

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
BACHARELADO EM BIOTECNOLOGIA**

CATHERINE DALL'AGNOL KRAUSE

**EVOLUÇÃO MOLECULAR DE GENES ENVOLVIDOS COM GRAVITROPISMO EM
PLANTAS**

**Porto Alegre
2019**

CATHERINE DALL'AGNOL KRAUSE

**EVOLUÇÃO MOLECULAR DE GENES ENVOLVIDOS COM GRAVITROPISMO EM
PLANTAS**

Trabalho de Conclusão de Curso submetido como requisito parcial para a obtenção do grau de Bacharel em Biotecnologia com ênfase em Bioinformática na Universidade Federal do Rio Grande do Sul.

Orientador Metodológico
Prof.^a Dra. Andreia Carina Turchetto Zolet

**Porto Alegre
2019**

CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Krause, Catherine Dall'Agnol
EVOLUÇÃO MOLECULAR DE GENES ENVOLVIDOS COM
GRAVITROPISMO EM PLANTAS / Catherine Dall'Agnol
Krause. -- 2019.
38 f.
Orientadora: Adnréia Carina Turchetto-Zolet.

Trabalho de conclusão de curso (Graduação) --
Universidade Federal do Rio Grande do Sul, Instituto
de Biociências, Curso de Biotecnologia:
Bioinformática, Porto Alegre, BR-RS, 2019.

1. Arquitetura de plantas. 2. Gravitropismo. 3.
Filogenia. 4. Estrutura gênica. I. Turchetto-Zolet,
Adnréia Carina, orient. II. Título.

Instituto de Biociências
Bacharelado em Biotecnologia

Evolução molecular de genes envolvidos com gravitropismo em plantas
Catherine Dall'Agnol Krause

Banca examinadora:

Prof.^a Dra. Caroline Turchetto
UFRGS

Prof.^a Dra. Fernanda Lazzarotto
UFRGS

AGRADECIMENTOS

Gostaria de agradecer imensamente à minha querida Prof^a Orientadora, Andréia, por todo o esforço, dedicação, compreensão e disponibilidade, mesmo com uma linda filha recém nascida; à minha família, principalmente meus pais, Gláucia e Kiko, minha avó, Lidia, e meu dindo, Valdir Júnior, por serem meus orientadores de vida e me guiarem durante todos os percalços e confusões, por celebrarem comigo todas as conquistas, até as pequenas, e por me amarem incondicionalmente; aos meus amigos, por me fazerem rir em cada momento que ia começar a chorar, por estarem ao meu lado sempre e por me ajudarem a descontraír em meio a um período de tensão e nervosismo; aos meus colegas de barra, por compartilharem comigo esses momentos difíceis, mas também por todas as queridas lembranças durante estes anos de graduação; aos demais colegas do curso de Biotecnologia, pelo apoio e parceria; aos meus colegas do NEJ POA, que sempre me incentivaram e foram muito compreensivos; e aos meus Professores da graduação, por terem me ensinado tanto, dentro e fora da sala de aula, e por me fazerem chegar onde estou agora.

DEDICATÓRIA

Dedico este trabalho às pessoas pelas quais meu coração mais dói de saudade: meus avós Valdir, Edelhardt e Idelzuíte, e minha bisavó Claudina. Não consigo imaginar como eu seria se não tivesse crescido, convivido e aprendido tanto com eles. Meus eternos agradecimentos e amor a cada momento que fui agraciada de passar com pessoas tão incríveis.

RESUMO

Genes da família *LAZY* são reguladores do Transporte Polar de Auxina (TPA), tendo um papel fundamental no gravitropismo, ou seja, na detecção da gravidade e modulação do crescimento das raízes e parte aérea. Em *Arabidopsis thaliana*, mutantes *lazy* apresentam fenótipo de gravitropismo reduzido na parte aérea. Porém, estudos sobre a presença e função desses genes em plantas ainda são escassos. Neste trabalho, identificamos homólogos dos genes *LAZY* em 63 espécies de plantas e algas e procuramos compreender a sua história evolutiva. O resultado da busca por *BLAST* nos genomas das espécies estudadas retornou 282 potenciais sequências homólogas aos genes *LAZY*. A análise filogenética revelou o agrupamento dessas sequências em dois clados: o primeiro deles inclui as sequências homólogas ao gene *LAZY5* de *Arabidopsis*, e o segundo inclui aos demais, sendo dividido em dois subgrupos (genes *LAZY2-4* e *LAZY1* e *6*). Essa análise permitiu concluir que esta família gênica é específica de plantas e que se diversificaram a partir de diferentes eventos de duplicação ao longo da evolução das espécies, sendo o clado que agrupa as *LAZY5* o mais divergente. Dessa maneira, os genes *LAZY* estão relacionados com os grandes eventos evolutivos que estão levando às mudanças na arquitetura das plantas terrestres.

Palavras-chave: Arquitetura de plantas, gravitropismo, filogenia, estrutura gênica.

ABSTRACT

Genes of the *LAZY* family are regulators of Polar Auxin Transport (PAT), thereby they have an essential role in gravitropism, which is the detection of gravity and modulation of root and shoot growth. In *Arabidopsis thaliana*, *lazy* mutants display a reduced shoot gravitropism phenotype. However, further studies about the presence and function of these genes are still scarce. In this paper, we have identified *LAZY* homologous genes in 52 species of plants, and we have sought to understand its evolutionary history. *BLAST* results have returned 282 potential *LAZY* homologous sequences in the studied species. Phylogenetic analysis has revealed the grouping of these sequences in two clades: the first one includes *Arabidopsis LAZY5* homologous sequences and the second, which is divided into two sub-groups (*LAZY2-4*; *LAZY1* and *6*) includes the others. This analysis allow us to conclude that this gene family is plant-specific and that it's genes have diverged from different duplication events during species evolution, being the *LAZY5* more divergent clade. Thus, the *LAZY* genes are related to the main evolutionary events that led to changes in land plants architecture.

Keywords: Plant architecture, gravitropism, phylogeny, genetic structure.

SUMÁRIO

1	INTRODUÇÃO	8
2	JUSTIFICATIVA	11
3	OBJETIVOS	12
3.1	Objetivo geral	12
3.2	Objetivos específicos	12
4	METODOLOGIA	13
4.1	Fontes de dados e recuperação de sequências	13
4.2	Alinhamento e análises Filogenéticas	14
4.3	Análise comparativa de organização e estrutura gênica	15
4.4	Análise de motivos e domínios proteicos	15
4.5	Análise de expressão <i>in silico</i>	15
5	RESULTADOS E DISCUSSÃO	17
5.1	Fontes de dados e recuperação de sequências	17
5.2	Alinhamento e análises Filogenéticas	20
5.3	Análise comparativa de organização e estrutura gênica	24
5.4	Análise de motivos e domínios proteicos	25
5.5	Análise de expressão <i>in silico</i>	27
6	CONCLUSÕES E PERSPECTIVAS	31
	REFERÊNCIAS	32
	ANEXOS	36

1 INTRODUÇÃO

As plantas se adaptaram de maneira a habitar os mais distintos ecossistemas terrestres. Como são organismos sésseis, o ambiente onde estão inseridas pode influenciar fortemente o seu crescimento. Dessa forma, é de extrema importância ter a capacidade de sentir e responder ao que está acontecendo ao seu redor, guiando o crescimento de seus órgãos até para encontrar regiões que favoreçam a captação de luz solar para a fotossíntese, e de água e nutrientes para as raízes, por exemplo (Su *et al.*, 2017). Para contemplar esta necessidade, múltiplos mecanismos de estratégias de crescimento - os chamados tropismos, foram desenvolvidos ao longo da evolução das plantas (Vandenbrink e Kiss, 2019).

O crescimento de órgãos em plantas se dá pela combinação de divisão celular nos meristemas e alongamento de células em regiões sub-apicais (Su *et al.*, 2017). Um dos tropismos de maior influência para a orientação desse crescimento, que vai determinar a arquitetura final da planta, é o gravitropismo - crescimento em resposta à gravidade (Chen *et al.*, 1999). O gravitropismo se mostra ainda mais importante durante o estágio de crescimento inicial, logo após a germinação, onde a gravidade é a única percepção no subsolo (Žádníková *et al.*, 2015). Com o uso desta estratégia, as plantas são capazes de sentir a gravidade e alterar a direção do crescimento de suas raízes e parte aérea através do gravitropismo positivo e negativo, respectivamente. Já os órgãos laterais, emergem dos órgãos primários de maneira comum, mas logo em seguida se realinham de acordo com um ângulo chamado ângulo de referência gravitrópico (do inglês "*gravitropic set-point angle*", GSA). O GSA é bem variado, sendo espécie- e órgão-específico, e pode ser modificado por fatores externos ou hormonais (Su *et al.*, 2017). Este processo consiste em três etapas: detecção da gravidade, transdução de sinal e crescimento assimétrico das células (Vandenbrink e Kiss, 2019).

O primeiro indício sobre a existência de um mecanismo de tal tipo foi através do experimento de Knight, em 1806, quando ele tentou cultivar plântulas em uma roda de jardinagem, posicionada perpendicularmente ao chão, de maneira a alterar o vetor da

gravidade (Knight, 1800). Próximos à Knight, Charles e Francis Darwin também se dedicaram ao estudo do movimento em resposta à luz em plantas com o trabalho "*The Power of Movement in Plants*", onde foi hipotetizado um mecanismo de sinalização, capaz de se mover do local onde foi recebido o estímulo externo até o local de resposta à ele (Darwin e Darwin, 1880). Esta teoria abriu os olhos de outros cientistas que, eventualmente, descobriram o sistema hormonal das plantas.

Tanto o gravitropismo quanto o fototropismo têm como mecanismo básico o crescimento assimétrico, ou crescimento diferencial, das células (Žádníková *et al.*, 2015). O fitormônio auxina exerce um papel fundamental para a coordenação do sinal deste crescimento, através do seu acúmulo ou gradiente de distribuição entre as células. Este hormônio é o único até então que possui um sistema de transporte polar (transporte polar de auxina, TPA), cujo mecanismo principal é o transporte direcionado célula-célula através de transportadores de influxo e efluxo que garantem um controle mais preciso de distribuição e, assim, permitem a formação de gradientes (Petrasek e Friml, 2009). Em algumas espécies, como o arroz e *Arabidopsis*, genes envolvidos com o TPA já foram identificados e caracterizados. Devido à importância da auxina para o gravitropismo, e do último para determinação da arquitetura das plantas, reguladores do TPA são potenciais alvos de estudo para controle deste mecanismo de crescimento.

Em 1996, Abe *et al.* identificaram um padrão de crescimento diferente em algumas cultivares de arroz. Por não terem um alongamento vertical das folhas, começaram a ser chamadas de plantas preguiçosas. Ao estudar estas plantas, descobriram os genes responsáveis por este comportamento e os nomearam genes *lazy* (do inglês, preguiçosos). Posteriormente, estes genes também foram identificados em milho (Dong *et al.* 2013) e em *Arabidopsis thaliana* (Yoshihara e Iino, 2007, Taniguchi *et al.*, 2017). O estudo destes genes em outras plantas é escasso.

Mais especificamente, em *Arabidopsis*, foram seis genes reguladores identificados parte da família multigênica *LAZY* (Taniguchi *et al.*, 2017.). Para facilitar a compreensão, os genes At5g14090, At1g17400, At1g19115, At1g72490, At3g24750 e At3g27025 serão

referidos neste trabalho como *LAZY1*, *LAZY2*, *LAZY3*, *LAZY4*, *LAZY5* e *LAZY6*, respectivamente. Yoshihara e Spalding, em 2017, conduziram um experimento com mutantes homocigotos para cinco dos genes *LAZY* em *Arabidopsis*. Tanto mutantes simples quanto mutantes de alta-ordem, com mais de um gene mutado, foram crescidos no escuro e tiveram seus fenótipos avaliados. Eles demonstraram que os mutantes, principalmente de alta-ordem, têm um mecanismo de gravitropismo defeituoso e a orientação das plântulas foi horizontal ou vertical para baixo.

