) Retenção foliar diferencial em soja provocada por percevejos (Heteroptera: Pentatomidae). *An da Soc Entomológica do Bras* 24:401–404.
- Sosa-Gómez DR, Omoto C (2012) Resistência a inseticidas e outros agentes de controle em artrópodes associados à cultura da soja. In: *Soja - Manejo integrado de insetos e outros artrópodes-praga*. Embrapa, Brasília, pp 673–724
- Sosa-Gómez DR, Silva JJ (2010) Neotropical brown stink bug (*Euschistus berus*) resistance to metamidophos in Paraná, Brazil. *Pesqui Agropecu Bras* 45:767–769.
- Sosa-Gómez DR, Silva JJ, Costa F, et al (2005) Population structure of the Brazilian Southern green stink bug, *Nezara viridula* L. (Heteroptera: Pentatomidae). *J Insect Sci* 5:23.
- Souza E deS. (2013) Complexo de percevejos da soja: reflexos sobre a produtividade de genótipos e determinação dos tipos de resistência a *Euschistus berus* (Fabricius) (Hemipter: Pentatomidae). Thesis, Universidade Estadual Paulista Júlio de Mesquita Filho - Campus Botucatu
- Souza RF, Toledo JFF (1995) Genetic analysis of soybean resistance to stinkbug. *Brazilian J Genet* 18:593–598.
- Specht J, Chase K, Macrander M, et al (2001) Soybean response to water: A QTL analysis of drought tolerance. *Crop Sci* 41:493–509.

- Spindel J, Begum H, Akdemir D, et al (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet* 11:e1004982. doi: 10.1371/journal.pgen.1004982
- Stuber CW, Lincoln SE, Wolff DW, et al (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. *Genetics* 132:823–839.
- Stupar RM (2010) Into the wild: The soybean genome meets its undomesticated relative. *PNAS* 107:21947–21948. doi: 10.1073/pnas.1016809108
- Sun D, Li W, Zhang Z, et al (2006) Quantitative trait loci analysis for the developmental behavior of Soybean (*Glycine max* L. Merr.). *Theor Appl Genet* 112:665–673.
- Sun S, Kim M, Van K, et al (2013) QTLs for resistance to Phomopsis seed decay are associated with days to maturity in soybean (*Glycine max*). *Theor Appl Genet* 128:2029–2038.
- Swaminathan S, Abeyssekara N, Liu M, et al (2016) Quantitative trait loci underlying host responses of soybean to *Fusarium virguliforme* toxins that cause foliar sudden death syndrome. *Theor Appl Genet* 129:495–506.
- Takeshima R, Hayashi T, Zhu J, et al (2016) A soybean quantitative trait locus that promotes flowering under long days is identified as FT5a, a FLOWERING LOCUS T ortholog. *J Exp Bot* 67:5247–5258. doi: 10.1093/jxb/erw283
- Tasma IM, Lorenzen LL, Green DE, Shoemaker RC (2001) Mapping genetic loci for flowering time, maturity, and photoperiod insensitivity in soybean. *Mol Breed* 8:25–35.
- Teng W, Han Y, Du Y, et al (2009) QTL analyses of seed weight during the development of soybean (*Glycine max* L. Merr.). *Heredity (Edinb)* 102:372–380.
- Terasawa F (1982) FT Pesquisa & Sementes - Relatório Técnico. Ponta Grossa
- Terry LI, Chase K, Jarvik T, et al (2000) Soybean quantitative trait loci for resistance to insects. *Crop Sci* 40:375–382.
- Thoday J (1961) Location of polygenes. *Nature* 191:368–370.
- Toledo AM (2005) Indução de isoflavonóides em soja por *Piezodorus guildinii* e efeito de extratos e de isoflavonóides na preferência alimentar de *Nezara viridula* (hemiptera: pentatomidae). Dissertation, Universidade Estadual Paulista
- Toojinda T, Tragroonrung S, Vanavichit A, et al (2005) Molecular breeding for rainfed lowland rice in the Mekong region. *Plant Prod Sci* 8:330–377.
- Trumper EV, Edelstein JD (2008) El complejo de chinches fitofagos en soja: revision y avances en el estudio de su ecología y manejo. INTA
- Tucker D, Maroof M, Mideros S, et al (2010) Mapping quantitative trait loci for partial resistance to *Phytophthora sojae* in a soybean interspecific cross. *Crop Sci* 50:628–635.
- Umbeck P, Johnson G, Barton K, Swain W (1987) Genetically transformed cotton (*Gossypium hirsutum* L.) plants.
- USDA (2017) United States Department of Agriculture. <https://www.usda.gov/oce/commodity/wasde/>. Accessed 12 Jan 2017
- Vaeck M, Reynaerts A, Höfte H, et al (1987) Transgenic plants protected from insect attack. *Nature* 328:33–37.
- Van Duyn JW, Turnipseed SG, Maxwell JD (1971) Resistance in soybeans to the Mexican bean beetle. *Crop Sci* 11:572–573.
- Van Duyn JW, Turnipseed SG, Maxwell JD (1972) Resistance in soybeans to the Mexican bean beetle II. *Crop Sci* 12:561–562.

