POLLYANNA CAPOBIANGO DA FONSECA

GENOME-WIDE ASSOCIATION ANALYSIS FOR SOYBEAN BACTERIAL PUSTULE RESISTANCE, FUNCTIONAL ANALYSIS, AND GENE NETWORKS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2019

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Câmpus Viçosa

T
F676g
2019

Fonseca, Pollyanna Capobiango da, 1991-
Genome-wide association analysis for soybean bacterial pustule resistance, functional analysis, and gene networks / Pollyanna Capobiango da Fonseca. – Viçosa, MG, 2019.
vi, 41 f. : il. (algumas color.) ; 29 cm.

Texto em inglês.
Orientador: Felipe Lopes da Silva.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. *Xanthomonas axonopodis* pv. *glycines*. 2. *Glycine max*. 3. Estudo de Associação Genômica Ampla. I. Universidade Federal de Viçosa. Departamento de Fitotecnia. Programa de Pós-Graduação em Fitotecnia. II. Título.

CDD 22. ed. 632.32

POLLYANNA CAPOBIANGO DA FONSECA

GENOME-WIDE ASSOCIATION ANALYSIS FOR SOYBEAN BACTERIAL PUSTULE RESISTANCE, FUNCTIONAL ANALYSIS, AND GENE NETWORKS

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Fitotecnia, para obtenção do título de *Magister Scientiae*.

APROVADA: 19 de fevereiro de 2019.

Eveline Teixeira Caixeta

Gaspar Mafone

Lucas Lima Verardo
(Coorientador)

Felipe Lopes da Silva
(Orientador)

AGRADECIMENTOS

Senhor, agradeço pelas dificuldades, me fizeram Te ouvir com mais disciplina e humildade.

Agradeço ao professor Felipe Lopes da Silva, pela confiança e ensinamentos. Aos professores Tiago Mendes e Lucas Verardo, pelo tempo e aprendizado. Aos professores Márcio Barbosa e José Rogério de Oliveira, pelo uso de equipamentos/instalações.

Agradeço também aos membros do Programa Soja, pela amizade, colaboração e convivência. Ao CNPq, pela concessão da bolsa de estudos, e ao Departamento de Fitotecnia, pela oportunidade de realização do curso. Assim como aos pesquisadores citados, os quais desejo coragem para perseverar em suas pesquisas.

Aos amigos, Itamar, Luciana, João Filipe, Ana, July e Matheus, muito obrigada.

À família – Cristina, principalmente – que me fez acreditar no impossível através da imaginação e dos sonhos, que um dia se tornaram realidade. Desde os velhos tempos, com a minha mente vou a mil lugares e a imaginação me dá forças pra voar.

"Para onde poderia eu escapar do teu Espírito?
Para onde poderia fugir da tua presença?
Se eu subir aos céus, lá estás;
se eu fizer a minha cama na sepultura, também lá estás.
Se eu subir com as asas da alvorada
e morar na extremidade do mar,
mesmo ali a tua mão direita me guiará e me susterá.
Mesmo que eu diga que as trevas me encobrirão,
e que a luz se tornará noite ao meu redor,
verei que nem as trevas são escuras para ti.
A noite brilhará como o dia,
pois para ti as trevas são luz."

Sl 139:7-12

SUMMARY

ABSTRACT

FONSECA, Pollyanna Capobiango da, M.Sc., Universidade Federal de Viçosa, February, 2019. **Genome-wide association analysis for soybean bacterial pustule resistance, functional analysis, and gene networks.** Adviser: Felipe Lopes da Silva. Co-advisers: Tiago Antônio de Oliveira Mendes and Lucas Lima Verardo.

Bacterial pustule of soybean (*Xanthomonas axonopodis* pv. *glycines*) is a foliar disease leading to symptoms such as yellow halos. Potential soybean protein-coding genes related to bacterial pustule infection were investigated aiming future application on soybean breeding programs. XAG2440^{p} and XAG2447 strains were used to inoculate 109 Brazilian cultivars on greenhouse, and the phenotyping data used on association analysis and gene mining. Network analysis was proceeded with two different association analysis, so that genes and transcription factors (TFs) with more connections were highlighted. Cultivars inoculated with XAG2447 and XAG2440^{p} not presented and presented 82.57% of susceptibility symptoms, respectively. Cultivars genotyped by 6,000 SNPs were used on the association analysis revealing 13 SNPs significantly associated with soybean bacterial pustule. Of these, eight are inside genes – 3 inside coding regions of LOC100779077, histone-lysine N-methyltransferase SUVR4, and ABC transporter B family member 9 proteins, and 5 outside of genes. Despite the fact that association analysis showed different marker regions in linkage disequilibrium, Gene-TFs network provided evidence that key TFs might regulate genes on both association analysis, for example, the highly connected ERF011, WRKY8, WRKY33, and MYB30. In addition, shikimate metabolic process, regulation of peroxisome size, response to ozone, and triglyceride catabolic process appear to play an important role on soybean immunity against the foliar pathogen. These results confirmed that disease screening symptoms of XAG2447 are less aggressive than of XAG2440^{p} and that soybean pustule resistance is a highly complex trait with defense-related protein-coding candidate genes and TFs possibly acting upon soybean defense mechanisms.

RESUMO

FONSECA, Pollyanna Capobiango da, M.Sc., Universidade Federal de Viçosa, fevereiro de 2019. **Associação genômica ampla para resistência da soja à pústula-bacteriana, análise funcional e redes gênicas.** Orientador: Felipe Lopes da Silva. Coorientadores: Tiago Antônio de Oliveira Mendes e Lucas Lima Verardo.

A pústula bacteriana (*Xanthomonas axonopodis* pv. *glycines*) é uma doença foliar da cultura da soja que leva à formação de halos amarelos. Genes potenciais da soja relacionados à infecção do hospedeiro pela pústula bacteriana foram investigados visando futura aplicação em programas de melhoramento de soja. As estirpes XAG2440^{p} e XAG2447 foram utilizadas para inocular 109 cultivares nacionais de soja em casa de vegetação e os dados de fenotipagem empregados na análise de associação genômica ampla e mineração gênica. A análise de redes foi realizada com dados de associação genômica, de modo que genes e fatores de transcrição (FTs) com maior número de sítios de ligação foram destacados. Cultivares inoculadas com XAG2447 e XAG2440^{p} não apresentaram e apresentaram 82,57% de grau de sintomas, respectivamente. Cultivares genotipadas por 6.000 marcadores foram utilizadas na análise de associação, que revelaram 13 SNPs significativamente associados à infecção por *Xanthomonas axonopodis* pv. *glycines*. Destes, 8 estão dentro de genes - 3 dentro de regiões codificadoras das proteínas LOC100779077, histona-lisina N-metiltransferase SUVR4 e transportador ABC membro da família 9, e 5 fora de genes. A análise de associação genômica de dois experimentos mostraram diferentes regiões em desequilíbrio de ligação. No entanto, a rede Gene-FTs forneceu evidências de que FTs chave podem regular genes procedentes de ambos conjuntos gênicos, por exemplo, os FTs altamente conectados ERF011, WRKY8, WRKY33 e MYB30. Além disso, os processos de metabolismo do shikimato, tamanho do peroxissomo, resposta ao ozônio e catabolismo de triglicerídeo parecem desempenhar um papel importante na resposta imune da soja contra este patógeno foliar. Os resultados apresentados mostram que a estirpe XAG2447 é menos agressiva no *screening* de cultivares e que a resistência da soja à pústula bacteriana é uma característica altamente complexa, com genes candidatos e TFs possivelmente apresentando papéis importantes nos mecanismos de defesa da soja.

GENERAL INTRODUCTION

The soybean [*Glycine max* (L.) Merril] is an annual diploid species, originated from the ancestral *Glycine soja* (Seib. Et Zucc.) and cultivated since approximately 5,000 years ago in China. Expansion of the cultivated area and domestication by Chinese farmers contributed to the increase of the cycle variability and development of landraces, ancestral cultivars that have relatively high genetic diversity (WILSON, 2008). Along with the introduction of Asian landraces to present-day cultivated areas, 78% of rare alleles were lost due to genetic bottleneck and the reduction of effective population size. Besides, soybean diversity was reduced by 50% compared to the ancestral (HYTEN et al., 2006) certainly contributing to the reduction of genetic diversity present on the cultivars currently developed. Thus, selective pressures play a very important role in the formation of the phenotype and genetic diversity (MEYER; PURUGGANAN, 2013), as well as the evolutionary processes mentioned above.

The observations considered above are of even greater importance when we investigate the co-evolution of plants and pathogens, which immune and resistance mechanisms present in the plant can be supplanted by the pathogen. In bacteria, for example, highly virulent strains need to be able to adapt to the local microbiota and provoke an infectious process in the host cells. Soybean bacterial pustule (*Xanthomonas axonopodis* pv. *glycines*), one of the main bacterial diseases occurring in soybean, causes yellowish spots on the leaves, that coalesce with disease progression (PARK et al., 2008). The main control method for bacterial pustule infection is the use of resistant cultivars. However, it has been observed the increase of its incidence and aggressiveness on soybean cultivated areas.

The recessive rxp gene, located on soybean chromosome 17 (NARVEL, 2001), is the only resistance gene extensively studied for bacterial pustule. However, subsequent studies have shown that resistance to bacterial pustule is a quantitative trait. For example, Goradia *et al.* (2009), established five severity classes. Besides, Kim *et al.* (2010a) indicate other chromosomes as important in plant infection and defense, for instance, chromosomes 5, 4, 10 and 19. Thus, the identification of new genes and resistance mechanisms related to the defense process of soybean to bacterial pustule is of extreme importance for the development of resistant cultivars.

One of the tools suitable and widely used for this purpose is the Genome-wide association studies (GWAS), based on linkage disequilibrium and Simple Linear Regression for the identification of significant single-base polymorphisms (SNPs). The advancement of GWAS was possible due to the development of easily automated and high efficient genotyping technologies (SONG et al., 2013), as *microarrays*.

Specifically for soybeans, the SoySNP6K (*Illumina Infinium BeadChip*) with 6,000 SNPs was used to construct a high density linkage map based on a single population and emphasizing the segregation data of each SNP marker with a success rate of 96.75%, which makes this map very accurate in terms of distribution of markers and (AKOND et al., 2013). In another study, SoySNP6K was used to discover genomic regions of importance for quantitative traits, such as seed weight, plant height, and soybean seed production through the association of haplotypes on several environments (CONTRERAS-SOTO et al., 2017). Thus, it was verified the efficacy of these markers to distinguish haplotypes.

Therefore, the functional analysis of allelic variants and the identification of genes possibly associated is of great importance for the development of soybean cultivars bacterial resistant. On the other hand, network analysis may help to clarify which mechanisms are associated with host defense process (SHEPPARD; GUTTMAN; FITZGERALD, 2018), besides making possible the study of the interactions of transcription factors regulating the expression of these genes.

Thus, the objectives of this study were: *i.* Soybean cultivars phenotyping and GWAS analysis for soybean bacterial pustule *ii*. Assign a biological function to the identified SNPs *iii*. Gene networks analysis.

REFERENCES

AKOND, M. et al. A SNP-Based Genetic Linkage Map of Soybean Using the SoyS - NP6K Illumina Infinium BeadChip Genotyping Array. **Plant Genetics, Genomics, and Biotechnology**, v. 1, n. 3, p. 80–89, 2013.

BOS, J. M.; OMMEN, S. R.; ACKERMAN, M. J. **Genetics and Genomics of Soybean**. [s.l: s.n.].

CONTRERAS-SOTO, R. I. et al. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-Based haplotype analysis. **PLoS ONE**, v. 12,

n. 2, p. 1–22, 2017.

GORADIA, L.; HARTMAN, G. L.; DANIEL, S. L. Evaluation of glyphosate-tolerant soybean cultivars for resistance to bacterial pustule. **European Journal of Plant Pathology**, v. 124, n. 2, p. 331–335, 2009.

HYTEN, D. L. et al. Impacts of genetic bottlenecks on soybean genome diversity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 45, p. 16666–16671, 2006.

KIM, D. H. et al. Fine mapping of a resistance gene to bacterial leaf pustule in soybean. **Theoretical and Applied Genetics**, v. 120, n. 7, p. 1443–1450, 2010.

MEYER, R. S.; PURUGGANAN, M. D. Evolution of crop species: genetics of domestication and diversification. **Nature Reviews Genetics**, v. 14, n. 12, p. 840–852, 2013.

NARVEL, J. M. Molecular Mapping of Rxp Conditioning Reaction to Bacterial Pustule in Soybean. **Journal of Heredity**, v. 92, n. 3, p. 267–270, 2001.

PARK, H. J. et al. Avirulence gene diversity of Xanthomonas axonopodis pv. glycines isolated in Korea. **Journal of Microbiology and Biotechnology**, v. 18, n. 9, p. 1500–1509, 2008.

SHEPPARD, S. K.; GUTTMAN, D. S.; FITZGERALD, J. R. Population genomics of bacterial host adaptation. **Nature Reviews Genetics**, v. 19, n. 9, p. 549–565, 2018.

SONG, Q. et al. Development and Evaluation of SoySNP50K, a High-Density Genotyping Array for Soybean. **PLoS ONE**, v. 8, n. 1, p. 1–12, 2013.