Com a revolução das técnicas de sequenciamento de última geração e, por conseguinte, aumento exponencial de dados genômicos, proteicos e de expressão, os estudos evolutivos sobre famílias gênicas se tornaram mais acessíveis. Ao longo da evolução, os genomas passam por diversos eventos de mutação, além de perdas e ganhos de material genético, muitas vezes modificando o número de cópias de genes (NCG). Estas modificações desempenham papéis muito importantes na história evolutiva, visto que causam significativas consequências, podendo ser relacionadas a doenças ou até mesmo a vantagens adaptativas (Ames *et al.*, 2011). Estudos comparativos de NCG se tornaram mais populares, visto que aliam rapidez e confiabilidade e geram dados relevantes sobre os passos evolutivos de ganhos e perdas de material genético associados a mudanças fenotípicas que tornam mais clara a nossa compreensão sobre a evolução funcional. Dessa maneira, estudos evolutivos sobre os genes *LAZY* podem trazer novas visões sobre os eventos de duplicação que originaram novas famílias, ordens e espécies de plantas terrestres.

2 JUSTIFICATIVA

Pouco se sabe sobre arquitetura de plantas e grande parte dos estudos já feitos foca na regulação do crescimento por hormônios vegetais. A família *LAZY* foi estudada apenas na planta modelo *Arabidopsis thaliana* e, ainda assim, não há maiores informações sobre sua regulação, mecanismo de atividade ou participação em rotas de sinalização. Outros estudos sobre a presença e função desses genes em plantas ainda são escassos. Esta família gênica é um alvo potencial para aplicações biotecnológicas, visto que pode ser responsável pela determinação da orientação do crescimento de plantas.

3 OBJETIVOS

3.1 Objetivo geral

O objetivo deste trabalho é identificar e caracterizar genes da família *LAZY* em plantas e entender sua história evolutiva.

3.2 Objetivos específicos

- Identificar genes homólogos à família *LAZY* em Viridiplantae;
- Caracterizar os genes identificados quanto a sua estrutura e organização;
- Caracterizar a sequência proteica codificada pelos genes *LAZY* quanto a presença de domínio e motivos conservados;
- Compreender o padrão evolutivo destes genes nas espécies selecionadas;
- Entender os mecanismos de ganho e perda de genes dentro da família *LAZY*.

4 METODOLOGIA

4.1 Fontes de dados e recuperação de sequências

Para recuperar as sequências homólogas dos genes *LAZY* nos genomas de plantas disponíveis, primeiramente foi feito um estudo *in silico* das sequências *LAZY* já descritas em *A. thaliana*. Neste organismo modelo, já havia um estudo que identificou e descreveu seis genes: *LAZY1* - AT5G14090, *LAZY2* - AT1G17400, *LAZY3* - AT1G72490, *LAZY4* - AT1G19115, *LAZY5* - AT3G24750 e *LAZY6* - AT3G27025 (Yoshihara e Spalding, 2017). Estas sequências foram recuperadas do banco de dados Phytozome (<http://www.phytozome.net/>) e, posteriormente, foram utilizadas como *queries* (iscas) para recuperar os homólogos nas outras 62 espécies de plantas com genoma e proteoma disponíveis no banco de dados de genomas Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>) através da ferramenta *BLAST* (BLASTp). Dentre estas espécies encontram-se clorófitas ou algas (*Chlamydomonas reinhardtii*, *Dunaliella salina*, *Volvox carteri*, *Coccomyxa subellipsoidea C-169*, *Micromonas pusilla CCMP1545*, *Micromonas pusilla sp* and *Ostreococcus lucimarinus*); a hepática *Marchantia polymorpha*; os musgos *Physcomitrella patens* e *Sphagnum fallax*; a licófitas *Selaginella moellendorffii*; a única espécie restante da família Amborellaceae, *Amborella trichopoda*; as angiospermas monocotiledôneas *Ananas comosus*, *Musa acuminata*, *Spirodela polyrhiza* (aquática) e *Zostera marina* (marinha); as gramíneas da família Poaceae *Brachypodium distachyon*, *Brachypodium stacei*, *Oryza sativa*, *Oropetium thomaeum*, *Panicum hallii*, *Panicum virgatum*, *Setaria italica*, *Setaria viridis*, *Sorghum bicolor* e *Zea mays*; a herbácea *Aquilegia caerulea*; o amaranto *Amaranthus hypochondriacus*; as asterídeas *Daucus carota*, *Mimulus guttatus*, *Solanum lycopersicum* e *Solanum tuberosum*; as crassuláceas *Kalanchoe fedtschenkoi* e *Kalanchoe laxiflora*; o eucalipto *Eucalyptus grandis*; a parreira *Vitis vinifera*; as Malpighiales *Linum usitatissimum*, *Manihot esculenta*, *Populus trichocarpa*, *Ricinus communis* e *Salix purpurea*; as laranjas *Citrus sinensis* e *Citrus clementina*; as frutas tropicais *Carica papaya* e *Theobroma cacao*; o algodão *Gossypium raimondii*; as brassicáceas *Arabidopsis halleri*, *Arabidopsis lyrata*, *Boechera stricta*, *Brassica oleracea*

capitata, *Brassica rapa*, *Capsella grandiflora*, *Capsella rubella* e *Eutrema salsugineum*; e as fabáceas *Cucumis sativus*, *Fragaria vesca*, *Glycine max*, *Malus domestica*, *Medicago truncatula*, *Phaseolus vulgaris*, *Prunus persica*, *Trifolium pratense*.

Todas as sequências de DNA genômico, CDS (*Coding sequence*) e de aminoácidos correspondentes foram recuperadas e utilizadas nas análises seguintes. Para facilitar a identificação das espécies, foi adicionado um código de três letras para cada táxon representando a primeira letra do gênero e as duas primeiras letras da espécie (por exemplo, *Arabidopsis thaliana* foi representada por Ath); apenas duas espécies precisaram de uma letra adicional para distingui-las e foram elas: *Ananas comosus* - Acom e *Brachypodium stacei* - Bsta.

4.2 Alinhamento e análises Filogenéticas

As sequências recuperadas foram alinhadas utilizando o algoritmo de alinhamento MUSCLE (Edgar, 2004), implementado no MEGA 7 (Kumar *et al.*, 2016) e utilizadas para a realização de análises filogenéticas pelo método bayesiano no programa BEAST v. 1.8.4. O alinhamento foi codificado e os blocos ambíguos foram removidos antes da realização da análise filogenética. Os programas jModelTest2 (Darriba *et al.*, 2012) e Prottest (Abascal, Zardoya e Posada, 2005) foram utilizados para selecionar os melhores modelos de substituição de nucleotídeo e aminoácidos, respectivamente. O *birth-death process* foi selecionado como prior para a análise bayesiana no BEAST v. 1.8.4 (Drummond *et al.*, 2012) e foi executado por 100.000.000 gerações de cadeias de Markov e Monte Carlo (MCMC) para sequências de aminoácidos e nucleotídeos. O programa Tracer 1.6 (Rambaut *et al.*, 2018) foi utilizado para verificar os dados obtidos pela convergência das cadeias de Markov e Monte Carlo (tamanhos adequados de amostras efetivas > 200) após os primeiros 10% das gerações serem deletados como burn-in. Em seguida, foi utilizado o TreeAnnotator para o cálculo da árvore consenso e

dos valores de probabilidade posterior. O FigTree apresenta ótimas ferramentas de edição e foi utilizado para obter uma melhor visualização da árvore gerada.

4.3 Análise comparativa de organização e estrutura gênica

Para análises posteriores, as sequências obtidas foram organizadas a partir de características estruturais, como organização em éxons e íntrons e tamanho, e a nível proteico, por presença de motivos e domínios. Como fonte destas informações foi utilizado o próprio banco de dados do *Phytozome*. Para algumas espécies selecionadas (*Eucalyptus grandis*, *Oryza sativa*, *Zea mays*, *Medicago truncatula* e *Setaria italica*), foi utilizado o software *Gene Structure Display Server* (GSDS - <http://gsds.cbi.pku.edu.cn>) como método de visualização.

4.4 Análise de motivos e domínios proteicos

Para informações adicionais sobre domínios e motivos proteicos das sequências, foram utilizados os programas PROTTER (Omasits *et al.*, 2013 - <http://wlab.ethz.ch/protter/>), TMHMM 2.0 (Krogh *et al.*, 2001 - <http://www.cbs.dtu.dk/services/TMHMM/>) e *Multiple Expectation maximization for Motif Elicitation* - MEME (Bailey *et al.*, 2006 - <http://meme-suite.org/>), e sobre distribuição de nucleotídeos e aminoácidos, o WEBLOGO (Crooks *et al.*, 2004 - <https://weblogo.berkeley.edu/logo.cgi>). Para refinar os dados resultantes do MEME, foram utilizados outros dois pacotes (Tomtom e GOMo) adjacentes ao programa.

4.5 Análise de expressão gênica *in silico*

Para avaliar a expressão dos genes da família *LAZY*, tanto em *Arabidopsis thaliana* quanto em outras espécies selecionadas (*Glycine max*, *Oriza sativa* e *Zea mays*) e

determinar especificidade de tecido e intensidade de expressão, foram utilizados os dados de RNA-seq disponíveis no banco de dados GENEVESTIGATOR (Hruz *et al.*, 2008 - <https://genevestigator.com/gv/>). A ferramenta utilizada para esta análise foi a *Hierarchical clustering*, utilizando os coeficientes de Pearson para normalização e clusterização dos dados. Como resultado são obtidos mapas de calor que representam a intensidade potencial de expressão dos genes em tecidos e estágios de desenvolvimento específicos.

5 RESULTADOS E DISCUSSÃO

5.1 Fontes de dados e recuperação de sequências

Até hoje, apenas seis genes da família *LAZY* foram caracterizados na planta modelo *A. thaliana*. Estas sequências foram utilizadas como *queries* para realizar a busca por BLAST nas outras 62 espécies (6 algas e 56 plantas terrestres) com genoma e proteoma depositados no banco de dados *Phytozome*. Para cada uma das sequências da família, foi realizada uma busca independente, portanto alguns genes recuperados para uma das *queries* (iscas) também foram obtidos utilizando outra sequência. Dessa forma, foram recuperadas sequências potenciais homólogas a *LAZY* em 52 das 62 espécies avaliadas, sendo 94 para *LAZY1*, 104 para *LAZY2*, 112 para *LAZY3*, 100 para *LAZY4*, 44 para *LAZY5* e 21 para *LAZY6*. Removendo aquelas sequências que foram redundantes nas buscas (retornaram como resultado em mais de uma busca com iscas diferentes), foram totalizadas 282 sequências.

A distribuição dos genes candidatos entre as espécies está apresentada na Tabela 1. As espécies que apresentaram um maior número de possíveis homólogos foram *Brassica oleracea capitata* (11 genes), *Brassica rapa* (11 genes) e *Glycine max* (13 genes), as três sendo plantas poliplóides. As espécies que não apresentaram nenhum homólogo foram as algas (*Micromonas pusilla*, *Chlamydomonas reinhardtii*, *Coccomyxa subellipsoidea*, *Dunaliella salina*, *Micromonas sp*, *Volvox carteri*), além da hepática *Marchantia polymorpha*, a licófito *Selaginella moellendorffii* e os musgos *Physcomitrella patens* e *Sphagnum fallax*.

Além das espécies citadas acima que não apresentaram nenhum homólogo, para *LAZY1*, *Zostera marina* também não apresentou. Para *LAZY2* e *LAZY3*, *Spirodela polyrhiza*. Para *LAZY4*, *Brachypodium stacei* e *Spirodela polyrhiza*. *LAZY5* não apresentou homólogos em 1 das 8 plantas da família Brassicaceae (*Brassica oleracea capitata*), em nenhuma das 10 Poáceas, em *Amborella trichopoda*, em *Amaranthus hypochondriacus*, nas Crassuláceas, em *Musa acuminata* e em *Spirodela polyrhiza*. Já a *LAZY6*, está presente somente nas Brassicáceas, na Malvácea *Gossypium raimondii*,

nas Fabáceas *Glycine max*, *Trifolium pratense* e *Medicago truncatula*, nas Crassuláceas, em *Manihot esculenta*, em *Solanum tuberosum* e em *Vitis vinifera*.