- Van Ooijen J, Voorrips R, Jansen T, Maliepaard C (2002) MapQTL 4.0, software for the calculation of QTL positions on genetic maps.
- Veiga RFA, Rossetto CJ, Razera LF, et al (1999) caracterização morfológica e agrônômica do cultivar de soja “Iac-100.” IAC - Bol. Técnico 177 - Inst. Agrônômico 23p.
- Venzon M, Rippasati JG, Ferreira JAM, Viríssimo JH (1999) Controle biológico de percevejos-da-soja no Triângulo Mineiro. *Ciência e Agrotecnologia* 23:70–78.
- Verneti F de J, Verneti Jr. F de J (2009) Genética da soja: Caracteres qualitativos e diversidade genética. Brasília - DF
- Vieira AJD, De Oliveira DA, Soares TCB, et al (2006) Use of the QTL approach to the study of soybean trait relationships in two populations of recombinant inbred lines at the F7 and F8 generations. *Brazilian J Plant Physiol* 18:281–290. doi: 10.1590/S1677-04202006000200004
- Villas-Bôas GL, Gazzoni DL, Oliveira MCN, et al (1990) Efeito de diferentes populações de percevejos sobre o rendimento e seus componenetes, características agrônômicas e qualidade de sementes de soja. *Bol. Pesqui. 1 - Embrapa-CNPSO* 43p.
- Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93:77–78.
- Vos P, Hogers R, M B, et al (1995) A new technique for DNA fingerprinting. *Nucleic Acids Res* 44:388–396.
- Walker D, Boerma H, All J, Parrott W (2002) Combining cry1Ac with QTL alleles from PI 229358 to improve soybean resistance to lepidopteran pests. *Mol Breed* 9:43–51.
- Walker D, Monteros M, Yates J (2010) Molecular breeding. In: *Genetics, genomics and breeding of soybean*. Science Publishers, Enfield, NH, pp 123–167
- Walker DR, Narvel JM, Boerma HR, et al (2004) A QTL that enhances and broadens Bt insect resistance in soybean. *Theor Appl Genet* 109:1051–1057. doi: 10.1007/s00122-004-1714-9
- Wang D, Graef GL, Procopiuk AM, Diers BW (2004) Identification of putative QTL that underlie yield in interspecific soybean backcross populations. *Theor Appl Genet* 108:458–467.
- Wang T, Ma X, Li Y, et al (2011) Changes in yield and yield components of single-cross maize hybrids released in China between 1964 and 2001. *Crop Sci* 51:512–525. doi: 10.2135/cropsci2010.06.0383
- Wang X, Jiang G, Green M, et al (2014) Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. *Mol Genet Genomics* 289:935–949.
- Warrington C V, Zhu S, Parrott W a, et al (2008) Seed yield of near-isogenic soybean lines with introgressed quantitative trait loci conditioning resistance to corn earworm (*Lepidoptera: Noctuidae*) and soybean looper (*Lepidoptera: Noctuidae*) from PI 229358. *J Econ Entomol* 101:1471–7.
- Watanabe S, Harada K, Abe J (2012) Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breed Sci* 61:531–43. doi: 10.1270/jsbbs.61.531
- Watanabe S, Hideshima R, Zhengjun X, et al (2009) Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics* 182:1251–1262. doi: 10.1534/genetics.108.098772
- Watanabe S, Xia Z, Hideshima R, et al (2011) A map-based cloning strategy employing a residual heterozygous line reveals that the GIGANTEA gene is involved in soybean maturity and flowering. *Genetics* 188:395–407. doi: 10.1534/genetics.110.125062
- Weng C, Yu K, Anderson T, Poysa V (2007) A quantitative trait locus influencing tolerance to *Phytophthora* root rot in the soybean cultivar “Conrad.” *Euphytica* 158:81–86.