ARTICLE 1 - GENOME-WIDE ASSOCIATION ANALYSIS IDENTIFIES NEW CAUSAL VARIANT LOCI AND CANDIDATE GENES FOR SOYBEAN BACTERIAL PUSTULE RESISTANCE

INTRODUCTION

Brazilian soybean [*Glycine max* (L.) Merril] agro-industrial complex is one of the largest in the world, generating products such as bran intended for animal feed and oil used in biodiesel production, among many other industrial uses. Economic loss only in result of disease incidence from 2010 to 2014 accounted $60.66 USD per acre in the United States and Ontario, Canada (ALLEN et al., 2017). Because of that, to study disease resistance genes and defense mechanisms against damaging pathogens is crucial for the development of highly productive cultivars. Bacterial pustule, caused by the gram-negative phytobacterium *Xanthomonas axonopodis* pv. *glycines*, is one of the most important bacterial soybean pathogens. Small yellowish halos are the symptoms produced by its infection, which coalesce with disease progression, reaching necrosis and premature defoliation of soybean leaves (PARK et al., 2008). Severity as high as 70% occurs when environmental conditions, such as high temperature and humidity, required for colonization and pathogen development are attained (ZINSOU et al., 2015). Although new treatment methods are being continuously tested, such as thymol nanoemulsion, which reduces bacterial activity and potentiates soybean growth (KUMARI et al., 2018), the use of resistant cultivars is the main method used for bacterial pustule prevention. Despite this fact, it has been observed an increase in the incidence of bacterial pustule in cultivars considered disease resistant. This outcome could be the result of virulent strains present on the environment and/or resistance suppression, augmenting the importance of studying soybean resistance genes.

Meanwhile, continued population growth rises the demand for soy-based products. It requires high-efficient strategies of structural organization and market planning aside from technologies advanced enough to enable productivity and competitive advantage of this annual diploid species cultivars. One of these technologies is the use of molecular breeding with the aim to develop varieties with desirable agronomic traits, such as disease resistant crops. In addition, understanding the genetic basis of plant-pathogen response and plant defense mechanisms, including better elucidate resistant genes, is essential to a target-specific and effective molecular breeding program.

The only source of resistance extensively studied against *Xanthomonas axonopodis* pv. *glycines* is the *Rxp* gene, located on soybean chromosome 17 (NARVEL, 2001). It is not known if the recessive *rxp* gene negatively regulates defense mechanisms of soybean or if the bacterium requires these resistance genes in recessivity for its development and reproduction (KIM et al., 2010a). Despite the existence of host-pathogen interactions considering only one gene, several subsequent studies have shown that resistance to bacterial pustule is a complex trait. It means that small effect genes may regulate bacterial pustule resistance with high environmental influence. For example, on the study of glyphosate resistant soybean characterization of bacterial pustule aggressiveness developed by Goradia et al. (2009), five severity classes were determined using chlorosis intensity and number of pustules as points of reference. Another study used RNA-seq in near-isogenic lines resistant and susceptible to bacterial pustule, revealing 1978 up-regulated and 783 down-regulated genes (KIM et al., 2011). Those results indicate the existence of various genes controlling resistance to bacterial pustule, although it could not be inferred which plays the most important roles in plant defense to be applied on molecular breeding programs.

Based on linkage disequilibrium of the phenotype with SNP marker in a great number of individuals (MEUWISSEN; HAYES; GODDARD, 2001), Genome-wide association study (GWAS) is a statistical tool that allows the identification of significant single-base polymorphisms (SNPs) by Simple Linear Regression. It can allow the identification of core genes and SNP haplotypes controlling a trait of interest, to be introduced on elite breeding lines used in crop improvement. Numerous studies have been developed using GWAS, for different pathosystems, as northern leaf blight in maize (POLAND et al., 2011), spot blotch in barley (GYAWALI et al., 2018), witches' broom disease and frosty pod rot disease in cacao (MCELROY et al., 2018), stem rot in soybean (IQUIRA; HUMIRA; FRANÇOIS, 2015) and cyst nematode (VUONG et al., 2015) also in soybean. GWAS data has been used on the identification of genes and QTLs (quantitative trait loci) (FANG et al., 2017), and more recently, on the development of metabolic networks (KANG et al., 2018), and the association between biological processes involved in the characterization of a given phenotype (LIU et al., 2017).

Bacterial pustule is one of the diseases required by the Brazilian Department of Agriculture to be screened on DUS (Distinctness, Uniformity, and Stability) tests before the cultivars are released. Inoculations are recommended to be proceeded with strains from *Instituto Biológico de Campinas*, XAG2440^{p} and XAG2447. Thus, the objective of this study was to investigate the genetic basis of soybean resistance to bacterial pustule

through cultivars phenotyping with recommended strains and association analysis. Followed by the characterization of associated SNPs and indication of nearby protein-coding genes of possible great importance for the development of resistant cultivars in a breeding program.

MATERIAL AND METHODS

Location and experimental design

A total of 109 soybean cultivars from various Brazilian breeding programs (Supplementary Table S1) were sown in 200 mL plastic cups filled with substrate and placed in a greenhouse located at the Diogo Alves de Melo Experimental Field, Viçosa, MG (648 m altitude; 20°45' S latitude; and 42°52' W longitude). The experiment was conducted using completely randomized design from May to July 2018 with three replications. Greenhouse conditions were maintained about 70% of humidity and 28°C.

Bacterial strains, inoculation and evaluation

Strains XAG2440^{p} (pathovar) and XAG2447 of *Xanthomonas axonopodis* pv. *glycines* were obtained from Instituto Biológico de Campinas. They were cultured in 523 medium plates (KADO; HESKETT, 1970) at 28°C for 48h, and suspended on 10mM $MgCl_2$ solution until OD_{600} of 0.3 (10^8 CFU/mL) prior to inoculation. Phenotyping experiments were independent for each strain, but the same inoculation protocols were applied. A humid chamber was simulated inside the greenhouse 24h in advance until 24h after inoculation, so that a higher humidity could be accessed and favored soybean colonization by the pathogen (Figure 1A). Inoculation and evaluation of symptoms proceeded as described by Goradia et al. (2009), where a needle of 2.9 cm x 0.65 mm was used for perforation, counting a total of 28 spots per leaflets (Figure 1B). On the surface of the R3/R4 trifoliolate leaves, 3.5 mL inoculum suspension was pulverized utilizing an atomizer on each plant. As a control, five cultivars were pulverized with 10mM $MgCl_2$ solution. Assignments of scores varying from 1 to 5 were performed nine days after inoculation, so that higher scores represented greater severity of symptoms, according to methodology developed by Goradia et al. (2009) (Table 1). Scores were analyzed through ranking and mode.

Table 1. Scale of *Xanthomonas axonopodis* pv. *glycines* severity symptoms according with evaluation methodology developed by Goradia et al. (2009).

Scores	Symptom
1	Resistant
2	Moderately resistant
3	Moderately susceptible
4	Susceptible
5	Highly susceptible

Genotyping and association analysis

All 109 cultivars were genotyped by 6,000 SNPs. Data were obtained with the SoySNP6K BeadChip (AKOND et al., 2013), a high-throughput genotyping platform by Illumina (Illumina Inc., San Diego, USA). Genotyping was run by Deoxi Biotechnology Ltda®, in Aracatuba, São Paulo, Brazil, and the same data set used on other studies, for instance, GWAS for agronomic traits (CONTRERAS-SOTO et al., 2017). Software Tassel 5 (BRADBURY et al., 2007) was used on data processing, with a maximum of 10% of lost genotypes (call rate), and greater than 5% of minor allele frequency (MAF).

Prediction of genotypic values in soybean cultivars used on this study was proceeded using Linear Mixed Model with Software Selegen (DEON, 2016), and its results applied on Genome-wide association analysis through Genome Association Prediction Tool, the GAPIT package, in software R (LIPKA et al., 2012). Population structure and kinship coefficient matrix used on the association analysis was inferred as described by CONTRERAS-SOTO et al. (2017). Manhattan plots were obtained with qqman 0.1.4 R package (TURNER, 2018).

Haplotype block and annotation of nearby genes

Statistically significant SNPs were mapped into linkage disequilibrium blocks with Haploview 4.2 (BARRETT et al., 2005), and the extension of blocks estimated. Proteins within block were annotated with Augustus 3.2.2 (STANKE; MORGENSTERN, 2005) using Arabidopsis genome as reference, and Gene Ontology retrieved with AgBase-GOanna (MCCARTHY et al., 2006).

Flanking sequences of each significant SNP associated with bacterial pustule, with 2000 bp downstream and upstream, was compared with annotation reference (Williams 82 assembly V2.0, NCBI RefSeq assembly accession: GCF_000004515.5) (O'LEARY

et al., 2016) using translated BLAST (blastx) (JOHNSON et al., 2008). Then, Open Reading Frames (ORFs) were identified with the web tool Expert Protein Analysis System (ExPASy) (GASTEIGER et al., 2003) and sequences alignment made with Clustal Omega Multiple Sequence Alignment (SIEVERS et al., 2011) to verify the identity of the sequence compared with database information.

RESULTS

Cultivars screening for bacterial pustule

The set of 109 cultivars was inoculated with XAG2440^{p} and XAG2447, *Xanthomonas axonopodis* pv. *glycines* strains, in separated experiments. Since scores values designed for each plant had a short extent of values, the ranking of the average scores of three repetition for each cultivar suited better the descriptive analysis of the non-parametric data. Considering the inoculation of XAG2447, the average score of the cultivars ranged from 1 to 1.67, with 76.15% of the cultivars classified as resistant. The remaining cultivars presented a few symptoms, and consequently were classified as moderately resistant, such as FUNDACEP 57 RR, CD 249 RR STS, CD 2630 RR, and CD 233 RR. In contrast, XAG2440^{p} inoculation resulted on a wider range of scores (Table 2, Figure 1C, 1D), with 17.43% of the cultivars categorized as resistant/moderately resistant, 35.78% as moderately susceptible, and 46.79% as susceptible/highly susceptible, with TMG 7161 RR and TMG115 RR presenting least severe symptoms. Also, the reference resistant cultivar of the Brazilian Department of Agriculture, BRS 133, was found to be resistant with XAG2447 screening, but susceptible with XAG2440^{p} strain inoculation. Complete data available on Supplementary Table S2.

Table 2. Ranking of scores for XAG2440^{p} strain screening evaluation.

Ranking XAG2440^{p} scores	% Individuals	Score average range	Cultivars mode score
Resistant - Moderately Resistant	17.43	1.33 - 2.66	2.66
Moderately susceptible	35.78	3.00 - 3.66	3.00
Susceptible - Highly susceptible	46.79	4.00 - 4.66	4.00

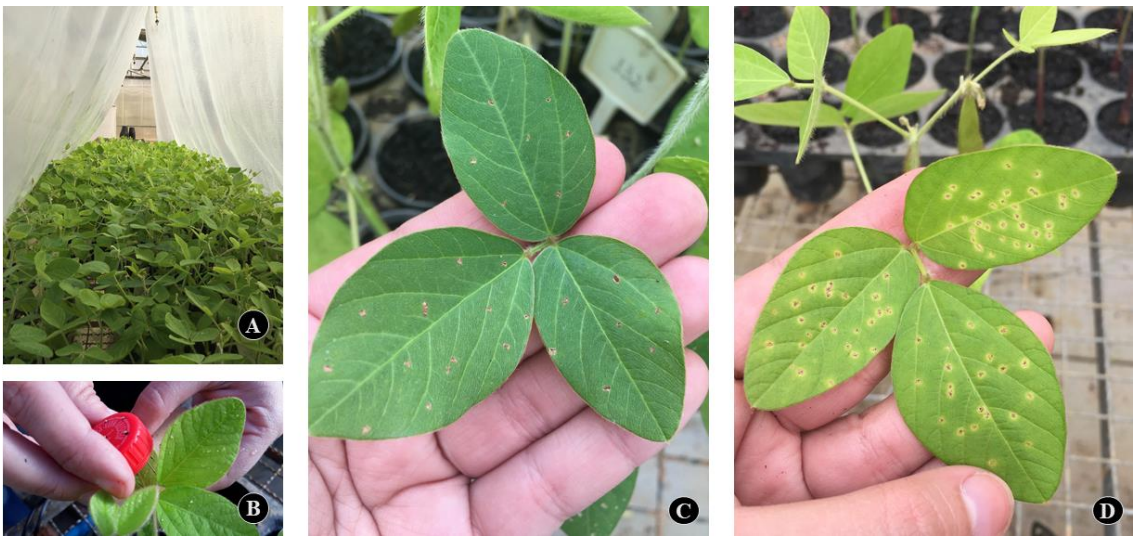


Figure 1. Inoculation and symptoms of soybean bacterial pustule. (A) Humid chamber simulation. (B) Needle perforation for inoculation procedure. Representative cultivars of a resistant TMG 7161 RR (C) and a susceptible genotype CD 2860 (D) symptoms of trifoliolate leaves of soybean to XAG2440^{p} strain.

GWAS

After genotyping data processed, 3807 polymorphic SNPs were found to attend MAF > 0.05 and call rate < 0.1 in the population. Distribution of these SNPs was on average 190.35 markers per chromosome, with a minimum of 140 markers on chromosome 1 and the maximum of 266 on chromosome 13, the most well represented. Associations were regarded as significant when p-values were < 0.003, equivalent to $-\log_{10}$(p-value) ≥ 2.5, a threshold found to be the most adequate due to the low p-values encountered in the analysis of this skewed distribution of phenotyping scores. Because there is no notable variability on soybean genotypes response to XAG2447 strain, statistical parameters adjustment would probably not overcome the biological bottleneck of response diversity, not preventing the occurrence of type I error in its association analysis. Therefore, GWAS was only conducted with strain XAG2440^{p} phenotyping data, resulting on 13 non-redundant SNP markers across six chromosomes significantly associated with bacterial pustule (Table 3). From 13 SNPs, five are located on chromosome 18, three on chromosome 13, two on chromosome 3, and one on chromosomes 5, 10, and 15, as illustrated by the Manhattan plot (Figure 2).

Table 3. Significantly associated markers identified on association analysis of soybean bacterial pustule with strain XAG2440^{p}.

