

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE AGRONOMIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FITOTECNIA

RESSEQUENCIAMENTO GENÔMICO DE SOJA (*Glycine max* (L.) Merr.) E
IDENTIFICAÇÃO DE GENES RELACIONADOS À TOLERÂNCIA AO
ALAGAMENTO DO SOLO

Gabriele Casarotto
Mestre em Agronomia/UFSM

Tese apresentada como um dos requisitos
à obtenção do Grau de Doutor em Fitotecnia
Ênfase em Melhoramento e Biotecnologia Vegetal

Porto Alegre (RS), Brasil
Fevereiro de 2017

CIP - Catalogação na Publicação

Casarotto, Gabriele
Ressequenciamento genômico de soja (*Glycine max*
(L.) Merr.) e identificação de genes relacionados à
tolerância ao alagamento do solo / Gabriele
Casarotto. -- 2017.
85 f.

Orientador: JOSÉ FERNANDES BARBOSA NETO.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Faculdade de Agronomia, Programa de
Pós-Graduação em Fitotecnia, Porto Alegre, BR-RS, 2017.

1. Soja. 2. Montagem do genoma. 3. Hipóxia. 4.
Expressão gênica. I. BARBOSA NETO, JOSÉ FERNANDES,
orient. II. Título.

GABRIELE CASAROTTO
Engenheira Agrônoma - UFSM
Mestre em Agronomia - UFSM

TESE

Submetida como parte dos requisitos
para obtenção do Grau de

DOCTOR EM FITOTECNIA

Programa de Pós-Graduação em Fitotecnia
Faculdade de Agronomia
Universidade Federal do Rio Grande do Sul
Porto Alegre (RS), Brasil

Aprovado em: 22.02.2017
Pela Banca Examinadora

Homologado em: 30.05.2017
Por

JOSÉ FERNANDES BARBOSA NETO
Orientador - PPG Fitotecnia

CHRISTIAN BREDEMEIER
Coordenador Substituto do
Programa de Pós-Graduação em
Fitotecnia

SAMUEL CORDEIRO VITOR MARTINS
PPG Fitotecnia/UFRGS

MARCELO GRAVINA DE MORAES
Faculdade de Agronomia/UFRGS

ANTONIO COSTA DE OLIVEIRA
UFPel

CARLOS ALBERTO BISSANI
Diretor da Faculdade
de Agronomia

As pessoas importantes na minha vida:

Meus pais, Valmir Luiz (*in memoriam*) e Enir.
Meus irmãos, Daniel e Julie.

Dedico este trabalho!

AGRADECIMENTOS

A Deus, pela vida, tranquilidade e força para realização deste trabalho.

À Universidade Federal do Rio Grande do Sul e ao Programa de Pós-Graduação em Fitotecnia, pela oportunidade de realização do Doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsa de estudos.

Ao meu orientador, professor José Fernandes Barbosa Neto, pela confiança, compreensão, paciência e dedicação. Guardo com carinho os conhecimentos e o exemplo pessoal e profissional demonstrado.

Aos professores André Luís Thomas e José Antônio Martinelli, pela disponibilidade e estímulo para conclusão deste trabalho.

Ao pesquisador Daniel da Rosa Farias, pela disponibilidade e auxílio nas análises.

A minha família, minha mãe Enir, meus irmãos Daniel e Julie e cunhados Juliana e Alízio, pelo amor, carinho, compreensão e incentivo. Vocês são meu porto seguro, minha fortaleza! Também, à pequena Antônia, minha sobrinha e afilhada, por deixar os dias mais alegres com seu lindo sorriso.

Ao meu namorado Tiago, por acreditar em mim, pela confiança, carinho e amor. Espero dividir inúmeras conquistas contigo!

Aos meus primos Flávio, Débora, Denise e Jorge, pelos conselhos, pelo carinho e por estarem sempre dispostos a ajudar. Amo vocês!

Aos meus familiares e amigos, que sempre apoiaram e compreenderam a minha ausência em momentos especiais de nossas vidas.

Aos colegas e amigos, Gerarda, Francisco, Ana Paula, Carolina, Yohana, Jéssica, Andriéli, Ernandes e Luan, pelos momentos de trabalho e descontração.

Aos colegas do Programa de Pós-graduação em Fitotecnia, pelo coleguismo e amizade.

A todos que de alguma forma colaboraram com a concretização deste trabalho.

Muito Obrigada!

RESSEQUENCIAMENTO GENÔMICO DE SOJA (*Glycine max* (L.) Merr.) E IDENTIFICAÇÃO DE GENES RELACIONADOS À TOLERÂNCIA AO ALAGAMENTO DO SOLO¹

Autor: Gabriele Casarotto

Orientado: José Fernandes Barbosa Neto

RESUMO

O cultivo da soja (*Glycine max* (L.) Merr.) em áreas de várzea surgiu como alternativa à monocultura de arroz, uma vez que cultivares com diferentes níveis de tolerância ao alagamento do solo foram identificadas. As tecnologias de sequenciamento de nova geração apresentam como vantagens o elevado volume de informações obtidas e uma visão direta na variação do DNA. Assim, é possível identificar polimorfismos e diferenciar genótipos quanto à característica de interesse. O objetivo deste trabalho foi ressequenciar e montar o genoma de duas cultivares de soja contrastantes para o caráter tolerância ao alagamento do solo, avaliar a expressão de genes candidatos e identificar alterações não sinônimas nas sequências gênicas. O DNA total das cultivares Introdução 27 (tolerante) e BRS 154 (sensível) foi extraído de folhas jovens. Ao DNA fragmentado e purificado foram adicionados indexadores e adaptadores conhecidos. Fragmentos com ~300pb foram selecionados e então realizado o sequenciamento pareado. As leituras de sequenciamento foram submetidas à verificação de qualidade das bases e a montagem foi realizada com os montadores ABySS, SOAPdenovo e a combinação Velvet e SSPACE. Para análise de RT-qPCR foram escolhidos como candidatos à tolerância ao encharcamento do solo genes relacionados a rotas de captação de energia (*ENO*, *ADHI*, *ALAT2* e *GLB1*), formação de estruturas associadas ao estresse (*XETP* e *LBD41*) e detoxificação celular (*APX2*). O alinhamento dos *scaffolds* gerados pelo montador ABySS e o genoma de referência da cv. Williams 82 foi realizado para reconstruir a sequência destes genes e as alterações não sinônimas foram identificadas. O sequenciamento gerou em média 111 milhões de leituras pareadas. A distribuição da qualidade das bases nas leituras foi elevada ($e \geq 28$) e adaptadores residuais não foram identificados. O programa ABySS mostrou maior eficiência na montagem dos genomas das cultivares Introdução 27 e BRS 154, gerando *scaffolds* maiores, com tamanho total de 1,024Gb e 1,020Gb, respectivamente. A taxa de alinhamento global das leituras com o genoma de referência foi de 97%. De maneira geral, a cultivar Introdução 27 apresentou maior expressão relativa de todos os genes avaliados. Alteração não sinônima entre os genótipos estudados foi constatada apenas no gene *APX2* e diferenças na região promotora associada ao estresse foi constatada apenas no gene *ALAT2*. Dessa forma, não foi possível explicar as diferenças de expressão observadas, sugerindo uma maior complexidade do caráter estudado.

¹Tese de Doutorado em Fitotecnia, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. (85f.) Fevereiro, 2017.

GENOMIC RESEQUENCING OF SOYBEAN (*Glycine max* (L.) Merr.) AND IDENTIFICATION OF GENES RELATED TO SOIL FLOODING TOLERANCE²

Author: Gabriele Casarotto

Adviser: José Fernandes Barbosa Neto

ABSTRACT

Soybean (*Glycine max* (L.) Merr.) in lowland areas emerged as alternative to rice monoculture, since cultivars with different levels of soil flooding tolerance have been identified. The next generation sequencing technologies generates high volume of data and a direct view of DNA variation. So, it is possible to identify polymorphisms and differentiate genotypes for traits of interest. The objective of this work was to resequence and assemble the genome of two contrasting soybean cultivars to soil flooding, to evaluate the expression of candidate genes and to identify non-synonymous changes in these gene sequences. Whole DNA of Introduction 27 and BRS 154 cultivars was extracted from young leaves. Indexers and adapters were added to fragmented and purified DNA. Fragments with ~300bp were selected and then pair-end sequencing was performed. Reads of sequencing were submitted to bases quality check and the assembly was performed with the ABySS, SOAPdenovo assemblers and the combination of Velvet and SSPACE. Candidate genes to soil flooding tolerance were submitted to RT-qPCR analysis. These genes were related to energy uptake routes (*ENO*, *ADH1*, *ALAT2* and *GLB1*), formation of stress-associated structures (*XETP* and *LBD41*) and cell detoxification (*APX2*). The alignment of the scaffolds generated by the ABySS assembler and the cv. Williams 82 was performed to reconstruct the sequence of these genes and the non-synonymous changes were identified. Sequencing generated on average 111 million pair-end reads. The distribution of base quality in the reads was high ($e \geq 28$) and residual adapters were not identified. The ABySS program showed higher efficiency in assembling the genomes of Introduction 27 and BRS 154 cultivars, generating larger scaffolds, with a total size of 1.024Gb and 1.020Gb, respectively. The overall rate of alignment to reads with reference genome was 97%. In general, Introduction 27 cultivar presented the highest relative expression of all evaluated genes. Non-synonym alteration among the studied genotypes was observed only in the *APX2* gene and differences in the promoter region associated with stress were found only in the *ALAT2* gene. This way, was not possible to explain the differences in expression observed, suggesting a larger complexity in the studied trait.

²Doctoral thesis in Plant Science, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. (85p.) February, 2017.

SUMÁRIO

	Página
1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	3
2.1 Origem e domesticação da soja cultivada	3
2.2 O cultivo da soja no Brasil	4
2.3 Soja na várzea e tolerância ao alagamento do solo	4
2.4 Alterações em plantas de soja em resposta ao alagamento do solo.....	5
2.4.1 Alterações morfológicas e injúrias	6
2.4.2 Alterações no metabolismo energético	7
2.4.3 Sinalização mediada por etileno	10
2.4.4 Alterações na homeostase celular	11
2.4.5 Hemoglobina não simbiótica e óxido nítrico sob hipóxia.....	11
2.4.6 Interação com outros estresses	13
2.5 Tecnologias de sequenciamento de DNA	14
2.5.1 Identificação de genes por meio de sequenciamento	16
2.5.2 Sequenciamento do genoma da soja e bancos de dados	17
3 MATERIAL E MÉTODOS	18
3.1 Sequenciamento do genoma da soja	18
3.1.1 Extração de DNA, preparação das bibliotecas e sequenciamento	18
3.1.2 Análises das <i>reads</i> , montagem dos genomas e identificação de polimorfismos.....	19
3.2 Análise de expressão de genes candidatos à tolerância ao encharcamento do solo	20
3.2.1 Obtenção do material experimental	20
3.2.2 Extração de RNA e síntese de cDNA	21
3.2.3 RT-qPCR	22
3.2.4 Análise da expressão relativa dos genes candidatos	23
3.3 Análise da sequência dos genes candidatos	24
3.4 Análise da sequência promotora e elementos regulação em cis.....	25

	Página
4 RESULTADOS E DISCUSSÃO	26
4.1 Análise do sequenciamento e montagem dos genomas	26
4.2 RT-qPCR.....	30
4.2.1 Amplificação e estabilidade dos genes de referência	30
4.2.2 Expressão relativa dos genes candidatos à tolerância ao alagamento.....	35
4.2.2.1 Expressão relativa de genes relacionados ao metabolismo energético	35
4.2.2.2 Expressão relativa de genes relacionados a modificações estruturais	41
4.2.2.3 Expressão relativa de genes relacionados a detoxificação de EROS.....	44
4.3 Análise das sequências dos genes candidatos à tolerância ao alagamento	45
4.4 Análise da sequência promotora e elementos regulação em cis.....	57
5 CONCLUSÕES.....	63
6 REFERÊNCIAS BIBLIOGRÁFICAS	64

RELAÇÃO DE TABELAS

	Página
1. Genes de referência e candidatos à tolerância ao alagamento do solo, locus gênico, sequência dos primers (5'-3') utilizados para a análise de RT-qPCR e temperatura de pareamento (Tm). UFRGS, Porto Alegre, RS, Brasil. 2017.....	22
2. Genes candidatos analisados por sequência gênica. UFRGS, Porto Alegre, RS, Brasil. 2017.....	24
3. Resumo dos resultados da montagem dos genomas das cultivares de soja Introdução 27 e BRS 154, gerados por meio de diferentes programas montadores. UFRGS, Porto Alegre, RS, Brasil. 2017.	28
4. Variações genômicas detectadas no alinhamento entre os genomas das cultivares BRS 154, Introdução 27 e Williams 82. UFRGS, Porto Alegre, RS, Brasil. 2017.	46
5. Elementos de regulação <i>cis</i> localizados na região promotora dos genes candidatos a tolerância ao alagamento do solo na cultivar de referência Williams 82. UFRGS, Porto Alegre, RS, Brasil. 2017.....	60

RELAÇÃO DE FIGURAS

	Página
1. Alterações no metabolismo energético de plantas em resposta ao alagamento do solo e deficiência de oxigênio. Com o estresse são induzidas rotas de catabolismo de açúcares, glicólise, fermentação e síntese de aminoácidos não tricarboxílicos. Adaptado de: Bailey-Serres & Voeselek (2008) e António <i>et al.</i> (2016). UFRGS, Porto Alegre, RS, Brasil. 2017.....	9
2. Modulação dos níveis de óxido nítrico (NO), realizada por meio de hemoglobina não simbiótica de classe 1 (nsHb1), em resposta a hipóxia. Adaptado de Riquelme & Hinrichsen (2015). UFRGS, Porto Alegre, RS, Brasil. 2017.	13
3. Distribuição da qualidade das bases nas leituras provenientes do sequenciamento de bibliotecas pair-end, com 101 pares de bases (pb) de comprimento, dos genótipos de soja Introdução 27 (esquerda) e BRS 154 (direita). As áreas em verde, amarelo e vermelho nos gráficos indicam elevada (escore \geq 28), intermediária (20<escore<28) e baixa (escore \leq 20) qualidade das bases, respectivamente. UFRGS, Porto Alegre, RS, Brasil. 2017.	27
4. Curvas de dissociação dos genes de referência <i>ELF1-β</i> (A), <i>UBI</i> (B) e <i>ACTB</i> (C) e candidatos à tolerância ao alagamento do solo, <i>ENO</i> (D), <i>ADHI</i> (E), <i>ALAT2</i> (F), <i>GLB1</i> (G), <i>XETP</i> (H), <i>LBD41</i> (I) e <i>APX2</i> (J), em folhas e raízes de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. UFRGS, Porto Alegre, RS, Brasil. 2017.	31
5. Variação na abundância de transcritos. Os pontos indicam o Ct médio estimado e as barras verticais o intervalo de confiança ($\alpha=0,05$) para os genes de referência Fator de Elongação 1- β (<i>ELF1-β</i>), Ubiquitina (<i>UBI</i>) e Actina B (<i>ACTB</i>). UFRGS, Porto Alegre, RS, Brasil. 2017.	32
6. Coeficientes de estabilidade dos genes de referência Fator de Elongação 1- β (<i>ELF1-β</i>), Ubiquitina (<i>UBI</i>) e Actina B (<i>ACTB</i>) de soja expressos em folhas e raízes, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Coeficientes calculados utilizando os algoritmos Delta Ct, BestKeeper, NormFinder e geNorm. Quanto menor o valor do coeficiente, maior a estabilidade do gene. UFRGS, Porto Alegre, RS, Brasil. 2017.....	34
7. Expressão relativa do gene enolase (<i>ENO</i>) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha=0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.	37

8. Expressão relativa do gene álcool desidrogenase (*ADHI*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017..... 38
9. Expressão relativa do gene alanina aminotransferase 2 (*ALAT2*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017..... 40
10. Expressão relativa do gene hemoglobina 1 (*GLBI*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017..... 41
11. Expressão relativa do gene (*LBD41*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017..... 42
12. Expressão relativa do gene (*XETP*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017..... 43
13. Expressão relativa do gene ascorbato peroxidase 2 (*APX2*) em folhas e raízes de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017..... 45
14. Frequência de indels e dos diferentes tipos de substituições nos genótipos BRS 154 e Introdução 27 em comparação com a cultivar de referência Williams 82 e entre si. UFRGS, Porto Alegre, RS, Brasil. 2017..... 47
15. Posição dos éxons preditos dos genes Glyma.19g190900 (A) e Glyma.03g190500 (B), codificadores da enzima Enolase, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017. 49
16. Posição dos éxons preditos dos genes Glyma.04g240800 (A) e Glyma.06g122600 (B), codificadores da enzima Álcool Desidrogenase 1, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017..... 51
17. Posição dos éxons preditos dos genes Glyma.01g026700 (A) e Glyma.02g038100 (B), codificadores da enzima alanina aminotransferase 2, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017..... 52
18. Posição dos éxons preditos do gene Glyma.11g121800, codificador da Hemoglobina 1, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017..... 53

	Página
19. Posição dos éxons preditos dos genes Glyma.08g365100 (A) e Glyma.18g297100 (B), codificadores da proteína LBD41, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.	54
20. Posição dos éxons preditos dos genes Glyma.16g045000 (A) e Glyma.19g106700 (B), codificadores da proteína XETP, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.	56
21. Posição dos éxons preditos do gene Glyma.12g073100, codificador da enzima Ascorbato Peroxidase 2, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.	57

1 INTRODUÇÃO

A soja (*Glycine max* (L.) Merr.) é a principal leguminosa cultivada mundialmente. O cultivo da espécie em áreas de várzea surgiu como uma alternativa à monocultura do arroz irrigado, buscando com isso diminuir a incidência de pragas, doenças, manejar o arroz vermelho e incrementar os ganhos econômicos do sistema de produção. Contudo, o ambiente de várzea proporciona, com os alagamentos frequentes, injúrias na soja, que por sua vez são variáveis em função do estágio de desenvolvimento da cultura e a duração do período de estresse. A tolerância ao alagamento do solo é uma característica complexa que envolve a privação de oxigênio para o sistema radicular e as modificações subsequentes que ocorrem no solo. A difusão e a solubilidade do oxigênio em solo alagado são reduzidas e o aumento da atividade microbiana consome rapidamente o oxigênio disponível, impossibilitando a respiração aeróbica. Para evitar o colapso energético causado pela hipóxia, genes relacionados com as vias de catabolismo de açúcares, glicolítica e de fermentação são induzidos. As respostas desencadeadas pelo estresse podem ser acionadas pela detecção direta da concentração de oxigênio ou indiretamente por alterações na homeostase celular, redução dos níveis de ATP, redução do pH do citosol ou ainda pelo aumento da concentração interna de espécies reativas de oxigênio (EROS), espécies reativas de nitrogênio (ERN) ou em resposta ao fitormônio etileno. Desse modo, essa via de sinalização é complexa e ainda não foi totalmente elucidada.

Entender as respostas moleculares de plantas ao estresse tem sido objetivo de inúmeros estudos e a utilização do sequenciamento genômico como ferramenta para a

identificação de genes envolvidos com essas respostas tem se mostrado uma ferramenta eficiente. As tecnologias de sequenciamento de nova geração (NGS) permitem a identificação de variações no DNA, como a detecção de polimorfismos de nucleotídeo único (*single nucleotide polymorphisms* ou *SNPs*) e inserções ou deleções (*indels*), além do desenvolvimento de marcadores e sítios *splicing* alternativo de forma rápida e com menor custo.

Desse modo, os objetivos deste trabalho foram sequenciar e montar o genoma de duas cultivares de soja, contrastantes para o caráter tolerância ao encharcamento do solo, verificar o perfil de expressão genes candidatos e identificar possíveis polimorfismos entre as sequências destes genes nos dois genótipos.

2 REVISÃO DE LITERATURA

2.1 Origem e domesticação da soja cultivada

A soja cultivada [*Glycine max* (L.) Merrill] é uma planta autógama, membro da família Fabaceae. É uma das culturas de grãos com maior importância mundial devido a sua composição rica em óleo e proteína, sendo utilizada como matéria-prima na indústria e como fonte de nutrientes e gordura na alimentação humana e animal. Originária do Oriente, evidências históricas indicam o Nordeste da China como centro de origem da cultura (Hymowitz, 1970). Em desacordo, análises filogenéticas utilizando marcadores moleculares do tipo microssatélites e análise de diversidade de nucleotídeos colocam a região Sul da China como provável local de surgimento da cultura (Guo *et al.*, 2010). Dessa forma, existem três hipóteses de origem da soja (*G. max*), sendo elas as hipóteses de origem única, múltipla e complexa (Sedivy *et al.*, 2017).

A hipótese de uma única origem sugere como ancestral direto da soja cultivada um agrupamento monomórfico da soja selvagem *Glycine soja* Siebold & Zucc. (Sedivy *et al.*, 2017). A hipótese de múltiplas origens, por sua vez, apoiada na diversidade de plantas de soja, sugere vários esforços independentes de domesticação. Já a hipótese complexa projeta uma nova perspectiva sobre a domesticação da cultura da soja, que ocorreu em um período prolongado e em múltiplos locais e que esses primeiros indivíduos domesticados podem ter desaparecido ou ter sido integrados na soja domesticada (Sedivy *et al.*, 2017).

A hipótese única é a mais amplamente aceita na comunidade científica. Ambas as espécies, *G. soja* e *G. max*, possuem evidências genômicas de duplicação e embora possuam

características distintas, exibem o mesmo número de cromossomos ($2n = 40$ cromossomos) e são capazes de cruzar e produzir descendentes férteis (Kim *et al.*, 2012). Diversos estudos têm evidenciado a uniformidade das cultivares de soja (Bonato *et al.*, 2006; Hyten *et al.*, 2006; Priolli *et al.*, 2002) indicando que poucos acessos contribuíram para a base genética de cultivares atuais. As perdas de alelos favoráveis, nos gargalos da domesticação da cultura reduziram em até 50% a diversidade genética (Hyten *et al.*, 2006). As bases genéticas estreitas dificultam o processo de melhoramento, dessa forma, *G. soja* é uma importante fonte de alelos para a soja cultivada.

2.2 O cultivo da soja no Brasil

O cultivo da soja ganhou espaço no mercado brasileiro durante as décadas de 1970 e 1980 (Dros, 2004). Atualmente, é a cultura que mais se destaca, tanto em área plantada quanto em produção, que na safra 2015/2016 foi de aproximadamente 33,25 milhões de hectares e 95,43 milhões de toneladas (Companhia Nacional de Abastecimento, 2017). O cultivo é distribuído em todas as regiões do país, sendo a Centro-Sul responsável por aproximadamente 91% da produção. Os maiores produtores nacionais são o Mato Grosso, Paraná, Rio Grande do Sul, Goiás e Mato Grosso do Sul, respectivamente (Companhia Nacional de Abastecimento, 2017).

2.3 Soja na várzea e tolerância ao alagamento do solo

Solos de várzea são caracterizados por serem hidromórficos, com uma camada superficial pouco espessa e subsuperficial impermeável, geralmente associada a um relevo plano ou ondulado, ocupando 5,4 milhões de hectares no Rio Grande do Sul (Pinto *et al.*, 1999). Devido a drenagem deficiente, os espaços porosos ficam saturados com água, o que acarreta uma série de alterações na dinâmica solo-planta (Bailey-Serres; Voesenek, 2008). Este ambiente exerce uma influência negativa sobre o crescimento e o desenvolvimento de

inúmeras espécies vegetais. Algumas culturas como a soja e milho, no entanto, possuem a capacidade, se manejadas adequadamente, de superar essas alterações garantindo elevada produtividade. O cultivo da soja em rotação com o arroz em áreas de várzea vem crescendo e ocupou cerca de 270 mil hectares no Rio Grande do Sul na safra 2015/2016 (Instituto Rio Grandense do Arroz, 2016). Trabalhos avaliando o desempenho diferencial de genótipos de soja sob condição de alagamento foram realizados (Thomas *et al.*, 2000; Pires *et al.*, 2002; Missio *et al.*, 2010) e identificados genótipos adaptados a condição de saturação do solo. Nesse sentido, esses estudos podem ser utilizados na elucidação dos mecanismos envolvidos no caráter tolerância ao alagamento do solo, para a recomendação de cultivares de soja e como suporte em programas de melhoramento da cultura.

2.4 Alterações em plantas de soja em resposta ao alagamento do solo

As injúrias causadas pelo alagamento do solo variam com o estágio de desenvolvimento da cultura. Durante a germinação, a rápida absorção de água pode levar ao rompimento físico e deterioração da semente da semente (Sayama *et al.*, 2009) e após a emissão da raiz primária, os danos variam com a intensidade e prolongamento do estresse (Komatsu *et al.*, 2009). Desse modo, o cultivo preferencial da espécie ocorre em regiões livres de inundações, devido a redução no crescimento e às perdas significativas na produtividade de grãos que esse estresse pode acarretar. Ainda, a redução no rendimento da cultura é acentuada quando o período de alagamento é ampliado (Rhine *et al.*, 2010). No entanto, cultivares de soja com nenhuma ou pequena alteração na produtividade de grãos sob alagamento foram identificadas (Rhine *et al.*, 2010), bem como a existência de variabilidade genética para esta característica foi relatada (Thomas *et al.*, 2000; Pires *et al.*, 2002).

A privação de oxigênio e a impossibilidade de obter energia por meio da via de fosforilação oxidativa é um dos principais entraves para a tolerância ao alagamento do solo

em plantas (Irfan *et al.*, 2010). Devido a essa incapacidade, ocorre a indução de genes relacionados às rotas de degradação de reservas, glicólise e fermentação, visando manter o metabolismo energético e garantir a sobrevivência durante o período estressante (Voeseck & Bailey-Serres, 2015). No entanto, a eficiência energética destes processos é menor, o que não garante a permanência da planta sob essas circunstâncias, sugerindo que o mecanismo de tolerância ao alagamento do solo é ainda mais complexo.

