UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

HISTÓRIA EVOLUTIVA DAS OSMOTINAS EM PLANTAS E SEU PAPEL NA RESPOSTA À SECA EM SOJA [*Glycine max* (L.) Merrill]

Giulia Ramos Faillace

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular)

Orientadora: Maria Helena Bodanese Zanettini

Porto Alegre

Abril

2019

INSTITUIÇÃO E FONTE FINANCIADORA

Este trabalho foi desenvolvido nos laboratórios de (i) Genética Molecular Vegetal e (ii) Cultura de Tecidos e Transformação Genética de Plantas do Departamento de Genética da Universidade Federal do Rio Grande do Sul.

Recebeu apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Chamada INCT – MCTI/CNPq/CAPES/FAPs nº 16/2014, Ativos 369 Biotecnológicos Aplicados a Seca e Pragas em Culturas Relevantes para o Agronegócio (INCT Biotec 370 Seca-Pragas) [88887.136360/2017-00 - 465480/2014-4].

“Até o que parece imperceptível se torna visível
aos olhos de quem se propõe a ver”

Giulia R. Faillace

AGRADECIMENTOS

Por diversas vezes, a caminho da universidade, pensei nos meus agradecimentos. Realmente aguardei muito por esse momento, e agora que ele chegou as palavras insistem em escapar. Apesar de estar aqui, agora, em busca de palavras, acredito que o ato de agradecer deveria estar presente em todos os momentos da nossa vida. Agradecer é muito mais que dizer um obrigado, é sentir-se grato, não só pelos bons momentos, mas por aqueles de dificuldade e muito aprendizado. A vida por si só já é motivo para agradecer e durante o período do doutorado fui colocada à prova quanto ao que sentia em relação a mim mesma e aos processos da vida. Aprendi a ser grata pelos bons e maus momentos, a tirar o melhor resultado das experiências vividas, a ter coragem, persistir, a confiar e acreditar em algo maior, a olhar para dentro. Este período do doutorado foi muito mais que aprender e aprofundar o conhecimento na área da biologia, foi aprender e aprofundar o conhecimento sobre mim mesma, sobre a mente e o coração. E é de coração que agradeço a vida e as pessoas que por ela passaram. Pessoas que direta ou indiretamente fizeram parte deste processo, e que de uma forma ou de outra estavam ali para me ensinar alguma coisa. A vida transcorre e evolui nas nossas relações, conheço mais de mim mesma através do meu olhar para com o outro e do olhar do outro para com a minha pessoa. Estamos todos conectados, fazendo parte de um mesmo jardim, onde algumas sementes germinam e outras não, algumas flores já estão abertas e outras em forma de botão. Cada um tem o seu processo, o seu tempo, somos diferentes, diversos, únicos, e é nas nossas individualidades que nos complementamos e integramos o conhecimento na sua forma mais ampla. Obrigada a todos que complementaram o meu processo até o fechamento deste ciclo, vocês contribuíram muito para torna-lo mais completo e integral.

LISTA DE ABREVIATURAS

ABA: do inglês, A*bscisic Acid*, ácido abscísico

AS: Ácido Salicílico

CRISPR: do inglês, C*lustered Regularly Interspaced Short Palindromic Repeats*, Repetições Palindrômicas Curtas Agrupadas e Regularmente Interespaçadas

DNA: do inglês, *Deoxyribonucleic Acid*, ácido desoxirribonucléico

DREB: do inglês, *Dehydration Responsive Element Binding*

EDTA: do inglês, E*thylenediamine Tetraacetic Acid*, ácido etilenodiamino tetra-acético

EMB48: EMBRAPA 48

GFP: do inglês, *Green Fluorescent Protein*, proteína fluorescente verde

GPCRs: do inglês, *G Protein-Coupled Receptors*, receptores acoplados à proteína G

Ig: imunoglobulina

kDa: do inglês, *kiloDaltons*, quilodaltons

MAPK: do inglês, *Mitogen-Activated Protein Kinase*, proteína-quinase ativada por mitógeno

MeJA: metil jasmonato

PEG: polietileno glicol

PHO36: do inglês, *seven trans-membrane domain receptor-like polypeptides*

pI: ponto isoelétrico

PR: do inglês, *Pathogenesis-Related protein*, proteína relacionada à patogênese

PR5K: do inglês, *PR5-like receptor kinase*, receptor de proteína quinase relacionada à patogenese

P35S CaMV: do inglês, Cauliflower Mosaic Virus 35S Promoter, promotor 35S do vírus mosaico de couve-flor

REDDD: arginina, glutamato e três aspartatos

TLPs: do inglês, *Thaumatin-like Proteins*, proteínas tipo taumatinas

UBQ3-P: do inglês*, Ubiquitin 3 Promoter from A. thaliana*, promotor constitutivo Ubiquitina 3 de Arabidopsis

3D: tridimensional

RESUMO

As plantas são constantemente expostas a estresses bióticos e abióticos. Em resposta a estas adversidades expressam diversas proteínas de defesa, dentre elas as proteínas da família *Thaumatins-like* (TLPs), também conhecidas por *Pathogenesis-related protein* 5 (PR5). As TLPs podem ser encontradas em diferentes espécies vegetais e são denominadas de acordo com a sua atividade biológica. Dentre elas destacam-se as osmotinas, proteínas multifuncionais descobertas inicialmente em células de tabaco, adaptadas a baixos potenciais hídricos. Ao longo do tempo, osmotinas de diferentes espécies vegetais vem sendo identificadas e relacionadas com a tolerância a estresses. Apesar disso, sua origem e diversificação ao longo da evolução da família das TLPs continua desconhecida. Para elucidar esta questão, o presente trabalho se baseou no método Bayesiano e foi construída uma árvore filogenética contendo 722 sequências de 32 espécies de Viridiplantae. O agrupamento de proteínas já caracterizadas como osmotinas com outras putativas osmotinas, permitiu a identificação de sua origem a partir das espermatófitas. Os resultados indicam que a separação filogenética e expansão gênica decorreram do surgimento e organização de motivos exclusivos do grupo das osmotinas, seguidos de duplicações em bloco e em tandem. Além disso, a partir da árvore filogenética gerada, foram confirmadas quatro osmotinas de soja já caracterizadas e identificada uma nova osmotina denominada GmOLPc. Inúmeros trabalhos tem demonstrado que as osmotinas desempenham um importante papel na tolerância à seca em diferentes espécies vegetais. No presente estudo, foi investigada a estrutura e o papel das osmotinas de soja em cultivares contrastantes na resposta à seca. A osmotina de soja denominada P21-like apresentou potencial eletrostático semelhante às osmotinas já caracterizadas como promotoras de tolerância à seca de *Nicotiana tabacum* e *Solanum nigrum*. As análises de expressão gênica das diferentes osmotinas de soja indicaram sua relação com a resposta ao estresse hídrico. Além disso, foi observado um aumento significativo de transcritos dos gene codificadores das osmotinas GmOLPc e P21-like nas folhas e raízes, respectivamente, da cultivar tolerante EMBRAPA 48. Este resultado reforça o possível papel dessas proteínas na tolerância à seca.

ABSTRACT

Plants are constantly exposed to biotic and abiotic stresses. In response to these adversities, they express several defense proteins, including the Thaumatins-like family proteins (TLPs), also known as Pathogenesis-related protein 5 (PR5). TLPs can be found in different plant species and are named according to their biological activity. Among them are the osmotins, multifunctional proteins initially discovered in tobacco cells adapted to low water potentials. Over time, osmotins from different plant species have been identified and related to stress tolerance. Despite this, its origin and diversification throughout the evolution of the TLP family remains unknown. To elucidate this question, the present work was based on the Bayesian method and a phylogenetic tree was constructed containing 722 sequences from 32 species of Viridiplantae. The grouping of proteins already characterized as osmotins with other putative osmotins allowed the identification of their origin from spermatophytes. The results indicate that the phylogenetic separation and gene expansion resulted from the emergence and organization of exclusive motifs of the osmotin group, followed by block and tandem duplications. In addition, from the phylogenetic tree generated, four previously characterized soybean osmotins were confirmed and a new osmotin named GmOLPc was identified. Numerous studies have shown that osmotins play an important role in drought tolerance in different plant species. In the present study, the structure and role of soybean osmotins in response to drought were investigated in soybean contrasting cultivars. The so-called P21-like soybean osmotin presented electrostatic potential similar to the osmotins of *Nicotiana tabacum* and *Solanum nigrum* already characterized as drought tolerance promoters. The gene expression analysis of the different soybean osmotins indicated their relationship with the response to water stress. In addition, a significant increase in transcripts of the GmOLPc and P21-like encoding osmotins in the leaves and roots, respectively, of the tolerant cultivar EMBRAPA 48 was observed. This result reinforces the possible role of these proteins in drought tolerance.

SUMÁRIO

Capítulo I

INTRODUÇÃO

OBJETIVOS

1. INTRODUÇÃO

1.1 Osmotinas e *thaumatins-like*

As plantas são constantemente desafiadas por diversos estresses abióticos e bióticos durante o seu desenvolvimento. Dentre os agentes bióticos encontram-se bactérias, fungos, vírus e insetos, enquanto que os fatores abióticos incluem seca, salinidade, frio, calor, radiação ultra-violeta e ausência ou excesso de nutrientes (Hakim et al., 2018). Em resposta a esses fatores adversos, as plantas desenvolveram um complexo sistema de defesa que inclui alterações metabólicas e a expressão de diversas proteínas (Anzlovar e Dermastia, 2003). Dentre elas destacam-se as osmotinas, proteínas multifuncionais descobertas inicialmente em células de tabaco (*Nicotiana tabacum* var Wisconsin 38) adaptadas a baixos potenciais hídricos (Singh et al., 1987). Por também apresentarem sua expressão induzida em plantas de *Nicotiana silvestre* submetidas à infecção por vírus e ferimentos, posteriormente as osmotinas foram classificadas como proteínas relacionadas à patogênese (PR) (Neale et al., 1990).

Nas plantas, as proteínas PR se dividem em 17 famílias nomeadas PR-1 à PR-17 (Sels et al., 2008). A classificação de proteínas PR se baseia em seu agrupamento dentro das famílias, de acordo com a sequência de aminoácidos que compartilham, relações serológicas e enzimáticas ou atividades biológicas (Van Loon et al., 1994). As osmotinas, por apresentarem sequência e estrutura muito similar às thaumatinas, proteínas doces encontradas nos frutos do arbusto africano *Thaumatococcus daniellii*, foram classificadas dentro da família das PR-5, também conhecida por *Thaumatin-like proteins* (TLPs) (Stintzi et al., 1991). Embora as taumatinas sejam proteínas muito doces, nenhuma outra TLP apresenta a mesma característica (Velazhahan et al., 1999). Sugere-se que a ausência de sabor doce das TLPs está associada ao baixo conteúdo de lisina na sequência de aminoácidos dessas proteínas e/ou à mudança de duas sequências tripeptídicas em relação às taumatinas (Viktorova et al., 2012; Batalia et al., 1996).

As TLPs têm sido encontradas em diversas plantas, em resposta a algum estresse ou estádio de desenvolvimento, situações onde se pode observar seu

acúmulo e presença em outros tecidos antes não contemplados (Velazhahan et al., 1999). Dentre os seus papéis fisiológicos e de desenvolvimento, incluem-se a formação de órgãos florais, amadurecimento de frutos, germinação de sementes e senescência (Cao et al., 2016). Trabalhos com diferentes espécies vegetais têm demonstrado que a expressão de TLPs pode ser induzida por diferentes vias hormonais, tais como ácido salicílico, etileno, ácido jasmônico e ácido abscísico (Misra et al., 2015; Jami et al., 2007). Geralmente, as TLPs são proteínas altamente solúveis que se acumulam em altos níveis em tecidos específicos, em compartimentos subcelulares, sob condições apropriadas ou secretadas para o meio extracelular. Geralmente, TLPs ácidas tendem a ser extracelulares, enquanto que as básicas são encontradas no vacúolo (ou em vesículas) da célula vegetal (Velazhahan et al., 1999). Neste sentido, podem ser classificadas dentre três subclasses de acordo com seu ponto isoelétrico (pI): ácida, neutra ou básica (Min et al., 2004). Análises da sequência de aminoácidos na porção N-terminal de diferentes TLPs de tabaco, demonstraram também uma clara diferença entre as isoformas ácida, neutra e básica (Koiwa et al., 1994).

De acordo com Breiteneder (2004) as TLPs podem ser classificadas em três grupos, (i) aquelas produzidas em resposta à infecção por patógenos, (ii) aquelas produzidas em resposta a estresses osmóticos, também chamadas de osmotinas, e (iii) as com atividade antifúngica, presentes nas sementes de cereais. Diversas TLPs descobertas em plantas foram denominadas de acordo com a sua atividade biológica, dentre estas destacam-se as permatinas, osmotinas, proteínas alergênicas e as com domínio kinase.

As permatinas são proteínas antifúngicas capazes de inibir o desenvolvimento de esporos e fungos através da permeabilização da membrana plasmática. Essas proteínas promovem o extravasamento do citoplasma via formação de poros, sugerindo uma interação física entre as permatinas e a membrana plasmática de fungos suscetíveis à ação dessas proteínas (Min et al., 2004). Exemplos de permatinas ocorrem em alta concentração em sementes de cereais, como a zeamatina de milho (*Zea mays*), hordomatina de cevada (*Hordeum vulgare* L.) e avematina de aveia (*Avena sativa* L.) (Liu et al., 2010).

As osmotinas e as osmotinas-like inicialmente foram definidas como proteínas básicas e neutras encontradas no vacúolo das células vegetais, em resposta a estresses osmóticos (Liu et al., 2010). Porém, diferentemente das classificações propostas, sabe-se atualmente que podemos encontrar osmotinas e/ou osmotinas-like com característica ácida, como a GmOLPa de soja (*Glycine max* L.), e que respondem tanto a estresses abióticos quanto bióticos, facilitando a compartimentalização de íons ou solutos, assim como apresentando atividade antifúngica (Kumar et al., 2015).

Muitas TLPs têm sido reportadas como componentes alergênicos em frutas e pólens de coníferas (Hoffmann-Sommergruber, 2002; Breiteneder, 2004). Dentre as TLPs alergênicas de alimentos incluem-se a proteína Pru av2 de cereja (*Prunus avium* L.), Cap a1 de pimentão (*Capsicum annuum* L.), Act c2 de kiwi (*Actinidia chinensis* Planch.) e Mal d2 de maça (*Malus x domestica*) (Liu et al., 2010). Dentre as TLPs alergênicas encontradas em pólen estão a Jun a3 do cedro da montanha (*Juniperus ashei* J. Buchholz), Cup a3 do cipreste do Arizona (*Cupressus arizonica* Greene) e Cry j3 do cedro japonês (*Cryptomeria japonica* L. f.) (Liu et al., 2010). Suas capacidades alergênicas se devem à presença de motivos na estrutura da proteína capazes de se ligarem a imunoglobulina (Ig)E e provocarem sintomas alérgicos em pessoas sensíveis a esses epítopos (Breiteneder 2004).

Em *Arabdopsis thaliana* foi encontrada uma outra forma de TLP, denominada PR5K (*PR5-like receptor kinase*). Esta proteína apresenta um domínio extracelular com sequência similar as TLPs ácidas, um domínio central transmembrana e uma porção quinase serina/treonina intracelular que está envolvida em respostas de defesa. A semelhança entre o domínio extracelular e as TLPs sugere uma possível interação com alvos microbianos (Wang et al., 1996).

1.2 História evolutiva da família Thaumatin-like

As TLPs são amplamente distribuídas, podendo ser encontradas em diferentes espécies e organismos, incluindo animais invertebrados (nematoides e insetos),

fungos e plantas (Liu et al., 2010). Nos animais invertebrados ocorrem nos nematoides *Caenorhabditis elegans* e *Caenorhabditis briggsae*, e em quatro ordens de insetos: Coleoptera (*Diaprepes* and *Biphyllus*), Hemiptera (*Toxoptera*), Hymenoptera (*Lysiphlebus*), e Orthoptera (*Schistocerea*). Nos fungos estão presentes em algumas espécies de basidiomicotas e ascomicotas. Nas plantas as TLPs são universais, podendo ser encontradas em algumas espécies de briófitas, gimnospermas e angiospermas (Liu et al. 2010). Estudos de análise genômica mostram que nas plantas elas ocorrem como uma família de múltiplos genes (Zhao and Su, 2010).

Apesar da alta similaridade observada entre as TLPs, mesmo uma pequena mudança na sequência de aminoácidos pode levar à diversidade de funções observada dentro desta família de proteínas. Adams et al. (2017) demonstraram que a mudança para três resíduos de triptofano na sequência de aminoácidos de proteínas da subfamília PR-5d em solanaceae conferiu uma característica específica de ligação à celulose a essas proteínas, sugerindo um ganho de função a esse subgrupo dentro da família das *Thaumatins-like*.

A ampla distribuição das TLPs, em conjunto com sua diversidade de funções, o crescente interesse por sua aplicabilidade biotecnológica e a recente disponibilização de genomas de diferentes espécies em banco de dados, tem contribuído para a realização de estudos sobre a origem, evolução e diversificação desta grande família gênica em eucariotos (Cao et al., 2016). Estudos filogenéticos também contribuem para a classificação das TLPs de acordo com a sua origem, estrutura e perfil de expressão.

Na análise filogenética conduzida por Shatters et al. (2006), dois clados distintos, com um ancestral comum para insetos e nematoides, foram formados. O mesmo autor sugere que a relação parafilética das sequências de insetos com as sequências de nematoides, bem como os grupos de insetos, indica uma herança de um único gene dentro dos insetos e dentro de cada ordem, seguida por uma duplicação gênica dentro desses grupos. Neste mesmo trabalho foi realizada uma comparação filogenética mais completa utilizando as sequências de TLPs encontradas nos animais invertebrados e nas espécies vegetais *Arabidopsis*

thaliana, *Oryza sativa* e *Pinus taeda*. Além dessas foram adicionadas à análise duas sequências descritas como taumatinas de *Thaumatococcus danielli*, uma descrita como osmotina de *Nicotiana tabacco*, e uma sequência *Triticum aestivum* que representa uma variante estrutural das *thaumatins-like*. Nesta análise, Shatters et al. (2006) demonstraram que as sequências dos animais formam um clado único, separado do das plantas. Nas plantas por sua vez foram formados 10 clados distintos, cada um contendo no mínimo uma sequência de *A. thaliana* e uma de *O. sativa*. Neste mesmo sentido Zhao e Su (2010), sugerem que a família gênica das TLPs em plantas se originou a partir de 10 genes ancestrais, antes da divergência entre mono e dicotiledôneas há cerca de 130-240 milhões de anos atrás. Além disso, as análises realizadas por Shatters et al. (2006) não deram suporte para a separação de nomenclatura entre osmotinas e *thaumatins-like*. Essa inconsistência na nomenclatura dos diferentes tipos de TLPs tem gerado confusão na classificação dessas proteínas em um ou mais grupos.

Para melhor decifrar a evolução da família das TLPs, Liu et al. (2010) utilizaram 118 sequências de TLPs para a análise filogenética, selecionadas a partir dos fungos ascomicotas e basidiomicotas, musgos, gimnospermas, angiospermas, nematoides e insetos. A árvore gerada indica que a família das TLPs é altamente divergente, com a possibilidade de nove grupos distintos. O primeiro grupo consiste apenas de sequências de ascomicotas. Devido a sua baixa similaridade e origem diferente, Liu et al. (2010) sugeriram que estas sequências não devam ser consideradas TLPs. Basidiomicotas, nematoides e insetos formaram três clados distintos, enquanto as plantas se dividiram em cinco grupos com sequências de gimnospermas e angiospermas, sugerindo um mínimo de cinco genes bastante diversos em sequência e função, presentes no ancestral comum das plantas com sementes. Os mesmos autores sugerem um ancestral comum a plantas, fungos e insetos, datando de 1 bilhão de anos atrás.

Petre et al. (2011) por sua vez conduziram uma análise com 598 sequências de TLPs a partir de 100 espécies, incluindo fungos (basidiomicotas e ascomicotas), invertebrados (nematoides e insetos) e plantas. A filogenia desses eucariotos revelou três principais grupos monofiléticos. O grupo I consistiu de

subclados específicos para fungos (basisiomicota e ascomicota) e plantas. O grupo II incluiu subclados específicos para os invertebrados (nematoides e insetos, em subclados distintos) e para as plantas. O grupo III por sua vez agrupou apenas sequências de plantas vasculares. Assim como observado por Shatters et al. (2006), as análises de Petre et al. (2011) não suportam a separação de nomenclatura entre osmotinas e *thaumatins-like*, o que gera confusão semântica na literatura.

De acordo com as análises filogenéticas conduzidas por Cao et al. (2015), as TLPs de Arabidopsis, arroz, álamo, milho, *Physcomitrella* e *Chlamydomona*, se dividem em seis grupos. Os mesmos autores sugerem que o mais recente ancestral comum das Viridiplantae apresenta cinco genes de TLP. *Chlamydomona*, por apresentar apenas uma sequência teria perdido quatro desses genes. A expansão da família das TLPs teria ocorrido depois da divergência das embriófitas conduzindo para as plantas vasculares.

Apesar de algumas diferenças entre as análises e as conclusões dos diferentes autores, em geral, os mesmos sugerem uma importante participação dos eventos de duplicação em tandem e da co-evolução entre plantas e patógenos para a expansão da família das TLPs. Além disso, concluem que o processo de evolução dessa família é altamente conservativo por manter certa similaridade nas sequências e na estrutura das proteínas nos diferentes organismos.

1.3 Estrutura e atividade antifúngica

Análises comparativas da estrutura primária da osmotina, TLP básica encontrada em tabaco, e muitas outras TLPs, revelam diversas características interessantes (Min et al., 2004). A maioria das TLPs descritas, incluindo as osmotinas, apresentam peso molecular entre 24-26 kDa, uma alanina localizada no sítio de clivagem e 16 resíduos de cisteína, responsáveis por formar oito pontes dissulfeto que estabilizam a proteína contra alterações de pH, proteases e desnaturação por altas temperaturas (Cao et al., 2016). Pequenas TLPs com baixo peso molecular (17 kDa) e apenas 10 resíduos de cisteína, que formam

cinco pontes dissulfeto, também podem ser encontradas em monocotiledôneas e coníferas (Petre et al., 2011).

Estudos de cristalografia mostram uma estrutura bastante conservada entre as proteínas da família das TLPs, podendo-se observar claramente a formação de três domínios e uma fenda localizada entre os domínios I e II (Figura 1) (Leone et al., 2006). Esta fenda pode apresentar natureza ácida/eletronegativa, neutra ou básica/eletropositiva para ligação de diferentes ligantes/receptores. As taumatinas, além de apresentarem um laço a mais no domínio II, possuem uma fenda básica, onde resíduos de lisina parecem exercer um importante papel na característica adocicada destas proteínas, através da interação com possíveis receptores presentes nas papilas gustativas de mamíferos (Min et al., 2004; Slootstra et al., 1995). Realmente, foi demonstrado que concentrações baixas de taumatinas são capazes de estimular as células gustativas de macaco-rhesus, indicando uma forte ligação dessas proteínas aos receptores de membrana destas células (Velazhahan et al., 1999).

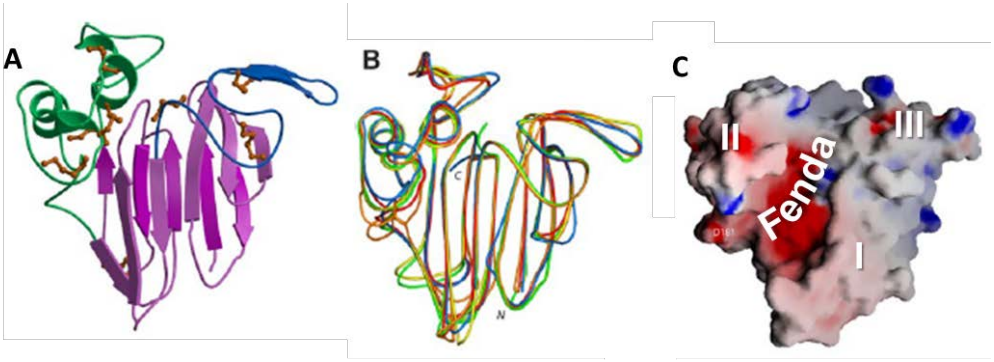


Figura 1 – **A**. Representação da estrutura das TLPs. Domínio I, II e III, em *roxo*, *verde* e *azul*, respectivamente. Pontes dissulfeto em *laranja*. **B**. Sobreposição de diferentes TLPs. TLP de banana (*vermelha*), thaumatina (*azul*), osmotina (*amarelo*), PR-5d (*verde*) e zeamatina (*laranja*). **C**. Representação geral da fenda entre os domínios. (Adaptada a partir de Leone et al., 2006)

Na maioria das TLPs, diferentemente das taumatinas, observa-se uma fenda de natureza ácida entre os domínios I e II. Quando ácida, a fenda geralmente é composta por um motivo REDDD (arginina, glutamato e três aspartatos) de resíduos de aminoácido que exercem um importante papel antifúngico (Figura 2) (Ramos et al., 2015). Esta estrutura tem sido citada como um centro catalítico que

promove a hidrólise de moléculas poliméricas de glucanos, através de sua ligação à β-1,3-glucanos, componentes da parede celular de fungos (Trudel et al., 1998). Entretanto, a presença de uma fenda ácida nas TLPs não está diretamente relacionada com a sua capacidade antifúngica. Menu-Bouaouiche et al. (2003) demonstraram que TLPs de maça e cereja, com uma fenda altamente ácida, não apresentavam atividade antifúngica. Além disso, há trabalhos que relatam a existência de TLPs antifúngicas desprovidas de atividade glucanase, o que leva a inferir que outras estruturas podem estar relacionadas na determinação da atividade antifúngica (Min et al., 2004; Leone et al., 2006).

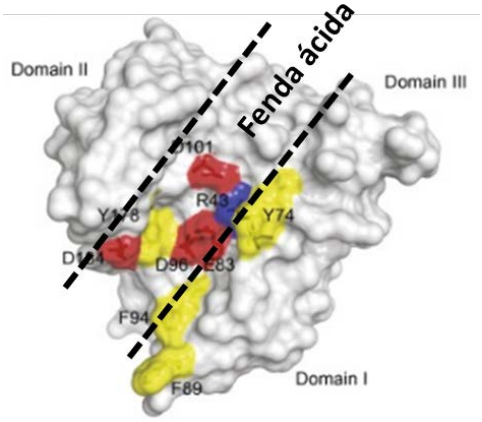


Figura 2 – Representação da distribuição do motivo REDDD na fenda das TLPs. Resíduos básicos, ácidos e aromáticos em *azul*, *vermelho* e *amarelo*, respectivamente, na volta e dentro da fenda. (Adaptada a partir de Ramos et al., 2015).