SNP	Allele	Chr[1]	Position (bp)[2]	MAF[3]	r^{2*}	-$\log_{10}$ (p-value)
Gm18_4188718_A_C	A/C	18	4188718	0.087156	0.203414	3.337415
Gm10_48426970_A_G	A/G	10	48426970	0.073394	0.197765	3.212595
Gm13_34818193_C_T	T/C	13	34818193	0.155963	0.187272	2.977933
Gm18_3954704_C_T	T/C	18	3954704	0.114679	0.184138	2.907099
Gm13_898111_T_G	T/G	13	898111	0.105505	0.177376	2.753018
Gm13_34946643_T_C	T/C	13	34946643	0.142202	0.176397	2.730578
Gm03_46592189_A_G	A/G	3	46592189	0.467890	0.175837	2.717708
Gm03_46889507_T_C	T/C	3	46889507	0.463303	0.175584	2.711906
Gm18_54979_G_A	A/G	18	54979	0.110092	0.172587	2.642848
Gm18_122382_A_G	A/G	18	122382	0.110092	0.172587	2.642848
Gm18_4324818_G_T	T/G	18	4324818	0.082569	0.171225	2.611343
Gm15_6925513_T_C	T/C	15	6925513	0.339450	0.169561	2.572738
Gm05_33176582_G_A	A/G	5	33176582	0.220183	0.168298	2.543374

[1] Chromosome

[2] SNP position on *Glycine max* genome assembly version Glyma.Wm82.a1 (Gmax1.01)

[3] Minor allele frequency (MAF)

[*] Coefficient of Determination Model with SNP

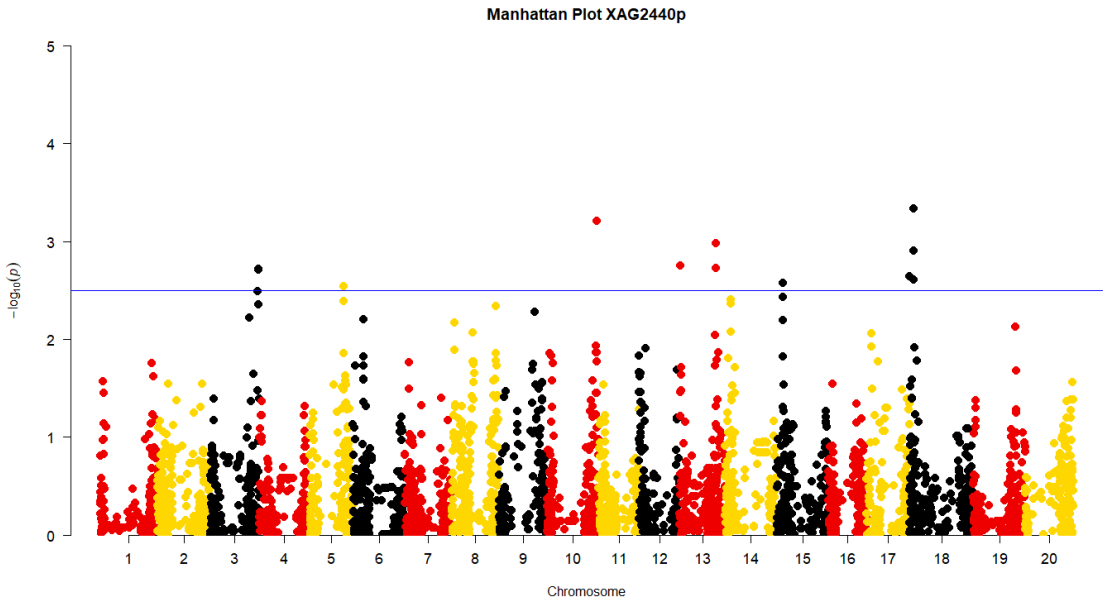


Figure 2. Manhattan plot of XAG2440^{p} GWAS.

The most significantly SNPs associated with bacterial pustule were Gm18_4188718_A_C on chromosome 18 and Gm10_48426970_A_G on chromosome 10. They contributed, respectively, to 20.34% and 19.78% of the phenotypic diversity, and exhibit low values of MAF, indicating their low allele frequency. For chromosome 3, Gm03_46592189_A_G and Gm03_46889507_T_C are about 297318 bp distant and contribute, separately, to 17.5% of the phenotypic diversity. Each of them had approximately, 0.46 of MAF, the greatest values of the subset. Linkage disequilibrium (LD) blocks determination by Haploview showed Gm03_46592189_A_G and Gm03_46889507_T_C within a disequilibrium block ranged from Gm03_46337339_T_C until Gm03_47039930_A_G, with 702 kb (Figure 3). Arabidopsis correspondent and defense-related genes identified within the LD block are VIRF INTERACTING PROTEIN 4 (AT5G28040; GO:0042742), RPI3 (AT3G04790; GO:0042742), PEPTIDE-N-GLYCANASE 1 (AT5G49570; GO:0010188), PATHOGENESIS-RELATED 4 (AT3G04720; GO:0042742, GO:0050832, GO:0080027), and NON-RESPONDING TO OXYLIPINS 7 (AT1G64790; GO:0042742, GO:0045087), among other defense-related genes. Chromosome 3 was the only to form a LD block to have significant associated SNP within it.

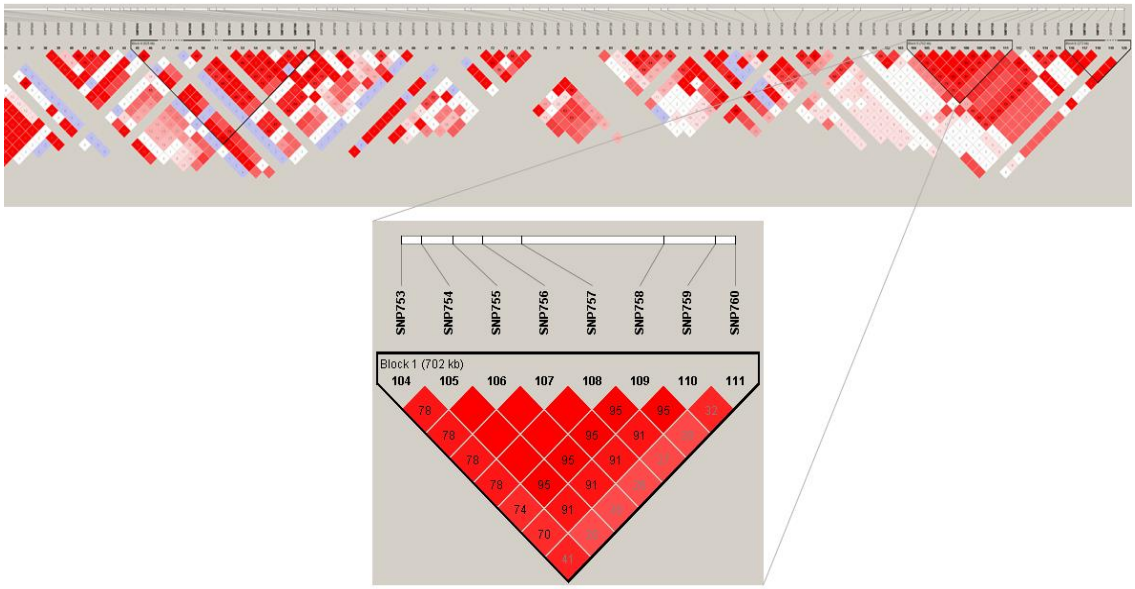


Figure 3. LD block from Gm03_46337339_T_C (SNP753) to Gm03_47039930_A_G (SNP760), containing chromosome 3 markers Gm03_46592189_A_G (SNP757) and Gm03_46889507_T_C (SNP758) significantly associated with bacterial pustule, LD in r^2.

Table 4. Nearby significant SNPs sequence annotation and its gene position (inside or outside of gene and coding region). Includes Resistant and Susceptible allele-cultivars average score and score difference between them.

SNP	Chr[1]	Position (bp)[2]	Annotation	Susceptible allele	Susceptible score	Resistant allele	Resistant score	Score difference	SNP-gene[3]	SNP-ORF[4]
Gm18_4188718_A_C	18	4188718	uncharacterized protein LOC100797777	AA	3.57	CC	2.66	0.91	inside gene	outside ORF
Gm10_48426970_A_G	10	48426970	No significant similarity found	GG	3.56	AA	2.4	1.16	-	-
Gm13_34818193_C_T	13	34818193	No significant similarity found	TT	3.56	CC	2.88	0.68	-	-
Gm18_3954704_C_T	18	3954704	protein DJ-1 homolog D	CC	3.58	TT	2.66	0.92	inside gene	outside ORF
Gm13_898111_T_G	13	898111	uncharacterized protein LOC100779077	TT	3.58	GG	2.7	0.88	inside gene	inside ORF
Gm13_34946643_T_C	13	34946643	dynamin-related protein 3A isoform X2	CC	3.56	TT	2.9	0.66	inside gene	outside ORF
Gm03_46592189_A_G	3	46592189	histone-lysine N-methyltransferase SUVR4	GG	3.72	AA	3.27	0.45	inside gene	inside ORF
Gm03_46889507_T_C	3	46889507	putative spliceosomal protein U1A	CC	3.71	TT	3.27	0.44	1042 bp	outside ORF
Gm18_54979_G_A	18	54979	pheophytinase, chloroplastic	GG	3.57	AA	2.8	0.77	inside gene	outside ORF
Gm18_122382_A_G	18	122382	splicing factor 1 isoform X2	AA	3.57	GG	2.8	0.77	1283 bp	outside ORF
Gm18_4324818_G_T	18	4324818	KH domain-containing protein SPIN1 isoform X2	GG	3.56	TT	2.72	0.84	inside gene	outside ORF
Gm15_6925513_T_C	15	6925513	ABC transporter B family member 9	CC	3.7	TT	3.12	0.58	inside gene	inside ORF
Gm05_33176582_G_A	5	33176582	No significant similarity found	GG	3.6	AA	3.31	0.29	-	-

[1] Chromosome

[2] SNP position on *Glycine max* genome assembly version Glyma.Wm82.a1 (Gmax1.01)

[3] SNP position according with the nearest gene

[4] SNP position according with the nearest ORF

Gene annotation

Higher average scores from each allele were designated to susceptible haplotypes, and resistant haplotypes with lower symptoms scores. That information was associated with annotation of the 13 SNPs flanking sequences on Table 4. No significant similarity was found between Gm05_33176582_G_A, Gm10_48426970_A_G, and Gm13_34818193_C_T flanking sequences and soybean gene annotation at NCBI, and that set contained markers that had both the largest and smallest phenotypic score difference between variation allele. The Gm18_4188718_A_C, Gm18_3954704_C_T, Gm13_34946643_T_C, Gm03_46889507_T_C, Gm18_54979_G_A, Gm18_122382_A_G, Gm18_4324818_G_T markers are outside of open reading frames.

Markers that are both within a gene and in coding regions are Gm13_898111_T_G, Gm03_46592189_A_G, and Gm15_6925513_T_C. They code, respectively, for uncharacterized protein LOC100779077, histone-lysine N-methyltransferase SUVR4, and ABC transporter B family member 9. Gm15_6925513_T_C is located at 237 bp or 79 codons from the beginning of a 501 bp/ 166 codons ORF, 3'5' frame 1, third base of the respective codon. Nucleotide base variation T/C did not change the translated amino acid, which continues to be the polar Lysine (Lys). As for Gm03_46592189_A_G and Gm13_898111_T_G, located on ORFs 5'3', frames 3 and 2, and first and second base of the codon, respectively, the produced amino acid changed with allele variation. Marker Gm03_46592189_A_G is at 208 bp from the beginning of LOC100802372 gene, and nucleotide change A/G altered the amino acid produced to Threonine (Thr)/Alanine (Ala), a shift from a polar amino acid to a hydrophobic one. Considering the Gm13_898111_T_G marker, 1334 bp from the initiation of a 2451 bp ORF, allele variation T/G modified amino acid production to Phenylalanine (Phe)/ Cysteine (Cys), a shift from hydrophobic to a polar amino acid, similarly to Gm03_46592189_A_G.

DISCUSSION

Soybean screening with *Xanthomonas axonopodis* pv. *glycines*

Completely randomized design, being the simplest of all experimental designs, permitted an adequate analysis of cultivars symptoms and pathogen aggressiveness on a homogeneous experimental condition. Regardless of XAG2440^{p} and XAG2447 bacterial

strains have been inoculated using the same design and methodology, their scores symptoms were substantially different. As presented on Table 2, 82.57% of the cultivars inoculated with XAG2440^{p} showed some degree of susceptibility, while none of the cultivars inoculated with XAG2447 developed those symptoms.

A differentiated response of XAG2440^{p} and XAG2447 could be due to some different processes. For example, the pathogen type III secretion system (TTSS) could be more intense on XAG2440^{p}. TTSS is encoded by the avirulence gene *avrBs3* homolog *avrXg1* (ATHINUWAT et al., 2009), associated with aggressiveness by the pathogen, and can possibly be the reason for more severe symptoms caused by XAG2440^{p} infection. In addition, plant defense mechanisms could be repressed by recessive resistant genes (KIM et al., 2010b) or inactivation of the resistant gene *rxp* (PARK et al., 2008), and make the studied cultivars more susceptible. A second explanation would be a mutation on the extracellular diffusible factor (DSF). To better colonize and overcome plant defense responses, bacteria developed the signaling mechanism of quorum sensing, where only from a minimal number of cells occurs effective infection and manifestation of virulence by the pathogen (PAPENFORT; BASSLER, 2016). The RpfF gene is responsible for the expression of the DSF signaling molecule on *Xanthomonas axonopodis* pv. *glycines*, and is related to the processes of pathogen colonization (THOWTHAMPITAK et al., 2008). Thereby, rpfF mutants, as could be XAG2447, have its virulence reduced by interruption of quorum sensing signaling.

