University of São Paulo
"Luiz de Queiroz" College of Agriculture

Revealing the genetic architecture of soybean resistance to stink bug complex

Emanoel Sanches Martins

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2021

Emanoel Sanches Martins
Bachelor in Agronomy

Revealing the genetic architecture of soybean resistance to stink bug complex

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof Dr **JOSÉ BALDIN PINHEIRO**

Thesis presented to obtain the degree of Doctor in Science. Area: Genetics and Plant Breeding

Piracicaba
2021

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

Martins, Emanoel Sanches

Revealing the genetic architecture of soybean resistance to stink bug complex / Emanoel Sanches Martins. -- versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

79 p.

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

1. *Glycine max* 2. Mapeamento de QTL 3. GWAS 4. QTL 5. Resistência a insetos 6. Mapeamento genético I. Título

DEDICATORY

To my beloved parents, Edivaldo and Marislei.

ACKNOWLEDGEMENTS

This work ends a great journey. The paths to achieve this was challenging, but the rewards and experiences throughout this journey have made it easier. During this journey, I had the support of several people and institutions that facilitated the paths to reach this end, which I would like to thank.

To God for giving me the strength to keep going.

To my parents, Edivaldo and Marislei, my brothers Edicarlos, Edimari, Emilly, to encourage and to support me, always being an escape valve, I love you.

The “Luiz de Queiroz” College of Agriculture, University of São Paulo (ESALQ/USP), especially the Department of Genetics and the Graduate Program in Genetics and Plant Breeding, for the opportunity offered and all the structure, so that I could develop professionally and academically.

To my advisor and friend Professor José Baldin Pinheiro for the countless opportunities offered in almost four years. I am grateful for his teachings, leadership and guidance dedicated to me and this work, always focused on preparing me professionally. As well as the example of a person who has always shown himself, I will always remember his advice.

To Professor Edward Alexander Lipka and the entire Lipka Lab, for indisputably collaborating and adding value to this journey and this work. The time I spent in his laboratory was of great value, thanks for the learning opportunity and the experiences gained in his laboratory.

To Professor Antônio Augusto Franco Garcia and Roberto Fristche-Neto. Two professors of whom I had the honor to meet and discuss various aspects of science, I am grateful for all the academic and life advice they provided in our discussions.

To Doctor Kaio Olimpio, a great friend, for support this work and for all advices during this journey. I am gratefully to work and to learn with you.

To the entire Laboratory of Genetic Diversity and Plant Breeding, in the last four years this laboratory had become my home and I am sure that the support of all of you and all the moments lived in the laboratory collaborated towards to finalize this work. Here I also extend my thanks to all the ESALQ employees who helped us in our work, especially field employees.

To the Genetics and Plant Breeding Group “Professor Roland Vencovsky” (GVENCK) for the opportunity to join an excellent team of professionals and develop new paths in my professional career.

To my friends Carolina, Dyrson, Elesandro, Gabriel, Getúlio, Gleicy, Jessica, João Vitor, Maiara, Patrícia, Pedro, Rafael and Willian. I am especially grateful for the family that we became, being in our spare time, people who were always willing to provide pleasant moments of joy, relaxation and leisure.

My girlfriend Thaise, a great woman, who became a great supporter in this final journey. I am very grateful for all your patience and companionship during this time.

I gratefully acknowledge the financial support of following institutions Fundação de Amparo à Pesquisa do Estado de São Paulo, grant numbers 2019/13010-0, 2018/12399-9, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, grant number 140586/2018-0 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

To everyone who in any way contributed to this work.

SUMMARY

RESUMO

Revelando a arquitetura genética da resistência da soja ao complexo de percevejo

Danos causados pelo complexo de percevejos na soja têm ultrapassado a cifra de milhões de dólares anuais. Com extensas áreas de cultivo, o prejuízo devido a estes insetos tem aumentado, reduzindo os ganhos na produção. Dentre as alternativas de controle, as cultivares resistentes ao complexo de percevejo podem ser uma alternativa ao uso de inseticidas. Sabe-se que tal resistência é de natureza quantitativa e, portanto, controlada por diversos genes. Desta maneira, estratégias que visam esclarecer o controle destas pragas se tornam preponderantes para o futuro da soja. Mediante tal fato, o objetivo deste trabalho consistiu em identificar e elucidar a arquitetura genética da resistência da soja ao complexo de percevejos. No primeiro estudo, buscamos mapear regiões de resistência e elucidar as interações entre QTLs e ambientes envolvidas na resistência ao complexo de percevejos da soja, por meio de uma população de linhagens recombinantes endogâmicas (RILs). Para tanto, foi utilizado uma população de 256 RILs desenvolvida a partir do cruzamento entre os genótipos IAC-100 (resistente) e CD-215 (suscetível). Os experimentos foram realizados na área experimental da ESALQ no município de Piracicaba, SP, durante as safras 2012/2013, 2013/2014, 2014/2015, 2015/2016, 2016/2017, 2017/2018 e 2018/2019, em um delineamento experimental de alfa-látice 10 x 26, com três repetições. Foram avaliadas 9 características associados a resistência ao complexo de percevejos e ao desempenho agronômico. Para genotipagem das RILs, foi utilizada a técnica de genotipagem por sequenciamento (Genotyping-by-Sequencing). Posteriormente, com os dados obtidos e as análises realizadas, encontramos um total de sete QTLs significativos sendo 5 deles considerados como QTLs estáveis e com potencial para uso na seleção assistida por marcadores. Adicionalmente, neste primeiro capítulo desenvolvemos um estudo de epistasia utilizando o software SPAEML, onde concluímos que o tamanho da população estudada (n=256) é muito pequena para quantificar efeitos epistáticos. No segundo estudo, desenvolvemos um trabalho com objetivo de validar as regiões encontradas pelo mapeamento de QTL por meio de um estudo de associação genômica ampla. Foi avaliado um painel de melhoramento de soja composto de 299 linhagens, obtido de um cruzamento multiparental. Essa população foi avalida durante os anos de 2018/2019 e 2019/2020, na área experimental da ESALQ em um delineamento de alfa-látice 16x19 com três repetições. Foram avaliados caracteres que tiveram QTLs significativos no mapeamento. Os dados genotípicos foram obtidos via genotipagem por sequenciamento. Com os dados obtidos, obtivemos as médias ajustadas via modelos mistos e posteriormente realizamos GWAS para validação das regiões de resistência ao complexo de percevejos. Foi encontrado um total de 22 QTNs sendo três regiões validando os QTLs encontrados no mapeamento anterior. Para os nossos estudos, os cromossomos 1, 6 e 15 aparecem como regiões com possíveis candidatos genes de resistência e com grande potencial de auxiliarem os programas de melhoramento na tomada de decisões quanto a busca de cultivares resistentes.

Palavras-chave: *Glycine max*, GWAS, Mapeamento de QTLs, Resistência a insetos, Validação de QTLs

ABSTRACT

Revealing the genetic architecture of soybean resistance to the stink bug complex

Damage caused by the stink bug complex in soybean has exceeded the figure of millions of dollars annually. Cultivating extensive areas, the damage due to these insects has increased, reducing the gains in production. Among the insects' control, resistant cultivars to the stink bug complex may be an alternative to the use of insecticides. It is known that this resistance has a quantitative nature and is thus controlled by several genes. In this way, strategies that aim to clarify the control of these insects become overpowering for the future of soybeans. To this end, the objective of this project was to identify and elucidate the genetic architecture of soybean resistance to the stink bug complex. In the first study, we seek to map regions of resistance and elucidate the interactions between QTLs and environments involved in resistance to the stink bug complex on soybean, in a population of recombinant inbred lines (RILs). For this purpose, a population of 256 RILs were used, developed from the cross between IAC-100 (resistant) and CD-215 (susceptible). The experiments were carried out at the ESALQ experimental area in the city of Piracicaba, SP, during the seasons 2012/2013, 2013/2014, 2014/2015, 2015/2016, 2016/2017, 2017/2018, and 2018/2019. We use an alpha-lattice 10 x 26, with three replications as the experimental design. Nine traits associated with resistance to the stink bug complex and agronomic performance were evaluated. For genotyping of RILs, the genotyping technique by sequencing (Genotyping-by-Sequencing) was used. Subsequently, with the data obtained and the analyzes carried out, we found a total of seven significant QTLs, 5 of which were considered as stable QTLs and with potential for use in marker assisted selection. Additionally, in this first study, we developed an epistasis study using the SPAEML software, in which we conclude that the size of the studied population (n = 256) is too small to quantify epistatic effects. In the second study, we developed a work aimed to validate the regions found by the QTL mapping through a genomic wide association study. A soybean breeding panel composed of 299 lines, obtained from a multiparental crossing, was evaluated. This population was evaluated during the years 2018/2019 and 2019/2020, in the experimental area of ESALQ in a 16x19 alpha-lattice design with three replications. We evaluated traits that had significant QTLs in the QTL mapping. Genotypic data were obtained by genotyping by sequencing. With the data obtained, we calculated the means adjusted through mixed models and subsequently performed the GWAS to validate the regions of resistance to the stinkbug complex. A total of 22 QTNs were found, with three regions validating the QTLs found in the previous mapping. For our studies, chromosomes 1, 6, and 15 appear as regions with possible candidate resistance genes and with great potential to assist breeding programs to decide for resistant cultivars.

Keyword: *Glycine max*, GWAS, QTL mapping, Resistance to insects, QTL validation

LIST OF FIGURES

LIST OF TABLES

1. INTRODUÇÃO

A soja [*Glycine max* (L) Merrill] é uma planta autógama (2n=40 cromossomos), sendo, dentre as leguminosas, a espécie mais cultivada em todo o mundo (Da Fonseca Santos et al. 2018). Possui centro de origem na região leste da China, com seu centro primário localizado no centro-sul deste país, e o centro secundário, local onde a soja foi domesticada, localizado na região da Manchúria (Hymowitz 1970). Apesar disso, algumas evidências moleculares corroboram com a tese de que a domesticação da soja foi realizada no sul da China (Ding et al. 2008). Desta maneira, é coerente afirmar que a domesticação da soja ocorreu nas regiões Norte e Sul da China (Pratap et al. 2012).

No Brasil, a cultura da soja é responsável por quase 60% da área cultivada no país (CONAB 2021). Porém, com essa extensa área de cultivo, problemas como o ataque de insetos e fungos se tornam recorrentes para essa cultura. Com relação aos insetos, a soja é atacada por algumas espécies desfolhadoras até sugadoras. Entretanto, com o avanço da transgenia para insetos desfolhadores, o ataque de insetos sugadores como percevejos tem sido um dos maiores problemas da cultura nos últimos anos (Vivan and Degrande 2011; Guedes et al. 2012).

Dentre os percevejos, os que mais se destacam são três pentatomídeos fitófagos: *Euschistus heros* (Fabricius), ou percevejo marrom, que ataca a soja geralmente de Novembro a Abril, sendo que, no restante do ano entra em estado de dormência; *Nezara viridula* (Linnaeus) que, ao contrário do percevejo marrom, pode ser encontrado o ano todo, utilizando outras plantas como hospedeiras na entressafra da soja; *Piezodorus guildinii* (Westwood) ou percevejo verde-pequeno, que também utiliza outras plantas como hospedeiras na entressafra da soja, e é considerado o mais prejudicial, causando maiores danos à qualidade dos grãos e maior retenção foliar. Esses três percevejos compõem o chamado “complexo de percevejos” da soja (Corrêa-Ferreira et al. 2009).

Os percevejos podem iniciar sua colonização na soja a partir da fase de pré-florescimento ou florescimento (R1 a R2), com aumento populacional até o período final de granação (R6), quando atingem seu nível populacional máximo (Panizzi et al. 2012). Durante o período reprodutivo, os estádios R5 (início de enchimento dos grãos) e R6 são considerados os estádios mais críticos, sendo a soja mais

suscetível ao ataque, pois os percevejos se alimentam dos grãos (Corrêa-Ferreira and Panizzi 1999). O estádio R5 apresenta maior intensidade de danos, quando comparado com os outros estádios de infestação, e apesar dos percevejos adultos serem mais visíveis nas plantas, as ninfas também têm participação nos danos causados, semelhante ou até de maior intensidade que os adultos (Corrêa-Ferreira et al. 2009).

Buscando diminuir os efeitos do complexo de percevejos nas lavouras, diversos métodos são utilizados, dentre estes os controles químicos e biológicos, onde as aplicações de controle são baseadas no nível da população de percevejos e na fase de desenvolvimento da cultura (Corrêa-Ferreira and de Azevedo 2002). Entretanto, tais métodos podem representar maiores despesas para a produção da soja, sendo que normalmente o controle químico dos percevejos possui custos elevados. Em um comparativo, o uso de inseticidas aumentou de 6,04 % para 13,8% do custo de produção de soja no país da safra de 2009/2010 para a de 2014/2015 (Lantmann 2014). Sendo que, por safra, o controle de percevejos exige até seis milhões de litros de inseticidas (Ereno 2011).

Além disso, segundo Corrêa-Ferreira et al., (2009), o controle de percevejos por meio de inseticidas tem sido pouco eficiente, devido à resistência dos percevejos aos produtos e também à diminuição na eficácia dos mesmos em populações suscetíveis. Por quase quatro décadas, meados de 1960 a meados de 2010, os inseticidas pertencentes ao grupo dos organofosforados e ciclodieno (Endosulfam) eram os mais utilizados, sendo que, as repetidas utilizações dessas moléculas favoreceram a evolução da resistência do percevejo marrom a estes produtos, principalmente produtos à base de endosulfam (Sosa-Gómez and Omoto 2012). Atualmente, os ingredientes ativos utilizados têm sido do grupo dos piretróides e neonicotinóides, sendo que alguns produtos à base de organofoforados e endosulfam estão sendo proibidos, pois são nocivos ao ambiente e não possuem seletividade a insetos. A falta de novas moléculas pode favorecer o surgimento de novas populações resistentes, acrescentando-se ainda que, na prática, os dois grupos piretróides e neonicotinóides estão sendo resumidos a apenas um, considerando-se que suas misturas são mais utilizadas que os produtos isoladamente.

O uso de inseticidas também é relatado causando prejuízos ambientais, quando usado de forma incorreta, podendo ter também um potencial nocivo à saúde humana (Belo et al. 2012). Nos últimos anos, os malefícios causados pelos inseticidas a abelhas têm tido um grande destaque, que são consideradas uma importante ferramenta na produção agrícola (Potts et al. 2010). Por exemplo, um dos inseticidas utilizados nos últimos anos foi o Imidacloprido, pertencente ao grupo dos neonicotinóides, com um uso considerável nas áreas de produção. Diversos trabalhos relatam os danos causados por neonicotinóides a abelhas melíferas. Estes, em doses e concentrações subletais, apesar de algumas vezes não levarem à morte das abelhas, alteram o comportamento e funcionamento da colônia (Lambin et al. 2001; El Hassani et al. 2008; Teeters et al. 2012). Além disso, podem afetar a aprendizagem e memória destes insetos, prejudicando as respostas aos estímulos alimentícios (Bortolotti et al. 2003; Decourtye et al. 2004a, b).

Assim, uma das alternativas para os programas de manejo integrado de pragas (MIP), é a utilização de cultivares resistentes à percevejos, podendo substituir ou ser utilizada de forma integrada com o controle químico (McPherson et al. 2007). A utilização de plantas resistentes é considerada ideal devido a diversos fatores. Inicialmente tem-se a diminuição no custo de produção e maior segurança para o produtor e o consumidor; não poluem e nem causam distúrbios ecológicos; permitem que agricultores de baixa renda possam cultivar soja, pois reduzem a aplicação de defensivos; e por fim essas cultivares não perdem sua eficácia quando considerados baixos níveis populacionais dos insetos (Rossetto et al. 1981; Lopes et al. 1997).

A resistência da soja a percevejo é considerada um caráter poligênico, de natureza quantitativa (Godoi and Pinheiro 2009). Desta maneira, trabalhos que elucidem a arquitetura genética da resistência da soja ao complexo de percevejo beneficiariam toda a cadeia relacionada à cultura. Alguns trabalhos com esse objetivo já foram realizados dentro do Laboratório de Diversidade Genética e Melhoramento da Escola Superior de Agricultura “Luiz de Queiroz”. Santos (2012) e Moller (2017), desenvolveram estudos buscando identificar QTLs associados a resistência ao complexo de percevejo. No trabalho de Santos (2012), constatou-se uma alta correlação genética entre as características associadas à resistência aos percevejos, identificando 14 QTLs, sendo seis relacionados a características

agronômicas e oito para resistência ao complexo de percevejos. Por sua vez, no trabalho de Moller (2017), foram encontrados 60 QTLs, sendo 29 QTLs associados às características de resistência aos percevejos e 31 QTLs relacionados às características agronômicas avaliadas, pelo mapeamento por intervalo múltiplo. Quando trabalhando com abordagem multivariada, foi possível identificar 20 posições genômicas, afetando as diferentes características analisadas. Entretanto, nesses estudos, não foram levados em consideração a estabilidade dos QTLs nos múltiplos ambientes avaliados.

Este nosso trabalho foi realizado com o intuito de ilustrar melhor o desafio por trás do estudo da resistência genética ao complexo de percevejos e está dividido em dois capítulos. O primeiro consta do Mapeamento de QTLs utilizando uma população de linhagens endogâmicas recombinantes (RILs), avaliados ao longo de sete safras, na cidade de Piracicaba-SP. Neste estudo, pôde-se verificar a dificuldade dos melhoristas em trabalhar com caracteres de natureza quantitativa, onde geralmente o controle é feito por um grande número de genes de pequenos efeitos, corroborando com os achados deste trabalho. No segundo capítulo, utilizamos do estudo de associação genômica ampla (GWAS) para validar as regiões encontradas no Mapeamento de QTLs. Neste trabalho, foi utilizado uma população com *background* diferente da população de RILs, composto por linhagens advindas de cruzamento multiparental. Esses estudos permitiram validar algumas regiões no genoma da soja e alguns possíveis genes candidatos também foram discutidos ao longo do trabalho.