Todos os potenciais homólogos recuperados para LAZY5 foram exclusivos, assim como os genes recuperados para LAZY6 em Brassicáceas.

Tabela 1. Dados taxonômicos, número de genes homólogos à LAZY por espécie e distribuição por clados de acordo com a filogenia.

Familia	Espécie	Acrônimo	Nº de genes	Clado		
				I	Ila	IIb
Chlamydomonadaceae	<i>Chlamydomonas reinhardtii</i>	Cre	0	0	0	0
Dunaliellaceae	<i>Dunaliella salina</i>	Dsa	0	0	0	0
Volvocaceae	<i>Volvox carteri</i>	Vca	0	0	0	0
Coccomyxa	<i>Coccomyxa subellipsoidea</i>	Csu	0	0	0	0
Mamiellaceae	<i>Micromonas pusilla</i>	Mpu	0	0	0	0
Mamiellaceae	<i>Micromonas sp.</i>	Msp	0	0	0	0
Bathycoccaceae	<i>Ostreococcus lucimarinus</i>	Olu	0	0	0	0
Marchantiaceae	<i>Marchantia polymorpha</i>	Mpo	0	0	0	0
Funariaceae	<i>Physcomitrella patens</i>	Ppa	0	0	0	0
Sphagnaceae	<i>Sphagnum fallax</i>	Sfa	0	0	0	0
Selaginellaceae	<i>Selaginella moellendorffii</i>	Smo	0	0	0	0
Musaceae	<i>Musa acuminata</i>	Mac	4	0	3	1
Bromeliaceae	<i>Ananas comosus</i>	Acom	4	1	2	1
Araceae	<i>Spirodela polyrhiza</i>	Spo	1	0	0	1
Zosteraceae	<i>Zostera marina</i>	Zmar	2	0	2	0
Amborellaceae	<i>Amborella trichopoda</i>	Atr	2	0	1	1
Poaceae	<i>Brachypodium distachyon</i>	Bdi	4	0	3	1
Poaceae	<i>Brachypodium stacei</i>	Bsta	4	0	3	1
Poaceae	<i>Oryza sativa</i>	Osa	3	0	2	1
Poaceae	<i>Oropetium thomaeum</i>	Oth	2	0	1	1
Poaceae	<i>Panicum hallii</i>	Pha	4	0	3	1
Poaceae	<i>Panicum virgatum</i>	Pvi	7	0	6	1
Poaceae	<i>Setaria italica</i>	Sit	3	0	2	1
Poaceae	<i>Setaria viridis</i>	Svi	3	0	2	1
Poaceae	<i>Sorghum bicolor</i>	Sbi	4	0	3	1
Poaceae	<i>Zea mays</i>	Zma	4	0	3	1
Amaranthaceae	<i>Amaranthus hypochondriacus</i>	Ahy	4	0	2	2
Phrymaceae	<i>Mimulus guttatus</i>	Mgu	5	1	3	1
Crassulaceae	<i>Kalanchoe fedtschenkoi</i>	Kfe	3	0	2	1
Crassulaceae	<i>Kalanchoe laxiflora</i>	Kla	8	0	4	4
Linaceae	<i>Linum usitatissimum</i>	Lus	9	2	6	1
Euphorbiaceae	<i>Manihot esculenta</i>	Mes	8	1	3	4
Euphorbiaceae	<i>Ricinus communis</i>	Rco	5	1	2	2
Salicaceae	<i>Populus trichocarpa</i>	Ptr	7	1	3	3
Salicaceae	<i>Salix purpurea</i>	Spu	7	1	3	3
Rutaceae	<i>Citrus sinensis</i>	Csi	5	1	2	2
Rutaceae	<i>Citrus clementina</i>	Ccl	8	4	2	2
Caricaceae	<i>Carica papaya</i>	Cpa	3	1	1	1
Ranunculaceae	<i>Aquilegia coerulea</i>	Aco	3	1	1	1
Apiaceae	<i>Daucus carota</i>	Dca	6	1	2	3
Solanaceae	<i>Solanum lycopersicum</i>	Sly	6	2	3	1
Solanaceae	<i>Solanum tuberosum</i>	Stu	5	1	1	2
Myrtaceae	<i>Eucalyptus grandis</i>	Egr	6	1	2	3
Vitaceae	<i>Vitis vinifera</i>	Vvi	4	1	1	2
Malvaceae	<i>Gossypium raimondii</i>	Gra	9	2	3	4
Malvaceae	<i>Theobroma cacao</i>	Tca	5	1	2	2
Brassicaceae	<i>Arabidopsis halleri</i>	Aha	6	1	3	2
Brassicaceae	<i>Arabidopsis lyrata</i>	Aly	6	1	3	2
Brassicaceae	<i>Arabidopsis thaliana</i>	Ath	6	1	3	2
Brassicaceae	<i>Boeheria stricta</i>	Bst	6	1	3	2
Brassicaceae	<i>Brassica oleracea capitata</i>	Bol	11	0	7	4
Brassicaceae	<i>Brassica rapa</i>	Bra	11	1	6	4
Brassicaceae	<i>Capsella grandiflora</i>	Cgr	6	1	3	2
Brassicaceae	<i>Capsella rubella</i>	Cru	6	1	3	2
Brassicaceae	<i>Eutrema salsugineum</i>	Esa	6	1	3	2
Cucurbitaceae	<i>Cucumis sativus</i>	Csa	4	1	1	2
Rosaceae	<i>Fragaria vesca</i>	Fve	4	1	2	1
Rosaceae	<i>Prunus persica</i>	Ppe	5	1	2	2
Rosaceae	<i>Malus domestica</i>	Mdo	7	1	2	4
Fabaceae	<i>Medicago truncatula</i>	Mtr	6	2	1	3
Fabaceae	<i>Phaseolus vulgaris</i>	Pvu	7	1	3	3
Fabaceae	<i>Glycine max</i>	Gma	13	3	4	6
Fabaceae	<i>Trifolium pratense</i>	Tpr	5	1	1	3

5.2 Alinhamento e análises Filogenéticas

Para uma melhor compreensão da história evolutiva desta família gênica, foi construída uma árvore filogenética utilizando as sequências proteicas dos genes homólogos recuperados.

O primeiro passo para a reconstrução de uma filogenia é fazer um alinhamento múltiplo global com todas as sequências. Após uma primeira tentativa, foi percebido que as sequências eram muito longas e possuíam blocos de aminoácidos muito variáveis. Para lidar com esta situação, foi utilizado o software GBlocks (Castresana, 2000 - <http://molevol.cmima.csic.es/castresana/Gblocks.html>) para fazer uma análise da ambiguidade e escolha dos blocos a serem removidos. Após a remoção dos blocos indicados pelo GBlocks, foi realizada a análise filogenética.

A análise filogenética foi realizada utilizando dois conjuntos de dados: um deles utilizando todas as sequências recuperadas com as seis sequências de arabis como iscas e o outro excluindo as sequências *LAZY5* e suas homólogas. O segundo conjunto de dados foi utilizado porque foi percebido que o gene *LAZY5* era significativamente diferente dos demais genes caracterizados, até mesmo após a remoção dos blocos mais variáveis. Dessa forma, sua presença na análise estava afetando esta etapa e prejudicando o *score* do alinhamento com as demais sequências. Para obter melhor resolução na filogenia e utilizar maior fragmento do gene na análise, foi optada pela remoção do *LAZY5* e seus homólogos exclusivos (aqueles que haviam sido recuperados utilizando apenas este gene como *query*).

Após a mineração dos dados, o resultado do alinhamento foi utilizado como entrada no programa BEAST (métodos bayesianos) para construção de árvores filogenéticas. Estas últimas foram analisadas pelo programa TreeAnnotator para cálculo da árvore consenso e, assim, foi obtida uma árvore de boa resolução que apresenta dois principais clados. As figuras 1 e 2 foram construídas utilizando o FigTree e indicam as árvores consenso para o alinhamento com e sem *LAZY5*, respectivamente.

A primeira observação que se pode fazer é que existiram dois grandes eventos evolutivos: o primeiro, que originou a separação do grupo no qual o gene *LAZY5* está

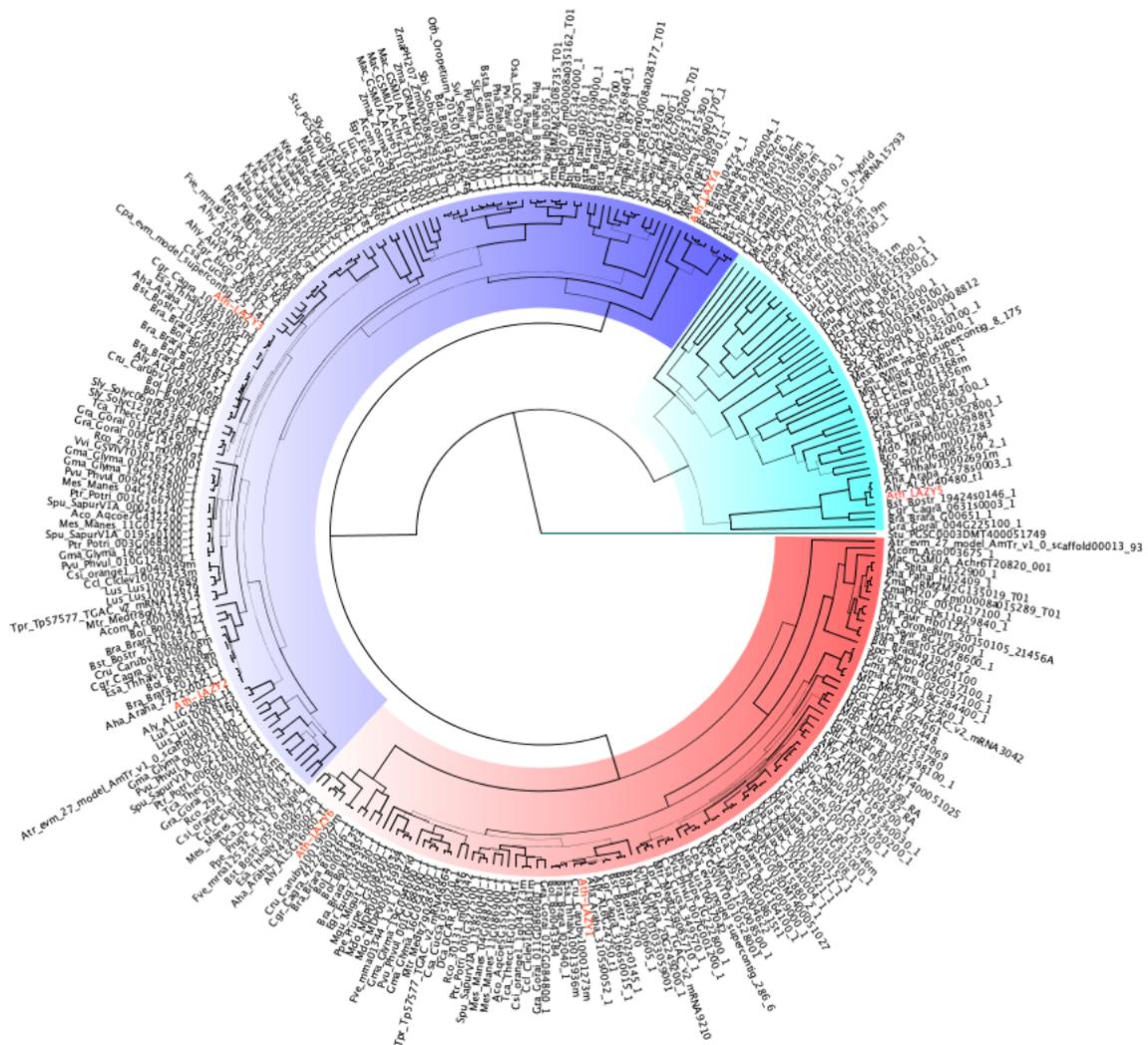
inserido, formando o Clado I; e o segundo, que originou a separação do grupo que abrange os genes *LAZY2*, *LAZY3* e *LAZY4* daquele que abrange *LAZY1* e *LAZY6*, formando os sub-clados IIa e IIb, respectivamente.