Ibeas et al. (2000) demonstraram a relação casual entre fosfomanoproteínas, a ligação de osmotinas à parede celular de leveduras e sua citotoxicidade. Fosfomanoproteínas são componentes da parede celular de *Saccharomyces cerevisiae* juntamente com os glucanos, responsáveis por determinar as propriedades da superfície celular, como hidrofobicidade e carga elétrica da célula (Ibeas et al., 2000). A deficiência de fosfomanoproteínas na parede celular de *S. cerevisiae* reduz a carga negativa da parede em até 90% e aumenta a resistência à osmotina, indicando a importância desses resíduos para a suscetibilidade das leveduras à proteína (Ibeas et al., 2000). Baseando-se na ligação entre TLPs e componentes da parede celular de fungos e leveduras, Ibeas et al. (2000)

sugerem que a provável função destes componentes seja capturar as TLPs do ambiente, aumentando a taxa de difusão da proteína ao longo da parede celular e sua atividade citotóxica. Neste sentido, Ramos et al. (2015) sugerem que a interação inicial da osmotina-like de *Calotropis procera* (CpOsm) e a célula de *Fusarium solani* pode ocorrer entre as cargas positivas da superfície da proteína e a superfície negativa da célula do fungo. Esta interação iônica não específica direcionaria a CpOsm para a superfície celular do fungo, promovendo o aumento da sua concentração na volta do esporo e sua difusão através da membrana celular, afetando sua permeabilidade e causando a perda do conteúdo citoplasmático. De acordo com Salzman et al. (2004) a ligação entre a osmotina e as fosfomanas é favorecida pela presença de Ca^{+2}, que se liga na fenda negativa da proteína e nas moléculas negativas de fosfomanas da superfície da leveduras, formando uma corrente e aproximando a osmotina da membrana plasmática da levedura. Este efeito é bloqueado quando EGTA (agente quelante de Ca^{+2}) sequestra Ca^{+2}, ou quando K^{+} compete com a osmotina pelos sítios de ligação aos manosefosfatos na superfície celular das leveduras (Figura 3).

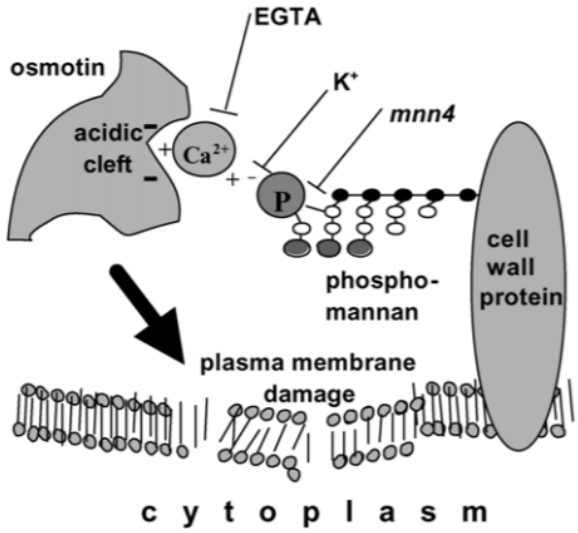


Figura 3 – Representação do mecanismo da atividade antifúngica proposta por Salzman et al. (2004). A figura mostra o Ca^{+2} intermediando a ligação entre a osmotina e as moléculas de fosfomanas presentes na parede celular de leveduras. Esta ligação permite a aproximação da osmotina à membrana plasmática destes organismos. Este efeito é bloqueado quando EGTA sequestra Ca^{+2}, ou quando K^{+} compete com a osmotina pelos sítios de ligação aos manosefosfatos na superfície celular das leveduras. Também pode ser bloqueado

em mutantes mm4, que não apresentam o resíduo final de mannosephosphate. O exato mecanismo de ação das osmotinas em relação ao dano causado na membrana plasmática ainda não é compreendido. (Retirada de Salzman et al., 2004).

Trabalhos também sugerem que a atividade antifúngica de algumas TLPs pode estar associada ao domínio I dessas proteínas (Chen et al., 1999; Mani et al., 2011). Outros sugerem que seu mecanismo de ação está associado à ligação a componentes de membrana específicos e ativação da transdução de sinais nas células fúngicas. Yun et al. (1998), por exemplo, demonstraram que a osmotina pode ativar o sistema de sinalização por MAPK (*mitogen-activated protein kinase*) em leveduras, induzindo mudanças na parede celular que aumentam a citotoxicidade dessas proteínas antifúngicas. Foi demonstrado também que a osmotina pode induzir morte celular programada em leveduras de *S. cerevisiae,* através do acúmulo de espécies reativas de oxigênio, por meio da supressão da sinalização de genes responsivos a estresse (RAS2/cAMP) (Narasimhan et al., 2001). Este processo ocorre a partir da ligação da osmotina a receptores polipeptídicos com sete domínios transmembrana (*seven trans-membrane domain receptor-like polypeptides* - PHO36) que apresentam características semelhantes aos receptores acoplados a proteínas G (*G protein-coupled receptors* - GPCRs) e regulam o metabolismo de fosfatos e lipídeos. A presença destes receptores torna as leveduras de *S. cerevisiae* suscetíveis à ação dessas proteínas (Narasimhan et al., 2005). Interessantemente, estes são os mesmos receptores mencionados anteriormente para a taumatina nas papilas gustativas, sugerindo que a ligação entre GPCR e TLP possa ser uma característica conservada desta família de proteínas (Liu et al., 2010). A interação com receptores e moléculas específicas pode explicar a especificidade da ação antifúngica de algumas TLPs (Ibeas et al., 2000; Min et al., 2004; Vitali et al., 2006).

Apesar dos diferentes mecanismos que têm sido propostos para explicar o efeito antifúngico dessas proteínas, nenhum estudo o elucidou completamente. O entendimento dos mecanismos de ação antifúngica das osmotinas pode ajudar a elucidar os mecanismos de ação das mesmas durante o estresse hídrico.

1.4 Papel das osmotinas durante a seca

As osmotinas foram inicialmente descobertas durante respostas a estresses osmóticos e descritas como proteínas osmoprotetoras, por manterem a osmolaridade celular através da compartimentalização de solutos ou por alterações metabólicas e estruturais (Abdin et al. 2011; Chowdhury et al. 2017). Inúmeros estudos vêm sendo realizados para determinar o papel fisiológico das osmotinas nos estresses abióticos (Goel et al., 2010).

Estudos mostram que a expressão de osmotinas pode ser induzida em cultura de células e raízes de tabaco em resposta ao tratamento com ácido abscísico (ABA) e polietileno glicol (PEG), que simula uma resposta ao estresse hídrico (Hakim et al., 2018). Além disso, trabalhos têm demonstrado a indução de *thaumatins-like* através da superexpressão de fatores de transcrição associados a respostas de defesa e a possível relação da expressão dessas proteínas com a tolerância à seca (Moon et al., 2014; Muoki et al., 2012). Gutha e Reddy (2008) demonstraram que a superexpressão de um importante fator de transcrição da família DREB (*C-repeat binding factor/dehydration responsive element binding factor*), associado com a tolerância e regulação de genes responsivos à seca, foi capaz de induzir a expressão da osmotina em tabaco.

A expressão de osmotinas em plantas transgênicas, como tomate, tabaco, algodão, soja, sésamo, cenoura, amora e arroz, tem levado à tolerância das mesmas à seca (Chowdhury et al. 2017; Le et al., 2018; Anoon et al., 2014; Weber et al., 2014; Das et al., 2011; Goel et al., 2010; Parkhi et al., 2009; Barthakur et al., 2001). Alguns fatores como o acúmulo de prolina (atua como osmorregulador e antioxidante), do conteúdo relativo de água, clorofila, expansão foliar, diminuição da abertura estomática, da peroxidação lipídica e malondialdeído (marcador de estresse oxidativo e degradação de membrana), aumento da atividade de enzimas antioxidantes, metabólitos secundários (fenólicos e flavonóides) e estabilidade da membrana plasmática, têm sido observadas em associação com a resposta da osmotina em relação à tolerância ao estresse hídrico (Hakim et al., 2018). Estes resultados indicam que a osmotina promove a proteção da maquinaria fotossintética, mantém a osmolaridade celular,

reduz a perda de água, o estresse oxidativo, os danos de membrana e estimula a expressão de genes envolvidos na regulação positiva da via de biossíntese de metabólitos (Chowdhury et al. 2017; Zhang et al., 2004; Kumar et al., 2015). Apesar da osmotina atuar à jusante da expressão de alguns genes envolvidos na resposta a estresse, sabe-se que essas proteínas não atuam como fatores de transcrição, pois não apresentam motivos de ligação ao DNA (Abdin et al., 2011).

Um dos possíveis mecanismos de ação dessas proteínas se dá através das cargas negativas que ocorrem na fenda das *thaumatins-like* que possibilitam a ligação com proteínas de membrana com carga positiva, como canais de íons e água, que eventualmente possa resultar em um aumento do fluxo de água através da membrana (Batalia et al., 1996). Apesar disso, assim como para a atividade antifúngica, o exato mecanismo dessas proteínas quanto ao aumento da tolerância a estresses osmóticos como a seca, ainda não foi elucidado.

1.5 Soja

A soja [*Glycine max* (L.) Merrill] é a espécie de maior importância econômica a nível mundial. A soja é um grão muito versátil que dá origem a produtos e subprodutos muito usados pela agroindústria, indústria química e de alimentos. A relevância desta espécie na agricultura é decorrente de sua capacidade de fixação de nitrogênio atmosférico, por meio da simbiose com microrganismos, de sua utilização para a alimentação humana e animal, além de servir de matéria prima para a produção de biocombustíveis (Reetz et al. 2012).

No Brasil, o plantio da soja, em larga escala, teve início em 1960 e, atualmente, é o principal responsável pelo crescente volume de exportações do agronegócio e o consequente avanço da economia nacional. O Brasil é hoje o segundo maior produtor desta oleaginosa e já está competindo com os Estados Unidos pelo título de maior exportador em nível mundial (Conab 2018).

A seca é um dos principais fatores ambientais que mais contribuem para a perda na produção de soja (Câmara e Heiffig, 2000). Estudos em casa de vegetação e no campo mostraram que a seca leva a uma redução significativa na produção do grão (24~50%) em diferentes locais e períodos (Ku et al., 2013).

Longos e curtos períodos de deficiência hídrica acarretam na diminuição da produção, ocasionando graves problemas econômicos e sociais. No Brasil, a seca prolongada durante o período de cultivo da soja vem tornando-se cada vez mais frequente (Brando et al., 2010). Esta situação pode tornar-se ainda mais dramática, de acordo com as previsões de mudanças climáticas, que apontam para um aumento na frequência, severidade e duração dos períodos de seca (Chen et al., 2016).

Com o intuito de mitigar os impactos gerados pela deficiência hídrica, ferramentas de biologia molecular, para a identificação e transferência de genes responsáveis pela tolerância à seca, vem sendo utilizadas no desenvolvimento de cultivares mais tolerantes (Shin et al., 2015; Guimarães-Dias et al., 2012; Manavalan et al., 2009; Pathan et al., 2007). Além disso, a recente aplicação de tecnologias contemporâneas de edição do genoma, como CRISPR, vem se mostrando bastante eficaz na melhoria de características agronômicas. Um dos pré-requisitos dessa abordagem é a identificação dos principais atores e vias genéticas subjacentes à resposta da planta ao estresse hídrico (Hua et al., 2018). Uma estratégia que tem se mostrado promissora para desenvolvimento de cultivares mais tolerantes a estresses abióticos está baseada na superexpressão de proteínas PR, como as osmotinas (Ahmed et al., 2013).

Até o momento foram identificadas quatro osmotinas em soja, denominadas P21, GmOLPa, GmOLPb e P21e (P21-like) (Tachi et al., 2009; Onishi et al., 2006; Graham et al., 1992). A P21 foi a primeira osmotina a ser identificada em folhas de soja, cujas plantas foram crescidas em casa de vegetação sem tratamentos de estresse, apenas sob condições naturais. A proteína foi isolada de folhas de plantas com 60 dias da cultivar Williams 82. Sua forma madura apresentou 202 aminoácidos e ponto isoelétrico de 4.6 (Graham et al., 1992).

Quatorze anos depois, uma segunda osmotina também com caráter ácido foi identificada em soja e denominada GmOLPa (Onishi et al., 2006). Esta com 201 aminoácidos e ponto isoelétrico de 4.4, foi predita como uma proteína extracelular por não apresentar a porção C-terminal, responsável por direcionar as proteínas para o vacúolo. Isolada de plantas da cultivar Enrei em estádio V3, submetidas a

estresse por sal, foi identificada em raízes de plantas não tratadas e tratadas, e em folhas e caules após 48h do tratamento com NaCl. Onishi et al (2006) relataram que a indução da transcrição de GmOLPa no caule e nas folhas ocorre devido ao acúmulo de Na^+/Cl^- nesses órgãos, após a saturação da capacidade de armazenamento desses íons na raiz. Os mesmos autores demonstraram que o gene pode ser também induzido após tratamento com ABA nas raízes e por seca nos três órgãos avaliados (raiz, caule e folha). Análises de cis-elementos à jusante do gene GmOLPa revelaram motivos de resposta a ABA e à seca, como ABRE, MYB/MYC e LTRE. Apesar disso, Onishi et al. (2006) sugerem que a indução deste gene seja primeiramente via uma resposta transcricional independente de ABA, pois neste tratamento quando comparado com os tratamentos por sal e seca, o gene foi levemente induzido apenas nas raízes.

Em 2009, Tachi et al., identificaram a terceira e quarta osmotina de soja (GmOLPb e P21-like) na cultivar Enrei submetida a tratamento por sal, metil jasmonato (MeJA) e ácido salicílico (AS). Diferentemente das outras osmotinas identificadas em soja, a GmOLPb corresponde a uma proteína neutra (pI 6.0), que apresenta uma porção C-terminal, possivelmente relacionada ao direcionamento para o vacúolo. Análises da estrutura 3D das três osmotinas de soja, P21, GmOLPa e GmOLPb, mostraram a presença dos três domínios e da fenda ácida, característica das proteínas pertencentes à família das *thaumatins-like*. Além disso, Tachi et al. (2009) observaram diferenças no potencial eletrostático na superfície das três proteínas, no sentido que a GmOLPb apresenta mais pontos de carga positiva na sua superfície do que a P21 e a GmOLPa, a mais carregada negativamente. Os autores sugeriram que as cargas negativas de GmOLPa podem facilitar a sua atividade antifúngica, através da interação de cátions inorgânicos presentes em toda a superfície da proteína. Neste mesmo estudo, foi observada a indução da expressão, principalmente, de GmOLPb e P21-like nas folhas de plantas tratadas com MeJA, e de GmOLPa nas folhas baixas de plantas tratadas com SA. Os três genes foram induzidos nas folhas de plantas tratadas com sal, porém em momentos diferentes. P21-like respondeu nos estágios mais iniciais e subsequentemente o GmOLPb. GmOLPa, respondeu nos estágios mais

tardios do estresse. Interessantemente, todos os três apresentavam expressão basal nas raízes de plantas não tratadas (Tachi et al., 2009).

A maioria dos estudos sobre a sinalização de respostas ao estresse hídrico se concentrou no estresse salino, principalmente porque as respostas das plantas ao sal e à seca estão intimamente relacionadas e os mecanismos se sobrepõem (Zhu, 2002). Como demonstrado pelos trabalhos citados anteriormente os três genes de osmotinas de soja respondem ao estresse por sal, porém seu papel na resposta ao estresse hídrico e sua relevância na tolerância de plantas à seca continua desconhecido.

2. OBJETIVOS

2.1 Objetivo geral

Investigar a origem e diversificação das osmotinas em plantas e entender seu papel na resposta à seca em soja [*Glycine max* (L.) Merrill].

2.2 Objetivos específicos

a) Investigar a ocorrência de genes que codificam osmotinas em plantas;
b) Elucidar o relacionamento filogenético e a origem evolutiva das osmotinas dentro da família das *Thaumatins-like*
c) Caracterizar as sequências gênicas e proteicas dos potenciais homólogos das osmotinas;
d) Investigar o papel das osmotinas de soja na resposta à seca.

Capítulo II
ARTIGO CIENTÍFICO 1

Genome-wide analysis and evolution of plant thaumatin-like proteins: a focus on the origin and diversification of osmotins

---

Aceito para publicação na Molecular Genetics and Genomics

Genome-wide analysis and evolution of plant thaumatin-like proteins: a focus on the origin and diversification of osmotins

Giulia Ramos Faillace[1], Andreia Carina Turchetto-Zolet[1], Frank Lino Guzman Escudero[2], Luisa Abruzzi de Oliveira-Busatto[1], Maria Helena Bodanese-Zanettini[1]*

[1] *Programa de Pós-Graduação em Genética e Biologia Molecular and Instituto Nacional de Ciência e Tecnologia: Biotec Seca-Pragas, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), 91501-970, Porto Alegre, RS, Brazil*

[2] *Programa de Pós-Graduação em Biologia Celular e Molecular, Centro de Biotecnologia (CBiot), Universidade Federal do Rio Grande do Sul (UFRGS), 91501-970, Porto Alegre, RS, Brazil*

*Author for correspondence:
Maria Helena Bodanese-Zanettini
Tel: +55 51 33086725
Email: maria.zanettini@ufrgs.br

Abstract

Osmotin is an important multifunctional protein related to plant stress responses and is classified into the thaumatin-like protein (TLP) family. Using genome-wide and phylogenetic approaches, we investigated osmotin origin and diversification across plant TLP evolution. Genomic and protein *in silico* analysis tools were also accessed and considered for the study conclusions. Phylogenetic analysis including a total of 722 sequences from 32 Viridiplantae species allowed the identification of an osmotin group that includes all previously characterized osmotins. Based on the phylogenetic tree results, it is evident that the osmotin group emerged from spermatophytes. Phylogenetic separation and gene expansion could be accounted for by an exclusive motif composition and organization that emerged and was maintained following tandem and block duplications as well as natural selection. The TLP family conserved residues and structures that were also identified in the sequences of the osmotin group, thus suggesting their maintenance for

defense responses. The gene expression of Arabidopsis and rice putative osmotins reinforces its roles during stress response.

Keywords: Evolution, Osmotin, Phylogenetic Analysis, PR-5, Thaumatin-like

Introduction

Osmotin is a protein first discovered in tobacco cells adapted to a low osmotic potential environment. Notably, osmotin is a multifunctional protein that acts as an osmoregulator and also provides plants with protection against pathogens (Abdin et al., 2011). The osmotin/osmotin-like proteins (OLPs) are known to facilitate the compartmentation of ions or solutes and exhibit antifungal activities (Kumar et al., 2015). Osmotin is classified into the pathogenesis-related protein family 5 (PR-5), a family of proteins with high sequence similarity to thaumatin, a sweet-tasting protein from the West African shrub *Thaumatococcus danielli* (Cao et al., 2016). Most typical proteins from the PR-5 family—also called thaumatin-like proteins (TLPs)—have molecular masses ranging from 21-26 kDa, and generally possess 16-cysteines residues. These residues can form eight disulfide linkages that are related to their structural stability across various pH conditions, proteases, and heat induced denaturation (Cao et al., 2016). Seven TLP structures have been solved to date, revealing a strongly conserved 3D organization with three domains and a characteristic cleft between domains I and II. This cleft may have an acidic, neutral, or basic nature for binding to different ligands/receptors. The acid cleft is known to confer antifungal activity due to the REDDD motif (arginine, glutamic acid, and three aspartic acid residues), a highly conserved amino acids that are dispersed in the primary sequence (Petre et al., 2011).

In spite of the high sequence similarity of TLPs, even a small change in the amino acids of these proteins leads to diverse functions (Kumar et al., 2015). TLPs are involved in plant defense systems against various biotic and abiotic stresses, such as pathogen attack, wounding, drought, salinity, and freezing. In addition, they have also been implicated in physiological and developmental roles, including floral organ formation, fruit ripening, seed germination, and senescence (Cao et al., 2016). Permatins are TLPs believed capable of creating transmembrane pores. Examples of permatins that occur in cereal seeds include the permatin from oat (*Avena sativa*) and the zeamatin from maize (*Zea mays* L.). Other

TLPs exhibit a binding ability to the Ig-E of allergic individuals through the allergenic motifs present in their protein structures. PruAV2 from cherry (*Prunus avium* L.) and MalD2 from apple (*Malus domestica*), are reported as allergenic TLPs (Liu et al., 2010). TLPs with kinase activity also exist, such as PR5K from *Arabidopsis thaliana*. This protein possesses both an extracellular TLP domain and an intracellular kinase domain related to a family of protein-serine/threonine kinases involved in the expression of self-incompatibility and disease resistance (Wang et al., 1996).

TLPs are widely distributed in plants, including green algae, bryophytes, gymnosperms, and angiosperms. Their antibiotic activities and physiological functions have aroused interest for crop improvement due to their biotechnological applicability. This emerging interest has contributed to the development of phylogenetic studies aimed at understanding the origin and distribution of the TLP superfamily (Shatters et al., 2006; Liu et al., 2010; Petre et al., 2011; Cao et al., 2016). In a phylogenetic study of plants, insects, and nematode proteins, Shatters et al. (2006) suggested that plant TLPs (from *A. thaliana*, *Oryza sativa*, and six proteins from *N. tabacum*, *Pinus taeda*, *Triticum aestivum*, *T. danielli*, and *Z. mays*) are divided into 10 clades exclusive to plant sequences. On the other hand, Liu et al. (2010) demonstrated that plant TLPs (*A. thaliana*, *O. sativa*, *Pinus monticola*, *Picea glauca*, and *Physcomitrella patens*) are divided into five exclusive plant groups across a total of ten groups formed in the phylogenetic tree of plants, fungi, and animals (insects and nematodes). Moreover, Petre et al. (2011) performed a phylogenetic study on 598 thaumatin domains of TLP sequences from 100 eukaryote species, including 410 sequences from 18 plant species and 188 sequences from fungi and invertebrates. The neighbor-joining tree constructed from these eukaryote sequences revealed three major monophyletic groups—one exclusive to plants and another two including animals, fungi, and plant-specific subclades. In these TLPs studies, different names/annotation using for plant TLPs (mainly osmotin and thaumatin-like) causes confusion regarding the classification of these proteins in a single group or in subgroups. To date, no phylogenetic studies have attempted to identify osmotin origin and diversification throughout Viridiplantae TLP evolution to clarify this semantic confusion. We hypothesized that osmotins emerged from a duplication event in ancestral land plants, and that they have subsequently diversified into a separate group along Viridiplantae TLP evolution. To test this hypothesis, the present study explored available data from public databases. A total of

722 sequences presented the thaumatin domain (THN, the signature of TLPs), including sequences from previously characterized thaumatin-like and osmotin proteins. These were selected from 32 Viridiplantae species for the TLP phylogenetic tree reconstruction using a Bayesian method. Analyses of gene and protein structure, gene duplication, and *in silico* expression patterns were also accessed and considered in the present study.

Materials and methods

Database search and sequence retrieval

The previously identified osmotin-like protein from *Solanum nigrum* (AAL87640) (Castillo et al., 2005) and the thaumatin code domain (Pfam-PF00314) present in all thaumatin-like proteins were used as queries for BLASTp search using the public database default. All 52 Viridiplantae species available in the public databases (Phytozome v.12.0, http://www.phytozome.net/; Congenie, http://congenie.org/) were screened for TLPs, and 26 species representing the different taxonomic groups were selected for the analysis. The queries sequence from *S. nigrum* and five previously characterized thaumatin-like protein sequences (*N. tabacum* CAA43854, *Petunia x hybrida* AAK55411, *A. sativa* AAB02259, *P. avium* P50694, and *M. domestica* Q9FSG7) were obtained from the NCBI database (http://www.ncbi.nlm.nih.gov) and included in the phylogenetic analysis, resulting in a total of 32 species. All sequences selected from BLASTp share the thaumatin domain, suggesting that they are included in the thaumatin-like family. The protein isoforms were excluded to refine the analysis. Taxa terminologies were abbreviated using the first letter of the genus and the first two letters of the species name (e.g. Ath corresponds to *Arabidopsis thaliana*). Information regarding the selected protein sequences is presented in Table S1.

Multiple-sequence alignments and phylogenetic analyses

The retrieved full-length coding sequences (cds) were translated to amino acid sequences and aligned using the muscle algorithm from MEGA7 software (Molecular Evolutionary Genetics Analysis) (Kumar et al., 2016). The thaumatin domain sequence of *A. thaliana*, *N. tabacum*, and *S. nigrum* were identified by the Simple Modular Architecture Research Tool (Letunic and Bork, 2017) and used as a reference to determine the thaumatin domain region in the other aligned sequences. Only the thaumatin domain sequence was selected

for phylogenetic analysis (File S1). The sequences were inspected manually and ProtTest 3.4 (Abascal et al., 2005) was used to select the amino acid substitution model for Bayesian analysis. The final data set included a total of 722 sequences from 32 species. The phylogeny was reconstructed using the Bayesian method in BEAST 1.8 (Drummond et al., 2012). The WAG+G model was the optimal model for protein sequences dataset according ProtTest (Abascal et al., 2005). The birth-death process was selected as a tree prior to Bayesian analysis using 100,000,000 generations performed with Markov chain Monte Carlo (MCMC) algorithms. Tracer 1.6 (Rambaut et al., 2014) was used to effectivity verify the obtained data by the convergence of Markov chains and adequate effective sample sizes (>200) following the first 10% of generations being deleted as burn-in. The TreeAnnotator (BEAST 1.8 package) was used to access the maximum clade credibility of the consensus tree. Trees were visualized and edited using FigTree v1.4.3 software (http:// tree.bio.ed.ac.uk/software/figtree/). Statistical support for the clades was determined by accessing the Bayesian posterior probability. Further sequence editions were performed by visual analysis of sequence organization in the reconstructed tree.

Structure analysis of putative osmotins

The gene and protein structure of sequences clustered in the osmotin group were analyzed. The intron/exon structures and organization were determined using the Gene Structure Display Server (GSDS) program (Hu et al., 2015). The subcellular location was predicted using the TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al., 2000). Protein structures were determined using the Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de/) and conserved residues were accessed by a web-based sequence logo-generating application (WebLogo—Weblogo.berkeley.edu). To identify shared motifs and structural divergences among thaumatin-like proteins, the MEME online tool (http://meme.nbcr.net/meme/intro.html) was used. Full-length protein sequences were subjected to the MEME tool using the following parameters: number of repetitions: any; maximum number of motifs: 20; minimum motif width: 6; and maximum motif width: 80. Motif localization was identified in the inferred 3D Arabidopsis protein structures based on protein homology modeling in the Swiss-model database (Dong et al., 2018). The theoretical isoelectric point of proteins

and their acid cleft, as well as their molecular weight, were obtained using the online computational tool IPC (Isoeletric Point Calculation) (Kozlowski, 2016).

Duplication pattern of putative osmotins

The duplication pattern of sequences clustered in the osmotin group was accessed in Plaza 3.0 database (http://bioinformatics.psb.ugent.be/plaza/). Synteny detection and duplication patterns were accessed using MCScanX software (http://chibba.pgml.uga.edu/mcscan2/) for *A. thaliana* and *O. sativa* possessing the genome at the pseudomolecule level (1 pseudomolecule = 1 chromosome). All gene sequences from each species were compared against themselves (intra species analysis) using all-vs-all BLASTp with parameters V=5, B=5, E-value<1e-10 with the output format set as tabular (-m 8). The resulting BLAST hits were incorporated along with the chromosome coordinates for all gene sequences as an input for MCScanX analysis. The resulting hits (68,367 and 47,743 for *O. sativa* and *A. thaliana*, respectively) were classified into five duplications patterns: singletons, dispersed duplicates, proximal duplicates, tandem duplicates, and segmental/WGD duplicates, depending on their copy number and genomic distribution. The analyses were conducted as previously described by Wang et al. (2012).