- Williams J, Kubelik A, Livak K, et al (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res* 18:6531–6535.
- Winter S, Shelp B, Anderson T, et al (2007) QTL associated with horizontal resistance to soybean cyst nematode in *Glycine soja* PI464925B. *Theor Appl Genet* 114:461–472.
- Wu C, Sun S, Nimmakayala P, et al (2004) A BAC- and BIBAC-based physical map of the soybean genome. *Genome Res* 14:319–326.
- Xavier A, Muir WM, Rainey KM (2016) Assessing predictive properties of genome-wide selection in soybeans. 6:2611–2616. doi: 10.1534/g3.116.032268
- Xia Z, Tsubokura Y, Hoshi M, et al (2007) An integrated high-density linkage map of soybean with RFLP, SSR, STS, and AFLP markers using a single F2 population. *DNA Res* 14:257–269. doi: 10.1093/dnares/dsm027
- Xia Z, Watanabe S, Yamada T, et al (2012) PNAS Plus: Positional cloning and characterization reveal the molecular basis for soybean maturity locus E1 that regulates photoperiodic flowering. *Proc Natl Acad Sci* 109:E2155–E2164. doi: 10.1073/pnas.1117982109
- Yamada T, Funatsuki H, Hagihara S, et al (2009) A major QTL, qPDH1, is commonly involved in shattering resistance of soybean cultivars. *Breed Sci* 59:435–440.
- Yan H, Wang H, H C, et al (2015) Detection and fine-mapping of Soybean mosaic virus resistance genes via linkage and association analysis in soybean. *J Integr Plant Biol* 57:722.
- Yang Z, Xin D, Liu C, et al (2013) Identification of QTLs for seed and pod traits in soybean and analysis for additive effects and epistatic effects of QTLs among multiple environments. *Mol Genet Genomics* 288:651–667.
- Yao D, Liu ZZ, Zhang J, et al (2015) Analysis of quantitative trait loci for main plant traits in soybean. *Genet Mol Res* 4:6101–6109.
- Yesudas CR, Sharma H, Lightfoot DA (2010) Identification of QTL in soybean underlying resistance to herbivory by Japanese beetles (*Popillia japonica*, Newman). *Theor Appl Genet* 121:353–362. doi: 10.1007/s00122-010-1314-9
- Yu J, Buckler E. (2006) Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 17:155–160.
- Yuan J, Njitti V, Meksem K, et al (2002) Quantitative trait loci in two soybean recombinant inbred line populations segregating for yield and disease resistance. *Crop Sci* 42:271–277.
- Yue Y, Liu N, Jiang B, et al (2017) A single nucleotide deletion in J encoding *GmELF3* confers long juvenility and is associated with adaption of tropic soybean. 10:656–658. doi: 10.1016/j.molp.2016.12.004
- Zeng Z (2001) Statistical methods for mapping quantitative trait loci. Raleigh, North Carolina
- Zeng Z (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc Natl Acad Sci* 90:10972–10976.
- Zeng Z-B, Kao C, Basten C (1999) Estimating the genetic architecture of quantitative traits. *Genet Res (Camb)* 74:279–289.
- Zerbino MS (2010) Manejo de chinches en soja. *INIA* 23:24–27.
- Zhang D, Chen H, Wang H, et al (2010) Identification of genomic regions determining flower and pod numbers development in soybean (*Glycine max* L.). *J Genet Genomics* 37:545–556.
- Zhang G, Gu C, Wang D (2009) Molecular mapping of soybean aphid resistance genes in PI 567541B. *Theor Appl Genet* 118:473–482. doi: 10.1007/s00122-008-0914-0

