



UNIVERSIDADE
ESTADUAL DE LONDRINA

VALÉRIA STEFANIA LOPES

CARACTERIZAÇÃO DA FAMÍLIA DE GENES HSP20 EM
Glycine max



UNIVERSIDADE ESTADUAL DE LONDRINA
EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - SOJA

VALÉRIA STEFANIA LOPES

CARACTERIZAÇÃO DA FAMÍLIA DE GENES HSP20 EM
Glycine max

Londrina
2012

VALÉRIA STEFANIA LOPES

CARACTERIZAÇÃO DA FAMÍLIA DE GENES HSP20 EM
Glycine max

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia, do Departamento de Bioquímica e Biotecnologia, da Universidade Estadual de Londrina, PR, como requisito para a obtenção do título de Mestre.

Orientadora: Dra. Francismar Corrêa Marcelino
Guimarães

Co-orientador: Prof. Dr. André Luiz Martinez de
Oliveira

Londrina
2012

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da
Universidade Estadual de Londrina**

Dados Internacionais de Catalogação-na-Publicação (CIP)

L864c Lopes, Valéria Stefania.
Caracterização da família de genes HSP20 em *Glycine max* / Valéria
Stefania Lopes. – Londrina, 2012.
169 f. : il.

Orientador: Francismar Corrêa Marcelino-Guimarães.

Co-orientador: André Luiz Martinez de Oliveira.

Dissertação (Mestrado em Biotecnologia) – Universidade Estadual de
Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em
Biotecnologia, 2012.

Inclui bibliografia.

1. Biotecnologia vegetal – Teses. 2. Soja – Efeito do estresse – Teses. 3.
Genética vegetal – Teses. 4. Proteínas de choque térmico – Teses. 5. *Glycine
max* – Teses. I. Marcelino-Guimarães, Francismar Corrêa. II. Oliveira,
André Luiz Martinez de. III. Universidade Estadual de Londrina. Centro de
Ciências Exatas. Programa de Pós-Graduação em Biotecnologia. IV. EMBRAPA.
V. Título.

CDU 663.1

VALÉRIA STEFANIA LOPES

CARACTERIZAÇÃO DA FAMÍLIA DE GENES HSP20 EM *Glycine max*

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia, do Departamento de Bioquímica e Biotecnologia, da Universidade Estadual de Londrina, PR, como requisito para a obtenção do título de Mestre.

BANCA EXAMINADORA

Dra. Francismar C. Marcelino Guimarães
Embrapa Soja/UEL– Londrina - PR

Profa. Dra. Mayra C. da C. Gallo de Carvalho
UENP – Bandeirantes - PR

Dr. Waldir Pereira Dias
Embrapa Soja – Londrina - PR

Londrina, 22 de março de 2012.

Aos meus queridos e amados pais, Stéfano e Sueli (in memoriam), ao meu irmão, Alex, e ao meu marido, Lucas, pelo amor ilimitado e incondicional... É por vocês que nunca desisto.

AGRADECIMENTOS

Agradeço à Universidade Estadual de Londrina e à Embrapa Soja, por me proporcionarem as condições necessárias para a realização deste trabalho;

À Capes, pelo apoio financeiro imprescindível;

À Dra. Francismar Corrêa Marcelino Guimarães, por ter me recebido no laboratório de Biotecnologia Vegetal desde a iniciação científica, pela orientação no trabalho, pelas críticas e pelo exemplo profissional;

Ao meu co-orientador, Prof. Dr. André Luíz Martinez de Oliveira, pela colaboração, correções e disponibilidade;

À Profa. Dra. Mayra Costa da Cruz Gallo de Carvalho, que tanto me ajudou desde as etapas iniciais deste trabalho (companheira de bancada), também pela orientação, atenção, paciência e pela amizade;

Aos pesquisadores da Embrapa Soja, Dr. Eliseu Binneck e Dr. Waldir Pereira Dias, pelas sugestões e ajuda no decorrer do trabalho;

Aos Analistas do Laboratório, Cesar Silveira, Danielle Cristina Gregório da Silva, Marcia Kamogae Kuwahara, Nilson Vieira, Renan Milagres e Silvana Rothenbach Marin pela atenção e amizade.

Aos amigos do Laboratório de Biotecnologia Vegetal da Embrapa Soja, por sempre me ajudarem e tornarem agradável e divertida a rotina de laboratório: Adriana Polizel, Alan Pereira, Aguida Morales, Amanda Paiva, André Passianotto, Carolina Rocha, Cibelle Engels, Cynara Romero, Elton Gargioni, Euziane Lino, Fabiana Rodrigues, Gislaine Vasquez, Idenize Orsini, João Vitor Maldonado, Juliane Marinho, Juliana Leite, Juliana Marcolino, Larissa Giroto, Lívia Nogueira, Lizandra Catelli, Luana Darben, Luciano Aoyagi, Máisa Pinheiro, Maria Cecília, Mayla Molinari, Mayra Carvalho, Michelle Rincão, Renata Fuganti, Salvador Lima, Selma Pereira (*in memoriam*), Tânia Hitomi (*in memoriam*), Thiago Nakayama, e Vera Lúcia Pierotte;

Aos amigos do Mestrado em Biotecnologia, Bruna Caroline, Francieli Bortoluzzi, Gesiele Barros, Leonardo Cardoso, Lucas Bissoqui e Patrícia Bittencourt, alunos da turma 2010, pelo companheirismo durante o período de disciplinas e por tornarem divertidas as aulas.

À Maria Aparecida dos Santos Letrari (Di), da divisão de referência da Biblioteca Central da UEL, pela ajuda nas correções e normalização da dissertação.

A minha AMADA FAMÍLIA:

Ao meu marido, Lucas Panicio Caitar, pelo grande amor, apoio, paciência, dedicação, confiança, companheirismo;

Meus pais, Stéfano Lopes e Sueli Ruiz Lopes (*in memorian*), pelo amor e dedicação...

Meu irmão, Alex Stéfano Lopes, pela força e por nunca me deixar desistir;

Meus avós, Antônio Lopes Hernandez (*in memorian*) e Hortência Paranzini Hernandez, pelo amor e preocupação com meu futuro;

Meu querido tio, Mauridez Lopes (Tio Tuta), pelo exemplo de persistência;

Aos meus sogros, José Melero Caitar e Aparecida Josefina Panicio Caitar, pelo apoio, caronas e entregas na secretaria de pós-graduação do CCE.

Aos meus tios e tias, em especial para tia Irany Lopes Dezoti, pelos bons conselhos e apoio.

Todo o meu carinho e agradecimento!

***"Que o teu orgulho e objetivo consistam em pôr
no teu trabalho, algo que se assemelhe a um
milagre".***

Leonardo da Vinci

LOPES, Valéria Stefania. **Caracterização da Família de Genes Hsp20 em *Glycine max***. 2012. 169f. Dissertação (Mestrado em Biotecnologia) – Universidade Estadual de Londrina, Londrina - PR.

RESUMO

As pequenas proteínas de choque térmico (HSP20) são frequentemente associadas com a resposta das plantas ao estresse causado por fatores abióticos e, mais recentemente, têm sido também associadas a resposta aos estresses bióticos. Nas plantas, os genes Hsp20 representam a classe mais abundante dentre as proteínas de choque térmico, mas ainda pouco se conhece sobre essa família de genes em soja. Devido à suas aparentes multifuncionalidades, essas proteínas são alvos promissores para o desenvolvimento de variedades agrícolas melhor adaptadas a condições de estresses abióticos e bióticos, até mesmo quando combinados. Dessa forma, o presente trabalho realizou a caracterização molecular *in silico* das regiões codificadoras e reguladoras da família de genes codificadores para HSP20 de soja, com foco em sua distribuição no genoma, localização subcelular, divisão em subfamílias, estrutura secundária e regulação frente a estresses bióticos e abióticos, além da identificação de padrões de cis elementos, potencialmente envolvidos na resposta da soja a nematóides. Após a prospecção de genes anotados em bancos de dados do genoma da soja como Hsp20, foram obtidos 76 modelos gênicos. Dentre estes, apenas 52 modelos gênicos fizeram parte dos potenciais candidatos a GmHsp20 (*Glycine max*-Hsp20), devido as suas características estruturais, de cis elementos e de expressão. Em seguida, foram identificados, a partir das análises *in vivo*, 45 genes como Hsp20 de soja, distribuídos em 11 subfamílias. Para cada uma dessas, foi possível observar padrões de estrutura secundária específicos. Dentre os 45 genes GmHsp20 responsivos ao estresse de calor, 5 foram também responsivos ao estresse de frio e outros 5 ao estresse por infecção pelo nematóide *M. javanica*. Além disso, foram observados mais dois genes responsivos ao estresse biótico, mas não ao choque térmico. Obtiveram-se modelos operacionais de promotores para os genes responsivos a cada tipo de estresse analisado. Entre os cis elementos identificados nos Hsp20, que respondem a infecção por *M. javanica*, estão os Wbox, CAAT box, ABRE e MYB, além do elemento HSE/Heat. Os promotores responsivos ao estresse biótico seguiram padrões de composição e distribuição de cis elementos, descritos na literatura como relacionados a esse tipo de estresse e outros. Tais resultados irão auxiliar na geração de tecnologias de expressão dirigida, ainda mais avançadas, e novos genótipos de soja cada vez mais adaptados a condições combinadas de estresse.

Palavras-chave: *Glycine max* 1. Estresse abiótico e biótico 2. Família de genes Hsp20 3. Modelos operacionais de promotores 4. Nematóides 5.

LOPES, Valéria Stefania. **Characterization of the Hsp20 gene family in Glycine max**. 2012. 169p. Dissertation (Master's Degree in Biotechnology) – Universidade Estadual de Londrina, Londrina -PR.

ABSTRACT

The small heat shock proteins (HSP20) are often associated in plant stress response caused by abiotic factors and, more recently, have also been associated with response to biotic stresses. The Hsp20 genes represent, in plants, the most abundant class among the heat shock proteins, but little is known about this gene family in soybean. Due their apparent multifunctionality, these proteins are promising targets to the crop varieties development for better conditions adapted to biotic and abiotic stresses, even when they are combined. Thus, the present study conducted an in silico molecular characterization of regulatory and coding regions of HSP20 genes from soybean, focus in its genome distribution, subcellular localization, division into subfamilies, secondary structure and regulation under biotic and abiotic stresses, besides the identification patterns to cis elements potentially involved in the response to nematodes. After the exploration of Hsp20 genes annotation in soybean genome databases, 76 gene models were obtained. After in silico analysis, just 52 gene models were part of the GmHsp20 potential candidates due to their structural characteristics of cis elements and expression profile. In addition, based on in vivo analysis, 45 soybean Hsp20 genes were identified, distributed in 11 subfamilies, for which is possible to observe a specific secondary structure for each one. Among the 45 GmHsp20 genes heat stress responsive, 5 genes were cold stress responsive and other five were nematode infection by *M. javanica* responsive. Moreover, two genes were observed being responsive to biotic stress, but they weren't responsive to thermal shock. Operational Models of Hsp20 promoters were obtained to responsive genes to each stress condition examined in this study. Among the identified cis elements in Hsp20 soybean genes that were responsive to *M. javanica* infection were W box, CAAT box, ABRE and MYB, besides the HSE / Heat element. Promoters responsive to biotic stress in soybean follows composition and distribution standards of cis elements, as described in the literature to be related to this type of stress. These results, such as responsive genes and promoters to many different stresses, can assist in generation of expression directed technologies even more advanced and new soybean genotypes more adapted to under combined stress conditions.

Keywords: *Glycine max* 1. 2 Biotic and Abiotic Stress. 3 Hsp20 Gene Family. Operational Models of Hsp20 Promoter 4. Nematodes 5.

LISTA DE FIGURAS

- Figura 1** - Micrografia eletrônica de varredura de um de nematóide de galhas-fêmea (parasita de plantas), *Meloidogyne arenaria*.....38
- Figura 2** - Ciclo de vida do fitonematóide endoparasita sedentário. Os juvenis J2 migram entre as células vegetais, facilitada pela digestão da lamela média, e modificam a fisiologia celular para a formação de células gigantes e galhas. As fêmeas adultas depositam seus ovos em massa proteica39
- Figura 3** - Formação de galha. As raízes de soja tornam-se engrossadas na região do parasitismo como resultado do processo de infecção pelo nematóide *M. javanica*, que provoca reações de hiperplasia e hipertrofia nos tecidos.....40
- Figura 4** - Diagrama esquemático ilustrando os principais componentes de sinalização e ativação de resposta da planta ao estresse biótico e abiótico. Os estresses de elevação/diminuição da temperatura e ataque de patógenos (1) provocam aumento na fluidez da membrana e permite um fluxo intracelular de íons de Ca^{2+} (2). Esses mesmos agentes de estresse desencadeiam nas células, por diferentes vias, a formação de espécies reativas de oxigênio (ROS) e metabólitos onde a ROS é precursora que tem efeito oxidativo nos lipídeos da membrana, levando ao mesmo processo de ativação de canais de cálcio (seta vertical vermelha). A maior concentração de Ca^{2+} ativa as calmodulinas (CaM), que através da interação com as proteínas quinases (CBK) (3), promove a fosforilação (P) dos fatores de transcrição, como por exemplo os HSFs (4). TFs ativados se ligam aos promotores de genes responsivos a estresse, nesse caso genes codificadores para HSP20 (5), causando o deslocamento das histonas e facilitando o acesso da RNA polimerase II (6). Como consequência, há alterações nos níveis de expressão dos genes Hsp20 as quais participam das respostas celulares e fisiológicas, realizando assim a manutenção da homeostase do organismo vegetal.....43

Figura 5 -	Fluxograma das etapas conduzidas para a realização das análises in silico dos genes Hsp20 de soja.	65
Figura 6 -	Representação esquemática do delineamento experimental instalado com o genótipo BRS133, susceptível a nematóide de galha, em fitotron para exposição a condições de estresse abiótico.	66
Figura 7 -	Peneiras de 60 meshs (superior) e 500 meshs (inferior), utilizadas para recuperar os indivíduos de <i>Meloidogyne javanica</i> a partir de raízes de soja trituradas em liquidificador.	67
Figura 8 -	A. Abertura do orifício ao lado do colo da planta para a inoculação de nematóides. B. Inoculação das amostras de soja com indivíduos de <i>Meloidogyne javanica</i>	68
Figura 9 -	Representação esquemática do delineamento experimental instalado com o genótipo BRS133, susceptível a nematóide de galha, em fitotron para inoculação do nematóide <i>M. javanica</i>	69
Figura 10a -	Logo obtido no MEME/MAST para o Modelo de Higger Markov referente ao domínio ACD ($3.3e-1785$) encontrado nas proteínas GmHsp20 preditas. O E-value do motivo é uma estimativa do número esperado de motivos com a mesma largura e número de ocorrências que estariam presentes em um conjunto de tamanho similar de sequências aleatórias. A altura dos símbolos em cada pilha no motivo indica a conservação da sequência de aminoácidos nessa posição.	78
Figura 10b -	Mapa gráfico obtido pela análise no programa MEME na busca pelo domínio ACD em cada sequência de aminoácidos. As barras azuis indicam o posicionamento do domínio ACD em cada um dos 76 modelos gênicos candidatos a Hsp20 de <i>Glycine max L.</i> Na escala o número 0 indica a região N-terminal e o 400 a região C-terminal. Os valores de p-valeu são encontrados logo na coluna (lowest p-valeu) a direita da coluna de glymas.(name).	79
Figura 11a -	Logos de Modelo de Higger Markov obtidos no MEME/TOMTOM baseados em sequências de consenso dos promotores das 53 Hsp20 preditas para a soja com região a	

montante disponível. Os motivos obtidos pela análise MEME foram plotados de acordo com suas posições dentro dos locais e suas seqüências consenso foram apresentadas como gráficos usando HMM. As ocorrências de motivo são classificadas por P-value e alinhados uns com os outros. O E-valor de motivo é uma estimativa do número esperado de motivos com a mesma largura e número de ocorrências que estariam presentes em um conjunto de tamanho similar de seqüências aleatórias. A altura de símbolos em cada pilha no motivo indica a conservação da seqüência nessa posição. As seqüências foram manualmente destacadas para indicar os HSEs perfeitos (caixa vermelha) consenso reconhecido: (nGAAnnTTCnnGAAn ou nTTCnnGAAnnTTCn); caixa amarela, mostrando módulo imperfeito do HSE. Três motivos foram encontrados, sendo os elementos HSE perfeitos representados pelos motivos 1 (Matches to Query 1) e 3 (Matches to Query 3) na figura.83

Figura 11b - Mapa gráfico obtido pela análise no programa MEME, na busca pelos cis elementos HSE, em cada seqüência de aminoácidos. As barras azul claras e vermelhas indicam o posicionamento dos elementos HSE da região promotora em cada um dos 53 modelos gênicos candidatos a Hsp20 de Glycine max L, que apresentam região da montante disponível no Phytozome. Na escala, o número 0 indica a região -1 pb à montante, e 500, a região -500 pb à montante do gene. Os valores de p-value são encontrados logo na coluna (lowest p-value) à direita da coluna de glymas.(name). Apenas os motivos encontrados na fita positiva foram considerados, obedecendo a orientação de leitura do gene. O HMM para cada motivo pode ser visualizado na figura 11a84

Figura 12 - Ilustração da predição dos sítios de ligação para o fator de transcrição de choque térmico nas regiões promotoras dos modelos gênicos de soja: Glyma02g45810, Glyma13g27590 e Glyma15g11360 pelo programa MatInspector. Para estes

modelos gênicos, os HSEs foram encontrados em posições fora dos padrões normalmente encontrados para as Hsp20. Foram analisados até 1500 pb a montante do sitio de início transcrição predito. As marcas em marrom no decorrer das fitas indicam a posição dos HSE em cada sequência. Na escala o número 1 indica o -1 pb ao sítio de início da transcrição e o número 1500 indica -1500 pb. Apenas os motivos encontrados na fita positiva foram considerados, obedecendo a orientação de leitura do gene.85

Figura 13 - Ilustração da predição dos sítios de ligação para o fator de transcrição de choque térmico nas regiões promotoras dos 53 modelos gênicos de soja, que apresentam região da montante disponível no Phytozome, pelo MatInspector. Foram analisados até 500 pb a montante do sitio de início transcrição predito. As marcas em marrom no decorrer das fitas indicam a posição dos HSE em cada sequência. Na escala o número 1 indica o -1 pb ao sítio de início da transcrição e o número 500 indica -500 pb. Apenas os motivos encontrados na fita positiva foram considerados, obedecendo a orientação de leitura do gene.88

Figura 14 - Heatmap para experimentos de microarranjo, baseados em perfis de expressão de genes Hsp20 durante estresse de calor em plantas de soja, obtido no banco de dados de expressão Genevestigator. Expressão de genes Hsp20 durante estresse de calor são apresentados como mapas de calor em vermelho/preto/verde em gradiente usando meta-análise ferramenta em <http://www.genevestigator.ethz.ch> Genevestigator. A cor vermelha mais escura corresponde a mais forte expressão. A existência de mais de uma probe para um gene explica a presença em alguns casos do mesmo glyma estar repetido duas ou mais vezes. Os retângulos azuis marcam os cinco genes inclusos para as análises in silico.91

Figura 15 - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e

temperatura de pareamento) dos primers em DNA genômico de soja. Nesta figura estão os resultados dos testes com os oligonucleotídeos para os modelos gênicos V74 (glyma20g01930), V48 (glyma13g24510), V52 (glyma14g04970), V64 (glyma16g04270), V63 (glyma16g01440) e V53 (glyma14g06900). Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para cada par de primers, e na segunda linha os primers testados seguidos dos controles negativos de reação (NO). Foi utilizado o marcador molecular de 50 pb (Ladder 50 pb – Invitrogen). b. Ladder de 50 pb – Invitrogen. Na horizontal, estão indicadas os tamanhos, em pares de base, das principais bandas do marcador. 98

Figura 16 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de 1 primer específicos para 1 modelo gênico: V73 (glyma19g01440). Na primeira linha do cabeçalho de cada foto, está anotada a codificação do primer estabelecidas neste trabalho (Vn^o) e o modelo gênico correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4° (tratamento de frio), Inoc 4dpi (tratamento de infecção por M.javanica após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por M.javanica após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por M.javanica após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por M.javanica após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 50 pares de base (Ladder 50 pb – Invitro). 99

Figura 17 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de seis primers

específicos para seis modelos gênicos: V3 (glyma02g41150), V9 (glyma04g38530), V13 (glyma06g05740), V17(glyma07g04860), V21 (glyma07g32070) e V26 (glyma08g07340), que demonstram a indução desses genes sob estresse de calor facilmente visualizada pela análise qualitativa. Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos primers estabelecidas neste trabalho (Vnº) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4º (tratamento de frio), Inoc 4dpi (tratamento de infecção por M.javanica após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por M.javanica após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por M.javanica após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por M.javanica após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 100 (Ladder 100 pb – Invitro) pares de base na primeira foto e o marcador de 50 pb (Ladder 50 pb – Invitrogen) para as demais fotos. 101

Figura 18 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de dois primers específicos para dois modelos gênicos: V16 (glyma07g03450) e V62 (glyma15g41880), que demonstram a indução desses genes sob estresse de calor facilmente visualizada pela análise qualitativa. Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos primers estabelecidas neste trabalho (Vnº) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4º (tratamento de frio), Inoc 4dpi (tratamento

de infecção por *M.javanica* após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por *M.javanica* após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por *M.javanica* após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por *M.javanica* após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 50 pb (Ladder 50 pb – Invitrogen.)..... 102

Figura 19 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de quatro primers específicos para quatro modelos gênicos: V40 (*glyma12g06210*), V50 (*glyma13g36220*), V52 (*glyma07g03450*) e V64 (*glyma15g41880*). Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos primers estabelecidas neste trabalho (Vnº) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4º (tratamento de frio), Inoc 4dpi (tratamento de infecção por *M.javanica* após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por *M.javanica* após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por *M.javanica* após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por *M.javanica* após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 100 pares de base (Ladder 100 pb – Invitro) para o gel de V52 e o marcador de 50 pb (Ladder 50 pb – Invitrogen) para as demais fotos. 103

Figura 20 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de quatro primers específicos para quatro modelos gênicos: V1 (*glyma01g26570*) e V76 (*glyma20g35650*). Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos primers

estabelecidas neste trabalho (Vn^o) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4° (tratamento de frio), Inoc 4dpi (tratamento de infecção por M.javanica após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por M.javanica após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por M.javanica após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por M.javanica após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 100 (Ladder 100 pb – Invitro) pares de base para o gel de V1 o marcador de 50 pb (Ladder 50 pb – Invitrogen) para o gel de V76. 105

Figura 21 - Posição dos cis elementos identificados pelos programas de predição na região a montante -500 pb de cada um dos 8 genes Hsp20 responsivos ao estresse biótico, identificados neste trabalho e na literatura. As caixas azuis indicam os elementos TATAbox, as vermelhas os HSE, as verde desoldado os CAAt box, as verdes os elementos TA-rich e as cinza os elementos Wbox. Os genes identificados pelos glyma16g04270 e glyma18g41650 são responsivos ao estresses biótico, mas não ao calor. 115

Figura 22 - Ilustração da estrutura de elementos comuns da região promotora de genes Hsp20 de soja – Modelo mínimo. O programa Frameworker Genomatix foi usado para procurar o principal elemento comum entre os genes. O modelo demonstra diversos padrões com pelo mesmo dois HSE, um em cada orientação. 120

Figura 23 - Ilustração do consenso de todos os modelos de estrutura de elementos comuns da região promotora de genes Hsp20 de soja – Modelo frio. O programa Frameworker Genomatix foi usado para procurar os principais elementos comuns entre os genes e suas posições de alta frequência. Para o modelo de

posição dos HSE na orientação negativa o valor de p-valeu foi de 9,37E-01. Para o modelo dos elementos ABRE, MYBS, MYCL e HEAT o valor de p-valeu foi de 9,57E-04. Outro modelo sem elementos mandatório resultou na inclusão os elementos L1BX, AHBP, GTBX, VTBP e HEAT com valor de p-valeu de 4,789E-06. 121

Figura 24 - Ilustração do consenso de todos os modelos de da estrutura de elementos comuns da região promotora de genes Hsp20 de soja – Modelo biótico. O programa Frameworker Genomatix foi usado para procurar os principais elementos comuns entre os genes e suas posições de alta frequência. Para o modelo de posição dos HSE na orientação negativa o valor de p-valeu foi de 0,0381311. Para o modelo dos elementos AHBP, GTBX, VTBP (TATA box vertebrate), PTBT (TATA box plant) o valor de p-valeu foi de 1,67E-04. Para o modelo Wbox com valor de p-valeu foi de 5,02E-01. Outro modelo CAAT box com valor de p-valeu foi de 0,167564..... 123

Figura 25 - Árvore filogenética e distribuição em subfamílias dos 45 GmHsp20 identificados em soja. A árvore foi construída com base no alinhamento da sequência de resíduos de aminoácidos que formam o domínio alfa-cristalino (ACD) dos organismos: Glycine max (Hsp20 de soja em análise estão destacadas em vermelho), At Arabidopsis thaliana, Os Oryza sativa, Vv Vitis vinífera, Ta Triticum aestivum, Ps Pisum sativum, Le Solanum lycopersicum, Zm Zea mays, Lp Lycopersicon peruvianum, Cd Cynodon dactylon, So Saccharum officinarum, Sb Sorghum bicolor, Hv Hordeum vulgare. As GmHsp20 foram distribuídas em 11 das 16 subfamílias já descritas. Nenhuma das GmHsp20 candidatas foram enquadradas para as subfamílias núcleo/citoplasmáticas CVI, CVII, CVIII, CIX e CX. Para nomenclatura das subfamílias RE (retículo endoplasmático), P (plastídios/cloroplasto), Px (peroxissomo), MI e MII (mitocôndria). 132

Figura 26 - Resultados positivos para a predição de peptídeo sinal. a. Resultado positivo para o Glyma10g32000 na predição de peptídeo sinal. No SignalP foram obtidos resultados positivos e bastante semelhantes para as GmHSP20, Glyma16g33130, Glyma19g01440 e Glyma20g35650, agrupadas junto com o Glyma10g32000 na subfamília ER na filogenia. b. Resultado positivo para o Glyma12g01580 na predição de peptídeo sinal, agrupado na subfamília MI. c. Resultado positivo para o Glyma04g40790 na predição de peptídeo sinal, agrupado na subfamília P. O Glyma06g14000 e Glyma18g10760, também agrupados na subfamília P, apresentaram resultado positivo e bastante semelhante ao obtido para o Glyma04g40790. Nos gráficos são atribuídas pontuações (scores) para a predição de peptídeo sinal (S-score), o C-score é um provável de local de clivagem. Y-score combina o resultado de C-score com S-score para indicar com melhor precisão um provável sítio de clivagem..... 135

Figura 27 - Diagrama para o resultado, com alta probabilidade, positivo (a) e de baixa probabilidade (b). a. Resultado com alta probabilidade, a predição da presença de domínio transmembrana no Glyma10g32000. Na legenda é possível verificar as linhas que indicam a predição de regiões de domínio transmembrana (vermelho), regiões da proteína que provavelmente estejam dispostas para o lado externo da membrana (rosa) e do lado interno (azul). O Glyma19g01440 também apresentou um resultado muito semelhante para a predição de domínio transmembrana obtido para o Glyma10g32000. b. Diagrama obtido do resultado, mas com baixa probabilidade para positivo, a predição de domínio transmembrana para o Glyma16g33130. O Glyma20g35650 também apresentou um resultado bastante semelhante. As GmHSP20, codificadas pelos Glyma16g33130, Glyma19g01440 Glyma10g32000 e Glyma20g35650, formam a subfamília ER na filogenia..... 137

- Figura 28** - Diagrama ilustrando os padrões de estrutura secundária preditos para as proteínas das subfamílias de GmHSP20. As posições de íntrons estão apontadas por setas em vermelho. As setas horizontais azuis delimitam regiões de formação de estruturas secundárias β -pregueadas. Espirais em verde são utilizadas para delimitar as regiões de possíveis formações de estruturas secundárias em α -hélices. Os retângulos em azul e amarelo delimitam as regiões preditas para os domínios α -cristalinos e peptídeos sinal, respectivamente. 139
- Figura 29** - Mapa de loci e coordenadas de eventos de duplicações dos genes parálogos candidatos a GmHsp20 nos cromossomos 17-20. As identificações de cada grupo de ligação são indicadas no topo de cada barra. Na primeira fileira estão apenas os cromossomos em que os GmHsp20 candidatos foram mapeados e apresentam valores de identidade para sequência de nucleotídeos acima de 70% com outros genes, caracterizando uma possível duplicação gênica. Os genes possivelmente duplicados estão ligados por traços e a porcentagem de identidade anotada na mesma cor do traço. Os retângulos vermelhos indicam as duplicações gênicas segmentares. 142
- Figura 1S** - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e temperatura de pareamento) dos primers de detecção de infecção de plantas por *M. javanica*. Para o teste de especificidade foram utilizados DNA genômico de soja e de *M. javanica*. O tamanho do fragmento esperado é de 945 pb. Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para os primers, e na segunda linha as amostras com o controle negativo da reação (NO). Foi utilizado o marcador molecular de 1 kb plus (Ladder de 1 kb plus – life technologies). b. Ladder de 1 kb plus – life technologies. Na horizontal estão indicadas os tamanhos em pares de base das principais bandas do marcador 161

- Figura 2S** - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR, com 945 pb, para confirmação da infecção das plantas de soja dos tratamentos bióticos pelo nematóide *M. javanica*. Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para os primers, e na segunda linha as amostras com o controle negativo da reação (NO). Foi utilizado o marcador molecular de 1 kb plus (Ladder de 1 kb plus – life technologies). b. Ladder de 1 kb plus – life technologies. Na horizontal estão indicadas os tamanhos em pares de base das principais bandas do marcador 161
- Figura 3S** - Eletroforese em gel de agarose 1% para a análise de integridade das amostras de RNA total extraídas dos tratamentos de estresse biótico e abiótico. Na primeira foto (superior) estão os RNAs das amostras de BRS133 de 1 a 9, estresse abiótico, e de 10 a 21, estresse biótico (infecção por *M. javanica*). Na segunda foto (abaixo) estão as amostras de RNA somente do experimento de estresses por infecção por *M.javanica*. As amostras de 22 a 33 são oriundas da cultivar PI595099 e de 34 a 45 de BRS133 (experimento de validação). Para avaliar a integridade do RNA a banda 28S deve ser quase duas vezes mais intensa que a banda 18S as duas bandas em boa intensidade e definição (NARDINELLI, 2008) 162
- Figura 4S** - Eletroforese em gel de agarose 1,2% dos produtos da reação de PCR para a confirmação de eliminação de DNA genômico da amostra após tratamento com DNase I. O tamanho do fragmento esperado é de 520 pb quando template for DNA genômico e 440 pb quando template for RNA. As amostras de 1 a 9 e 10 a 21 são dos tratamentos de estresses abióticos e bióticos, respectivamente. Foram utilizados o controle de DNA genômico de soja e de reação (NO)..... 162
- Figura 5S** - Eletroforese em gel de agarose 1,2% dos produtos da reação de PCR com o primer para o gene β -actina para a análise da qualidade do cDNA sintetizado. O tamanho do fragmento

esperado é de 520 pb quando template for DNA genômico e 440 pb quando template for RNA 163

Figura 6S - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e temperatura de pareamento) dos primers em DNA genômico de soja. Nesta figura estão os resultados dos testes com os oligonucleotídeos para os modelos gênicos V74 (glyma20g01930), V48 (glyma13g24510), V52 (glyma14g04970), V64 (glyma16g04270), V63 (glyma16g01440) e V53 (glyma14g06900). Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para cada par de primers, e na segunda linha os primers testados seguidos dos controles negativos de reação (NO). Foi utilizado o marcador molecular de 50 pb (Ladder 50 pb – life Technologies) 163

Figura 7S - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e temperatura de pareamento) dos primers em DNA genômico de soja. Nesta figura estão os resultados dos testes com os oligonucleotídeos para os modelos gênicos V74 (glyma20g01930), V48 (glyma13g24510), V52 (glyma14g04970), V64 (glyma16g04270), V63 (glyma16g01440) e V53 (glyma14g06900). Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para cada par de primers, e na segunda linha os primers testados seguidos dos controles negativos de reação (NO). Foi utilizado o marcador molecular de 50 pb (Ladder 50 pb – life Technologies) 166

LISTA DE GRÁFICOS

- Gráfico 1** - Resultado da análise de expressão relativa por RT-qPCR dos genes candidatos Hsp20 no genótipo suscetível BRS133 nos tratamentos de estresse abiótico (calor/vermelho e frio/azul) e biótico (4dpi/cinza claro e 8dpi M.javanica/cinza escuro). Estão plotados apenas os genes foram significativos ao nível de 5%. As escalas são diferentes para cada tipo de estresse, podendo apresentar valor negativos. Spots alinhados em escalas positivos indicam genes que foram induzidos significativamente e alinhados em escalas negativas foram reprimidos significativamente. As barras de erro indicam em cada uma das escalas em que está os spots a margem de erro para mais ou para menos. Os glymas seguem em linha reta na horizontal para cada um dos gráficos/tratamentos. 109
- Gráfico 2** - Resultado da análise de expressão relativa por RT-qPCR dos genes candidatos Hsp20 no genótipo resistente PI595099 e suscetível no tratamento de estresse biótico (4 e 8dpi M.javanica). Estão plotados apenas os genes foram significativos ao nível de 5% em algum dos tratamentos. Na legenda, cada cor indica uma cultivar submetida a um dos tratamento por infecção do nematóide. Apenas os tratamentos em que a alteração no nível de expressão genica foi significativo estão marcadas com asterisco logo acima. As barras de erro indicam na escala a margem de erro para mais ou para menos. 111

LISTA DE TABELAS

- Tabela 1** - Resultados da análise in silico para a predição dos critérios mínimos utilizados na seleção dos candidatos potenciais a Gmhsp20. Estão apresentados candidatos potenciais em negrito aqueles que possuem domínio ACD, pelo menos um elemento HSE e expressão previamente observada em ao menos um dos bancos investigados. Quando a predição de elementos HSE não foi possível devido à indisponibilidade de sequência no Phytozome, mas quando houve observância dos outros critérios, o modelo gênico foi considerado como candidato potencial. Os resultados de ocorrência e posicionamento dos elementos HSE são apresentados em quatro colunas distintas que correspondem aos programas utilizados na análise. A sigla “nd” significa que a região à montante não está disponível no Phytozome. 92
- Tabela 2** - Ocorrência e posicionamento dos cis elementos: CCAAT, TATABox, Wbox e TA-rich na região à montante (- 500pb) dos 47 modelos gênicos identificados como candidatos potenciais a GmHsp20. As posições são estimadas em relação ao sítio de início da transcrição. Os resultados são apresentados para os três programas utilizados nas análises de cis elementos. Os demais elementos foram obtidos pelo programa PlantCare. A sigla “nd” significa que a região à montante não está disponível no Phytozome. 116
- Tabela 3** - Localização cromossômica, características físico-químicas, padrão de íntrons e localização subcelular para as GmHsp20. As abreviação na coluna de localização subcelular predita estão identificadas para as organelas: citoplasma (cito), peroxíssomo (Px), cloroplasto (clorop), mitocôndria (mito) e reticulo endoplasmático (R.E.)..... 125
- Tabela 1S** -Lista de oligonucleotídeos desenhados especificamente para cada modelo gênico candidato a Hsp20 de soja para uso nas análises de PCR qualitativo e RT- Qpcr 160

SUMÁRIO

1	INTRODUÇÃO	27
2	REVISÃO DE LITERATURA	29
2.1	SOJA: ASPECTOS GERAIS E ECONÔMICOS	29
2.2	ESTRESSES ENFRENTADOS PELA CULTURA	30
2.2.1	Estresses Abióticos	31
2.2.2	Estresses Bióticos	34
2.3	NEMATÓIDES NA CULTURA DA SOJA	34
2.3.1	Aspectos Gerais Relacionados ao Gênero <i>Meloidogyne</i> – <i>M. Javanica</i>	36
2.4	SOBREPOSIÇÃO DE VIAS DE SINALIZAÇÃO EM RESPOSTA AO ESTRESSE ABIÓTICO E BIÓTICO EM PLANTAS.....	41
2.5	AS HSPS E OS FATORES DE TRANSCRIÇÃO DE CHOQUE TÉRMICO.....	47
2.5.1	Família HSP20	49
2.6	AS HSPS NA INTERAÇÃO MOLECULAR PLANTA-PATÓGENO	51
2.7	BIOINFORMÁTICA: ARMAZENAMENTO E ANÁLISE DE INFORMAÇÕES.....	54
3	OBJETIVO GERAL	59
3.1	OBJETIVOS ESPECÍFICOS.....	59
4	MATERIAL E MÉTODOS	60
4.1	ANÁLISES “IN SILICO”	60
4.1.1	Busca por Modelos Gênicos Hsp20 em Soja	60
4.1.2	Domínios Conservados e Estrutura Secundária das GmHsp20.....	60
4.1.3	Busca por HSE e Outros cis Elementos na Região Regulatória dos Modelos Gênicos	61
4.1.4	Seleção de Modelos Gênicos para Análise “in vivo” e Desenho dos Oligonucleotídeos Iniciadores.....	62
4.1.5	Busca por Modelos de Promotores	63
4.1.6	Relação Filogenética e Distribuição em Subfamílias.....	63
4.1.7	Análise de Localização Cromossômica e Duplicação Gênica	64
4.2	ANÁLISES “IN VIVO”	66

4.2.1	Material Biológico e Instalação do Experimento para Análises “in vivo”	65
4.2.2	Tratamento – Estresse Abiótico (Temperatura).....	66
4.2.3	Investigação do Perfil Heat Shock em Tecido Específico – Folha.....	66
4.2.4	Preparo do Inóculo e Inoculação das Plantas	67
4.2.5	Tratamento – Estresse Biótico (Infecção <i>M. javanica</i>).....	68
4.2.6	Análise Comparativa dos Genes Responsivos ao Estresse por Infecção pelo Nematóide em Genótipo Resistente	69
4.2.7	Preparo das Amostras, Extração de DNA e RNA Total.....	69
4.2.8	Síntese de cDNA.....	73
4.2.9	Detecção da Expressão das GmHsp20s por PCR Qualitativo	74
4.2.10	Análise do Nível de Expressão - RTq-PCR.....	75
5	RESULTADOS E DISCUSSÃO	78
5.1	IDENTIFICAÇÃO DE GMHSP20 EM SOJA.....	78
5.1.1	Estudo da Região Codificadora.....	78
5.1.2	Estudo da Região Promotora	81
5.1.3	Análise de Dados de Expressão Digital das Hsp20 em Soja	90
5.2	ANÁLISES “IN VIVO” - INDUÇÃO DA EXPRESSÃO DOS HSP20 CANDIDATOS NOS DIFERENTES ESTRESSES	97
5.2.1	Análises Qualitativas	97
5.2.2	Análises Quantitativas.....	106
5.3	IDENTIFICAÇÃO DE ELEMENTOS REGULATÓRIOS.....	112
5.4	IDENTIFICAÇÃO DE MODELOS OPERACIONAIS DE PROMOTORES.....	119
5.5	ANÁLISES DOS GENES CARACTERIZADOS COMO HSP20 DE SOJA.....	123
5.5.1	Região Codificadora	124
5.5.2	Subfamílias de GmHsp20.....	130
5.5.3	Predição de Estruturas Secundárias e Domínios Funcionais para as Subfamílias das GmHSP20	138
5.5.4	Localização Cromossômica e Duplicação de Genes GmHsp20	140
6	CONCLUSÕES	146
	REFERÊNCIAS.....	148

DADOS SUPLEMENTARES	160
----------------------------------	------------

1 INTRODUÇÃO

As proteínas de choque térmico (*Heat Shock Proteins* – HSPs) foram inicialmente identificadas em *Drosophila melanogaster* em resposta ao choque térmico (RITOSSA, 1962 *apud* CASHIKAR et. al., 2005). As HSPs são divididas em proteínas de alta massa molecular, compreendendo as HSP100, HSP90, HSP70/DnaK, HSP60/GroE, e de baixa massa molecular, as HSP20 ou *small heat shock proteins* (sHSPs) de 15-42 kDa (SARKAR et. al, 2009).

Nas plantas, há cerca de 4 vezes mais Hsp20 que nos animais. Essa diversificação gênica e funcional dos genes *heat shock*, nas plantas, pode ser uma consequência da biologia sésil das mesmas. As HSP20 vegetais são codificadas por famílias nucleares multigênicas e são localizadas em diferentes compartimentos celulares (WATERS et al., 1996). Em *Arabidopsis* ssp., 19 genes para Hsp20 foram categorizados em 12 subfamílias, em função da localização subcelular e homologia (SCHARF et al., 2001; MA et al., 2006; SIDDIQUE et. al., 2008; WATERS et. al., 2008). A principal característica das proteínas HSP20 é a presença de uma sequência de aminoácidos, evolutivamente conservada, contendo de 80-100 resíduos, chamada de domínio α -cristalino (α -*crystallin domain* – ACD), localizado na região C-terminal (CASHIKAR et. al., 2005). Além disso, a região à montante da sequência codificante das Hsp20 apresenta, comumente, várias repetições do domínio 5' nGAAnnTTCnnGAA n 3' (HSE, *heat shock element*), que são reconhecidas e ativadas pelos fatores de transcrição (FT) *heat shock factors* ou HSF. (SARKAR et al., 2009).

As HSP20 são chaperonas moleculares independentes de ATP que formam grandes complexos oligoméricos com proteínas desnaturases, de 200-800kDa e agem impedindo a ação das desnaturases, tanto em células procarióticas como em células eucarióticas (LEE e VIERLING, 2000; CASHIKAR et. al., 2005). Desempenham, ainda, função de auxílio a outras chaperonas, na formação da conformação nativa de cadeias polipeptídicas nascentes e reorganização da conformação nativa de proteínas desnaturadas (TOROK et al., 2001; CASHIKAR et. al., 2005).

As Hsp20s são frequentemente associadas à resposta de plantas ao estresse por calor, frio, metais pesados, espécies reativas de oxigênio e outros (SUN et al., 2002). Mais recentemente, também têm sido associadas à resposta à infecção

por nematóides (ESCOBAR *et al.*, 2003; FUGANTI *et al.*, 2010; van OOIJEN *et al.*, 2010; KANDOTH *et al.*, 2011), bactérias (MAIMBO *et al.*, 2007; GAROFALO *et al.*, 2009) e outros patógenos. Ainda não se conhece os fatores específicos que desencadeiam a expressão dessas proteínas durante o estresse biótico, mas as próprias alterações metabólicas decorrentes do ataque de patógenos podem gerar estímulos indutores similares aos encontrados nas vias de estresse abiótico (GRUNDLER e BOCKENHOFF, 1997; BARCALA *et al.*, 2008)

Dada a diversificação funcional das HSP20, essas proteínas são tidas como alvo ideais para auxiliar a geração de cultivares agrícolas tolerantes a uma série de condições de estresses abióticos e bióticos. Isso é especialmente importante na cultura da soja se considerarmos sua ampla distribuição no Brasil (CONAB, 2011). Até o momento, foi identificado um único gene Hsp20 de soja, cuja expressão varia em função da reação (resistência ou suscetibilidade) das plantas à infecção por *Meloidogyne javanica* (FUGANTI *et al.*, 2010). A caracterização da família Hsp20 em soja deverá auxiliar na identificação de outros genes candidatos potenciais para o desenvolvimento de cultivares mais produtivas sob condições desfavoráveis. Neste sentido, este trabalho teve como principal objetivo a caracterização molecular da família de genes Hsp20 em soja, com foco em sua distribuição no genoma, localização subcelular, filogenia e regulação frente a estresses bióticos e abióticos.

2 REVISÃO DE LITERATURA

2.1 SOJA: ASPECTOS GERAIS E ECONÔMICOS

A soja (*Glycine max*) é uma planta industrial por excelência, devido às suas características de cultivo e vantagens em relação a outras culturas, como a composição química, o que faz dela a *commodity* mais produzida e consumida na alimentação e indústria mundial (QUEIROZ *et al.*, 2006; CONAB, 2011).

Essa importância tem alicerce nos inúmeros subprodutos oriundos da soja, dentre os quais o óleo, amplamente utilizado no consumo humano, e o farelo, utilizado para o consumo animal, são os principais (OLIVEIRA *et al.*, 2009). O óleo, oriundo da primeira etapa de processamento dos grãos da soja, o esmagamento, pode ser do tipo comestível, empregado na fabricação de margarinas, óleos de cozinha, maionese, gorduras vegetais e produtos farmacêuticos, ou utilizado na indústria como biocombustível (OLIVEIRA *et al.*, 2009).

Ao longo dos anos, a cultura da soja apresentou uma evolução significativa no Brasil, tanto em produção quanto em produtividade, partindo de uma produção de menos de 9,8 milhões de toneladas em 1977/78, atingindo cerca de 75,32 milhões de toneladas em 2010/11 (CONAB, 2011). Tal avanço decorreu em parte pelo aumento da área destinada ao cultivo, mas o impacto significativo foi resultante, principalmente, das tecnologias geradas por Universidades e Institutos de Pesquisas. Esses estudos contribuíram para o desenvolvimento de cultivares superiores, geneticamente adaptadas a diferentes regiões edafo-climáticas, além de melhorias nas tecnologias e condições de cultivo. Diversas dessas cultivares foram desenvolvidas objetivando minimizar os efeitos dos diferentes estresses, bióticos e abióticos, que interferem na produtividade da cultura (POPP *et al.*, 2003).

No âmbito socioeconômico, a cadeia produtiva da soja garante ao Brasil mais de cinco milhões de empregos, diretos e indiretos, e participa com pelo menos 16% dos 35% do PIB (Produto Interno Bruto) gerado pela agroindústria (EMBRAPA, 2010). Atualmente, o país ocupa lugar de destaque no cenário mundial de produção de soja, como grande exportador, com valor total das exportações de US\$ 17,1 bilhões anuais, e segundo maior produtor. Em 2011, o Brasil respondeu por 28,44% da produção mundial de soja (EMBRAPA/USDA, 2011). Essa

leguminosa é o produto agropecuário líder na pauta de exportações nacionais (CARDOZO e PALMEIRA, 2006; EMBRAPA, 2011). Na safra 2010/11, o agronegócio da soja participou com 45,58% do total de grãos produzidos no Brasil (CONAB, 2011).

Na safra 2010/2011, a produtividade média da soja brasileira foi de 3.106 Kg/ha, chegando a alcançar cerca de 3.190 Kg/ha no Estado de Mato Grosso, o maior produtor brasileiro de soja, com produção de 20,4 milhões de toneladas, e área plantada de 6,4 milhões de hectares. O Paraná, segundo Estado produtor, com área plantada de 4,6 milhões de ha, responde pela produção de 15,4 milhões de toneladas, com produtividade média de 3.360 kg/ha (CONAB, 2011; EMBRAPA, 2011).

Apesar de todos esses aspectos positivos da sojicultura brasileira, há de se considerar que os níveis nacionais de produtividade hoje alcançados ainda podem e devem ser melhorados. A cultura ainda enfrenta desafios para crescer de modo competitivo e sustentável, para atender a demanda interna, conquistar e manter espaço no mercado externo (CARDOZO e PALMEIRA, 2006).

2.2 ESTRESSES ENFRENTADOS PELA CULTURA

O conceito de estresse é definido como todos os efeitos, sejam bióticos ou abióticos, que levam a uma série de mudanças nas características morfológicas, fisiológicas, bioquímicas e moleculares, tendo como consequência o dano às plantas (TAIZ *et al.*, 2009; WANG *et al.*, 2001).

Os fatores que desencadeiam estresse podem ser divididos em função de sua origem como bióticos e abióticos, e podem causar perdas importantes na produtividade das culturas em todo o mundo (KNIGHT e KNIGHT, 2001 *apud* SINGH *et al.*, 2002). Os fatores de estresse abiótico geralmente ganham maior importância com a expansão territorial das culturas, como é o caso da soja, cuja expansão tem levado ao aumento dos fatores de influência sobre a produtividade, destacando-se os edafo-climáticos.

O estresse abiótico é a principal causa de perdas na produção agrícola em todo o mundo, inclusive no Brasil, reduzindo os rendimentos em mais de 50%, principalmente nas mais importantes culturas (BRAY *et al.*, 2000 *apud* RODRÍGUEZ *et al.*, 2005).

Os estresses bióticos também apresentam elevado poder de dano às culturas e sua abrangência e intensidade dependem da região, genótipos utilizados e condições ambientais durante a safra. No Brasil, nos últimos anos, os principais fatores biológicos que vem causando perdas na cultura da soja de ampla abrangência e/ou intensidade, são as doenças causadas por nematóides e a Ferrugem Asiática da Soja (FUGANTI *et al.*, 2004; SANTANA, 2007; BETTIOL E MORANDI, 2009). Outras doenças, como oídio, podridão de *Phytophthora*, mancha alvo, etc. são menos expressivas e são de fácil controle, até o momento, com o emprego principalmente de variedades resistentes ou defensivos.

2.2.1 Estresses Abióticos

Inevitavelmente, as plantas precisam interagir com os fatores climáticos. Estresses ambientais, como a seca, salinidade, frio, calor e substâncias químicas oriundas das atividades humanas causam efeitos adversos, muitas vezes similares, sobre o crescimento das plantas e a produtividade das culturas (WANG *et al.*, 2003). Todos esses fatores podem agir simultaneamente e acabam por colocar a planta sob estresses combinados, causando lesão celular e a produção de estresses secundários, tais como os osmóticos e oxidativos (WANG *et al.*, 2003; MITTLER *et al.*, 2006).

Dentre os estresses abióticos citados, a temperatura é um dos fatores limitantes, mais importantes, para a produtividade e adaptação das culturas, especialmente quando os extremos de temperatura coincidem com os estágios críticos de desenvolvimento da planta. Assim, consideráveis perdas econômicas nas lavouras são causadas, anualmente, por extremos de temperaturas (LONG e ORT, 2010). As plantas são muitas vezes expostas, diária e sazonalmente, a grandes flutuações de temperatura (EFEOGLU, 2009; LONG e ORT, 2010). Essas mudanças bruscas, dependendo da região, podem ser frequentes no decorrer do dia e causar danos irreversíveis. O estresse por temperatura afeta o metabolismo e a biologia celular das plantas, especialmente as membranas celulares, e muitos processos fisiológicos básicos tais como a fotossíntese, a respiração e às relações hídricas (WAHID *et al.*, 2007)

A temperatura ótima de crescimento das plantas varia entre as espécies, mas está em uma faixa geral de 10°C a 40°C. Para a cultura da soja, as

condições ótimas de temperatura para o desenvolvimento estão no intervalo entre 20°C e 30°C. Qualquer temperatura acima ou abaixo desse intervalo gera estresse sobre a atividade metabólica das plantas (TRESHOW, 1970 *apud* EFEOGLU, 2009; FINKA *et al.*, 2010). O desenvolvimento das plantas ocorre através de reações bioquímicas, que são realizadas e controladas por enzimas. Falando a nível molecular, a atividade das enzimas, além do equilíbrio e a velocidade das reações relacionadas a elas, é diretamente dependente da temperatura, acordo com a constante de Michaelis-Menten (Km). Assim, atividade nula ou muito baixa é encontrada quando as enzimas estão submetidas a temperaturas abaixo ou muito acima da ótima para sua atividade (NELSON e COX, 2010). Nas plantas, a baixa temperatura pode resultar, imediatamente, em restrições mecânicas, mudanças nas atividades de macromoléculas e na redução do potencial osmótico celular (XIONG *et al.*, 2002).

