

ANÁLISE TRANSCRIPTÔMICA DE EIXOS EMBRIONÁRIOS DE SOJA  
(*GLYCINE MAX*) DURANTE A GERMINAÇÃO

**DANIEL BELLIENY RABELO**

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE  
DARCY RIBEIRO – UENF

CAMPOS DOS GOYTACAZES – RJ  
OUTUBRO 2016

ANÁLISE TRANSCRIPTÔMICA DE EIXOS EMBRIONÁRIOS DE SOJA  
(*GLYCINE MAX*) DURANTE A GERMINAÇÃO

**DANIEL BELLIENY RABELO**

Tese apresentada ao Centro de Biociências e Biotecnologia, da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para obtenção do título de Doutor em Biociências e Biotecnologia

Orientador: Dr. Thiago Motta Venancio

CAMPOS DOS GOYTACAZES – RJ  
OUTUBRO 2016

## AGRADECIMENTOS

A Regina Bellieny, que fez com que toda a trajetória fosse possível. Os ensinamentos de caráter, responsabilidade e disciplina jamais serão esquecidos. A ela este trabalho é dedicado, bem como todas as vitórias conquistadas e vindouras.

Ao professor Thiago Venancio, pela paciência, pelos ensinamentos e pelo zelo. Seu empenho e determinação em entregar sempre o melhor trabalho possível faz com que todos à sua volta busquem constantemente o aperfeiçoamento.

À professora Elenir e os colaboradores de seu laboratório: Eduardo Oliveira e Elane Ribeiro pelo apoio na etapa experimental, sem a qual este projeto não seria possível.

Aos colegas de laboratório: Lupis, Filipe, Rajesh, Kanhu, Gustavo e Hemmanoel pela contribuição com idéias e discussões. Uma equipe harmoniosa é essencial para o desenvolvimento individual. E o ambiente aqui criado é capaz de fazer emergir o melhor de cada um.

À professora Olga Machado, e as secretárias Beatriz Almeida e Marlene dos Santos pelo apoio incansável a todos os estudantes da Pós-Graduação em Biotecnologia e Biotecnologia.

Aos financiadores FAPERJ, CAPES e CNPq pelo suporte ao projeto.

A Anne Braga, pela amizade, pelo apoio e dedicação que foram fundamentais em todos os momentos.

Aos padrinhos Carlos Bellieny e Fátima Bellieny, que acreditaram sempre e ajudaram de forma decisiva na minha formação.

## SUMÁRIO

AGRADECIMENTOS .....	i
ÍNDICE DE ABREVIACÕES.....	iv
ÍNDICE DE FIGURAS E TABELAS.....	vi
RESUMO .....	x
ABSTRACT .....	xi
<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 Origem, classificação e história da domesticação da soja.....	1
1.2 A soja na agricultura atual .....	2
1.3 Germinação em soja.....	3
1.4 O processo de nodulação .....	5
1.5 Abordagens genômicas no estudo da soja.....	6
1.5.1 Sintenia, duplicação e evolução de genomas de plantas.....	7
1.6 Desenvolvimento embrionário e regulação hormonal da germinação .....	9
1.7 Fatores de transcrição (TFs).....	12
1.7.1 Principais famílias de fatores de transcrição em plantas.....	17
<b>2 JUSTIFICATIVA .....</b>	<b>18</b>
<b>3 OBJETIVOS .....</b>	<b>19</b>
3.1 Objetivos gerais.....	19
3.2 Objetivos específicos .....	19
<b>4 METODOLOGIA.....</b>	<b>20</b>
4.1 Material vegetal, germinação e purificação de RNA .....	20
4.2 Mapeamento dos fragmentos ao genoma de referência .....	21
4.3 Análise de expressão gênica diferencial .....	22
4.4 Análise de ortologia e regiões de sintenia .....	23
4.5 Análise de fatores de transcrição e enriquecimento de termos de Gene Ontology .....	23
<b>5 RESULTADOS E DISCUSSÃO .....</b>	<b>24</b>
5.1 Análise geral do transcriptoma de eixos embrionários de soja durante a germinação .....	24

<b>5.2</b>	<b>Identificação de genes preferencialmente expressos e análise da transição do perfil transcricional em 24 HAI.....</b>	<b>41</b>
<b>5.3</b>	<b>Genes envolvidos com homeostase da parede celular, proteínas quinases e fatores de transcrição são regulados positivamente durante a germinação .....</b>	<b>44</b>
<b>5.4</b>	<b>Ativação transcricional de genes do metabolismo primário durante a germinação .....</b>	<b>51</b>
<b>5.4.1</b>	<b>Metabolismo de carboidratos .....</b>	<b>51</b>
<b>5.4.2</b>	<b>Metabolismo de lipídios.....</b>	<b>54</b>
<b>5.4.3</b>	<b>Metabolismo de aminoácidos .....</b>	<b>55</b>
<b>5.5</b>	<b>Biossíntese e sinalização de fitormônios durante a germinação.....</b>	<b>62</b>
<b>6</b>	<b>CONCLUSÕES .....</b>	<b>66</b>
<b>7</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>67</b>

## ÍNDICE DE ABREVIACÕES

Tg – Teragrama

FPKM – fragmentos por kilobase milhão

TF – *transcription factor*

WGD – *whole genome duplication*

WGT – *whole genome triplication*

Mya – *million years ago*

ABA – *abscisic acid*

GA – *gibberellic acid*

NCED – *9-cis-epoxycarotenoid dioxygenase*

GID1 – *gibberellin insensitive dwarf 1*

HUB1 – *histone monoubiquitination 1*

PIF – *phytochrome interacting factor*

BR – *brassinosteroids*

ETR – *ethylene response*

EIN – *ethylene insensitive*

ERS – *ethylene response sensor*

CTR – *constitutive triple response*

DET – *de-etiolated*

BRI1 – *brassinosteroid insensitive 1*

MYB – *myeloblastosis*

AP2 – *apetala*

DOF – *DNA-binding with one finger*

EREBP – *ethylene responsive element binding protein*

ARF – *auxin response factor*

IAA – *indole-3-acetic acid*

bZIP – *basic region leucine zipper*

HAI – *hours after imbibition*

GO – *gene ontology*

TMK1 – *transmembrane kinase 1*  
ACO – *aminocyclopropane-1-carboxylate oxidase*  
CYP – *cytochrome P450*  
LBD – *LOB domain*  
RFO – *raffinose family oligosaccharide*  
PFK – *phosphofructokinase*  
GAPDH – *glyceraldehyde-3-phosphate dehydrogenase*  
TIM/TPI – *triose-phosphate isomerase*  
G6PDH – *glucose-6-phosphate dehydrogenase*  
FBPase – *fructose 1,6-bisphosphatase*  
CSY – *citrate synthase*  
IDH – *isocitrate dehydrogenase*  
MDH – *malate dehydrogenase*  
ADH – *alcohol dehydrogenase*  
LDH – *lactate dehydrogenase*  
TAG – *triacylglycerols*  
ICL – *isocitrate lyase*  
MSY – *malate synthase*  
PEPCK – *phosphoenolpyruvate carboxykinase*  
AlaAT – *alanine aminotransferase*  
AspAT – *aspartate aminotransferase*  
GABA – *gamma-aminobutyric acid*  
GAD1 – *glutamate decarboxylase*  
DAHP – *3-deoxy-D-arabino-hepturosonate-7-phosphate*

## ÍNDICE DE FIGURAS E TABELAS

<b>Figura 1</b> <i>Árvore filogenética incluindo gêneros de monocotiledôneas e dicotiledôneas. Círculos azuis representam eventos de duplicação. O círculo preto representa um evento de triplicação de genoma. (Adaptado de Van de Peer (2011))</i> .....	9
<b>Figura 2</b> <i>Ilustração esquemática do funcionamento da ativação diferencial de vários genes (A-E) por um fator de transcrição genérico codificado pelo gene X. (<a href="http://scienceblogs.com/pharyngula/2006/07/09/transcription-factors-and-morp/">http://scienceblogs.com/pharyngula/2006/07/09/transcription-factors-and-morp/</a>)</i> .....	14
<b>Figura 3</b> <i>Densidade da expressão gênica, medida em FPKM, para cada um dos estágios da germinação: dry (eixo seco), 3HAI (embrião 3 horas após a embebição), 6HAI (embrião 6 horas após a embebição), 12HAI (embrião 12 horas após a embebição), 24HAI (embrião 24 horas após a embebição).</i> ....	26
<b>Figura 4</b> <i>A similaridade/divergência entre o perfil de expressão gênica das amostras sequenciadas é representada pela distância JS (Jensen-Shannon). Maiores valores representam maior divergência entre os perfis das amostras.</i> ....	27
<b>Figura 5</b> <i>A similaridade/divergência entre o perfil de expressão gênica das amostras sequenciadas é representada pela distância JS (Jensen-Shannon). Maiores valores representam maior divergência entre os perfis das amostras.</i> ....	28
<b>Figura 6</b> <i>Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x, e semente 3 horas após embebição (3HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica</i> .....	29
<b>Figura 7</b> <i>Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x e semente 6 horas após embebição (6HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.</i> .....	30
<b>Figura 8</b> <i>Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x, e semente 12 horas após embebição (12HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.</i> .....	31
<b>Figura 9</b> <i>Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas</i>	



amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica. ....32

**Figura 10** Comparação da expressão gênica entre as amostras de semente 3 horas após germinação (3HAI) no eixo x e semente 6 horas após embebição (6HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica. ....33

**Figura 11** Comparação da expressão gênica entre as amostras de semente 3 horas após embebição (3HAI) no eixo x e semente 12 horas após embebição (12HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica. ....34

**Figura 12** Comparação da expressão gênica entre as amostras de semente 3 horas após embebição (3HAI) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica. ....35

**Figura 13** Comparação da expressão gênica entre as amostras de semente 6 horas após embebição (6HAI) no eixo x e semente 12 horas após embebição (12HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica. ....36

**Figura 14** Comparação da expressão gênica entre as amostras de semente 6 horas após embebição (6HAI) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica. ....37

**Figura 15** Comparação da expressão gênica entre as amostras de semente 12 horas após embebição (12HAI) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica. ....38

**Figura 16** Padrões globais de expressão gênica nos eixos embrionários de soja durante a germinação (embrião seca, 3HAI, 6HAI, 12HAI e 24HAI). Os padrões de expressão foram agrupados por HCL (hierarchical clustering). Apenas genes expressos (FPKM >1) estão representados. Os valores de expressão foram convertidos em z-score, e as intensidades de expressão variam entre -2 e 2. (HAI: Hours after imbibition) .....40

**Figura 17** (A) Imagens do material biológico utilizado nos sequenciamentos. Da esquerda para a direita: embrião seco, 3HAI, 6HAI, 12HAI e 24HAI (B) Log<sub>2</sub> da contagem total de genes preferencialmente expressos no contexto da germinação em cada tempo (dry, 3HAI, 6HAI, 12HAI, 24HAI). Acima das barras, gráficos-donut mostram os valores absolutos das contagens de genes preferencialmente expressos separados por intervalo de intensidade de expressão (<5; <20; <100; >100). (C) Para cada amostra, peso seco e peso fresco dos eixos embrionários foram medidos. Os gráficos de caixa denotam a variação encontrada em 20 indivíduos organizados em duas raias: a primeira (esquerda) sendo peso fresco, e a segunda (direita) o peso seco. (D) As barras indicam as contagens de genes regulados positivamente em progressão nos intervalos do experimento. ....43

**Figura 18** Mapa de calor de todos os genes associados com vias do metabolismo primário (baseado em anotações do banco de dados KEGG) expressos nos eixos de sementes de soja durante a germinação. As colunas representam as amostras analisadas no estudo, da esquerda para a direita (eixo seco, 3HAI, 6HAI, 12HAI e 24HAI). Os valores de expressão em FPKM foram convertidos em z-score e a intensidade de cor varia de -2 a 2.....57

**Figura 19** Principais vias metabólicas e enzimas discutidas no estudo. Níveis transcricionais de genes-chave estão representados nos mapas de calor. As famílias representadas estão dispostas numericamente seguindo a ordem: 1) Invertase; 2) Sacarose sintase (SuSy); 3) Celulose sintase; 4) Hexoquinase; 5) glicose-6-fosfato desidrogenase; 6) fosfofrutoquinase (PFK); 7) Frutose 1,6-bifosfato (FBPase); 8) Difosfato-frutose-6-fosfato 1-fosfofrutoquinase (PFP); 9) Piruvato quinase (PK); 10) Piruvato decarboxilase; 11) Álcool desidrogenase (ADH); 12) Lactato desidrogenase (LDH); 13) Piruvato desidrogenase (PDH-E2); 14) Citrato sintase (CSY); 15) Isocitrato desidrogenase (IDH); 16) Isocitrato liase (ICL); 17) Malato sintase (MSY); 18) malato desidrogenase (MDH); 19) Fosfoenolpiruvato carboxiquinase (PEPCK); 20) Triacilglicerol (TAG) lipase; 21) Glicerol quinase; 22) Glicerol-3-fosfato desidrogenase (GPDHC1); 23) COMATOSE (CTS); 24) Glutamato decarboxilase (GAD); 25) Gamma-amino-N-butirato transaminase (GABA-T). As 5 colunas em cada mapa de calor

representam os tempos analisados no experimento. Da esquerda para a direita: eixo seco, 3HAI, 6HAI, 12HAI e 24HAI. (HAI – hours after imbibition).....58

**Figura 20** Últimas etapas da biossíntese e principais vias de sinalização de hormônios promotores da germinação: Giberelina (A) e etileno (B). Representações esquemáticas seguem o modelo KEGG ([http://www.genome.jp/kegg/document/help\\_pathway.html](http://www.genome.jp/kegg/document/help_pathway.html)), exceto pelo padrão de cores. Formas elípticas representam famílias gênicas. Retângulos verdes representam vias conectadas. Mapas de calor representam os níveis transcricionais (FPKM) dos genes expressos (FPKM >1) pertencentes às famílias mostradas nas vias. As cinco colunas nos mapas de calor representam as amostras analisadas, da esquerda para a direita: eixo seco, 3HAI, 6HAI, 12HAI e 24HAI. (HAI – hours after imbibition).....65

**Tabela 1** Média do número fatores de transcrição presentes em genomas vegetais em comparação com a média total de genes de cada grupo (Adaptado de Jin, Zhang et al. (2014). .....15

**Tabela 2** Distribuição de reads mapeados no genoma de referência da Soja: Estatísticas gerais do sequenciamento da germinação de eixos embrionários da soja. Cada linha da coluna representa um ponto na escala temporal da germinação, iniciando em 0 horas (semente seca) até 24 horas após embebição em água. ....39

**Tabela 3** Enriquecimento de Termos GO (Gene Ontology) entre os genes diferencialmente expressos durante a germinação. ....48

**Tabela 4** Níveis transcricionais (em FPKM) de enzimas representadas na figura 20.....60

## RESUMO

A soja (*Glycine max*) é uma das mais importantes culturas de plantio na agricultura mundial, e constitui uma fonte crucial de proteína e óleo na alimentação humana e animal. Nesse contexto, a germinação é considerada uma etapa de crucial interesse sob a perspectiva agrônômica. No entanto, padrões transcricionais globais subjacentes às transições morfológicas e fisiológicas no eixo embrionário da soja durante a germinação permanecem inexplorados. No presente estudo, reportamos a análise de ~1,58 bilhões de *reads* obtidos por metodologia RNASeq de eixos embrionários de soja em 5 estágios da germinação (eixo seco, 3, 6, 12 e 24 horas após embebição em água). Os resultados mostram uma transição considerável no perfil transcricional iniciada após 12 horas de embebição, na qual são observados 730 genes preferencialmente expressos (99,6% do total de genes preferencialmente expressos em todas as 5 condições analisadas). Os dados corroboram a ativação de processos críticos para a germinação, como glicólise, ciclo do ácido cítrico e remodelação da parede celular. Após 3 horas de embebição ocorre a regulação positiva de quinases e 81 fatores de transcrição, particularmente da família LOB, sugerindo que regulação transcricional e pós-transcricional desempenham papel central logo após a embebição. Mobilização de lipídios e ciclo do glioxilato encontram-se também transcricionalmente ativos nos eixos embrionários, sugerindo que o catabolismo local de reservas lipídicas contribuem com a produção de energia no órgão durante a germinação. Os dados apresentam também evidências que suportam a inativação do ácido abscísico e a regulação positiva de vias relacionadas com hormônios promotores da germinação, como: giberelina, etileno e brassinosteróides.

## ABSTRACT

Soybean (*Glycine max*) is a major legume crop worldwide, providing a critical source of protein and oil. In this context, germination is considered to be a step of critical relevance by the agronomic perspective. Nevertheless, the global transcriptional patterns underlying the morphological and physiological changes of soybean embryonic axes during germination remain unknown. Here we report the analysis of ~1.58 billion RNA-Seq reads from soybean embryonic axes at five germination stages. Our results show a remarkable transition in the overall transcription pattern, which starts after 12 hours after imbibition featuring 730 preferentially expressed genes (99,6% of all preferentially expressed genes in all analyzed conditions). Also strongly supports the early activation of processes that are critical for germination, such as glycolysis, Krebs cycle and cell wall remodeling. Strikingly, only 3 hours after imbibition there is a preferential up-regulation of protein kinases and transcription factors, particularly from the LOB domain family, implying that transcriptional and post-transcriptional regulation play major roles early after imbibition. Lipid mobilization and glyoxylate pathways are also transcriptionally active in the embryonic axes, strongly suggesting that the local catabolism of oil reserves in the embryonic axes contributes to energy production during germination. We also present evidence supporting abscisic acid inactivation and the up-regulation of pathways related with the germination promoting hormones gibberellin, ethylene and brassinosteroids.

# 1 INTRODUÇÃO

## 1.1 Origem, classificação e história da domesticação da soja

O início da domesticação da soja é frequentemente associado a um intervalo de tempo ocorrido há aproximadamente 5000-9000 anos em diversos estudos (McKie e Anderson 1967; Carter, Nelson et al. 2004; Qi, Li et al. 2014). Contudo, sabe-se há mais de 40 anos que nenhum traço de espécies leguminosas foram encontrados em sítios arqueológicos do período Neolítico (que comporta o intervalo de tempo citado anteriormente) no norte da China, onde acredita-se ter se iniciado o cultivo desta espécie (Ho 1969). As informações biogeográficas e históricas conhecidas suportam que a soja tenha sido domesticada inicialmente no território que atualmente corresponde à China, em um período próximo ao século XVII a.C. durante a dinastia Shang (Ho 1969). Mais adiante, entre os séculos XI e VII a.C., como resultado da expansão da dinastia Zhou, a cultura da soja alcançou a Coreia, Japão e países do sudeste asiático (Hymowitz 1970).

Curiosamente, embora o início da domesticação da soja seja atribuído à região norte da China, um estudo recente sugeriu, com base em análises filogenéticas e de distribuição geográfica, que a origem monofilética de todos os cultivares de soja tenha ocorrido na região sul daquele país (Guo, Wang et al. 2010). Assim, até a terceira década do século XX, a Ásia permaneceu como o continente com a maior produção de soja; sendo China, Indonésia, Japão e Coreia os principais produtores em escala global (Piper e Morse 1910; Burtis 1950). Entre as décadas de 1940 e 1950 os EUA se tornaram os maiores produtores de soja do planeta, e ao final da década de 1960 a área plantada de soja no mundo ultrapassou os 28 milhões de hectares em aproximadamente 25 países (Hymowitz 1970).

A nomenclatura correta a ser atribuída à soja foi tópic de intenso debate entre taxonomistas no passado (Lawrence 1949; Paclt 1949). A controvérsia se deu devido a uma sobreposição presente na obra de Linnaeus (1753), em que a soja recebeu o nome *Phaseolus max* (p.725), e também o nome de *Dolichos soja* (p.727). Nesse contexto, o trabalho de Hermann (1962) contribuiu para a consolidação da nomenclatura *Glycine max* atribuída à soja, além de reduzir de 323 para 18 o número de espécies, subespécies e variedades do gênero *Glycine* (Hermann 1962; Hymowitz 1970).

## 1.2 A soja na agricultura atual

Sementes acumulam proteínas, carboidratos e lipídeos, que são consumidos para sustentar as etapas iniciais do desenvolvimento pós-germinativo (Schmidt, Barbazuk et al. 2011). Durante o desenvolvimento da semente, o acúmulo de reservas é regulado por uma complexa rede de componentes genéticos, fisiológicos e ambientais (Golombek, Rolletschek et al. 2001; Brocard-Gifford, Lynch et al. 2003; Weber, Borisjuk et al. 2005; Fait, Angelovici et al. 2006). Além disso, a produção de sementes é essencial na economia mundial, sendo importante para diversas finalidades, como a produção de biocombustíveis e alimentação humana e animal.

A soja (*Glycine max* Merr.) é a leguminosa mais cultivada no mundo, constituindo uma importante fonte de proteína e óleo. A produção mundial de soja e o rendimento por hectare aumentaram de forma constante ao longo do século passado com o melhoramento das técnicas agrícolas e o desenvolvimento de cultivares mais produtivos e adaptados a diferentes latitudes. A soja é um componente chave da alimentação humana e animal, sendo responsável por mais da metade da produção de oleaginosas do mundo (Ash 2015). Em

termos de massa de grãos produzidos, a soja é o quarto cultivo mais importante do mundo e ocupa nos Estados Unidos o segundo lugar em termos de área plantada (FAOSTAT 2016).

A produção mundial de soja é de mais de 255,000 Tg por ano, dos quais mais de 95% são provenientes de apenas sete países (Ainsworth, Yendrek et al. 2012). Embora a cultura tenha sido domesticada primeiramente na China e ausente do Novo Mundo até a colonização europeia, atualmente Estados Unidos, Brasil e Argentina são responsáveis por 80% da produção mundial de soja. O rendimento da soja nesses três países tem aumentado constantemente ao longo das últimas duas décadas (Ainsworth, Yendrek et al. 2012), frequentemente a reboque do dramático aumento na área plantada (Ainsworth, Yendrek et al. 2012). A produção comercial de soja no Brasil começou na década de 1960, superando 40 milhões de toneladas em 2002. O Brasil atualmente é o segundo maior produtor de soja do mundo ao lado dos EUA, com um rendimento de quase 3 toneladas por hectare (FAOSTAT 2016).

Análises históricas de cultivares lançados ao longo dos últimos 90 anos indicam que um dos principais fatores que impactam o rendimento é o aumento do número de sementes por planta (Morrison, Voldeng et al. 2000; Jin, Liu et al. 2010). Em contrapartida, patógenos e doenças impõem severas limitações à produção mundial. Embora as plantas sejam naturalmente resistentes à diversos estresses bióticos e abióticos, eventualmente patógenos e pragas conseguem se estabelecer, causando enormes prejuízos econômicos. As estratégias de controle tem sido essencialmente baseadas em programas de melhoramento genético e/ou utilização de pesticidas (Wally e Punja 2010).

### **1.3 Germinação em soja**

Em várias espécies vegetais, a dormência imposta pelo tegumento da semente cria



uma barreira física que impossibilita a absorção de água, extensão da radícula, e a saída de inibidores da germinação por parte da semente, impedindo que seja disparada a sequência de eventos da germinação (Kelly, Van Staden et al. 1992; Li e Foley 1997; Welbaum, Bradford et al. 1998). O processo denominado como pós-maturação (*after-ripening*) ocorre quando sementes maduras tornam-se não-dormentes, e adquirem a capacidade de germinar após o período de estocagem em ambiente seco (Probert 2000; Finch-Savage e Leubner-Metzger 2006). A pós-maturação depende diretamente da temperatura e da umidade do ambiente. O processo ocorre, em geral, em ambientes com umidade entre 8-15%; em sementes oleaginosas, porém, a umidade ideal para a pós-maturação tende a ser mais baixa se comparada com sementes ricas em amido (Probert 2000; Bazin, Batlla et al. 2010).

Por definição, o escopo de eventos compreendidos como germinação tem início com a captação de água pela semente quiescente não-dormente, e se encerra com a protrusão da radícula (Bewley e Black 1994; Koornneef, Bentsink et al. 2002). A germinação constitui um processo complexo, que comumente é dividido em três fases: embebição, aumento de atividade metabólica e iniciação do crescimento. A embebição é acompanhada por uma perda considerável de solutos pelas células (Hoekstra, Golovina et al. 1999). Ao passo que a citada perda de solutos acelera a germinação reduzindo a concentração de inibidores germinativos no interior das sementes (Matilla, Gallardo et al. 2005), tal evento causa também dano às membranas e compartimentos celulares em decorrência da rápida captação de água (Powell e Matthews 1979). Como meio de compensar o dano causado durante a desidratação, estocagem e reidratação, a semente aciona mecanismos de reparo durante a embebição, incluindo o reparo de membranas e de DNA (Bray e West 2005; Ogé, Bourdais et al. 2008). Morfologicamente, a germinação se encerra com a emersão da radícula e ruptura do tegumento, sendo o desenvolvimento subsequente denominado crescimento da plântula

ou crescimento pós-germinativo (Bewley e Black 1994).

No âmbito molecular, sabe-se que a semente seca madura contém mRNAs, estocados durante a maturação, referidos como mRNAs longevos (*long-lived mRNAs*). Esses mRNAs perduram na semente mesmo após o processo de dissecação e permanecem íntegros até a semente atingir o estado quiescente. Sendo assim, o aparato de síntese proteica utiliza estes mRNAs longevos para tradução nos estágios iniciais da germinação (Dure e Waters 1965; Waters e Dure 1966; Raghavan 2000). Mais de 10.000 mRNAs longevos diferentes foram identificados em análises transcriptômicas em *Arabidopsis* (Nakabayashi, Okamoto et al. 2005; Kimura e Nambara 2010; Okamoto, Tatematsu et al. 2010). Quantidades similares também foram reportadas em arroz e cevada (Howell, Narsai et al. 2009; Sreenivasulu, Borisjuk et al. 2010).

## 1.4 O processo de nodulação

Uma característica importante das leguminosas é a sua capacidade de estabelecer simbiose com bactérias fixadoras de nitrogênio (família Rhizobiaceae) (Stewart 1966; Schmutz, Cannon et al. 2010), num processo denominado nodulação. O desenvolvimento de nódulos depende complexas interações entre os simbiontes, sendo a morfologia destas estruturas frequentemente distinta em diferentes leguminosas (Jones, Kobayashi et al. 2007; Dilworth, James et al. 2008). Na interação entre *Medicago truncatula* e *Sinorhizobium meliloti*, os primeiros processos de sinalização envolvem flavonóides derivados de plantas e fatores Nod derivados de Rhizobium, induzindo a divisão celular no córtex radicular interno e levando à formação do primórdio nodular (Maunoury, Redondo-Nieto et al. 2010). Tem sido evidenciado nos últimos anos que a assimilação de nitrogênio e o metabolismo de carbono

são intimamente coordenados por uma complexa rede transcricional e enzimática (Nunes-Nesi, Fernie et al. 2010). Esta simbiose baseia-se na alocação de carbono para o simbionte *Bradyrhizobium japonicum* em troca de nitrogênio fixado. A partir de análises de ortologia foram identificadas em soja 28 nodulinas e 24 genes reguladores da nodulação (Schmutz, Cannon et al. 2010).

## 1.5 Abordagens genômicas no estudo da soja

A recente publicação do genoma da soja (Schmutz, Cannon et al. 2010) abriu um novo leque de perspectivas para exploração de novos aspectos bioquímicos e fisiológicos. Dentre estas iniciativas, estudos em larga escala vem sendo conduzidos para estudar o transcriptoma, proteoma e metaboloma de diferentes tecidos e condições fisiológicas da espécie (Libault, Farmer et al. 2010).

O genoma de soja tem um tamanho estimado de 978 megabases. Dos 46.430 genes preditos com alta confiança à data da publicação, 34.073 (73%) apresentaram ortólogos em outras angiospermas, sendo classificados em 12.523 famílias (Schmutz, Cannon et al. 2010). Expansões em tandem de famílias de genes são comuns na cultura da soja e incluem NBS-LRR, proteínas F-box, proteínas de resposta a auxina, e proteínas contendo outros domínios comumente encontrados em grandes famílias em genomas vegetais (Schmutz, Cannon et al. 2010; Bellieny-Rabelo, Oliveira et al. 2013).

Diversos estudos tem lançado mão de metodologias de sequenciamento de segunda geração (e.g. RNA-Seq) para explorar o transcriptoma de soja em diversos tecidos e condições fisiológicas (Libault, Farmer et al. 2010; Severin, Woody et al. 2010; Jones e Vodkin 2013; Wang, Cao et al. 2014; Prince, Joshi et al. 2015; Song, Prince et al. 2016). Dada a importância da soja na agricultura mundial e os progressos na área de bioquímica e

biologia molecular, torna-se crucial a associação entre a genômica e bioinformática com o melhoramento genético, visando o desenvolvimento direcionado de novos cultivares (Specht, Hume et al. 1999). Apesar do sucesso dos programas de melhoramento genético clássico, é imperativa a necessidade de se compreender as bases bioquímicas de características vantajosas que são selecionadas pelos melhoristas a fim de identificar potenciais alvos para futuras melhorias e extrapolação para outras espécies (Ainsworth, Yendrek et al. 2012).