Referências

Belo MS da SP, Pignati W, Dores EFG de C, et al (2012) Uso de agrotóxicos na produção de soja do estado do Mato Grosso: um estudo preliminar de riscos ocupacionais e ambientais. Rev Bras Saúde Ocup 37:78–88. https://doi.org/10.1590/S0303-76572012000100011

Bortolotti L, Montanari R, Marcelino J, et al (2003) Effects of sub-lethal imidacloprid doses on the homing rate and foraging activity of honey bees. Bull Insectology 56:63–67

CONAB (2021) Acompanhamento da Safra Brasileira. Cia Nac Abast Acompan da Safra Bras 7:1–89

Corrêa-Ferreira BS, de 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

Corrêa-Ferreira BS, Panizzi AR (1999) Percevejos da Soja e seu Manejo. Circ técnica/Embrapa CNPSo, ISSN 0100-6703 n24 45

Da Fonseca Santos M, Möller M, Clough SJ, Pinheiro JB (2018) Heritability of agronomic traits correlated with reduced stink bug damage in an F2:3 soybean population derived from IAC-100. J Crop Improv 32:1–18. https://doi.org/10.1080/15427528.2017.1370404

Decourtye A, Armengaud C, Renou M, et al (2004a) Imidacloprid impairs memory and brain metabolism in the honeybee (Apis mellifera L.). Pestic Biochem Physiol 78:83–92. https://doi.org/10.1016/J.PESTBP.2003.10.001

Decourtye A, Devillers J, Cluzeau S, et al (2004b) Effects of imidacloprid and deltamethrin on associative learning in honeybees under semi-field and laboratory conditions. Ecotoxicol Environ Saf 57:410–419. https://doi.org/10.1016/J.ECOENV.2003.08.001

Ding Y, Zhao T, Gai J (2008) Genetic diversity and ecological differentiation of Chinese annual wild soybean (Glycine soja). Biodivers Sci 16:133. https://doi.org/10.3724/sp.j.1003.2008.07286

El Hassani AK, Dacher M, Gary V, et al (2008) Effects of Sublethal Doses of Acetamiprid and Thiamethoxam on the Behavior of the Honeybee (Apis mellifera). Arch Environ Contam Toxicol 54:653–661. https://doi.org/10.1007/s00244-007-9071-8

ERENO D (2011) Armadilhas biológicas: percevejo-marrom da soja pode ser monitorado com feromônio sintético desenvolvido pela Embrapa. In: Pesqui. FAPESP. https://www.embrapa.br/documents/1355163/2088709/1111_21_clipping_revistaFapesp.pdf/e21cc4d1-85fe-4105-9092-19388a099d9e. Accessed 31 Mar 2018

Godoi CRC de, Pinheiro JB (2009) Genetic parameters and selection strategies for soybean genotypes resistant to the stink bug-complex. Genet Mol Biol 32:328–336. https://doi.org/10.1590/S1415-47572009000200020

GUEDES JVC, ARNEMANN JA, STÜRMER GR, et al (2012) Percevejos da soja: novos cenários, novo manejo. Rev. Plantio Direto 12:28–34

Hymowitz T (1970) On the domestication of the soybean. Econ Bot 24:408–421. https://doi.org/10.1007/BF02860745

Lambin M, Armengaud C, Raymond S, Gauthier M (2001) Imidacloprid-induced facilitation of the proboscis extension reflex habituation in the honeybee. Arch Insect Biochem Physiol 48:129–134. https://doi.org/10.1002/arch.1065

Lantmann Á (2014) Gastos com defensivos | Soja Brasil - Soja Brasil. http://www.projetosojabrasil.com.br/artigo-gastos-com-defensivos/. Accessed 31 Mar 2018

Lopes ECA, Destro D, Montalván R, et al (1997) Genetic gain and correlations among traits for stink bug resistance in soybeans. Euphytica 97:161–166. https://doi.org/10.1023/A:1003042520035

McPherson RM, Buss GR, Roberts PM (2007) Assessing Stink Bug Resistance in Soybean Breeding Lines Containing Genes from Germplasm IAC-100. J Econ Entomol 100:1456–1463. https://doi.org/10.1093/jee/100.4.1456

Moller M (2017) Indentificação de QTLs em soja associados a resistência aos percevejos e a caracteres agronômicos utilizando a abordagem de mapeamento multivariado. Escola Superior de Agricultura Luiz de Queiroz

PANIZZI AR, BUENO AF, SILVA FAC (2012) Insetos que atacam vagens. In: Soja: manejo integrado de insetos e outros artrópodes-praga, EMBRAPA. Brasilia, pp 335–420

Potts SG, Biesmeijer JC, Kremen C, et al (2010) Global pollinator declines: trends, impacts and drivers. Trends Ecol Evol 25:345–353. https://doi.org/10.1016/J.TREE.2010.01.007

Pratap A, Gupta SK, Kumar J, Solanki RK (2012) Soybean. In: Technological Innovations in Major World Oil Crops. Springer New York, pp 293–321

Rossetto CJ, Lourenção AL, Igue T, Miranda MAC De (1981) Picadas de alimentação de Nezara viridula em cultivares e linhagens de soja de diferentes

graus de suscetibilidade. Bragantia 40:109–114. https://doi.org/10.1590/S0006-87051981000100010

Santos M da F (2012) Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevejos Piracicaba 2012. Escola Superior de Agricultura Luiz de Queiroz

Sosa-Gómez DR, Omoto C (2012) Resistência a Inseticidas e outros Angentes de Controle em Artrópodes Associados à Cultura da Soja. In: Soja-Manejo Integrado de Insetos e outros Artrópodes-praga. pp 673–724

Teeters BS, Johnson RM, Ellis MD, Siegfried BD (2012) Using video-tracking to assess sublethal effects of pesticides on honey bees (Apis mellifera L.). Environ Toxicol Chem 31:1349–1354. https://doi.org/10.1002/etc.1830

VIVAN LM, DEGRANDE PE (2011) Pragas da soja. Fundação MT Bol Pesqui Soja 11:239–297

2. UNCOVERING THE GENETIC ARCHITECTURE OF STINK BUG COMPLEX RESISTANCE IN SOYBEAN

ABSTRACT

Mitigating the effects of the stink bug on soybean fields has been one of the major bottlenecks for breeding and chemical companies. Nowadays, the development of resistant cultivars, which may complement or even be an alternative to the use of insecticides, appears as a promising strategy to reduce yield losses, as well as mitigate the environmental impact. For quantitative traits as stink bugs resistance, the development of resistant cultivars is a challenge faced by the breeders. Thus, the better comprehension of the genetic architecture of this trait, as well as the identification of QTLs underlying the resistance, can aid breeding efforts and assist with selection in the breeding programs. Therefore, this project aimed to elucidate the genetic architecture of the soybean resistance to the stink bug complex through QTL analysis. We obtained data from one RIL population (n = 256) grown during seven field seasons at ESALQ in Piracicaba, Brazil. The experimental design was an alpha-lattice of 10x26, with three replicates each. We evaluated nine traits related to resistance to stink bugs in soybean and agronomic traits. A total of 2,037 SNPs obtained from Genotyping By Sequencing (GBS) were used for the statistical analysis. Best linear unbiased estimators (BLUEs) of the traits from mixed models fitted across all seven field seasons were used as the response variable in QTL by environment analysis. The Li and Ji methods were used to adjust for the genome-wide type I error rate at $\alpha = 0.05$. Additionally, these BLUEs were used in a stepwise model selection procedure to search for an epistatic QTL using the SPAEML software. Among the evaluated traits, we observed seven significant additive QTLs, most of them in traits related to the resistance to the stink bug complex. We also confirmed via simulation study that the sample size of this RIL population is too small to rigorously quantify epistatic associations. The additive QTLs found in this work will be studied in future research to uncover the genetic architecture of stink bug complex resistance in soybean.

Keywords: Genetic resistance; Linkage mapping; Resistance to insects; Statistical analysis;

2.1. Introduction

Soybean [*Glycine max* (L) Merrill] is one of the most cultivated species with great importance in several agroindustry sectors. According to Conab (2021),

the Brazilian harvested area increased 4.2% from the previous crop season (2019/2020), totaling 38.5 million hectares, breaking the record production reaching 135.4 million tons, an increase of 8.5% compared to the last season. However, with the advances in the cultivated area, and due to the high incidence of monoculture systems, the soybean crop is currently very vulnerable to several pathogens and insect pests, leading to significant losses in production (Boerma and Walker 2005). Among the insects that attack soybeans, stink bugs are those that cause the most considerable economic losses, which can damage grain quality and can also compromise crop productivity. In an estimation carried out in Mato Grosso, the state with the highest production of soybean in Brazil, stink bugs cause a reduction of up to 30% in grain yield and higher than 50% for seed germination losses (VIVAN and DEGRANDE 2011). According to Guedes et al. (2012), the losses caused by stink bugs can reach up to 125 kg ha^{-1}, when considered one stink bug per linear meter.

One of the alternatives to control the stink bug complex is the use of resistant cultivars, which can replace chemical control or can be used in combination with them (McPherson et al. 2007a). The use of resistant plants is ideal due to several factors, such as the decrease in the cost of production, the safety for producers and consumers, no associated pollution or ecological disturbances, and their ability to allow low-income farmers to grow soybeans. This results in a reduction in the application of pesticides, and finally, resistant cultivars do not lose their effectiveness in low levels of insect populations (Rossetto et al. 1981; Lopes et al. 1997; Pinheiro et al. 2005).

The resistance of soybean to stink bugs is a polygenic trait, presenting a quantitative nature (Godoi and Pinheiro 2009). In this sense, QTL (Quantitative Trait Loci) mapping becomes a valuable tool, helping breeding programs with the identification and selection of resistant plants through marker-assisted selection (MAS). The use of MAS is especially useful because phenotyping is laborious and difficult to apply in commercial breeding programs.

Therefore, the identification of QTLs involved in the resistance of soybean to stink bug may contribute to the greater elucidation of the genetic architecture of this trait. In soybean, previous studies found QTLs for resistance to chewing insects (Rector et al. 1998, 1999, 2000; Terry et al. 2000; Narvel et al. 2001; Walker et al. 2004; Boerma and Walker 2005; Guo et al. 2005, 2006; Zhu et al. 2006, 2008; Zhao

et al. 2008; Yesudas et al. 2010) however, there is a lack of information related to QTLs involved in resistance to stink bugs.

Routine breeding programs aim to evaluate their genotypes in different locations and years, taking into account their stability in multiple environments. Thus, data from multiple environments, multi-environment trials (METs) are critical for studying genotype by environment interaction (G × E) and make it possible to elucidate QTL by environment interactions (QTL × E) to find stable QTLs (Jiang and Zeng 1995). Studies of this nature are usually more informative and present essential elements to the breeders, supporting the MAS (Dekkers and Hospital 2002; Boer et al. 2007; Malosetti et al. 2008).

For the studies involving MET, the use of statistics through linear mixed models becomes a powerful tool, allowing flexibility to the variance-covariance structures when considering random effects (SMITH et al. 2005; Smith et al. 2007; Malosetti et al. 2008; Pastina et al. 2012; Margarido et al. 2015). A series of studies reported that this approach is best for the analysis of G × E (Piepho 1997; Cullis et al. 1998; Smith et al. 2001, 2007; Smith et al. 2005) and QTL × E (Piepho 2000; Verbyla et al. 2003; Malosetti et al. 2004, 2008; Malosetti et al. 2006; Boer et al. 2007). However, there are no studies involving QTL x E of soybean resistance to the stink bug complex. Because of the great importance of this crop, this research aims to elucidate the interactions between QTLs and environments, identifying stable genomic regions that uncover the resistance to the stink bug complex, and perform a epistatic study of the QTLs found. Furthermore, the search for putatively epistatic loci controlling stink bug resistance could potentially contribute to a better understanding of its genetic architecture and identify more targets for marker-assisted selection, then we performed also an epistatic study to elucidate this resistance.

2.2. Materials and Methods

2.2.1. Population

For the QTL mapping, we evaluated 256 recombinant inbred lines (RILs) that compose part of the germplasm of the Laboratory of Genetic Diversity and Breeding at ESALQ / USP. This population was developed from a biparental cross between soybean genotypes IAC-100 (resistant to the stink bug complex) and CD-215

(susceptible). The parents and commercial cultivars AS3730 and Produza were included as controls, totaling 260 evaluated genotypes.

2.2.2. Phenotypic Date

For the QTL mapping, we used the data collected from 2012/2013, 2013/2014, 2014/2015, 2015/206, 2016/2017, 2017/2018, and 2018/2019 crop seasons at Piracicaba-SP. The term “environment” was used to described the crop seasons evaluated at this research. The experimental design 10 x 26 alpha-lattice with three replicates, where the plot consisted of four rows of five meters in length, spaced 0.5 m between rows, and containing 18 seeds per linear meter was used.

No chemical control was applied to control insects to allow the natural infestation of stink bugs. To monitor the prevalence of these stink bugs in the area, the beat cloth method was applied from the flowering to the full maturation (Panizzi et al. 1977).

The evaluated traits are described in Table 1. The tolerance trait was calculated as $TOL = \left(1 - \left(\frac{GY - WHS}{GY}\right)\right) x\ 100$. To compute the weight of healthy seeds (WHS), the seed that was not damaged by stink bugs was considered. After the harvest, the seeds were passed through a spiral, where the empty, green, and malformed grains were separated by gravity and centrifugal forces. After this processing, the data were taken in kg ha^{-1} (Rocha et al. 2014).

Table 1. Traits evaluated to agronomic performance (AP) and stink bug resistance (SBR) at the RILs population on soybean.

Trait	Abbreviations	Measurement	Type
Number of days for flowering	NDF	days	AP
Number of days to maturity	NDM	days	SBR
Grain filling period	GFP	days	SBR
Plant height at maturity	PHM	cm	AP
Lodging	L	score	AP
Grain yield	GY	kg ha^{-1}	AP
One hundred seeds weight	HSW	grams	SBR
Weight of healthy seeds	WHS	kg ha^{-1}	SBR
Tolerance	TOL	%	SBR

2.2.3. Phenotypic Analysis

A model that best explains the genetic effect of the phenotypic data was adjusted using the mixed model approach, considering the stage-wise approach. The first model included the effect of each genotype in each environment as follows:

$$y_{irs} = \mu + P_s + B_{r/s} + G_i + \underline{\varepsilon}_{irs} \qquad \text{(model 1)}$$

where: y_{irs} is the phenotype of genotype *i*, in block r, of the replicate *s*, in environment *j*; μ is the intercept; P_s is the fixed effect of the replicate s, where s = 1, 2, and 3; $B_{r/s}$ is the random effect of block *r* within the replicate s, where $B_{r/s}$ ~ N(0, σ^2_{rs})).; G_i is the genetic fixed effect of genotype *i*; ε_{irs} is the residual. Assume that the residuals are independent and identically distributed (ε_i ~ N(0, σ^2_e)).

After we had the adjusted means to each environment, we proceeded to QTL mapping and the variance/covariance structures (VCOV) selection. We compared different variance/covariance structures (VCOV), to explain the genetic effect of each genotype by the environment interaction. To compare the VCOV matrix we used the model as follow:

$$y_{ij} = \mu + L_j + \underline{G}_{ij} + \underline{\varepsilon}_{ij} \qquad \text{(model 2)}$$

where: y_{ij} is the phenotype of genotype *i* in environment *j*; μ is the intercept; L_j is the environment fixed effect; G_{ij} is the genetic effect of genotype *i* in environment *j*, where G_{ij} ~ N(0,$\sigma^2 G$), where **G** is the variance-covariance structure of genetic within environment effect; ε_{ij} is the residual. Assume that the residuals are independent and identically distributed (ε_i ~ N(0, σ^2_e)).

2.2.4. Genotyping

We extracted DNA from the young leaf DNA of RILs population, using the CTAB method (hexadecyltrimethylammonium bromide) (Doyle and Doyle 1990). In the mapping population, we included the IAC-100 and CD-215 parents, for genotyping. In summary, the DNA extracted was digested by HindIII enzyme using the genotyping by sequencing (GBS), following the protocol described by Elshire et al. (2011). We conducted this step at the Roy J. Carver Biotechnology Center of the University of Illinois Urbana-Champaign. The reads were aligned to the William 82, Gmax_275_Wm82. a2. V4 genome. The SNP calling was performed using the

TASSEL bioinformatic pipeline, setting as parameter: minor allele frequency (MAF) ≥ 0.05; minimum coefficient of inbreeding of 0.9 and call rate of < 0.8. The final data was with a total 2037 SNPs, which was used to perform the QTL mapping.

2.2.5. Linkage Map

To construct the Genetic Map, estimate the distance, and find the best markers order in their respective linkage groups (LG), the OneMap package in R was used (MARGARIDO et al. 2007). The map considers the genomic information available from SoyBase (http://www.soybase.org), using the markers previously mapped within the 20 LGs of soybean (Grant et al. 2010). The markers were found and initially designated to their respective LGs, according to the database, and finally ordered within each LG. We calculated the distances by the mapping function of Kosambi, (1944), expressed in centiMorgans (cM).

2.2.6. QTL Mapping

First, we adjusted the means by model 1. With the adjusted means to each environment, we performed a QTL mapping (second step). To detect significant QTLs, we used the $\alpha = 0.05$ "logarithm of the odds" (LOD) threshold, using the method proposed by Li and Ji (2005), a Bonferroni correction based on the effective number of independent tests. The model used was equivalent to the multiple interval mapping (MIM). We performed these analyses in the Genstat software (VSN International 2017). We can express the QTL x environment model by:

$$y_{ij} = \mu_j + \sum_{q \in Q} \alpha_{jq} x_{iq} + \varepsilon_{ij} + e_{ij} \qquad \text{(model 3)}$$

where: y_{ij} is the environment *j* mean for genotype *I,* $\mu_j$is the environment *j* intercept, Q is the set of QTLs, *q = 1*, …, *Q,* $\alpha_{jq}$is the effect of QTL *q* for trait *j,* $x_{iq}$is the genetic predictor of QTL *q* for genotype *I,* ε_{ij} is the genetic residual of trait *j* for genotype *i* (or residual if unit errors are omitted), e_{ij} is the unit error of trait *j* for genotype *i.*

2.2.7. Stepwise Model Selection Procedure

Additionally, we performed a stepwise model selection procedure to identify non-additive sources of genetic variation that can contribute to the traits evaluated within each environment. To do this, we used the stepwise procedure for constructing an additive and epistatic multi-locus model (SPAEML) (Chen et al. 2019). This approach implemented to identify the best multi-locus linear model that combines additive and epistatic effects

$$y_i = \mu + \sum_{j \in I} \beta_j x_{ij} + \sum_{(u,v) \in U} \gamma_{uv} x_{iu} x_{iv} + \varepsilon_i \qquad \text{(model 4)}$$

where: y_i is the observed trait value of the i^{th} genotype; μ is the grand mean; β_j is the additive effect of the j^{th} marker; x_{ij} is the observed genotype of the j^{th} marker of the i^{th} genotype numerically coded as, e.g., 0 for *aa* and 2 for *AA*; γ_{uv} is the two-way epistatic term between the u^{th} and the v^{th} marker; x_{iu} and x_{iv} are the observed genotypes for the u^{th} and the v^{th} markers, both of which are numerically coded in the same manner as x_{ij}; I is a subset of the m markers with additive effects included in the model; U is another subset of markers with two-way epistatic effects included in the model; ε_i represents a normally distributed random error term.

To investigate the performance of SPAEML in these data, we conducted a simulation study. Specifically, we simulated the same number and effects of QTLs that were identified in the QTL x environment study. We used previous studies in maize (Lipka et al. 2013; Chen and Lipka 2016; Chen et al. 2019) and the decay rate of LD for soybean that is slower than maize (Hyten et al. 2007; Kaler et al. 2020) to develop a criterion for quantifying the true positive detection of the simulated quantitative trait nucleotides (QTNs). Specifically, a true positive was defined as a signal detected by SPAEML located within a surrounding ±300 kb of a QTN. Detection rates of SPAEML, defined as the proportion of times that a detected additive QTN, located within ±300 kb of any of the simulated QTN, was correctly specified in the SPAEML model as an additive signal, or misspecified as epistatic.

2.3. Results

2.3.1. Phenotypic Analysis

There were 9 traits evaluated in the present study, being five traits related to stink bug resistance as GFP, NDM, HSW, WHS, and TOL, and the other four traits (GY, LO, NDF, and PHM) related to agronomic traits. The residual distribution of the

256 RILs and the two parents CD 215 and IAC-100 for every trait was normal under the Shapiro-Wilk test (Appendix A). The phenotypic distribution was also plotted to figure out how parents are contrasting to each trait (Figure 1). We have traits contrasting for the two parents in HSW, LO, NDF, and NDM. For the trait GFP, the two parents had the same class. To other traits as GY, PHM, TOL, and WHS the parents were classified in a different class, but were not much contrasting.

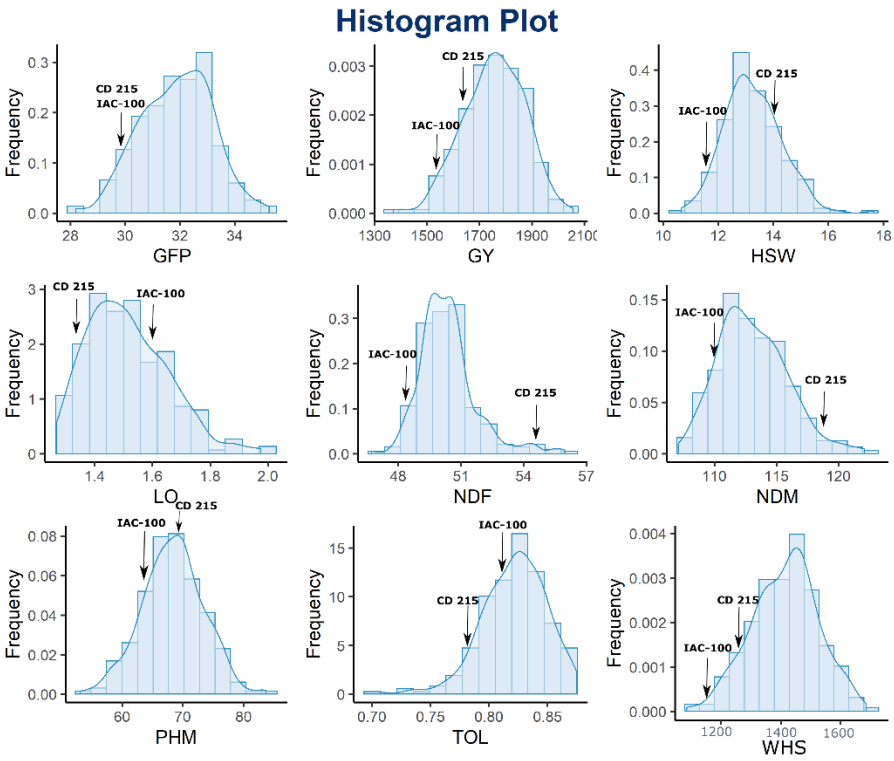


Figure 1. Frequency distribution to the traits evaluated in 256 RILs derived from CD 215 and IAC-100 crossing. GFP: grain filling period in days, GY: grain yield in kg ha^{-1}, HSW: one hundred seeds weight in grams, LO: lodging in score from 1 to 5, NDF: number of days to flowering, NDM: number of days to maturity, PHM: Plant height at maturity in cm, TOL: tolerance in %, WHS: weight of healthy seeds in kg ha^{-1}

The correlation coefficients between the traits evaluated are plotted in figure 2, all these correlations are related to the average across the seven-crop season. The majority of correlations in this present study were significant positive correlations. We had a high correlation between GFP and NDM, GY and WHS, NDF and NDM, PHM with GY and WHS, TOL and WHS. For trait HSW many of the correlations were not significant, being only to LO, PHM, and TOL significant negative correlations.