Those cited processes are based on gene-for-gene resistance for the *rxp* gene (SCI et al., 2008). However, as confirmed phenotypically by Goradia et al. (2009) and molecularly by Kim et al. (2010b), soybean bacterial pustule response can show different levels of resistance. In addition, different strains of *Xanthomonas axonopodis* pv. *glycines* are able to cause a distinguished range of symptoms, with various resistant candidate genes probable to be the reason of such variation. A hypothesis that fits a more complex disease response, and complements the mechanisms cited, is an influence of the pathogen molecular pattern-triggered immunity (PTI), which is a non-specific and first step host immune defense system block against pathogen invasion (KANG et al., 2012). Then, a second step would be the detection of specific bacterial effector proteins by, for example, DSF, through effector-triggered immunity (ETI) mechanism. DSF molecules promote parasitism, causing hypersensitive reaction on the plant. Its detection occurs through resistance genes, that encodes for leucine-rich repeat and nucleotide-binding domains (KIM et al., 2011). Other genes regulating virulence by the pathogen are *XagR*, a *luxR* homolog controlling at least 79 infection-related genes, and pip, a gene encoding proline

iminopeptidase (CHATNAPARAT et al., 2012), which act differently depending on the infection stage and could possibly respond to several resistance genes. Also, near-isogenic lines of soybean resistant and susceptible to bacterial pustule sequenced using RNA-seq showed that PTI and ETI are important defense mechanisms to *Xanthomonas axonopodis* pv. *glycines* (KIM et al., 2011).

Thus, the difference of pathogenicity levels between XAG2440^{p} and XAG2447 strains could be the result of a diverse range of processes interacting with each other, which must be investigated in future studies. In addition, it is important that plant scientists are aware of the variability of symptoms caused by XAG2440^{p} and XAG2447 strains when testing bacterial pustule resistance on DUS tests required by Brazilian Department of Agriculture for the launch of new cultivars.

Association analysis and biological functions

The most significantly associated markers with bacterial pustule (Gm18_4188718_A_C and Gm10_48426970_A_G) were among those with lower MAF values, which reflect its low frequency on the studied cultivars, due to recent mutation events or its low fitness reducing reproductive success, for example. On the other hand, in chromosome 3, Gm03_46592189_A_G and Gm03_46889507_T_C markers have the highest values of MAF, therefore allele changes may not substantially influence soybean response to bacterial pustule. The equality of MAF values of markers in chromosome 3 makes sense once those alleles are segregating together in a linkage disequilibrium block of 702 kb (Figure 2). In this region are located many genes and transcription factors coding for defense-related proteins, suggesting to be a promising region to be further investigated.

Scores average difference between haplotypes identified as resistance and susceptible have a minimum extent and are very similar (Table 4). Because GWAS returns markers in linkage disequilibrium with a region or gene governing that trait, it is not possible to affirm that a single nucleotide modification of the allele, and its subsequent amino acid change, can be a direct cause of phenotypic variation, unless the identified SNP is located inside the protein coding region of a gene (RAFALSKI, 2002). The Gm13_898111_T_G, Gm03_46592189_A_G, and Gm15_6925513_T_C markers fit this specificity and their respective genes codes for uncharacterized protein LOC100779077, histone-lysine N-methyltransferase SUVR4, and ABC transporter B family member 9. Because LOC100779077 is an uncharacterized protein, we cannot infer about its function

or connection with plant defense mechanisms. Histone-lysine N-methyltransferase SUVR4 is highly related with chromatin structure modulation (THORSTENSEN et al., 2006), possibly regulating transposons damage with post-translational silencing histone H3K9 through methylation, an epigenetic control (VEISETH et al., 2011). No direct evidence of SUVR4 relationship with plant defense was reported to this date, however histone lysine methyl transferases are known for regulating ETI and influence on plant immunity pathways, correlated with lipid, cuticular wax, and carotenoid biosynthesis, for example (LEE et al., 2016). ABC transporter B family member 9 is an ATP-binding cassette transporter of molecules across membranes, proteins that can participate on resistance and parasitism processes (HWANG et al., 2016). During infection, for example, ABC transporters on bacteria can use efflux pumps to carry virulence factors or quorum sensing signaling molecules and overcome plant defenses (DU et al., 2018). In plants, ABC transporters play an important role on the transportation of secondary metabolites, often related with plant defense, like alkaloids and phenolic compounds (YAZAKI, 2006). On potato, ABC transporter subfamily, pleiotropic drug resistance proteins, is related to pathogen response regulation through foliar secretion exudates with antimicrobial activity (RUOCCO et al., 2011).

Five out of 13 SNPs were inside genes, but outside its coding region, coding for uncharacterized protein LOC100797777, protein DJ-1 homolog D, dynamin-related protein 3A isoform X2, pheophytinase (chloroplastic), and KH domain-containing protein SPIN1 isoform X2. Protein DJ1 family is related with plant stress response on the yeast homolog Hsp31 through detoxifying capabilities (MELVIN et al., 2017), thus probably presenting the same function on plants, since it is a highly conserved protein (XU et al., 2010). Dynamin-related protein 3A is associated with peroxisome replication in arabidopsis (LINGARD et al., 2008), an organelle involved on molecular signaling that accumulates endogenous reactive oxygen species (ROS) (NYATHI; BAKER, 2006). On the event of host cell infection, ROS act on cell response and plant-pathogen interaction with ozone playing an important role to trigger hypersensitive response (WASZCZAK; CARMODY; KANGASJARVI, 2018). Chloroplastic pheophytinase is reported to be related to plant senescence (CHEN et al., 2017; WANG et al., 2018). SPIN1 protein is linked to RNA-binding activity and KH domain with regulation of flowering time in rice (VEGA-SANCHEZ et al., 2008) and heat stress control in arabidopsis (GUAN et al., 2013). No associations with plant defense response were found for these proteins, same as for uncharacterized protein LOC100797777.

The searched sequence size of 2000 bp upstream and downstream was not enough to find genes nearby markers Gm10_48426970_A_G, Gm13_34818193_C_T, Gm05_33176582_G_A, while for markers Gm03_46889507_T_C and Gm18_122382_A_G, LOC100500002 putative spliceosomal protein U1A and splicing factor 1 isoform X2 were found more than 1kb far from the SNPs, both related with post-transcriptional gene regulation through alternative splicing process (HERZEL et al., 2017). So, these two SNPs can be inserted in the regulatory region of these genes, promoter or termination or enhancer, that alter the transcription through splicing.

CONCLUSIONS

Association analysis highlighted significant SNPs positioned on different locations compared with genes identified in previous studies, like *rxp* gene on chromosome 17. This could be due to the inoculation of a genetically diverse strain of the pathogen, activating distinct soybean pustule resistance pathways. Such drastically modification on genetic patterns governing bacterial pustule resistance was not expected. Evidences presented on this study suggest a much more complex defense system than previously thought.

Markers Gm13_898111_T_G, Gm03_46592189_A_G, and Gm15_6925513_T_C and other defense-related candidate genes identified to be associated with soybean bacterial pustule on this study should be investigated simultaneously and separately. By focusing on the comprehension of resistance mechanisms and host-pathogen interactions through the expression of important genes, we might be able to develop molecular markers with the purpose of application on marker-assisted selection and molecular breeding programs.

Retrieve biological information of small effects single SNPs resulted from association analysis and connect its pathways was important for the identification of core genes regulating soybean response to *Xanthomonas axonopodis* pv. *glycines* complex pathosystem defense process. Because GWAS is based on single tests for each marker, a better understanding of gene-gene and gene-protein interactions, post-transcriptional gene regulation and epigenetics control would make unprecedented advancements of GWAS interpretation, and improve its applications. Thereby, the complementary use of other methodologies, such as network analysis, may possibly offer biological insights to elucidate the results of soybean pustule resistance association analysis.

REFERENCES

AKOND, M. et al. A SNP-Based Genetic Linkage Map of Soybean Using the SoySNP6K Illumina Infinium BeadChip Genotyping Array. **Journal of Plant Genome Sciences**, v. 1, n. 3, p. 80–89, 2013.

ALLEN, T. W. et al. Soybean Yield Loss Estimates Due to Diseases in the United States and Ontario, Canada, from 2010 to 2014. **Plant Health Progress**, v. 18, n. 1, p. 19–27, 2017.

ATHINUWAT, D. et al. *Xanthomonas axonopodis* pv. *glycines* Soybean Cultivar Virulence Specificity Is Determined by *avrBs3* Homolog *avrXg1*. **Phytopathology**, v. 99, n. 8, p. 996–1004, 2009.

BARRETT, J. C. et al. Haploview: Analysis and visualization of LD and haplotype maps. **Bioinformatics**, v. 21, n. 2, p. 263–265, 2005.

BRADBURY, P. J. et al. TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics**, v. 23, n. 19, p. 2633–2635, 2007.

CHATNAPARAT, T. et al. XagR, a LuxR Homolog, Contributes to the Virulence of *Xanthomonas axonopodis* pv. *glycines* to Soybean. **Molecular Plant-Microbe Interactions**, v. 25, n. 8, p. 1104–1117, 2012.

CHEN, J. et al. Suppressor of Overexpression of CO 1 Negatively Regulates Dark-Induced Leaf Degreening and Senescence by Directly Repressing Pheophytinase and Other Senescence-Associated Genes in Arabidopsis. **Plant Physiology**, v. 173, n. 3, p. 1881–1891, 2017.

CONTRERAS-SOTO, R. I. et al. A genome-wide association study for agronomic traits in soybean using SNP markers and SNP-Based haplotype analysis. **PLoS ONE**, v. 12, n. 2, p. 1–22, 2017.

D. TURNER, S. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. **Journal of Open Source Software**, v. 3, n. 25, p. 731, 2018.

DEON, M. Software Selegen-REML/BLUP: A useful tool for plant breeding. **Crop Breeding and Applied Biotechnology**, v. 16, n. 4, p. 330–339, 2016.

DU, D. et al. Multidrug efflux pumps: structure, function and regulation. **Nature Reviews Microbiology**, v. 16, n. 9, p. 523–539, 2018.

FANG, C. et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. **Genome Biology**, v. 18, n. 1, 2017.

GASTEIGER, E. et al. ExPASy: The proteomics server for in-depth protein knowledge and analysis. **Nucleic Acids Research**, v. 31, n. 13, p. 3784–3788, 2003.

GORADIA, L.; HARTMAN, G. L.; DANIEL, S. L. Evaluation of glyphosate-tolerant soybean cultivars for resistance to bacterial pustule. **European Journal of Plant Pathology**, v. 124, n. 2, p. 331–335, 2009.

GUAN, Q. et al. A KH domain-containing putative RNA-binding protein is critical for heat stress-responsive gene regulation and thermotolerance in arabidopsis. **Molecular Plant**, v. 6, n. 2, p. 386–395, 2013.

GYAWALI, S. et al. Genome wide association studies (GWAS) of spot blotch resistance

at the seedling and the adult plant stages in a collection of spring barley. **Molecular Breeding**, v. 38, n. 5, 2018.

HERZEL, L. et al. Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. **Nature Reviews Molecular Cell Biology**, v. 18, n. 10, p. 637–650, 2017.

HWANG, J. U. et al. Plant ABC Transporters Enable Many Unique Aspects of a Terrestrial Plant's Lifestyle. **Molecular Plant**, v. 9, n. 3, p. 338–355, 2016.

IQUIRA, E.; HUMIRA, S.; FRANÇOIS, B. Association mapping of QTLs for sclerotinia stem rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. **BMC Plant Biology**, v. 15, n. 1, p. 1–12, 2015.

JOHNSON, M. et al. NCBI BLAST: a better web interface. **Nucleic acids research**, v. 36, n. Web Server issue, p. 5–9, 2008.

KADO, C.; HESKETT, M. Selective media for isolation of Agrobacterium, Corynebacterium, Erwinia, Pseudomonas, and Xanthomonas. **Phytopathology**, v. 60, n. 6, p. 969–979, 1970.

KANG, J. H. et al. Regulatory gene network from a genome-wide association study for sow lifetime productivity traits. **Animal Genetics**, v. 49, n. 3, p. 254–258, 2018.

KANG, Y. J. et al. Genome-wide mapping of NBS-LRR genes and their association with disease resistance in soybean. **BMC Plant Biology**, v. 12, 2012.

KIM, D. H. et al. Fine mapping of a resistance gene to bacterial leaf pustule in soybean. **Theoretical and Applied Genetics**, v. 120, n. 7, p. 1443–1450, 2010a.

KIM, K. H. et al. RNA-Seq analysis of a soybean near-isogenic line carrying bacterial leaf pustule-resistant and-susceptible alleles. **DNA Research**, v. 18, n. 6, p. 483–487, 2011.

KIM, K. S. et al. Fine mapping the soybean aphid resistance gene Rag1 in soybean. **Theoretical and Applied Genetics**, v. 120, n. 5, p. 1063–1071, 2010b.

KUMARI, S. et al. Thymol nanoemulsion exhibits potential antibacterial activity against bacterial pustule disease and growth promotory effect on soybean. **Scientific Reports**, v. 8, n. 1, p. 1–12, 2018.

LEE, S. et al. Global regulation of plant immunity by histone lysine methyl transferases. **The Plant Cell**, v. 28, p. 1640–1661, 2016.

LINGARD, M. J. et al. Arabidopsis PEROXIN11c-e, FISSION1b, and DYNAMIN-RELATED PROTEIN3A Cooperate in Cell Cycle-Associated Replication of Peroxisomes. **the Plant Cell Online**, v. 20, n. 6, p. 1567–1585, 2008.

LIU, Y. et al. Network-assisted analysis of GWAS data identifies a functionally-relevant gene module for childhood-onset asthma. **Scientific Reports**, v. 7, n. 1, p. 1–10, 2017.

MCCARTHY, F. M. et al. AgBase: A functional genomics resource for agriculture. **BMC Genomics**, v. 7, p. 1–13, 2006.

MCELROY, M. S. et al. Prediction of Cacao (Theobroma cacao) Resistance to Moniliophthora spp. Diseases via Genome-Wide Association Analysis and Genomic Selection. **Frontiers in Plant Science**, v. 9, p. 1–12, 2018.