### 1.5.1 Sintenia, duplicação e evolução de genomas de plantas

A maioria das espécies de angiospermas descendem de ancestrais que sofreram duplicações integrais de genoma, seja através de autoploidia ou alopoliploidia. Evidências recentes demonstram que possivelmente todas as angiospermas tenham sofrido pelo menos um evento de poliploidização, sugerindo que a hibridização é um fator importante na geração de diversidade no grupo (Soltis e Soltis 2009). Apesar do alto potencial deletério das alterações de ploidia em decorrência do agudo desbalanço genético e metabólico (Otto e Whitton 2000), evidências genômicas fortíssimas indicam uma alta prevalência deste tipo de fenômeno em plantas (Semon e Wolfe 2007). Dado o alto número de duplicações genômicas, alguns cientistas argumentam que a poliploidização ofereça novas oportunidades para sucesso evolutivo. Foi defendido por Crow e Wagner (2006) que a duplicação do genoma poderia reduzir o risco de extinção através de redundância funcional, robustez mutacional, e taxas aumentadas de evolução e adaptação (Van de Peer, Maere et al. 2009). Contudo, é importante notar que em 500-600 milhões de anos de evolução dos vertebrados, não mais que duas (ou três em teleósteos) duplicações genômicas completas foram detectadas até o momento (Van de Peer, Maere et al. 2009), indicando que a frequência dos eventos de poliploidização observada em plantas não ocorre em vertebrados.

A inferência de que duas ou mais regiões genômicas evoluíram de uma única região ancestral é conhecida como sintenia. Análises de sintenia tem sido gradativamente mais exploradas no âmbito da pesquisa em genômica de plantas. Tal ocorre devido ao aumento significativo de genomas de plantas sequenciados desde o advento das plataformas de sequenciamento de segunda geração (Tang, Wang et al. 2008; Schnable, Freeling et al. 2012; Akhunov, Sehgal et al. 2013). Eventos de escala genômica que geram duplicações gênicas, desde pequenas (apenas um gene ou pequenos segmentos) até grandes (cromossomos inteiros ou genomas completos) são elementos de grande relevância no processo evolutivo, na medida em que expandem famílias gênicas e fazem emergir novas funções e padrões de expressão gênicos (Lynch e Conery 2000; Semon e Wolfe 2007; Bellieny-Rabelo, Oliveira et al. 2013). Por meio de estudos de genômica comparativa, quatro eventos de duplicação de genoma inteiro (*Whole Genome Duplication: WGD*) e uma triplicação de genoma inteiro (*Whole Genome Triplication: WGT*) foram identificados desde a linhagem basal das plantas até o gênero *Glycine* (Figura 1). Tais eventos citados ocorreram há ~319 milhões de anos atrás (*Million Years Ago: mya*) ao ancestral das espermatófitas, ~192mya ao ancestral das angiospermas, ~130mya ao ancestral das dicotiledôneas (triplicação de genoma), ~58mya ao ancestral das leguminosas e ~13mya ao ancestral do gênero *Glycine* (Jaillon, Aury et al. 2007; Jiao, Wickett et al. 2011; Severin, Cannon et al. 2011). Imediatamente após um evento de duplicação gênica a pressão evolutiva sobre as duas cópias é comumente reduzida, o que pode ocasionar à progressiva perda de um dos parálogos. Contudo, há ainda a possibilidade de ocorrer divergência funcional entre os parálogos (seja por subfuncionalização ou neofuncionalização), o que pode levar à conservação de ambos no genoma, gerando uma fonte rica para possíveis inovações biológicas (Abrouk, Murat et al. 2010; Wang, Wang et al. 2012; Rensing 2014).

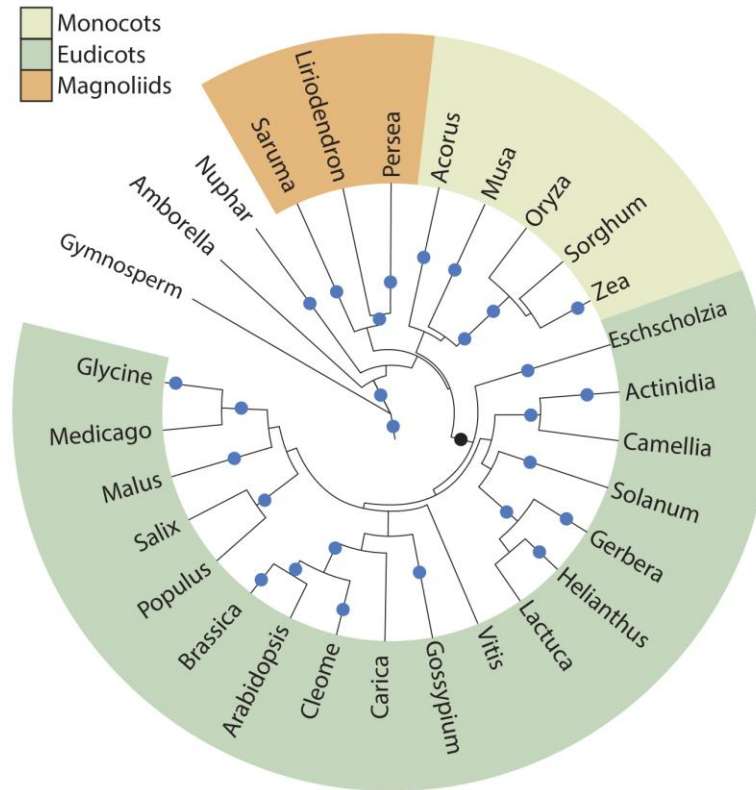


Figura 1 Árvore filogenética incluindo gêneros de monocotiledôneas e dicotiledôneas. Círculos azuis representam eventos de duplicação. O círculo preto representa um evento de triplicação de genoma. (Adaptado de Van de Peer (2011))

## 1.6 Desenvolvimento embrionário e regulação hormonal da germinação

A divisão clássica dos estágios do desenvolvimento embrionário de leguminosas compreende as seguintes fases: (1) globular, (2) cordiforme, (3) cotilédone, (4) maturação, (5) dormência (Walbot 1971). A maturação representa uma fase crucial à preparação para a germinação, quando a semente produz grande quantidade de proteínas de reserva, lectinas, e inibidores de tripsina (Millerd 1975; Orthofer 1978; Thomas 1993). O órgão da semente

responsável pelo armazenamento de reservas recebe o nome de cotilédone. As moléculas estocadas são fundamentais para a sobrevivência da semente sob condições adversas de temperatura e umidade (Goldberg, Hoschek et al. 1981; Seo, Nambara et al. 2009). Quando encontradas as condições ambientais favoráveis, a germinação é desencadeada, causando a retomada dos processos metabólicos da semente, levando a uma fase de absorção de água, e a subsequente emergência da radícula (Finch-Savage e Leubner-Metzger 2006; Holdsworth, Finch-Savage et al. 2008).

A manutenção da dormência, bem como o término do processo – permitindo o início da germinação – são controladas majoritariamente pelos hormônios ácido abscísico (*abscisic acid*: ABA) e ácido giberélico (*gibberellic acid*: GA), , respectivamente. ABA é um sesquiterpenóide endógeno de sementes – independente do estado de dormência – relacionado aos processos de fechamento estomatal, desenvolvimento da semente, tolerância à dessecação e mecanismos de resposta a estresse (Wareing e Saunders 1971; Nambara e Marion-Poll 2005). Em contrapartida, GA's compõem um grupo de diterpenóides formado por ~100 moléculas, no qual poucas atuam como hormônios bioativos no controle de processos como: florescimento, expansão foliar, alongação de caule e germinação (Yamaguchi 2008). Nas duas últimas décadas diversos trabalhos foram capazes de identificar componentes envolvidos na interação regulatória entre ABA-GA – por exemplo, NCED (*9-cis epoxy-carotenoid dioxygenase*), GA2ox (*gibberellin 2-oxidase*), GID1 (*gibberellin insensitive dwarf 1*), HUB1 (*histone monoubiquitination 1*), PIF's (*phytocrome-interacting factors*) – elucidando aspectos importantes sobre sua ativação/desativação e biossíntese (Ni, Tepperman et al. 1998; Ueguchi-Tanaka, Ashikari et al. 2005; Liu, Koornneef et al. 2007; Seo, Nambara et al. 2009; Nonogaki 2014). Atualmente acredita-se que a regulação hormonal por ação de ABA-GA seja um processo dependente mais da proporção entre as concentrações

dos hormônios do que de sua concentração absoluta independente (Finch-Savage e Leubner-Metzger 2006).

Embora GA e ABA sejam os fitormônios mais bem descritos no que se refere à regulação da germinação, outras moléculas tem sido reportadas com importante atuação no balanço hormonal deste processo, como etileno (Karssen, Zagorski et al. 1989; Koornneef, Bentsink et al. 2002) e brassinosteróides (Steber e McCourt 2001). Uma vez observado que muitas espécies vegetais produzem etileno durante a germinação, e também que fatores antagônicos à dormência estimulam a produção deste hormônio, suspeitou-se que a produção do etileno durante o início da germinação poderia contribuir com a quebra do estado de dormência (Matilla 2000). Os primeiros estudos que apontaram tal associação foram realizados nas espécies: *Trifolium subterraneum* (Esashi e Leopold 1969), *Arachis hypogea* (Ketring e Morgan 1971; Ketring e Morgan 1972), *Spergula arvensis* (Van Staden, Olatoye et al. 1973) e *Avena fatua* (Adkins e Ross 1981). A resposta ao etileno em *Arabidopsis* é mediada por cinco receptores conhecidos como ETR1 (Ethylene Response 1), ETR2 (Ethylene Response 2), EIN4 (Ethylene Insensitive 4), ERS1 (Ethylene Response Sensor 1), e ERS2 (Ethylene Response Sensor 2) (Chang, Kwok et al. 1993; Hua e Meyerowitz 1998; Hua, Sakai et al. 1998; Sakai, Hua et al. 1998). Embora a ação destes receptores não seja completamente conhecida (Wilson, Kim et al. 2014), estudos genéticos demonstram que na ausência de etileno os receptores regulam positivamente um inibidor da via de sinalização do etileno conhecido como CTR1 (Constitutive Triple Response 1) (Kieber, Rothenberg et al. 1993). Brassinosteróides (BRs) constituem uma família com mais de 40 hormônios esteróides encontrados em diversas espécies vegetais, em praticamente todos os tecidos; embora mais proeminentemente em sementes e pólen (Mandava 1988; Arteca 1995; Schmidt, Altmann et al. 1997; Yokota 1997; Clouse e Sasse 1998). A ação dos



brassinosteróides foi descrita em sementes de tabaco, nas quais o hormônio foi responsável pela indução da ruptura do endosperma (Leubner-Metzger 2001). Em *Arabidopsis* também foi proposta a ação de brassinosteróides como antagonistas de ABA durante a germinação. Tal hipótese é baseada na observação de que plantas mutantes para os genes DET2 (esteróide- $\alpha$ -redutase, necessário à síntese de BRs) (Noguchi, Fujioka et al. 1999) e BRI1 (um receptor de BRs) (Friedrichsen, Joazeiro et al. 2000), apresentaram sensibilidade aumentada ao ABA, reduzindo drasticamente a taxa de germinação das sementes (Steber e McCourt 2001).

## **1.7 Fatores de transcrição (TFs)**

Famílias gênicas que desempenham a função de TFs são caracterizadas por possuírem um domínio de ligação ao DNA que lhes confere a capacidade de se ligar a regiões regulatórias próximas a genes, regulando sua transcrição (Pabo e Sauer 1992). As funções dos TFs sempre despertam grande interesse científico devido ao seu papel em processos relacionados ao desenvolvimento, diferenciação celular e respostas ao ambiente. O extenso grupo de genes codificadores de TFs em cada genoma é também uma potencial fonte de inovações evolutivas importantes para o surgimento e desenvolvimento de características cruciais das leguminosas, tais como a associação com bactérias fixadoras de nitrogênio (Ramalingam, Vera Cruz et al. 2003; Miller, Galiba et al. 2006; Francia, Barabaschi et al. 2007; Libault, Joshi et al. 2009). TFs constituem o grupo mais numeroso e diversificado de proteínas envolvidas na transcrição gênica. TFs são os principais reguladores da transcrição, formando coletivamente complexas redes regulatórias que determinam os programas transcricionais tecido- e condição-dependentes (Ratcliffe e Riechmann 2002).

A mudança relativamente recente de uma perspectiva centrada no *gene*, para uma centrada no *genoma*, favorece em especial o estudo dos TFs, uma vez que estes compreendem uma fração substancial dos genomas eucarióticos e são agrupados em um número relativamente grande de famílias gênicas, classificadas pelo tipo de domínio de ligação ao DNA que codificam (Riechmann e Ratcliffe 2000). Tradicionalmente, TFs tem sido descritos como proteínas modulares que contêm uma variedade de domínios para: ligação ao DNA, ativação, ligação de moléculas sinalizadoras e interação com outras proteínas (Schwechheimer, Zourelidou et al. 1998). Os domínios de ligação ao DNA encontram-se altamente conservados ao longo da história evolutiva dos eucariotos, podendo, frequentemente, ser identificados a partir da estrutura primária da proteína. Em contraste, os domínios de ativação são pouco conservados, sendo sua detecção computacional frequentemente elusiva (Triezenberg 1995).

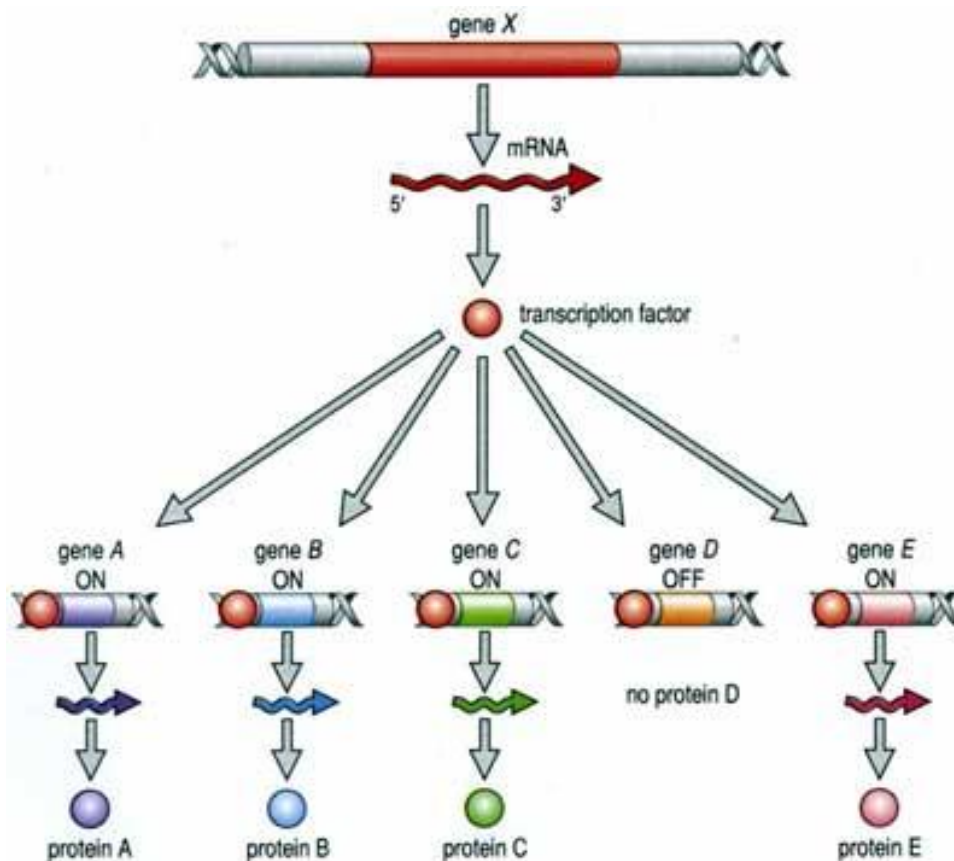


Figura 2 Ilustração esquemática do funcionamento da ativação diferencial de vários genes (A-E) por um fator de transcrição genérico codificado pelo gene X. (<http://scienceblogs.com/pharyngula/2006/07/09/transcription-factors-and-morp/>)

Relações evolutivas entre TFs são usualmente detectadas através de alinhamento de sequências de homólogos próximos pela busca de perfis de seqüências com matrizes de PSI-BLAST (Schäffer, Wolf et al. 1999), ou *hidden Markov models* (Eddy 2011) para domínios de ligação a DNA conservados. Estas metodologias permitem o mapeamento dos genes que codificam TFs, tornando viável estimar os tamanhos destas famílias em vários genomas, de forma semi-automática (Tabela 1). O estudo dos domínios de ligação a DNA permite também avaliar as relações evolutivas entre os TFs. Estudos independentes em *E. Coli* (Pérez-Rueda e Collado-Vides 2000; Babu e Teichmann 2003), archaea (Aravind e Koonin 1999), plantas e animais (Morgenstern e Atchley 1999; Riechmann e Ratcliffe 2000; Ledent e Vervoort 2001)

demonstraram que os conjuntos de TFs obtêm seus domínios de ligação a DNA de um repertório ancestral relativamente pequeno e muito bem conservado (Babu, Luscombe et al. 2004).

*Tabela 1 Média do número fatores de transcrição presentes em genomas vegetais em comparação com a média total de genes de cada grupo (Adaptado de Jin, Zhang et al. (2014).*

Lineage	Species	Gene	TF (%)	Family
Chlorophyta	10	10 550	141 (1.34)	35
Bryophyta	1	32 273	1079 (3.34)	53
Lycopodiophyta	1	22 271	665 (2.99)	54
Coniferophyta	1	71 158	1851 (2.60)	55
Basal Magnoliophyta	1	26 846	900 (3.35)	58
Monocot	15	34 017	1701 (5.00)	58
Eudicot	38	34 798	1861 (5.35)	58

Várias evidências indicam que a proliferação de diferentes famílias de TFs tem contribuído substancialmente para a diversidade morfológica nos animais. A generalidade dessa observação foi reforçada pela evidência da massiva expansão e diversificação de várias famílias de TFs em *A. thaliana*. Muitos destes incluem domínios de ligação ao DNA bem caracterizados, como MADS-box e MYB, conhecidos por participarem de funções biológicas específicas de plantas, incluindo desenvolvimento floral e diferenciação meristemal (Riechmann e Ratcliffe 2000; Lespinet, Wolf et al. 2002). Dentre as principais (mais numerosas) famílias gênicas de TFs detectadas em *A. thaliana*, muitas são especificamente encontradas em plantas, como por exemplo: AP2 (APETALA2), Dof, EREBP (*ethylene-responsive-element binding protein*), NAC, WRKY, ARF-Aux/IAA (*auxin response*

*factor*). Alguns outros grupos não muito numerosos em animais ou leveduras tem sido consideravelmente expandidos na linhagem vegetal, como por exemplo: MYB (esta constitui a maior família de TFs em *A. thaliana*), MADS e bZIP (*basic-region leucine-zipper*) (Riechmann e Ratcliffe 2000).

Há aproximadamente 10 anos estimou-se que mais de 99% dos TFs preditos em leguminosas ainda careciam de caracterização funcional (Udvardi, Kakar et al. 2007). Após esta publicação, 13 genomas de leguminosas foram montados em sua completude (ou bastante próximo dela) (Dash, Campbell et al. 2016), aumentando ainda mais o hiato entre informação gênica e comprovação experimental de função. Diante deste cenário, investigações computacionais e evolutivas são essenciais para auxiliar na predição funcional deste vasto número de TFs conhecidos (Libault, Joshi et al. 2009). Inicialmente, vários TFs foram identificados baseados em similaridade de sequência entre genes de leguminosas e de *A. thaliana* conhecidos por resultarem em fenótipo mutante específico. TFs de leguminosas associados ao controle de meristemas florais foram descobertos dessa forma: LjFLO, PsFLO/LFY, PsPEAM4, e MtPIM (Hofer, Turner et al. 1997; Berbel, Navarro et al. 2001; Taylor, Hofer et al. 2002; Dong, Zhao et al. 2005; Benlloch, d'Erfurth et al. 2006). TFs envolvidos especificamente na interação com bactérias rhizobium, responsáveis pela fixação de nitrogênio, foram identificados. Entre estes TFs estão os genes NSP1 e NSP2, que codificam dois TFs do tipo GRAS que são essenciais para a infecção dos pelos radiculares em *M. truncatula*. Os pelos radiculares dos mutantes que não possuem os genes não são infectados pelos simbiossitos (Catoira, Galera et al. 2000; Oldroyd e Long 2003). Em *Lotus japonicus*, outra importante espécie de leguminosa, TFs associados especificamente ao processo de nodulação também foram identificados: NIN, ASTRAY, MtERN (Schauser, Roussis et al. 1999; Nishimura, Ohmori et al. 2002; Middleton, Jakab et al. 2007).

### 1.7.1 Principais famílias de fatores de transcrição em plantas

Atualmente são conhecidas diversas famílias de TFs exclusivamente encontradas em genomas de plantas. Dentre estas, estão AP2/EREBP (Riechmann e Meyerowitz 1998), NAC (Aida, Ishida et al. 1997), os fatores de resposta à auxina (ARFs) (Guilfoyle, Hagen et al. 1998) e WRKY (Eulgem, Rushton et al. 2000). Alguns domínios presentes em reguladores transcricionais específicos de plantas são caracterizados por serem inovações surgidas nesta linhagem. A família AP2, por exemplo, possui um novo módulo de reconhecimento de DNA utilizando uma estrutura folha- $\beta$  presente apenas em plantas (Allen, Yamasaki et al. 1998). Outras famílias, por sua vez, parecem ser relacionadas filogeneticamente a famílias existentes antes das plantas, como por exemplo as proteínas GRAS. Com base no arranjo similar da arquitetura de domínios funcionais, acredita-se que GRAS possua relação filogenética distante com a família STATS de animais (Richards, Peng et al. 2000).

Desde a clonagem da primeira proteína WRKY (SPF1) em *Ipomea batatas* (L) (Ishiguro e Nakamura 1994), uma quantidade expressiva de genes WRKY vem sendo identificada experimentalmente em diversas espécies vegetais, tais como trigo (Rushton, Macdonald et al. 1995), *A. thaliana* (de Pater, Greco et al. 1996), tabaco (Yang, Chen et al. 1999), arroz (Zhang, Xie et al. 2004), e uva (Marchive, Mzid et al. 2007). Atualmente, as WRKY estão entre as mais bem descritas famílias de fatores de transcrição em plantas. E participam de processos que vão desde a resposta a estresses bióticos (Pandey e Somssich 2009), a embriogênese (Lagace e Matton 2004), desenvolvimento do tricoma (Johnson, Kolevski et al. 2002), sinalização hormonal (Miao e Zentgraf 2007), senescência foliar (Zhou, Jiang et al. 2011) entre outras. Sabe-se que domínio WRKY contém cerca de 60 aminoácidos constituído por um *core* conservado (WRKYGQK) na porção N-terminal, além de um *motif zinc-finger-like* em sua extremidade C-terminal (Eulgem, Rushton et al. 2000). Proteínas

WRKY possuem afinidade com o *motif* W-box (TTGACT/C), embora outros sítios de ligação com afinidade a WRKY tenham sido recentemente identificados (Cai, Qiu et al. 2008).

A primeira proteína da família (*v-myb*) foi identificada no genoma do vírus da mieloblastose aviária (Klempnauer, Gonda et al. 1982). Na linhagem vegetal, o gene C1 de *Zea mays*, responsável pela regulação da pigmentação por antocianina, foi o primeiro gene MYB descrito experimentalmente (Paz-Ares, Ghosal et al. 1987). A estrutura das proteínas da família apresenta comumente de 1 a 4 repetições do domínio MYB que podem atuar sinergicamente ou individualmente na ligação ao DNA. Cada uma das repetições possui cerca de 52 aminoácidos contendo 3 resíduos de triptofano (ou resíduo alifático) espaçados regularmente formando um núcleo hidrofóbico na região do domínio (Kanei-Ishii, Sarai et al. 1990). Proteínas MYB são usualmente classificadas em 4 grupos de acordo com o número de repetições encontradas do domínio MYB: MYB1R, R2R3-MYB, MYB3R e MYB4R (Dubos, Stracke et al. 2010). Integrantes do grupo R2R3-MMYB são específicas de plantas e, também, o grupo mais abundante na linhagem, com mais 100 representantes conhecidas em monocotiledôneas e dicotiledôneas (Wilkins, Nahal et al. 2009).

## 2 JUSTIFICATIVA

As novas tecnologias de sequenciamento de segunda e terceira geração vem permitindo a elucidação de diversos genomas vegetais, desafiadores em virtude de seu tamanho, ploidia e complexidade de regiões repetitivas. Boa parte destas tecnologias pode ainda ser aplicado ao estudo de transcriptomas em diversas condições, visando compreender a dinâmica e complexidade regulatória. O estudo adequado destes conjuntos de dados tem potencial de revelar *loci* com potencial biotecnológico, facilitando sua

aplicabilidade em experimentos no laboratório e no campo. Trabalhos realizados nos últimos anos investigaram de forma detalhada vários genomas e transcriptomas de soja em diversos tecidos e condições fisiológicas (Libault, Farmer et al. 2010; Schmutz, Cannon et al. 2010; Severin, Woody et al. 2010; Jones e Vodkin 2013; Wang, Cao et al. 2014; Whaley, Sheridan et al. 2015; Yuan, Li et al. 2016), fornecendo valiosa e extensa fonte de dados para análise.

Apesar de diversos estudos transcriptômicos em soja terem revelado aspectos importantes acerca da fisiologia desta espécie, nenhum destes conjuntos de dados compreendeu tecidos de soja durante a germinação. Comparados aos micro-arranjos (*microarrays*), RNA-Seq pode, em teoria, detectar quase todos os transcritos de uma determinada amostra biológica, o que dificilmente acontece com os microarranjos, restritos a detecção de genes/transcritos já conhecidos (Chen, Wang et al. 2011).

### **3 OBJETIVOS**

#### **3.1 Objetivos gerais**

- Sequenciar o transcriptoma de eixos embrionários de sementes de soja durante vários estágios da germinação;
- Investigar os principais eventos bioquímicos e regulatórios ativados em eixos embrionários de soja durante a germinação.

#### **3.2 Objetivos específicos**

- Observar padrões globais no perfil transcricional dos eixos embrionários de soja durante a germinação;



- Correlacionar dados transcricionais com eventos morfológicos visíveis;
- Identificar genes diferencialmente expressos durante a germinação, com especial foco sobre os genes envolvidos com funções regulatórias;
- Investigar a expressão de genes responsáveis pela reativação de vias metabólicas essenciais à produção energética para manutenção do crescimento do eixo embrionário;
- Descrever a ação regulatória desempenhada por genes pertencentes às vias de biossíntese e sinalização por fitormônios no eixo embrionário.

## 4 METODOLOGIA

### 4.1 Material vegetal, germinação e purificação de RNA

As sementes de soja (cultivar BRS284) foram dispostas diretamente em placas de Petri de 15cm, contendo 2g de algodão molhado em 30 mL de água destilada estéril. As placas foram colocadas em câmara climática a 28 °C com foto período 12/12 horas e 50-65% de umidade relativa. As sementes foram retiradas das placas após 3 horas, 6 horas, 12 horas e 24 horas após a embebição (as sementes dormentes também foram utilizadas posteriormente). Após a coleta das sementes, os eixos embrionários foram separados dos cotilédones. Para cada amostra, 20 eixos foram coletados para a extração de RNA. O RNA total foi extraído utilizando-se o *Rneasy plant mini kit* (Qiagen). A qualidade da extração de RNA foi avaliada com uso de gel de agarose 1%, espectrômetro *Picodrop Microliter* e *Agilent Bioanalyzer 2100*. A preparação de amostras para o RNA-Seq foi conduzida no LQFPP, em colaboração com a Profa. Antônia Elenir A. Oliveira. Para cada amostra biológica, 2 réplicas

tiveram seus mRNAs sequenciados. As amostras de RNA foram preparadas utilizando o Kit de preparação de amostras *TruSeq RNA v2* (Illumina, cat. RS-122-2001) e submetidas ao sequenciamento *paired-end* (2x 100bp) em plataforma *Illumina HiSeq 2000* da LaCTAD (UNICAMP, Campinas, Brasil).

## 4.2 Mapeamento dos fragmentos ao genoma de referência

Inicialmente, fragmentos de leitura (reads) tiveram as sequencias de adaptadores aparadas utilizando o software Trimmomatic (Bolger, Lohse et al. 2014). Análises gerais de qualidade do sequenciamento foram realizadas utilizando o software FastQC (<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc>). Estas análises permitiram a identificação de bases enviesada ou de baixa qualidade nas 10 primeiras posições dos fragmentos. Estas regiões foram aparadas visando manter apenas sequencias com alta qualidade de resolução.