A tolerância ao alagamento exige a ativação de outros processos que consomem poder redutor ($\text{NADH}+\text{H}^+$ - nicotinamida adenina dinucleotídio reduzida) e colaboram com a conservação do processo glicolítico. Dentre estes processos estão compreendidos a produção de aminoácidos não tricarboxílicos e a oxidação do óxido nítrico por meio de hemoglobinas não simbióticas (Dat *et al.*, 2004; Bailey-Serres & Voeseck, 2008). Ainda, a cascata de sinalização que desencadeia as respostas ao alagamento do solo em plantas, possui como participantes ativos espécies reativas de oxigênio (EROS) e o fitormônio etileno, envolvido tanto na formação de aerênquima e raízes adventícias, bem como no desenvolvimento da parte aérea sob esta condição (Voeseck & Sasidharan, 2013).

A elucidação dos mecanismos envolvidos na resposta das plantas ao estresse por alagamento do solo é imprescindível e tem sido foco de inúmeros estudos com a cultura da soja (Komatsu *et al.*, 2010; Khatoon *et al.*, 2012; Nanjo *et al.*, 2014) e com as mais variadas espécies (Huynh *et al.*, 2005; Magneschi & Perata, 2009; Sasidharan *et al.*, 2013). Desse modo, são descritas neste tópico as principais alterações e mecanismos envolvidos na tolerância de plantas de soja ao encharcamento do solo.

2.4.1 Alterações morfológicas e injúrias

Alterações no sistema radicular são as principais modificações reportadas sob condição de encharcamento do solo em soja. Plântulas de soja, de apenas dois dias, submetidas ao alagamento apresentaram redução do crescimento radicular e hipocótilo em

12 horas após a exposição ao tratamento (Nanjo *et al.*, 2011). Além da redução do crescimento, os autores verificaram a redução no conteúdo de antocianinas, que foi associada a coloração verde menos intensa das plântulas submetidas ao estresse em comparação com as não tratadas. Com o prolongamento do estresse, além do comprimento da raiz principal e hipocótilo, o número e a massa fresca das raízes também são afetados (Komatsu *et al.*, 2009).

Os efeitos individuais causados pelo alagamento e pela privação de oxigênio também foram estudados e além dos danos ao sistema radicular serem superiores durante o alagamento, a redução também foi significativa no crescimento da parte aérea das plantas de soja (Katoon *et al.*, 2012). Dessa forma, a saturação do solo demonstra efeito prejudicial ao desenvolvimento da cultura da soja. O alagamento prolongado durante 15 dias em estádios de desenvolvimento vegetativo (seis e oito nós) e reprodutivo (vagens formadas com aproximadamente dois cm) em diferentes cultivares de soja evidenciou respostas diferentes quanto a produção de massa seca total das plantas (Fante *et al.*, 2010). Além disso, os autores observaram a formação de raízes adventícias e hipertrofia caulinar, que de acordo com os resultados de Thomas *et al.* (2005), podem ser formadas rapidamente após o estresse.

A redução do crescimento, mesmo em estádios iniciais, pode ter consequências negativas sobre a produtividade de grãos (Shi *et al.*, 2008). As maiores perdas de produtividade foram verificadas, no entanto, quando o alagamento coincide com o período de enchimento de grãos, podendo acarretar redução de até 39% da produtividade de grãos de soja (Rhine *et al.*, 2010). Em contrapartida, os mesmos autores identificaram genótipos capazes de tolerar pequenos períodos de estresse e manter os níveis produtivos, sendo uma alternativa para áreas sujeitas a saturação.

2.4.2 Alterações no metabolismo energético

O alagamento do solo reduz a disponibilidade de oxigênio, além da difusão do oxigênio em água ser reduzida em até 10^4 vezes, e limita as trocas gasosas entre as plantas e

a atmosfera (Irfan *et al.*, 2010). Em condição de hipóxia, a concentração de oxigênio no solo varia entre 0,1 e 20,9%. No entanto, rapidamente o oxigênio pode ser consumido, pela atividade da microbiota e um ambiente anóxico (sem oxigênio) é gerado (Bailey-Serres & Voesenek, 2008). A baixa ou nenhuma disponibilidade de oxigênio para a respiração aeróbica gera um déficit energético.

Uma molécula de glicose, na via de respiração aeróbica é completamente oxidada gerando um saldo de 38 moléculas de ATP - adenosina trifosfato. Na ausência de oxigênio, ocorre a redução da síntese de ATP e uma crise energética é instaurada, podendo ocasionar a morte celular e conseqüentemente da planta (Dat *et al.*, 2004). Como alternativa de sobrevivência ocorrem alterações no metabolismo, como o aumento das vias da glicólise e da fermentação anaeróbica (Dat *et al.*, 2004; Bailey-Serres & Voesenek, 2008).

Na glicólise, uma molécula de glicose é quebrada produzindo duas moléculas de piruvato e essa reação gera apenas duas moléculas de ATP e duas moléculas de NADH+H⁺ (Figura 1). A manutenção desta rota requer o consumo deste NADH+H⁺. A regeneração de NAD⁺ ocorre por meio das vias de fermentação láctica e etanólica, que consomem poder redutor produzindo lactato e etanol, respectivamente (Figura 1). Os genes codificadores das enzimas piruvato descarboxilase (*PDC*), álcool desidrogenase (*ADH*), que atuam na rota de fermentação alcoólica foram expressos 71 a 10.000 vezes mais, respectivamente, em hipocótilo e raízes de soja, 12 horas após as plantas serem submetidas a condição de alagamento em comparação com a condição controle (Komatsu *et al.*, 2009). O gene codificador da enzima lactato desidrogenase (*LDH*), atuante na fermentação láctica, também possui sua expressão aumentada e associada à capacidade de resposta de plantas de soja ao alagamento do solo (Komatsu *et al.*, 2009). Esses resultados sugerem a similaridade da cascata de genes ativados pelo alagamento entre soja e as demais culturas.

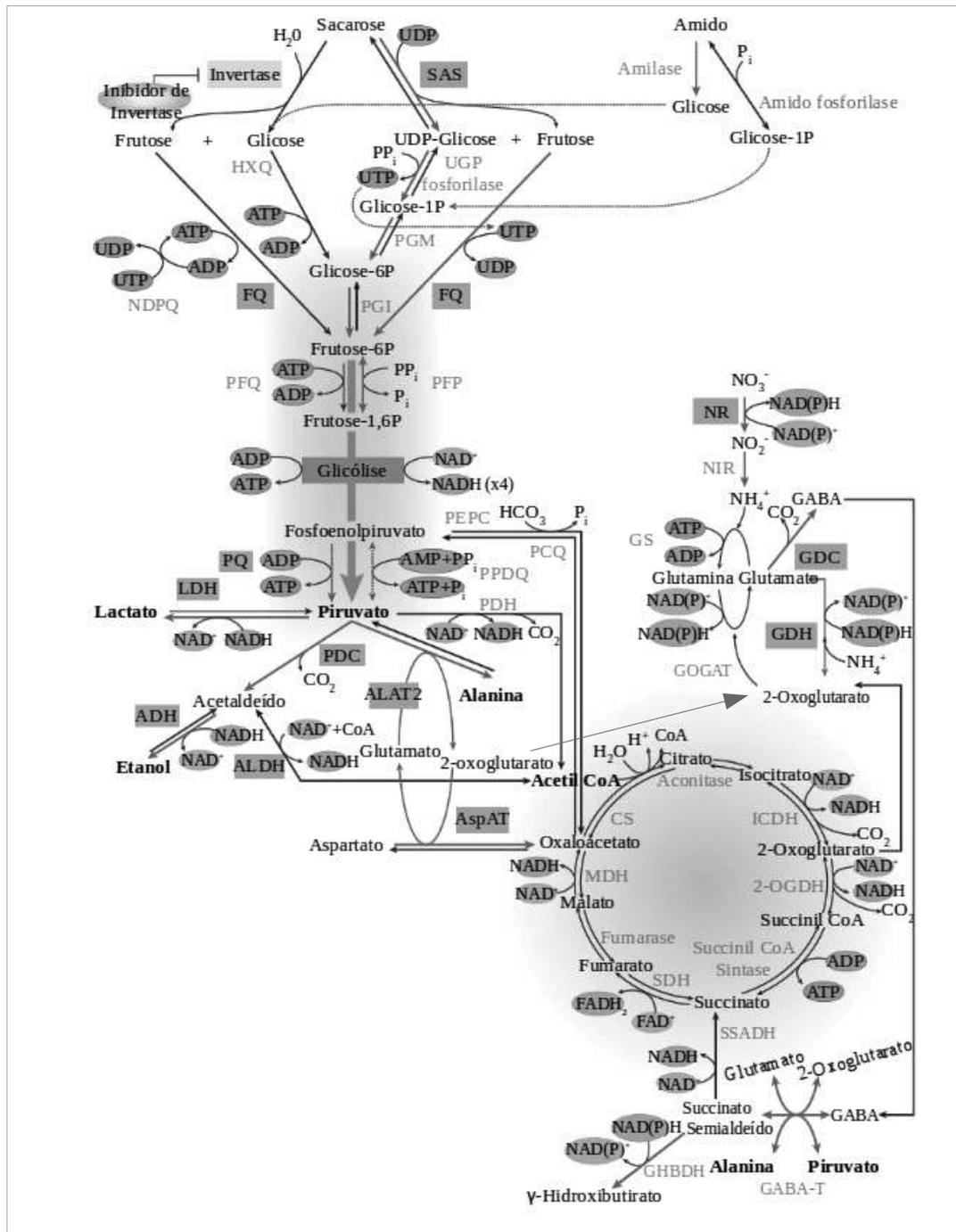


FIGURA 1. Alterações no metabolismo energético de plantas em resposta ao alagamento do solo e deficiência de oxigênio. Com o estresse são induzidas rotas de catabolismo de açúcares, glicólise, fermentação e síntese de aminoácidos não tricarboxílicos. Adaptado de: Bailey-Serres & Voisenek (2008) e António *et al.* (2016). UFRGS, Porto Alegre, RS, Brasil. 2017. Abreviaturas: SAS, sacarose sintase; UGP fosforilase, UDP-glicose fosforilase; PGM, fosfoglicose mutase; HXQ, hexoquinase; PGI, fosfoglicose isomerase; NDPQ, nucleosídeo difosfato quinase; FQ, frutoquinase; PFP, fosfofrutoquinase dependente de P_i ; PFQ, fosfofrutoquinase dependente de ATP; PEPC, fosfoenolpiruvato carboxilase; PCQ, fosfoenolpiruvato carboxilase quinase; PQ, piruvato quinase; PPDQ, piruvato P_i diquinase; PDH, piruvato desidrogenase; LDH, lactato desidrogenase; PDC, piruvato descarboxilase; ADH, álcool desidrogenase; ALDH, acetaldeído desidrogenase; ALAT2, alanina aminotransferase 2; AspAT, aspartato aminotransferase; CoA, coenzima A; ICDH, isocitrato desidrogenase; 2-OGDH, 2-oxoglutarato desidrogenase; SDH, succinato desidrogenase; MDH, malato desidrogenase; CS, citrato sintase; SSADH, Succinato semialdeído desidrogenase; NADH, nicotinamida desidrogenase; GABAT, GABA-transaminase; GDH, glutamato desidrogenase; GDC, glutamato descarboxilase; GOGAT, glutamina 2-oxoglutarato aminotransferase NAD(P)H dependente; GS, glutamina sintase; NIR, nitrito redutase; NR, nitrato redutase.

A reciclagem de NAD^+ por meio da fermentação, no entanto, é insuficiente para atender a demanda energética que a manutenção do metabolismo celular requer. Neste sentido, formas alternativas são exploradas pela planta para conseguir manter o metabolismo e sobreviver ao período adverso. A formação de alanina a partir do ácido pirúvico mediada por aminotransferases é uma dessas formas alternativas (Figura 1). O maior acúmulo de alanina em comparação a outros aminoácidos formados no ciclo dos ácidos tricarbóxicos foi verificado em *Lotus japonicus* (Rocha *et al.*, 2010). Em concordância com esses achados, em soja, a maior expressão do gene Alanina aminotransferase 2 (*ALAT2*) também foi associada ao estresse (Valliyodan *et al.*, 2014).

2.4.3 Sinalização mediada por etileno

Entre as modificações metabólicas durante o período de alagamento do solo está o aumento da produção do hormônio vegetal etileno, um importante sinalizador do estresse, que atua também na formação de aerênquima, raízes adventícias e alongação celular, garantindo a captação de oxigênio externo (Voesenek & Sasidharan, 2013). Devido a volatilidade, o etileno é rapidamente percebido por receptores celulares e acarreta a ativação de genes fatores de transcrição responsivos ao etileno (ERF) (Voesenek & Sasidharan, 2013). O acúmulo de etileno, assim como de espécies reativas de oxigênio e nitrogênio, em células corticais e da epiderme acarretam a morte celular programada, facilitando a formação de aerênquimas e raízes adventícias (Irfan *et al.*, 2010). A formação de aerênquima, mediada pelo etileno, ainda envolve a expressão das enzimas Expansina, Xiloglucano Endotransglicosilase, Pectina, entre outras enzimas associadas a morte celular programada, bem como a interação de etileno com auxina (Irfan *et al.*, 2010). Também, o rápido crescimento do pecíolo sob água foi dependente da interação entre os fitormônios etileno, ácido abscísico e ácido giberélico em *Rumex sp.* (Chen *et al.*, 2010).

2.4.4 Alterações na homeostase celular

A geração de espécies reativas de oxigênio (EROS) é resultado do metabolismo aeróbio. No entanto, existe um equilíbrio entre sua produção e consumo, mantendo a estabilidade celular (Apel & Hirt, 2004). As alterações no conteúdo de O₂, ocasionadas pelo alagamento do solo, são rapidamente percebidas pelas plantas, que alteram o metabolismo, geram EROS e modificam esse equilíbrio, levando ao acúmulo destas moléculas (Irfan *et al.*, 2010). EROS, tem se mostrado de extrema importância na indução da enzima álcool desidrogenase e produção de etanol (Dat *et al.*, 2004). Ainda, existem genes codificadores de fatores de transcrição sensíveis ao status redox e que foram associados a resposta a hipóxia em arábida (Bailey-Serres & Chang, 2005). Desse modo, o acúmulo de EROS em pequenas quantidades atua como sinalizador de hipóxia e é requerido também para o processo de morte celular programada e formação de aerênquima, que facilita as trocas gasosas para as raízes (Voesenek; Bailey-Serres, 2015).

O desbalanço, no entanto, entre a produção e o consumo de EROS afeta a sobrevivência da célula e, posteriormente, da planta inteira, principalmente pela oxidação de lipídios, proteínas de membrana e ácidos nucleicos. A proteção das células aumenta com o acúmulo de metabólitos antioxidantes como prolina, ascorbato e glutatona e com a indução das enzimas endógenas superóxido dismutase, catalase, peroxidase (Irfan *et al.*, 2010). Várias isoformas de enzimas antioxidantes, como ascorbato peroxidase estão distribuídas nos diferentes compartimentos celulares e sob inundação tem sua transcrição reduzida (Shi *et al.*, 2008). Ainda em soja, foi reportada a mudança no nível de proteínas relacionadas a regulação redox, como ascorbato oxidase e peroxidase, glutatona peroxidase e outras peroxidases (Komatsu *et al.*, 2013).

2.4.5 Hemoglobina não simbiótica e óxido nítrico sob hipóxia

Fitoglobinas são proteínas de plantas que possuem um domínio heme, similares a

hemoglobina, e que apresentam moderada a muito alta afinidade com o oxigênio (Hill *et al.*, 2016). Essas proteínas podem ser classificadas em três tipos: as simbióticas, encontradas em nódulos de espécies que fazem fixação simbiótica de nitrogênio; as não-simbióticas, presentes em diferentes partes da planta, ou seja, não associadas a nodulação e as truncadas, que por sua vez são similares as não simbióticas, mas que apresentam tamanho reduzido (Riquelme & Hinrichsen, 2015).

Fitoglobinas ou hemoglobinas não simbióticas (nsHb) são amplamente distribuídas no reino vegetal, sendo encontrada tanto em espécies leguminosas quanto não leguminosas. O gene *AtnsHb1*, codificador de hemoglobina não simbiótica de classe 1 em *Arabidopsis thaliana*, foi induzido por hipóxia e apresenta alta afinidade com oxigênio (Trevaskis *et al.*, 1997). Estudos sugerem também que genes *nsHb1* em plantas apresentam grande importância na manutenção do metabolismo energético (Figura 2) sob condição de privação de oxigênio (Igamberdiev *et al.*, 2011).

Sob condição de hipóxia, óxido nítrico (NO) é produzido e hemoglobinas não simbióticas de classe 1 (nsHb1) aparentam estar envolvidas no controle dos níveis deste composto (Igamberdiev & Hill, 2004). A privação do oxigênio, ocasionada pelo encharcamento do solo, acarreta a redução da síntese de ATP e aumento na produção de NADH (Figura 2). Ocorre também o acúmulo de nitrato que é convertido em óxido nítrico e a indução na expressão de *nsHb1*. Dessa forma, a manutenção do metabolismo redox e estado energético celular é realizada pela *nsHb1* por meio da oxidação do NO e NAD(P)H (Figura 2).

Em esforço para melhorar a tolerância ao encharcamento do solo na cultura do milho, plantas foram transformadas superexpressando o gene exógeno *VHb* (AY278220), uma hemoglobina da espécie *Vitreoscilla stercoraria* (Du *et al.*, 2016). As plantas transformadas mostraram melhoria na capacidade de crescimento sob condição de estresse, apresentando maior crescimento da parte aérea e radicular, em comparação com plantas normais. Além

disso, exibiram aumento da expressão do gene *ADHI* e maior atividade da enzima peroxidase (Du *et al.*, 2016). Dessa forma, a elucidação do mecanismo pelo qual as hemoglobinas auxiliam na tolerância ao alagamento do solo em plantas aparenta grande importância e um tema ainda a ser explorado.

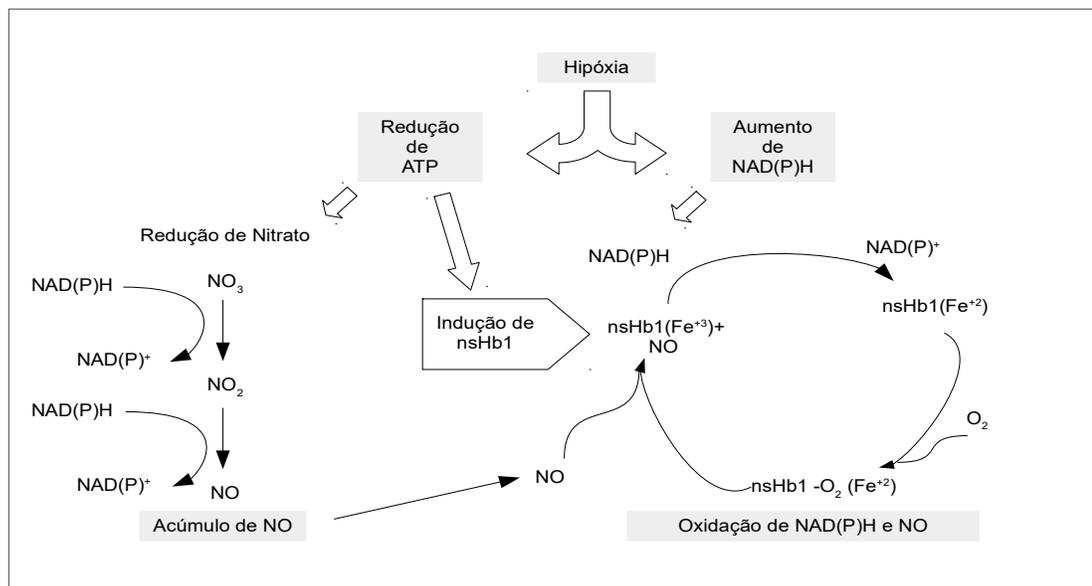


FIGURA 2. Modulação dos níveis de óxido nítrico (NO), realizada por meio de hemoglobina não simbiótica de classe 1 (nsHb1), em resposta a hipóxia. Adaptado de Riquelme & Hinrichsen (2015). UFRGS, Porto Alegre, RS, Brasil. 2017.

2.4.6 Interação com outros estresses

O estresse por excesso de água não ocorre de forma isolada. O estresse salino, ocasionado principalmente pelo aumento do sódio (Na^+) e cloro (Cl^-) e o aumento de amônio (NH_4^+) e ferro (Fe^{2+}) no solo podem atingir níveis tóxicos para as culturas (Thomas & Lange, 2014). O alagamento, de forma geral, afeta a seletividade da membrana e o transporte desses compostos para as células. Neste sentido, a tolerância pode estar associada a manutenção da permeabilidade seletiva da membrana, a presença de um sistema de excreção, ou a capacidade de armazenar estes compostos (Barrett-Lennard, 2003).

O aumento da concentração de Cl^- e Na^+ na parte aérea de plantas é resultado da redução do crescimento e/ou aumento do transporte do íon (Barrett-Lennard, 2003). Em milho, os maiores teores de sódio foram encontrados em genótipo sensível ao alagamento, o

que pode ser determinante para a maior sensibilidade ao estresse (Bispo, 2011). O teor de sódio, portanto, pode ser considerado como elemento importante na manifestação da tolerância de um genótipo ao encharcamento do solo. Em condições salinas, a nutrição da planta com amônio estimula o influxo de sódio, aumentando a toxicidade associada com o acúmulo desse íon, mesmo em genótipos tolerantes de soja (Thomas, 2009). Desse modo, a planta dispõe de três diferentes tipos de tolerância à salinidade: a tolerância ao estresse osmótico, a exclusão de sódio pelas raízes e a tolerância do tecido ao acúmulo de sódio. O menor teor de sódio encontrado no tecido da planta tolerante deve-se provavelmente ao uso de algum destes mecanismos.

Em solos alagados a nitrificação é interrompida pela deficiência de oxigênio e ocorre o acúmulo de amônio (Thomas & Sodek, 2005). A maioria das espécies desenvolve sintomas de toxicidade quando cultivadas em solos com moderada a alta quantidade de amônio, que acarreta a despolarização da membrana e aumento da condutância das células, alterando a absorção de nutrientes (Mak *et al.*, 2006). A presença de nitrato (mais facilmente reduzido) durante a hipóxia no cultivo de soja aumenta a tolerância a esse estresse, o que não acontece quando a fonte de nitrogênio adicionada é o amônio (Thomas & Sodek, 2006). A diferença entre plantas sensível e tolerante está atrelada a capacidade de detoxificar ou absorver em menor quantidade o amônio ou ainda, na vantagem da utilização do nitrato como acceptor de elétrons, regenerando NADH, por meio do ciclo Hb/NO (Figura 2), possibilitando a manutenção da glicólise.

2.5 Tecnologias de sequenciamento de DNA

A determinação da sequência de DNA representa uma ferramenta importante no estudo de diversos fenômenos biológicos. A metodologia descrita por Sanger (1977) foi a primeira metodologia a ser automatizada e capaz de gerar grande quantidade de dados (Connel *et al.*, 1987), permitindo avanços no estudo de genomas inteiros ou de partes de

genomas. Os fragmentos sequenciados pela metodologia de Sanger eram caracterizados por um número pequeno de sequências com tamanho de até 1.000 bases. Os primeiros projetos de sequenciamento de plantas como de arábido (*Arabidopsis Genome Initiative*, 2000) e arroz (*International Rice Genome Sequencing Project*, 2005) utilizaram a metodologia de Sanger.

O surgimento de novas tecnologias de sequenciamento, conhecidas como sequenciamento de nova geração (NGS) a partir de 2005 possibilitou ainda maior geração de dados, com menor custo, maior eficiência e rapidez (Varshney *et al.*, 2009). Inúmeras empresas lançaram suas plataformas de sequenciamento, a 454 Roche; Illumina, Solexa. Ainda, outras tecnologias, como a SOLiD da Applied Biosystems, IonTorrent, PacBio e Helicos Biosciences foram disponibilizadas.

Caracterizado pelo grande número de sequências com tamanho médio entre 35 e 250 bases, o NGS estimulou avanços na área de armazenamento e análise de dados biológicos (Mardis, 2008). A grande quantidade de sequências curtas, aliada a problemas inerentes a cada uma das metodologias desafiam a montagem de genomas, principalmente genomas complexos, com grande número de sequências repetitivas, sendo necessário o desenvolvimento de algoritmos computacionais capazes de analisar essas sequências (Hamilton & Buell, 2012). De maneira geral, a combinação de programas montadores é realizada para contornar essa problemática (Farias, 2013).

O sequenciamento de nova geração tem oportunizado a descoberta de polimorfismos, análise de expressão gênica e genotipagem de populações, estudos evolutivos e de respostas a estresses bióticos e abióticos (Shedure & Ji, 2008). Ainda, o ressequenciamento de genomas é uma ferramenta importante para caracterizar a variação genética existente (Bentley, 2006).

2.5.1 Identificação de genes por meio de sequenciamento

O sequenciamento genômico como estratégia de identificação de genes em plantas apresenta como principais vantagens o elevado volume de informações obtidas e uma visão direta na variação do DNA (Xu *et al.*, 2014). É possível identificar a partir dessa estratégia, variantes alélicas, polimorfismos de seqüências, tais como polimorfismos de um único nucleotídeo (*SNPs*) e as inserções/eliminações (*indels*) (Varshney *et al.*, 2009, Huang *et al.*, 2011). O desenvolvimento de marcadores para seleção assistida, diferenciações entre populações próximas e auxílio à clonagem baseada em mapeamento, passaram a ser beneficiadas pelas informações obtidas no sequenciamento genômico. Além disso, com o advento dos sequenciamentos de nova geração (NGS) tornou-se mais corriqueiro a identificação de níveis diferenciais de metilação do DNA, modificações de histonas, identificação de *splicing* alternativo e análise de interações proteína-DNA (Delseny *et al.*, 2010; Mastrangelo *et al.*, 2012).