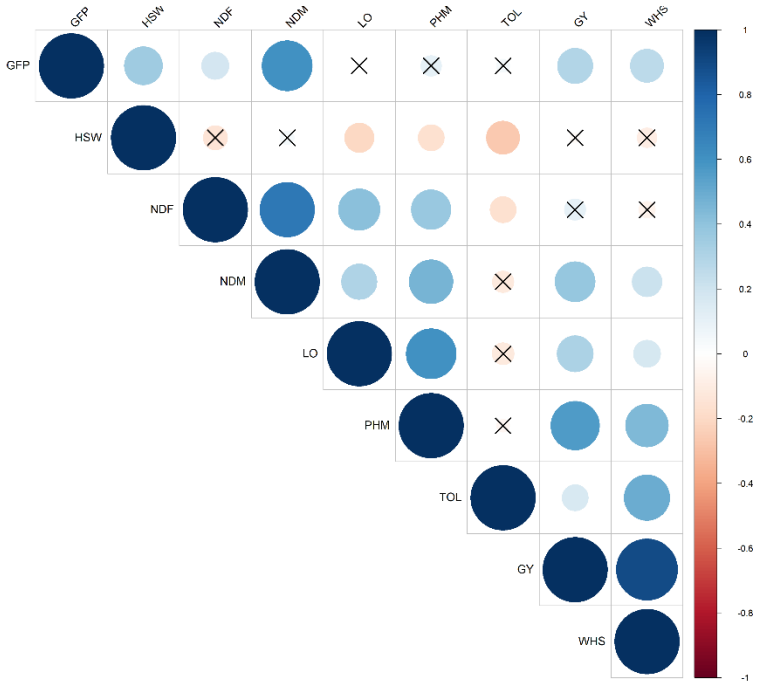


Figure 2. Phenotypic correlation between traits evaluated in 256 RILs derived from CD 215 and IAC-100 crossing. The blue colors means positive correlation, and the red color means negative correlation. The significant correlations are the bullets without “X”.

About the natural stink bug infestation, it occurred above the recommended in all crop seasons (Figure 3). In the first year (2012/2013), the natural pressure of the stink bug was the lowest among the harvests evaluated, with the maximum number of stink bugs in the area of 2.5 stink bugs per linear meter. Crop season 2017/2018 was the year with the highest pressure, with an increase from 0.56 stinkbug/m to 12.5 stinkbug/m. In the last year of evaluation (2018/2019), we observed 8.6 stink bug/m.

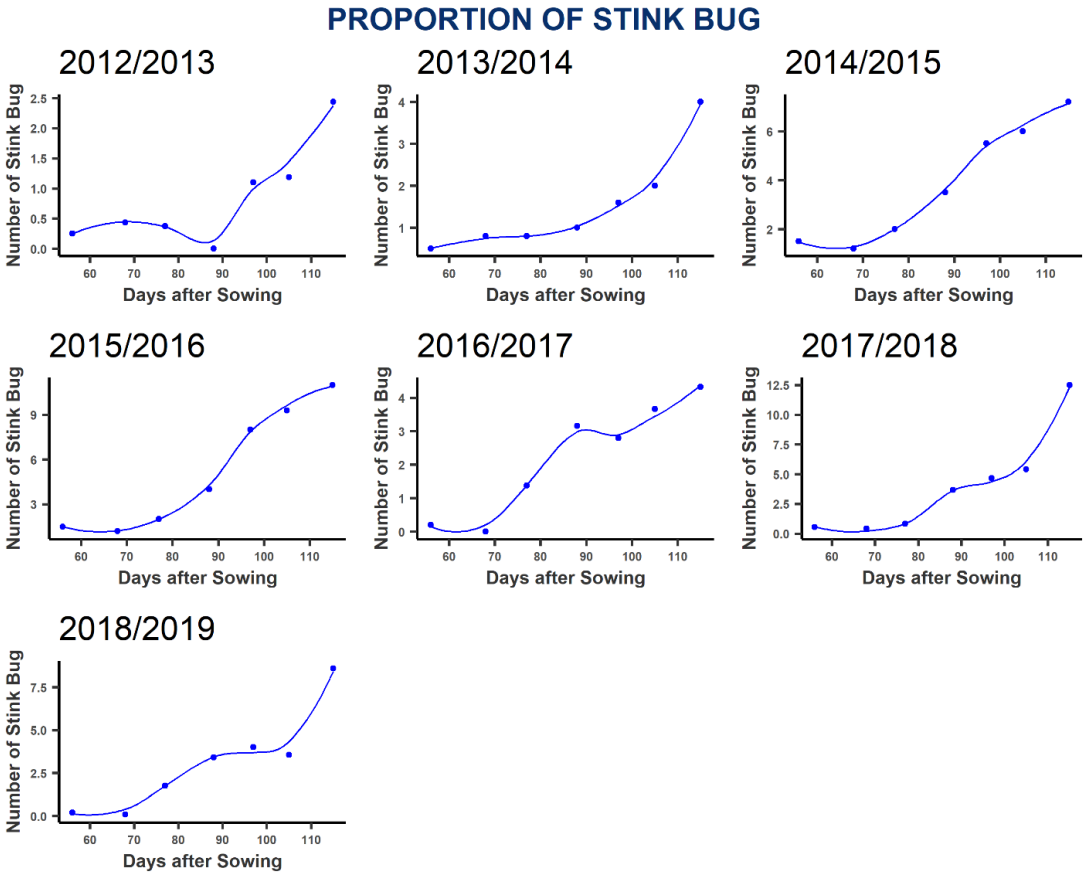


Figure 3. The occurrence of stink bugs after soybean sowing, for each crop season evaluated. To each crop season we applied the beat cloth method considering one linear meter and 10 samples at the trials from the flowering to the full maturation.

2.3.2. Construction of Genetic Linkage Map

We constructed a genetic linkage map covering 1445.10 cM using 2,037 SNPs and the Kosambi function, with an average marker spacing of less than 10 cM across 20 chromosomes (Table 2). However, we have some linkage groups with gaps of more than 30 cM (Appendix B). Our first attempt was to split some groups into two or three linkage groups as LG D1, E, K, and N. Nonetheless, we had two LG (A1 and C1) that had still lack of polymorphic markers where we maintain the groups with more than 30 cM of gap.

Table 2. Linkage map summary from the RILs population on soybean.

Linkage Group	Chromosome	SNPs	Length (cM)
A1	5	3	92.94
A2	8	59	24.00
B1	11	12	4.40
B2	14	166	48.38
C1	4	3	50.84
C2	6	56	97.08
D1a	1	139	153.98
D1b	2	58	61.86
D2	17	487	136.79
E_1	15a	191	155.97
E_2	15b	108	80.46
E_3	15c	106	29.31
F	13	135	97.78
G	18	4	31.80
H	12	154	98.40
I	20	112	63.02
J	16	3	52.57
K_1	9a	68	5.78
K_2	9b	57	29.75
L	19	5	9.29
M	7	14	34.82
N_1	3a	17	2.74
N_2	3b	52	46.42
O	10	28	36.72
Total		2037	1445.10

2.3.3. QTL X E Analysis

Among the evaluated traits, QTLs were observed only for GFP, HSW, TOL, GY, and WHS (Table 3). We found a total of seven QTLs, most of them in traits related to the resistance to the stink bug complex. In summary, we found QTLs in linkage groups A1, D1a, D1b, D2, E_1, and E_3, for some linkage groups we have overlapped regions as in D1a and E_3 (Figure 4).

The first trait related to the stink bug resistance that we found significant QTLs was the grain filling period, which corresponds to the stages that start the grain filling (R5) and ending of the grain filling (R7), a period in which there is generally an increase in stink bugs in the field. We found one significant QTL, GFP@D1b_48.36, located on chromosome 2 (LG D1_b). This QTL presented some stability, being stable across 3 years, and explained less than 10% of the phenotypic variation.

Table 3. QTLs X E detected in the "IAC100 x CD215" RILs population for the evaluated traits.

QTL	Locus ID	Add.eff.[a]	Prob[b]	R^{2c}	CI[d]
2012/2013					
GFP@D1b_48.36	GM02_47071187	0.92	0.00	6.60	27.41-61.86
HSW@E1_33.06*	GM15_10675840	0.27	0.00	2.90	9.06-57.06
HSW@D2_65.09*	GM17_15539823	0.29	0.00	3.30	46.15-84.03
WHS@E3_9.55	GM15_50203966	114.66	0.01	13.70	4.89-14.21
2013/2014					
GFP@D1b_48.36	GM02_47071187	0.54	0.05	1.60	27.41-61.86
HSW@E1_33.06*	GM15_10675840	0.27	0.00	2.90	9.06-57.06
HSW@D2_65.09*	GM17_15539823	0.29	0.00	3.40	46.15-84.03
WHS@E3_9.55	GM15_50203966	64.38	0.03	10.70	4.89-14.21
2014/2015					
GY@D1a_27.18	C1P27	137.97	0.00	9.90	20.36-34.00
HSW@E1_33.06*	GM15_10675840	0.27	0.00	1.50	9.06-57.06
HSW@D2_65.09*	GM17_15539823	0.29	0.00	1.80	46.15-84.03
WHS@E3_9.55	GM15_50203966	234.25	0.00	22.10	4.89-14.21
2015/2016					
GFP@D1b_48.36	GM02_47071187	0.16	0.00	3.60	27.41-61.86
HSW@E1_33.06*	GM15_10675840	0.27	0.00	2.60	9.06-57.06
HSW@D2_65.09*	GM17_15539823	0.29	0.00	3.10	46.15-84.03
TOL@A1_0.00	GM05_412602	0.02	0.00	6.00	0.00-24.71
WHS@E3_9.55	GM15_50203966	76.92	0.01	17.50	4.89-14.21
2016/2017					
GY@D1a_27.18	C1P27	131.41	0.00	8.00	20.36-34.00
HSW@E1_33.06*	GM15_10675840	0.27	0.00	2.30	9.06-57.06
HSW@D2_65.09*	GM17_15539823	0.29	0.00	2.70	46.15-84.03
WHS@D1a_27.18	C1P27	128.96	0.00	9.80	16.54-37.82
WHS@E3_9.55	GM15_50203966	167.96	0.00	16.60	4.89-14.21
2017/2018					
GY@D1a_27.18	C1P27	102.20	0.00	15.70	20.36-34.00
HSW@E1_33.06*	GM15_10675840	0.27	0.00	3.30	9.06-57.06
HSW@D2_65.09*	GM17_15539823	0.29	0.00	3.90	46.15-84.03
TOL@A1_0.00	GM05_412602	0.01	0.01	4.70	0.00-24.71
WHS@D1a_27.18	C1P27	80.54	0.00	10.80	16.54-37.82
2018/2019					
HSW@E_33.06	GM15_10675840	0.27	0.00	6.10	9.06-57.06
HSW@D2_65.09	GM17_15539823	0.29	0.00	7.10	46.15-84.03
WHS@D1a_27.18	C1P27	58.05	0.04	4.30	16.54-37.82

[a] Add.eff., Additive effect for each QTL
[b] Prob, Probability ($p \leq 0.05$) for each QTL
[c] R^2, percentage of phenotypic variation explained by a QTL
[d] CI, the confidence interval for the map location of a QTL
* QTLs considered stable in our analyses

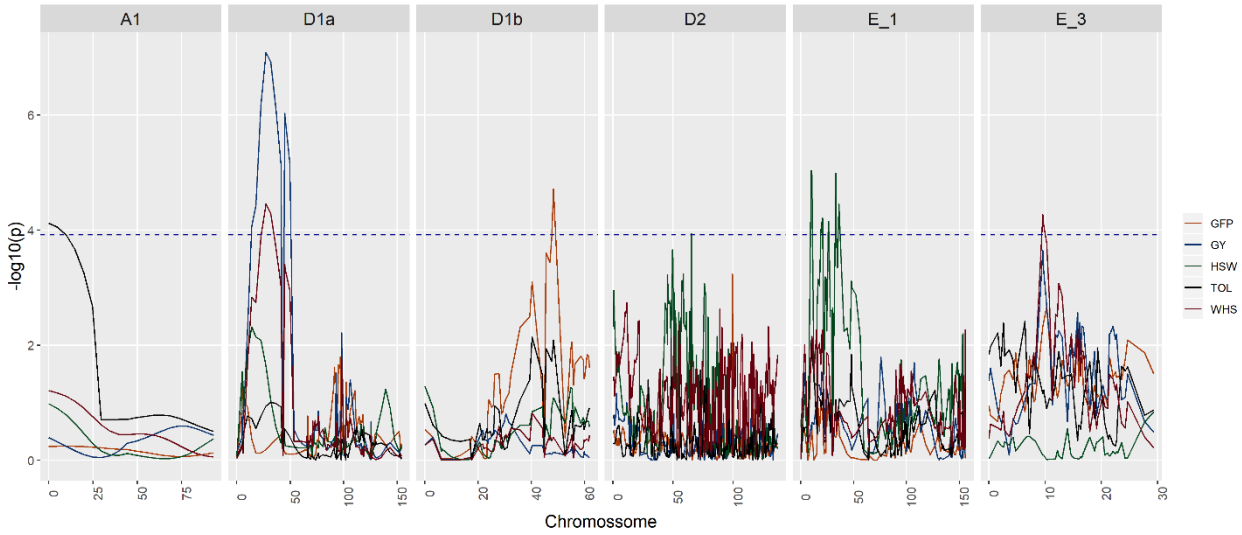


Figure 4. -Log10$_p$ of traits that had significant QTLs in our analyses in order to linkage groups (A1, D1a, D1b, D2, E_1, E_3).

For the one hundred seeds' weight, significant QTLs were also found, being two QTL on chromosome 15 (LG E). These two QTLs reveal as stable QTLs across every year. The QTL HSW@E_33.06 demonstrated an additive effect of 0.27 gr and have the biggest R^2 at the last year affording 6.1%. The second QTL to this trait was HSW@D2_65.09, which had an additive effect of 0.29 and explained 7.1% of phenotypic variation, being the biggest R^2 again at the last year.

Another trait related to stink bug resistance is tolerance, that is the proportion of lost seeds from stink bug damage. For this we found one QTL, TOL@A1_0.00 in chromosome 5 (LG A1), at the crop seasons 2014/2015 and 2017/2018. These QTLs presented interaction with the environment, had an additive effect on average 0.15%, and explained 6.0% of the phenotypic variation.

One of the main traits for evaluating resistance is the weight of health seeds. These seeds we considered as seeds without stink bug damage. For this trait, the analysis evidenced two QTLs, one in chromosome 1 (LG D1a) and the other in chromosome 15 (LG E). Despite these QTLs presented some environmental interaction, we considered these regions also as stables QTLs, where at least in 3 years these QTLs were significant. The WHS@E_9.55QTL was significant at the crop seasons 2012/2013, 2013/2014, 2014/2015, 2015/2016, and 2016/2017. This QTL was the one that explained the greater part of phenotypic variation, explaining more than 10% in each year. The other QTL was WHS@D1a_27.18 being significant

in 2016/205, 2017/2018, and 2018/2019, and explaining close to 10% of the phenotypic variation.

Finally, the last trait that we had a QTL significant was grain yield, which we consider in our research as an agronomic trait, but it is an important economic trait to select resistant cultivars. We found QTL significant in 2014/2015, 2016/2017, and 2017/2018, in chromosome 1 (LG D1a). This QTL explained the phenotypic variation between 8.0% in 2016/2017, and 15.7% in 2017/2018, with an additive effect between 102.20 kg ha^{-1} and 137.97 kg ha^{-1}.

2.3.4. Epistatic Study

In the epistatic study, we saw that generally, the results from epistasis and QTL mapping were not similar (data not published). Some QTLs detected as epistatic were not the same or close with the detected QTL on QTL mapping, and none of the finals results from SPAEML on these data contained additive effects. These results motivated our simulation study, where we explored the performance of SPAEML, in which the true quantitative trait nucleotides (QTNs) underlying simulated traits, as well as their effect sizes, were known.

To simulate the phenotypes, we used the same effects found in the QTL mapping, and the same model (only additive effects). The QTLs/heritabilities identified in the QTL studies were used to define the effect sizes and heritabilities of the simulated traits. Since we simulated only additive signals, our perspective was that in the results we just had additive signals or a high rate of additive signals.

With the simulated phenotypes, we proceeded with the SPAEML. If some models had epistatic results, we defined this as “detected as epistatic”, and remembering if these happen, would be a wrong result since we just simulated additive QTN. To develop these criteria for quantifying the true additive signals we based on previous studies in maize (Lipka et al. 2013; Chen and Lipka 2016; Chen et al. 2019) and the decay rate of LD for soybean.

On the results, we can see that the rate of detected-as-a-epistatic effect was higher than detected-as-additive, and this happened for almost all scenarios (Figure 5). In this case, these results were misspecified, because we have just simulated additive signals. Only to GFP and LO the SPAEML detect “true” additive signals, with

a rate of 2% and 1%, considering the QTN simulated. For some other traits, SPAEML identified additive effect but outside of the window that was considered as a false-positive. One of the limitations in this work might have been the population size, which is too small and this can affect the accuracy of SPAEML (Chen et al. 2019).

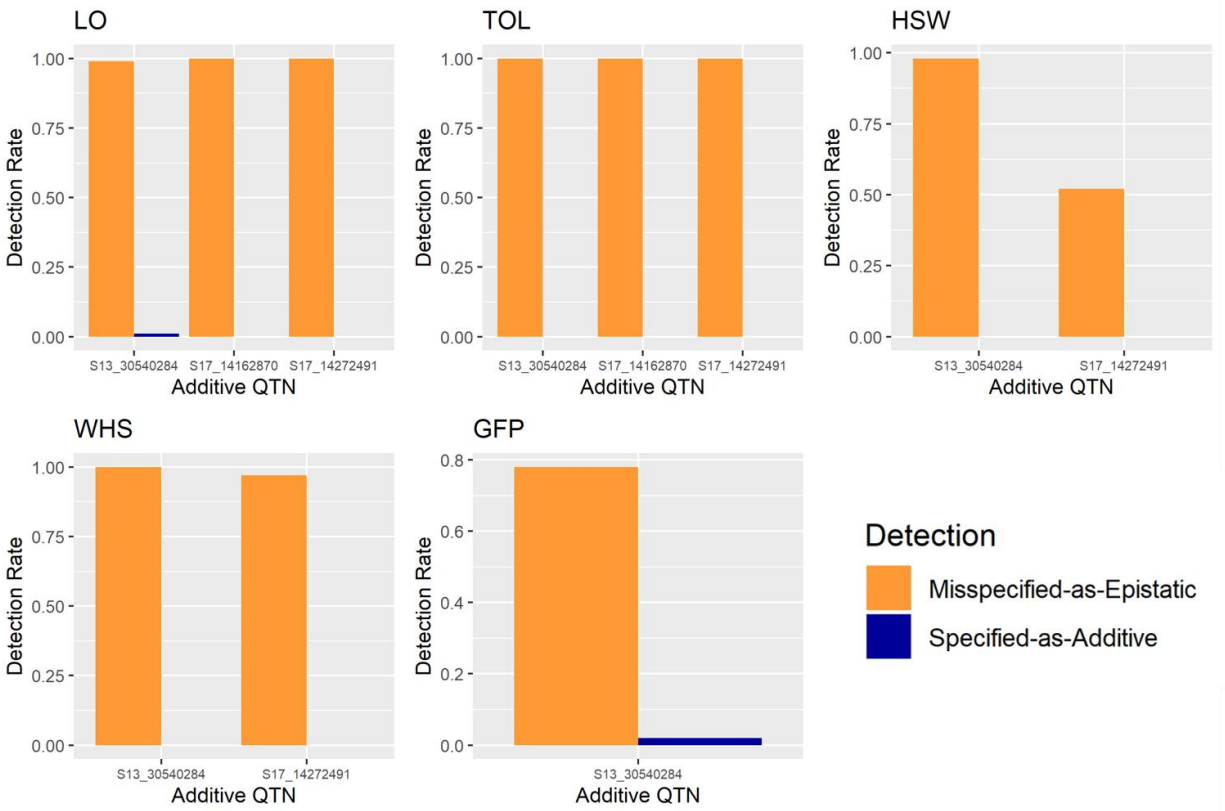


Figure 5. Detection rate for the traits grain filling period (GFP), one hundred seeds weight (HSW), lodging (LO), tolerance (TOL), and weight of healthy seeds (WHS). We simulated the same number of QTLs found on the QTL x E study. Detection rates of SPAEML, defined as the proportion of times that a detected additive QTN located within ±300 kb of any of the simulated QTN was correctly specified in the SPAEML model as an additive, or misspecified as epistatic.