- Zhang H, Li C, Davis EL, et al (2016) Genome-wide association study of resistance to soybean cyst nematode (*Heterodera glycines*) HG Type 2.5.7 in wild soybean (*Glycine soja*). *Front Plant Sci* 7:1–11. doi: 10.3389/fpls.2016.01214
- Zhang J, Li W, Zhang L, et al (2013) QTL mapping of soybean resistance to whitefly (*Bemisia tabaci gennadius*) under multi-environment conditions. *Aust J Crop Sci* 7:1212–1218.
- Zhang J, Song Q, Cregan PB, et al (2015) Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4
- Zhang WK, Wang YJ, Luo GZ, et al (2004) QTL mapping of ten agronomic traits on the soybean (*Glycine max* L. Merr.) genetic map and their association with EST markers. *Theor Appl Genet* 108:1131–1139. doi: 10.1007/s00122-003-1527-2
- Zhao C, Takeshima R, Zhu J, et al (2016) A recessive allele for delayed flowering at the soybean maturity locus E9 is a leaky allele of FT2a, a *FLOWERING LOCUS T* ortholog. *BMC Plant Biol* 1–15. doi: 10.1186/s12870-016-0704-9
- Zhao G, Wang J, Han Y, et al (2008) Identification of QTL underlying the resistance of soybean to pod borer, *Leguminivora glycinivorella* (Mats.) obraztsov, and correlations with plant, pod and seed traits. *Euphytica* 164:275–282. doi: 10.1007/s10681-008-9728-z
- Zhi-Wei Z, Zhe-Ming C, Jing-Song Z, Ting-Ting W (2007) Genetic structure analyses of different populations of grass carp (*Ctenopharyngodon idella*) using the TRAP technique. *Chinese J Agric Biotechnol* 4:27–32.
- Zhou P, Tan Y, He Y, et al (2003) Simultaneous improvement for four quality traits of Zhenshan 97, an elite parent of hybrid rice, by molecular marker-assisted selection. *Theor Appl Genet* 106:326–331.
- Zhu S, Walker DR, Boerma HR, et al (2006) Fine mapping of a major insect resistance QTL in soybean and its interaction with minor resistance QTLs. *Crop Sci* 46:1094–1099. doi: 10.2135/cropsci2005.06-0109
- Zhu S, Walker DR, Boerma HR, et al (2008) Effects of defoliating insect resistance QTLs and a cry1Ac transgene in soybean near-isogenic lines. *Theor Appl Genet* 116:455–463. doi: 10.1007/s00122-007-0682-2
- Zhu Y, Song Q, Hyten D, et al (2003) Single-nucleotide polymorphisms in soybean. *Genetics* 163:1123–1134.
- Zou F, Fine JP, Hu J, Lin DY (2004) An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait loci. *Genetics* 168:2307–2316.
- Zuo Q, Hou J, Zhou B, et al (2013) Identification of qtls for growth period traits in soybean using association analysis and linkage mapping. *Plant Breed* 132:317–323. doi: 10.1111/pbr.12060

APÊNDICES

APÊNDICE A. Marcadores SSR polimórficos entre os genitores IAC-100 e CD-215 selecionados para a genotipagem da população F₂. Nome do SSR, grupo de ligação (GL), número do cromossomo da soja e posição (cM) de acordo com o ConsensusMap 4.0 (Soybase)

SSR	GL	Cromossomo	Posição (cM) ¹	SSR	GL	Cromossomo	Posição (cM) ¹
Satt184	D1A	1	16,13	Satt253	H	12	63,3
Satt468	D1A	1	52,41	Satt222	H	12	63,75
AW781285	D1A	1	56,59	Satt586	F	13	33,70
Satt147	D1A	1	92,53	Satt030	F	13	34,39
Satt129	D1A	1	93,50	Satt146	F	13	37,14
Satt387	N	3	43,69	Satt114	F	13	43,90
Satt194	C1	4	13,37	Satt334	F	13	51,20
Satt195	C1	4	60,39	AW756935	F	13	98,17
Satt684	A1	5	5,85	Satt318	B2	14	60,77
Satt050	A1	5	42,62	Satt474	B2	14	63,36
AW734043	C2	6	5,85	Satt575	E	15	3,67
Satt432	C2	6	37,43	Satt699	E	15	71,85
Satt281	C2	6	38,90	Satt263	E	15	75,75
Satt557	C2	6	102,23	Satt185	E	15	76,27
Satt460	C2	6	106,99	Satt452	E	15	76,94
Satt307	C2	6	109,96	Satt045	E	15	77,29
Satt540	M	7	34,27	Satt716	E	15	79,62
Satt463	M	7	46,27	Satt132	J	16	43,69
Satt536	M	7	58,17	Satt622	J	16	46,11
Satt551	M	7	89,45	Satt582	D2	17	47,18
Satt480	A2	8	22,81	Satt514	D2	17	73,91
Sat_409	A2	8	32,94	Satt310	D2	17	90,28
Satt102	K	9	31,63	Satt386	D2	17	110,43
Satt001	K	9	43,24	Satt038	G	18	7,89
Sat_044	K	9	53,24	Satt309	G	18	10,10
Satt196	K	9	83,24	Satt324	G	18	35,43
Satt173	O	10	53,41	Satt012	G	18	63,44
Satt262	O	10	54,06	Satt191	G	18	89,37
Satt243	O	10	107,31	Satt523	L	19	25,57
Satt197	B1	11	49,07	Satt166	L	19	45,99
Satt453	B1	11	108,41	Satt156	L	19	48,86
Satt666	H	12	5,76	Satt229	L	19	78,26
Satt635	H	12	8,56	Satt571	I	20	14,97
Satt469	H	12	55,17	Satt239	I	20	29,61