In order to complement the *A. thaliana* gene duplication analysis, TAIR10 transposable elements annotation (http://www.arabidopsis.org/) was accessed to identify possible transposable elements surrounding the osmotin gene.

Gene ontology annotation and gene expression data mining

The QuickGO Annotation list (https://www.ebi.ac.uk/QuickGO/annotations) and the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/analyses_search_blast.shtml) were accessed for Arabidopsis and rice putative osmotin genes ontology (GO) annotations including molecular function, biological process, and the cellular component.

In order to gain insights regarding osmotin gene expression under stress, expression data (RNA-seq) from *O. sativa* and *A. thaliana* putative osmotins were searched in the Rice eFP Browser (http://bar.utoronto.ca/efprice/cgi-bin/efpWeb.cgi?dataSource=ricestress_rma) in the Rice Expression Profile Database (RiceXPro) (http://ricexpro.dna.affrc.go.jp/data-set.html), and in ePlant database (http://bar.utoronto.ca/eplant/), respectively.

Results

Annotation and phylogenetic analyses

Public databases were screened for thaumatin-like sequences, and a total of 1518 sequences from 46 species were retrieved. Seven algae genomes available in Phytozome were analyzed; however, only *Chlamydomonas reinhardtii* presented one TLP sequence. For the executability of the next analyses, 26 species representing the different taxonomic groups were selected and added to the six previously characterized thaumatin-like protein sequences, resulting in a total of 850 sequences. After protein alignment, some sequences showed missing data (multiple gaps). These sequences were excluded, resulting in a total of 722 sequences being used for the analysis (Table S1). The group of surveyed species included representatives of algae (one species), mosses (two species), pinales (two species), monocots (six species), and eudicots (20 species). Only one putative thaumatin-like sequence was retrieved for the chlorophyte algae *C. reinhardtii*. In the bryophyta *P. patens*, six putative thaumatin-like sequences were identified, while 18 putative thaumatin-like sequences were found for *Sphagnum fallax*. One member of the vascular plants and a unique species that represents the lycopodiophyta group—*Selaginella moellendorffii*—presents 14 putative thaumatin-like sequences. A total of 13 and 39 putative thaumatin-like sequences were identified for *Piceae abies* and *Pinus taeda* (pinales), respectively. For monocots, a minimum of 12 and a maximum of 34 putative thaumatin-like sequences were found per species. The number of sequences in eudicots ranged from 15 (*Amborella trichopoda*, most basal lineage in the clade of angiosperms) to 57 (*Glycine max*) in putative thaumatin-like sequences (Figure 1).

The phylogenetic tree reconstructed with the selected 722 putative thaumatin-like sequences allowed the identification of a group supported by posterior probabilities that clustered all previously characterized osmotins, which was named the osmotin group (Figure 2 and Figure S1).

The osmotin group clustered all previously characterized and other putative osmotins from pinales, monocots, and eudicots species selected for this study (Figure 3). This indicates that the duplication that had originated the osmotin group likely occurred in the Spermatophyta ancestor. The number of sequences per species in this group ranged from

one (*A. thaliana* and *Spirodela polyrhiza*) to twenty-six (*P. taeda*) (Figures 1 and 3). Two subgroups were formed in the osmotin group: the first includes the two pinales representatives, while the second includes the sequences of all analyzed angiosperms (Figure 3). The majority of *P. abies* and *P. taeda* sequences grouped in the osmotin group (Figure 1). In the monocot subgroups, the previously characterized permatin of *A. sativa*, zeamatin of *Z. mays* (Vigers et al., 1991), and osmotin of *O. sativa* (Medina and Quatrano, 1996) were identified. Based on the phylogenetic tree, it was also possible to identify the orthologous sequences of the three monocots characterized as permatin, zeamatin, and osmotin (Bdi_Bradi4g05440_1, Sit_Seita_2G365800_1, Zma_GRMZM2G010048_T01, respectively). The eudicot representatives, the most basal species in the clade of Angiosperms (*A. trichopoda*), and one sequence of the monocot *S. polyrhiza* formed a subgroup. This subgroup also includes eight previously characterized osmotins: Gma_Glyma_11G025600_1_GmOLPa, Gma_Glyma_05G204600_1_P21 (Onishi et al., 2006), Sni_AAL87640 (AF450276)_SnOLP (Campos et al., 2002), Nta_CAA43854(X61679)_OSM (Kumar and Spencer, 1992), Stu_PGSC0003DMT400007869_OSM (Castillo et al., 2005), Phy_AAK55411(AF376058)_osmotin (Kim et al., 2002), Gma_Glyma_01G217700_1_GmOLPb (Tachi et al., 2009), and Ath_AT4G11650_1_OSM (Capelli et al., 1997). The Sni_SnOLP, Nta_OSM, Stu_OSM, and Phy_osmotin grouped in a Solanaceae subgroup, suggesting a possible duplication in the basis of Solanaceae. The orthologous proteins of these sequences were identified in *Solanum tuberosum* and *Solanum lycopersicum*. The characterized proteins from *G. max* (GmOLPa, GmOLPb, and P21) were dispersed in three distinct subgroups. The *A. thaliana* osmotin (Ath_OSM) was a unique sequence of this species found in the osmotin group. This sequence was grouped in a subgroup of Brassicaceae that presents an orthologous sequence from *Arabidopsis lyrata*.

Another three proteins, PruAV2 and MalD2—previously characterized as allergenic (Inschlag et al., 1998; Gao et al., 2005)—and PR5K (Wang et al., 1996) from *A. thaliana*, which presents a kinase domain, were included in other groups that diverged following the emergence of the osmotin group (Figures 1 and S1).

Structure analyses

Gene and protein structures were examined to explore possible mechanisms of osmotin evolution and diversity. For this purpose, the exon–intron organization pattern, protein primary structure, and conserved residues of the thaumatin domain were analyzed.

Exon–intron organization analysis revealed that nearly all pinales sequences were disrupted by at least one intron, except for Pta_PITA_000000597, which did not present introns (Figure 4a). Fourteen and six putative pinales osmotin sequences presented one and two introns, respectively. The other sequences presented more than two introns, while most monocot and eudicot putative osmotin sequences did not present introns. Five and three monocot sequences presented one and two introns, respectively (Figure 4b). Fifteen and two eudicot sequences presented one and two introns, respectively (Figure 4c). No introns were observed for *S. polyrhiza*, *Medicago truncatula*, *Salix purpurea*, *Mimulus guttatus*, and *Solanum lycopersicum* putative osmotin sequences.

The protein length of putative osmotins varied from 199 to 843 amino acids (Table 1). The molecular weight (MW) was relative to protein length, which varied from 21.21 to 93.57 KDa. Nearly all sequences with a length and MW greater than 290 amino acids and 30 KDa, respectively, presented other domains beyond the thaumatin (THN) domain, including an extra THN domain, phosphotransferases (STYKc, S_TKc), kinases (SCOP d1qpca, Pkinase), and leucine-rich repeat (LRR) domains (Table 1). Six sequences with two THN domains were observed in *P. taeda*, while one sequence was observed in *S. tuberosum*. These sequences did not present transmembrane portions (Table 1). Sequences with THN and phosphotransferase/kinase domains were observed in *P. taeda*, *P. abies*, *Setaria italica*, *Z. mays*, *Brachypodium distachyon*, and *O. sativa*. Nearly all of these sequences presented a transmembrane portion. In monocots, the sequences with THN and phosphotransferase/kinase domains formed a subgroup in the phylogenetic tree (Figure S1). A unique THN domain is generally not accompanied by a transmembrane portion. Apart from THN, any other domain was observed in eudicot sequences of the osmotin group, with the exception of Aco_Aqcoe3G114200_1, which presents an internal repeat (RPT1). Other domains, such as phosphotransferases and kinases, were also observed in non-putative osmotin sequences along the phylogenetic tree (Table S2 and Figure S1). In the non-putative osmotin sequences, only eudicots presented THN and

phosphotransferase/kinase domains. These sequences formed a group in the phylogenetic tree in which the previously characterized PR5K protein is included. A transmembrane portion was also observed in all of these sequences (Table S2 and Figure S1).

Most putative osmotin sequences presented a signal peptide. These proteins were predicted to be targeted to the secretory pathway, with the exception of Zma_GRMZM2G002555_T01 and Pta_PITA_000020282, which were predicted to be targeted to mitochondria (Table 1). Sequences that did not have a signal peptide were predicted to be targeted to a secretory pathway, chloroplast, or mitochondria.

The total isoelectric point (pI) of the putative osmotin proteins varied from 3.83 to 8.32, though most of these sequences were predicted not to present a net charge in acid pH (pI<7.00) (Table 1 and Figure S2b). The non-putative osmotins exhibited a similar range of pI values (Figure S2).

The presence of conserved amino acids described for the thaumatin domain was searched in the aligned sequences of the osmotin group. Sixteen cysteine residues, REDDD and FF hydrophobic motifs, and the acid cleft previously described in previous publications (Petre et al., 2011; Ahmed et al., 2013) were identified in similar positions (Figure 5). Notably, some putative osmotin sequences did not present all REDDD residues (Figure 5 and Table 1). In Aco_Aqcoe1G274500_1, Aco_Aqcoe1G274400_1, Pab_MA_10428085g0010, Pab_MA_69685g0010, and Pab_MA_10327089g0010, the REDDD residues were altered by other amino acids in the same alignment position (Table 1). In some cases, a change in REDDD residues was accompanied by an alteration in acid cleft pI value (pI<7.00 to pI>7.00). This situation was observed for Pta_PITA_000043543, Zma_GRMZM2G010048_T01, and Mgu_Migut_E01128_1, which presented pI values of 8.58, 7.38, and 9.91, respectively (Table 1). A cleft with pI>7.00 was also observed in non-putative osmotin sequences (Table S2). Most of the putative and non-putative osmotin sequences presented an acid cleft with pI values of 3.45 or 3.83 (Table S2).

To identify structural features in the proteins, MEME was used for motif searching (Figures 6 and S3). Combined block diagrams of signatures generated on the basis of the MEME analysis indicated that thaumatin protein architectures are generally broadly conserved in the motifs “e” (burnt yellow), “d” (purple), “i” (dark orange), “f” (pool

green), “b” (light blue), and “a” (red) (Figure S3). Protein sequences localized in the phylogenetic tree before osmotin group emergence were incomplete in motif composition, as observed for the bryophytes and the green algae *C. reinhardtii*. Alternatively, they were complete with motifs organized as e-h-d-i-f-b-g-c-a-t, with the “t” motif being exclusive to such sequences (Figure S3c). The thaumatin sequences localized after the osmotin group emergence presented the e-h(most)/l-d-i-f-b-g(most)/p-c-a-q/none-j motif composition (Figure S3a). Some sequences that present other domains beyond THN possess additional motifs identified as “s”, “m”, “n”, “o”, and “r”. The “o” motif is exclusive to thaumatin sequences with the kinase domain, which were localized following osmotin group emergence in the phylogenetic tree, including the previously characterized thaumatin PR5K of *A. thaliana*. Most putative osmotin sequences presented the motifs e-l-d-i-f-b-k/g-a-q-j (Figures 6 and S3b), with the “k” motif being exclusive to such proteins. Some pinales and monocot putative osmotin sequences that presented other domains beyond THN also possess the motifs s-m-n-r.

Motif “t” and the motifs “c” and “h” of non-putative osmotins were observed both before and after osmotin group emergence, respectively. Additionally, the motifs “k” and “l” of putative osmotins were identified in the inferred 3D proteins structures Ath_AT5G40020_1, Ath_AT4G11650_1_OSM, and Ath_AT2G28790_1 (Figure S4). Motifs “t”, “l”, and “h” were observed in domain I, while motifs “k” and “c” were identified in domain II of the Arabidopsis proteins.

Gene duplication pattern

In order to study the contribution of gene duplication events in the expansion of the osmotin group, Plaza 3.0 software was accessed using the ID of putative osmotin sequences (Table 1). Some species that were not present in the Plaza database were described as “not available”. *A. trichopoda*, *B. distachyon*, *S. lycopersicum*, and *S. tuberosum* putative osmotin sequences were predicted to be tandem duplicated, while most *A. lyrata*, *E. grandis*, *G. max*, *M. esculenta*, and *M. truncatula* sequences were predicted to be tandem and block duplicated (Table 1).

The *O. sativa* and *A. thaliana* genomes (Phytozome v.12.0) were selected and analyzed by MCScanX software to better understand the duplication patterns of putative osmotin genes.

In the whole genome of *O. sativa*, 6,567 (15,56%) and 3,706 (8,78%) segmentally (collinear genes in collinear blocks) and tandem duplicated genes were found, respectively. Regarding putative *O. sativa* osmotins, four genes were found to be segmentally duplicated, all which are located on duplicated segments of chromosomes 3 and 12 (Figure 7). The other two *O. sativa* putative osmotin genes, located in chromosomes 3 and 1, were classified as tandem and dispersed duplications, respectively. Tandem duplicates are consecutive genes repetition in the genome and are presumed to arise through unequal crossing over or localized transposon activities. Dispersed duplicates are neither adjacent to each other in the genome nor within homologous chromosome segments and may result from transposition events. In the whole genome of *A. thaliana*, 10,367 (37,81%) and 8,039 (29,32%) genes were dispersed and segmentally duplicated, respectively. Moreover, BLASTp hits from the unique Arabidopsis osmotin gene matched with other four thaumatin-like genes (AT1G75050, AT1G75040, AT1G75030, and AT1G77700), though the osmotin was classified as dispersed duplication due to their low similarity. Distant single gene transposition may explain the widespread existence of dispersed duplicates within and among genomes (Wang et al., 2011). Evaluation of the chromosome region where the Arabidopsis osmotin gene is located confirmed the presence of several transposable elements (Table S3).

Gene ontology annotation and gene expression pattern

The gene ontology (GO) assessment provided insights regarding the function of putative Arabidopsis and rice osmotins (Table S4). According to these data, the putative osmotins participate in abiotic and biotic stress responses. The LOC_Os01g02310 rice gene also was predicted to be involved in other biological processes and molecular functions. This result corroborates with the presence of other domains in this protein.

The relative expression profiles of putative osmotins under stress conditions were investigated in *O. sativa* and *A. thaliana* using the RNA-seq BAR database and RiceXPro (Figures 8 and 9). The putative rice osmotins LOC_Os01g02310, LOC_03g46060, and LOC_Os12g43450 were highly expressed during drought stress (Figure 8a). LOC_03g46060 and LOC_12g43490 were highly expressed during salt stress, while LOC_03g46070 was highly expressed during salt and cold stresses. Upregulation was also observed for LOC_12g45960, LOC_03g46070, LOC_03g46060, LOC_12g43490, and

LOC_Os12g43450 putative osmotins in inoculated whole leaf with *Magnaporthe oryzae* (Figure 8b). The *A. thaliana* osmotin was predicted to be highly expressed at different time points during drought, salt, and osmotic stresses (Figure 9a). This osmotin was also observed to be highly expressed during challenge with three types of fungi (Figure 9b).

Discussion

With the aim to identify the origin and diversification of osmotins throughout the evolution of the TLP superfamily in Viridiplantae, the present study was performed using genomic resources for 26 species. A total of 716 sequences were retrieved and added to six previously characterized TLPs from six different species. The phylogenetic tree generated using Bayesian analysis included a total of 722 sequences from 32 plant species, revealing distinct groups that suggest a complex pattern of molecular evolution for this superfamily. These groups of genes have likely originated by distinct gene duplication events during plant evolution. Liu et al. (2010) suggested that the large number of groups in the TLP phylogeny may have evolved through multiple rounds of gene duplication. Tandem and block duplications have been also observed by Cao et al. (2016) during the expansion of the TLP gene family. Additionally, the presence of multiple members from the same species in the same group or subgroup suggests that gene duplication events continued to occur throughout the evolution of plant species (Shatters et al., 2006). The distribution of sequences from plants of distinct taxa among phylogenetic groups shows a highly complex mechanism of gene gains and losses during TLP evolution. Confirming data reported by Cao et al. (2016), *C. reinhardtii* and *P. patens* exhibited the lowest number of TLP sequences (one and six, respectively). These results suggest that the expansion of the TLP family occurred after the plant land conquest. An expressive increase in number of TLP sequences in monocots and eudicots has also been observed. Moreover, Petre et al. (2011) reported an important gene expansion in the transition from bryophytes to tracheophytes. However, it is important to highlight that this study did not include the bryophyte *S. fallax*, which has 18 thaumatin-like sequences according to our data.

According to the phylogenetic tree, an osmotin group clustering all previously characterized osmotins was shown to have emerged from Spermatophyte taxon onwards during the expansion of the TLP superfamily (Figures 1 and 2). This phylogenetic separation could be accounted for by the diversification of some amino acid residues

represented by the different motif compositions and organization shared by the putative osmotins (e-l-d-i-f-b-k/g-a-q-j) (Figures 6 and S3b). Amino acid substitutions followed by natural selection may have resulted in a new group of proteins and adaptive functional alterations for some plant PR proteins (Liu et al., 2010). Moreover, the motif analysis contributed to understanding the phylogenetic tree topology and structural characteristics of putative osmotin sequences, as well as the identification of a specific osmotin motif ("k"). Notably, this motif—alongside motifs "t" and "c", which were identified in non-putative osmotins before and after osmotin group emergence, respectively—were localized in domain I of the Arabidopsis proteins (Figure S4). Previous studies have suggested that the antifungal activity of some TLPs could be associated to the domain I of these proteins (Mani et al., 2011; Chen et al., 1999). The osmotin protein could possess many domains that may individually fulfill their own function or in combination with neighbors (Hakim et al., 2017). Some specific motifs were also identified for proteins with phosphotransferases/kinases domains beyond THN, such as s-m-n-r for some pinales and monocot putative osmotins, as well as s-m-n-o-r for eudicot non-putative osmotins. This kinase domain has been cited as the most useful neighbor or partner of osmotin, which is involved in the phosphotransfer reaction and is considered essential for most eukaryotic cell signaling and regulatory procedures (Hakim et al., 2017). These structures were generally accompanied by a transmembrane domain, as observed for the previously characterized PR5K protein. According to Wang et al. (1996), the functional PR5K receptor kinase has an extracellular THN domain responsible for interaction with microbial targets, a central transmembrane-spanning domain, and an intracellular kinase domain related to a family of protein-serine/threonine kinases involved in the expression of self-incompatibility and disease resistance. Both the osmotin protein and kinase-like protein adjust a wide variety of cellular procedures; consequently, the modular nature of osmotin suggests that it serves multiple roles (Hakim et al., 2017).

Notably, the shared motifs among nearly all thaumatin-like proteins were related to the conserved amino acid residues characteristic of the TLP family (REDDD motif in the acid cleft). As demonstrated by Petre at al. (2011), the amino acids forming the TLP acidic cleft are under negative selection for maintaining antifungal activity. On the other hand, some exposed amino acids are under positive selection to avoid pathogen enzyme inhibitors or protease recognition. In the present study, it was observed that nearly all thaumatin-like

proteins—including putative osmotins—conserved a pI<7.00 in the cleft region, which implies that the cleft charge is stabilized in acid pH environments (Tables 1 and S2). The acid cleft was predicted to favor the antifungal activity of these proteins (Salzman et al., 2004). Regarding this aspect, high expression of the Arabidopsis osmotin during fungal challenge was observed (Figure 9b). Previous studies have demonstrated an upregulation of Arabidopsis osmotin during *Alternaria brassicicola* and *Pseudomonas syringae* pv. *tomato* infection (Mukherjee et al., 2010; Mohr and Cahill, 2007). Rice gene expression data also indicates an upregulation of putative osmotins during *M. oryzae* challenge (Figure 8b). Moreover, osmotins are also known to be expressed during abiotic stresses (Kumar et al., 2015). Putative Arabidopsis and rice osmotins were expressed during drought, salt, and osmotic stresses (Figures 8 and 9). Notably, Jiang et al. (2007) observed a differential expression of osmotin in Arabidopsis roots under salt stress. Transgenic *Solanum tuberosum* plants overexpressing the osmotin encoding gene from Arabidopsis provides evidence for the involvement of this protein in the mechanisms of salt tolerance (Evers et al., 1999). Upregulation was also observed in five putative rice osmotins (LOC_Os03g46070, LOC_Os03g45960, LOC_Os12g43490, LOC_Os12g43450, and LOC_Os03g46060) during drought stress in available RNAseq data from Zong et al. (2013) (Table S3).

Most putative osmotins were observed to present total pI >7. This could be also related to the fact that most of these proteins were predicted to be targeted to the secretory pathway. This subcellular location was also proposed by Lehtonen et al. (2014) and Petre et al. (2011), and the results of the present study reinforce the participation of these proteins in defense mechanisms. As previously suggested, angiosperms display defense mechanisms against pathogen infection that are conserved. Some genes related to defense were acquired by duplication events and later underwent diversification and functional specialization (León and Montesano, 2017). As consequence of plant-pathogen co-evolution, the size of some multigene families involved in resistance increased greatly, as observed for the TLP family (Petre et al., 2011).

Gene duplications have had an important contribution to the expansion and emergence of new groups inside the thaumatin-like gene family. Polyploidy or large segment duplication is one manner in which family member copy numbers increase, though tandem or local

gene duplication due to unequal crossover or conversion events may also be an important mode of family member expansion (Cannon et al., 2004). In the present study, an uneven chromosome distribution of putative osmotin genes accompanied by tandem and block duplications was observed for different plant species. Tandem and block duplications have been also highlighted by Cao et al. (2016) during the expansion of the TLP gene family. According to Cannon et al. (2004), the Arabidopsis TLP gene family has been estimated to have undergone five large-scale segmental duplications and two local duplication events. As such, it was demonstrated that the Arabidopsis osmotin gene was dispersed, and locally duplicated (likely due to transposable element action). Associated to TLP family evolution, gene intron gains and losses were identified (Table S1). The number of introns per gene varied among putative osmotins. Briefly, it was observed that nearly all pinales sequences were disrupted by at least one intron, while most monocots and eudicots did not present introns (Figure 4). According to Koonin (2006), introns are important components of eukaryotic genes, and their gain or loss affect the complexity of genetic structure. It has been hypothesized that intron loss could lead to more efficient transcription and subsequent gene expression (Wang et al., 2014). This is an interesting hypothesis for genes related to abiotic and biotic stress responses, such as those observed for putative osmotin-encoding genes.

In conclusion, the present study provides a robust phylogenetic analysis of TLPs in Viridiplantae and new information regarding osmotin evolution. In addition, the analyses of gene and protein structure facilitated an improved understanding of the emergence and expansion of the osmotin group formed during TLP evolution. This is the first study to report a monophyletic group for osmotins and its origin in Spermatophytes. Our results indicate that tandem and block duplications events lead to the gene expansion and diversification of the osmotin group. Conserved residues related to defense response were maintained during TLP evolution, while different amino acids represented by specific motif compositions provided the emergence of new groups, including the osmotin group.

Abbreviations

TLP: thaumatin-like proteins

OLPs: osmotin-like proteins

PR-5: protein family 5

AA: amino acids

CDS: coding sequences

WGD: whole-genome duplication

Availability of data and materials

The data analyzed in this study are available on Phytozome v.12.0, Congenie, and the NCBI database. See Table S1 for accession numbers.

Phytozome v.12.0 Database. http://www.phytozome.net/. Accessed 8 May 2017.

Congenie Database. http://congenie.org/. Accessed 16 May 2017.

NCBI Database. http://www.ncbi.nlm.nih.gov. Accessed 16 May 2017.

The Arabidopsis Information Resource (TAIR10). http://www.arabidopsis.org/. Accessed 9 July 2017.

The software used in the present study are available at the links below:

Molecular Evolutionary Genetics Analysis (MEGA) - https://www.megasoftware.net/

Simple Modular Architecture Research Tool (SMART) - http://smart.embl-heidelberg.de/

ProtTest 3.4 - http://darwin.uvigo.es/software/prottest2_server.html

Bayesian Evolutionary Analysis Sampling Trees (BEAST) - http://beast.community/

Figtree - http:// tree.bio.ed.ac.uk/software/figtree/

Gene Structure Display Server (GSDS) - http://gsds.cbi.pku.edu.cn/

TargetP 1.1 Server - http://www.cbs.dtu.dk/services/TargetP/

WebLogo - https://weblogo.berkeley.edu/logo.cgi

WGmapping available in Plaza 3.0 - http://bioinformatics.psb.ugent.be/plaza/

MCScanX software - http://chibba.pgml.uga.edu/mcscan2/

Swiss-model - https://swissmodel.expasy.org/

Acknowledgements

We thank the two anonymous reviewers for their helpful comments on the manuscript.

Author contributions

Participated in the study design: GRF, LAO-B, ACT-Z, and MHB-Z. Performed the *in silico* analyses: GRF. Performed gene duplication analysis: FLGE. Performed the phylogenetic analysis: GRF and ACT-Z. Wrote the paper: GRF. Revised the paper: LAO-B, ACT-Z, and MHB-Z. Supervised and coordinated the study: MHB-Z. All authors read and approved the final manuscript.

Compliance with ethical standards

Funding

This study was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and INCT – MCTI/CNPq/CAPES/FAPs nº 16/2014, Ativos Biotecnológicos Aplicados a Seca e Pragas em Culturas Relevantes para o Agronegócio (INCT Biotec Seca-Pragas) [88887.136360/2017-00 - 465480/2014-4]. The funds provided by these funding institutions have been used to pay the stipends of GRF and LAO-B, respectively.

Conflict of interest

GRF declares that she has no conflict of interest.

ACT-Z declares that she has no conflict of interest.

FLGE declares that he has no conflict of interest.

LAO-B declares that she has no conflict of interest.