No entanto, a identificação de genes por sequenciamento genômico tem desvantagens. A montagem de um genoma é processo complexo e dificultado, particularmente quando o genoma é grande e rico em seqüências repetidas e elementos transponíveis e visto que pequenos fragmentos são gerados pelas tecnologias NGS (Hamilton & Buell, 2012). A dificuldade aumenta quando não está disponível um genoma de referência da espécie ou de espécies aparentadas. A qualidade dos dados, e não apenas o programa montador, tem um efeito sobre a qualidade de um genoma montado, o grau de contiguidade de um conjunto de dados varia muito entre montadores e genomas diferentes e a correção de uma montagem também varia (Salzberg *et al.*, 2012). A qualidade dos dados também pode estar relacionada com a profundidade do sequenciamento, ou seja, o número de vezes que uma região foi sequenciada. Dessa forma, quanto maior a profundidade mais precisos são os dados gerados (Sims *et al.*, 2014).

2.5.2 Sequenciamento do genoma da soja e bancos de dados

O genoma da soja, baseado na cultivar Williams 82, foi completamente sequenciado e montado pela primeira vez em 2010. Neste esforço, foram sequenciados 1,1 gigabases (Gb) de genoma, sendo que 950 megabases foram montadas e ancoradas, a maior parte em 20 cromossomos (Schmutz *et al.*, 2010). Ainda, uma segunda montagem foi realizada, com cerca de 978Mb montados. A disponibilidade de um genoma de referência impulsionou estudos de ressequenciamento (Lam *et al.*, 2010; Zhou *et al.*, 2015) e sequenciamento *de novo* (Li *et al.*, 2014) de soja e espécies aparentadas, bem como a descoberta de SNPs (Wu *et al.*, 2010) e variações estruturais entre genótipos (Anderson *et al.*, 2014).

A grande quantidade de dados gerados requer banco de dados que possibilitem a armazenagem e manipulação das sequências. Além dos bancos de dados universais como o *GeneBank* (<http://www.ncbi.nlm.nih.gov/genbank/>) e *Phytozome* (<http://phytozome.jgi.doe.gov/pz/portal.html>), estão disponíveis bancos de dados específicos (<http://soybase.org/>; <http://www.plantgdb.org/GmGDB/>) que contém informações sobre mapas genéticos, expressão de genes, transcriptomas, metiloma, entre outras informações específicas para a cultura da soja.

3 MATERIAL E MÉTODOS

3.1 Sequenciamento do genoma da soja

3.1.1 Extração de DNA, preparação das bibliotecas e sequenciamento

O DNA total dos genótipos de soja (*Glycine max* (L.) Merrill) Introdução 27 (I27) e BRS 154 foi extraído de folhas jovens através de protocolo CTAB (brometo de cetiltrimetilamônio) modificado (Lodhi *et al.*, 1994; Lefort & Douglas, 1999). A extração consistiu na maceração de 100mg de tecido foliar em microtubos com nitrogênio líquido e adição de 600µL de tampão de extração. Em seguida, foi adicionado PVPP e as amostras foram homogeneizadas e incubadas a 60°C durante 25 min. Após período de resfriamento dos microtubos em temperatura ambiente (TA) foram adicionados 600µL clorofórmio:álcool isoamílico (CIA 24:1) e misturados por inversão durante 4 min. As amostras foram centrifugadas a 10.000g em TA e o sobrenadante transferido para novo microtubo, ao qual foram adicionados 250µL de solução precipitadora de NaCl (5M) e 750µL de isopropanol gelado. As amostras foram homogeneizadas e incubadas a 4°C por 20 min para a formação do pellet de DNA e posteriormente centrifugadas a 10.000g durante 10 min. O pellet foi lavado com 1.000µL de etanol 75% gelado e seco durante 10 min em estufa com circulação de ar a 37°C. O pellet foi ressuscitado em 50µL de TE (pH 8,0) + RNase (20µg/ml) e incubado a 50°C durante 15 min.

A quantificação das amostras e a qualidade das mesmas foi estimada utilizando fluorímetro modelo Qubit® 2.0 (Invitrogen) e espectrofotômetro modelo Genesys 2®

(Thermo Spectronic), respectivamente. Para o sequenciamento a concentração de DNA mínima foi de 50ng/μL e a razão entre as absorbâncias, nos comprimentos de onda de 260 e 280nm, entre 1,8 e 2,0, respectivamente. Atendidos os requisitos de quantidade e qualidade, as duas amostras foram enviadas para sequenciamento no Centro de Genômica Funcional da Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ/USP.

O DNA foi fragmentado para o sequenciamento utilizando o kit de preparação de amostras *Nextera DNA* (Illumina). Para a construção das bibliotecas pareadas foram utilizados inicialmente 20μL de cada uma das amostras de DNA, em uma concentração de 2,5ng/μL. O DNA fragmentado foi purificado com o kit *QIAquick PCR purification* (Qiagen). Os fragmentos purificados foram amplificados com o kit *Nextera PCR Master Mix*, onde foram adicionados indexadores e adaptadores conhecidos, requeridos para a geração de *clusters* e o sequenciamento propriamente dito. A seleção de fragmentos de tamanho adequado (~300pb) foi realizada utilizando o kit *AMPure XP beads*. As bibliotecas foram quantificadas por meio PCR em tempo real, utilizando o kit *Kapa Library Quantification* e então realizado o sequenciamento em equipamento HiSeq 2500 versão 4 (<http://www.illumina.com/products/hiseq-sbs-kit-v4.html>).

3.1.2 Análises das *reads*, montagem dos genomas e identificação de polimorfismos

As *reads* geradas pelo sequenciador foram submetidas à verificação de qualidade das bases e identificação de adaptadores remanescentes por meio do programa FastQC. Somente *reads* com elevada qualidade foram utilizadas para a montagem dos genomas das cultivares de soja I27 e BRS 154. A determinação da cobertura teórica do genoma (CT) alcançada pelo sequenciamento foi calculada pela expressão: $CT = \frac{(TR \times NC)}{(TRG)}$, onde TR é o tamanho das *reads*, NC é o número de *clusters*, ambos após a verificação da qualidade e limpeza e TRG é o tamanho do genoma de referência de soja, que corresponde a 978Mb montadas.

A montagem foi realizada com os programas montadores ABySS (Simpson *et al.* 2009), SOAPdenovo (Li *et al.*, 2010) e a combinação dos programas Velvet (Zerbino & Birney, 2008) e SSPACE (Boetzer *et al.*, 2011), tornando possível a escolha dos melhores resultados. As leituras de sequenciamento e os *scaffolds* foram alinhados com o genoma de referência da espécie *Glycine max* (L.) Merril cv. Williams 82, disponível no *National Center of Biotechnology Information*() por meio dos programas Bowtie 2 (Langmead & Salzberg, 2012) e BLAST (Altschul *et al.*, 1990). Desse modo, foi possível a varredura dos genomas e a identificação de variações nucleotídicas, tanto *SNPs* (single nucleotide polymorfisms) quanto *indels* (inserções e deleções).

3.2 Análise de expressão de genes candidatos à tolerância ao encharcamento do solo

3.2.1 Obtenção do material experimental

O experimento foi conduzido na casa de vegetação de vidro do Departamento de Plantas de Lavoura da Universidade Federal do Rio Grande do Sul entre os meses de abril e maio de 2016 em delineamento inteiramente casualizado, com três repetições. Sementes de soja das cultivares I27 e BRS 154 foram semeadas em copos plásticos (200ml) contendo uma mistura de substrato comercial e vermiculita, na proporção 3:1. Os copos foram mantidos sob iluminação natural e umidade próxima a capacidade de campo até as plantas atingirem o estágio de desenvolvimento V1, quando foram aplicados os tratamentos. Nesse estágio, metade das plantas foi submetida ao estresse, que consistiu na manutenção de uma lâmina de água de 2cm acima do nível dos copos e a outra metade foi mantida em capacidade de campo (controle). As plantas foram mantidas sob estas condições por 0, 24, 48 e 96 horas. Em cada um destes tempos amostras radiculares e da parte aérea foram coletadas, imediatamente congeladas em nitrogênio líquido e armazenadas a -80°C.

3.2.2 Extração de RNA e síntese de cDNA

O RNA total foi extraído dos tecidos radiculares e foliares utilizando o reagente Trizol (Invitrogen), de acordo com as instruções do fabricante. A extração consistiu na maceração de 100mg de tecido em nitrogênio líquido e, em seguida, adição de 1.000µL de Trizol. Então, as amostras foram homogeneizadas e centrifugadas a 10.000rpm durante 15 minutos a 4°C. O sobrenadante foi transferido para um novo microtubo de 1,5ml e incubado durante 5 min em temperatura ambiente (TA). Após, foram adicionados 200µL de CIA (24:1) a cada tubo e esses agitados vigorosamente no vórtex durante 15 seg, seguidos de 5 min de incubação em TA e centrifugação a 10.000rpm durante 15 min a 4°C. A fase aquosa resultante da centrifugação foi transferida para um novo microtubo. Em seguida, foi adicionada solução precipitadora de RNA (0,8M citrato de sódio e 1,2M NaCl) e isopropanol na quantidade correspondente a metade do volume da fase aquosa transferida para o novo tubo. Posteriormente, os tubos foram invertidos para misturar a solução, e incubados durante 10 min em TA e centrifugados a 10.000 rpm durante 15 min a 4°C. O sobrenadante oriundo do processo de centrifugação foi descartado e o precipitado lavado com etanol 70% e centrifugado a 8.000 rpm durante 5 min a 4°C. O sobrenadante foi descartado e o pellet permaneceu durante 10 min em estufa com circulação de ar a 30°C, para secagem. Por fim, o pellet foi ressuspenso em 35µL de água livre de RNase, incubado durante 10 min a 55°C, centrifugado a 8.000 rpm durante 5 min a 4°C para precipitar o material insolúvel e o sobrenadante transferido para um novo microtubo.

A concentração de RNA foi mensurada em espectrofotômetro Genesys 2 (Thermo Spectronic) e, posteriormente, as amostras foram submetidas ao tratamento com DNase I (Invitrogen), eliminando contaminantes de DNA que poderiam influenciar nas análises posteriores. A síntese de cDNA foi realizada utilizando a enzima transcriptase reversa M₋MLV (Invitrogen), segundo protocolo fornecido de empresa fabricante.

3.2.3 RT-qPCR

Foram escolhidos os genes fator de alongação 1- β (*ELF1- β*), actina 11 (*ACTB*) e ubiquitina (*UBI*) como genes de referência endógenos, os quais permaneceram estáveis sob a condição de encharcamento do solo em estudos anteriores (Byefield *et al.*, 2006; Jian *et al.*, 2008; Nakayama *et al.*, 2014; Valliyodan *et al.*, 2014). Os genes hemoglobina 1 (*GLB1*), alanina aminotransferase 2 (*ALAT2*), enolase (*ENO*), proteína contendo domínio *LOB 41* (*LBD41*), xiloglucano endotransglicosilase parcial (*XETP*), álcool desidrogenase 1 (*ADH1*) e ascorbato peroxidase 2 (*APX2*), por sua vez, apresentarem resposta diferencial sob esta condição (Nakayama *et al.*, 2014; Valliyodan *et al.*, 2014) e foram selecionados como genes candidatos a tolerância ao alagamento do solo. A lista dos genes e a sequência dos primers utilizados neste estudo estão na Tabela 1.

TABELA 1. Genes de referência e candidatos à tolerância ao alagamento do solo, locus gênico, sequência dos *primers* (5'-3') utilizados para a análise de RT-qPCR e temperatura de pareamento (*Tm*). UFRGS, Porto Alegre, RS, Brasil. 2017.

Gene	Locus gênico	Sequência <i>forward</i>	Sequência <i>reverse</i>	Tm
<i>ELF1-β^A</i>	Glyma.02g276600	GTTGAAAAGCCAGGGGACA	TCTTACCCCTTGAGCGTGG	64
<i>ACTB^B</i>	Glyma.15g050200	GAGCTATGAATTGCCTGATGG	CGTTTCATGAATTCCAGTAGC	62
<i>UBI^C</i>	Glyma.10g251900	AGCTATTCGCAGTCCCAAAT	CAGAGACGAACCTTGAGGAGA	64
<i>GLB1^C</i>	Glyma.19g190900	CCATTTTAGAACCGGCGTAGC	TAGCCGGTGACCACATTTCTG	65
<i>ALAT2^C</i>	Glyma.04g240800	TTGCAAGTGATCCACGACACA	AAAGGCTTTGGTGGCATGG	64
<i>ENO^C</i>	Glyma.01g026700	ATCTCAATTGCATCGCTGCC	CATCGGAGCAAGTCAAATCCA	64
<i>LBD41^C</i>	Glyma.11g121800	AAAACGTCTTGAAAGGCGCA	CTTTGGACACGTGGCGTATGT	64
<i>XETP^C</i>	Glyma.18g297100	GAGTGAAATCCCATTCGACCA	CGACCAATCCGTTTTCTCCA	64
<i>ADH1^D</i>	Glyma.16g045000	CCTGTGAGGGACTGAGTCTCTTTC	ACAGCAGCAAAGGTCTATCATTCT	64
<i>APX2^D</i>	Glyma.12g073100	CACGGTGCCATAATATTCTCTC	CAACCCAACCTCAATCATCATCAC	62

ELF1 β : fator de alongação 1- β ; *ACT*: actina 11; *UBI*: ubiquitina; *GLB1*: hemoglobina 1; *ALAT2*: alanina aminotransferase 2; *ENO*: enolase; *LBD41*: proteína contendo domínio LOB 41; *XETP*: xiloglucano endotransglicosilase parcial; *ADH1*: álcool desidrogenase 1; *APX2*: ascorbato peroxidase 2. Sequências dos primers foram obtidas de: A (Jian *et al.*, 2008), B (Byefield *et al.*, 2006), C (Valliyodan *et al.*, 2014), D (Nakayama *et al.*, 2014).

Para a reação de RT-qPCR (reverse transcription - quantitative *polimerase chain*

reaction) foram utilizados 10 μ L da amostra de cDNA (2ng/ μ L), 0,4 μ L de cada *primer* (10 μ M), 2 μ L de 1X *SYBER Green Master Mix* (Invitrogen), 2 μ L tampão 10X, 0,5 μ L dNTPs (10mM), 1,2 μ L de MgCl₂ (50mM), 0,2 μ L de *ROX Reference Dye* (Invitrogen), 0,1 μ L de *Taq Platinum* (Invitrogen) e água ultrapura autoclavada para completar o volume de reação de 20 μ L. As amplificações foram realizadas em placas contendo 96 poços PCR-96M2 – HS – C[®] (Axygen) com selador MicroAmp[®] Optical Adhesive Film (Applied Biosystems) e avaliadas em Termociclador 7300 Real Time Systems (Applied Biosystems) seguindo as condições: 50°C por 2 min; 95°C por 10 min; 40 ciclos de 95°C por 15 segundos, 62°C a 65°C (dependendo do par de *primer*) por 1 min e 72°C por 30 s. Os níveis de expressão dos genes endógenos e alvos foram analisados em triplicatas de cada uma das 3 repetições biológicas. Concomitantemente, uma amostra em branco (apenas água) foi utilizada para descartar contaminação e as curvas de dissociação confirmaram a especificidade da amplificação.

3.2.4 Análise da expressão relativa dos genes candidatos

A estabilidade dos genes referência foi avaliada a partir dos valores de *Ct* (cycle threshold) obtidos na reação de RT-qPCR (Wang *et al.*, 2015), utilizado a ferramenta RefFinder (Xie *et al.*, 2012) (<http://fulxie.0fess.us/?ckattempt=1>), que integra os programas computacionais geNorm, Normfinder, BestKeeper e método *Ct*, para comparar e ranquear genes de referência. A classificação mais abrangente do coeficiente de estabilidade fornecida pelo programa, foi utilizada para a escolha dos genes normalizadores.

Para os genes alvo foram calculadas a média, desvio padrão e o intervalo de confiança dos valores de *Ct* de cada tratamento. O ajuste das curvas de amplificação e eficiência da RT-qPCR foi realizada através do software LinRegPCR v.12.2. Valores de R>0,99, com eficiência entre 1,8 e 2 e números de pontos maiores que 4 foram aceitos (Tuomi *et al.*, 2010). Por sua vez, a análise de expressão relativa foi realizada através da fórmula de

Dussault & Pouliot (2006), onde o $\Delta\Delta Ct = (Ct_{\text{alagado}} - Ct_{\text{referência}}) - (Ct_{\text{controle}} - Ct_{\text{referência}})$, sendo o $\Delta\Delta Ct$ a expressão relativa do gene, e a aplicação do resultado em $2^{-(\Delta\Delta Ct)}$ fornece a dimensão de variação.

3.3 Análise da sequência dos genes candidatos

As sequências dos genes candidatos (Tabela 2) diferencialmente expressos analisados por meio de *RT-qPCR*, disponíveis na base de dados da cultura da soja para a cultivar de referência Williams 82 (W82) (<https://www.soybase.org/>) foram alinhadas, por meio da ferramenta BLAST (Altschul *et al.*, 1990), com os *scaffolds* montados pelo programa Abyss, com a finalidade de identificar as sequências nas cultivares em estudo. Foram consideradas verdadeiras sequências com identidade superior a 95%.

TABELA 2. Genes candidatos analisados por sequência gênica. UFRGS, Porto Alegre, RS, Brasil. 2017.

Gene	Locus gênico	Ortólogo em <i>Arabidopsis thaliana</i>	Descrição	Função
ENO	Glyma.19g190900 Glyma.03g190500	AT2G36530.1	Enolase	Metabolismo energético; glicólise
ADH1	Glyma.04g240800 Glyma.06g122600	AT1G77120.1	Álcool desidrogenase 1	Fermentação
ALAT2	Glyma.01g026700 Glyma.02g038100	AT1G70580.1	Alanina aminotransferase 2	Formação de aminoácidos
GLB1	Glyma.11g121800	AT2G16060.1	Hemoglobina 1	Hemoglobina não simbiótica
LBD41	Glyma.18g297100 Glyma.08g365100	AT3G02550.1	Proteína 41 contendo domínio LOB	Formação de raízes adventícias; regulação da transcrição; DNA-ligado
XETP	Glyma.16g045000 Glyma.19g106700	AT5G13870.1	Xilogucano endotransglicosilase	Síntese e organização da parede celular; formação de aerênquima
APX2	Glyma.12g073100	AT1G07890	Ascorbato peroxidase 2	Detoxificação de peróxidos

Após a recuperação das sequências, elas foram alinhadas com a proteína predita pela sequência de referência, por meio do aplicativo FGENESH+, que faz a anotação de genes

com base na similaridade com proteínas conhecidas (Solovyev, 2007). Desse modo, as sequências gênicas foram recuperadas e identificadas as alterações não sinônimas, ou seja, alterações que acarretam mudança na sequência da proteína dos genes alvo.

3.4 Análise da sequência promotora e elementos regulação em *cis*

Com a finalidade de recuperar as sequências promotoras foram utilizados 1Kb flanqueando o sítio de início da transcrição dos genes candidatos à tolerância ao alagamento (Tabela 2). Polimorfismos entre as sequências gênicas das cultivares de soja W82, BRS 154 e I27 foram identificados por meio da ferramenta BLAST. Para a identificação de motivos de regulação *cis*, presentes nas sequências promotoras dos genes alvo, foi utilizado o banco de dados PlantCARE (*Plant Cis-Acting Regulatory Elements*), que contém dados de elementos *cis* que atuam na regulação da expressão gênica em plantas (Lescot *et al.*, 2002).

4 RESULTADOS E DISCUSSÃO

4.1 Análise do sequenciamento e montagem dos genomas

O sequenciamento genômico gerou em média 111,58 milhões de *reads* pareadas, com comprimento médio de 101 pares de base (pb), característica comum nas tecnologias de nova geração (NGS) (Varsheny *et al.*, 2009). Uma vez que o sequenciamento por síntese não é livre de erros, seja durante a construção das bibliotecas ou do sequenciamento propriamente dito, os arquivos contendo as *reads* foram disponibilizados em formato fastq, que contém as sequências e a qualidade de inserção de cada base. Durante o sequenciamento, inserções são menos frequentes que deleções e substituições, respectivamente (Minoche *et al.*, 2011). Assim, a identificação e limpeza de bases com baixa qualidade, ou seja, inseridas de forma errônea na sequência é fundamental para a montagem eficiente dos genomas.

A análise de qualidade, realizada por meio do programa FastQC, evidenciou pequena redução da qualidade em posições terminais nas sequências pareadas de ambas as cultivares (Figura 3), o que é comumente identificado no tipo de tecnologia de sequenciamento utilizado. No entanto, os escores de qualidade foram, em geral, superiores a 28 (Figura 3), indicando alta qualidade e descartando a necessidade de pré-processamento das leituras. Dessa forma, a cobertura teórica alcançada pelo sequenciamento foi de aproximadamente 20 vezes.

A montagem do genoma das cultivares Introdução 27 (I27) e BRS 154 foi realizada utilizando os programas montadores ABySS, SOAP*denovo* e a combinação Velvet e

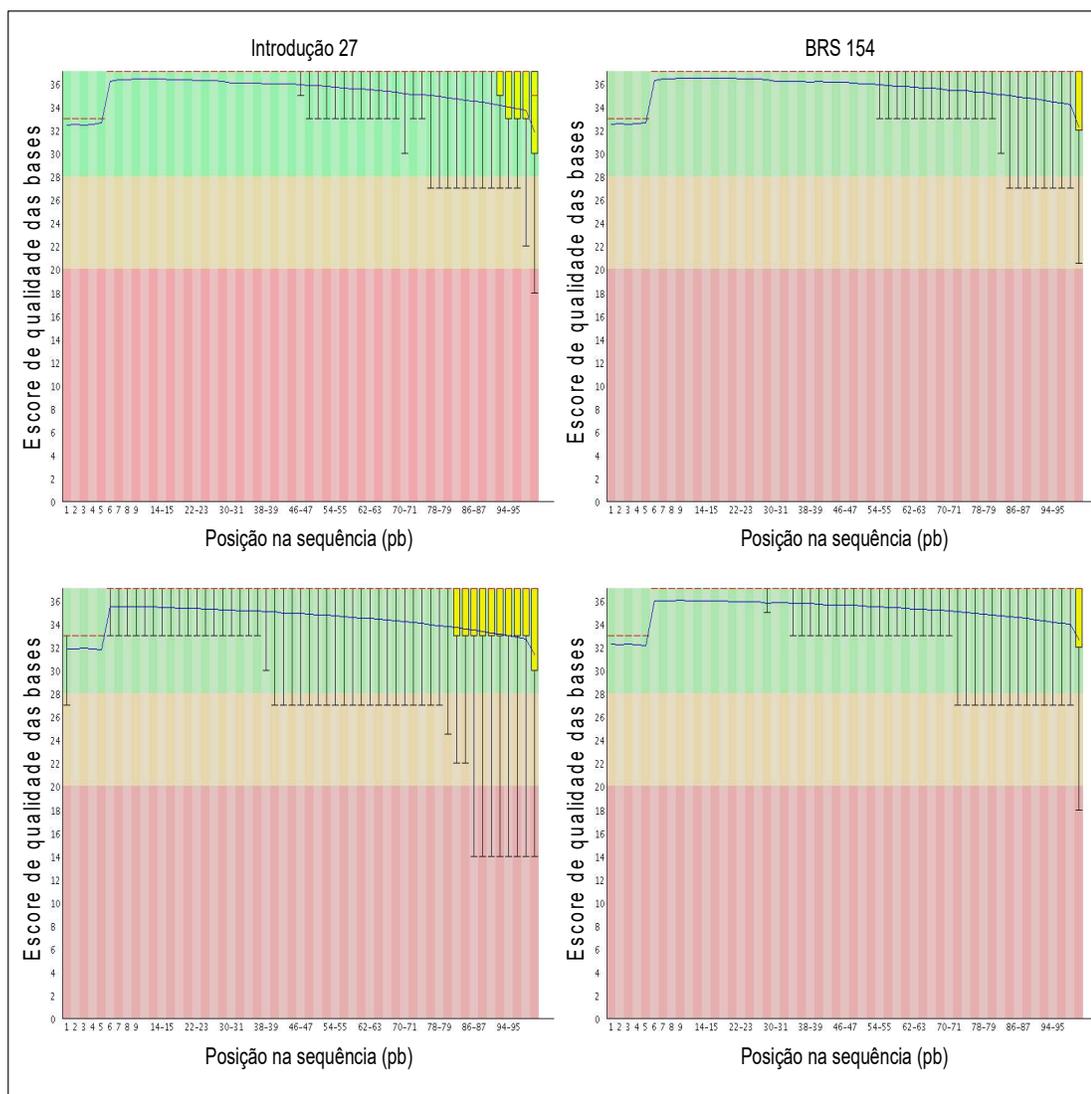


FIGURA 3. Distribuição da qualidade das bases nas leituras provenientes do sequenciamento de bibliotecas pair-end, com 101 pares de bases (pb) de comprimento, dos genótipos de soja Introdução 27 (esquerda) e BRS 154 (direita). As áreas em verde, amarelo e vermelho nos gráficos indicam elevada ($\text{escore} \geq 28$), intermediária ($20 < \text{escore} < 28$) e baixa ($\text{escore} \leq 20$) qualidade das bases, respectivamente. UFRGS, Porto Alegre, RS, Brasil. 2017.

SSPACE, o que permitiu a escolha dos melhores resultados (Tabela 3). Para a cultivar I27 a montagem realizada pelos programas ABySS, SOAPdenovo e Velvet/SSPACE gerou respectivamente 4,39, 2,26 e 0,41 milhões de *contigs*, com tamanho médio de 230pb, 362pb e 219pb. Destes contigs, 44,2%, 43% e 0% foram montados em *scaffolds*, totalizando 4,18, 1,83 e 0,42 milhões de *scaffolds*, com comprimento médio de 245pb, 459pb e 219pb, nessa ordem para os montadores ABySS, SOAPdenovo e Velvet/SSPACE. Esses programas

produziram 4,21, 2,11 e 0,29 milhões de contigs, com tamanho médio de 240pb, 398pb e 215pb, respectivamente, para a cultivar BRS 154. O programa SOAPdenovo montou a maior proporção de contigs em *scaffolds* (57%), seguido pelos programas AByss (37,8%) e Velvet/SSPACE (0%).