A velocidade das reações bioquímicas aumenta à medida que há aumento na temperatura. Contudo, temperaturas acima da faixa limite, 37°C, diminuem gradativamente a velocidade das reações catalisadas pelas enzimas. A alta temperatura influencia na integridade da estrutura terciária desses catalizadores, causando sua desnaturação ou mesmo inativação. Quando uma enzima é inativada, o processo ao qual ela está relacionada é reprimido ou reduzido (NELSON e COX, 2010, FINKA *et al.*, 2010).

Nas plantas, a reação de fotossíntese é uma importante via afetada pela temperatura ambiental. A fotossíntese é realizada na faixa entre 10 e 30°C em plantas de clima temperado. Sob temperaturas abaixo dessa faixa, a fotossíntese pode ser limitada pela disponibilidade de fosfato no cloroplasto. À medida que a temperatura aumenta a taxa de fotossíntese aumenta, mas quando as temperaturas ótimas de reação são ultrapassadas, a taxa fotossintética volta a decrescer e o sistema de transporte fotossintético de elétrons é afetado (TAIZ *et al.*, 2009; TRESHOW, 1970 *apud* EFEOGLU, 2009). Uma das estratégias realizadas pelas plantas, em resposta às altas temperaturas, é o fechamento dos estômatos para diminuir a taxa de evapotranspiração de água, mas que tem como consequência a diminuição nas trocas gasosas influenciando nas taxas fotossintéticas e fotorespiratória (TAIZ *et al.*, 2009).

Estudos sobre a termotolerância demonstram que um complexo de vias pode regular os diversos processos de resposta, tais como os fitormônais

através do ácido salicílico e ácido abscísico (HUA, 2009). O estresse de temperatura também pode modificar drasticamente as propriedades físicas das membranas biológicas, pelo aumento da degeneração oxidativa dos lipídios (DASH *et al.*, 2002; MATOS *et al.*, 2007). A extraordinária sensibilidade membranas biológicas a estímulos ambientais das tem sido demonstrada em diversos organismos. As membranas podem adaptar sua composição lipídica, até certo ponto, em resposta as condições ambientais (SAIDI *et al.*, 2010). Evidências comprovam que dispositivos sensoriais presentes nas membranas são capazes de detectar sinais específicos e coordenar níveis de expressão de conjuntos de genes-chave (LOS e MURATA, 2004). O estado físico dos lípidos da membrana plasmática também pode regular a atividade de proteínas membranares, como os canais de íons, translocadores de moléculas pequenas, aquaporinas e receptores quinase-associados (HOHMANN, 2003).

Atualmente, os avanços na área de pesquisa sobre estresses abióticos tem foco na obtenção de variedades tolerantes à esses estresses. As estratégias em grande parte estão embasadas principalmente na expressão de genes que codificam enzimas que metabolizam ROS (reactive oxygen species), genes para proteínas de choque térmico (HSPs), e genes que codificam fatores de transcrição, como DREB (Dehydration responsive element binding) 1A e 2A, relacionados à resposta seca, estresse osmótico e calor. Para a soja, as pesquisas atuais voltadas para o desenvolvimento de cultivares tolerantes ao estresse abiótico, têm como foco principal a seca, embora outros estresses também estejam sendo estudados, como temperatura e salinidade (FRAIRE-VELÁZQUEZ *et al.*, 2011).

2.2.2 Estresses Bióticos

Além do estresse abiótico, as plantas como integrantes de sistemas biológicos sofrem o ataque de pragas e diferentes patógenos. Para a soja, já foram descritas mais de 50 doenças provocadas por fitopatógenos (fungos, nematóides, bactérias e vírus), que infectam as plantas de soja em diferentes estágios de desenvolvimento (BONATO, 2000; YORINORI, 2000).

Cerca de 15% a 20% das perdas anuais na produção brasileira de soja são devido às doenças, sendo algumas capazes de ocasionar perdas de quase 100% (EMBRAPA, 2007). Dentre estas, as causadas pelo ataque de nematóides

parasitos de raízes, pertencentes a diferentes gêneros, tais como *Meloidogyne* spp., *Heterodera* spp., *Pratylenchus* spp. e *Rotylenchus* spp. e outros, contribuem para a queda do rendimento da sojicultura, especialmente nas regiões tropicais e subtropicais (DIAS et. al., 2010).

Estima-se que 10,6% das perdas anuais da produção agrícola mundial, são causadas pelo parasitismo desses organismos (BARKER e KOENNING, 1998), em torno de US\$ 125 bilhões anualmente apenas pela ação dos fitonematóides endoparasitas sedentários (CHITWOOD, 2003). Esses fitopatógenos causam não somente danos diretos às plantas, reduzindo o crescimento e/ou tornando-as improdutivas, mas também indiretos pela sua interação com outros organismos, como fungos e bactérias, que tem sua entrada facilitada pelo sistema radicular danificado (FERRAZ et. al., 2001).

2.3 NEMATÓIDES NA CULTURA DA SOJA

Os nematóides, também denominados “vermes”, pertencentes ao filo Nematoda, possuem morfologia arredondada não-segmentada e simetria bilateral (DECKER E SVESHNIKOVA, 1989 *apud* PAZHAVARICAL, 2009). Acredita-se que o filo Nematoda tenha surgido antes do Cambriano (~750-650 Ma) o que contribuiu para sua ampla diversificação. Esses organismos exibem uma excelente adaptabilidade aos mais diferentes nichos ecológicos, desde regiões polares geladas a desertos, de vales de montanhas, e até mesmo água salgada e doce, o que os coloca como bem sucedidos evolutivamente (DE LEY e BLAXTER, 2002; WEISCHER E BROWN, 2000 *apud* PAZHAVARICAL, 2009).

Os nematóides variam muito em tamanho, apresentando de menos de 100 µm a alguns centímetros. A grande maioria é de vida-livre, alimentando-se de bactérias, protistas, fungos, de outros nematóides e de matéria orgânica do solo, enquanto que alguns precisam de outros organismos vivos, como animais ou plantas para completar seu ciclo de vida (parasitas). Centenas de espécies de nematóides são de grande importância para o ser humano, pois parasitam animais vertebrados ou invertebrados (incluindo humanos) ou plantas (PARKINSON *et al.*, 2004; SCHOLL e BIRD, 2005).

Alguns deles apresentam elevada importância econômica por causarem severos danos às plantas. Diversas espécies são capazes de parasitar a

soja em todo o mundo, sendo que mais de 100 espécies de nematóides já foram descritas, envolvendo cerca de 50 gêneros. No Brasil, os gêneros de nematóides mais prejudiciais à cultura têm sido os formadores de cisto (*Heterodera glycines*), os formadores de galhas (*Meloidogyne* spp.), o causador de lesão radicular (*Pratylenchus brachyurus*) e o reniforme (*Rotylenchulus reniformis*) (NICKLE, 1991 *apud* PAZHAVARICAL, 2009). O estresse provocado diretamente pela invasão do parasita é ainda aumentado por fatores secundários decorrentes de sua instalação tais como o estresse hídrico, sintomas de deficiências minerais (DIAS *et al.*, 2010) e favorecimento da instalação de outros agentes parasitas, como fungos, vírus e bactérias.

O nematóide reniforme, *R. reniformes*, parasito da soja em praticamente todos os países produtores do mundo, apresenta danos variáveis em função de fatores bióticos e abióticos. O nematóide de cisto da soja, *H. glycines* possui grande importância econômica em países como a China, Japão e Estados Unidos, bem como a Argentina e Brasil. Com base na média anual de produtividade da sojicultura americana, os cálculos de perdas devido ao parasitismo do nematóide de cisto para os anos de 2006 a 2009 resultaram em um prejuízo avaliado em 1,286 bilhões de dólares (KOENNING e WRATHER, 2010 *apud* KANDOTH *et al.*, 2011). Efeitos depressivos na produtividade em lavouras de soja brasileira, mesmo sem sintomas aparentes de danos, têm sido relatados nos estados de São Paulo, Paraná e Rio Grande do Sul. Os prejuízos podem atingir cerca de 400 kg/ha em lavouras formadas por cultivares suscetíveis ao *H. glycines* (GARCIA *et al.*, 2005 *apud* DIAS *et al.*, 2009).

O parasitismo por *Pratylenchus brachyurus* figura até o momento em segundo lugar em termos de importância para a cultura da soja no Brasil, o que tornam as atuais e crescentes pesquisas referentes a tal grupo de nematóides essenciais (FERRAZ, 2006). Recentemente, a ocorrência de *P. brachyurus* tem ganhado importância, tanto pelos danos à cultura quanto pela ampla disseminação e alta incidência do patógeno em áreas produtoras do Estado de Mato Grosso, região importante no cultivo da soja.

Os nematóides formadores de galha, *Meloidogyne* spp. são uma constante preocupação aos sojicultores do Brasil, há pelo menos 50 anos, por causarem elevadas perdas diretas, por redução da produção, e indiretas por limitação ao uso agrícola das áreas infestadas (FERRAZ *et al.* 2001). Segundo

levantamento realizado por Wrather e colaboradores (1997), os danos causados pelos nematóides à cultura da soja podem atingir 15% da produção, o que, em termos globais representaria uma perda de cerca de 78 milhões de dólares. As meloidoginoses apresentam uma distribuição populacional abrangente em áreas, como o norte do Rio Grande do Sul, sudoeste e norte do Paraná, sul e norte de São Paulo e sul do Triângulo Mineiro. Atualmente, também são um crescente problema em lavouras do Mato Grosso do Sul e Goiás na região central do país (ÁVILA *et al.*, 2007).

2.3.1 Aspectos Gerais Relacionados ao Gênero *Meloidogyne* – *M. javanica*

As primeiras observações de campo das meloidoginoses da soja no Brasil tiveram início na década de 50, com estudos da eficiência do nematicida *Downfume* W-10 no controle dos nematóides ocorrentes na região paulista de Campinas. Depois de diversos outros estudos, as publicações de Lordello em 1954, descreveram as espécies *M. javanica* e *M. incognita* como as principais em importância no país (FERRAZ *et. al.* 2001).

A partir dos anos 70, intensificaram-se os estudos objetivando o levantamento das ocorrências do nematóide de galha nas mais diferentes regiões e a avaliação de técnicas de controle. Dez anos depois iniciaram-se os trabalhos na linha de melhoramento genético, visando à obtenção de genótipos resistentes. Desde então, têm-se trabalhos cada vez mais elaborados, envolvendo métodos avançados de genética e biologia molecular, vêm permitindo um melhor conhecimento das interações *Meloidogyne*/soja. (FERRAZ *et. al.* 2001).

O gênero *Meloidogyne* é representado por noventa espécies (CARNEIRO *et al.*, 2008). Até o presente momento, sabe-se da ocorrência de várias espécies pertencentes ao gênero *Meloidogyne* no Brasil, como *M. coffeicola*, *M. exigua*, *M. incognita*, *M. paranaensis*, *M. javanica* e outros. No entanto, *M. javanica* e *M. incognita* representam o maior problema para a cultura da soja, principalmente, no Norte do Rio Grande do Sul, Sudoeste e Norte do Paraná, Sul e Norte de São Paulo e Sul do Triângulo Mineiro, e na região Central do país onde vários focos têm sido detectados. A ocorrência de nematóides é mais comum nas regiões subtropicais e tropicais sendo possível dizer que a expansão da cultura, inicialmente

introduzida nas regiões temperadas, certamente vem alterando a realidade das meloidoginoses na soja (FERRAZ et. al., 2001).

O gênero *Meloidogyne* é cosmopolita e tem ampla gama de hospedeiros, incluindo monocotiledôneas e dicotiledôneas, plantas cultivadas (perenes, semi-perenes e anuais) e daninhas, pertencentes a diferentes famílias botânicas. Sua capacidade reprodutiva varia nas tantas e tão distintas espécies vegetais (FERRAZ, et. al. 2001). Contudo, pode-se dizer que quatro características biológicas adaptativas são responsáveis pelo sucesso do parasita na infecção e estabelecimento nos mais diferentes ambientes. A primeira consiste na elaborada estratégia de parasitismo, pela indução à transformação das células das raízes da planta hospedeira em tecido nutridor diferenciado, que permite seu desenvolvido e reprodução. A segunda característica é definida pela morfologia das fêmeas de *Meloidogyne*, com forma corpórea arredonda (Figura 1) e comportamento sedentário, possibilitando conter dois ovários bem desenvolvidos para formar grande quantidade de ovos (em média 400 ovos) ao longo de período variável de quatro a seis semanas sob condições favoráveis (FERRAZ, et. al. 2001). Outra vantagem adaptativa é que os ovos ficam aglomerados em uma massa gelatinosa junto aos corpos das fêmeas, interna ou externamente às raízes, protegidos em meio à substância gelatinosa produzidas pelas próprias fêmeas. Por fim, no gênero *Meloidogyne* podem ocorrer diferentes modalidades reprodutivas, que vão da reprodução sexuada a reprodução assexuada por partenogênese meiótica (dependente de macho) ou mitótica (independente de macho). Nessa última condição, a formação de ovos independente da presença de machos representa uma vantagem apreciável pela economia de energia que seria despendida a procura de um parceiro sexual (FERRAZ, et. al. 2001).

Figura 1 - Micrografia eletrônica de varredura de um de nematóide de galhas-fêmea (parasita de plantas), *Meloidogyne arenaria*.



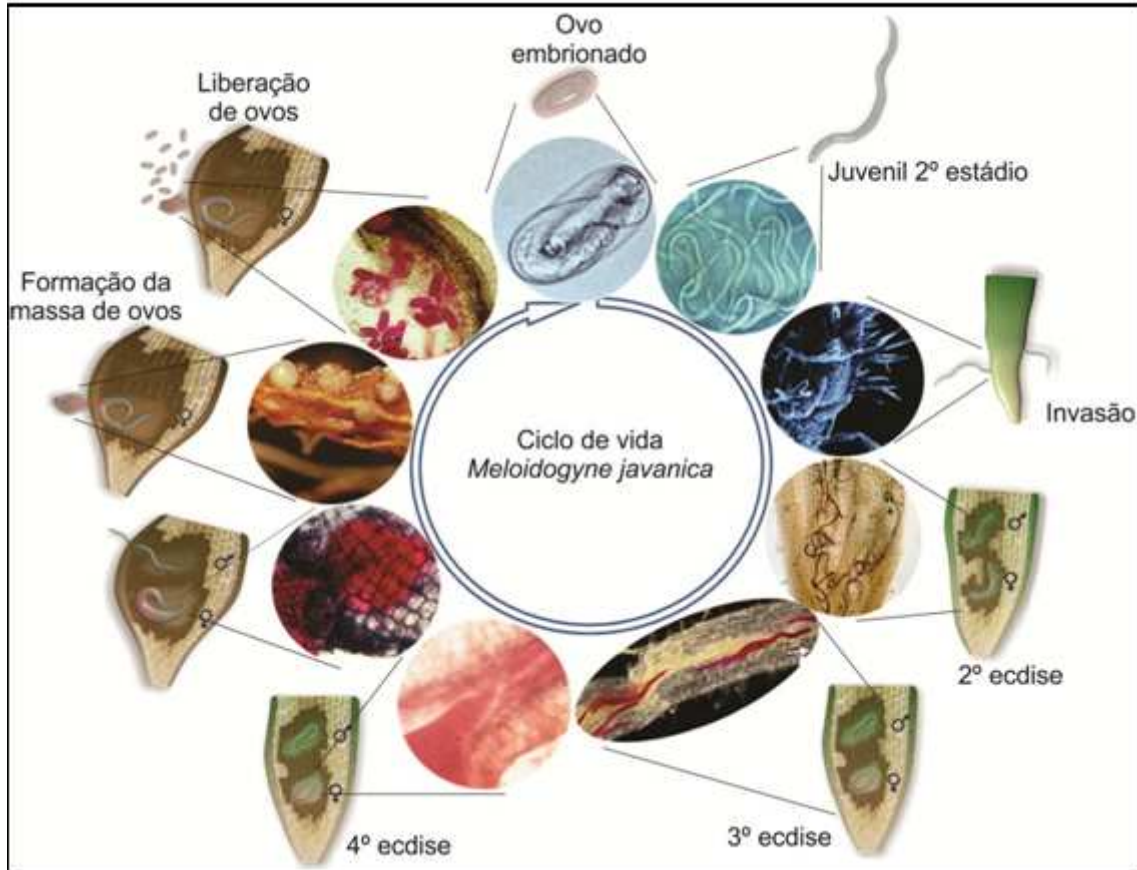
Fonte: Foto por Eisenback JD (Virginia Polytechnic Institute e State University, Blacksburg, Virginia, EUA) (BLUMENTHAL e DAVIS, 2004).

As espécies deste gênero apresentam, de forma generalizada, o ciclo de vida de 22 a 25 dias, aproximadamente, dependendo da temperatura do solo. As temperaturas ótimas estão entre 25 e 30°C, tendo sua atividade metabólica reduzida em extremos de temperatura (≥ 40 e ≤ 5 °C) (CHRISTIE, 1976).

O ciclo de vida consiste de seis estádios de desenvolvimento: ovo, quatro fases larvais (juvenis) e adulto (Figura 2). Ainda dentro dos ovos os juvenis sofrem a primeira ecdise transformando-se em juvenis de segundo estágio (J_2) ou fase infectiva, que eclodem de seus ovos e passam a migrar no solo (não se alimentam até o estabelecimento do parasitismo, e suas reservas energéticas são corpúsculos de lipídeos) até a penetração na raiz, que se dá na base da coifa, na zona meristemática da planta hospedeira. A coifa é encontrada por quimiotaxismo, uma vez que possuem apurado sistema sensorial que percebe exudados vegetais. Após a penetração nas raízes, os patógenos migram pelo cilindro vascular, pelo emprego de força mecânica do estilete e degradação enzimática da parede celular e lamela média, pela atividade de celulasas, pectinases e proteases, para formação do sítio de alimentação para o fitopatógeno, as células gigantes (GHEYSEN e FENOLL, 2002; ABAD *et al.*, 2003).

O sexo do nematóide é definido no quarto estágio de desenvolvimento, juvenis - J_4 , alcançado após ecdises do estágio J_2 para J_3 , e desse para J_4 . A definição do sexo sofre influência das condições de interação molecular planta-nematóide. Quando favorável, ocorre diferenciação para fêmea adulta, e quando em condições desfavoráveis de parasitismo, J_4 se diferencia em macho adulto, vermiforme, que abandona o hospedeiro (WILLIAMSON e GLEASON, 2003).

Figura 2 - Ciclo de vida do fitonematóide endoparasita sedentário. Os juvenis J_2 migram entre as células vegetais, facilitada pela digestão da lamela média, e modificam a fisiologia celular para a formação de células gigantes e galhas. As fêmeas adultas depositam seus ovos em massa protéica.



Fonte: (Modificado de <http://www.nematoides.com.br/wordpress/index.php/sobre-nematoides/genero-Meloidogyne/>).

A infecção pelos nematóides de galhas na raiz implica no desenvolvimento de células especializadas para alimentação, no cilindro vascular do periciclo, endoderme e áreas do córtex, chamadas células gigantes ou nutridoras. Essas células são muito grandes, com citoplasma denso e granuloso (indicativo de um metabolismo ativo), múltiplos núcleos poliplóides, e uma espessa parede celular com inúmeras invaginações (BARCALA et. al.; 2010). Os sintomas são provocados em função da penetração dos juvenis de segundo estágio (J_2), em células que ainda não estão diferenciadas, e pela injeção de secreções produzidas pelas glândulas esofagianas do nematóide. Tais secreções contêm substâncias que atuam como fitormônios e incitam modificações no ciclo celular, no balanço hormonal, na sinalização celular, na expressão gênica e, conseqüentemente, afetam

a fisiologia da raiz e induzem a formação de múltiplos sítios de alimentação (FERRAZ, et. al., 2001; BIRD *et al.*, 2003;).

Outro dano distinto, mas também decorrente do mesmo evento, a infecção, é a incitação de reações hiperplásicas e hipertróficas de células na região parasitada, fazendo com que as raízes tenham uma aparência engrossada. O processo culmina nas chamadas galhas na raiz (Figura 3), que podem levar a compressão de elementos de vasos do xilema, resultando na desorganização do cilindro vascular, e conseqüentemente, a interferência na absorção e transporte interno de solutos e fotoassimilados (CARNEIRO *et al.*, 1999).

Figura 3 - Formação de galha. As raízes de soja tornam-se engrossadas na região do parasitismo como resultado do processo de infecção pelo nematóide *M. javanica*, que provoca reações de hiperplasia e hipertrofia nos tecidos.



Fonte: (<http://www.nematoides.com.br/wordpress/index.php/sobre-nematoides/genero-Meloidogyne/>).

A presença de galhas nas raízes são a máxima expressão sintomatológica do parasitismo pelas espécies do gênero *Meloidogyne* spp., que além das alterações de ordem estruturais e morfológicas das plantas, culminam em drásticas mudanças funcionais. Assim, nas lavouras de soja onde há espécimes de *Meloidogyne*, observam-se a formação de clareiras de plantas com porte reduzido, clorose, manchas necróticas, de crescimento lento e desigual e que murcham nas horas mais quentes do dia, normalmente dispostas em manchas circulares (reboleiras). (COSTAMILAN, 2000; FERRAZ, et. al., 2001; EMBRAPA 2007).

2.4 SOBREPOSIÇÃO DE VIAS DE SINALIZAÇÃO EM RESPOSTA AO ESTRESSE ABIÓTICO E BIÓTICO EM PLANTAS

No decorrer da evolução, as plantas desenvolveram estratégias de adaptação e superação a diferentes tipos de estresse abióticos e bióticos (FRAIRE-VELÁZQUEZ *et al.*, 2011).

A percepção de estresses e a transdução de sinais são as etapas iniciais na determinação da resposta da planta. Esses sinais estão relacionados à ativação de canais iônicos e cascatas de sinalização das quais fazem parte as quinases, com a produção de espécies reativas de oxigênio (ROS) e o acúmulo dos hormônios ácido salicílico (SA), etileno (ET), ácido jasmônico (JA) e ácido abscísico (ABA). Em consequência, há uma série de alterações simultâneas ao nível transcricional, que resultam nas respostas metabólicas de defesa (SINGH *et al.*, 2002; FRAIRE-VELÁZQUEZ *et al.*, 2011). O complexo controle transcricional da expressão de genes responsivos ao estresse é categórico para a defesa da planta. O mesmo pode ser ativado por diferentes formas de estresse ambiental ou por patógenos (genes de resistência), mas culminam na sobreposição de vias (FRAIRE-VELÁZQUEZ *et al.*, 2011).

O genoma dos eucariotos é constituído, em sua maior parte, por regiões reguladoras da função gênica, que atuam de forma conjunta para determinar uma infinidade de fenótipos possíveis, a partir de um número de genes muitas vezes menor (SINGH *et al.*, 2002). Tais regiões reguladoras possuem sítios de ligação de proteínas, um aparato cerne, que ativam a expressão basal do gene. No caso de genes envolvidos em respostas a situações específicas ou anormais, essas regiões contam com aparato melhor elaborado, para ativação em níveis diferentes. Os promotores e os fatores de transcrição que se ligam a eles, são capazes de controlar a resposta em situações de estresses, por múltiplas vias de sinalização (SINGH *et al.*, 2002; FRAIRE-VELÁZQUEZ *et al.*, 2011).

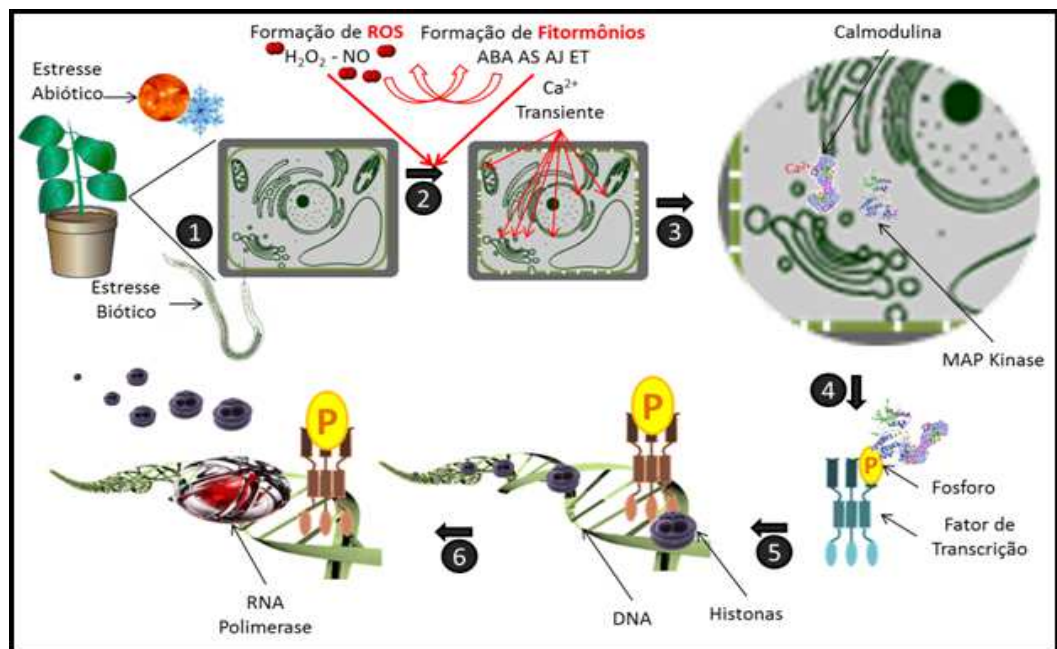
Dentro do conjunto de genes induzidos e reprimidos, em resposta aos mais variados estresses, podem ser encontrados: genes cujas proteínas agem diretamente no metabolismo da planta contra os estresses; e genes cujos transcritos ou proteínas por eles codificados regulam a expressão e a sinalização gênica, como por exemplo, os fatores de transcrição (ZHU, 2002 *apud* ZHU, 2006).

A regulação da expressão gênica nos seres vivos pode ocorrer tanto ao nível transcricional quanto pós-transcricional, sendo o primeiro o principal ponto de regulação. Mas o produto da expressão gênica, seja ele um transcrito ou uma proteína, é apenas parte de uma via de sinalização que, em geral, requer a participação de outros componentes tais como sinalizadores iônicos, hormônios não protéicos e outros. As vias de sinalização nas plantas são geralmente estudadas, utilizando técnicas que permitem acesso global ao conjunto de genes expressos (transcritoma), em resposta a uma determinada condição.

Muitos estudos vêm demonstrando que os padrões de expressão de genes em resposta a diferentes tipos de estresses, mesmo quando ativados por diferentes vias, são regulados em algum momento por uma sobreposição de sinais (SINGH et al, 2002). Diversas investigações têm demonstrado que espécies reativas de oxigênio e íons de Cálcio (Ca^{2+} - forma mais comum na natureza) são mensageiros secundários na resposta inicial aos estresses abióticos e bióticos. Por exemplo, os níveis de Ca^{2+} citosólico aumentam nas células vegetais, em resposta a diversas condições ambientais inóspitas, incluindo o ataque de patógenos e estresses osmótico, hídrico, por frio e ferimentos (FRAIRE-VELÁZQUEZ et al., 2011).

O acúmulo de Ca^{2+} no espaço intracelular sinaliza simultaneamente para a indução de inúmeras vias bioquímicas, que fazem parte da resposta de defesa da planta, entre as quais estão as reguladas pela atividade das proteínas MAP (*mitogen-activated protein*), quinases Ca^{2+} dependentes e calmodulinas (FRAIRE-VELÁZQUEZ et al., 2011). Estudos sobre a relação da concentração de Ca^{2+} , mediando a atividade das calmodulinas, tentam explicar a hipótese, ainda não confirmada, da fosforilação e a consequente ativação de fatores de transcrição (TFs) (LIU et al., 2008; ZHANG et al., 2009). A ativação dos TFs, pela fosforilação, possibilita ligação dos mesmos a sequências específicas na região promotora de genes e aumenta a expressão destes pela exclusão de histonas (Figura 4) (GELLI et al., 1999).

Figura 4 - Diagrama esquemático ilustrando os principais componentes de sinalização e ativação de resposta da planta ao estresse biótico e abiótico. Os estresses de elevação/diminuição da temperatura e ataque de patógenos (1) provocam aumento na fluidez da membrana e permite um fluxo intracelular de íons de Ca^{2+} (2). Esses mesmos agentes de estresse desencadeiam nas células, por diferentes vias, a formação de espécies reativas de oxigênio (ROS) e metabólitos onde a ROS é precursora que tem efeito oxidativo nos lipídeos da membrana, levando ao mesmo processo de ativação de canais de cálcio (seta vertical vermelha). A maior concentração de Ca^{2+} ativa as calmodulinas (CaM), que através da interação com as proteínas quinases (CBK) (3), promove a fosforilação (P) dos fatores de transcrição, como por exemplo os HSFs (4). TFs ativados se ligam aos promotores de genes responsivos a estresse, nesse caso genes codificadores para HSP20 (5), causando o deslocamento das histonas e facilitando o acesso da RNA polimerase II (6). Como consequência, há alterações nos níveis de expressão dos genes Hsp20 as quais participam das respostas celulares e fisiológicas, realizando assim a manutenção da homeostase do organismo vegetal.



Aparentemente, a sinalização do estresse de calor está diretamente relacionada com a concentração de íons de Ca^{2+} no interior da célula, para ativação das calmodulinas e/ou MAP quinases e posterior fosforilação dos HSFs (*Heat Transcription Factors*). Segundo Sangwan e colaboradores (2002), o fluxo de Ca^{2+} tem se mostrado essencial na ativação de proteínas (MAP) quinases, sob estresse de calor. A permeabilidade de íons através da membrana plasmática é profundamente afetada pelo estresse de calor, de forma que a concentração de cálcio livre no citoplasma pode ser rapidamente alterada em função do estresse e,

assim, sinalizar para a indução de proteínas de resposta a choque térmico (SWINDELL *et al.*, 2007). Segundo Xiong e colaboradores (2002), quando a planta é submetida a baixas temperaturas ocorre influxo de Ca^{2+} através da membrana plasmática e seu acúmulo no interior celular, possibilitando a aclimatação da mesma ao ambiente.

A ativação de quinases, pela alteração da concentração de Ca^{2+} durante o estresse biótico, tem sido relatada. Os resultados de pesquisas recentes sugerem que o ataque de insetos herbívoros desencadeie a ativação de vias relacionadas à quinases, através do fluxo de Ca^{2+} (Arimura e Maffei, 2010). Estudos com tomates resistentes ao fitopatógeno *Fusarium oxysporum* demonstram que logo após o reconhecimento do gene de avirulência/elicitador do patógeno, é desencadeada uma resposta de defesa com acúmulo significativo nas concentrações de cálcio intracelular, com ativação de proteínas de fosforilação Ca^{2+} dependentes e geração de ROS (DEY *et al.*, 2010). Ainda, segundo o mesmo autor, o fluxo intracelular de cálcio é essencial para a regulação do mecanismo de sinalização de defesa de plantas de tomate ao patógeno *Fusarium oxysporu*.

Sabe-se também que as plantas, durante os estresses abióticos ou após o reconhecimento de patógenos pelos genes R (resistência), respondem com um acúmulo de espécies reativas de oxigênio (FRAIRE-VELÁZQUEZ *et al.*, 2011). Além dos íons cálcio, é bem conhecido também que o peróxido de hidrogênio (H_2O_2) e óxido nítrico (NO) funcionam como mensageiros secundários em plantas sob estresses biótico e abiótico (NEILL *et al.*, 2002; GOULD *et al.*, 2003). A produção de H_2O_2 em plantas submetidas à elevadas temperaturas é bem conhecida. Segundo Saidi e colaboradores (2011), recentemente foi estabelecido que a produção de H_2O_2 contribui para a tradução do sinal em plantas submetidas a choque de calor, induzindo a expressão de HSPs (*Heat Shock Proteins*). Os estresses causados pelo calor, frio ou seca, resultam no aumento da produção de ROS pelo aparato fotossintético (MITTLER, 2006). Em *Arabidopsis*, estudos utilizando mutantes deficientes para a síntese de NO, resultaram na ausência ou ineficiência de resposta de termotolerância. A partir desses resultados foi possível concluir que, nas plantas selvagens, a termotolerância é alcançada por meio da sinalização NO dependente, decorrente do estresse por choque térmico (XUAN *et al.*, 2010). Ainda, foi demonstrado que a expressão de uma Hsp20 em *Arabidopsis* (AtHsp18.2) foi alterada em função da disponibilidade de NO celular, sendo inibida no mutante

deficiente para a produção de NO e induzida no mesmo mutante tratado com um doador de NO (XUAN *et al.*, 2010).

Em situação de estresse biótico, as vias iniciais de sinalização para resposta de defesa da planta a patógenos estão claramente associadas à abertura de canais de iônicos de Ca^{2+} na membrana plasmática, e produção de espécies reativas de oxigênio (DEY *et al.*, 2010). A resposta de hipersensibilidade das plantas contra o ataque de patógenos leva a formação de ROS, principalmente pela ação de NADPH oxidases na membrana plasmática, causando o aumento na concentração de ROS no citoplasma. Tais espécies reativas de oxigênio também são formadas nas mitocôndrias e cloroplastos. Espécies reativas de oxigênio podem ser formadas durante a fosforilação oxidativa, com a passagem de um elétron da ubiquinona para O_2 , tornando esse reduzido. O radical superóxido O_2^- livre na célula é altamente reativo e danoso; e após sua formação, são rapidamente convertidos a peróxido de hidrogênio (H_2O_2), pela enzima superóxido dismutase (SOD) (NELSON e COX, 2010). Apesar de causar danos em altas concentrações, as ROS estão envolvidas em várias vias de proteção e regulam o nível de expressão de genes de defesa (NEILL *et al.*, 2002). Evidências recentes demonstraram que a NADPH oxidase pode ser ativada por sinais de alteração na concentração de Ca^{2+} (FRAIRE-VELÁZQUEZ *et al.*, 2011).

As vias metabólicas ativadas nas plantas em resposta a fatores de estresse abióticos e bióticos ainda não são totalmente compreendidas, e pouco se sabe sobre os pontos de convergência de sinais que integram e sobrepõem essas respostas. Aparentemente, as concentrações de ânions de Ca^{2+} no interior da célula e a formação de ROS funcionam como mensageiros secundários na modulação da atividade de proteínas específicas. Todavia, estão presentes em todos os tipos de resposta das plantas ao estresse. Entre essas proteínas específicas, as quinases e os fatores de transcrição estão entre os mais relatados (FRAIRE-VELÁZQUEZ *et al.*, 2011).

As vias moleculares que levam a expressão das Hsp (*heat shock proteins*), por exemplo, não são totalmente compreendidas (SWINDELL *et al.*, 2007). Entretanto, a percepção do choque térmico está acoplada a múltiplas vias de transdução de sinais que levam à ativação de HSFs (*Heat shock transcription factors*), os quais por sua vez reconhecem *cis* elementos especialmente presentes

nos genes que respondem a choque térmico, induzindo sua expressão (SWINDELL *et al.*, 2007).

Independentemente do mecanismo de ação das calmodulinas ou Map quinases sobre a ativação dos HSFs, é fato que a presença de tais proteínas na célula é fundamental para a expressão dos genes Hsps. Estudos de estresse de calor utilizando plantas de *Arabidopsis* CaM3 (calmodulinas) mutantes e superexpressando o gene, demonstram que essa proteína é essencial para a resposta ao estresse. O nocaute do gene calmodulina 3 em *Arabidopsis* levou a uma clara redução na termotolerância sob tratamento de choque térmico a 45 °C durante 50 minutos. A superexpressão dessa proteína, ou silenciamento, em comparação ao tipo selvagem, resultou no aumento e na redução da termotolerância, respectivamente (ZHANG *et al.*, 2009). Liu e colaboradores (2008), analisando a resposta de ativação dos HSFs e o acúmulo de Hsps em plantas de *Arabidopsis* *cbk3* (*calmodulinas binding quinase*) mutantes, observaram que mesmo com o aumento da temperatura não houve resposta ao estresse, o que é normalmente observado em plantas não mutantes.

Gould e colaboradores (2003) e Konigshofer e colaboradores (2008) demonstraram que os HSFs podem ser ativados tanto por NO quanto por H₂O₂, por vias ainda não bem elucidadas. No caso das proteínas de choque térmico isso é especialmente importante porque elas desempenham um papel excepcionalmente amplo nos processos celulares, respondendo a estresses térmicos e não térmicos, incluindo os estresses bióticos. Atualmente é sabido que as HSPs são funcionalmente diversificadas e podem responder a estresses específicos, como por exemplo, o gene Hsp18.0-CI de arroz é induzido em resposta ao frio e ao calor, o Hsp18.6-CIII é induzido em resposta a salinidade, a desidratação e ao calor, já o Hsp17.4-CI responde a diferentes tipos de estresse, sendo induzido em condições de alta salinidade, desidratação, calor, e também em função da inoculação com *Magnaporthe grisea* (SARKAR *et al.*, 2009). Como os HSFs são fosforilados e ativados durante o choque térmico ainda não está bem esclarecido, mas provavelmente exista uma relação com o aumento na atividade de calmodulinas e MAP quinases (SAIDI *et al.*, 2011).

2.5 AS HSPs E OS FATORES DE TRANSCRIÇÃO DE CHOQUE TÉRMICO

As plantas não são capazes de evitar a exposição a fatores adversos devido a sua natureza sésil, mas podem suplantá-los através da evolução de vários mecanismos morfológicos e fisiológicos, que lhes permitem sobreviver ao estresse. Todos os tipos de estresses por uma via ou outra acabam causando disfunção em proteínas e enzimas, e que em algum momento levam à indução de um grupo de proteínas chamadas proteínas de choque térmico (HSPs) (AL-WHAIBI, 2010).

As proteínas HSPs foram inicialmente identificadas em *Drosophila melanogaster*, em resposta ao choque térmico (RITOSSA, 1962 *apud* CASHIKAR et. al., 2005). Essas proteínas estão agrupadas em cinco subfamílias conservadas, divididas em função de seu peso molecular aproximado: (1) HSP70/DnaK, (2) HSP90/HtpG, (3) HSP100/ClpB, (4) HSP60/GroE e (5) HSP20 ou pequenas proteínas de choque térmico (small heat shock proteins - sHsps) de 15 – 42 kDa (MOSELEY, 1997; SARKAR et. al, 2009).

A complexidade das famílias de genes HSP em plantas tem dificultado a definição de suas funções durante o estresse, porque muitos membros da família também são essenciais para o crescimento e desenvolvimento, mesmo em situações normais. As HSPs possuem um papel importante na biogênese de proteínas celulares, auxiliando no dobramento e “controle de qualidade” (KOTAK et al., 2007). Até o momento, sabe-se que as HSPs funcionam como chaperonas moleculares que contribuem para a sobrevivência das plantas durante o estresse (FINKA et al., 2010). Em relação à função durante estresse térmico, as investigações das HSP100 e HSP20 estão mais avançadas (KOTAK et al., 2007).

As chaperonas moleculares participam no dobramento de cadeias polipeptídicas nascentes e no redobramento de proteínas desnaturadas durante condições de estresse. Podem prevenir a agregação irreversível de proteínas cujas superfícies hidrofóbicas estão expostas pela desnaturação, e contribuem, portanto, para a homeostase celular (AL-WHAIBI, 2010). Elas reconhecem e se ligam a proteínas quando estas não estão localizadas nos compartimentos celulares apropriados. As cadeias polipeptídicas recém-sintetizadas podem dobrar-se espontaneamente sem o auxílio de chaperonas moleculares, mas algumas proteínas de alto peso molecular e complexos multi-proteicos requerem o auxílio de

chaperonas moleculares e co-chaperonas para alcançar sua conformação final e estabilidade (CHEN e SHIMAMOTO, 2011). Outra informação relevante é que essas proteínas não são altamente específicas para seus substratos.

Ainda, algumas HSP podem funcionar como ubiquitinas (peso molecular de 8,5 KDa), reconhecendo e solubilizando proteínas que tenham sido irreversivelmente danificadas para o processo de degradação por enzimas proteolíticas. As HSPs agem também regulando a homeostase proteica, a fluidez da membrana e, ainda, prevenindo ou atrasando a morte celular durante o estresse térmico (AL-WHAIBI, 2010).

A transcrição de genes que codificam para proteínas de choque térmico (HSP) é controlada por proteínas chamadas de fatores reguladores de transcrição de choque térmico (*Heat Shock Factors* - HSFs). Os HSFs são componentes terminais da cadeia de transmissão de sinais, mediando a ativação de genes Hsps, os quais são responsivos ao estresse por calor e a um grande número de outros tipos de estresse. Estes fatores de transcrição reconhecem sequências de ligação palindrômicas (pentanucleotídeo 5'GAAnnTTCnnGAA3' / invertido 5'TTCnnGAAnnTTC3') conhecidas como HSE (*Heat Shock Element*). Os HSE são conservados nas regiões promotoras de genes induzíveis por choque térmico (von KOSKULL-DÖRING *et al.*, 2007). A maior parte das plantas apresentam em torno de 21 HSFs, como no caso de *Arabidopsis*, cada um tendo o seu papel na regulação da expressão dos genes Hsps, mas que cooperam em todas as fases de respostas de stress (desencadeamento, manutenção e recuperação) (KOTAK *et al.*, 2007; SCHARF *et al.*, 2012). Contudo, esse número de HSFs é muito maior em algumas plantas, como é o caso da soja que apresenta 52 genes HSF (SCHARF *et al.*, 2012).

Ao longo dos anos, as espécies *Arabidopsis thaliana* e *Sycopersicum esculentum* (tomate) foram as plantas que tiveram seu sistema de HSFs mais conhecido (KOTAK *et al.*, 2007). A compreensão das características de estrutura funcional e o conhecimento em detalhes de seus domínios de oligomerização permitiu a classificação de três classes de HSFs em plantas: classe A, B e C. Aparentemente, os fatores da classe A são dominantes durante a resposta ao estresse, devido ao seu grande potencial de ativação dos genes HSP e ao seu acúmulo durante ciclos repetidos de estresse e recuperação (von KOSKULL-DÖRING *et al.*, 2007).

2.5.1 FAMÍLIA HSP20

As pequenas proteínas de choque térmico ou HSP20 são predominantemente produzidas em plantas principalmente sob estresse de calor, e apresentam massa molecular de 15-42 kDa quando analisadas em gel de poli-acrilamida por eletroforese desnaturante (SUN *et al.*, 2002). Essas proteínas vegetais são codificadas por famílias nucleares multigênicas. Em plantas superiores existem pelo menos 20 genes Hsp20, mas 40 tipos já foram identificados em uma única espécie (AL-WHAIBI, 2010). Esse número pode ser quatro vezes menor quando se trata de organismos animais. Acredita-se que nas plantas, as Hsp20 tenham sofrido uma grande diversificação funcional para atender a condições nas quais os animais, diferente das plantas, poderiam simplesmente escapar (WATERS *et al.*, 1996; WATERS *et al.*, 2008).

Além disso, é interessante notar que nenhuma das proteínas HSP20 localizadas em organelas de plantas possui homólogas em animais, fungos, ou mesmo em algas verdes (WATERS e VIERLING, 1999 apud WATERS *et al.*, 2008); nesses organismos, as HSP20 parecem estar localizadas apenas no citoplasma/núcleo enquanto as plantas possuem diversas subfamílias de HSP20, que estão distribuídas em diversos compartimentos celulares. Em *Arabidopsis*, 19 genes codificadores de HSP20 foram categorizados em 12 subfamílias, de acordo com a localização celular e filogenia, enquanto foram encontrados 36 Hsp20 em *Populus trichocarpa* e 23 em *Oryza sativa* (SCHARF *et al.*, 2001; SIDDIQUE *et al.*, 2008; WATERS *et al.*, 2008).

O domínio α -cristalino (α -crystallin domain – ACD), característico a todos os membros da família de HSP20, possui em média 80 a 100 aminoácidos conservados na porção C-terminal das HSP20 (CASHIKAR *et al.*, 2005). Este domínio é dividido em duas regiões designadas consenso I e II, separadas por uma região hidrofóbica variável em tamanho. A região consenso I (27 aminoácidos) N-terminal, consiste na sequência conservada Pro- $X_{(14)}$ -Gly-Val-Leu; e a região consenso II (29 aminoácidos) C-terminal, possui a sequência conservada Pro- $X_{(14)}$ /Val/Leu/Ile-Val/Leu/Ile. O ACD é precedido por uma região N-terminal variável em tamanho e de considerável diversidade na sua sequência (CASHIKAR *et al.*, 2005; WATERS *et al.*, 2008). Dentro do domínio ACD das HSP20 é formada uma estrutura secundária típica, geralmente, com a formação de 6 a 8 configurações

folha- β (estrutura β – pregueada). As configurações β são numeradas ao longo da cadeia de aminoácidos no sentido amino-caboxiterminal, sendo encontradas dentro do ACD normalmente as cadeias $\beta 2$ a $\beta 9$ (SIDDIQUE et al, 2008).

Cada classe das HSPs tem uma função característica. As HSP20 são proteínas que tem atividade independente de ATP, por isso não participam de atividades que demandam energia, como o dobramento de proteínas recém-sintetizadas e redobramento de proteínas desnaturadas ou inativas por dobramento inadequado. No entanto, elas foram catalogadas dentro do grupo de chaperonas moleculares pela sua capacidade de reconhecer e se ligar às proteínas desnaturadas ou com dobramento inadequado e, assim, impedir interações inapropriadas que levariam à formação irreversível de agregados, precipitação e degradação dos mesmos por proteinases (SUN *et al.*, 2002).

As Hsp20 são algumas das representantes das proteínas chamadas de ubiquitinas. Elas formam grandes complexos hetero oligoméricos que variam em tamanho de 200-800kDa e também agem marcando agregados já formados de proteínas desnaturadas, ou de dobramento inadequado irreversível, para a ação das proteínas de degradação tanto em células procarióticas como em células eucarióticas (AL-WHAIBI, 2010). A atuação das HSP20 consiste em manter, quando ainda possível, os substratos proteicos em um estado íntegro e passível para o dobramento, ou marcar e permitir a ação das proteinases quando necessário e, além disso, podem direcionar proteínas para outros locais na célula (SUN *et al.*, 2002; AL-WHAIBI, 2010).

Estudos sobre a relação das HSP20 com a resposta ao estresse de calor nas plantas têm revelado que mesmo em espécies divergentes, como plantas C3, C4, MAC (Metabolismo Ácido das Crassuláceas), monocotiledôneas, e outras espécies, há forte relação entre o estresse de choque térmico e a expressão de tais proteínas, sendo concomitantemente observado um acúmulo de HSP20 em organelas, como cloroplastos e mitocôndrias. Há ainda a hipótese dessas proteínas também estarem relacionadas com o controle de qualidade e manutenção de integridade da membrana, especialmente sob condições de estresse (NAKAMOTO e VÍGH, 2007).

2.6 AS HSPs NA INTERAÇÃO MOLECULAR PLANTA-PATÓGENO

O conhecimento de como acontecem as interações moleculares planta/patógeno é essencial, considerando a possibilidade de minimizar e mesmo impedir prejuízos causados por patógenos às diversas culturas economicamente importantes (HINES e ZAHN, 2009). Sabe-se que na resposta ao estresse biótico, estão envolvidos genes que codificam proteínas de resistência (genes R) nas plantas, os quais mediam o reconhecimento de proteínas patógeno-específicas, codificadas por genes de avirulência (genes Avr), ou elicitores (mecanismo de resistência gene-a-gene) (van OOIJEN *et al.*, 2010). Após o reconhecimento da proteína Avr, os genes R são ativados e as proteínas por eles codificadas iniciam as respostas de defesa.

A resposta de resistência da planta exige a interação de várias outras proteínas, além das codificadas pelos genes R, as quais são denominadas proteínas auxiliares na resposta de defesa, e incluem proteínas envolvidas na geração de ROS, quinases e proteínas de choque térmico. (BUCHANAM *et al.*, 2000; van OOIJEN *et al.*, 2010).

Recentemente, evidências genéticas revelaram que as chaperonas desempenham um papel fundamental na imunidade da planta (SHIRASU, 2009). A atividade das proteínas chaperonas de choque térmico durante o estresse biótico, por exemplo, tem se mostrado importante para a estabilidade e acúmulo de proteínas R, e assim para a coordenação de toda a cascata de sinalização de defesa (BOTER *et al.*, 2007).

Maimbo e colaboradores (2007) estudaram as interações compatíveis e incompatíveis de plantas de *Nicotiana tabacum* com cepas patogênicas e não patogênicas da bactéria *Ralstonia solanacearum*, indutoras de resposta de hipersensibilidade. Através da técnica de differential display-PCR, foi realizado o isolamento e a caracterização funcional de genes relacionados à interação tabaco/*R. solanacearum*. Dentre os fragmentos diferencialmente regulados nas folhas de tabaco inoculadas com *R. solanacearum*, foi encontrada uma sequência similar às Hsp20 de *Medicago truncatula*, *Cucumis melo* e uma proteína hipotética de *Arabidopsis*. Em tabaco, esse fragmento isolado tem sequência correspondente ao gene codificador da proteína NtsHSP17, com comprovada resposta a indução por estresse de calor. Os resultados obtidos por PCR em tempo real, logo após o

silenciamento desse gene em tabaco e inoculação do patógeno, demonstraram que a expressão de genes relacionados com a defesa, como proteínas de ligação a elementos de resposta ao etileno e genes relacionados à patogênese PR1 (patógeno-relacionada) e PR4, foi comprometida. O silenciamento do gene NtsHsp17, codificador para uma NtHSP20, também resultou no aumento da virulência de *R. solanacearum*, inclusive da cepa não patogênica. Estes resultados indicam que essa HSP20 pode ter importante relação na resposta de imunidade de plantas de tabaco quando atacadas por *R. solanacearum*.

Resultados semelhantes foram encontrados por Garofalo e colaboradores (2009), que analisaram a expressão de proteínas Hsp20 durante as interações compatíveis e incompatíveis de plantas de laranja e pimenta com os patógenos bacterianos *Xanthomonas axonopodis* pv. citri e *Xanthomonas campestris* pv. vesicatoria, respectivamente. Os perfis de expressão das Hsp20, nas diferentes situações, estiveram relacionados ao tempo decorrido após o tratamento de inoculação das plantas e à compatibilidade da interação, sugerindo que tais proteínas estejam relacionadas à resposta basal de defesa das plantas.