MHB-Z declares that she has no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21: 2104–2105

Abdin MZ, Kiran U, Alam A (2011) Analysis of osmotin, a PR protein as metabolic modulator in plants. Bioinformation 5: 336

Ahmed NU, Park JI, Jung HJ, Chung MY, Cho YG, Nou IS (2013) Characterization of thaumatin-like gene family and identification of *Pectobacterium carotovorum* subsp. *carotovorum* Inducible Genes in *Brassica oleracea*. Plant Breed Biotech 1: 111-121

Campos MA, Ribeiro SG, Rigden DJ, Monte DC, Grossi de Sa MF (2002) Putative pathogenesis-related genes within *Solanum nigrum* var. americanum genome: isolation of two genes coding for PR5-like proteins, phylogenetic and sequence analysis. Physiol Mol Plant Pathol 61: 205-216

Cannon SB, Mitra A, Baumgarten A, Young ND, May G (2004) The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. BMC Plant Biol 4: 10

Cao J, Lv Y, Hou Z, Li X, Ding L (2016) Expansion and evolution of thaumatin-like protein (TLP) gene family in six plants. Plant Growth Regul 79: 299–307

Capelli N, Diogon N, Greppin H, Simon P (1997) Isolation and characterization of a cDNA clone encoding an osmotin-like protein from *Arabidopsis thaliana*. Gene 191: 51-6

Castillo RA, Herrera C, Ghislain M, Gebhardt C (2005) Organization of phenylalanine ammonia lyase (PAL), acidic PR-5 and osmotin-like (OSM) defence-response gene families in the potato genome. Mol Genet Genomics 274: 168-179

Chen WP, Chen PD, Liu DJ, Kynast R, Friebe B, Velazhahan R, Muthukrishnan S, Gill BS (1999) Development of wheat scab symptoms is delayed in transgenic wheat plants that constitutively express a rice thaumatin-like protein gene. Theor Appl Genet 99: 755–760

Dong W, Vannozzi A, Chen F, Hu Y, Chen Z, Zhang L (2018) MORC domain definition and evolutionary analysis of the MORC gene family in green plants. Genome biology and evolution 10: 1730-1744

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969–1973

Emanuelsson O, Nielsen H, Brunak S, Von Herijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300: 1005-1016

Evers D, Overney S, Simon P, Greppin H, Hausman JF (1999) Salt tolerance of *Solanum tuberosum* L.: Overexpressing an heterologous osmotin-like protein. Biol Plant 42:105-12

Gao ZS, Weg WE, Schaart JG, Arkel G, Breiteneder H, Hoffmann-Sommergruber K, Gilissen LJ (2005) Genomic characterization and linkage mapping of the apple allergen genes Mal d 2 (thaumatin-like protein) and Mal d 4 (profilin). Theor Appl Genet 111: 1087-1097

Hakim, Ullah A, Hussain A, Shaban M, Khan AH, Alariqi M, Gul S, Jun Z, Lin S, Li J, Jin S, Munis MFH (2017) Osmotin: a plant defense tool against biotic and abiotic stresses. Plant Physiology and Biochemistry 123: 149-159

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31: 1296-1297

Inschlag C, Hoffmann-Sommergruber K, O'Riordain G, Ahorn H, Ebner C, Scheiner O, Breiteneder H (1998) Biochemical characterization of Pru a 2, a 23-kD thaumatin-like protein representing a potential major allergen in cherry (*Prunus avium*). Int Arch Allergy Immunol 116: 22-28

Jiang Y, Yang B, Harris NS, Deyholos MK (2007) Comparative proteomic analysis of NaCl stress-responsive proteins in Arabidopsis roots. Journal of Experimental Botany 58: 3591–3607

Kim H, Mun JH, Byun BH, Hwang HJ, Kwon YM, Kim SG (2002) Molecular cloning and characterization of the gene encoding osmotin protein in Petunia hybrid. Plant Sci 162: 745-752

Koonin EV (2006) The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate. Biol Direct 1: 22

Kozlowski LP (2016) IPC–isoelectric point calculator. Biology direct 11: 55

Kumar V, Spencer ME (1992) Nucleotide sequence of an osmotin cDNA from the *Nicotiana tabacum* cv. white burley generated by the polymerase chain reaction. Plant Mol Biol 18: 621-622.

Kumar SA, Kumari PH, Kumar GS, Mohanalatha C, Kishor PK (2015) Osmotin: a plant sentinel and a possible agonist of mammalian adiponectin. Frontiers in Plant Science 6: 163

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33: 1870-1874

Lehtonen MT, Takikawa Y, Rönnholm G, Akita M, Kalkkinen N, Ahola-Iivarinen E, et al (2014) Protein secretome of moss plants (*Physcomitrella patens*) with emphasis on changes induced by a fungal elicitor. J Proteome Res 13: 447–459

León IP, Montesano M (2017) Adaptation Mechanisms in the Evolution of Moss Defenses to Microbes. Frontiers in Plant Science 8: 366

Letunic I, Bork P (2017) 20 years of the SMART protein domain annotation resource. Nucleic Acids Research 46: D493-D496

Liu JJ, Sturrock R, Ekramoddoullah AKM (2010) The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Rep 29: 419–436

Mani T, Sivakumar KC, Manjula S (2012) Expression and functional analysis of two osmotin (PR5) isoforms with differential antifungal activity from *Piper colubrinum*: prediction of structure–function relationship by bioinformatics approach. Molecular biotechnology 52: 251-261

Medina J, Quatrano RS (1996) Characterization of a rice cDNA (Accession No. L76377) encoding an Osmotin-like protein (PGR96-058). Plant Physiol 111: 1354

Mohr PG, Cahill DM (2007) Suppression by ABA of salicylic acid and lignin accumulation and the expression of multiple genes, in Arabidopsis infected with *Pseudomonas syringae* pv. *tomato*. Functional & integrative genomics 7: 181-191

Mukherjee AK, Carp MJ, Zuchman R, Ziv T, Horwitz BA, Gepstein S (2010) Proteomics of the response of *Arabidopsis thaliana* to infection with *Alternaria brassicicola*. Journal of proteomics 73: 709-720

Onishi M, Tachi H, Kojima T, Shiraiwa M, Takahara H (2006) Molecular cloning and characterization of a novel salt-inducible gene encoding an acidic isoform of PR-5 protein in soybean. Plant Physiol Biochem 44: 574–580

Petre B, Major I, Rouhier N, Duplessis S (2011) Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar. BMC Plant Biology 11: 33

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6. Available from http://tree.bio.ed.ac.uk/software/tracer/

Salzman RA, Koiwa H, Ibeas JI, Pardo JM, Hasegawa PM, Bressan RA (2004) Inorganic cations mediate plant PR5 protein antifungal activity through fungal Mnn1-and Mnn4-regulated cell surface glycans. Molecular plant-microbe interactions 17: 780-788

Shatters RG, Boykin LM, Lapointe SL, Hunter WB, Weathersbee III AA (2006) Phylogenetic and structural relationships of the PR5 gene family reveal an ancient multigene family conserved in plants and select animal taxa. J Mol Evol 63:12–29

Tachi H, Yamada KF, Kojima T, Shiraiwa M, Takahara H (2009) Molecular characterization of a novel soybean gene encoding a neutral PR-5 protein induced by high-salt stress. Plant Physiol Biochem 47: 73–79

Vigers AJ, Roberts WK, Selitrennikoff CP (1991) A new family of plant antifungal proteins. Mol Plant Microbe Interact 4: 315-23

Wang H, Devos KM, Bennetzen JL (2014) Recurrent Loss of Specific Introns during Angiosperm Evolution. PLOS Genetics 10: e1004843

Wang X, Zafian P, Choudhary M, Lawton M (1996) The PR5K receptor protein kinase from *Arabidopsis thaliana* is structurally related to a family of plant defense proteins. Proc Natl Acad Sci USA 93: 2598-602

Wang Y, Wang X, Tang H, Tan X, Ficklin SP , Feltus FA, Paterson AH (2011) Modes of Gene Duplication Contribute Differently to Genetic Novelty and Redundancy, but Show Parallels across Divergent Angiosperms. Plos One 6: e28150

Wang Y, Tang H, DeBarry JD, Xu T, Li J, Wang X, Lee T, Jin H, Marler B, Kissinger HGJC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research 40: e49-e49

Zong W, Zhong X, You J, Xiong L (2013) Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress. Plant molecular biology 81: 175-188

Table

Table 1

Sequences according tree position	Duplication pattern	Lenght	MW (kDa)	pI	Signal peptide	Subcelular location[1]	Transmembrane portion	Domains[2]	REDDD motif					pI acid cleft
Stu_PGSC0003DMT400007869_OSM	tandem	247	26.60	7.2	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Sly_Solyc08g080640_1_1	tandem	247	26.64	7.04	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Phy_AAK55411(AF376058)_osmotin	not available	246	26.40	7.05	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Stu_PGSC0003DMT400007870	tandem	246	26.66	6.16	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Sly_Solyc08g080650_1_1	tandem	246	26.70	6.16	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Sly_Solyc08g080620_1_1	tandem	226	24.58	7.33	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Nta_CAA43854(X61679)_OSM	not available	245	26.69	7.25	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Stu_PGSC0003DMT400007868	tandem	247	26.62	6.04	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Stu_PGSC0003DMT400007865	tandem	226	24.61	5.16	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Sni_AAL87640(AF450276)_SnOLP	not available	246	26.63	5.56	YES	S (1)	NO	THN	R	E	D	D	D	2.95
Stu_PGSC0003DMT400034395	any information	216	23.58	6.79	YES	S (1)	NO	THN	R	E	D	E	D	3.53
Stu_PGSC0003DMT400007905	tandem	250	27.33	5.83	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Sly_Solyc08g080660_1_1	tandem	250	27.48	5.84	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Sly_Solyc08g080670_1_1	tandem	250	27.49	5.5	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Stu_PGSC0003DMT400007906	tandem	250	27.44	5.5	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Egr_Eucgr_D01888_1	tandem and block	241	26.17	6.5	YES	S (5)	NO	THN	R	E	D	D	D	3.45
Sly_Solyc12g056390_1_1	tandem	227	24.98	5.15	NO	S (1)	NO	THN	R	E	D	D	D	3.45
Stu_PGSC0003DMT400010886	tandem	227	25.00	5.15	NO	S (1)	NO	THN	R	E	D	D	D	3.45
Mgu_Migut_E01125_1	not available	229	24.80	5.85	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Stu_PGSC0003DMT400010890	tandem	426	46.43	6.31	YES	S (1)	NO	THN;THN	R	E	D	D	D	3.45
Spu_SapurV1A_2357s0010_1	not available	225	24.29	4.26	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Spu_SapurV1A_0507s0110_1	not available	225	24.29	4.26	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Mes_Manes_02G025100_1	tandem and block	226	24.65	4.83	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Mes_Manes_02G028300_1	tandem and block	225	24.43	5.17	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_D01887_1	tandem and block	227	24.28	4.53	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Egr_Eucgr_D01893_1	tandem and block	223	23.76	4.97	YES	S (5)	NO	THN	R	E	D	N	D	4.03
Egr_Eucgr_D01899_1	tandem and block	211	22.53	4.64	NO	_ (3)	NO	THN	R	E	D	N	D	4.03
Spu_SapurV1A_0507s0100_1	not available	246	26.47	6.08	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Spu_SapurV1A_2357s0020_1	not available	246	26.53	6.08	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Spu_SapurV1A_0098s0100_1	not available	246	26.48	6.46	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Spu_SapurV1A_3598s0010_1	not available	246	26.42	6.73	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Pvu_Phvul_002G155500_1	not available	242	26.10	5.87	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Gma_Glyma_01G217700_1_GmOLPb	tandem and block	249	27.21	6.06	NO	S (4)	NO	THN	R	E	D	D	D	3.45
Mtr_Medtr5g010635_1	tandem and block	241	26.22	5.57	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Mes_Manes_01G064300_1	tandem and block	247	27.11	5.6	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Mes_Manes_01G064200_1	tandem and block	247	27.12	6.47	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Mes_Manes_01G064400_1	tandem and block	217	23.26	4.57	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Ath_AT4G11650_1_OSM	block	244	26.63	5.77	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Aly_AL6G46240_t1	tandem and block	244	26.64	5.57	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Cgr_Cagra_1912s0003_1	not available	244	26.58	5.77	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Aly_AL6G46230_t1	tandem and block	223	24.27	5.51	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Cgr_Cagra_1912s0002_1	not available	223	24.41	6.93	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Bra_Brara_B02758_1	any information	244	26.35	4.83	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Mgu_Migut_E01123_1	not available	230	24.92	5.23	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Aco_Aqcoe7G040700_1	not available	201	21.62	5.2	NO	_ (5)	NO	THN	R	E	D	D	D	3.45
Aco_Aqcoe7G040600_1	not available	224	24.25	5.53	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Aco_Aqcoe3G114200_1	not available	292	31.57	6.42	YES	S (1)	NO	THN;RPT1	R	E	D	D	D	3.45
Aco_Aqcoe7G040400_1	not available	237	26.01	6.44	YES	S (1)	NO	THN	R	E	D	D	D	3.76
Atr_evm_27_model_AmTr_v1_0_scaffold00032_18	tandem	231	24.79	6.92	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Atr_evm_27_model_AmTr_v1_0_scaffold00032_17	tandem	217	23.29	5.8	NO	C (4)	NO	THN	A	E	D	D	D	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00032_19	tandem	199	21.21	4.46	YES	S (1)	NO	THN	R	E	D	E	D	4.29
Egr_Eucgr_D01904_1	tandem and block	225	24.20	7.01	YES	S (1)	NO	THN	R	E	D	D	D	3.83

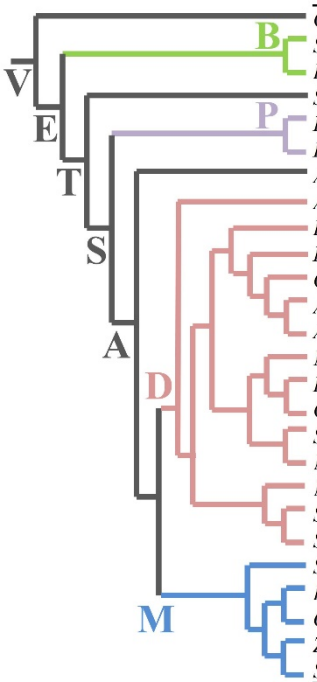
Egr_Eucgr_L02566_1	any information	218	23.41	6.67	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_D01892_1	tandem and block	225	24.19	7.24	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_H03864_1	tandem	225	24.20	7.34	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_D01898_1	tandem and block	225	24.25	7.34	YES	S (1)	NO	THN	R	E	V	D	D	4.03
Egr_Eucgr_H03865_1	tandem	225	24.15	7.13	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_L02568_1	any information	223	24.29	7.24	NO	C (5)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_H03863_1	tandem	212	22.84	6.88	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_D01894_1	tandem and block	243	25.83	5.18	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Egr_Eucgr_D01900_1	tandem and block	231	24.79	6.86	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_L03623_1	any information	227	24.25	6.86	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_L01962_1	any information	225	24.20	7.34	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Spu_SapurV1A_0183s0030_1	not available	226	24.66	6.72	YES	S (1)	NO	THN	R	E	D	D	D	3.76
Spu_SapurV1A_0271s0330_1	not available	226	24.61	6.73	YES	S (1)	NO	THN	R	E	D	D	D	3.76
Spu_SapurV1A_0271s0340_1	not available	225	24.79	7.04	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Spu_SapurV1A_0183s0040_1	not available	225	24.43	5.1	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Spu_SapurV1A_0271s0350_1	not available	225	24.56	6.43	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Spu_SapurV1A_0183s0050_1	not available	225	24.69	7.04	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Pvu_Phvul_002G286500_1	not available	222	23.77	4.78	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Pvu_Phvul_002G286600_1	not available	221	23.70	4.68	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Gma_Glyma_05G204600_1_P21	tandem and block	224	23.90	5.21	YES	S (2)	NO	THN	R	E	D	D	D	3.83
Gma_Glyma_05G204800_1	tandem and block	222	23.81	7	YES	S (2)	NO	THN	R	E	D	D	D	3.83
Spo_Spipo32G0003200	any information	223	23.93	6.72	YES	S (4)	NO	THN	R	E	D	D	D	3.83
Aco_Aqcoe1G274500_1	not available	268	29.89	6.26	NO	S (1)	NO	THN	Q	G	H	N	Q	4.93
Aco_Aqcoe1G274400_1	not available	257	28.80	6.01	YES	S (1)	NO	THN	Q	G	H	N	Q	4.93
Aco_Aqcoe3G114100_1	not available	263	28.86	7.28	NO	_ (4)	YES	THN	R	E	D	D	D	3.45
Mes_Manes_02G028400_1	tandem and block	213	23.20	4.67	YES	S (1)	NO	THN	R	E	D	D	D	3.42
Bra_Brara_A01323_1	any information	231	25.18	6.73	NO	S (1)	NO	THN	R	E	D	D	D	4.18
Aly_AL7G30420_t1	block	231	25.09	6.7	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Cgr_Cagra_15158s0009_1	not available	231	25.26	6.5	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Spu_SapurV1A_0761s0080_1	not available	227	24.18	5.59	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Mes_Manes_02G025200_1	tandem and block	223	24.16	5.07	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_E00560_1	tandem and block	235	25.39	5.89	YES	S (3)	NO	THN	R	E	D	D	D	3.83
Egr_Eucgr_E00561_1	tandem and block	208	22.41	6.14	NO	C (3)	NO	THN	S	E	D	E	D	4.29
Mtr_Medtr8g096900_1	tandem and block	248	27.22	7.2	NO	M (3)	YES	THN	R	E	D	D	D	3.83
Sly_Solyc12g056360_1_1	tandem	229	25.51	5.19	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Stu_PGSC0003DMT400010891	tandem	216	23.65	4.86	NO	_ (5)	NO	THN	R	E	D	E	D	3.53
Sly_Solyc12g056380_1_1	tandem	229	25.46	5.29	NO	S (1)	NO	THN	R	E	G	N	D	3.8
Gma_Glyma_01G217600_1	tandem and block	223	23.89	4.58	YES	S (1)	NO	THN	R	D	D	D	D	3.39
Gma_Glyma_11G025600_1_GmOLPa	block	224	23.86	4.44	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Pvu_Phvul_002G155400_1	not available	225	23.85	4.25	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Mtr_Medtr5g010640_1	tandem and block	236	25.32	4.95	YES	S (1)	NO	THN	R	E	D	Q	D	4.03
Mgu_Migut_A00799_1	not available	227	25.10	7.18	YES	S (1)	NO	THN	K	E	D	E	E	3.93
Mgu_Migut_E01128_1	not available	223	24.26	8.29	YES	S (1)	NO	THN	K	E	D	E	T	9.91
Sit_Seita_5G078100_1	tandem	458	50.05	5.35	YES	S (1)	YES	THN;STYKc	R	E	D	D	D	3.83
Sit_Seita_5G078300_1	tandem	409	44.90	5.99	YES	S (1)	YES	THN;SCOP d1qpca	R	E	D	D	D	3.83
Osa_LOC_Os01g02310_1	tandem	627	68.67	5.88	YES	S (1)	YES	THN;S_TKc	R	E	D	D	D	4.03
Bdi_Bradi2g01217_1	tandem	585	64.79	6.74	YES	S (1)	YES	THN;Pkinase	R	E	D	N	S	3.8
Zma_GRMZM2G002555_T01	any information	645	71.07	6.37	YES	M (5)	YES	THN;Pkinase	L	E	D	G	Q	3.8
Zma_GRMZM2G435592_T02	any information	653	71.83	6.31	YES	S (2)	YES	THN;S_TKc	R	E	D	D	N	3.59
Bdi_Bradi2g01146_1	tandem	612	68.42	6.21	YES	S (1)	YES	THN;S_TKc	R	D	D	N	-	4.2
Bdi_Bradi2g01200_2	tandem	634	69.87	6.86	YES	S (2)	NO	THN;S_TKc	R	E	D	D	-	3.59
Sit_Seita_5G078200_1	tandem	623	68.39	5.71	YES	S (1)	YES	THN;S_TKc	S	E	D	D	T	2.95
Bdi_Bradi2g01228_1	tandem	421	46.37	5.99	YES	S (1)	NO	THN;S_TKc	R	E	D	D	-	3.59
Sit_Seita_9G148300_1	tandem and block	227	23.70	4.26	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Zma_GRMZM2G039639_T01	block	230	23.80	3.94	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Bdi_Bradi1g13060_1	tandem	239	25.44	4.45	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Bdi_Bradi1g13070_1	tandem	234	24.49	7.07	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Osa_LOC_Os03g46070_1	tandem and block	229	23.76	4.37	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Osa_LOC_Os03g45960_1	tandem and block	232	24.09	5.77	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Sit_Seita_2G365800_1	any information	231	24.19	6.92	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Zma_GRMZM2G374971_T01_Zeamatin	any information	227	23.99	6.73	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Zma_GRMZM2G010048_T01	block	233	23.86	6.2	YES	S (5)	NO	THN	R	E	D	H	T	7.38
Osa_LOC_Os03g46060_1_OSM	tandem and block	222	22.76	6.51	YES	S (1)	NO	THN	R	E	D	D	G	4.03
Asa_AAB02259(U57787)_permatin	not available	228	23.67	7.2	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Bdi_Bradi4g05440_1	tandem	227	23.85	5.2	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Osa_LOC_Os12g43490_1	tandem and block	258	26.51	5.08	YES	S (3)	NO	THN	R	E	D	D	D	3.83

Osa_LOC_Os12g43450_1	tandem and block	238	25.00	7.2	YES	S (1)	NO	THN	R	E	D	D	D	4.18
Zma_GRMZM2G092474_T01	any information	220	23.06	7.07	YES	S (2)	NO	THN	R	E	D	D	D	3.83
Sit_Seita_5G355700_1	any information	222	22.89	4.59	YES	S (1)	NO	THN	R	E	D	D	D	4.18
Bdi_Bradi4g03290_1	tandem	226	23.22	6.45	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Bdi_Bradi4g05430_2	tandem	256	26.32	6.67	NO	_ (3)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000010735	any information	233	24.81	7.07	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000010737	any information	234	24.75	7.2	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000010739	any information	510	54.05	7.1	YES	S (1)	NO	THN;THN	R	E	D	D	D	3.83
Pta_PITA_000010740	any information	234	24.78	7.07	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000000597	any information	208	21.88	4.07	NO	_ (4)	NO	THN	R	E	D	E	D	3.92
Pta_PITA_000041533	any information	444	46.22	5.53	YES	S (2)	NO	THN;THN	R	E	D	D	D	3.45
Pta_PITA_000069097	any information	234	23.91	4.02	YES	S (2)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000070827	any information	379	38.94	4.14	YES	S (2)	NO	THN;THN	R	E	D	D	D	3.83
Pab_MA_10429511g0010	any information	245	26.03	4.77	NO	_ (1)	NO	THN	R	E	D	D	D	3.83
Pab_MA_6505g0010	any information	245	26.01	4.89	NO	S (4)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000024408	any information	242	25.71	5.77	YES	S (1)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000066768	any information	235	2,511	4.87	NO	_ (5)	NO	THN	R	E	D	D	D	3.83
Pta_PITA_000087912	any information	567	62.21	8.32	YES	S (1)	YES	THN;S_TKc	R	E	D	D	D	3.45
Pta_PITA_000058482	any information	644	71.08	7.31	YES	S (1)	YES	THN;S_TKc	R	E	D	D	D	3.45
Pta_PITA_000002550	any information	240	25.98	7.04	YES	S (1)	NO	THN	R	E	D	D	-	3.8
Pta_PITA_000069801	any information	475	51.92	7.55	NO	S (4)	NO	THN;S_TKc	R	E	D	D	D	3.45
Pta_PITA_000020282	any information	594	64.98	7.94	YES	M (4)	YES	THN;S_TKc	R	E	D	D	D	3.83
Pta_PITA_000091230	any information	324	34.78	8.17	YES	S (1)	NO	THN	R	E	D	D	D	3.45
Pta_PITA_000078522	any information	252	26.53	4.37	NO	_ (3)	NO	THN	S	E	D	N	-	3.09
Pab_MA_10435621g0020	any information	499	54.20	8.16	NO	S (3)	YES	THN;S_TKc	R	E	D	D	D	3.45
Pab_MA_3795g0010	any information	242	25.77	4.5	NO	S (1)	NO	THN	R	E	D	E	D	3.53
Pab_MA_10428085g0010	any information	843	93.57	7.39	NO	_ (3)	YES	THN;RPT 1; LRR_8; S_TKc	-	K	G	N	E	6.56
Pta_PITA_000018832	any information	508	53.83	4.66	YES	S (1)	NO	THN;THN	K	E	L	G	D	4.49
Pta_PITA_000093937	any information	510	53.88	4.47	YES	S (1)	NO	THN;THN	R	E	F	G	D	4.03
Pta_PITA_000004949	any information	698	76.23	6.37	NO	_ (2)	YES	THN;S_TKc	R	E	A	G	D	3.7
Pta_PITA_000037146	any information	667	72.60	6	NO	S (3)	YES	THN;S_TKc	R	E	A	G	D	3.8
Pta_PITA_000091324	any information	838	91.49	5.22	YES	S (2)	YES	THN;THN;S_TKc	R	E	D	G	Y	4.4
Pab_MA_473307g0010	any information	553	61.35	5.44	NO	S (4)	YES	THN;STYKc	R	E	G	G	N	4.4
Pab_MA_69685g0010	any information	406	44.10	5.31	NO	_ (5)	YES	THN	E	K	E	Y	E	6.64
Pta_PITA_000043543	any information	630	70.60	6.5	NO	M (3)	YES	THN;STYKc	H	K	G	Y	D	8.58
Pta_PITA_000038798	any information	508	55.58	5.69	YES	S (1)	NO	THN;THN	R	G	Y	D	T	3.53
Pta_PITA_000038163	any information	237	25.65	6.38	YES	S (1)	NO	THN	R	G	Y	N	T	4.19
Pab_MA_10327089g0010	any information	237	25.45	5.8	YES	S (2)	NO	THN	K	G	Y	E	N	3.96
Pta_PITA_000000202	any information	214	22.11	3.83	NO	C (4)	NO	THN	R	E	D	D	D	3.83
Pab_MA_133779g0010	any information	232	24.33	4.16	YES	S (1)	NO	THN	R	E	D	D	D	3.83

[1]Subcellular locations are numbered according to the TargetP server, from one to five, where one indicates the strongest prediction. C, chloroplast; M, mitochondria; S, secretory pathway; _, any other location

[2]Domains predicted by SMART: THN, Thaumatin; RPT1, Internal Repeat 1; STYKc, phosphotransferases, possible dual-specificity Ser/Thr/Tyr kinase; SCOP d1qpca, protein-kinase like; S_TKc, phosphotransferases, serine or threonine-specific kinase; Pkinase, protein kinase domain; LRR_8, leucine-rich repeat domain.

Figures



Species	Osmotin group	Total of Thaumatin-likes
Chlamydomonas reinhardtii	-	1
Sphagnum fallax	-	18
Physcomitrella patens	-	6
Selaginella moellendorffii	-	14
Pinus taeda	26	39
Picea abies	9	13
Amborella trichopoda	3	15
Aquilegia coerulea	7	28
Eucalyptus grandis	18	50
Brassica rapa	2	34
Capsella grandiflora	3	26
Arabidopsis lyrata	3	25
Arabidopsis thaliana	1	22
Medicago truncatula	3	37
Phaseolus vulgaris	4	25
Glycine max	5	57
Salix purpurea	13	53
Manihot esculenta	7	36
Mimulus guttatus	4	32
Solanum tuberosum	10	27
Solanum lycopersicum	8	28
Spirodela polyrhiza	1	12
Brachypodium distachyon	9	26
Oryza sativa	6	27
Zea mays	6	34
Setaria italica	6	31

Fig. 1 Number of total TLP sequences and putative osmotins according to the phylogenetic tree for 26 species selected from the Phytozome v.12.0 and Congenie databases. Species are organized according to their taxonomic distribution in the species tree. Species in *green*, *purple*, *red*, and *blue* represent the bryophytes, pinales, eudicots, and monocots, respectively. The names of internal nodes are abbreviated (V=Viridiplantae, E=Embryophyte, B=Bryophytes, T=Tracheophytes, S=Spermatophytes, P=Pinales, A=Angiosperm, M=Monocots, and D=Eudicots). The previously characterized sequences selected from the NCBI database (*Solanum nigrum*, *Nicotiana tabacum*, *Petunia x hybrida*, *Avena sativa*, *Prunus avium*, and *Malus domestica*) were not included in the figure.