Em seguida, as sequencias foram alinhadas ao genoma de referência (versão 1.1) (Schmutz, Cannon et al. 2010) tornando possível a montagem dos fragmentos nos *loci* conhecidos, bem como o mapeamento dos transcritos ativos nos estágios do experimento. Esta etapa foi realizada utilizando os softwares Bowtie 2.1.0 (Langmead e Salzberg 2012) e Tophat2 (Kim, Pertea et al. 2013) utilizando o parâmetro “--b2-very-sensitive” (-D 20 -R 3 -N 0 -L 20 -i S,1,0.50). Com os dados do sequenciamento alinhados, foi possível então estimar abundancia relativa de transcritos normalizados em FPKM (*Fragments Per Kilobase of transcript/gene per Million mapped reads*) e realizar testes para estimar expressão diferencial para cada um deles. Como forma de estabelecer um limiar de confiabilidade para os dados de expressão, genes com o valor de FPKM igual ou superior a 1 foram considerados “expressos”. Tal etapa de normalização dos dados foi realizada utilizando o software Cufflinks

(Trapnell, Roberts et al. 2012).

### 4.3 Análise de expressão gênica diferencial

A ferramenta Cuffdiff (Trapnell, Roberts et al. 2012) foi utilizada para realizar as seguintes análises: a) de genes regulados positivamente na progressão temporal da germinação e; b) de genes preferencialmente expressos em estágios individuais do experimento. Genes regulados positivamente foram filtrados em ambiente R (<http://www.r-project.org/>) pela ferramenta *cummeRbund* (Goff, Trapnell et al. 2012), utilizando o método *getSig()* (*alpha=0.1, level='genes'*). Genes preferencialmente expressos foram encontrados pelo método *csSpecificity()*, também presente no *cummeRbund*, que computa a distância *Jensen-Shannon*; esta distância fica sempre entre 0 (para genes expressos constitutivamente) e 1 (genes exclusivamente expressos). Genes com valores entre 0.7 e 1 foram considerados preferencialmente expressos.

Os padrões transcricionais foram então agrupados por HCL (*hierarchical clustering*) (Eisen, Spellman et al. 1998) e os respectivos mapas de calor (*heatmaps*) foram gerados pelo software Gtools (Perez-Llamas e Lopez-Bigas 2011). Alguns grupos de dados foram expressos na forma de *z-score* em vez de seus valores absolutos em FPKM. A normalização por *z-score* indica quantos desvios padrão um elemento está de sua média; e pode ser calculado pela fórmula:  $z = (X - \mu) / \sigma$ ; na qual “z” é o *z-score*, “X” é o valor do elemento, “ $\mu$ ” representa a média da população, e “ $\sigma$ ” é o desvio padrão.

Outras análises adicionais foram realizadas por meio de *scripts* desenvolvidos *in house*, escritos em linguagem Perl (<https://perl.org/>). A submissão do conjunto completo de dados deste trabalho foi feita ao banco de dados GEO (<https://www.ncbi.nlm.nih.gov/geo/>) sob o identificador: GSE83481.

#### 4.4 Análise de ortologia e regiões de sintonia

A busca por genes ortólogos foi realizada com os softwares BLASTp e MCL. No primeiro passo o software BLASTp (Altschul, Madden et al. 1997) foi executado sobre os conjuntos completos (*all vs. all*) de proteínas preditas das espécies envolvidas na análise, obtidas do banco de dados do DOE-JGI (*Department of Energy Joint Genome Institute*; em <http://phytozome.jgi.doe.gov/>). Para cada comparação par-a-par, foi aplicado o filtro de E-value máximo de 1e-05 e cobertura mínima de 50% da menor proteína do par. Em seguida o software MCL (Dongen 2000) foi utilizado para realizar o agrupamento das proteínas em potenciais grupos de genes ortólogos (parâmetro: -l 3.0). Para evitar redundância nos dados, apenas a proteína mais longa codificada por cada gene foi utilizada nos casos em que informações de *splicing* alternativo estavam disponíveis no banco de dados.

#### 4.5 Análise de fatores de transcrição e enriquecimento de termos de Gene Ontology

As anotações das famílias de fatores de transcrição em soja foram obtidas da plataforma online PlantTFDB (Jin, Zhang et al. 2014) e inspecionadas manualmente. O enriquecimento de termos GO foi conduzido na plataforma agriGO (FDR < 0,01) (Du, Zhou et al. 2010). A lista de genes expressos (FPKM > 1) foi utilizada como *background* da análise estatística. As listas de termos GO enriquecidos foram consolidadas utilizando a ferramenta SEACOMPARE (presente no *website* do agriGO). A redundância de termos GO enriquecidos foi removida usando REVIGO (Supek, Bošnjak et al. 2011).

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Análise geral do transcriptoma de eixos embrionários de soja durante a germinação

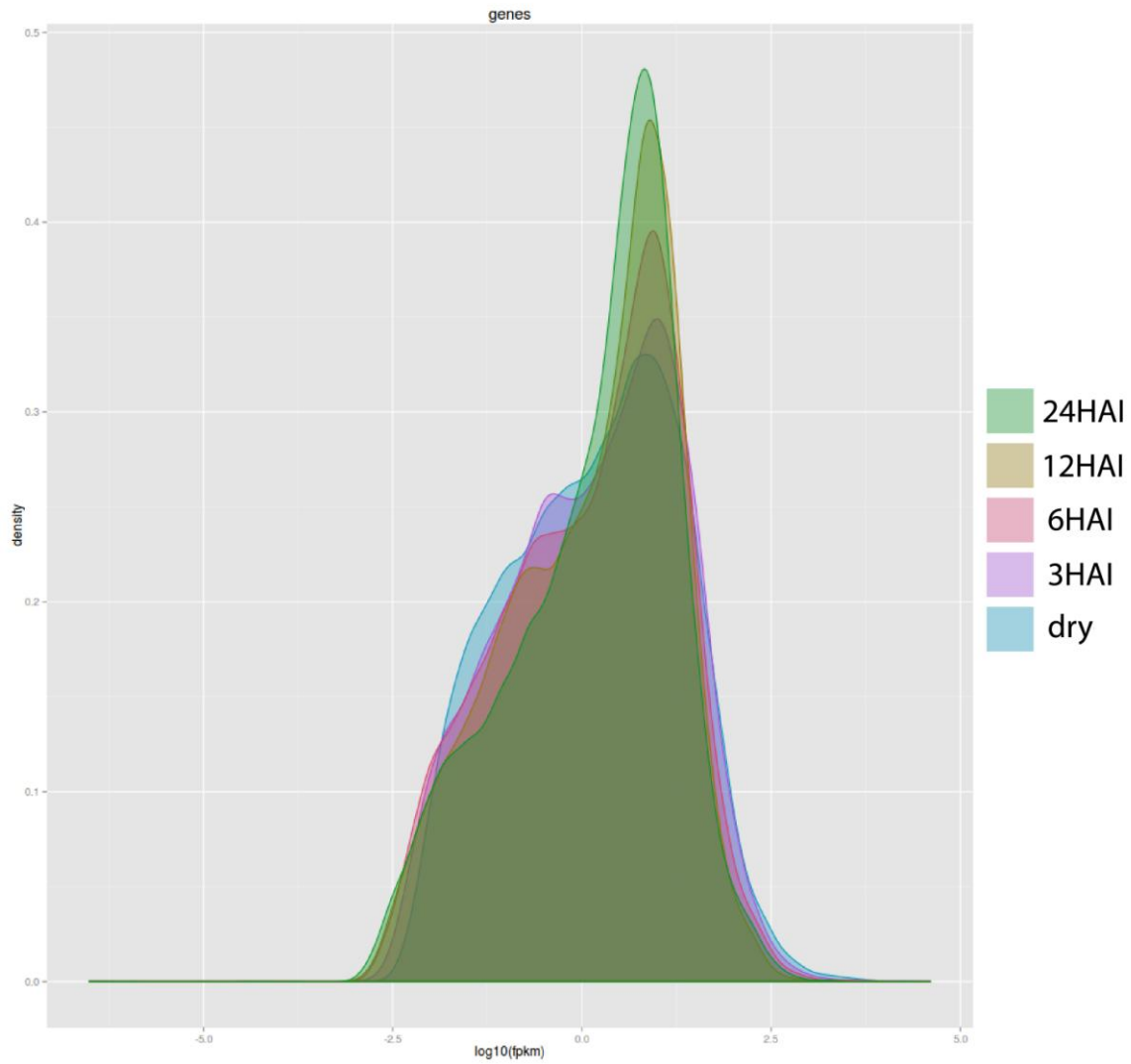
A densidade da expressão gênica encontrada em cada um dos estágios da germinação do eixo embrionário da soja está representada a seguir (Figura 3) – eixo seco (“dry”), com 3 (“3HAI”), 6 (“6HAI”), 12 (“12HAI”) e 24 horas (“24HAI”) após a embebição. O resultado evidencia que a intensidade da expressão gênica tende a aumentar de forma gradativa no decorrer da linha temporal analisada pelo experimento. Tendo a amostra 24HAI (24 horas após a germinação) com a maior densidade de transcritos com altos níveis de expressão.

Para fins de controle da qualidade do sequenciamento, a similaridade/divergência do padrão de expressão entre as amostras (Figura 4) e, também, entre as respectivas réplicas (Figura 5) foi analisada. Com base no método estatístico Jensen-Shannon (*JS distance*) foi possível estimar a divergência entre os transcriptomas. Em cada matriz, cada um dos estágios é comparado a todos os outros. E tonalidades mais escuras representam maior divergência entre as partes comparadas. O resultado confirma a qualidade do sequenciamento para cada par, baseado na similaridade observada entre as réplicas biológicas, bem como a esperada divergência entre as amostras mais distantes na escala temporal.

Foram avaliados os índices de correlação da expressão gênica entre as amostras sequenciadas representados em forma de dispersão (Figuras 6–15). A comparação é feita par-a-par e cada ponto do gráfico representa o valor encontrado para a intensidade de expressão de cada gene (valor convertido em  $\log_{10}$ ). Quanto mais dispersos os pontos se

encontram em relação à linha diagonal pontilhada, mais divergentes serão os perfis de expressão considerados. Além de apontar para a esperada diferença de padrão entre amostras mais distantes na linha temporal, o resultado enfatiza, principalmente, a distância crítica entre o padrão de expressão gênica na amostra de 24HAI se comparada às demais.

Um total de 242-372 milhões de reads *paired-end* foram sequenciados por estágio (Tabela 1). Em termos de profundidade de sequenciamento o conjunto total de dados deste trabalho (~1,58 bilhões de *reads*) é aproximadamente 2,4 vezes maior que outro transcriptoma de soja publicado recentemente (Wang, Cao et al. 2014). A qualidade dos transcriptomas sequenciados neste trabalho é ilustrada pela alta taxa de mapeamento de *reads* no genoma de referência: ~1,45 bilhões (91,8%) foram mapeados com sucesso, dos quais ~1,03 bilhões (~71%) foram mapeados em um único locus (Tabela 2). Foram encontrados 33.305 genes expressos em pelo menos um estágio (Figura 16) (FPKM > 1), que correspondem a ~61,5% dos genes preditos no genoma de referência da soja (Schmutz, Cannon et al. 2010).



*Figura 3 Densidade da expressão gênica, medida em FPKM, para cada um dos estágios da germinação: dry (eixo seco), 3HAI (embrião 3 horas após a embebição), 6HAI (embrião 6 horas após a embebição), 12HAI (embrião 12 horas após a embebição), 24HAI (embrião 24 horas após a embebição).*

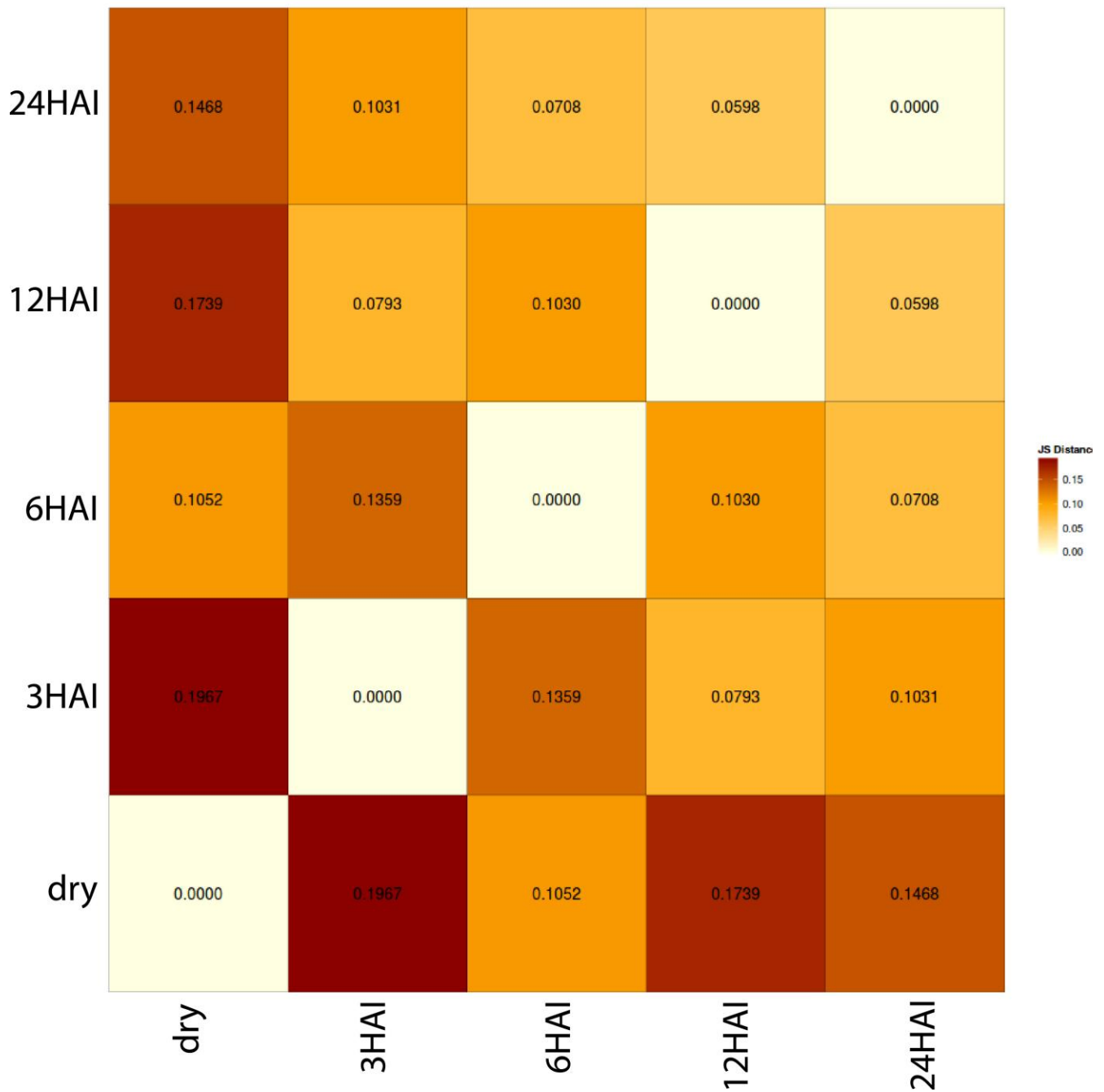


Figura 4 A similaridade/divergência entre o perfil de expressão gênica das amostras sequenciadas é representada pela distância JS (Jensen-Shannon). Maiores valores representam maior divergência entre os perfis das amostras.



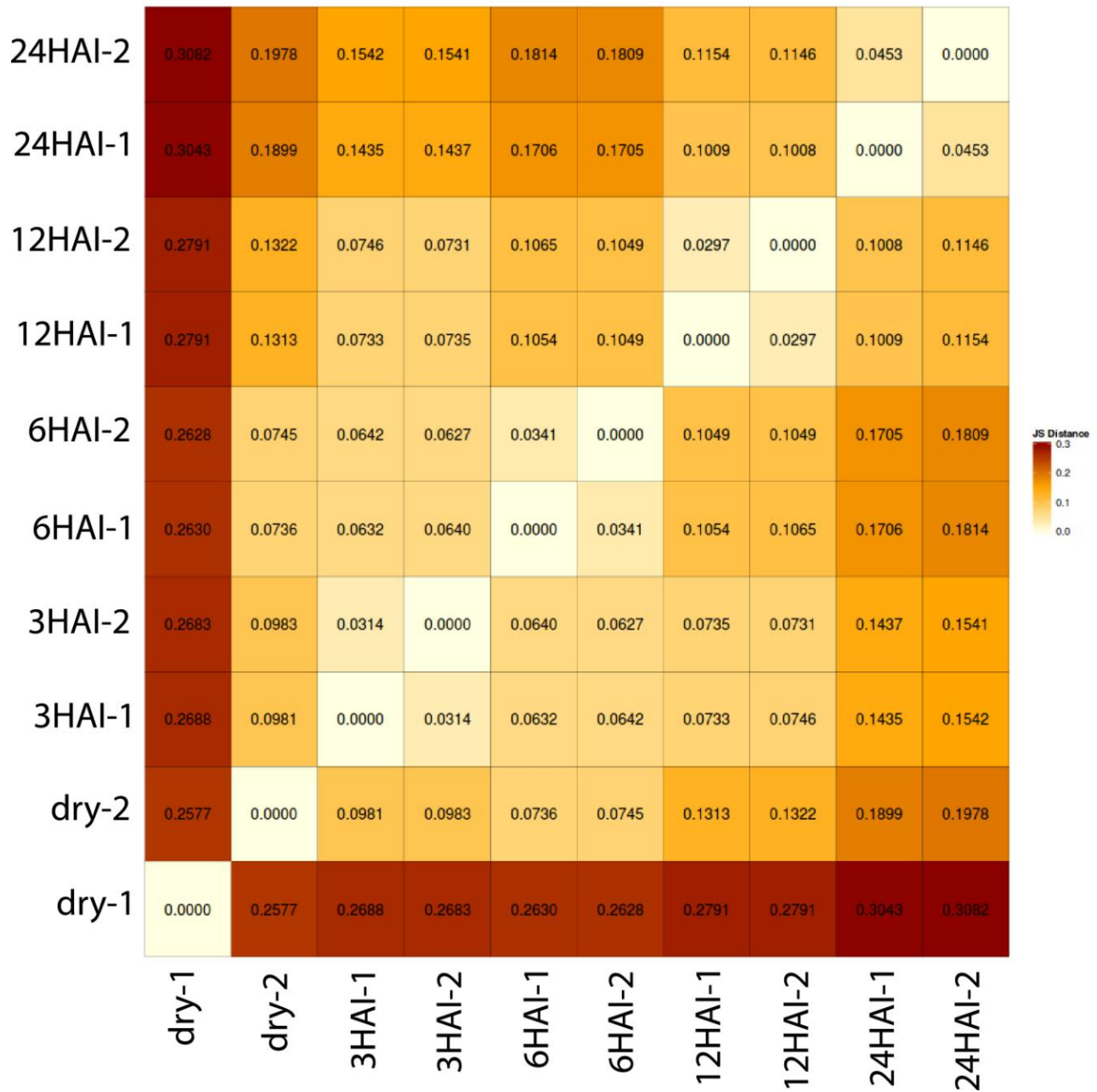
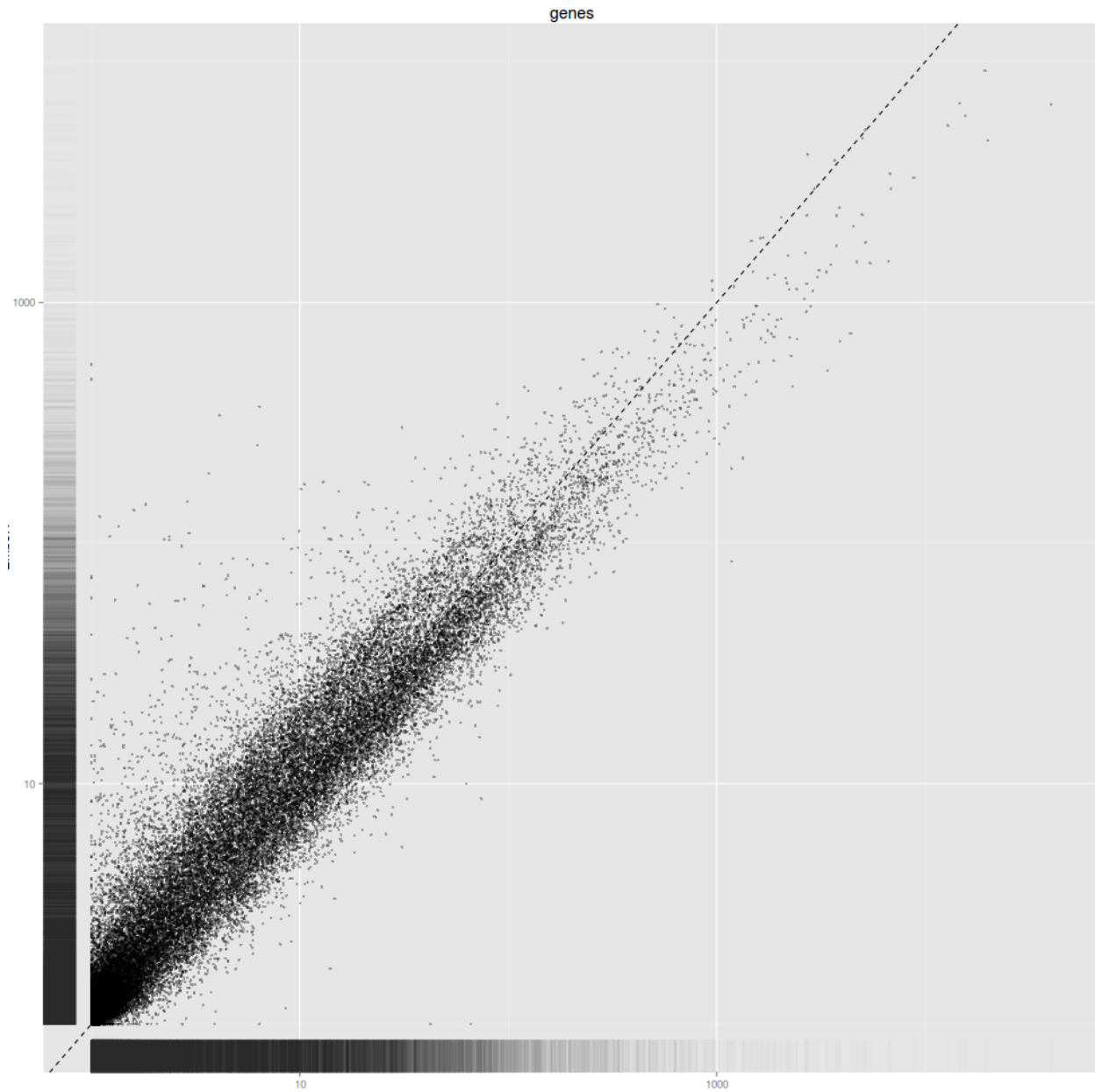
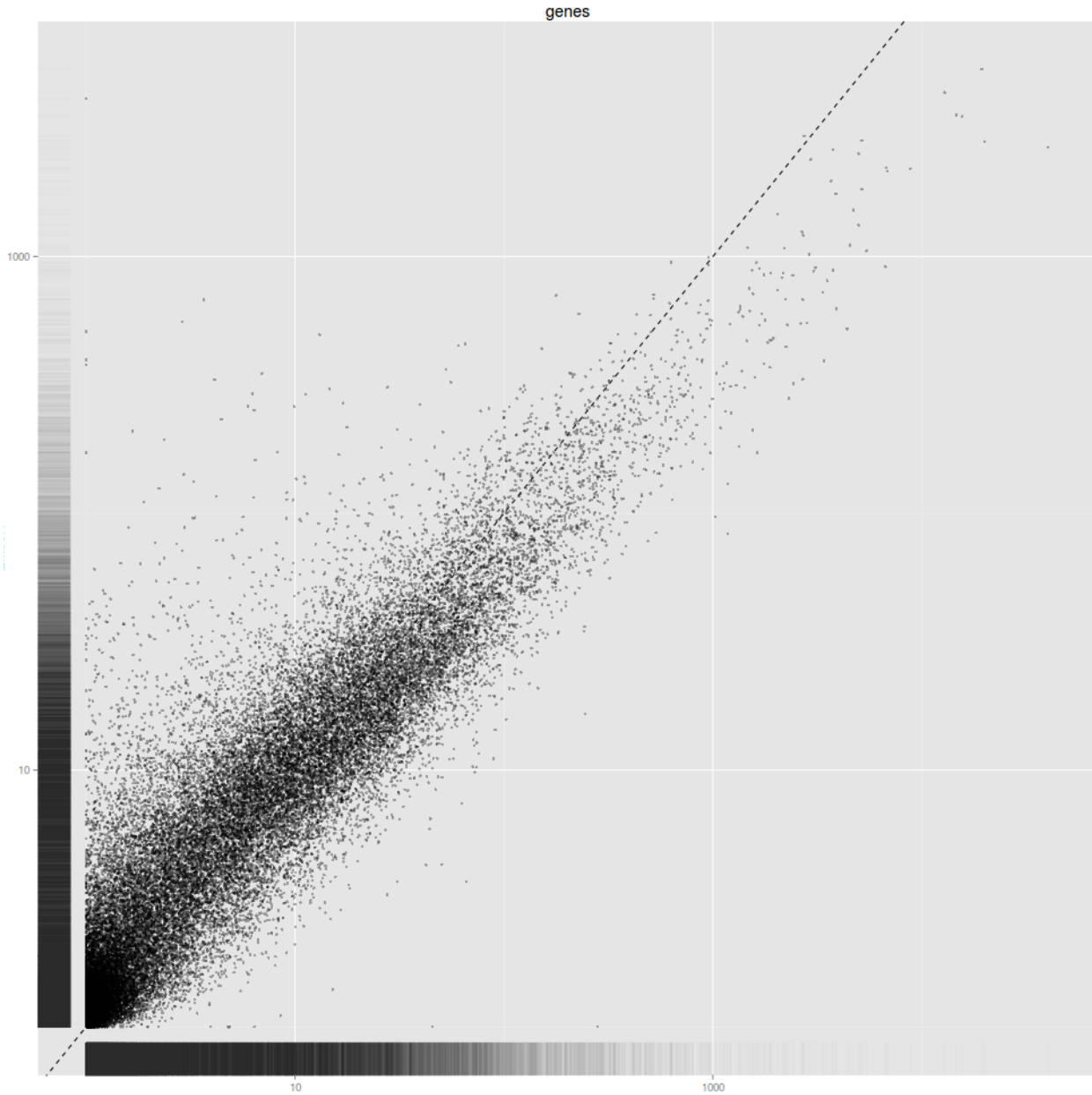


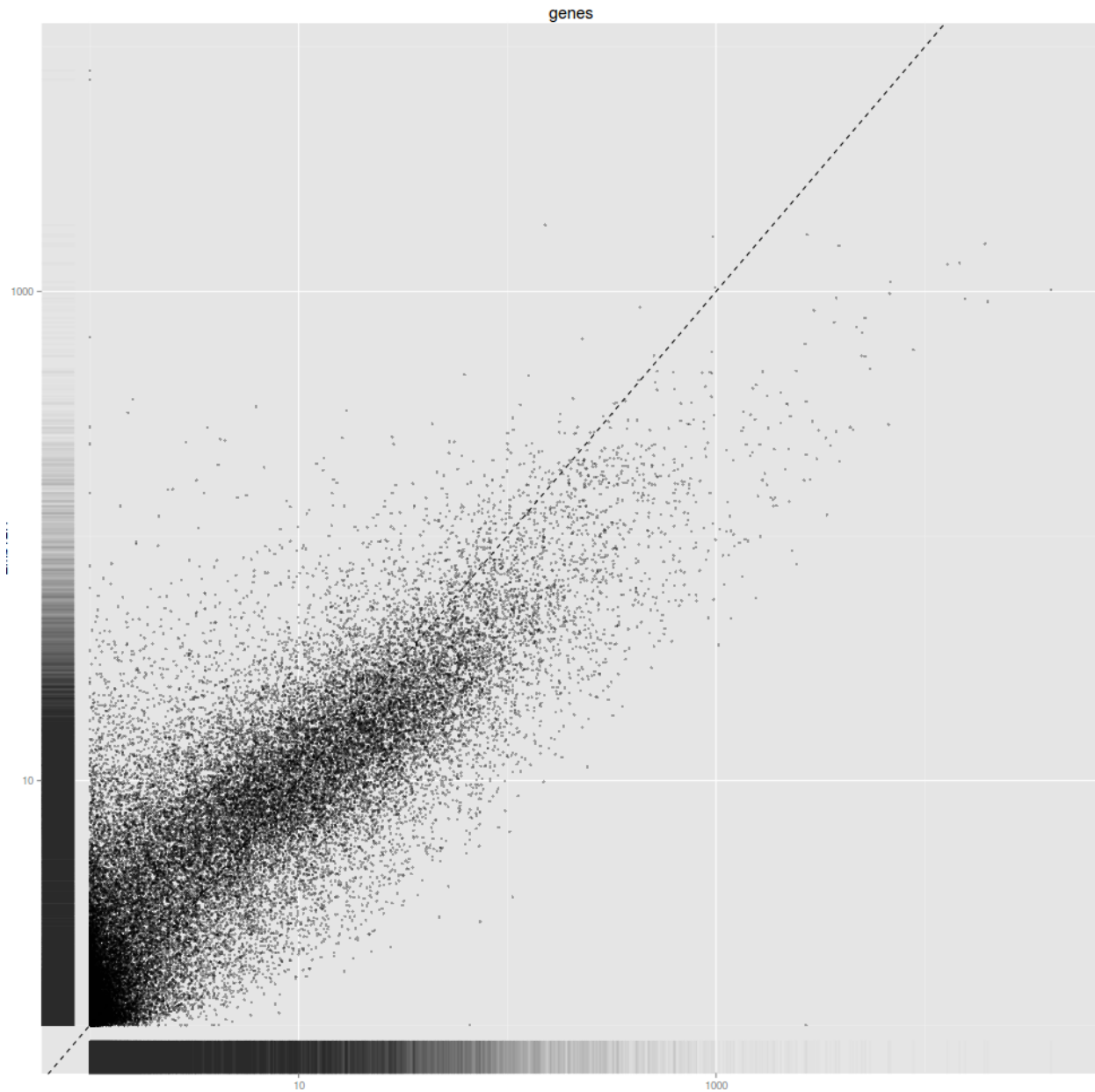
Figura 5 A similaridade/divergência entre o perfil de expressão gênica das amostras sequenciadas é representada pela distância JS (Jensen-Shannon). Maiores valores representam maior divergência entre os perfis das amostras.