¹Posição do marcador SSR em centimorgan (cM) obtida a partir do mapa da soja (ConsensusMap 4.0)

ANEXOS

ANEXO A. Combinações de *primers* utilizadas para os marcadores AFLP e TRAP. Os *primers* AFLP possuem três nucleotídeos seletivos e a denominação 'E' indica o *primer* correspondente ao iniciador EcoRI e 'M' indica o *primer* correspondente ao iniciador MseI. No caso dos *primers* dos marcadores TRAP, ARB representa os *primers* arbitrários obtidos na literatura e KB representam os *primers* fixos desenhados por Möller (2010). Na coluna tamanho dos alelos pode-se verificar o tamanho, em pares de bases (pb), obtido para os 67 alelos dos marcadores AFLP (denominados 'A') e para os 31 alelos dos marcadores TRAP (denominados 'T') identificados como polimórficos entre os genitores IAC-100 e CD-215.

Nome Marcador	Combinação <i>Primer</i>	Tamanho Alelo	Nome Marcador	Combinação <i>Primer</i>	Tamanho Alelo (pb)
A1		156	A50		650
A2		118	A51	E-AAG/M-CTG	646
A3		116	A52		210
A4	E-AAC/M-CAT	96	A53		450
A5		82	A54	E-ACT/M-CTG	270
A6		80	A55		140
A7		76	A56		132
A8		370	A57		340
A9		210	A58		250
A10		204	A59	E-AAG/M-CAG	220
A11	E-ACT/M-CAT	194	A60		140
A12		192	A61		110
A13		188	A62		86
A14		186	A63	E-AAG/M-CAA	380
A15		110	A64		250
A16		92	A65		140
A17		360	A66	E-ACC/M-CTG	300
A18		342	A67		110
A19	E-AAG/M-CTA	210	T1	ARB1/KB10	760
A20		186	T2		710
A21		184	T3		480
A22		182	T4	ARB1/KB16	150
A23		164	T5		120
A24	E-AAC/M-CTT	380	T6	ARB1/KB44	850
A25		376	T7		550
A26		372	T8		240
A27		530	T9		800
A28		520	T10		650
A29	E-ACC/M-CTC	204	T11	ARB1/KB53	648
A30		200	T12		500
A31		110	T13		240
A32		80	T14		142
A33		380	T15	ARB1/KB65	840
A34		340	T16		760
A35	E-ACA/M-CAT	302	T17		260
A36		210	T18	ARB2/KB66	222
A37		158	T19		220
A38		150	T20		156
A39	E-AAG/M-CTC	450	T21	ARB3/KB10	500
A40		152	T22		170
A41		352	T23	ARB3/KB40	130
A42	E-ACT/M-CTC	350	T24	ARB4/KB44	260
A43		104	T25	ARB5/KB44	340
A44		88	T26		290
A45		80	T27		360
A46		400	T28	ARB5/KB56	310
A47	E-AAC/M-CAG	390	T29		260
A48		350	T30		158
A49		98	T31		130

ANEXO B. Sequências dos *primers* fixos e aleatórios utilizados na amplificação dos marcadores TRAP (*Target Region Amplified Polymorphism*). A denominação em ambos os *primers* consiste no nome utilizado neste trabalho. *Primers* aleatórios: GenBank consiste no número da sequência que foi utilizada para a síntese do *primer* aleatório e a palavra-chave utilizada para a busca da sequência. *Primers* fixos: a identificação de cada *primer* e a referência onde foram obtidas as sequências dos mesmos.