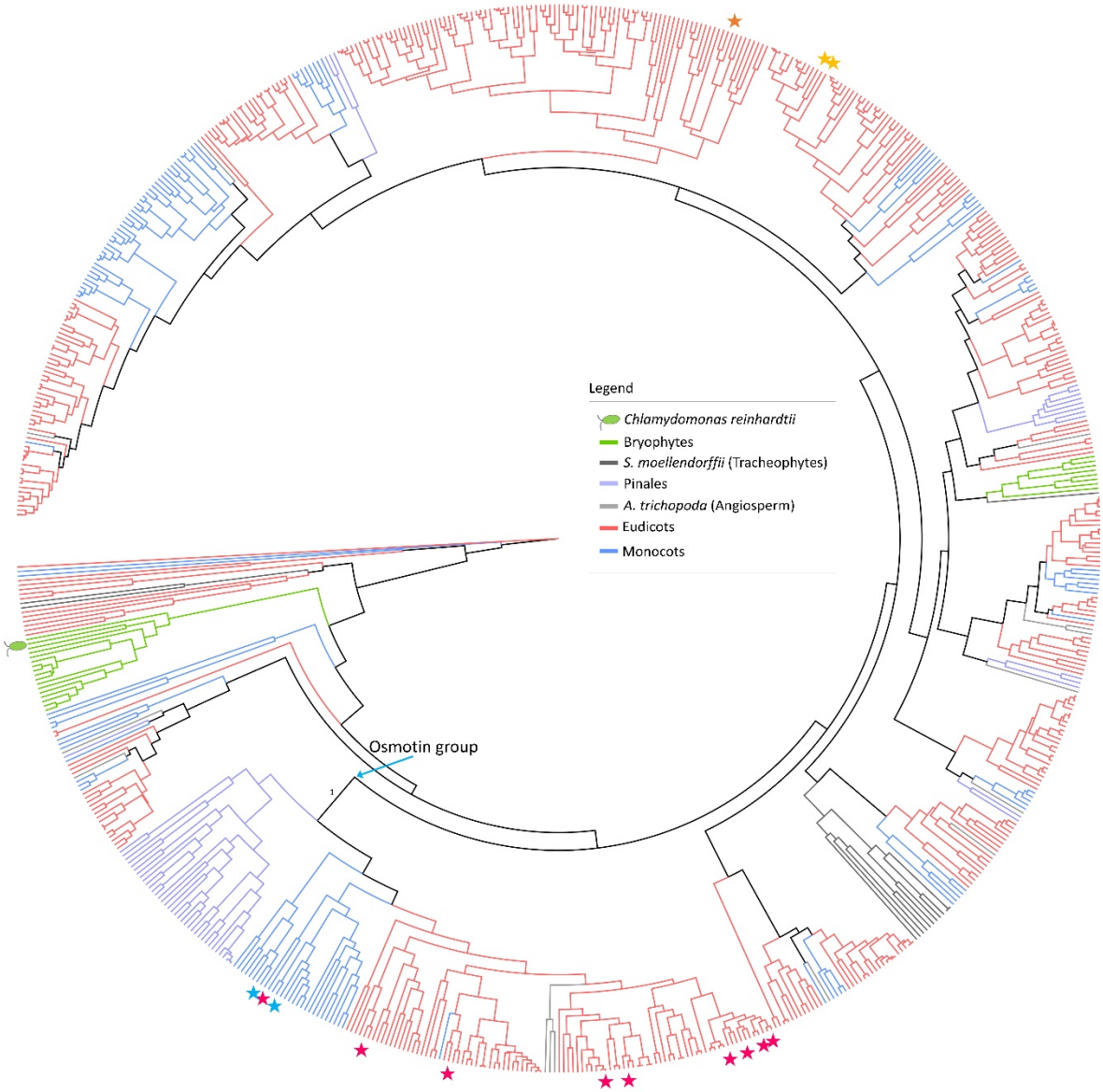


Fig. 2 Phylogenetic tree comprising 722 thaumatin domains from 32 plant species. *Pink*, *blue*, *yellow*, and *orange* stars indicate the previously characterized permatin, osmotin, allergenic, and kinase proteins, respectively. *Blue* arrow indicates the osmotin group. For clarity, protein names and posterior probabilities are not indicated, but can be retrieved in Figure S1.

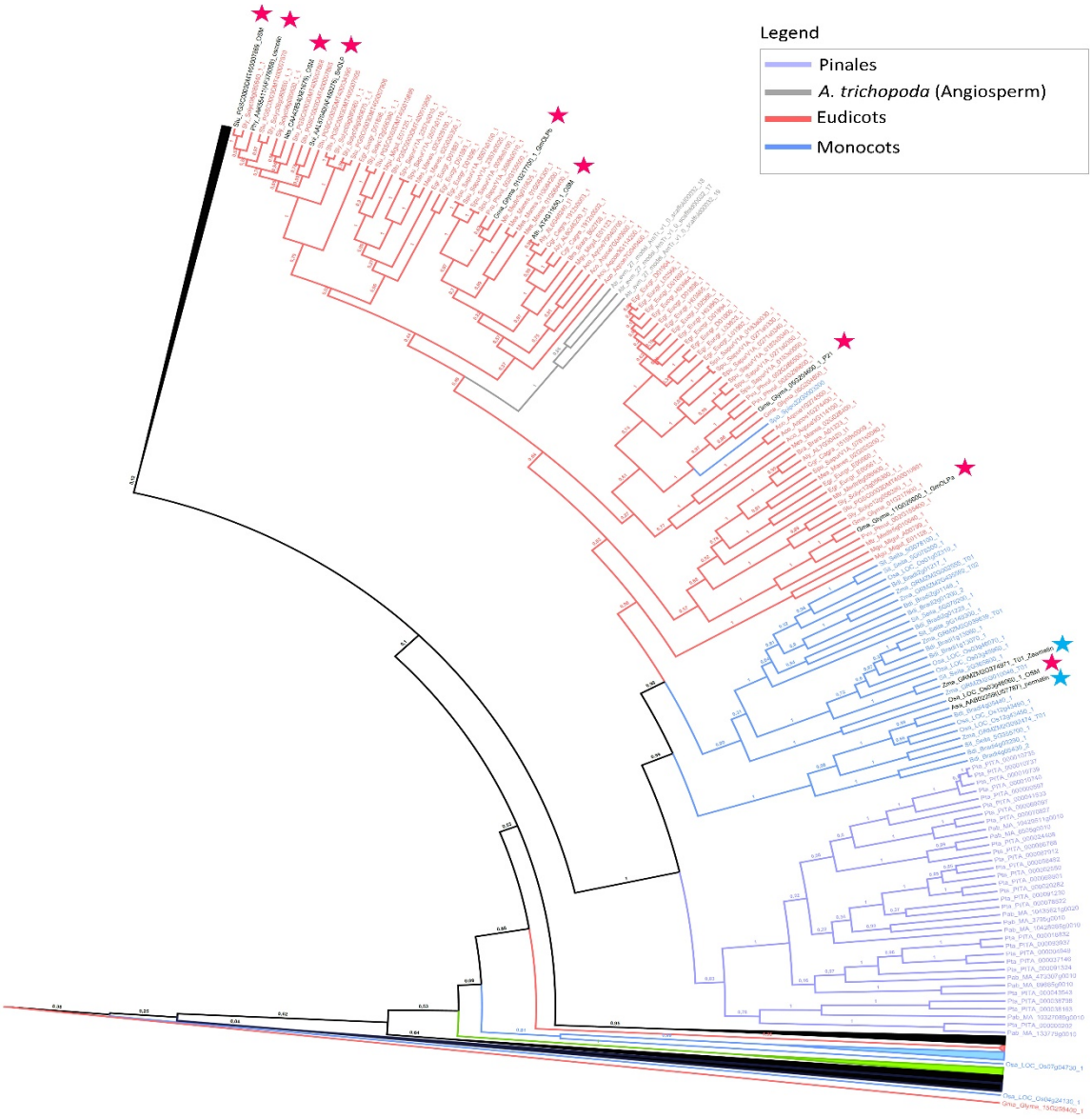
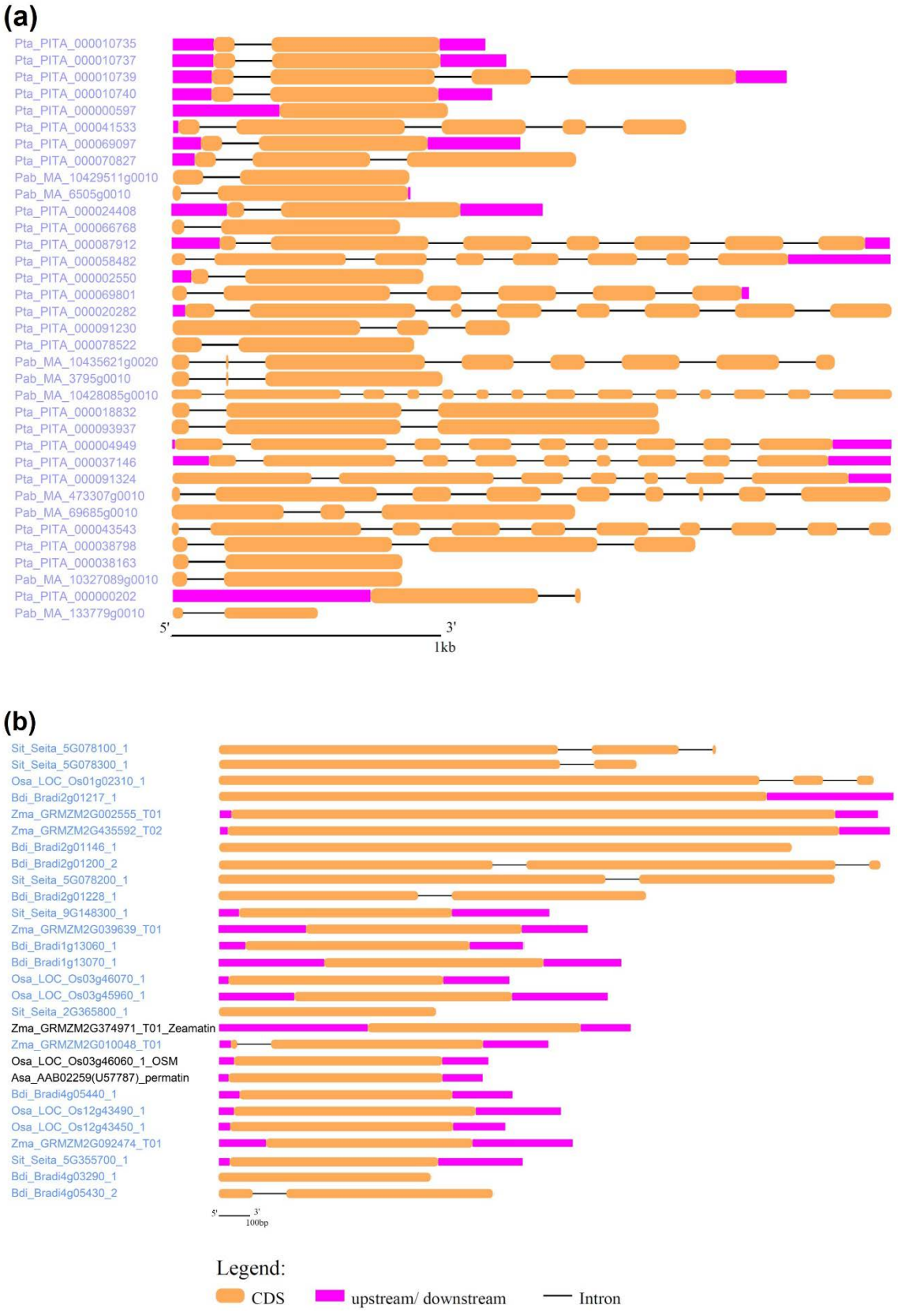


Fig. 3 Detailed representation of the phylogenetic relationship among thaumatin domains within the osmotin group. Branch lengths are proportional to phylogenetic distances. Previously characterized permatins and osmotins are indicated by *blue* and *pink* stars, respectively.


(a)
Pta_PITA_000010735
Pta_PITA_000010737
Pta_PITA_000010739
Pta_PITA_000010740
Pta_PITA_000000597
Pta_PITA_000041533
Pta_PITA_000069097
Pta_PITA_000070827
Pab_MA_10429511g0010
Pab_MA_6505g0010
Pta_PITA_000024408
Pta_PITA_000066768
Pta_PITA_000087912
Pta_PITA_000058482
Pta_PITA_000002550
Pta_PITA_000069801
Pta_PITA_000020282
Pta_PITA_000091230
Pta_PITA_000078522
Pab_MA_10435621g0020
Pab_MA_3795g0010
Pab_MA_10428085g0010
Pta_PITA_000018832
Pta_PITA_000093937
Pta_PITA_000004949
Pta_PITA_000037146
Pta_PITA_000091324
Pab_MA_473307g0010
Pab_MA_69685g0010
Pta_PITA_000043543
Pta_PITA_000038798
Pta_PITA_000038163
Pab_MA_10327089g0010
Pta_PITA_000000202
Pab_MA_133779g0010
5'
3'
1kb
(b)
Sit_Seita_5G078100_1
Sit_Seita_5G078300_1
Osa_LOC_Os01g02310_1
Bdi_Bradi2g01217_1
Zma_GRMZM2G002555_T01
Zma_GRMZM2G435592_T02
Bdi_Bradi2g01146_1
Bdi_Bradi2g01200_2
Sit_Seita_5G078200_1
Bdi_Bradi2g01228_1
Sit_Seita_9G148300_1
Zma_GRMZM2G039639_T01
Bdi_Bradi1g13060_1
Bdi_Bradi1g13070_1
Osa_LOC_Os03g46070_1
Osa_LOC_Os03g45960_1
Sit_Seita_2G365800_1
Zma_GRMZM2G374971_T01_Zeamatin
Zma_GRMZM2G010048_T01
Osa_LOC_Os03g46060_1_OSM
Asa_AAB02259(U57787)_permatin
Bdi_Bradi4g05440_1
Osa_LOC_Os12g43490_1
Osa_LOC_Os12g43450_1
Zma_GRMZM2G092474_T01
Sit_Seita_5G355700_1
Bdi_Bradi4g03290_1
Bdi_Bradi4g05430_2
5'
3'
100bp
Legend:
CDS
upstream/ downstream
Intron

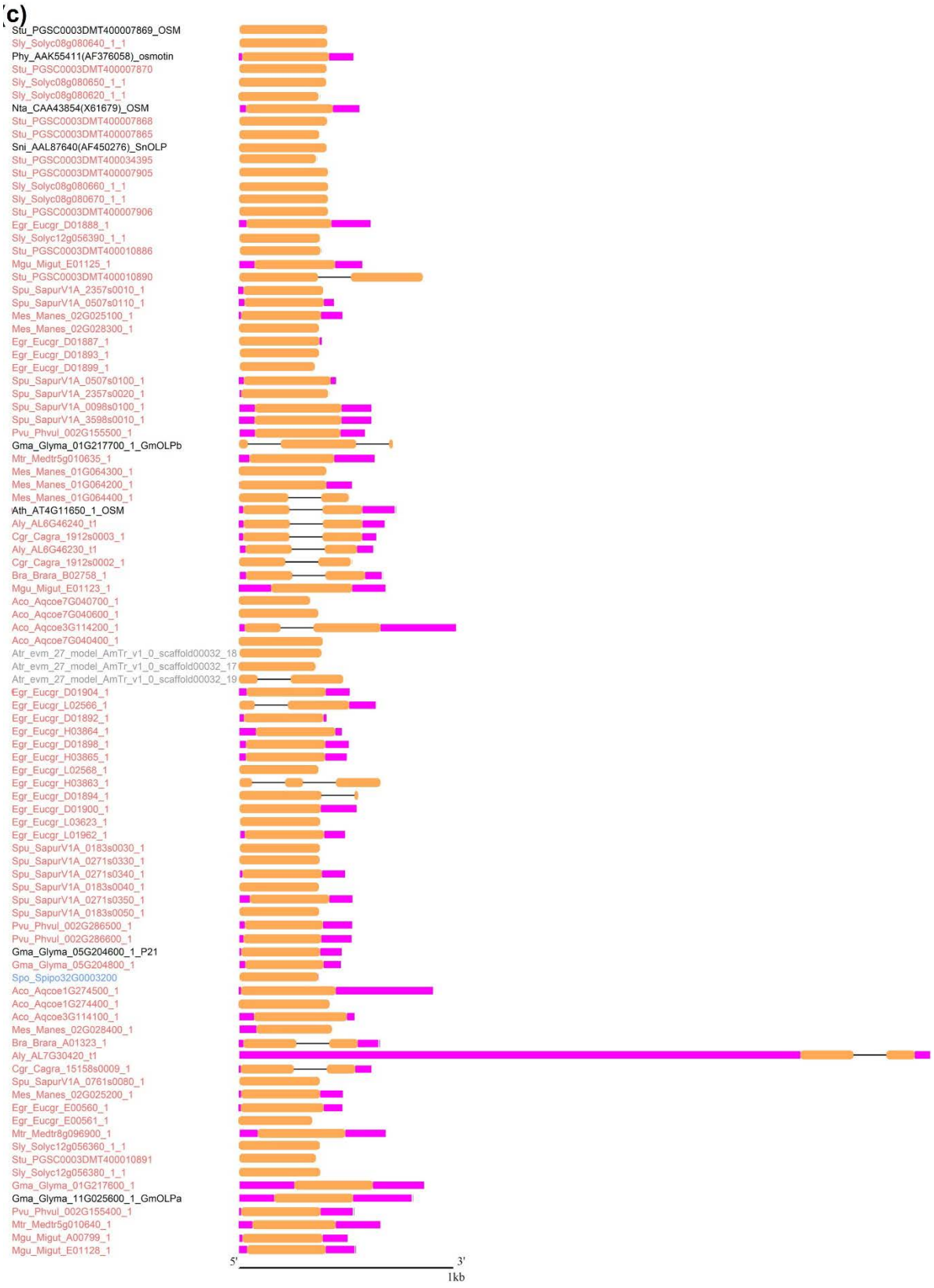


Fig. 4 Gene structure of putative osmotins according to the phylogenetic tree organization. **(a)** Pinales, **(b)** monocots, and **(c)** eudicots, and *Spirodela polyrhiza*.

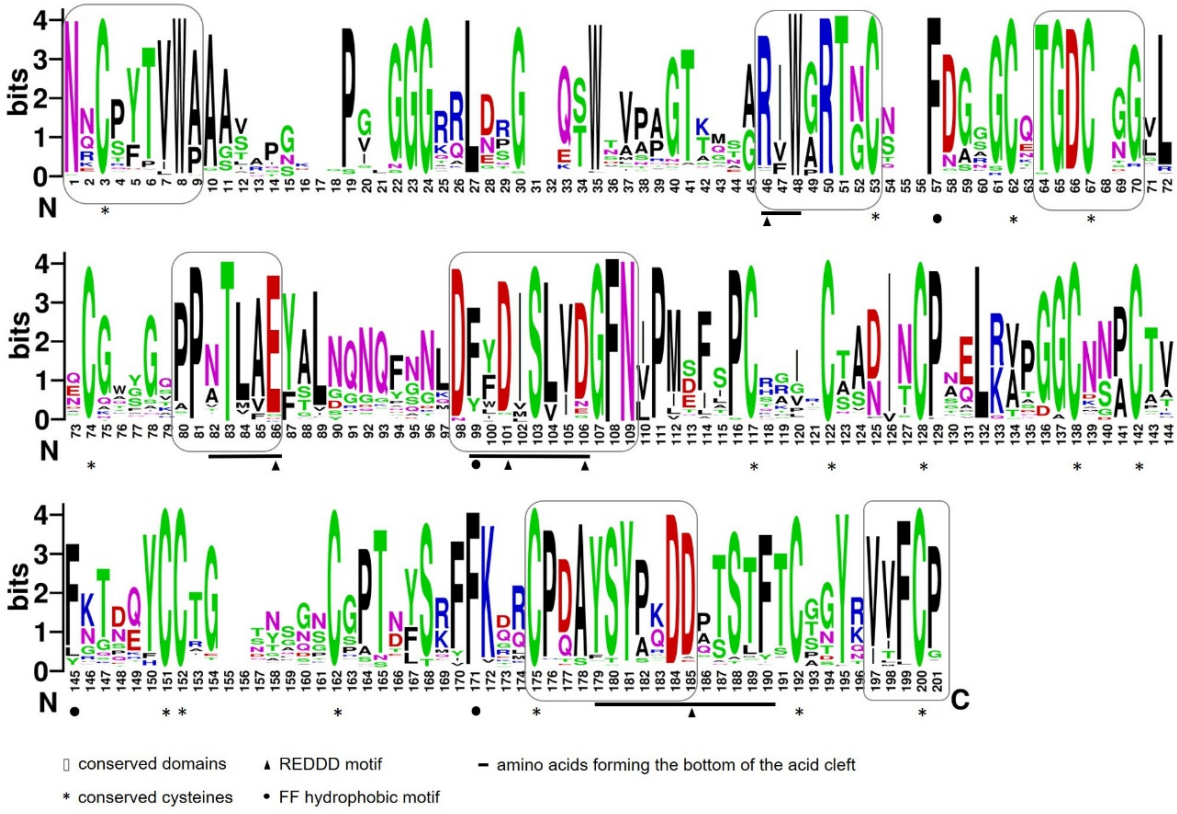
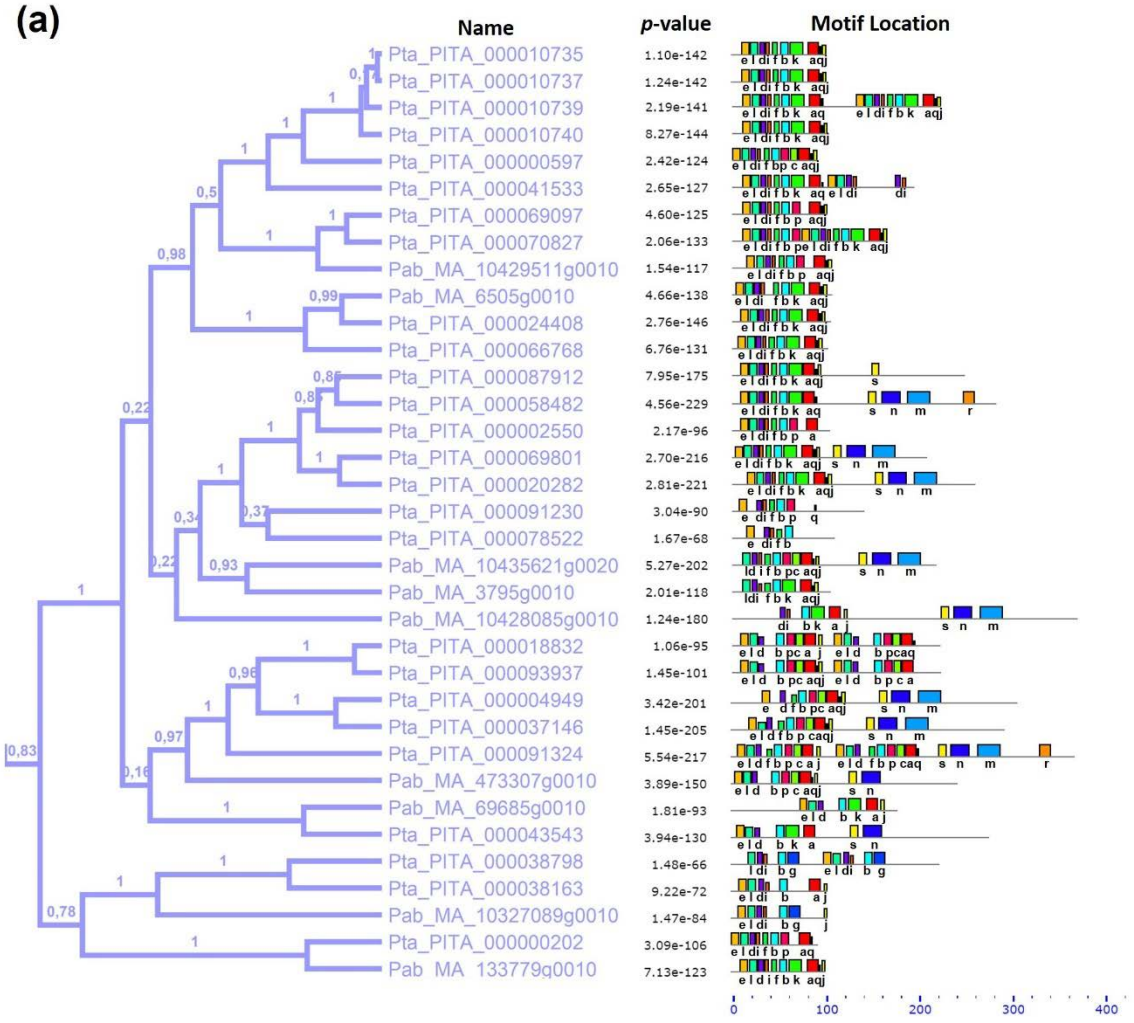
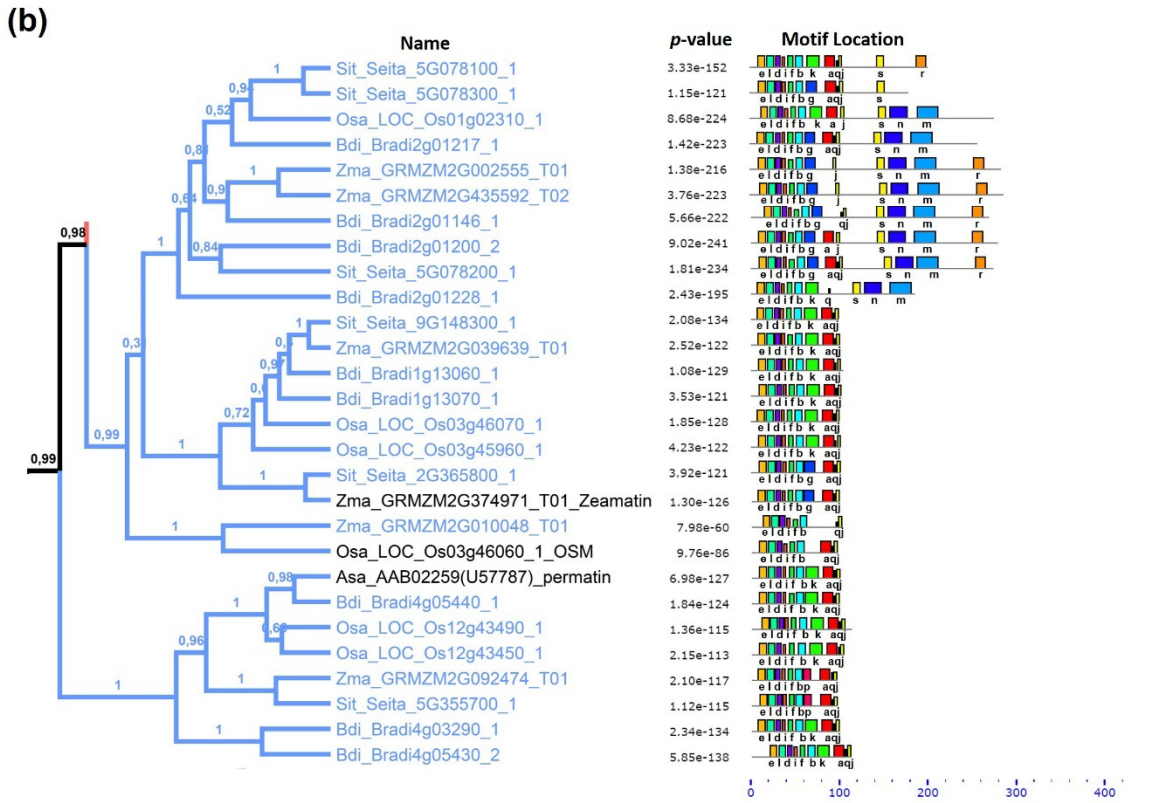
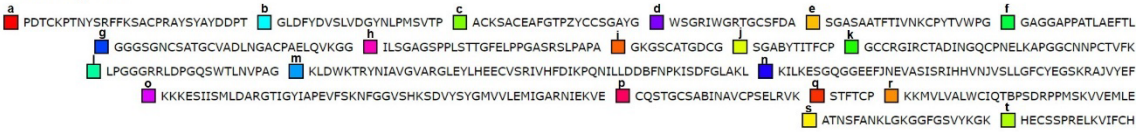


Fig. 5 Logo of thaumatin domain sequence alignments of the osmotin group. The logo alignment residue numeration does not represent the canonical domain sequence numbering. Conserved residues and protein secondary structure are indicated according to Petre et al. (2011). Amino acid frequency was analyzed using WebLogo 2.8.2 online software.