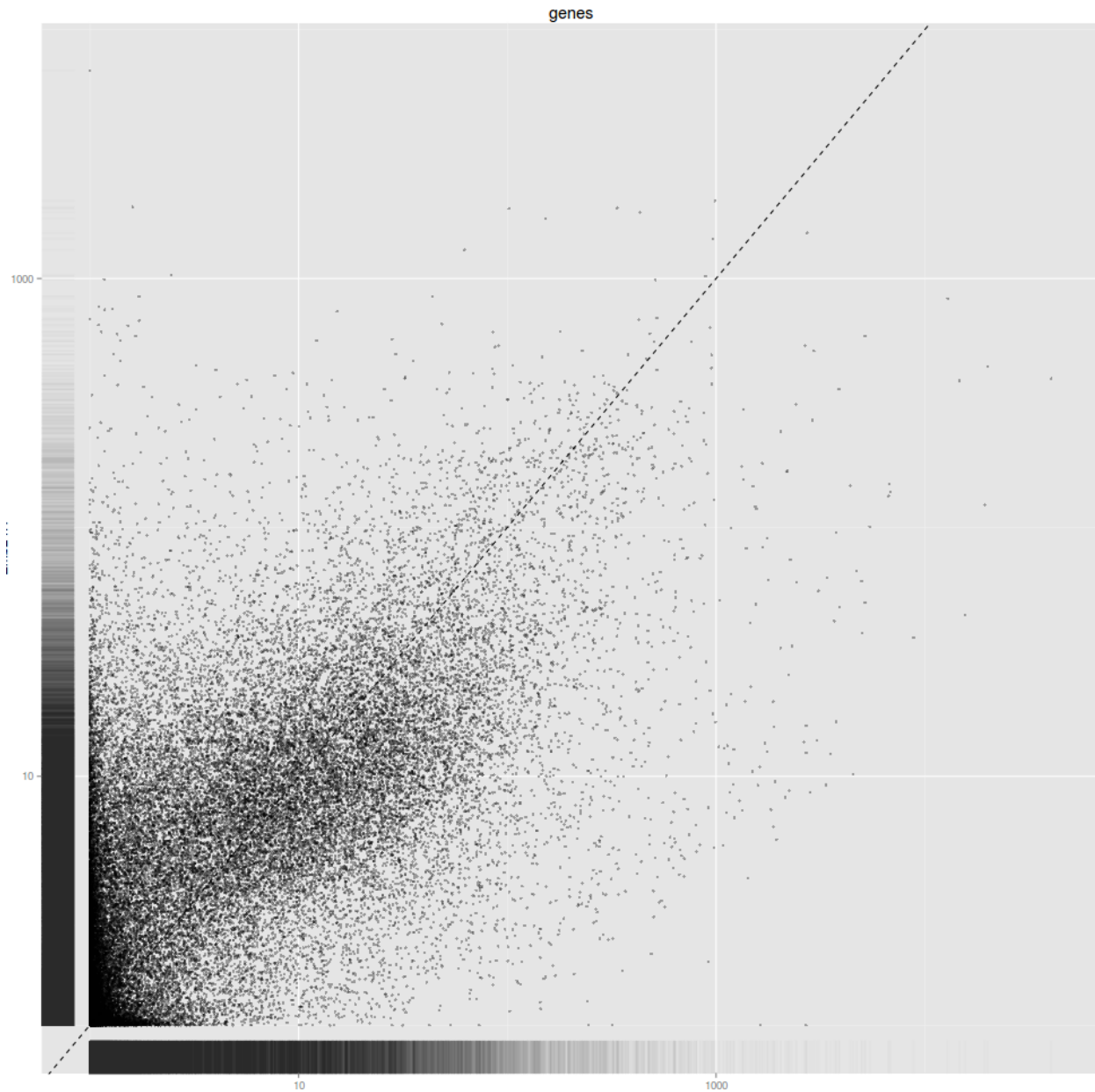
*Figura 6 Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x, e semente 3 horas após embebição (3HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica*



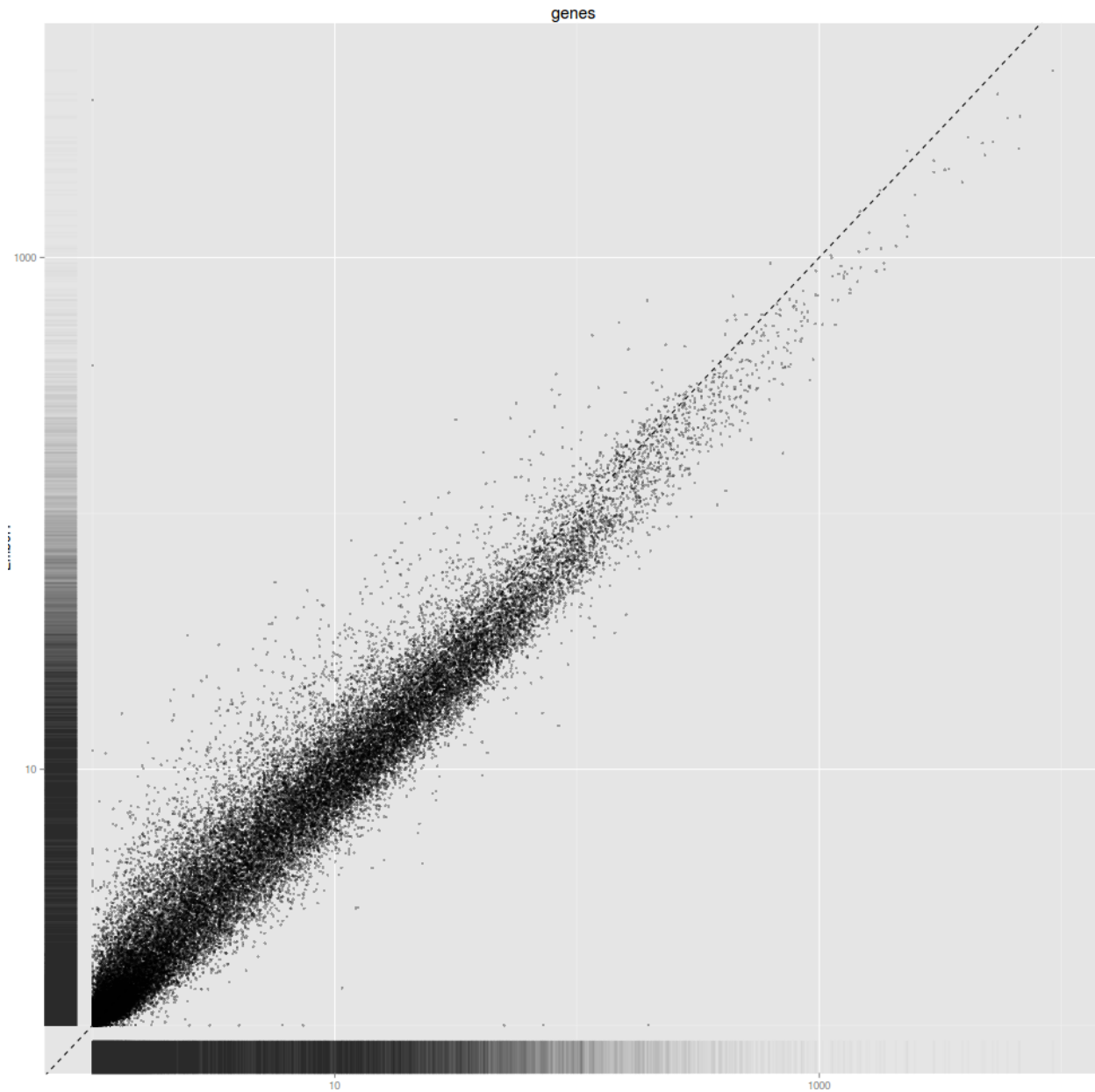
*Figura 7 Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x e semente 6 horas após embebição (6HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*



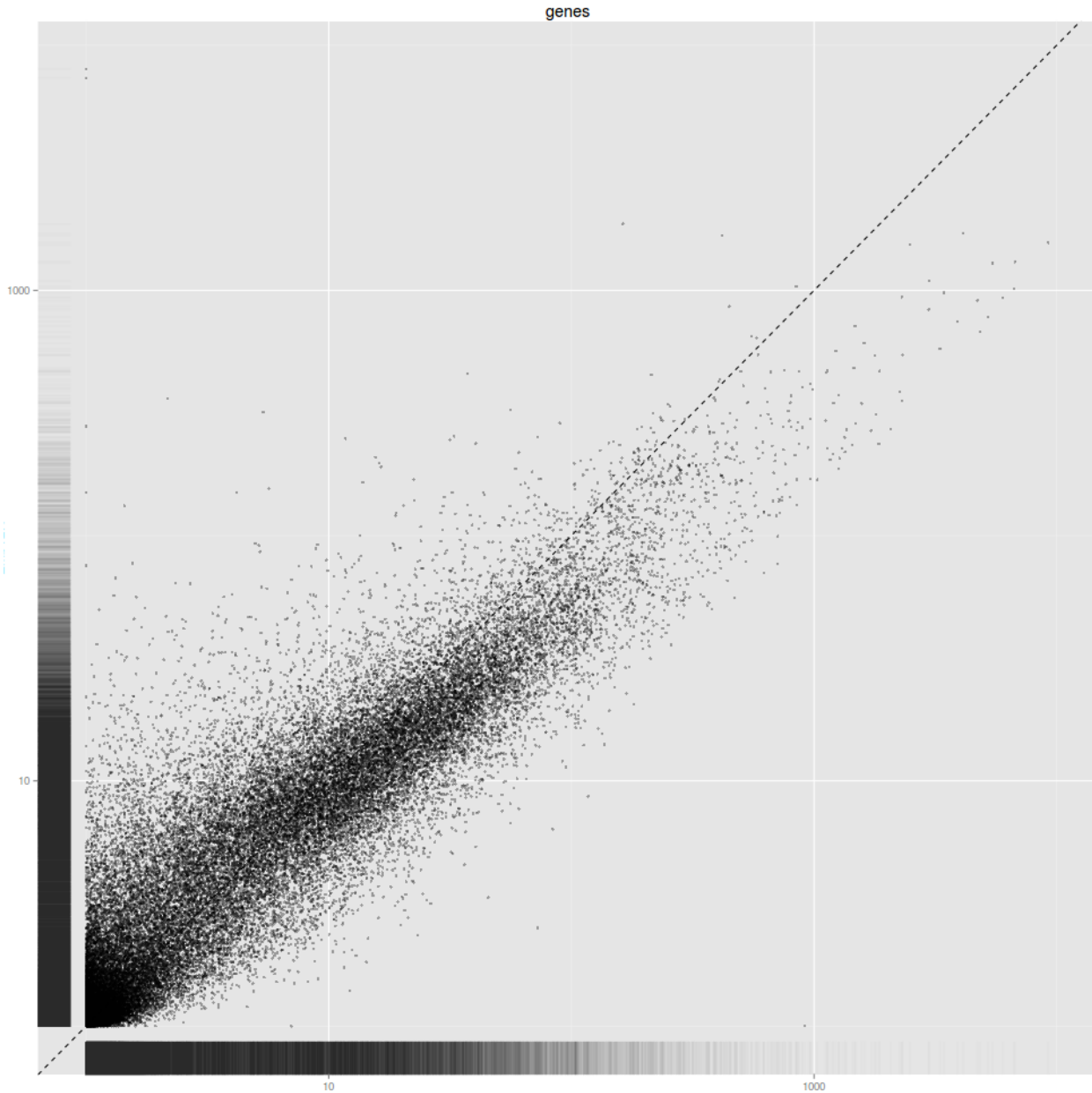
*Figura 8 Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x, e semente 12 horas após embebição (12HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*



*Figura 9 Comparação da expressão gênica entre as amostras de eixo seco (dry) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*

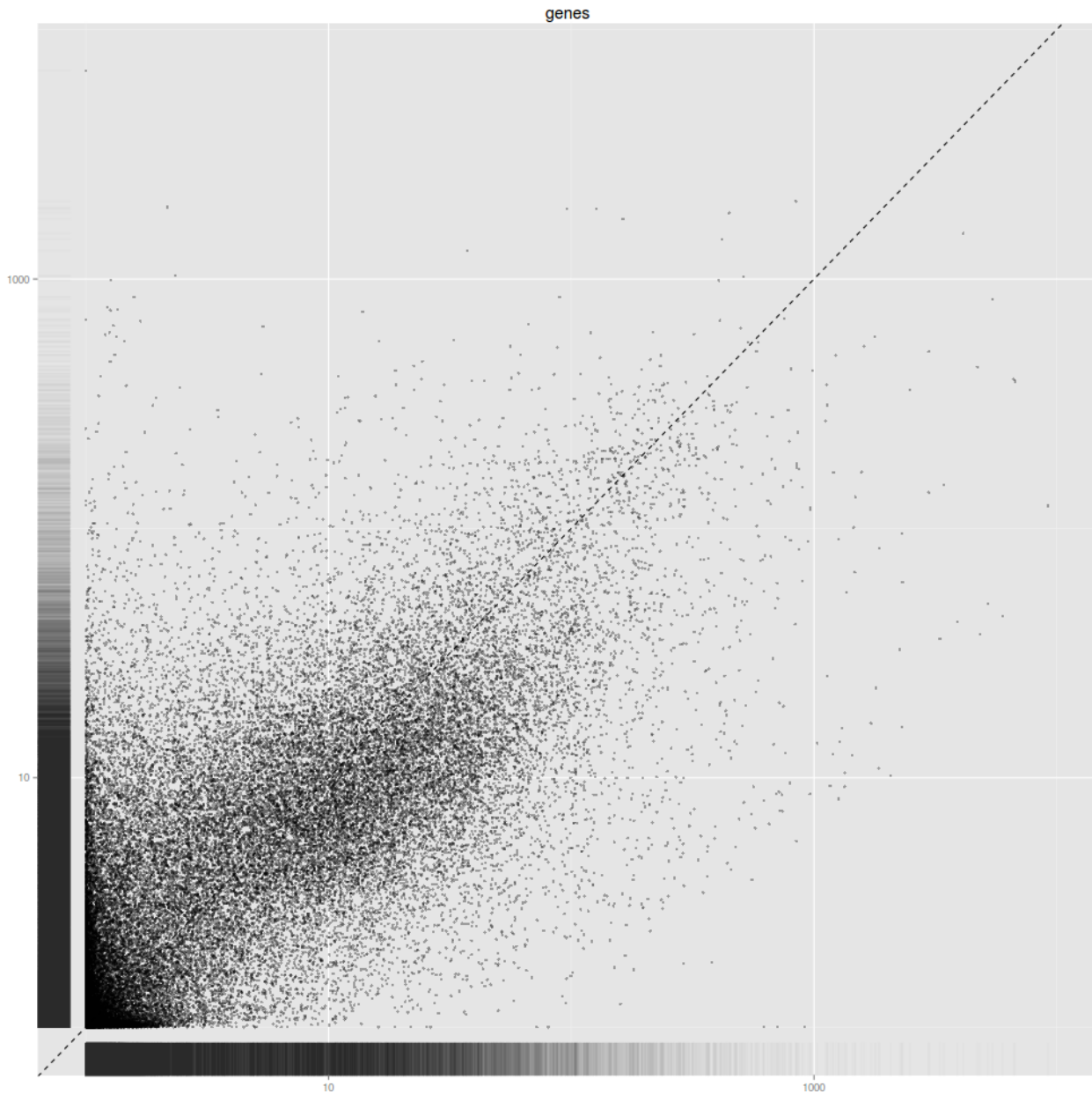


*Figura 10 Comparação da expressão gênica entre as amostras de semente 3 horas após germinação (3HAI) no eixo x e semente 6 horas após embebição (6HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*



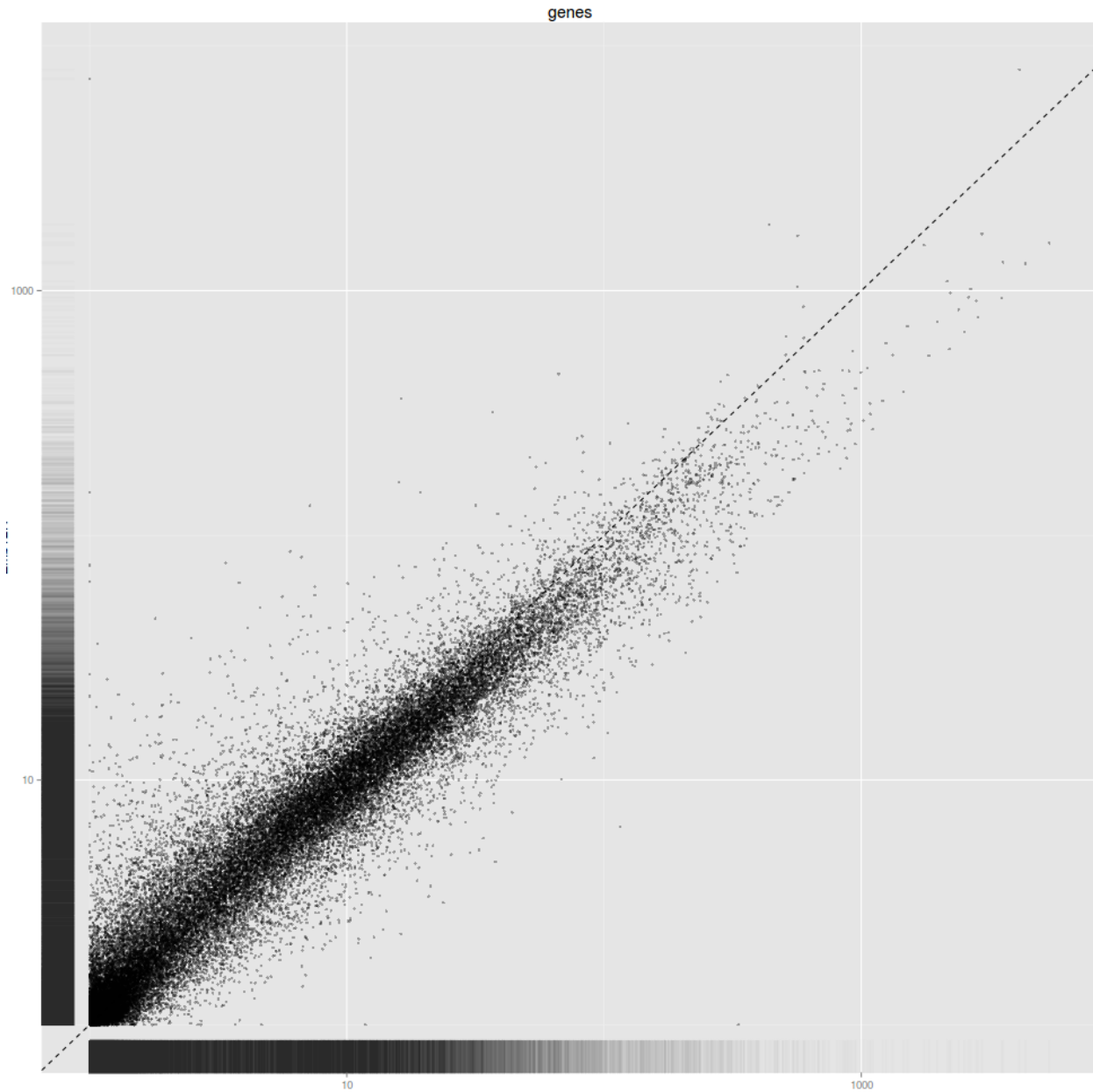
*Figura 11 Comparação da expressão gênica entre as amostras de semente 3 horas após embebição (3HAI) no eixo x e semente 12 horas após embebição (12HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*



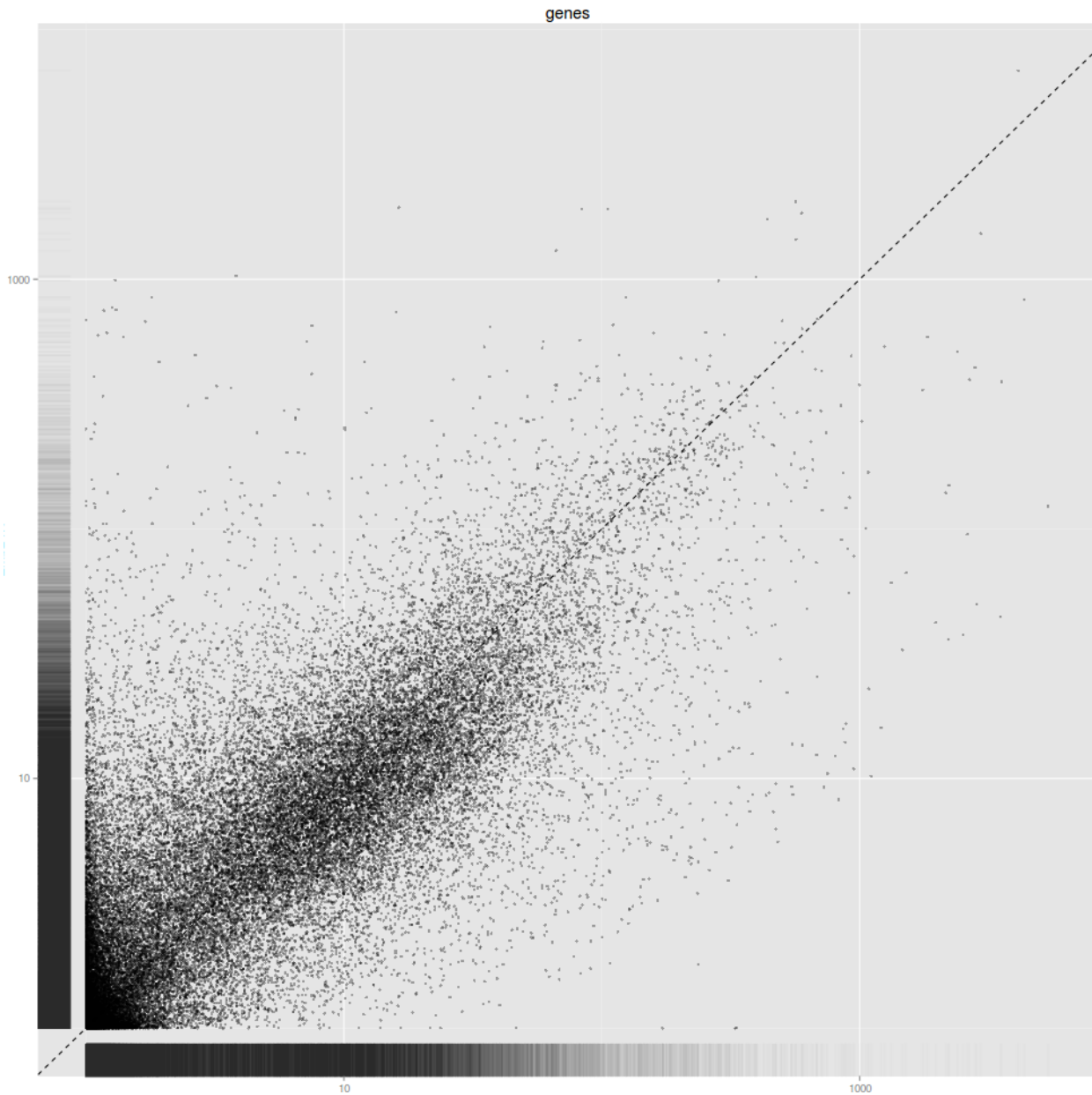


*Figura 12 Comparação da expressão gênica entre as amostras de semente 3 horas após embebição (3HAI) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*

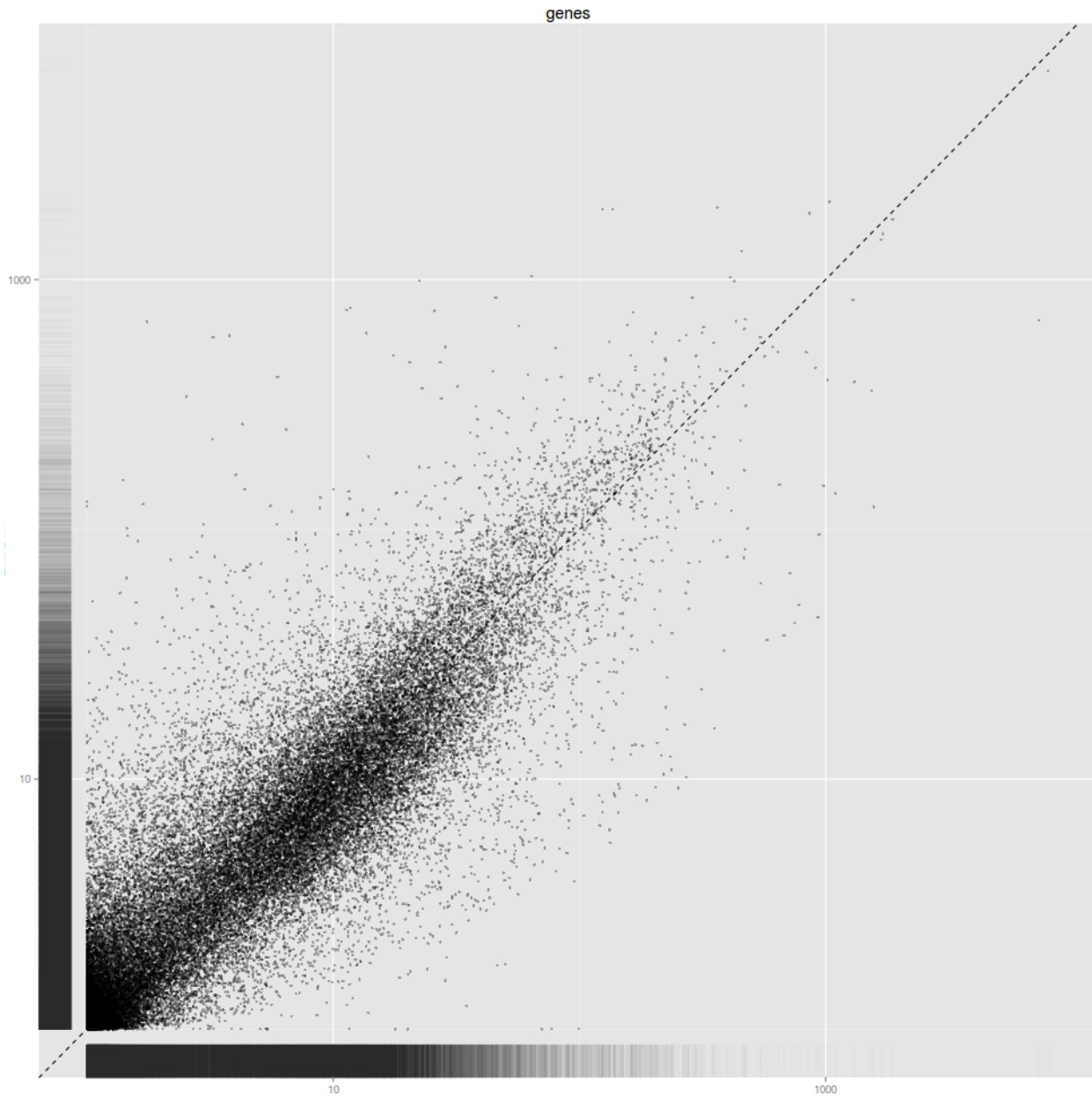




*Figura 13 Comparação da expressão gênica entre as amostras de semente 6 horas após embebição (6HAI) no eixo x e semente 12 horas após embebição (12HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*



*Figura 14 Comparação da expressão gênica entre as amostras de semente 6 horas após embebição (6HAI) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*