MELVIN, P. et al. Methylglyoxal detoxification by a DJ-1 family protein provides dual

abiotic and biotic stress tolerance in transgenic plants. **Plant Molecular Biology**, v. 94, n. 4–5, p. 381–397, 2017.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. **Genetics**, v. 157, n. 4, p. 1819–189, 2001.

MOORE, J. H.; ASSELBERGS, F. W.; WILLIAMS, S. M. Bioinformatics challenges for genome-wide association studies. **Bioinformatics**, v. 26, n. 4, p. 445–455, 2010.

NARVEL, J. M. Molecular Mapping of Rxp Conditioning Reaction to Bacterial Pustule in Soybean. **Journal of Heredity**, v. 92, n. 3, p. 267–270, 2001.

NYATHI, Y.; BAKER, A. Plant peroxisomes as a source of signalling molecules. **Biochimica et Biophysica Acta - Molecular Cell Research**, v. 1763, n. 12, p. 1478–1495, 2006.

O'LEARY, N. A. et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. **Nucleic Acids Research**, v. 44, n. D1, p. D733–D745, 2016.

PAPENFORT, K.; BASSLER, B. L. Quorum sensing signal-response systems in Gram-negative bacteria. **Nature Reviews Microbiology**, v. 14, n. 9, p. 576–588, 2016.

PARK, H. J. et al. Avirulence gene diversity of Xanthomonas axonopodis pv. glycines isolated in Korea. **Journal of Microbiology and Biotechnology**, v. 18, n. 9, p. 1500–1509, 2008.

POLAND, J. A. et al. Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. **Proceedings of the National Academy of Sciences**, v. 108, n. 17, p. 6893–6898, 2011.

RAFALSKI, A. Applications of single nucleotide polymorphism in crop genetics. **Current opinion in plant biology**, v. 5, p. 94–100, 2002.

RUOCCO, M. et al. Four potato (Solanum tuberosum) ABCG transporters and their expression in response to abiotic factors and Phytophthora infestans infection. **Journal of Plant Physiology**, v. 168, n. 18, p. 2225–2233, 2011.

SCI, J. C. et al. Marker-Assisted Foreground and Background Selection of Near Isogenic Lines for Bacterial Leaf Pustule Resistant Gene in Soybean. **Crop Science**, v. 2008, n. 4, p. 263–268, 2008.

SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, n. 539, 2011.

STANKE, M.; MORGENSTERN, B. AUGUSTUS: A web server for gene prediction in eukaryotes that allows user-defined constraints. **Nucleic Acids Research**, v. 33, n. SUPPL. 2, p. 465–467, 2005.

THORSTENSEN, T. et al. The Arabidopsis SUVR4 protein is a nucleolar histone methyltransferase with preference for monomethylated H3K9. **Nucleic Acids Research**, v. 34, n. 19, p. 5461–5470, 2006.

THOWTHAMPITAK, J. et al. Role of rpfF in virulence and exoenzyme production of Xanthomonas axonopodis pv. glycines, the causal agent of bacterial pustule of soybean. **Phytopathology**, v. 98, n. 12, p. 1252–1260, 2008.

VEGA-SANCHEZ, M. E. et al. SPIN1, a K Homology Domain Protein Negatively

Regulated and Ubiquitinated by the E3 Ubiquitin Ligase SPL11, Is Involved in Flowering Time Control in Rice. **the Plant Cell Online**, v. 20, n. 6, p. 1456–1469, 2008.

VEISETH, S. V. et al. The SUVR4 histone lysine methyltransferase binds ubiquitin and converts H3K9me1 to H3K9me3 on transposon chromatin in arabidopsis. **PLoS Genetics**, v. 7, n. 3, 2011.

VUONG, T. D. et al. Genetic architecture of cyst nematode resistance revealed by genome-wide association study in soybean. **BMC Genomics**, v. 16, n. 1, p. 1–13, 2015.

WANG, Q. L. et al. Metabolic reprogramming in chloroplasts under heat stress in plants. **International Journal of Molecular Sciences**, v. 19, n. 3, p. 9–11, 2018.

WASZCZAK, C.; CARMODY, M.; KANGASJARVI, J. Reactive Oxygen Species in Plant Signaling. **Annual Review of Plant Biology**, v. 69, n. 1, p. 209–236, 2018.

XU, X. M. et al. The Arabidopsis DJ-1a protein confers stress protection through cytosolic SOD activation. **Journal of Cell Science**, v. 123, n. 10, p. 1644–1651, 2010.

YAZAKI, K. ABC transporters involved in the transport of plant secondary metabolites. **FEBS Letters**, v. 580, n. 4, p. 1183–1191, 2006.

ZINSOU, V. A. et al. Occurrence and characterization of *Xanthomonas axonopodis* pv. *glycines*, causing bacterial pustules on soybean in Guinea Savanna of Benin. **African Crop Science Journal**, v. 23, n. 3, p. 203–210, 2015.

SUPPLEMENTARY DATA

Supplementary Table S1. Soybean cultivars screened with *Xanthomonas axonopodis* pv. *glycines* XAG2440^{p} and XAG2447 strains.

Soybean cultivars			
CD 216	5D 711 RR	BRS 284	CD 246
CD 215	BRS 243 RR	BMX FORÇA RR	CD 251 RR
BMX TURBO	FUNDACEP 57 RR	CD 236 RR	CD 237 RR
FUNDACEP 61 RR	5D 690 RR	CD 2630 RR	BRSMG 68
CD213RR	BMX POTENCIA	FUNDACEP 55 RR	BRS 256 RR
IGRARA516 RR	BRS 246 RR	IGRARA628 RR	CD 2860
CD 225RR	5D 660 RR	CD 248 RR	CD 243 RR
FUNDACEP 63RR	CD 206 RR	CD 235 RR	CD 244 RR
BRS 230	FUNDACEP 58 RR	IGRARA626 RR	M-SOY 7901
CD 5969	FUNDACEP 53 RR	CD 252	TMG115RR
M-SOY 6101	CD 249 RR STS	CD 233 RR	M-SOY 8001
EMGOPA 304 (Campeira)	CD 239 RR	CD 221	M9144RR
BRSMG RENASCENÇA	CD 214	CD 2737 RR	CD 234RR
BMX MAGNA RR	BRS 259	CD 218	CD 205
CD 250 RRSTS	FUNDACEP 39	CD FAPA 220	CD 2792 RR
CD 238 RR	BRS 258	MGBR 48 Garimpo	UFV 16
CD 224	CD 232	EMBRAPA 59	CD 253
BRS 262	M 7211 RR	5G770 RR	FUNDACEP 59 RR
CD 208	BRS 268	BRS 213	BRSMT (PINTADO)
BRS 282	TMG 1066 RR	FUNDACEP 33	
BRS 185	M 6707 RR	CD254RR	
BRS 133	CD 226RR	M7908 RR	
CD 206	CD 201	CAC 1	
EMBRAPA 48	CD 242RR	CD 245 RR	
CD 217	CD 228	IGRARA518 RR	
BRS 257	CD 230RR	CD 219 RR	
CD 229 RR	5G830RR	CD 2840	
BRS 184	M7639RR	CD 2800	
ANTA 82	CD 240RR	P98Y51	
BRS 232	TMG 7161RR	P98Y70	

Supplementary Table S2. Cultivars and symptoms scores ranking for XAG2440^{p} strain

Ranking XAG2440^{p}	Average score range	Cultivars mode score	% individuals	Cultivars
Resistant - Moderately Resistant	1.33 - 2.66	2.66	17.43	5G770 RR, BRS 243 RR, BRS 258, CD 214, CD 234RR, CD 236 RR, CD 2630 RR, CD 2792 RR, CD 2800, FUNDACEP 53 RR, FUNDACEP 55 RR, FUNDACEP 58 RR, FUNDACEP 63RR, IGRARA626 RR, M 6707 RR, M 7211 RR, TMG 1066 RR, TMG 7161RR, TMG115RR
Moderately susceptible	3.00 - 3.66	3.00	35.78	5D 660 RR, 5D 711 RR, ANTA 82, BMX TURBO, BRS 184, BRS 230, BRS 246 RR, BRS 256 RR, BRS 268, BRSMT (PINTADO), CAC 1, CD 201, CD 206, CD 219 RR, CD 221, CD 224, CD 225RR, CD 228, CD 237 RR, CD 239 RR, CD 243 RR, CD 244 RR, CD 245 RR, CD 249 RR STS, CD 250 RRSTS, CD 251 RR, CD 5969, FUNDACEP 57 RR, FUNDACEP 59 RR, IGRARA516 RR, IGRARA518 RR, M7639RR, M7908 RR, M9144RR, MGBR 48 Garimpo, M-SOY 7901, M-SOY 8001, P98Y51, P98Y70
Susceptible - Highly susceptible	4.00 - 4.66	4.00	46.78	5D 690 RR, 5G830 RR, BMX FORÇA RR, BMX MAGNA RR, BMX POTENCIA, BRS 133, BRS 185, BRS 213, BRS 232, BRS 257, BRS 259, BRS 262, BRS 282, BRS 284, BRSMG 68, BRSMG RENASCENÇA, CD 205, CD 206 RR, CD 208, CD 215, CD 216, CD 217, CD 218, CD 226 RR, CD 229 RR, CD 230 RR, CD 232, CD 233 RR, CD 235 RR, CD 238 RR, CD 240 RR, CD 242 RR, CD 246, CD 248 RR, CD 252, CD 253, CD 2737 RR, CD 2840, CD 2860, CD/FAPA 220, CD213 RR, CD254 RR, EMBRAPA 48, EMBRAPA 59, EMGOPA 304 Campeira, FUNDACEP 33, FUNDACEP 39, FUNDACEP 61 RR, IGRARA628 RR, M-SOY 6101, UFV 16

ARTICLE 2 - NETWORK ANALYSIS OF CANDIDATE GENES AND TRANSCRIPTION FACTORS REGULATING SOYBEAN RESPONSE TO BACTERIAL PUSTULE

INTRODUCTION

Phytobacterium *Xanthomonas axonopodis* pv. *glycines* is one of the most important bacterial soybean pathogens, causing bacterial pustule. As bacterial pustule has been proved to be a genetically complex disease by greenhouse conditions (GORADIA; HARTMAN; DANIEL, 2009), and molecularly with RNA-seq (KIM et al., 2011) and genome-wide association analysis (Article 1), it is essential to identify genes of core importance and establish the relationship among them to initiate the understanding of soybean defense mechanisms. Unravel plant defense mechanisms against pathogens is fundamental to cultivar improvement so that disease resistance is conducted in a targeted and effective way. By studying genes associated to plant-pathogen interactions and its functional relationship through network analysis, we can better understand biological mechanisms associated with host defense processes (LIN et al., 2017; SHEPPARD; GUTTMAN; FITZGERALD, 2018) information that can help to increase gains in productivity on soybean breeding programs.

Genome-wide association (GWAS) is a tool used to explain the genetic variance correspondent to a trait, in which each SNP marker significantly associated with the studied trait usually contributes in small proportions for the overall phenotype (BOYLE; LI; PRITCHARD, 2017). Associated markers have these small effects that individually or collectively would hardly explain a complex trait (MANOLIO et al., 2009), such as soybean bacterial pustule. Thus, explore the biological meaning for each small effect and relate those effects altogether in function with annotated pathway databases can be useful to connect relevant information missed on association analysis. For example, molecular mechanisms on plant-pathogen interaction, besides mining potential major resistance genes in order to be incorporated into future cultivars.

Significant variants may be identified on multiple genes in combination with enhancers, repressor, and promoters, making association data analysis more complex (CANNON; MOHLKE, 2018). Thus, the investigation of candidate genes and their functional role may be more advantageous in terms that small effects markers might poorly explain the genotype-phenotype variation (MOORE; ASSELBERGS;

WILLIAMS, 2010). Relating individual parts of a complex system through biological networks is an approach to better understand the relationship between these components and make their complex interactions more informative (KAO et al., 2017), which can certainly increase the power of association analysis with biological evidence. In addition, associated markers are usually enriched in the recognition sequences of transcription factors (TFs) related with regulatory processes (MAURANO et al., 2012). So, identifying key TFs regulating genes provenient of association analysis and its connections can certainly help the understanding of biological interactions surrounding the complexity of defense response regulatory process.

A study of 84 soybean agronomical traits developed by FANG et al. (2017) resulted in 245 significant SNPs. Then, network analysis was used to group blocks in linkage disequilibrium associated with each phenotypic trait of interest and identify important genes governing those traits. On another study, a regulatory gene network was defined for sow lifetime productivity traits, mainly through the identification of overrepresented TFs and the core genes associated with them (KANG et al., 2018). Network analysis could also be applied to the integration of different biological data, such as gene expression and genome-wide association, as used with Dupuytren's disease (BECKER et al., 2016).

Two previous association studies for soybean bacterial pustule using the same genotyped cultivars found evidence of different sets of markers in linkage disequilibrium (Supplementary Table S1) (BARBOSA, 2017; Article 1). A network analysis (gene-biological process and gene-TF) approach was used with the objective to better understand the divergent data found in these association analyses, and clarify the molecular mechanisms of soybean defense to *Xanthomonas axonopodis* pv. *glycines*.

MATERIAL AND METHODS

Biological process network

Sets of significantly associated SNP markers from experiments 1 (BARBOSA, 2017) and 2 (Article 1), were entitle as GWAS 1 and GWAS 2, respectively. These association data was used on the assignment to genes with a multi-gene approach on NCBI soybean genome browser (RefSeq assembly accession: GCF_000004515.5) (O'LEARY et al., 2016). Soybean genes within 10484 bp upstream and 7742 bp downstream from the significant marker identified were included on the network analysis.

The distance was defined according to the furthermost genes identified from the significant loci and applied to the remaining markers. Then, correspondent Arabidopsis genes were searched with Soybase (GRANT et al., 2009), owing to the fact that Arabidopsis annotated pathways are larger and more reliable than of soybean.