Em plantas de tabaco foi demonstrado que a resistência contra *Fusarium oxysporum* e nematóides do gênero *Meloidogyne* mediada pela resposta de hipersensibilidade é fortemente reduzida quando uma classe específica de Hsp20 é silenciada. Em ambos os casos, proteínas Hsp20 interagem de forma específica com as proteínas codificadas pelos genes R específicos (I-2 e Mi-1, respectivamente), desencadeando a resposta de hipersensibilidade (van OOIJEN *et al.*, 2010). Este é o único relato demonstrando a interação física entre uma HSP20 e um gene R do tipo NB-ARC-LRR (*nucleotide binding sites-leucine-rich regions*) em tomate.

Alguns estudos recentes também sugerem a participação das Hsp20 na resposta a estresse biótico em soja. Kandoth e colaboradores (2011), estudando os mecanismos de resistência da soja ao nematóide *Heterodera glycines*, compararam os perfis de expressão gênica de sincícios em desenvolvimento em duas linhagens de soja diferenciadas apenas pelo gene Rhg1 (resistência à *H. glycines*). Dentre os 1.447 genes que foram diferencialmente expressos entre as duas linhagens, 4 correspondem a genes para proteínas HSP20: Glyma14g06910, Glyma10g32000, Glyma08g07340 e Glyma04g05720; e outros para enzimas que levam ao aumento dos níveis de espécies reativas de oxigênio. Esses resultados

demonstram que uma complexa rede de eventos moleculares ocorre durante a resistência mediada por Rhg1, levando a uma resposta de defesa contra um patógeno das raízes de soja e que, possivelmente, as Hsp20 desempenhem importante papel na resistência.

Fuganti e colaboradores (2004), utilizando marcadores microsatélites, identificaram um QTL ligado a resistência ao patógeno *Meloidogyne javanica* a partir de uma população segregante de soja. O sequenciamento do gene mapeado no QTL revelou alta similaridade com um gene da família Hsp20, o GmHps 17.6-L. Fuganti e colaboradores (2010) avaliaram os níveis de expressão do gene GmHps 17.6-L em genótipos de soja resistentes e suscetíveis ao nematóide *M. javanica* e observaram haver indução no nível de expressão desse gene nos genótipos resistentes, quando comparados aos genótipos suscetíveis. Tais dados indicam que essa proteína pode participar de forma categórica na complexa resposta de resistência da soja à esse patógeno, porém ensaios futuros são necessários para comprovar essa hipótese.

Outros estudos, também realizados com soja, vêm sugerindo que o posicionamento dos *cis* elementos de resposta a choque térmico nos promotores das Hsp está diretamente relacionado a sua ativação durante estresse biótico. Escobar e colaboradores (2003) analisaram a influência das alterações de posição de elementos HSE (Heat Shock Element) na região promotora de um gene Hsp20 de girassol HaHsp17.7G4 sobre a sua expressão nas células gigantes, formadas após a infecção pelo nematóide *Meloidogyne incognita* em plantas de tabaco transgênicas. Os resultados obtidos pelos autores demonstraram que a Hsp20 HaHsp17.7 responde à infecção pelo nematóide e que a presença de um único elemento HSE já é suficiente para sua indução, desde que o mesmo ocorra até - 83 pares de base (pb) do sítio de início da transcrição. Quando o elemento HSE é posicionado a distâncias maiores que - 83 pb do sítio de início da transcrição, o gene Hsp20 HaHsp17.7 deixa de ser induzido em resposta a infecção pelo nematóide.

Posteriormente, a análise comparativa dos promotores de HaHsp17.7 (- 83 pb G4) e dois outros Hsp20 de girassol, que também respondem à infecção pelos nematóides, HaHsp18.6 (G2 contem dois HSE, um proximal, - 49 pb) e HaHsp17.6 (G1 contem um HSE, mais distante, - 108 pb), em plantas de *Arabidopsis* transformadas pela técnica de *floral deep*, corroborou com os dados obtidos por Escobar e colaboradores (2003), ratificando a importância da presença

de um único núcleo de elemento HSE distantes até - 83 pb do sítio de início da transcrição (BARCALA *et al.*, 2008). Para essas análises foram utilizadas construções com o gene marcador *Gus* sob regulação de diferentes configurações de posições e números de elementos HSE no promotor. Além disso, no mutante G1, a distância entre o HSE e o TATAbox estavam muito maior do que a HSE I em G2 ou G4 (49 pb em comparação a 12 pb no G2 e 21 pb no G4). Elementos CAATbox também foram preditos e foram encontrados entre as matrizes de HSE no promotor de G2 (posição de CAAT box - 177 pb e - 72 pb) e G4 (posição de CAATbox - 141 pb). Em contraste, com o CAATbox foi identificado no promotor posicionado - 67 pb, antes do elemento HSE.

Os resultados demonstraram que o promotor G2 foi ativado após a infecção do nematóide em plantas de tabaco transgênicas, enquanto o G1 não foi induzido após a infecção do nematóide. Além disso, aparentemente o promotor G4 responde apenas a nematóides de galha, pois não foi ativado nas plantas infectadas por nematóide de cisto. Esses resultados indicam principalmente que a disposição dos elementos HSE na região promotora, e ainda a sua configuração em conjunto com outros elementos, como TATAbox e CAATbox, exerce importante influência na resposta desses genes sob estresse biótico. Além disso, essa configuração parece ter influência na expressão desses genes em diferentes espécies de patógenos.

2.7 BIOINFORMÁTICA: ARMAZENAMENTO E ANÁLISE DE INFORMAÇÕES

De acordo com a definição do Centro Nacional de Informações sobre Biotecnologia dos EUA– NCBI (2001); Bioinformática é o campo da ciência onde a Biologia, a Ciência da Computação e a Tecnologia da Informação se reúnem para formar uma só disciplina. Existem três vertentes principais dentro da bioinformática: a) o desenvolvimento de novos algoritmos e estatísticas que permitam avaliar relações entre os membros de grandes conjuntos de dados; b) a análise e interpretação de vários tipos de dados, incluindo sequências de nucleotídeos, aminoácidos, domínios de proteínas, estruturas de proteínas; c) o desenvolvimento e implementação de ferramentas que permitem o acesso e a gestão eficiente dos diferentes tipos de informação.

O termo “Biologia Molecular Computacional” pode ser usado, para melhor entendermos esse novo campo da ciência, ou mesmo como sinônimo para

Bioinformática. De forma resumida, esse termo descreve qualquer uso de computadores e programas computacionais para aplicar métodos matemáticos e estatísticos para analisar e caracterizar a estrutura, função e variabilidade de componentes moleculares, como sequências de DNA/RNA (genes), de aminoácidos (proteínas) e qualquer outra informação de origem biológica (SANTOS, 2004).

As tecnologias de sequenciamento de nova geração, ou plataformas de sequenciamento de alta performance (454, Solexa e SOLiD), têm colocado a bioinformática como um conhecimento de fronteira na biologia molecular; o que também pode ser visto na proteômica com os espectrômetros de massa cada vez mais modernos, utilizados para o sequenciamento massivo de cadeias de aminoácidos.

O uso de tais tecnologias provocou uma rápida expansão do volume de dados de sequências biológicas (nucleotídeos e aminoácidos) geradas nos últimos anos, resultando no acúmulo de informações como dados de expressão de genes e de proteínas, interações moleculares, perfil metabólico/bioquímico, estruturas moleculares, filogenia, entre outras, gerando uma significativa demanda por recursos de manipulação e análise computacional especializada, que devem ser especialmente pensados para as novas formas de obtenção da informação. Toda essa imensa quantidade de informações, oriundas de tipos de pesquisa e análise diferentes em inúmeros laboratórios de todo o mundo, precisa ser organizada e colocada à disposição dos pesquisadores. As ferramentas de bioinformática permitem que esses dados sejam armazenados e organizados em banco de dados, racionalizando o acesso para processamento, análise e interpretação dos dados biológicos em redes inter-relacionadas, interativas e amigáveis (LUSCOMBE *et al.*, 2001; QUEIROZ, 2002; De CARVALHO e SILVA, 2010).

A criação dos bancos de dados acelera a investigação em outras áreas como a medicina, a biotecnologia, a agronomia, etc. Um banco de dados pode ser definido como um conjunto de dados, estruturado de forma que possa ser utilizado com eficiência simultaneamente por vários usuários dentro de uma organização (SILBERCHATZ *et al.*, 2006 *apud* CAITAR, 2010).

Os BDBM's (banco de dados de biologia molecular) estão classificados de acordo com as informações biológicas neles contidas, em bancos primários, secundários e especializados. Os bancos primários são aqueles que armazenam informação de sequências, como de nucleotídeos e proteínas. Os

principais bancos dessa categoria são *European Molecular Biology Laboratory* (EMBL)/*European Bioinformatics Institute* (EBI) (mais antigo), *National Center for Biotechnology Information* (NCBI, GenBank) e o *DNA Database of Japan* (DDBJ), todos formam o *International Nucleic Acid Sequence Data Library* (XIONG, 2006).

Os BDBMs secundários armazenam resultados de análises feitas a partir de dados primários. Os bancos SWISS-Prot e o PIR (*Protein Information Resources*) são exemplos de BDBs secundários que armazenam informações referentes às proteínas, como anotações de função da proteína nos organismos, estrutura secundária e terciária, além de literatura associada. Por fim, os BDBMs especializados atendem a um interesse particular de pesquisa, como por exemplo, um organismo em específico ou tipo de dado de proteínas. Os BDs Flybase, HIV *sequence database*, Soybase são de organismos em específico e o BD PROSITE é de dados de proteínas (XIONG, 2006).

Atualmente, a maioria desses BDs apresentam atualizações constantes e uma excepcional interligação entre os mesmos, o que permite uma melhor busca e análise das informações. Um exemplo disso é o banco de dados Phytozome, que é um projeto conjunto do *Department of Energy's Joint Genome Institute* e do *Center for Integrative Genomics*. Nele está contida uma grande quantidade de dados de sequenciamento e anotações genômicas de 25 espécies de plantas, dentre elas a soja. Com acesso livre, esse banco de dados também é integrado a outros BDs, como Pfam, KOG, KEGG e atribuições PANTHER e anotações à disposição do público a partir RefSeq, UniProt, TAIR, JGI.

A correspondência entre bancos de dados tem provado ser um método extremamente útil para encontrar a função de um gene, cuja sequência foi recentemente adquirida, mas sua função permanece desconhecida. Por exemplo, se em uma análise de busca por similaridade, utilizando uma sequência desconhecida obtem-se como retorno uma ou mais sequências, cuja função já é conhecida, a atividade bioquímica ou estrutura da sequência desconhecida pode ser inferida, caso exista elevada similaridade entre as sequências. Porém, nenhum retorno pode ser o resultado obtido em uma análise de busca por similaridade, o que indica que provavelmente se trata de uma sequência ainda não descrita (REHM, 2001).

Atualmente, os milhões de dados armazenados e disponibilizados à comunidade científica nos BDs, e o conhecimento já existente sobre os sistemas biológicos, tem permitido o avanço de uma nova metodologia de pesquisa, os

ensaios *in silico*. Essa abordagem computacional tem poupado tempo e também custos, que seriam despendidos durante uma triagem experimental, fornecendo uma seleção mais apurada de candidatos potenciais, como genes e proteínas para as análises de interesse (XIE *et al.*, 2007). Essa metodologia possibilita que a pesquisa se inicie em etapas mais avançadas, tendo como base as informações disponíveis, e permitindo o desenvolvimento de ensaios *in vitro* dirigidos.

Os bancos de dados de expressão gênica digital são bons exemplo para os casos em que o tempo de fase experimental da pesquisa pode ser significativamente poupado. O Geninvestigator, agora na versão V4, é um banco de dados de expressão digital, disponível via web, que permite explorar a regulação da expressão de genes de interesse em centenas de condições experimentais, tipos de tecidos, doenças e genótipos; Ou mesmo, quais genes estão regulados em uma situação de interesse. Atualmente, tem sido um dos mais importantes bancos dessa categoria (expressão digital), usado por pesquisadores de inúmeras universidades e empresas em todo o mundo. Isso é devido, principalmente à qualidade das informações disponíveis, já que são dados curados, e ao acesso cada vez mais facilitado, graças às periódicas atualizações e *redesign* da interface gráfica (HRUZ *et al.*, 2008).

Nos últimos anos, o número de bases de dados e ferramentas on-line para análise de sequências e experimentos *in silico* aumentou exponencialmente. Programas on-line, como MEME (*multiple Em for motif elicitation*), MatInspector (Genomatix), Place e PlantCare têm sido amplamente utilizados para análises de motivos proteicos e também elementos *cis* regulatórios (SARKAR *et al.*, 2009; LIN *et al.*, 2011). O pacote Genomatix ainda disponibiliza ferramentas para a obtenção de modelos de disposição e composição de elementos *cis* regulatórios nas regiões promotoras dos genes de interesse, tendo como base, dados armazenados em bancos interligados e análises estatísticas. Outras ferramentas on-line, como SignalP Server, desenvolvido por Petersen e colaboradores (2011), podem ser utilizadas para a predição *in silico*, a partir de um Modelo de Hidden Markov (HMM), de peptídeos sinais em sequências de aminoácidos com diferentes aplicações (OUYANG *et al.*, 2009). O software on-line TMHMM server desenvolvido por Krogh e colaboradores (2000), é utilizado na predição e localização de domínios transmembrana em sequencias peptídicas, e tem tido bastantes aplicações. Os diagramas gerados como resultados no TMHMM

indicam a probabilidade da presença de um domínio transmembrana ao longo da cadeia de aminoácidos. Quando há alta probabilidade da presença de um domínio, é possível definir sua ocorrência por meio do diagrama, e identificar as regiões transmembrana, tanto intra quanto extracelulares.

Outras ferramentas como o Phyre vêm sendo amplamente utilizadas pela comunidade científica (média de 4.150 submissões diárias) para predição de estrutura e função de uma proteína a partir da submissão de uma sequência de aminoácidos (KELLEY e STERNBERG, 2009). O servidor Phyre dispõe de um banco de dados de estruturas secundárias e terciárias de proteínas conhecidas, e que é atualizado a cada submissão no Protein Data Bank (PDB) de uma nova estrutura proteica. A predição da estrutura de uma proteína de interesse ocorre pela submissão da sequência de aminoácidos que é comparada contra a base de dados e a estrutura de uma proteína homóloga serve de parâmetro para estabelecer o perfil estrutural da proteína em questão (Bennett-Lovsey *et al.*, 2008). Além de programas para análises de sequência na busca de elementos *cis* regulatórios, domínios funcionais e estruturas, atualmente, muitos bancos tem proporcionado a investigação digital da expressão gênica. Tais bancos são alimentados por resultados experimentais de investigações globais do transcrito nas mais variadas condições. Assim, o perfil de expressão de um gene ou conjunto de genes de interesse pode ser obtido avaliando-se sua expressão digital. Além disso, as mudanças nas interfaces gráficas de sites de análise e bancos de dados biológicos, para o uso de profissionais que não tenham conhecimentos profundos de programação e bancos de dados, estão tornando mais simples o acesso e a obtenção dos resultados (LOPES e DIAS, 2010).

Tudo isso torna a pesquisa cada vez mais dinâmica e interligada, combinando os resultados obtidos pela comunidade científica e fazendo com que todo o conhecimento seja aproveitado ao máximo.

3 OBJETIVO GERAL

O presente trabalho pretendeu, identificar membros e fazer uma caracterização *in silico* das regiões codificadoras e reguladoras dos genes da família Hsp20 de soja, a fim de estabelecer os padrões de *cis* elementos potencialmente envolvidos na resposta aos estresses abiótico e biótico.

3.1 OBJETIVOS ESPECÍFICOS

- ✓ Caracterizar *in silico* a família de GmHsp20 (*Glycine max* – Hsp20) quanto a distribuição no genoma, localização subcelular e divisão em subfamílias;
- ✓ Caracterizar a família de GmHsp20, quanto a presença de elementos conservados na região codante e reguladora;
- ✓ Caracterizar *in vivo* os perfis de expressão das Hsp20 em raízes de soja, frente a indução por estresses abióticos (calor e frio) e biótico (infecção pelo nematóide *Meloidogyne javanica*);
- ✓ Estabelecer modelos operacionais para os promotores dos GmHsp20, candidatos induzidos por calor, frio e, principalmente pelo estresse biótico (infecção por *M. javanica*);
- ✓ Avaliar níveis de expressão diferenciados dos genes responsivos ao estresse biótico entre plantas de soja suscetíveis e resistentes ao patógeno submetidas a infecção por *M. javanica*.

4 MATERIAL E MÉTODOS

4.1 ANÁLISES “*IN SILICO*”

4.1.1 Busca por Modelos Gênicos Hsp20 em Soja

O modelo estatístico que representa a melhor sequência de um domínio é conhecido como perfil do Modelo de Markov (Hidden Markov Model - HMM). O HMM que representa o domínio das HSP20 (PF00011) foi obtido pelo site Pfam (<http://pfam.sanger.ac.uk/family/PF00011#tabview=tab4>) e empregado para a busca de modelos gênicos anotados no genoma da soja como Hsp20. A submissão da sequência perfil HMM utilizando a ferramenta BlastP contra os bancos de dados Phytozome (<http://www.phytozome.net/>) e Superfamily (<http://supfam.cs.bris.ac.uk/SUPERFAMILY/>) considerou todos os modelos gênicos com e-value ≤ 0.001 e foi seguida da remoção de sequências redundantes entre os dois bancos de dados. Como estratégia adicional de busca, foi utilizada a palavra-chave “Hsp20” para a identificação de possíveis modelos gênicos anotados como Hsp20 em soja, que poderiam ter sido ignorados quando o perfil HMM foi utilizado, devido à presença do domínio Hsp20 incompleto.

As sequências peptídicas e 500 pb à montante do sítio de início da transcrição de todos os modelos gênicos identificados (76 modelos) foram obtidas com o auxílio da ferramenta BioMart (<http://www.phytozome.net/biomart/martview/1510e95e6cd635e0cbbd85843a3f7d54>)

4.1.2 Domínios Conservados e Estrutura Secundária das GmHsp20

A presença do domínio conservado alfa cristalino (*alpha crystallin domain* –ACD) nas proteínas preditas foi determinada com o software MEME (<http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>), utilizando como parâmetros: repetição por sequência igual a 1, número máximo de motivos encontrados igual a 1, configuração para motivo ideal de 80 a 100 aminoácidos (CASHIKAR et. al., 2005) . A sequência do motivo obtido foi submetida nos programas InterProScan (<http://www.ebi.ac.uk/Tools/InterProScan/>) e PROSITE (<http://www.expasy.org>), para confirmação de identidade.

A sequência protéica dos genes preditos da família GmHsp20 foi utilizada na ferramenta EXPASY PROTPARAM (<http://www.expasy.org/tools/protparam.html>) para obtenção das informações sobre peso molecular, ponto isoelétrico teórico (PI), composição de resíduos de aminoácidos e índice de instabilidade [sendo considerado um índice >40 como instável] (GURUPRASAD *et al.*, 1990)]. As anotações de localização cromossômica predita, número e tamanho de íntrons em pb foram obtidos no banco de dados Phytozome.

Para as predições de peptídeos sinal, domínio transmembrana e localização subcelular foram utilizados os softwares SignalP (<http://www.cbs.dtu.dk/services/SignalP/>), WoLF, PSORT (<http://wolfpsort.org/>) (<http://psort.nibb.ac.jp/>), PREDOTAR (<http://urgi.versailles.inra.fr/predotar.html/>), TargetP (<http://www.cbs.dtu.dk/services/TargetP/>), e TMHMM 2.0 (<http://www.cbs.dtu.dk>).

4.1.3 Busca por HSE e Outros *cis* Elementos na Região Regulatória dos Modelos Gênicos

Nem todos os 76 modelos gênicos possuem atualmente a região 5'UTR predita. Desta forma, foi possível obter apenas 53 sequências com informação disponível para pelo menos 500pb à montante do sitio de início da transcrição. Na busca por *cis* elementos nessas sequências foram utilizadas as bases de dados PlantCare (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>), PLACE (<http://www.dna.affrc.go.jp/PLACE/signalup.html>), MEME (confirmado no TOMTOM que permite a pesquisa por busca de similaridade do motivo encontrado nos bancos de dados: Jaspar, Transfac e Uniprobe) e MatInspector – Genomatix (http://www.genomatix.de/cgi-bin/matinspector_prof/mat_fam.pl?s=f3538718a86b1d874576e65b1b106706).

4.1.4 Seleção de Modelos Gênicos para Análise “*in vivo*” e Desenho dos Oligonucleotídeos Iniciadores

Para as análises *in vivo*, foram selecionados os modelos gênicos que obedeciam aos critérios: (a) expressão evidenciada em pelo menos um dos bancos de dados de expressão gênica de soja: Genosoja (<http://www.bioinfo.cnpsa.embrapa.br>), SoyBase (Glycine max RNA-Seq Atlas) (<http://soybase.org/soyseq/>), Geninvestigator (<https://www.geneinvestigator.com/gv/user/serveApplet.jsp?heapSize=500&exprMethod=rma>); (b) presença de domínio ACD na região codificadora; (c) presença de pelo menos um elemento HSE (*heat shock element*) na região à montante do sítio de início da transcrição. Quando a expressão do modelo gênico com domínio ACD predito foi evidenciada nos bancos de dados disponíveis para a soja, mas a presença de elementos HSE não foi avaliada devido à indisponibilidade da região regulatória, o modelo gênico foi incluído para análise *in vivo*.

O desenho dos oligonucleotídeos específicos para as análises de RT-PCR e RTq-PCR (two-step) foi realizado, utilizando o software Primer3Plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) e Vector NTI Advance™ (Invitrogen), visando como alvo de amplificação segmentos de 85 – 150bases na região 5’UTR, quando possível. Essa estratégia foi adotada para facilitar a amplificação específica de cada elemento da família de Hsp20. Para essa família gênica em específico, o desenho de *primers* na região 5’UTR foi mais eficiente na busca pela especificidade, já que existe uma alta conservação na região próximas ao 3’UTR, pela presença do domínio ACD. As principais configurações para o desenho dos oligonucleotídeos específicos, foram o tamanho do fragmento, a porcentagem de CG de no mínimo 40%, tamanho dos *primers* de 18 a 22 pb e PoliX máximo de 3. Todos os oligonucleotídeos específicos obtidos foram contrastados contra o banco de dados de *Glycine max* no Phytozome, para certificar a especificidade da amplificação. A lista dos oligonucleotídeos utilizados neste trabalho (sequencia e tamanho do fragmento amplificado), bem como os modelos gênicos alvo-específicos é apresentada na Tabela 1S.

4.1.5 Busca por Modelos de Promotores

Para a busca por modelos combinatórios de *cis* elementos nos promotores dos genes Hsp20 foi utilizado o software FrameWorker – Genomatix (<http://www.genomatix.de/cgi-bin/gems/launch.pl?s=3084bde478ae75857e7900731bf9e9c1;GEMS=1;TASK=framework>). Foram construídos modelos para a caracterização da região promotora mínima dos genes Hsp20 de soja. Após os resultados das análises de expressão *in vivo*, por RT-qPCR, de dados obtidos de literatura e de bancos de expressão, buscou-se também por modelos combinatórios de *cis* elementos nos promotores dos Hsp20 que se mostraram responsivos a estresses biótico e abiótico. A construção do modelo para os Hsp20 responsivos a estresse biótico incluiu as identificadas no presente estudo e os modelos gênicos Glyma14g06910.1, Glyma10g32000.1, Glyma08g07340.1, Glyma04g05720.1 (KANDOTH *et al.*, 2011) e Glyma13g24460.1 (FUGANTI *et al.*, 2010), que foram previamente identificadas em soja como responsivos a estresse biótico. Todas as etapas do estudo para a obtenção de modelos gênicos de GmHsp20 está apresentada no Fluxograma (Figura 5).

4.1.6 Relação Filogenética e Distribuição em Subfamílias

A categorização das HSP20 de soja em subfamílias foi possível através do uso da informação da predição de localização subcelular (SCHARF *et al.*, 2001) e relação filogenética entre os candidatos, utilizando tanto a sequência de aminoácidos de domínios proteicos quanto a região regulatória, quando disponível.

Para análise da relação filogenética utilizou-se a sequência de aminoácidos correspondente aos ACD identificados pelo MEME, para cada proteína selecionada para a análise *in vivo* e predita como HSP20 de soja, as GmHSP20. Também foram utilizadas as sequências dos domínios ACD de HSP20 de diversos organismos (*Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinífera*, *Triticum aestivum*, *Pisum sativum*, *Solanum lycopersicum*, *Zea mays*, *Lycopersicon peruvianum*, *Cynodon dactylon*, *Saccharum officinarum*, *Sorghum bicolor*, *Hordeum vulgare*) para os quais já se conhece a distribuição entre as subfamílias de todas, ou algumas, HSP20. Através do *software* ClustalX (versão 2.1) realizou-se o alinhamento múltiplo par-a-par de todas as sequências, e a partir deste alinhamento foi construída uma

árvore filogenética (software Mega 5.0) não-enraizada pelo método *neighbor-joining* (SAITOU e NEI, 1987) com *bootstrap* (1.000 repetições). A estrutura secundária das possíveis GmHSP20 foi predita utilizando o software Phyre² (<http://www.sbg.bio.ic.ac.uk/phyre2>). Também foram anotados os tamanhos das ORF (Open Reading Frame) e o número e tamanho de íntrons preditos, disponíveis no Phytozome. A ORF é uma sequência de DNA, delimitada por um códon de início de leitura e um códon de parada.

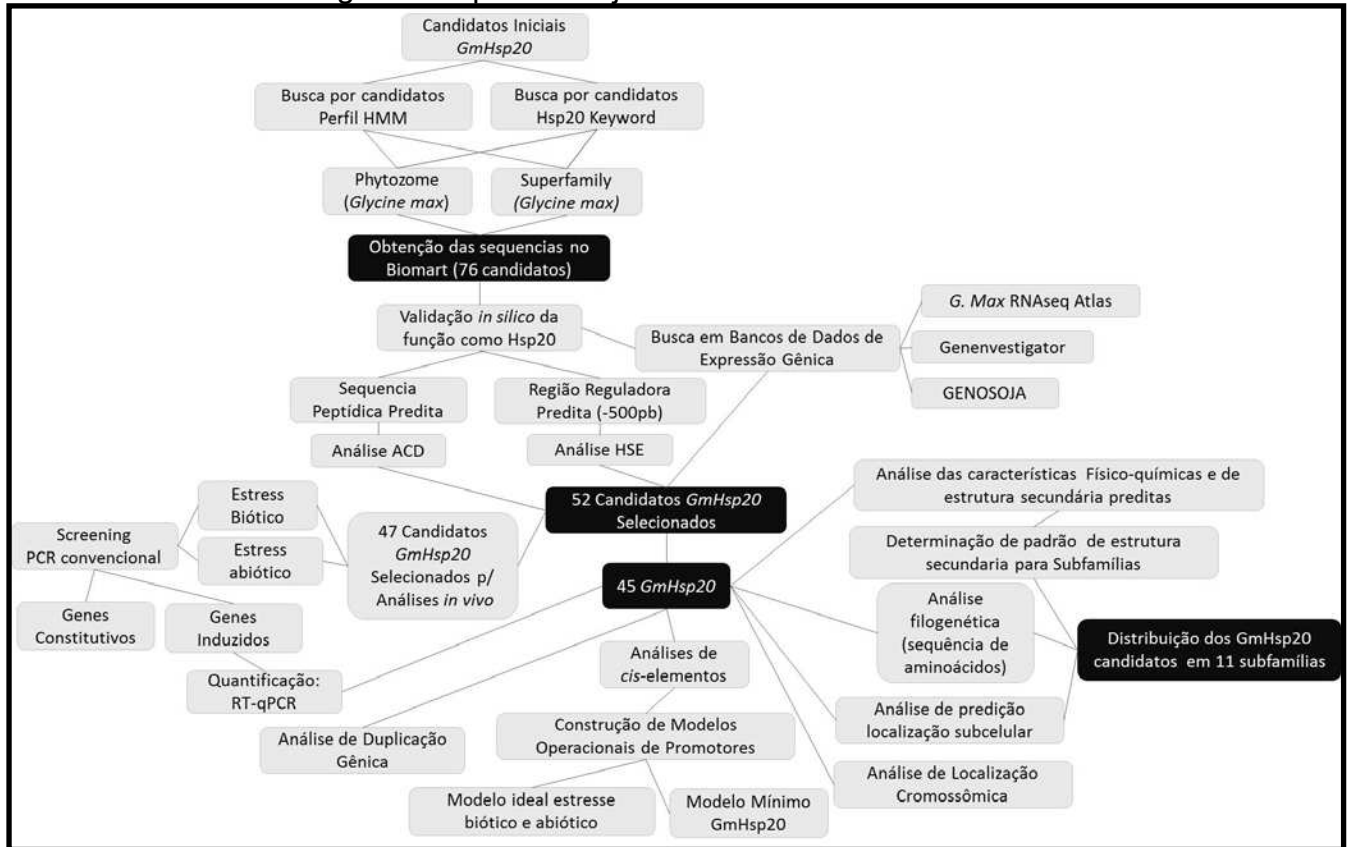
A partir das estruturas secundárias preditas para as HSP20 estudadas e da análise filogenética para distribuição em subfamílias, foram construídos modelos específicos para cada subfamília das HSP20 de soja. Nessa análise, foram consideradas a presença de íntrons, a posição do domínio ACD, o número de regiões β -pregueada e presença/posição de peptídeo sinal (van MONTFORT *et al.*, 2001).

4.1.7 Análise de Localização Cromossômica e Duplicação Gênica

Para a análise de duplicação gênica, as sequências proteicas das 47 GmHSP20 candidatas foram analisadas pela ferramenta blastp suite-2sequences-NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Os parâmetros de seleção foram mínimo de 70% identidade e mínimo de 80% de cobertura. Os genes GmHsp20 foram plotados nos cromossomos utilizando o software MapChart, através da informação de suas posições cromossômicas, disponível no Phytozome, e do tamanho em pares de base de cada um dos 20 cromossomos da soja.

Para estabelecer as possíveis duplicações dos genes GmHsp20 no genoma da soja, foram utilizadas informações fornecidas no trabalho de Schmutz e colaboradores (2010) a respeito das duplicações ocorridas durante a evolução da soja, e os resultados de identidade e cobertura.

Figura 5 - Fluxograma das etapas conduzidas para a realização das análises *in silico* dos genes Hsp20 de soja.



4.2 ANÁLISES “IN VIVO”

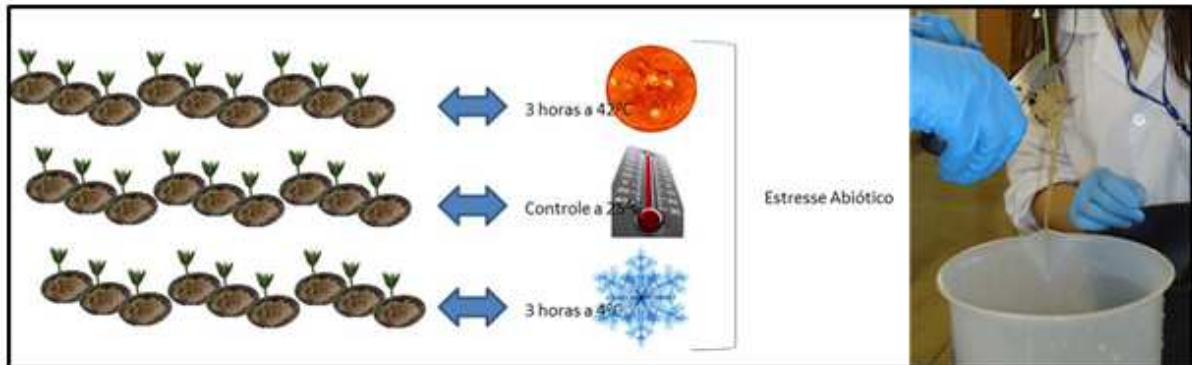
4.2.1 Material Biológico e Instalação do Experimento para Análises “in vivo”

O genótipo de soja, BRS133, suscetível à infecção pelo nematóide *Meloidogyne javanica*, foi disponibilizado para este experimento pelo Banco Ativo de Germoplasma (BAG) da Embrapa Soja. As sementes da cultivar foram tratadas com fungicida, semeadas em papel de germinação e mantidas na câmara de germinação do Laboratório de Qualidade de Sementes da Embrapa Soja. Três dias após a emergência, cada plântula foi transplantada para um tubete plástico com 5 cm de diâmetro e 25 cm de altura contendo areia autoclavada. O experimento foi conduzido em fitotron, tendo a temperatura controlada em $25^{\circ}\text{C} \pm 3^{\circ}\text{C}$, umidade em $90\% \pm 10\%$ e regime de luminosidade de 16 horas luz e 8 horas escuro. Os tratamentos tiveram início após as plantas atingirem o estágio de desenvolvimento em V3 (formação do terceiro trifólio).

4.2.2 Tratamento – Estresse Abiótico (Temperatura)

O delineamento experimental para os tratamentos de estresses abióticos foi inteiramente casualizado, com três réplicas para cada uma das três repetições do tratamento. Os tratamentos utilizados foram choque térmico por três horas a 4°C (SARKAR *et al.*, 2009; SWINDELL *et al.*, 2007) (tratamento de frio), a 42°C (ZHOU *et al.*, 2011; SWINDELL *et al.*, 2007) (tratamento de calor) e 25°C (tratamento controle), totalizando 27 tubetes com uma planta em cada (Figura 6). Para as coletas, o sistema radicular de cada planta foi retirado e lavado, um a um, utilizando água destilada na temperatura respectiva a cada tratamento, para eliminar o excesso de areia. Em seguida, foi retida e descartada a parte aérea e, imediatamente, a raiz foi envolta em papel alumínio, devidamente identificado e armazenado em caixa de isopor contendo nitrogênio líquido.

Figura 6 - Representação esquemática do delineamento experimental instalado com o genótipo BRS133, susceptível a nematóide de galha, em fitotron para exposição a condições de estresse abiótico.



4.2.3 Investigação do Perfil *Heat Shock* em Tecido Específico – Folha

Para análise de expressão dos modelos gênicos candidatos a Hsp20 de soja que não foram induzidos nos tecidos radiculares sob estresse de calor, característica marcante dos genes da família Hsp20, foi instalado um novo experimento para coleta de folhas de plantas de soja do genótipo BRS133 submetidas ao choque térmico. Essa estratégia visou à investigação da possível indução desses genes de forma tecido-específica e permitindo assim o seu enquadramento dentro da família Hsp20. O delineamento experimental para o tratamento de estresse por choque térmico foi inteiramente casualizado, com três

repetições por tratamento (plantas), submetidas por três horas a 42°C (ZHOU *et al.*, 2011; SWINDELL *et al.*, 2007) e 25°C (tratamento controle). Para as coletas, da parte aérea foram retiradas três folhas de cada planta e, imediatamente, envoltas em papel alumínio, devidamente identificado e armazenado em caixa de isopor contendo nitrogênio líquido.

4.2.4 Preparo do Inóculo e Inoculação das Plantas

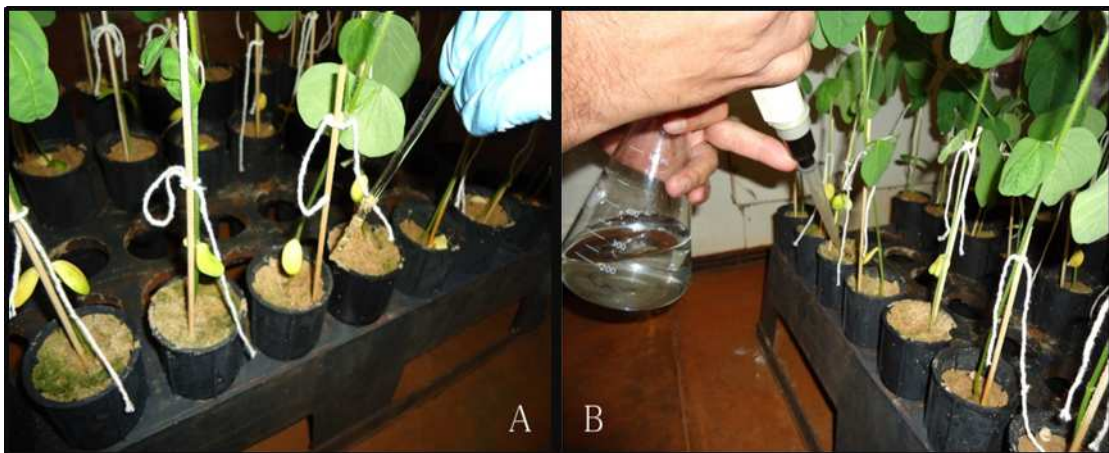
Para a obtenção do inóculo, uma população de *Meloidogyne javanica* foi multiplicada na casa de vegetação do Setor de Nematologia da Embrapa Soja, em plantas de amendoim, por três meses. As raízes das plantas de soja foram processadas de acordo com a técnica de Boneti e Ferraz (1981), lavadas, cortadas em pedaços e trituradas em liquidificador, para extração dos nematóides. Posteriormente, as amostras foram passadas em peneiras de 60 e 500 meshes e, após enxague com água corrente, o conteúdo retido na peneira de 500 mesh (ovos, larvas e adultos do nematóide) foi recuperado em beakers (Figura 8). Após três dias em câmara de eclosão, os J₂ viáveis foram coletados a cada dia, durante três dias, em frasco erlenmeyer. O material obtido na extração e eclosão (incubado em câmara de eclosão) foi quantificado em lâmina de Peters, utilizando 1mL da suspensão para a projeção do número total de indivíduos para que o inóculo fosse ajustado para conter 125 indivíduos/mL (LOPES, 2009). A partir de uma única massa de ovos até 500 juvenis podem infectar uma ou mais raízes próximas e iniciar um novo ciclo de vida (DROPKIN, 1989 *apud* TEREFE *et al.*, 2009).

Figura 7 - Peneiras de 60 meshes (superior) e 500 meshes (inferior), utilizadas para recuperar os indivíduos de *Meloidogyne javanica* a partir de raízes de soja trituradas em liquidificador.



Cada planta foi inoculada com 4mL de larvas juvenis (J₂) de *M. javanica* com o auxílio de pipeta automática, junto às raízes, através de um pequeno orifício aberto ao lado da raiz da planta (Figura 9), totalizando cerca de 500 nematóides por planta.

Figura 8 - A. Abertura do orifício ao lado do colo da planta para a inoculação de nematóides. B. Inoculação das amostras de soja com indivíduos de *Meloidogyne javanica*.

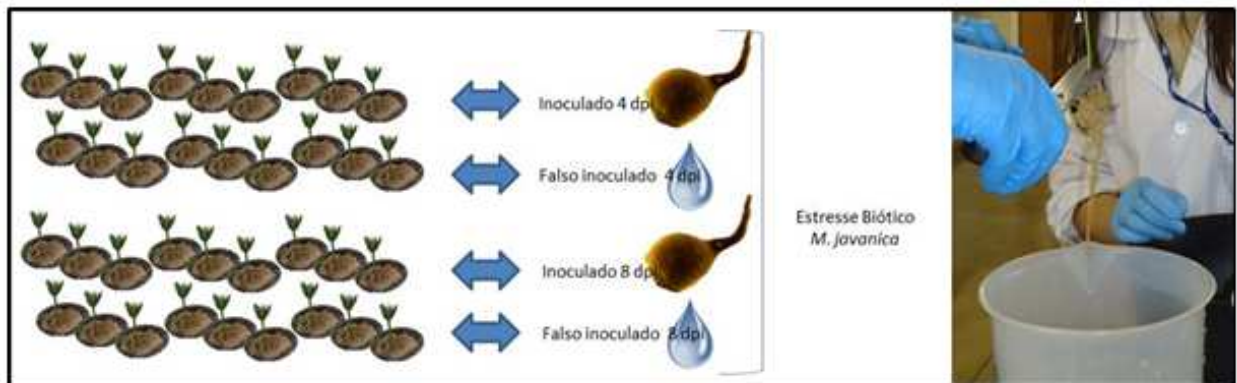


Todas as amostras coletadas foram mantidas em ultrafreezer a -80° C até a continuidade dos procedimentos de extração de mRNA.

4.2.5 Tratamento – Estresse Biótico (Infecção *M. javanica*)

O delineamento experimental para o tratamento de estresse biótico por infecção pelo nematóide *M. javanica* foi conduzido inteiramente casualizado, com três réplicas para cada uma das três repetições do tratamento, e uma estrutura de tratamentos com dois fatores: tempo de inoculação (dias após inoculação - 4dpi e 8 dpi) (KANDOTH *et al.*, 2011), e infecção (inoculado - INOC e falso inoculado - MOCK) (Figura 7), totalizando 36 tubetes com uma planta cada. A metodologia de coleta ocorreu da mesma forma para o estresse abiótico.

Figura 9 - Representação esquemática do delineamento experimental instalado com o genótipo BRS133, susceptível a nematóide de galha, em fitotron para inoculação do nematóide *M. javanica*.



4.2.6 Análise Comparativa dos Genes Responsivos ao Estresse por Infecção pelo Nematóide em Genótipo Resistente

Para análise comparativa entre os genótipos suscetível e resistente a infecção pelo patógeno, foi instalado um experimento com a cultivar PI505099, resistente ao patógeno. Essa estratégia visou a análises de expressão relativa dos genes que na cultivar suscetível foram induzidos pela infecção. O delineamento experimental e a coleta seguiram o mesmo procedimento aplicado no tratamento de estresse biótico para a BRS133.

4.2.7 Preparo das Amostras, Extração de DNA e RNA Total

Cada uma das três repetições dos blocos, composta por três réplicas, foram maceradas separadamente utilizando-se pistilo, almofariz e nitrogênio líquido. Após maceradas, as amostras foram distribuídas em microtubos de 1,5mL e armazenadas em ultrafreezer a -80°C .

O DNA das amostras de todos os tratamentos (estresse biótico - infecção pelo nematóide *Meloidogyne javanica*) foi extraído para diagnóstico molecular (técnica de PCR - Polimerase Chain Reaction) e confirmação da presença do parasita nos tecidos de raiz. O conjunto de oligonucleotídicos específicos para a detecção da espécie *M. javanica*, com *amplicon* de 945bp (*foward*_5'CAAACCCACGCGGCTTCGGC3' e *reverse*_5'TGGGGGTGCCCTTCCGTCAA3'), foi desenhado utilizando o programa

computacional Primer-BLAST/NCBI (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome).

O DNA de fêmeas do nematóide *M. javanica* foi extraído, utilizando protocolo de extração modificado de Madani e colaboradores (2005), para servir de controle positivo nas reações de PCR de confirmação da infecção nas raízes. Cerca de 50 fêmeas adultas foram coletadas diretamente das raízes, com auxílio de lupa e espátula, em tubo eppendorf de 1,5mL contendo 20 µL de água bidestilada e autoclavada.

Em seguida, adicionou-se 12 µL de proteinase K (600 µg/ml) e a amostra foi macerada dentro dos tubos com auxílio de pistilo. Na etapa seguinte, foram adicionados 120 µL de 1x Buffer PCR (*Invitrogen*). A amostra foi incubada durante 1 hora a temperatura de 65°C, seguida pela incubação por 15 minutos a 95°C. Em seguida, adicionou-se 150µL de clorofórmio e a amostra foi agitada por suaves inversões durante 5 minutos. Após centrifugação a 18000g por 5 minutos foi transferido o sobrenadante para um tubo novo e precipitou-se o DNA com 0,2 mL de isopropanol gelado. Na etapa seguinte, a amostra foi misturada por inversões e incubada durante 10 min a -20°C. Após centrifugação a 18000g por 15 minutos a 4°C, o sobrenadante foi descartado e adicionou-se 0,4 mL de etanol 70%. Após centrifugação a 18000g por 5 minutos a 4°C o etanol foi removido e o precipitado final ressuspendido em 30µL de água Milli-Q.

Para a reação de padronização de PCR de detecção da infecção pelo nematóide foram misturados 1 µL de DNA de soja e nematóides a [30ng/µL], 18,8µL de água Milli-Q, 2,5 µL de tampão de PCR [10x], 1 µL de dNTPs Mix a [2,5mM], 1 µL de MgCl₂ a [50mM], 0,5 µL PCR primer, *forward* e *reverse*, a [10µM] e 0,2 µL Taq DNA Polimerase [5U/µL] (Kit PCR *Invitrogen*), tendo como volume final de reação 25 µL. As misturas de reação foram incubadas em termociclador para a amplificação dos fragmentos utilizando-se o programa: 95°C por 1', seguido de 30 ciclos de 95°C por 30", variado°C por 30" e 72°C por 1' e um passo adicional de 72°C por 7 minutos. As reações de PCR foram feitas em duplicata, sendo testadas seis diferentes temperaturas de pareamento para o par de oligonucleotídeos (54, 56, 58, 60, 62 e 64°C) (Figura 1S).

Em seguida as amostras foram aplicadas em gel de eletroforese, preparado com agarose 1,5%, tampão TAE 1x (108g [2 moles] de Tris base, 57,1 ml ácido acético glacial, 100 ml de solução de EDTA dissódico 0,5 M (pH 8.0), água

destilada para o volume de 1 litro) e corado com 25 μ L de brometo de etídio (10 mg/mL). Para a aquisição e armazenamento das imagens foi utilizado o equipamento transluminador e fotodocumentador Loccus.

Para a extração de DNA dos nematóides, diretamente das raízes infectadas no experimento, foi utilizado o protocolo descrito por Rahman e colaboradores (2010) e a integridade avaliada por eletroforese.

As raízes maceradas em cadinho com nitrogênio líquido e distribuídas em tubos de 1,5mL foram acrescidas de 1mL de tampão de extração de DNA [4% (w/v) CTAB, 100 mM Tris-HCl (pH 8,0), 1,4 M NaCl, 50% mM Na-EDTA (pH 8,0) e 1 (w/v) DTT] para cada 50 mg (peso fresco) de tecido da raiz.

As amostras foram homogeneizadas em vortex por 5 minutos à temperatura ambiente. Na etapa seguinte, foram incubadas por 30 minutos a 55°C. Após centrifugação a 18000g em temperatura ambiente por 5 minutos transferiu-se o sobrenadante para um tubo novo de 2 mL. A seguir adicionou-se 200 μ g de RNase A por mL de tampão de extração e incubou-se por 15 minutos a 37°C. Após período de incubação, adicionou-se igual volume de fenol: clorofórmio: álcool Isoamílico [25:24:1]. Após centrifugação a 18000g por 5 minutos a 4°C transferiu-se o sobrenadante para um tubo novo de 2 mL, e foi adicionado volume igual de clorofórmio: álcool Isoamílico [24:1]. As amostras foram vigorosamente agitadas em vortex, centrifugadas a 18000g durante 5 minutos a 4°C durante 5 min, e em seguida transferiu-se o sobrenadante para um tubo novo de 2mL. Para a precipitação do DNA adicionou-se 350 μ L de isopropanol gelado, misturou-se por inversões. Após incubação a -80°C por 1h as amostras foram descongeladas sem agitação a 4°C. Em seguida, foram centrifugadas a 20000g durante 20 minutos a 4°C, descartou-se o sobrenadante e adicionou-se 1mL de etanol 70% (v/v). Após centrifugação a 14000g em temperatura ambiente por 5 minutos, descartou-se novamente o sobrenadante e o precipitado foi ressuscitado em 20 μ L de água Milli-Q. A concentração e qualidade das amostras foram determinadas em espectrofotômetro (A260/280 e A260/230) utilizando NanoDrop® ND-1000 espectrofotômetro (Produtos Nanodrop, Wilmington, DE, EUA).

Para a reação de PCR de confirmação de infecção foram utilizadas as condições de reação descritas anteriormente e temperatura de anelamento de 64°C. Em seguida, as amostras foram aplicadas em gel de eletroforese, preparado com agarose 1,5%, tampão TAE 1x (108g [2 moles] de Tris base, 57,1 ml ácido

acético glacial, 100 ml de solução de EDTA dissódico 0,5 M (pH 8.0), água destilada para o volume de 1 litro) e corado com brometo de etídio (10 mg/mL). Para a aquisição e armazenamento das imagens foi utilizado o equipamento transluminador e fotodocumentador Loccus (Figura 2S).

O RNA total das raízes coletadas nos tratamentos com *Meloidogyne javanica* (infectada) e sadia, bem como nos experimentos de estresse abióticos, nos diferentes tempos, foi extraído utilizando-se o reagente TRIZOL (*Invitrogen*, Carlsbad, CA, EUA), de acordo com instruções do fabricante. O Trizol é uma solução monofásica de fenol e guanidina isotiocianato, que permite a extração de RNA vegetal em uma etapa única (CHOMCZYNSKI e SACCHI, 1987). Durante a homogeneização ou lise da amostra, o Trizol mantém a integridade do RNA enquanto rompe as células e dissolve os componentes celulares.

As raízes maceradas em cadinho com nitrogênio líquido e distribuídas em tubos de 1,5mL foram acrescidas do reagente Trizol (1 ml/100 mg tecido). A amostra foi homogeneizada em vortex por 5 minutos à temperatura ambiente. Em seguida, adicionou-se 0,2 mL de clorofórmio para cada 1 mL de Trizol, agitou-se vigorosamente por 15 segundos e incubou-se por 3 minutos. Após centrifugação a 18000g por 15 minutos a 4°C transferiu-se a fase líquida para um novo microtubo e precipitou-se o RNA com 0,5 mL de isopropanol para cada 1mL de Trizol usado inicialmente. A adição de clorofórmio ao tampão contendo RNA solubiliza os lipídios da membrana celular e nuclear, e quando seguida da centrifugação, separa a solução em fases aquosa e orgânica, permitindo a sua remoção.

As amostras foram incubadas a temperatura ambiente por 10 minutos e fez-se nova centrifugação a 18000g por 10 minutos a 4°C para precipitar o RNA. O sobrenadante foi removido e o precipitado lavado com etanol 75%. O etanol foi removido por centrifugação a 14000g por 5 minutos a 4°C e o precipitado final diluído em 50µL de água tratada com DEPC (dietilpirocarbonato) autoclavada. Todas as soluções e utensílios para os procedimentos foram preparadas e tratados com água DEPC autoclavada, a qual atua inativando RNAses através da degradação dos resíduos de histidina presentes na enzima.

A concentração de RNA total das amostras foi determinada em espectrofotômetro (nanodrop) a 230 nm e a integridade visualizada por eletroforese em gel de agarose 1% contendo brometo de etídeo (EtBr), em tampão SB pH 8,5

(Figura 3S). Após a extração, todas as amostras de RNA foram mantidas em freezer -80°C .

4.2.8 Síntese de cDNA

Em uma etapa previa à síntese de cDNA, as amostras de RNA foram tratadas com a enzima Desoxiribonuclease I (Kit *Invitrogen* - DNase I), a fim de eliminar qualquer molécula de DNA presente na amostra. Para a reação, foram utilizados 2 μg de RNA total, 1 μL do tampão de reação da enzima (10x DNaseI Reaction Buffer), 1 μL de enzima DNaseI [1U/ μL] e água DEPC para completar o volume de 10 μL . A reação foi incubada à temperatura ambiente durante 15 minutos, seguido da adição de 1 μL de solução EDTA [25mM]. A reação foi incubada a 65°C por 10 minutos.