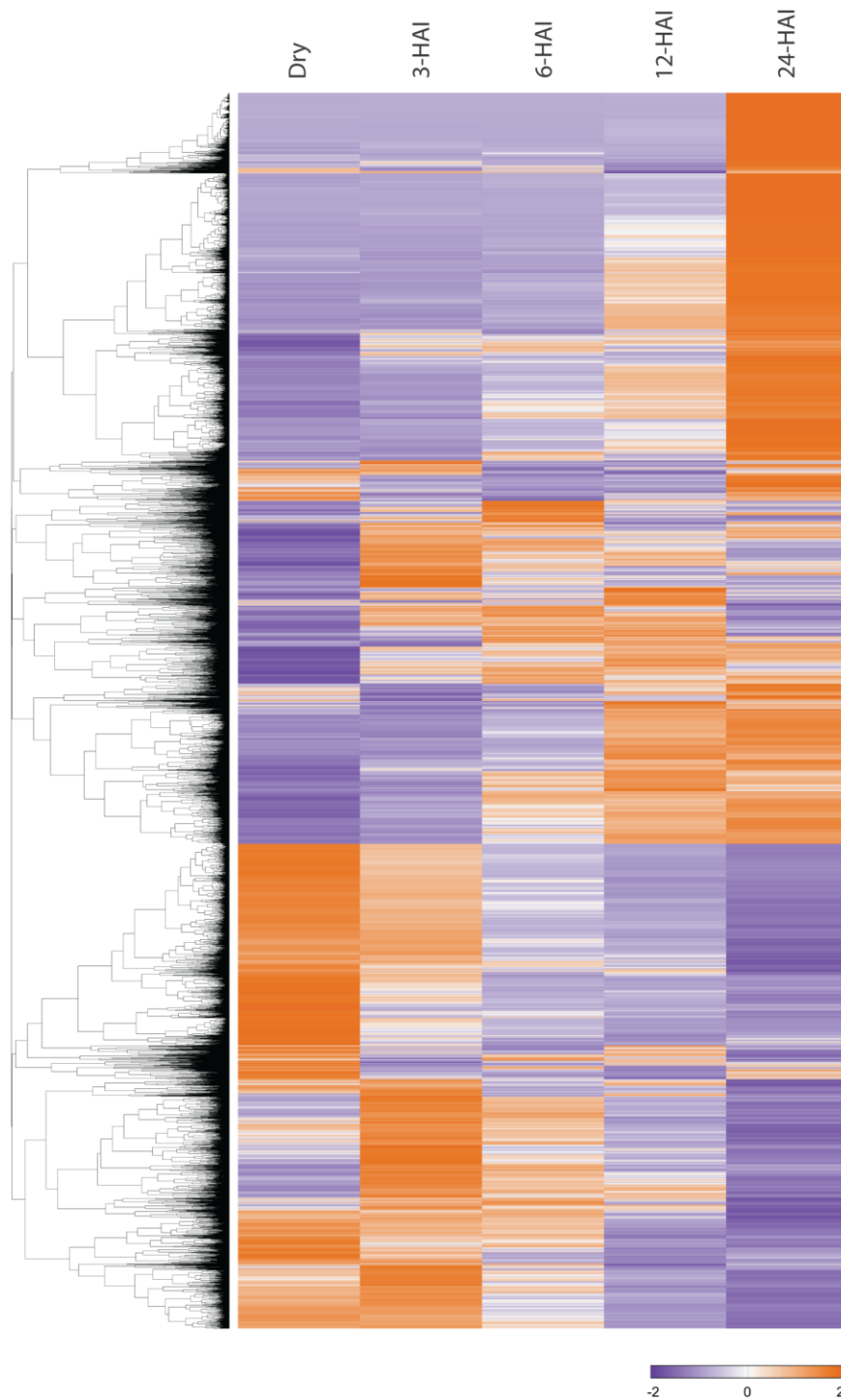


*Figura 15 Comparação da expressão gênica entre as amostras de semente 12 horas após embebição (12HAI) no eixo x e semente 24 horas após embebição (24HAI) no eixo y. Cada gene mapeado nas respectivas amostras é representado por um ponto no gráfico. Maior dispersão em relação à linha pontilhada indica menor correlação entre os perfis de expressão gênica.*

*Tabela 2 Distribuição de reads mapeados no genoma de referência da Soja: Estatísticas gerais do sequenciamento da germinação de eixos embrionários da soja. Cada linha da coluna representa um ponto na escala temporal da germinação, iniciando em 0 horas (semente seca) até 24 horas após embebição em água.*

Amostra	Reads não mapeados	Múltiplos alinhamentos	Alinhamentos únicos	Total de reads mapeados
Dry	21.888.018	85.630.708	157.064.044	242.694.752
3HAI	16.427.943	71.875.156	182.185.569	254.060.725
6HAI	19.798.969	87.809.774	211.697.839	299.507.613
12HAI	17.874.585	76.704.465	200.731.192	277.435.657
24HAI	24.183.233	97.381.607	275.170.244	372.551.851

\*HAI, hours after imbibition.



*Figura 16 Padrões globais de expressão gênica nos eixos embrionários de soja durante a germinação (embrião seca, 3HAI, 6HAI, 12HAI e 24HAI). Os padrões de expressão foram agrupados por HCL (hierarchical clustering). Apenas genes expressos (FPKM >1) estão representados. Os valores de expressão foram convertidos em z-score<sub>exc</sub>, e as intensidades de expressão variam entre -2 e 2. (HAI: Hours after imbibition)*

## **5.2 Identificação de genes preferencialmente expressos e análise da transição do perfil transcricional em 24 HAI.**

Embora a literatura clássica aborde de forma ampla aspectos morfológicos e bioquímicos relacionados à germinação de sementes, as transições subjacentes em seu perfil transcricional permanecem ainda pouco compreendidas. Os resultados iniciais sugerem que a amostra de 24 horas após embebição (HAI – *hours after imbibition*) se distancia consideravelmente das demais no que tange a seu perfil transcricional. Para melhor compreender tal diferença, foram analisados os conjuntos de genes preferencialmente expressos em cada estágio da germinação (Figura 17A-B). Com base em tal investigação viu-se que 99.6% do total de genes preferencialmente expressos (730 de 733) ocorrem em 24-HAI (Figura 17B). Intuitivamente este resultado poderia ser associado a uma transição morfológica de proporção igualmente drástica. Contudo, a medição de peso seco destes eixos embrionários (Figura 17C) revela que o aumento de volume observado nos eixos decorre apenas da assimilação de água pelas células nesse estágio (Bewley, Bradford et al. 2012). A regulação positiva na progressão da germinação segue a mesma tendência observada na análise de genes preferencialmente expressos (Figura 17D), recapitulando a divergência nos perfis transcricionais entre 12 e 24HAI. Os resultados evidenciam uma transição marcante no perfil transcricional dos eixos embrionários de soja ocorrida após as 12 HAI. Embora fosse intuitiva a associação deste evento a transições morfológicas significativas, tal consequência não é observada no peso seco dos eixos embrionários (Figura 17); o que está de acordo com o conhecimento atual sobre o mecanismo de crescimento do eixo embrionário. O crescimento do órgão ocorre sob ação majoritária da entrada água e expansão celular (Bewley, Bradford et al. 2012). Somado a isso, a presença

significativa de genes relacionados a fotossíntese (GO:0015979) no conjunto de genes preferencialmente expressos em 24HAI sugere que a transição do perfil transcricional neste período constitui uma fase de preparação para o estabelecimento da plântula e para seu crescimento autotrófico.

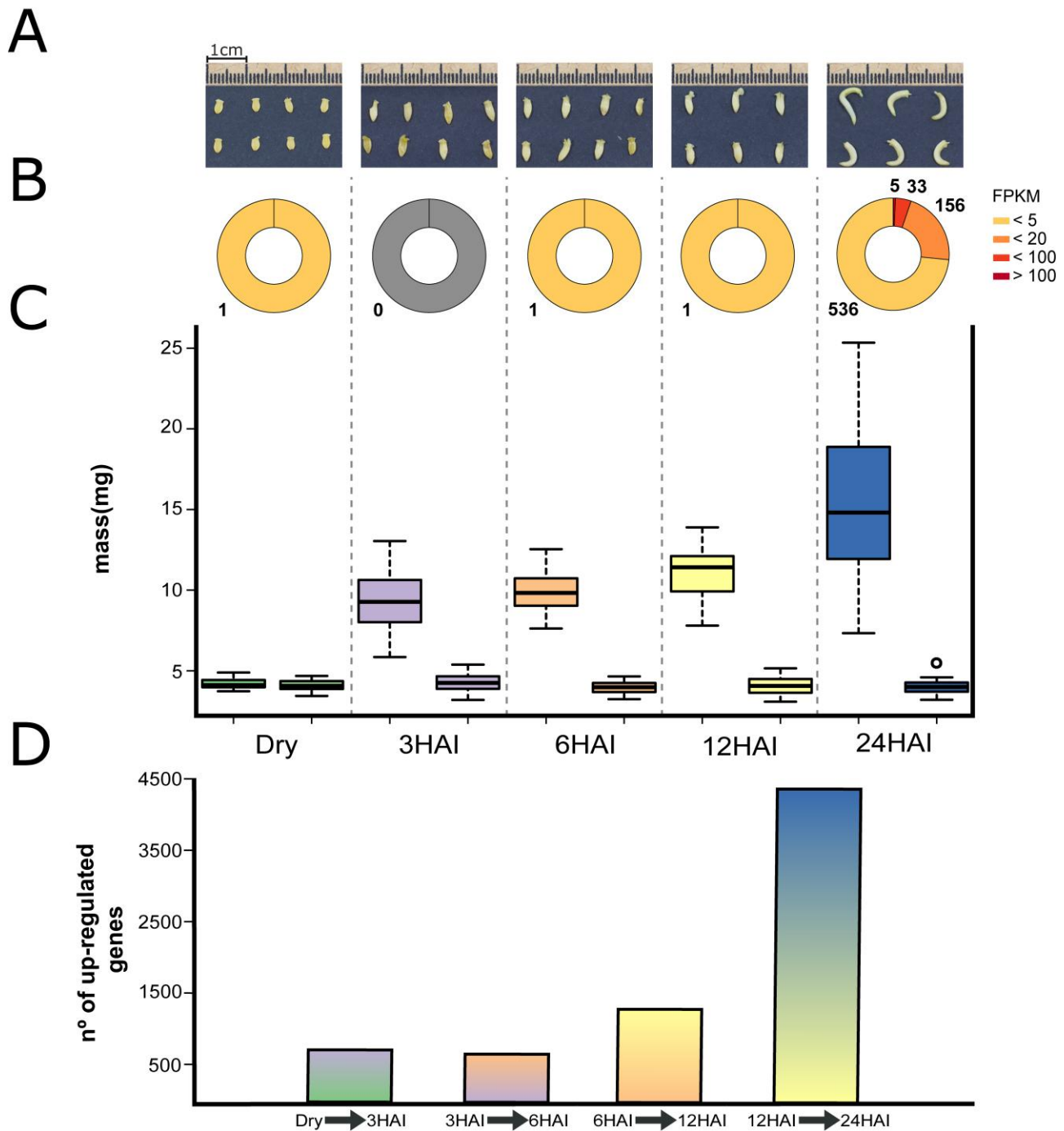


Figura 17 (A) Imagens do material biológico utilizado nos sequenciamentos. Da esquerda para a direita: embrião seco, 3HAI, 6HAI, 12HAI e 24HAI (B) Log<sub>2</sub> da contagem total de genes preferencialmente expressos no contexto da germinação em cada tempo (dry, 3HAI, 6HAI, 12HAI, 24HAI). Acima das barras, gráficos-donut mostram os valores absolutos das contagens de genes preferencialmente expressos separados por intervalo de intensidade de expressão (<5; <20; <100; >100). (C) Para cada amostra, peso seco e peso fresco dos eixos embrionários foram medidos. Os gráficos de caixa denotam a variação encontrada em 20 indivíduos organizados em duas raías: a primeira (esquerda) sendo peso fresco, e a segunda (direita) o peso seco. (D) As barras indicam as contagens de genes regulados positivamente em progressão nos intervalos do experimento.



### 5.3 Genes envolvidos com homeostase da parede celular, proteínas quinases e fatores de transcrição são regulados positivamente durante a germinação

Foram detectados 730, 674, 1307 e 4394 genes regulados positivamente nos intervalos eixo seco/3HAI, 3HAI/6HAI, 6HAI/12HAI, 12HAI/24HAI, respectivamente. Utilizando esse dado, foi conduzida uma análise de enriquecimento de termos GO com o objetivo de investigar as tendências mais conspícuas no curso na germinação. A análise de enriquecimento de termos GO apontou um aspecto marcante na germinação da soja: a regulação positiva preferencial de genes codificadores de proteínas quinases (GO:0004672) no intervalo eixo seco/3HAI (Tabela 3). Um dos genes da família das quinases (Glyma06g15270) codifica um receptor de brassinosteróides, homólogo ao ATBRI1 de *Arabidopsis thaliana* (At4g39400), e encontra-se regulado positivamente (4x) em 3HAI. Nesse contexto, foi demonstrado que a regulação de BR via ATBRI1 é capaz de promover a germinação, na medida em que atua como antagonista aos efeitos de ABA (Steber e McCourt 2001). Viu-se ainda que o gene Glyma06g15270 é positivamente regulado em presença de BR exógeno e é capaz de restaurar o fenótipo *dwarf mutant bri1-5* em *A. thaliana* (Wang, Sun et al. 2014). O gene Glyma12g22660 (também da família das quinases) apresenta níveis de expressão relativamente altos (FPKM >20); este gene é homólogo ao gene THESEUS1 de *A. thaliana* (THE1/At5g54380), que é induzido por BR e necessário para a ocorrência normal do processo de alongação celular (Guo, Li et al. 2009). THESEUS1 ainda atenua a inibição do crescimento do hipocótilo em alguns mutantes de parede celular (Hematy, Sado et al. 2007) e compõe um sistema de verificação na parede celular de células em processo de expansão (Hematy, Sado et al. 2007). Curiosamente, o gene Glyma12g22660 é regulado positivamente em 3HAI (eixo seco → 3HAI) e também em 24HAI (12HAI → 24HAI). Tal padrão de ativação

(em 3 e 24HAI) coincide com: a) a segunda fase de absorção de água pela semente; b) com a ativação massiva de genes relacionados à parede celular (Tabela 3). Assim sendo, os resultados sugerem que a sinalização por BR via BRI1 e THESEUS1 constitui um importante sistema responsável pela regulação da expansão celular durante a germinação. Outro gene da família das quinases fortemente induzido em 3HAI é o Glyma10g33240, um receptor de etileno cujo homólogo em *A. thaliana* atua em plântulas (Sakai, Hua et al. 1998; Kim, Helmbrecht et al. 2011) e em sementes sob estresse salino (Wilson, Kim et al. 2014). Um homólogo (Glyma14g39290) da TMK1 (*transmembrane kinase*) de *A. thaliana* (At1g66150) também integra este grupo; sabe-se atualmente que a TMK1 é induzida por GA (Cao, Cheng et al. 2006). Por último, foi observada a regulação positiva precoce do gene Glyma07g10695, que codifica uma proteína quinase e possui um homólogo em *A. thaliana* (At1g18390); este já descrito como atuante na via de sinalização de ABA (Lim, Yang et al. 2015). As observações relativas à família das quinases sugerem que as cascatas de sinalização envolvendo fosforilação, incluindo as vias de sinalização por fitormônios, desempenham função crucial nos estágios iniciais da germinação.

O único termo GO enriquecido em todos os estágios analisados foi GO:0016705, que é composto por diversas famílias de oxidoredutases, incluindo: 1) homólogos de ACCox, ACO1 e ACO4, que catalisam a última etapa da biossíntese do etileno. Vale ainda ressaltar que 5/6 genes ACO4 regulados ultrapassam a marca de 10 FPKM somente após 6HAI, indicando que o etileno seja mais relevante no estágio tardio da germinação; 2) 47 genes da família citocromo P450 (CYP), que são induzidos principalmente entre 12 e 24HAI; 3) GA oxidases (8 genes); 4) 28 oxoglutarato/ferro-dependente dioxigenases (Tabela 3). Além disso, encontrou-se ainda um termo GO relacionado (GO:0055114 – processo de oxido-redução) enriquecido entre genes induzidos em 6, 12 e 24HAI, que inclui: aldeído desidrogenases,

lipoxigenases e peroxidases (Tabela 3). Embora distintas evolutiva e estruturalmente, todas estas famílias estão associadas ao metabolismo e homeostase da parede celular (Wang, Ji et al. 2009; Nelson, Flematti et al. 2010; Daniel, Pavkov-Keller et al. 2015).

Um notável enriquecimento de termos GO relacionados a fotossíntese (GO:0009579 – tilacóide; GO:0015995 - biossíntese de clorofila) também foi observado (Tabela 3). Baseado nisso, consolidamos estes termos sob o termo geral relacionado à fotossíntese (GO:0015979). Este grupo amplo compreende várias subunidades de fotossistemas I e II, incluindo genes codificadores de proteínas do complexo antena e de proteínas de ligação à clorofila (e.g. Glyma01g28810 e Glyma05g25810, respectivamente).

O aspecto mais marcante presente no conjunto de genes regulados positivamente logo após a embebição (eixo seco/3HAI) é a presença massiva de reguladores transcricionais DNA-dependentes (GO:0006355) (Tabela 3) sendo a vasta maioria destes classificados como TFs no banco de dados PlantTFDB (Jin, Zhang et al. 2014). Com o objetivo de aprofundar a análise de TFs, observou-se que 53% (2000/3717) de todos os TFs de soja são expressos em algum momento da germinação. Destes, 432 são regulados positivamente em pelo menos um intervalo de tempo analisado. De fato, 3HAI é o único estágio que apresenta enriquecimento de TFs (81 genes;  $p\text{-value} = 8 \times 10^{-8}$ ; Teste Exato de Fisher), confirmando os resultados obtidos por enriquecimento de termos GO. Notamos ainda a grande quantidade relativa de genes da família LBD, que constitui 1/9 do total de TFs regulados positivamente em 3HAI. A presença significativa de TFs da família LBD (LOB domain) (Shuai, Reynaga-Pena et al. 2002) no estágio inicial da germinação (eixo seco-3HAI) é coerente com a discussão anterior acerca da sinalização por BR, uma vez que BR induz a expressão de fatores de transcrição LBD em *A. thaliana*. O LBD, por sua vez, reprime a acumulação de BRs por meio da regulação da expressão da enzima BAS1 (Bell, Lin et al.

2012). Assim sendo, LBDs são parte de um mecanismo de retroalimentação negativo envolvido na homeostase de BR (Ryu e Hwang 2013). Contudo, devido aos baixos níveis de expressão observados para os genes homólogos a BAS1, é provável que este mecanismo esteja inativo nos eixos embrionários de soja durante a germinação. Adicionalmente, sabe-se que LBDs promovem a transcrição de uma expansina em *A. thaliana* e, presumivelmente, outros genes associados ao afrouxamento da parede celular (Lee, Kim et al. 2013). Com isso, sugere-se que TFs LBD sejam integrantes de um sistema BAS1-independente responsável pela regulação da remodelação da parede celular durante a germinação em soja.

Tabela 3 Enriquecimento de Termos GO (Gene Ontology) entre os genes diferencialmente expressos durante a germinação.

GO Term	Level	Description	Dry-3HAI	3-6HAI	6-12HAI	12-24HAI	FDR (dry-3HAI)	Num (dry-3HAI)	FDR (3-6HAI)	Num (3-6HAI)	FDR (6-12HAI)	Num (6-12HAI)	FDR (12-24HAI)	Num (12-24HAI)
GO:0005576	C	extracellular region	■			■	3,7E-9	17					1,7E-12	44
GO:0005618	C	cell wall	■	■		■	5,3E-7	14	5,4E-3	9			1,2E-12	43
GO:0048046	C	apoplast	■	■		■	5,3E-7	10	6,4E-3	6			6,1E-9	23
GO:0006073	P	cellular glucan metabolic process	■			■	3,2E-6	15					8,8E-10	43
GO:0006355	P	regulation of transcription, DNA-dependent	■				3,2E-6	67						
GO:0045449	P	regulation of transcription	■				3,2E-6	67						
GO:0030312	C	external encapsulating structure	■			■	3,7E-6	14					3,3E-10	44
GO:0016762	F	xyloglucan:xyloglucosyl transferase activity	■			■	5,6E-6	10					2,6E-8	23
GO:0019222	P	regulation of metabolic process	■				8,3E-6	68						
GO:0065007	P	biological regulation	■				3,0E-5	90						
GO:0006350	P	transcription	■				3,2E-5	69						
GO:0050794	P	regulation of cellular process	■				3,4E-5	86						
GO:0050789	P	regulation of biological process	■				6,2E-5	86						
GO:0016758	F	transferase activity, transferring hexosyl groups	■			■	1,0E-4	28					1,2E-5	99
GO:0016740	F	transferase activity	■				1,1E-4	119						
GO:0016757	F	transferase activity, transferring glycosyl groups	■			■	1,1E-4	32					2,1E-4	114
GO:0031012	C	extracellular matrix	■			■	1,3E-4	7					4,5E-3	13
GO:0044421	C	extracellular region part	■			■	1,3E-4	7					4,5E-3	13
GO:0003700	F	transcription factor activity	■				1,0E-3	35						
GO:0016747	F	transferase activity, transferring acyl groups other than amino-acyl groups	■				1,3E-3	18						
GO:0005976	P	polysaccharide metabolic process	■			■	1,3E-3	17					7,9E-9	74
GO:0016746	F	transferase activity, transferring acyl groups	■				1,4E-3	19						
GO:0016705	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	■			■	2,4E-3	22	7,2E-3	21	5,3E-3	34	2,2E-4	85
GO:0004672	F	protein kinase activity	■				3,5E-3	54						
GO:0030528	F	transcription regulator activity	■				3,5E-3	46						
GO:0016070	P	RNA metabolic process	■				8,8E-3	70						
GO:0004866	F	endopeptidase inhibitor activity	■				9,6E-3	6						
GO:0005975	P	carbohydrate metabolic process		■	■	■			5,2E-7	55	1,4E-7	90	3,6E-11	224
GO:0044282	P	small molecule catabolic process		■					1,1E-5	18				
GO:0044262	P	cellular carbohydrate metabolic process		■		■			3,4E-5	34			2,9E-5	121
GO:0006096	P	glycolysis		■					5,2E-5	12				
GO:0055114	P	oxidation reduction		■	■	■			3,6E-4	63	1,5E-4	113	5,4E-20	365
GO:0046164	P	alcohol catabolic process		■					4,9E-4	13				
GO:0006546	P	glycine catabolic process		■		■			5,4E-4	5			7,8E-3	8
GO:0016052	P	carbohydrate catabolic process		■					1,7E-3	13				
GO:0005996	P	monosaccharide metabolic process		■					2,0E-3	14				
GO:0004553	F	hydrolase activity, hydrolyzing O-glycosyl compounds		■		■			2,3E-3	24	8,4E-13	58	4,8E-13	115
GO:0044248	P	cellular catabolic process		■					2,8E-3	23				
GO:0016798	F	hydrolase activity, acting on glycosyl bonds		■		■			3,4E-3	24	7,0E-13	61	5,9E-13	120
GO:0006091	P	generation of precursor metabolites and energy		■		■			4,1E-3	14			2,0E-4	51
GO:0005829	C	cytosol		■					5,4E-3	7				
GO:0044445	C	cytosolic part		■					5,4E-3	7				
GO:0016491	F	oxidoreductase activity		■		■			6,1E-3	67	1,4E-3	121	4,6E-22	425

GO:0016706	F	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors						7,1E-3	12				
GO:0004857	F	enzyme inhibitor activity						7,2E-3	10			2,8E-6	37
GO:0009310	P	amine catabolic process						7,6E-3	5				
GO:0015995	P	chlorophyll biosynthetic process								2,4E-8	12	5,2E-6	14
GO:0015979	P	photosynthesis								4,7E-8	28	1,1E-22	74
GO:0015994	P	chlorophyll metabolic process								6,5E-8	12	6,1E-6	15
GO:0016851	F	magnesium chelatase activity								8,1E-8	11	4,3E-7	15
GO:0051002	F	ligase activity, forming nitrogen-metal bonds								8,1E-8	11	4,3E-7	15
GO:0070008	F	serine-type exopeptidase activity								4,0E-7	14	1,3E-7	23
GO:0004180	F	carboxypeptidase activity								1,3E-6	14	9,4E-7	23
GO:0003824	F	catalytic activity								1,8E-4	537	4,3E-8	1565
GO:0008238	F	exopeptidase activity								1,9E-4	15	7,4E-4	26
GO:0042440	P	pigment metabolic process								3,5E-4	12		
GO:0005506	F	iron ion binding								4,1E-4	37	4,3E-10	103
GO:0020037	F	heme binding								6,1E-4	32	4,6E-8	84
GO:0009341	C	beta-galactosidase complex								6,3E-4	7		
GO:0009579	C	thylakoid								6,3E-4	16	2,4E-11	42
GO:0034357	C	photosynthetic membrane								6,3E-4	15	3,3E-10	39
GO:0046906	F	tetrapyrrole binding								6,9E-4	32	7,3E-8	84
GO:0004565	F	beta-galactosidase activity								7,5E-4	7		
GO:0015925	F	galactosidase activity								7,5E-4	7		
GO:0033013	P	tetrapyrrole metabolic process								1,2E-3	13		
GO:0016787	F	hydrolase activity								1,6E-3	185	1,1E-5	523
GO:0018130	P	heterocycle biosynthetic process								3,9E-3	17		
GO:0005507	F	copper ion binding										4,5E-20	58
GO:0016614	F	oxidoreductase activity, acting on CH-OH group of donors										1,0E-12	98
GO:0000785	C	chromatin										1,2E-12	41
GO:0032993	C	protein-DNA complex										1,2E-12	41
GO:0009765	P	photosynthesis, light harvesting										2,0E-12	23
GO:0016616	F	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor										5,0E-12	94
GO:0019684	P	photosynthesis, light reaction										9,6E-12	27
GO:0034622	P	cellular macromolecular complex assembly										3,6E-11	62
GO:0006323	P	DNA packaging										4,4E-10	41
GO:0071103	P	DNA conformation change										6,8E-9	43
GO:0005694	C	chromosome										7,2E-9	51
GO:0034621	P	cellular macromolecular complex subunit organization										1,0E-8	64
GO:0006629	P	lipid metabolic process										1,3E-8	155
GO:0004601	F	peroxidase activity										1,8E-8	41
GO:0016684	F	oxidoreductase activity, acting on peroxide as acceptor										1,8E-8	41
GO:0048037	F	cofactor binding										1,8E-8	154
GO:0050662	F	coenzyme binding										1,0E-7	120
GO:0006979	P	response to oxidative stress										2,1E-7	37
GO:0005509	F	calcium ion binding										4,2E-7	37
GO:0016209	F	antioxidant activity										7,3E-7	47
GO:0030599	F	pectinesterase activity										9,8E-7	25



## 5.4 Ativação transcricional de genes do metabolismo primário durante a germinação

### 5.4.1 Metabolismo de carboidratos

Numa perspectiva ampla, os resultados corroboram a noção de que sementes nos estágios iniciais de germinação abrigam duas subpopulações de mRNAs distintas: transcritos que tiveram função crucial no estágio tardio da maturação da semente (antes da germinação); e transcritos expressos *de novo* na semente após a embebição que serão necessários aos eventos metabólicos essenciais ao início da germinação (Figura 18).

No contexto da reativação do metabolismo primário, sabe-se que sacarose e oligossacarídeos da família das rafinoses (RFO – *raffinose family oligosaccharides*) são os principais carboidratos solúveis disponíveis em sementes de soja (Kuo, Doehlert et al. 1990) e os principais acumulados após o processo de dessecação (Obendorf, Zimmerman et al. 2009). As concentrações de RFO declinam consideravelmente durante a germinação, em concomitância com a perda de tolerância à dessecação e o aumento nos níveis de açúcares redutores (Koster e Leopold 1988). Em termos de produção de energia, as invertases (EC 3.2.1.26; sacarose → frutose + glicose) são as maiores responsáveis pelo consumo de sacarose. Identificamos a expressão de 10 invertases expressas, das quais 6 foram induzidas em 24HAI; 4 destas apresentaram níveis de expressão relativamente altos (FPKM >10), com picos de expressão ocorrendo em 24HAI (Figura 19; Tabela 4). Baseado em tais observações, os resultados apontam para uma demanda crescente de monossacarídeos derivados da hidrólise da sacarose na germinação tardia e, presumivelmente, nos estágios iniciais do estabelecimento da plântula. Enquanto invertases são frequentemente associadas a condições de alta demanda energética (e.g. expansão celular), a atividade da sacarose



sintase (SuSy; EC 2.4.1.13; sacarose + UDP  $\leftrightarrow$  UDP-glicose + frutose) tem sido associada à biossíntese de polissacarídeos estruturais ou de reserva (Winter e Huber 2000). Embora reversível, a atividade de SuSy favorece a quebra da sacarose na maioria dos tecidos heterotróficos, disponibilizando precursores como UDP-glicose para a biossíntese de polissacarídeos estruturais e de reserva (Winter e Huber 2000; Salerno e Curatti 2003). A sacarose sintase é capaz de associar-se a membranas e prover um canal de UDP-glicose do citoplasma até os complexos celulose sintase (Amor, Haigler et al. 1995; Winter e Huber 2000; Haigler, Ivanova-Datcheva et al. 2001), que interagem com os microtúbulos e direcionam a deposição de celulose e alongação celular (Paredes, Somerville et al. 2006). Foi observado que 14, dos 19 genes que codificam celulose sintases, são regulados positivamente em 24HAI (Tabela 3). Apesar dos altos níveis de expressão de invertases e SuSys na germinação, sabe-se que as concentrações de sacarose não sofrem redução ao longo da germinação (Koster e Leopold 1988; Kuo, Doehlert et al. 1990). Com base nisto, os resultados apontam para uma constante renovação do conteúdo de sacarose no eixo embrionário, sugerindo um papel central para o consumo da sacarose durante a germinação em soja; o que é corroborado também pela ativação precoce de genes das vias da glicólise e respiração celular, que podem ser alimentadas pela degradação da sacarose.