Biological processes network interaction analysis of Arabidopsis correspondent genes of both experiments was performed with ClueGO Cytoscape plug-in v2.5.3 (BINDEA et al., 2009), used to group genes according to common GO-biological processes. The statistics used was a kappa score of 0.3, which permitted specific visualization of the network structure functional groups.

Gene-TF network

Transcription start sequence was obtained from 2 kbp upstream to 0.2 kbp downstream of each soybean gene with the same assembly version used on gene assignment. TFs binding sites related with selected soybean genes were predicted in *Arabidopsis thaliana* with a threshold of $1x10^{-5}$ p-value on Plant Transcriptional Regulatory Map web tool (JIN et al., 2015). Identified TFs were then analyzed with the Biological Networks Gene Ontology tool (BiNGO), a Cytoscape plug-in (MAERE; HEYMANS; KUIPER, 2005). By BiNGO, TFs set were analyzed with Benjamini and Hochberg False Discovery Rate multiple test correction and 0.05% p-value, and return overrepresented GO-terms biologically related. Then, a literature review of bacterial diseases most related biological processes was proceded, so that important and reliable TFs were considered in the analysis. Genes and key TFs selected of each GWAS were merged on Cytoscape v3.7.0 (SHANNON et al., 2003), and analyzed through the NetworkAnalyzer emphasizing those with more TF biding sites on the gene-TF network. Besides, selected TFs in common between experiments and candidate genes were highlighted through literature, providing a functional overview.

RESULTS

Biological process network

Gene mining for GWAS 1 and GWAS 2 data resulted on 29 and 34 soybean genes identified, respectively, and of those, 24 and 30 genes correspondent to Arabidopsis (Supplementary Table S2). Nine genes lacked information about correspondent on Arabidopsis and consequently were not considered in the analysis. Even with a lower

number of genes compared with soybean, Arabidopsis annotated pathways database are much more informative.

Thus, biological process network (Figure 1) revealed 23 genes accounting for 17 different processes. Monocarboxylic acid biosynthetic process and actin cytoskeleton organization were related, respectively, with four and two genes encompassing GWAS 1 and GWAS 2 data. Nine other processes were related with genes from GWAS 1, including response to molecule of bacterial origin and shikimate metabolic process. Six other biological processes were connected with genes from GWAS 2, as regulation of peroxisome size and response to ozone (Supplementary Table S3).

Gene-TF network

Plant Transcriptional Regulatory Map search identified 445 and 490 unique TF. Overrepresented GO-terms with BiNGO returned 220 and 254 biological processes for GWAS 1 and 2, respectively. Processes of defense response to bacterium (GO:0042742, p-value $1.83x10^{-10}$) and response to bacterium (GO:0009617, p-value $2.41x10^{-9}$) were prioritized, while twenty-five other processes related with plant defense were partially or completely covered, such as Induced systemic resistance, jasmonic acid mediated signaling pathway (GO:0009864), Regulation of immune effector (GO:0002697), and Positive regulation of immune system process (GO:0002684). The literature review-based selection of TFs within those biological processes resulted on 11 TFs for GWAS 1 genes and the same 11 plus 3 different TFs for genes from GWAS 2. Individual gene-TF network for GWAS 1 and 2 (Figure S1 and S2) was designed, and a merged gene-TF network (Figure 2) of both association analysis with 15 genes correspondent to GWAS 1 and 19 to GWAS 2, besides the 14 TFs (Supplementary Table S4).

DISCUSSION

Biological processes

Correspondent Arabidopsis genes were used in the Biological processes network analysis. It was not found a correspondent for nine soybean genes, two of them soybean model genes having within its sequence the significantly associated SNP marker. Identified biological processes possibly related with plant resistance are shikimate metabolic process, regulation of peroxisome size, response to ozone, lipid transport, response to molecule of bacterial origin, and other processes metabolic related.

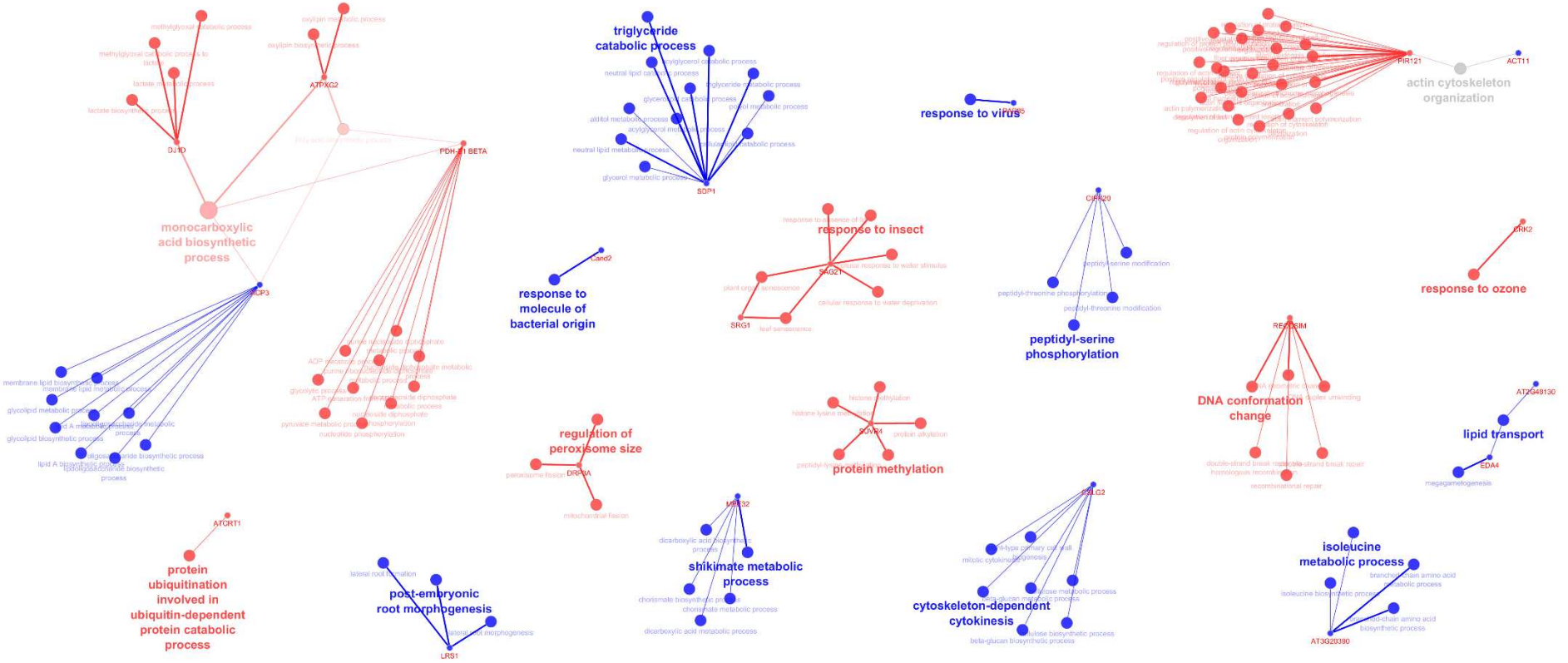


Figure 1. Biological processes network for GWAS 1 – blue (BARBOSA, 2017) and GWAS 2 – red (Article 1) candidate genes. Candidate genes identified are labeled in red. Biological processes in gray are equally related, in number of genes, to both GWAS. Biological processes in light red are also related to both processes, but more closely to GWAS 1.

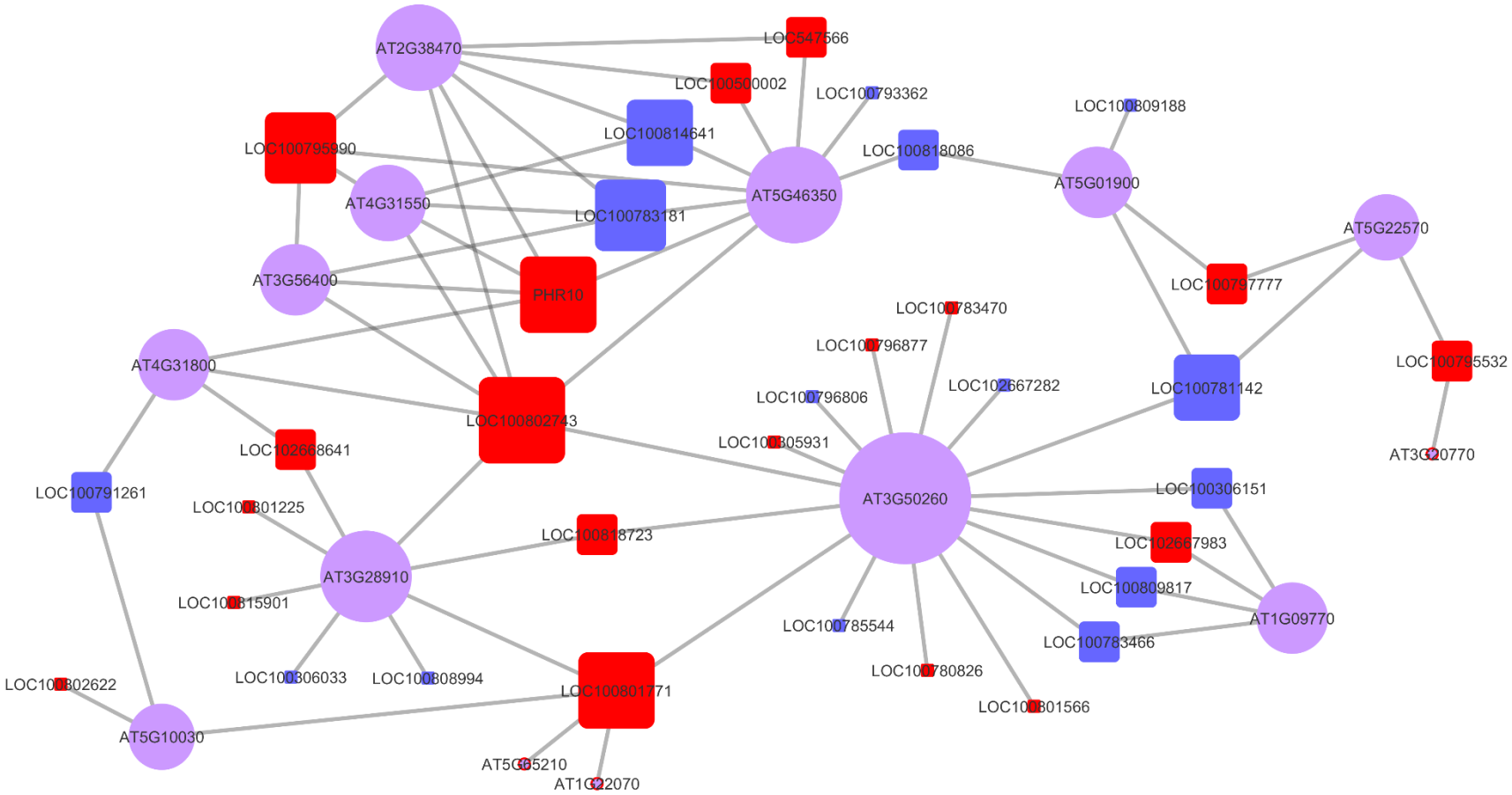


Figure 2. Gene-transcription factor (TF) network. GWAS 1 – blue (BARBOSA, 2017) and GWAS 2 – red (Article 1) candidate genes are the rectangles, and TFs in purple circles. TFs unique to GWAS 2 are circled in red. Node size increased according to the number of connections (number of TF binding sites).

Salicylic acid (SA), a signaling phytohormone related with plant immunity, is synthesized through isochorismate (IC) and phenylalanine ammonium lyase (PAL) pathways using chorismite, a precursor produced in the shikimate pathway (BERENS et al., 2017). Shikimate metabolic process is present on the biological process network through gene MEE32 from GWAS 1.

Phenylpropanoid metabolic pathway is regulated by PAL and produces cell-wall phenolic compounds involved in plant defense, for example, modulating watermelon resistance to Fusarium wilt (CHANG et al., 2014). This study concluded that cell wall reinforcement would provide a mechanical resistance against the pathogen with phenolic compounds and lignin production through peroxidase activity. Lignin production through peroxidase activity is also reported to be related with the formation of a structural barrier in shallots against *Xanthomonas axonopodis* pv. *allii* (YANTI, 2015).

The presence of gene DRP3A - regulation of peroxisome size in GWAS 2, on the biological processes network, might be due to the accumulation of enzymes, as peroxidases, or peroxisome growing for a subsequent division. Peroxisomes are organelles containing oxidases that produce reactive oxygen species (ROS), hormones, and signaling molecules related with plant defense (KAO; GONZALEZ; BARTEL, 2017; NYATHI; BAKER, 2006), and is dependent of actin cytoskeleton for its cellular disposition (LUIS, SANDALIO, CORPAS, PALMA, 2006). Actin cytoskeleton organization is another process present the biological process network through ACT11 and PIR121, genes from GWAS 1 and 2, respectively, which could be linked to a more regulatory function.

Another indicative of a possible important role played by peroxisomes on soybean defense against bacterial pustule is that peroxisome protein import in Arabidopsis is regulated by E3 ubiquitin ligase SP1 (PAN; SATKOVICH; HU, 2016). This gene is apparently similar to LOC100783181 E3 ubiquitin-protein ligase KEG-like, not included on the biological process network because of the lack of an Arabidopsis orthologous. E3 ubiquitin ligases are reported to be related to plant defense in rice (YOU et al., 2016) and Arabidopsis (DENG et al., 2017).

Response to ozone biological process, represented by CRK2 gene from GWAS 2, might also be related to peroxisomal activity. Ozone is decomposed on numerous ROS that act as effector-triggered immunity (ETI) signaling though apoplast accumulation (KACHROO; KACHROO, 2009) and trigger cell death by anti-microbial hypersensitive response (WASZCZAK; CARMODY; KANGASJARVI, 2018), a defense strategy to avoid pathogen colonization of cells not yet infected.