Marcadores TRAP - *Primers* Aleatórios

Denominação	GenBank	Sequência do <i>primer</i>	Palavras-chave
KB10	BM094898	5'-GATGATTGGGTTCTGTGTCG-3'	<i>Glycine max</i> and <i>jasmonic acid</i>
KB16	GR860132	5'-ATCCTCTTCCCAAACAAAGC-3'	<i>Glycine max</i> and <i>sistemine</i>
KB40	GR860209	5'-TCTGTCGTTTGTTCCTTGATC-3'	<i>Glycine max</i> and <i>betaine</i>
KB44	BG882005	5'-ACACTGAATGGCTGGTGAAG-3'	<i>Glycine max</i> and <i>urease ubíqua</i>
KB53	BM093720	5'-CTCGTTTGGTTCCATTGTTGT-3'	<i>Glycine max</i> and <i>polyphenol oxidase</i>
KB56	BI095220	5'-CACTGAATGGCTTCTGAAGG-3'	<i>Glycine max</i> and <i>urease</i>
KB65	GR860149	5'-TGTGCCACTTAGGTCACCAG-3'	<i>Glycine max</i> and <i>WRKY27</i>
KB66	AB126934	5'-AAGCTGTAGCCCGTTCCA-3'	<i>Phytoalexins</i>

Marcadores TRAP - *Primers* Fixos

Denominação	Identificação	Sequência do <i>primer</i>	Referência
ARB1	Ga5	5'-GGAACCAAACACATGAAGA-3'	(Liu et al. 2005)
ARB2	Odd-26	5'-CTATCTCTCGGGACCAAAC-3'	(Zhi-Wei et al. 2007)
ARB3	Arbi 2	5'-GACTGCGTACGAATTGAC-3'	(Alwala et al. 2006)
ARB4	T04	5'-CGTAGTGATCGAATTCTG-3'	(Liu et al. 2005)
ARB5	T03	5'-CGTAGCGCGTCAATTATG-3'	(Liu et al. 2005)

ANEXO C. Estimativas da variância fenotípica ($\hat{\sigma}_F^2$), variância genotípica ($\hat{\sigma}_G^2$) e variância residual ($\hat{\sigma}_E^2$) das variáveis avaliadas, obtidas por Santos (2012). Caracteres de Resistência: **PEG**: período de granação, **RF**: retenção foliar, **NVP**: número de vagens por planta, **IPDV**: índice percentual de dano nas vagens, **NS**: número de sementes, **PSM**: peso de sementes manchadas, **PSB**: peso de sementes boas e **PCS**: peso de cem sementes. Caracteres Agronômicos: **NDF**: número de dias para o florescimento, **APF**: altura da planta no florescimento, **NDM**: número de dias para a maturidade, **APM**: altura da planta na maturidade, **AC**: acamamento, **VA**: valor agronômico e **PG**: produtividade de grãos.

Caracteres	$\hat{\sigma}_F^2$	$\hat{\sigma}_G^2$	$\hat{\sigma}_E^2$
Resistência			
PEG	4,69 (3,94 - 5,68)	3,36 (2,63 - 4,44)	4,00 (5,53 - 4,57)
RF	0,009 (0,007 - 0,011)	0,002 (0,001 - 0,008)	0,021 (0,019 - 0,024)
NVP	548,60 (460,67 - 664,48)	382,81 (297,86 - 510,31)	497,36 (438,86 - 569,46)
IPDV	0,003 (0,002 - 0,004)	0,001 (0,001 - 0,002)	0,004 (0,004 - 0,005)
NS	2016,42 (1693,25 - 2442,36)	1385,66 (1074,55 - 1855,34)	1892,29 (1669,71 - 2162,79)
PSM	2,96 (2,49 - 3,59)	1,19 (0,77 - 2,09)	5,31 (4,68 - 6,06)
PSB	23,88 (20,05 - 28,92)	16,58 (12,88 - 22,13)	21,91 (19,33 - 25,04)
PCS	1,97 (1,66 - 2,39)	1,48 (1,17 - 1,93)	1,48 (1,30 - 1,69)
Agronômicos			
NDF	3,36 (2,82 - 4,07)	2,80 (2,27 - 3,54)	1,67 (1,47 - 1,91)
APF	8,58 (7,20 - 10,39)	6,96 (5,61 - 8,85)	4,86 (4,32 - 5,61)
NDM	15,71 (13,20 - 19,03)	12,72 (10,26 - 16,20)	8,98 (7,92 - 10,26)
APM	21,86 (18,36 - 26,48)	19,68 (16,21 - 26,48)	6,54 (5,77 - 7,48)
AC	0,014 (0,012 - 0,017)	0,001 (0,000 - 1,184)	0,039 (0,034 - 0,044)
VA	0,007 (0,006 - 0,009)	0,004 (0,003 - 0,006)	0,008 (0,007 - 0,009)
PG	33,19 (27,87 - 40,20)	23,53 (18,39 - 31,20)	28,97 (25,57 - 33,12)