Foi observada também a ativação precoce das vias da glicólise e ciclo de Krebs, que podem ser parcialmente alimentadas pela degradação de glicose. Entre os genes expressos nestas vias, estão: 12 hexoquinases; 12 fosfofrutoquinases (PFK); 13 gliceraldeído 3-fosfato desidrogenases (GAPDH); triosefosfato isomerases (TIM) e piruvato quinases (PKs) (Figura 19 e Tabela 4). O gene At2g21170 de *A. thaliana* codifica uma proteína TIM (anotada como TPI) que é necessária à transição do crescimento heterotrófico para o autotrófico nos estágios pós-germinativos. Mutações nesse gene resultam em acumulação deletéria de

diidroxicetona fosfato e desenvolvimento anormal do cloroplasto (Chen e Thelen 2010). No presente trabalho, foram encontrados 4 ortólogos do gene At2g21170 em soja, dos quais Glyma19g36990 e Glyma03g34300 são expressos em níveis relativamente altos na germinação, sugerindo que estes sejam igualmente críticos na transição para a modalidade autotrófica de crescimento na soja.

A glicose pode ainda ser consumida pela via das pentoses fosfato, num processo intimamente dependente da atividade da enzima glicose-6-fosfato desidrogenase (G6PDH). Os dados revelaram a presença de 8 G6PDHs (Figura 19 e Tabela 4). Foi observada também a expressão de 3 frutose-1,6-bifosfatases (FBPase), 7 PFKs dependentes de pirofosfato (PPF) (Figura 19 e Tabela 4). No que concerne à produção de Acetil-CoA, foi observada a expressão de 5 genes codificadores da subunidade E2 do complexo piruvato desidrogenase (PDC – *pyruvate dehydrogenase complex*), 19 genes da subunidade E1 e 7 genes da subunidade E3. Sendo o ciclo do ácido cítrico um dos principais mecanismos de consumo de Acetil-CoA, viu-se que 13 genes que codificam citrato sintases (CSY – *citrate synthase*), 6 genes da isocitrato desidrogenase (IDH – *isocitrate dehydrogenase*) e 8 genes da malato desidrogenase (MDH – *malate dehydrogenase*) apresentam níveis relativamente altos de expressão durante a germinação (Figura 19 e Tabela 4).

Embora a semente possua mecanismos de rápida incorporação de água durante a germinação, a disponibilidade de oxigênio em seu interior é comumente baixa (Rosental, Nonogaki et al. 2014). Tal circunstância exige a rápida ativação de genes associados à fermentação. Os dados confirmam a expressão em altos níveis (FPKM >100) de piruvato decarboxilases (PDC – *pyruvate decarboxylase*), álcool desidrogenases (ADH – *alcohol dehydrogenase*), e lactato desidrogenases (LDH – *lactate dehydrogenase*) (Figura 19 e Tabela 4).

## 5.4.2 Metabolismo de lipídios

Triacilgliceróis (TAGs – *triacylglycerols*) armazenados em corpos lipídicos estão comumente localizados próximos aos glioxissomos, organelas especializadas em converter ácidos graxos derivados dos TAGs em açúcares solúveis via ciclo do glioxilato (Graham 2008; Kelly, Quettier et al. 2011). Embora os ciclos do glioxilato e do ácido cítrico compartilhem várias enzimas, o primeiro contorna as etapas de decarboxilação do segundo pela ação de duas enzimas: ICL, responsável pela quebra do isocitrato em glioxilato e succinato; e MSY, que condensa glioxilato e acetil-CoA em malato. O malato é então convertido em oxaloacetato pela MDH, que é subsequentemente convertido a fosfoenolpiruvato (PEP) pela PEPCK (Figura 19; Tabela 4) (Graham 2008). Curiosamente, os níveis de expressão das ICL são moderados (FPKM < 5), enquanto os níveis de expressão de MSYs são relativamente altos (FPKM > 30) (Figura 19; Tabela 4). MDH e PEPCK também são expressos durante a germinação, sendo que 3 genes MDH e 3 PEPCK atingem seu valor máximo de expressão em 24HAI (Figura 19; Tabela 4). Soma-se a isso os genes expressos de TAG-lipases e COMATOSE, corroborando a ocorrência de transporte de acil-CoA até os peroxissomos/glioxissomos. A mobilização de TAGs gera também quantidades relevantes de glicerol, que pode ser utilizado pela glicólise na forma de dihidroxicetona fosfato (DHAP), por meio da lançadeira glicerol-3-fosfato, que por sua vez depende da ação das enzimas glicerol quinase e glicerol-3-fosfato desidrogenase (G3PD) (Weitbrecht, Muller et al. 2011); ambos os genes são expressos durante a germinação (Figura 19; Tabela 4). É amplamente aceito que a mobilização de reservas lipídicas é crítica nos estágios pós-germinativos (Fait, Angelovici et al. 2006; Graham 2008). Contudo, alguns estudos tem demonstrado que a  $\beta$ -oxidação e o ciclo do glioxilato determinam o potencial da germinação em diversas condições (Pinfield-Wells, Rylott et al. 2005; Pracharoenwattana, Cornah et al.

2005). Os resultados do presente estudo confirmam a ativação precoce de genes ligados à mobilização de lipídios na germinação dos eixos embrionários de soja, corroborando a recente constatação de que tanto proteínas de reserva quanto corpos lipídicos localizados nas células do embrião são catabolizados localmente durante a germinação (Rosental, Nonogaki et al. 2014).

### 5.4.3 Metabolismo de aminoácidos

Proteínas de reserva são fundamentais como fonte de aminoácidos para síntese proteica e produção de energia, particularmente glutamato e aspartato, que são os aminoácidos mais abundantes nas sementes de soja (FAOSTAT 2016). Ambos (glutamato e aspartato) são substratos das enzimas AlaAT e AspAT, que são ativadas durante a embebição e provavelmente atuam de forma importante nas vias respiratórias (Miyashita, Dolferus et al. 2007; Weitbrecht, Muller et al. 2011). Evidências recentes ressaltam aspectos importantes da acumulação de aspartato na germinação, como: a influência negativa da presença de ABA sobre as concentrações de AspAT, ou o rápido aumento nos níveis de transcrição de AspAT após a embebição (Rosental, Nonogaki et al. 2014). O aminoácido não-proteico GABA também possui importante função no metabolismo energético das sementes durante a germinação (Weitbrecht, Muller et al. 2011; Rosental, Nonogaki et al. 2014). Níveis altos de GABA são metabolizados durante a germinação em *A. thaliana* (Fait, Angelovici et al. 2006), e castanha (Desmaison e Tixier 1986). Além disso, a acumulação de GABA em sementes durante a germinação está relacionada à atividade de GAD (Guo, Yang et al. 2012). GAD sintetiza GABA a partir de glutamato, e constitui a primeira enzima do sistema GABA. GABA é então metabolizado por GABA-T, que gera semi-aldeído succínico que

participará do ciclo do ácido cítrico na forma de succinato (Figura 19; Tabela 4). A expressão de GAD e GABA-T evidencia a atividade do sistema GABA durante a germinação de soja.

Embora seja bem estabelecido que a maior parte dos aminoácidos usados na síntese de novo de proteínas sejam provenientes da degradação de proteínas de reserva (Rosental, Nonogaki et al. 2014), nossos resultados fornecem evidências transcricionais para a ativação de algumas vias de biossíntese de aminoácidos durante a germinação em soja. Uma das etapas fundamentais na biossíntese de aminoácidos aromáticos (i.e. fenilalanina, tirosina e triptofano) é a produção de corismato, que ocorre pela via do shiquimato. Esta se inicia com uma reação catalisada pela enzima 3-deoxi-7-fosfoeptulonato sintase (DAHP sintase), que consiste na conversão de PEP e eritrose-4-fosfato (produzida pela via das pentoses fosfato) em 3-deoxi-D-arabino-hept-2-ulosonato 7-fosfato (Tzin e Galili 2010). A DAHP sintase conecta o metabolismo de carboidratos à via do shiquimato, que irá em última instância, produzir corismato, um metabólito central para a biossíntese de aminoácidos aromáticos, salicilato e folato (Tzin e Galili 2010). Os níveis observados de expressão de DAHP sintase (FPKM > 37) (Material Suplementar) suportam a utilização de carbono proveniente da via das pentoses fosfato para a via do shiquimato no eixo embrionário. Embora os genes da via do shiquimato sejam provenientes do DNA nuclear, os processos bioquímicos a ela relativos (a exemplo do que ocorre em outras vias biossintéticas) ocorrem nos plastídeos, que são notadamente organelas majoritariamente não-funcionais em sementes secas. Contudo, um estudo recente demonstrou que a biogênese de plastídeos é disparada na fase inicial da germinação de soja (Yu, Li et al. 2015). Assim sendo, especula-se que plastídeos recém-formados são prontamente equipados com a via do shiquimato, e, presumivelmente, de outras vias biossintéticas.

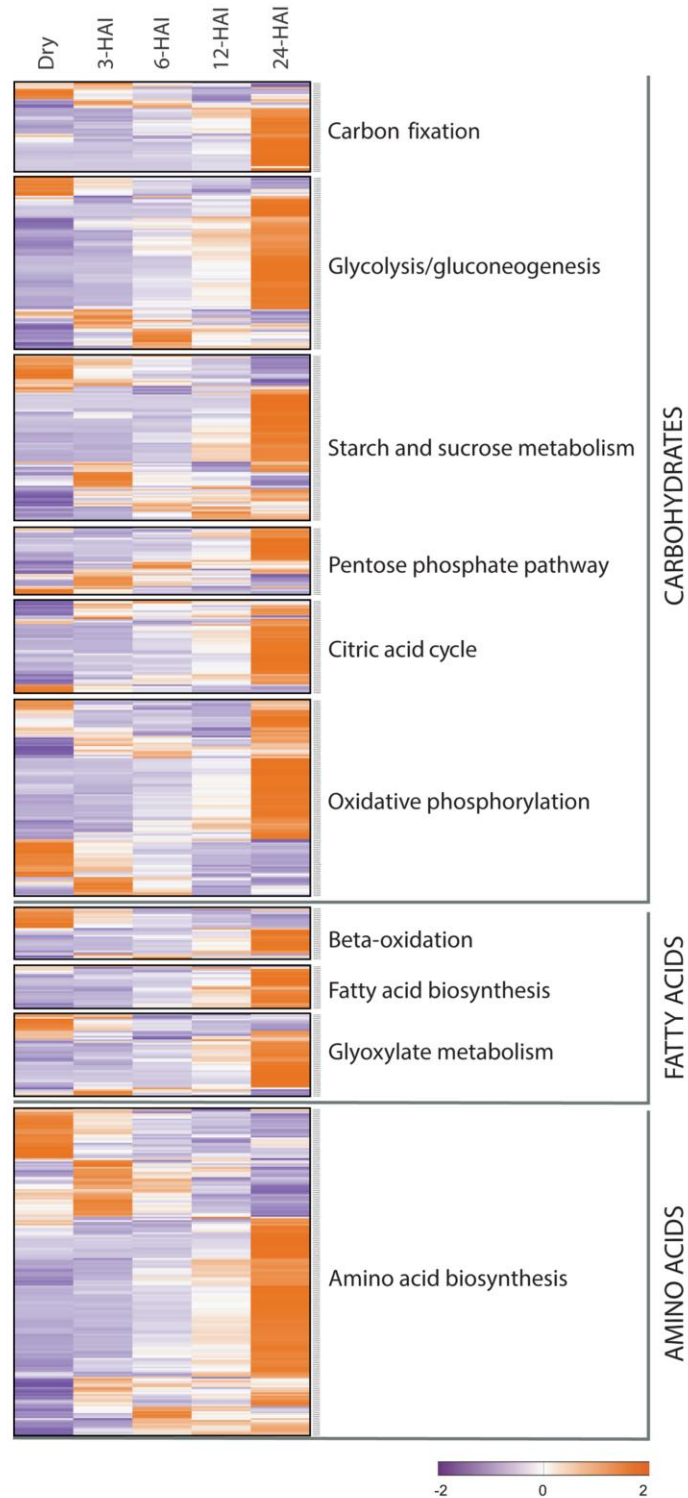


Figura 18 Mapa de calor de todos os genes associados com vias do metabolismo primário (baseado em anotações do banco de dados KEGG) expressos nos eixos de sementes de soja durante a germinação. As colunas representam as amostras analisadas no estudo, da esquerda para a direita (eixo seco, 3HAI, 6HAI, 12HAI e 24HAI). Os valores de expressão em FPKM foram convertidos em z-score e a intensidade de cor varia de -2 a 2

*Figura 19 Principais vias metabólicas e enzimas discutidas no estudo. Níveis transcricionais de genes-chave estão representados nos mapas de calor. As famílias representadas estão dispostas numericamente seguindo a ordem: 1) Invertase; 2) Sacarose sintase (SuSy); 3) Celulose sintase; 4) Hexoquinase; 5) glicose-6-fosfato desidrogenase; 6) fosfofrutoquinase (PFK); 7) Frutose 1,6-bifosfato (FBPase); 8) Difosfato-frutose-6-fosfato 1-fosfofrutoquinase (PFP); 9) Piruvato quinase (PK); 10) Piruvato decarboxilase; 11) Álcool desidrogenase (ADH); 12) Lactato desidrogenase (LDH); 13) Piruvato desidrogenase (PDH-E2); 14) Citrato sintase (CSY); 15) Isocitrato desidrogenase (IDH); 16) Isocitrato liase (ICL); 17) Malato sintase (MSY); 18) malato desidrogenase (MDH); 19) Fosfoenolpiruvato carboxiquinase (PEPCK); 20) Triacilglicerol (TAG) lipase; 21) Glicerol quinase; 22) Glicerol-3-fosfato desidrogenase (GPDHC1); 23) COMATOSE (CTS); 24) Glutamato decarboxilase (GAD); 25) Gamma-amino-N-butirato transaminase (GABA-T). As 5 colunas em cada mapa de calor representam os tempos analisados no experimento. Da esquerda para a direita: eixo seco, 3HAI, 6HAI, 12HAI e 24HAI. (HAI – hours after imbibition).*

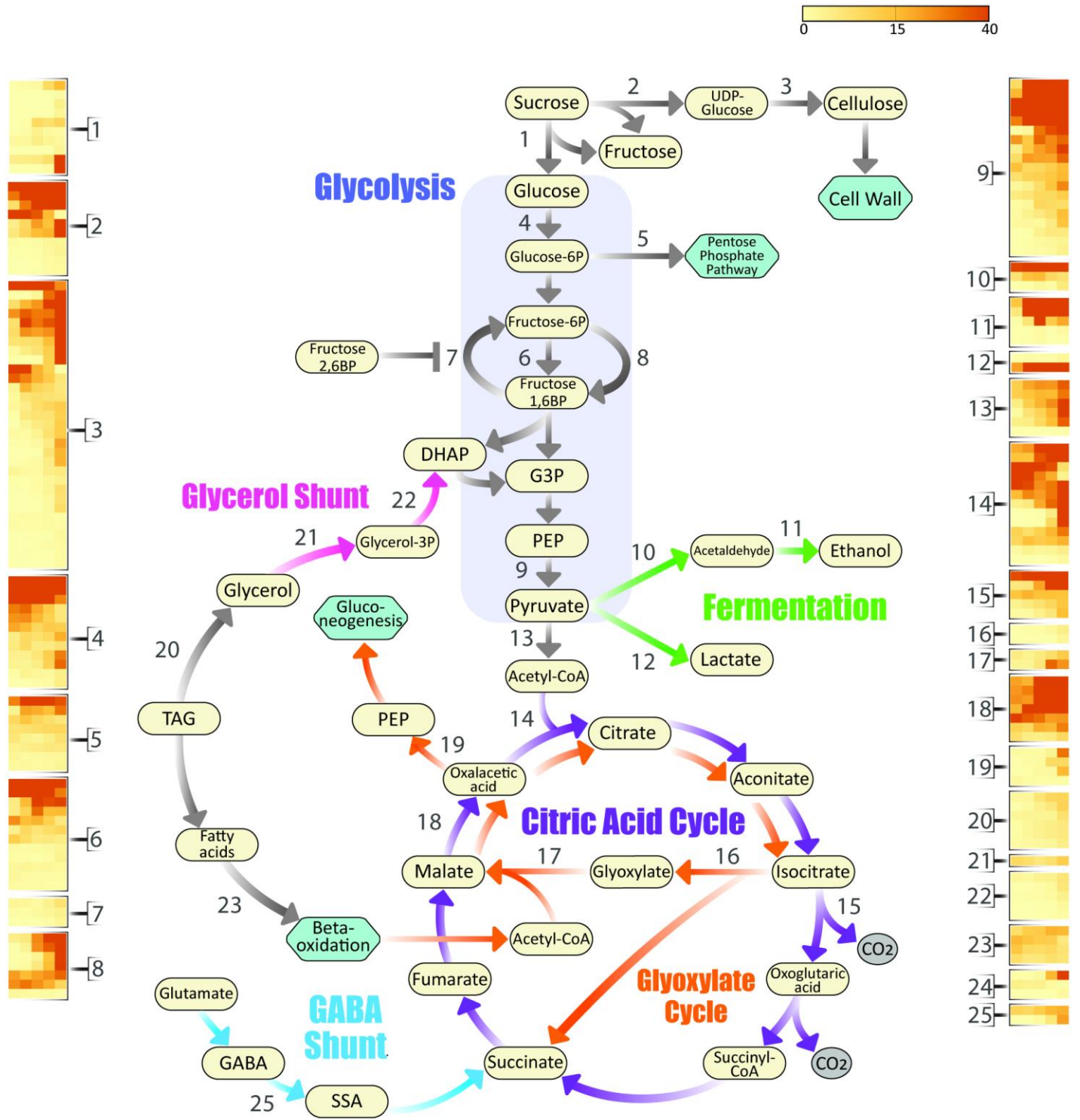




Tabela 4 Níveis transcricionais (em FPKM) de enzimas representadas na figura 19.

INVERTASE (1)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma01g41990	0,155884	1,12449	1,62902	8,92596	15,5262
Glyma11g03360	0	0,0530837	0,0232149	0,136483	1,00758
Glyma17g34590	0	0	0	0,030462	4,26399
Glyma07g01090	0,0536806	0,0278757	0,0433867	0,0543608	2,6726
Glyma08g20490	0,0504611	0,618056	9,0339	7,73426	10,1626
Glyma15g02850	0,341239	1,03192	3,46846	1,74293	1,34562
Glyma13g42530	2,06772	2,18439	4,71715	3,02722	1,96884
Glyma17g34570	0	0,00666405	0,00453551	0,065572	2,98891
Glyma05g04290	4,4419	4,51021	5,57004	7,64027	39,9113
Glyma17g14750	0	0,0170921	0,0300085	1,33284	54,5588
CELLULOSE SYNTHASE (3)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma01g01780	0,492254	1,04942	1,1117	1,30649	0,759724
Glyma01g44281	0,0775137	2,77317	5,02345	6,52205	12,3545
Glyma02g36720	34,1165	27,0126	17,6597	12,8089	6,50624
Glyma02g47076	0,80353	0,468891	0,656896	1,26107	0,863006
Glyma02g08920	0,201918	0,715104	1,27208	4,66229	17,3219
Glyma03g37550	0,112424	0,147188	0,095656	0,0747455	1,31579
Glyma04g07220	24,2027	33,9149	28,7214	29,5576	49,4098
Glyma05g26840	45,3758	48,6748	17,72	10,9514	0,753654
Glyma05g32100	7,97539	16,5844	23,4704	41,0629	99,0907
Glyma06g48260	0,312207	0,242082	2,64001	30,7207	161,552
Glyma06g07320	8,38256	25,0896	21,319	24,5299	69,0892
Glyma08g15380	6,09318	7,73694	8,24048	15,3087	53,6418
Glyma08g44310	0,0515237	0,173173	0,288434	0,337477	2,13671
Glyma08g44320	0,176157	0,0878046	0,762985	2,74484	7,50017
Glyma09g15620	1,77774	4,77594	6,94589	11,8919	49,0233
Glyma09g05630	0,00851908	0,0176714	0,00899012	0,110118	1,31197
Glyma10g33300	2,77818	1,59526	1,50391	4,1785	5,34496
Glyma10g36790	0,0442903	0,0476379	0,275008	1,83065	7,2903
Glyma11g21190	0,492992	0,34879	0,4714	11,2203	68,4833
Glyma11g01230	0,387184	5,84369	7,11983	7,00537	11,5881
Glyma12g31780	0,0396004	0,0482721	0,307761	6,99803	16,9019
Glyma12g31800	0,0115924	0,0372033	0,0603908	1,69125	8,38098
Glyma12g31810	1,30532	1,15758	3,34915	7,46585	11,9354
Glyma12g31830	0,426075	0,320623	0,961263	2,49414	3,82875
Glyma12g31840	0,132605	0,0634979	0,266449	0,872274	1,52619
Glyma12g36570	2,27378	4,09062	8,13517	9,87424	11,9851
Glyma13g27250	1,86152	6,60866	7,9182	7,02709	6,71206
Glyma15g43040	7,52463	40,7897	29,8838	35,8811	67,5907
Glyma16g20880	0,159955	0,500015	1,62236	4,70453	20,4997
Glyma17g08000	311,249	491,408	301,404	130,459	25,3499
Glyma19g40170	3,03122	17,2591	10,423	8,44823	5,86922
FBPase (7)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma07g30110	0,835989	0,630106	1,72967	2,64708	3,78463
Glyma18g41940	0,0375527	0,285854	0,430856	0,923718	2,75236
Glyma16g28310	2,7166	2,47062	1,34736	3,60678	6,9798
SUCROSE SYNTHASE (SuSy) (2)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma17g05067	29,6475	64,1092	116,842	123,544	92,4288
Glyma03g37441	0,910833	0,689262	0,952889	2,40451	2,66541
Glyma09g29710	0,310476	0,28708	0,510622	2,07374	3,22134
Glyma14g39070	0,0324188	0,0815828	0,307381	1,18996	1,84901
Glyma13g17421	8,13792	28,5553	119,594	130,715	82,7969
Glyma15g20180	5,9562	10,1645	15,679	17,7673	43,6757
Glyma09g08550	3,99209	9,48789	22,333	14,726	48,5632
Glyma03g37441	0,910833	0,689262	0,952889	2,40451	2,66541
Glyma19g40041	93,1219	95,528	152,245	115,634	46,8237
Glyma15g16171	66,7099	56,9943	35,3832	15,9227	1,35517
Glyma02g40740	0,0820006	0,0463469	0,1745	1,14787	1,60445
HEXOKINASE (4)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma05g35890	3,09558	2,82864	4,52143	11,0628	32,066
Glyma08g03730	3,76453	3,68847	6,29555	11,6884	28,2398
Glyma07g12190	19,8456	16,0456	9,60189	12,2472	11,7058
Glyma01g01060	241,433	176,933	146,483	123,824	61,6367
Glyma01g01070	241,433	176,933	146,483	123,824	61,6367
Glyma08g21450	186,432	168,793	95,9771	75,093	26,7513
Glyma01g43650	24,0927	25,4417	16,1513	11,0362	6,36887
Glyma17g37720	4,62156	8,76048	7,71479	9,53933	18,3299
Glyma11g01820	28,3891	32,0441	19,8389	12,6283	6,20544
Glyma12g02451	5,44117	9,08135	5,05653	5,34888	7,05774
Glyma11g10130	0,650094	3,33029	2,29277	1,925	4,26382
Glyma14g40440	0,047694	0,130403	0,461631	1,07366	4,78049
GLUCOSE-6-PHOSPHATE DEHYDROGENASE (G6PDH) (5)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma16g06850	3,22432	21,0335	18,8079	15,3158	29,6284
Glyma19g41450	3,51181	4,02177	4,99907	5,73877	5,65078
Glyma03g38870	2,61428	1,59151	1,6678	2,12239	2,32946
Glyma18g52110	9,53564	13,6603	10,0609	8,07369	6,13172
Glyma19g24250	27,3077	112,914	83,6291	48,0318	35,9354
Glyma07g01630	6,29193	3,63713	2,63598	4,53912	7,35572
Glyma02g10710	7,04559	7,43668	8,50301	9,56554	9,94901
Glyma19g22690	1,74209	1,91549	1,4743	1,93742	2,84244
PHOSPHOFRUCTOKINASE (PFK) (6)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma18g21720	3,87931	9,38084	6,7809	8,38097	9,81634
Glyma06g09320	5,04978	15,9193	66,2237	37,3251	39,0398
Glyma07g39960	286,16	173,924	137,771	125,407	144,75
Glyma04g09180	4,95804	10,7245	31,6056	15,6974	16,2945
Glyma20g33785	1,97102	2,64627	2,17354	2,36441	1,81687
Glyma09g00230	2,25147	1,40791	1,07818	0,981361	0,774413
Glyma10g33820	11,8403	8,3595	5,40236	5,95084	4,77647
Glyma13g42990	1,10103	1,02474	0,907537	0,632988	0,470164
Glyma07g01710	1,60423	1,0492	1,31511	1,90596	2,85883
Glyma07g15170	192,335	97,1953	67,4082	40,2386	28,2151
Glyma01g00870	24,8638	15,2802	17,7654	11,6086	5,63001
Glyma08g38450	0,733373	0,463724	0,360915	0,934917	3,4599