Monocarboxylic acid biosynthetic – genes DJ1D, ATPXG2, and PDH-E1 BETA from GWAS 1 and ACP3 from GWAS 2 – process includes membrane lipid biosynthesis which could also be related with a structural obstruction for pathogen invasion. Lipid transport process – EDA4 and AT2G48130 genes of GWAS 1 – can be related to phospholipids transport, contributing to the physical barrier hypothesis. However, it could also be related with the transport of fatty acid triglycerides and its catabolic process. Fatty acids products are precursors of various defense strategies, including cuticle formation, and SA, ethylene (ET), and jasmonic acid (JA) pathways. For example, oxylipins are biochemical products of fatty acids oxidation on the peroxisomes, and its activation through different pathways contribute to the oxylipin JA immune response or directly act as a defense signal (KACHROO; KACHROO, 2009).

Concerning response to molecule of bacterial origin, gene Cand2 from GWAS 1 might participate in quorum sensing signaling (JIN et al., 2012). Some processes did not present a clear relationship with bacterial pustule-soybean interaction, such as response to virus, post-embryonic root morphogenesis, and response to insect. Other processes seem to be related with a regulatory function, thus became difficult to predict its role on a defense process, as DNA conformation change, cytoskeleton-dependent cytokinesis, isoleucine metabolic process. Finally, possibly related to post-translation regulation were protein ubiquitination involved in catabolic process, protein methylation, and peptidyl-serine phosphorylation.

Gene-TF network

Although GWAS 1 and GWAS 2 biological processes connected in an interesting manner, soybean gene-TF network provided more insights into bacterial pustule association data. Gene-TF network highlighted TFs regulating the transcription of genes from both association studies, only three of them present exclusively on GWAS 2. TFs with more connections were AT3G50260 (Ethylene-responsive transcription factor ERF011), AT5G46350 (WRKY8 DNA-binding protein 8), AT2G38470 (WRKY33 DNA-binding protein 33), and AT3G28910 (Transcription factor MYB30).

Ethylene-responsive transcription factors ERF 2 and ERF 4 were found to contribute to immune plant response on Arabidopsis to *Fusarium oxysporum* (MCGRATH et al., 2005) and ERF 9 to *Botrytis cinerea* (MARUYAMA et al., 2013), both necrotrophic fungus pathogens. WRKY transcription factors act biding on numerous genes loci regulating gene expression on transcriptional level due to stress response

(PHUKAN; JEENA; SHUKLA, 2016). Pathogen molecules recognition by plant receptor on molecular pattern-triggered immunity (PTI) activates TF WRKY 33 biding on ET pathway key promoters in Arabidopsis (BIRKENBIHL et al., 2017). On the other hand, TF WRKY 8 is induced by wounding and related with SA and JA-mediated host basal defense against pathogen (CHEN; ZHANG; YU, 2010). Transcription factor MYB30 regulates defense process and root elongation through ROS signaling, a pathway that comprises genes related with fatty acid transport (MABUCHI et al., 2018). Those results suggest that TFs more connected in the network are directly related with the studied trait, thus endorsing a common TFs regulation of multiple different genes activated on each respective GWAS, contributing for bacterial pustule resistant and susceptible soybean phenotypes.

Genes with more connections from GWAS 1 were LOC100783181 (E3 ubiquitin-protein ligase KEG-like), with the polymorphic marker located inside the gene, a subunit of the E3 ubiquitin ligase complex, LOC100814641 (F-box/kelch-repeat protein SKIP30), and the gene related with actin cytoskeleton organization LOC100781142 (actin-3). Highly connected genes from GWAS 2 were PHR10 (MYB-CC domain-containing transcription factor PHR10), LOC100802743 peroxygenase 2, and LOC100801771 ABC transporter B family member 9. ATP-binding cassette (ABC) transporters are reported to be involved in plant defense in potato (RUOCCO et al., 2011), possibly through efflux pumps activity (DU et al., 2018). Peroxygenase 2 is an enzyme acting on lipid catalysis and molecular signaling, as described in the biological process network, and PHR10 apparently is connected with transcription factor family MYB. Therefore, even though different sets of genes were previously revealed with the soybean bacterial pustule association studies, they appear to be connected by defense-related biological processes and common TFs reinforcing the complexity of the trait.

CONCLUSIONS

Different significantly associated loci to bacterial pustule of soybean were assigned to nearby genes and its gene-biological process and gene-TF network obtained. Biological process network resulted on GO terms linking GWAS 1 and 2 on a biochemical context, for example, regulation of peroxisome size, response to ozone, and triglyceride catabolic process. Gene-TF network provided key TF regulating genes of both studies, such as the highly connected ERF011, WRKY8, WRKY33, and MYB30. Once bacterial pustule is a complex disease, it should to be influenced by several genes

and in a great extent by environmental conditions, which might explain different loci results showed in the association analysis. Six candidate genes regulating soybean defense response were highlighted (e.g. LOC100783181, LOC100814641, LOC100781142, PHR10, LOC100802743, LOC100801771), and its subsequent validation processes should be systematically applied.

REFERENCES

BARBOSA, R. M. **Associação genômica para resistência de soja à pústula bacteriana**. Dissertação (Mestrado). Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, 2017.

BECKER, K. et al. Meta-analysis of genome-wide association studies and network analysis-based integration with gene expression data identify new suggestive loci and unravel a Wnt-centric network associated with Dupuytren's disease. **PLoS ONE**, v. 11, n. 7, p. 1–18, 2016.

BERENS, M. L. et al. Evolution of Hormone Signaling Networks in Plant Defense. **Annual Review of Phytopathology**, v. 55, n. 1, p. 401–425, 2017.

BINDEA, G. et al. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. **Bioinformatics**, v. 25, n. 8, p. 1091–1093, 2009.

BIRKENBIHL, R. P. et al. Induced Genome-Wide Binding of Three Arabidopsis WRKY Transcription Factors during Early MAMP-Triggered Immunity. **The Plant Cell**, v. 29, p. 20–38, 2017.

BOYLE, E. A.; LI, Y. I.; PRITCHARD, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. **Cell**, v. 169, n. 7, p. 1177–1186, 2017.

CANNON, M. E.; MOHLKE, K. L. Deciphering the Emerging Complexities of Molecular Mechanisms at GWAS Loci. **American Journal of Human Genetics**, v. 103, n. 5, p. 637–653, 2018.

CHANG, T. H. et al. Cell wall reinforcement in watermelon shoot base related to its resistance to Fusarium wilt caused by Fusarium oxysporum f. sp. niveum. **Journal of Agricultural Science**, v. 153, n. 2, p. 296–305, 2014.

CHEN, L.; ZHANG, L.; YU, D. Wounding-Induced WRKY8 Is Involved in Basal Defense in Arabidopsis. **Molecular Plant-Microbe Interactions**, v. 23, n. 5, p. 558–565, 2010.

DENG, F. et al. Expression and regulation of ATL9, an E3 ubiquitin ligase involved in plant defense. **PLoS ONE**, v. 12, n. 11, p. 1–18, 2017.

DU, D. et al. Multidrug efflux pumps: structure, function and regulation. **Nature Reviews Microbiology**, v. 16, n. 9, p. 523–539, 2018.

FANG, C. et al. Genome-wide association studies dissect the genetic networks underlying agronomical traits in soybean. **Genome Biology**, v. 18, n. 1, 2017.

GORADIA, L.; HARTMAN, G. L.; DANIEL, S. L. Evaluation of glyphosate-tolerant

soybean cultivars for resistance to bacterial pustule. **European Journal of Plant Pathology**, v. 124, n. 2, p. 331–335, 2009.

GRANT, D. et al. SoyBase, the USDA-ARS soybean genetics and genomics database. **Nucleic Acids Research**, v. 38, n. SUPPL.1, p. 843–846, 2009.

JIN, G. et al. Two G-protein-coupled-receptor candidates, Cand2 and Cand7, are involved in Arabidopsis root growth mediated by the bacterial quorum-sensing signals N-acyl-homoserine lactones. **Biochemical and Biophysical Research Communications**, v. 417, n. 3, p. 991–995, 2012.

JIN, J. et al. An arabidopsis transcriptional regulatory map reveals distinct functional and evolutionary features of novel transcription factors. **Molecular Biology and Evolution**, v. 32, n. 7, p. 1767–1773, 2015.

KACHROO, A.; KACHROO, P. Fatty Acid–Derived Signals in Plant Defense. **Annual Review of Phytopathology**, v. 47, n. 1, p. 153–176, 2009.

KANG, J. H. et al. Regulatory gene network from a genome-wide association study for sow lifetime productivity traits. **Animal Genetics**, v. 49, n. 3, p. 254–258, 2018.

KAO, P. Y. P. et al. Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions. **Biochimica et Biophysica Acta - General Subjects**, v. 1861, n. 2, p. 335–353, 2017.

KAO, Y.-T.; GONZALEZ, K. L.; BARTEL, B. Peroxisome function, biogenesis, and dynamics in plants. **Plant Physiology**, v. 176, p. 162–177, 2017.

KIM, K. H. et al. RNA-Seq analysis of a soybean near-isogenic line carrying bacterial leaf pustule-resistant and-susceptible alleles. **DNA Research**, v. 18, n. 6, p. 483–487, 2011.

LIN, J. R. et al. PGA: Post-GWAS analysis for disease gene identification. **Bioinformatics**, v. 34, n. 10, p. 1786–1788, 2017.

LUIS, SANDALIO, CORPAS, PALMA, B. Reactive Oxygen Species and Reactive Nitrogen Species in Peroxisomes: Production , Scavenging and Role. **Plant Physiology**, v. 141, p. 330–335, 2006.

MABUCHI, K. et al. MYB30 links ROS signaling, root cell elongation, and plant immune responses. **Proceedings of the National Academy of Sciences**, v. 115, n. 20, p. E4710–E4719, 2018.

MAERE, S.; HEYMANS, K.; KUIPER, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. **Bioinformatics**, v. 21, n. 16, p. 3448–3449, 2005.

MANOLIO, T. A. et al. Finding the missing heritability of complex diseases. **Nature**, v. 461, p. 747–753, 2009.

MARUYAMA, Y. et al. The Arabidopsis transcriptional repressor ERF9 participates in resistance against necrotrophic fungi. **Plant Science**, v. 213, p. 79–87, 2013.

MAURANO, M. T. et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. **Science**, v. 337, p. 1190–1195, 2012.

MCGRATH, K. C. et al. Repressor- and Activator-Type Ethylene Response Factors Functioning in Jasmonate Signaling and Disease Resistance Identified via a Genome-Wide Screen of Arabidopsis Transcription Factor Gene Expression. **Plant Physiology**, v.

139, p. 949–959, 2005.

MOORE, J. H.; ASSELBERGS, F. W.; WILLIAMS, S. M. Bioinformatics challenges for genome-wide association studies. **Bioinformatics**, v. 26, n. 4, p. 445–455, 2010.

NYATHI, Y.; BAKER, A. Plant peroxisomes as a source of signalling molecules. **Biochimica et Biophysica Acta - Molecular Cell Research**, v. 1763, n. 12, p. 1478–1495, 2006.

O'LEARY, N. A. et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. **Nucleic Acids Research**, v. 44, n. D1, p. D733–D745, 2016.

PAN, R.; SATKOVICH, J.; HU, J. E3 ubiquitin ligase SP1 regulates peroxisome biogenesis in *Arabidopsis*. **Proceedings of the National Academy of Sciences**, v. 113, n. 46, p. E7307–E7316, 2016.

PHUKAN, U. J.; JEENA, G. S.; SHUKLA, R. K. WRKY Transcription Factors: Molecular Regulation and Stress Responses in Plants. **Frontiers in Plant Science**, v. 7, p. 1–14, 2016.

RUOCCO, M. et al. Four potato (Solanum tuberosum) ABCG transporters and their expression in response to abiotic factors and Phytophthora infestans infection. **Journal of Plant Physiology**, v. 168, n. 18, p. 2225–2233, 2011.

SHANNON, P. et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. **Genome Research**, v. 13, n. 11, p. 2498–2504, 2003.

SHEPPARD, S. K.; GUTTMAN, D. S.; FITZGERALD, J. R. Population genomics of bacterial host adaptation. **Nature Reviews Genetics**, v. 19, n. 9, p. 549–565, 2018.

WASZCZAK, C.; CARMODY, M.; KANGASJARVI, J. Reactive Oxygen Species in Plant Signaling. **Annual Review of Plant Biology**, v. 69, n. 1, p. 209–236, 2018.

YANTI, Y. Peroxidase Enzyme Activity of Rhizobacteria-introduced Shallots Bulbs to Induce Resistance of Shallot towards Bacterial Leaf Blight (Xanthomonas Axonopodis pv Allii). **Procedia Chemistry**, v. 14, p. 501–507, 2015.

YOU, Q. et al. An E3 Ubiquitin Ligase-BAG Protein Module Controls Plant Innate Immunity and Broad-Spectrum Disease Resistance. **Cell Host and Microbe**, v. 20, n. 6, p. 758–769, 2016.