TABELA 3. Resumo dos resultados da montagem dos genomas das cultivares de soja Introdução 27 e BRS 154, gerados por meio de diferentes programas montadores. UFRGS, Porto Alegre, RS, Brasil. 2017.

Parâmetros	AbySS	SOAPdenovo	Velvet/SSPACE
----- Introdução 27 -----			
Número de <i>contigs</i>	4.387.028	2.265.179	412.714
Tamanho total de <i>contigs (pb)</i>	1.007.202.447	817.018.714	90.511.800
Maior <i>contig (pb)</i>	24.813	17.478	17.854
Menor <i>contig (pb)</i>	62	3	125
Tamanho médio de <i>contig (pb)</i>	230	361	219
Número de <i>Scaffolds</i>	4.178.769	1.827.451	412.713
Tamanho total de <i>scaffolds (pb)</i>	1.024.945.366	839.272.440	90.511.826
Maior <i>scaffold (pb)</i>	47.402	17.478	17.854
Menor <i>scaffold (pb)</i>	63	100	125
Tamanho médio de <i>scaffold (pb)</i>	245	459	219
N50 (<i>pb</i>)	1.864	1.108	219
L50	103.224	200.471	166.903
<i>Contigs</i> montados em <i>scaffolds (%)</i>	44,2	43	0
<i>Contigs</i> não montados em <i>scaffolds (%)</i>	55,8	57	100
----- BRS 154 -----			
Número de <i>contigs</i>	4.208.711	2.112.811	295.178
Tamanho total de <i>contigs (pb)</i>	1.008.911.936	820.996.688	63.507.669
Maior <i>contig (pb)</i>	35.525	14.599	23.143
Menor <i>contig (pb)</i>	62	4	125
Tamanho médio de <i>contig (pb)</i>	240	389	215
Número de <i>Scaffolds</i>	4.051.063	1.565.985	295.175
Tamanho total de <i>scaffolds (pb)</i>	1.019.782.672	848.178.547	63.507.745
Maior <i>scaffold (pb)</i>	38.314	23.589	23.143
Menor <i>scaffold (pb)</i>	63	100	125
Tamanho médio de <i>scaffold (pb)</i>	252	542	215
N50 (<i>pb</i>)	1.836	1.731	215
L50	109.040	129.577	119.655
<i>Contigs</i> montados em <i>scaffolds (%)</i>	37,8	57	0
<i>Contigs</i> não montados em <i>scaffolds (%)</i>	62,2	43	100

A qualidade da montagem compreende a precisão e a contiguidade, ou seja, o comprimento dos *contigs* ou *scaffolds* (Paszewick & Studholme, 2010). De maneira geral, o programa ABySS mostrou maior eficiência na montagem de ambos os genomas, gerando *scaffolds* maiores, com tamanho total de 1,024Gb e 1,020Gb, respectivamente para os genomas I27 e BRS 154 (Tabela 3). O tamanho dos *scaffolds* é desuniforme e as estatísticas L50 e N50 fornecem informação adicional sobre a contiguidade da montagem. O L50 fornece o número de *scaffolds* que somados medem exatamente a metade do genoma montado e o N50 é o tamanho do menor *scaffold* do L50, e quanto maior o N50 melhor (Yandell & Ence, 2012). Avaliando as estatísticas N50 e L50 o programa ABySS apresentou melhor desempenho, sendo seguido pelos montadores SOAPdenovo e Velvet/SSPACE. A combinação destes últimos montadores foi menos eficiente, gerando *scaffolds* com N50 menor e L50 maior (Tabela 3).

O primeiro relato de sequenciamento genômico da cultura foi em 2010, quando o genoma da cultivar Williams 82 (W82), foi sequenciado pelo método de Sanger e montado. Esse trabalho resultou na montagem de 950Mb ancoradas em 20 cromossomos, representando 85% do genoma predito da cultura, com 8x de cobertura (Schmutz *et al.*, 2010). A montagem atualizada do genoma da cultivar W82, utilizado como referência neste estudo, compreende 978Mb e 1.169 *scaffolds* adicionais não montados. Os respectivos valores de N50 e L50 desta montagem foram 48Mb e 10 (NCBI, 2015). O alinhamento dos *reads* de sequenciamento, obtidos para as cultivares I27 e BRS 154, com o genoma de referência da soja revelou um alinhamento global de 97%, para ambas as cultivares, indicando que boa parte do genoma foi abrangido pelo sequenciamento.

Além do genoma da cultivar W82, está disponível também o genoma da cultivar Enrei. O sequenciamento foi realizado utilizando duas tecnologias NGS (HiSeq2000 e Roche 454 FLX Titanium) e a montagem resultou em um genoma de 928Mb, montado em contigs,

com 22,2x de cobertura (Shimomura *et al.*, 2015).

Estudos utilizando o sequenciamento genômico para acessar a diversidade genética, padrões de seleção, domesticação, melhoramento e características agronômicas tem sido realizados em *Glycine max* e *Glycine soja*, parental selvagem da cultura (Lam *et al.*, 2010; Li *et al.*, 2014; Zhou *et al.*, 2015). Dessa forma, o sequenciamento genômico de qualidade fornece base para análise de variações genéticas importantes na compreensão de processos como a tolerância ao alagamento do solo.

4.2 RT-qPCR

A análise de PCR em tempo real foi realizada com o intuito de verificar a expressão diferencial de genes candidatos à tolerância ao alagamento do solo entre os genótipos sensível (BRS 154) e tolerante (Introdução 27) de soja. Para isso, foram cultivadas plantas com e sem o estresse, durante 0, 24, 48 e 96 horas. Para cada condição, tempo e cultivar foram coletadas três amostras de tecido das partes aérea e radicular, as quais foram submetidos à extração de RNA, síntese de cDNA e RT-qPCR.

Os genes de referência Actina B (*ACTB*), Ubiquitina (*UBI*) e Fator de alongação 1- β (*ELF1- β*) e alvo enolase (*ENO*), álcool desidrogenase 1 (*ADH1*), alanina aminotransferase 2 (*ALAT2*), hemoglobina 1 (*GLB1*), xiloglucano endotransglicosilase parcial (*XETP*), proteína contendo domínio LOB 41 (*LBD41*) e ascorbato peroxidase 2 (*APX2*) foram escolhidos com base na literatura. Desse modo, foi realizada a análise de especificidade, estabilidade dos genes de referência e expressão relativa dos genes alvo.

4.2.1 Amplificação e estabilidade dos genes de referência

A especificidade das amplificações dos genes de referência utilizados nesse estudo pode ser comprovada através de um pico único nas respectivas curvas de dissociação para a Actina B (*ACTB*), Ubiquitina (*UBI*) e Fator de alongação 1- β (*ELF1- β*) (Figura 4).

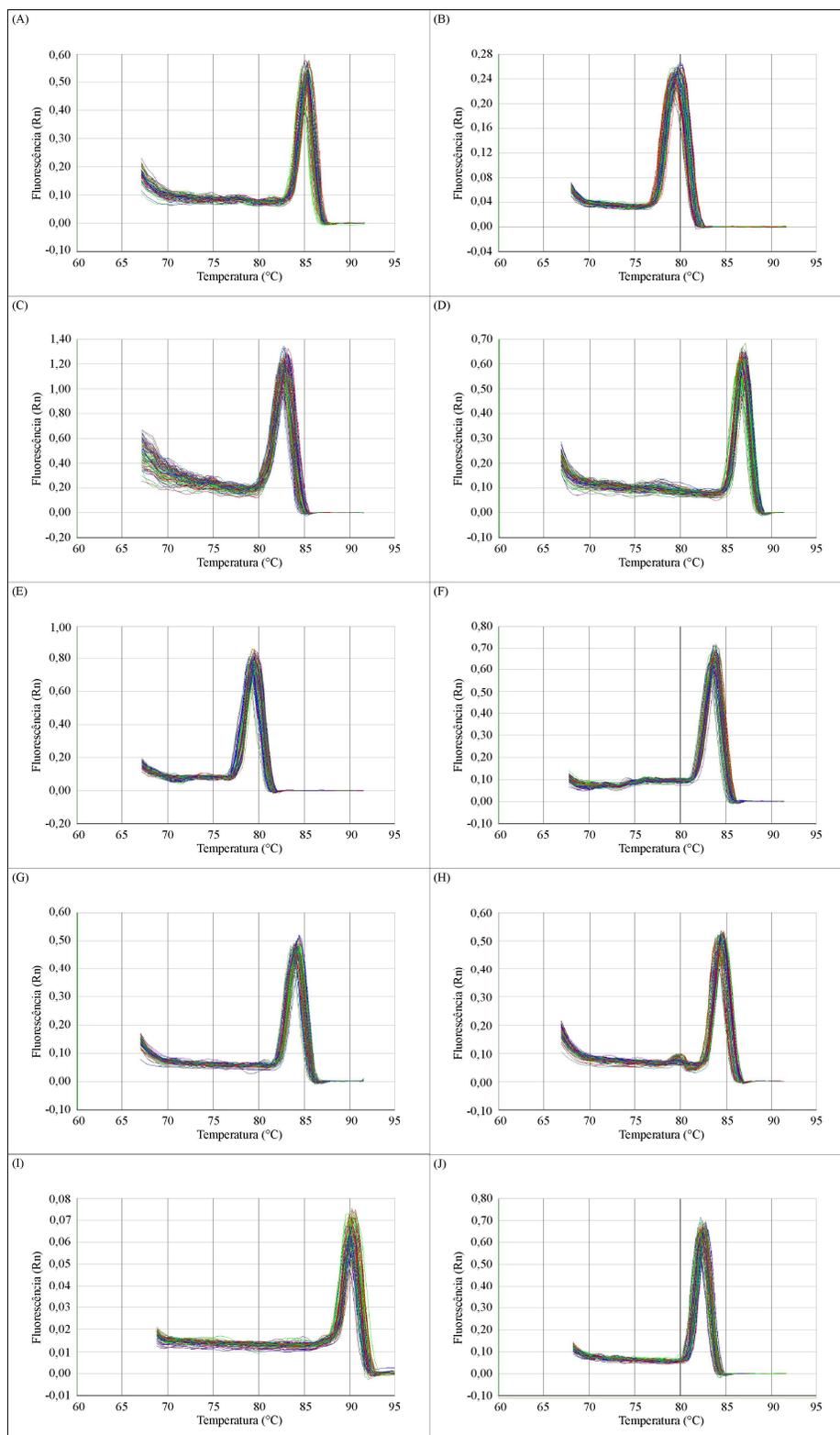


FIGURA 4. Curvas de dissociação dos genes de referência *ELF1-β* (A), *UBI* (B) e *ACTB* (C) e candidatos à tolerância ao alagamento do solo, *ENO* (D), *ADHI* (E), *ALAT2* (F), *GLB1* (G), *XETP* (H), *LBD41* (I) e *APX2* (J), em folhas e raízes de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. UFRGS, Porto Alegre, RS, Brasil. 2017.

As curvas de dissociação mostraram a temperatura na qual ocorre a desnaturação das fitas de DNA e picos de mesma intensidade (tamanho) em uma mesma temperatura, indicando elevada especificidade da reação (Figura 4), o que assegura que os *primers* utilizados anelaram apenas nos genes *ACTB*, *UBI* e *ELF1- β* e que não ocorreu a amplificação de outros fragmentos, ou seja, fragmentos inespecíficos, apoiando utilização dos resultados da análise (Deprez *et al.*,2002). A amplificação do gene *ACTB* apresentou maior quantidade de transcritos com menor número de ciclos na reação de *PCR* em tempo real demonstrando ter uma elevada presença constitutiva, resultando em menor valor de *Ct* médio em comparação aos outros dois genes de referência (Figura 5). No entanto, a amplificação desse gene para os diferentes tempos e condição de alagamento demonstrou elevada variação, o que resultou em uma maior amplitude do intervalo de confiança (Figura 5).

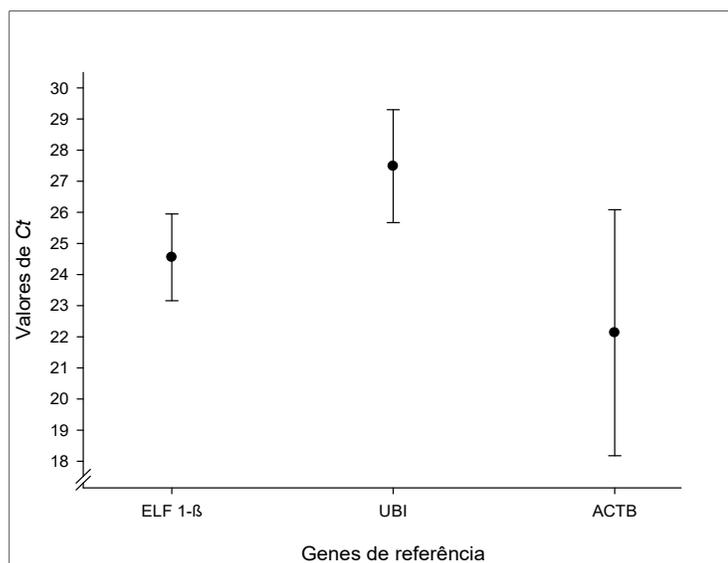


FIGURA 5. Variação na abundância de transcritos. Os pontos indicam o *Ct* médio estimado e as barras verticais o intervalo de confiança ($\alpha = 0,05$) para os genes de referência Fator de Elongação 1- β (*ELF1- β*), Ubiquitina (*UBI*) e Actina B (*ACTB*). UFRGS, Porto Alegre, RS, Brasil. 2017.

Os genes *ELF1- β* e *UBI* apresentaram *Cts* superior à *ACTB*, apresentando, desta forma, menor quantidade nas condições avaliadas. Entretanto esses dois genes demonstraram menor variação frente aos tratamentos em que foram avaliados resultando em

menor amplitude do intervalo de confiança. O valor médio de Ct para os genes *ELF1- β* , *UBI* e *ACTB* foi de 24,55, 27,48 e 22,13, respectivamente (Figura 5).

A estabilidade dos genes considerados como referência, aferida pelo programa computacional RefFinder, que utiliza um compilado de quatro algoritmos, indicou maior estabilidade para o gene *ELF1- β* para os tratamentos utilizados neste estudo (Figura 6). A análise de estabilidade obtida pelo cálculo do Delta Ct demonstrou maior coeficiente de estabilidade para o gene *ACTB* e menor para *ELF1- β* , sendo que o *UBI* apresentou comportamento intermediário, tendo valores respectivos de 4,611, 3,188 e 3,716 (Figura 6A). É importante ressaltar que quanto menor o valor do coeficiente de estabilidade menor é a influência dos tratamentos utilizados no estudo sobre sua expressão e mais estável o gene nas condições avaliadas (Xie *et al.*, 2012). Portanto, quanto menor o valor do coeficiente de estabilidade, mais indicado o gene se torna para utilização como normalizador. A verificação da estabilidade dos genes normalizadores realizada pelos algoritmos BestKeeper, NormFinder e Genorm apresentaram comportamento similar, como maior estabilidade para *ELF1- β* , seguido por *UBI* e *ACTB* (Figura 6B, C e D).

Os genes *ELF1- β* e *ACTB* quando avaliados para condição de alagamento em estudo com a cultura da soja se apresentaram como mais estáveis (Nakayama *et al.*, 2014). No presente estudo, apenas *ELF1- β* apresentou elevada estabilidade, enquanto que *ACTB* foi menos estável que *UBI*. A utilização de dois genes para a normalização da expressão relativa foi necessária, melhorando o desvio padrão e a visualização das diferenças na expressão dos genes alvo (Nakayama *et al.*, 2014). Desse modo, a partir da avaliação de estabilidade, os genes *ELF1- β* e *UBI* foram utilizados para a normalização da expressão relativa dos genes candidatos envolvidos com a tolerância ao alagamento.

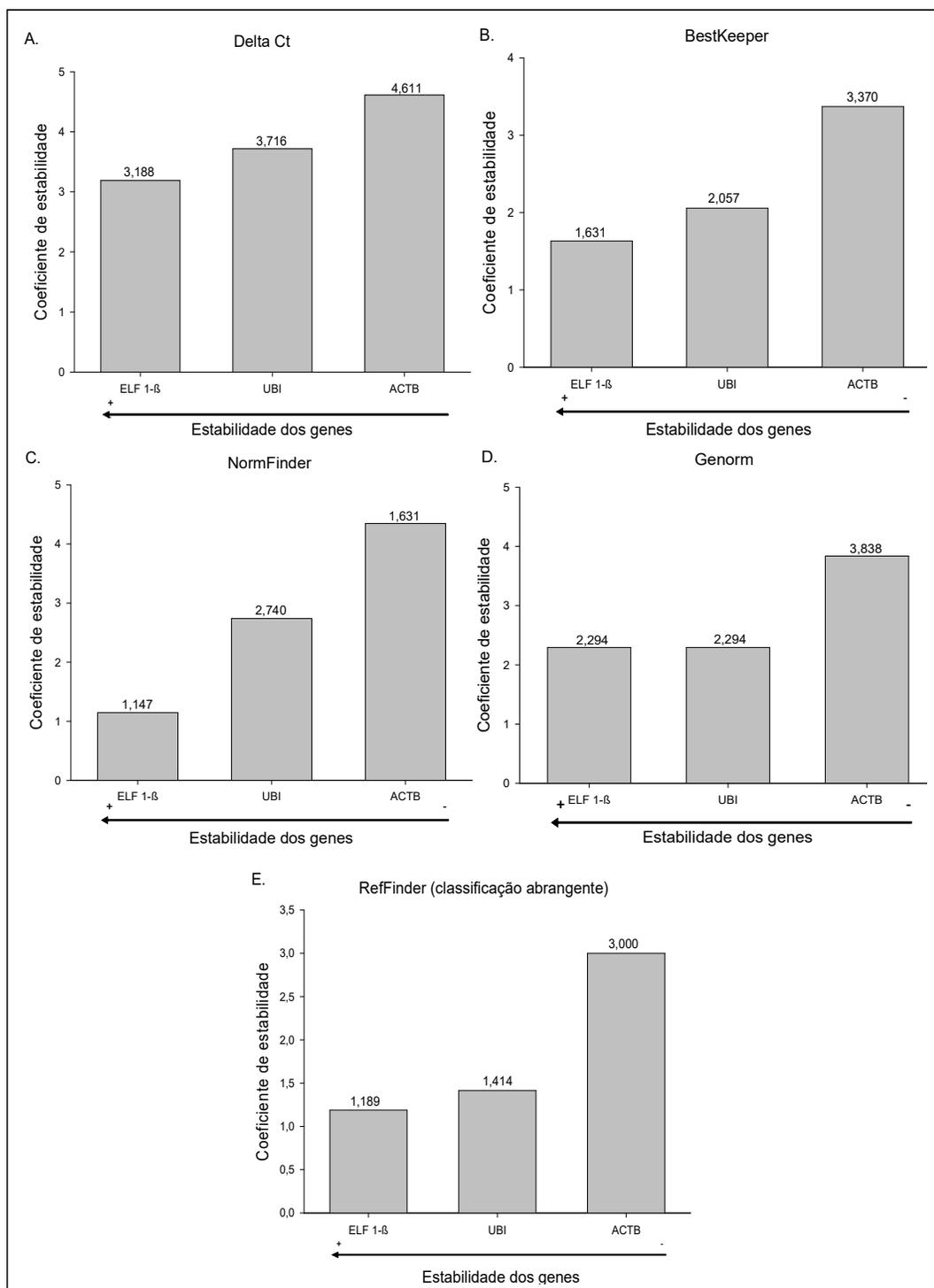


FIGURA 6. Coeficientes de estabilidade dos genes de referência Fator de Elongação 1- β (*ELF1- β*), Ubiquitina (*UBI*) e Actina B (*ACTB*) de soja expressos em folhas e raízes, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Coeficientes calculados utilizando os algoritmos Delta Ct, BestKeeper, NormFinder e geNorm. Quanto menor o valor do coeficiente, maior a estabilidade do gene. UFRGS, Porto Alegre, RS, Brasil. 2017.

4.2.2 Expressão relativa dos genes candidatos à tolerância ao alagamento

Análise de *RT-qPCR* foi utilizada para comparar o perfil de expressão de genes alvo em diferentes tecidos (parte aérea e raiz) e tempos de alagamento/controle em soja. A técnica $\Delta\Delta C_t$ utiliza genes de referência, ou seja, aqueles que não mudam sua expressão com as condições experimentais para comparar genes de interesse. A quantidade de transcritos das cultivares Introdução 27 e BRS 154 nos diferentes tempos de alagamento ou controle (0, 24, 48 e 96 horas) foi comparada com a abundância no tempo 0 (zero) horas da cultivar BRS 154, considerada sensível ao estresse neste trabalho. Assim, foi possível comparar os efeitos do alagamento nos diferentes tempos e como o mesmo afetou o aumento e/ou redução da transcrição dos genes alvo em cada uma das cultivares.

Os genes alvo *ENO*, *ADHI*, *ALAT2*, *GLB1*, *LBD41*, *XETP* e *APX2* foram escolhidos por sua participação em rotas metabólicas, como captação de energia, formação de aerênquima e raízes adventícias e degradação de EROS, afetadas pelo encharcamento do solo, estudados em trabalhos com a cultura da soja (Nakayama *et al.*, 2014; Valliyodan *et al.*, 2014). De maneira geral, o estresse aumentou a expressão de todos os genes mencionados, em especial, para a cultivar I27.

4.2.2.1 Expressão relativa de genes relacionados ao metabolismo energético

Sob alagamento a planta sofre a privação de oxigênio e altera o metabolismo com a finalidade de captar energia para manutenção das funções biológicas e sobrevivência (Irfan *et al.*, 2010). A expressão de genes envolvidos no metabolismo energético anaeróbico está, desse modo, correlacionada à maior captação de energia e tolerância ao estresse. O gene Enolase (*ENO*) codifica uma enzima glicolítica responsável pela conversão de 2-fosfoglicerato a fosfoenolpiruvato, o qual origina, por sua vez piruvato (Hossain *et al.*, 2013). Essa etapa da glicólise é chave para a planta superar períodos de estresse. Nesse sentido avaliar a expressão deste gene sob condição de alagamento é importante.

A expressão do gene *ENO* no tecido aéreo da cultivar BRS 154 não foi alterada significativamente durante as primeiras 48 horas, seja na condição controle ou alagada (Figura 7A). Na avaliação realizada em 96 horas foi evidenciada uma redução da expressão de até 90%, em comparação com o período 0 (zero). No entanto, a redução em ambas as condições indicou que o alagamento não contribuiu para alterar a expressão de *ENO* na parte aérea da cultivar BRS 154. Em contrapartida, a expressão deste gene em folhas da cultivar I27 no tempo inicial (zero) era baixa, em torno de 0,14 vezes a expressão da cultivar BRS 154 e após 48 horas de alagamento ocorreu um aumento da quantidade de transcritos, sendo que a expressão relativa aumentou para 0,71 vezes (Figura 7A). Dessa forma, o alagamento alterou de forma diferencial a expressão a expressão de *ENO* apenas 48 horas após o alagamento, uma vez que a expressão do mesmo em 48 horas na condição controle foi inferior (0,21 vezes), não diferindo do tempo zero (Figura 7A). Neste contexto, para a parte aérea o alagamento induziu, em geral, apenas modestos ganhos em expressão deste gene no genótipo tolerante, evidenciando baixa resposta do mesmo nos tecidos não atingidos diretamente pelo estresse.

No tecido radicular a expressão do gene *ENO* para a cultivar sensível (BRS 154) não diferiu dentro de cada tempo de avaliação entre as condições controle e alagada, exceto na avaliação de 96 horas, quando a expressão relativa na condição controle foi duas vezes superior à alagada (Figura 7B). Desse modo, o prolongamento do estresse pode reduzir a expressão deste gene nas raízes da cultivar BRS 154. Já para a cultivar tolerante, os resultados mostraram um pico de expressão, ou seja, um aumento de, aproximadamente, 14 vezes na expressão de *ENO* na raiz em 48 horas após o alagamento (Figura 7B). A diferença da expressão entre as cultivares, principalmente o aumento de expressão no tecido radicular da cultivar I27 pode ser atribuída ao efeito do alagamento e a capacidade da cultivar tolerante em responder adaptativamente ao estresse. Entretanto, quando avaliado cultivares contrastantes de soja para a característica tolerância ao alagamento do solo o nível de

expressão do gene *ENO* aumentou na cultivar sensível ao estresse cerca de 1,64 vezes, e reduzido para 0,60 vezes na tolerante, apenas no período de 24 horas após o início do estresse, quando comparado a condição controle (Valliyodan *et al.* 2014). Contudo, o mecanismo de tolerância da cultivar tolerante é baseado em modificação estrutural e não ajustes do metabolismo energético (Valliyodan *et al.* 2014).

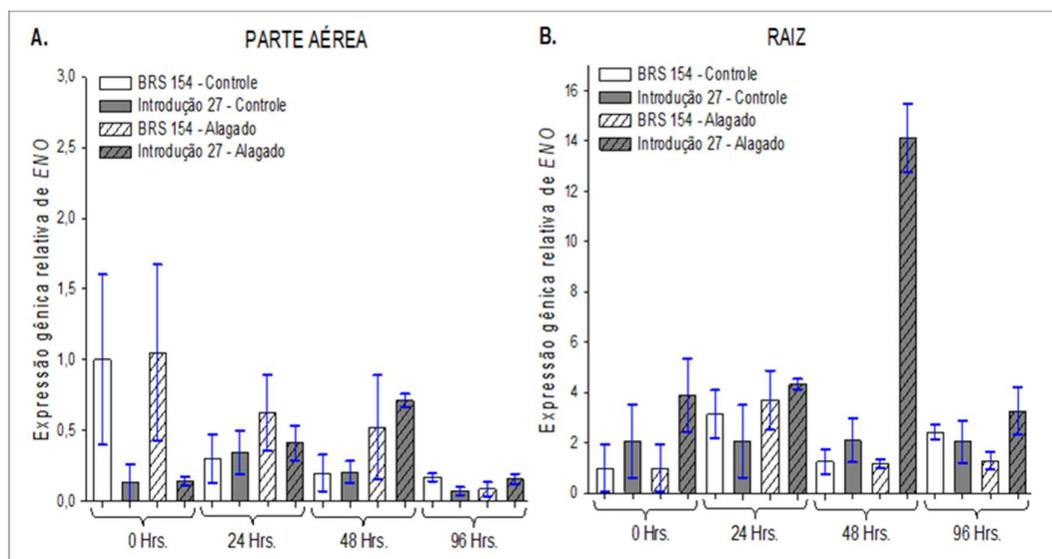


FIGURA 7. Expressão relativa do gene enolase (*ENO*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha=0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.