(a)
Name
p-value
Motif Location
Pta_PITA_000010735
Pta_PITA_000010737
Pta_PITA_000010739
Pta_PITA_000010740
Pta_PITA_000000597
Pta_PITA_000041533
Pta_PITA_000069097
Pta_PITA_000070827
Pab_MA_10429511g0010
Pab_MA_6505g0010
Pta_PITA_000024408
Pta_PITA_000066768
Pta_PITA_000087912
Pta_PITA_000058482
Pta_PITA_000002550
Pta_PITA_000069801
Pta_PITA_000020282
Pta_PITA_000091230
Pta_PITA_000078522
Pab_MA_10435621g0020
Pab_MA_3795g0010
Pab_MA_10428085g0010
Pta_PITA_000018832
Pta_PITA_000093937
Pta_PITA_000004949
Pta_PITA_000037146
Pta_PITA_000091324
Pab_MA_473307g0010
Pab_MA_69685g0010
Pta_PITA_000043543
Pta_PITA_000038798
Pta_PITA_000038163
Pab_MA_10327089g0010
Pta_PITA_000000202
Pab_MA_133779g0010
1.10e-142
1.24e-142
2.19e-141
8.27e-144
2.42e-124
2.65e-127
4.60e-125
2.06e-133
1.54e-117
4.66e-138
2.76e-146
6.76e-131
7.95e-175
4.56e-229
2.17e-96
2.70e-216
2.81e-221
3.04e-90
1.67e-68
5.27e-202
2.01e-118
1.24e-180
1.06e-95
1.45e-101
3.42e-201
1.45e-205
5.54e-217
3.89e-150
1.81e-93
3.94e-130
1.48e-66
9.22e-72
1.47e-84
3.09e-106
7.13e-123
0
100
200
300
400


(b)
Name
p-value
Motif Location
Sit_Seita_5G078100_1
Sit_Seita_5G078300_1
Osa_LOC_Os01g02310_1
Bdi_Bradi2g01217_1
Zma_GRMZM2G002555_T01
Zma_GRMZM2G435592_T02
Bdi_Bradi2g01146_1
Bdi_Bradi2g01200_2
Sit_Seita_5G078200_1
Bdi_Bradi2g01228_1
Sit_Seita_9G148300_1
Zma_GRMZM2G039639_T01
Bdi_Bradi1g13060_1
Bdi_Bradi1g13070_1
Osa_LOC_Os03g46070_1
Osa_LOC_Os03g45960_1
Sit_Seita_2G365800_1
Zma_GRMZM2G374971_T01_Zeamatin
Zma_GRMZM2G010048_T01
Osa_LOC_Os03g46060_1_OSM
Asa_AAB02259(U57787)_permatin
Bdi_Bradi4g05440_1
Osa_LOC_Os12g43490_1
Osa_LOC_Os12g43450_1
Zma_GRMZM2G092474_T01
Sit_Seita_5G355700_1
Bdi_Bradi4g03290_1
Bdi_Bradi4g05430_2
3.33e-152
1.15e-121
8.68e-224
1.42e-223
1.38e-216
3.76e-223
5.66e-222
9.02e-241
1.81e-234
2.43e-195
2.08e-134
2.52e-122
1.08e-129
3.53e-121
1.85e-128
4.23e-122
3.92e-121
1.30e-126
7.98e-60
9.76e-86
6.98e-127
1.84e-124
1.36e-115
2.15e-113
2.10e-117
1.12e-115
2.34e-134
5.85e-138
0
100
200
300
400


Motif legend
a PDTCKPTNYSRFFKSACPRAYSYAYDDPT
b GLDFYDVSLVDGYNLPMSVTP
c ACKSACEAFGTPZYCCSGAYG
d WSGRIWGRTGCSFDA
e SGASAATFTIVNKCPYTVWPG
f GAGGAPPATLAEFTL
g GGGSGNCSATGCVADLNGACPAELQVKGG
h ILSGAGSPPLSTTGFELPPGASRSLPAPA
i GKGSCATGDCG
j SGABYTITFCP
k GCCRGIRCTADINGQCPNELKAPGGCNNPCTVFK
l LPGGGRRLDPGQSWTLNVPAG
m KLDWKTRYNIAVGVARGLEYLHEECVSRIVHFDIKPQNILLDDBFNPKISDFGLAKL
n KILKESGQGGEEFJNEVASISRIHHVNJVSLLGFCYEGSKRAJVYEF
o KKKESIISMLDARGTIGYIAPEVFSKNFGGVSHKSDVYSYGMVVLEMIGARNIEKVE
p CQSTGCSABINAVCPSELRVK
q STFTCP
r KKMVLVALWCIQTBPSDRPPMSKVVEMLE
s ATNSFANKLGKGGFGSVYKGK
t HECSSPRELKVIFCH

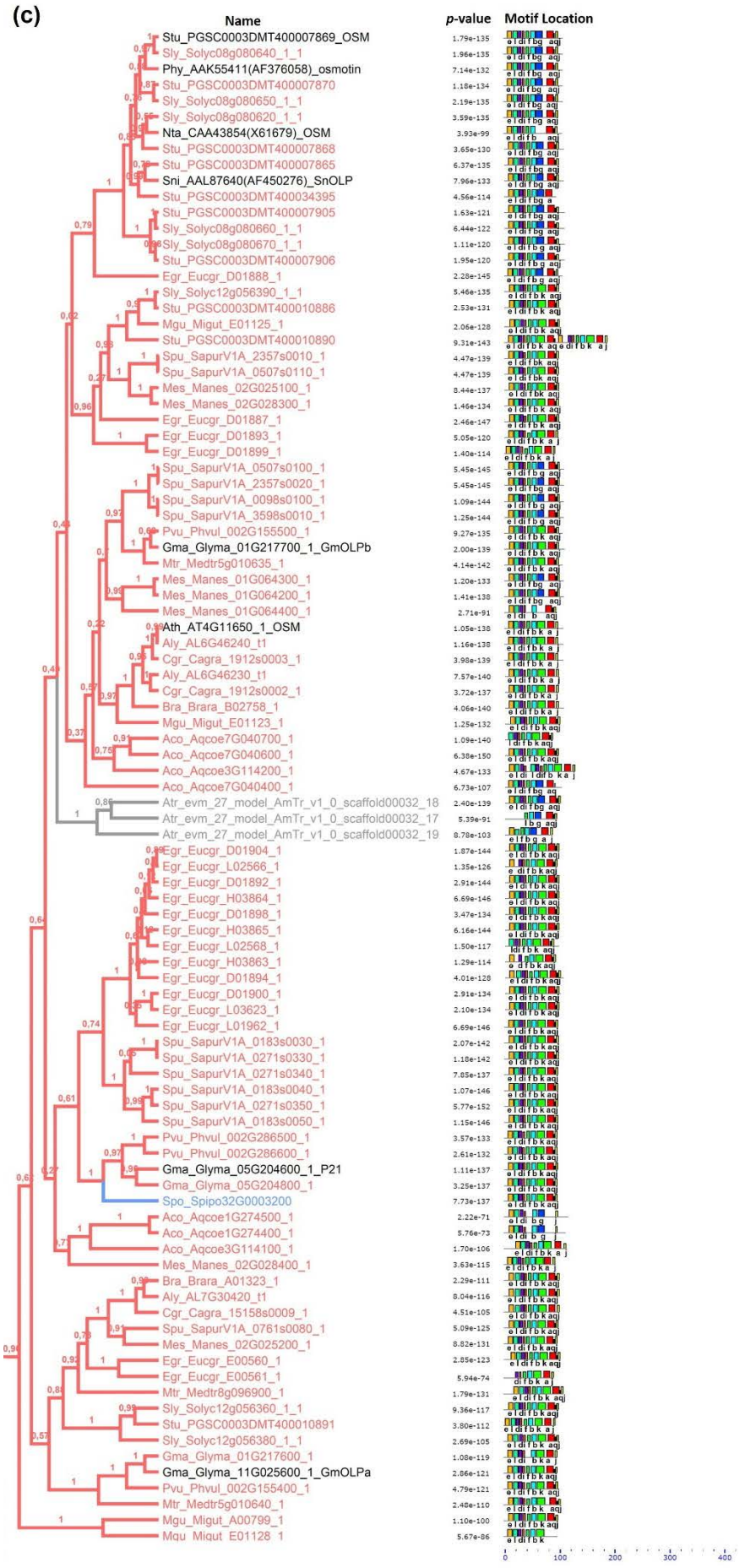

(c)
Name
p-value
Motif Location
Stu_PGSC0003DMT400007869_OSM
Sly_Solyc08g080640_1_1
Phy_AAK55411(AF376058)_osmotin
Stu_PGSC0003DMT400007870
Sly_Solyc08g080650_1_1
Sly_Solyc08g080620_1_1
Nta_CAA43854(X61679)_OSM
Stu_PGSC0003DMT400007868
Stu_PGSC0003DMT400007865
Sni_AAL87640(AF450276)_SnOLP
Stu_PGSC0003DMT400034395
Stu_PGSC0003DMT400007905
Sly_Solyc08g080660_1_1
Sly_Solyc08g080670_1_1
Stu_PGSC0003DMT400007906
Egr_Eucgr_D01888_1
Sly_Solyc12g056390_1_1
Stu_PGSC0003DMT400010886
Mgu_Migut_E01125_1
Stu_PGSC0003DMT400010890
Spu_SapurV1A_2357s0010_1
Spu_SapurV1A_0507s0110_1
Mes_Manes_02G025100_1
Mes_Manes_02G028300_1
Egr_Eucgr_D01887_1
Egr_Eucgr_D01893_1
Egr_Eucgr_D01899_1
Spu_SapurV1A_0507s0100_1
Spu_SapurV1A_2357s0020_1
Spu_SapurV1A_0098s0100_1
Spu_SapurV1A_3598s0010_1
Pvu_Phvul_002G155500_1
Gma_Glyma_01G217700_1_GmOLPb
Mtr_Medtr5g010635_1
Mes_Manes_01G064300_1
Mes_Manes_01G064200_1
Mes_Manes_01G064400_1
Ath_AT4G11650_1_OSM
Aly_AL6G46240_t1
Cgr_Cagra_1912s0003_1
Aly_AL6G46230_t1
Cgr_Cagra_1912s0002_1
Bra_Brara_B02758_1
Mgu_Migut_E01123_1
Aco_Aqcoe7G040700_1
Aco_Aqcoe7G040600_1
Aco_Aqcoe3G114200_1
Aco_Aqcoe7G040400_1
Atr_evm_27_model_AmTr_v1_0_scaffold00032_18
Atr_evm_27_model_AmTr_v1_0_scaffold00032_17
Atr_evm_27_model_AmTr_v1_0_scaffold00032_19
Egr_Eucgr_D01904_1
Egr_Eucgr_L02566_1
Egr_Eucgr_D01892_1
Egr_Eucgr_H03864_1
Egr_Eucgr_D01898_1
Egr_Eucgr_H03865_1
Egr_Eucgr_L02568_1
Egr_Eucgr_H03863_1
Egr_Eucgr_D01894_1
Egr_Eucgr_D01900_1
Egr_Eucgr_L03623_1
Egr_Eucgr_L01962_1
Spu_SapurV1A_0183s0030_1
Spu_SapurV1A_0271s0330_1
Spu_SapurV1A_0271s0340_1
Spu_SapurV1A_0183s0040_1
Spu_SapurV1A_0271s0350_1
Spu_SapurV1A_0183s0050_1
Pvu_Phvul_002G286500_1
Pvu_Phvul_002G286600_1
Gma_Glyma_05G204600_1_P21
Gma_Glyma_05G204800_1
Spo_Spipo32G0003200
Aco_Aqcoe1G274500_1
Aco_Aqcoe1G274400_1
Aco_Aqcoe3G114100_1
Mes_Manes_02G028400_1
Bra_Brara_A01323_1
Aly_AL7G30420_t1
Cgr_Cagra_15158s0009_1
Spu_SapurV1A_0761s0080_1
Mes_Manes_02G025200_1
Egr_Eucgr_E00560_1
Egr_Eucgr_E00561_1
Mtr_Medtr8g096900_1
Sly_Solyc12g056360_1_1
Stu_PGSC0003DMT400010891
Sly_Solyc12g056380_1_1
Gma_Glyma_01G217600_1
Gma_Glyma_11G025600_1_GmOLPa
Pvu_Phvul_002G155400_1
Mtr_Medtr5g010640_1
Mgu_Migut_A00799_1
Mgu_Migut_E01128_1
0
100
200
300
400

Fig. 6 Architecture of conserved protein motifs in putative osmotin sequences identified using the MEME tool. Each motif is represented by a colored block with a letter below. The lengths and positions of the blocks correspond to the lengths and positions of motifs in the individual protein sequences. The height of each block is proportional to its −log (*p* value), truncated at the height corresponding to a motif with a *p* value of 1e −10. The gene names and combined *p* values are shown on the left side of the figure. The scale indicates the lengths of the proteins as well as the motifs. **(a)** Pinales, **(b)** monocots, **(c)** eudicots, and *Spirodela polyrhiza*.

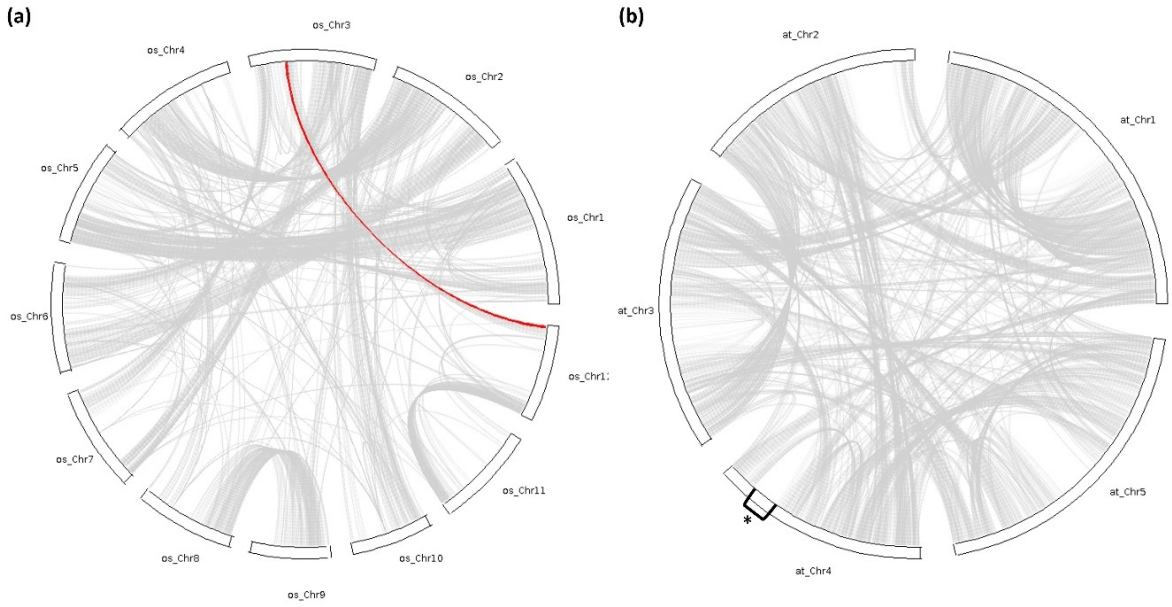


Fig. 7 Collinear gene pairs for osmotin genes on **(a)** 12 *O. sativa* chromosomes and **(b)** 5 *A. thaliana* chromosomes. *Grey* lines indicate the collinear gene pairs in the whole genome, while *red* lines indicate the collinear gene pairs for osmotin genes. An asterisk (*) indicates the non-duplicated region where *A. thaliana* osmotin is localized.

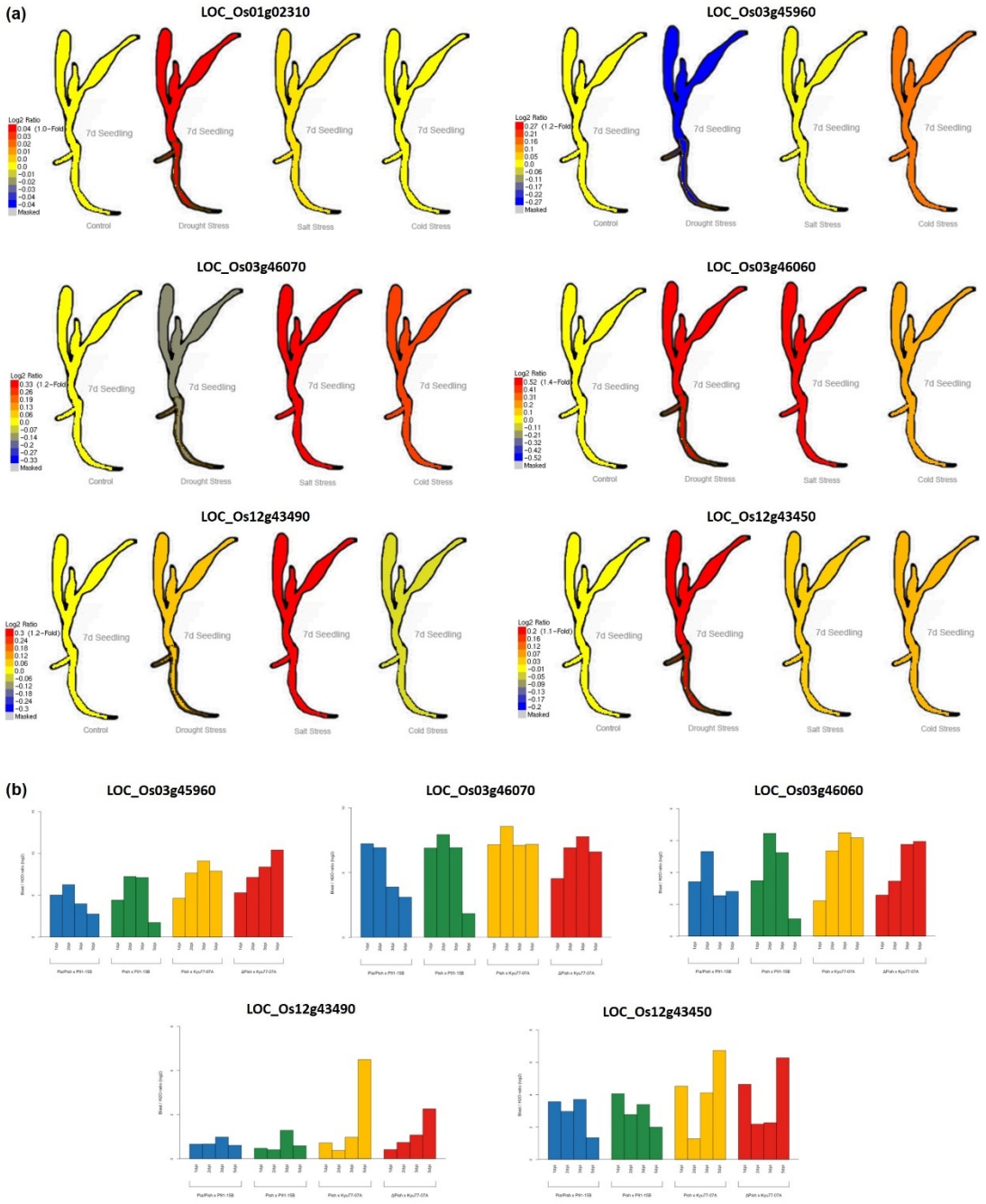


Fig. 8 Relative expression profile of *O. sativa* putative osmotin genes. **(a)** As plotted in the bar scale, *red* and *blue* colors indicate up- and downregulated genes (Log 2 ratio) under abiotic stresses. **(b)** Signal intensity data based on 75 percentile normalization and log2 transformation for the average relative value to H_2O treatment (blast / H_2O) of 3 replicates among susceptible (Pish, *Oryza sativa* cv Nipponbare - P91-15B, *Magnaporthe oryzae* strain; ΔPish, *Oryza sativa* cv Nipponbare - Kyu77-07A, *Magnaporthe oryzae* strain) and resistant lines (Pia, Pish- P91-15B, Pish-Kyu77-07A).

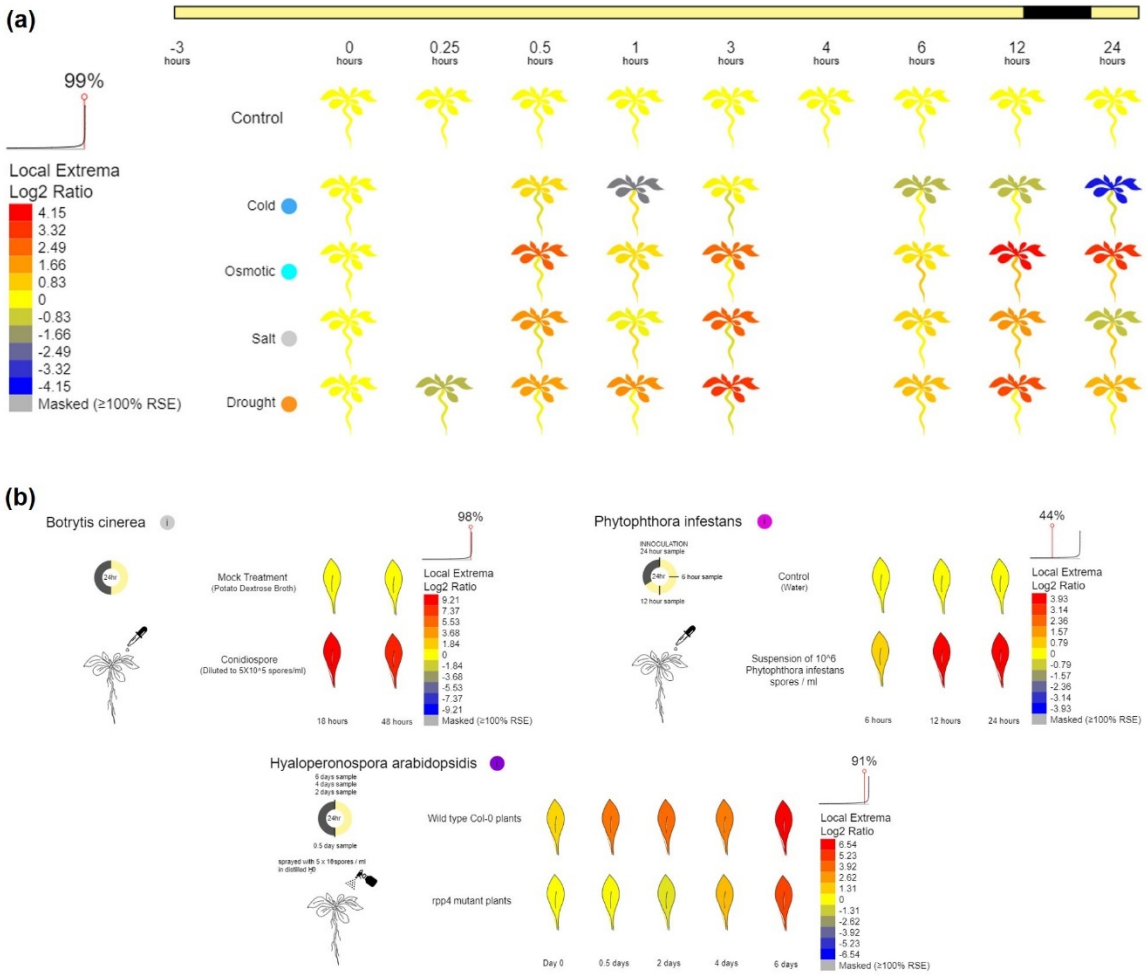


Fig. 9 Relative expression profile of a putative *A. thaliana* osmotin gene. As plotted in the bar scale, *red* and *blue* color indicate up- and downregulated genes (Log 2 ratio) under **(a)** abiotic stresses and **(b)** biotic stresses.

Supplementary material

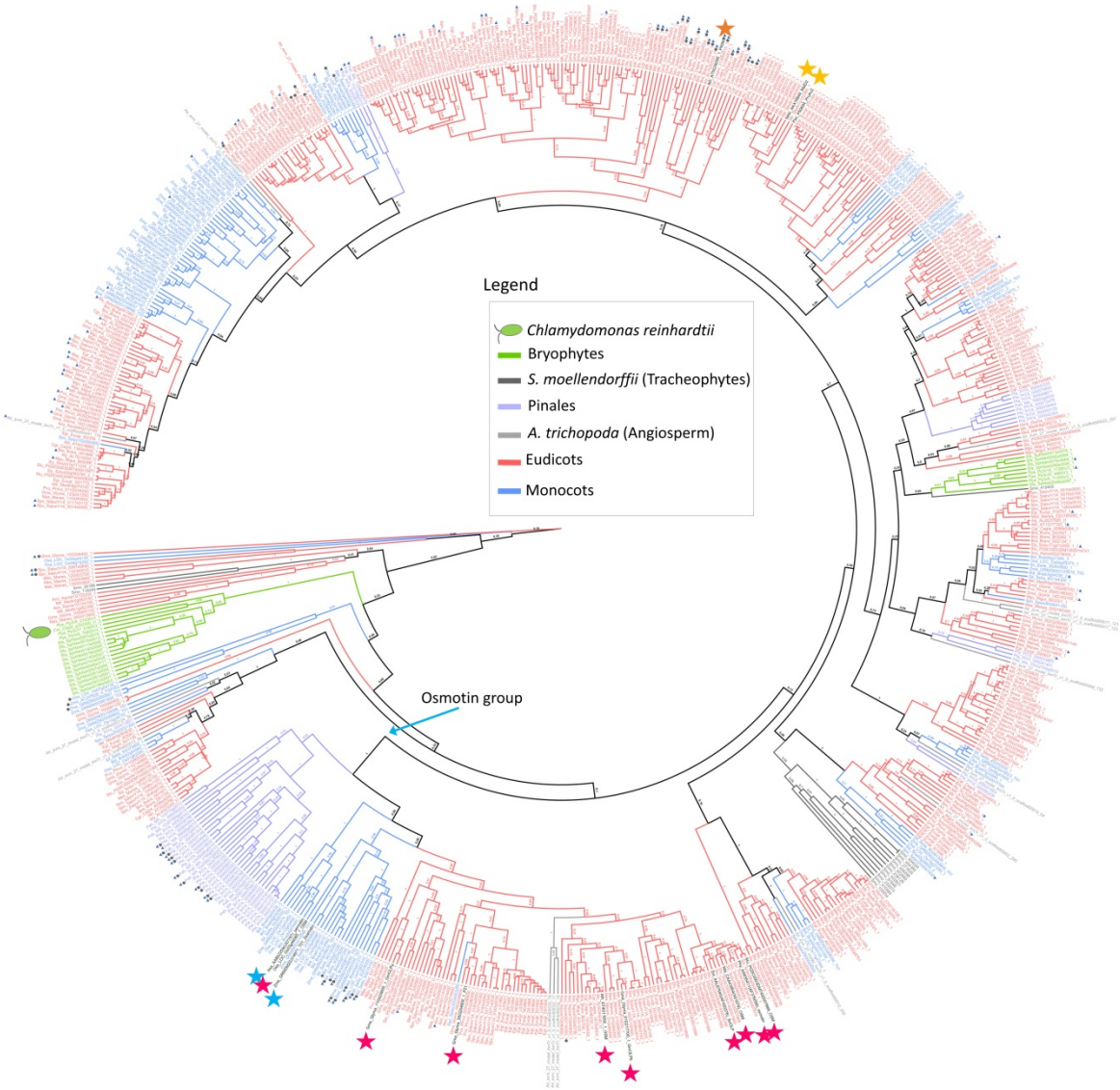


Fig. S1 Phylogenetic tree comprising 722 thaumatin domains from 32 plant species with labels and posterior probability values. Branch lengths are proportional to phylogenetic distances. *Pink*, *blue*, *yellow*, and *orange* stars indicate the previously characterized permatin, osmotin, allergenic, and kinase proteins, respectively. *Blue* arrow indicates the osmotin group. The small *blue* triangle and *dark grey* circle next to the labels indicate the presence of a transmembrane domain and different domains beyond THN, respectively.