2.4. Discussion

We performed a QTL mapping x environment study of the IAC 100 x CD215 population to reach a better understanding of resistance to stink bugs in soybean. The cultivar IAC-100 is considered a genotype resistant to stink bug and defoliating insects (ROSSETO et al. 1995; Piubelli et al. 2005) due to several mechanisms, among them a shorter cycle for grain filling and the ability to abort the damaged

pods, replacing them with new ones. Also, IAC-100 has a high content of isoflavones being correlated with the resistance mechanisms of this cultivar, being isoflavones a fundamental role in the mechanisms of plants for resistance to insects (Rao et al. 1990; Piubelli et al. 2003, 2005). This cultivar has also been used in breeding programs in the USA as a source of resistance to stink bugs and defoliating insects (McPherson and Buss 2007; McPherson et al. 2007b).

In this study we evaluated a total of nine traits, being evaluated agronomic traits, and also traits related to stink bug resistance. About agronomic traits, we had just QTL significant to grain yield (GY) in linkage group D1a. These QTL had a QTL x environment interaction, explaining on average more than 10% of the phenotypic variation. QTL associated with grain yield in linkage groups D1a were also related by Orf et al. (1999), but at the interval 63.52cM and 65.52cM.

We consider as traits related to stink bug resistance the grain filling period (GFP), that when the cultivar has a shorter period this is considered as a resistance mechanism; one hundred seeds weight (HWS), where we expect that resistant cultivars have a lower weight of hundred seed; the weight of healthy seeds (WHS), that was a trait developed in our lab (Rocha et al. 2014), we used the seed that was not damaged by stink bugs to weight; tolerance (TOL) that was the proportion of the losses seeds when we passed in a spiral.

Considering these traits related to stink bug resistance we found a total of six QTLs, most of them located on Chromosome 15 and 17, linkage groups E and D2 respectively. We have also a QTL found in Chromosome 5 to tolerance, and in Chromosome 2 to grain filling period. Other research also found QTLs related to these traits in these Chromosomes. For instance, to one hundred seeds weight, a major QTL was found in chromosome 17 (D2) accounted for 9.4–20.9 % of phenotypic variation of this trait, placed in position 45.45cM (Kato et al. 2014), close to QTL found in this present research. Other research found QTLs related to HSW in chromosome 8. Han et al. (2012) studying three different biparental populations in three environments, found two QTLs on chromosome 8 across the three environments. HSW is a component trait of seed yield, with relation to seed size (Burris et al. 1973; Smith and Camper 1975). To resistance we focusing on seeds with small seed size so that the stink bug cannot reach the seeds, but with a high yield.

In the literature we have some QTLs founds to insect resistance. For instance, Rector et al. (2000), found a major QTL to corn borer, Komatsu et al. (2005) found a QTL to the common cutworm, and Li et al. (2007) found the gene of resistant aphid, all of them were located on satt463 in chromosome 7. We have some QTLs to insect resistance in chromosome 15, for instance, the QTLs of corn borer were located at three different populations on chromosome 15 (Terry et al. 2000; Boerma and Walker 2005). In Chromosome 17, we have QTLs to Japanese beetle resistance (Yesudas et al. 2010), corn earworm (Terry et al. 2000), and whitefly resistance (Zhang et al. 2013). Another important player in the defense response of soybean to insect attack is the Isoflavones. In the literature has been proposed that these compounds are part of the defense against herbivorous insects, which are produced and accumulate after insect attack to cause toxicity or as repellent to insects (War et al. 2012). In this way, QTLs related to flavonoids were also found in linkage group D1b, A1, and E (Meng et al. 2011, 2016a). The QTLs found in this research were located in these linkage groups and also in the linkage group D2. All of these chromosomes are related with other QTLs to insect resistance. These regions might be a potential source to validate QTLs related to insect resistance and to use in MAS.

Additionally, in our findings, we had some stables QTLs. For the marker assisted selection be applied in a breeding program, the QTLs found must present stability across environments and appear in different genetic backgrounds (Brummer et al. 1997). In our research, we considered stable QTLs those that are significant across at least three years of the seven years evaluated (Brummer et al. 1997; Kato et al. 2014; Qi et al. 2014). In this way, QTLs to GFP, HSW, and WHS could be considered as stable regions which indicates that this QTLs might have potential to be included or to be useful in a breeding program. It is clear that the traits evaluated in this work follow a quantitative nature, being detected few QTLs, which generally explain a few portions of the phenotypic variation. Further, the few markers presenting stability, which just to the trait HSW having QTLs stable across all years evaluated, corroborate with this affirmative as well.

Furthermore, the QTLs found at the single QTLs by environment analysis also had an overlap in the confidence interval. QTL mapping suggested the region for

these traits on LG E_3 and LG D1a are overlapped. Therefore, genes underlying these QTLs might be linked or pleiotropic.

We performed an epistatic study in this population, from which we can understand the interaction effect between alleles at two or more genomic loci. Most results found in this present study were misspecified as epistatic. In this way, we confirmed via simulation study that the sample size (n=256) of this RIL population is too small to rigorously quantify epistatic associations. Chen et al. (2019) also observed that at n = 300, SPAEML was more likely to misspecify additive QTN as epistatic and identify only one locus contributing to an epistatic QTN. So, one of the limitations in this work for epistatic study might be the population size.

In summary, the present study sought to understand the genetic architecture of stink bug resistance in soybean. We found some additive QTLs that can underlie this resistance in soybean. We studied traits related to stink bug resistance, some of the QTLs found in these traits might be linked or pleiotropic. We also confirmed via simulation study that the sample size of this RIL population is too small to rigorously quantify epistatic associations. In this research, we sought to add the evaluation in multi-environments. With this information, we expected to generate better information in these regions, such as the identification of stable QTLs, a better understanding of the genetic architecture for the resistance to stink bug complex on soybean, markers to use in marker-assisted selection, and aid in the development of resistant cultivars. This work potentiates the studies regarding the stink bug complex, bringing information that can contribute to several soybean breeding programs. This can help in the definition of techniques and the selection of strategies, having a great influence on the generation of cultivars that are more productive and resistant to stink bugs. The additive QTLs found in this work will be studied in future research to validate the genetic architecture of stink bug complex resistance on soybean.

References

Boer MP, Wright D, Feng L, et al (2007) A mixed-model quantitative trait loci (QTL) analysis for multiple-environment trial data using environmental covariables for QTL-by-environment interactions, with an example in maize. Genetics 177:1801–13. https://doi.org/10.1534/genetics.107.071068

Boerma HR, Walker DR (2005) Discovery and utilization of QTLs for insect resistance in soybean. In: Genetics of Adaptation. Springer-Verlag, Berlin/Heidelberg, pp 181–189

Brummer EC, Graef GL, Orf J, et al (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Sci 37:370–378. https://doi.org/10.2135/cropsci1997.0011183X003700020011x

Burris JS, Edje OT, Wahab AH (1973) Effects of Seed Size on Seedling Performance in Soybeans: II. Seedling Growth and Photosynthesis and Field Performance 1 . Crop Sci 13:207–210. https://doi.org/10.2135/cropsci1973.0011183x001300020017x

Chen AH, Ge W, Metcalf W, et al (2019) An assessment of true and false positive detection rates of stepwise epistatic model selection as a function of sample size and number of markers. Heredity (Edinb) 122:660–671. https://doi.org/10.1038/s41437-018-0162-2

Chen AH, Lipka AE (2016) The use of targeted marker subsets to account for population structure and relatedness in genome-wide association studies of maize (Zea mays L.). G3 Genes, Genomes, Genet 6:2365–2374. https://doi.org/10.1534/g3.116.029090

CONAB (2021) Acompanhamento da Safra Brasileira. Cia Nac Abast Acompan da Safra Bras 7:1–89

Cullis B, Gogel B, Verbyla A, Thompson R (1998) Spatial Analysis of Multi-Environment Early Generation Variety Trials. Biometrics 54:1. https://doi.org/10.2307/2533991

Dekkers JCM, Hospital F (2002) MULTIFACTORIAL GENETICSTHE USE OF MOLECULAR GENETICS IN THE IMPROVEMENT OF AGRICULTURAL POPULATIONS. Nat Rev Genet 3:22–32. https://doi.org/10.1038/nrg701

Doyle J., Doyle J. (1990) Isolation of plant DNA fresh tissue. Focus (Madison) 12:13–15

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Godoi CRC de, Pinheiro JB (2009) Genetic parameters and selection strategies for soybean genotypes resistant to the stink bug-complex. Genet Mol Biol 32:328–336. https://doi.org/10.1590/S1415-47572009000200020

Grant D, Nelson RT, Cannon SB, et al (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Res 38:D843–D846. https://doi.org/10.1093/nar/gkp798

GUEDES JVC, ARNEMANN JA, STÜRMER GR, et al (2012) Percevejos da soja: novos cenários, novo manejo. Rev. Plantio Direto 12:28–34

Guo B, Sleper DA, Arelli PR, 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. https://doi.org/10.1007/s00122-005-0031-2

Guo B, Sleper DA, Lu P, et al (2006) Erratum (Retracted Article): QTLs associated with resistance to soybean cyst nematode in soybean meta-analysis of QTL locations - Retraction (Crop Science 45 (202-208)). Crop Sci 46:202. https://doi.org/10.2135/cropsci2005.04-0036

Han Y, Xie D, Teng W, et al (2012) QTL underlying developmental behaviour of 100-seed weight of soybean. Plant Breed 131:600–606. https://doi.org/10.1111/j.1439-0523.2012.01987.x

Hyten DL, Choi IY, Song Q, et al (2007) Highly variable patterns of linkage disequilibrium in multiple soybean populations. Genetics 175:1937–1944. https://doi.org/10.1534/genetics.106.069740

Jiang C, Zeng ZB (1995) Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:

Kaler AS, Gillman JD, Beissinger T, Purcell LC (2020) Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. Front Plant Sci 10:1794. https://doi.org/10.3389/fpls.2019.01794

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. https://doi.org/10.1007/s00122-014-2304-0

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. https://doi.org/10.2135/cropsci2004.0744

Kosambi DD (1944) THE ESTIMATION OF MAPS DISTANCES FROM RECOMBINATION VALUES. Eugenics 12:172–175

Li J, Ji L (2005) Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. Heredity (Edinb) 95:221–227. https://doi.org/10.1038/sj.hdy.6800717

Li Y, Hill CB, Carlson SR, et al (2007) Soybean aphid resistance genes in the soybean cultivars Dowling and Jackson map to linkage group M. Mol Breed 19:25–34. https://doi.org/10.1007/s11032-006-9039-9

Lipka AE, Gore MA, Magallanes-Lundback M, et al (2013) Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 Genes, Genomes, Genet 3:1287–1299. https://doi.org/10.1534/g3.113.006148

Lopes ECA, Destro D, Montalván R, et al (1997) Genetic gain and correlations among traits for stink bug resistance in soybeans. Euphytica 97:161–166. https://doi.org/10.1023/A:1003042520035

MALOSETTI M., BOER MP., BINK MCAM., EEUWIJK FA (2006) Multi-trait QTL analysis based on mixed models with parsimonious covariance matrices. In: WORLD CONGRESS ON GENETICS APPLIED TO LIVESTOCK PRODUCTION, 8. Belo Horizonte, p 25

Malosetti M, Ribaut JM, Vargas M, et al (2008) A multi-trait multi-environment QTL mixed model with an application to drought and nitrogen stress trials in maize (Zea mays L.). Euphytica 161:241–257. https://doi.org/10.1007/s10681-007-9594-0

Malosetti M, Voltas J, Romagosa I, et al (2004) Mixed models including environmental covariables for studying QTL by environment interaction. Euphytica 137:139–145. https://doi.org/10.1023/B:EUPH.0000040511.46388.ef

Margarido GRA, Pastina MM, Souza AP, Garcia AAF (2015) Multi-trait multi-environment quantitative trait loci mapping for a sugarcane commercial cross provides insights on the inheritance of important traits. Mol Breed 35:175. https://doi.org/10.1007/s11032-015-0366-6

Margarido GRA, Souza AP, Garcia AAF (2007) OneMap: software for genetic mapping in outcrossing species. Hereditas 144:78–79. https://doi.org/10.1111/j.2007.0018-0661.02000.x

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. https://doi.org/10.1093/jee/100.3.962

McPherson RM, Buss GR, Roberts PM (2007a) Assessing Stink Bug Resistance in Soybean Breeding Lines Containing Genes from Germplasm IAC-100. J Econ Entomol 100:1456–1463. https://doi.org/10.1093/jee/100.4.1456

McPherson RM, Buss GR, Roberts PM (2007b) Assessing Stink Bug Resistance in Soybean Breeding Lines Containing Genes from Germplasm IAC-100. J Econ Entomol 100:1456–1463. https://doi.org/10.1093/jee/100.4.1456

Meng F, Han Y, Teng W, et al (2011) QTL underlying the resistance to soybean aphid (Aphis glycines Matsumura) through isoflavone-mediated antibiosis in soybean cultivar “Zhongdou 27.” Theor Appl Genet 123:1459–1465. https://doi.org/10.1007/s00122-011-1680-y

Meng S, He J, Zhao T, et al (2016) Detecting the QTL-allele system of seed isoflavone content in Chinese soybean landrace population for optimal cross design and gene system exploration. Theor Appl Genet 129:1557–1576. https://doi.org/10.1007/s00122-016-2724-0

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. https://doi.org/10.2135/cropsci2001.1931

Orf JH, 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. https://doi.org/10.2135/cropsci1999.3961642x

Panizzi AR, Corrêa BS, Gazzoni DL, et al (1977) Insetos da Soja no Brasil. Bol Técnico 20

Pastina MM, Malosetti M, Gazaffi R, et al (2012) A mixed model QTL analysis for sugarcane multiple-harvest-location trial data. Theor Appl Genet 124:835–849. https://doi.org/10.1007/s00122-011-1748-8

Piepho H-P (1997) Analyzing Genotype-Environment Data by Mixed Models with Multiplicative Terms. Biometrics 53:761. https://doi.org/10.2307/2533976

Piepho H-P (2000) A Mixed-Model Approach to Mapping Quantitative Trait Loci in Barley on the Basis of Multiple Environment Data. Genetics 156:

Pinheiro JB, Vello NA, Rossetto CJ, Zucchi MI (2005) Potential of soybean genotypes as insect resistance sources. Biotechnology 294–301. https://doi.org/10.12702/1984-7033.v05n03a06

Piubelli GC, Hoffmann-Campo CB, De Arruda IC, et al (2003) Flavonoid increase in soybean as a response to Nezara viridula injury and its effect on insect-feeding preference. J Chem Ecol 29:1223–1233. https://doi.org/10.1023/A:1023889825129

Piubelli GC, Hoffmann-Campo CB, Moscardi F, et al (2005) Are Chemical Compounds Important for Soybean Resistance to Anticarsia gemmatalis? J Chem Ecol 31:1509–1525. https://doi.org/10.1007/s10886-005-5794-z

Qi Z, Han X, Hou M, et al (2014) QTL analysis of soybean oil content under 17 environments. Can J Plant Sci 94:245–261. https://doi.org/10.4141/CJPS2013-219

Rao K V., Chattopadhyay SK, Reddy GC (1990) Flavonoids with mosquito larval toxicity. J Agric Food Chem 38:1427–1430. https://doi.org/10.1021/jf00096a028

Rector BG, All JN, Parrott WA, Boerma HR (1999) Quantitative Trait Loci for Antixenosis Resistance to Corn Earworm in Soybean. Crop Sci 39:531. https://doi.org/10.2135/cropsci1999.0011183X003900020038x

Rector BG, All JN, Parrott WA, Boerma HR (1998) Identification of molecular markers linked to quantitative trait loci for soybean resistance to corn earworm. TAG Theor Appl Genet 96:786–790. https://doi.org/10.1007/s001220050803

Rector BG, All JN, Parrott WA, Boerma HR (2000) Quantitative Trait Loci for Antibiosis Resistance to Corn Earworm in Soybean. Crop Sci 40:233. https://doi.org/10.2135/cropsci2000.401233x

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. https://doi.org/10.1590/1984-70332014v14n3a26

ROSSETO CJ, GALLO PB, RAZERA LF, et al (1995) Mechanisms of resistance to stink bug complex in the soyabean. An Socidedade Entomol Bras 24:517–522

Rossetto CJ, Lourenção AL, Igue T, Miranda MAC De (1981) Picadas de alimentação de Nezara viridula em cultivares e linhagens de soja de diferentes graus de suscetibilidade. Bragantia 40:109–114. https://doi.org/10.1590/S0006-87051981000100010

Smith A, Cullis B, Thompson R (2001) Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. Biometrics 57:1138–1147. https://doi.org/10.1111/j.0006-341X.2001.01138.x

SMITH AB, CULLIS BR, THOMPSON R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143:449. https://doi.org/10.1017/S0021859605005587

Smith AB, Stringer JK, Wei X, Cullis BR (2007) Varietal selection for perennial crops where data relate to multiple harvests from a series of field trials. Euphytica 157:253–266. https://doi.org/10.1007/s10681-007-9418-2

Smith TJ, Camper HM (1975) Effects of Seed Size on Soybean Performance. Agron J 67:681–684. https://doi.org/10.2134/agronj1975.00021962006700050025x

Terry LI, Chase K, Jarvik T, et al (2000) Soybean Quantitative Trait Loci for Resistance to Insects. Crop Sci 40:375. https://doi.org/10.2135/cropsci2000.402375x

VERBYLA AP., ECKERMAN PJ., THOMPSON R., CULLIS BR (2003) The analysis of quantitative trait loci in multi-environment trials using a multiplicative mixed model. Aust J Agric Res 54:1395–1408

VIVAN LM, DEGRANDE PE (2011) Pragas da soja. Fundação MT Bol Pesqui Soja 11:239–297

VSN International (2017) Genstat for Windows 19th Edition. In: VSN Int. Hemel Hempstead, UK. https://genstat.kb.vsni.co.uk/knowledge-base/whats-new19th/. Accessed 13 Sep 2019

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. https://doi.org/10.1007/s00122-004-1714-9

War AR, Paulraj MG, Ahmad T, et al (2012) Mechanisms of plant defense against insect herbivores. Plant Signal. Behav. 7:1306

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. https://doi.org/10.1007/s00122-010-1314-9

Zhang J, Li W, Zhang L, et al (2013) QTL mapping of soybean resistance to whitefly (Bemisia Tabaci Gennadius) under multi-environment conditions

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. https://doi.org/10.1007/s10681-008-9728-z

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. https://doi.org/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. https://doi.org/10.1007/s00122-007-0682-2

Declarations

Fundings
This study was financed by the following institutions Fundação de Amparo à Pesquisa do Estado de São Paulo, grant numbers 2019/13010-0, 2018/12399-9, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, grant number 140586/2018-0 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Acknowledgements
The authors gratefully acknowledge the financial support of following institutions Fundação de Amparo à Pesquisa do Estado de São Paulo, grant numbers 2019/13010-0, 2018/12399-9, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, grant number 140586/2018-0 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Also, we would like to thank the Dr. Kaio de Oliveira and Professor Alexander Edward Lipka for scientific support. Finally, we wish like to thank you “Luiz de Queiroz” College of Agriculture and the University of Illinois at Urbana-Champaign for providing the necessary facilities.