DIPHOSPHATE-FRUCTOSE-6-PHOSPHATE 1-PHOSPHOTRANSFERASE (PFK) (8)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma12g15951	1,95892	1,50818	1,06758	1,18201	0,559003
Glyma20g01010	0,34239	3,13767	24,7546	33,9593	59,5508
Glyma15g11890	0,51591	1,5616	5,59626	15,4044	42,0954
Glyma07g20400	4,33082	5,74016	26,0209	44,9527	77,946
Glyma17g01361	25,0867	29,8442	24,1977	18,6135	9,82261
Glyma07g39380	22,2109	15,7556	21,4225	24,6829	32,3436
Glyma09g01050	0,262906	0,450398	3,08133	14,4047	50,4612
PYRUVATE DECARBOXYLASE (PDC) (10)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma15g06185	7,0435	14,6045	15,8909	9,27122	7,30352
Glyma07g18570	38,5511	202,244	263,629	81,7306	51,1792
Glyma01g29190	0,0107197	0,432838	2,09467	0,946674	5,33371
ALCOHOL DEHYDROGENASE (ADH) (11)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma14g27940	0,206718	6,69651	48,9246	12,3086	19,6974
Glyma14g24860	0,142251	0,228091	1,05807	0,0208876	0
Glyma06g12780	13,6467	77,6071	303,352	132,034	99,9743
Glyma04g41990	5,94654	80,9798	349,124	89,6585	67,1216
Glyma04g39190	0,0189845	0,356407	1,2302	0,197542	1,65859
PYRUVATE DEHYDROGENASE (PDH - E2) (13)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma10g35960	11,934	14,4071	16,9192	16,0853	29,4357
Glyma20g31630	11,8963	20,1285	24,2907	23,1258	33,8479
Glyma07g39330	4,25122	4,79473	9,24102	15,2266	33,4585
Glyma01g20720	0,113369	0,0691453	0,0396801	0,0280645	0,203173
Glyma20g24030	1,29661	6,93087	10,6447	17,2064	46,9168
Glyma16g00590	4,4988	11,827	17,1864	24,6456	50,2937
ISOCITRATE DEHYDROGENASE (IDH) (15)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma02g40820	5,64614	6,32762	20,643	48,2822	132,937
Glyma14g39160	37,343	72,6609	75,9432	99,6342	183,177
Glyma15g42100	3,27145	5,26054	7,34487	9,93613	13,6181
Glyma05g08770	13,9236	17,8349	14,7384	14,7983	14,2233
Glyma08g17080	1,47163	1,47135	3,85759	6,33673	15,3561
Glyma15g42100	3,27145	5,26054	7,34487	9,93613	13,6181
ISOCITRATE LYASE (ICL) (16)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma12g10780	0,0964106	0,600604	0,722843	2,06891	4,32749
Glyma06g45950	0,35925	0,198412	0,223275	0,196756	2,30681
MALATE DEHYDROGENASE (MDH) (18)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma06g34190	24,6062	22,709	54,0622	78,122	134,919
Glyma05g01010	10,8637	9,55228	18,1505	16,8812	27,9377
Glyma07g30430	7,91437	4,82005	4,38213	5,37766	9,09446
Glyma17g10880	16,8795	54,272	79,119	56,0607	49,8573
Glyma02g08010	20,846	17,3641	47,5841	91,6188	280,157
Glyma11g04720	58,908	34,4358	30,8443	22,1574	23,4463
Glyma12g19520	23,1742	27,9353	57,8741	93,6685	237,287
GLYCEROL KINASE (21)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma07g03230	9,21476	7,9159	5,91878	7,56704	11,7933
GLYCEROL-3-PHOSPHATE DEHYDROGENASE (GPDHC1) (22)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma12g01380	0,0590567	0,193613	0,593084	1,90852	6,24067
Glyma11g20840	2,37412	1,87171	3,29737	3,69806	6,51
Glyma10g15900	0,244216	0,592569	0,750667	1,5893	3,6575
Glyma03g28970	0,104682	0,116913	0,249517	0,313483	1,90471
Glyma02g33340	0,0382097	0,0344492	0,0542307	0,374521	1,10645
GLUTAMATE DECARBOXYLASE (GAD) (24)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma11g33280	0,299323	0,313214	1,64936	16,5397	500,851
Glyma08g09670	4,22349	3,38258	3,7693	1,48884	0,498156
Glyma02g40040	0,00885193	0	0,0244257	0,0885585	1,81669
PYRUVATE KINASE (PK) (9)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma19g00870	8,91371	110,947	259,859	107,703	56,0169
Glyma01g40860	0,193354	0,500082	1,14381	1,6398	3,66656
Glyma10g32230	7,56115	10,4451	11,4569	13,3022	16,3748
Glyma20g27300	0,0516529	0,459237	1,62876	2,4742	3,35362
Glyma20g35400	8,50012	22,5238	20,7071	20,3903	32,2292
Glyma10g34490	23,4814	12,6048	32,3869	25,5217	32,1467
Glyma10g40110	5,3464	6,58991	9,1681	9,04474	11,8394
Glyma20g33060	0,827428	2,3191	29,7114	37,3726	74,7767
Glyma13g21360	4,11522	13,2309	21,3858	26,0471	41,1399
Glyma09g23150	14,8788	9,4341	9,61715	11,975	15,6708
Glyma07g35110	53,4043	59,8323	54,674	47,7001	51,3304
Glyma19g37420	1,11935	2,05592	9,79397	8,81392	8,19802
Glyma20g30430	1,32464	3,64616	5,58138	8,06157	15,1516
Glyma10g37210	2,11996	3,83964	8,18854	14,0221	26,3527
Glyma05g09310	16,4324	43,71	105,199	60,6788	56,1451
Glyma20g02980	27,5865	62,7122	48,9197	42,8643	43,9578
Glyma10g07480	49,8529	47,2961	50,0143	47,4413	58,8822
Glyma03g34740	0,130751	0,400028	3,89039	5,65891	5,38229
Glyma16g20980	1,76067	2,09235	1,90348	4,95672	14,2242
LACTATE DEHYDROGENASE (LDH) (12)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma17g14950	0,316234	0,896658	0,614879	0,546282	2,44105
Glyma05g04520	21,5936	83,349	108,206	75,2937	86,8904
CITRATE SYNTHASE (CSY) (14)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma15g42301	15,6233	26,3552	22,0772	29,2518	48,6894
Glyma08g42400	10,0384	14,874	12,2381	14,1889	17,7406
Glyma08g16770	51,1789	37,8025	24,7205	29,1252	59,1609
Glyma02g45790	88,7628	47,2256	27,2703	18,394	3,40173
Glyma18g12393	1,63855	3,63141	2,49378	2,86156	4,02127
Glyma14g03000	60,4081	98,8159	55,8835	36,2053	7,97824
Glyma03g30810	10,8254	7,98881	11,6071	16,7078	29,4474
Glyma15g15010	0	0,0118502	0	0,21181	1,83445
Glyma09g04000	19,3039	23,4257	43,9153	72,4082	157,872
Glyma01g03530	3,79201	5,67413	19,6849	31,8986	76,4111
Glyma15g42140	0,0172923	0,0669347	0,339433	9,42378	93,8083
Glyma02g04120	14,3684	13,9459	34,8179	55,8577	144,415
Glyma08g17010	0,188274	0,0707325	0,20225	4,9162	103,776
MALATE SYNTHASE (MSY) (17)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma05g03100	5,0441	3,58693	5,42655	7,85598	12,2369
Glyma17g13730	3,04398	2,73343	6,03114	32,4328	24,1176
PHOSPHOENOLPYRUVATE CARBOXYKINASE (PEPCK) (19)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma01g02330	0,192126	0,865254	1,29447	8,8107	9,06523
Glyma04g09510	0,152074	0,416924	1,63736	1,66691	16,1518
Glyma06g09610	1,97973	1,05034	0,673942	0,540348	0,480799
Glyma09g33650	1,26512	0,917323	0,878132	3,1293	30,0814
TRIACYLGLYCEROL LIPASE (TAG Lipases) (20)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma17g13600	0,02409	1,21712	3,71145	5,84981	10,6573
Glyma15g08745	0,65117	2,31691	3,40369	4,61636	5,69745
Glyma13g30456	2,38655	3,07168	2,82332	2,98125	7,39962
Glyma15g08730	0,86777	1,44419	1,96732	2,2294	4,3335
Glyma15g08760	0,65117	2,31691	3,40369	4,61636	5,69745
Glyma05g02950	2,07051	2,19489	6,6977	9,76014	10,4263
COMATOSE (CTS) (23)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma05g31270	13,7904	8,91956	8,50461	9,3632	6,95356
Glyma08g14485	16,2381	14,8907	12,2038	11,6575	5,63688
Glyma11g38160	8,10035	5,9006	6,02818	7,6431	6,51137
Glyma18g02110	7,97188	12,3579	12,8056	17,7837	16,5647
GAMMA-AMINO-N-BUTYRATE TRANSAMINASE (GABA-T) (25)					
ID	Dry	3HAI	6HAI	12HAI	24HAI
Glyma12g02510	10,4818	9,28659	12,4214	16,3412	18,2577
Glyma11g10190	0,232561	0,436433	1,61449	6,36761	21,739

## 5.5 Biossíntese e sinalização de fitormônios durante a germinação

A dessecação da semente bem como a manutenção do estado de dormência são regulados positivamente por ABA (Hilhorst 1995). Foram encontrados, com níveis relativamente altos de expressão, dois parálogos da família ABA *Insensitive3/Viviparous1* (ABI3/VP1) (Glyma08g47240 e Glyma18g38490) (FPKM > 60). Estes genes são conhecidos pela sua atuação como TFs na mediação de respostas a ABA (Holdsworth, Finch-Savage et al. 2008) (Material Suplementar). Genes que codificam AIP2 (*ABI3 interacting protein*) (Glyma17g33631 e Glyma14g12380), responsáveis pela degradação dos ABI3, também foram encontrados entre os genes expressos na germinação (Material Suplementar). Integrantes de outra família (CYP707A2) importante para realização do catabolismo de ABA também foram encontrados expressos nos períodos iniciais da germinação (e.g. Glyma09g41960 e Glyma16g20490). Em *A. thaliana* e tomate, a expressão de ABI3 é reprimida logo após a germinação por um sinal derivado de GA (Bassel, Mullen et al. 2006). Este parece ser o caso também em soja, visto que ambos os genes permanecem com níveis de expressão relativamente altos até 12HAI, quando sofrerem drástica redução em 24HAI (~9x). Além disso, a degradação de ABI3 é promovida pela AIP2 E3 ligase (Zhang, Garreton et al. 2005), para a qual foram encontrados dois homólogos expressos. Curiosamente, o TF ABI5, que atua *downstream* a ABI3 (Lopez-Molina, Mongrand et al. 2002), também foi identificado como expresso no curso da germinação, apresentando declínio em seu nível de expressão em 24HAI. Tal como ABI3, ABI5 também é regulado por degradação proteica (Garcia, Lynch et al. 2008). Os resultados indicam que a proteólise mediada por ubiquitina destes TFs precede a repressão transcricional, permitindo uma regulação precisa dos efeitos da sinalização por ABA.

O ácido giberélico (GA – *gibberellic acid*) constitui um dos principais agentes na promoção da germinação em sementes, atuando como antagonista do hormônio ABA (Colebrook, Thomas et al. 2014). Nesse contexto, foram encontrados diversos genes codificadores de proteínas que integram vias de biossíntese e sinalização de GA com níveis relativamente altos de expressão, como por exemplo: 2 SLY1 (*SLEEPY 1*) e 3 GID1 (*gibberellin-insensitive dwarf1*) (Figura 20). Também são expressos durante a germinação genes das famílias GA3ox1 – responsável pela síntese da forma ativa de GA em soja – DELLA e PIF4 (Figura 20). A ligação de GA ao receptor GID1 promove a formação de uma superfície de ligação a DELLA (Murase, Hirano et al. 2008; Shimada, Ueguchi-Tanaka et al. 2008). O complexo GA-GID1-DELLA é então reconhecido por SLY1, que pertence à família F-box. O complexo formado, origina uma E3-ubiquitina-ligase conhecida como SCF<sup>SLY1</sup>. A interação com a SCF<sup>SLY1</sup> resulta na ubiquitinação e subsequente proteólise de DELLA, permitindo assim a liberação das respostas a GA (Tyler, Thomas et al. 2004). Os dados permitiram a identificação da expressão de homólogos SLY1 e GID1. Entre os genes GID1, 3 são transcritos em níveis relativamente altos nos eixos secos (Figura 20). Curiosamente, os três pertencem ao subgrupo GID1B, que são induzidos durante o período de pós-maturação na ausência de biossíntese de GA (Hauvermale, Tuttle et al. 2015). Sabe-se que proteínas GID1B apresentam maior afinidade por GA do que os outros subgrupos da família (GID1A e GID1C) (Nakajima, Shimada et al. 2006). Além disso, sabe-se também que GID1A-C são filogeneticamente distintos de GID1B e apresentam padrões de expressão também distintos durante a germinação (Voegelé, Linkies et al. 2011). Os resultados encontrados são coerentes com a divergência funcional entre GID1A-C e GID1B, visto que os dois genes GID1C tem seus níveis de expressão progressivamente aumentados no curso da germinação (Figura 20), enquanto GID1B são regulados negativamente entre 0 e 24HAI.

Uma vez que o nível de expressão de *GID1B* não é capaz de compensar a ausência de *GID1A-C* (Voegelé, Linkies et al. 2011), sugere-se que a expressão aumentada dos genes *GID1C* é essencial à mediação da resposta canônica a GA na germinação de soja. A alta concentração de *GID1B* nas sementes secas e sua alta afinidade por GA pode ser interpretada como parte de um mecanismo de detecção mais sensível para captar níveis muito baixos de GA presentes logo após a embebição (Hauvermale, Tuttle et al. 2015).

Embora não sejam tão bem descritos na germinação quanto ABA e GA, o etileno e brassinosteróides (BR) são reconhecidos por sua ação promotora neste processo (KeÇpczyński e KeÇpczyńska 1997; Steber e McCourt 2001). Os dados confirmam a expressão de genes-chave para a biossíntese de etileno: *ACS* (*aminocyclopropane-1-carboxylic acid synthase*) e *ACCoX* (*aminocyclopropanecarboxylate oxidase*) (Figura 20). Uma vez sintetizado, a molécula de etileno pode ser reconhecida por 5 diferentes tipos de receptores de membrana: *EIN4* (*ethylene insensitive*), *ETR1* (*ethylene resistant*), *ETR2*, *ERS1* (*ethylene response sensor*) e *ERS2* (Corbineau, Xia et al. 2007). O presente trabalho permitiu identificar a expressão de genes pertencentes a todos os grupos, exceto *ERS2* (Figura 20; Material Suplementar). A produção de etileno se inicia logo após a embebição da semente e promove a quebra da dormência e germinação por meio da regulação da via de sinalização por ABA (Beaudoin, Serizet et al. 2000; Kucera, Cohn et al. 2005; Corbineau, Xia et al. 2007). O principal intermediário da produção de etileno é o S-AdoMet (*S-adenosyl-methionine*) que é sintetizado pela S-AdoMet sintetase; em seguida o S-AdoMet é convertido em ACC pela ACC sintase (*ACS*). A síntese do etileno é catalisada então pela enzima ACO, que é descrita como fundamental para o processo de germinação (Linkies e Leubner-Metzger 2012). Tal característica é confirmada em nosso estudo, dada a detecção da expressão dos genes-chave da biossíntese de etileno. No que concerne à sinalização por etileno, os

resultados confirmam a regulação por retroalimentação na via durante a germinação, dada a expressão considerável de reguladores negativos, como genes das famílias CTR1, e F-box (EBF1) (Figura 20) que promovem a degradação dos fatores de transcrição EIN3/EIL (Potuschak, Lechner et al. 2003).

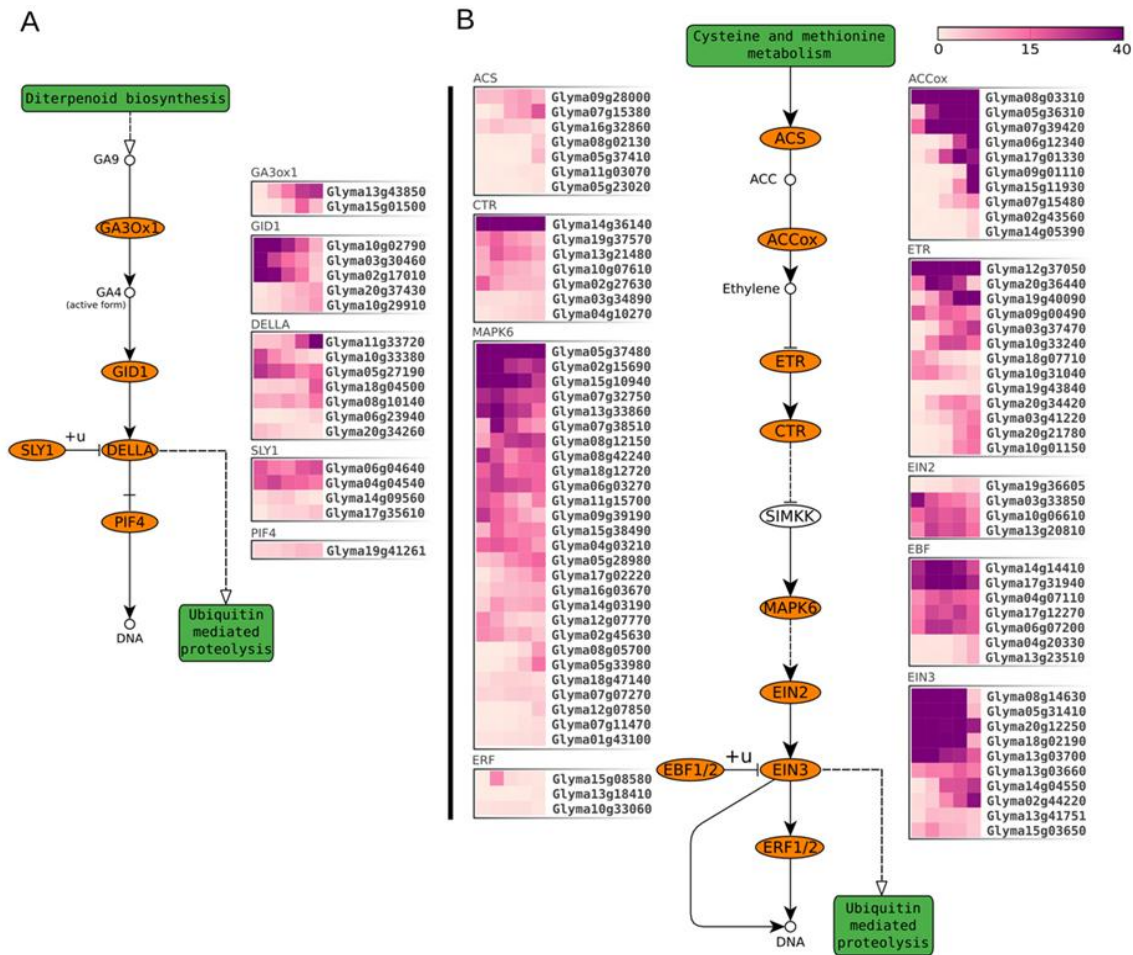


Figura 20 Últimas etapas da biossíntese e principais vias de sinalização de hormônios promotores da germinação: Giberelina (A) e etileno (B). Representações esquemáticas seguem o modelo KEGG ([http://www.genome.jp/kegg/document/help\\_pathway.html](http://www.genome.jp/kegg/document/help_pathway.html)), exceto pelo padrão de cores. Formas elípticas representam famílias gênicas. Retângulos verdes representam vias conectadas. Mapas de calor representam os níveis transcricionais (FPKM) dos genes expressos (FPKM >1) pertencentes às famílias mostradas nas vias. As cinco colunas nos mapas de calor representam as amostras analisadas, da esquerda para a direita: eixo seco, 3HAI, 6HAI, 12HAI e 24HAI. (HAI – hours after imbibition)

## 6 CONCLUSÕES

1. Foi encontrada um enriquecimento notável de TFs entre os genes regulados positivamente na germinação, especialmente pertencentes à família LOB. Tal fato sugere que um amplo conjunto de TFs precocemente ativados desempenham papel decisivo disparando o programa transcricional necessário à germinação.
2. Observou-se uma forte evidência para a regulação transcricional de diversas vias metabólicas primárias envolvidas na mobilização de reservas e produção de ATP, tais como a ativação precoce de vias respiratórias, e a degradação local de reservas lipídicas para alimentar a via do glioxilato, a qual está frequentemente associada ao período pós-germinativo. Essas reservas podem atuar como uma fonte de ATP local, contribuindo com a capacidade de germinação dos eixos embrionários na ausência de cotilédones.
3. Além disso, diversos genes envolvidos na biossíntese e sinalização de hormônios promotores da germinação, como GA, etileno e BR, estão regulados positivamente em estágios distintos da germinação. Genes envolvidos na degradação e regulação negativa de ABA foram ativadas, corroborando a noção de que a razão GA:ABA é crítica para a germinação.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

- Abrouk, M., F. Murat, et al. (2010). "Palaeogenomics of plants: synteny-based modelling of extinct ancestors." Trends Plant Sci **15**(9): 479-487.
- Adkins, S. W. and J. D. Ross (1981). "Studies in wild oat seed dormancy I. The role of ethylene in dormancy breakage and germination of wild oat seeds (*Avena fatua* L.)." Plant Physiol **67**(2): 358-362.
- Aida, M., T. Ishida, et al. (1997). "Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant." Plant Cell **9**(6): 841-857.
- Ainsworth, E. A., C. R. Yendrek, et al. (2012). "Accelerating yield potential in soybean: potential targets for biotechnological improvement." Plant Cell Environ **35**(1): 38-52.
- Akhunov, E. D., S. Sehgal, et al. (2013). "Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat." Plant Physiol **161**(1): 252-265.
- Allen, M. D., K. Yamasaki, et al. (1998). "A novel mode of DNA recognition by a  $\beta$ -sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA." The EMBO Journal **17**(18): 5484-5496.
- Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-3402.
- Amor, Y., C. H. Haigler, et al. (1995). "A membrane-associated form of sucrose synthase and its potential role in synthesis of cellulose and callose in plants." Proceedings of the National Academy of Sciences **92**(20): 9353-9357.
- Aravind, L. and E. V. Koonin (1999). "DNA-binding proteins and evolution of transcription regulation in the archaea." Nucleic Acids Res **27**(23): 4658-4670.
- Arteca, R. (1995). "Brassinosteroids." Plant Hormones Physiology, biochemistry and molecular biology, Kluwer Academic Press, Dordrecht, The Netherlands: 206-213.
- Ash, M. (2015). "U.S. Soybean Shipments Surge but New Sales are Slowing."
- Babu, M. M., N. M. Luscombe, et al. (2004). "Structure and evolution of transcriptional regulatory networks." Curr Opin Struct Biol **14**(3): 283-291.
- Babu, M. M. and S. A. Teichmann (2003). "Evolution of transcription factors and the gene regulatory network in *Escherichia coli*." Nucleic Acids Res **31**(4): 1234-1244.



- Bassel, G. W., R. T. Mullen, et al. (2006). "ABI3 expression ceases following, but not during, germination of tomato and Arabidopsis seeds." Journal of experimental botany **57**(6): 1291-1297.
- Bazin, J., D. Batlla, et al. (2010). "Role of relative humidity, temperature, and water status in dormancy alleviation of sunflower seeds during dry after-ripening." J Exp Bot: erq314.
- Beaudoin, N., C. Serizet, et al. (2000). "Interactions between abscisic acid and ethylene signaling cascades." The Plant Cell **12**(7): 1103-1115.
- Bell, E. M., W.-c. Lin, et al. (2012). "Arabidopsis lateral organ boundaries negatively regulates brassinosteroid accumulation to limit growth in organ boundaries." Proceedings of the National Academy of Sciences **109**(51): 21146-21151.
- Belliény-Rabelo, D., A. E. Oliveira, et al. (2013). "Impact of whole-genome and tandem duplications in the expansion and functional diversification of the F-box family in legumes (Fabaceae)." PLoS One **8**(2): e55127.
- Benlloch, R., I. d'Erfurth, et al. (2006). "Isolation of mtpim proves Tnt1 a useful reverse genetics tool in *Medicago truncatula* and uncovers new aspects of AP1-like functions in legumes." Plant Physiol **142**(3): 972-983.
- Berbel, A., C. Navarro, et al. (2001). "Analysis of PEAM4, the pea AP1 functional homologue, supports a model for AP1-like genes controlling both floral meristem and floral organ identity in different plant species." The Plant Journal **25**(4): 441-451.
- Bewley, J. D. and M. Black (1994). *Seeds*. Seeds, Springer: 1-33.
- Bewley, J. D., K. Bradford, et al. (2012). Seeds: physiology of development, germination and dormancy, Springer Science & Business Media.
- Bolger, A. M., M. Lohse, et al. (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." Bioinformatics: btu170.
- Bray, C. M. and C. E. West (2005). "DNA repair mechanisms in plants: crucial sensors and effectors for the maintenance of genome integrity." New Phytologist **168**(3): 511-528.
- Brocard-Gifford, I. M., T. J. Lynch, et al. (2003). "Regulatory networks in seeds integrating developmental, abscisic acid, sugar, and light signaling." Plant Physiol **131**(1): 78-92.
- Burtis, E. (1950). "World soybean production and trade." by KS Markley, Interscience Publishers, Inc., New York **1**: 61-108.
- Cai, M., D. Qiu, et al. (2008). "Identification of novel pathogen-responsive cis-elements and their binding proteins in the promoter of OsWRKY13, a gene regulating rice disease

- resistance." Plant Cell Environ **31**(1): 86-96.
- Cao, D., H. Cheng, et al. (2006). "Gibberellin mobilizes distinct DELLA-dependent transcriptomes to regulate seed germination and floral development in Arabidopsis." Plant Physiol **142**(2): 509-525.
- Carter, T., R. Nelson, et al. (2004). "Soybeans: Improvement Production and Uses." Agron. Mongr **16**: 303-416.
- Catoira, R., C. Galera, et al. (2000). "Four genes of *Medicago truncatula* controlling components of a Nod factor transduction pathway." Plant Cell **12**(9): 1647-1665.
- Chang, C., S. F. Kwok, et al. (1993). "Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators." SCIENCE-NEW YORK THEN WASHINGTON- **262**: 539-539.
- Chen, G., C. Wang, et al. (2011). "Overview of available methods for diverse RNA-Seq data analyses." Sci China Life Sci **54**(12): 1121-1128.
- Chen, M. and J. J. Thelen (2010). "The plastid isoform of triose phosphate isomerase is required for the postgerminative transition from heterotrophic to autotrophic growth in Arabidopsis." The Plant Cell **22**(1): 77-90.
- Clouse, S. D. and J. M. Sasse (1998). "BRASSINOSTEROIDS: Essential Regulators of Plant Growth and Development." Annu Rev Plant Physiol Plant Mol Biol **49**: 427-451.
- Colebrook, E. H., S. G. Thomas, et al. (2014). "The role of gibberellin signalling in plant responses to abiotic stress." Journal of experimental biology **217**(1): 67-75.
- Corbineau, F., Q. Xia, et al. (2007). "Ethylene, a key factor in the regulation of seed dormancy." "ONE ROTTEN APPLE SPOILS THE WHOLE BARREL": THE PLANT HORMONE ETHYLENE, THE SMALL MOLECULE AND ITS COMPLEXITY: 113.
- Crow, K. D. and G. P. Wagner (2006). "What is the role of genome duplication in the evolution of complexity and diversity?" Mol Biol Evol **23**(5): 887-892.
- Daniel, B., T. Pavkov-Keller, et al. (2015). "Oxidation of Monolignols by Members of the Berberine Bridge Enzyme Family Suggests a Role in Plant Cell Wall Metabolism." Journal of Biological Chemistry **290**(30): 18770-18781.
- Dash, S., J. D. Campbell, et al. (2016). "Legume information system (LegumeInfo. org): a key component of a set of federated data resources for the legume family." Nucleic acids research **44**(D1): D1181-D1188.
- de Pater, S., V. Greco, et al. (1996). "Characterization of a zinc-dependent transcriptional

- activator from Arabidopsis." Nucleic Acids Res **24**(23): 4624-4631.
- Desmason, A. M. and M. Tixier (1986). "Amino acids content in germinating seeds and seedlings from *Castanea sativa* L." Plant Physiol **81**(2): 692-695.
- Dilworth, M. J., E. K. James, et al. (2008). Nitrogen-fixing leguminous symbioses, Springer Science & Business Media.
- Dong, Z.-c., Z. Zhao, et al. (2005). "Floral patterning in *Lotus japonicus*." Plant Physiol **137**(4): 1272-1282.
- Dongen, S. M. (2000). Graph clustering by flow simulation.
- Du, Z., X. Zhou, et al. (2010). "agriGO: a GO analysis toolkit for the agricultural community." Nucleic Acids Res **38**(Web Server issue): W64-70.
- Dubos, C., R. Stracke, et al. (2010). "MYB transcription factors in Arabidopsis." Trends Plant Sci **15**(10): 573-581.
- Dure, L. and L. Waters (1965). "Long-lived messenger RNA: evidence from cotton seed germination." Science **147**(3656): 410-412.
- Eddy, S. R. (2011). "Accelerated Profile HMM Searches." PLoS Comput Biol **7**(10): e1002195.
- Eisen, M. B., P. T. Spellman, et al. (1998). "Cluster analysis and display of genome-wide expression patterns." Proceedings of the National Academy of Sciences **95**(25): 14863-14868.
- El Yahyaoui, F., H. Küster, et al. (2004). "Expression profiling in *Medicago truncatula* identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program." Plant physiology **136**(2): 3159-3176.
- Esashi, Y. and A. Leopold (1969). "Dormancy regulation in subterranean clover seeds by ethylene." Plant Physiol **44**(10): 1470.
- Eulgem, T., P. J. Rushton, et al. (2000). "The WRKY superfamily of plant transcription factors." Trends Plant Sci **5**(5): 199-206.
- Fait, A., R. Angelovici, et al. (2006). "Arabidopsis seed development and germination is associated with temporally distinct metabolic switches." Plant Physiol **142**(3): 839-854.
- FAOSTAT. (2016). Retrieved 21/09/2016, from <http://faostat.fao.org/>.
- Finch-Savage, W. E. and G. Leubner-Metzger (2006). "Seed dormancy and the control of germination." New Phytol **171**(3): 501-523.
- Francia, E., D. Barabaschi, et al. (2007). "Fine mapping of a HvCBF gene cluster at the frost resistance locus Fr-H2 in barley." Theoretical and Applied Genetics **115**(8): 1083-1091.

- Friedrichsen, D. M., C. A. Joazeiro, et al. (2000). "Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase." Plant Physiol **123**(4): 1247-1256.
- Garcia, M. E., T. Lynch, et al. (2008). "A small plant-specific protein family of ABI five binding proteins (AFPs) regulates stress response in germinating Arabidopsis seeds and seedlings." Plant molecular biology **67**(6): 643-658.
- Goff, L., C. Trapnell, et al. (2012). "cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data." R package version **2**(0).
- Goldberg, R. B., G. Hoschek, et al. (1981). "Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis." Developmental biology **83**(2): 201-217.
- Golombek, S., H. Rolletschek, et al. (2001). "Control of storage protein accumulation during legume seed development." Journal of Plant Physiology **158**(4): 457-464.
- Graham, I. A. (2008). "Seed storage oil mobilization." Annu. Rev. Plant Biol. **59**: 115-142.
- Guilfoyle, T., G. Hagen, et al. (1998). "How does auxin turn on genes?" Plant Physiol **118**(2): 341-347.
- Guo, H., L. Li, et al. (2009). "Three related receptor-like kinases are required for optimal cell elongation in Arabidopsis thaliana." Proceedings of the National Academy of Sciences **106**(18): 7648-7653.
- Guo, J., Y. Wang, et al. (2010). "A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences." Ann Bot **106**(3): 505-514.
- Guo, Y., R. Yang, et al. (2012). "Accumulation of  $\gamma$ -aminobutyric acid in germinated soybean (*Glycine max* L.) in relation to glutamate decarboxylase and diamine oxidase activity induced by additives under hypoxia." European Food Research and Technology **234**(4): 679-687.
- Haigler, C. H., M. Ivanova-Datcheva, et al. (2001). "Carbon partitioning to cellulose synthesis." Plant Mol Biol **47**(1-2): 29-51.
- Hauvermale, A. L., K. M. Tuttle, et al. (2015). "Loss of Arabidopsis thaliana seed dormancy is associated with increased accumulation of the GID1 GA hormone receptors." Plant and Cell Physiology: pcv084.
- Hematy, K., P. E. Sado, et al. (2007). "A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis." Curr Biol **17**(11): 922-931.