É interessante observar que o clado I é composto majoritariamente por espécies eudicotiledôneas, tendo apenas um representante da classe monocotiledônea (*Ananas comosus*).

O clado IIa é o grupo mais heterogêneo, pois tanto as monocotiledôneas quanto as eudicotiledôneas apresentaram diversidade de representantes, sendo entre um e seis em mono (média de 2,5 genes) e entre um e sete em eudicotiledôneas (média de 3 genes). *LAZY2* e *LAZY3* são mais próximas entre si quando comparadas à *LAZY4*, que foi significativamente alocada em um grupo formado apenas por espécies da família Brassicaceae (Aly, Bra, Bol, Aha, Esa, Bst, Cru e Cgr).

Analisando o clado IIb, é possível perceber que todas as espécies de monocotiledôneas apresentam apenas um representante deste grupo, com exceção de *Zostera marina* que não apresenta nenhum. Já as eudicotiledôneas apresentam entre 1 e 6 genes agrupados neste grupo, sendo a média e a mediana iguais a 2 genes. Além disto, *LAZY6* e os seus homólogos em Bst, Esa, Aly, Cru, Cgr, Bra e Bol, foram significativamente separados das demais 92 sequências de IIb, formando um grupo exclusivo da família Brassicaceae. Estes são indicativos de que, além do evento que originou o gene ancestral ao clado IIb, ocorreram duplicações mais recentes que a divisão das classes mono e eudicotiledôneas. Os grupos formados exclusivamente por espécies de Brassicaceae comentados acima (tanto em IIa como em IIb) sugerem que ocorreram eventos de duplicação ainda mais recentes dentro desta família.

Este padrão de agrupamento por famílias também pode ser observado para as espécies de Fabaceae (Mtr, Pvu, Gma e Tpr) e de Poaceae (Bdi, Bsta, Osa, Oth, Pha, Pvi, Sit, Svi, Sbi e Zma).



0.3

Figura 1. Árvore filogenética da família LAZY, utilizando as 282 sequências homólogas à LAZY em 52 espécies de plantas. Apenas probabilidades posteriores superiores a 0,5 estão presentes e aquelas maiores a 0,8 estão representadas com as linhas grossas. Três principais grupos foram formados e estão indicados por cores diferentes, sendo o clado I em verde, IIa em azul e IIb em vermelho.

5.3 Análise comparativa de organização e estrutura gênica

Para tentar compreender melhor os eventos evolutivos que ocorreram nesta família gênica, foi analisada a quantidade de éxons e íntrons presentes nos genes homólogos identificados por BLAST. Estes dados foram obtidos a partir das informações sobre as sequências no próprio banco de dados do *Phytozome*. Foram encontradas sequências cujas quantidades de íntrons variam de 0 (em *Amborella trichopoda* e *Spirodela polyrhiza*), a 8 (em *Fragaria vesca*), sendo a mediana (valor encontrado em maior número de vezes) igual a 4 (143 sequências). A Tabela S1 apresenta um resumo destas características.

A Figura 3 mostra a organização gênica realizada no programa GSDS (*Gene Structure Display Server*), para algumas espécies de plantas selecionadas como representativas (*Arabidopsis thaliana*, *Eucalyptus grandis*, *Medicago truncatula*, *Oryza sativa*, *Zea mays* e *Setaria italica*). Como podemos observar, a estrutura dos genes nestes organismos não teve muitas variações, com 22 sequências apresentando 3 íntrons, 4 sequências com 2 íntrons (em Egr, Ath e Mtr), 1 sequência com 4 íntrons (Mtr) e 1 sequência com apenas 1 íntron (Egr).

A divisão dos clados nesta análise foi coerente com aqueles gerados pela filogenia, sendo o I (inferior) aquele que engloba o gene *LAZY5*; o IIa (superior), os genes *LAZY2-4*; e o IIb (meio), os genes *LAZY1* e 6. O clado IIa foi aquele que manteve-se mais homogêneo em relação à organização estrutural, já que todos os genes englobados possuem 3 íntrons; e o clado I foi o mais heterogêneo, já que, de seus 4 genes, dois possuem 2 íntrons, um 3 íntrons e um 4 íntrons.

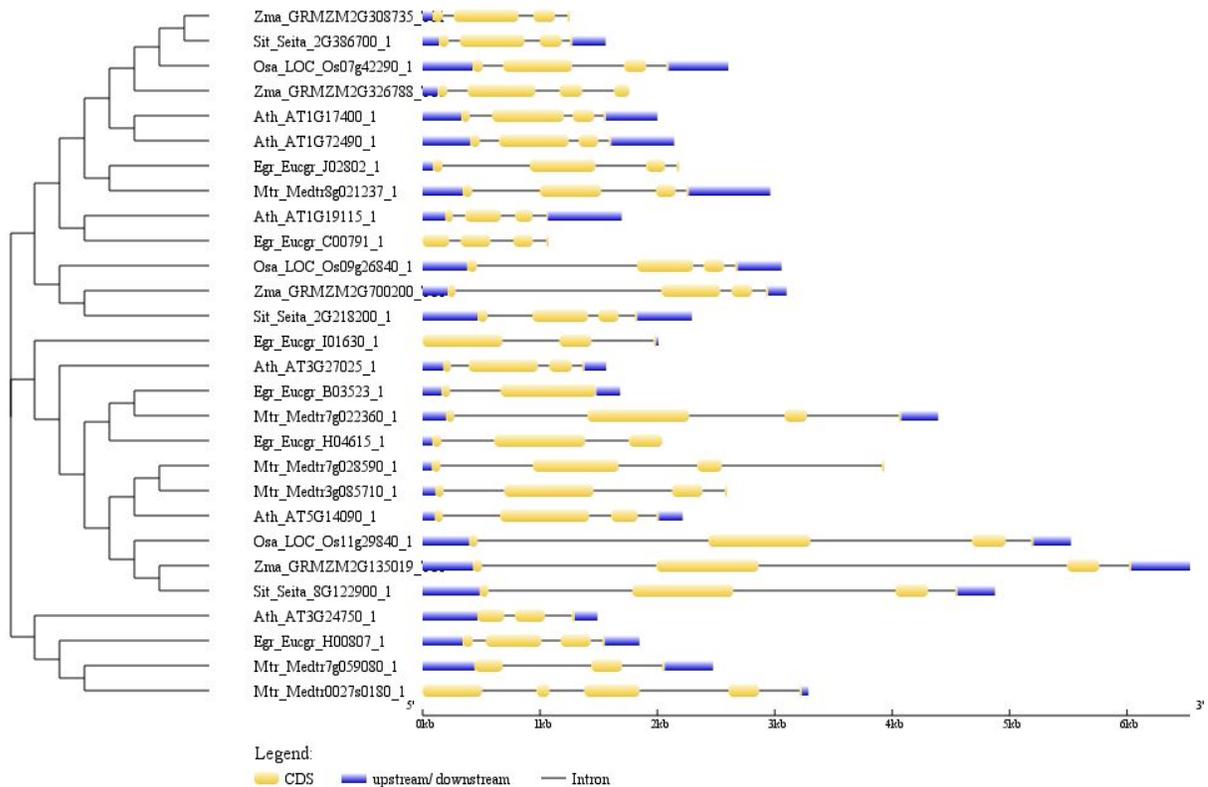


Figura 3: Organização estrutural dos genes LAZY de *Arabidopsis thaliana* e os homólogos à família em *Eucalyptus grandis*, *Oryza sativa*, *Zea mays*, *Medicago truncatula* e *Setaria italica* obtido a partir de suas sequências e da árvore filogenética (construída previamente pelo MEGA X) utilizando o software GSDS.

5.4 Análise de motivos e domínios proteicos

Seguindo a avaliação da estrutura dos genes homólogos, foram realizadas análises que avaliaram as seguintes propriedades: tamanho da proteína, distribuição dos resíduos, presença de motivos conservados e de domínios transmembrana putativos. De maneira geral, o tamanho das proteínas variou de 164 a 564, com a média mantendo-se em 289 aminoácidos.

Para verificar a representatividade dos resíduos, foi gerado um gráfico, a partir do alinhamento das sequências proteicas de todos os genes homólogos às LAZY recuperados, que mostra a distribuição deles em cada posição da sequência (Figura 4).

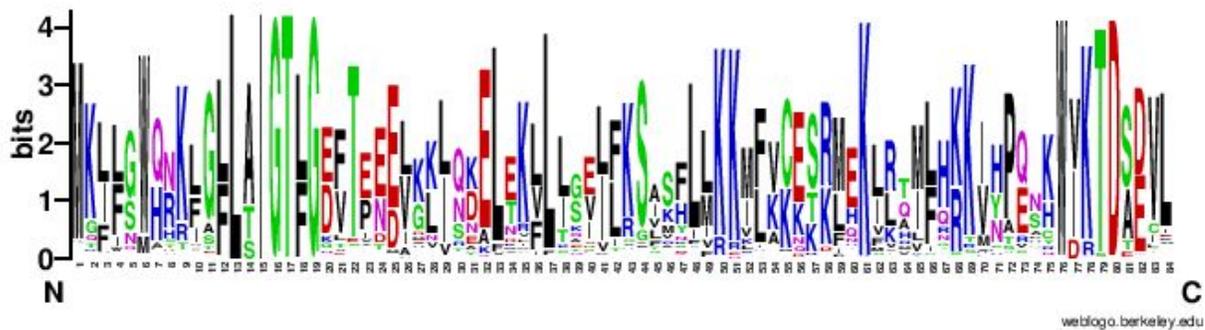


Figura 4: Representatividade de resíduos por posição das sequências proteicas codificadas pelos genes homólogos à *LAZY*, gerado com o programa *WebLogo*.

Para verificar a presença de domínios transmembrana putativos, e assim ter uma noção da posição intracelular das proteínas *LAZY*, foram realizadas duas análises. A primeira delas, feita com o programa *Protter*, revelou que nenhuma das proteínas apresentou domínios transmembrana. A segunda, utilizando o programa *TMHMM*, ratificou estes dados, indicando que não houve presença de hélices transmembrana.

Para refinar a busca por motivos proteicos, além das informações estruturais obtidas a partir do banco de dados, foi utilizado o software *MEME*. Foi feita uma varredura por todas as sequências CDS dos genes da família *LAZY* e homólogos (totalizando 282), buscando por qualquer número de repetições por sequência. Foram encontrados 3 motivos, que tiveram entre 131 e 248 repetições e apresentaram uma diferença estatística significativa ($E\text{-value} \leq 2.6e-718$). Em seguida, estes motivos foram submetidos ao programa Tomtom e GOMo para serem comparados à motivos publicados e identificar possíveis funções, respectivamente. Nenhum deles apresentou associação significativa a um termo GO (de *Gene Ontology*), nem a qualquer motivo já caracterizado no banco de dados do *MEME*. Os principais resultados podem ser acessados na Tabela S2 em anexo.