SUPPLEMENTARY DATA

Supplementary Table S1. Genome-wide association data (significant markers) of soybean bacterial pustule studies

	SNP marker	Chrm*	Adjusted p-value	Location 2.1**
GWAS 1 SNP markers				
SNP752	Gm03_46214163_C_T	3	3.581796	44213517
SNP2159	Gm08_42793022_A_G	8	3.557384	43619289
SNP750	Gm03_46055685_C_T	3	3.473926	44055029
SNP1021	Gm05_682648_A_G	5	3.340338	2357871
SNP681	Gm03_38069022_A_G	3	3.335659	36042575
SNP2473	Gm10_2437001_A_G	10	3.33313	2445007
SNP3378	Gm13_29677928_G_T	13	3.284479	30875555
SNP3416	Gm13_32875289_C_A	13	3.198286	34087365
SNP662	Gm03_36634361_G_A	3	3.158909	34612416
SNP3354	Gm13_27665585_A_G	13	3.151448	28859734
SNP660	Gm03_36438792_G_T	3	3.047569	34416830
GWAS 2 SNP markers				
SNP4571	Gm18_4188718_A_C	18	3.337415	4207722
SNP2697	Gm10_48426970_A_G	10	3.212595	49022659
SNP3436	Gm13_34818193_C_T	13	2.977933	36026888
SNP4570	Gm18_3954704_C_T	18	2.907099	3977064
SNP3210	Gm13_898111_T_G	13	2.753018	20966535
SNP3437	Gm13_34946643_T_C	13	2.730578	36155338
SNP757	Gm03_46592189_A_G	3	2.717708	44591609
SNP758	Gm03_46889507_T_C	3	2.711906	44889124
SNP4520	Gm18_54979_G_A	18	2.642848	54979
SNP4521	Gm18_122382_A_G	18	2.642848	122380
SNP4573	Gm18_4324818_G_T	18	2.611343	4344194
SNP3847	Gm15_6925513_T_C	15	2.572738	6939367
SNP1151	Gm05_33176582_G_A	5	2.543374	33444202

*Chrm: chromosome

**Location 2.1: NCBI soybean genome assembly

Supplementary Table S2. Arabidopsis genes included on the biological processes network, Arabidopsis correspondent (AT), and soybean (LOC and GLYMA) identification.

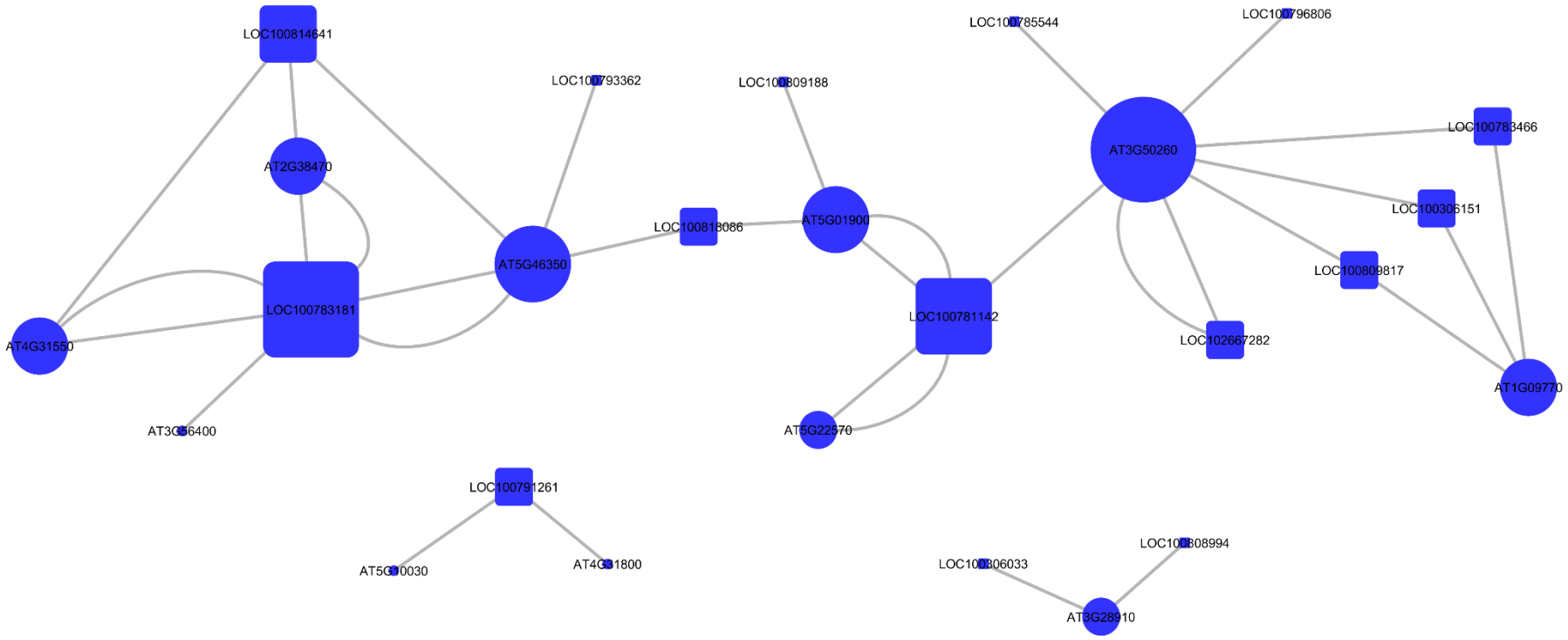
Gene	Arabidopsis	Soybean	
		LOC	GLYMA
DJ1D	AT3G02720.1	LOC100795300	GLYMA_18G046000
ATPXG2	AT5G55240.1	LOC100802743	GLYMA_03G249900
PDH-E1 BETA	AT1G30120.1	LOC100780826	GLYMA_05G141000
ACP3	AT1G54630.1	LOC100784363	GLYMA_03G242600
ATCRT1	AT5G56340.1	LOC100784876	GLYMA_18G000200
SDP1	AT5G04040.1	LOC100791261	GLYMA_03G130900
Cand2	AT3G05010.1	LOC100783834	GLYMA_03G242700
DRP3A	AT4G33650.2	LOC100807752	GLYMA_13G256100
LRS1	AT3G05090.1	LOC100788971	GLYMA_03G244500
PAP85	AT3G22640.1	LOC100787186	GLYMA_10G028300
SAG21	AT4G02380.1	LOC547566	GLYMA_03G253200
SRG1	AT1G17020.1	LOC100801225	GLYMA_15G090100
SUVR4	AT3G04380.1	LOC100802372	GLYMA_03G249800
MEE32	AT3G06350.1	LOC100788604	GLYMA_03G242400
CSLG2	AT4G24000.1	LOC100809188	GLYMA_13G174300
AT3G20390	AT3G20390.1	LOC100785544	GLYMA_08G316200
RECQSIM	AT5G27680.1	LOC100797777	GLYMA_18G048900
CIPK20	AT5G45820.1	LOC100807927	GLYMA_13G228500
EDA4	AT2G48140.1	LOC100809817	GLYMA_10G028200
AT2G48130	AT2G48130.1	LOC100306151	GLYMA_10G028100
CRK2	AT1G70520.1	LOC100801131	GLYMA_18G046100
ACT11	AT3G12110.1	LOC100781142	GLYMA_03G144800
PIR121	AT5G18410.2	LOC100814483	GLYMA_03G253000

Supplementary Table S3. GWAS experiment group, genes and its correspondent biological process related.

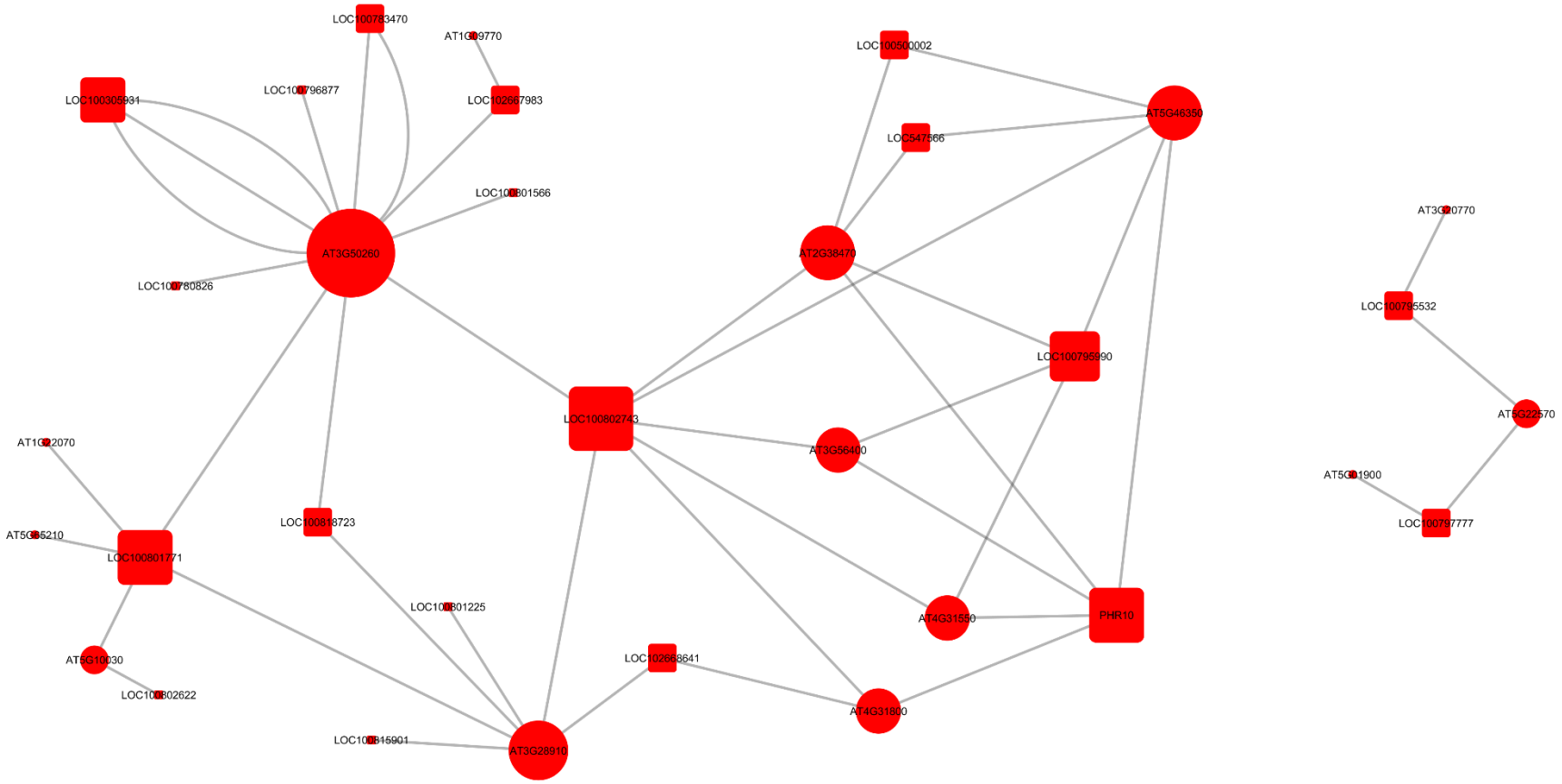
Group	Genes	Biological Process
GWAS 1+2	DJ1D, ATPXG2, PDH-E1 BETA, ACP3	Monocarboxylic acid biosynthetic process
GWAS 1+2	PIR121, ACT11	Actin cytoskeleton organization
GWAS 1	Cand2	Response to molecule of bacterial origin
GWAS 1	SDP1	Triglyceride catabolic process
GWAS 1	PAP85	Response to virus
GWAS 1	MEE32	Shikimate metabolic process
GWAS 1	LRS1	Post-embryonic root morphogenesis
GWAS 1	CIPK20	Peptidyl-serine phosphorylation
GWAS 1	CSLG2	Cytoskeleton-dependent cytokinesis
GWAS 1	AT3G20390	Isoleucine metabolic process
GWAS 1	EDA4, AT2G48130	Lipid transport
GWAS 2	ATCRT1	Protein ubiquitination involved in catabolic process
GWAS 2	DRP3A	Regulation of peroxisome size
GWAS 2	CRK2	Response to ozone
GWAS 2	SAG21, SRG1	Response to insect
GWAS 2	SUVR4	Protein methylation
GWAS 2	RECQSIM	DNA conformation change

Supplementary Table S4. Transcription factors

GWAS	FT genes	Description
GWAS 1+2	AT3G56400	Probable WRKY transcription factor 70
GWAS 1+2	AT4G31800	WRKY18 transcription factor 18
GWAS 1+2	AT3G50260	Ethylene-responsive transcription factor ERF011
GWAS 1+2	AT5G01900	Probable WRKY transcription factor 62
GWAS 1+2	AT5G22570	Probable WRKY transcription factor 38
GWAS 1+2	AT1G09770	CDC5 cell division cycle 5
GWAS 1+2	AT2G38470	WRKY33 DNA-binding protein 33
GWAS 1+2	AT5G46350	WRKY8 DNA-binding protein 8
GWAS 1+2	AT5G10030	Transcription factor TGA4
GWAS 1+2	AT4G31550	WRKY11 DNA-binding protein 11
GWAS 1+2	AT3G28910	Transcription factor MYB30
GWAS 2	AT5G65210	TGA1 bZIP transcription factor family protein
GWAS 2	AT3G20770	EIN3 Ethylene insensitive 3 family protein
GWAS 2	AT1G22070	TGA3 transcription factor TGA3



Supplementary Figure S1. Gene-transcription factor (TF) network for GWAS 1. Candidate genes are the rectangles and TFs in circles. Node size increased according to the number of connections (number of TF binding sites).



Supplementary Figure S2. Gene-transcription factor (TF) network for GWAS 2. Candidate genes are the rectangles and TFs in circles. Node size increased according to the number of connections (number of TF binding sites).

CONCLUSIONS

The results presented in this study suggest that XAG2447, one of the strains most used on bacterial pustule screening in the assays of distinguishability, homogeneity, and stability of soybean cultivars, presents a lower symptomatic response compared to strain XAG2440^{p}.

Soybean response to bacterial pustule is a complex character, now also studied with genomic association analysis. Markers on linkage disequilibrium and candidate genes identified possibly contribute to the activation of soybean defense mechanisms.

GWAS of two different experiments resulted in markers located at distinct positions, therefore different candidate genes to be studied. However, most genes of both experiments are regulated by the same transcription factors. In addition, gene networks analysis suggests the existence of a relationship between biological processes related to the genes of these different sets.

The validation of selected genes and transcription factors should be performed through gene expression analysis so that the process of infection and defense of soybean to bacterial pustule can be better understood.