Ao final da glicólise são geradas duas moléculas de piruvato. O piruvato é um importante componente da via energética sob condição de hipóxia, pois é utilizado como substrato tanto na fermentação láctica, quanto etanólica. Nesta última, o gene álcool desidrogenase 1 (*ADHI*) desempenha papel fundamental na formação do etanol e seria um gargalo nas respostas da planta em ambiente alagado. A participação deste gene na resposta de plantas sob alagamento e/ou hipóxia foi estudada em várias culturas, como arroz, milho, soja e arábida (Sachs *et al.*, 1996; Ismond *et al.*, 2003; Ismail *et al.*, 2009; Komatsu *et al.*, 2009). Os resultados deste trabalho também evidenciaram a importância de *ADHI* para a tolerância ao alagamento em soja. A cultivar I27, em geral, apresentou maior expressão

deste gene, tanto na parte aérea, quanto radicular sob estresse (Figura 8). O aumento da expressão relativa de *ADHI* ocorreu nas primeiras 24 horas de alagamento, sendo 5,49 (Figura 8A) e 42,92 (Figura 8B) vezes superior ao genótipo sensível, respectivamente em folhas e raízes. Esses resultados contradizem o estudo anterior que verificou maior expressão de *ADHI* (10,17 vezes) e genes relacionados com rotas energéticas sob condição de alagamento em cultivar sensível (Valliyodan *et al.*, 2014).

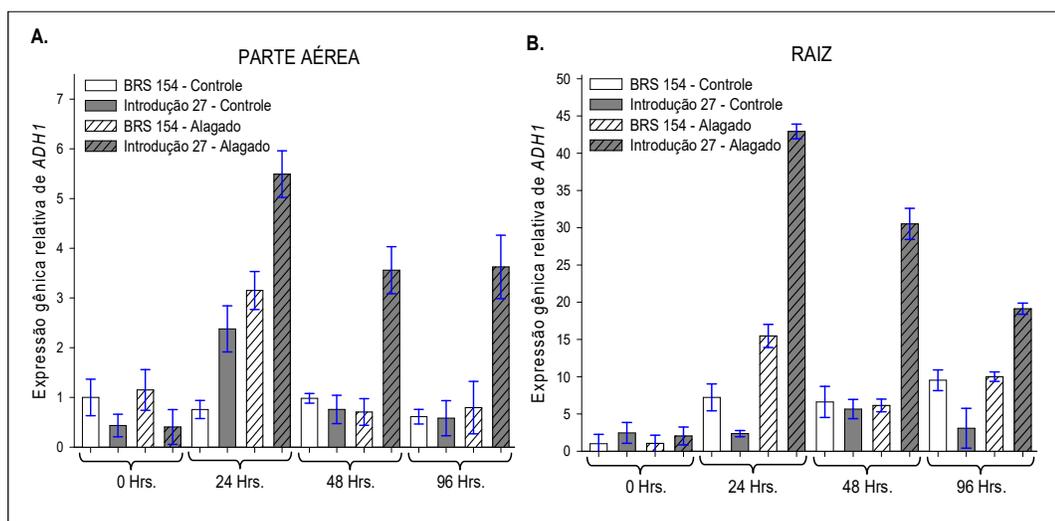


FIGURA 8. Expressão relativa do gene álcool desidrogenase (*ADHI*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.

Em estudos com proteômica a análise de proteínas diferencialmente expressas na cultivar Enrei, sob 72 horas de alagamento, evidenciou aumento de 1,5 e 3,6 vezes na concentração das enzimas enolase e álcool desidrogenase 1, respectivamente (Khatoon *et al.*, 2012). A cultivar em questão apresenta moderada tolerância ao alagamento do solo (Nanjo *et al.*, 2014), corroborando com os resultados do presente trabalho, que identificou maiores níveis de transcritos em cultivar tolerante de soja (Figuras 7 e 8). No entanto, a resposta positiva de apenas um ou poucos genes ou mecanismos não indica necessariamente que o genótipo seja tolerante ao estresse e mantenha seu desenvolvimento, tornando necessária a investigação mais completa do mecanismo de resposta ao alagamento.

Embora rotas anaeróbicas sejam ativadas, a energia garantida seria insuficiente para a manutenção das atividades metabólicas da planta e sobrevivência. Vias alternativas, seja para formação de NADH ou ciclagem de oxigênio estão envolvidas na resposta de plantas de soja em ambiente alagado, entre elas a via de formação de alanina (Rocha *et al.*, 2010) e o ciclo do óxido nítrico e hemoglobinas não simbióticas (Riquelme & Hinrichsen, 2015). Assim, outros genes candidatos, envolvidos na manutenção do metabolismo energético como resposta ao alagamento e que tiveram sua expressão investigada foram alanina aminotransferase 2 (*ALAT2*) e hemoglobina não simbiótica *GLB1*.

A expressão diferencial do gene *ALAT2* foi evidenciada em cultivar tolerante e sensível de soja, apenas em 24 horas, não sofrendo alteração em 72, 168 e 240 horas de alagamento (Valliyodan *et al.*, 2014). A expressão deste gene, em geral, apresentou diferença significativa somente para a cultivar I27, nas avaliações de 48 e 96 horas de alagamento, para a parte aérea e 48 horas para o sistema radicular (Figura 9). A expressão relativa na parte aérea foi inferior a expressão nas raízes, indicando maior participação deste gene no tecido diretamente afetado pelo alagamento. Em trabalho avaliando a expressão de *ALAT2* em soja sob alagamento foi verificada diferença apenas no primeiro dia de alagamento, sendo valores maiores no genótipo sensível e redução para o tolerante, 1,38 e -1,24 vezes quando comparado ao controle, respectivamente (Valliyodan *et al.*, 2014). No mesmo trabalho é relatada o incremento de 3,86 vezes na expressão do gene *ALAT1* na cultivar tolerante, a partir do prolongamento do estresse até o décimo dia (Valliyodan *et al.*, 2014).

Na ausência de alagamento, a expressão relativa de *GLB1*, em geral, foi superior na cultivar BRS 154, cerca de 1 a 1,5 vezes, não diferindo do tempo zero (Figura 10). O alagamento estimulou a expressão deste gene na parte aérea da cultivar BRS 154, sendo o pico de expressão (1,82 vezes) em 96 horas de alagamento. Para a cultivar I27 esse estímulo foi evidenciado durante a avaliação de 48 horas e a expressão permaneceu sem alteração até 96 horas de alagamento, sendo os valores de expressão relativa para estes tempos 1,85 e 1,97

vezes respectivamente (Figura 10A). Esse modesto aumento sob alagamento foi identificado na parte aérea. No tecido radicular, o aumento da expressão de *GLB1* na cultivar I27 foi mais evidente, sendo 8,34, 7,29 e 3,80 vezes maior que a expressão da cultivar sensível em 24, 48 e 96 horas de avaliação, respectivamente (Figura 10B). Os níveis transcricionais e de proteínas hemoglobina não simbiótica foram regulados positivamente pelo alagamento em fases iniciais de desenvolvimento da cultivar Enrei (Komatsu *et al.*, 2009). Em apenas 12 horas de alagamento a quantidade de transcritos foi 95 vezes superior na condição alagada em comparação com o a condição controle.

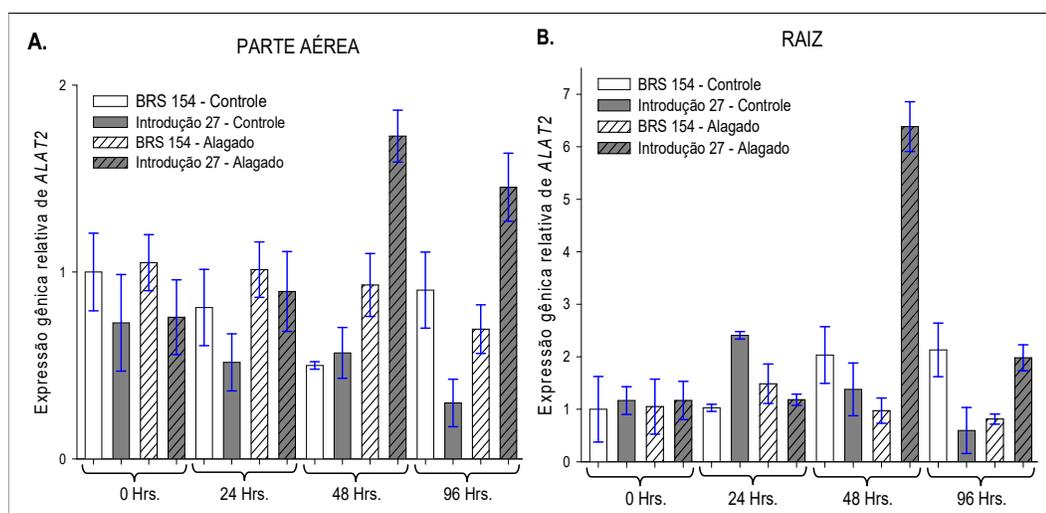


FIGURA 9. Expressão relativa do gene alanina aminotransferase 2 (ALAT2) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha=0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.

No período de 24 horas após o início do alagamento é relatado os maiores indícios da presença de hemoglobina não simbiótica quando comparado a condição controle, superando em mais de 1000 vezes o valor inicial (Komatsu *et al.* 2009). Estes resultados corroboram com o presente trabalho e inferem para a elevada importância de hemoglobina não simbiótica na resposta de plantas de soja ao alagamento do solo. No entanto, os resultados obtidos no presente trabalho também demonstram redução na expressão relativa

do gene *GLB1* na cultivar I27 no período mais longo do estresse (Figura 10B), sugerindo que este participe de uma resposta inicial ao alagamento.

O aumento no nível transcricional dos genes *ENO*, *ADHI*, *ALAT2*, *GLB1*, principalmente na cultivar tolerante, evidencia que a mesma é capaz de superar o encharcamento do solo através da reorganização do metabolismo energético. Alam *et al.* (2010), avaliando o proteoma de soja sob alagamento durante 3 e 7 dias de estresse, verificaram a presença de isoformas das enzimas Enolase e Álcool Desidrogenase, evidenciando a importante participação das mesmas na resposta das plantas de soja ao estresse. Ainda, a indução dos genes das vias glicolítica e de fermentação, bem como o aumento dos níveis de hemoglobina não simbiótica foram verificados por Komatsu *et al.* (2009), apoiando os resultados deste experimento.

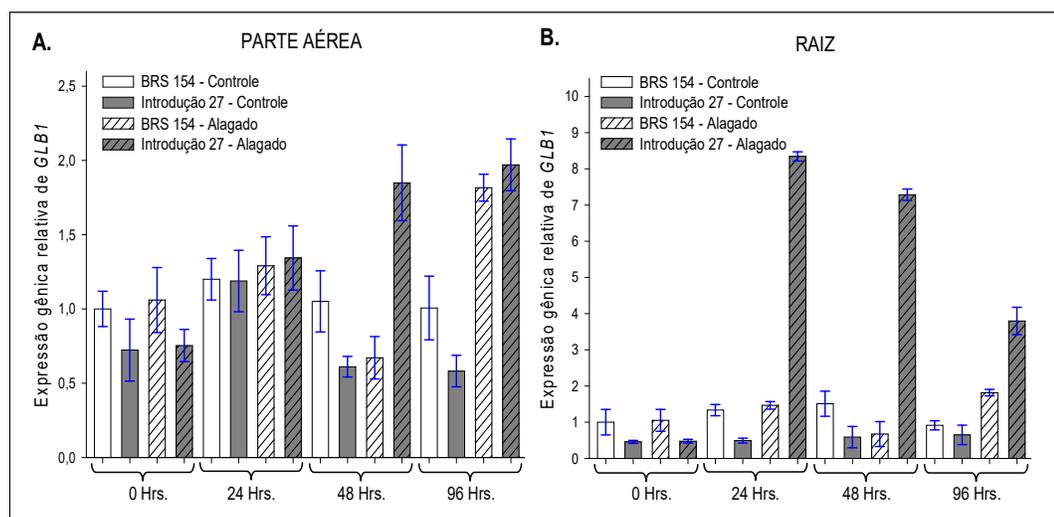


FIGURA 10. Expressão relativa do gene hemoglobina 1 (*GLB1*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha=0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.

4.2.2.2 Expressão relativa de genes relacionados a modificações estruturais

O alagamento acarreta redução da síntese de ATP, em decorrência da privação de oxigênio, e ocasiona a alteração do metabolismo energético. No entanto, também ocorrem modificações estruturais, como a formação de aerênquima e raízes adventícias, que podem

facilitar a difusão de oxigênio e prolongar a sobrevivência das plantas. A avaliação da expressão relativa dos genes *LBD41* (Figura 11) e *XETP* (Figura 12), com efeitos sobre modificações estruturais na cultura da soja, revelaram a participação dos mesmos na resposta ao alagamento na cultivar tolerante.

O gene *LBD41* se mostrou responsivo ao estresse na parte aérea e raízes de soja das cultivares avaliadas (Figura 11). O aumento da expressão, visualizado principalmente na cultivar tolerante (I27), já nas primeiras 24 de alagamento. Na parte aérea o aumento foi de aproximadamente 4, 12 e 8 vezes, respectivamente para 24, 48 e 96 horas de encharcamento do solo (Figura 11A). Em raízes, o alagamento acarretou aumento geral da expressão relativa em ambas as cultivares (Figura 11B). Vinte quatro horas após o alagamento a expressão de *LBD41* era similar para as cultivares BRS 154 e I27 (7,65 e 8,68 vezes). Na avaliação de 48 horas a expressão deste gene se manteve elevada apenas na cultivar tolerante (13,88 vezes) e sofreu redução na cultivar sensível (2,16 vezes). Após 96 horas ocorreu um novo aumento da expressão de *LBD41* na cultivar sensível BRS 154 (6,87 vezes), enquanto na cultivar tolerante I27 o nível de expressão permaneceu elevado, cerca de 13 vezes (Figura 11B).

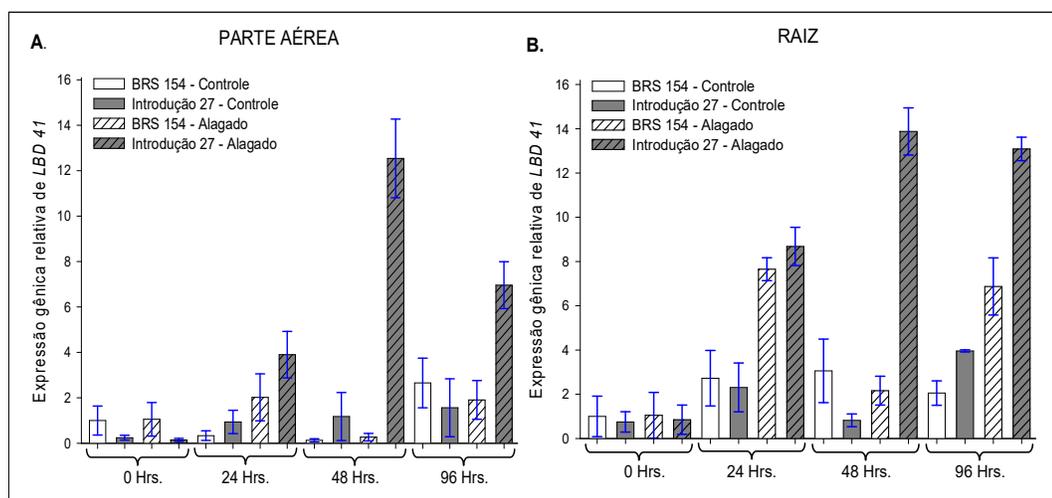


FIGURA 11. Expressão relativa do gene (*LBD41*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha=0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.

A família de genes fatores de transcrição *LOD* (*lateral organ boundary*), a qual faz parte o *LBD41*, é específica de plantas e dentre suas atribuições está a formação de raízes laterais em arábidoopsis (Okushima *et al.*, 2007). Também, a maior expressão de genes da família *LOB* foi associada a maior formação de raízes adventícias em cultivar tolerante de soja, onde o gene *LBD41* apresentou maior expressão com o prolongamento do estresse (Valliyodan *et al.*, 2014).

O gene xiloglucano endotransglicosilase parcial (*XETP*), codifica uma enzima que atua na degradação e relaxamento da parede celular. No presente experimento *XETP* se mostrou influenciado pelo alagamento tanto em tecido foliar, quanto radicular (Figura 12). Em geral, a maior expressão de *XETP* foi em raízes da cultivar I27, em ambas as condições avaliadas (Figura 12B). O alagamento acarretou elevada expressão deste gene em todos os períodos de alagamento, em especial nas 24 horas, quando a expressão foi cerca de 24 vezes superior e mantida elevada até a avaliação de 96 horas (Figura 12B). Maiores níveis de expressão de *XETP* foram identificados em cultivar sensível de soja por Valliyodan *et al.* (2014), que relataram também a expressão diferencial apenas em 24 horas de alagamento.

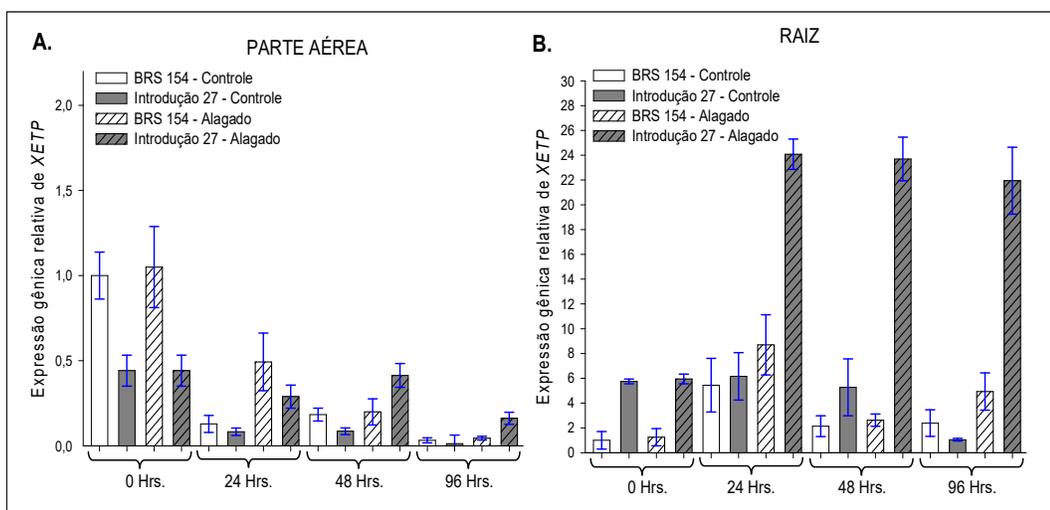


FIGURA 12. Expressão relativa do gene (*XETP*) na parte aérea e raiz de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha = 0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.

A análise de expressão dos genes relacionados à formação de estruturas adaptativas evidenciou maior expressão dos genes *LBD41* e *XETP* sob condição de alagamento no sistema radicular da cultivar tolerante (Figuras 11 e 12). Este resultado foi oposto às expectativas, uma vez que a cultivar I27 não é caracterizada pela elevada formação de aerênquimas e raízes adventícias. No entanto, os maiores níveis de expressão relativa destes dois genes após 24 horas de alagamento e a persistência durante as 96 horas de avaliação (Figuras 11 e 12), sugerem que mecanismos de alteração estruturais podem estar atuando, concomitantemente na resposta de tolerância ao alagamento apresentado pelo cultivar I27.

4.2.2.3 Expressão relativa de genes relacionados a detoxificação de EROS

Dentre os genes candidatos relacionados com a detoxificação de EROS responsivos ao alagamento do solo na cultura da soja foi evidenciada a expressão diferencial do gene *APX2* entre os genótipos avaliados, principalmente no tecido radicular (Figura 13). De maneira geral, o alagamento acarretou uma redução na expressão relativa do gene no tecido radicular da cultivar sensível e um aumento nas raízes da cultivar I27 (Figura 13B). O aumento da expressão na cultivar tolerante foi de aproximadamente quatro vezes em 48 horas a partir do início do estresse e três vezes em 96 horas (Figura 13B). A enzima codificada por este gene participa da rota de detoxificação de EROS, como o peróxido de hidrogênio (Kausar *et al.*, 2012). Neste sentido, a maior expressão do gene *APX2*, que codifica esta enzima, pode contribuir para as respostas adaptativas da planta diante do estresse.

Estudos anteriores mostraram a redução da expressão de proteínas e atividade de peroxidases sob condição de alagamento em soja (Shi *et al.*, 2008; Alam *et al.*, 2010; Wang *et al.*, 2016). A redução dos níveis de *APX2* pode, nestes casos, estar associada ao acúmulo de espécies reativas de oxigênio. Assim, os resultados deste trabalho sugerem que o aumento no nível de transcritos na cultivar I27 e a manutenção de níveis mais elevados com a

continuidade do período de alagamento (Figura 13B) podem desempenhar um papel importante na resposta da cultivar tolerante frente ao estresse oxidativo acarretado pelo alagamento do solo.

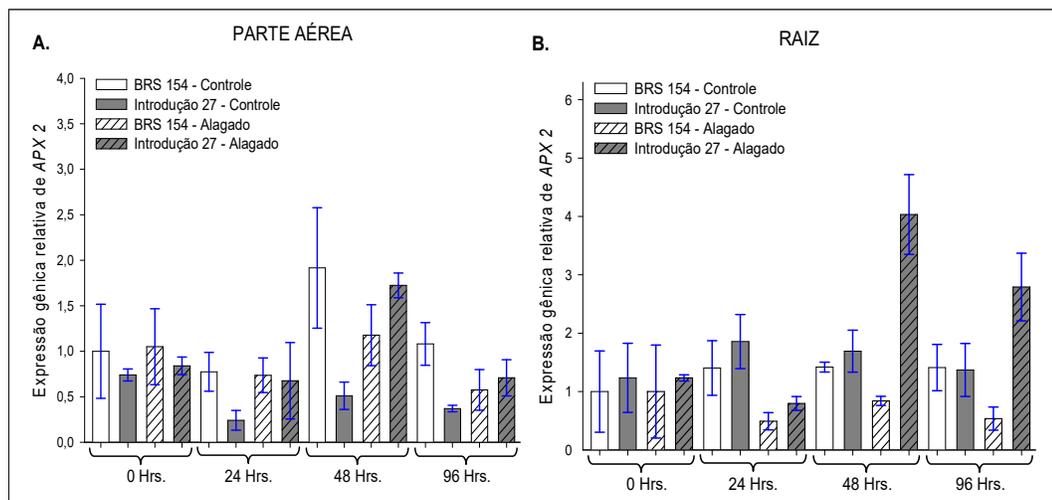


FIGURA 13. Expressão relativa do gene ascorbato peroxidase 2 (*APX2*) em folhas e raízes de plantas de soja, das cultivares Introdução 27 e BRS 154 submetidas a 0, 24, 48 e 96 horas de alagamento ou condição controle. Barras verticais indicam o intervalo de confiança ($\alpha=0,05$). UFRGS, Porto Alegre, RS, Brasil. 2017.

4.3 Análise das sequências dos genes candidatos à tolerância ao alagamento

Os genomas das cultivares estudadas foram comparados entre si e com o genoma de referência da Cultivar Williams 82 para determinar as variações existentes entre os mesmos. As variações foram filtradas com profundidade mínima de 10 leituras e qualidade superior a 30. O maior número de variações foi detectado entre as cultivares BRS 154 e I27, entre a cultivar BRS 154 e Williams 82 e entre I27 e Williams 82, respectivamente, sendo mais frequente o polimorfismo de nucleotídeo único ou *SNP* (Tabela 4). Em ordem, foram encontrados 224.591, 117.149 e 20.952 *SNPs* nestas comparações. A razão entre transições e transversões (TS/TV), ou seja, o número de substituições A↔G e T↔C em relação ao número de substituições A↔T, A↔C, G↔T e G↔C foi de 1,64, 1,78 e 2,08, respectivamente entre as comparações BRS154 x I27, BRS 154 x Williams 82 e I27 x Williams 82 (Tabela 4). A maior frequência de transições também pode ser constatada

através dos gráficos de frequência de substituições, para cada uma das comparações (Figura 14). Os valores de TS/TV detectados neste estudo foram ligeiramente inferiores aos encontrados em estudo com cultivares de soja contrastantes para a resistência ao Vírus do Mosaico Amarelo, onde foi encontrada aproximadamente 2 vezes mais transições que transversões (Yadav *et al.*, 2015).

TABELA 4. Variações genômicas detectadas no alinhamento entre os genomas das cultivares BRS 154, Introdução 27 e Williams 82. UFRGS, Porto Alegre, RS, Brasil. 2017.

Genomas	Número de <i>Indels</i>	Número de <i>SNPs</i>	TS/TV*
BRS 154 x Williams 82	16.236	117.149	1,78
Introdução 27 x Williams 82	2.257	20.952	2,08
BRS 154 x Introdução 27	18.750	224.591	1,64

* Razão entre substituições do tipo transição e transversão.