As amostras de RNA tratadas foram submetidas a uma avaliação pela técnica de PCR utilizando oligonucleotídeos desenhados para o gene β -actina. O intuito dessa reação foi de verificar a eficiência da reação de tratamento das amostras de RNA com a enzima DNaseI. Esse par de iniciadores é complementar a dois exons diferentes, assim quando na presença de moléculas de DNA genômico, o fragmento amplificado apresentou tamanho igual a 520 pb e quando apenas na presença de RNA (cDNA) apresentou fragmento amplificado igual a 440 pb, já que um íntron com tamanho de 80 pb é retirado após a etapa de processamento do RNA mensageiro (Figura 4S).

A concentração de RNA tratado foi determinada, em espectrofotômetro (nanodrop) a 230 nm, utilizando como branco, uma reação preparada com os mesmos reagentes da digestão por DNaseI, mas sem a adição de RNA.

Para a etapa de síntese de cDNA, foi utilizado o Kit SuperScript III (*Invitrogen*). Foram utilizados na reação 500ng de RNA tratado, 1 μL de 50 μM Primer Oligo (dT)₂₀, 1 μL de 10mM dNTP mix e água DEPC para completar um volume final de 10 μL . A reação foi incubada a 65°C por 5 minutos e logo em seguida colocada no gelo por pelo menos 1 minuto. Em seguida foram adicionados 2 μL de 10x RT buffer, 4 μL de 25mM MgCl_2 , 2 μL de 0,1M DTT, 1 μL de 40U/ μL RNase OUTTM e 1 μL de 200U/ μL SuperScriptTM III. A reação foi incubada a 50°C durante 50 minutos, depois a 85°C por 5 minutos e colocada no gelo para resfriar. A transcriptase

reversa (RT) é uma enzima de origem viral, um retrovírus, que tem a capacidade de produzir uma molécula de DNA fita simples a partir da cópia de uma molécula de RNA mensageiro. Após a síntese, a RT deixa um gancho no final 3'OH livre da molécula, possibilitando o acesso da DNA polimerase e formação de segunda fita de DNA (Griffiths, 2002). O RNAm utilizado como molde foi degradado pela ação da RNase H, com a adição de 1µl da enzima seguida de incubação a 37°C por 20 minutos.

As amostras de cDNA foram submetidas a mesma avaliação de qualidade que as amostras de RNA quando tratadas com DNaseI (Figura 5S). Por fim, as amostras foram armazenadas em freezer a -20°C.

4.2.9 Detecção da Expressão das GmHsp20s por PCR Qualitativo

Previamente às análises de expressão, foram realizados testes por PCR qualitativo para os 47 modelos gênicos selecionados, avaliando-se a especificidade e melhor temperatura de pareamento dos 47 oligonucleotídeos iniciadores desenhados. Para realizar os testes foi utilizado 60 ng de DNA genômico de soja. Para as misturas de reação foram utilizados 2 µL de DNA a [30ng/µL], 17,8 µL de água Milli-Q, 2,5 µL de tampão de PCR [10x], 1 µL de dNTPs Mix a [2,5mM], 1 µL de MgCl₂ a [50mM], 0,5 µL PCR primer, *forward* e *reverse*, a [10µM] e 0,2 µL Taq DNA Polimerase [5U/µL], tendo como volume final de reação 12,5 µL.

As misturas de reação foram incubadas em termociclador para a amplificação dos fragmentos utilizando-se o seguinte programa: 94°C por 1', seguido de 30 ciclos de 94°C por 45", (variação)°C por 30" e 72°C por 1' e um passo adicional de 72°C por 10 minutos. As reações de PCR foram feitas em duplicata e sempre acompanhadas de reações controle (branco), sendo testadas três diferentes temperaturas de pareamento (60, 62 e 64°C) para cada oligonucleotídeo.

Os produtos da amplificação foram analisados por eletroforese convencional, sendo aplicado como marcador junto as amostras 4 µl de *ladder* de 50 pb (*Invitrogen*), usado para inferir o tamanho do fragmento obtido na PCR (Figura 6S). Os géis de agarose 1,5%, foram preparados com tampão TAE 1x, e corado com brometo de etídio (10 mg/mL). Para a aquisição e armazenamento das imagens foi utilizado o equipamento transluminador e fotodocumentador Locus.

Uma vez determinadas as condições ideais de amplificação, os 47 iniciadores foram utilizados em PCR qualitativa para todos os tratamentos, a fim de identificar modelos gênicos com expressão não induzida pelo choque térmico, e identificar genes responsivos ao choque térmico e/ou ao estresse biótico (Figura 7S). Para cada gene e tratamento analisado, foram preparadas triplicatas com três repetições biológicas na placa de PCR com 96 poços. As misturas de reação foram preparadas com: 2 μL de cDNA [diluição 1:2], 6,1 μL de água Milli-Q, 2,5 μL de tampão de PCR [10x], 0,7 μL de dNTPs Mix a [2,5mM], 1 μL de MgCl_2 a [50mM], 1 μL PCR iniciador, *forward* e *reverse*, a [10 μM] e 0,2 μL de Taq DNA Polimerase [5U/ μL] – Kit *Invitrogen*, tendo como volume final de reação 12,5 μL . Todas as PCRs tiveram como controle reação, reações negativas onde não adicionou-se cDNA ou qualquer outro tipo de molécula molde.

As misturas de reação foram incubadas em termociclador para a amplificação dos fragmentos utilizando-se o seguinte programa: 94°C por 1', seguido de 30 ciclos de 94°C por 45", 60°C por 30" e 72°C por 1' e um passo adicional de 72°C por 10 minutos. Os produtos da amplificação foram analisados por eletroforese convencional, sendo aplicado como marcador junto as amostras 4 μL de *ladder* de 50 pb, usado para inferir o tamanho do fragmento obtido na PCR. Os géis de agarose 1,5%, foram preparados com tampão TAE 1x, e corado com brometo de etídio (10 mg/mL). Para a aquisição e armazenamento das imagens foi utilizado o equipamento transluminador e fotodocumentador (Loccus).

4.2.10 Análise do Nível de Expressão - RTq-PCR

A análise do perfil de expressão dos modelos gênicos foi realizada pela técnica de RT-qPCR. A quantificação relativa, para cada gene e tratamento analisado, foi realizada utilizando triplicatas técnicas de cada uma das três repetições biológicas. As reações de PCR em tempo real foram conduzidas em termociclador ABI PRISM® 7500 Sequence Detection System (Applied Biosystems) utilizando o kit SYBR® Green PCR Master Mix (Applied Biosystems), conforme as instruções do fabricante. Como normalizador foi utilizado o gene de β -actina (acesso nº: GMU60500), validado por Stolf-Moreira e colaboradores (2011), como controle endógeno para normalizar as quantidades iniciais de cDNA oriundo de experimentos com plantas de soja

A eficiência da reação de amplificação para cada par de iniciadores dos genes estudados e do normalizador foi estimada através de uma série de diluições do cDNA (10 a 10^{-4}), onde foram testadas diferentes concentrações finais dos primers ($0,14 \mu\text{M}$, $0,16 \mu\text{M}$, $0,2 \mu\text{M}$, $0,24 \mu\text{M}$). Os valores de Ct obtidos foram plotados em função do logaritmo das diluições, e o índice de inclinação da reta foi utilizado para calcular a eficiência de cada sistema de acordo com a fórmula $E = [10^{-1/\text{slope}}]^{-1}$. Os parâmetros de ciclagem utilizados para as reações foram: ativação da enzima “hotstart” a 95°C por 10 min, seguidos de 40 ciclos de 95°C por 15 seg para o pareamento e extensão a 60°C por 1 min, sendo coletados os dados nesta última etapa. Adicionalmente foi conduzida uma etapa de desnaturação lenta com elevação da temperatura de $1^\circ\text{C}/\text{min}$ para a construção da curva de dissociação e verificação de possível amplificação inespecífica. Considerando que os fragmentos gerados apresentam temperatura de desnaturação específica é possível detectar a presença de amplificação inespecífica em uma mesma reação.

A reação de quantificação relativa foi realizada em um volume final de $12,5 \mu\text{L}$, adicionando-se $6,25 \mu\text{L}$ de SYBR Green PCR Master Mix 2X (composto pelas enzimas uracil DNA-N-glicosilase e AmpliTaq Gold, MgCl_2 , dNTPs, KCl e o fluoróforo SYBR green); de $0,14$ a $0,24 \mu\text{M}$ de cada *iniciador* R e F; $1 \mu\text{L}$ de cDNA na concentração 10^{-1} , $1 \mu\text{L}$ fluoróforo ROX na concentração 1:10 e água milliQ para completar o volume final. Os parâmetros de ciclagem para as reações de amplificação foram: 50°C por 2 min, 95°C por 10 min, seguidos de 40 ciclos de 95°C por 2 min, 60°C por 1 min, sendo coletados os dados nesta última etapa.

A determinação dos níveis de expressão dos genes alvos foi realizada pela quantificação relativa utilizando a expressão $2^{-\Delta\Delta\text{Ct}}$ (LIVAK e SCHMITTGEN, 2001). Para cada tratamento foi detectado o valor de Ct (Cycle threshold), tanto para o gene alvo quanto para o normalizador. Este valor representa o ponto em que o sinal de amplificação é detectado. O valor de RQ será calculado utilizando-se o método ΔCt . Assim, o nível de expressão gênica (RQ) é calculado pela subtração do Ct da amostra alvo de cada tratamento pelo Ct do controle endógeno, para normalizar a reação. Assim, obtém-se o valor de ΔCt . O valor de ΔCt do tratamento é subtraído do valor do ΔCt do amostra controle (calibrador), encontrando o valor de $\Delta\Delta\text{Ct}$. Este valor será empregado na fórmula do nível de expressão (RQ), $x^{(-\Delta\Delta\text{Ct})}$ onde “x” representa a somatória da eficiência do gene alvo (se $100\%=1$) e do controle endógeno (se $100\%=1$) (LIVAK e SCHMITTGEN,

2001). As eficiências dos genes alvo e endógeno devem ser próximas de 100%, não obrigatoriamente, pode-se trabalhar com eficiências menores, no entanto, elas devem ser próximas. Para as análises estatísticas será aplicado o teste T com nível de significância igual a 5%.

Os dados de expressão relativa foram submetidos no programa SigmaPlot 12.0 para a construção dos gráficos de dispersão utilizando a ferramenta ScattarPlot Horizontal Error Bar.

5 RESULTADOS E DISCUSSÃO

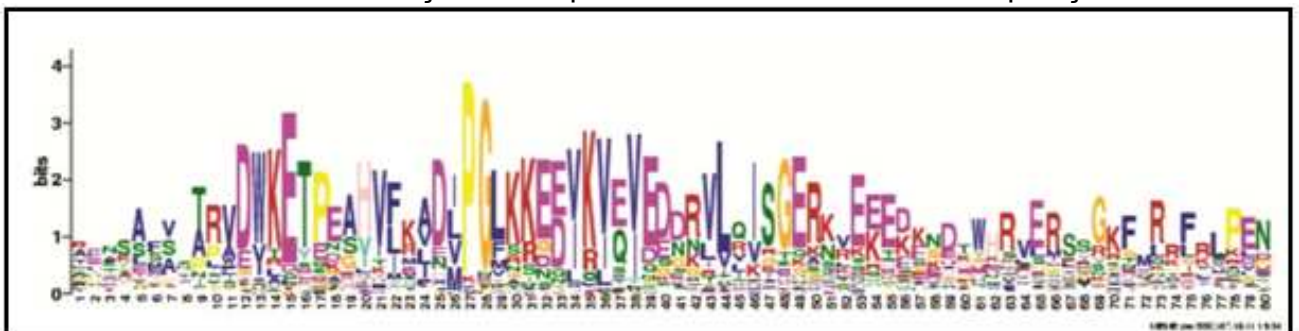
5.1 IDENTIFICAÇÃO DE GMHSP20 EM SOJA

5.1.1 Estudo da Região Codificadora

As duas estratégias de busca por genes Hsp20 no genoma da soja, busca por similaridade da sequência perfil HMM e busca por palavra-chave, retornaram 73 modelos gênicos (sequências preditas de genes) a partir do banco de dados *Superfamily* e 74 a partir do banco de dados do *Phytozome*. No total, 76 modelos gênicos preditos foram identificados como candidatos para Hsp20 no genoma da soja.

Todos os candidatos inicialmente identificados e analisados por métodos estatísticos apresentaram uma sequência conservada de 80 aminoácidos, similar ao domínio Alfa Cristalino, com valores de *p-value* bastante significativos e *E-value* geral de $3.3e^{-1785}$ (Figura 10a).

Figura 10a - Logo obtido no MEME/MAST para o Modelo de Higger Markov referente ao domínio ACD ($3.3e^{-1785}$) encontrado nas proteínas GmHsp20 preditas. O *E-value* do motivo é uma estimativa do número esperado de motivos com a mesma largura e número de ocorrências que estariam presentes em um conjunto de tamanho similar de sequências aleatórias. A altura dos símbolos em cada pilha no motivo indica a conservação da sequência de aminoácidos nessa posição.



O domínio α -cristalino ou ACD, característico a todos os membros da família de HSP20, foi descoberto há quase um século e caracterizado em proteínas estruturais cristalinas presentes na lente dos olhos de vertebrados (MIIRNER 1894 *apud* De JONG *et al.*, 1993). Posteriormente, Wistow (1985)

constatou a presença do mesmo domínio identificado nas proteínas cristalinas de vertebrados, na região C-terminal de sequências de HSP20. O domínio que ficou conhecido como ACD consiste em uma sequência conservada de aproximadamente 80 aminoácidos que compõem a estrutura terciária de uma subunidade α -cristalina. Esse domínio está presente em muitas proteínas, entre elas as Hsp20 (De JONG *et al.*, 1993). O posicionamento do domínio ACD nas 76 HSP20 preditas foi identificado, em todos os casos, por valores significativos de *p-value* ($5.59e-88 \leq p\text{-value} \leq 3.03e-08$) (Figura 10b).

Figura 10b - Mapa gráfico obtido pela análise no programa MEME na busca pelo domínio ACD em cada sequência de aminoácidos. As barras azuis indicam o posicionamento do domínio ACD em cada um dos 76 modelos gênicos candidatos a Hsp20 de *Glycine max L.* Na escala o número 0 indica a região N-terminal e o 400 a região C-terminal. Os valores de *p-value* são encontrados logo na coluna (*lowest p-value*) a direita da coluna de glymas.(*name*).

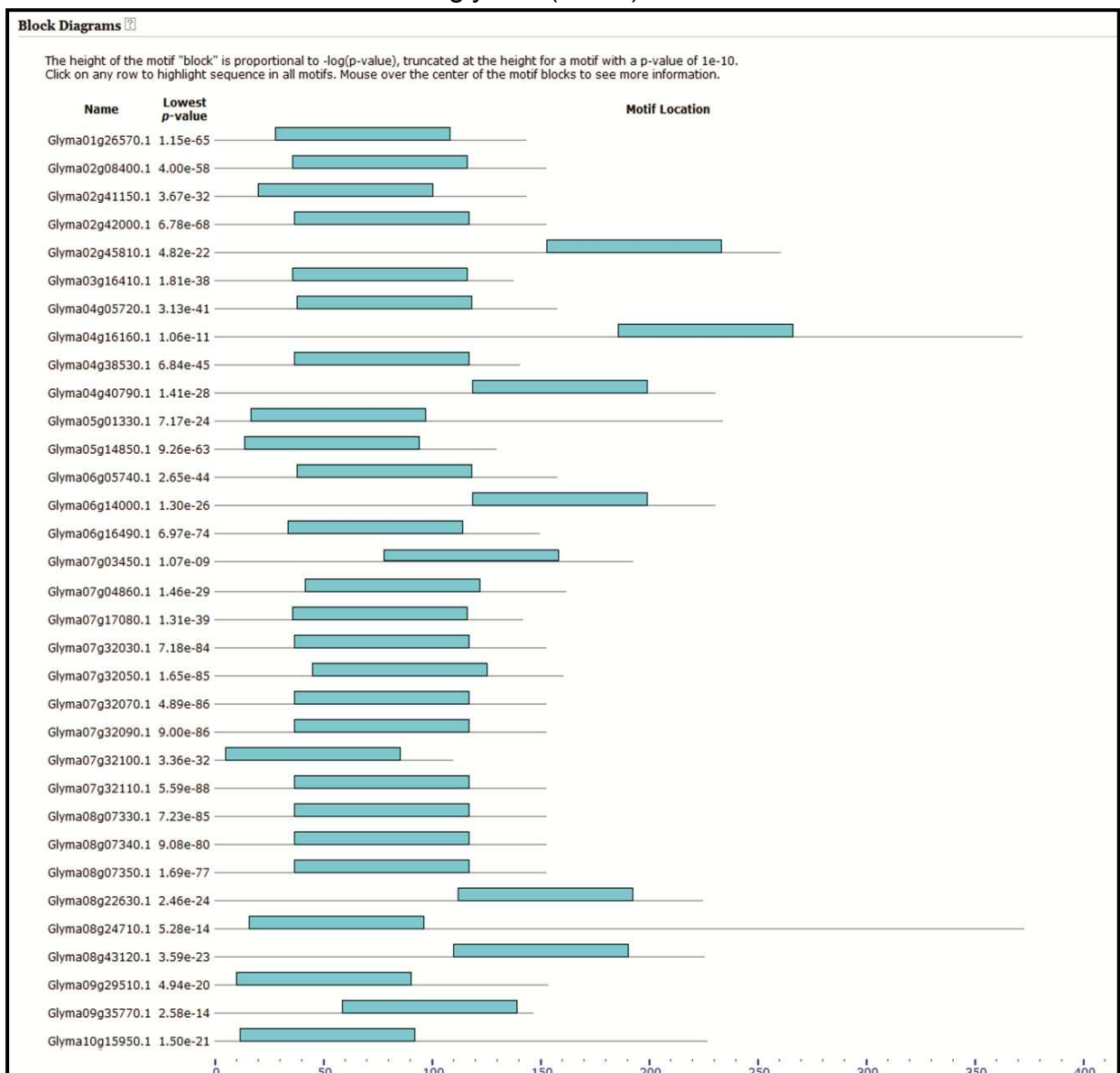
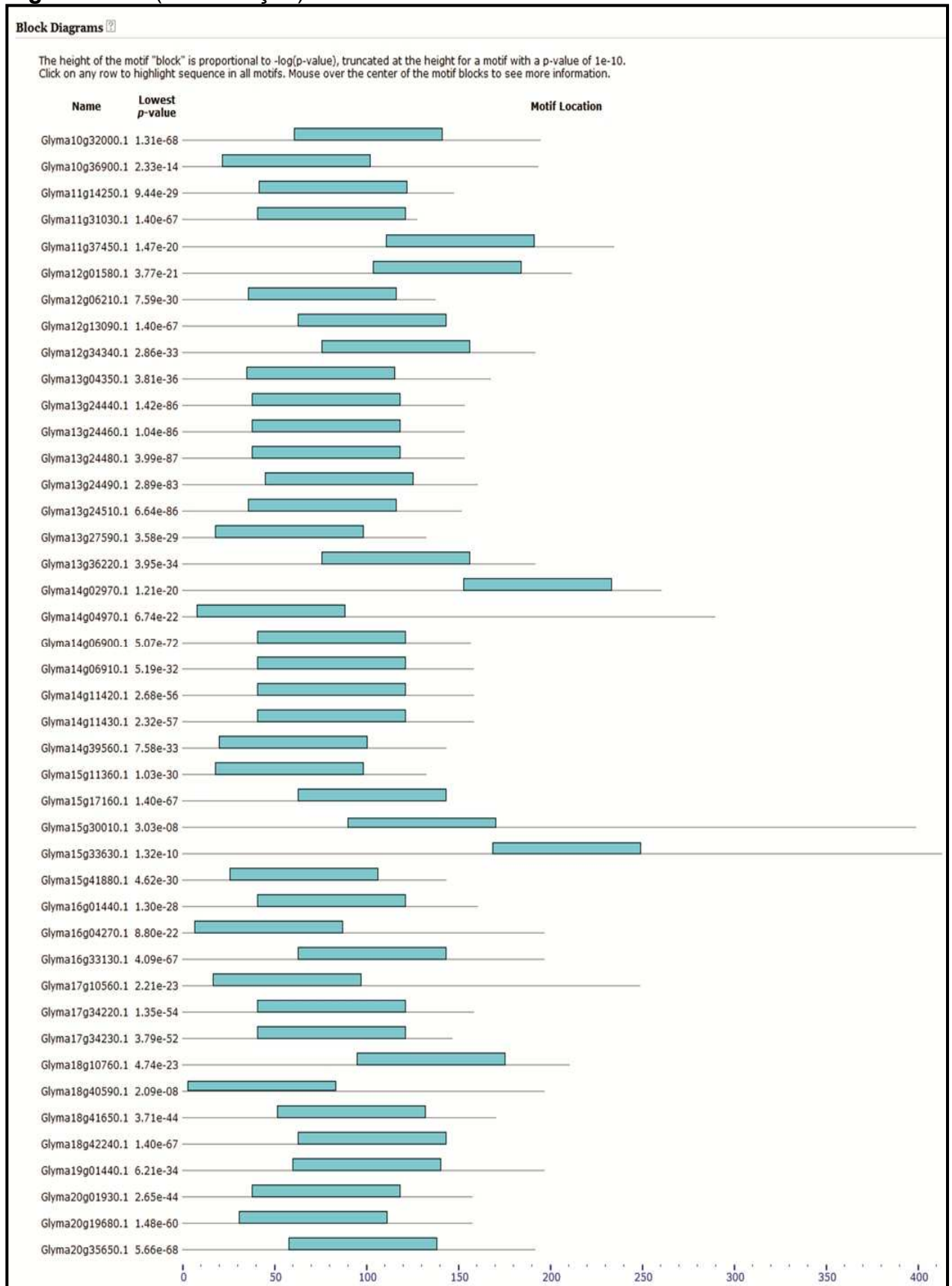


Figura 10b – (Continuação)



Quanto à posição do ACD nas proteínas Hsp20 de soja preditas, foi observado que a conservação na região C-terminal não se manteve. A configuração de posição do ACD, entre o centro da proteína e o braço N-terminal, não foi descrita na literatura em genes Hsp20 de *Arabidopsis* e arroz (SIDDIQUE et al., 2008; OUYANG et al., 2009; SARKAR et al., 2009). Dois genes com perfil de expressão sob choque térmico, já comprovado por ensaios de microarranjo (Geninvestigator – WESTON et al., 2011): glyma13g27590 e glyma15g11360, apresentaram seu ACD na região entre o centro da proteína e o braço N-terminal, nunca na extremidade N-terminal. Em *Arabidopsis* e arroz, duas espécies nas quais as HSP20 são bem estudadas, o domínio ACD se mantém em uma região conservada C-terminal em todas as subfamílias (SIDDIQUE et al., 2008, SARKAR et al., 2009; OUYANG et al., 2009). BASHA e colaboradores (2006) investigaram o papel do domínio C-terminal-cristalino e do braço de N-terminal em relação as variáveis interações das Hsp20 com outras proteínas. Essa pesquisa desmontou inicialmente que a variação da região N-terminal, que precede o ACD, em comprimento e sequência de aminoácidos é normalmente encontrada, e que contribui de forma importante para a diversidade estrutural entre diferentes subfamílias de Hsp20. Algumas proteínas preditas analisadas nesse trabalho, com configuração de posição do ACD na extremidade N-terminal já foram descritas como não induzidas sob choque térmico em experimentos de microarranjo (Geninvestigator – WESTON et al., 2011), como por exemplo, os modelos gênicos: glyma14g04970 e glyma16g04270. Segundo BASHA e colaboradores (2006), o braço N-terminal das Hsp20, além de contribuir para a diversidade estrutural das Hsp20, é importante tanto para sua atividade quanto para a especificidade de interação. Assim, a posição variável do domínio ACD nas proteínas HSP20 de soja, mesmo na região entre o centro da proteína e o braço N-terminal não impede sua indução sob estresse de calor. Contudo, quando o ACD é encontrado na extremidade N-terminal, deixando a região N-terminal insignificante, a proteína não demonstra ser induzida sob estresse de calor, provavelmente porque não tem atividade de HSP20.

5.1.2 Estudo da Região Promotora

As análises de predição de núcleos HSE na região de 500 pb à montante do sítio de início da transcrição foram realizadas apenas para os modelos

gênicos com a região 5'UTR predita disponível no *Phytozome*. Dos 76 modelos gênicos para GmHsp20, apenas 53 deles apresentavam suas regiões 5'UTR preditas e disponíveis no *Phytozome*, e por isso tiveram suas sequências à montante analisadas. Desses 53 modelos gênicos, 49 apresentaram pelo menos um elemento HSE predito, dentro do intervalo de pares de bases estabelecido como padrão das Hsp20, por ao menos um dos programas utilizados na análise (Figuras 11a,11b e 13). Além dos 49 com HSE, foram encontradas três exceções ao padrão de disposições dos elementos HSE até -500 pb. Os modelos gênicos Glyma13g27590, Glyma02g45810 e Glyma15g11360 apresentaram os *cis* elementos para ligação dos fatores de choque térmico muito acima do filtro utilizado para a seleção das Hsp20 de soja. Apenas o programa MatInspector detectou a presença desses elementos nessas posições (Figura 12).

Figura 11a - Logos de Modelo de Higger Markov obtidos no MEME/TOMTOM baseados em sequências de consenso dos promotores das 53 Hsp20 preditas para a soja com região a montante disponível. Os motivos obtidos pela análise MEME foram plotados de acordo com suas posições dentro dos locais e suas sequências consenso foram apresentadas como gráficos usando HMM. As ocorrências de motivo são classificadas por *P-value* e alinhados uns com os outros. O E-valor de motivo é uma estimativa do número esperado de motivos com a mesma largura e número de ocorrências que estariam presentes em um conjunto de tamanho similar de sequências aleatórias. A altura de símbolos em cada pilha no motivo indica a conservação da sequência nessa posição. As sequências foram manualmente destacadas para indicar os HSEs perfeitos (caixa vermelha) consenso reconhecido: (nGAAnnTTCnnGAAn ou nTTCnnGAAnnTTCn); caixa amarela, mostrando módulo imperfeito do HSE. Três motivos foram encontrados, sendo os elementos HSE perfeitos representados pelos motivos 1 (Matches to Query 1) e 3 (Matches to Query 3) na figura.

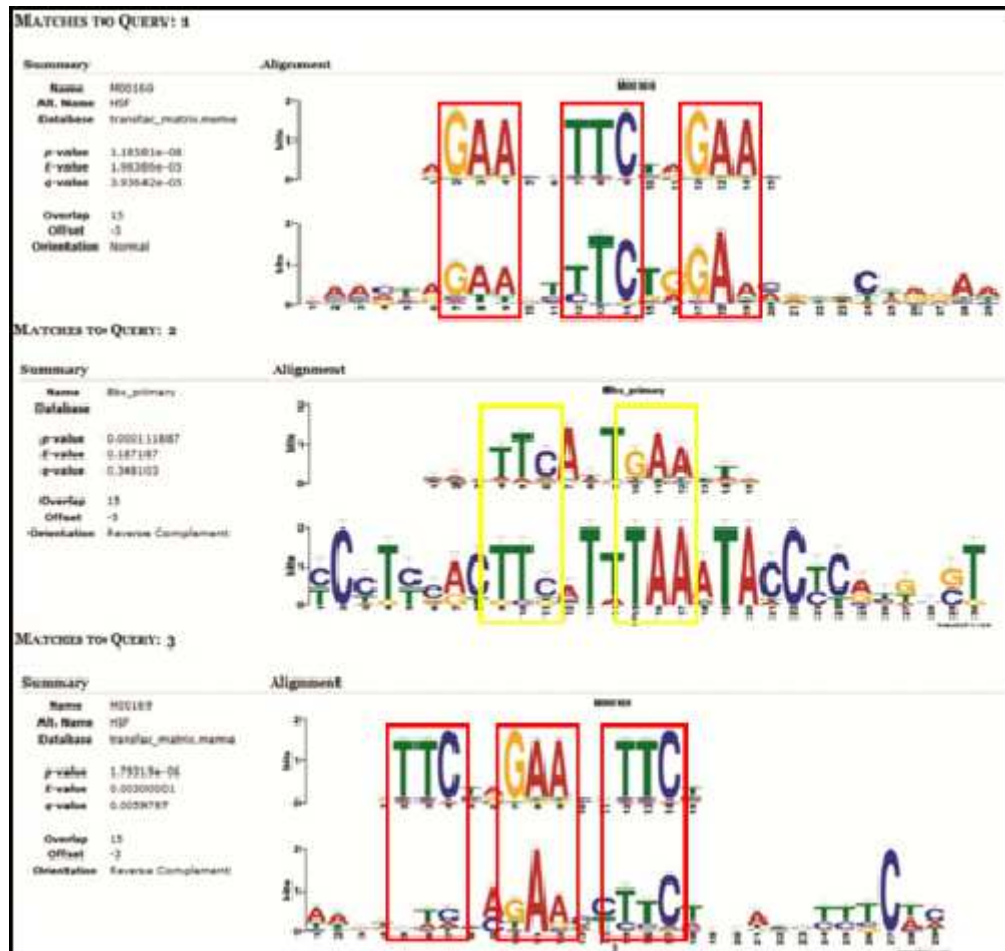


Figura 11b - Mapa gráfico obtido pela análise no programa MEME, na busca pelos cis elementos HSE, em cada sequência de aminoácidos. As barras azul claras e vermelhas indicam o posicionamento dos elementos HSE da região promotora em cada um dos 53 modelos gênicos candidatos a Hsp20 de *Glycine max L*, que apresentam região da montante disponível no Phytozome. Na escala, o número 0 indica a região -1 pb à montante, e 500, a região -500 pb à montante do gene. Os valores de *p-value* são encontrados logo na coluna (*lowest p-value*) à direita da coluna de glymas.(*name*). Apenas os motivos encontrados na fita positiva foram considerados, obedecendo a orientação de leitura do gene. O HMM para cada motivo pode ser visualizado na figura 11a.

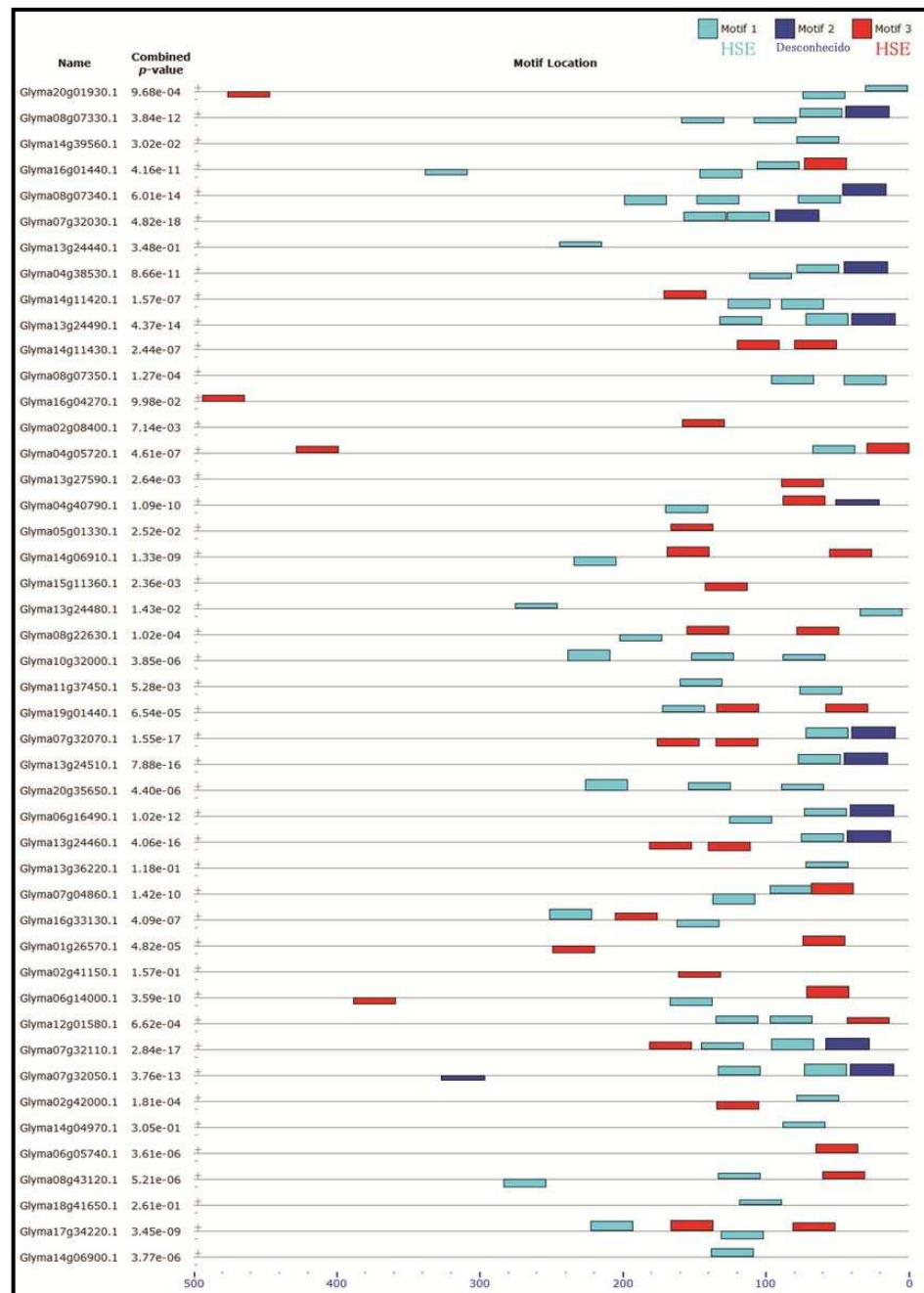
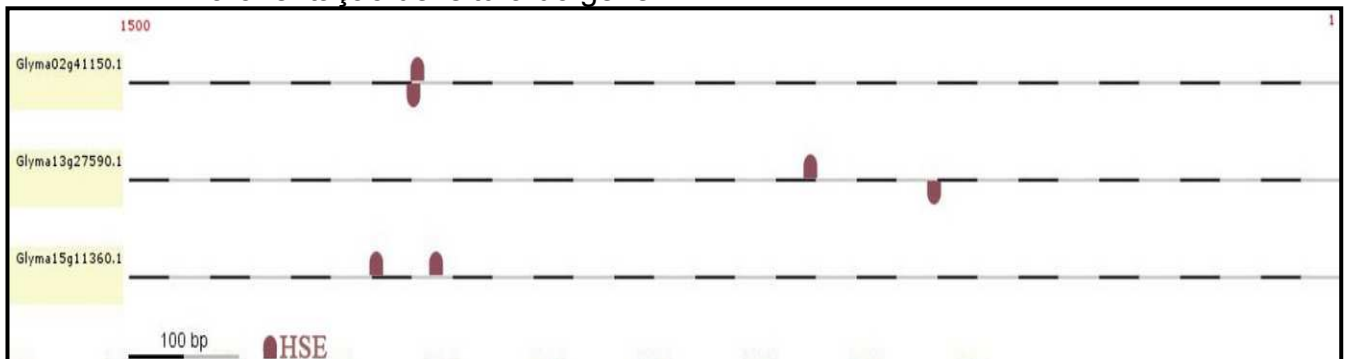


Figura 12- Ilustração da predição dos sítios de ligação para o fator de transcrição de choque térmico nas regiões promotoras dos modelos gênicos de soja: Glyma02g45810, Glyma13g27590 e Glyma15g11360 pelo programa Matisnpector. Para estes modelos gênicos, os HSEs foram encontrados em posições fora dos padrões normalmente encontrados para as Hsp20. Foram analisados até 1500 pb a montante do sitio de início transcrição predito. As marcas em marrom no decorrer das fitas indicam a posição dos HSE em cada sequência. Na escala o número 1 indica o -1 pb ao sítio de início da transcrição e o número 1500 indica -1500 pb. Apenas os motivos encontrados na fita positiva foram considerados, obedecendo a orientação de leitura do gene.



Os elementos HSE são sequências consenso 5'nGAAnnTTCnnGAAn3' encontradas na região promotora, geralmente acima do TATAbox dos genes responsivos a choque térmico. Uma homologia de, pelo menos, 70 a 80% com a sequência consenso HSE é suficiente para induzir a transcrição dos genes responsivos ao choque térmico, e tais elementos podem estar repetidos de uma a três vezes nessa região, e localizados até -350pb à montante (RASCHKE *et al.*, 1988; CZARNECKA, *et al.*, 1989). Na soja, os *cis* elementos HSE foram também identificados ($p \text{ value} \leq 0.001$) em repetições de 1 a 3 por sequência em 49 modelos gênicos (Tabela 1) dentro do intervalo de -500 pb à montante, dos quais 4 contém domínios HSE incompletos ($p \text{ value} > 0.001$) de acordo com as análises realizadas utilizando o pacote de programas MEME.

É interessante notar que nos modelos gênicos o núcleo HSE está, de forma geral, distante até -150 bases à montante do sítio de início da transcrição enquanto para outros o domínio aparece até -500 bases à montante, salvo exceções acima de -500 pb (Figuras 11a, 11b e 13). O artigo mais recente sobre caracterização de Hsp20 de soja analisou as sequências codificadoras e reguladoras de apenas dois genes, GmHsp18.5-C e GmHsp17.9-D. Nesse trabalho, as sequências foram obtidas utilizando a informação para duas proteínas de choque

térmico previamente identificadas (CZARNECKA *et al.*, 1984 *apud* RASCHKE *et al.*, 1988). Os autores utilizaram essas sequências como base para o desenho de sondas, que foram utilizadas para a identificação de proteínas de choque térmico homólogas no genoma da soja. Duas a nove repetições de elementos HSE foram encontradas, em torno de -130 bases à montante do sítio de iniciação da transcrição para as duas únicas sequências estudadas (RASCHKE *et al.*, 1988). Em arroz, os núcleos HSE foram encontrados variando em número de 2 a 9 dentro da região promotora dos genes Hsp20 induzidos por tratamento térmico (OUYANG *et al.*, 2009).

O número de repetições de núcleos HSE parece não ter forte influência sobre a expressão dos genes Hsp20 em resposta ao estresse de calor. Entretanto, estudos atuais sugerem que o posicionamento de tais elementos em relação ao promotor cerne TATAbox exerça forte influência sobre sua ativação em resposta a estresse biótico (ESCOBAR *et al.*, 2003). Nesses promotores têm sido demonstrado que o posicionamento do primeiro núcleo de elementos HSE parece ter importante relação com a expressão de Hsp20 em resposta a infecção por nematóides. Escobar e colaboradores (2003) demonstraram que o Hsp20 HaHsp17.7G4 responde a infecção por nematóide e que a presença de um único elemento HSE já é suficiente para sua indução, desde que o mesmo ocorra até -83pb do sítio de início da transcrição. Esses resultados foram comprovados por transformação em *Arabidopsis thaliana* com construções contendo diferentes configurações de posição e número de núcleos HSE fusionado ao gene *gus* (BARCALA *et al.*, 2008). Os resultados da busca por *cis* elementos HSE, na região promotora dos Hsp20 preditos em soja, demonstraram que, em geral, esses elementos se concentram dentro de uma região em torno de -150 pb, mas aparentemente, existem algumas exceções em que o único elemento HSE da região regulatória localiza-se a cerca de -500 pb (Glyma02g41150) ou mais (Figura 11a, 11b e 13 e Tabela 1). Alguns modelos gênicos de soja já foram identificados como responsivos a infecção por nematóides: Glyma08g07340, Glyma14g06910 e Glyma10g32000 pelo nematóide *H. glycines* (KANDOTH *et al.*, 2011), e Glyma13g24460 encontrado sob estresse por infecção pelo nematóide *M. javanica* (FUGANTI *et al.*, 2010). Nestes modelos, o primeiro elemento HSE encontra-se distante até -88pb. Já o Glyma04g05720, também identificado como responsivo a infecção por nematóides (KANDOTH *et al.*, 2011), apresenta o primeiro núcleo HSE

na posição -48 bases. Nesse sentido, parece que o posicionamento dos elementos HSE nos Hsp20 de soja responsivos ao estresse por infecção com nematóide obedece ao padrão previamente identificado, e conseqüentemente, outros Hsp20 de soja, para os quais ainda não há evidências de indução por estresse devido a nematóides, foram identificados utilizando o padrão estabelecido por Escobar e colaboradores (2003), e Barcala e colaboradores (2008).

Figura 7 - Ilustração da predição dos sítios de ligação para o fator de transcrição de choque térmico nas regiões promotoras dos 53 modelos gênicos de soja, que apresentam região da montante disponível no Phytozome, pelo Matisnpector. Foram analisados até 500 pb a montante do sitio de início transcrição predito. As marcas em marrom no decorrer das fitas indicam a posição dos HSE em cada sequência. Na escala o número 1 indica o - 1 pb ao sítio de inicio da transcrição e o número 500 indica -500 pb. Apenas os motivos encontrados na fita positiva foram considerados, obedecendo a orientação de leitura do gene

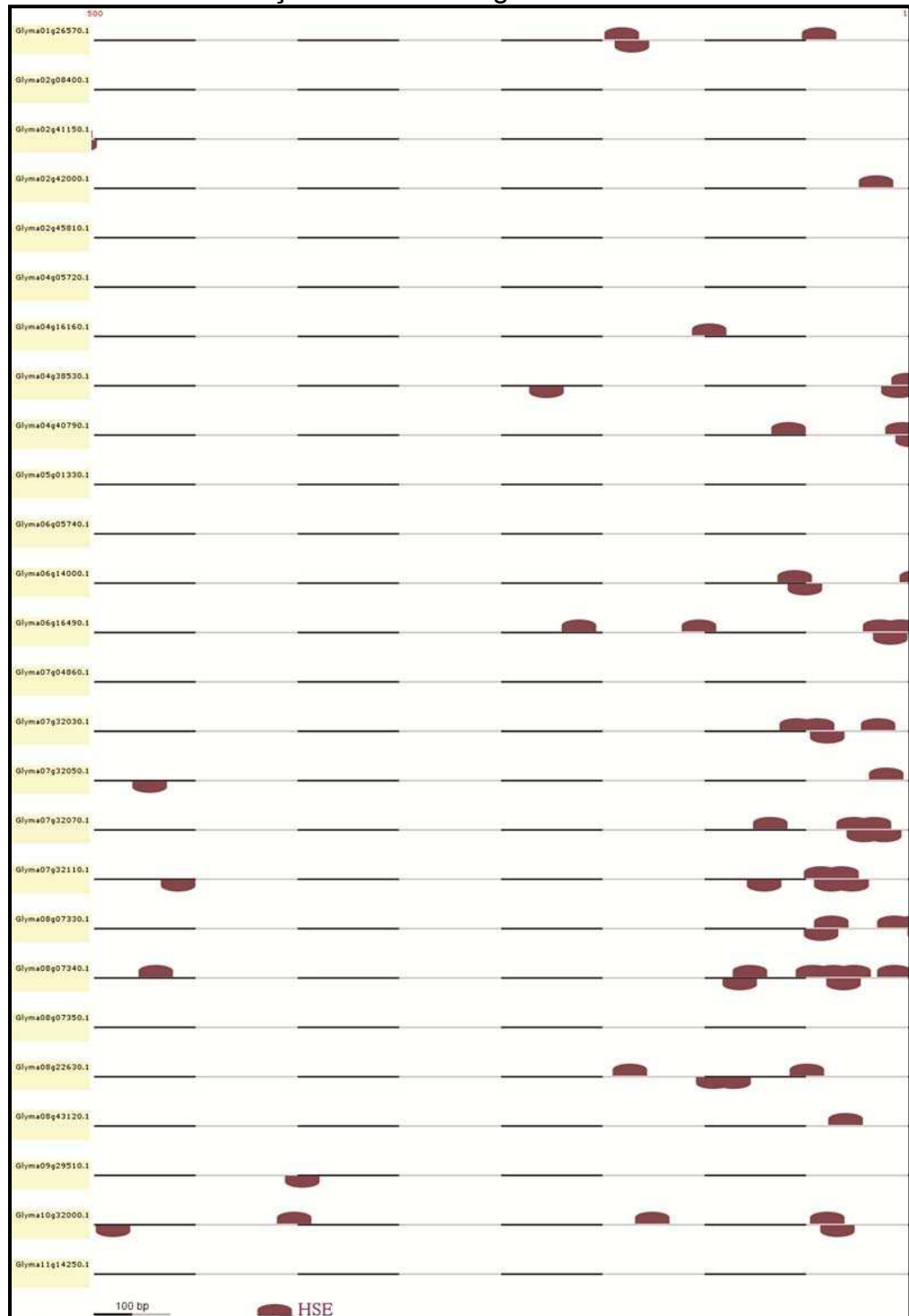
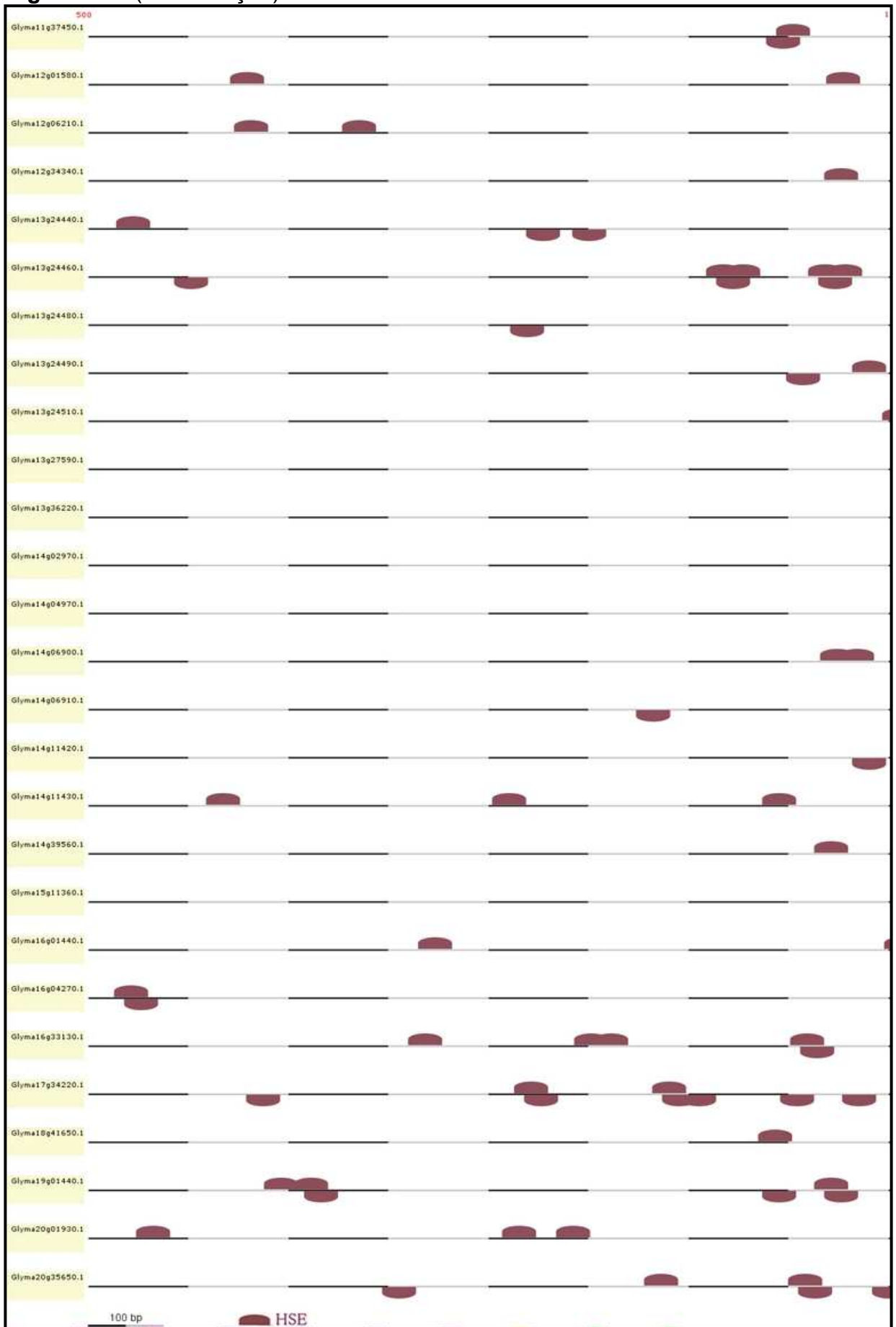


Figura 13 – (Continuação)



5.1.3 Análise de Dados de Expressão Digital das Hsp20 em Soja

Para a investigação dos reais genes Hsp20 de soja, foi ainda realizada uma análise de expressão digital para os 76 modelos gênicos candidatos em três bancos de dados de expressão de soja objetivando filtrar entre os candidatos aqueles cuja expressão não tenha sido observada em nenhum outro experimento prévio e que, portanto, possam representar erros de montagem do genoma da soja ou genes não funcionais. A abordagem foi baseada em análise de perfil de expressão digital de ESTs identificadas no atlas de expressão de soja (*Glycine // A//* RNA-Seq Atlas) (SEVERIN *et al.*, 2010), dados de expressão de bibliotecas subtrativas em estresses de seca e fatores bióticos depositados no Banco de Dados do Projeto Genosoja-Consórcio Nacional para Estudo do Genoma da Soja ([// A// P://bioinfo03.ibi.unicamp.br/soja/](http://bioinfo03.ibi.unicamp.br/soja/)) e dados de pelo menos 140 experimentos de microarranjo disponíveis no Genevestigator (Hruz *et al.*, 2008), que incluem análises de expressão realizadas em diferentes fases do desenvolvimento da soja, com diferentes genótipos e sob diferentes estresses bióticos e abióticos. Um sinal positivo foi atribuído ao modelo gênico, quando o mesmo apresentou nível significativo de expressão nos bancos de dados utilizado na busca (Tabela 1). A seleção por nível de expressão, nos três bancos de dados utilizados, resultou na eliminação de 24 modelos gênicos.

Após as análises de predição de domínio ACD, predição de HSE na região montante e identificação do nível de expressão em experimentos prévios disponíveis em BDs, apenas 52 genes dos 76 inicialmente identificados como genes candidatos foram classificados como potenciais GmHsp20 (Tabela 1), incluindo os 3 modelos gênicos com HSE acima de -500 pb à montante para as análises *in silico*, e outros 2 modelos gênicos com indução da expressão sob estresse de calor em soja, glyma07g32090 e glyma18g10760 (Figura 14).

As demais análises *in silico* realizadas neste trabalho consideraram apenas os 52 potenciais candidatos em soja, e *in vivo*, foram analisados 47 potenciais candidatos em soja (Tabela 1).

Tabela 1 - Resultados da análise *in silico* para a predição dos critérios mínimos utilizados na seleção dos candidatos potenciais a Gmhs20. Estão apresentados candidatos potenciais em negrito aqueles que possuem domínio ACD, pelo menos um elemento HSE e expressão previamente observada em ao menos um dos bancos investigados. Quando a predição de elementos HSE não foi possível devido à indisponibilidade de sequência no Phytozome, mas quando houve observância dos outros critérios, o modelo gênico foi considerado como candidato potencial. Os resultados de ocorrência e posicionamento dos elementos HSE são apresentados em quatro colunas distintas que correspondem aos programas utilizados na análise. A sigla “nd” significa que a região à montante não está disponível no Phytozome.