3. VALIDATION OF QTL TO STINK BUG COMPLEX RESISTANCE IN SOYBEAN

ABSTRACT

Regarding the insects that attack soybean, stink bugs are those that cause the greatest yield losses. To deal with these losses the development of resistant cultivars is a great alternative that could substitute or be used together with insecticides. One of the steps to find resistant cultivar is to understand the genetic architecture of the resistance, using the approach as GWAS and QTL mapping. To achieve more accurate results and have marker more reliable to use in a breeding program, the validation of regions of interest is substantial information. This project aimed to identify regions involved in resistance in soybean to stink bug complex through genome-wide association study, and validating the regions found in linkage mapping through GWAS. To perform this, we used a soybean breeding panel population, composed of 299 inbred lines and checks from a multiparental cross. Data from two agricultural years was used. The experiments were carried out at the ESALQ Experimental Station in the city of Piracicaba, State of São Paulo, Brazil. The experimental design was an alpha-lattice 16x19 with three replicates each. We evaluated just traits that have significant QTLs, at the QTL mapping study. To obtain the genotypic data, genotyping-by-sequencing (GBS) was used. The phenotypic data was analyzed through mixed models and later the study of GWAS was carried out to validate the QTLs found in the RILs population. We found a total of 22 SNPs, related to the traits evaluated, three of them validated in chromosome 1, and 15. Additionally the QTL in chromosome 6 is close to another region found earlier in this population. It is clear that the chromosome 1, 6, and 15 plays a crucial role at the resistance to stink bug on soybean, having regions that underlie the genetic architecture of this resistance.

Keywords: genetic architecture, GWAS, insects, resistant cultivars.

3.1. INTRODUCTION

Soybean is one of the most important cultivated crops in the world, and a major agricultural commodity in Brazil. The importance of soybean passes since to be a primary source of protein and oil around the world, until human food products (Ortega et al. 2016; Wilcox 2016). Soybean was domesticated 3000-5000 years ago in China, being introduced from China to the USA in 1765, and later to others countries as Argentina and Brazil. As a commercial crop, soybean started its expansion in Brazil in 1970 from south of Brazil, and has been expanding in area and importance since this year, occupying areas for entire Brazil (dos Santos Silva et al.

2017). As a result of this expansion, damage caused by pests like stink bugs has also increased.

Among the stink bugs species, species of pentatomidae that cause more damage in Brazil soybean are the *Euchistus heros*, *Nezara viridula*, commonly known as “Southern green stink bug”, and *Piezodorus guildinii* known as “Red-banded stink bug”. These three stink bugs compound the called “stink bug complex”. The stink bugs are considered one of the worst pests, causing direct damage to feeding pods and injecting salivary secretions, which cause irreversible damage to the seeds as in its germination and developing, generating yield losses consequently (Tood and Turnipseed 1974; Panizzi and Slansky 1985).

Insect damage is considered a limiting factor for soybean production, generating high yield losses, affecting leaves, vascular sap, pods, seeds, roots (Chang and Hartman 2017). To manage this, the farmers generally use as a tool application of insecticides, but with a limited number of new molecules, and a fast evolution in insecticide resistance, turns the control of stink bug so hard (Bentivenha et al. 2018). Therefore, an appropriate tactic to reduce problems with these pests is the use of genetic resistance, which could be used as complementary to the use of insecticides and is an important part of integrated pest management strategy (Painter 1951; Smith 2005; Chang and Hartman 2017).

The development of resistant cultivars depends on phenotyping lines in the field, but the development of protocols to this is laborious turning the progress on breeding programs difficult. Since the advent of genetic markers based on DNA, and its use to elucidate the genetic architecture of several traits, many researchers are using DNA regions to help breeders to develop cultivars with regions of interest to determinate traits. QTL mapping is one of the first approaches developed to find regions that could be related to a trait and dissect the genetic architecture of the trait. Based on a linkage map, genotype, and phenotype, QTL mapping identifies regions that explain part of the variation of the traits (Lynch and Walsh 1998). With the advance in low-cost platforms to sequence the genomes of crops, new approaches arose as genomic wide association studies (GWAS), assisting the development of marker assisted selection (MAS). GWAS is based on the recombination events from a population, occurred in an evolutionary time, in which its linkage blocks are smaller than the populations developed in QTL mapping, providing higher mapping resolution

compared to QTL mapping. These two techniques have been applied in soybean breeding and helped breeders to face their challenges (Zhang et al. 2017; Liu et al. 2017; Ghione et al. 2021; Yang et al. 2021).

Currently, a lot of researches has focused on the validation of QTL and/or high-resolution mapping (Landi et al. 2005; Sallam et al. 2016). A validation stage of the mapped QTLs is important as their positions and effects may be imprecise. It should be kept in mind that factors that affect the accuracy of the mapped QTLs, such as the mapping approaches, population size, genetic marker nature, and G x E, may affect the usefulness of the linkage between markers and QTL that is essential for MAS (Melchinger et al. 1998). The QTL validation usually refers to the verification that the QTL effect is present in different genetic backgrounds, where the researcher can rule out statistical errors (Langridge et al. 2001). The effects of these QTLs are confirmed in experiments and both QTL and markers are validated in relevant germplasm. This next step "post QTL mapping" is important and fully required to integrate and exploit molecular genetic research results in conventional breeding, making MAS much more effective with higher reliability than found with QTL.

In this study, we performed a GWAS in a breeding panel aiming to elucidate the genetic architecture of resistance to stink bug complex on soybean, focusing on validating the regions found at QTL mapping population, and searching for new candidate regions that could underlie the resistance to stink bug complex on soybean.

3.2. Materials and Methods

3.2.1. Population

To validate the QTLs found in the RILs population, we performed a genomic wide association study (GWAS). The GWAS population, was composed by 299 inbred lines from the soybean breeding panel of Diversity genetic and Plant Breeding laboratory of the University of São Paulo, which originated from the same germplasm source as the RILs. This population was derived from the commercial cultivars BRS-133, CD-215, Conquista, Dowling, IAC-100, and Pintado, recombined with other PI genotypes (PI-plant introduction): PI 200487 (Kinoshita), PI 471904 (Orba), PI 200526 (Shiranui), and PI 459025 (Bing Nan). It can be inferred that within this

soybean breeding panel the same parents of the RILs population can also be found, which indicates the possibility of an existent variability for stink bug resistance. Furthermore, progenies resulting from these crosses were recombined with nine populations that have the variability for stink bug resistance. These genotypes underwent two selection cycles for stink bugs complex resistance and a selection cycle for grain yield, totaling 299 inbred lines. In this experiment, five commercial checks were used in the field: IAC 100, CD215, AS3730, Produza, and NS7300.

3.2.2. Phenotypic data

Data from 2018/2019 and 2019/2020 crop seasons were used. All trials were developed at the same farm for the linkage analysis in Piracicaba-SP. The experimental design was a 16 x 19 alpha-lattice with three replicates. The experimental plots consisted of two lines of four meters long, spaced 0.5 meters between rows and 18 seeds per linear meter.

No chemical control was applied to control insects in order to allow the natural infestation of stink bugs. To monitor the prevalence of these stink bugs in the area, the beat cloth method was applied from the flowering to the full maturation (Panizzi et al. 1977).

With the linkage analyzes carried out, we decided to evaluate just traits that have significant QTLs from the first analysis. We evaluated GFP - Grain filling period in days; GY - Grain yield in kg ha^{-1}; HSW - Healthy seeds weight in grams; TOL- tolerance in percentage, which was calculated as $TOL = \left(1 - \left(\frac{GY-WHS}{GY}\right)\right) \; x \; 100$; WHS - Weight of a hundred seeds in kg ha^{-1}. To compute the HSW the seeds that were not damaged by stink bugs were considered. After the harvest, the seeds passed through a spiral, where the empty, green, and malformed grains were separated by gravity and centrifugal forces. After grain processing, the data was taken in kg ha^{-1} (Rocha et al. 2014). To leaf retention, we just had the opportunity to phenotype at the crop season 2019/2020, when we had a high pressure of stink bug in the field.

3.2.3. Phenotypic analysis

The proposed model to analyze the phenotypic data was divided into two steps. The initial model we performed for each year considering the alpha-lattice design, with the objective to run an individual GWAS to each year, being:

$$y_{irs} = \mu + P_s + \underline{B}_{r/s} + G_i + \underline{\varepsilon}_{irs} \qquad \text{(model 1)}$$

where: y_{irs} is the phenotype of genotype *i*, in block r, of the replicate s; μ is the intercept; P_s is the effect of the replicate *s*, where s = 1, 2, and 3; $B_{\underline{r}}$ is the random effect of block *r* within the replicate *s*, where $B_{\underline{r}}$ ~ N(0, $\sigma^2_{B\underline{r}}$); G_i is the fixed genetic effect of genotype *i*; ε_{irs} is the residue. Assume that the residues are independent and identically distributed (ε_i ~ N(0, σ^2_e)).

After we take out the adjusted means to each genotype, we carried out the analysis over two years :

$$y_{ij} = \mu + G_i + \underline{L}_j + \underline{\varepsilon}_{ij} \qquad \text{(model 2)}$$

where: y_{ij} is the phenotype of genotype *i* in environment *j*; μ is the intercept; G_i is the fixed genetic effect of genotype *i*; L_j is the random effect to each year *j*, where (L_j ~ N(0, σ^2_j) ε_{ij} is the residue. Assume that the residues are independent and identically distributed (ε_i ~ N(0, σ^2_e)).

Furthermore, we also performed a full model (one step), considering the model 1 and 2 together, to have the BLUEs to each genotype and run the GWAS. Since the correlation to this model and the model performed in two steps was too high, 0.99 to every evaluated trait. This can happen because the adjusted means have close accuracy at the fitting models, which turn the weight models unnecessary. Therefore, we decided to follow with the two steps analyses to run GWAS analyzes over two years.

In the meantime, that we ran the linear mixed model considering the full model, we used this to measure some descriptive analyzes. With a low unbalanced condition over two years in these trials we measured the heritability using the following expression:

$$H^2 = \frac{V_g}{V_g + \frac{V_{gl}}{r} + \frac{V_\varepsilon}{rl}}$$

where: H^2 is the heritability; V_g is the genotypic variance, V_{gl} is the variance of the interaction G x L, V_ε is the residual variance, r is the number of replications,

and l is the number of years. Finally, to detect statistical significance to the random effects, we performed likelihood ratio tests (LRT). To perform the LRT we used the lme4 package (Bates et al. 2015), to each source of variation we performed two models (full and restrict), aiming to test the full model against the restrict model where the best model was the one that maximized the likelihood function. To detect significance to fixed effects we performed an ANOVA.

3.2.4. Genotyping

Genomic DNA was extracted from one week-old seedling leaf tissue using the DNeasy Plant Mini Kit (Qiagen®, Germany). For genotyping this population, the technique of genotyping-by-sequencing (GBS) was used, following the protocol described by (Elshire et al. 2011). The genotyping was carried out at the Functional Genomics Core facility of Sao Paulo University in Piracicaba-SP. The reads were aligned to the William 82 (version: Gmax_275_Wm82. a2. V4 genome). A total of 288 genotypes were genotyped and digested by the NsiI enzyme. The SNP calling, we performed using the TASSEL bioinformatic pipeline using minor allele frequency (MAF) ≥ 0.05; and call rate of < 0.9. The final data had 7,230 SNPs, which were used to perform the GWAS.

3.2.5. Genomic Wide Association Studies

The validation of the linkage analysis was done through GWAS. The association analysis was performed using the GAPIT computational package (Lipka et al. 2012). The Fixed and random model Circulating Probability Unification (FarmCPU; LIU et al., 2016) were used to perform the GWAS analyzes. The FarmCPU uses a multilocus model to control false positives. With this, a multiple loci linear mixed model (MLMM) is performed divided into two stages: first in a fixed-effect model (FEM), contain single markers test, one at a time, to avoid the confounding between kinship and the test marker, where are included pseudo QTNs as covariates to control false positives. Pseudo QTNs are estimated by the random estimated model (REM), and they are used to define the kinship, this kind of process

prevents an over-fitting problem in FEM. FEM and REM were run iteratively until no change on pseudo-QTNs. FEM and REM are modeled as described below:

FEM: $y_i = M_{i1}b_1 + M_{i2}b_2 + \cdots + M_{it}b_t + S_{ij}d_j + \varepsilon_i$

REM: $y_i = u_i + \varepsilon_i$

where y_i is the observation of the i[th] individual in both models; $M_{i1}, M_{i2}, \ldots, M_{it}$, are the genotypes of t pseudo QTNs; $b_1, b_2, \ldots, b_t$, are the effects of the pseud QTNs; S_{ij} is the genotype of i[th] individual j[th] genetic marker; d_j is the effect of the j[th] genetic marker; u_i is the total genetic effect of the i[th] individual; and finally, ε_i the residual having a ε_i ~ N(0, σ_e^2).

The GWAS was carried out using 7,230 SNP, and the adjusted means from the phenotypic model 2. For correction of the multiple tests and to consider an SNP as significant, we used the Bonferroni multiple test correction (Weir 1996) and additionally permutation tests using 1000 iterations, with 5% of global significance for type 1 error. Moreover, to each significant SNP, we calculated the proportion of variance explained (PVE), being $PVE = \left(\frac{V_{qtn}}{V_i}\right)^2$, where V_{qtn} is $2freq(1 - freq)effect^2$, and the V_i is the phenotypic variance of the BLUEs to each trait.

To obtain chromosome physical lengths (bp) we used the Glyma.Wm.82.a4 reference genome through SoyBase (www.soybase.org) calculating the genome-wide inter-marker distance and chromosome-wide densities. The correlation squared - r^2, which is the correlation to the square between the presence and absence of alleles at different loci (Hill and Robertson 1968), was calculated as a measure of DL, with the R package synbreed (Wimmer et al. 2012). Only significant r^2 values ($P < 0.001$), calculated according to Remington et al. (2001), were considered informative. The extension of the LD decay was measured as the chromosomal distance when (r^2) dropped to half its maximum value (Huang et al. 2010).

In order to validate the regions found in QTL mapping we considered regions co-detected in both analysis, where we aimed to find regions on GWAS that were inside of the confidence interval (CI) from QTL mapping (Zhang et al. 2019). To search candidate resistance genes we examined regions close to the significant SNP, considering the LD decay region, based on the soybean genome assembly version 2.0 from Williams 82 at the Phytozome database assembly (Ghione et al. 2021).

3.3. Results

3.3.1. Phenotypic Analysis

To all traits we observed an extensive phenotypic variation, for instance to WHS we had a range from 5.00 kg ha^{-1} to 2360.00 kg ha^{-1}(Table 1). The coefficient of the variation (CV) varied from 9.51 % (HSW) to 42.67% (WHS). The high value of CV to WHS was inflated due to the high pressure of stink bugs in the field in the second year. Generally, to all traits, we noticed a high genetic variance effect with the exception of WHS where the G x Y component was higher than the genetic component. The heritability was considered high, being 48.51%, 78.49%, 74.62%, 71.80%, and 53.97% to GFP, HSW, LR, TOL, and WHS respectively. This could indicate that most part of the phenotypic variance can be genetic what turns association studies appropriate to dissect the genetic architecture of the traits evaluated.

Table 4. Descriptive summary and variance components of evaluated traits in a soybean breeding panel.

Traits	Range	Mean	CV (%)[a]	Variance Components				H^2 (%)[e]
				Genetic	G x Y[b]	B x Y[c]	B[d]	
GFP	20.00-48.00	32.69	10.66	3.00	2.32	0.48	-	48.51
HSW	5.10-21.73	13.43	9.51	1.91	0.50	0.19	-	78.49
LR	1.00-5.00	3.69	17.83	0.42	-	-	0.10	74.62
TOL	0.34-94.19	20.30	31.13	45.99	22.80	12.10	-	71.80
WHS	5.00-2360.00	450.92	42.67	20512.18	22646.09	10346.84	-	53.97

[a] CV(%): variation coefficient in %; [b] GxY: variance component of interaction genotype by year; [c] BxY: variance component of interaction block by year; [d] B: variance component of block; [e] H^2: heritability.

To all traits we observed a significant p-value (Table 5) for the likelihood ratio test, corroborating with our evidence that there is a genetic component controlling the genetic architecture to stink bug resistance. Considering that we are evaluating a set of quantitative traits, and this set is related to stink bug resistance, we can find some lines that present tolerance to the stink bug complex. For instance, the LQ198 could be an interesting genotype to work as a resistant cultivar, this line shows 29 days to GFP, 1,161.81 kg ha^{-1} to WHS, scoring 2 to LR, 58% to TOL, and 11.20 gr to HSW

(Appendix C). Despite that we also had lines that are susceptible to the stink bug complex, for example, LQ186 had 144 kg ha^{-1} to WHS, which is our main trait to evaluate resistant lines. This kind of values corroborate with our idea that this population has variability to study stink bug resistance.

Table 5. Likelihood ratio test (LRT) of random effects for the evaluated traits in a soybean breeding panel.

Trait	LRT		P.value	
	G	G x Y	G	G x Y
GFP	-5127.10	-5114.10	2.20E-16**	3.47E-07**
HSW	-3265.60	-3244.30	2.20E-16**	6.72E-11**
LR	-1153.20	-	2.20E-16**	-
TOL	-5473.20	-5361.60	2.20E-16**	2.20E-16**
WHS	-12451.00	-12382.00	2.20E-16**	2.20E-16**

3.3.2. Linkage Disequilibrium

GWAS was accomplished using a set of 7,230 SNPs with MAF > 0.05, and a missing rate < 10%, distributed over the 20 chromosomes in soybean. Genome-wide LD decay at the breeding panel was estimated. The overall LD decay for all chromosomes was estimated when half of the maximum r^2 was achieved. In our study, r^2 decreased while the distance increased, an average LD across all chromosomes decayed to r^2 = 0.225 (half of its maximum value) in approximately 7,900 kb, considering euchromatic and heterochromatic regions together (Figure 6). This value indicates a strong LD existed in the population.

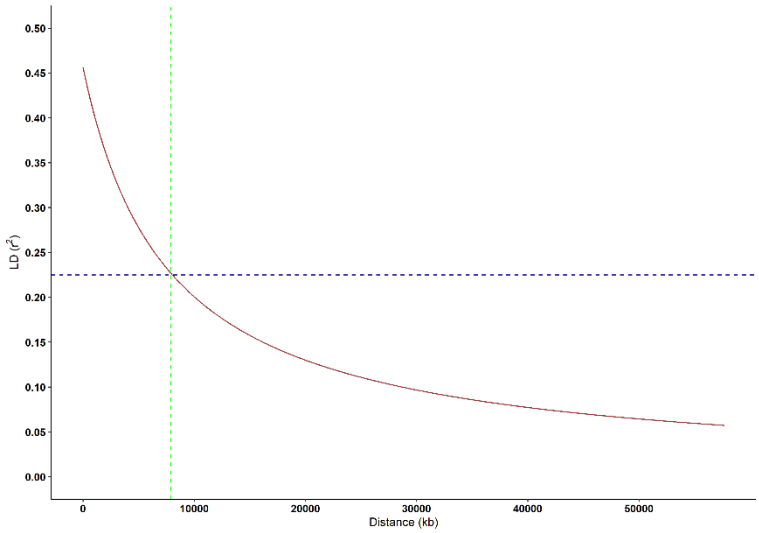


Figure 6. Linkage disequilibrium (LD) decay across soybean genome.

3.3.3. Genome-Wide Association Studies

We performed the GWAS using the BLUEs of 288 genotypes over two years, accounting for both population structure and family relatedness. We identified a total of 22 SNPs in different chromosomes associated with the evaluated traits, considering the Bonferroni threshold and the permutation test threshold. The Q-Q plot performed to all traits, that represent the expected and observed probability of getting association of SNPs with phenotype, indicated an effective control for false-positive associations (Figure 7). Furthermore, the model fitted in GWAS showed that all -log10(p) observed are similar to the expected with exception of that significant associations.