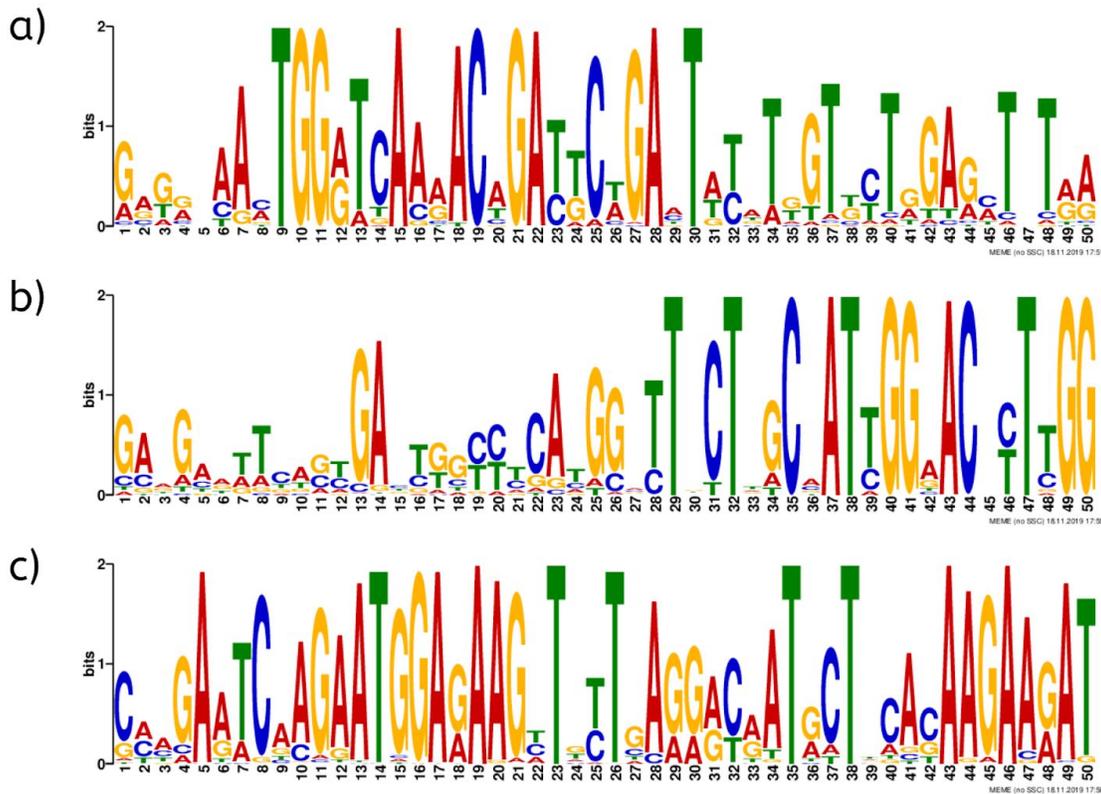


Figura 5. Motivos proteicos identificados nas sequências de *LAZY* e seus homólogos utilizando o MEME. a) Motivo 1, com E -value $3.1e-1011$, 248 repetições e 50 resíduos de comprimento; b) motivo 2, com E -value $8.1e-817$, 221 repetições e 50 resíduos de comprimento; c) motivo 3, com E -value $2.6e-718$, 131 repetições e 50 resíduos de comprimento.

5.5 Análise de expressão gênica *in silico*

Usando dados públicos de RNA-seq disponíveis no *GENEVESTIGATOR*, foram feitas análises de expressão com o objetivo de avaliar padrões de expressão em tecidos e estágios de desenvolvimento, buscando identificar potenciais funções para cada um dos genes *LAZY*. Foram encontradas amostras para todas as seis sequências em *A. thaliana* (AT1G19115, AT3G24750, AT1G72490, AT1G17400, AT3G27025, AT5G14090). Em *O. sativa*, todas as três sequências identificadas por BLAST (LOC_Os09g26840, LOC_Os11g29840, LOC_Os07g42290) tiveram amostras, assim como dez das treze em *G. max* (Glyma_02G097100, Glyma_10G298100, Glyma_20G249200, Glyma_18G284400, Glyma_16G054300, Glyma_09G071100,

Glyma_19G094600, Glyma_19G263200, Glyma_03G264200, Glyma_16G009400) e todas as quatro em *Z. mays* (Zm00001d020620/GRMZM2G135019_T01, Zm00001d049174/GRMZM2G308735_T01, Zm00001d047263/GRMZM2G326788_T01, Zm00001d022133/GRMZM2G700200_T01). Foram analisadas 25 partes anatômicas e 9 estágios de desenvolvimento em *A. thaliana*; 23 partes anatômicas e 7 estágios de desenvolvimento em *G. max*; 13 partes anatômicas e 9 estágios de desenvolvimento em *O. sativa* e 70 partes anatômicas e 7 estágios de desenvolvimento em *Z. mays*. Os dados resumidos encontram-se na Tabela 2 e os resultados das análises encontram-se nas figuras suplementares 1 a 4.

As análises mostraram que nem todas as *LAZY* em *A. thaliana* têm o mesmo padrão de expressão nas diferentes partes anatômicas e que os genes agrupados em um mesmo clado na filogenia tiveram padrões semelhantes. Os genes *LAZY1* e *LAZY6* apresentaram uma intensidade de expressão maior nas partes anatômicas relacionadas a inflorescência e desenvolvimento do hipocótilo, enquanto os genes *LAZY2*, *LAZY3*, *LAZY4* e o próprio *LAZY5* apresentaram maior intensidade nos tecidos de raiz. Os padrões de expressão nos estágios de desenvolvimento condizem com estes resultados, mostrando que os genes *LAZY1* e *LAZY6* apresentam uma intensidade, e por conseguinte, expressão mais constitutiva. Os demais genes apresentam intensidade média nos estágios de germinação e plântula e baixa ou zerada nos estágios seguintes.

As análises em *G. max*, *O. stiva* e *Z. mays* ratificaram também os agrupamentos verificados em *A. thaliana*, visto que os genes homólogos à *LAZY1* e 6 apresentaram maior intensidade nos tecidos relacionados à inflorescência e de forma constitutiva ao longo dos estágios de desenvolvimento e os homólogos à *LAZY2*, 3 e 4 com maior intensidade nos tecidos relacionados ao desenvolvimento de raízes e nos estágios de germinação e plântula. À exceção disto está o gene Glyma.19G094600 de *G. max*, que foi classificado no clado II, junto às *LAZY1* e 6 na filogenia, mas que nesta análise foi agrupado junto aos genes homólogos do clado I, apresentando maior intensidade de

expressão nos tecidos de desenvolvimento da radícula, hipocótilo e caule e no estágio de germinação.

Estes resultados são condizentes com as conclusões de Yoshihara e Spalding nos seus estudos com mutantes *LAZY* em *Arabidopsis* em 2017, quando inseriram o gene repórter GUS (β -glucuronidase) em uma região controlada por promotores dos genes *LAZY* para analisar os padrões normais de expressão. Tanto em plântulas crescidas na luz ou no escuro, os padrões foram similares, tendo os genes *LAZY2-5* mais expressos nos cotilédones, hipocótilo e raiz e os genes *LAZY1* e *LAZY6* mais expressos na parte aérea das plântulas, principalmente no meristema apical e petíola.

Relacionando às análises filogenéticas, os resultados ratificam as hipóteses de eventos evolutivos levantadas, visto que os genes se mantiveram agrupados conforme os clados identificados. Dado que os genes *LAZY2-4* (clado IIa) estão mais próximos ao gene *LAZY5* (clado I), apresentando padrões de expressão em tecidos e em etapas do desenvolvimento semelhantes, enquanto os genes *LAZY1* e *LAZY6* (clado IIb) divergiram, também é possível especular que houve pressão evolutiva, e assim uma seleção positiva, que acarretaram em mudanças funcionais nas proteínas codificadas por estes genes.

Tabela 2. Dados de expressão dos genes *LAZY* em *A. thaliana*, *O. sativa*, *G. max* e *Z. mays*. Informações obtidas a partir da análise de amostras de RNA-seq destas espécies disponíveis no banco de dados *GENEVESTIGATOR*.

Espécie	Gene (Clado)	Partes anatômicas	Estágios do desenvolvimento
<i>Arabidopsis thaliana</i>	AT5G14090 - LAZY1 (1b)	inflorescência, meristema apical, hipocótilo	semente germinada, plântula, roseta jovem, roseta madura, bolting, flor jovem, flor madura, flores e silicas, silicas maduras
	AT1G17400 - LAZY2 (1a)	desenvolvimento da raiz (endoderme, protoplasto, columela, coifa, zona de alongamento, raiz primária)	semente germinada e plântula
	AT1G172490 - LAZY3 (1a)	desenvolvimento da raiz (endoderme, protoplasto, columela, coifa, zona de alongamento, raiz primária)	semente germinada e plântula
	AT1G19115 - LAZY4 (1a)	desenvolvimento da raiz (endoderme, protoplasto, columela, coifa, zona de alongamento, raiz primária)	semente germinada e plântula
	AT3G24750 - LAZY5 (l)	desenvolvimento da raiz (endoderme, protoplasto, columela, coifa, zona de alongamento, raiz primária)	semente germinada e plântula
	AT3G27025 - LAZY6 (1b)	inflorescência, meristema apical, hipocótilo	semente germinada e plântula
<i>Oryza sativa</i>	LOC_Os09g26840 (1a)	calo, radícula, zonas de maturação e alongamento, inflorescência, raiz inflorescência, broto	semente germinada, plântula, roseta jovem, roseta madura, brotamento, flor jovem, flor madura, flores e silicas, silicas maduras
	LOC_Os11g29840 (1b)	inflorescência, broto	plântula, inflorescência
	LOC_Os07g42290 (1a)	calo, radícula, raiz	plântula, perfilhamento
	Glyma_02G097100 (1b)	hipocótilo, radícula, internós, broto	plântula, perfilhamento
<i>Glycine max</i>	Glyma_10G298100 (1b)	hipocótilo, radícula, internós, broto	crescimento do broto principal
	Glyma_20G249200 (1b)	hipocótilo, radícula, internós, broto	crescimento do broto principal
	Glyma_18G284400 (1b)	hipocótilo, radícula, internós, broto	crescimento do broto principal
	Glyma_16G054300 (1b)	hipocótilo, radícula, internós, broto	germinação, crescimento do broto principal, formação da inflorescência, florescimento, formação do fruto, desenvolvimento do grão, amadurecimento
	Glyma_09G071100 (1a)	radícula, zona de maturação, hipocótilo, caule	germinação
	Glyma_19G094600 (1b)	radícula, zona de maturação, hipocótilo, caule	germinação
	Glyma_19G263200 (1a)	radícula, zona de maturação, hipocótilo, caule	germinação
	Glyma_03G284200 (1a)	radícula, zona de maturação, hipocótilo, caule	germinação
	Glyma_16G009400 (1a)	radícula	germinação
	Zm00001d020620 / GRMZM2G135019_T01 (1b)	plântula, coleótilo, raiz, radícula, maturação, córtex, inflorescência, pericarpo, broto, nó, folhas	germinação, alongamento de caule, formação da inflorescência, anese (maturação da flor), formação do fruto, desenvolvimento do grão
<i>Zea mays</i>	Zm00001d049174 / GRMZM2G308735_T01 (1a)	plântula, coleótilo, raiz, radícula, maturação, córtex, inflorescência, pericarpo, broto, nó, folhas	germinação, alongamento de caule, formação da inflorescência, anese (maturação da flor), formação do fruto, desenvolvimento do grão
	Zm00001d047263 / GRMZM2G326788_T01 (1a)	raiz, radícula, zona de maturação, esteia	germinação
	Zm00001d022133 / GRMZM2G700200_T01 (1a)	radícula, nó, raiz	germinação
	Zm00001d020620 / GRMZM2G135019_T01 (1b)	plântula, coleótilo, raiz, radícula, maturação, córtex, inflorescência, pericarpo, broto, nó, folhas	germinação, alongamento de caule, formação da inflorescência, anese (maturação da flor), formação do fruto, desenvolvimento do grão

6 CONCLUSÕES E PERSPECTIVAS

Levando em consideração todas as análises realizadas e observações levantadas neste trabalho, é possível concluir que os genes *LAZY* têm relação direta com os grandes eventos evolutivos que estão relacionados à mudanças na arquitetura das plantas terrestres. Algas e outras plantas mais antigas na história evolutiva não possuem estes genes, visto que as funções que eles desempenham não eram necessários para estas espécies. Conforme surgiram estas necessidades, os genes *LAZY* surgiram, se duplicaram e se diversificaram, estando presentes em todas as espécies de angiospermas disponíveis no banco de dados utilizado.

Como perspectiva futura para este trabalho, uma análise que pode agregar informações ao entendimento da história evolutiva desta família gênica é a de pressão seletiva, utilizando programas como o *CODEML* em *PAML* (v.4.9) (YANG, 2007) para computar as taxas de substituição não sinônimas para substituições sinônimas (dN/dS ou ω) e calcular homogeneidade e seleção positiva para cada sítio das sequências, verificando o tipo de seleção agindo sobre estes genes (positiva, neutra ou purificadora).