Da mesma forma, as alterações do tipo inserções e deleções (*indels*) foram detectadas em maior número na comparação entre os genótipos sensível e tolerante de soja estudados neste trabalho (Tabela 4). Foram detectados 18.750, 16.236 e 2.257 *indels* entre as cultivares BRS 154 e I27, BRS 154 e Williams 82 e I27 e Williams 82, respectivamente. O comprimento de *indels* oscilou entre 1 e 18pb em todas as comparações e de maneira geral, existe uma elevada frequência de *indels* pequenos, tanto inserções quanto deleções de até 5pb, além de uma maior frequência de inserções (Figura 14).

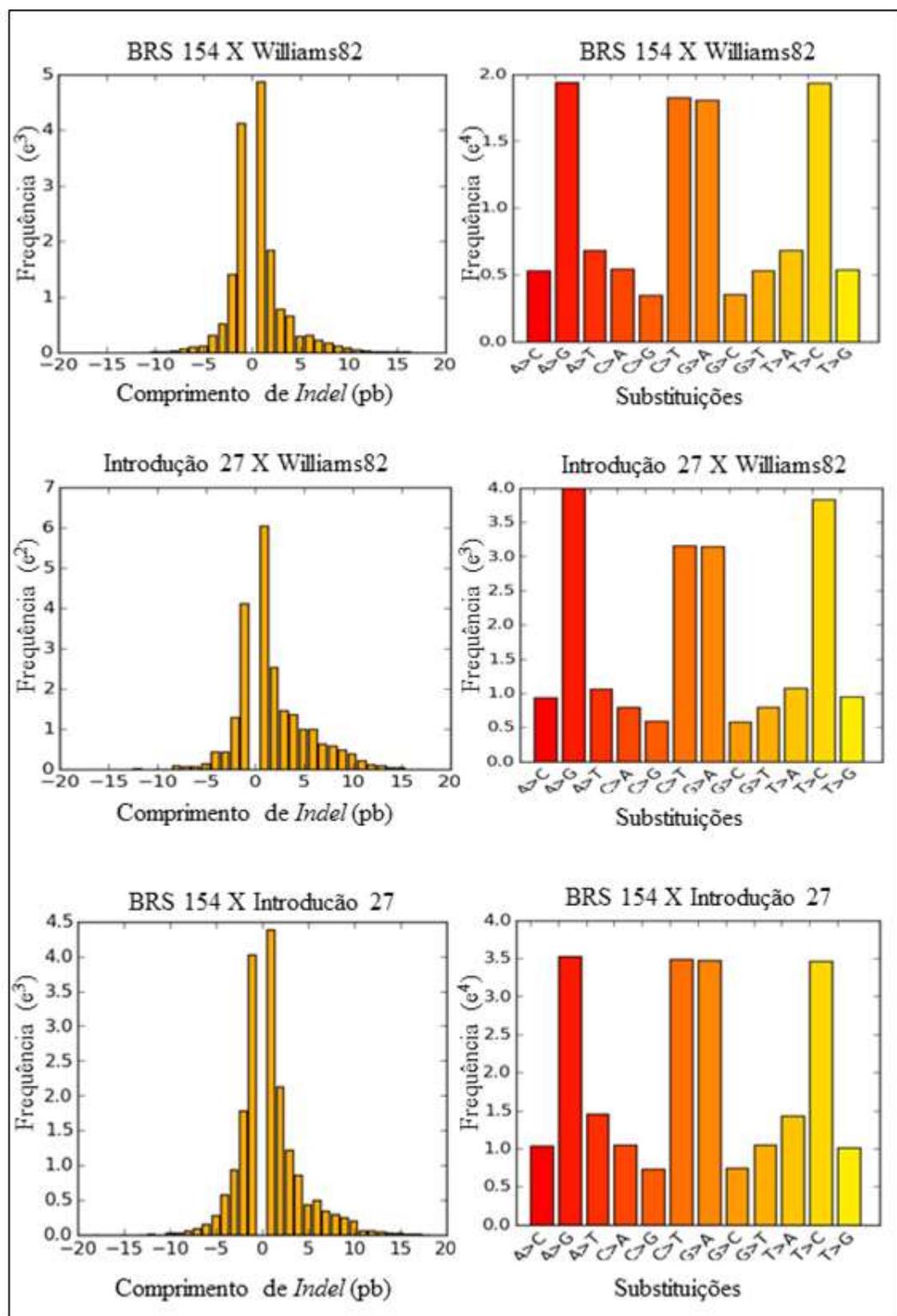


FIGURA 14. Frequência de *indels* e dos diferentes tipos de substituições nos genótipos BRS 154 e Introdução 27 em comparação com a cultivar de referência Williams 82 e entre si. UFRGS, Porto Alegre, RS, Brasil. 2017.

Para recuperar a sequência dos genes candidatos à tolerância ao alagamento do solo foi realizado o alinhamento das sequências gênicas de referência, disponíveis na base de

dados da soja e *scaffolds* e então, as sequências reconstruídas foram comparadas com as proteínas codificadas pelos respectivos genes. Desse modo, foi possível identificar as alterações de nucleotídeos que acarretam mudanças nas proteínas ou na posição dos éxons dentro da sequência gênica.

O gene Enolase (Glyma.19g190900) possui 4737pb de sequência e codifica uma proteína com 444 aminoácidos. A sequência codificadora deste gene, na cultivar de referência Williams 82, apresenta 17 éxons, códon de iniciação na posição 70, códon de terminação na posição 4.128 e sinal poliA na posição 4.582 (Figura 15A). A sequência recuperada no genótipo BRS 154 apresenta uma inserção de 8pb na posição 26 e outras pequenas alterações na porção não codificadora do gene, que mudam apenas a posição dos éxons dentro da sequência e não acarretam alteração na sequência da proteína. Foram recuperados todos os 17 éxons e os códons de iniciação, terminação e sinal poliA foram localizados nas posições 78, 4.134 e 4.589, respectivamente. Para a cultivar Introdução 27, o sequenciamento não recuperou toda a sequência gênica. O fragmento não sequenciado está posicionado no éxon 14 e dentro porção sequenciada nenhuma alteração não sinônima foi verificada (Figura 15A).

O gene Glyma.03g190500, parálogo do gene Glyma.19g190900, contém uma sequência de 4.502pb de comprimento e também codifica uma Enolase com 444 aminoácidos e 98% de identidade. A sequência codificadora na cultivar Williams 82 está localizada entre as posições 66 e 4.179 e o sinal de término de transcrição na posição 4.193 (Figura 15B). O sequenciamento recuperou apenas 39% do gene Glyma.03g190500 da cultivar BRS 154 e nenhum polimorfismo foi verificado. No entanto, a predição da proteína nas regiões sequenciadas e coincidentes com o gene de referência apresentou diferenças, como maior comprimento do éxon 1 e ausência do segundo éxon (Figura 15B).

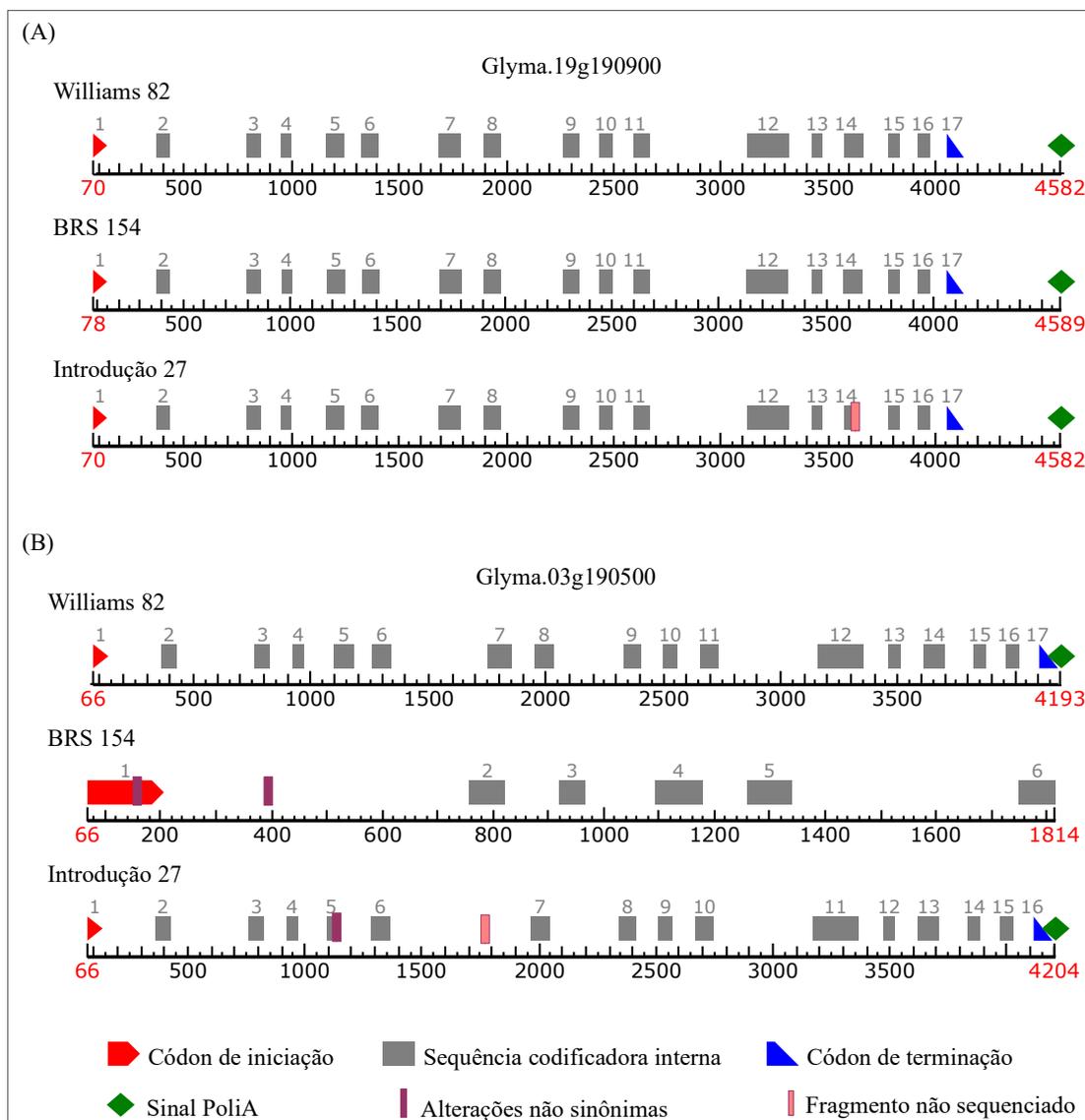


FIGURA 15. Posição dos éxons preditos dos genes Glyma.19g190900 (A) e Glyma.03g190500 (B), codificadores da enzima Enolase, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.

Esse resultado controverso é explicado porque uma pequena fração do gene foi montada e com isso muitas marcas foram perdidas. Já para a cultivar Introdução 27 foi recuperada uma sequência de 4.513pb, com uma cobertura de 99% da sequência de referência. Uma inserção de 20pb, na posição 1.111 acarretou a redução dos aminoácidos codificados pelo éxon 5. Ainda, uma falha na reconstrução do gene, entre as posições 1.764 e 1.805, não permitiu a identificação de um éxon. Sendo assim, foram identificados 16 éxons na cultivar I27, posicionados de forma diferencial na sequência gênica (Figura 15B), uma

sequência codificadora de 1.182pb e 393 aminoácidos.

Os genes Glyma.04g240800 e Glyma.06g122600 codificam a enzima álcool desidrogenase 1. O primeiro possui 2.999pb, que é transcrito em um RNAm de 1.143 bases e traduzido em uma proteína de 380 aminoácidos. A sequência codificadora na cultivar de referência Williams 82 está locada entre as posições 2.938 e 173, respectivamente entre os sinais de início e término de transcrição e foram identificados 10 éxons codificando a sequência proteica (Figura 16A). Na cultivar I27 foi recuperada uma sequência de 3005pb e não foram identificadas alterações não sinônimas, e sim apenas alterações de posicionamento. No entanto, na cultivar BRS 154, foi reconstruído apenas 59% do gene e alterações sinônimas foram identificadas, mas devido ao grande número de bases não sequenciadas, inferências sobre essas alterações não são precisas (Figura 16A).

A análise da sequência de 3.294pb de Glyma.06g122600 e seu produto proteico é mais dispendiosa, uma vez que este gene pode ser decodificado em quatro sequências de aminoácidos diferentes, com 381, 337, 349 e 373 aminoácidos, respectivamente, na sequência da proteína. O quadro de leitura na cultivar Williams 82 inicia com o sinal de transcrição, na posição 63, e termina com o sinal poliA, na posição 2.291 (Figura 16B). O sequenciamento e montagem recuperaram 3.305pb na cultivar BRS 154 e 3.306pb na cultivar I27. Na cultivar sensível, o quadro de leitura está contido entre as posições 63 e 3.002, para todas as predições realizadas e nenhuma alteração acarretou alteração na sequência de aminoácidos dessas proteínas (Figura 16B). Por sua vez, as sequências de proteínas preditas para a cultivar tolerante apresentaram uma redução no número de aminoácidos na sequência, pois um fragmento do éxons 9, 8, 9 e 8, respectivamente, não são traduzidos (Figura 16B).

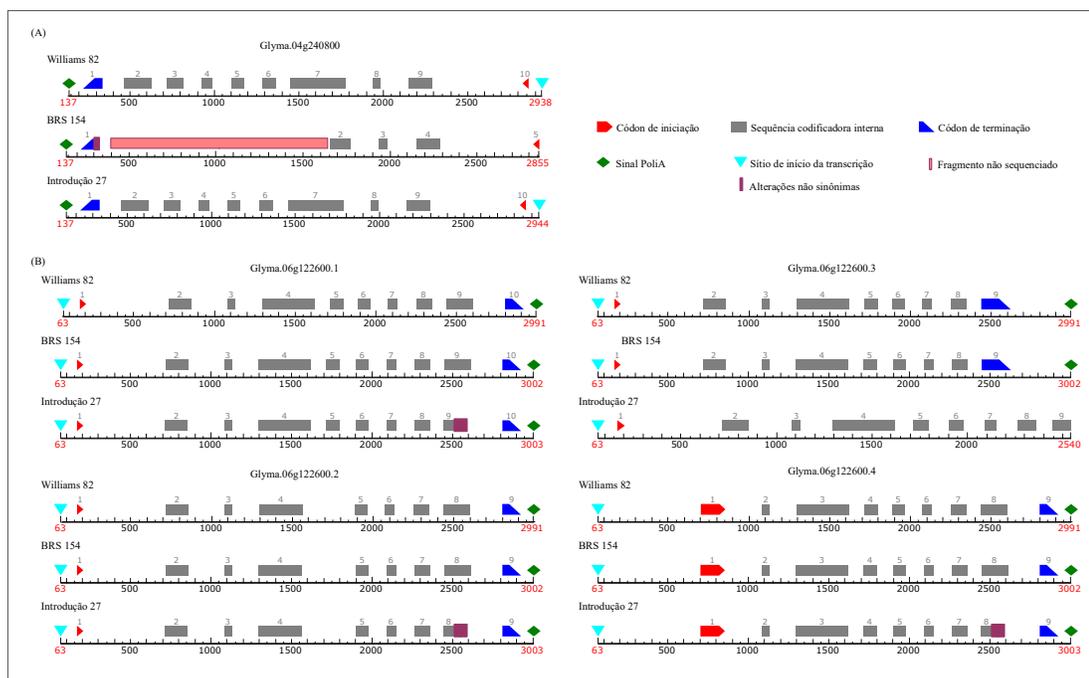


FIGURA 16. Posição dos éxons preditos dos genes Glyma.04g240800 (A) e Glyma.06g122600 (B), codificadores da enzima Álcool Desidrogenase 1, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.

A Alanina aminotransferase 2 é codificada pelos genes Glyma.01g026700 e Glyma.02g038100, com 7499pb e 7812pb de comprimento, respectivamente. O primeiro é composto por 13 éxons que codificam uma proteína com 481 aminoácidos. O sinal de início de transcrição está locado na posição 6.976 e o sinal poliA na posição 173 (Figura 17A). A análise de Glyma.01g026700 na Cultivar BRS 154 não identificou alterações na sequência de nucleotídeos que acarretam alteração de aminoácidos. No entanto, a sequência gênica não foi totalmente recuperada pelo sequenciamento, sendo possível a identificação da sequência codificadora até a posição 5.071. A cultivar I27 apresentou 454 aminoácidos, e alterações nos éxons 3 e 5, embora em regiões flanqueando fragmentos não sequenciados, o que dificulta inferências concretas (Figura 17A). Glyma.02g038100 também possui 13 éxons, codificando uma proteína similar de 481 aminoácidos. Para a cultivar I27 não foi identificada alteração na sequência codificadora, apenas na posição do sinal de início da transcrição e para a cultivar BRS 154 foram recuperados apenas 12 éxons, sendo que não foi recuperada a sequência do éxon faltante (Figura 17B).

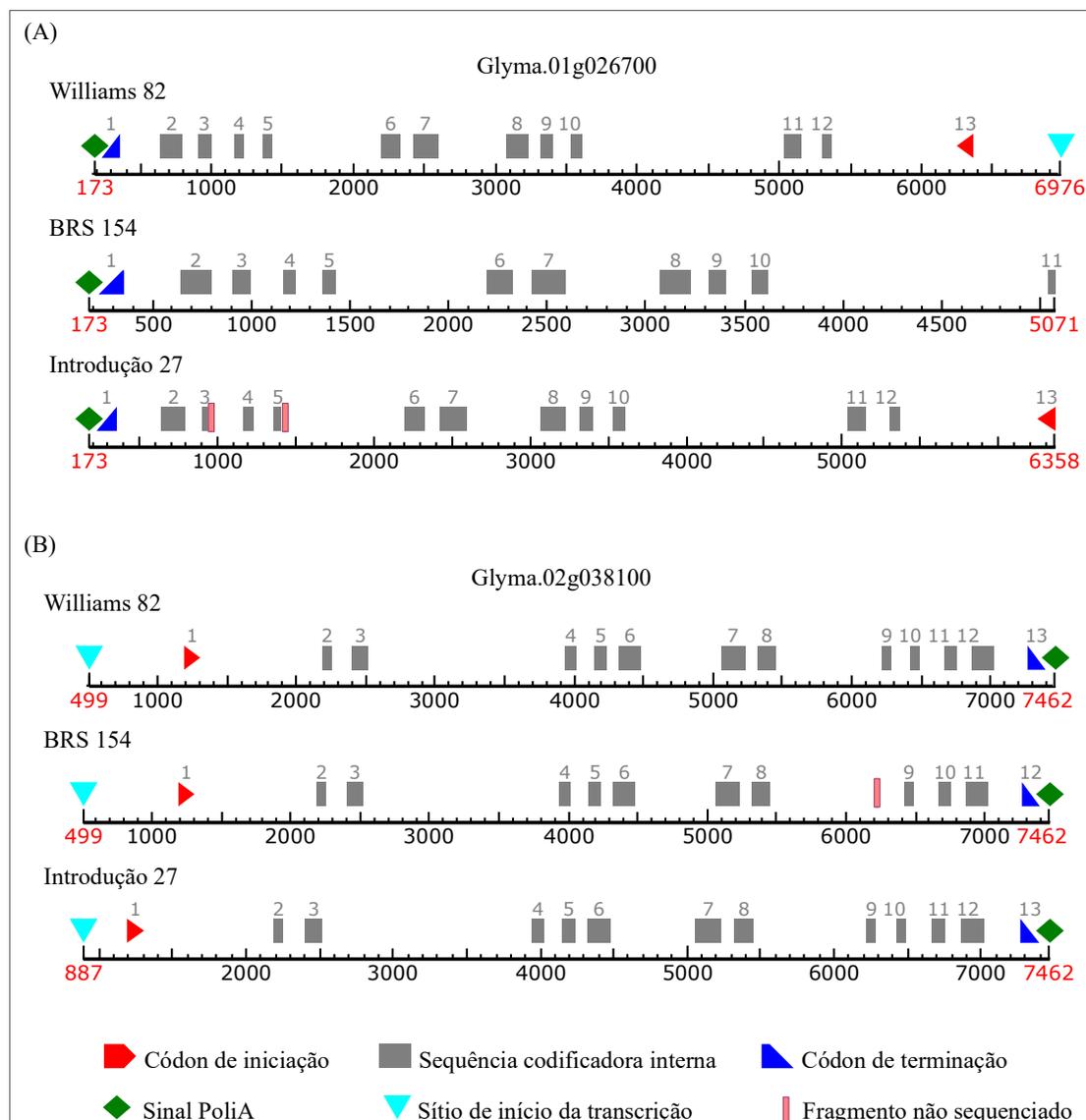


FIGURA 17. Posição dos éxons preditos dos genes Glyma.01g026700 (A) e Glyma.02g038100 (B), codificadores da enzima alanina aminotransferase 2, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.

O gene Hemoglobina 1 (Glyma.11g121800) apresenta 1466pb e é composto de quatro éxons, que são transcritos em uma molécula de RNAm de 486 bases, traduzidas em uma proteína de 161 aminoácidos. Os sinais de início de finalização da transcrição estão localizados nas posições 139 e 1.250, respectivamente (Figura 18). Este gene foi recuperado integralmente pelo sequenciamento e não apresentou diferença na sequência entre os genótipos avaliados e a referência Williams 82 (Figura 18). Sendo assim, a sequência da proteína por eles codificada apresenta 100% de identidade.

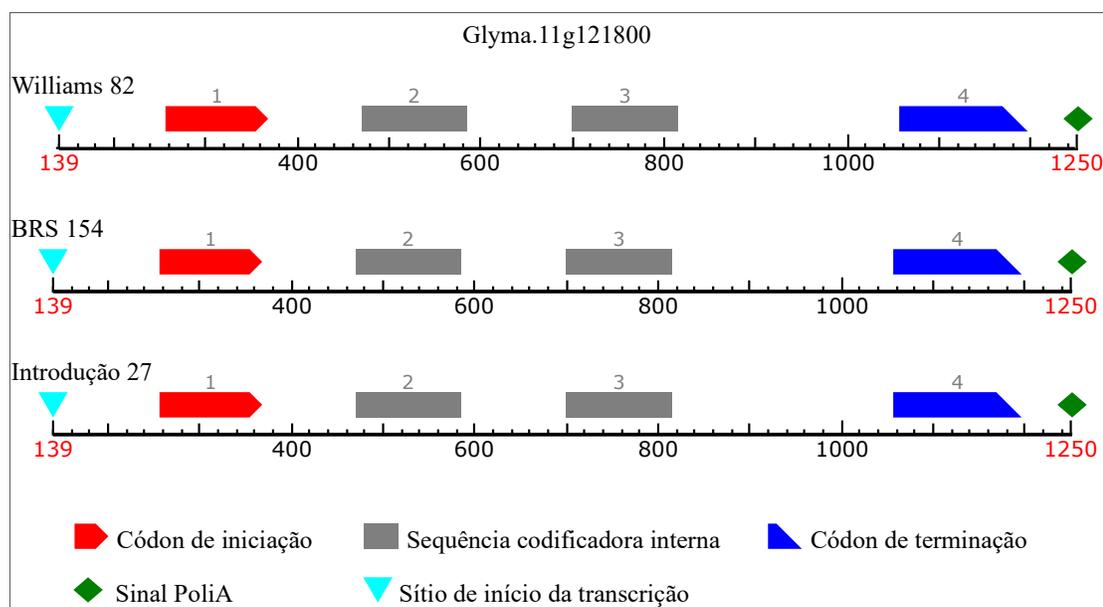


FIGURA 18. Posição dos éxons preditos do gene Glyma.11g121800, codificador da Hemoglobina 1, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.

O gene Glyma.08g365100, codificador da proteína LBD41, apresenta 1.870pb e dois éxons distribuídos entre as posições 21 e 1.506, respectivamente os pontos de sinalização de final e início da transcrição. Este gene foi parcialmente recuperado em ambas as cultivares sequenciadas. Desse modo, nas cultivares BRS 154 e I27 não foi possível identificar o fator de iniciação da transcrição e também o início do éxon dois, que contém o códon de iniciação (Figura 19A). A análise do parálogo, Glyma.18g297100, indicou que as sequências das cultivares BRS 154 e I27, embora não totalmente montadas, não apresentam diferenças nas regiões codificadoras (Figura 19B) e, dessa forma, nenhuma alteração na sequência de aminoácidos foi verificada. A sequência gênica de Glyma.18g297100 possui 1.449pb, e os sinais de transcrição e poliA, respectivamente, nas posições 80 e 1.286 (Figura 19B).