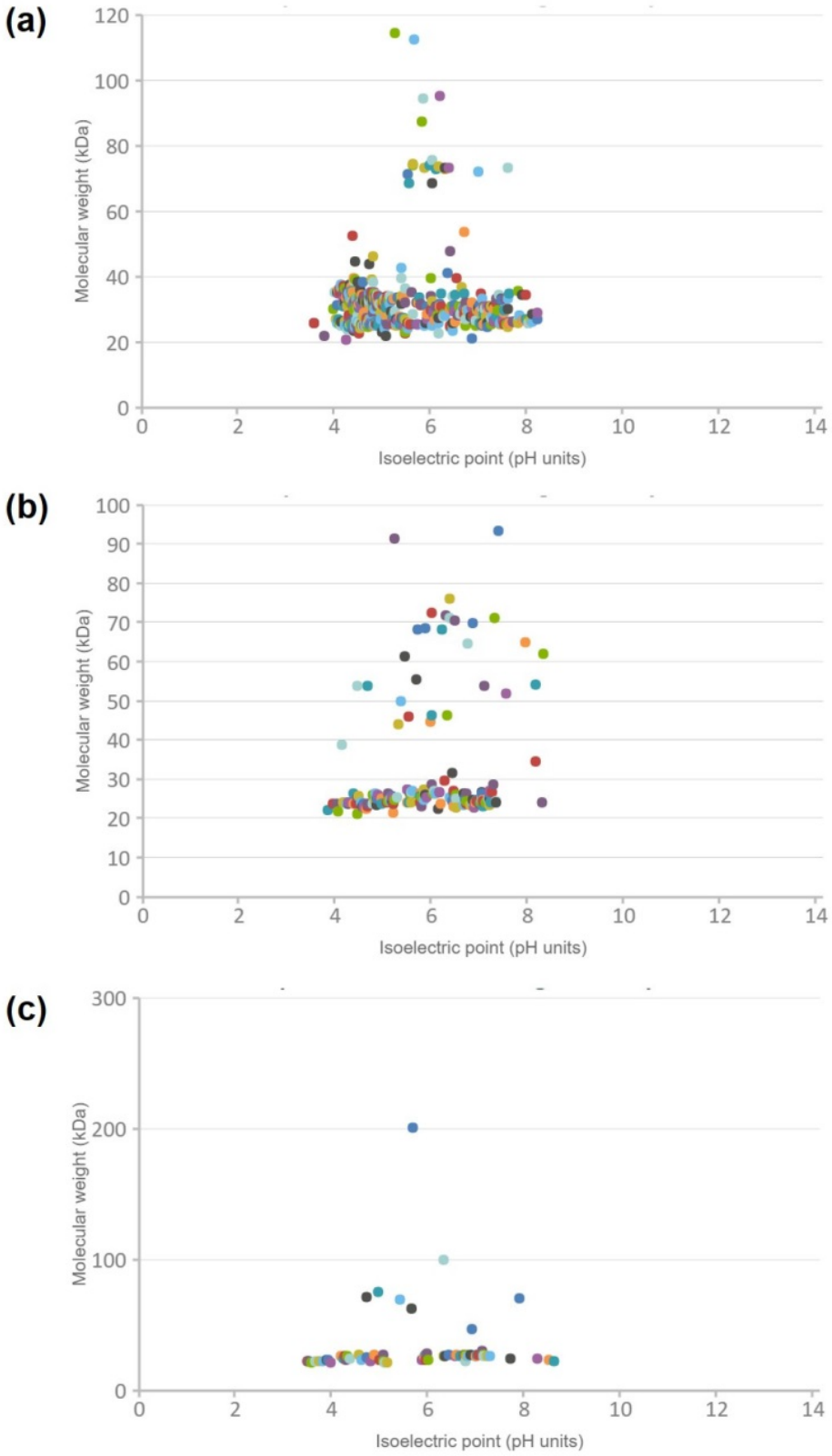
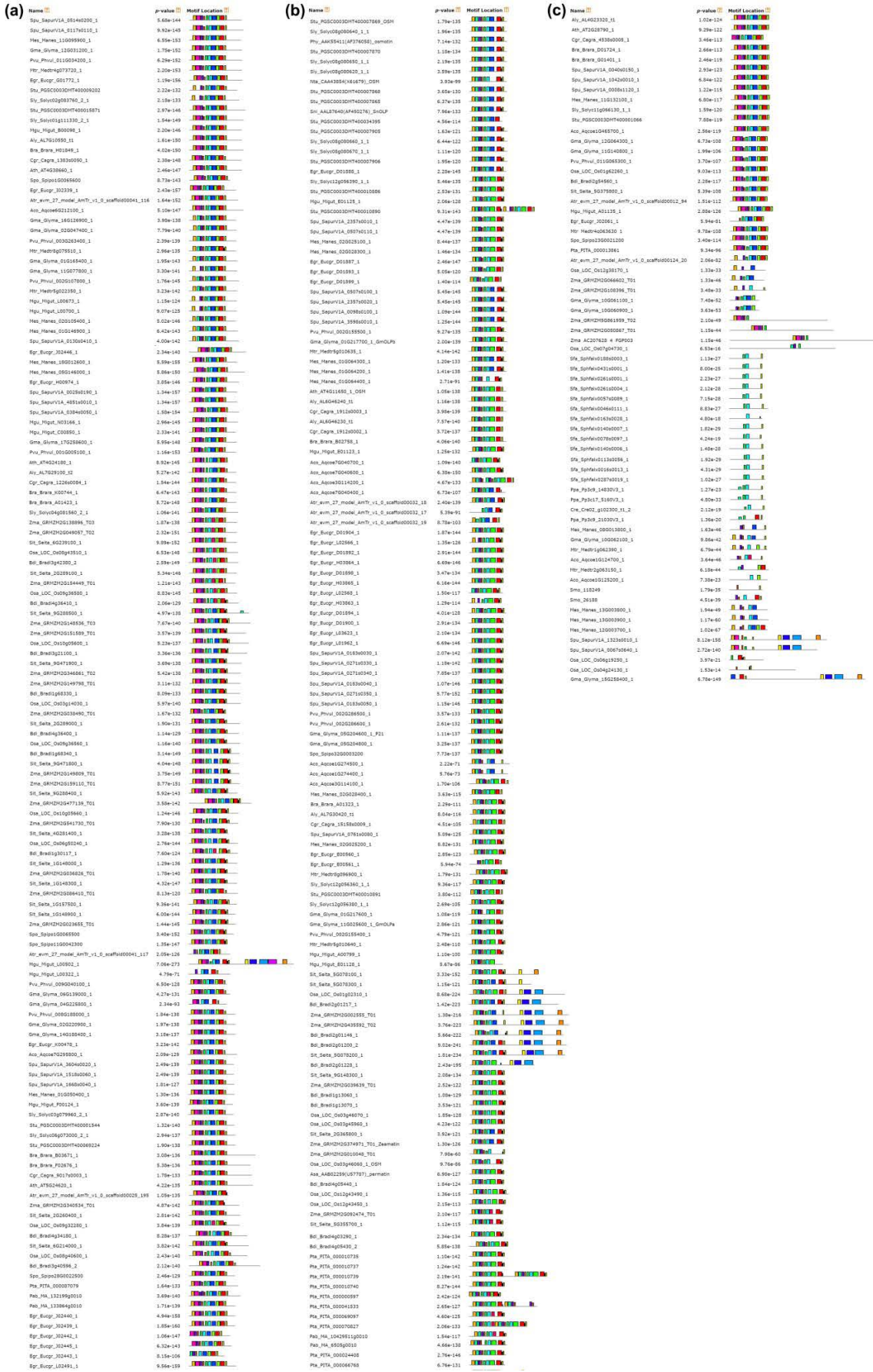
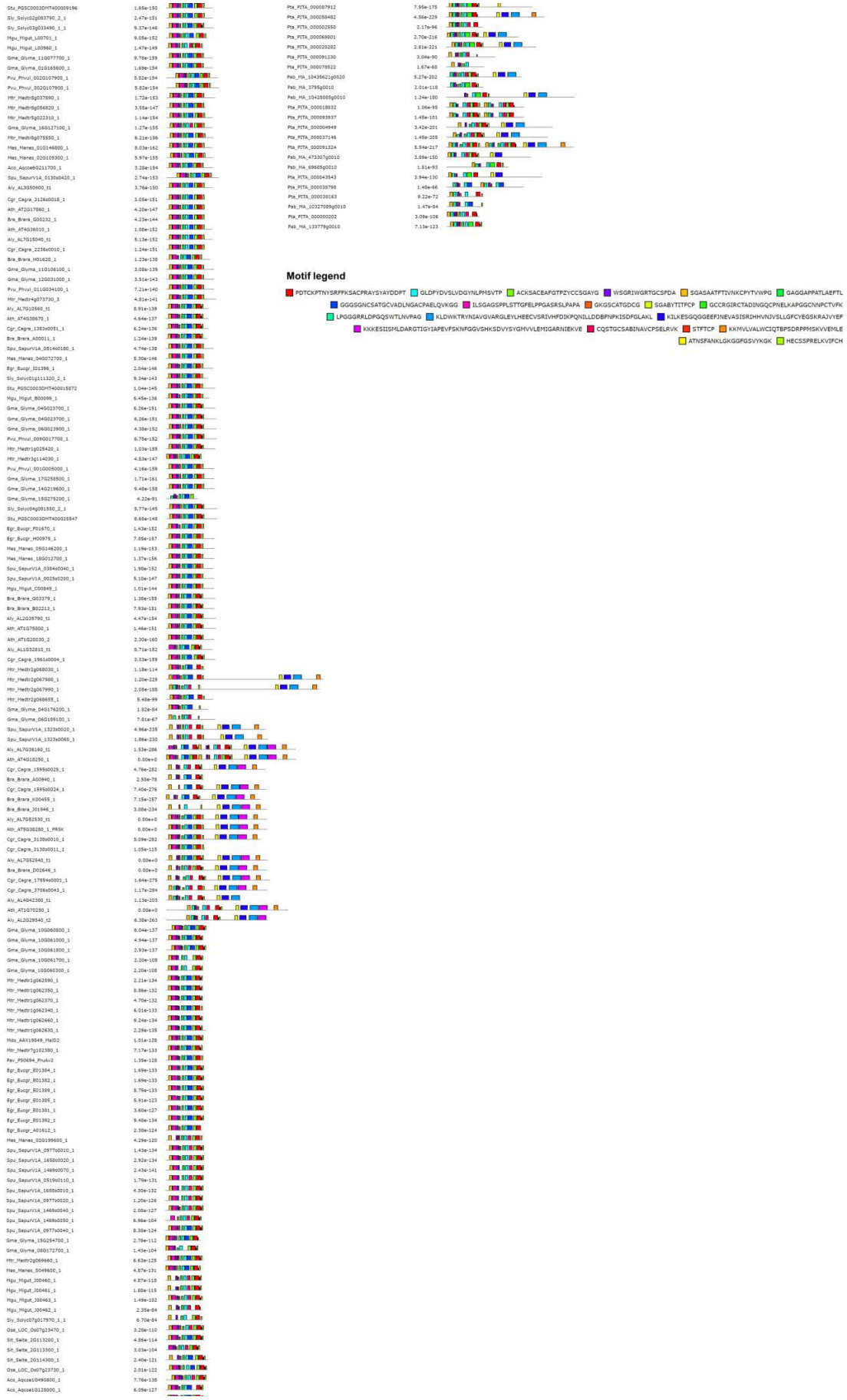
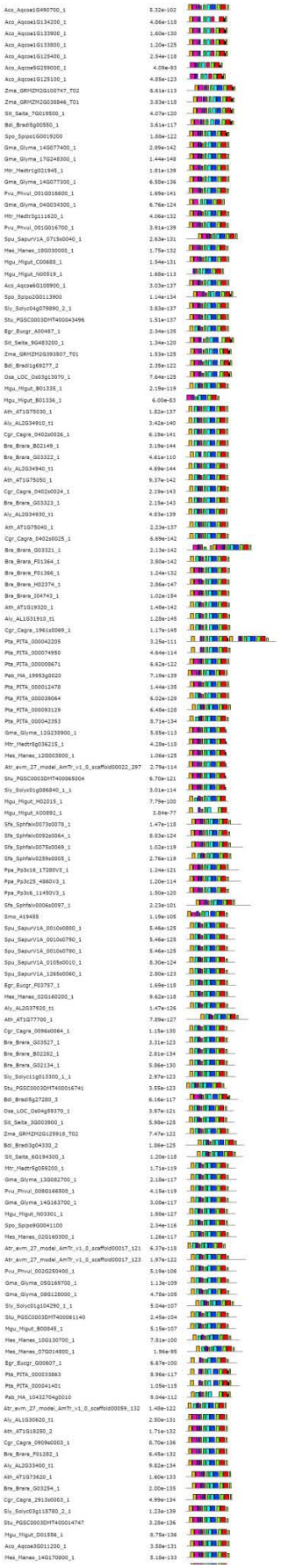


Fig. S2 Isoelectric points *vs* molecular weight for protein sequences **(a)** after osmotin group emergence, **(b)** for the osmotin group, and **(c)** before osmotin group emergence.


(a)
Name
p-value
Motif Location
Spu_SapurV1A_0514s0200_1 5.68e-144
Spu_SapurV1A_0117s0110_1 9.92e-145
Hes_Hanes_110095900_1 6.59e-153
Gma_Glyma_12G031200_1 1.75e-152
Pvu_Phvul_011G034200_1 6.29e-152
Htr_Hedtr4g073720_1 2.20e-153
Egr_Eucgr_G01772_1 1.19e-156
Stu_PGSC0003DHT400009202 2.22e-132
Sly_Solyc02g083760_2_1 2.19e-133
Stu_PGSC0003DHT400015871 2.97e-146
Sly_Solyc01g111330_2_1 1.94e-149
Mgu_Higut_B00098_1 2.20e-146
Aly_ALTG10950_t1 1.61e-150
Bra_Brara_H01849_1 4.02e-150
Cpr_Cagra_13836s0050_1 2.38e-148
Ath_AT4G38960_1 2.46e-147
Spo_Spipo15G0065600 8.73e-143
Egr_Eucgr_I02339_1 2.43e-157
Atr_evm_27_model_AmTr_v1_0_scaffold00041_116 1.64e-152
Aco_Aqcoe6G212100_1 5.10e-147
Gma_Glyma_16G126900_1 3.98e-138
Gma_Glyma_02G047400_1 7.79e-140
Pvu_Phvul_003G263400_1 2.39e-139
Htr_Hedtr8g079510_1 2.96e-135
Gma_Glyma_01G165400_1 1.95e-143
Gma_Glyma_11G077800_1 3.30e-141
Pvu_Phvul_002G107800_1 1.76e-145
Htr_Hedtr5g022350_1 3.23e-142
Mgu_Higut_L00673_1 1.15e-124
Mgu_Higut_L00700_1 9.07e-123
Hes_Hanes_02G105400_1 5.02e-146
Hes_Hanes_01G146500_1 6.42e-143
Spu_SapurV1A_0130s0410_1 4.00e-142
Egr_Eucgr_I02446_1 2.34e-140
Hes_Hanes_18G012600_1 5.99e-155
Hes_Hanes_05G146000_1 5.86e-150
Egr_Eucgr_H00974_1 3.05e-146
Spu_SapurV1A_0025s0190_1 1.34e-157
Spu_SapurV1A_4851s0010_1 1.34e-157
Spu_SapurV1A_0304s0050_1 1.58e-154
Mgu_Higut_N03166_1 2.96e-145
Mgu_Higut_C00850_1 2.33e-141
Gma_Glyma_17G258600_1 5.95e-148
Pvu_Phvul_001G005100_1 1.16e-153
Ath_AT4G24180_1 8.92e-145
Aly_ALTG29100_t2 5.27e-142
Cpr_Cagra_1226s0004_1 1.84e-144
Bra_Brara_k00744_1 6.47e-143
Bra_Brara_A21423_1 5.72e-148
Sly_Solyc04g081560_2_1 1.06e-141
Zma_GRMZM2G138896_T03 1.87e-138
Zma_GRMZM2G049057_T02 2.32e-151
Sit_Seita_6G239100_1 9.89e-152
Osa_LOC_Os06g43510_1 6.53e-148
Bdi_Brad3g42300_2 2.59e-149
Sit_Seita_2G289100_1 5.34e-146
Zma_GRMZM2G154449_T01 1.21e-143
Osa_LOC_Os06g36580_1 8.83e-145
Bdi_Brad4g36410_1 2.06e-129
Sit_Seita_9G288500_1 4.97e-138
Zma_GRMZM2G148536_T03 7.67e-140
Zma_GRMZM2G151689_T01 3.67e-139
Osa_LOC_Os10g05600_1 5.23e-137
Bdi_Brad3g21100_1 3.36e-136
Sit_Seita_9G471900_1 3.69e-138
Zma_GRMZM2G346861_T02 5.42e-138
Zma_GRMZM2G149798_T01 3.11e-132
Bdi_Brad1g68330_1 8.09e-133
Osa_LOC_Os03g14030_1 5.97e-140
Zma_GRMZM2G038490_T01 1.67e-132
Sit_Seita_2G289000_1 1.90e-131
Bdi_Brad4g36400_1 1.14e-129
Osa_LOC_Os06g36560_1 1.16e-140
Bdi_Brad1g68340_1 3.14e-149
Sit_Seita_9G471800_1 4.04e-148
Zma_GRMZM2G149809_T01 3.75e-149
Zma_GRMZM2G159110_T01 8.77e-151
Sit_Seita_9G288400_1 5.92e-143
Zma_GRMZM2G477139_T01 3.58e-142
Osa_LOC_Os10g05660_1 1.24e-146
Zma_GRMZM2G541730_T01 7.90e-130
Sit_Seita_4G381400_1 3.28e-138
Osa_LOC_Os06g50240_1 2.76e-144
Bdi_Brad1g30117_1 7.60e-124
Sit_Seita_1G148000_1 1.29e-136
Zma_GRMZM2G036826_T01 1.78e-140
Sit_Seita_1G148300_1 4.32e-147
Zma_GRMZM2G086410_T01 8.13e-120
Sit_Seita_1G157500_1 9.36e-141
Sit_Seita_1G148900_1 6.00e-144
Zma_GRMZM2G023655_T01 1.44e-145
Spo_Spipo1G0068500 3.40e-152
Spo_Spipo1G0042300 1.35e-147
Atr_evm_27_model_AmTr_v1_0_scaffold00041_117 2.05e-126
Mgu_Higut_L00502_1 7.06e-273
Mgu_Higut_L00322_1 4.79e-71
Pvu_Phvul_009G040100_1 6.50e-128
Gma_Glyma_06G139000_1 4.27e-131
Gma_Glyma_04G226800_1 2.34e-93
Pvu_Phvul_008G188000_1 1.84e-138
Gma_Glyma_02G220900_1 1.97e-138
Gma_Glyma_14G188400_1 3.18e-137
Egr_Eucgr_K00478_1 3.23e-142
Aco_Aqcoe7G295800_1 2.09e-129
Spu_SapurV1A_3604s0020_1 2.49e-139
Spu_SapurV1A_1518s0060_1 2.49e-139
Spu_SapurV1A_1660s0040_1 1.81e-127
Hes_Hanes_01G050400_1 1.30e-136
Mgu_Higut_F00124_1 3.60e-139
Sly_Solyc03g079960_2_1 2.07e-140
Stu_PGSC0003DHT400001944 1.32e-140
Sly_Solyc06g073000_2_1 2.94e-137
Stu_PGSC0003DHT400069224 1.90e-138
Bra_Brara_B03671_1 3.08e-136
Bra_Brara_F02676_1 5.38e-136
Cpr_Cagra_9017s0003_1 1.76e-133
Ath_ATEG24620_1 4.22e-135
Atr_evm_27_model_AmTr_v1_0_scaffold00025_195 1.05e-135
Zma_GRMZM2G340534_T01 4.87e-142
Sit_Seita_2G260400_1 2.81e-142
Osa_LOC_Os09g32280_1 3.04e-139
Bdi_Brad4g34180_1 8.28e-137
Sit_Seita_6G214000_1 3.82e-142
Osa_LOC_Os08g40600_1 2.43e-140
Bdi_Brad3g40596_2 2.12e-140
Spo_Spipo28G0022500 2.46e-129
Pta_PITA_000037079 1.64e-133
Pab_MA_1321999g0010 3.69e-140
Pab_MA_1338864g0010 1.71e-139
Egr_Eucgr_I02440_1 4.94e-158
Egr_Eucgr_I02439_1 1.85e-160
Egr_Eucgr_I02442_1 1.06e-147
Egr_Eucgr_I02445_1 6.32e-143
Egr_Eucgr_I02443_1 8.15e-106
Egr_Eucgr_L02491_1 9.56e-159
(b)
Name
p-value
Motif Location
Stu_PGSC0003DHT400007869_OSH 1.79e-135
Sly_Solyc08g080640_1_1 1.96e-135
Phy_AAK55411(AF376058)_osmotin 7.14e-132
Stu_PGSC0003DHT400007870 1.18e-134
Sly_Solyc08g080650_1_1 2.19e-135
Sly_Solyc08g080620_1_1 3.59e-138
Nta_CAA43854(X61679)_OSH 3.93e-99
Stu_PGSC0003DHT400007868 3.65e-130
Stu_PGSC0003DHT400007865 6.37e-135
Sni_AAL87640(AF450276)_SnOLP 7.96e-133
Stu_PGSC0003DHT400034395 4.56e-114
Stu_PGSC0003DHT400007905 1.63e-121
Sly_Solyc08g080660_1_1 6.44e-122
Sly_Solyc08g080670_1_1 1.11e-120
Stu_PGSC0003DHT400007906 1.95e-120
Egr_Eucgr_D01888_1 2.28e-145
Sly_Solyc12g096390_1_1 5.46e-139
Stu_PGSC0003DHT400010886 2.63e-131
Mgu_Higut_E01125_1 2.06e-128
Stu_PGSC0003DHT400010890 9.31e-143
Spu_SapurV1A_2357s0010_1 4.47e-139
Spu_SapurV1A_0507s0110_1 4.47e-139
Hes_Hanes_02G025100_1 8.44e-137
Hes_Hanes_02G028300_1 1.46e-134
Egr_Eucgr_D01887_1 2.46e-147
Egr_Eucgr_D01893_1 5.05e-120
Egr_Eucgr_D01899_1 1.40e-114
Spu_SapurV1A_0507s0100_1 5.45e-145
Spu_SapurV1A_2357s0020_1 5.45e-145
Spu_SapurV1A_0098s0100_1 1.09e-144
Spu_SapurV1A_3596s0010_1 1.26e-144
Pvu_Phvul_002G155500_1 9.27e-135
Gma_Glyma_01G217700_1_GmOLPs 2.00e-139
Htr_Hedtr5g010635_1 4.14e-142
Hes_Hanes_01G064300_1 1.20e-133
Hes_Hanes_01G064200_1 1.41e-138
Hes_Hanes_01G064400_1 2.71e-91
Ath_AT4G11650_1_OSH 1.05e-138
Aly_ALEG46240_t1 1.16e-138
Cpr_Cagra_1913s0003_1 3.99e-139
Aly_ALEG46230_t1 7.57e-140
Cpr_Cagra_1912s0002_1 3.72e-137
Bra_Brara_B02758_1 4.06e-140
Mgu_Higut_E01123_1 1.25e-132
Aco_Aqcoe7G046700_1 1.09e-140
Aco_Aqcoe7G040600_1 6.38e-130
Aco_Aqcoe3G114200_1 4.67e-133
Aco_Aqcoe7G040400_1 6.73e-107
Atr_evm_27_model_AmTr_v1_0_scaffold00032_18 2.40e-139
Atr_evm_27_model_AmTr_v1_0_scaffold00032_17 5.39e-91
Atr_evm_27_model_AmTr_v1_0_scaffold00032_19 8.78e-103
Egr_Eucgr_D01904_1 1.87e-144
Egr_Eucgr_L02566_1 1.35e-126
Egr_Eucgr_D01892_1 2.91e-144
Egr_Eucgr_H03864_1 6.69e-146
Egr_Eucgr_D01898_1 3.47e-134
Egr_Eucgr_H03865_1 6.16e-144
Egr_Eucgr_L02568_1 1.50e-117
Egr_Eucgr_H03863_1 1.29e-114
Egr_Eucgr_D01894_1 4.01e-128
Egr_Eucgr_D01900_1 2.91e-134
Egr_Eucgr_L03623_1 2.10e-134
Egr_Eucgr_L01962_1 6.69e-146
Spu_SapurV1A_0163s0030_1 2.07e-142
Spu_SapurV1A_0271s0330_1 1.18e-142
Spu_SapurV1A_0271s0340_1 7.05e-137
Spu_SapurV1A_0163s0040_1 1.07e-146
Spu_SapurV1A_0271s0350_1 5.77e-152
Spu_SapurV1A_0163s0050_1 1.15e-146
Pvu_Phvul_002G286500_1 3.57e-133
Pvu_Phvul_002G286600_1 2.61e-132
Gma_Glyma_09G204600_1_PZ1 1.11e-137
Gma_Glyma_09G204800_1 3.25e-137
Spo_Spipo32G0033200 7.73e-137
Aco_Aqcoe1G274500_1 2.22e-71
Aco_Aqcoe1G274400_1 9.76e-73
Aco_Aqcoe3G114100_1 1.70e-106
Hes_Hanes_02G028400_1 3.63e-115
Bra_Brara_A01323_1 2.29e-111
Aly_ALTG30420_t1 8.04e-116
Cpr_Cagra_1515s0009_1 4.51e-105
Spu_SapurV1A_0761s0080_1 5.09e-125
Hes_Hanes_02G025200_1 8.82e-131
Egr_Eucgr_E00560_1 2.65e-123
Egr_Eucgr_E00561_1 5.94e-74
Htr_Hedtr8g096900_1 1.79e-131
Sly_Solyc12g056360_1_1 9.36e-117
Stu_PGSC0003DHT400010891 3.80e-112
Sly_Solyc12g056380_1_1 2.69e-105
Gma_Glyma_01G217600_1 1.08e-119
Gma_Glyma_11G025600_1_GmOLPs 2.86e-121
Pvu_Phvul_002G155400_1 4.79e-121
Htr_Hedtr5g010640_1_1 2.48e-116
Mgu_Higut_A00799_1 1.10e-100
Mgu_Higut_E01128_1 9.67e-96
Sit_Seita_5G078100_1 3.31e-152
Sit_Seita_5G078300_1 1.15e-121
Osa_LOC_Os01g02310_1 8.68e-224
Bdi_Brad2g01217_1 1.42e-223
Zma_GRMZM2G002555_T01 1.38e-216
Zma_GRMZM2G435592_T02 3.76e-223
Bdi_Brad2g01148_1 3.66e-222
Bdi_Brad2g01200_2 9.02e-241
Sit_Seita_5G078200_1 1.81e-234
Bdi_Brad2g01228_1 2.43e-195
Sit_Seita_9G148300_1 2.08e-134
Zma_GRMZM2G039639_T01 2.52e-122
Bdi_Brad1g13060_1 1.08e-129
Bdi_Brad1g13070_1 3.53e-121
Osa_LOC_Os03g46070_1 1.65e-128
Osa_LOC_Os03g45960_1 4.23e-122
Sit_Seita_2G365800_1 3.92e-121
Zma_GRMZM2G374971_T01_Zeamatin 1.30e-126
Zma_GRMZM2G010048_T01 7.98e-60
Osa_LOC_Os03g46060_1_OSH 9.76e-86
Ase_AAB02259(U57797)_permatin 6.98e-127
Bdi_Brad4g05440_1 1.84e-124
Osa_LOC_Os12g43490_1 1.36e-115
Osa_LOC_Os12g43450_1 2.15e-113
Zma_GRMZM2G092474_T01 2.10e-117
Sit_Seita_9G355700_1 1.12e-115
Bdi_Brad4g03290_1 2.34e-134
Bdi_Brad4g05430_2 5.85e-138
Pta_PITA_000010735 1.10e-142
Pta_PITA_000010737 1.24e-142
Pta_PITA_000010739 2.19e-141
Pta_PITA_000010740 8.27e-144
Pta_PITA_000000597 2.42e-124
Pta_PITA_000041533 2.65e-127
Pta_PITA_000069097 4.60e-125
Pta_PITA_000070827 2.06e-133
Pab_MA_10429511g0010 1.54e-117
Pab_MA_6505g0010 4.66e-138
Pta_PITA_000024408 2.76e-146
Pta_PITA_000066766 6.76e-131
(c)
Name
p-value
Motif Location
Aly_ALEG23320_t1 1.02e-124
Ath_AT1G28790_1 9.29e-122
Cpr_Cagra_4930s0005_1 3.46e-113
Bra_Brara_D01724_1 2.66e-113
Bra_Brara_G01401_1 2.46e-119
Spu_SapurV1A_0040s0150_1 2.93e-123
Spu_SapurV1A_1042s0010_1 6.84e-122
Spu_SapurV1A_0008s1120_1 1.22e-115
Hes_Hanes_11G132100_1 6.80e-117
Sly_Solyc11g066130_1_1 1.59e-120
Stu_PGSC0003DHT400001066 7.68e-119
Aco_Aqcoe1G465700_1 2.56e-119
Gma_Glyma_12G064300_1 6.73e-108
Gma_Glyma_11G140800_1 1.99e-106
Pvu_Phvul_011G063300_1 3.70e-107
Osa_LOC_Os01g62260_1 9.03e-113
Bdi_Brad2g54960_1 2.28e-117
Sit_Seita_5G375800_1 5.39e-108
Atr_evm_27_model_AmTr_v1_0_scaffold00012_94 1.91e-112
Mgu_Higut_A01125_1 2.08e-126
Egr_Eucgr_J02061_1 5.94e-81
Htr_Hedtr4g063630_1 9.78e-108
Spo_Spipo23G0021200 3.40e-114
Pta_PITA_000013861 9.34e-96
Atr_evm_27_model_AmTr_v1_0_scaffold00124_20 2.06e-82
Osa_LOC_Os12g38170_1 1.33e-33
Zma_GRMZM2G066602_T01 1.33e-46
Zma_GRMZM2G108396_T01 3.48e-33
Gma_Glyma_10G061100_1 7.40e-52
Gma_Glyma_10G060900_1 3.63e-53
Zma_GRMZM5G861959_T02 2.10e-49
Zma_GRMZM2G085067_T01 1.15e-44
Zma_AC207628_4_FGP003 1.15e-46
Osa_LOC_Os07g04730_1 6.53e-16
Sfa_Sohfaliv0188s0003_1 1.13e-27
Sfa_Sohfaliv0431s0001_1 8.00e-25
Sfa_Sohfaliv0261s0001_1 2.23e-27
Sfa_Sohfaliv0261s0004_1 2.12e-28
Sfa_Sohfaliv0057s0089_1 7.15e-28
Sfa_Sohfaliv0046s0111_1 8.83e-27
Sfa_Sohfaliv0163s0028_1 4.80e-18
Sfa_Sohfaliv0148s0007_1 1.82e-29
Sfa_Sohfaliv0078s0097_1 4.24e-19
Sfa_Sohfaliv0148s0006_1 1.40e-28
Sfa_Sohfaliv0113s0058_1 1.92e-29
Sfa_Sohfaliv0016s0013_1 4.31e-29
Sfa_Sohfaliv0287s0019_1 1.02e-27
Ppa_Pp3c9_14830V3_1 1.27e-23
Ppa_Pp3c17_5160V3_1 4.80e-33
Cre_Cre02_g102300_t1_2 2.12e-19
Ppa_Pp3c9_21030V3_1 1.36e-20
Hes_Hanes_08G013800_1 1.63e-46
Gma_Glyma_10G062100_1 9.86e-42
Htr_Hedtr1g062390_1 6.79e-44
Aco_Aqcoe1G124700_1 3.64e-46
Htr_Hedtr2g063150_1 6.18e-44
Aco_Aqcoe1G125200_1 7.38e-23
Smo_118249 1.79e-35
Smo_26158 4.51e-39
Hes_Hanes_13G003900_1 1.94e-49
Hes_Hanes_13G003960_1 1.17e-60
Hes_Hanes_12G003700_1 1.02e-67
Spu_SapurV1A_1323s0010_1 8.12e-158
Spu_SapurV1A_0067s0640_1 2.72e-140
Osa_LOC_Os06g19250_1 3.97e-21
Osa_LOC_Os04g24130_1 1.53e-14
Gma_Glyma_15G258400_1 6.78e-149