REFERÊNCIAS

- ABASCAL, F; ZARDOYA, R and POSADA, D. ProtTest: selection of best-fit models of protein evolution. **Bioinformatics**. 21(9):2104-2105, 2005.
- ABE, K; TAKAHASHI, H and SUJI, H. Lazy Gene (la) Responsible for Both an Agravitropism of Seedlings and Lazy Habit of Tiller Growth in Rice (*Oryza sativa* L.). **Journal of Plant Research**. 109:381-386, 1996.
- AMES, R; MONEY, D; GHATGE, V; WHELAN, S and LOVELL, S. Determining the evolutionary history of gene families. **Bioinformatics**. 28(1):48-55, 2011.
- ARMENGOT, L; MARQUÈS-BUENO, M and JAILLAIS, Y. Regulation of polar auxin transport by protein and lipid kinases. **Journal of Experimental Botany**. 67(14):4015-4037, 2016.
- AWADI, A; BEN SLIMEN, H; SMITH, S; KNAUER, F; MAKNI, M and SUCHENTRUN, F. Positive selection and climatic effects on MHC class II gene diversity in hares (*Lepus capensis*) from a steep ecological gradient. **Scientific Reports**. 8(1), 2018.
- BAILEY, T; WILLIAMS, N; MISLEH, C and LI, W. MEME: discovering and analyzing DNA and protein sequence motifs. **Nucleic Acids Research**. 34(Web Server):W369-W373, 2006.
- BELINKY, F; BABENKO, V; ROGOZIN, I and KOONIN, E. Purifying and positive selection in the evolution of stop codons. **Scientific Reports**. 8(1), 2018.
- CHAPMAN, M; LEEBENS-MACK, J and BURKE, J. Positive Selection and Expression Divergence Following Gene Duplication in the Sunflower CYCLOIDEA Gene Family. **Molecular Biology and Evolution**. 25(7):1260-1273, 2008.
- CASTRESANA, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. **Molecular Biology and Evolution**. 17:540-552, 2000.
- CHEN, R; ROSEN, E and MASSON, P. Gravitropism in Higher Plants. **Plant Physiology**. 120:343-350, 1999.
- CROOKS, G; HON, G; CHANDONIA, J and BRENNER, S. WebLogo: A Sequence Logo Generator. **Genome Research**. 14(6):1188-1190, 2004.
- DARRIBA, D; TABOADA, G; DOALLO, R and POSADA, D. jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**. 9(8):772-772, 2012.
- DARWIN, C and DARWIN, F. **The power of movement in plants**. London, 1880.

DONG, W; VANNOZZI, A; CHEN, F; HU, Y; CHEN, Z and ZHANG, L. MORC Domain Definition and Evolutionary Analysis of the MORC Gene Family in Green Plants. **Genome Biology and Evolution**. 10(7):1730-1744, 2018.

DONG, Z; JIANG, C; CHEN, X; ZHANG, T; DING, L; SONG, W; LUO, H; LAI, J; CHEN, H; LIU, R; ZHANG, X and JIN, W. Maize LAZY1 Mediates Shoot Gravitropism and Inflorescence Development through Regulating Auxin Transport, Auxin Signaling, and Light Response. **Plant Physiology**. 163(3):1306-1322, 2013.

DRUMMOND, AJ; SUCHARD, MA; XIE, D and Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. **Molecular Biology and Evolution**. 29:1969–1973, 2012.

EDGAR, R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**. 32(5):1792-1797, 2004.

GOLDMAN, N and YANGF, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. **Molecular Biology and Evolution**. 11(5):725-736, 1994.

GUO, H.-S; ZHANG, Y.-M; SUN, X.-Q; LI, M.-M; HANG, Y.-Y and XUE, J.-Y. Evolution of the KCS gene family in plants: the history of gene duplication, sub/neofunctionalization and redundancy. **Molecular Genetics and Genomics**. 291(2):739–752, 2015.

HRUZ, T; LAULE, O; SZABO, G; WESSENDORP, F; BLEULER, S; OERTLE, L; WIDMAYER, P; GRUISSEM, W and ZIMMERMANN, P. Genevestigator v3: A reference expression database for the meta-analysis of transcriptomes. **Adv Bioinformatics**. 2008:420747, 2008.

KLINGENBERG, C. Evolution and development of shape: integrating quantitative approaches. **Nature Reviews Genetics**. 11(9):623-635, 2010.

KNIGHT, T. **On the Direction of the Radicle and Germen during the Vegetation of Seeds**. Londres: Proceedings of the Royal Society of London, 1(0):218-220, 1800.

KROGH, A; LARSSON, B; VON HEIJNE, G and SONNHAMMER, E. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. **Journal of Molecular Biology**. 305(3):567-580, 2001.

KUMAR, S; STECHER, G and TAMURA, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. **Molecular Biology and Evolution**. 33(7):1870-1874, 2016.

- MAGADUM, S; BANERJEE, U; MURUGAN, P; GANGAPUR, D and RAVIKESAVAN, R. Gene duplication as a major force in evolution. **Journal of Genetics**. 92(1):155-161, 2013.
- NARAMOTO, S. Polar transport in plants mediated by membrane transporters: focus on mechanisms of polar auxin transport. **Cell Biology**. 40:8-14, 2017.
- NIELSEN, R and YANG, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. **Genetics**. 148(3):929-36, 1998.
- OMASITS, U; AHRENS, C; MÜLLER, S and WOLLSCHIED, B. Protter: interactive protein feature visualization and integration with experimental proteomic data. **Bioinformatics**. 30(6):884-886, 2013.
- PETRASEK, J and FRIML, J. Auxin transport routes in plant development. **Development**. 136(16):2675-2688, 2009.
- RAMBAUT, A; DRUMMOND, A; XIE, D; BAELE, G and SUCHARD, M. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. **Systematic Biology**. 67(5):901-904, 2018.
- SEGATTO, A; THOMPSON, C and FREITAS, L. Molecular evolution analysis of WUSCHEL-related homeobox transcription factor family reveals functional divergence among clades in the homeobox region. **Development Genes and Evolution**. 226(4):259-268, 2016.
- SU, S; GIBBS, N; JANCEWICZ, A and MASSON, P. Molecular Mechanisms of Root Gravitropism. **Current Biology**. 27(17):R964-R972, 2017.
- TALAVERA, G and CASTRESANA, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. **Systematic Biology**. 56:564-577, 2007.
- TANIGUCHI, M; FURUTANI, M; NISHIMURA, T; NAKAMURA, M; FUSHITA, T; IJIMA, K; BABA, K; TANAKA, H; TOYOTA, M; TAsAKA, M and MORITA, M. The Arabidopsis LAZY1 Family Plays a Key Role in Gravity Signaling within Statocytes and in Branch Angle Control of Roots and Shoots. **The Plant Cell**. 29(8):1984-1999, 2017.
- TEALE, W and PALME, K. Naphthylphthalamic acid and the mechanism of polar auxin transport. **Journal of Experimental Botany**. 69(2):303-312, 2017.
- VANDENBRINK, J and KISS, J. Plant responses to gravity. **Seminars in Cell & Developmental Biology**. 92:122-125, 2019.

WONG, W; YANG, Z; GOLDMAN, N and NIELSEN, R. Accuracy and Power of Statistical Methods for Detecting Adaptive Evolution in Protein Coding Sequences and for Identifying Positively Selected Sites. **Genetics**. 168(2):1041-1051, 2004.

YANG, Z. Maximum Likelihood Estimation on Large Phylogenies and Analysis of Adaptive Evolution in Human Influenza Virus A. **Journal of Molecular Evolution**. 51(5):423-432, 2000.

YANG, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. **Molecular Biology and Evolution**. 24(8):1586-1591, 2007.

YOSHIHARA, T and IINO, M. Identification of the Gravitropism-Related Rice Gene LAZY1 and Elucidation of LAZY1-Dependent and -Independent Gravity Signaling Pathways. **Plant and Cell Physiology**. 48(5):678-688, 2007.

YOSHIHARA, T; SPALDING, E and IINO, M. AtLAZY1 is a signaling component required for gravitropism of the *Arabidopsis thaliana* inflorescence. **The Plant Journal**. 74(2):267-279, 2013.

YOSHIHARA, T and SPALDING, E. LAZY genes mediate the effects of gravity on auxin gradients and plant architecture. **Plant Physiology**. 175:959-969, 2017.

YUAN, Y; ZHANG, Q; ZENG, S; GU, L; SI, W; ZHANG, X; TIAN, D; YANG, S and WANG, L. Selective sweep with significant positive selection serves as the driving force for the differentiation of japonica and indica rice cultivars. **BMC Genomics**. 18(1), 2017.

ŽÁDNÍKOVÁ, P; SMET, D; ZHU, Q; STRAETEN, D and BENKOVÁ, E. Strategies of seedlings to overcome their sessile nature: auxin in mobility control. **Frontiers in Plant Science**. 6, 2015.

ANEXOS

Figura S1. Intensidade e padrão de expressão gênica em *Arabidopsis thaliana*, em (a) diferentes tecidos e (b) diferentes estágios de desenvolvimento.

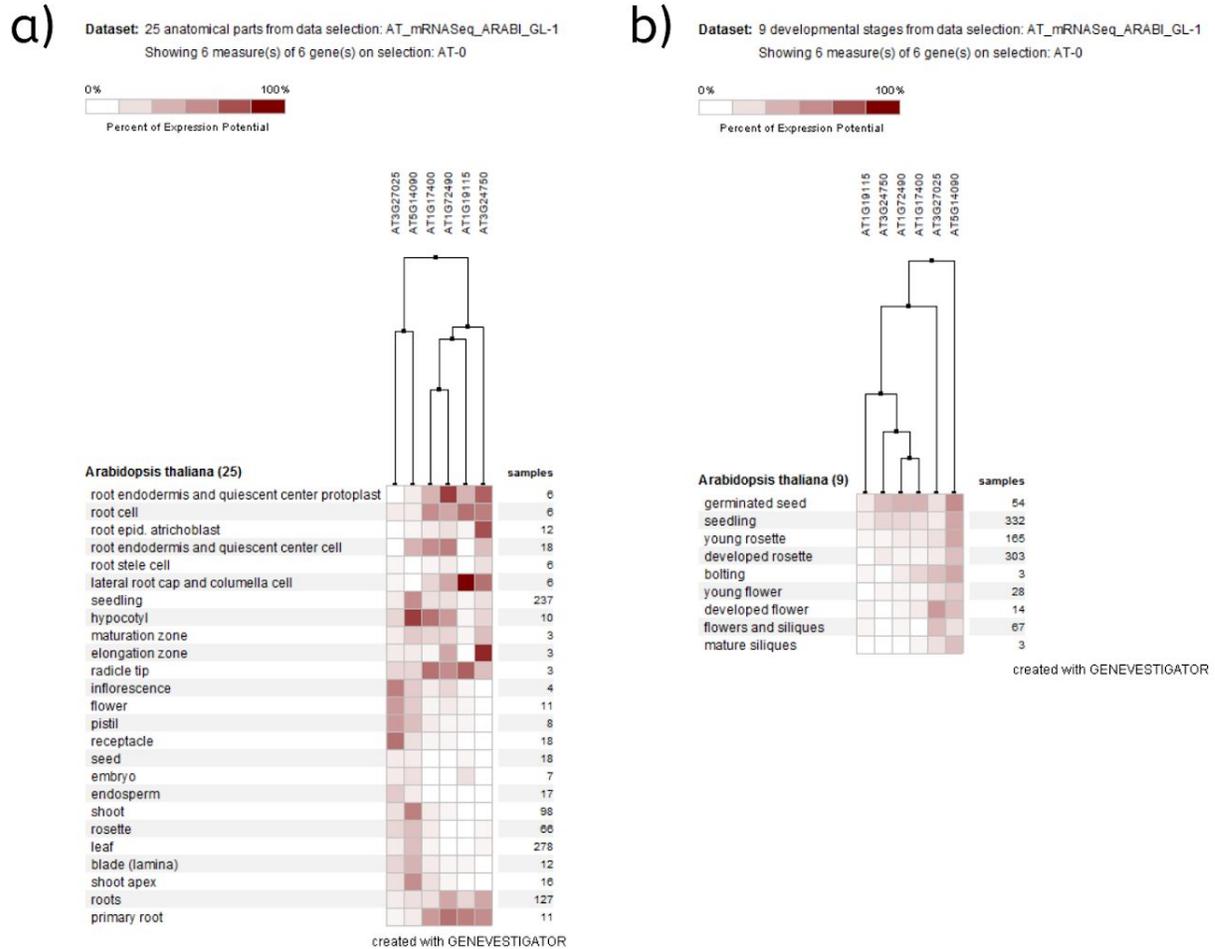


Figura S2. Intensidade e padrão de expressão gênica em *Glycine max*, em (a) diferentes tecidos e (b) diferentes estágios de desenvolvimento.