- Hermann, F. J. (1962). A revision of the genus Glycine and its immediate allies, US Dept. of Agriculture.
- Hilhorst, H. W. (1995). "A critical update on seed dormancy. I. Primary dormancy." Seed Science Research **5**(02): 61-73.
- Ho, P.-t. (1969). "The loess and the origin of Chinese agriculture." The American Historical Review **75**(1): 1-36.
- Hoekstra, F., E. Golovina, et al. (1999). "Imbibitional leakage from anhydrobiotes revisited." Plant Cell Environ **22**(9): 1121-1131.
- Hofer, J., L. Turner, et al. (1997). "UNIFOLIATA regulates leaf and flower morphogenesis in pea." Current Biology **7**(8): 581-587.
- Holdsworth, M. J., W. E. Finch-Savage, et al. (2008). "Post-genomics dissection of seed dormancy and germination." Trends Plant Sci **13**(1): 7-13.
- Howell, K. A., R. Narsai, et al. (2009). "Mapping metabolic and transcript temporal switches during germination in rice highlights specific transcription factors and the role of RNA instability in the germination process." Plant Physiol **149**(2): 961-980.
- Hua, J. and E. M. Meyerowitz (1998). "Ethylene responses are negatively regulated by a receptor gene family in *Arabidopsis thaliana*." Cell **94**(2): 261-271.
- Hua, J., H. Sakai, et al. (1998). "EIN4 and ERS2 are members of the putative ethylene receptor gene family in *Arabidopsis*." Plant Cell **10**(8): 1321-1332.
- Hymowitz, T. (1970). "On the domestication of the soybean." Economic Botany **24**(4): 408-421.
- Ishiguro, S. and K. Nakamura (1994). "Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and  $\beta$ -amylase from sweet potato." Molecular and General Genetics MGG **244**(6): 563-571.
- Jaillon, O., J. M. Aury, et al. (2007). "The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla." Nature **449**(7161): 463-467.
- Jiao, Y., N. J. Wickett, et al. (2011). "Ancestral polyploidy in seed plants and angiosperms." Nature **473**(7345): 97-100.
- Jin, J., X. Liu, et al. (2010). "Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in Northeast China." Field Crops Research **115**(1): 116-123.

- Jin, J., H. Zhang, et al. (2014). "PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors." Nucleic Acids Res **42**(Database issue): D1182-1187.
- Johnson, C. S., B. Kolevski, et al. (2002). "TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor." Plant Cell **14**(6): 1359-1375.
- Jones, K. M., H. Kobayashi, et al. (2007). "How rhizobial symbionts invade plants: the Sinorhizobium-Medicago model." Nat Rev Microbiol **5**(8): 619-633.
- Jones, S. I. and L. O. Vodkin (2013). "Using RNA-Seq to profile soybean seed development from fertilization to maturity." PLoS One **8**(3): e59270.
- Kanei-Ishii, C., A. Sarai, et al. (1990). "The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the myb protooncogene product." Journal of Biological Chemistry **265**(32): 19990-19995.
- Karszen, C. M., S. Zagorski, et al. (1989). "Key role for endogenous gibberellins in the control of seed germination." Ann Bot **63**(1): 71-80.
- KeÇpczyński, J. and E. KeÇpczyńska (1997). "Ethylene in seed dormancy and germination." Physiologia Plantarum **101**(4): 720-726.
- Kelly, A. A., A.-L. Quettier, et al. (2011). "Seed storage oil mobilization is important but not essential for germination or seedling establishment in Arabidopsis." Plant Physiol **157**(2): 866-875.
- Kelly, K., J. Van Staden, et al. (1992). "Seed coat structure and dormancy." Plant growth regulation **11**(3): 201-209.
- Ketring, D. and P. Morgan (1971). "Physiology of oil seeds II. Dormancy release in Virginia-type peanut seeds by plant growth regulators." Plant Physiol **47**(4): 488-492.
- Ketring, D. and P. Morgan (1972). "Physiology of oil seeds IV. Role of endogenous ethylene and inhibitory regulators during natural and induced afterripening of dormant Virginia-type peanut seeds." Plant Physiol **50**(3): 382-387.
- Kieber, J. J., M. Rothenberg, et al. (1993). "CTR1, a negative regulator of the ethylene response pathway in Arabidopsis, encodes a member of the raf family of protein kinases." Cell **72**(3): 427-441.
- Kim, D., G. Pertea, et al. (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biol **14**(4): R36.

- Kim, H., E. E. Helmbrecht, et al. (2011). "Ethylene receptor ETHYLENE RECEPTOR1 domain requirements for ethylene responses in Arabidopsis seedlings." Plant Physiol **156**(1): 417-429.
- Kimura, M. and E. Nambara (2010). "Stored and neosynthesized mRNA in Arabidopsis seeds: effects of cycloheximide and controlled deterioration treatment on the resumption of transcription during imbibition." Plant Mol Biol **73**(1-2): 119-129.
- Klempnauer, K.-H., T. J. Gonda, et al. (1982). "Nucleotide sequence of the retroviral leukemia gene v-myb and its cellular progenitor c-myb: the architecture of a transduced oncogene." Cell **31**(2): 453-463.
- Koornneef, M., L. Bentsink, et al. (2002). "Seed dormancy and germination." Curr Opin Plant Biol **5**(1): 33-36.
- Koster, K. L. and A. C. Leopold (1988). "Sugars and desiccation tolerance in seeds." Plant Physiol **88**(3): 829-832.
- Kucera, B., M. A. Cohn, et al. (2005). "Plant hormone interactions during seed dormancy release and germination." Seed Science Research **15**(4): 281-307.
- Kuo, T. M., D. C. Doehlert, et al. (1990). "Sugar metabolism in germinating soybean seeds evidence for the sorbitol pathway in soybean axes." Plant Physiol **93**(4): 1514-1520.
- Ladrera, R., D. Marino, et al. (2007). "Reduced carbon availability to bacteroids and elevated ureides in nodules, but not in shoots, are involved in the nitrogen fixation response to early drought in soybean." Plant Physiol **145**(2): 539-546.
- Lagace, M. and D. P. Matton (2004). "Characterization of a WRKY transcription factor expressed in late torpedo-stage embryos of *Solanum chacoense*." Planta **219**(1): 185-189.
- Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.
- Lawrence, G. (1949). "Name of the soybean." Science (New York, NY) **110**(2865): 566-566.
- Ledent, V. and M. Vervoort (2001). "The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis." Genome Res **11**(5): 754-770.
- Lee, H. W., M. J. Kim, et al. (2013). "LBD18 acts as a transcriptional activator that directly binds to the EXPANSIN14 promoter in promoting lateral root emergence of Arabidopsis." The Plant Journal **73**(2): 212-224.
- Lespinet, O., Y. I. Wolf, et al. (2002). "The role of lineage-specific gene family expansion in the

- evolution of eukaryotes." Genome Res **12**(7): 1048-1059.
- Leubner-Metzger, G. (2001). "Brassinosteroids and gibberellins promote tobacco seed germination by distinct pathways." Planta **213**(5): 758-763.
- Li, B. and M. E. Foley (1997). "Genetic and molecular control of seed dormancy." Trends Plant Sci **2**(10): 384-389.
- Libault, M., A. Farmer, et al. (2010). "Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection." Plant Physiol **152**(2): 541-552.
- Libault, M., A. Farmer, et al. (2010). "An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants." Plant J **63**(1): 86-99.
- Libault, M., T. Joshi, et al. (2009). "Legume transcription factor genes: what makes legumes so special?" Plant Physiol **151**(3): 991-1001.
- Lim, C. W., S. H. Yang, et al. (2015). "The *AtLRK10L1. 2*, *Arabidopsis* ortholog of wheat *LRK10*, is involved in ABA-mediated signaling and drought resistance." Plant Cell Rep **34**(3): 447-455.
- Linkies, A. and G. Leubner-Metzger (2012). "Beyond gibberellins and abscisic acid: how ethylene and jasmonates control seed germination." Plant cell reports **31**(2): 253-270.
- Linnaeus, C. v. (1753). "Species plantarum, 2 vols." Salvius, Stockholm.
- Liu, Y., M. Koornneef, et al. (2007). "The absence of histone H2B monoubiquitination in the *Arabidopsis* *hub1* (*rdo4*) mutant reveals a role for chromatin remodeling in seed dormancy." Plant Cell **19**(2): 433-444.
- Lopez-Molina, L., S. Mongrand, et al. (2002). "ABI5 acts downstream of ABI3 to execute an ABA-dependent growth arrest during germination." The Plant Journal **32**(3): 317-328.
- Lynch, M. and J. S. Conery (2000). "The evolutionary fate and consequences of duplicate genes." Science **290**(5494): 1151-1155.
- Mandava, N. B. (1988). "Plant growth-promoting brassinosteroids." Annual review of plant physiology and plant molecular biology **39**(1): 23-52.
- Marchive, C., R. Mzid, et al. (2007). "Isolation and characterization of a *Vitis vinifera* transcription factor, *VvWRKY1*, and its effect on responses to fungal pathogens in transgenic tobacco plants." J Exp Bot **58**(8): 1999-2010.
- Marino, D., P. Frendo, et al. (2007). "Nitrogen fixation control under drought stress. Localized or systemic?" Plant Physiol **143**(4): 1968-1974.



- Matilla, A., M. Gallardo, et al. (2005). "Structural, physiological and molecular aspects of heterogeneity in seeds: a review." Seed Science Research **15**(02): 63-76.
- Matilla, A. J. (2000). "Ethylene in seed formation and germination." Seed Science Research **10**(2): 111-126.
- Maunoury, N., M. Redondo-Nieto, et al. (2010). "Differentiation of symbiotic cells and endosymbionts in *Medicago truncatula* nodulation are coupled to two transcriptome-switches." PLoS One **5**(3): e9519.
- McKie, J. and K. L. Anderson (1967). "The soybean book."
- Miao, Y. and U. Zentgraf (2007). "The antagonist function of Arabidopsis WRKY53 and ESR/ESP in leaf senescence is modulated by the jasmonic and salicylic acid equilibrium." Plant Cell **19**(3): 819-830.
- Middleton, P. H., J. Jakab, et al. (2007). "An ERF transcription factor in *Medicago truncatula* that is essential for Nod factor signal transduction." Plant Cell **19**(4): 1221-1234.
- Miller, A. K., G. Galiba, et al. (2006). "A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-A m 2 in *Triticum monococcum*." Molecular Genetics and Genomics **275**(2): 193-203.
- Miller, A. (1975). "Biochemistry of legume seed proteins." Annual review of plant physiology **26**(1): 53-72.
- Miyashita, Y., R. Dolferus, et al. (2007). "Alanine aminotransferase catalyses the breakdown of alanine after hypoxia in *Arabidopsis thaliana*." The Plant Journal **49**(6): 1108-1121.
- Morgenstern, B. and W. R. Atchley (1999). "Evolution of bHLH transcription factors: modular evolution by domain shuffling?" Mol Biol Evol **16**(12): 1654-1663.
- Morrison, M. J., H. D. Voldeng, et al. (2000). "Agronomic changes from 58 years of genetic improvement of short-season soybean cultivars in Canada." Agronomy Journal **92**(4): 780-784.
- Murase, K., Y. Hirano, et al. (2008). "Gibberellin-induced DELLA recognition by the gibberellin receptor GID1." Nature **456**(7221): 459-463.
- Nakabayashi, K., M. Okamoto, et al. (2005). "Genome-wide profiling of stored mRNA in *Arabidopsis thaliana* seed germination: epigenetic and genetic regulation of transcription in seed." The Plant Journal **41**(5): 697-709.
- Nakajima, M., A. Shimada, et al. (2006). "Identification and characterization of *Arabidopsis* gibberellin receptors." The Plant Journal **46**(5): 880-889.

- Nambara, E. and A. Marion-Poll (2005). "Abscisic acid biosynthesis and catabolism." Annu. Rev. Plant Biol. **56**: 165-185.
- Nelson, D. C., G. R. Flematti, et al. (2010). "Karrikins enhance light responses during germination and seedling development in *Arabidopsis thaliana*." Proc Natl Acad Sci U S A **107**(15): 7095-7100.
- Ni, M., J. M. Tepperman, et al. (1998). "PIF3, a phytochrome-interacting factor necessary for normal photoinduced signal transduction, is a novel basic helix-loop-helix protein." Cell **95**(5): 657-667.
- Nishimura, R., M. Ohmori, et al. (2002). "The novel symbiotic phenotype of enhanced-nodulating mutant of *Lotus japonicus*: astray mutant is an early nodulating mutant with wider nodulation zone." Plant and cell physiology **43**(8): 853-859.
- Noguchi, T., S. Fujioka, et al. (1999). "Arabidopsis det2 is defective in the conversion of (24R)-24-methylcholest-4-en-3-one to (24R)-24-methyl-5 $\alpha$ -cholestan-3-one in brassinosteroid biosynthesis." Plant Physiol **120**(3): 833-840.
- Nonogaki, H. (2014). "Seed dormancy and germination-emerging mechanisms and new hypotheses." Front Plant Sci **5**: 233.
- Nunes-Nesi, A., A. R. Fernie, et al. (2010). "Metabolic and signaling aspects underpinning the regulation of plant carbon nitrogen interactions." Mol Plant **3**(6): 973-996.
- Obendorf, R. L., A. D. Zimmerman, et al. (2009). "Accumulation of Soluble Carbohydrates during Seed Development and Maturation of Low-Raffinose, Low-Stachyose Soybean." Crop Science **49**(1): 329.
- Ogé, L., G. Bourdais, et al. (2008). "Protein repair L-isoaspartyl methyltransferase1 is involved in both seed longevity and germination vigor in *Arabidopsis*." Plant Cell **20**(11): 3022-3037.
- Okamoto, M., K. Tatematsu, et al. (2010). "Genome-wide analysis of endogenous abscisic acid-mediated transcription in dry and imbibed seeds of *Arabidopsis* using tiling arrays." The Plant Journal **62**(1): 39-51.
- Oldroyd, G. E. and S. R. Long (2003). "Identification and characterization of nodulation-signaling pathway 2, a gene of *Medicago truncatula* involved in Nod factor signaling." Plant Physiol **131**(3): 1027-1032.
- Orthofer, F. (1978). "Processing and utilization." Soybean Physiology, Agronomy, and Utilization. Norman, AG (ed.): 219-246.

- Otto, S. P. and J. Whitton (2000). "Polyploid incidence and evolution." Annu Rev Genet **34**: 401-437.
- Pabo, C. O. and R. T. Sauer (1992). "Transcription factors: structural families and principles of DNA recognition." Annual review of biochemistry **61**(1): 1053-1095.
- Paclt, J. (1949). "Nomenclature of the soybean." Science **109**(2831): 339-339.
- Pandey, S. P. and I. E. Somssich (2009). "The role of WRKY transcription factors in plant immunity." Plant Physiol **150**(4): 1648-1655.
- Paredes, A. R., C. R. Somerville, et al. (2006). "Visualization of cellulose synthase demonstrates functional association with microtubules." Science **312**(5779): 1491-1495.
- Paz-Ares, J., D. Ghosal, et al. (1987). "The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators." The EMBO Journal **6**(12): 3553.
- Perez-Llamas, C. and N. Lopez-Bigas (2011). "Gitoools: analysis and visualisation of genomic data using interactive heat-maps." PLoS One **6**(5): e19541.
- Pérez-Rueda, E. and J. Collado-Vides (2000). "The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12." Nucleic Acids Res **28**(8): 1838-1847.
- Pinfield-Wells, H., E. L. Rylott, et al. (2005). "Sucrose rescues seedling establishment but not germination of Arabidopsis mutants disrupted in peroxisomal fatty acid catabolism." The Plant Journal **43**(6): 861-872.
- Piper, C. V. and W. J. Morse (1910). "Soy Bean History, Varieties And Field Studies."
- Potuschak, T., E. Lechner, et al. (2003). "EIN3-dependent regulation of plant ethylene hormone signaling by two Arabidopsis F box proteins: EBF1 and EBF2." Cell **115**(6): 679-689.
- Powell, A. A. and S. Matthews (1979). "The influence of testa condition on the imbibition and vigour of pea seeds." J Exp Bot **30**(1): 193-197.
- Pracharoenwattana, I., J. E. Cornah, et al. (2005). "Arabidopsis peroxisomal citrate synthase is required for fatty acid respiration and seed germination." Plant Cell **17**(7): 2037-2048.
- Prince, S. J., T. Joshi, et al. (2015). "Comparative analysis of the drought-responsive transcriptome in soybean lines contrasting for canopy wilting." Plant Sci **240**: 65-78.
- Probert, R. J. (2000). "The role of temperature in the regulation of seed dormancy and

- germination." Seeds: the ecology of regeneration in plant communities **261**: 292.
- Qi, X., M. W. Li, et al. (2014). "Identification of a novel salt tolerance gene in wild soybean by whole-genome sequencing." Nat Commun **5**: 4340.
- Raghavan, V. (2000). Seed germination. Developmental Biology of Flowering Plants, Springer: 7-24.
- Ramalingam, J., C. Vera Cruz, et al. (2003). "Candidate defense genes from rice, barley, and maize and their association with qualitative and quantitative resistance in rice." Molecular Plant-Microbe Interactions **16**(1): 14-24.
- Ratcliffe, O. J. and J. L. Riechmann (2002). "Arabidopsis transcription factors and the regulation of flowering time: a genomic perspective." Current issues in molecular biology **4**: 77-92.
- Rensing, S. A. (2014). "Gene duplication as a driver of plant morphogenetic evolution." Curr Opin Plant Biol **17**: 43-48.
- Richards, D. E., J. Peng, et al. (2000). "Plant GRAS and metazoan STATs: one family?" Bioessays **22**(6): 573-577.
- Riechmann, J. L. and E. M. Meyerowitz (1998). "The AP2/EREBP family of plant transcription factors." Biological chemistry **379**: 633-646.
- Riechmann, J. L. and O. J. Ratcliffe (2000). "A genomic perspective on plant transcription factors." Curr Opin Plant Biol **3**(5): 423-434.
- Rosental, L., H. Nonogaki, et al. (2014). "Activation and regulation of primary metabolism during seed germination." Seed Science Research **24**(01): 1-15.
- Rushton, P. J., H. Macdonald, et al. (1995). "Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of  $\alpha$ -Amy2 genes." Plant Mol Biol **29**(4): 691-702.
- Ryu, H. and I. Hwang (2013). "Brassinosteroids in plant developmental signaling networks." Journal of Plant Biology **56**(5): 267-273.
- Sakai, H., J. Hua, et al. (1998). "ETR2 is an ETR1-like gene involved in ethylene signaling in Arabidopsis." Proceedings of the National Academy of Sciences **95**(10): 5812-5817.
- Salerno, G. L. and L. Curatti (2003). "Origin of sucrose metabolism in higher plants: when, how and why?" Trends Plant Sci **8**(2): 63-69.
- Schäffer, A. A., Y. I. Wolf, et al. (1999). "IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices." Bioinformatics

15(12): 1000-1011.

- Schauser, L., A. Roussis, et al. (1999). "A plant regulator controlling development of symbiotic root nodules." Nature **402**(6758): 191-195.
- Schmidt, J., T. Altmann, et al. (1997). "Brassinosteroids from seeds of *Arabidopsis thaliana*." Phytochemistry **45**(7): 1325-1327.
- Schmidt, M. A., W. B. Barbazuk, et al. (2011). "Silencing of soybean seed storage proteins results in a rebalanced protein composition preserving seed protein content without major collateral changes in the metabolome and transcriptome." Plant Physiol **156**(1): 330-345.
- Schmutz, J., S. B. Cannon, et al. (2010). "Genome sequence of the palaeopolyploid soybean." Nature **463**(7278): 178-183.
- Schnable, J. C., M. Freeling, et al. (2012). "Genome-wide analysis of syntenic gene deletion in the grasses." Genome Biol Evol **4**(3): 265-277.
- Schwechheimer, C., M. Zourelidou, et al. (1998). "Plant transcription factor studies." Annu Rev Plant Biol **49**(1): 127-150.
- Semon, M. and K. H. Wolfe (2007). "Consequences of genome duplication." Curr Opin Genet Dev **17**(6): 505-512.
- Seo, M., E. Nambara, et al. (2009). "Interaction of light and hormone signals in germinating seeds." Plant Mol Biol **69**(4): 463-472.
- Severin, A. J., S. B. Cannon, et al. (2011). "Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy." Plant Cell **23**(9): 3129-3136.
- Severin, A. J., J. L. Woody, et al. (2010). "RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome." BMC Plant Biol **10**: 160.
- Shimada, A., M. Ueguchi-Tanaka, et al. (2008). "Structural basis for gibberellin recognition by its receptor GID1." Nature **456**(7221): 520-523.
- Shuai, B., C. G. Reynaga-Pena, et al. (2002). "The lateral organ boundaries gene defines a novel, plant-specific gene family." Plant Physiol **129**(2): 747-761.
- Soltis, P. S. and D. E. Soltis (2009). "The role of hybridization in plant speciation." Annu Rev Plant Biol **60**: 561-588.
- Song, L., S. Prince, et al. (2016). "Genome-wide transcriptome analysis of soybean primary root under varying water-deficit conditions." BMC Genomics **17**: 57.
- Specht, J., D. Hume, et al. (1999). "Soybean yield potential—a genetic and physiological

- perspective." Crop Science **39**(6): 1560-1570.
- Sreenivasulu, N., L. Borisjuk, et al. (2010). "Barley grain development: toward an integrative view." International Review of Cell and Molecular Biology **281**: 49-89.
- Steber, C. M. and P. McCourt (2001). "A role for brassinosteroids in germination in Arabidopsis." Plant Physiol **125**(2): 763-769.
- Stewart, W. D. P. (1966). "Nitrogen fixation in plants." Nitrogen fixation in plants.
- Supek, F., M. Bošnjak, et al. (2011). "REVIGO summarizes and visualizes long lists of gene ontology terms." PLoS One **6**(7): e21800.
- Tang, H., X. Wang, et al. (2008). "Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps." Genome Res **18**(12): 1944-1954.
- Taylor, S. A., J. M. Hofer, et al. (2002). "PROLIFERATING INFLORESCENCE MERISTEM, a MADS-box gene that regulates floral meristem identity in pea." Plant Physiol **129**(3): 1150-1159.
- Thomas, T. L. (1993). "Gene expression during plant embryogenesis and germination: an overview." Plant Cell **5**(10): 1401.
- Trapnell, C., A. Roberts, et al. (2012). "Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks." Nat Protoc **7**(3): 562-578.
- Triezenberg, S. J. (1995). "Structure and function of transcriptional activation domains." Curr Opin Genet Dev **5**(2): 190-196.
- Tyler, L., S. G. Thomas, et al. (2004). "DELLA proteins and gibberellin-regulated seed germination and floral development in Arabidopsis." Plant physiology **135**(2): 1008-1019.
- Tzin, V. and G. Galili (2010). "The biosynthetic pathways for shikimate and aromatic amino acids in Arabidopsis thaliana." The arabidopsis book: e0132.
- Udvardi, M. K., K. Kakar, et al. (2007). "Legume transcription factors: global regulators of plant development and response to the environment." Plant Physiol **144**(2): 538-549.
- Ueguchi-Tanaka, M., M. Ashikari, et al. (2005). "GIBBERELLIN INSENSITIVE DWARF1 encodes a soluble receptor for gibberellin." Nature **437**(7059): 693-698.
- Van de Peer, Y. (2011). "A mystery unveiled." Genome Biol **12**(5): 1.
- Van de Peer, Y., S. Maere, et al. (2009). "The evolutionary significance of ancient genome duplications." Nat Rev Genet **10**(10): 725-732.
- Van Staden, J., S. Olatoye, et al. (1973). "Effect of light and ethylene upon cytokinin levels in

- seed of *Spergula arvensis*." J Exp Bot **24**(4): 662-666.
- Voegelé, A., A. Linkies, et al. (2011). "Members of the gibberellin receptor gene family *GID1* (*GIBBERELLIN INSENSITIVE DWARF1*) play distinct roles during *Lepidium sativum* and *Arabidopsis thaliana* seed germination." Journal of experimental botany **62**(14): 5131-5147.
- Walbot, V. (1971). "RNA metabolism during embryo development and germination of *Phaseolus vulgaris*." Developmental biology **26**(3): 369-379.
- Wally, O. and Z. K. Punja (2010). "Genetic engineering for increasing fungal and bacterial disease resistance in crop plants." GM Crops **1**(4): 199-206.
- Wang, J., Q. Ji, et al. (2009). "Overexpression of a cytosol-localized rhamnose biosynthesis protein encoded by *Arabidopsis* *RHM1* gene increases rhamnose content in cell wall." Plant Physiol Biochem **47**(2): 86-93.
- Wang, L., C. Cao, et al. (2014). "RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications." BMC plant biology **14**(1): 1.
- Wang, L., C. Cao, et al. (2014). "RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications." BMC Plant Biol **14**: 169.
- Wang, M., S. Sun, et al. (2014). "Isolation and characterization of the brassinosteroid receptor gene (*GmBRI1*) from *Glycine max*." International journal of molecular sciences **15**(3): 3871-3888.
- Wang, Y., X. Wang, et al. (2012). "Genome and gene duplications and gene expression divergence: a view from plants." Ann N Y Acad Sci **1256**: 1-14.
- Wareing, P. and P. Saunders (1971). "Hormones and dormancy." Annual review of plant physiology **22**(1): 261-288.
- Waters, L. C. and L. S. Dure (1966). "Ribonucleic acid synthesis in germinating cotton seeds." Journal of molecular biology **19**(1): 1-11.
- Weber, H., L. Borisjuk, et al. (2005). "Molecular physiology of legume seed development." Annu Rev Plant Biol **56**: 253-279.
- Weitbrecht, K., K. Müller, et al. (2011). "First off the mark: early seed germination." J Exp Bot **62**(10): 3289-3309.
- Welbaum, G., K. Bradford, et al. (1998). "Biophysical, physiological and biochemical

- processes regulating seed germination." Seed Science Research **8**(02): 161-172.
- Whaley, A., J. Sheridan, et al. (2015). "RNA-seq analysis reveals genetic response and tolerance mechanisms to ozone exposure in soybean." BMC genomics **16**(1): 1.
- Wilkins, O., H. Nahal, et al. (2009). "Expansion and diversification of the Populus R2R3-MYB family of transcription factors." Plant Physiol **149**(2): 981-993.
- Wilson, R. L., H. Kim, et al. (2014). "The Ethylene Receptors ETHYLENE RESPONSE1 and ETHYLENE RESPONSE2 Have Contrasting Roles in Seed Germination of Arabidopsis during Salt Stress." Plant Physiol **165**(3): 1353-1366.
- Winter, H. and S. C. Huber (2000). "Regulation of sucrose metabolism in higher plants: localization and regulation of activity of key enzymes." Critical Reviews in Plant Sciences **19**(1): 31-67.
- Yamaguchi, S. (2008). "Gibberellin metabolism and its regulation." Annu Rev Plant Biol **59**: 225-251.
- Yang, P., C. Chen, et al. (1999). "A pathogen-and salicylic acid-induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter." The Plant Journal **18**(2): 141-149.
- Yokota, T. (1997). "The structure, biosynthesis and function of brassinosteroids." Trends Plant Sci **2**(4): 137-143.
- Yu, X., A. Li, et al. (2015). "How membranes organize during seed germination: three patterns of dynamic lipid remodelling define chilling resistance and affect plastid biogenesis." Plant, cell & environment **38**(7): 1391-1403.
- Yuan, S., R. Li, et al. (2016). "RNA-Seq analysis of differential gene expression responding to different rhizobium strains in soybean (*Glycine max*) roots." Frontiers in Plant Science **7**: 721.
- Zhang, X., V. Garreton, et al. (2005). "The AIP2 E3 ligase acts as a novel negative regulator of ABA signaling by promoting ABI3 degradation." Genes & Development **19**(13): 1532-1543.
- Zhang, Z. L., Z. Xie, et al. (2004). "A rice WRKY gene encodes a transcriptional repressor of the gibberellin signaling pathway in aleurone cells." Plant Physiol **134**(4): 1500-1513.
- Zhou, X., Y. Jiang, et al. (2011). "WRKY22 transcription factor mediates dark-induced leaf senescence in Arabidopsis." Mol Cells **31**(4): 303-313.