To GWAS we just considered traits that have significant QTLs identified on the QTL mapping analysis and a new evaluated trait that was leaf retention. The GWAS analysis was based on the adjusted means phenotypic data over two years. We found QTNs to all evaluated traits (Figure 8, Table 6).

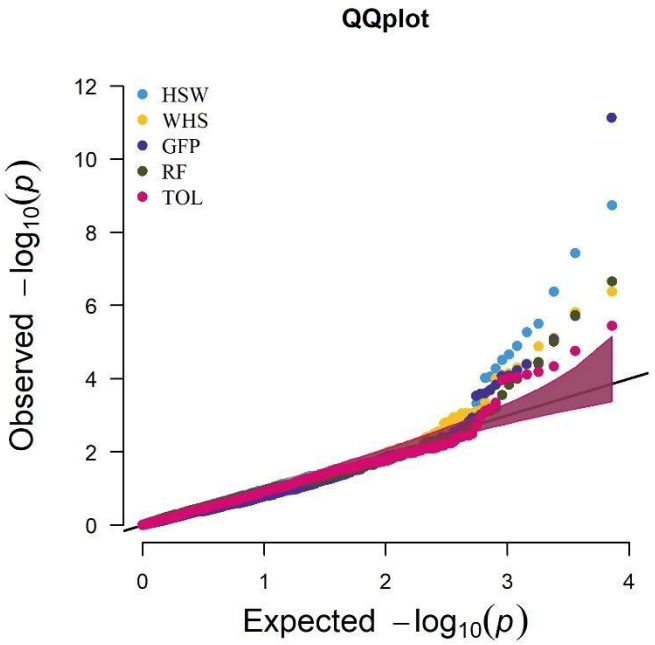


Figure 7. Quantile-Quantile (QQ) plot to the evaluates traits. The shadow area indicates a 95% confidence interval.

To the grain filling period, we identified associated regions on chromosomes 6 and 11 (GM06_23658447 and GM11_5665644). In chromosome 11, we identified a high peak that can indicate an interesting region to future studies. This QTN was located at 5,665,644 bp in chromosome 11, with an effect of 1.44 days, and had a PVE of 5.48%. For the other QTN (GM06_23658447), this showed an effect of -0.92 days and had a lower PVE than that found in chromosome 11, being 2.66.

To one hundred seeds weight we found a total of 7 QTNs, where 5 of them were significant under the conservative Bonferroni threshold (p< 6.915e-06) and the other two regions were significant when the permutation test threshold was used (p < 2.011118e-05). At this trait, we had a range of PVE from 1.34% to 3.70%, and the effect of each QTN varied from -0.57gr to 0.43 gr.

The leaf retention, which is one of the damages caused by a stink bug, we found three regions associated, two of which were found in chromosome 11, at the same chromosome to GFP, and one in chromosome 15. Their effects on the scored index varied from -0.21 to 0.16, and the PVE had a range from 1.93% (GM11_346007) to 2.53% (GM11_17678972).

The weight of healthy seed had 4 significant SNPs, being in chromosomes GM01, GM 09, GM13, and GM19. The highest peak was found in chromosome 13,

SNP GM13_13822483, with an effect of 64.08 kg.ha^{-1}, with also the highest PVE, explaining 8.18% of the variation. Furthermore, to validate regions associated with stink bug resistance we also considered the peaks in chromosome GM15 and GM17, which appear close to the permutation test threshold, p<1.701956e-05.

For tolerance, we found a total of seven regions significant, one was significant under the Bonferroni threshold, and the others to the permutation test (p < 2.434118e-05). The highest peak was found to the SNP GM13_13822483, the same QTN found on WHS, this SNP had a PVE of 5.27%, and an effect to increase the tolerance on average 20%. The other SNPs had PVE lower than 2.64%.

Table 6. Quantitative trait nucleotides associated with grain filling period (GFP), one hundred seeds weight (HSW), leaf retention (LR), tolerance (TOL), and weight of healthy seeds (WHS), according to the GWAS results to the soybean breeding panel.

Trait	SNP[a]	CHR[b]	POS(bp)[c]	P.value	MAF[d]	Effect	PVE (%)[e]
GFP	GM06_23658447	GM06	23658447	1.95E-06	0.10	-0.92	2.66
	GM11_5665644	GM11	5665644	7.33E-12	0.09	1.44	5.48
HSW	GM7_26900513	GM07	26900513	4.20E-07	0.06	-0.56	1.71
	GM7_38091330	GM07	38091330	1.82E-09	0.14	-0.57	3.70
	GM9_37481600	GM09	37481600	3.11E-06	0.20	0.35	1.83
	GM15_14732104	GM15	14732104	1.27E-05	0.12	0.43	1.81
	GM16_3708919	GM16	3708919	3.68E-08	0.49	-0.36	2.91
	GM19_10589608	GM19	10589608	5.36E-06	0.06	-0.51	1.34
LR	GM11_346007	GM11	346007	9.72E-06	0.13	-0.21	1.93
	GM11_17678972	GM11	17678972	2.19E-07	0.24	-0.19	2.53
	GM15_11506064	GM15	11506064	1.84E-06	0.35	0.16	2.26
TOL	GM01_28779056	GM01	28779056	7.65E-05	0.07	0.02	0.00
	GM10_49714291	GM10	49714291	6.47E-05	0.41	0.02	2.64
	GM11_16023802	GM11	16023802	8.99E-05	0.26	-0.01	2.64
	GM13_13822483	GM13	13822483	3.60E-06	0.46	0.02	5.27
	GM17_35883142	GM17	35883142	1.74E-05	0.08	0.02	2.64
	GM20_30337441	GM20	30337441	4.60E-05	0.23	0.01	2.64
WHS	GM01_32236644	GM01	32236644	1.29E-05	0.29	-38.15	2.42
	GM09_45694446	GM09	45694446	7.63E-06	0.17	-48.39	2.70
	GM13_13822483	GM13	13822483	4.15E-07	0.45	64.08	8.18
	GM15_37213844	GM15	37213844	7.77E-05	0.13	-34.59	1.08
	GM17_8074629	GM17	8074629	7.16E-05	0.27	-27.61	1.20
	GM19_17399139	GM19	17399139	1.52E-06	0.06	64.68	1.87

[a]SNP: Single-nucleotide polymorphisms; [b]CHR: chromosome; [c]POS: Position in base pairs; [d]MAF: Minor allele frequency; [e]PVE: Proportion of variance explained by SNP-trait association

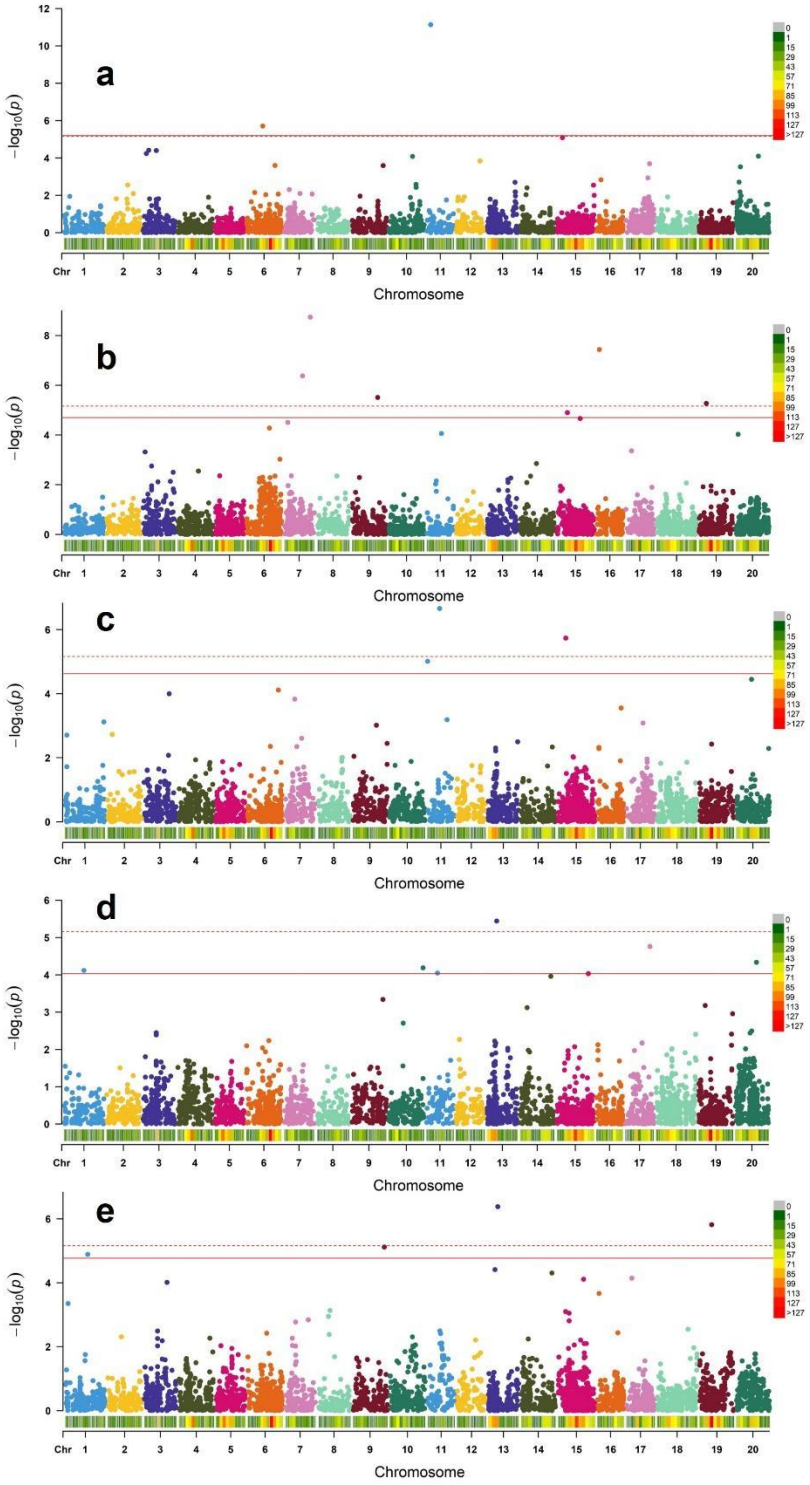


Figure 8. GWAS to five traits on soybean related to stink bug resistance. The traits to each analyzis are: (a) Grain filling period (GFP); (b) One hundred seeds weight (HSW); (c) Leaf retention (LR); (d) Tolerance (TOL); (e) Weight of healthy seeds (WHS). In the Manhattan plots the solid line indicates the Bonferroni threshold (p< 6.91e-06) and the dashed line denotes the permutation test threshold.

3.3.4. QTL Validation to stink bug resistance in soybean

To validate the regions associated to stink bug resistance at GWAS, we combined our linkage analysis and GWAS, searching for co-detected SNP regions by both analyses (Table 4). We identified three regions that were co-detected in both analyses. These SNPs were in chromosomes 1 and 15. We considered co-detected regions when the SNP loci from GWAS analysis were inside of the QTLs intervals at the linkage analysis. These validated regions were to one hundred seeds weight (HSW) and weight of health seeds (WHS). To HSW the regions in chromosome 15 had a weak PVE (<2%), with a R^2 ranging from 1.5 to 6.1 to GWAS and linkage analysis respectively. To WHS we identified regions in chromosomes 1 and 15. We also had a weak PVE in both chromosomes (<3%), but we had a higher R^2, in chromosome 1 variated from 4.3 to 10.8, and in 15 from 10.5 to 16.8. The region detected in chromosome 15 looks like to be linked to both traits, HSW and WHS, with a higher R^2 to WHS.

Table 7. Co-detected SNP loci regions by linkage analysis in the first study and GWAS

Trait	Approach	Locus ID	CHR [a]	Add. eff.[b]	Prob[c]	R^{2d}	PVE[e]	POS (bp) [f]
HSW	QTL mapping	GM15_10675840	GM15	0.27	0	1.5~6.1	-	8410421~14801197
	GWAS	GM15_14732104	GM15	0.43	1.27E-05	-	1.81	14732104
WHS	QTL mapping	GM15_50203966	GM15	64.38	0.03	10.7~16.8	-	9992000~48519467
	GWAS	GM15_37213844	GM15	-34.59	7.77E-05	-	1.08	37213844
	QTL mapping	C1P27	GM01	80.54	0	4.3~10.8	-	20042588~37672903
	GWAS	GM01_32236644	GM01	-38.15	1.29E-05	-	2.42	32236644

[a]CHR: chromosome; [b] Add.eff: additive-effect of significant SNP; [c]Prob: Probability of signicant SNP; [d]R^2: percentage of variation explained by QTL from linkage mapping; [e]PVE: Proportion of variance explained by SNP-trait association from GWAS; [f]POS: Position in base pairs to each marker found in linkage mapping and GWAS

3.4. Discussion

Mitigating the effects of the stink bug complex on farms has been one of the major bottlenecks for breeding and insecticide companies. The development of resistant cultivars, which may complement or even be an alternative to the use of insecticides, appears as a potential strategy to reduce losses to the farmer, as well as reduce the impact on the environment. This research tried to reveal the genetic architecture of the stink bug resistance in soybean. We evaluated a breeding panel that one of the founders is the IAC-100 that has been used as a source of resistance in breeding programs, very known as resistant to stink bug complex, presenting mechanisms as small grain filling period (ROSSETO et al. 1995; McPherson et al. 2007a; Sabljic et al. 2020). In our study, we did not notice a completely resistant cultivar but is valuable that some genotypes present some level of tolerance to stink bug as LQ 198. In Brazil, as far we know, until this moment we have few resistant cultivars available, where many of them are older cultivars with no agronomic values, with exception of the new releases from EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária) with the block technology.

We know that the resistance to stink bug complex presents a quantitative nature (Godoi and Pinheiro 2009), which turns the work of the breeders harder to discover regions associated with the target trait. The traits evaluated in this research confirm this affirmative. We evaluated traits that are related to stink bug resistance (Rocha et al. 2014; Da Fonseca Santos et al. 2018), that follows a quantitative nature with high heritability, greater than 48%. Nevertheless, we found few significant regions associated to each trait, most of them with small effect, this kind of situation express the intricacy to work with the resistance to stink bug on soybean.

The LD plays an important role in the application of GWAS, and can determine the resolution of association mapping (Zhu et al. 2008; Vuong et al. 2015). It is known that genetic diversity, selection, founding events, and other factors affect the extend of LD (Flint-Garcia et al. 2003). The LD decay of the soybean genome in this study was approximately 7,900 Kb, which is longer when compared with others crops like maize and rice (Mather et al. 2007; Vuong et al. 2015). Previous studies have revealed that this tends to occur in self-pollinated crops such as soybean. Kaler et al. (2020), comparing maize and soybean, observed that the cross-pollinated had a faster LD decay than soybean due to the higher recombination rate. Furthermore,

Lam et al. (2010) working with wild and cultivated soybeans demonstrate that both had high LD, with a higher level to cultivated soybean. Moreover, the LD decay also varies across chromosomes and regions as heterochromatic and euchromatic regions. In a previous study, it was identified that the LD decay had a range from 360 Kb in the heterochromatic region, and 9,600 Kb in the euchromatic region (Hwang et al. 2014). In our study, we analyze the LD decay over the 20 chromosomes and considering the heterochromatic and euchromatic regions together. The power to identify SNP is related to factor as LD, what implies in the mapping resolution. Species with slow LD decay, require also a low marker density to identify associations between marker and phenotype (Zhang et al. 2015; Vuong et al. 2015). Therefore, we are working with 7,230 polymorphic markers and a slow LD decay, which assure we have acceptable coverage of the LD blocks and reasonable power to identify regions of the large and small effect associated with the evaluated traits.

We sought to evaluate the traits that had significant QTLs at the linkage analysis. For all traits evaluated, GFP, HSW, LR, TOL, and WHS, we discovered significant regions. To GFP, QTNs were identified in GM06 and GM11. Few researchers have been worked with this trait, many of them worked, generally, with the entire reproductive period stage (R1~R7). Significant QTNs were also identified to the reproductive period stage in chromosomes 6 and 11. Zhang et al., (2015), found significant SNPs in position 16,723,946 bp and 26,933,523 bp to GM06, to GM11 was in 1,395,042 bp and 17,274,491 bp. The closest SNP was, approximately, 3 Mb of both SNPs revealed in this research. Moreover, a putative gene, Glyma.06g218100, was found close (300 Kb) to SNP in chromosome 6. This gene belongs to the alpha/beta-hydrolases superfamily protein, which plays crucial role in the development of plants, and in the cellular lipid metabolic process (Mindrebo et al. 2016). Despite this region was not near to the SNPs found to another work, we are sure that the GM06 have an important role to the genetic architecture to grain filling period and growth stage, consequently this chromosome could be more studied to assist breeding programs.

Leaf retention was a new trait evaluated in this population. We had the opportunity to phenotype due to the high pressure of stink bugs, where the whole field presented symptoms of leaf retention. Leaf retention is a physiological disturb where the leaves remain green after pod maturation, motivating delays on maturity.

We found two QTNs at the same chromosome as GFP, chromosome 11, but not close to the regions on GFP. Another trait evaluated and that had significant regions at this chromosome was tolerance (TOL), close ~ 1.5 Mb of the region to LF. Considering the longer extension of the LD decay in this genome it is reasonable to consider that this might be a high LD block. Chang; Hartman (2017), characterizing insect resistance regions in the USDA Soybean Germplasm, found some regions to defoliate insect in chromosome 11, near ~ 200kb of the region found in this research, in a high LD region at this chromosome. As far we know, any QTL to stink bug resistance was detected at this region, additional studies will be needed to support our preliminary evidence as a region to insect resistance.

To attend our goals, we validated some regions that were found at the linkage analysis performed in our first study. We co-detected a total of three regions in both linkage analysis and GWAS, two of them in chromosome 15, and one in chromosome 1. The two regions in chromosome 15 are related to the traits one hundred seeds weight (HSW), and weight of healthy seeds (WHS). These two are so distant, approximately 23,000 Kb, three times of our LD decay in the soybean genome. Despite these two regions are not close at the GWAS, the linkages analyses showed that these regions belonging in the same LD block having an overlap at the confidence interval (CI) to these markers. It is known that the chromosome 15 had one of the major insect resistance QTL, which contributes 26% of the antibiotic and 20% of the antixenotic effect to corn earworm resistance on soybean. Moreover, the two regions are also near to the Pb locus that controls the tip of the pubescence (Hulburt et al. 2004; Parrott et al. 2008; Ortega et al. 2016). We know that the Pb locus provides some antixenosis resistance typically when is conferred the sharp trichome, which deters insects from feeding, presenting the genotype an ability of non-preference (Hulburt et al. 2004). Furthermore, chromosome 15 is also related to QTL to isoflavones. Isoflavones are a great player when we talk about plant-insect interaction, playing as defensive agents, helping in antixenosis and antibiotic resistance (Piubelli et al. 2003; Kubo 2006; Hohenstein et al. 2019). Two QTLs were found in chromosome 15 to isoflavone, the two are inside of CI in this work, and one of them is close 3Mb from the QTN found to HSW. Additionally, 3 genes were described related to these regions, Glyma15g15200, Glyma15g41040, Glyma15g41130, belonging to glycosyl hydrolases, cyclin family

protein, and auxin-responsive protein families respectively. Auxin responsive plays essential roles in diverse aspects of plant development, auxin acts by modifying the plant defense responses, indicating and modulating the levels of auxin to mediate the insect specificity and activate the host defense (Erb et al. 2012).

Another validated region was in chromosome 1, where we found the QTN at position 32236644 bp*,* and the CI to linkage analysis was 20042588~37672903. Chromosome 1 also had regions that explain the genetic architecture of isoflavones. Two regions and four genes (Glyma01g16250, Glyma01g16370, Glyma01g36091, Glyma01g36110) were described as being related to seed isoflavone at this chromosome (Meng et al. 2016b). Withal, we had a gene that is close 1 Mb to the QTN in GWAS that is an abscisic acid receptor, highly expressed in roots and pods (Libault et al. 2010; Severin et al. 2010). Abscisic acid (ABA) was related in some crops to play roles as defense attacks to herbivores (Erb et al. 2012). In maize, ABA was related inducing defense response, the jasmonate (JA) core pathway that is the major signal in plant-insect interaction, increasing the levels of JA during insect attacks (Adie et al. 2007).