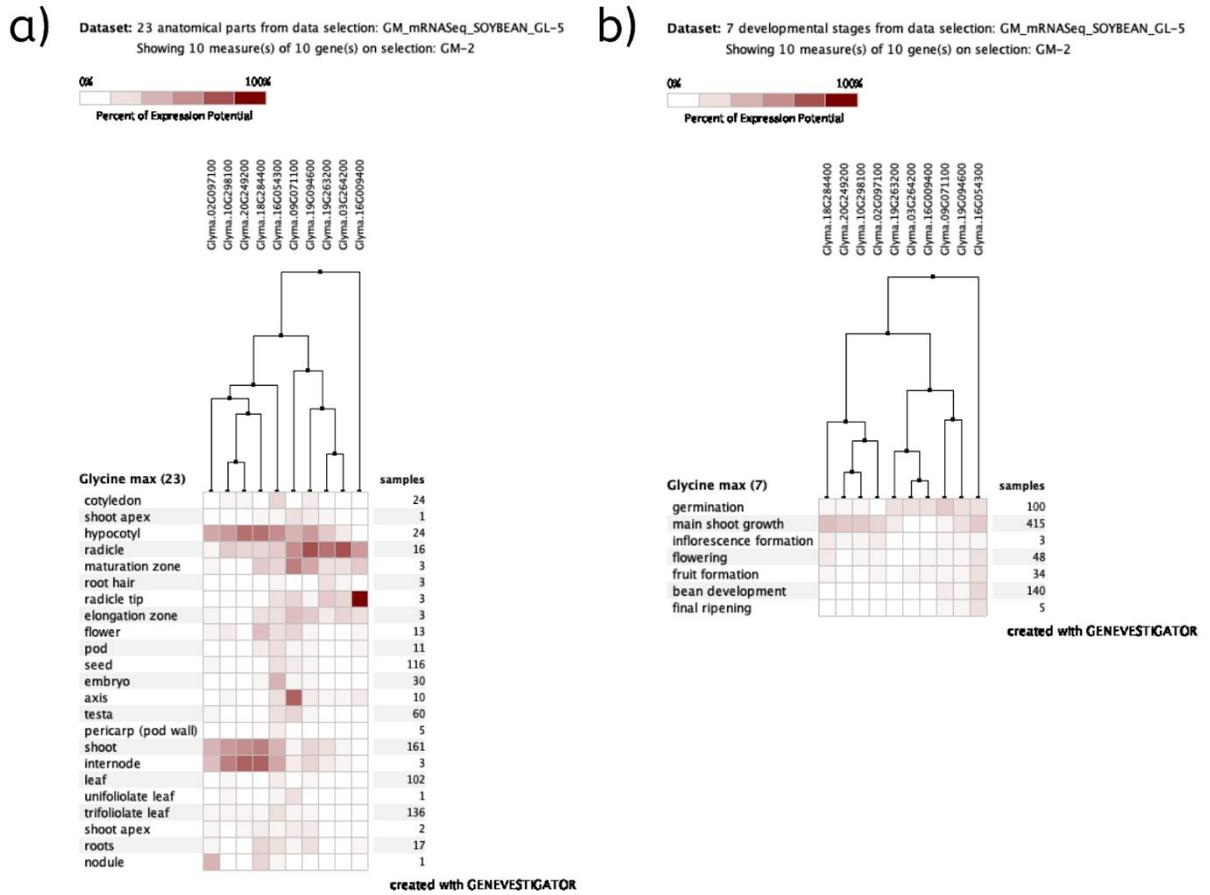


Figura S3. Intensidade e padrão de expressão gênica em *Oryza sativa*, em (a) diferentes tecidos e (b) diferentes estágios de desenvolvimento.

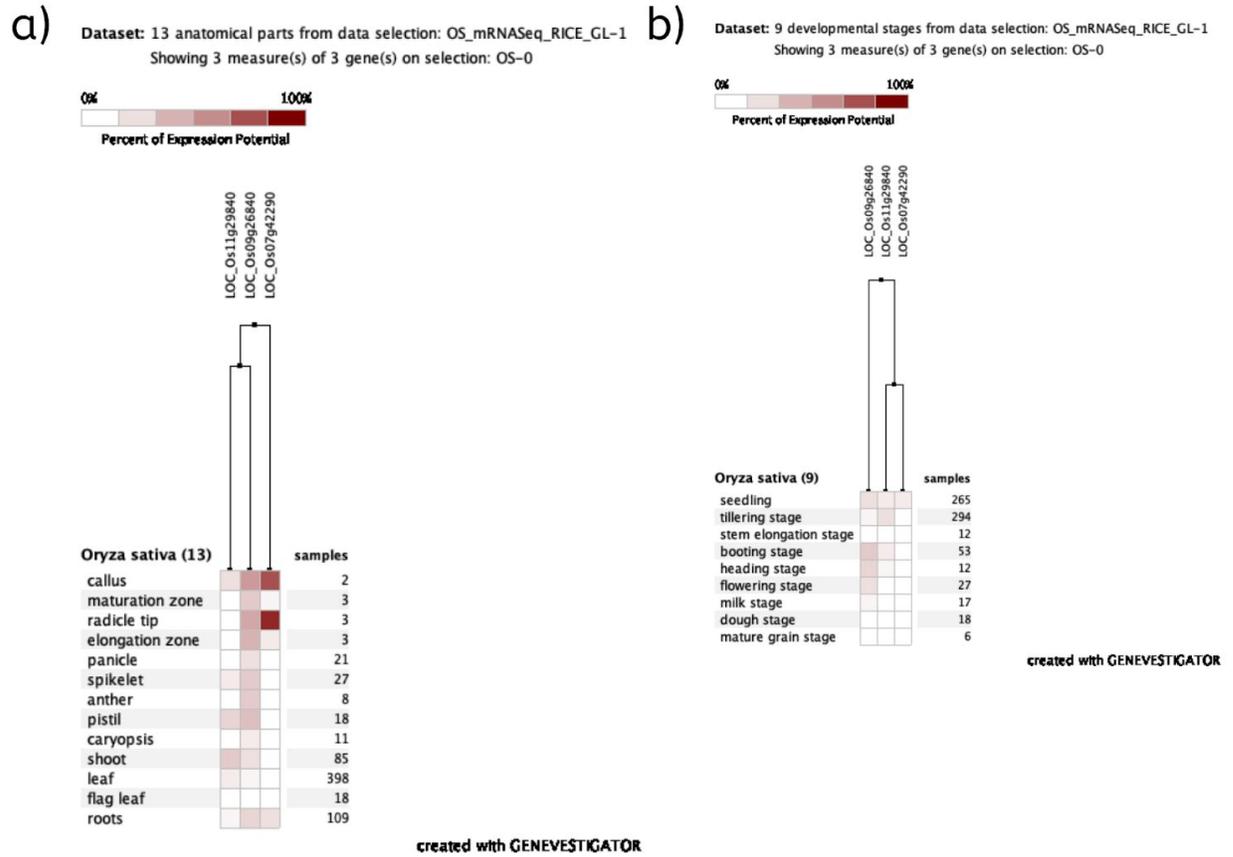


Figura S4. Intensidade e padrão de expressão gênica em *Zea mays*, em (a) diferentes tecidos e (b) diferentes estágios de desenvolvimento.

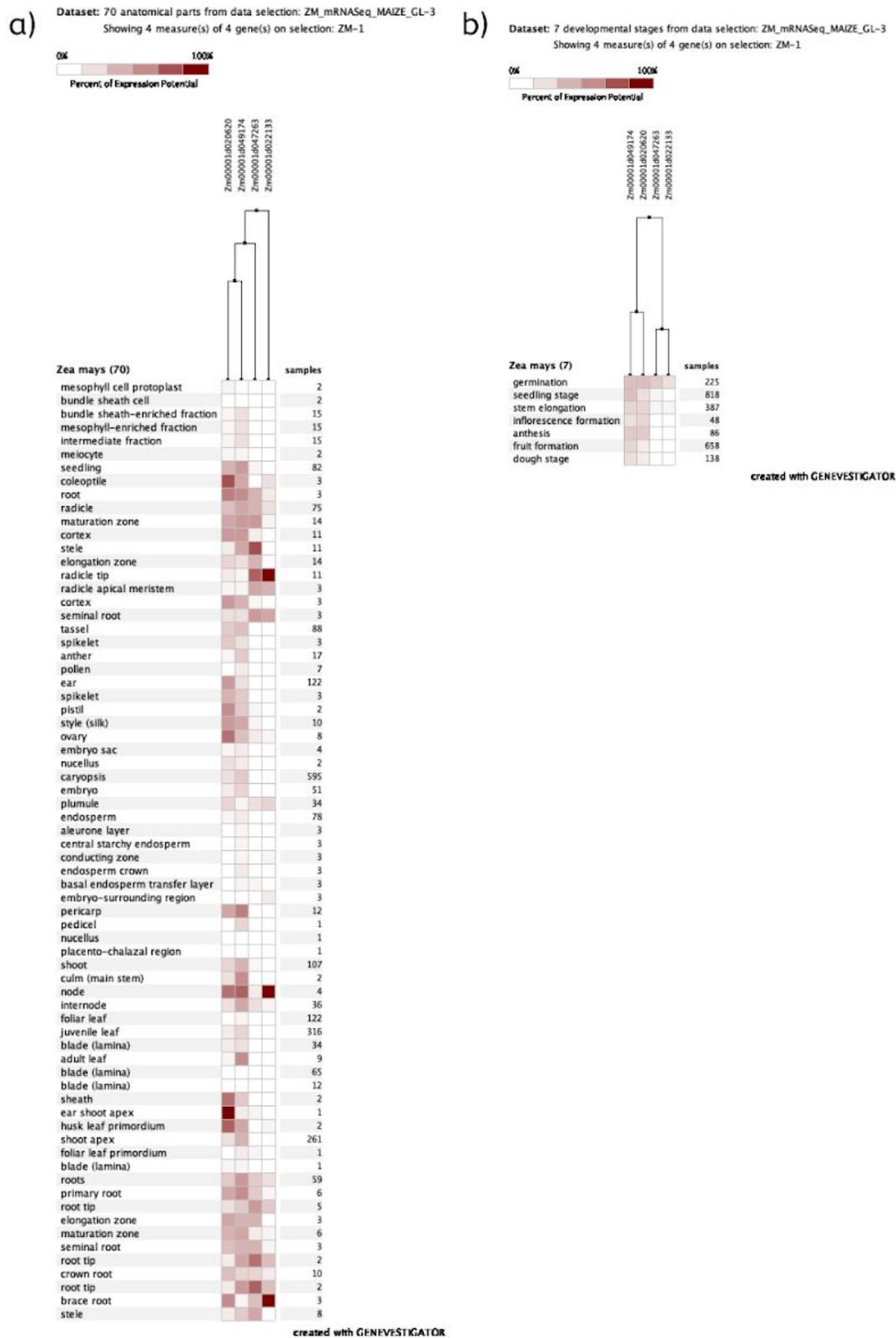


Tabela S1. Dados taxonômicos completos, com número de genes, número de íntrons e comprimento da sequência proteica por espécie dos homólogos à LAZY recuperados por BLAST e distribuição por clados de acordo com a filogenia.