Nº.	Nome do Transcrito	Posição de ACD Predita (¶ A)	Região regulatória (500pb) montante	Posição Predita de Ancoragem de HSE na Região Regulatória (¶ A)					Nível de Expressão nos BDs		
				MatIsnpector	MEME		Place	PlantCare	Genosoja	Soybean Soy-seq	Genestigator
				HSE*	HSE1	HSE2	HSE	HSE			
1	*Glyma01g26570.1	29-108	dispon.	-240 -143 -58	_____	-59	_____	_____	+	+	+
2	*Glyma02g08400.1	37-116	dispon.	_____	_____	-143	_____	_____	-	+	-
3	*Glyma02g41150.1	21-100	dispon.	-510	_____	_____	_____	_____	+	+	+
4	*Glyma02g42000.1	38-117	dispon.	-115	-63	_____	_____	_____	-	+	-
5	*Glyma02g45810.1	154-233	dispon.	-1142	_____	_____	_____	_____	-	-	+
6	Glyma03g16410.1	37-116	não dispon..	nd	nd	nd	nd	nd	-	-	-
7	*Glyma04g05720.1	39-118	dispon.	-44	-52	-15 -413	_____	_____	+	+	-
8	Glyma04g16160.1	187-266	dispon.	-197	_____	_____	_____	_____	-	-	-
9	*Glyma04g38530.1	38-117	dispon.	-99	-63	_____	_____	-244 -364	+	-	-
10	*Glyma04g40790.1	120-199	dispon.	-158 -102	_____	-73	_____	_____	+	+	+
11	Glyma05g01330.1	18-97	dispon.	_____	_____	-151	_____	_____	-	-	-
12	Glyma05g14850.1	15-94	não dispon..	nd	nd	nd	nd	nd	-	-	-
13	*Glyma06g05740.1	39-118	dispon.	-64	_____	-50	_____	_____	+	+	-
14	*Glyma06g14000.1	120-199	dispon.	-155 -95	_____	-56	_____	_____	+	+	-
15	*Glyma06g16490.1	35-114	dispon.	-261 -202 -113 -103 -50	-58	_____	-47	-413	+	+	-
16	*Glyma07g03450.1	79-158	não dispon..	nd	nd	nd	nd	nd	+	-	-

Os modelos gênicos com asterisco (*Glyma) são os potenciais candidatos a Hsp20 de soja. O modelos gênicos em negrito (***Glymas**) foram selecionados, como descritos na metodologia, para o desenho de oligonucleotídeos e análises *in vivo*.

Tabela 1 – Continuação.

Nº.	Nome do Transcrito	Posição de ACD Predita (ll A)	Região regulatória (500pb) montante	Posição Predita de Ancoragem de HSE na Região Regulatória (ll A)					Nível de Expressão nos BDs		
				MatInspector HSE*	MEME HSE1 HSE2		Place HSE	PlantCare HSE	Genosoja	Soybean Soy-seq	Geninvestigador
17	*Glyma07g04860.1	43-122	dispon.	-84	-82	-53	_____	_____	+	-	-
18	Glyma07g17080.1	37-116	não dispon..	nd	nd	nd	nd	nd	-	-	-
19	*Glyma07g32030.1	38-117	dispon.	-154 -144 -114	-112 -142	_____	_____	-18 -143	+	+	-
20	*Glyma07g32050.1	46-125	dispon.	-110 -60 -40	-58 -118	_____	_____	_____	+	+	-
21	*Glyma07g32070.1	38-117	dispon.	-167 -126 -116	-43	_____	_____	-322	+	-	-
22	*Glyma07g32090.1	38-117	não dispon..	nd	nd	nd	nd	nd	-	-	+
23	Glyma07g32100.1	6-85	não dispon..	nd	nd	nd	nd	nd	-	-	-
24	*Glyma07g32110.1	38-117	dispon.	-142 -132 -83 -73	-81 -130	-166	_____	_____	-	+	-
25	*Glyma08g07330.1	38-117	dispon.	-137 -106 -96 -63	-47	_____	-134	-85 -402 -95 -228	+	-	-
26	*Glyma08g07340.1	38-117	dispon.	-469 -177 -146 -136 -126 -106 -65	_____	_____	_____	-125	+	+	-
27	*Glyma08g07350.1	38-117	dispon.	-74 -53 -33 -23 -8	_____	_____	_____	-22 -387	+	+	-
28	*Glyma08g22630.1	113-192	dispon.	-236 -149 -52	_____	-63 -140	_____	_____	-	+	-
29	Glyma08g24710.1	17-96	não dispon..	nd	nd	nd	nd	nd	-	-	-
30	*Glyma08g43120.1	111-190	dispon.	-130	-118	-45	_____	_____	+	-	-
31	Glyma09g29510.1	11-90	dispon.	_____	_____	_____	_____	-337	-	-	-
32	Glyma09g35770.1	60-139	não dispon..	nd	nd	nd	nd	nd	-	-	-
33	Glyma10g15950.1	13-92	não dispon..	nd	nd	nd	nd	nd	-	-	-
34	*Glyma10g32000.1	62-141	dispon.	-401 -225 -139 -75	-73 -137 -223	_____	_____	_____	+	+	-
35	Glyma10g36900.1	23-102	não dispon..	nd	nd	nd	nd	nd	-	-	-
36	*Glyma11g14250.1	43-122	dispon.	-147 -64	_____	_____	_____	_____	-	+	+

37	Glyma11g31030.1	42-121	não dispon..	nd	nd	nd	nd	nd	-	-	-
38	*Glyma11g37450.1	112-191	dispon.	_____	-145	_____	_____	_____	+	+	+

Os modelos gênicos com asterisco (*Glyma) são os potenciais candidatos a Hsp20 de soja. Os modelos gênicos em negrito (***Glymas**) foram selecionados, como descritos na metodologia, para o desenho de oligonucleotídeos e análises *in vivo*.

Tabela 1 – Continuação.

Nº.	Nome do Transcrito	Posição de ACD Predita (A)	Região regulatória (500pb) montante	Posição Predita de Ancoragem de HSE na Região Regulatória (A)					Nível de Expressão nos BDs				
				MatsInspector			MEME		Place	PlantCare	Genosoja	Soybean Soy-seq	Genes tigator
				HSE*			HSE1	HSE2	HSE	HSE			
39	*Glyma12g01580.1	105-184	dispon.	-420 -122 -84	-82 -120	-28	_____	_____	+	+	-		
40	*Glyma12g06210.1	37-116	dispon.	-418 -364	_____	_____	_____	_____	+	+	+		
41	Glyma12g13090.1	64-143	não dispon..	nd	nd	nd	nd	nd	-	-	-		
42	Glyma12g34340.1	77-156	dispon.	-123	_____	_____	_____	_____	-	-	-		
43	Glyma13g04350.1	36-115	não dispon..	nd	nd	nd	nd	nd	-	-	-		
44	*Glyma13g24440.1	39-118	dispon.	-477	-229	_____	_____	-107	+	+	+		
45	*Glyma13g24460.1	39-118	dispon.	-182 -172 -131 -121	-60	_____	_____	_____	+	+	-		
46	*Glyma13g24480.1	39-118	dispon.	_____	-234	_____	_____	-138	+	-	-		
47	*Glyma13g24490.1	46-125	dispon.	-109 -59	-57 -117	_____	_____	-281	+	+	-		
48	*Glyma13g24510.1	37-116	dispon.	-94 -84 -64	-62	_____	_____	-84	+	+	+		
49	*Glyma13g27590.1	19-98	dispon.	-656	_____	_____	_____	_____	-	-	+		
50	*Glyma13g36220.1	77-156	dispon.	-42	-57	_____	_____	-41	-	-	+		
51	Glyma14g02970.1	154-233	dispon.	_____	_____	_____	_____	-388	-	-	-		
52	*Glyma14g04970.1	9-88	dispon.	_____	-73	_____	_____	_____	+	-	-		
53	*Glyma14g06900.1	42-121	dispon.	-125 -115	-123	_____	_____	_____	+	+	-		
54	*Glyma14g06910.1	42-121	dispon.	-88	_____	-40 -154	_____	_____	-	+	-		
55	*Glyma14g11420.1	42-121	dispon.	_____	_____	-156	_____	-465 -432	+	+	-		

56	*Glyma14g11430.1	42-121	dispon.	-432 -289 -154	_____	-65 -105	_____	-462	+	+	-
57	*Glyma14g39560.1	21-100	dispon.	-128 -65	-63	_____	_____	_____	+	-	-
58	*Glyma15g11360.1	19-98	dispon.	-1119 -1179	_____	_____	_____	_____	-	-	+
59	Glyma15g17160.1	64-143	não dispon..	nd	nd	nd	nd	nd	-	-	-
60	Glyma15g30010.1	91-170	não dispon..	nd	nd	nd	nd	nd	-	-	-

Os modelos gênicos com asterisco (***Glyma**) são os potenciais candidatos a Hsp20 de soja. O modelos gênicos em negrito (***Glymas**) foram selecionados, como descritos na metodologia, para o desenho de oligonucleotídeos e análises *in vivo*.

Tabela 1 – Continuação

Nº.	Nome do Transcrito	Posição de ACD Predita (ll A)	Região regulatória (500pb) montante	Posição Predita de Ancoragem de HSE na Região Regulatória (ll A)					Nível de Expressão nos BDs		
				MatIsnpector	MEME		Place	PlantCare	Genosoja	Soybean Soy-seq	Genenes tigator
				HSE*	HSE1	HSE2	HSE	HSE			
61	Glyma15g33630.1	170-249	não dispon..	nd	nd	nd	nd	nd	-	-	-
62	*Glyma15g41880.1	27-106	não dispon..	nd	nd	nd	nd	nd	+	-	-
63	*Glyma16g01440.1	42-121	dispon.	-326 -93 -57	-91	-58	_____	_____	+	-	-
64	*Glyma16g04270.1	8-87	dispon.	-478	_____	-479	_____	_____	-	+	-
65	*Glyma16g33130.1	64-143	dispon.	-331 -248 -238 -140 -52	-236	-190	_____	_____	+	+	-
66	Glyma17g10560.1	18-97	não dispon..	nd	nd	nd	nd	nd	-	-	-
67	*Glyma17g34220.1	42-121	dispon.	-278 -209	-207	-66 -151	_____	-390	+	+	-
68	Glyma17g34230.1	42-121	não dispon..	nd	nd	nd	nd	nd	-	-	-
69	*Glyma18g10760.1	96-175	não dispon..	nd	nd	nd	nd	nd	-	-	+
70	Glyma18g40590.1	4-83	não dispon..	nd	nd	nd	nd	nd	-	-	-
71	*Glyma18g41650.1	53-132	dispon.	-156	_____	_____	_____	_____	-	+	+
72	Glyma18g42240.1	64-143	não dispon..	nd	nd	nd	nd	nd	-	-	-
73	*Glyma19g01440.1	61-140	dispon.	-403 -388 -128 -52	-157	-43 -119	_____	_____	+	+	-
74	*Glyma20g01930.1	39-118	dispon.	-467 -284 -257 -52	-15	_____	_____	-234	+	+	-

75	Glyma20g19680.1	32-111	não dispon..	nd	nd	nd	nd	nd	-	-	-
76	*Glyma20g35650.1	59-138	dispon.	-213 -141 -76	-74 -139 -211	_____	_____	-311 -486	+	+	-

Os modelos gênicos com asterisco (*Glyma) são os potenciais candidatos a Hsp20 de soja. O modelos gênicos em negrito (***Glymas**) foram selecionados, como descritos na metodologia, para o desenho de oligonucleotídeos e análises *in vivo*.

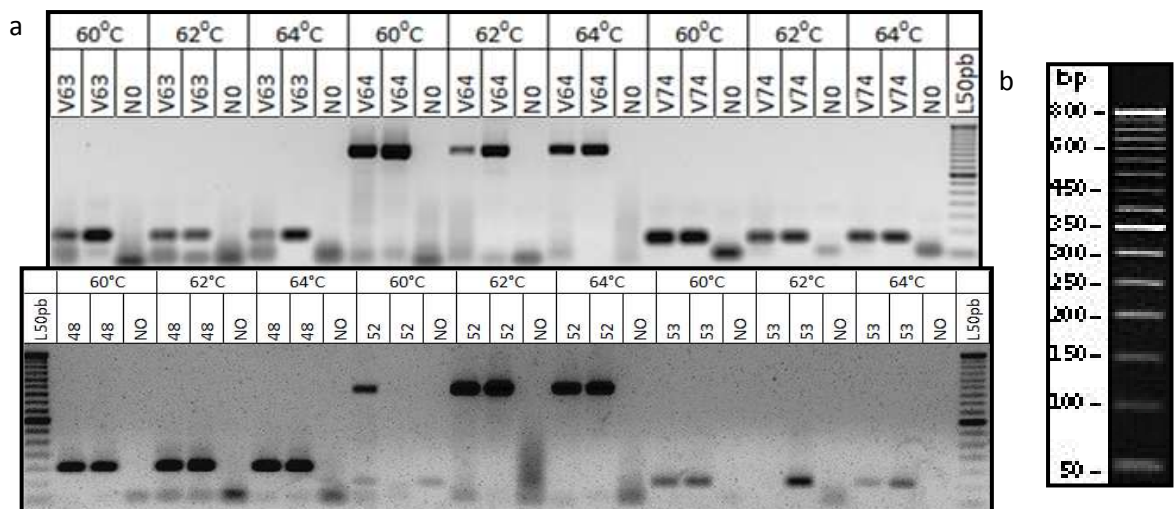
5.2 ANÁLISES “*IN VIVO*” - INDUÇÃO DA EXPRESSÃO DOS HSP20 CANDIDATOS NOS DIFERENTES ESTRESSES

Visando identificar genes candidatos a Hsp20 em soja responsivos a estresse biótico e abiótico, foram avaliadas plantas submetidas a estresse de frio, calor e infecção pelo nematóide *M. javanica* por PCR convencional, para um *screening* inicial da ativação de tais genes frente a estes estresses, seguida da quantificação via PCR quantitativo (RT-qPCR).

5.2.1 Análises Qualitativas

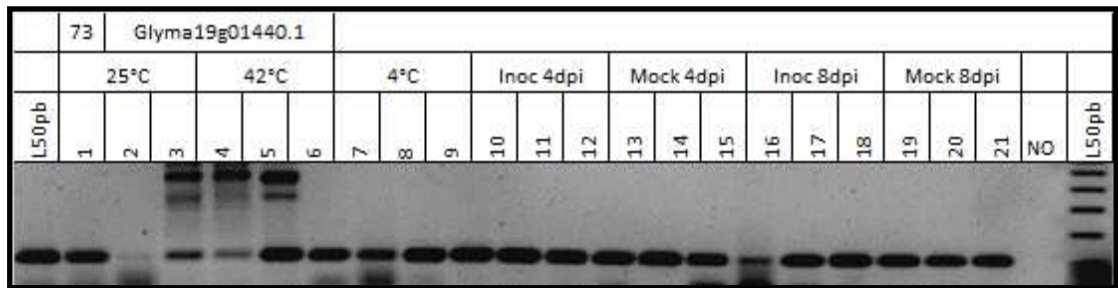
Inicialmente todos os oligonucleotídeos desenhados foram testados quando a especificidade e temperatura de pareamento ideal em análises de PCR convencional. Conforme exemplificado na Figura 15, pode ser observado que a temperatura de 60^oC para o primer V74 (Glyma20g01930) e 64^oC para o primer V48 (Glyma13g24510) foram as ideais. Nenhum dos oligonucleotídeos apresentou bandas inespecíficas utilizando o DNA genômico como alvo. Os primers V52 e V64, desenhados para os modelos gênicos Glyma14g04970 e Glyma16g04270, respectivamente, pareiam um em cada exon. Esse resultado confirma a presença de um íntron de 527bp dentro do gene identificado como Glyma14g04970 e outro de 457pb dentro do gene Glyma16g04270, previsto conforme a amplificação do cDNA, que resultou nas bandas de tamanhos esperado 105 bp para V52 e 116pb para V64 (Figura 15).

Figura 9 - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e temperatura de pareamento) dos *primers* em DNA genômico de soja. Nesta figura estão os resultados dos testes com os oligonucleotídeos para os modelos gênicos V74 (glyma20g01930), V48 (glyma13g24510), V52 (glyma14g04970), V64 (glyma16g04270), V63 (glyma16g01440) e V53 (glyma14g06900). Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para cada par de *primers*, e na segunda linha os *primers* testados seguidos dos controles negativos de reação (NO). Foi utilizado o marcador molecular de 50 pb (*Ladder* 50 pb – *Invitrogen*). b. *Ladder* de 50 pb – *Invitrogen*. Na horizontal, estão indicadas os tamanhos, em pares de base, das principais bandas do marcador.



Embora nenhum íntron tenha sido predito, no banco de dados Phytozome, para o gene identificado pelo glyma19g01440, foram observados duas bandas não específicas na eletroforese dos produtos de PCR com cDNA para as três repetições de estresse de choque térmico a 42°C (Figura 16).

Figura 16- Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de 1 *primer* específicos para 1 modelo gênico: V73 (glyma19g01440). Na primeira linha do cabeçalho de cada foto, está anotada a codificação do *primer* estabelecidas neste trabalho (Vn^o) e o modelo gênico correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4°C (tratamento de frio), Inoc 4dpi (tratamento de infecção por *M.javanica* após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por *M.javanica* após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por *M.javanica* após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por *M.javanica* após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 50 pares de base (*Ladder* 50 pb – *Invitro*).



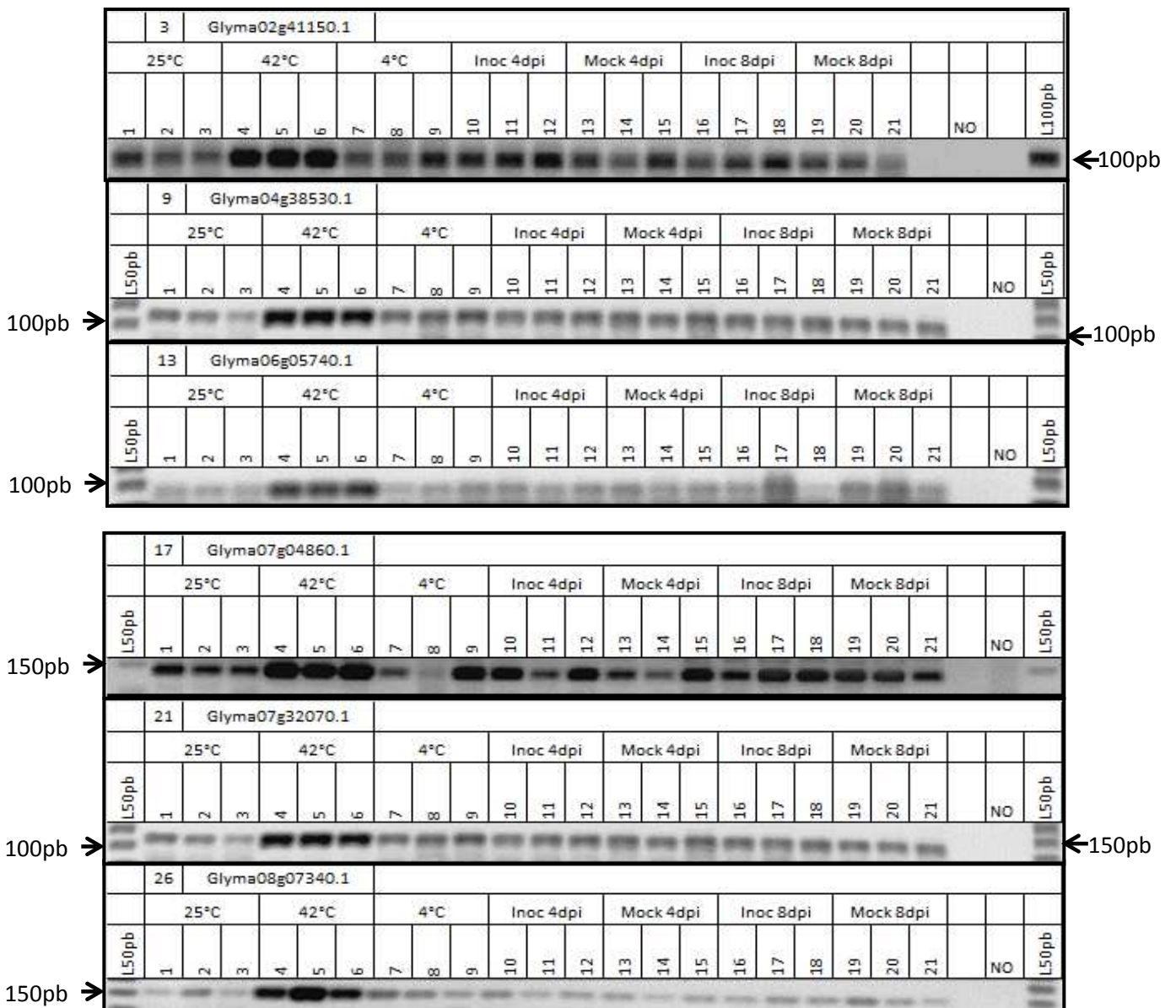
A ocorrência de mais de uma banda para um mesmo gene, em uma determinada condição de estresse, pode ser devido a presença de íntrons no gene, e que *splicings* alternativos ocorrem após a transcrição desse gene, em condição específica, de estresse de calor. A expressão de variantes de *splicing* alternativo de seis genes Hsp20 foi observada em arroz (SARKAR *et al.*, 2009). Segundo Barbazuk e colaboradores (2008) esse tipo de mecanismo já foi observado sob condições específicas de estresse de frio e seca combinado com salinidade.

Em trigo, foi observado que o fator de transcrição WDREB2 pareceu ser controlado por níveis diferenciados de transcritos e também por *splicing* alternativo. Esse fator de transcrição é ativado por vários estresses abióticos para produzir três transcritos diferentes por *splicing* alternativo e que são expressos diferencialmente. Isso porque sua expressão é regulada através de duas vias diferentes, ABA (ácido abscísico) dependente na resposta seca e a salinidade, e ABA independente em resposta ao frio. Sugere-se que mudanças significativas de fatores de *splicing* ocorrem, sob condições de estresse abiótico, e que isso afeta os padrões de *splicing* dos transcritos do gene Wdreb2 (Egawa *et al.*, 2006).

Um total de 40 dos 47 genes que foram analisados *in vivo* apresentaram indução acentuada sob choque térmico, facilmente visualizada em gel de agarose após a eletroforese dos produtos da reação de RT-PCRs convencional, enquanto nenhum para estresse de frio (inconclusivo) e nematóide. É importante salientar que tais dados diferiram quando as amostras foram analisadas via PCR quantitativo, cuja sensibilidade é maior quando comparada ao PCR convencional.

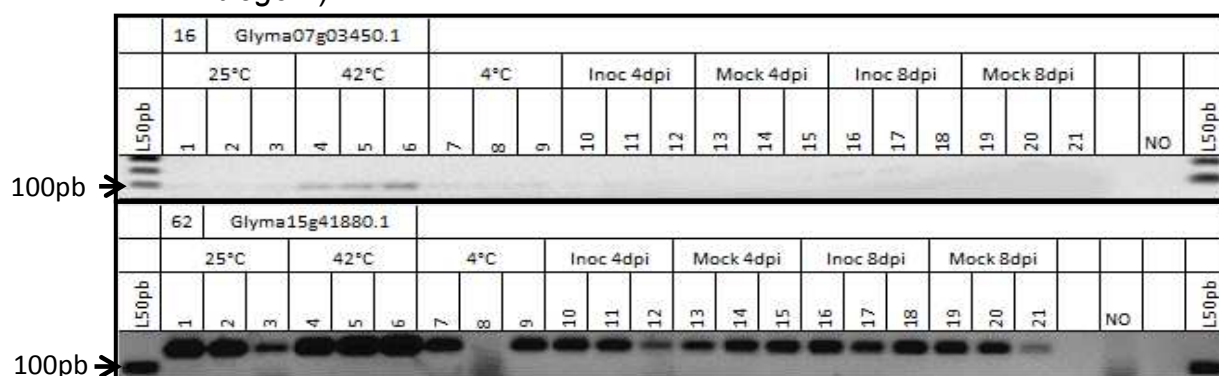
A ativação de 40 genes em resposta ao estresse térmico, sugere que as estratégias adotadas para a seleção *in silico* dos Gmhsp20 candidatos foram eficientes, já que apenas 7 genes não tiveram indução em tecidos de raiz no tratamento de choque térmico (Figura 17, e Figura 7S). Sarkar e colaboradores (2009) e Ouyang e colaboradores (2009) obtiveram resultados semelhantes em testes qualitativos de Hsp20 em amostras arroz sob tratamento de choque térmico em diferentes intensidades.

Figura 10 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de seis *primers* específicos para seis modelos gênicos: V3 (glyma02g41150), V9 (glyma04g38530), V13 (glyma06g05740), V17(glyma07g04860), V21 (glyma07g32070) e V26 (glyma08g07340), que demonstram a indução desses genes sob estresse de calor facilmente visualizada pela análise qualitativa. Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos *primers* estabelecidas neste trabalho (Vn°) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4° (tratamento de frio), Inoc 4dpi (tratamento de infecção por *M.javanica* após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por *M.javanica* após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por *M.javanica* após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por *M.javanica* após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 100 (Ladder 100 pb – *Invitro*) pares de base na primeira foto e o marcador de 50 pb (Ladder 50 pb – *Invitrogen*) para as demais fotos.



A inclusão dos genes, sem região regulatória disponível para as análises *in vivo*, resultou na identificação de outros dois GmHsp20: Glyma07g03450 (16) e Glyma15g41880 (62) (Figura 18), cuja a expressão foi ativa em resposta ao estresse térmico.

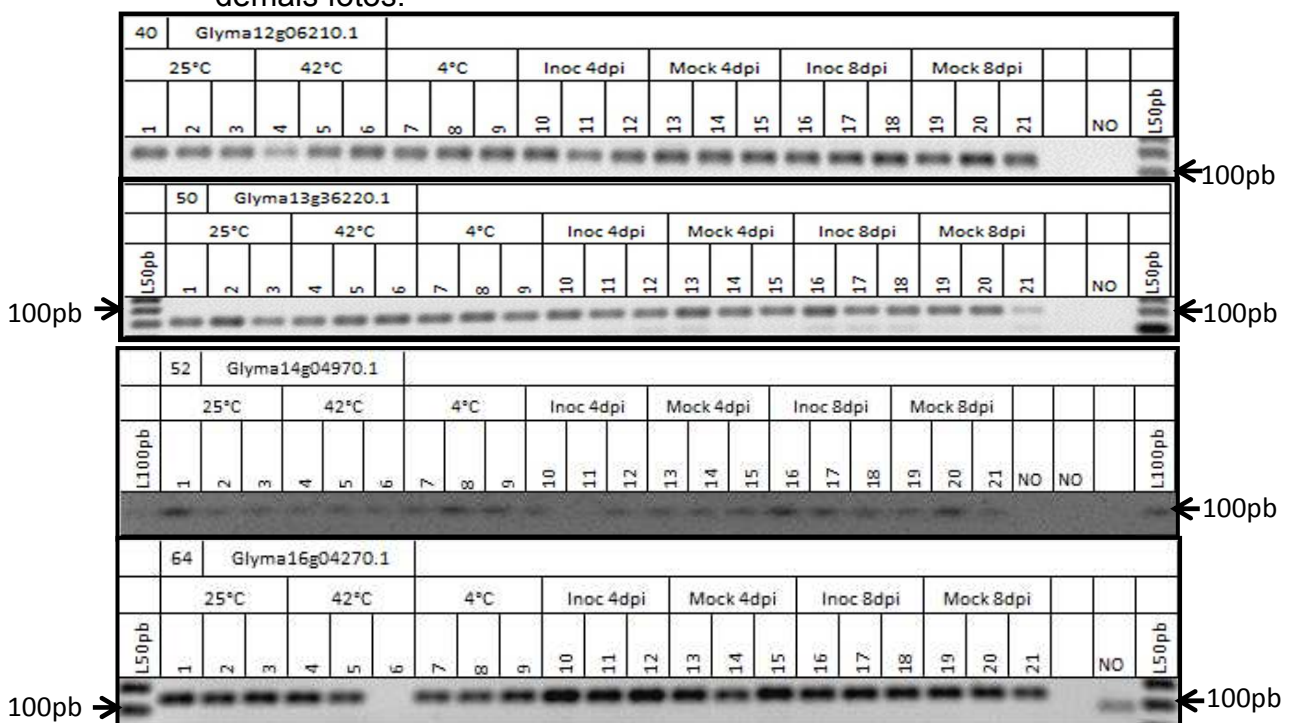
Figura 11 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de dois *primers* específicos para dois modelos gênicos: V16 (glyma07g03450) e V62 (glyma15g41880), que demonstram a indução desses genes sob estresse de calor facilmente visualizada pela análise qualitativa. Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos *primers* estabelecidas neste trabalho (Vn°) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4° (tratamento de frio), Inoc 4dpi (tratamento de infecção por *M.javanica* após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por *M.javanica* após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por *M.javanica* após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por *M.javanica* após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 50 pb (*Ladder 50 pb – Invitrogen.*).



Os resultados das análises de RT-PCR convencional conduzidas para os tratamentos dos ensaios *in vivo*, demonstram que os modelos gênicos V52 (Glyma14g04970) e V64 (Glyma16g04270) não tiveram alteração significativa na expressão em resposta ao estresse de choque térmico em tecido radicular que pudesse ser observada pela análise visual quando comparada ao controle (25°C) (Figura 19). Nas análises de *cis* elementos HSE na região promotora, os genes V52 e V64 apresentam o primeiro HSE nas posições -73 pb (MEME p-valeu acima de 0,001) e -478 pb à montante (MatInspector), respectivamente.

Ainda, as análises das sequências peptídicas demonstram que o ACD está localizado na região da extremidade N-terminal e não na região C-terminal, nem mesmo entre as regiões N-terminal e central, como normalmente pode ser encontrado nas proteínas HSP20 (WISTOW, 1985 *apud* De JONG *et al.*, 1993). A ausência de indução sob estresse de calor, somada ao posicionamento do domínio ACD na extremidade N-terminal nesses dois modelos gênicos, indicam que eles não codificam para HSP20 verdadeiras em soja.

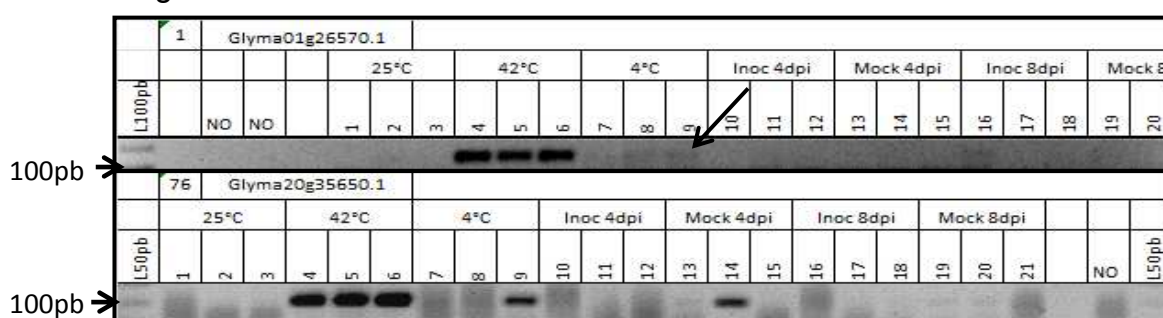
Figura 12 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de quatro *primers* específicos para quatro modelos gênicos: V40 (glyma 12g06210), V50 (glyma13g36220), V52 (glyma07g03450) e V64 (glyma15g41880). Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos *primers* estabelecidas neste trabalho (Vn^o) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4° (tratamento de frio), Inoc 4dpi (tratamento de infecção por *M.javanica* após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por *M.javanica* após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por *M.javanica* após 8 dias de inoculação), Mock 8 dpi (tratamento falso inoculado por *M.javanica* após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 100 pares de base (*Ladder 100 pb – Invitro*) para o gel de V52 e o marcador de 50 pb (*Ladder 50 pb – Invitrogen*) para as demais fotos.



Outros cinco genes: V2 (glyma02g08400), V36 (glyma11g14250), V40 (glyma12g06210), V50 (glyma13g36220) e V71 (glyma18g41650) também mantiveram padrão de expressão semelhante entre as amostras de choque térmico e o controle no RT-PCR convencional. Tais genes, entretanto, possuem elementos HSE preditos e apresentam domínio ACD C-terminal, o que os caracteriza como Hsp20. Nesse caso, a ausência de indução sob choque térmico nas raízes de soja possivelmente indique que esses três modelos gênicos codifiquem para HSP20 com função não relacionada ao estresse de calor ou que são induzidas por este estresse em outros tecidos da planta que não a raiz. Em arroz, por exemplo, a expressão de quatro genes OsHsp20 é detectada em condições de 25⁰C, mas não sob choque térmico. O resultado dessa hipótese pode ser conferido no item seguinte - Análises Quantitativas.

O tratamento de frio parece ter tido influência muito menor que o de calor sobre a expressão das GmHsp20 preditas, já que nas análises qualitativas não foi possível notar, seguramente, nenhuma indução a 4⁰C. Os resultados obtidos via RT-PCR convencional para o estresse de frio não puderam ser considerados como conclusivos (Figura 20), devido a baixa sensibilidade e níveis baixos de ativação. Resultado semelhante foi encontrado por SARKAR e colaboradores (2009) em experimento com plantas de arroz submetidas ao estresse de frio (4⁰C por 3h), onde apenas três OsHsp20 (Hsp16.9C-CI, Hsp18.0-CI e Hsp23.6-MII) foram induzidas.

Figura 20 - Eletroforese em gel de agarose 1,2% dos produtos das reações de RT-PCR convencional tendo como alvo o cDNA. Nesta figura estão os resultados das análises de quatro *primers* específicos para quatro modelos gênicos: V1 (glyma01g26570) e V76 (glyma20g35650). Na primeira linha do cabeçalho de cada foto, estão anotadas as codificações dos *primers* estabelecidas neste trabalho (Vn^o) e os modelos gênicos correspondentes. Na segunda linha, foram delimitadas as repetições biológicas (números na terceira linha) de cada tratamento, sendo 25°C (tratamento controle de temperatura), 42°C (tratamento de choque térmico), 4°C (tratamento de frio), Inoc 4dpi (tratamento de infecção por *M.javanica* após 4 dias de inoculação), Mock 4dpi (tratamento falso inoculado por *M.javanica* após 4 dias de inoculação), Inoc 8dpi (tratamento de infecção por *M.javanica* após 8 dias de inoculação), Mock 8dpi (tratamento falso inoculado por *M.javanica* após 8 dias de inoculação) e o controle negativo da reação (NO). Foi utilizado o marcador molecular de 100 (Ladder 100 pb – *Invitro*) pares de base para o gel de V1 o marcador de 50 pb (Ladder 50 pb – *Invitrogen*) para o gel de V76.



Em relação ao estresse biótico, a indução ou repressão de nenhum gene pôde ser confirmada utilizando a técnica de RT-PCR convencional. O tratamento de estresse biótico pode ter uma influência ainda menor sobre a expressão das GmHsp20 preditas. Essa aparente discrepância pode ser explicada em função da própria diferença na amplitude dos estresses aplicados. No choque térmico, a planta toda foi submetida a condição de estresse de forma que todas as células da raiz deveriam estar respondendo ao estresse. Já no estresse biótico, somente algumas células da raiz são afetadas pela infecção pelo nematóide o que pode dificultar o acesso aos transcritos correspondentes aos genes responsivos a infecção.

Além disso, a obtenção de resultados de expressão gênica diferencial pela técnica de PCR convencional possibilita apenas uma comparação entre amostras por meio da visualização de géis de eletroforese. Assim, essa metodologia funciona relativamente bem quando as diferenças entre concentração

de transcritos do gene de interesse são grandes entre as amostras, mas não tem bons resultados quando há apenas diferenças sutis. Isso acontece porque a PCR convencional é analisada somente após o término da reação, na fase platô, onde a reação já não ocorre exponencialmente. Ao final da reação, a eficiência já não é de 100%, pois pode haver escassez de reagentes (como *primers* e nucleotídeos), e a enzima DNA polimerase pode ter sua atividade diminuída ou totalmente inibida por subprodutos formados durante a reação. Além disso, é preciso correr as amostras em gel de eletroforese e o resultado obtido, é analisado visualmente ou por detectores de diferença de intensidade pouco sensíveis. Todo esse processo apresenta muitos interferentes que tornam variável a correlação entre a quantidade inicial de alvo com a quantidade de alvo amplificado no decorrer dos ciclos de amplificação (BRASIL *et al.*, 2010).

O RT-qPCR, por outro lado, possibilita o monitoramento da reação ciclo a ciclo. A emissão de fluorescência ocorre a cada ciclo de amplificação, sendo diretamente relacionada com o aumento da quantidade do produto gerado. Os dados são coletados na fase exponencial da reação, onde a quantidade de fragmento alvo amplificado é proporcional à quantidade de material genético utilizado no início da reação. Além disso, o uso do gene normalizador possibilita o ajuste da quantidade de cDNA de cada amostra para que então a comparação entre as amostras tratadas e controles possa ser feita (MARCELINO, 2006).

5.2.2 Análises Quantitativas

Os resultados das análises de eficiência, com a padronização da melhor concentração de cada um dos 47 oligonucleotídeos específicos para os 47 modelos gênicos candidatos, estiveram entre 80 e 99,82%, com curvas de dissociação mostrando apenas um pico bem definido. Somente o oligonucleotídeo específico para o modelo gênico 13g36220 (V50) não pôde ser aplicado para as análises de RT-qPCR, pois seus resultados de eficiência não estiveram dentro dos padrões lineares aceitáveis. O gene normalizador apresentou eficiência de 93,98% quando utilizada a concentração final em cada reação de 0,32 μ M dos oligonucleotídeos. A concentração final de oligonucleotídeos específicos nas

reações de RT-qPCR para os 47 genes Hsp20 candidatos esteve no intervalo entre 0,12 μ M e 0,2 μ M para o *forward* e *reverse*.

Os resultados de expressão do estresse de calor relativo ao controle nas análises de RT-qPCR foram coerentes com os obtidos pela técnica de PCR convencional, com 40 genes induzidos dos 47 avaliados. A expressão dos genes Hsp20 durante o estresse de calor é muito acentuada, assim pôde ser avaliada mesmo pela técnica convencional, resultado apresentado no item anterior. Esses 40 genes, confirmados e inseridos como responsivos ao choque térmico, somados aos 5 genes já descritos por experimentos analisados por microarranjos disponíveis no Geninvestigator, formam a família de GmHsp20 de soja descrita nesse trabalho. Os demais 7 modelos gênicos candidatos, não foram induzidos por choque térmico. São os glymas 02g08400, 11g14250, 12g06210, 14g04970, 16g04270, 18g41650 e 13g36220. Com exceção dos glymas 16g04270 e 18g41650, que foram responsivos ao estresse por infecção pelo nematóide *M javanica*, é provável que os outros cinco genes logo acima apresentem uma expressão constitutiva em órgãos vegetativos, estando envolvidos em atividade de manutenção celular ou em outra condição específica, que não a de choque térmico.

O *primer* V50 não funcionou para as análises de RT-qPCR, mas seguiu o padrão obtido nas análises de PCR convencional para os genes que não foram induzidos sob choque térmico. Além disso, todos os outros genes descritos como não induzidos nas análises de PCR convencional demonstraram o mesmo padrão no RT-qPCR. Considerando essas comparações, foi inferido que o gene 13g36220 provavelmente, também não responde a estresse de choque térmico. Além disso, as análises de expressão gênica digital em experimentos de microarranjos de plantas de soja submetidas a choque térmico não foram conclusivas quanto a sua indução. Assim, esse gene não foi incluso nas análises *in silico* de distribuição em subfamílias, estrutura secundária, duplicação gênica e outras. Contudo, caso seja posteriormente comprovado um perfil de indução de expressão sob estresse de calor, esse gene será o único representante da subfamília CVI de soja de acordo com resultado de análises investigativas à parte realizadas para esse trabalho.

A existência de genes controlados por promotores tecido-específicos já foi comprovada. Swindell e colaboradores (2007), por exemplo,

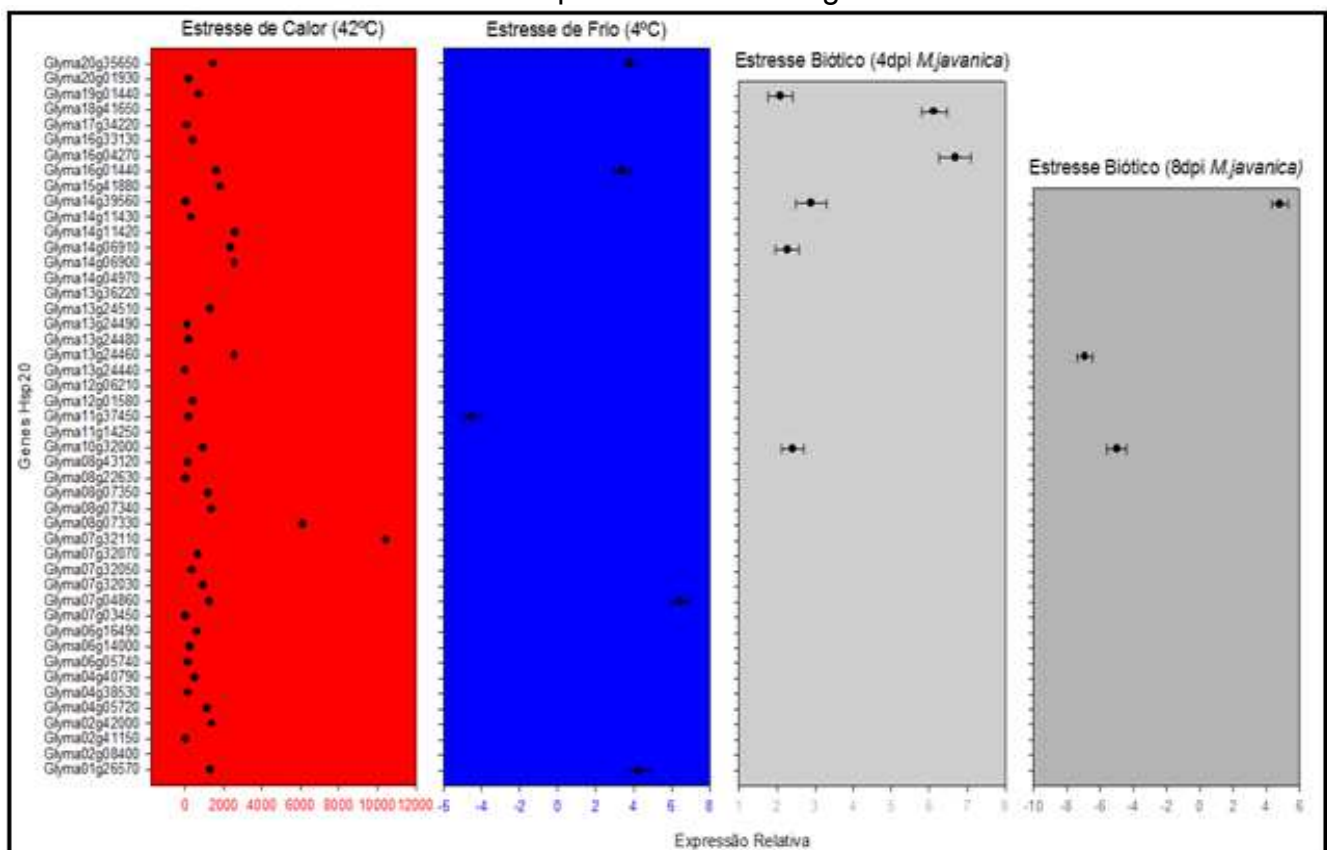
analisaram os padrões de expressão de Hsp20 de *Arabidopsis*, sob estresse de calor, em diferentes tecidos, e obtiveram padrões de expressão diferentes quando comparados entre em tecidos vegetais distintos. Sarkar e colaboradores (2009) observaram a expressão de dois genes de arroz, OsHsp18.8-CV em flores e OsHsp19.0-CII em pistilo, sugerindo que a expressão destes genes Hsp20 pode ser regulada diferencialmente entre os tipos de tecidos. Considerando esses resultados em pesquisas anteriores, análises de RT-qPCR de tecidos foliares submetidos ao choque térmico foram conduzidas para observar se a ausência de indução nas raízes desses 7 genes se mantém nas folhas de soja. Os resultados dessa análise, realizada apenas por RT-qPCR, demonstram que não há discrepância de indução dos genes Hsp20 entre os tecidos de raiz e folha na soja sob estresse de choque térmico – calor.

Para os demais estresses, frio e infecção por *M.javanica*, a indução ou repressão de genes pôde ser notada apenas por RT-qPCR, devido à sensibilidade da técnica. Sob estresses de frio, a análise de expressão dos 47 genes resultou na identificação de 5 genes responsivos, todos os cinco também induzidos sob estresses de calor. Os genes induzidos são identificados pelos glymas 01g26570, 07g04860, 16g01440 e 20g35650.1, e o gene reprimido 11g37450. As análises semi-quantitativas conduzidas por Sarkar e colaboradores (2009) em arroz, resultaram em 3 genes Hsp20 relacionados ao frio. Contudo, os autores não obtiveram resultados compatíveis nas análises de microarranjo. A análise por blastP, das proteínas de soja responsivas a estresse por frio, contra o banco de dados de proteínas de arroz no Phytozome resultou em 69,1% de identidade entre a proteína codificada pelo gene (glyma01g26570) em soja e a proteína HSP16. 9C-CI de arroz. Outra proteína codificada pelo gene (glyma07g4860) apresentou 49,1% de identidade com a proteína HSP18.0-CI de arroz, todas induzidas sob estresse de frio.

Os resultados de RT-qPCR para os tratamentos de estresse biótico por infecção pelo nematóide *M.javanica*, resultaram na identificação de 7 genes responsivos, cinco deles também induzidos sob estresses de calor (Gráfico 1). Os genes identificados pelos glymas 10g32000, 13g24460, 14g06910, 14g39560 e 19g01440 foram induzido em pelo menos um dos tratamentos de estresse biótico e ao estresse de calor, enquanto os genes identificados pelos glymas 16g04270 e

18g41650 tiveram indução no tratamento de 4 dpi, mas não foram induzidos por choque térmico.

Gráfico 1 - Resultado da análise de expressão relativa por RT-qPCR dos genes candidatos Hsp20 no genótipo suscetível BRS133 nos tratamentos de estresse abiótico (calor/vermelho e frio/azul) e biótico (4dpi/cinza claro e 8dpi *M.javanica*/cinza escuro). Estão plotados apenas os genes foram significativos ao nível de 5%. As escalas são diferentes para cada tipo de estresse, podendo apresentar valor negativos. *Spots* alinhados em escalas positivas indicam genes que foram induzidos significativamente e alinhados em escalas negativas foram reprimidos significativamente. As barras de erro indicam em cada uma das escalas em que está os *spots* a margem de erro para mais ou para menos. Os glymas seguem em linha reta na horizontal para cada um dos gráficos/tratamentos



O gene identificado pelo glyma10g32000 foi induzido no tratamento de estresse biótico em 4 dpi e reprimido em 8 dpi, e o glyma14g06910 foi induzido em 4

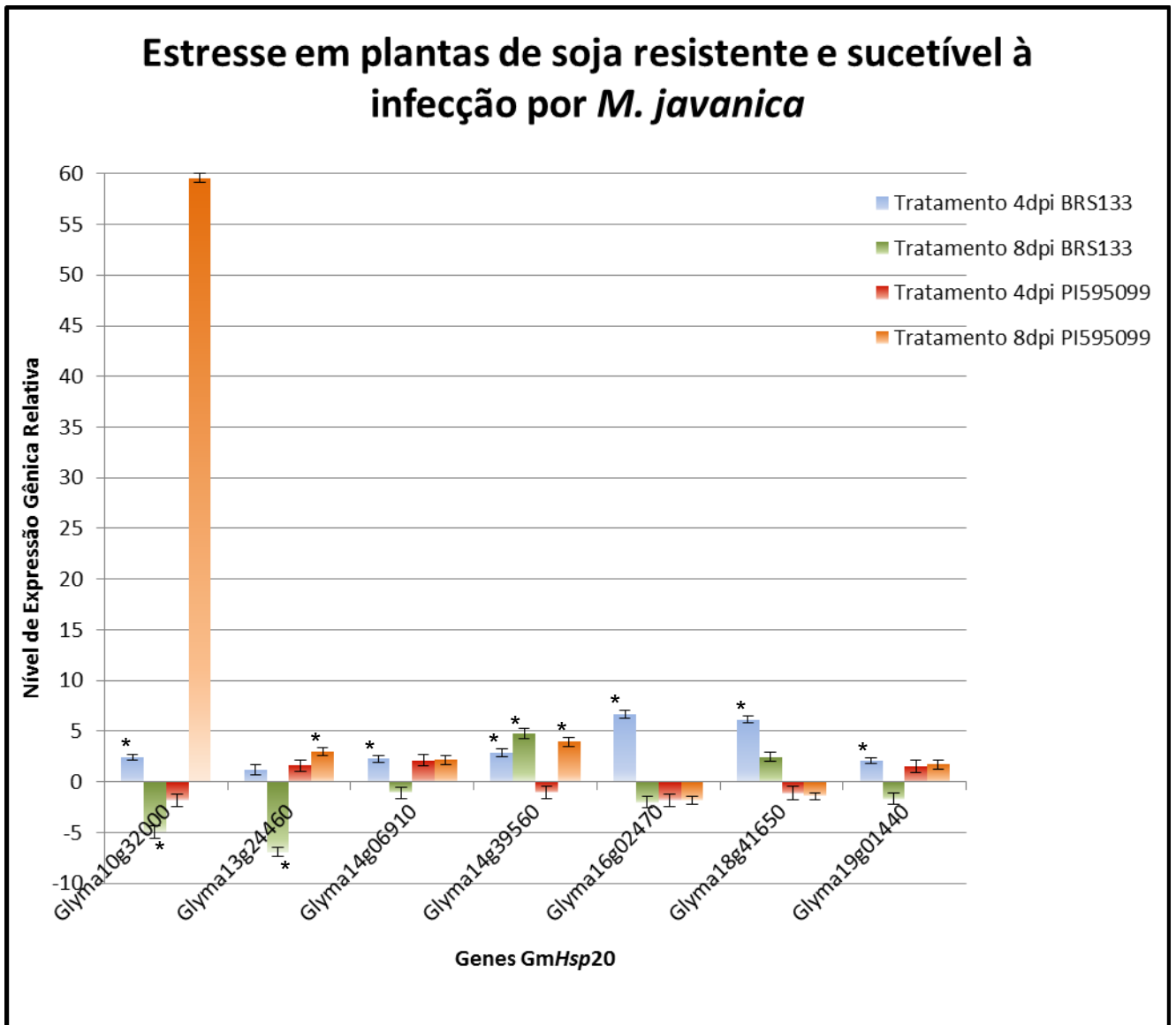
dpi. Esses dois e mais outros dois genes Hsp20 de soja, glyma04g05720, glyma08g 07340 foram descritos por Kandath e colaboradores (2011) como responsivos a infecção por *H. glycines*. Em resposta a estresse biótico em plantas de arroz de 4 dpi por *M. grisea*, foi observada a indução de 4 genes

Hsp17.4-CI, Hsp18.0-CI, Hsp16.9A-CI, Hsp18.0-CII, enquanto outros 4 genes, todos pertencentes a subfamília MI foram reprimidos Hsp22.3-CVI, Hsp16.6-CVIII, Hsp18.8-CX e Hsp17.8-CXI (SARKAR *et al.*, 2009). Após a análise por blastN foi observado que o gene Hsp17.4-CI de arroz é homólogo ao gene identificado pelo glyma14g06910 de soja. Esse resultado sugere que as funções desses genes em resposta a infecção por patógenos podem ter sido mantidas, mesmo depois da divergência entre as espécies durante o processo evolutivo.