Additionally, Ghione et al., (2021), working with GWAS to find markers related to stink bug resistance on soybean, had also found regions associated in chromosome 6 and 15, close to the regions found in our research. Likewise, several kinds of research were performed to discover the genetic architecture to stink bug resistance on soybean at the Genetic Diversity and Plant Breeding Laboratory at the University of São Paulo. Santos, (2012), had found QTLs in chromosomes 6 and 15. To chromosome 6 was detected three QTLs, one of them was the SNP BARC-066175–19,800 (Gm06:18,736,715.0.18737142) close to the SNP found in this research to the trait GFP. Regarding chromosome 15, Santos (2012) also found three QTLS, being all of them, BARC-050109–09,389 (Gm15:10,948,749.0.10949249), BARC-028607–05,972 (Gm15:11,650,801.0.11651354) and BARC-054023–12,243 (Gm15:14,778,781.0.14779195), close to the validate region to HSW. Furthermore, Moller (2017), found QTLs in the same chromosomes to the traits GFP (Chr 15), WHS (Chr 6), and HSW (Chr 15). These three authors corroborate with our findings, that chromosomes 6 and 15 play crucial roles in resistance to stink bug on soybean.

The chromosomes 1, 6, and 15 had regions that underlie the resistance to stink bug complex on soybean. Moreover, to the best of our knowledge, we developed the first study to validate these regions as QTNs to stink bug resistance. In addition, some genes corroborate our findings, and they are also indicated as candidate genes to control this resistance. This kind of research answer some open questions about the genetic architecture to stink bug resistance, helping breeders to work with one of the worst insects that attack soybean. Nevertheless, additional information is important to continue the efforts to discover this genetic architecture. Functional analysis of genes, together with transcriptomic, and metabolomic, could eventually be used to elucidate the molecular basis of the stink bug resistance in soybean. For instance, these omics could finally provide genes and pathways for the host response to the pathogen, and additionally metabolites involved in the regulation of soybean resistance to stink bug complex.

References

Adie B, Chico JM, Rubio-Somoza I, Solano R (2007) Modulation of plant defenses by ethylene. J. Plant Growth Regul. 26:160–177

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48. https://doi.org/10.18637/jss.v067.i01

Bentivenha JPF, Canassa VF, Baldin ELL, et al (2018) Role of the Rutin and Genistein Flavonoids in Soybean Resistance to Piezodorus guildinii (Hemiptera: Pentatomidae). Arthropod Plant Interact 12:311–320. https://doi.org/10.1007/s11829-017-9578-5

Chang HX, Hartman GL (2017) Characterization of insect resistance loci in the USDA soybean germplasm collection using genome-wide association studies. Front Plant Sci 8:670. https://doi.org/10.3389/fpls.2017.00670

Da Fonseca Santos M, Möller M, Clough SJ, Pinheiro JB (2018) Heritability of agronomic traits correlated with reduced stink bug damage in an F2:3 soybean population derived from IAC-100. J Crop Improv 32:1–18. https://doi.org/10.1080/15427528.2017.1370404

dos Santos Silva FC, Sediyama T, de Cássia RTO, et al (2017) Economic importance and evolution of breeding. In: Soybean Breeding. Springer International Publishing, pp 1–16

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLoS One 6:e19379. https://doi.org/10.1371/journal.pone.0019379

Erb M, Meldau S, Howe GA (2012) Role of phytohormones in insect-specific plant reactions. Trends Plant Sci. 17:250–259

Flint-Garcia SA, Thornsberry JM, Edward IV SB (2003) Structure of Linkage Disequilibrium in Plants. Annu. Rev. Plant Biol. 54:357–374

Ghione CE, Lombardo LA, Vicentin IG, Heinz RA (2021) Association mapping to identify molecular markers associated with resistance genes to stink bugs in soybean. Euphytica 217:1–12. https://doi.org/10.1007/s10681-021-02768-1

Godoi CRC de, Pinheiro JB (2009) Genetic parameters and selection strategies for soybean genotypes resistant to the stink bug-complex. Genet Mol Biol 32:328–336. https://doi.org/10.1590/S1415-47572009000200020

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231. https://doi.org/10.1007/BF01245622

Hohenstein JD, Studham ME, Klein A, et al (2019) Transcriptional and chemical changes in soybean leaves in response to long-term aphid colonization. Front Plant Sci 10:310. https://doi.org/10.3389/fpls.2019.00310

Huang X, Wei X, Sang T, et al (2010) Genome-wide asociation studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967. https://doi.org/10.1038/ng.695

Hulburt DJ, Boerma HR, All JN (2004) Effect of Pubescence Tip on Soybean Resistance to Lepidopteran Insects. J Econ Entomol 97:621–627. https://doi.org/10.1093/jee/97.2.621

Hwang EY, Song Q, Jia G, et al (2014) A genome-wide association study of seed protein and oil content in soybean. BMC Genomics 15:1–12. https://doi.org/10.1186/1471-2164-15-1

Kaler AS, Gillman JD, Beissinger T, Purcell LC (2020) Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. Front Plant Sci 10:1794. https://doi.org/10.3389/fpls.2019.01794

Kubo I (2006) Chapter 4 New concept to search for alternate insect control agents from plants. Adv. Phytomedicine 3:61–80

Lam HM, Xu X, Liu X, et al (2010) Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. Nat Genet 42:1053–1059. https://doi.org/10.1038/ng.715

Landi P, Sanguineti MC, Salvi S, et al (2005) Validation and characterization of a major QTL affecting leaf ABA concentration in maize. Mol Breed 15:291–303. https://doi.org/10.1007/s11032-004-7604-7

Langridge P, Lagudah ES, Holton TA, et al (2001) Trends in genetic and genome analyses in wheat: a review. Aust J Agric Res 52:1043–1077. https://doi.org/10.1071/AR01082

Libault M, Farmer A, Brechenmacher L, et al (2010) Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to Bradyrhizobium japonicum infection. Plant Physiol 152:541–552. https://doi.org/10.1104/pp.109.148379

Lipka AE, Tian F, Wang Q, et al (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liu N, Li M, Hu X, et al (2017) Construction of high-density genetic map and QTL mapping of yield-related and two quality traits in soybean RILs population by RAD-sequencing. BMC Genomics 18:1–13. https://doi.org/10.1186/s12864-017-3854-8

Liu X, Huang M, Fan B, et al (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLoS Genet 12:e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer

Mather KA, Caicedo AL, Polato NR, et al (2007) The extent of linkage disequilibrium in rice (Oryza sativa L.). Genetics 177:2223–2232. https://doi.org/10.1534/genetics.107.079616

McPherson RM, Buss GR, Roberts PM (2007) Assessing Stink Bug Resistance in Soybean Breeding Lines Containing Genes from Germplasm IAC-100. J Econ Entomol 100:1456–1463. https://doi.org/10.1093/jee/100.4.1456

Melchinger AE, Utz HF, Schon CC (1998) Quantitative Trait Locus (QTL) Mapping Using Different Testers and Independent Population Samples in Maize Reveals Low Power of QTL Detection and Large Bias in Estimates of QTL Effects. Genetics 149:383–403

Meng S, He J, Zhao T, et al (2016) Detecting the QTL-allele system of seed isoflavone content in Chinese soybean landrace population for optimal cross design and gene system exploration. Theor Appl Genet 129:1557–1576. https://doi.org/10.1007/s00122-016-2724-0

Mindrebo JT, Nartey CM, Seto Y, et al (2016) Unveiling the functional diversity of the alpha/beta hydrolase superfamily in the plant kingdom. Curr. Opin. Struct. Biol. 41:233–246

Moller M (2017) Indentificação de QTLs em soja associados a resistência aos percevejos e a caracteres agronômicos utilizando a abordagem de mapeamento multivariado. Escola Superior de Agricultura Luiz de Queiroz

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. https://doi.org/10.1007/s00122-015-2658-y

Painter R (1951) Insect resistance in crop plants

Panizzi AR, Corrêa BS, Gazzoni DL, et al (1977) Insetos da Soja no Brasil. Bol Técnico 20

Panizzi AR, Slansky F (1985) Review of Phytophagous Pentatomids (Hemiptera: Pentatomidae) Associated with Soybean in the Americas. Florida Entomol 68:184. https://doi.org/10.2307/3494344

Parrott W, Walker D, Zhu S, et al (2008) Genomics of Insect-Soybean Interactions. In: Genetics and Genomics of Soybean. Springer New York, pp 269–291

Piubelli GC, Hoffmann-Campo CB, De Arruda IC, et al (2003) Flavonoid increase in soybean as a response to Nezara viridula injury and its effect on insect-feeding preference. J Chem Ecol 29:1223–1233. https://doi.org/10.1023/A:1023889825129

Remington DL, Thornsberry JM, Matsuoka Y, et al (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci U S A 98:11479–11484. https://doi.org/10.1073/pnas.201394398

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. https://doi.org/10.1590/1984-70332014v14n3a26

ROSSETO CJ, GALLO PB, RAZERA LF, et al (1995) Mechanisms of resistance to stink bug complex in the soyabean. An Socidedade Entomol Bras 24:517–522

Sabljic I, Barneto JA, Balestrasse KB, et al (2020) Role of reactive oxygen species and isoflavonoids in soybean resistance to the attack of the southern green stink bug. PeerJ 8:e9956–e9956. https://doi.org/10.7717/peerj.9956

Sallam A, Arbaoui M, El-Esawi M, et al (2016) Identification and Verification of QTL Associated with Frost Tolerance Using Linkage Mapping and GWAS in Winter Faba Bean. Front Plant Sci 7:1098. https://doi.org/10.3389/fpls.2016.01098

Santos M da F (2012) Mapeamento de QTL e expressão gênica associados à resistência da soja ao complexo de percevejos Piracicaba 2012. Escola Superior de Agricultura Luiz de Queiroz

Severin AJ, Woody JL, Bolon YT, et al (2010) RNA-Seq Atlas of Glycine max: A guide to the soybean transcriptome. BMC Plant Biol 10:1–16. https://doi.org/10.1186/1471-2229-10-160

Smith C (2005) Plant resistance to arthropods molecular and conventional approaches

Tood JW, Turnipseed SG (1974) Effects of Southern Green Stink Bug Damage on Yield and Quality of Soybeans123. J Econ Entomol 67:421–426. https://doi.org/10.1093/jee/67.3.421

Vuong TD, Sonah H, Meinhardt CG, et al (2015) Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. BMC Genomics 16:1–13. https://doi.org/10.1186/s12864-015-1811-y

Weir BS (Bruce S. (1996) Genetic data analysis II : methods for discrete population genetic data. Sinauer Associates

Wilcox JR (2016) World distribution and trade of soybean. In: Soybeans: Improvement, Production, and Uses. wiley, pp 1–14

Wimmer V, Albrecht T, Auinger HJ, Schön CC (2012) Synbreed: A framework for the analysis of genomic prediction data using R. Bioinformatics 28:2086–2087. https://doi.org/10.1093/bioinformatics/bts335

Yang L, Tian Y, Liu Y, et al (2021) QTL mapping of qSCN3-1 for resistance to soybean cyst nematode in soybean line Zhongpin 03-5373. Crop J 9:351–359. https://doi.org/10.1016/j.cj.2020.08.003

Zhang D, Zhang H, Chu S, et al (2017) Integrating QTL mapping and transcriptomics identifies candidate genes underlying QTLs associated with soybean tolerance to low-phosphorus stress. Plant Mol Biol 93:137–150. https://doi.org/10.1007/s11103-016-0552-x

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 16:1–11. https://doi.org/10.1186/s12864-015-1441-4

Zhang T, Wu T, Wang L, et al (2019) A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. Int J Mol Sci 20:. https://doi.org/10.3390/ijms20235915

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. https://doi.org/10.1007/s00122-007-0682-2

Declarations

Fundings

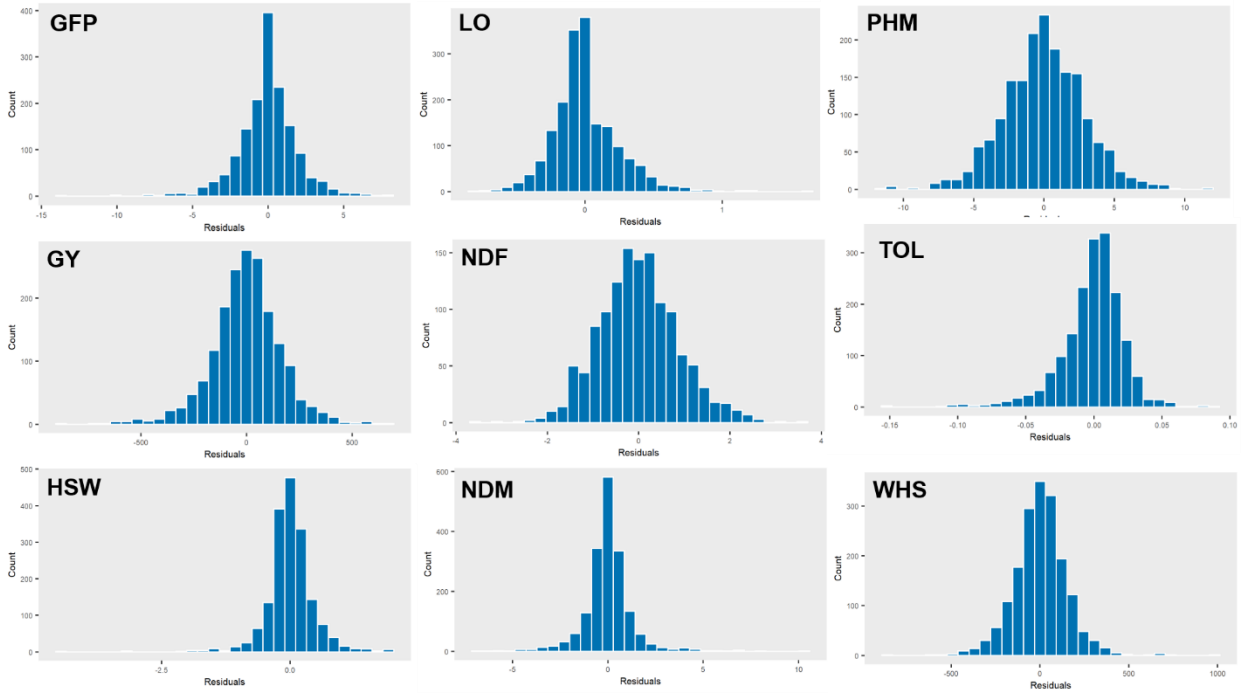
This study was financed by the following institutions Fundação de Amparo à Pesquisa do Estado de São Paulo, grant numbers 2019/13010-0, 2018/12399-9, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, grant number 140586/2018-0 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

Acknowledgements

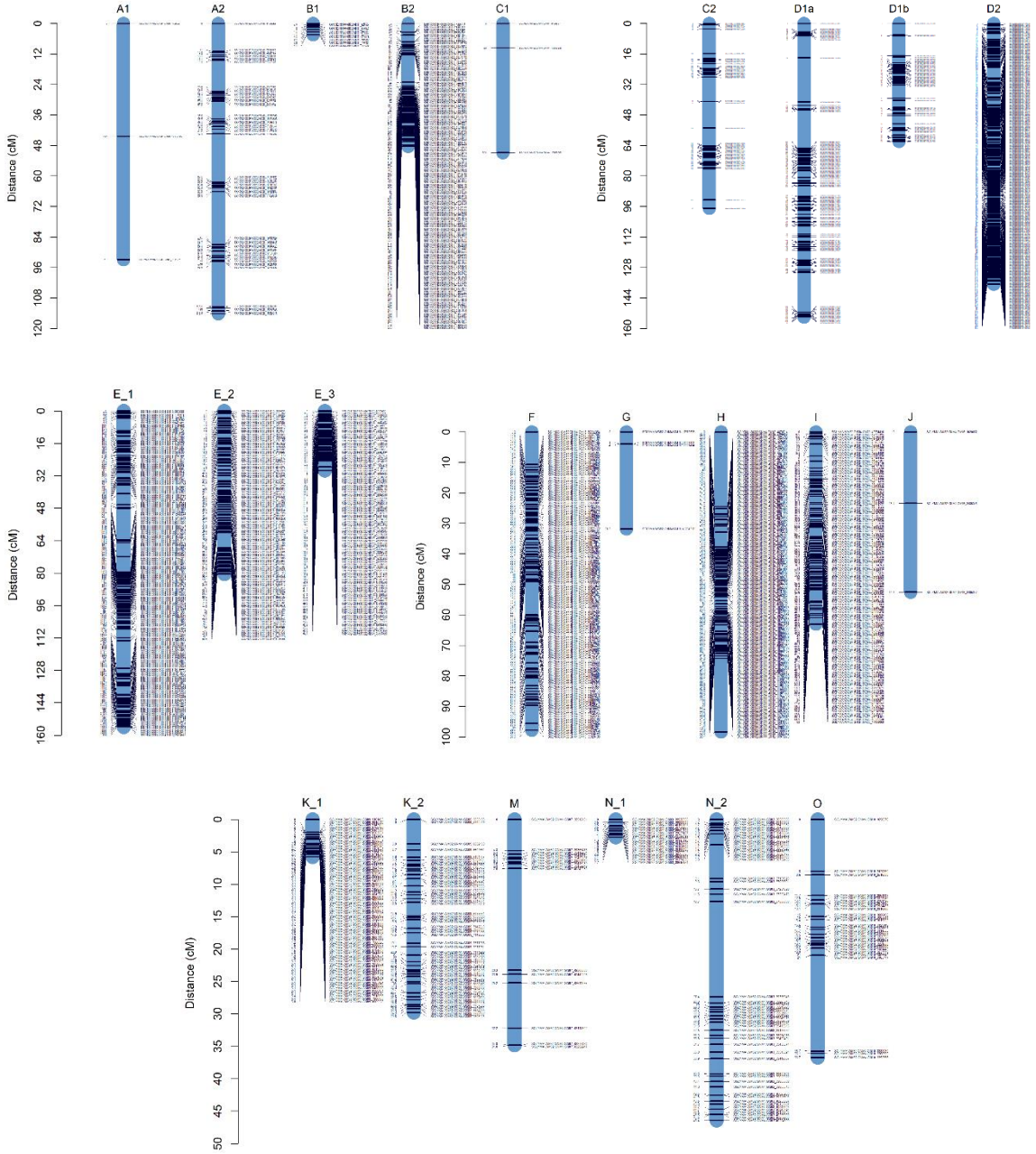
The authors gratefully acknowledge the financial support of following institutions Fundação de Amparo à Pesquisa do Estado de São Paulo, grant numbers 2019/13010-0, 2018/12399-9, Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, grant number 140586/2018-0 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Also, we would like to thank the Dr. Kaio de Oliveira and Professor Alexander Edward Lipka for scientific support. Finally, we wish like to thank you “Luiz de Queiroz” College of Agriculture and the University of Illinois at Urbana-Champaign for providing the necessary facilities.

APPENDICES

APPENDIX A. Residuals distributions of traits evaluated. Every trait was tested and present normal distribution under Shapiro-Wilks test.



APPENDIX B. The Genetic Linkage Map of soybean of 256 RILs derived from crossing between IAC 100 x CD-215.



APPENDIX C. Adjusted means over two years to the soybean breeding panel.