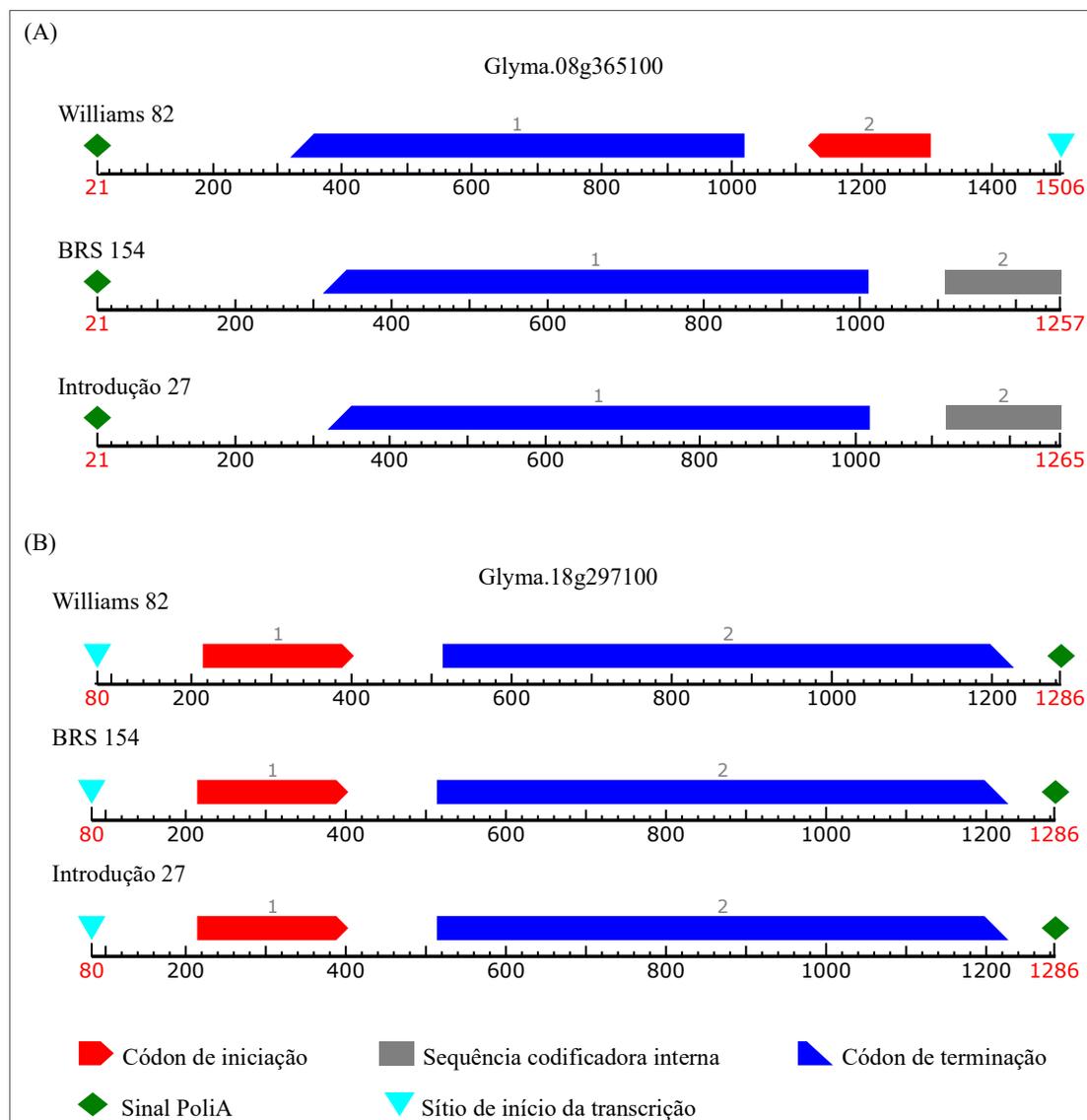


FIGURA 19. Posição dos éxons preditos dos genes Glyma.08g365100 (A) e Glyma.18g297100 (B), codificadores da proteína LBD41, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.

O gene *XETP* (Glyma.16g045000) possui 2.339pb de comprimento e 1.038pb transcritos, codificando uma proteína de 345 aminoácidos. O sequenciamento recuperou integralmente as sequências (2.340pb) e a análise da sequência identificou uma pequena alteração na posição dos sinais de transcrição, éxons e uma alteração de base, na sequência codificadora das cultivares BRS 154 e I27, que acarreta alteração na sequência da proteína em comparação com a cultivar Williams 82 (Figura 20A). Essa substituição de uma base, uma citosina por uma timina, na posição 2.174, muda o aminoácido de arginina para lisina.

Glyma.19g106700, por sua vez, possui 2.277pb e é transcrito e traduzido em proteína de 295 aminoácidos. Nenhuma alteração não sinônima foi detectada, apenas alterações que acarretam mudança na posição do éxon 4 e sinal poliA, em ambas as cultivares sequenciadas (Figura 20B).

O gene Glyma.12g073100, codificador da enzima Ascorbato Peroxidase 2, na cultivar de referência Williams 82, possui 2.711pb, e uma sequência codificadora de 9 éxons, localizados entre a posição 86 e 2.624, respectivamente entre o sinal poliA e o códon de iniciação de tradução. O gene *APX2* foi reconstituído totalmente pelo sequenciamento nas cultivares BRS 154 e I27 e a análise das sequências codificadoras indicou uma única alteração de base, na cultivar sensível em comparação a tolerante e a referência e que acarretou a alteração de um aminoácido na sequência proteica (Figura 21). Essa alteração está contida no éxon 8 e uma glutamina é substituída por uma prolina na sequência.

Embora o sequenciamento tenha recuperado de forma parcial ou total os genes candidatos, alterações na sequência gênica foram localizadas principalmente em regiões não codificantes. Essas alterações de maneira geral, acarretaram apenas a mudança na posição dos sinais de início de transcrição, códons de iniciação e terminação, éxons internos e sinal de término da transcrição ou sinal poliA. Desse modo, a interpretação dos resultados fica restrita e, nos casos em que a sequência não foi recuperada de forma integral, inferências são apenas especulativas.

Nos genes *XETP* e *APX2* foram identificadas alterações de uma única base que mudaram a sequência de aminoácidos nas proteínas. No primeiro gene ocorreu a substituição de uma arginina por uma lisina, em ambas as cultivares sequenciadas, em comparação com a referência. Esses dois aminoácidos possuem propriedades similares, suas moléculas são polares e básicas. Esse tipo de substituição é mais frequente e, em geral, não acarreta prejuízos (Lal & Verma, 2017). Por sua vez, a alteração em *APX2*, identificada no genótipo sensível BRS 154, foi uma substituição não conservativa, ou seja, as propriedades químicas

dos aminoácidos diferem.

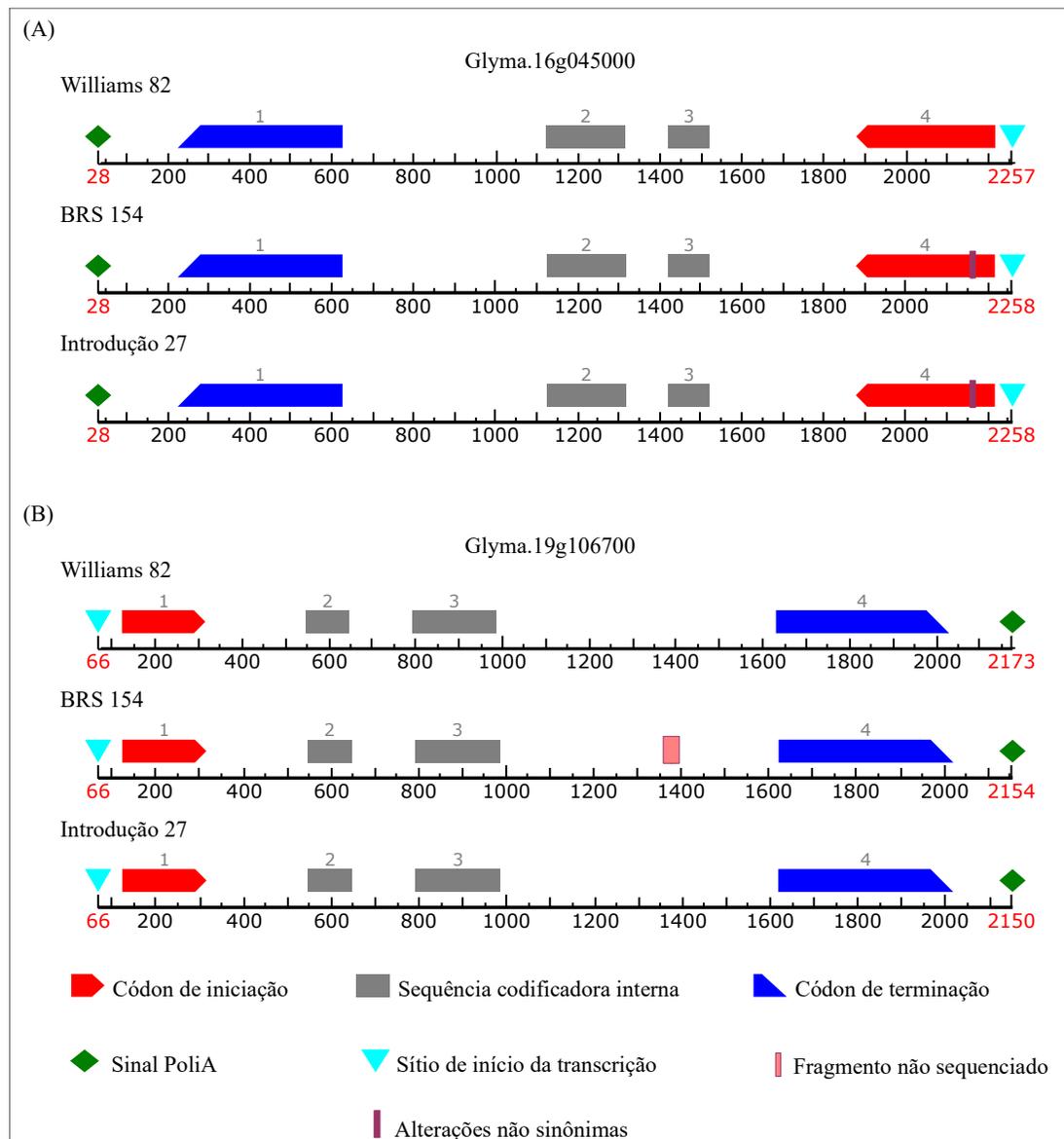


FIGURA 20. Posição dos éxons preditos dos genes Glyma.16g045000 (A) e Glyma.19g106700 (B), codificadores da proteína XETP, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.

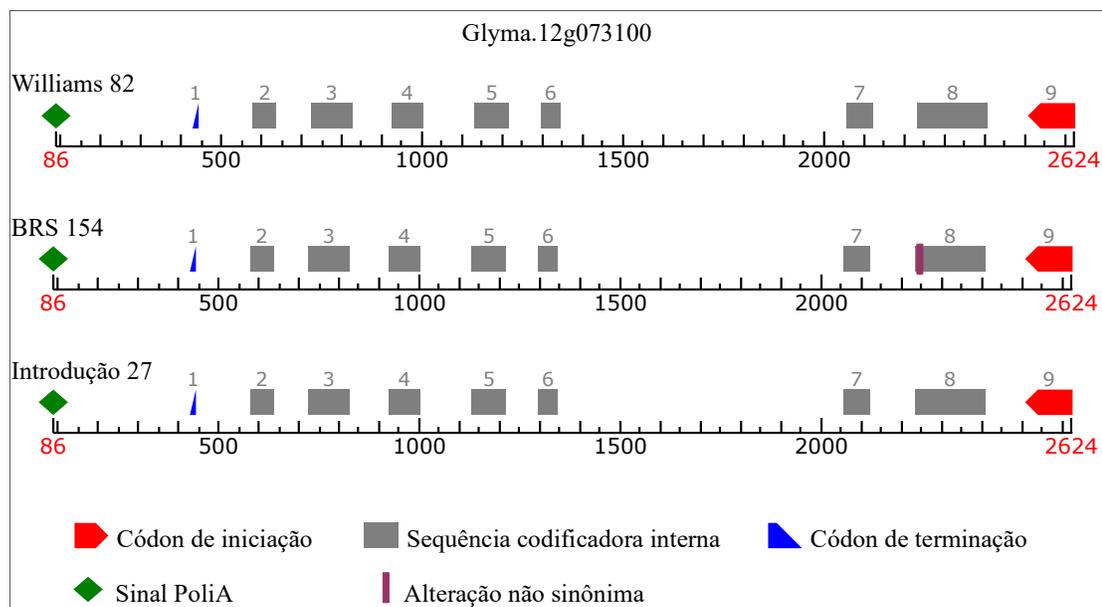


FIGURA 21. Posição dos éxons preditos do gene Glyma.12g073100, codificador da enzima Ascorbato Peroxidase 2, para as cultivares Williams 82, Introdução 27 e BRS154. UFRGS, Porto Alegre, RS, Brasil. 2017.

4.4 Análise da sequência promotora e elementos regulação em *cis*

Para a análise da região promotora foram utilizados 1Kb flanqueando o sítio de início de transcrição dos genes candidatos e por meio da ferramenta BLAST foi recuperada as regiões com identidade maior que 95% entre as cultivares em estudo e a mesma região na cultivar de referência. A região promotora de 1Kb de comprimento do gene codificador da enzima enolase Glyma.19g190900 não foi totalmente recuperada, sendo que a cobertura alcançada pelo sequenciamento foi de apenas 30% e 86% para os genótipos BRS 154 e I27, respectivamente. A similaridade da região promotora recuperada para BRS 154 foi de 97%, enquanto que para I27 foi de 100%, com a sequência de referência de W82. A análise de similaridade da sequência promotora do parálogo Glyma.03g190500, que também não foi totalmente remontada a partir dos dados do sequenciamento, evidenciou 99% de similaridade para ambas as cultivares em relação a cultivar de referência.

Da mesma forma, o sequenciamento recuperou apenas 40% e 59% da região promotora de *ADHI* (Glyma.04g240800) nos genótipos BRS154 e I27, com 94% e 99% de identidade das sequências com a mesma região na cultivar de referência, respectivamente.

A análise do promotor de Glyma.06g122600, recuperado totalmente em ambas as cultivares estudadas, indicou polimorfismos apenas entre a cultivar BRS 154 e W82, sendo que as sequências possuem 98% de identidade.

O gene Glyma.01g026700 possui 100% de cobertura e 99% de similaridade entre a cultivar W82 e ambas as cultivares em estudo. Já Glyma.02g038100 apresentou 100% de cobertura e similaridade entre cultivar BRS 154 e W82 e 78% de cobertura e 98% de identidade entre I27 e W82. A cobertura da região promotora de *GLBI* recuperada pelo sequenciamento foi de 100% e 96%, para os genótipos BRS 154 e I27, respectivamente, sendo que ambas as sequências possuem 99% de identidade com a sequência de referência.

Os genes Glyma.16g045000 e Glyma.19g106700 foram recuperados com cobertura e identidade com a região promotora de referência superior a 92%. As sequências reguladoras de Glyma.16g045000 para os genótipos BRS154 e I27 apresentaram 99% de identidade com a cultivar W82, enquanto que para Glyma.19g106700 foi de 96% e 99%, respectivamente. Na sequência promotora dos genes *LBD41* (Glyma.08g365100 e Glyma.18g297100) e *APX2* (Glyma.12g073100) não foram identificados polimorfismos entre as cultivares estudadas e a cultivar de referência.

A varredura das sequências promotoras dos genes candidatos ao alagamento do solo em soja no banco de dados de motivos *cis* de regulação PlantCARE possibilitou a identificação de elementos de regulação basal, específicos de resposta a estresses, luz, fitormônios e outros (Tabela 5). A base de bancos PlantCARE é específica de plantas e contém referências comprovadas experimentalmente da atuação desses elementos de regulação no controle da expressão gênica (Lescot *et al.*, 2002).

Entre os elementos de regulação encontrados nas sequências estudadas estão os basais CAAT-box e TATA-box, receptores do complexo de transcrição da RNA polimerase II e elementos de regulação associados ao metabolismo primário e/ou especializado como elemento responsivo à giberelina (motivo-GARE), elemento responsivo ao ácido abscísico

(ABRE), elemento responsivo ao etileno (ERE), elemento responsivo anaeróbico (ARE), elemento de choque térmico (HSE), elemento responsivo a baixa temperatura (LTR) (Tabela 5). Estes motivos, também foram identificados na região promotora de genes envolvidos na biossíntese e sinalização de etileno (*MnERF-B1*, *MnERF-B2* e *MnERF-B3*) em *Morus notabilis* e associados à resposta ao alagamento em *M. alba* (Shang *et al.*, 2014).

Elementos de regulação ARE foram identificados na região promotora de 9 dos genes estudados, exceto nos genes *ALAT2* (Glyma.01g026700), *XETP* (Glyma.19g106700) e *APX2* (Glyma.12g073100) (Tabela 5). Os elementos ARE, motivo-GT e motivo-GC, foram identificados e caracterizados na região promotora dos genes aldolase e *ADHI*, em milho (Walker *et al.*, 1987). Sequências similares foram identificadas, também em arábido (Liu *et al.*, 2005). Neste sentido, a presença dos elementos ARE na região promotora dos genes avaliados neste trabalho pode estar envolvida na resposta ao alagamento e hipóxia, uma vez que a expressão relativa dos mesmos foi alterada durante o estresse.

Elementos ERE foram identificados nos promotores dos genes *ADHI* (Glyma.06g122600) e *ALAT2* (Glyma.01g026700). Fatores de resposta ao etileno (ERF), também podem se ligar ao motivo-ATCTA (Hsu *et al.*, 2011; Gasch *et al.*, 2016). Motivo-ATCT foi identificado nos promotores dos genes *ENO* (Glyma.19g190900 e Glyma.03g190500), *ALAT2* (Glyma.01g026700), *LBD41* (Glyma.18g297100 e Glyma.08g365100), *XETP* (Glyma.16g045000) e *APX2* (Glyma.12g073100). Fujimoto *et al.* (2000) identificaram fatores com a capacidade de ligação com o motivo GCC-box, associado com a expressão gênica induzida por etileno e patógenos em arábido. O motivo GCC, similar ao G-box, é um potencial sítio de ligação de ERF (Gasch *et al.*, 2016).

TABELA 5. Elementos de regulação *cis* localizados na região promotora dos genes candidatos a tolerância ao alagamento do solo na cultivar de referência Williams 82. UFRGS, Porto Alegre, RS, Brasil. 2017.

Gene ¹	Locus gênico	Elementos de regulação
<i>ENO</i>	Glyma.19g190900	motivo-AAGAA; ACE; ARE ; elemento rico-AT; motivo-ATCT ; Box 4; Box I; CAAT-box; motivo-CGTCA; G-box ; HSE ; MBS ; MRE ; motivo-Skn-1; Sp1; TATA-box; motivo-TATCCAT/C; motivo-TGACG; Desconhecido 4; motivo-WUN; como-2-box; circadiano;
	Glyma.03g190500	ABRE ; ARE ; motivo-AT1; motivo-ATCT ; Box 4; Box I; Box III; CAAT-box; G-box ; motivo-GA; motivo-GARE; HSE ; I-box ; MRE ; TATA-box; repetições ricas-TC; elemento-TCA; Desconhecido 4; Desconhecido 6; como-2-box; circadiano;
<i>ADHI</i>	Glyma.04g240800	Motivo-AAGAA; ABRE ; ACE; ARE ; motivo-AT1; motivo-ATGCAAAT; Box 4; CAAT-box; CAT-box; G-box ; motivo-GATA; motivo-GT1; HSE ; LTR; motivo-Skn-1; TATA-box; repetições ricas-TC; elemento-TCA; motivo-TCT; Desconhecido 1; Desconhecido 3;
	Glyma.06g122600	5'UTR rico-Py; motivo-AAGAA; ABRE ; ARE ; sequência rica-AT; motivo-ATGCAAAT; Box 4; Box I; Box III; CAAT-box; CAT-box; CCAAT-box; ERE ; G-box ; motivo-GA; motivo-GCN4; motivo-GT1; MBS ; motivo-Skn-1; TATA-box; repetições ricas-TC; elemento-TCA; elemento-TGA; Desconhecido 1; Desconhecido 3; Desconhecido 4; circadiano;
<i>ALAT2</i>	Glyma.01g026700	5'UTR rico-Py; ACE; AE-box; motivo-ATCT ; motivo-ATGCAAAT; Box 4; Box I; Box III; CAAT-box; motivo-CGTCA; ERE ; motivo-GCN4; HSE ; I-box ; elemento-LAMP; MBS ; P-box; motivo-Skn-1; TATA-box; repetições ricas-TC; motivo-TGACG; Desconhecido 1; Desconhecido 3; Desconhecido 4; Desconhecido 6; circadiano; dOCT;
	Glyma.02g038100	5'UTR rico-Py; motivo-AAGAA; ABRE ; ACE; AE-box; ARE ; Box-W1; CAAT-box; CAT-box; motivo-CGTCA; G-box ; motivo-GA; I-box ; sítio-O ₂ ; motivo-Skn-1; TATA-box; motivo-TATCCAT/C; repetições ricas-TC; elemento-TCA; motivo-TGACG; Desconhecido 1; Desconhecido 3; Desconhecido 4; W-box ; box II; chs-CMA1a; circadiano;
<i>GLB1</i>	Glyma.11g121800	Motivo-AAGAA; ACE; AE-box; ARE ; Box-W1; CAAT-box; motivo-CGTCA; motivo-GCN4; elemento-LAMP; MBS ; MNF1; motivo-Skn-1; Sp1; TATA-box; repetições ricas-TC; elemento-TCA; motivo-TCCC; motivo-TGACG; Desconhecido 4; W-box ; circadiano;
<i>LBD41</i>	Glyma.18g297100	Sítio ligado 3-AF1; 5'UTR rico-Py; ABRE ; AC-II; ACE; ARE ; motivo-ATCT ; Box 4; CAAT-box; CAT-box; motivo-CGTCA; EIRE; G-box ; elemento-LAMP; LTR; MBS ; motivo-Skn-1; TATA-box; repetições ricas-TC; elemento-TCA; motivo-TGACG; Desconhecido 1; Desconhecido 2; Desconhecido 3; Desconhecido 4; circadiano;
	Glyma.08g365100	Motivo-AAGAA; AC-II; ARE ; sequência rica-AT; motivo-ATCT ; CAAT-box; motivo-CATT; EIRE; G-box ; motivo-GAG; motivo-GT1; I-box ; elemento-LAMP; LTR; MBS ; sítio O ₂ ; TATA-box; elemento-TCA; elemento-TGA; Desconhecido 1; Desconhecido 3; Desconhecido 4; Desconhecido 6;
<i>XETP</i>	Glyma.16g045000	Motivo-AAGAA; AE-box; ARE ; motivo-ATCT ; motivo-ATGCAAAT; Box 4; Box III; CAAT-box; motivo-GA; motivo-GATA; motivo-GT1; I-box ; TATA-box; motivo-TATCCAT/C; repetições ricas-TC; Desconhecido 13; Desconhecido 4; Desconhecido 8;
	Glyma.19g106700	Motivo-AAGAA; Box 4; CAAT-box; motivo-GA; motivo-GT1; elemento-LAMP; MRE ; motivo-Skn-1; TATA-box; repetições ricas-TC; Desconhecido 4; Desconhecido 8; chs-CMA2b;
<i>APX2</i>	Glyma.12g073100	5'UTR rico-Py; motivo-AAAC; motivo-AAGAA; ABRE ; AC-I; AC-II; AE-box; motivo-AT1; motivo-ATCT ; núcleo-AuxRR; Box 4; CAAT-box; CCAAT-box; motivo-CGTCA; G-box ; motivo-GAG; motivo-GARE; motivo-GCN4; Gap-box; HSE ; L-box; MRE ; como-MSA; sítio O ₂ ; motivo-Skn-1; Sp1; TATA-box; TATC-box; repetições ricas-TC; motivo-TGACG; Desconhecido 4; chs-Unit 1 m1; circadiano;

¹ *ENO*: enolase; *ADHI*: álcool desidrogenase 1; *ALAT2*: alanina aminotransferase 2; *GLB1*: hemoglobina 1; *LBD41*: proteína contendo domínio LOB 41; *XETP*: xiloglucano endotransglicosilase parcial; *APX2*: ascorbato peroxidase 2.*Elementos em negrito foram correlacionados com a resposta ao alagamento e/ou hipóxia.

As regiões promotoras de genes relacionados ao alagamento do solo também foram enriquecidas em motivos ABRE, G-box, W-box, MRE, MBS e I-box (Tabela 5). Os

elementos ABRE e G-box são locais de ligação de fatores de transcrição da família *bZIP*, W-box de fatores *WRKY*, MRE e MBS de fatores *MYB* que regulam também respostas sob estresses (Sheshadri *et al.*, 2016). A sequência correspondente ao motivo I-box atua na regulação dependente de hipóxia via fatores de transcrição *bZIP* (GASCH *et al.*, 2016). Além disso, foi relatada a associação entre fatores *WRKY* e as alterações metabólicas que permitem a sobrevivência do mutante de arroz para o gene *ADHI(rad)* (Mohanty *et al.*, 2016). Os autores evidenciaram que a expressão de *WRKY* foi induzida e que 61% dos genes ativado pelo estresse apresentavam motivo W-box na sequência promotora (Mohanty *et al.*, 2016).

A análise de elementos *cis* para identificação de genes responsivos ao estresse em soja identificou motivos ABRE, MYB ou MYC, sugerindo a atuação de fatores de transcrição do tipo *bZIP*, *MYB* e *bHLH* que se ligam a estes motivos (Mochida *et al.*, 2010). Em arroz, os promotores de genes com a expressão ativada ou reduzida no mutante (*rad*) foram enriquecidos com elementos *cis* que são associados a fatores de transcrição das famílias *MYB*, *bZIP*, *ERF* and *ARF* (fator de resposta a auxina) (Mohanty *et al.*, 2016). Em concordância, os fatores de transcrição *MYB* e *ERF* foram induzidos. Esses resultados, associados aos encontrados neste trabalho evidenciam a participação destes elementos reguladores na resposta ao estresse por alagamento.

A de privação de oxigênio induz a expressão de inúmeros genes, incluindo codificadores de proteínas e fatores de transcrição de choque térmico. Os genes *ENO* (Glyma.19g190900 e Glyma.03g190500), *ADHI* (Glyma.04g240800) e *APX2* apresentaram elementos *cis* HSE. Em *Arabidopsis*, o fator de choque térmico A2 (*HsfA2*), associado com aclimatação ao calor e outros estresses abióticos, foi induzido por anóxia e sua superexpressão acarretou melhoria na tolerância a anóxia e submersão (Banti *et al.*, 2010). A presença destes elementos, pode estar associada a manutenção da produção de energia e detoxificação celular em ambos os casos (calor e hipóxia).

Os genes *ADHI* e *PDCI* foram induzidos pela submergência em *Arabidopsis* e em plantas transgênicas expressando *OsSnRK1* (proteína quinase 1 relacionada a não fermentação da sacarose 1) de arroz (Cho *et al.*, 2012). A análise da sequência do promotor *PDCI* revelou um elemento *cis* altamente conservado (ACTGAT) para a ligação de fatores de transcrição da família *bZIP*, consistente com a localização de ligação de SnRK1 (Cho *et al.*, 2012). Em genes de soja diferencialmente expressos sob condição de déficit hídrico e alagamento foram identificados 24 genes apresentando motivo-ACGT (Chen *et al.*, 2016). Embora nenhum destes motivos tenha sido identificado nas sequências promotoras em estudo, esses trabalhos destacam a importância de compreender a regulação da resposta ao estresse.