Classificação	Família	Espécie	Acronimo	Nº de íntrons (intervalo; média)	Comprimento da proteína (intervalo; média)	Nº de genes	I	Clado IIa	IIb	
Alga	Chlamydomonadales	<i>Chlamydomonas reinhardtii</i>	Cre	-	-	-	-	-	-	
	Dunaliellaceae	<i>Dunaliella salina</i>	Dsa	-	-	-	-	-	-	
	Volvocales	<i>Volvox carterii</i>	Vca	-	-	-	-	-	-	
	Coccomyxa	<i>Coccomyxa subellipsoidea</i>	Csu	-	-	-	-	-	-	
	Mamiellales	<i>Micromonas pusilla</i>	Mpu	-	-	-	-	-	-	
	Mamiellales	<i>Micromonas sp.</i>	Msp	-	-	-	-	-	-	
	Bathycoccales	<i>Ostreococcus lucimarinus</i>	Olu	-	-	-	-	-	-	
	Hepática	Marchantiaceae	<i>Marchantia polymorpha</i>	Mpo	-	-	-	-	-	
Musgo	Funariaceae	<i>Physcomitrella patens</i>	Fpa	-	-	-	-	-	-	
	Sphagnaceae	<i>Sphagnum falax</i>	Sfa	-	-	-	-	-	-	
Licófito	Selaginellales	<i>Selaginella moellendorffii</i>	Smo	-	-	-	-	-		
Monocotiledónea	Musaceae	<i>Musa acuminata</i>	Mec	3 - 4; 3,4	315 - 398; 340,6	4	-	3	1	
	Bromeliaceae	<i>Ananas comosus</i>	Acom	2 - 4; 3,6	228 - 328; 296,6	4	1	2	1	
	Araceae	<i>Spirodela polytriza</i>	Spo	0	294	1	-	-	1	
	Zosteraceae	<i>Zostera marina</i>	Zmar	3 - 4; 3,6	256 - 304; 274,3	2	-	2	-	
	Amborellaceae	<i>Amborella trichopoda</i>	Atr	0 - 4; 3	267 - 272; 268,25	2	-	1	1	
	Poaceae	<i>Brachypodium distachyon</i>	Bdi	3 - 5; 4	248 - 393; 295,3	4	-	3	1	
	Poaceae	<i>Brachypodium stacei</i>	Bsta	4 - 5; 4,6	248 - 394; 305,6	4	-	3	1	
	Poaceae	<i>Oryza sativa</i>	Osa	2 - 5; 4	206 - 416; 292	3	-	2	1	
	Poaceae	<i>Oryzopsis latipes</i>	Ola	1 - 2; 1,5	269 - 385; 327	2	-	1	1	
	Poaceae	<i>Panicum hallii</i>	Pha	3 - 5; 4,3	249 - 426; 318	4	-	3	1	
	Poaceae	<i>Panicum virgatum</i>	Pvi	3 - 4; 4,3	264 - 616; 323	7	-	6	1	
	Poaceae	<i>Setaria italica</i>	Sit	3 - 5; 4,2	202 - 411; 285,7	3	-	2	1	
	Poaceae	<i>Setaria viridis</i>	Svi	4 - 5; 4,3	249 - 411; 313,6	3	-	2	1	
	Poaceae	<i>Sorghum bicolor</i>	Sbi	2 - 5; 4	209 - 419; 290,4	4	-	3	1	
	Poaceae	<i>Zea mays</i>	Zma	2 - 4; 3,5	193 - 413; 304,2	4	-	3	1	
	Eudicotiledónea	Amaranthaceae	<i>Amaranthus hypochondriacus</i>	Ahy	1 - 4; 2,3	273 - 378; 326,6	4	-	2	2
		Phymalaceae	<i>Mimulus guttatus</i>	Mgu	1 - 4; 2,2	253 - 328; 283	5	1	3	1
		Crassulaceae	<i>Kalanchoe fedtschenkoi</i>	Kfe	4	262 - 378; 320	3	-	2	1
		Crassulaceae	<i>Kalanchoe laxiflora</i>	Kla	2 - 4; 3,5	206 - 378; 277	8	-	4	4
		Linaceae	<i>Linum usitatissimum</i>	Lus	2 - 3; 2,5	193 - 387; 279	9	2	6	1
Euphorbiaceae		<i>Manihot esculenta</i>	Mes	3 - 5; 4	277 - 391; 330,2	8	1	3	4	
Euphorbiaceae		<i>Ricinus communis</i>	Rco	3 - 4; 3,7	241 - 364; 281,2	5	1	2	2	
Salicaceae		<i>Populus trichocarpa</i>	Ptr	2 - 4; 3,4	266 - 390; 317,2	7	1	3	3	
Salicaceae		<i>Salix purpurea</i>	Spu	3 - 4; 3,5	265 - 379; 295,2	7	1	3	3	
Rutaceae		<i>Citrus sinensis</i>	Csi	2 - 4; 3	197 - 373; 286,6	5	1	2	2	
Rutaceae		<i>Citrus clementina</i>	Ccl	2 - 4; 3,3	197 - 380; 287	8	4	2	2	
Caricaceae		<i>Carica papaya</i>	Cpa	4	270-368; 307,6	3	1	1	1	
Ranunculaceae		<i>Aquilegia coerulea</i>	Aco	4 - 5; 4,3	263 - 364; 303	3	1	1	1	
Apiaceae		<i>Daucus carota</i>	Dca	2 - 3; 2,3	173 - 377; 271	6	1	2	3	
Solanaceae		<i>Solanum lycopersicum</i>	Sly	3 - 5; 4,2	267 - 380; 301	6	2	3	1	
Solanaceae		<i>Solanum tuberosum</i>	Stu	2 - 5; 3	218 - 382; 276	5	1	1	2	
Myrtaceae		<i>Eucalyptus grandis</i>	Egr	2 - 4; 3	273 - 380; 314	6	1	2	3	
Vitaceae		<i>Vitis vinifera</i>	Vvi	4 - 5; 4,6	239 - 475; 389	4	1	1	2	
Malvaceae		<i>Gossypium hirsutum</i>	Gha	4 - 5; 4,5	276 - 366; 321	9	2	3	4	
Malvaceae		<i>Theobroma cacao</i>	Tca	4 - 6; 5	269 - 513; 351,3	5	1	2	2	
Brassicaceae		<i>Arabidopsis thaliana</i>	Ath	4 - 5; 4,3	176 - 357; 277,5	6	1	3	2	
Brassicaceae		<i>Arabidopsis lyrata</i>	Aly	4 - 5; 4,5	175 - 358; 277,6	6	1	3	2	
Brassicaceae		<i>Arabidopsis thaliana</i>	Ath	3 - 5; 4	174 - 358; 262	6	1	3	2	
Brassicaceae		<i>Boechera stricta</i>	Bst	3 - 5; 3,6	173 - 346; 281,5	6	1	3	2	
Brassicaceae		<i>Brassica oleracea capitata</i>	Bol	3 - 4; 3,2	175 - 378; 293	11	-	7	4	
Brassicaceae		<i>Brassica napus</i>	Bra	3 - 5; 3,8	173 - 346; 274,5	11	1	6	4	
Brassicaceae		<i>Capsella grandiflora</i>	Cgr	3 - 5; 3,8	171 - 357; 282,6	6	1	3	2	
Brassicaceae		<i>Capsella rubella</i>	Cru	1 - 4; 3,3	171 - 357; 263,8	6	1	3	2	
Brassicaceae		<i>Eutrema salsugineum</i>	Esa	3 - 4; 3,6	167 - 357; 264,3	6	1	3	2	
Cucurbitaceae		<i>Cucumis sativus</i>	Csa	2 - 5; 4	270 - 312; 286,6	4	1	1	2	
Rosaceae		<i>Fragaria vesca</i>	Fve	1 - 8; 4,3	179 - 468; 348,3	4	1	2	1	
Rosaceae		<i>Prunus persica</i>	Ppe	3 - 5; 4,3	268 - 487; 380,3	5	1	2	2	
Rosaceae		<i>Malus domestica</i>	Mdo	3 - 4; 3,5	260 - 403; 306,5	7	1	2	4	
Fabaceae		<i>Medicago truncatula</i>	Mtr	3 - 5; 4	173 - 342; 259	6	2	1	3	
Fabaceae		<i>Phaseolus vulgaris</i>	Pvu	4 - 5; 4,2	261 - 380; 322	7	1	3	3	
Fabaceae		<i>Glycine max</i>	Gma	4 - 5; 4,5	186 - 382; 297	13	3	4	6	
Fabaceae		<i>Trifolium pratense</i>	Tpr	4 - 5; 4,3	263 - 373; 300	5	1	1	3	
							278	43	131	103

Tabela S2. Motivos proteicos identificados com o programa *MEME*, motivos caracterizados (*Tomtom*) e termos de ontologia genética (*GOMo*) associados à eles.

Nome	Motivo (MEME)			Termo de Ontologia Genética (GOMo)	Motivo caracterizado associado (Tomtom)		
	E-value	Repetições (sítios)	Comprimento (resíduos)		Nome	E-value	p-value
Motivo 1	3.1e-1011	248	50	N/A	C2C2gata_tnt.GATA14_col_a_m1 (GATA14)	5.28e+00	6.05e-03
Motivo 2	8.1e-817	221	50	N/A	GRF_tnt.AIGRF6_col_a_m1 (AIGRF6)	4.87e+00	5.59e-03
					bHLH_tnt.bHLH80_colamp_a_m1 (bHLH80)	4.59e+00	5.26e-03
					ABI3VP1_tnt.AT5G25475_col_a_m1 (AT5G25475)	4.79e+00	5.50e-03
					bHLH_tnt.bHLH122_col100_a_m1 (bHLH122)	6.00e+00	6.88e-03
					bHLH_tnt.bHLH80_col_a_m1 (bHLH80)	6.60e+00	7.57e-03
					bHLH_tnt.bHLH122_colamp_a_m1 (bHLH122)	8.17e+00	9.36e-03
					NAC_tnt.NST1_colamp_a_m1 (NST1)	8.96e+00	1.03e-02
					bHLH_tnt.bHLH130_col_a_m1 (bHLH130)	9.39e+00	1.08e-02
					HB_tnt.LMI1_col_a_m1 (LMI1)	8.27e-01	9.48e-04
					MYBrelated_tnt.AT1G18960_col_a_m1 (AT1G18960)	8.65e-01	9.92e-04
					C2C2gata_tnt.GATA16_col_a_m1 (GATA16)	8.68e-01	9.95e-04
					Motivo 3	2.6e-718	131
Homeobox_tnt.ATHB6_colamp_a_m1 (ATHB6)	1.22e+00	1.40e-03					
Homeobox_tnt.ATHB13_colamp_a_m1 (ATHB13)	1.44e+00	1.65e-03					
Homeobox_tnt.HAT5_colamp_a_m1 (HAT5)	1.83e+00	2.09e-03					
Homeobox_tnt.ATHB7_col_a_m1 (ATHB7)	1.84e+00	2.11e-03					
HSF_tnt.HSFA6A_col_a_m1 (HSFA6A)	2.48e+00	2.84e-03					
Homeobox_tnt.ATHB20_colamp_a_m1 (ATHB20)	2.52e+00	2.89e-03					
Homeobox_tnt.HAT5_col_a_m1 (HAT5)	2.95e+00	3.38e-03					
Homeobox_tnt.ATHB20_col_a_m1 (ATHB20)	3.18e+00	3.65e-03					
C2C2gata_tnt.ZML2_col_a_m1 (ZML2)	3.40e+00	3.90e-03					
HB_tnt.ATHB5_col_a_m1 (ATHB5)	4.29e+00	4.92e-03					
C2H2_tnt.AT4G26030_col_a_m1 (AT4G26030)	5.53e+00	6.35e-03					
C2C2gata_tnt.GATA15_col_b_m1 (GATA15)	7.20e+00	8.26e-03					
Homeobox_tnt.ATHB13_col_a_m1 (ATHB13)	8.13e+00	9.32e-03					
NAC_tnt.NAC2_colamp_a_m1 (NAC2)	8.50e+00	9.75e-03					
MADS_tnt.AGL13_col_b_m1 (AGL13)	9.07e+00	1.04e-02					