Os genes identificados pelos glymas 14g39560 e 19g01440 tiveram indução nos tratamentos por infecção pelo nematóide *M.javanica* em 4 e 8 dpi e somente em 4 dpi, respectivamente. O glyma13g24460 foi reprimido no tratamento de 8 dpi na cultivar suscetível (BRS133), resultado discrepante ao obtido por Fuganti e colaboradores (2010), já que esse gene havia sido induzido mesmo no genótipo suscetível sob infecção por *M. javanica*. Os resultados de Fuganti e colaboradores (2010) apresentaram uma alteração de nível de indução entre o genótipo resistente, que teve maior indução em relação ao suscetível. Essa discrepância provavelmente pode ser explicada pela alteração no gene de referência utilizado na técnica de RT-qPCR no presente trabalho. Na pesquisa desenvolvida por Fuganti e colaboradores (2010), o gene 18S ribossomal foi utilizado como gene de referência, ou normalizador, enquanto no presente trabalho, foi utilizado o gene β -actina. Segundo Hruz e colaboradores (2011), embora muitos genes possam ser utilizados como normalizados em reações de RT-qPCR, o não estabelecimento de um melhor gene de referência para as análises em uma determinada espécie pode ser a causa de uma grande variação nos resultados experimentais.

Com o objetivo de avaliar se a alteração nos perfis de expressão dos 7 genes responsivos ao estresse biótico na cultivar suscetível poderia ser também observada na cultivar resistente (PI595099), plantas de soja resistentes infectadas com o nematoide foram também utilizadas nas análises de expressão desse conjunto de genes. Os resultados obtidos demonstram que apenas três genes foram induzidos significativamente na cultivar resistente no tratamento de 8 dpi, glyma10g3200, glyma13g24460 e glyma14g39560 (Gráfico 2).

Gráfico 2 - Resultado da análise de expressão relativa por RT-qPCR dos genes candidatos Hsp20 no genótipo resistente PI595099 e suscetível no tratamento de estresse biótico (4 e 8dpi *M.javanica*). Estão plotados apenas os genes foram significativos ao nível de 5% em algum dos tratamentos. Na legenda, cada cor indica uma cultivar submetida a um dos tratamento por infecção do nematóide. Apenas os tratamentos em que a alteração no nível de expressão genica foi significativo estão marcadas com asterisco logo acima. As barras de erro indicam na escala a margem de erro para mais ou para menos



Quando comparado ao resultado obtido por Fuganti e colaboradores (2010), o gene identificado pelo glyma13g24460 apresentou resultado semelhante, com a indução na cultivar resistente, no entanto não foi possível observar a indução na cultivar suscetível.

Aparentemente a indução das Hsp20 na cultivar resistente ocorre em período tardio em relação a cultivar suscetível. No tratamento de 8 dpi a cultivar suscetível apresenta dois (*glyma10g3200* e *glyma13g24460*) genes reprimidos, enquanto a resistente tem esses mesmos dois genes induzidos.

5.3 IDENTIFICAÇÃO DE ELEMENTOS REGULATÓRIOS

A predição de *cis* elementos na região regulatória dos GmHsp20 candidatos foi realizada utilizando os programas *MatInspector*, *Place* e *PlantCare*. Além dos elementos HSE, também buscou-se a identificação de elementos TATAbox, CCAAT, Wbox e TA *rich region*, os quais foram previamente reportados como frequentes nas regiões promotoras de Hsp20, bem como outros genes responsivos a estresses.

Esses elementos aparentemente agem como participante ou mesmo determinantes na ativação dos genes em situações específicas em outras espécies (YU *et al.*, 2001; THURAU *et al.*, 2003; BARCALA *et al.*, 2008; FUGANTI *et al.*, 2010). Os elementos TATAbox são normalmente encontrados em torno de -30 bases do sítio de início da transcrição nos eucariotos e dessa forma, para apenas 28 dos 45 Hsp20, e 7 dos genes não responsivos ao estresses de calor, foi identificada a presença de um provável elemento TATABox (Tabela 2). O fato de não ter sido encontrado esse elemento nos demais genes deve estar relacionado a erros nas sequências disponíveis no Phytozome, pois possivelmente todos os genes Hsp20 de soja apresentam o TATABox. Tal erro também pode ter ocorrido para a predição dos demais elementos regulatórios, como, por exemplo, os HSE, considerando que o nível de informatividade para regiões UTR's e promotoras em soja é menor comparado com a predição das ORFs.

O elemento Wbox funciona como elemento de ligação no DNA para os fatores de transcrição da família WRKY, os quais são conhecidamente importantes na indução de genes de defesa via elicitor (YU *et al.*, 2001). Na soja, genes da família WRKY vêm sendo relacionados à resposta de resistência a ferrugem asiática. O fenótipo de resistência mediado pelo gene *Rpp2* é alterado quando qualquer um de três fatores de transcrição (FT) WRKY específicos são silenciados. Além disso, o silenciamento dos FT WRKY impediu a indução de genes

normalmente presentes em resposta da ferrugem, entre os quais está o gene GmPR1 (relacionado a patógenese), que é caracterizado pela presença de múltiplos núcleos W-boxes na sua região promotora do GmPR1, provavelmente necessários para sua regulação pelo FT GmWRKY36 (PANDEY *et al.*, 2011).

A presença do elemento W-box foi também evidenciada na região promotora do gene NPR1 (*non-expressor of pathogenesis related genes*) de *Arabidopsis*, um regulador positivo basal da resposta de resistência a doenças em plantas. (YU *et al.*, 2001). De acordo com os resultados obtidos em nossas análises, o elemento W-box está presente em posições variadas em 30 genes dos 45 GmHsp20 e 7 não responsivos ao calor estudados (Tabela 2). Dos genes de GmHsp20 identificados neste trabalho e na literatura como responsivos a estresses por infecção de nematóides, os modelos gênicos Glyma04g05720 e Glyma08g07340 não apresentaram nenhum elemento Wbox. Já os outros genes: glyma10g32000, glyma13g24460, glyma14g06910 e glyma14g39560 apresentam um núcleo W-box em suas regiões reguladoras, embora em posições diferentes (-406 pb, -4 pb, -71 pb e -476 pb, respetivamente).

O elemento CAAT box, por outro lado, foi encontrado em todos os genes reportados como responsivos ao estresse por infecção pelo nematóide, aparecendo caracteristicamente entre os elementos HSE preditos (-56 a -472 pb). Recentemente tais elementos foram observados na região promotora do gene Hs1 pro-1 que confere resistência completa ao nematóide do cisto *Heterodera schachtii* (JUNG e LÖPTIEN, 1986 apud THURAU *et al.*, 2003). Os elementos CAAT box estão presentes em uma região que é essencial para regulação sítio- específica da expressão do gene que ocorre após infecção pelo nematóide (THURAU *et al.*, 2003). Nos GmHsp20, os elementos CAAT box preditos localizam-se nessa mesma região embora ainda não haja evidência experimental de sua função na indução de tais genes.

Os resultados obtidos por Barcala e colaboradores (2008) dão indícios que os genes Hsp20 responsivos ao estresse biótico apresentam uma configuração específica de HSE em conjunto com elementos TATAbox e CAATbox. O padrão de disposição dos elementos de ligação à fatores de transcrição na região promotora dos mutantes Hsp20 de *Arabidopsis* responsivos a infecção por nematóide, demonstra que o TATAbox deve estar preferencialmente entre – 12 bases

e – 21 bases à montante, seguido do núcleo HSE a cerca de – 83 bases e imediatamente à montante dos HSE, posicionado a cerca de – 84 bases a – 141 bases, em relação ao sítio de início da transcrição, os elementos CCAATbox. Nossos resultados para os GmHsp20, responsivos ao estresse biótico, indicam uma distribuição heterogênea entre os promotores que apresentam de dois a cinco núcleos CCAAT (Tabela 2). Tais elementos apresentam o mesmo padrão dos genes Hsp20 de *Arabidopsis* e girassol (THURAU *et al.*, 2003), com os CAAT box entre os elementos HSE e/ou imediatamente acima, e os Wbox mais distantes, acima dos HSEs. Para todos os genes, confirmados como Hsp20 de soja responsivos a estresse biótico, foram identificados HSE seguido de CAAT box. Para os genes glyma10g32000, glyma14g06910 e glyma14g39560 os resultados indicaram a presença desses dois elementos e um padrão de posições semelhante ao descrito anteriormente, sendo pelo menos um primerio HSE seguido de CCAAT e depois o Wbox mais distante (Figura 21).

Figura 13 - Posição dos *cis* elementos identificados pelos programas de predição na região a montante -500 pb de cada um dos 8 genes Hsp20 responsivos ao estresse biótico, identificados neste trabalho e na literatura. As caixas azuis indicam os elementos TATAbox, as vermelhas os HSE, as verde desdoadas os CAAT box, as verdes os elementos *TA-rich* e as cinza os elementos Wbox. Os genes identificados pelos glyma16g04270 e glyma18g41650 são responsivos

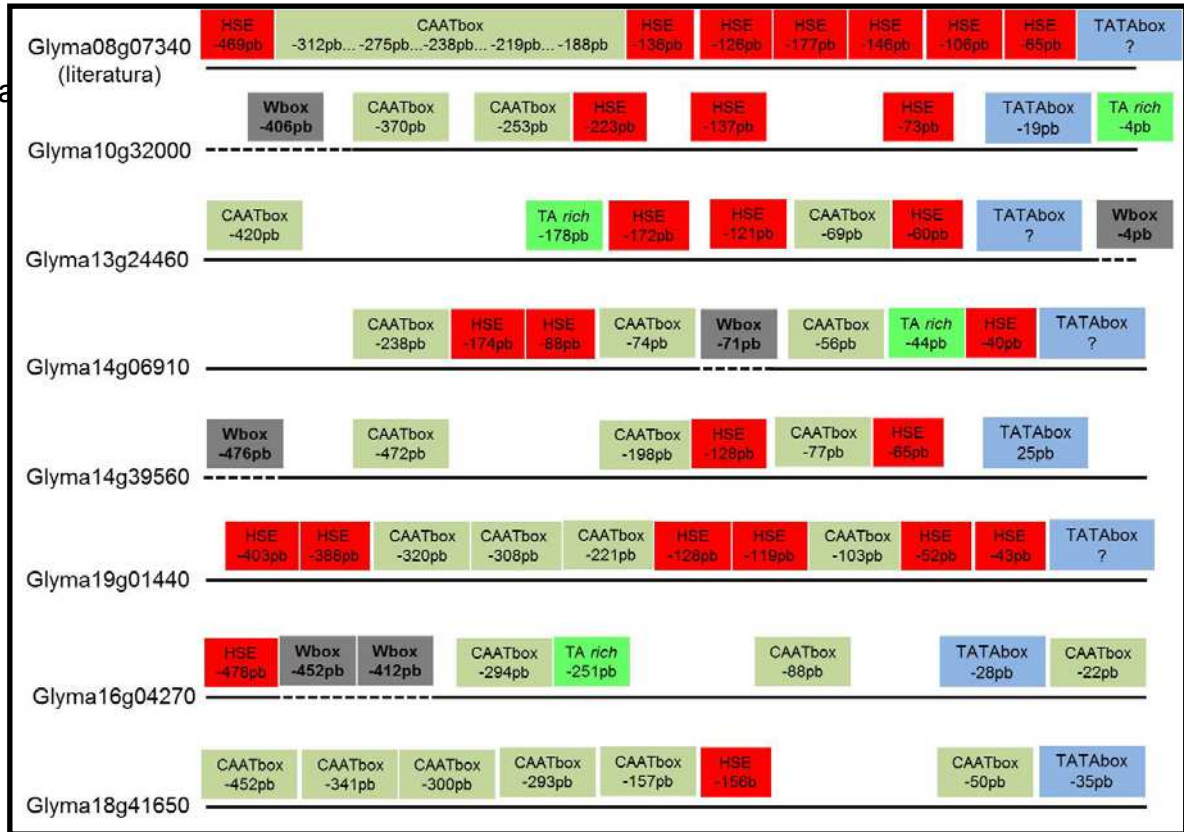


Tabela 2 - Ocorrência e posicionamento dos *cis* elementos: CCAAT, TATABox, Wbox e TA-rich na região à montante (- 500pb) dos 47 modelos gênicos identificados como candidatos potenciais a GmHsp20. As posições são estimadas em relação ao sítio de início da transcrição. ^aOs resultados são apresentados para os três programas utilizados nas análises de *cis* elementos. Os demais elementos foram obtidos pelo programa PlantCare. A sigla “nd” significa que a região à montante não está disponível no Phytozome.

Glymas	MatInspector – PLACE – PlantCARE			
	^a CCAAT	^a TATABox	^a Wbox	TA-rich region
Glyma01g26570	-14 -110 -280 -333 -377	-29	_____	_____
Glyma02g08400	_____	-14 -30	_____	_____
Glyma02g41150	-2 -105 -165 -166 -262 -356 -471 -472	_____	-17 -323	_____
Glyma02g42000	-86 -158 -290 -299 -447 -468 -473	-30	-263 -264	_____
Glyma02g45810	-215 -221 -240	_____	-31 -35	_____
Glyma04g05720	-372	_____	_____	-309
Glyma04g38530	-119 -120 -300 -488	_____	_____	_____
Glyma04g40790	-296 -395	-32	_____	_____
Glyma06g05740	-196 -253 -290 -409	-32	-120	_____
Glyma06g14000	-17 -186	-28	-5	_____
Glyma06g16490	-154 -229 -357	_____	-41 -170 -169 -396	_____
Glyma07g03450	nd	nd	nd	nd
Glyma07g04860	-36 -67 -306 -330 -406	-19	-315 -165 -493	_____
Glyma07g32030	-178 -301	_____	_____	_____
Glyma07g32050	-442	-25	_____	_____
Glyma07g32070	-401 -202	_____	_____	_____
Glyma07g32090	nd	nd	nd	nd
Glyma07g32110	-313	_____	-215	_____
Glyma08g07330	-491	-28	-306	_____
Glyma08g07340	-188 -219 -238 -275 -312	_____	_____	_____
Glyma08g07350	-85 -106 -116 -135 -249 -347	_____	_____	_____
Glyma08g22630	-13 -284	-40	-87	_____
Glyma10g32000	-253 -370	-19	-406	-4
Glyma11g14250	-42 -103 -247 -278 -425 -440	-8	_____	_____
Glyma11g37450	-109	-27	-96	_____

Glyma12g01580	-295 - 410	_____	-69 -292	_____
Glyma12g06210	-28 -130 -159	_____	_____	_____
Glyma13g24440	-215 -304 -494	-26	_____	_____
Glyma13g24460	-69 -420	_____	-4	-178
Glyma13g24480	-246 -333 -461	-57	-481 -377	_____
Glyma13g24490	-260	-24	-331 -437 -464	_____
Glyma13g24510	-156 - 477	_____	-198 -342 -398	_____
Glyma13g27590	-129 -202 -394	-24	-442	_____
Glyma13g36220	-326 -397	_____	-169 -252 -366	_____
Glyma14g04970	-46 -134 -252 -256 -454 -474 -488	-34	-97 -100 -336	_____
Glyma14g06900	-176 -218 -377 -493	-26	-288	_____
Glyma14g06910	-56 -74 -238	-44	-71	-44
Glyma14g11420	-274 - 312 - 366	-32	-240 -390	_____
Glyma14g11430	-163 -332 -422 -485	_____	-167	_____
Glyma14g39560	-77 -198 -472	-25	-476	_____
Glyma15g11360	-36 -182 -256	_____	_____	_____
Glyma15g41880	_____	_____	_____	_____
Glyma16g01440	-72 -76 -442	-24	_____	_____
Glyma16g04270	-22 -88 -294	-28	-412 -452	-251
Glyma16g33130	-271 -396	-26	_____	_____
Glyma17g34220	-128 -301 -331	-29	-222	_____
Glyma18g10760	nd	nd	nd	nd
Glyma18g41650	-50 -157 -293 -300 -341 -452	-35	_____	_____
Glyma19g01440	-103 -221 -308 -320	_____	_____	_____
Glyma20g01930	-207 -265 -345 -398 -433	_____	-131 -477	_____
Glyma20g35650	_____	_____	-21	_____

Diferente dos elementos TATAbox e Wbox, o elemento TA-rich consiste em uma região rica em elementos repetitivos TA à montante do sítio de início da transcrição que deve exercer função regulatória ou ainda alterar a distância entre *cis* elementos de modo a potencializar a expressão dos genes (WHITE *et al.*, 1994). O gene GmHsp17.6-L é fortemente induzido em plantas de soja resistentes a *M. javanica* e sua expressão relaciona-se ao número de repetições TA presentes no elemento TA rich (FUGANTI *et al.* 2010). O número de repetições TA parece também estar relacionado à força de indução da expressão do gene, como demonstrado para GmHsp17.6-L em resposta a infecção pelo nematóide *M. Javanica*. As plantas resistentes possuem 32 repetições TA enquanto as suscetíveis apresentam apenas 9 repetições (FUGANTI *et al.* 2010).

Entre as Hsp20 de soja foram encontradas regiões ricas em TA nas posições -178, -4, e -309pb dos genes ou glymas: glyma04g05720 (TA)₁₂, glyma13g24460 (TA)₉, glyma10g32000 (TA)₁₃, glyma14g06910 (TA)₆ e glyma16g04270 (TA)₈, respectivamente; sendo os quatro últimos comprovadamente induzidos em raízes de plantas de soja infectadas com nematóides pelos resultados de análise de expressão relativa por RT-qPCR, e todos responsivos ao estresse de calor, exceto o último, glyma16g04270. Assim, parece que a região rica em TA nas GmHsp20 possa funcionar de modo a potencializar a expressão do gene, como encontrado para a GmHsp17.6-L (Glyma13g24460) (FUGANTI *et al.*, 2010), mas também possa exercer função regulatória, já que sua presença foi observada apenas entre alguns dos genes GmHsp20 responsivos ao estresse biótico por infecção por nematóides.

As análises de predição de *cis* elemento HSE e de expressão *in vivo* dos 47 genes, antes candidatos a Hsp20, sob a infecção com *M. javanica* nos permitiram comprovar a relação previamente proposta por Barcala e colaboradores (2008) entre o posicionamento dos elementos HSE e sua influência na expressão de Hsp20 de soja em resposta ao estresse por infecção pelo nematóide *M. javanica*.

Todas as cinco Hsp20 de soja responsivas ao estresses biótico, por infecção pelo nematóide *M. javanica*, seguem o padrão estabelecido por Barcala e colaboradores (2008), com a posição do primeiro HSE até -83 pb à montante. Além da presença e posição dos demais *cis* elementos investigados. Os outros dois genes não caracterizados como Hsp20 (glyma16g04270 e glyma18g41650), mas que

também estão relacionados aos estresse por infecção pelo nematóide *M.javanica* seguem um padrão distinto em relação ao descrito por Barcala e colaboradores (2008).

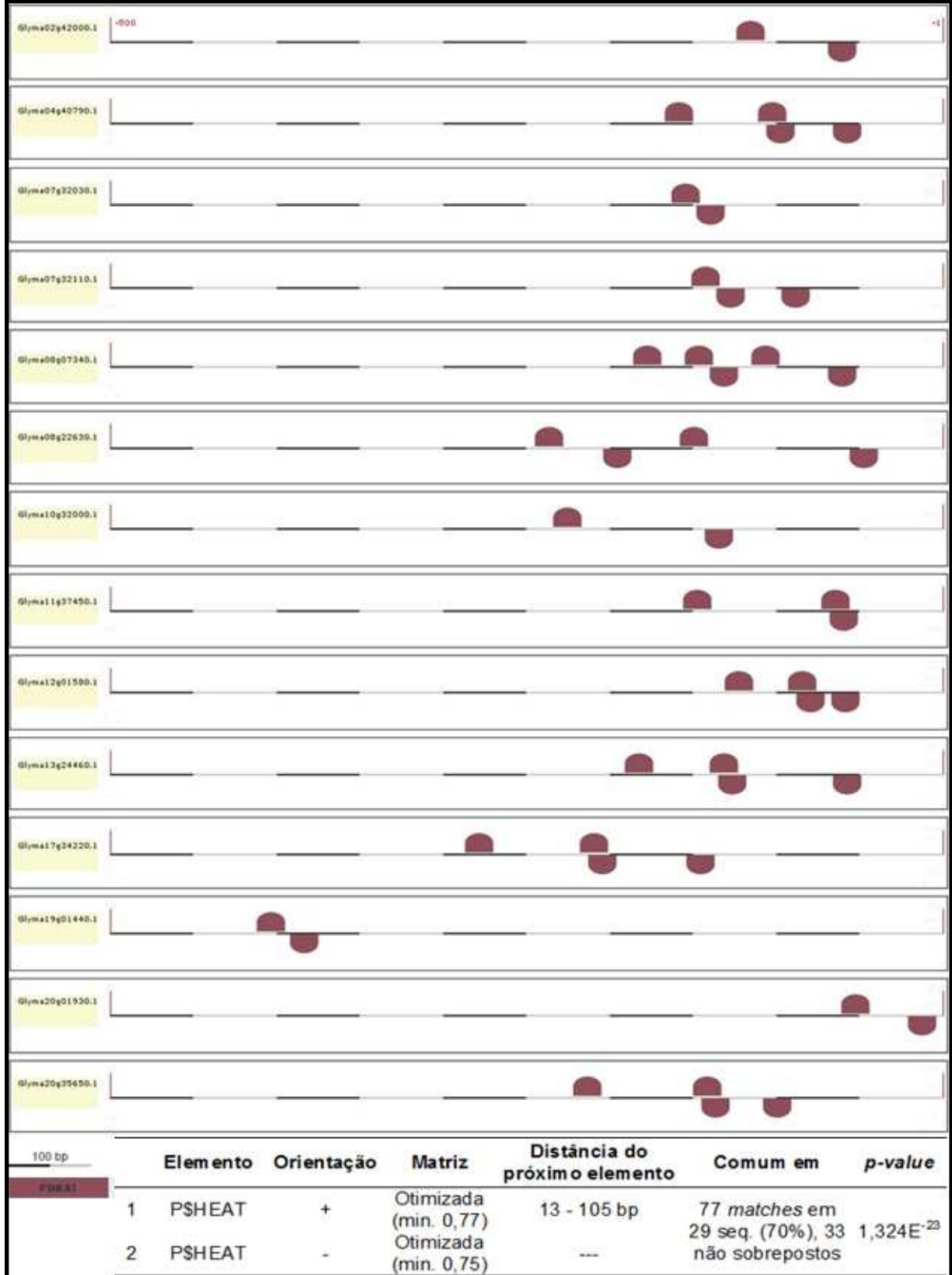
5.4 IDENTIFICAÇÃO DE MODELOS OPERACIONAIS DE PROMOTORES

A identificação de elementos *cis* regulatórios é primordial para a caracterização de promotores. Entretanto, a combinação de diferentes elementos em padrões que se repetem entre promotores responsivos a estímulos semelhantes que define sua estrutura básica e permite a comparação entre promotores distintos (FESSELE *et al.*, 2002; WERNER *et al.*, 2003). Nesse sentido, análises utilizando o software *FrameWorker* foram empregadas para procurar por padrões ou modelos operacionais que se repetem entre as comprovadas GmHsp20 (modelo mínimo), entre o grupo das responsivas a estresse de frio e ao estresse por infecção do *M.javanica*.

O primeiro modelo mínimo encontrado indica a presença de pelo menos um elemento HSE na fita positiva. Como o *FrameWorker* considera não apenas a presença de um elemento, mas também a conservação de seu posicionamento, este resultado corrobora com a análise realizada pelos programas de predição de posição de *cis* elementos, demonstrando haver variação no posicionamento dos elementos HSE entre os promotores Hsp20 de soja.

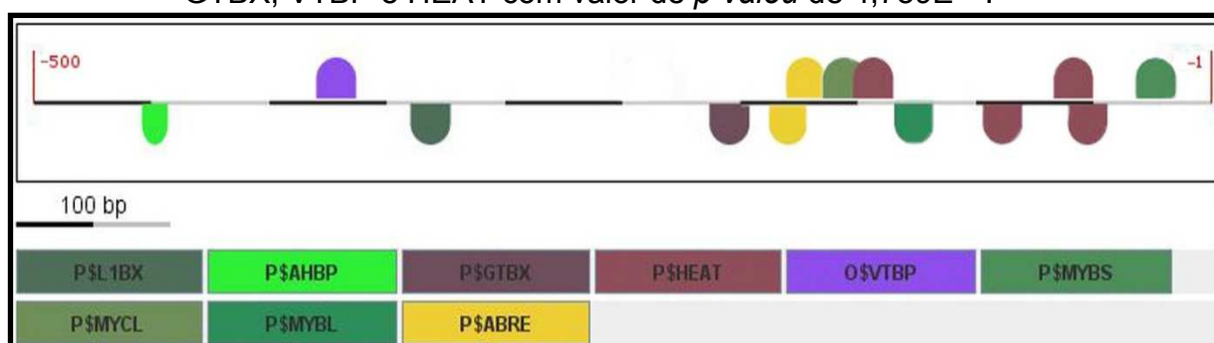
Os genes GmHsp20 apresentam seus elementos HSE na região à montante do gene, em geral, até -400 pb na orientação da fita positiva. Existem exceções com o primeiro HSE até cerca de -1100 bp na orientação da fita positiva. O modelo formado sugere a ocorrência de 14 padrões distintos entre as Hsp20, exemplificados pelos glymas da figura 22.

Figura 14 - Ilustração da estrutura de elementos comuns da região promotora de genes Hsp20 de soja – Modelo mínimo. O programa *Frameworker* Genomatix foi usado para procurar o principal elemento comum entre os genes. O modelo demonstra diversos padrões com pelo mesmo dois HSE, um em cada orientação.



A análise comparativa entre os promotores dos 5 glymas responsivos ao estresses de frio: 01g26570, 07g04860, 16g01440, 11g37450 e 20g35650, resultou na formação de seis modelos, um sem elemento mandatório e cinco contendo o elemento mandatório (HEAT). Nessa análise, foram observadas a presença de famílias de elementos comuns: L1BX, AHBP, GTBX, VTBP, MYBS, MYCL, MYBL, ABRE, além do elemento HEAT (Figura 23).

Figura 15 - Ilustração do consenso de todos os modelos de estrutura de elementos comuns da região promotora de genes Hsp20 de soja – Modelo frio. O programa *Frameworker* Genomatix foi usado para procurar os principais elementos comuns entre os genes e suas posições de alta frequência. Para o modelo de posição dos HSE na orientação negativa o valor de *p-valeu* foi de $9,37E^{-01}$. Para o modelo dos elementos ABRE, MYBS, MYCL e HEAT o valor de *p-valeu* foi de $9,57E^{-04}$. Outro modelo sem elementos mandatório resultou na inclusão os elementos L1BX, AHBP, GTBX, VTBP e HEAT com valor de *p-valeu* de $4,789E^{-06}$.



A L1 box (L1BX) foi identificado em *A. thaliana* como um elemento presente na região promotora essencial para a expressão do fator protodérmico (PROTODERMAL FACTOR1 - PDF1). Curiosamente, algumas sequências alvo semelhantes de L1 box tem sido identificadas na regiões à montante de genes patógeno relacionados (ABE *et al.*, 2001)

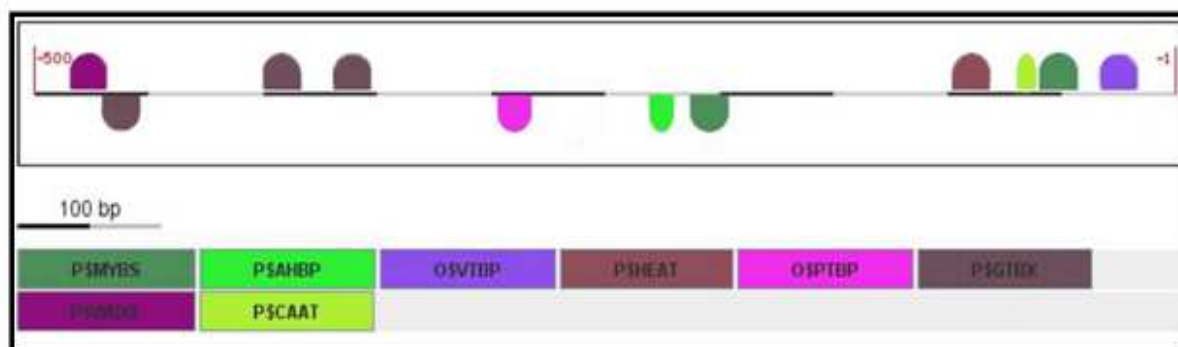
A primeira sequência caracterizada com domínio bZIP era do gene *EmBP1*. Esse domínio reconhece e se liga a sequencias consensos na região promotora de genes, os elementos do tipo ABRE (ácido abscísico - *ABA Responsive Element*) (MARALHÃES JÚNIOR, 2010). A sequencia consenso desse elemento regulatório apresentam uma sequência de 8 a 10 pb com um núcleo central ACGT (região G-box) (YAMAGUCHI-SHINOZAKI e SHINOZAKI, 2005). A presença desses *cis* elementos já foi descrita, principalmente em promotores de genes responsivos ao ABA. Em plantas de *Arabidopsis*, duas sequências ABRE são necessárias para a

expressão do gene *RD29B*, que foi identificado como responsivo a estresse por déficit hídrico. Atualmente, o promotor desse gene tem sido amplamente utilizado para o controle de expressão de genes de interesse, principalmente sobre estresse de déficit hídrico. Contudo, esse e outros genes que apresentam o elemento ABRE, além de responsivos ao déficit hídrico, também são induzidos quando a planta está sob estresse de frio e por ABA. Esses genes também são induzidos em plantas mutantes ABA-deficientes ou ABA-insensíveis por falta de água e frio (MARALHÃES JÚNIOR, 2010).

Outros elementos regulatórios encontrados nos promotores das *GmHsp20* também são sítios de reconhecimento e ligação para fatores de transcrição nas rotas de transdução de sinais ABA-dependente, como o MYB (*Myeloblastosis Oncogene*) e MYC (*Myelocytomatosis Oncogene*). A principal característica dos fatores de transcrição MYCL é a sua base estrutural hélice-alça-hélice (bHLH), em um trecho de 40 a 50 aminoácidos contendo duas alfa-hélices anfipáticas separadas por uma região de ligação de comprimento variável (MURRE *et al.*, 1994). A expressão do gene que codifica para o fator de transcrição *AtMYC2* de *Arabidopsis*, (ABA)-induzido, tem sido demonstrada em plantas sob o stress seca (LORENZO *et al.*, 2004). Atualmente, os fatores de transcrição da família MYB têm sido muito associados ao papel nas reações de defesa das plantas. Contudo, sua regulação em plantas, parecem estar também envolvida no controle do metabolismo secundário ou resposta a metabolitos secundários, tais como o metabolismo dos flavonóides e fenilpropanóides e da via das antocianinas (BUCK e ATCHLEY, 2003)

A análise comparativa entre os promotores dos 8 glymas responsivos a infecção pelo nematóide *M. javanica* e outros patógenos, como *H.glycines*: 08g07340, 10g32000, 13g24460, 14g06910, 14g39560 e 19g01440, responsivos a estresse de calor, e 16g04270 e 18g41650, não induzidos sob choque térmico, resultou na formação de três modelos, todos com elemento mandatório HEAT. Nessa análise foi observada a presença de famílias de elementos comuns: AHBP, GTBX, VTBP, MYBS, WBXF e CAAT além do elemento HEAT (Figura 24).

Figura 16 - Ilustração do consenso de todos os modelos de da estrutura de elementos comuns da região promotora de genes Hsp20 de soja – Modelo biótico. O programa *Frameworker* Genomatix foi usado para procurar os principais elementos comuns entre os genes e suas posições de alta frequência. Para o modelo de posição dos HSE na orientação negativa o valor de *p-valeu* foi de 0,0381311. Para o modelo dos elementos AHBP, GTBX, VTBP (TATA box *vertebrate*), PTBT (TATA box *plant*) o valor de *p-valeu* foi de $1,67E^{-04}$. Para o modelo Wbox com valor de *p-valeu* foi de $5,02E^{-01}$. Outro modelo CAAT box com valor de *p-valeu* foi de 0,167564.



Vários estudos demonstraram que em muitos promotores de genes de defesa de plantas há presença de elementos de Wbox (TTGACC / T), onde os fatores de transcrição WRKYs são capazes de reconhecer e ligar-se (YAMASAKI *et al*, 2005). A transcrição de genes WRKY é forte e rapidamente induzida em inúmeras espécies de plantas em resposta à infecção, ferimento por patógeno, ou estresses abióticos, como a seca, a adaptação ao frio ou calor tolerância induzida pelo frio e outros (RIZHSKY *et al*, 2002).

5.5 ANÁLISES DOS GENES CARACTERIZADOS COMO HSP20 DE SOJA

A caracterização, *in silico*, dos modelos gênicos anotados nos bancos de dados como Hsp20 de soja, consideraram aspectos específicos para enquadrá-los como possíveis candidatos a genes codificadores de Hsp20 de soja. Enquanto que, as análises *in vivo*, estabeleceram os genes que demonstram o perfil de expressão Hsp20, sendo induzidos sob estresses de choque térmico. Assim, apenas 45 genes fazem parte da família de Hsp de soja.

5.5.1 Região Codificadora

Nos GmHsp20, as ORFs preditas variaram em tamanhos de 426 a 3463 pb. Foi verificado que entre codantes de mesmo tamanho ou tamanhos muito próximos houve variação no número e tamanho de íntrons. (Tabela 3).

Segundo Schmutz e colaboradores (2010), a estrutura exon-íntron e a posição dos íntrons nos genes das plantas mostram elevada conservação entre as espécies. Genes com íntrons duplicados têm uma forte tendência a mantê-los, assim como o tamanho desses. Segundo Schmutz e colaboradores (2010) a taxa de perda de íntrons é mais lenta em *poplar* (0,4%) do que na soja (0,6%). Dentro da família de Hsp20 de soja caracterizada, nossas análises indicam que 32 (53,84%) do total de GmHsp20 não apresentam íntrons. Este é um percentual médio muito superior ao dos genes Hsp20 de arroz preditos (19,9% - 48,72% de genes sem íntrons) (JAIN *et al.* 2008; OUYANG *et al.*, 2009). Entre os genes de soja com íntrons, 10 (35,71%) apresentam um único íntron. Com base na presença e no número de íntrons, os genes podem ser divididos em três padrões. O padrão 1 consiste em genes sem íntrons, enquanto para o padrão 2 há apenas um íntron. Já para o padrão 3, menos comum, são encontrados mais de um íntron. Dentre os GmHsp20 de soja os padrões 1 e 2 foram prevalentes. Contudo, quando enquadrados no padrão número 3, de 6 genes, 5 apresentaram apenas dois íntrons. Uma consideração relevante é a presença de íntrons nas regiões não traduzidas, principalmente na posição 5', que acredita-se potencializar a tradução. Entre os 4 genes com íntrons preditos, dois o apresentam na região 5'UTR. O padrão global da posição dos íntrons pode ser utilizado como base, no estabelecimento das relações filogenéticas em uma família de genes. Isso será melhor apresentado, a seguir, nas seções de subfamílias de GmHsp20 e estrutura secundária.

Tabela 3 - Localização cromossômica, características físico-químicas, padrão de íntrons e localização subcelular para as GmHsp20. As abreviação na coluna de localização subcelular predita estão identificadas para as organelas: citoplasma (cito), peroxíssomo (Px), cloroplasto (clorop), mitocôndria (mito) e reticulo endoplasmático (R.E.)

Nome do Transcrito	Localização no Genoma - Cromossomo	Tamanho da ORF Predita (bp)	Íntrons - Tamanho (pb)	Peso Molecular Predito (KDa)	pI Predito	Índice de Instabilidade	Localização Subcelular Predita
Glyma01g26570	Crom01(D1a): 35,047,848...35,048,596	435	0	16,37	5.56	44.69	cito
Glyma02g41150	Crom02(D1b): 46,309,221...46,310,788	915	480	16,25	6.98	45.83	cito/Px
Glyma02g42000	Crom02(D1b): 47,097,366...47,098,060	462	0	17,29	6.76	52.45	cito
Glyma02g45810	Crom02(D1b): 49,961,013...49,964,816	3463	1018/110/268/378/903	28,63	7.63	52.30	cito
Glyma04g05720	Crom04(C1): 4,366,064...4,366,873	477	0	17,94	6.62	39.52	cito
Glyma04g38530	Crom04(C1): 44,921,535...44,922,271	426	0	15,91	8.86	41.78	cito
Glyma04g40790	Crom04(C1): 46,743,808...46,745,820	1741	1045	25,99	7.74	41.73	clorop
Glyma06g05740	Crom06(C2): 4,100,838...4,101,520	477	0	17,80	6.62	46.88	cito
Glyma06g14000	Crom06(C2): 11,060,463...11,062,336	1685	989	26,06	6.98	36.98	clorop
Glyma06g16490	Crom06(C2): 12,926,962...12,927,793	453	0	17,10	6.76	45.01	cito

Glyma07g03450	Crom07(M): 2,422,428...2,423,510	1083	440/61	21,62	7.02	34.40	mito
Glyma07g04860	Crom07(M): 3,586,439...3,587,416	705	129 (5'UTR)/87	18,21	6.97	43.81	cito
Glyma07g32030	Crom07(M): 36,962,519...36,963,220	462	0	17,61	6.86	41.05	cito
Glyma07g32050	Crom07(M): 36,977,969...36,978,814	486	0	18,51	5.99	44.25	cito
Glyma07g32070	Crom07(M): 36,987,613...36,988,240	462	0	17,51	5.72	43.37	cito
Glyma07g32090	Crom07(M): 36,995,756...36,996,604	462	0	17,54	5.33	42.12	cito
Glyma07g32110	Crom07(M): 37,000,908...37,001,669	462	0	17,52	5.70	38.95	cito
Glyma08g07330	Crom08(A2): 5,260,106...5,260,914	462	0	17,48	5.80	43.81	cito
Glyma08g07340	Crom08(A2): 5,263,766...5,264,591	462	0	17,42	5.52	40.85	cito
Glyma08g07350	Crom08(A2): 5,268,144...5,269,164	462	0	17,35	6.17	42.98	cito
Glyma08g22630	Crom08(A2): 17,182,793...17,184,632	1653	771/210 (3'UTR)	25,66	6.92	46.96	mito
Glyma08g43120	Crom08(A2): 42,987,975...42,988,809	758	97	25,35	9.52	37.34	clorop
Glyma10g32000	Crom10(O): 40,472,306...40,473,244	588	0	22,38	5.74	45.06	R. E.

Glyma11g37450	Crom11(B1): 38,558,601...38,559,879	1068	360	27,33	8.47	59.06	clorop
Glyma12g01580	Crom12(H): 941,609...942,895	949	310	23,81	6.93	49.62	mito
Glyma13g24440	Crom13(F): 27,839,030...27,840,066	465	0	17,53	5.98	41.51	cito
Glyma13g24460	Crom13(F): 27,847,751...27,848,553	465	0	17,59	5.70	43.70	cito
Glyma13g24480	Crom13(F): 27,853,841...27,854,840	465	0	17,53	5.98	42.86	cito
Glyma13g24490	Crom13(F): 27,862,859...27,863,745	486	0	18,50	7.93	43.24	cito
Glyma13g24510	Crom13(F): 27,869,587...27,870,386	459	0	17,39	6.86	47.59	cito
Glyma13g27590	Crom13(F): 30,780,322...30,781,372	714	312	15,23	5.32	38.21	cito
Glyma14g06900	Crom14(B2): 5,086,252...5,086,956	474	0	17,74	6.33	54.03	cito
Glyma14g06910	Crom14(B2): 5,087,964...5,088,623	480	0	17,91	6.86	41.92	cito
Glyma14g11420	Crom14(B2): 9,944,523...9,945,260	480	0	17,86	5.49	50.82	cito
Glyma14g11430	Crom14(B2): 9,947,884...9,948,637	480	0	17,99	5.79	43.81	cito
Glyma14g39560	Crom14(B2): 48,651,742...48,652,904	914	479	16,16	6.92	55.32	cito/Px

Glyma15g11360	Crom15(E):	8,325,524...8,326,426	647	245	15,38	5.20	44.98	cito
Glyma15g41880	Crom15(E):	49,087,958...49,088,389	432	0	16,42	5.86	48.18	cito
Glyma16g01440	Crom16(J):	1,023,807...1,024,860	814	129 (5'UTR)/199	17,97	6.16	46.98	cito
Glyma16g33130	Crom16(J):	36,161,134...36,161,989	594	0	22,51	6.40	35.08	R. E.
Glyma17g34220	Crom17(D2):	38,089,483...38,090,271	480	0	17,88	5.95	38.96	cito
Glyma18g10760	Crom18(G):	9,611,401...9,612,296	896	160/98	23,86	9.40	32.00	clorop
Glyma19g01440	Crom19(L):	1,063,015...1,063,949	594	0	22,32	5.88	49.08	R. E.
Glyma20g01930	Crom20(I):	1,455,228...1,456,036	477	0	17,92	6.62	41.17	cito
Glyma20g35650	Crom20(I):	43,885,724...43,886,516	579	0	21,96	6.24	39.47	R. E.

As análises de peso molecular teórico das sequências protéicas preditas de GmHSP20 resultaram em um intervalo de 15,23 KDa (133aa_ Glyma13g27590) a 28,63 KDa (144aa_ Glyma02g45810). No entanto, foi possível observar que algumas proteínas tiveram diferentes pesos moleculares, mesmo quando apresentaram o mesmo número de resíduos de aminoácidos. Isso pode ser explicado pelas diferenças na composição de aminoácidos entre as proteínas. Embora não demonstrado na tabela 3, a análise de predição de peso molecular foi expandida para todos os 76 modelos gênicos anotados como GmHsp20. Essa observação é relevante, considerando que as classes de HSPs são divididas tendo como critério principal o peso molecular, e as pequenas proteínas de choque térmico HSP20 apresentam predominantemente massa molecular de 15-42 kDa, e o limite máximo das Hsp20 de soja a cerca de 30 kDa. (SUN *et al.*, 2002). De acordo com a análise realizada, duas proteínas preditas correspondentes aos Glymas 15g30010 (45,08 kDa) e 15g33630 (48,42kDa), apresentaram peso molecular superior a faixa determinada para as HSP dessa classe, ambas sendo excluídas do conjunto de HSP20 selecionadas por não estarem em conformidade aos critérios utilizados nas análises anteriores (disposição de núcleos HSE e expressão digital).

Os resultados de predição de ponto isoelétrico ficaram entre 5.20 (Glyma15g11360) e 9.52 (Glyma08g43120), e os índices de instabilidade indicaram o Glyma18g10760 com pontuação 32.00, como mais estável, e o Glyma11g37450 com índice de 59.06, como a proteína mais instável. Esses resultados demonstram que o índice de instabilidade e *pI* aparentemente estejam relacionados, pois dentre as 10 GmHSP20 e que foram consideradas estáveis (índice de corte valores ≤ 40), a maioria também apresentou *pI* alcalino. Além disso, 67,30% das GmHSP20 são instáveis.

O termo “estabilidade” para proteína refere-se à tendência em manter sua conformação nativa. Sabe-se que características como *pI* e a estabilidade são determinadas, principalmente, pela composição de aminoácidos da proteína. As interações fracas (não covalentes) e ligações de dissulfeto, formadas entre os aminoácidos, auxiliam na conformação proteica, principalmente as interações hidrofóbicas estabelecidas por aminoácidos apolares (NELSON e COX, 2010). Guruprasad e colaboradores (1990) determinaram o índice de instabilidade para o programa ExPASyProtparam, utilizado neste estudo, após análises estatísticas

feitas com proteínas estáveis e não-estáveis. Os resultados da pesquisa desses autores sugerem que a instabilidade proteica *in vivo* possivelmente possa estar relacionada com a disposição de determinados dipeptídeos ao longo da sequência de aminoácidos.

Em termos funcionais, acredita-se que a instabilidade proteica seja uma característica comum entre proteínas que respondem a estresse. Isso porque a resposta ao estresse ocorre como um evento pontual na célula, e deve permitir o rápido *turnover* das proteínas relacionadas, voltando para a transcrição a regulação da presença da proteína no ambiente celular (COOKE *et al.*, 1979).

As células eucarióticas são organizadas em compartimentos que estão delimitados por membranas, muito semelhante à própria membrana citoplasmática da célula, e nesses compartimentos ou organelas existem mecanismos de controle do pH interno. Os resultados de predições de ponto isoelétrico das proteínas HSP20 agrupadas na subfamília P (cloroplasto) (próxima seção a ser apresentada e discutida), demonstram que elas tem pI na faixa entre 7.74 e 9.52, a faixa mais alta de pI encontrada nas análises. Organelas como cloroplasto, por exemplo, apresentam pH alcalino devido aos mecanismos de bombeamento de prótons através da membrana durante a cadeia de transporte de elétrons (ALBERTS *et al.*, 2002). Os resultados de pI para as proteínas agrupadas nas subfamílias mitocondriais estiveram na faixa entre 6.95 e 7.02. Para as proteínas da subfamília RE, os resultados de pI estiveram na faixa entre 5.5 e 6.5.

Os resultados de predição de peso molecular e pI se mostraram complementares aos de predição de domínios e demais análises para predição de Hsp20 em soja, corroborando várias das predições realizadas. O conhecimento, mesmo que teórico, sobre as características dessas proteínas, facilitam o isolamento e a caracterização *in vivo* daquelas HSP20 de interesse por técnicas de proteômica, como os géis bidimensionais.

5.5.2 Subfamílias de GmHsp20

Os resultados obtidos nas análises filogenéticas com o alinhamento das sequências peptídicas das 45 GmHSP20 (somente as responsivas ao estresse de calor) mostraram haver uma identidade mínima de 17,39% e máxima de 98,05%

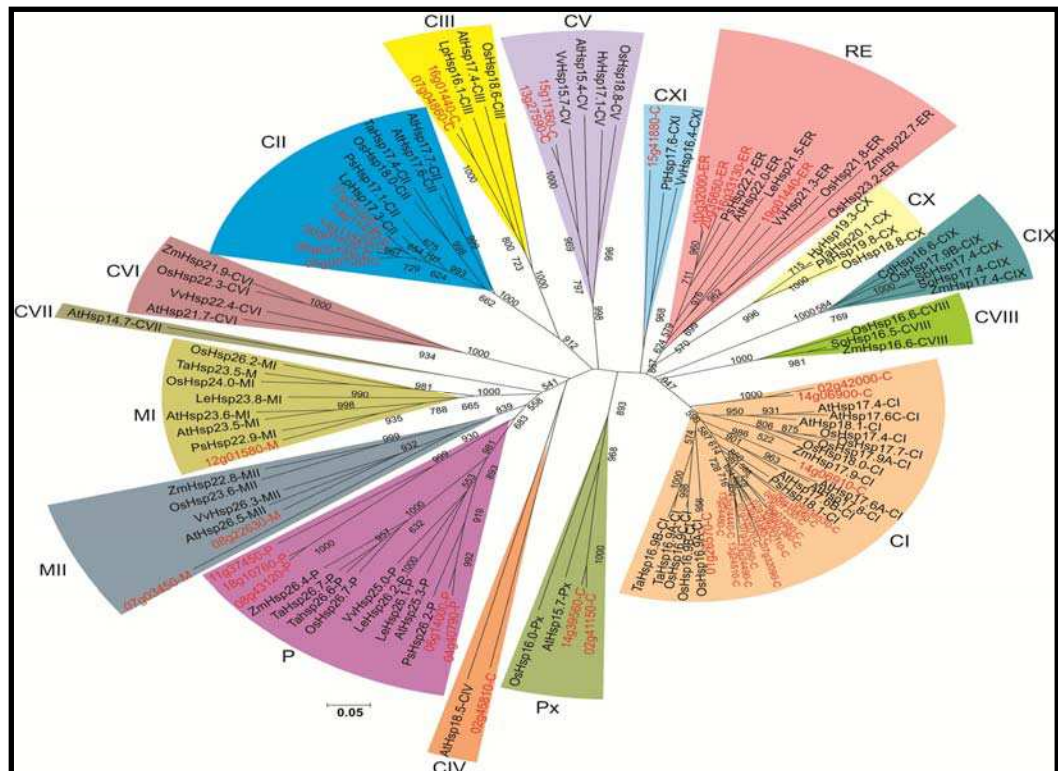
entre as GmHSP20, principalmente evidenciada na região C-terminal das sequências correspondentes ao domínio ACD. Comparando duas sequências de GmHsp20 de classes diferentes, GmHsp18.5-C e GmHsp17.9-D, Raschke e colaboradores (1988) obtiveram identidade de 38%, sendo maior principalmente na extremidade 3' da região codificadora.

Para examinar as relações evolutivas da família Hsp20 da soja e dividi-las em subfamílias, uma árvore filogenética não enraizada foi construída a partir do alinhamento das sequências do ACD das 45 GmHsp20. A inclusão de representantes de HSP20 de diversas espécies para todas as subfamílias já descritas possibilitou a diferenciação das GmHSP20 em 11 das 16 subfamílias já descritas em outras espécies, sendo 11 citoplasmáticas (CI, CII, CIII, CIV, CV, CVI, CVII, CVIII, CIX, CX e CXI), 2 mitocondriais (MI e MII), uma de cloroplasto (P), uma de retículo endoplasmático (ER) e uma de Peroxissomo (PX). A maioria das GmHsp20 foram distribuídas nas subfamílias CI (19 GmHsp20) e CII (6 GmHsp20), assim como observado em arroz (OUYANG *et al.*, 2009). Para as demais subfamílias citoplasmáticas CIII, CIV, CV, CVI, CVII, CVIII, CIX, CX e CIX foram distribuídas 2, 1, 2, 0, 0, 0, 0, 0 e 1 GmHsp20, respectivamente. No geral, a distribuição das GmHSP20 nas subfamílias foi coerente (para MI e ER) ou idêntica (para MII, P e Px) com a classificação de HSP20 previamente reportada para o arroz. O grande número de proteínas GmHSP20 classificadas nas subfamílias citoplasmáticas indica que o citoplasma possa ser o principal local de atuação das proteínas HSP20. No citoplasma é onde acontece a montagem das proteínas, logo após sua síntese pelo processo de tradução. Assim, é provável que essa concentração de HSP20 nas subfamílias citoplasmáticas esteja relacionada com sua função de impedir que essas proteínas realizem dobramentos inadequados ou interações inapropriadas formando agregados.