A regulação de genes sob estresse pode ocorrer em nível transcricional, pela ligação de fatores de transcrição aos elementos *cis* específicos presentes nas regiões flangeadoras 5' destes genes. Qualquer alteração que ocorra nesta região, como polimorfismos de nucleotídeo único e *indels*, alteram também a resposta ao estresse. A comparação das sequências promotoras das cultivares I27, BRS 154 e a cultivar de referência W82 evidenciou poucas alterações em motivos de regulação *cis* relacionados a resposta ao alagamento. Além disso, algumas regiões não foram completamente recuperadas pelo sequenciamento.

O gene *ALAT2* (Glyma.02g038100) na cultivar I27 apresentou um *SNP* coincidente com a posição do elemento W-box. No gene Glyma.06g122600, uma falha de 20pb impossibilitou a recuperação do motivo MBS na sequência promotora da cultivar BRS 154. A identificação de elementos *cis* na região promotora fornece informações sobre a regulação transcricional desencadeada em resposta ao alagamento e privação de oxigênio. A identificação dos elementos *cis* responsivos ao alagamento pode não ser capaz de explicar como os genes são regulados durante o estresse. No entanto, são informativos para estudos adicionais visando reconstruir a via regulatória.

5 CONCLUSÕES

O genoma das cultivares Introdução 27 e BRS 154 foi sequenciado e montado. As sequências possuem 1,024Gb e 1,020Gb de tamanho, respectivamente.

A expressão dos genes *ENO*, *ADHI*, *ALAT2*, *GLB1*, *LBD41*, *XETP* e *APX2*, relacionados com a tolerância ao alagamento do solo, de maneira geral, aumentou em maior proporção no tecido radicular em comparação com a parte aérea. A cultivar Introdução 27 apresentou maior expressão relativa sob condição de alagamento destes genes.

Foram identificadas alterações não sinônimas na sequência dos genes *XETP* e *APX2*. No entanto, apenas no gene *APX2* a diferença foi entre os genótipos sensível e tolerante de soja e acarretou a alteração de um aminoácido na sequência proteica.

Foram identificados elementos de regulação *cis*, ARE, ERE, motivo-ATCT, G-box, ABRE, W-box, MRE, MBS, I-box, HSE, associados à resposta a estresses. Entre as cultivares estudadas, apenas o gene *ALAT2* (Glyma.02g038100) na cultivar I27 apresentou um *SNP* coincidente com a posição do elemento W-box.

6 REFERÊNCIAS BIBLIOGRÁFICAS

- ALAM, I. *et al.* Proteome analysis of soybean roots under waterlogging stress at an early vegetative stage. **Journal of Biosciences**, New Delhi, v.35, n.1, p.49–62, 2010.
- ALTSCHUL, S. F. *et al.* Basic local alignment search tool. **Journal of Molecular Biology**, Amsterdam, v.215, n.3, p.403-410, 1990.
- ANDERSON, J. E. *et al.* A roadmap for functional structural variants in the soybean genome. **G3: Genes, genomes and genetics**, Bethesda, v.4, p.1307-1318, 2014.
- ANTÓNIO, C. *et al.* Regulation of primary metabolism in response to low oxygen availability as revealed by carbon and nitrogen isotope redistribution. **Plant Physiology**, Rockville, v.170, n.1, p.43-56, 2016.
- APEL K.; HIRT, H. Reactive oxygen species: metabolism, oxidative stress and signal transduction. **Annual Review of Plant Biology**, Palo Alto, v.55, n.1, p.373–99, 2004.
- ARABIDOPSIS GENOME INITIATIVE. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. **Nature**, Basingstoke, v.408, p.796–815, 2000.
- BAILEY-SERRES, J. CHANG, R. Sensing and signalling in response to oxygen deprivation in plants and other organisms. **Annals of Botany**, Oxford, v.96, n.4, p.507–518, 2005.
- BAILEY-SERRES, J.; VOESENEK, L. A. C. J. Flooding stress: acclimations and genetic diversity. **Annual Review of Plant Biology**, Palo Alto, v.59, n.1, p.313-339, 2008.
- BANTI, V. *et al.* The heat-inducible transcription factor HsfA2 enhances anoxia tolerance in *Arabidopsis*. **Plant Physiology**, Rockville, v.152, n.3, p.1471–1483, 2010.
- BARRETT-LENNARD E. G. The interaction between waterlogging and salinity in higher plants: causes, consequences and implications. **Plant and Soil**, The Hague, v.253, n.1, p.35–54, 2003.
- BENTLEY, D. R. Whole-genome re-sequencing. **Current Opinion in Genetics & Development**, London, v.16, p.545-552, 2006.
- BISPO, N. B. **Seleção de genótipos e análise da tolerância do milho (*Zea mays* L.) ao encharcamento do solo**. 2011. 64 f. Tese (Doutorado) - Programa de Pós-Graduação em Fitotecnia, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

- BOETZER, M. *et al.* Scaffolding pre-assembled contigs using SSPACE. **Bioinformatics**, Oxford, v.27, n.4, p.578-579, 2011.
- BONATO, A. L. V. *et al.* Genetic similarity among soybean (*Glycine max* (L) Merrill) cultivars released in Brazil using AFLP markers. **Genetics and Molecular Biology**, Ribeirão Preto, v.29, n.1, p.692-704, 2006.
- BYFIELD, G. E. *et al.* Two genes from soybean encoding soluble D9 stearoyl-ACP desaturases. **Crop Science**, New York, v.46, p.840-846, 2006.
- CHEN, X. *et al.* Endogenous abscisic acid as a key switch for natural variation in flooding-induced shoot elongation. **Plant Physiology**, Rockville, v.154, n.2, p.969-977, 2010.
- CHEN, W. *et al.* Identification and Comparative Analysis of Differential Gene Expression in Soybean Leaf Tissue under Drought and Flooding Stress Revealed by RNA-Seq. **Frontiers in Plant Science**, Lausanne, v.7, p.1-19, 2016.
- CHO, Y. H. *et al.* Regulatory functions of SnRK1 in stress-responsive gene expression and in plant growth and development. **Plant Physiology**, Rockville, v.158, n.4, p.1955-1964, 2012.
- COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da safra brasileira:** Grãos. Brasília, 2017. 162p. Disponível em:http://www.conab.gov.br/OlalaCMS/uploads/arquivos/17_01_11_11_30_39_boletim_graos_janeiro_2017.pdf Acesso em: 05 fev. 2017.
- CONNELL, C. *et al.* Automated DNA-sequence analysis. **Biotechniques**, London, v.5, p.342-348, 1987.
- DAT J. *et al.* Sensing and signaling during plant flooding. **Plant Physiology and Biochemistry**, Amsterdam, v.42, n.4, p.273-282, 2004.
- DELSENY, M; HSING, Y. I. High throughput DNA sequencing: the new sequencing revolution. **Plant Science**, Shannon, v.179, n.5, p.407-422, 2010.
- DEPREZ, R. H. L. *et al.* Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. **Analytical Biochemistry**, San Diego, v.307, p. 63-69, 2002.
- DROS, J. M. **Administrando os avanços da soja:** dois cenários de expansão do cultivo de soja na América do Sul. Amsterdã: AIDEnvironment, 2004. 71p.
- DU, H. *et al.* Overexpression of *Vitreoscilla* hemoglobin increases waterlogging tolerance in *Arabidopsis* and maize. **BMC Plant Biology**, London, v.16, n.1, p.16-35, 2016.
- DUSSAULT, A. A.; POULIOT, M. Rapid and simple comparison of messenger RNA levels using real-time PCR. **Biological Procedures Online**, London, v. 8, n. 1, p. 1-10, 2006.
- FANTE, C. A. *et al.* Respostas fisiológicas em cultivares de soja submetidas ao alagamento em diferentes estádios. **Bragantia**, Campinas, v.69, n.2, p.253-261, 2010.

FARIAS, D. R. **Desenvolvimento de ferramentas de bioinformática para montagem e prospecção de genomas selvagens visando o melhoramento de arroz**. 2013. 82 f. Tese (Doutorado) - Faculdade de Agronomia Eliseu Maciel da Universidade Federal de Pelotas, Universidade Federal de Pelotas, Pelotas, 2013.

FUJIMOTO, S. Y. *et al.* Arabidopsis ethylene-responsive element binding factors act as transcriptional activators or repressors of GCC box-mediated gene expression. **Plant Cell**, Rockville, v.12, n.3, p.393-340, 2000.

GASCH, P. *et al.* Redundant ERF-VII transcription factors bind to an evolutionarily conserved *cis*-motif to regulate hypoxia-responsive gene expression in Arabidopsis. **Plant Cell**, Rockville, v.28, n.1, p.160-180, 2016.

GUO, J. *et al.* Single origin and moderate bottleneck during domestication of soybean (*Glycine max*): Implications from microsatellites and nucleotide sequences. **Annals of Botany**, Oxford, v.106, n.1, p.505-514, 2010.

HAMILTON, J. P.; BUELL, R. Advances in plant genome sequencing. **The Plant Journal**, Oxford, v.70, n.1, p.77-190, 2012.

HILL, R. *et al.* Phytoglobin: a novel nomenclature for plant globins accepted by the globin community at the 2014 XVIII conference on Oxygen-Binding and Sensing Proteins [version 1; referees: 2 approved]. **F1000 Research**, London, v.5, n.212, p.1-8, 2016.

HOSSAIN, Z. *et al.*, 2013 Soybean proteomics for unraveling abiotic stress response mechanism. **Journal of Proteome Research**, Washington, v.12, n.1, p.4670-4684, 2013.

HSU, F. C. *et al.* Insights into hypoxic systemic responses based on analyses of transcriptional regulation in arabidopsis. **PLoS ONE**, San Francisco, v.6, n.12, p.1-19, 2011.

HUANG, X., *et al.* Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. **Nature Genetics**, New York, v.44, n.1, p.32-39, 2011.

HUYNH N. *et al.* Regulation of flooding tolerance of SAG12:ipt Arabidopsis plants by cytokinin. **Journal of Experimental Botany**, Oxford, v.56, n.415, p.1397-1407, 2005.

HYMOWITZ, T. On the domestication of the soybean. **Economic Botany**, Bronx, v.24, n.1, p.408-421, 1970.

HYTEN, D. L. *et al.* Impacts of genetic bottlenecks on soybean genome diversity. **Proceedings of the National Academy of Sciences of United States of America**, Washington, v.103, n.1, p.16666-16671, 2006.

IGAMBERDIEV, A. U. *et al.* Structural and functional properties of class 1 plant hemoglobins. **IUBMB Life**, London, v.63, n.3, p.146-15, 2011.

IGAMBERDIEV, A. U.; HILL, R. D. Nitrate, NO and haemoglobin in plant adaptation to hypoxia: an alternative to classic fermentation pathways. **Journal of Experimental Botany**, Oxford, v.55, n.408, p.2473-2482, 2004.

INSTITUTO RIO GRANDENSE DO ARROZ. **Soja em rotação com arroz**: evolução de área e produtividade. 2016. Disponível em: <<http://www.irga.rs.gov.br>> Acesso em: 05 fev. 2017.

INTERNATIONAL RICE GENOME SEQUENCING PROJECT. The map-based sequence of the rice genome. **Nature**, Basingstoke, v.436, p.793–800, 2005.

IRFAN, M. *et al.* Physiological and biochemical changes in plants under waterlogging. **Protoplasma**, New York, v.241, n.1, p.3–17, 2010.

ISMAIL, A. M. *et al.* Mechanisms associated with tolerance to flooding during germination and early seedling growth in rice (*Oryza sativa*). **Annals of Botany**, Oxford, v.103, n.2, p.197–209, 2009.

ISMOND, K. P. *et al.* Enhanced low oxygen survival in arabidopsis through increased metabolic flux in the fermentative pathway. **Plant Physiology**, Rockville, v.132, p. 1292–1302, 2003.

JIAN, B. *et al.* Validation of internal control for gene expression study in soybean by quantitative real-time PCR. **BMC Molecular Biology**, London, v.9 n.59, p.1-14, 2008.

KAUSAR, R. *et al.* Characterization of ascorbate peroxidase in soybean under flooding and drought stresses. **Molecular Biology Reports**, Dordrecht, v.39, p. 10573–10579, 2012.

KHATOON, A. *et al.* Analysis of response mechanism in soybean under low oxygen and flooding stresses using gel-base proteomics technique. **Molecular Biology Reports**, Dordrecht, v.39, n.12, p.10581–10594, 2012.

KIM M. J. *et al.* Tracing soybean domestication history: from nucleotide to genome. **Breeding Science**, Tokyo, v.61, n.1, p. 445–452, 2012.

KOMATSU, S. *et al.* A comprehensive analysis of the soybean genes and proteins expressed under flooding stress using transcriptome and proteome techniques. **Journal of Proteome research**, Washington, v.8, n.1, p.4766-4778, 2009.

KOMATSU S. *et al.* Comparative proteomics analysis of differentially expressed proteins in soybean cell wall during flooding stress. **Amino Acids**, Wien, v.39, n.5, p.1435–1449, 2010.

KOMATSU, S. *et al.* “Omics” techniques for identifying flooding-response mechanisms in soybean. **Journal of Proteomics**, Amsterdam, v.93, n.1, p.169–178, 2013.

LAL, D.; VERMA, M. Large-Scale Sequence Comparison. In: KEITH J. M. (Ed.). **Bioinformatics: Volume I: Data, Sequence Analysis, and Evolution**. New York: Springer Science, 2017. (Methods in Molecular Biology, v.1525)

LAM, H. M. *et al.* Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. **Nature Genetics**, New York, v.42, n.12, p.1053-1059, 2010.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, New York, v.9, n.4, p.357-359, 2012.

LEFORT, F.; DOUGLAS, G. C. An efficient micro-method of DNA isolation from mature leaves of four hardwood tree species *Acer*, *Fraxinus*, *Prunus* and *Quercus*. **Annals of Forest Science**, Paris, v.56, n.3, p.259-263, 1999.

LESCOT, M. *et al.* PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. **Nucleic Acids Research**, Oxford, v.30, n.1, p.325-327, 2002.

LI, R. *et al.* De novo assembly of human genomes with massively parallel short read sequencing. **Genome Research**, Cold Spring Harbor, v.20, p.265-272, 2010.

LI, Y. H. *et al.* De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. **Nature Biotechnology**, New York, v.32, n.10, 2014.

LIU, F. *et al.* Global transcription profiling reveals comprehensive insights into hypoxic response in arabidopsis. **Plant Physiology**, Rockville, v.137, n.3, p.1115-1129, 2005.

LODHI, M. A. *et al.* A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. **Plant Molecular Biology Reporter**, New York, v.12, n.1, p.6-13, 1994.

MAGNESCHI, L.; PERATA, P. Rice germination and seedling growth in the absence of oxygen. **Annals of Botany**, Oxford, v.103, n.2, p.181-196, 2009.

MAK, D. O. D. *et al.* Characterization of ammonia transport by the kidney Rh glycoproteins RhBG and RhCG. **American Journal of Physiology-Renal Physiology**, Bethesda, v.290, n.2, p.297-305, 2006.

MARDIS E. The impact of next-generation sequencing technology on genetics. **Trends in Genetics**, Cambridge, v.24, n.3, p.133-141, 2008.

MASTRANGELO A.M, *et al.* Improvement of Drought Resistance in Crops: From Conventional Breeding to Genomic Selection. In: VENKATESWARLU, B. *et al.* (Ed.). **Crop Stress and its Management: Perspectives and Strategies**. New York: Springer, 2012. p.225–259.

MINOCHE, A. E. *et al.* Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. **Genome Biology**, London, v.12, p.1-15, 2011.

MISSIO, E. L. *et al.* Desempenho de genótipos de soja em solo hidromórfico de várzea. **Pesquisa Agropecuária Gaúcha**, Porto Alegre, v.16, n.1, p.23-29, 2010.

MOCHIDA, K. *et al.* Genome-wide analysis of two-component systems and prediction of stress-responsive two component system members in soybean. **DNA Research**, Oxford, v.17, n.5, p.303–324, 2010.

MOHANTY, B. *et al.* Transcriptional regulatory mechanism of alcohol dehydrogenase 1-deficient mutant of rice for cell survival under complete submergence. **Rice**, New York, v.9, n.51, p.1-8, 2016.

NAKAYAMA, T. J. *et al.* Reference genes for quantitative real-time polymerase chain reaction studies in soybean plants under hypoxic conditions. **Genetics and Molecular Research**, Ribeirão Preto, v.13, n.1, p.860-871, 2014.

NANJO Y. *et al.* Analyses of flooding tolerance of soybean varieties at emergence and varietal differences in their proteomes. **Phytochemistry**, London, v.106, n.1, p.25-36, 2014.

NANJO, Y. *et al.* Transcriptional responses to flooding stress in roots including hypocotyl of soybean seedlings. **Plant Molecular Biology**, Dordrecht, v.77, n.1, p.129–144, 2011.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION. **Genome**. [Base de Dados]. [2017]. (Soybean) Disponível em: <<https://www.ncbi.nlm.nih.gov/genome/?term=soybean>>. Acesso em: 05 fev. 2017.

OKUSHIMA, Y. *et al.* ARF7 and ARF19 regulate lateral root formation via direct activation of LBD/ASL genes in Arabidopsis. **Plant Cell**, Rockville, v.19, n.1 p.118-130, 2007.

PASZIEWICK, K; STUDHOLME, D. J. De novo assembly of short sequence reads. **Briefings in Bioinformatics**, London, v.11, n.5, p. 457-472, 2010.

PINTO, L. F. S. *et al.* Caracterização dos solos de várzea. In: GOMES, A. D. S.; PAULETTO, E. A. (Ed.). **Manejo de solo e da água em áreas de várzea**. Pelotas: Embrapa Clima Temperado, 1999. p.11-36.

PIRES, J. L. F. *et al.* Adaptações morfofisiológicas da soja em solo inundado. **Pesquisa Agropecuária Brasileira**, Brasília, v.37, n.1, p.41-50, 2002.

PRIOLLI, R. H. G. *et al.* Characterization of Brazilian soybean cultivars using microsatellite markers. **Genetics and Molecular Biology**, Ribeirão Preto, v.25, n.1, p.185-193, 2002.

RHINE M. D. *et al.* Yield and nutritional responses to waterlogging of soybean cultivars. **Irrigation Science**, New York, v.28, n.1, p.135-142, 2010.

RIQUELME, A.; HINRICHSEN, P. Non-symbiotic hemoglobin and its relation with hypoxic stress. **Chilean Journal of Agricultural Research**, Santiago de Chile, v.75, n.1, p.80-89, 2015.

ROCHA, M. *et al.* Glycolysis and the tricarboxylic acid cycle are linked by alanine aminotransferase during hypoxia induced by waterlogging of *Lotus japonicus*. **Plant Physiology**, Rockville, v.152, p. 1501-1513, 2010.

SACHS, M. M. *et al.* Anaerobic gene expression and flooding tolerance in maize. **Journal of Experimental Botany**, Oxford, v.47, n.294, p. 1-15, 1996.

SALZBERG, S.L. *et al.* GAGE: A critical evaluation of genome assemblies and assembly algorithms. **Genome Research**, Cold Spring Harbor, v.22, n.3, p.557-567, 2012.

SASIDHARAN R. *et al.* Root transcript profiling of two *Rorippa* species reveals gene clusters associated with extreme submergence tolerance. **Plant Physiology**, Rockville, v.163, n.3, p.1277-1292, 2013.

- SAYAMA, T. *et al.* QTL analysis of seed-flooding tolerance in soybean (*Glycine max* [L.] Merr.). **Plant Science**, Shannon, v.176, p.514–521, 2009.
- SCHMUTZ, J. *et al.* Genome sequence of the palaeopolyploid soybean. **Nature**, Basingstoke, v.463, p.178-183, 2010.
- SEDIVY, E. J. *et al.* Soybean Domestication: The Origin, Genetic Architecture and Molecular Bases. **New Phytologist**, Oxford, n.1, p.1-15, 2017.
- SHANG, J. *et al.* Identification of the mulberry genes involved in ethylene biosynthesis and signaling pathways and the expression of MaERF-B2-1 and MaERF-B2-2 in the response to flooding stress. , Berlin, v.14, n.4, p.767–777, 2014.
- SHENDURE, J.; JI, H. Next-generation DNA sequencing. **Nature Biotechnology**, New York, v.26, p.1135–1145, 2008.
- SHESHADRI, S. A. *et al.* Stress-mediated *cis*-element transcription factor interactions interconnecting primary and specialized metabolism in planta. **Frontiers in Plant Science**, Lausanne, v.7, p.1-23, 2016.
- SHI, F. *et al.* Cytosolic ascorbate peroxidase 2 (cAPX2) is involved in the soybean response to flooding. **Phytochemistry**, London, v.69, n.6, p.1295–1303, 2008.
- SHIMOMURA, M. *et al.* The *Glycine max* cv. Enrei Genome for Improvement of Japanese Soybean Cultivars. **International Journal of Genomics**, New York, v.2015, n.1, p.1-8, 2015.
- SIMPSON, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. **Genome Research**, Cold Spring Harbor, v.19, p.1117–1123, 2009.
- SIMS, D. *et al.* Sequencing depth and coverage: key considerations in genomic analyses. **Nature Reviews. Genetics**, London, v.15, 2014.
- SOLOVYEV, V. V. Statistical approaches in Eukaryotic gene prediction. In: HANDBOOK of Statistical genetics. 3.ed. Chichester: Wiley-Interscience, 2007. 1616 p.
- THOMAS, A. L. *et al.* Aerenchyma formation and recovery from hypoxia of the flooded root system of nodulated soybean. **Annals of Botany**, Oxford, v.96, n.1, p.1191-1198, 2005.
- THOMAS, A. L., *et al.* Rendimento de grãos de cultivares de soja em solo de várzea. **Pesquisa Agropecuária Gaúcha**, Porto Alegre, v.6, n.1, p.107-112, 2000.
- THOMAS, A. L. Potencialização da absorção de sódio pelo amônio em soja com sistema radicular inundado. In: REUNIÃO DE PESQUISA DA SOJA DA REGIÃO SUL, 37., 2009, Porto Alegre. **Programa e resumos**. Porto Alegre: UFRGS, 2009.
- THOMAS, A. L.; LANGE, C. E. **Soja em solos de várzea do sul do Brasil**. Porto Alegre: Evangraf, 2014. 128 p.

THOMAS, A. L.; SODEK, L. Amino acid and ureide transport in the xylem of symbiotic soybean plants during short-term flooding of the root system in the presence of different sources of nitrogen. **Brazilian Journal of Plant Physiology**, Londrina, v.18, p.333-339, 2006.

THOMAS, A. L.; SODEK, L. Development of the nodulated soybean plant after flooding of the root system with different sources of nitrogen. **Brazilian Journal of Plant Physiology**, Londrina, v.17, p.291-297, 2005.

TREVASKIS, B. *et al.* Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. **Proceedings of the National Academy of Sciences of United States of America**, Washington, v.94, n.1, p.12230-12234, 1997.

TUOMI, J. M. *et al.* Bias in the C_q value observed with hydrolysis probe based quantitative PCR can be corrected with the estimated PCR efficiency value. **Methods**, San Diego, v.50, p.313-322, 2010.

VALLIYODAN, B. *et al.* Expression of Root-Related Transcription Factors Associated with Flooding Tolerance of Soybean (*Glycine max*). **International Journal of Molecular Sciences**, Basel, v.15, n.1 p.17622-17643, 2014.

VARSHNEY, R. K. *et al.* Next-generation sequencing technologies and their implications for crop breeding. **Trends in Biotechnology**, Barking, v.27, n.9, p.522-530, 2009.

VOESENEK L. A. C. J.; BAILEY-SERRES J. Flood adaptive traits and processes: an overview. **New Phytologist**, Oxford, v.206, n.1, p.57-73, 2015.

VOESENEK L. A. C. J., SASIDHARAN, R. Ethylene—and oxygen signalling—drive plant survival during flooding. **Plant Biology**, Stuttgart, v.15, n.3, p.426-435, 2013.

WALKER, J. C. *et al.* DNA sequences required for anaerobic expression of the maize alcohol dehydrogenase 1 gene. **PNAS**, Washington, v.84, n.19, p.664-668, 1987.

WANG, X. *et al.* Gel-free/label-free proteomic analysis of root tip of soybean over time under flooding and drought stresses **Journal of Proteomics**, Amsterdam, v.130, n.1, p.42-55, 2016.

WANG, X. *et al.* Identification of the valid reference genes for quantitative rt-pcr in annual ryegrass (*Lolium multiflorum*) under salt stress. **Molecules**, Basel, v.20, n.3, p.4833- 4847, 2015.

WU, X. *et al.* SNP discovery by high-throughput sequencing in soybean. **BMC Genomics**, London, v.11, p.1-10, 2010.

XIE, F. *et al.* miRDeepFinder: a miRNA analysis tool for deep sequencing of plant small RNAs. **Plant Molecular Biology**, Dordrecht, v.80, n.1, p.75-84, 2012.

XU, J. *et al.* Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. **BMC Plant Biology**, London, v.14, n.83, p.1471-2229, 2014.

YADAV, C. B. *et al.* Genome-Wide SNP Identification and Characterization in Two Soybean Cultivars with Contrasting Mungbean Yellow Mosaic India Virus Disease Resistance Traits. **PLoS ONE**, San Francisco, v.10, n.4, 2015.

YANDELL, M.; ENCE, D. A beginner's guide to eukaryotic genome annotation. **Nature Reviews. Genetics**, London, v. 13, p. 329-342, 2012.

ZERBINO, D. R.; BIRNEY, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. **Genome Research**, Cold Spring Harbor, v.18, p.821-829, 2008.

ZHOU, Z. *et al.* Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. **Nature Biotechnology**, New York, v.33, n.4, p.408-414, 2015.