Genotype	GFP	HSW	RF	TOL	WHS
LQ001	32.79771	15.63597	4.582882	0.137072	184.348
LQ002	34.46025	13.8916	4.257029	0.146812	331.2094
LQ003	32.53449	12.36146	3.363106	0.172462	490.6895
LQ004	35.2914	14.77198	3.47853	0.225391	565.5954
LQ005	34.29527	12.27831	2.739436	0.180171	773.1408
LQ006	30.3678	13.22114	3.582819	0.125458	244.7767
LQ007	33.6951	16.46814	3.00401	0.219854	409.3207
LQ008	38.0191	16.1212	4.015698	0.376998	821.1215
LQ009	35.04811	13.55981	1.942491	0.327429	763.1966
LQ010	32.72868	13.20649	4.070325	0.089464	181.2447
LQ011	29.39751	9.753306	4.422156	0.176544	316.5221
LQ012	28.89987	12.68034	4.809776	0.144149	249.6453
LQ013	35.01605	16.91068	4.096334	0.186497	338.0876
LQ014	34.30766	12.92486	3.710976	0.179038	414.9797
LQ015	34.41695	12.63179	3.170181	0.20275	395.0242
LQ016	34.69809	14.74493	4.624351	0.239518	448.8438
LQ017	33.50053	10.62857	4.506293	0.165289	358.7612
LQ018	33.63873	14.20019	4.248648	0.172526	470.2874
LQ019	32.64759	13.12425	3.602392	0.135268	344.4657
LQ020	30.59626	10.92505	3.265168	0.325208	770.2705
LQ021	30.22791	13.17497	4.038955	0.1802	409.7939
LQ022	32.44465	13.68096	2.215852	0.156057	487.4029
LQ023	31.13847	15.42464	3.262796	0.250086	512.9024
LQ024	32.92576	12.46944	4.544853	0.11227	222.6847
LQ025	33.92095	13.49521	4.366529	0.15768	336.1056
LQ026	33.40556	15.25191	4.93346	0.175935	488.2353
LQ027	34.07508	16.0502	4.449482	0.160776	273.4242
LQ028	31.43156	13.17806	2.977771	0.173867	430.2229
LQ029	31.13248	11.80563	4.400769	0.128723	323.2561
LQ030	34.47766	12.56753	4.476065	0.138858	311.774
LQ031	36.64877	13.02868	4.199815	0.125266	273.8423
LQ032	35.02196	11.4437	4.314628	0.140705	354.9709
LQ033	31.56693	12.34497	4.026219	0.171055	364.5519
LQ034	32.96897	14.45164	2.695766	0.16727	296.7511
LQ035	34.93962	13.7114	2.787356	0.167081	384.4523
LQ037	31.20907	13.03479	3.699845	0.112336	205.3324
LQ038	33.82721	11.31124	4.108421	0.169401	369.1704
LQ039	32.47546	14.0895	3.623036	0.090296	202.8599
LQ040	28.59568	12.10084	3.059097	0.115084	204.5793
LQ042	31.46358	13.38478	2.537291	0.220964	585.1539
LQ043	30.74446	14.37748	2.586978	0.262612	625.151
LQ044	29.12076	11.00753	3.754109	0.207337	537.5508
LQ045	32.12187	12.56604	4.6695	0.18009	379.491

Genotype	GFP	HSW	RF	TOL	WHS
LQ046	36.66898	15.73186	4.319998	0.237859	571.6676
LQ047	32.19517	13.79321	4.638882	0.150534	257.7806
LQ048	26.67668	13.98653	4.423255	0.175626	375.9391
LQ049	31.89858	12.97778	3.64924	0.171825	360.5272
LQ050	35.37046	12.20294	3.361747	0.211895	505.3031
LQ051	28.45906	13.92245	4.392542	0.173816	281.2482
LQ052	33.63578	16.42279	3.014986	0.269039	519.7012
LQ053	33.53019	14.41491	3.290713	0.164553	444.4683
LQ054	33.75618	13.35993	4.267959	0.162625	362.3542
LQ055	33.98416	16.92737	4.429471	0.123221	230.4686
LQ056	37.21122	15.15551	4.660191	0.13039	337.5782
LQ057	32.3755	10.43077	4.722321	0.167017	356.9271
LQ058	31.42653	11.90148	3.220666	0.133726	292.76
LQ059	35.90113	12.81465	4.775956	0.091533	148.3693
LQ060	32.06611	11.49828	3.446502	0.213106	475.5841
LQ061	34.43416	12.66648	2.609281	0.191735	408.138
LQ062	29.86156	13.67059	4.457745	0.109195	243.6644
LQ063	28.24741	12.95595	4.998587	0.08225	170.2014
LQ064	33.36994	15.44775	3.12623	0.169502	352.0827
LQ065	31.02524	11.28858	2.967877	0.340234	670.074
LQ066	33.77645	14.21557	4.084044	0.247408	484.9052
LQ067	28.41022	10.88701	3.387238	0.218503	409.339
LQ068	36.28059	15.71754	3.934322	0.248282	519.3451
LQ070	32.26829	14.43929	4.363144	0.25137	507.089
LQ071	33.75898	13.911	5.222733	0.120511	242.2618
LQ072	29.78902	11.52081	3.739059	0.141486	350.3959
LQ073	35.31238	16.03086	3.549588	0.233833	631.0218
LQ074	29.4271	12.80371	2.721066	0.220134	458.0705
LQ075	36.62532	12.15372	3.524133	0.313377	706.4116
LQ076	30.76037	11.95672	3.684812	0.205279	461.8304
LQ077	33.65883	12.51229	3.313629	0.141121	379.2968
LQ078	30.26964	13.33958	3.974994	0.182001	362.3586
LQ079	27.02113	13.01189	4.444785	0.17835	327.1638
LQ080	26.03064	11.04832	3.369266	0.21083	426.2221
LQ081	33.50879	14.5059	4.241847	0.106208	159.3687
LQ082	35.15599	12.56029	2.041425	0.191637	568.0216
LQ083	31.15356	12.2321	3.239267	0.135654	332.7785
LQ084	32.58362	13.30392	4.137267	0.128931	281.7664
LQ085	35.36597	13.28904	4.608072	0.158409	230.3242
LQ086	28.76246	13.98388	3.927447	0.145736	308.1024
LQ087	32.0627	13.42157	2.538912	0.219218	438.976
LQ088	33.40568	13.71356	3.436722	0.135515	279.8101
LQ089	30.64462	11.98036	3.666636	0.275019	425.39
LQ090	34.8888	12.64867	4.783782	0.134472	273.5515
LQ091	31.87197	13.68099	4.22288	0.180267	350.7627
LQ092	32.30445	12.88157	4.000641	0.150143	380.0279

Genotype	GFP	HSW	RF	TOL	WHS
LQ093	29.70588	15.16196	4.920633	0.1597	229.3118
LQ094	33.23857	11.75136	3.985951	0.207879	396.6055
LQ095	37.05263	15.55039	4.2825	0.284538	834.9462
LQ096	33.84599	14.30412	4.530756	0.201966	437.7125
LQ097	32.22489	15.23195	5.059049	0.154497	308.1595
LQ098	30.16596	14.11741	4.558291	0.191005	447.5196
LQ099	26.96376	11.1048	4.427276	0.159245	263.4873
LQ100	28.05794	12.048	4.166699	0.152267	207.3403
LQ101	31.96474	13.42291	3.966862	0.18864	377.3926
LQ102	31.05466	10.83171	3.010115	0.224246	490.2498
LQ103	32.97415	13.61544	3.821706	0.145373	250.6317
LQ104	31.16309	11.51634	4.570942	0.148056	333.2819
LQ105	35.92127	13.98246	4.327272	0.145916	360.7098
LQ106	33.35465	12.18051	3.09414	0.194688	454.709
LQ107	34.31776	13.19274	3.451888	0.164882	318.9637
LQ108	29.10863	12.24483	4.11144	0.125956	214.7083
LQ109	36.75647	12.95675	3.340935	0.156811	321.1443
LQ110	34.94987	16.59068	3.743024	0.267083	620.144
LQ111	33.48066	14.12941	2.521244	0.226453	540.7526
LQ112	32.69477	12.85589	2.977497	0.126717	392.5544
LQ113	33.3278	12.81911	4.243282	0.172033	489.3141
LQ114	34.45812	13.1676	3.949815	0.203458	518.7687
LQ115	31.8465	13.97099	3.1399	0.216612	549.7661
LQ116	33.36326	13.61289	2.951803	0.172026	396.2547
LQ117	35.83828	13.19491	4.1256	0.203337	451.0635
LQ118	38.61048	14.24701	4.585069	0.179127	425.5578
LQ119	37.37401	11.38275	2.887359	0.244732	573.9211
LQ120	35.43554	11.90921	4.759315	0.170414	360.4433
LQ121	33.62272	15.17324	3.81555	0.209425	481.9116
LQ122	32.62084	13.59616	4.207424	0.213471	428.5028
LQ123	34.6898	15.08838	4.666901	0.236678	596.5688
LQ124	32.93033	14.43581	4.09935	0.124997	256.909
LQ125	36.81123	14.00448	4.848853	0.144988	298.2214
LQ126	31.17047	11.95857	3.38566	0.234116	651.3586
LQ127	33.65658	12.76607	3.731922	0.144509	344.3405
LQ128	32.71596	11.65874	4.04969	0.222869	325.5911
LQ129	32.34513	14.9887	3.828837	0.275741	569.2616
LQ130	32.59837	10.82583	3.980585	0.196248	371.3475
LQ131	36.24652	14.70787	4.784168	0.242208	537.2072
LQ132	31.10235	10.34106	3.799036	0.131842	199.7716
LQ133	30.8076	13.80933	4.643595	0.163267	299.2324
LQ134	31.8754	13.28827	4.121411	0.110884	200.5424
LQ135	33.58328	11.40068	3.265021	0.138001	299.5219
LQ136	34.68068	13.56427	3.768605	0.213028	598.329
LQ137	31.12472	12.32216	3.878056	0.128625	203.4101
LQ138	33.00904	12.8274	4.75689	0.119138	320.5537

Genotype	GFP	HSW	RF	TOL	WHS
LQ139	34.41004	13.88244	3.68401	0.231149	568.9714
LQ140	33.26367	14.55344	3.721661	0.134076	306.7846
LQ141	36.17969	16.6634	3.375923	0.26589	740.0479
LQ142	32.21553	12.88941	4.204847	0.126547	206.9828
LQ143	31.33495	14.53409	4.191934	0.096163	165.6307
LQ145	33.29601	14.57697	4.577135	0.157721	304.7703
LQ146	35.72532	13.65083	3.649981	0.186766	381.2969
LQ147	31.57991	12.24018	4.541989	0.134367	261.7048
LQ148	30.36248	11.85371	1.877129	0.203918	570.3
LQ149	32.45489	11.78471	2.73928	0.164761	466.3509
LQ150	33.78066	12.81965	2.628208	0.180465	478.4813
LQ151	30.69708	14.60055	2.877328	0.193427	352.0127
LQ152	33.54953	17.05414	3.695521	0.150149	293.3466
LQ153	34.39558	14.60974	3.926554	0.27733	520.7644
LQ154	32.18322	13.6372	3.957375	0.145557	199.8794
LQ155	35.98371	15.96632	3.395072	0.345749	764.9649
LQ156	31.08667	14.61971	4.001683	0.117289	254.0665
LQ157	29.52856	14.16934	3.67205	NA	63.31869
LQ158	36.11053	14.96598	3.67181	0.235646	564.33
LQ159	26.78669	NA	4.715604	0.194307	233.157
LQ160	33.26674	12.11332	4.349102	0.137588	296.3195
LQ161	33.37724	12.22466	1.895286	0.255117	572.4521
LQ162	32.94439	12.18182	4.202351	0.192866	373.4273
LQ163	30.41711	11.67797	3.340573	0.155641	316.2428
LQ164	29.84582	13.49858	4.06639	0.170941	355.8276
LQ165	29.79801	12.37368	2.721873	0.202324	459.2407
LQ166	34.44176	12.34096	4.302187	0.131166	235.0626
LQ167	34.9968	12.11104	4.018597	0.133452	343.3207
LQ168	32.24105	12.4978	3.16481	0.25244	594.0587
LQ169	34.4454	15.17184	4.492232	0.206792	457.66
LQ170	34.31915	12.64747	4.427782	0.187354	369.823
LQ171	33.92993	14.63883	3.89719	0.291116	590.8443
LQ172	30.2	14.70834	3.672712	0.296077	368.7497
LQ173	36.03886	13.15518	2.540521	0.313388	910.4987
LQ174	36.75774	13.18121	3.214577	0.205709	500.4925
LQ175	35.58227	12.7902	3.078368	0.253669	757.6428
LQ176	33.67338	12.50914	2.391292	0.22091	490.9944
LQ177	33.67752	12.95155	2.512386	0.230684	545.2774
LQ178	35.22521	12.48933	3.107331	0.170805	481.1844
LQ180	37.04792	13.91552	4.337625	0.231682	490.1788
LQ181	35.12857	12.86221	3.816611	0.16748	381.4419
LQ182	33.66539	15.10054	4.44477	0.134771	324.7426
LQ183	32.82144	12.7303	2.496651	0.209852	483.6493
LQ184	30.71773	13.66579	3.390788	0.24477	515.1134
LQ185	32.17996	16.92408	4.235466	0.221893	513.013
LQ186	29.2162	14.54626	4.6528	0.107579	144.1098

Genotype	GFP	HSW	RF	TOL	WHS
LQ187	27.96576	10.8227	3.903683	0.130686	207.2568
LQ188	31.56847	13.05814	3.769163	0.198048	395.2196
LQ189	29.03365	16.07488	2.847464	0.172132	337.2198
LQ190	32.826	12.57333	4.373785	0.207905	456.08
LQ191	37.50006	13.73849	4.622686	0.187866	455.6662
LQ192	31.72817	14.28262	3.499024	0.233969	470.7262
LQ193	30.28549	14.81096	3.756635	0.150129	264.1181
LQ194	31.01528	12.94813	2.968479	0.151758	408.3045
LQ195	33.64201	14.76359	2.271811	0.186387	482.89
LQ196	32.59951	13.58349	3.634037	0.187978	407.03
LQ197	37.28019	17.20255	3.736907	0.328553	839.0604
LQ198	29.02736	11.20019	2.065015	0.589058	1161.819
LQ199	30.34284	12.88631	3.998566	0.127203	158.2193
LQ200	32.65088	13.81195	4.812786	0.152227	316.8601
LQ201	30.17897	11.0141	2.830872	0.301036	568.575
LQ202	36.2842	11.55131	4.236305	0.227926	588.6177
LQ203	35.26744	12.46005	2.405679	0.25547	689.5909
LQ204	27.70145	11.24564	3.055204	0.187311	395.109
LQ205	35.58909	15.47341	2.888935	0.253528	696.1921
LQ206	35.82356	11.17331	3.331026	0.188715	494.5564
LQ207	30.9794	11.80577	3.169995	0.112308	273.535
LQ208	34.23461	13.63139	2.684556	0.164148	399.6574
LQ209	34.668	13.10604	2.768557	0.204919	496.6496
LQ210	35.08145	16.08484	3.560265	0.255305	440.3503
LQ211	31.41069	14.30371	2.807148	0.325593	594.1707
LQ212	29.23699	10.66947	3.129083	0.45529	736.323
LQ213	30.82638	13.24973	4.248935	0.139127	315.2581
LQ214	33.29657	13.19275	3.022927	0.206916	628.8393
LQ215	38.98101	15.21103	3.850829	0.195168	383.5299
LQ216	32.42784	16.79034	3.119936	0.249157	540.9858
LQ217	35.0418	15.26107	3.965339	0.255566	432.8988
LQ218	30.68849	10.95425	4.415997	0.148892	333.4732
LQ219	34.43486	14.60759	2.596207	0.273135	474.92
LQ220	31.68349	10.57745	3.221777	0.19089	508.1068
LQ221	33.65677	12.70073	4.78751	0.112324	177.1216
LQ222	28.64973	12.1648	4.930639	0.10273	156.7358
LQ223	28.79188	12.93182	3.115546	0.131044	326.119
LQ224	30.7271	12.92036	3.876681	0.187162	398.894
LQ225	32.70256	11.77886	3.419471	0.203494	400.4446
LQ226	28.85478	12.60926	4.67787	0.211955	359.7316
LQ227	30.91075	14.07278	4.22134	0.139672	232.4193
LQ228	31.64167	13.68253	4.127367	0.178455	400.1585
LQ229	31.09169	13.27936	3.786259	0.22533	478.5126
LQ230	32.33979	13.21557	3.425566	0.285932	459.0064
LQ231	30.62977	13.53839	2.670401	0.154954	354.1114
LQ232	32.4922	15.75196	4.691444	0.141173	268.7969

Genotype	GFP	HSW	RF	TOL	WHS
LQ233	32.79733	12.34565	2.556403	0.173305	373.2142
LQ234	32.19091	13.71533	4.392691	0.141634	278.116
LQ235	28.87841	11.40849	3.826969	0.215935	534.9638
LQ236	32.18456	13.02165	2.666966	0.230375	646.1753
LQ237	34.80577	16.06824	3.695041	0.243828	523.4839
LQ238	30.68476	13.31758	4.56043	0.152097	200.6446
LQ239	35.36032	15.43987	3.38501	0.305583	821.6224
LQ240	33.82375	15.35695	3.976361	0.270475	722.5142
LQ241	32.01756	12.66874	3.891772	0.256137	505.6732
LQ242	30.14812	15.55152	3.330202	0.110846	301.2691
LQ243	30.88066	13.29604	4.621172	0.103753	244.0838
LQ244	36.1897	13.41187	4.323515	0.217727	614.0464
LQ245	28.94574	13.64427	3.287179	0.154474	340.84
LQ246	33.33592	11.10882	3.568874	0.27252	605.1694
LQ247	32.44814	11.53852	3.924078	0.147125	362.8053
LQ248	29.6405	11.11828	3.384611	0.139191	211.8438
LQ249	35.92437	13.24224	4.656212	0.151215	320.051
LQ250	30.48566	12.57116	3.709316	0.107257	232.6431
LQ251	30.30549	11.69043	3.829222	0.139891	261.3052
LQ252	32.47005	13.72909	1.883552	0.276443	626.0015
LQ253	29.41041	12.08882	3.839115	0.107324	262.6217
LQ254	34.73643	16.14066	3.067978	0.262073	648.4566
LQ255	32.58369	14.00118	4.118491	0.164544	337.4495
LQ256	36.89187	14.28335	4.203596	0.234773	536.3526
LQ257	31.87145	12.55176	2.739181	0.196038	485.0405
LQ258	34.20871	13.15155	4.634596	0.142988	301.6299
LQ259	31.11909	12.35756	4.494274	0.16197	399.2515
LQ260	31.86502	12.13432	4.073081	0.184711	468.5594
LQ261	31.42916	13.45786	3.324499	0.175821	352.9989
LQ262	30.33775	12.96666	4.701952	0.185712	373.0857
LQ263	33.30082	14.72357	4.000653	0.210722	453.9323
LQ264	32.83877	13.94936	3.985181	0.221253	454.0865
LQ265	33.41195	13.41626	2.654457	0.204944	423.2836
LQ266	31.0008	14.62949	3.492314	0.156695	245.7116
LQ267	34.34289	11.85531	2.826946	0.175386	563.5575
LQ269	32.42002	12.97045	4.798741	0.131974	370.408
LQ270	34.80334	13.56083	2.99095	0.218206	584.8021
LQ271	32.54753	13.16357	3.962567	0.121063	307.4716
LQ272	33.31231	12.69643	3.353435	0.17491	349.0563
LQ273	30.79919	12.34171	3.770387	0.101361	206.0689
LQ274	32.73506	15.32671	4.333181	0.181698	348.8213
LQ275	32.83875	12.96957	4.150081	0.072049	148.7939
LQ276	32.1553	14.92891	2.815163	0.125045	241.5441
LQ277	34.6524	11.78935	4.094921	0.213623	500.937
LQ278	34.75402	16.29154	3.935573	0.195465	427.3703
LQ279	34.44195	13.2763	3.302087	0.157759	387.0772

Genotype	GFP	HSW	RF	TOL	WHS
LQ280	27.65384	13.37859	3.155658	0.171059	309.2436
LQ281	30.2369	13.89944	3.938644	0.250825	291.4805
LQ282	31.82168	13.82982	3.912207	0.183914	423.1087
LQ283	26.02433	13.90709	3.874577	0.115529	193.3543
LQ284	33.76583	12.10618	4.067386	0.167993	338.5071
LQ285	28.96202	10.74716	4.583298	0.114034	288.5092
LQ286	33.77332	15.32273	4.370868	0.206378	440.5683
LQ287	29.79377	13.02034	4.817613	0.14319	252.666
LQ289	32.06934	12.10379	4.503638	0.189285	313.5338
LQ290	37.50319	15.55154	4.167116	0.144633	278.6378
LQ291	31.65206	13.12348	4.2445	0.23382	565.7791
LQ292	36.00353	13.11576	4.917003	0.147825	245.7144
LQ293	27.81428	14.0059	3.281738	0.179008	147.5275
LQ294	34.95805	13.71601	2.416585	0.247987	512.6105
LQ295	32.15532	12.10562	3.663774	0.253007	466.5706