Notavelmente, os membros de uma mesma subfamília de espécies distintas aparecem filogeneticamente mais relacionados do que membros de uma mesma espécie pertencentes a subfamílias distintas (Figura 25). Isso indica que há uma provável sintenia entre as HSP20 de soja, arroz, *Arabidopsis*, e as demais espécies utilizadas na análise filogenética das GmHsp20. Os genes Hsp20 provavelmente tiveram um ancestral comum que deu origem às diferentes subfamílias antes da diversificação entre essas espécies, e isso foi conservado

durante a evolução (OUYANG *et al.*, 2009). É interessante notar que o padrão de distribuição de íntrons parece ser conservado entre membros de uma mesma subfamília. Assim, para as proteínas classificadas como mitocondriais foi encontrado um padrão com pelo menos um íntron (padrão 2 e 3), para as localizadas no cloroplasto predomina o padrão 2, com exceção do glyma18g10760 que segue o padrão 3, e para as citoplasmáticas o padrão de íntron 1 é predominante (Tabela 3 e Figura 28).

Figura 17 - Árvore filogenética e distribuição em subfamílias dos 45 GmHsp20 identificados em soja. A árvore foi construída com base no alinhamento da sequência de resíduos de aminoácidos que formam o domínio alfa-cristalino (ACD) dos organismos: *Glycine max* (Hsp20 de soja em análise estão destacadas em vermelho), *At Arabidopsis thaliana*, *Os Oryza sativa*, *Vv Vitis vinifera*, *Ta Triticum aestivum*, *Ps Pisum sativum*, *Le Solanum lycopersicum*, *Zm Zea mays*, *Lp Lycopersicon peruvianum*, *Cd Cynodon dactylon*, *So Saccharum officinarum*, *Sb Sorghum bicolor*, *Hv Hordeum vulgare*. As GmHsp20 foram distribuídas em 11 das 16 subfamílias já descritas. Nenhuma das GmHsp20 candidatas foram enquadradas para as subfamílias núcleo/citoplasmáticas CVI, CVII, CVIII, CIX e CX. Para nomenclatura das subfamílias RE (retículo endoplasmático), P (plastídios/cloroplasto), Px (peroxissomo), MI e MII (mitocôndria).



Todas as proteínas HSP20 de plantas são codificadas por genes nucleares (OUYANG *et al.*, 2009). Esses genes são divididos em até 16 subfamílias, que estão relacionadas com o endereçamento e instalação das proteínas por eles codificadas em diferentes locais da célula. Além disso, dentro da classe das HSP20 de plantas existem aquelas que depois de sintetizadas participam como proteínas intrínsecas nas membranas (FRIEDBERG, 2003).

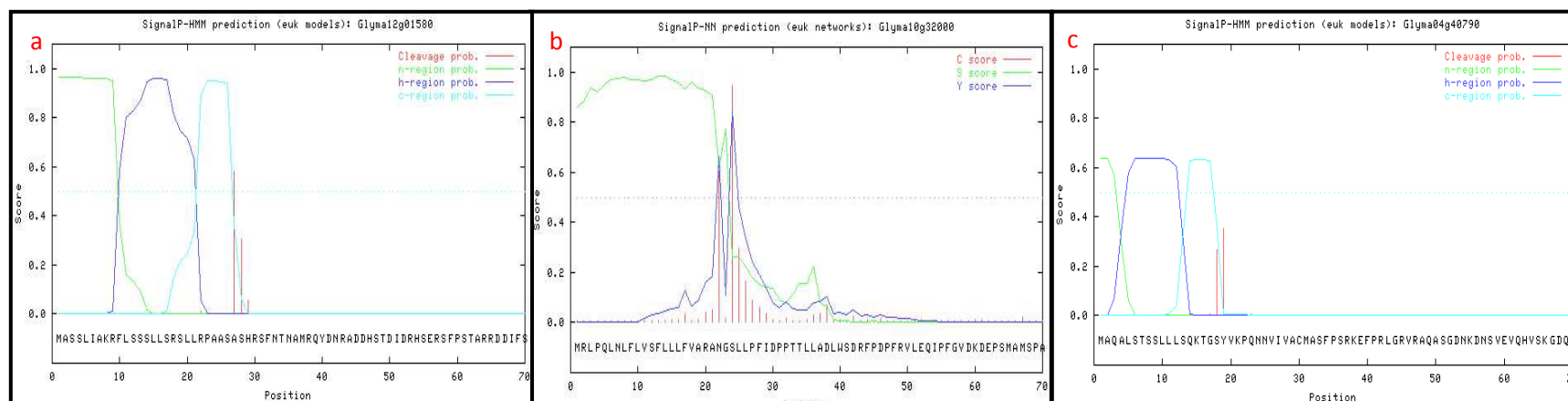
As análises em programas de predição de localização subcelular de proteínas aplicadas aos 45 genes identificaram 5 proteínas com localização predita para o cloroplasto (P), 4 proteínas com localização predita para o retículo endoplasmático, 3 mitocondriais (M) e 31 citoplasmáticas (C) (tabela 3). Esse resultado foi coerente com o obtido nas análises filogenéticas para o agrupamento das HSP20 de soja em subfamílias, como discutido anteriormente. Apenas a subdivisão da subfamília peroxissômica foi estabelecida a partir do alinhamento com proteínas Hsp20 experimentalmente localizadas nos peroxissomos, já que os programas de predição de localização celular e busca por peptídeo sinal não detectam diferenças entre proteínas citoplasmáticas e peroxissômica.

Todas as 45 proteínas foram analisadas pelos programas de predição de peptídeos sinal (SignalP) e predição de domínio transmembrana, TMHMM. Os resultados de predição de peptídeo sinal foram positivos para todos os genes agrupados nas análises filogenéticas na subfamília ER (Figura 26a). Esse resultado era esperado, já que normalmente as proteínas com função em outras organelas apresentam um sinal de endereçamento, o peptídeo sinal, para serem transportadas. Os peptídeos sinal governam o transporte e a localização das proteínas na célula. Muitos tipos diferentes de sinais de secreção são encontrados, no entanto em eucariotos o peptídeo sinal das proteínas tem como alvo muitas vezes o retículo endoplasmático, exatamente onde estas podem sofrer modificações pós-traducionais, tais como glicosilação, para em seguida serem secretadas (ALBERTS *et al.*, 2002).

Ainda, os resultados para a predição de peptídeo sinal demonstraram que o gene identificado pelo glyma12g01580, agrupado na subfamília MI, foi positivo para peptídeo sinal (Figura 26b). O que corrobora a análise filogenética que agrupou esse gene na subfamília MI. Os resultados de predição de peptídeo sinal também foram positivos para o Glyma04g40790 e Glyma06g14000 (Figura 14c). Na célula

existem diferentes tipos de sinais de endereçamento de proteínas para as organelas. Mas o principal é baseado na extensão contínua de sequência de aminoácidos com tamanho entre 15 e 60, denominado peptídeo sinal, que geralmente está localizado na região aminoterminal. Exatamente o padrão encontrado para os glymas 16g33130, 19g01440, 20g35650, 10g32000, 12g01580, 4g40790 e 6g14000 (ALBERTS *et al.*, 2002). Os peptídeos sinal tem função em marcar as proteínas que serão exportadas para determinados locais, como por exemplo, o ambiente extracelular ou organelas, como reticulo endoplasmático, as mitocôndrias, os cloroplastos e os peroxissomos, onde exercerão suas funções (ALBERTS *et al.*, 2002). Notavelmente, nenhuma das proteínas agrupadas na subfamília peroxissômica tiveram resultados positivos para a predição de peptídeos sinal.

Figura 18 - Resultados positivos para a predição de peptídeo sinal. **a.** Resultado positivo para o Glyma10g32000 na predição de peptídeo sinal. No SignalP foram obtidos resultados positivos e bastante semelhantes para as GmHSP20, Glyma16g33130, Glyma19g01440 e Glyma20g35650, agrupadas junto com o Glyma10g32000 na subfamília ER na filogenia. **b.** Resultado positivo para o Glyma12g01580 na predição de peptídeo sinal, agrupado na subfamília MI. **c.** Resultado positivo para o Glyma04g40790 na predição de peptídeo sinal, agrupado na subfamília P. O Glyma06g14000 e Glyma18g10760, também agrupados na subfamília P, apresentaram resultado positivo e bastante semelhante ao obtido para o Glyma04g40790. Nos gráficos são atribuídas pontuações (scores) para a predição de peptídeo sinal (S-score), o C-score é um provável de local de clivagem. Y-score combina o resultado de C-score com S-score para indicar com melhor precisão um provável sítio de clivagem.



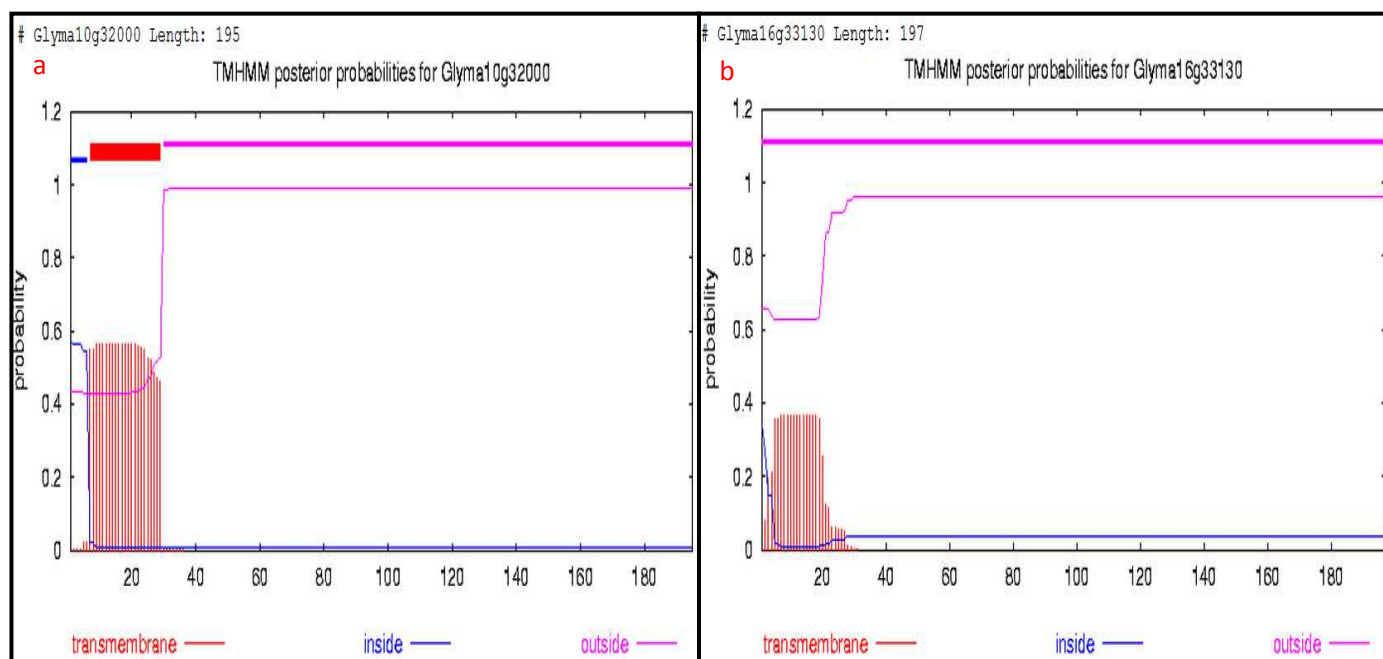
Os resultados de predição de domínio transmembrana para os genes, agrupados na subfamília RE, foram positivos com alta probabilidade para glyma10g32000 e glyma19g01440, mas tiveram baixa probabilidade para glyma16g33130, glyma20g35650 e glyma18g10760, não caracterizando como um resultado seguramente positivo (Figura 27a). Este padrão de posição de domínio transmembrana e peptídeo sinal, associado ao padrão obtido de estrutura secundária para as proteínas de RE, indicam que aquelas com resultado de alta probabilidade sejam proteínas transmembrana, apresentando uma cadeia polipeptídica que atravessa a bicamada lipídica em uma conformação de α -hélice, como a maioria das proteínas transmembrana. Os resultados de padrão de estrutura secundária para ER e demais subfamílias podem ser conferidos na próxima seção (Predição de estruturas secundárias e domínios funcionais para as subfamílias das GmHSP20).

Segundo Alberts e colaboradores (2002), proteínas que apresentam domínio transmembrana em regiões de estrutura secundária α -hélice, em geral são de passagem única através da membrana, ou seja, o peptídeo atravessa apenas uma vez. Além disso, segmentos da cadeia polipeptídica que atravessam a membrana são compostos principalmente de resíduos de aminoácidos com cadeia laterais apolares. As ligações peptídicas estabelecidas na bicamada em geral formam pontes de hidrogênio entre elas, e isso pode ser ainda mais reforçado caso a cadeia polipeptídica tenha uma estrutura secundária em α -hélice, exatamente o que foi encontrado após as análises de estrutura secundária para as proteínas HSP20 de RE.

As outras duas proteínas, agrupadas na análise filogenética para RE, apresentam baixa probabilidade para peptídeo transmembrana e não podem ser consideradas como proteínas transmembrana (Figura 27b). Segundo Alberts e colaboradores (2002), existem tipos de proteínas de membrana que se localizam inteiramente no citosol ou nas organelas associadas a uma das camadas bilipídica. Essa associação ocorre principalmente por meio de uma α -hélice exposta na superfície da proteína (ALBERTS *et al.*, 2002). Talvez seja o caso dessas duas proteínas HSP20, codificadas pelos genes Glyma16g33130 e Glyma20g35650, agrupadas na subfamília RE. Contudo, isso apenas pode ser afirmado caso novas

evidências, oriundas de análises adicionais, corroborem os resultados obtidos no TMHMM.

Figura 19 - Diagrama para o resultado, com alta probabilidade, positivo (a) e de baixa probabilidade (b). **a.** Resultado com alta probabilidade, a predição da presença de domínio transmembrana no Glyma10g32000. Na legenda é possível verificar as linhas que indicam a predição de regiões de domínio transmembrana (vermelho), regiões da proteína que provavelmente estejam dispostas para o lado externo da membrana (rosa) e do lado interno (azul). O Glyma19g01440 também apresentou um resultado muito semelhante para a predição de domínio transmembrana obtido para o Glyma10g32000. **b.** Diagrama obtido do resultado, mas com baixa probabilidade para positivo, a predição de domínio transmembrana para o Glyma16g33130. O Glyma20g35650 também apresentou um resultado bastante semelhante. As GmHSP20, codificadas pelos Glyma16g33130, Glyma19g01440 Glyma10g32000 e Glyma20g35650, formam a subfamília ER na filogenia.



Aparentemente, as estruturas α -hélice determinadas para as HSP20 de RE (porção N-terminal) são também domínios transmembrana. Isso pode ser melhor verificado nos resultados de estrutura padrão para cada subfamília na seção seguinte. As HSP20 endomembrana já foram descritas em *Arabidopsis* AtHSP22.0 e na soja GmHSP22.3 (LaFAYETTE *et al.* 1996; HELM *et al.*, 1995 *apud* OUYANG *et al.*, 2009). Até o momento, as observações das HSP20 endomembranares nessas plantas, como também em milho e cevada, sugerem que elas possam ter funções

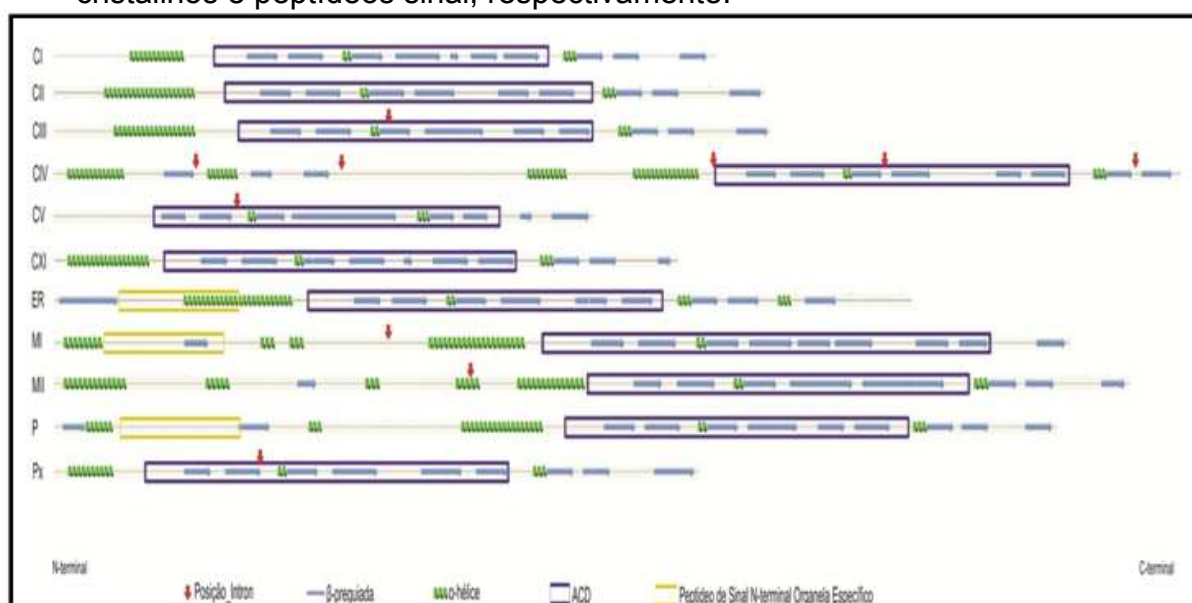
importantes na proteção da membrana durante o estresse, principalmente de calor (HELM *et. al.*, 1995).

5.5.3 Predição de Estruturas Secundárias e Domínios Funcionais para as Subfamílias das GmHSP20

As HSP20 apresentam uma estrutura secundária bem diferenciada entre as subfamílias. Contudo, o ACD presente em todas elas é principalmente constituído por estruturas de folha- β , variando apenas em número de acordo com a subfamília (MONTFORT *et al.* 2001). Os resultados de predição de modelos de estrutura secundária para cada subfamília de GmHSP20, demonstram que as subfamílias CI e CII apresentam diferença no tamanho da estrutura α -hélice na porção aminoterminal, e uma estrutura folha- β , sendo 7 segmentos folha- β para CI e 6 segmentos folha- β para CII (Figura 28). Esses padrões de estrutura secundária do domínio ACD para CI e CII foram muito similares aos obtidos em *Arabidopsis* por Siddique e colaboradores (2008).

A subfamília CIII também é bastante semelhante a CII, mas baseado nas análises de genes com íntrons, todos os representantes de CIII apresentam um íntron dentro do ACD e da terceira estrutura folha- β , padrão exato ao encontrado por Siddique e colaboradores em *Arabidopsis* (2008). A subfamília citoplasmática CIV é bem distinta das demais já relatadas principalmente na porção N-terminal, e com 5 íntrons. Padrão muito discrepante ao encontrado em *Arabidopsis* (SIDDIQUE *et al.*, 2008), e em arroz nenhuma proteína para CIV foi descrita (SARKAR *et al.*, 2009). Para a subfamília CV houve conservação em relação aos padrões encontrados em arroz e *Arabidopsis*, a presença de um íntron logo após a segunda estrutura folha- β dentro do domínio ACD é característica peculiar.

Figura 28- Diagrama ilustrando os padrões de estrutura secundária preditos para as proteínas das subfamílias de GmHSP20. As posições de íntrons estão apontadas por setas em vermelho. As setas horizontais azuis delimitam regiões de formação de estruturas secundárias β -pregueadas. Espirais em verde são utilizadas para delimitar as regiões de possíveis formações de estruturas secundárias em α -hélices. Os retângulos em azul e amarelo delimitam as regiões preditas para os domínios α -cristalinos e peptídeos sinal, respectivamente.



Não foi encontrada nenhuma proteína GmHsp20 para a subfamília CVI e CVII. Para *Arabidopsis* foram descritas apenas para CVI e CVII, e para arroz somente para CVI (SIDDIQUE *et al.*, 2008; SARKAR *et al.*, 2009).

Ainda não foram encontrados em *Arabidopsis* genes pertencentes às subfamílias citoplasmáticas CVIII, CIX, CX e CXI, entretanto genes membros dessas subfamílias já foram descritos em espécies como *Oryza sativa*, *Zea mays* e outras (SARKAR *et al.*, 2009). Assim como em *Arabidopsis*, na soja não foi possível identificar, entre os Hsp20, membros representantes para essas novas subfamílias. O padrão de estrutura secundária para CXI em soja é bastante similar ao de CI e CII no ACD, a alteração está apenas no tamanho da estrutura α -hélice na região N-terminal. De forma semelhante, Sarkar e colaboradores (2009) também reportaram que a estrutura exon-íntron e a estrutura secundária das proteínas HSP20 da subfamília CXI em arroz aproximam-se à encontrada para a subfamília CI.

A principal característica de estrutura secundária das GmHSP20 da subfamília de retículo endoplasmático é a presença de duas grandes estruturas em folha- β e α -hélice na região N-terminal, na mesma região onde foi predito também

um peptídeo sinal. Esse resultado de padrão e localização de peptídeo sinal é o mesmo encontrado em *Arabidopsis* (SIDDIQUE *et al.*, 2008). As duas subfamílias mitocondriais apresentam de 4 a 5 estruturas α -hélice de tamanhos variados e 1 folha- β pequena ao longo na região N-terminal. A única diferença encontrada foi no número de estruturas folha- β dentro do ACD, sendo 7 e 5 segmentos/estruturas para MI e MII, respectivamente. O padrão para MI e MII também é semelhante quanto à posição C-terminal do ACD e a presença de um íntron na região N-terminal fora do ACD. Essa posição não é encontrada em nenhuma outra subfamília HSP20. Contudo, foi possível prever a presença de peptídeo sinal apenas para MI. Esse padrão está muito semelhante ao encontrado em *Arabidopsis* por Siddique e colaboradores (2008), principalmente para MI. Os padrões para P e Px estão coerentes com os encontrados em *Arabidopsis* e podem ser verificados através do diagrama ilustrado na Figura 28. Esses resultados quando comparados com os obtidos por Siddique e colaboradores (2008), indicam que houve uma conservação bastante significativa nas sequências e conseqüentemente na estrutura das HSP20 dentro de cada subfamília.

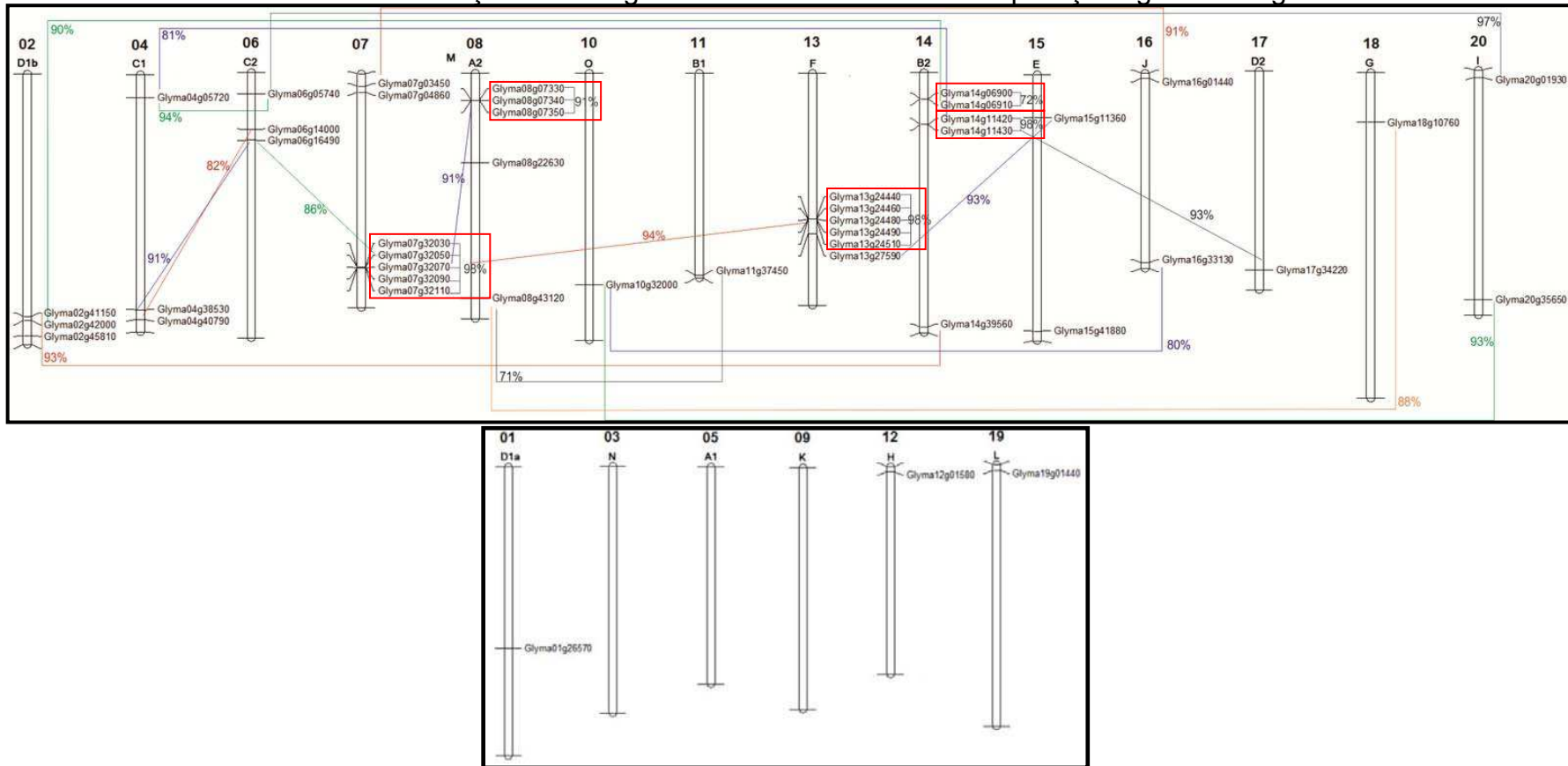
5.5.4 Localização Cromossômica e Duplicação de Genes GmHsp20

Para determinar a distribuição genômica dos genes GmHsp20, suas posições cromossômicas foram identificadas de acordo com o banco de dados Phytozome. Todos os 45 genes GmHsp20 estão organizados em 17 dos 20 cromossomos da soja, sendo os cromossomos 7 e 13 os que apresentam maior número de genes preditos de proteínas tipo HSP20 com 7 genes para o primeiro cromossomo e 6 para o último grupo de ligação. Na disposição geral dentro do contexto genômico, os GmHsp20 no genoma da soja foram encontrados em número de 1 gene no cromossomo 1 (D1a), 3 genes no cromossomo 2 (D1b), 3 no cromossomo 4 (C1), 3 no cromossomo 6 (C2), 7 no cromossomo 7 (M), 5 no cromossomo 8 (A2), 1 no cromossomo 10 (O), 1 no cromossomo 11 (B1), 1 no cromossomo 12 (H), 6 no cromossomo 13(F), 5 no cromossomo 14 (B2), 2 no cromossomo 15 (E), 2 no cromossomo 16 (J), 1 no cromossomo 17 (D2), 1 no cromossomo 18 (G), 1 no cromossomo 19 (L) e 2 no cromossomo 20 (I) (Tabela 3 e

Figura 29). Por outro lado, nenhum gene GmHsp20 está presente nos cromossomos 3, 5 e 9.

Um aspecto interessante sobre a distribuição genômica das Gmhsp20 a localização dos 45 genes, que em geral, ocorre nas extremidades cromossômicas, região onde a maior parte dos genes que codificam para proteínas são encontrados (SCHMUTZ *et al.*, 2010). Possivelmente, isso esteja associado as altas taxas de recombinação entre regiões de extremidades de cromátides homólogas, o que facilita a ocorrência de duplicações gênicas (SEE *et al.*, 2006). Dentre as principais forças para o processo de evolução de genomas, a duplicação gênica é uma das mais importantes. As famílias de genes são formadas durante a evolução através da duplicação cromossômica e também de duplicações que ocorrem ao longo do mesmo cromossomo, como é o caso da duplicação segmentar (CANNON *et al.* 2004). Pautados nos resultados obtidos por SCHMUTZ e colaboradores (2010) e nos resultados das análises de duplicação obtidos neste estudo, possivelmente, a evolução da família de genes Hsp20 no genoma da soja tenha envolvido um total de 21 duplicações gênicas, sendo 5 segmentares em 4 cromossomos. Este número é semelhante ao encontrado no genoma de arroz, onde um total de 23 genes OsHsp20 tem origem em eventos de duplicação (OUYANG *et al.*, 2009).

Figura 20 - Mapa de loci e coordenadas de eventos de duplicações dos genes parálogos candidatos a GmHsp20 nos cromossomos 17-20. As identificações de cada grupo de ligação são indicadas no topo de cada barra. Na primeira fileira estão apenas os cromossomos em que os GmHsp20 candidatos foram mapeados e apresentam valores de identidade para sequência de nucleotídeos acima de 70% cm outros genes, caracterizando uma possível duplicação gênica. Os genes possivelmente duplicados estão ligados por traços e a porcentagem de identidade anotada na mesma cor do traço. Os retângulos vermelhos indicam as duplicações gênicas segmentares.



Até o momento, a soja é o maior genoma de planta já sequenciado, com 1,1 gigabases. Esses dados permitiram além do conhecimento das sequências genômicas, a construção do mapa físico e genético da soja e facilitaram a realização de muitos estudos na genômica estrutural e funcional. Segundo Schmutz e colaboradores (2010), a soja é um organismo allotetraploplóide que passou por pelo menos dois eventos de duplicação genômica. Análises computacionais sugeriram ainda a ocorrência de rearranjos cromossômicos, às vezes envolvendo mais de dois cromossomos (SCHMUTZ *et al.*, 2010). Os eventos de duplicações do genoma resultaram em um genoma altamente duplicado, com quase 75% dos genes presentes em múltiplas cópias. A partir dos resultados de Schmutz e colaboradores (2010), foi possível inferir que o primeiro evento de duplicação, proto-soja (ancestral), ocorreu próximo das origens da linhagem de *Papilionoideae*, datada a cerca de 59 milhões de anos atrás. A segunda duplicação ocorre dentro da espécie *Glycine*, e estima-se ter ocorrido a 13 milhões de anos atrás. Em média, 61,4% dos genes homólogos são encontrados em blocos envolvendo apenas dois cromossomos; apenas 5,63% ao longo de três cromossomos, e 21,53% atravessando quatro cromossomos.

Os dados obtidos por Schmutz e colaboradores (2010) indicam que o cromossomo 14 possui 11,8% de seus genes conservados ao longo de outros três cromossomos (2, 13 e 17) e que os blocos restantes têm relação com fragmentos de outros 14 cromossomos no genoma. A análise de duplicação dos genes Hsp20 no genoma da soja permitiu a identificação de pelo menos 4 agrupamentos de genes parálogos Hsp20 localizadas no cromossomo 14, com os genes encontrados nos cromossomos 2, 4 e 17. Esses resultados estão de acordo com os apresentados por Schmutz e colaboradores (2010), que mostram que o cromossomo 14 possui elevada sintonia com o cromossomo 4, e principalmente com os cromossomos 2 e 17.

Segundo Schmutz e colaboradores (2010), na soja as expansões segmentares de famílias gênicas associadas com resposta basal de resistência são comuns, incluem famílias NBS-LRR, F-box, auxina-responsiva, proteínas e outros domínios comumente encontrados em grandes famílias. No caso das GmHsp20, essas duplicações segmentares parecem ter contribuído significativamente para o aumento no número de representantes da subclasse CI, que localizam-se nos

cromossomos 7, 8, 13 e 14. Em arroz, os membros da subfamília CI encontram-se também distribuídos em aglomerados de duplicações segmentares (OUYANG *et al.*, 2009). Considerando o conceito de parcimônia e os resultados obtidos por Schmutz e colaboradores (2010) a respeito da evolução do genoma da soja, essa conservação das proteínas HSP20 duplicadas dentro dos mesmos cromossomos, como observado nos genomas de arroz e da soja, provavelmente tem origem em processos de duplicação segmentar ocorrido antes dos eventos de divergência, na espécie ancestral, seguido pelas duplicações cromossômicas no próprio ancestral e também dentro das espécies (SCHLUETER *et al.*, 2008; SCHMUTZ *et al.*, 2010).

De acordo com os resultados de Ouyang e colaboradores (2009), por exemplo, os padrões de expressão dos genes Hsp20 de arroz, duplicados em segmentos, foram muito semelhantes dentro dos quatro segmentos. As duplicações segmentares representam uma fonte importante de nova variação genética e, dessa forma, podem estar relacionadas a origem de genes ou alelos de resposta a patógeno, onde as interações específicas proteína-proteína são necessárias para desencadear respostas de resistência. É interessante notar na figura de duplicação, que três dos seis GmHsp20, responsivos a infecção por nematóides, Glyma08g07340 (somente na literatura – responsivo a *H. glycines*) Glyma13g24460, Glyma14g06910 (literatura e análises de RT-qPCR), estão organizados em blocos de duplicações segmentares no genoma.

O cromossomo 20 é altamente homólogo ao cromossomo 10, mas ainda possui 7,08% de sua extensão gênica homóloga a fragmentos de outros quatro cromossomos (SCHMUTZ *et al.*, 2010). Essa relação pôde ser também observada para os genes de Hsp20, visto que um dos genes classificados na subfamília RE, presente no braço inferior do cromossomo 10, Glyma10g32000, foi duplicado para a mesma extremidade no cromossomo 20. Esse mesmo gene localizado no cromossomo 10, também tem uma alta identidade (80%) com Glyma16g33130 classificado na subfamília RE. Ainda, para o cromossomo 20 foi observada uma possível duplicação, com 97% de identidade, entre o Hsp20 Glyma20g01930 (localizado no braço superior) e o Glyma06g05740 do cromossomo 6, localizado na mesma extremidade. Entre os cromossomos 4 e 6, que apresentam alto grau de sintenia (Schmutz *et al.*, 2010), foram observadas três prováveis

duplicações, uma delas com 82% de similaridade envolvendo 2 dos 4 genes classificados na subfamília P.

6 CONCLUSÕES

As análises *in silico* realizadas para a seleção dos genes anotados nos bancos de dados como Hsp20 de soja demonstraram ser muito eficientes. O uso dos parâmetros de presença de domínio ACD na região C-terminal, presença dos *cis* elementos HSE e verificação do nível de expressão nos bancos de dados de expressão foi uma estratégia eficaz na otimização do tempo destinado às análises *in vivo*, pela seleção de candidatos mais prováveis. Inicialmente havia 76 modelos gênicos anotados como Hsp20 de soja e, após a seleção *in silico*, esse número caiu para 52 candidatos. Desses 52, apenas 47 genes foram analisados *in vivo*, já que os demais puderam ser validados como Gmhsp20 por meio dos dados de expressão digital disponíveis. Dos 47 analisados *in vivo*, apenas 7 não foram responsivos ao estresse de choque térmico – calor. Quando somados os cinco genes, não inclusos para as análises *in vivo*, aos 40 genes, identificados como responsivos ao estresses de calor, foram totalizados 45 genes Hsp20 na soja.

As análises de filogenia, localização subcelular e de estrutura secundária para os 45 GmHSp20 demonstraram que os mesmos estão distribuídos em 11 subfamílias, para as quais foi possível observar um padrão de estrutura secundária previamente estabelecido para HSP20. Além disso, as análises de filogenia demonstraram que os genes Hsp20 apresentam uma relevante conservação entre as plantas, principalmente observada entre os membros da subfamília CI.

A região promotora das GmHsp20 é minimamente definida pela presença de núcleos HSE, mas outros elementos reguladores são também importantes tais como o promotor cerne TATAbox, W-box, CCAAT box, ABRE, MYB e TA *rich region* para os estresse abióticos e bióticos foco nessa pesquisa.

Os tratamentos de estresses abióticos e bióticos tiveram efeitos significativos na alteração da expressão dos genes Hsp20 preditos. O tratamento de calor possibilitou a identificação dos Hsp20 por RT-PCR e as análises de RT-qPCR corroboraram esses resultados. Para o estresse de frio foram identificados 5 genes relacionados, todos eles Hsp20, através das análises de RT-qPCR, enquanto que a infecção pelo nematóide *M. javanica*, possibilitou a identificação de 7 genes responsivos a este estresse, sendo 2 exclusivamente expressos nessa condição. Em geral, na cultivar de soja suscetível (BRS133) esses genes foram induzidos no

tratamento de 4 dpi e reprimidos no tratamento de 8 dpi, enquanto que na cultivar de soja resistente (PI595099) os genes tiveram indução tardia no tratamento de 8 dpi.

Esses resultados são promissores para as pesquisas de melhoramento, considerando principalmente a busca pelos institutos de pesquisa por cultivares cada vez mais adaptadas aos mais diferentes estresses. Os genes e promotores identificados constituem importante fonte para o desenvolvimento de ferramentas biotecnológicas aplicadas ao desenvolvimento de variedades de soja mais adaptadas a estresses que acometem a cultura da soja no Brasil.

REFERÊNCIAS

- ABAD, P. et al. Root-knot nematode parasitism and host response: molecular basis of a sophisticated interaction. **Molecular Plant Pathology**, London, v. 4, n. 4, p. 217–224, 2003.
- ABE, M.; TAKAHASHI, T.; KOMEDA, Y. Identification of a cisregulatory element for L1 layer-specific gene expression, which is targeted by an L1-specific homeodomain protein. **The Plant Journal**, Oxford, v.26, p. 487–494, 2001.
- ALBERTS, B. et al. Compartimentos intracelulares e endereçamento de proteínas. **Molecular Biology of the Cell**, New York, p. 659–710, 2002.
- AL-WHAIBI, M. H. Plant heat-shock proteins: a mini review. **Journal of King Saud University (Science)**, Riyadh, v. 22, n. 6, 2010.
- ARIMURA, G.; MAFFEI, M. E. Calcium and secondary CPK signaling in plants in response to herbivore attack. **Biochemical and Biophysical Research Communications**, Orlando, n. 400, p. 455–460, 2010.
- ÁVILA, C. J.; JUNIOR, R. P. S. Diagnóstico da Ocorrência de Pragas de Solo em Mato Grosso do Sul. In: REUNIÃO SUL-BRASILEIRA SOBRE PRAGAS DE SOLO: PRAGAS-SOLO-SUL, 10., 2007, Dourados. **Anais...** Dourados: Embrapa, 2007. p. 17–50.
- BARBAZUK, W. B.; FU, Y.; MCGINNIS, K. M. Genome-wide analyses of alternative splicing in plants: Opportunities and challenges. **Genome Research**,. New York, v.18, p.1381–1392, 2008.
- BARCALA, M. et al. Distinct heat-shock element arrangements that mediate the heat shock, but not the late-embryogenesis induction of small heat-shock proteins, correlate with promoter activation in root-knot nematode feeding cells. **Plant Molecular Biology**, Dordrecht, v. 66, p. 151–164, 2008.
- BARCALA, M. et al. Early transcriptomic events in microdissected *Arabidopsis* nematode-induced giant cells. **The Plant Journal**, Oxford, v. 61, p. 698–712, 2010.
- BASHA, E.; FRIEDRICH, K. L.; VIERLING, E. The N-terminal Arm of Small Heat Shock Proteins Is Important for Both Chaperone Activity and Substrate Specificity. **The Journal of Biological Chemistry**, v. 281, n. 52, p. 39943–39952, 2006.
- BENNETT-LOVSEY, R. M. et al. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. **Proteins**, New York, v. 70, p. 611–625, 2008.
- BARKER, K. R.; KOENNING, S. R. Developing sustainable systems for nematode management. **Annual Review of Phytopathology**, Palo Alto, n. 36, p. 165–205, 1998.

BETTIOL, W.; MORANDI, M. A. B. **Biocontrole de doenças de plantas: uso e perspectivas**. Jaguariúna: Embrapa Meio Ambiente, 2009.

BIRD, D. M.; OPPERMAN, C. H.; DAVIES, K. G. Interactions between bacteria and plant-parasitic nematodes: now and then. **International Journal for Parasitology**, Oxford, v. 33, n. 11, p. 1269-1276, 2003.

BONATO, E. R. **Estresses em soja**. Passo Fundo: Embrapa Trigo, 2000.

BONETI, J. I. S.; FERRAZ, S. Modificações do método de Hussey e Baker para a extração de *Meloidogyne exigua* de cafeeiro. **Fitopatologia Brasileira**, Brasília, p. 553, 1981

BOTER, M. et al. Structural and functional analysis of SGT₁ reveals that its interaction with HSP90 is required for the accumulation of Rx, an R protein involved in plant immunity. **Plant Cell**, Rockville, v. 19, p. 3791–3804, 2007.

BRASIL, P. E. et al. ELISA versus PCR for diagnosis of chronic Chagas disease: systematic review and meta-analysis. **BMC Infectious Diseases**, London, v. 337, n. 10, 2010.

BUCHANAN, B. B.; GRUISSEM, W.; JONES, R. L. **Biochemistry e Molecular Biology of Plants**. Rockville, American Society of Plant Physiologists, 2000. p. 1044-1154.

BUCK, M. J.; ATCHLEY, W. R. Phylogenetic analysis of plant basic helixloop-helix proteins. **Journal of Molecular Evolution**, New York, v. 56, p. 742-750, 2003.

CAITAR, L. P. **Diferenças da abordagem na administração com foco na segurança do SGBD PostgreSQL e Oracle**. 2011. Monografia (Especialização em Análise de sistemas e banco de dados) - Universidade do Norte do Paraná, Londrina.

CANNON, S. B. et al. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. **BMC Plant Biology**, London, v. 10, n. 4, 2004.

CARDOZO, M. C.; PALMEIRA, M. E. Desafios de logística nas exportações brasileiras do complexo agronegocial da soja. **Observatorio de la Economía Latinoamericana**. n. 71, 2006.

CARNEIRO, R. G.; MAZZAFERA, P.; FERRAZ, L. C. C. B. Carbon partitioning in soybean infection with *Meloidogyne incógnita* and *M. javanica*. **Journal of Nematology**, College Park, v. 31, p. 348-355, 1999.

CARNEIRO, R. M. D. G. et al. Ocorrência de *Meloidogyne* spp. e fungos nematófagos em hortaliças no Distrito Federal, Brasil. **Nematologia Brasileira**, Campinas, v. 32, n. 8, p. 135-141, 2008.

CASHIKAR, A. G.; DUENNWALD, M.; LINDQUIST, S. L. A chaperone pathway in protein disaggregation: Hsp26 alters the nature of protein aggregates to facilitate

reactivation by Hsp104. **The Journal of Biological Chemistry**, Bethesda, v. 280, n. 25, p. 23869-23875, 2005.

CHEN, L.; SHIMAMOTO, K. Emerging roles of molecular chaperones in plant innate immunity. **Journal of General Plant Pathology**, Saga, n. 77, p. 1–9, 2011.

CHITWOOD, D. J. Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture-Agricultural Research Services. **Pest Management Science**, Sussex, n. 59, p. 748-753, 2003.

CHOMCZYNSKI, P.; SACCHI, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Analytical Biochemistry**, New York, v. 1, n. 162, p. 156-159, 1987.

COMPANHIA NACIONAL DE ABASTECIMENTO. CONAB. Safras. Disponível em: <<http://www.conab.gov.br/conteudos.php?a=1253&t=2>>. Acesso em: 3 out. 2011.

COSTAMILAN, L. M. Estresses ocasionados por doenças e por nematóides. In: BONATO, E. R. (Ed.). **Estresses em soja**. Passo Fundo: Embrapa Trigo, 2000. p.19-44.

COOKE, R. J.; OLIVER, J.; DAVIES, D. D. Stress and protein turnover in Lemna minor. **Plant Physiology**, Minneapolis, v. 64, p. 1109-1113, 1979.

CZARNECKA, E.; KEY, J. L.; GURLEY, W. B. Regulatory Domains of the Gmhsp1 7.5-E Heat Shock Promoter of Soybean. **Molecular and Cellular Biology**, Washington, v. 9, n. 8, p. 3457-3463, 1989.

DASH, S.; MOHANTY, N.; Response of seedlings to heat stress in cultivars of wheat: growth temperature-dependent differential modulation of photosystem 1 and 2 activity, and foliar anti-oxidant defense capacity. **Journal of Plant Physiology**, Stuttgart, n. 159, p. 49-59, 2002.

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

DE JONG, W. W.; LEUNISSEN, J. A.; VOORTE, C. E. M. P. Evolution of the α -Crystallin/Small Heat-Shock Protein Family. **Molecular Biology and Evolution**, Chicago, v. 10, n. 1, p. 103-126, 1993.

DE LEY, P.; BLAXTER, M. Systematic position and phylogeny. In: LEE, D. L. (Ed.). **The biology of nematodes**. London: Taylor & Francis, 2002. p. 1-30.

DEY, S.; GHOSE, K.; BASU, D. Fusarium elicitor-dependent calcium influx and associated ROS generation in tomato is independent of cell death. **European Journal of Plant Pathology**, Dordrecht, n. 126, p. 217–228, 2010.

DIAS, W. P. et al. Nematóides de cisto da soja: biologia e manejo pelo uso da resistência genética. **Nematologia Brasileira**, Piracicaba, v. 33, n. 1, p.1–16, 2009.

DIAS, W. P. et al. **Nematóides de soja: identificação e controle**. Londrina: Embrapa, 2010. (Circular Técnica, 76).

EGAWA, C. et al. Differential regulation of transcript accumulation and alternative splicing of a DREB2 homolog under abiotic stress conditions in common wheat. **Genes Genetic Systems**. v.81, p.77–91, 2006.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA - EMBRAPA. Soja em números (safra 2010/2011). Disponível em: <http://www.cnpso.embrapa.br/index.php?cod_pai=2&op_page=294>. Acesso em: 20 nov. 2011.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. EMBRAPA. **Tecnologia de produção de soja: região central do Brasil**. Londrina: Embrapa Soja, 2007.

EFEOGLU, B. Heat shock proteins and heat shock response in plants. **Gazi University Journal of Science**, Ankara, v. 22, n. 2, p. 67-75, 2009.

ESCOBAR, C. et al. The american phytopathological society induction of the Hahsp17.7G4 promoter by Root-Knot Nematodes: involvement of Heat-Shock elements in promoter activity in giant cells. **Molecular Plant-Microbe Interactions**, Saint Paul, v. 16, n. 12, p. 1062–1068, 2003.

FERRAZ, L. C. C. B. et al. **Relações parasito-hospedeiro nas Meloidoginoses da soja**. Embrapa Soja: Warta, 2001.

FERRAZ, L. C. C. B. O nematóide *Pratylenchus brachyurus* e a soja sob plantio direto. **Revista Plantio Direto**, Passo Fundo, v. 15, n. 95, p. 23-27, 2006.

FINKA, A.; MATTOO, R. U.; GOLOUBINOFF, P. Meta-analysis of heat- and chemically up regulated chaperone genes in plant and human cells. **Cell Stress and Chaperones**, Mansfield, n. 16, p. 15-31, 2010.

FRAIRE-VELÁZQUEZ, S.; RODRÍGUEZ-GUERRA, R.; SÁNCHEZ-CALDERÓN, L. Abiotic and biotic stress response crosstalk in plants. In: SHANKER, A. K.; VENKATESWARLU, B. **Abiotic stress response in plants: physiological, biochemical and genetic perspectives**. Croácia: InTech, 2011. p. 3-26.

FRIEDBERG, J. N. **An investigation of a heat shock transcription factor and the heat shock response under temperature stress in Alfafa**. 2003. Thesis (PhD in Molecular Genetics & Biotechnology) - University of Guelph, Guelph.

FUGANTI, R. et al. Identificação de marcadores moleculares de microssatélites para a seleção de genótipos de soja resistentes a *Meloidogyne javanica*. **Nematologia Brasileira**, Campinas, v. 28, n. 2, p. 125-130, 2004.

FUGANTI, R. et al. Size of AT(n) insertions in promoter region modulates Gmhsp17.6-L mRNA transcript levels. **Journal of Biomedicine and Biotechnology**, Washington, n. 10, p.1-9, 2010.

GAROFALO, C. G. et al. Expression analysis of small heat shock proteins during compatible and incompatible plant-pathogen interactions. **Advanced Studies in Biology**, Ruse, v. 1, n. 5, p. 197–205, 2009.

GELLI, A.; BLUMWALD, E. Hyperpolarization-activated Ca²⁺- permeable Channels in the Plasma Membrane of Tomato Cells. **Journal of Membrane Biology**, New York, n. 155, p. 35-45, 1997.

GRIFFITHS, A. J. F. **Introdução a genética**. 7. ed. Rio de Janeiro: Guanabara Koogan, 2002.

GHEYSEN, G.; FENOLL, C. Gene expression in nematode feeding sites. **Annual Review of Phytopathology**, Palo Alto, v. 40, p. 191–219, 2002.

GOULD, K. S. et al. Nitric oxide production in tobacco leaf cells: a generalized stress response. **Plant, Cell & Environment**, Logan, n. 26, p. 1851–1862, 2003.

GRUNDLER, F. M. W.; BÖCKENHOFF, A. Physiology of nematode feeding and feeding sites. In: FENOLL, C.; GRUNDLER, F. M. W.; OHL, S. A. (Ed.). **Cellular and molecular aspects of plant-nematode interactions**. Dordrecht: Kluwer Academic Publishers, 1997. p.107–119.

GURUPRASAD, K.; REDDY, B. V.; PANDIT, M. W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. **Protein Engineering**, Oxford, n. 4, p. 155–161, 1990.

HELM, K. W.; SCHMEITS, J.; VIÉRLING, E. An endomembrane-localized small heat-shock protein from *Arabidopsis thaliana*. **Plant Physiology**, Minneapolis, v. 107, p. 287-288, 1995.

HINES, P. J.; ZAHN, L. M. What's Bugging Plants? **Science**, Washington, v. 324, n. 5928, p. 741, 2009.

HOHMANN, S. Osmotic stress signaling and osmoadaptation in yeasts. **Microbiology and Molecular Biology Reviews**. v.66, p.300–372, 2003.

HRUZ, T. et al. Genevestigator V3: a reference expression database for the meta-analysis of Transcriptomes. **Advances in Bioinformatics**, New York, v 2008, p.1-5, 2008.

HRUZ, T. et al. RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. **BMC Genomics**, Washington, v. 12, n. 156, p. 1:14, 2011.

HUA, J. From freezing to scorching, transcriptional responses to temperature variations in plants. **Current Opinion in Plant Biology**, Saint Louis, n. 12, p. 568–573, 2009.

KANDOTH, P. K. et al. The Soybean Rhg1 Locus for Resistance to the Soybean Cyst Nematode *Heterodera glycines* Regulates the Expression of a Large Number of

Stress- and Defense-Related Genes in Degenerating Feeding Cells. **Plant Physiology**, Minneapolis, n. 155, p. 1960–1975, 2011.

JAIN, M.; KHURANA, P.; TYAGI, A. K.; KHURANA, J. P. Genome-wide analysis of intronless genes in rice and Arabidopsis. **Functional & Integrative Genomics**, Perth, n. 8, p. 69–78, 2008.

KOTAK, S. et al. Complexity of the heat stress response in plants. **Current Opinion in Plant Biology**, Saint Lois, n. 10, p. 310–316, 2007.

KELLEY, L. A.; STERNBERG, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. **Nature Protocols**, New Jersey, v. 4, n. 3. p. 363–371, 2009.

KROGH, A. et al. Predicting transmembrane protein topology with hidden markov model: application to complete genomes. **Journal of Molecular Biology**, London, n. 305, p. 567–580, 2001.

LAFAYETTE, P. R. et al. Molecular characterization of cDNAs encoding low-molecular-weight heat shock proteins of soybean. **Plant Molecular Biology**, Dordrecht, n. 30, p. 159–169, 1996.

LEE, G. J.; VIERLING, E. A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. **Plant Physiology**, Minneapolis, v. 122, n. 1, p. 189–198, 2000.

LIN, M. F.; JUNGREIS, I.; KELLIS, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. **Bioinformatics**, Oxford, v. 27, p. 275–282, 2011.

LIU, H. T. et al. The calmodulin-binding protein kinase 3 is part of heat-shock signal transduction in Arabidopsis thaliana. **Plant Journal**, Oxford, n. 55, p. 760–773, 2008.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-2\Delta\Delta CT}$ Method. **Methods**, San Diego, v. 25, n. 4, p. 402–408, 2001.

LONG, S. P.; ORT, D. R. More than taking the heat: crops and global change. **Current Opinion in Plant Biology**, Saint Lois, n. 13, p. 241–248, 2010.

LOPES, K. P.; DIAS, S. R. Comparação entre diferentes tecnologias em banco de dados para manipulação rápida e ubíqua de dados biológicos. **Anuário da produção de iniciação científica discente**, v. 13, n. 16, p. 243–255, 2010.

LOPES, M. J. C. **Aspectos histopatológicos e mudanças na expressão de genes em genótipos de soja resistente, durante a interação com *Meloidogyne javanica***. 2009. Tese (Doutorado em Agronomia) - Universidade Federal de Lavras, Lavras.

LORENZO, O.; CHICO, J. M.; SANCHEZ-SERRANO, J. J.; SOLANO, R. JASMONATE-INSENSITIVE1 encodes a MYC transcription factor essential to

- discriminate between different jasmonate-regulated defense responses in *Arabidopsis*. **The Plant Cell**, Rockville, v. 16, p. 1938–1950, 2004.
- LOS, D. A.; MURATA, N. Membrane fluidity and its roles in the perception of environmental signals. **Biochimica et Biophysica Acta**, Alberta, v. 1666, p. 142–157, 2004.
- LUSCOMBE, N. M.; GREENBAUM, D.; GERSTEIN, M. What is bioinformatics? a proposed definition and overview of the field. **Methods of Information in Medicine**, Stuttgart, v. 40, n. 4, p. 346–358, 2001.
- MA, C. et al. Identification and characterization of a stress-inducible and a constitutive small heat-shock protein targeted to the matrix of plant peroxisomes. **Plant Physiology**, Minneapolis, v. 141, n. 1, p. 47-60, 2006.
- MAGALHÃES JÚNIOR, A. M. **Ácido Abscísico e o estresse abiótico**. Pelotas. Embrapa Clima Temperado, 2010 (Documentos, 307).
- MAIMBO, M. et al. Induction of a small heat shock protein and its functional roles in *Nicotiana glauca* plants in the defense response against *Ralstonia solanacearum*. **Plant Physiology**, Minneapolis, v. 145, p. 1588–1599, 2007.
- MARCELINO, F. C. **Avaliação de resíduos de transgênicos em alimentos no Brasil e desenvolvimento de metodologias de análise**. 2006. Tese (Doutorado em Genética e Melhoramento) - Universidade Federal de Viçosa, Viçosa.
- MATOS, A. R. et al. Alternative oxidase involvement in coldstress response of *Arabidopsis thaliana* fad2 and FAD3+ cell suspensions altered in membrane lipid composition. **Plant and Cell Physiology**, Kyoto, n. 48, p. 856–865, 2007.
- MITTLER, R. Abiotic stress, the field environment and stress combination. **Trends in Plant Science**, Oxford, n. 11, p. 15–19, 2006.
- MOSELEY, P. L. Heat shock proteins and heat adaptation of the whole organism. **Journal of Applied Physiology**, Bethesda, n. 83, p. 1413-1417, 1997.
- MURRE, C. et al. Structure and function of helix–loop–helix proteins. **Biochimica et Biophysica Acta**, Alberta, n. 1218, p. 129–135, 1994.
- NAKAMOTO, H.; VÍGH, L. The small heat shock proteins and their clients. **Cellular and Molecular Life Sciences**, Basel, n. 64, p. 294–306, 2007.
- NELSON, D. L.; COX, M. M. **Principles of biochemistry de Lehninger**. 5. ed. Porto Alegre: Artmed, 2010.
- NEILL, S. J. et al. Hydrogen peroxide and nitric oxide as signalling molecules in plants. **Journal of Experimental Botany**, Oxford, n. 53, p. 1237–1247, 2002.
- OLIVEIRA, S. V.; REYS M. A. Estruturação e consolidação da produção do biodiesel: base de soja: no Rio Grande do Sul. **Revista Extensão Rural**, Porto Alegre, ano 16, n. 17, p. 93-116, 2009.

OUYANG, Y.; CHEN, J.; XIE, W.; WANG, L.; ZHANG, Q. Comprehensive sequence and expression profile analysis of Hsp20 gene family in rice. **Plant Molecular Biology**, Dordrecht, n. 70, p. 341–357, 2009.

PANDEY, A. K. et al. Functional analysis of the Asian Soybean Rust Resistance Pathway Mediated by Rpp2. **Molecular Plant-Microbe Interactions**, Saint Paul, v. 24, n. 2, p.194–206, 2011.

PARKINSON, J. et al. L. A transcriptomic analysis of the phylum Nematoda. **Nature Genetics**, New York, v. 36, n. 12, p. 1259-1267, 2004.

PAZHAVARICAL, S. **Investigations on the early stages of interactions between the nematodes *Meloidogyne javanica* and *Pratylenchus thornei* and two of their plant hosts**. 2009. Thesis (PhD) University of Western Sydney, Sydney.

PETERSEN, T. N. et al. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature Methods**, New York, n. 8, p. 785-786, 2011.

POPP, M. P.; DILLON, C. R.; KEISLING, T. C. Economic and weather influences on soybean planting strategies on heavy soils. **Agricultural Systems**, Essex, v. 76, p. 969-984, 2003.

QUEIROZ, S. C. N.; NOGUEIRA, R. T.; SCRAMIN, S. **Importância dos Fitoestrógenos, presentes na soja, para a saúde humana**. Embrapa meio ambiente: Jaguariúna, 2006.

QUEIROZ, A. **Introdução a bioinformática**. Natal: UFRN/CB/DBF, 2002.

RAHMAN, S. A. S. A. et al. In planta PCR-based detection of early infection of plant-parasitic nematodes in the roots: a step towards the understanding of infection and plant defence. **European Journal of Plant Pathology**, Dordrecht, n. 128, p. 343–351, 2010.

RASCHKE, E.; BAUMANN, B.; SCHOFFL, F. Nucleotide sequence analysis of soybean small heat shock protein genes belonging to two different multigene families. **Journal of Molecular Biology**, London, n. 11, p. 549-557, 1988.

REHM, B. H. A. Bioinformatic tools for DNA/protein sequence analysis, functional assignment of genes and protein classification. **Applied Microbiology and Biotechnology**, Berlin, n. 57, p. 579–592, 2001.

RIZHSKY, L.; LIANG, H.; MITTLER, R. The combined effect of drought stress and heat shock on gene expression in tobacco. **Plant Physiology**, Minneapolis, n. 130, p. 1143–1151, 2002.

RODRÍGUEZ, M.; CANALES, E.; BORRÁS-HIDALGO, O. Molecular aspects of abiotic stress in plants. **Biología Aplicada**, La Habana, v. 22, n. 1, 2005.

SAITOU, N.; NEI, M. The Neighbor-joining Method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, Chicago, n. 4, v. 4, p. 406-425, 1987.

SANTOS, E. C. **Uma introdução à Bioinformática através da análise de algumas ferramentas de software livre ou de código aberto utilizadas para o estudo de alinhamento de sequências**. 2004. Monografia (Especialização em Administração em Redes Linux) - Universidade Federal de Lavras, Lavras.

SAIDI, Y. et al. Membrane lipid composition affects plant heat sensing and modulates Ca²⁺-dependent heat shock response. **Plant Signaling & Behaviour**, Perth, v. 5, n. 12, p. 1530-1533, 2010.

SAIDI, Y.; FINKA, A.; GOLOUBINOFF, P. Heat perception and signalling in plants: a tortuous path to thermotolerance. **New Phytologist**, Cambridge, n. 190, p. 556–565, 2011.

SANGWAN, V. et al. Opposite changes in membrane fluidity mimic cold and heat stress activation of distinct plant MAP kinase pathways. **Plant Journal**, Oxford, n. 31, p. 629–638, 2002.

SANTANA, H. **Identificação de raças do Nematóide de cisto da soja [*Heterodera glycines* (ichinohe)] a partir de populações de campo e isolados Monocísticos e resistência de cultivares comerciais à raça 3**. 2009. Dissertação (Mestrado em Agronomia) - Universidade Estadual do Oeste do Paraná, Marechal Cândido Rondon.

SARKAR, N. K.; KIM, Y-K.; GROVER, A. Rice sHsp genes: genomic organization and expression profiling under stress and development. **BMC Genomics**, Washington, v. 10, n. 393, p. 1471-2164, 2009.

SCHARF, K. D.; SIDDIQUE, M.; VIERTLING, E. The expanding family of Arabidopsis thaliana small heat stress proteins and a new family of proteins containing alpha-crystallin domains (Acd proteins). **Cell Stress and Chaperones**, Mansfield, v. 6, n. 3, p. 225-237, 2001.

SCHARF, K. D.; BERBERICH, T.; EBERSBERGER, I.; NOVER, L. The plant heat stress transcription factor (Hsf) family: Structure, function and evolution? **Biochimica et Biophysica Acta**, Alberta, v. 1819, p. 104–119, 2012.

SCHOLL, E. H.; BIRD, D. M. Resolving tylenchid evolutionary relationships through multiple gene analysis derived from EST data. **Molecular Phylogenetics and Evolution**, Orlando, v. 36, n. 3, p. 536-545, 2005.

SCHMUTZ, J. et al. Genome sequence of the palaeopolyploid soybean. **Nature**, London, v. 463, p. 178-183, 2010.

SHIRASU, K. The HSP90-SGT1 Chaperone Complex for NLR Immune Sensors. **Annual Review of Plant Biology**, Palo Alto, v. 60, p. 139-164, 2009.

SCHLUETER, J. A.; BRIAN, E.; SCHEFFLER, S. J.; SHOEMAKER, R. C. Fractionation of Synteny in a Genomic Region Containing Tandemly Duplicated Genes across Glycine max, Medicago. **Journal of Heredity**, Washington, v. 99, n. 4, p. 390–395, 2008.

- SEE, D. R.; BROOKS, S.; NELSON, J. C.; BROWN-GUEDIRA, G.; FRIEBE, B.; GIL, B. S. Gene evolution at the ends of wheat chromosomes. **Proceedings of the National Academy of Sciences**, Allahabad, v. 103, n. 11, p. 4162–4167, 2006.
- SEVERIN, A. J. et al. RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. **BMC Plant Biology**, London, v. 10, p. 160, 2010.
- SIDDIQUE, M. et al. The plant sHSP superfamily: five new members in *Arabidopsis thaliana* with unexpected properties. **Cell Stress and Chaperones**, Mansfield, v. 13, n. 2, p. 183-197, 2008.
- SINGH, K. B.; FOLEY, R. C.; OÑATE-SÁNCHEZ, L. Transcription factors in plant defense and stress responses. **Current Opinion in Plant Biology**, Saint Louis, v. 5, p. 430–436, 2002.
- STOLF-MOREIRA, R. et al. Identification of reference genes for expression analysis by real-time quantitative PCR in drought-stressed soybean. **Pesquisa Agropecuária Brasileira**, Rio de Janeiro, v. 46, n. 1, p. 58-65, 2011.
- SUN, W.; MONTAGU, M. V.; VERBRUGGEN, N. Review: small heat shock proteins and stress tolerance in plants. **Biochimica et Biophysica Acta**, Alberta, v. 1577, p. 1–9, 2002.
- SWINDELL, W. R.; HUEBNER, M.; WEBER, A. P. Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. **BMC Genomics**, Washington, v. 8; n. 125, London, 2007.
- TAIZ, L.; ZEIGER, E. **Fisiologia vegetal**. 4. ed. Porto Alegre: Artmed, 2009.
- TEREFE, M.; TEFERA, T.; SAKHUJA, P. K. Effect of a formulation of *Bacillus firmus* on root-knot nematode *Meloidogyne incognita* infestation and the growth of tomato plants in the greenhouse and nursery. **Journal of Invertebrate Pathology**, San Diego, v. 100, p. 94–99, 2009.
- THURAU, T.; KLFLE, S.; JUNG, C.; CAI, D. The promoter of the nematode resistance gene Hs1 pro-1 activates a nematode-responsive and feeding site-specific gene expression in sugar beet (*Beta vulgaris* L.) and *Arabidopsis thaliana*. **Plant Molecular Biology**, Dordrecht, v. 52, p. 643–660, 2003.
- TOROK, Z. et al. Synechocystis HSP17 is an amphitropic protein that stabilizes heat-stressed membranes and binds denatured proteins for subsequent chaperone-mediated refolding. **Proceedings of the National Academy of Sciences**, Allahabad, v. 98, p. 3098-3103, 2001.
- VAN OOIJEN, G. et al. The small heat shock protein 20 RSI2 interacts with and is required for stability and function of tomato resistance protein I-2. **The Plant Journal**, Oxford, v. 63, p. 563–572, 2010.

- VON KOSKULL-DÖRING, P.; SCHARF, K.; NOVER, L. The diversity of plant heat stress transcription factors. **TRENDS in Plant Science**, Oxford, v. 12, n. 10, p. 452-457, 2007.
- VAN MONTFORT, R. L. M.; SLINGSBY, C.; VIERLING, E. Structure and function of the small heat shock protein/a-crystallin family of molecular chaperones. **Advances in Protein Chemistry**, New York, v. 59, p. 105–156, 2001.
- XIE, L.; WANG, J.; BOURNE, P. E. In Silico Elucidation of the Molecular Mechanism Defining the Adverse Effect of Selective Estrogen Receptor Modulators. **PLoS Computational Biology**, New York, v. 3, n. 11, p. 2324-2332, 2007.
- XIONG, L.; SCHUMAKER, K. S.; ZHU, J. K. Cell Signaling during Cold, Drought, and Salt Stress. **The Plant Cell**, Rockville, p. 166-183, 2002.
- XIONG, J. **Essential Bioinformatics**, Texas A & M University, 2006 , p. 10-16.
- XUAN, Y. et al. Nitric oxide functions as a signal and acts upstream of AtCaM3 in thermotolerance in Arabidopsis seedlings. **Plant Physiology**. v.153, p. 1895–1906, 2010.
- WAHID, A.; GELANI, S.; ASHRAF, M.; FOOLAD, M. R. Heat tolerance in plants: an overview. **Environmental and Experimental Botany**, Elmsford, v. 61, p. 199–223, 2007.
- WANG, W. X.; VINOCUR, B.; SHOSEYOV, O.; ALTMAN, A. Biotechnology of plant osmotic stress tolerance: physiological and molecular considerations. **Acta Horticulturae**, The Hague, v. 560, p. 285–292, 2001.
- WANG, W.; VINOCUR, B.; ALTMAN, A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. **Planta**, Berlin, v. 218, p. 1–14, 2003.
- WATERS, E. R.; LEE, G. J.; VIERLING, E. Review article: evolution, structure and function of the small heat shock proteins in plants. **Journal of Experimental Botany**, Oxford, v. 47, p. 325-338, 1996.
- WATERS, E. R.; AEVERMANN, B. D.; SANDERS-REED, Z. Comparative analysis of the small heat shock proteins in three angiosperm genomes identifies new subfamilies and reveals diverse evolutionary patterns. **Cell Stress and Chaperones**, Mansfield, v. 13, n. 2, p. 127-142, 2008.
- WESTON, D. et al. Comparative physiology and transcriptional networks underlying the heat shock response in *Populus trichocarpa*, *Arabidopsis thaliana* and *Glycine max*. **Plant, Cell and Environment**, Malden, v.34, p.1488–1506, 2011.
- WILLIAMSON, V. M.; KUMAR, A. Nematode resistance in plants: the battle underground. **Trends in Genetics**, Cambridge, v. 22, n. 7, p. 396-403, 2006.
- WISTOW, G. Domain structure and evolution in a-crystallins and small heat-shock proteins. **FEBS**, Oxford, v. 181, n. 1, 1985.

WRATHER, J. A. et al. Soybean disease loss estimates for the top 10 soybean producing countries in 1994. **Plant Disease**, Saint Paul, v. 81, n. 1, p. 107-110, 1997.

WHITE, R. J. et al. The TBP-binding repressor Dr1 differentially regulates RNA polymerases I, II and III. **Science**, Washington, v. 266, p. 448-450, 1994.

YAMAGUCHI-SHINOZAKI, K.; SHINOZAKI, K. Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters DREB2A, involved in drought-responsive gene expression. **The Plant Cell**, Rockville, v. 18, n. 5, p. 1292-1309, 2005.

YAMASAKIA, K. et al. Solution structure of an Arabidopsis WRKY DNA binding domain. **The Plant Cell**, Rockville, v. 17, p. 944–956, 2005.

YORINORI, J. T. Risco de surgimento de novas doenças na cultura da soja. In: CONGRESSO DE TECNOLOGIA E COMPETIVIDADE NO MERCADO GLOBAL, 1., 2000, Cuiabá. **Anais...** Cuiabá: Fundação MT, 2000. p. 165-169.

YU, D.; CHEN, C.; CHEN, Z. Evidence for an important role of WRKY binding proteins in the regulation of NPR1 gene expression. **The Plant Cell**, Rockville, v. 13, p. 1527–1539, 2001.

ZHANG, W. et al. Molecular and Genetic Evidence for the Key Role of AtCaM3 in Heat-Shock Signal Transduction in Arabidopsis. **Plant Physiology**, Minneapolis, v. 149, p. 1773-1784, 2009.

ZHOU, Y. et al. NnHSP17.5, a cytosolic class II small heat shock protein gene from *Nelumbo nucifera*, contributes to seed germination vigor and seedling thermotolerance in transgenic Arabidopsis. **Plant Cell Reports**, Berlin, v. 30, 2011.

ZHU, B. et al. Identification and characterization of a novel heat shock transcription factor gene, GmHsfA1, in soybeans (*Glycine max*). **Journal of Plant Research**, Tokyo, v. 119, p. 247–256, 2006.

DADOS SUPLEMENTARES

Tabela 1S -Lista de oligonucleotídeos desenhados especificamente para cada modelo gênico candidato a Hsp20 de soja para uso nas análises de PCR qualitativo e RT- qPCR.

n ^o	Glymas	PRIMER_F	PRIMER_R	TM °C	Amplicon
1	Glyma01g26570	GAATCCGAAGCACAAACACA	GGGTCCATGATGTTTCTGTTG	60	124 bp
2	Glyma02g08400	CCCTTCATCACCATTCACTG	TGGAAGATCAGCACGGAAG	59	136 bp
3	Glyma02g41150	ATGGCTGGCACTATCTTTGG	CTGTTGAGCCGGACCATTTC	60	85 bp
4	Glyma02g42000	AACGTCTTCGACCCAGTTTC	TGGCTATGGCACTACTTTTCG	59	95 bp
7	Glyma04g05720	CTCAGCACAGAACGAAGCTG	GATTCGGTGCAGAGTGTGA	60	138 bp
9	Glyma04g38530	CAACATTTGCTGCTATCCATTTC	AACATTGCTCCTTCGACCAC	60	114pb
10	Glyma04g40790	CCCAATGGCTCAAGCACTAT	AGCTCTCACCTTCCTAGCC	60	148 bp
13	Glyma06g05740	TACCGGCACACGATAACAAG	GAATGGAGAGTCCAAATCCATC	60	71pb
14	Glyma06g14000	ATGGCCCAAGCACTATCAAC	TTGGAAACTCCTTCCTCGAC	60	124 bp
15	Glyma06g16490	GCAGAGTTCAAAACACCAACC	AGCGAGAAAGGGTCGAAAAC	60	104 bp
16	Glyma07g03450	GCTTCCTCTTCCTTGATTGG	ACAGCAACTTCAGTGCCTACC	59	88pb
17	Glyma07g04860	GTGGGTTGCATTGTTGTGAG	CATGAGCACGAGAAGAAGGA	60	138 bp
19	Glyma07g32030	GCCGTAGGAGCAATGTGTTT	CCAATCCACACGAGTGCTTA	60	122 bp
20	Glyma07g32050	GTTTCGATCCTTTCTCACTGGAC	TTCAGGGAAGGAAGCTGAAG	60	85 bp
21	Glyma07g32070	GGAAGAGGTTAAGAAGCCTGATG	TTGGAACACAAGACCGAGAG	60	127pb
24	Glyma07g32110	AGTTTCTTCAGTGGCCCAAG	ACCCGAGTGCTCACAAATG	60	128 bp
25	Glyma08g07330	CACCCAATTTGCAGAGTTCC	TTAAAGGGCTCCCACACATC	60	137 bp
26	Glyma08g07340	GAGGATCATTGTATGAGTCGTG	CCAAACACACAATCTTTCAACC	59	150pb
27	Glyma08g07350	CAGAAAGAAGTGAAGGCATCG	CCACCGAAGAACTTGGAAAT	60	125 bp
28	Glyma08g22630	TTCCTTTCAACCATCAGCAA	GGAAGCACACACCCTTTGTT	60	133 bp
30	Glyma08g43120	TGCCCGCACTTCATCTAGTA	TGCACTTGCTGCTTTGACTT	60	147 bp
34	Glyma10g32000	AGAAATGAGGCTGCCACAAC	TGGAGCAGAGAACCATTTG	61	85 bp
36	Glyma11g14250	TCTCACCTTCGAGAACATCG	TCACCTGGTTGATTGGAAC	60	126pb
38	Glyma11g37450	CCACATCGAGGTATGTTTCGTC	GGTGCTTGTGTGTTTAGGC	59	117 bp
39	Glyma12g01580	GTCCTCTCTCATTGCGAAGC	TGGCGTTGGTGTGAAAG	60	104pb
40	Glyma12g06210	TGTCTCTTCCAAGTTGTTCACTC	AACACAGTTCATCGCACAC	59	118pb
44	Glyma13g24440	CATGGTGTGGAGAATTCCAG	GGATACGCGAGGTATTTAAAGG	59	133pb
45	Glyma13g24460	TTCCTGAAAATCGTGAGAGC	CCTTTGTCTTGTGTTCCAATG	58	71 pb
46	Glyma13g24480	CGTCGTATCAAACGAAGAA	CCGAAGAAACCTGGAATCAG	60	115 bp
47	Glyma13g24490	CCCAAACACTCTTTCTTCAGC	GCCTTGAATACGTGTGCTTC	59	111 bp
48	Glyma13g24510	GCAATGTGTTTCGATCCCTTC	GAATATCGGCCTTGAACACG	60	150 bp
50	Glyma13g36220	GCCAAAGGAATTTGTTCTGTG	GAGCAATTGTTAGGCTCTGATG	59	96pb
52	Glyma14g04970	GAAGTGCCACGCTCATTGTA	CGATCTGTGCTTACCATTG	60	105 bp
53	Glyma14g06900	AACGTCTTCGACCCATTCTC	TTGGCTATGGCACTGCTTTC	60	107 bp
54	Glyma14g06910	ACCCATTCCATGGCTTCC	TCTCCTTCCAGTCGATACGC	60.5	93pb
55	Glyma14g11420	ATTCAGGGTGATGGGTTTG	CGTGTGCGAGCACTGTATGT	60	109 bp
56	Glyma14g11430	CTTTCGGCAACCTTCAATTT	TGGGAGCACTGTACGCTTTG	60	140 bp
57	Glyma14g39560	ATGGCTGACGCTATCTTTGG	GCTATTGAGCCAGACCATTCTC	60	86pb
62	Glyma15g41880	ACCATTTCCGAGCCTTACTC	TCACATCCTCTTTGGCGAAC	60	135 bp
63	Glyma16g01440	GAAGTTGCCATTGCTTGAGAG	GCACCTTACATTCCCACCTATTC	59	102pb
64	Glyma16g04270	ACCCTTGAACCTTCACTTCC	CGTTTCCATCTATTTCTTTCG	58.5	116pb
65	Glyma16g33130	GGCATATCCGCACATACACA	TTGCTTTGGATGGGAAACC	60	101 bp
67	Glyma17g34220	TCGCAAAGACATGGATTTCA	CCATGTTGGAGCATTGTGTG	60	124 bp
71	Glyma18g41650	AACTCAACCCGACACACACC	AGAGGGTGGTGTGTTCTCT	60	140 bp
73	Glyma19g01440	CAGAATGGCTCGCACAATTA	CATGTCCCAGAGGGTGTCTAC	60	115 bp
74	Glyma20g01930	GCACACGATAACGAGAGAAGAG	GCAGAGTGTGAAGAATGGAGAG	59.5	82pb
76	Glyma20g35650	CCATCATTGTCTGCCACATC	AGAGACCCGTTTGTCTTGG	60	102 bp

Figura 1S -Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e temperatura de pareamento) dos *primers* de detecção de infecção de plantas por *M. javanica*. Para o teste de especificidade foram utilizados DNA genômico de soja e de *M. javanica*. O tamanho do fragmento esperado é de 945 pb. Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para os *primers*, e na segunda linha as amostras com o controle negativo da reação (NO). Foi utilizado o marcador molecular de 1 kb *plus* (*Ladder de 1 kb plus – life technologies*). b. *Ladder de 1 kb plus – life technologies*. Na horizontal estão indicadas os tamanhos em pares de base das principais bandas do marcador.

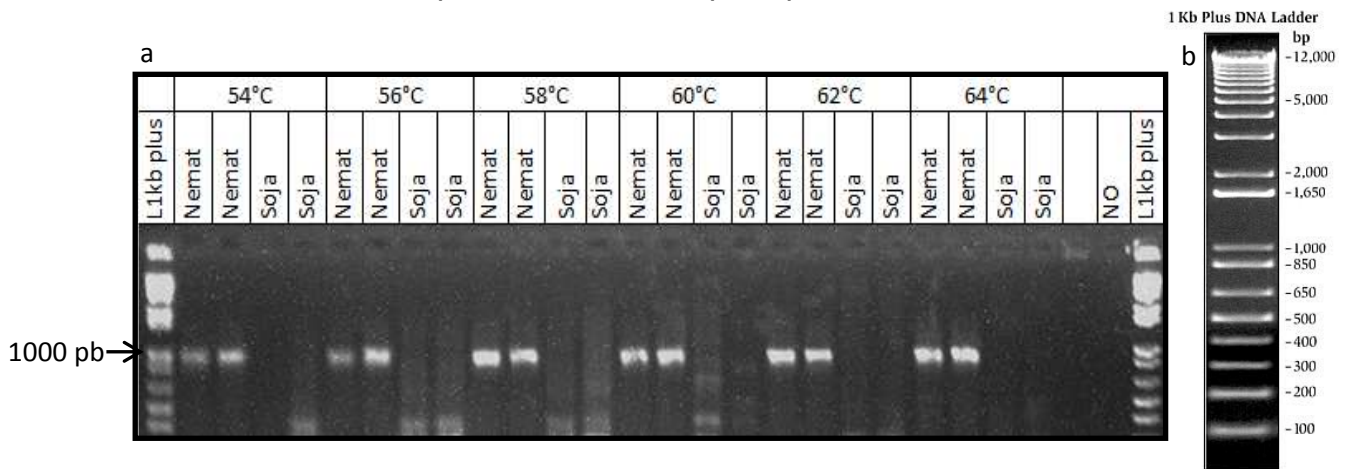


Figura 2S -Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR, com 945 pb, para confirmação da infecção das plantas de soja dos tratamentos bióticos pelo nematóide *M. javanica*. Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para os *primers*, e na segunda linha as amostras com o controle negativo da reação (NO). Foi utilizado o marcador molecular de 1 kb *plus* (*Ladder de 1 kb plus – life technologies*). b. *Ladder de 1 kb plus – life technologies*. Na horizontal estão indicadas os tamanhos em pares de base das principais bandas do marcador.

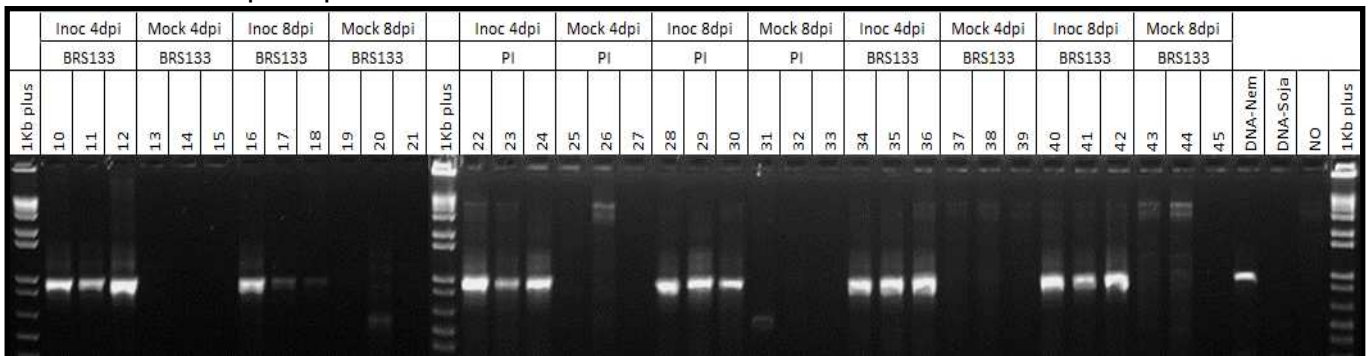


Figura 3S -Eletroforese em gel de agarose 1% para a análise de integridade das amostras de RNA total extraídas dos tratamentos de estresse biótico e abiótico. Na primeira foto (superior) estão os RNAs das amostras de BRS133 de 1 a 9, estresse abiótico, e de 10 a 21, estresse biótico (infecção por *M. javanica*). Na segunda foto (abaixo) estão as amostras de RNA somente do experimento de estresses por infecção por *M.javanica*. As amostras de 22 a 33 são oriundas da cultivar PI595099 e de 34 a 45 de BRS133 (experimento de validação). Para avaliar a integridade do RNA a banda 28S deve ser quase duas vezes mais intensa que a banda 18S as duas bandas em boa intensidade e definição (NARDINELLI, 2008).

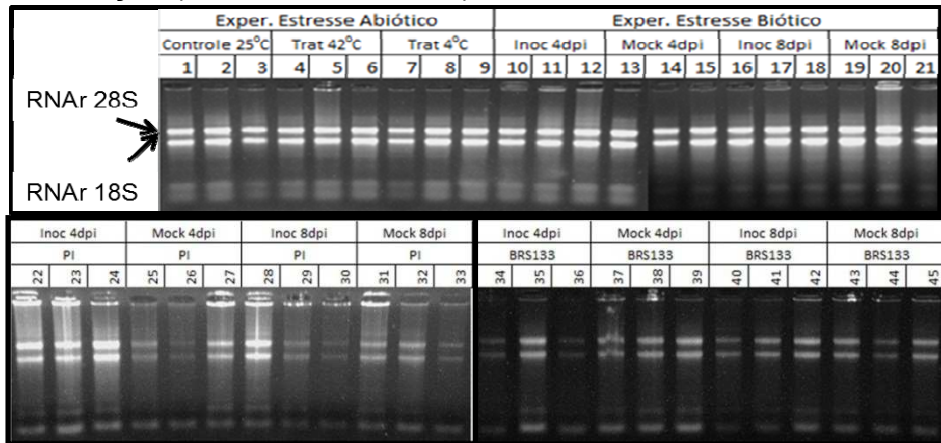


Figura 4S -Eletroforese em gel de agarose 1,2% dos produtos da reação de PCR para a confirmação de eliminação de DNA genômico da amostra após tratamento com DNase I. O tamanho do fragmento esperado é de 520 pb quando template for DNA genômico e 440 pb quando template for RNA. As amostras de 1 a 9 e 10 a 21 são dos tratamentos de estresses abióticos e bióticos, respectivamente. Foram utilizados o controle de DNA genômico de soja e de reação (NO).

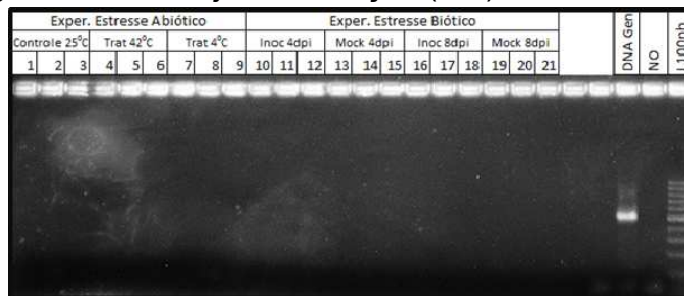


Figura 5S - Eletroforese em gel de agarose 1,2% dos produtos da reação de PCR com o primer para o gene β -actina para a análise da qualidade do cDNA sintetizado. O tamanho do fragmento esperado é de 520 pb quando template for DNA genômico e 440 pb quando template for RNA.

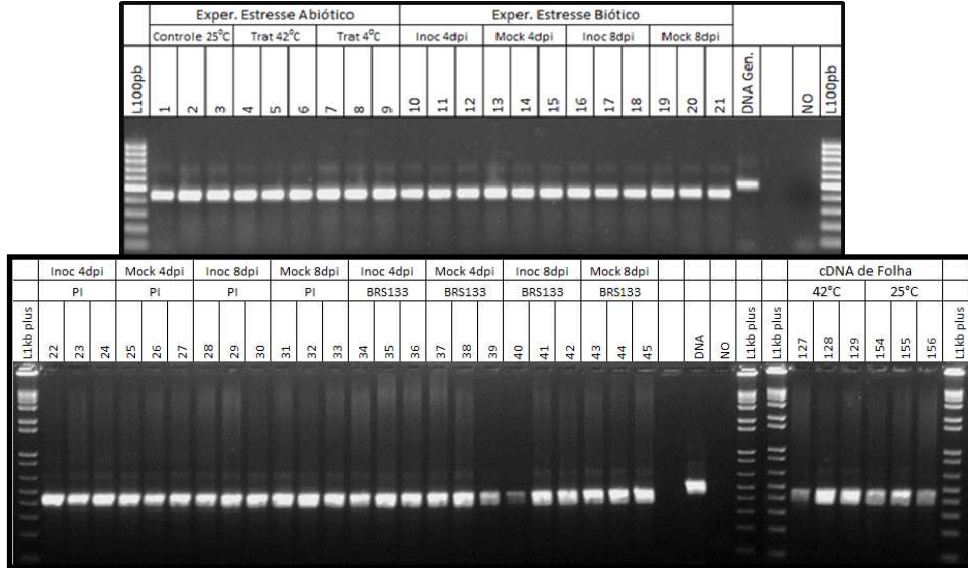


Figura 6S - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e temperatura de pareamento) dos *primers* em DNA genômico de soja. Nesta figura estão os resultados dos testes com os oligonucleotídeos para os modelos gênicos V74 (glyma20g01930), V48 (glyma13g24510), V52 (glyma14g04970), V64 (glyma16g04270), V63 (glyma16g01440) e V53 (glyma14g06900). Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para cada par de *primers*, e na segunda linha os *primers* testados seguidos dos controles negativos de reação (NO). Foi utilizado o marcador molecular de 50 pb (*Ladder* 50 pb – *life Technologies*).

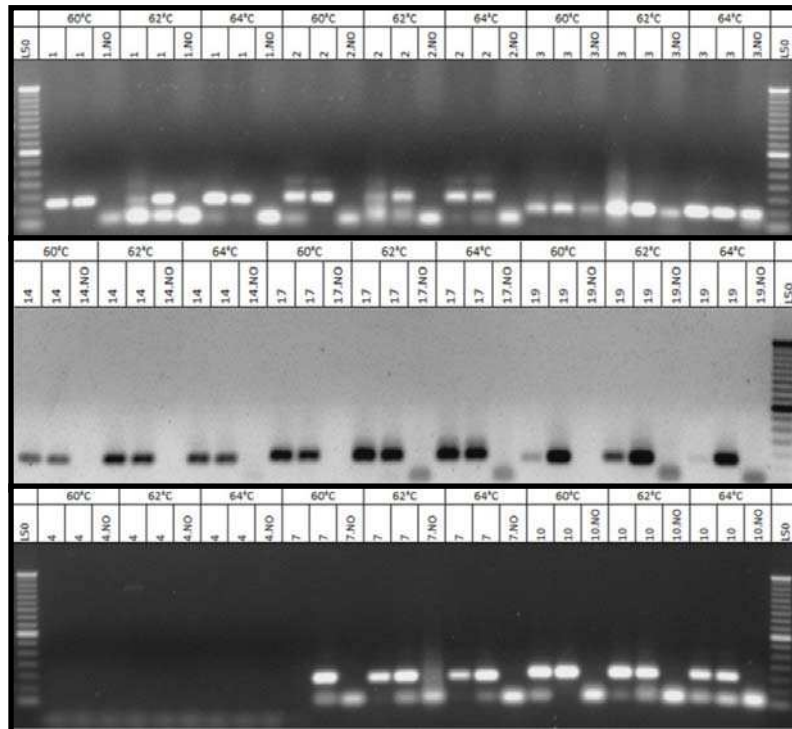


Figura 6S – (Continuação)

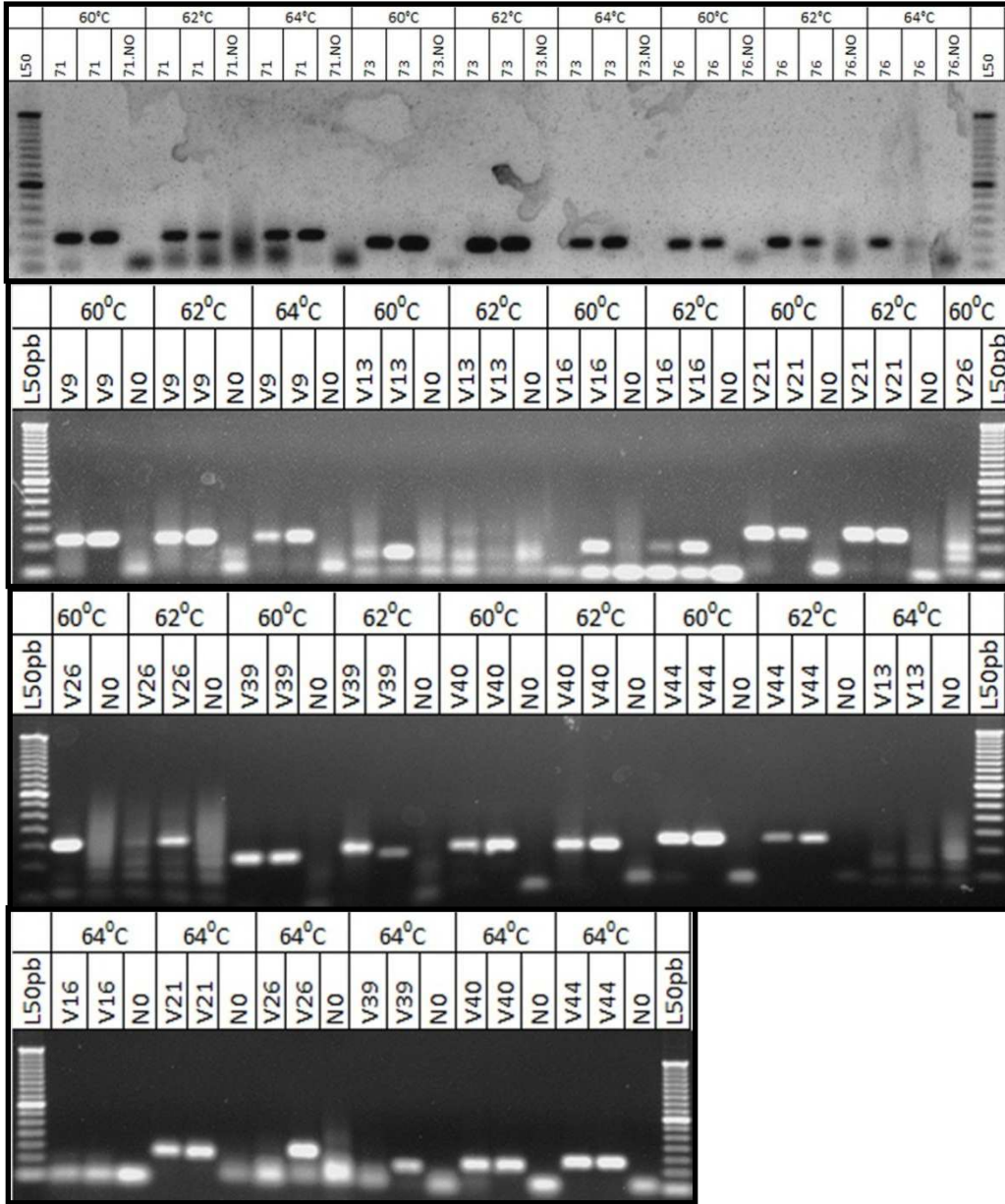


Figura 7S - Eletroforese em gel de agarose 1,2% dos produtos de reação de PCR para testes de amplificação (especificidade e temperatura de pareamento) dos *primers* em DNA genômico de soja. Nesta figura estão os resultados dos testes com os oligonucleotídeos para os modelos gênicos V74 (glyma20g01930), V48 (glyma13g24510), V52 (glyma14g04970), V64 (glyma16g04270), V63 (glyma16g01440) e V53 (glyma14g06900). Na primeira linha do cabeçalho de cada foto estão anotadas as temperaturas testadas para cada par de *primers*, e na segunda linha os *primers* testados seguidos dos controles negativos de reação (NO). Foi utilizado o marcador molecular de 50 pb (*Ladder 50 pb – life Technologies*).

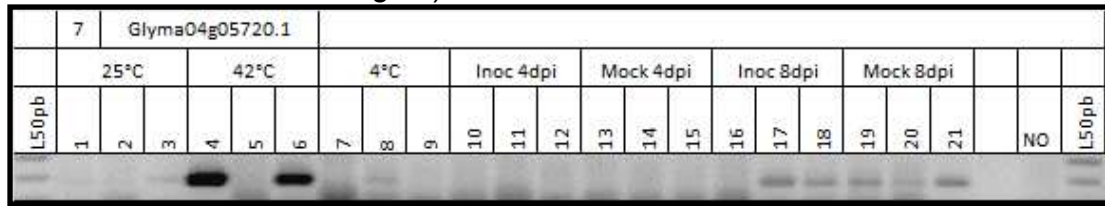


Figura 7S - (Continuação)

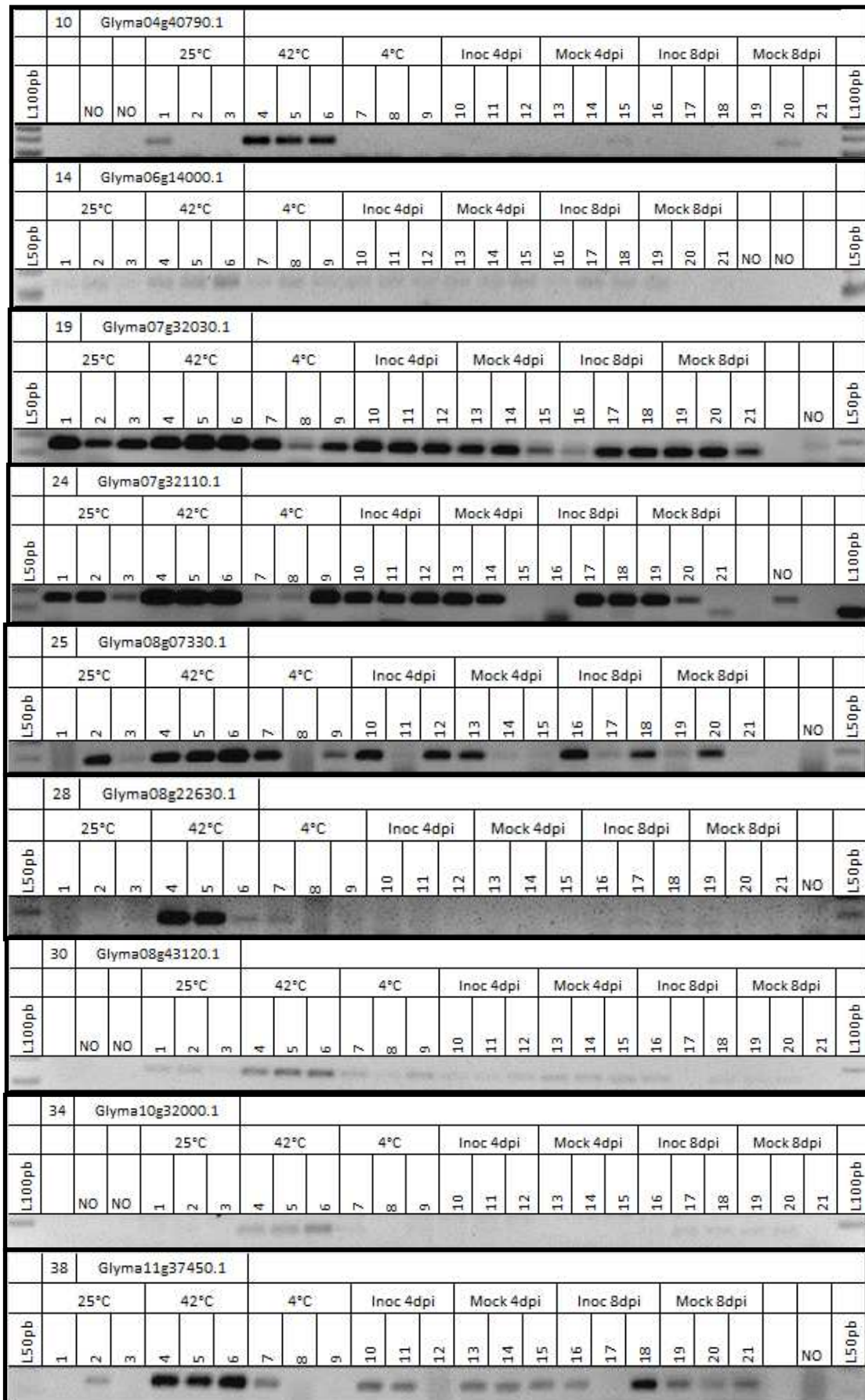


Figura 7S – (Continuação)

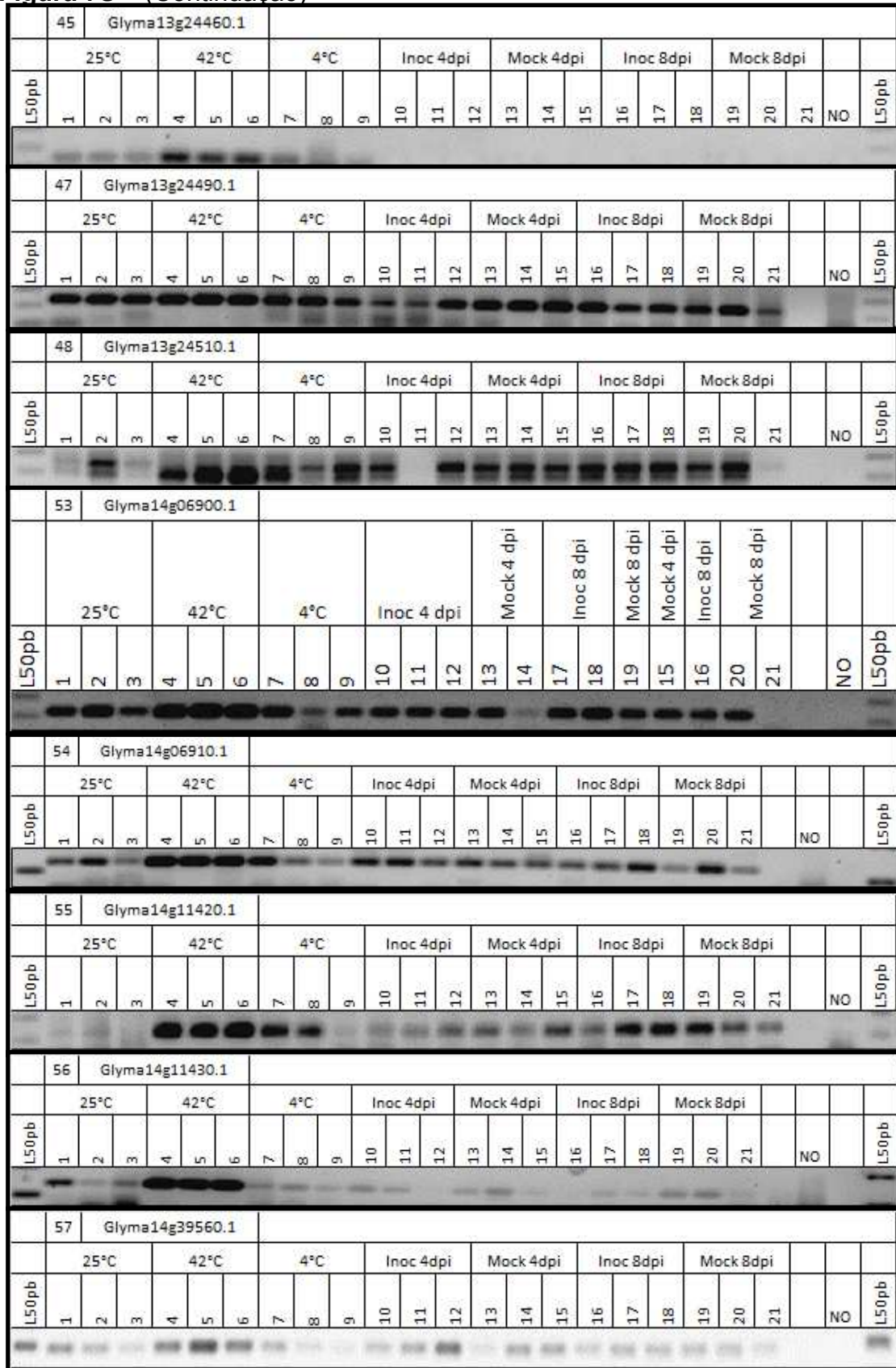


Figura 7S – (Continuação)