Motif legend
PDTCKPTNYSRFPKSACPRAYSYAYDDPT
GLDFYDVSLVDGYNLPMSVTP
ACKSACEAFGTPZYCCSGAYG
WSGRIWGRTGCSFDA
SGASAATFTIVNKCPYTVWPG
GAGGAPPATLAEFTL
GGGSGNCSATGCVADLNGACPAELQVKGG
TLSGAGSPPLSTTGFELPPGASRSLPAPA
GKGSCATGDCG
SGABYTTTFCP
GCCRGIRCTADINGQCPNELKAPGGCNNPCTVFK
LPGGGRRLDPGQSWTLNVPAG
KLDWKTRYNIAVGVARGLEYLHEECVSRIVHFDIKPQNILLDDBFNPKISDFGLAKL
KILKESGQGGEEFJNEVASISRIHHVNJVSLLGFCYEGSKRAJVYEF
KKKESIISMLDARGTIGYIAPEVFSKNFGGVSHKSDVYSYGMVVLEMIGARNIEKVE
CQSTGCSABINAVCPSELRVK
STTFCP
KKMVLVALWCIQTBPSDRPPMSKVVEMLE
ATNSFANKLGKGGPGSVYKGK
HECSSPRELKVIFCH


Aco_Aqcoe1G490700_1 5.32e-102
Aco_Aqcoe1G134200_1 4.56e-118
Aco_Aqcoe1G133900_1 1.60e-130
Aco_Aqcoe1G133800_1 1.20e-125
Aco_Aqcoe1G125400_1 2.54e-118
Aco_Aqcoe5G259000_1 4.09e-93
Aco_Aqcoe1G125100_1 4.85e-123
Zma_GRMZM2G100747_T02 6.61e-113
Zma_GRMZM2G038846_T01 3.03e-118
Sit_Seita_7G019500_1 4.07e-120
Bdi_Bradi5g00550_1 3.61e-117
Spo_Spipo10G0019200 1.80e-122
Gma_Glyma_14G077400_1 2.89e-142
Gma_Glyma_17G248300_1 1.44e-148
Mtr_Medtr1g021945_1 1.81e-139
Gma_Glyma_14G077300_1 6.50e-136
Pvu_Phvul_001G016600_1 1.69e-141
Gma_Glyma_04G034300_1 6.76e-124
Mtr_Medtr3g111620_1 4.06e-132
Pvu_Phvul_001G016700_1 3.91e-139
Spu_SapurV1A_0715s0040_1 2.63e-131
Mes_Manes_18G030000_1 1.75e-132
Mgu_Higut_C00688_1 1.54e-131
Mgu_Higut_N00519_1 1.68e-113
Aco_Aqcoe6G108900_1 3.03e-137
Spo_Spipo2G0013900 1.14e-134
Sly_Solyc04g079890_2_1 3.83e-137
Stu_PGSC0003DHT400043496 1.51e-137
Egr_Eucgr_A00487_1 2.34e-135
Sit_Seita_9G483200_1 1.34e-120
Zma_GRMZM2G393507_T01 1.53e-125
Bdi_Bradi1g69277_2 2.35e-122
Osa_LOC_Os03g13070_1 7.64e-125
Mgu_Higut_B01335_1 2.19e-119
Mgu_Higut_B01336_1 6.00e-63
Ath_AT1G75030_1 1.82e-137
Aly_AL2G34910_t1 3.42e-140
Cgr_Cagra_0402s0026_1 6.19e-141
Bra_Brara_B02149_1 3.19e-144
Bra_Brara_G03322_1 4.61e-110
Aly_AL2G34940_t1 4.69e-144
Ath_AT1G75050_1 9.37e-142
Cgr_Cagra_0402s0024_1 2.19e-143
Bra_Brara_G03323_1 2.19e-143
Aly_AL2G34930_t1 4.83e-139
Ath_AT1G75040_1 2.23e-137
Cgr_Cagra_0402s0025_1 6.69e-142
Bra_Brara_G03321_1 2.13e-142
Bra_Brara_F01364_1 3.60e-142
Bra_Brara_F01366_1 1.24e-132
Bra_Brara_H02374_1 2.86e-147
Bra_Brara_I04743_1 1.02e-154
Ath_AT1G19320_1 1.48e-142
Aly_AL1G31910_t1 1.26e-145
Cgr_Cagra_1961s0069_1 1.17e-145
Pta_PITA_000042205 3.25e-111
Pta_PITA_000074950 4.64e-114
Pta_PITA_000005671 6.62e-122
Pab_MA_19953g0020 7.19e-139
Pta_PITA_000012478 1.44e-135
Pta_PITA_000039084 6.02e-129
Pta_PITA_000093129 6.48e-128
Pta_PITA_000042353 8.71e-134
Gma_Glyma_12G238900_1 5.85e-113
Mtr_Medtr3g036215_1 4.28e-118
Mes_Manes_12G003800_1 1.06e-125
Atr_evm_27_model_AmTr_v1_0_scaffold00022_297 2.79e-114
Stu_PGSC0003DHT400065804 6.70e-121
Sly_Solyc01g086840_1_1 3.01e-114
Mgu_Higut_H02015_1 7.79e-100
Mgu_Higut_K00892_1 3.84e-77
Sfa_Sphfalx0073s0078_1 1.47e-118
Sfa_Sphfalx0092s0064_1 8.63e-124
Sfa_Sphfalx0075x0069_1 1.02e-119
Sfa_Sphfalx0259s0005_1 2.76e-119
Ppa_Pp3s16_17280V3_1 1.24e-121
Ppa_Pp3s25_4860V3_1 1.20e-114
Ppa_Pp3s6_11450V3_1 1.50e-120
Sfa_Sphfalx0006s0097_1 2.23e-101
Smo_419455 1.19e-105
Spu_SapurV1A_0010s0800_1 5.46e-125
Spu_SapurV1A_0010s0790_1 5.46e-125
Spu_SapurV1A_0010s0780_1 5.46e-125
Spu_SapurV1A_0105s0010_1 8.30e-124
Spu_SapurV1A_1265s0040_1 2.80e-123
Egr_Eucgr_F03757_1 1.49e-118
Mes_Manes_02G160200_1 9.62e-118
Aly_AL2G37920_t1 1.47e-126
Ath_AT1G77700_1 7.89e-127
Cgr_Cagra_0096s0364_1 1.15e-130
Bra_Brara_G03527_1 3.31e-123
Bra_Brara_B02282_1 2.81e-134
Bra_Brara_G02134_1 5.66e-130
Sly_Solyc11g013300_1_1 2.97e-123
Stu_PGSC0003DHT400016741 3.95e-123
Bdi_Bradi5g27280_3 6.16e-117
Osa_LOC_Os04g59370_1 3.97e-121
Sit_Seita_3G003900_1 5.90e-125
Zma_GRMZM2G125918_T02 7.47e-122
Bdi_Bradi3g04330_2 1.84e-125
Sit_Seita_6G194300_1 1.20e-118
Mtr_Medtr5g059200_1 1.71e-119
Gma_Glyma_13G082700_1 2.18e-117
Pvu_Phvul_008G166500_1 4.15e-119
Gma_Glyma_14G163700_1 3.08e-117
Mgu_Higut_N03301_1 1.88e-127
Spo_Spipo9G0041100 2.34e-116
Mes_Manes_02G160300_1 1.26e-117
Atr_evm_27_model_AmTr_v1_0_scaffold00017_121 6.37e-118
Atr_evm_27_model_AmTr_v1_0_scaffold00017_123 1.97e-122
Pvu_Phvul_002G250400_1 5.19e-106
Gma_Glyma_05G169700_1 1.13e-109
Gma_Glyma_08G128000_1 4.78e-105
Sly_Solyc01g104290_1_1 5.04e-107
Stu_PGSC0003DHT400061140 2.45e-104
Mgu_Higut_B00945_1 5.15e-107
Mes_Manes_10G130700_1 7.81e-100
Mes_Manes_07G014800_1 1.86e-95
Egr_Eucgr_G00607_1 6.67e-100
Pta_PITA_000033863 8.96e-117
Pta_PITA_000041401 1.05e-115
Pab_MA_10432704g0010 9.04e-112
Atr_evm_27_model_AmTr_v1_0_scaffold00099_132 1.48e-122
Aly_AL1G30620_t1 2.50e-131
Ath_AT1G18250_2 1.71e-132
Cgr_Cagra_0909s0003_1 8.70e-136
Bra_Brara_F01282_1 6.45e-132
Aly_AL2G33400_t1 9.82e-134
Ath_AT1G73620_1 1.60e-133
Bra_Brara_G03294_1 1.00e-135
Cgr_Cagra_2913s0003_1 4.99e-134
Sly_Solyc03g118780_2_1 1.23e-139
Stu_PGSC0003DHT400014747 3.28e-136
Mgu_Higut_D01556_1 8.75e-136
Aco_Aqcoe3G011200_1 3.58e-131
Mes_Manes_14G170800_1 5.18e-133

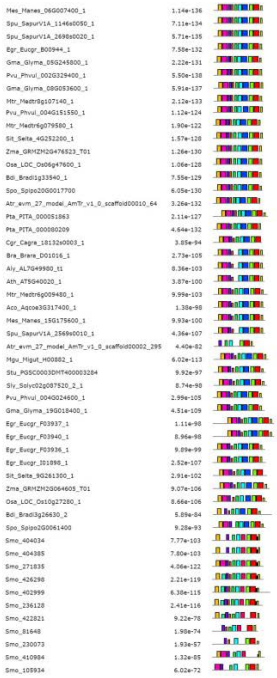


Fig. S3 Architecture of conserved protein motifs in protein sequences localized **(a)** after osmotin group emergence, **(b)** for the osmotin group, and **(c)** before osmotin group emergence according to the ordering of the phylogenetic tree. Each motif is represented by a colored block. The lengths and positions of the blocks correspond to the lengths and positions of motifs in the individual protein sequences. The height of each block is proportional to its −log (*p* value), truncated at the height corresponding to a motif with a *p* value of 1e −10. The gene names and combined *p* values are shown on the left side of the figure.

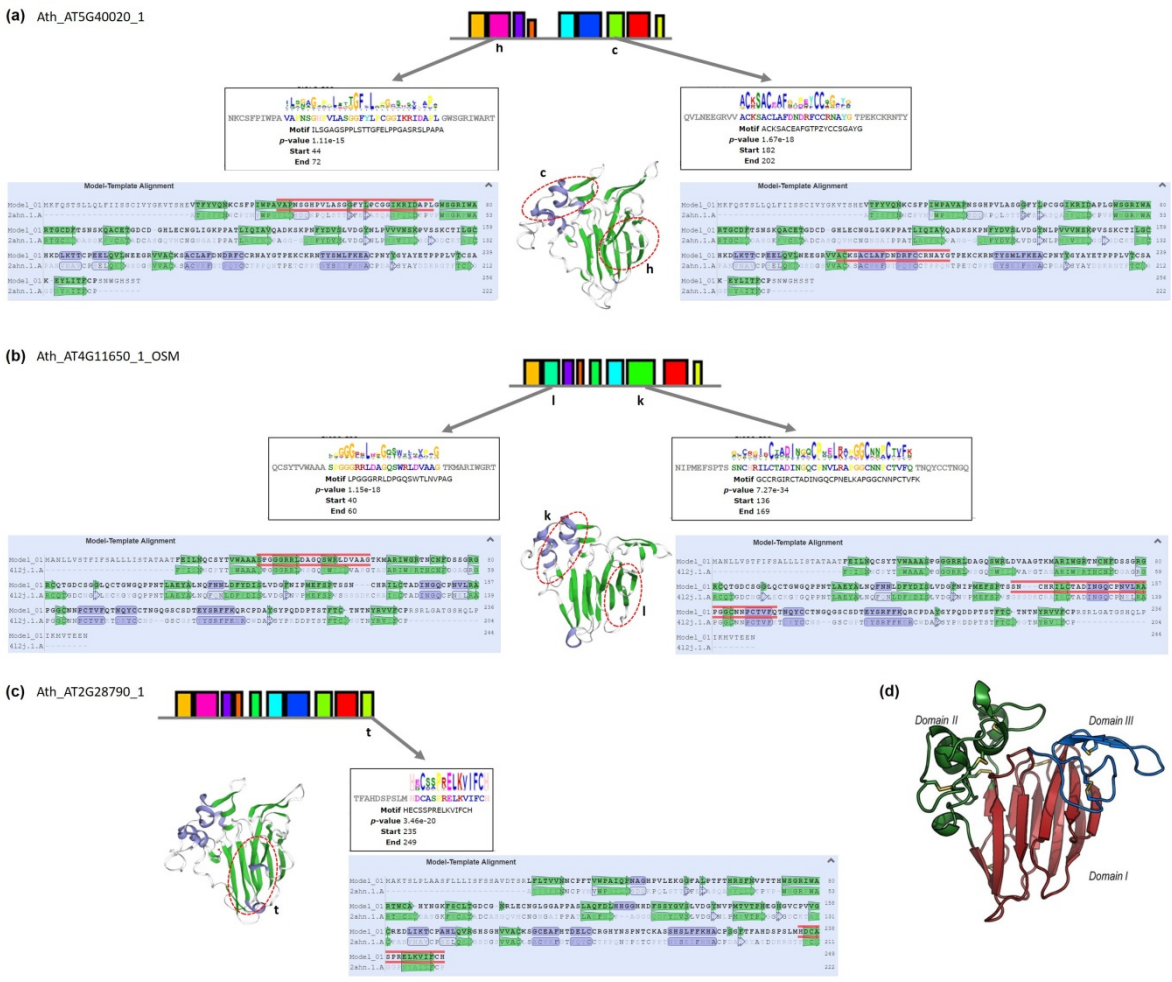


Fig. S4 Motif localization in homologous modeling of Arabidopsis protein for **(a)** non-putative osmotin localized after osmotin group emergence, **(b)** osmotin, and **(c)** non-putative osmotin localized before osmotin group emergence. The *green*, *purple*, and *red* background in the sequence model-template alignment represent the stranded β-sheet, α-helix, and motif sequences, respectively. The *red* circles in the 3D proteins represent the region of each motif. **(d)** Domain localization in TLP 3D protein.

Table S1 Summary of all plant, gene, and protein sequences employed in this study.

Classification	Family	Specie	Taxa terminologies	Protein ID	Score	E-value	Length(aa)	No. Introns (within CDS)	Database
Chlorophyta	Chlamydomonadaceae	*Chlamydomonas reinhardtii*	Cre	Cre02.g102300.t1.2	60,8	4,40E-11	226	6	Phytozome
Bryophyta	Funariaceae	*Physcomitrella patens*	Ppa	Pp3c16_17280V3.1	171	2,00E-50	303	5	Phytozome
				Pp3c17_5160V3.1	82,8	2,60E-18	232	2	Phytozome
				Pp3c25_4860V3.1	171,8	1,60E-50	313	5	Phytozome
				Pp3c6_11450V3.1	172,2	2,10E-51	248	3	Phytozome
				Pp3c7_8680V3.1*	48,5	1,30E-06	138	1	Phytozome
				Pp3c9_14830V3.1	45,8	1,90E-05	216	2	Phytozome
				Pp3c9_21030V3.1	47	7,20E-06	214	2	Phytozome
Bryophyta	Sphagnaceae	*Sphagnum fallax*	Sfa	Sfxfak0006s0097.1	139,8	2,10E-38	370	3	Phytozome
				Sfxfak0016s0013.1	63,2	6,90E-12	222	2	Phytozome
				Sfxfak0046s0111.1	59,3	2,40E-10	254	2	Phytozome
				Sfxfak0057s0089.1	74,3	5,30E-16	213	2	Phytozome
				Sfxfak0073s0078.1	167,2	3,80E-49	317	4	Phytozome
				Sfxfak0075s0069.1	175,3	3,30E-52	329	5	Phytozome
				Sfxfak0078s0097.1	53,1	2,10E-08	211	2	Phytozome
				Sfxfak0092s0064.1	164,1	1,20E-48	255	2	Phytozome
				Sfxfak0113s0056.1	63,9	3,10E-12	222	2	Phytozome
				Sfxfak0140s0006.1	67,4	1,90E-13	222	2	Phytozome
				Sfxfak0140s0007.1	77	6,70E-17	222	2	Phytozome
				Sfxfak0163s0028.1	47,4	2,60E-06	217	2	Phytozome
				Sfxfak0188s0002.1*	58,9	5,60E-11	136	1	Phytozome
				Sfxfak0188s0003.1	66,2	4,10E-13	220	2	Phytozome
				Sfxfak0259s0005.1	161	2,10E-47	261	2	Phytozome
				Sfxfak0261s0001.1	65,9	5,30E-13	217	2	Phytozome
				Sfxfak0261s0004.1	71,2	7,80E-15	217	2	Phytozome
				Sfxfak0287s0019.1	67,8	1,60E-13	231	2	Phytozome
				Sfxfak0431s0001.1	66,2	3,30E-13	215	2	Phytozome
				Sfxfak0431s0002.1*	34,7	2,00E-02	134	0	Phytozome
				Sfxfak0431s0003.1*	59,3	4,70E-11	137	1	Phytozome
Lycopodiophyta	Selaginellaceae	*Selaginella moellendorffii*	Smo	105934	117,1	3,70E-31	249	0	Phytozome
				118249	96,7	5,80E-24	218	0	Phytozome
				118611*	98,6	4,90E-25	167	0	Phytozome
				123421*	96,3	2,00E-24	145	0	Phytozome
				230073	97,8	3,60E-24	237	2	Phytozome
				236128	154,5	3,50E-45	257	1	Phytozome
				26188	99,8	4,30E-25	212	0	Phytozome
				271835	152,5	2,00E-44	260	1	Phytozome
				35487*	73,9	9,90E-17	116	0	Phytozome
				402999	160,6	3,50E-47	292	1	Phytozome
				404034	131,3	1,50E-36	242	2	Phytozome
				404385	131,3	1,50E-36	242	2	Phytozome
				410984	125,6	3,10E-34	264	1	Phytozome
				419455	163,3	9,70E-49	243	1	Phytozome
				422821	120,6	1,40E-32	241	1	Phytozome
				422825*	86,7	3,40E-20	249	7	Phytozome
				426298	162,5	3,50E-48	260	2	Phytozome
				81648	128,3	1,50E-35	230	0	Phytozome

Pinophyta (Gymnosperm)	Pinaceae	*Picea abies*	Pab	MA_6505g0010	253,06	3,51E-83	247	1	Congenie
				MA_189802g0010*	247,67	1,16E-81	206	0	Congenie
				MA_8772866g0010*	217,24	1,45E-69	220	1	Congenie
				MA_152803g0010*	213,77	1,96E-68	207	0	Congenie
				MA_133779g0010	213,39	5,18E-68	232	1	Congenie
				MA_10429511g0010	206,84	2,86E-65	245	1	Congenie
				MA_3795g0010	196,44	3,43E-61	242	2	Congenie
				MA_10435621g0020	185,65	1,65E-54	499	7	Congenie
				MA_10433458g0020*	173,71	2,76E-53	174	1	Congenie
				MA_473307g0010	180,26	4,38E-52	553	8	Congenie
				MA_19953g0020	169,09	9,18E-51	233	1	Congenie
				MA_132199g0010	165,62	2,30E-48	337	2	Congenie
				MA_10428085g0010	172,17	2,34E-48	884	13	Congenie
				MA_133864g0010	163,31	9,00E-48	302	2	Congenie
				MA_10327089g0010	160,23	2,52E-47	237	1	Congenie
				MA_10434201g0010*	158,3	5,34E-47	199	2	Congenie
				MA_5112537g0010*	157,15	5,72E-47	164	0	Congenie
				MA_10432704g0010	150,21	2,96E-43	262	1	Congenie
				MA_69685g0010	152,53	8,18E-43	406	2	Congenie
				MA_10484g0010*	138,66	4,40E-40	145	1	Congenie
				MA_4320679g0010*	139,81	4,56E-40	193	0	Congenie
Pinophyta (Gymnosperm)	Pinaceae	*Pinus taeda*	Pta	PITA_000024408	263,08	1,98E-87	242	1	Congenie
				PITA_000010737	248,05	1,14E-81	234	1	Congenie
				PITA_000010740	246,51	4,50E-81	234	1	Congenie
				PITA_000010735	244,2	3,30E-80	233	1	Congenie
				PITA_000020282	250,75	1,47E-78	594	7	Congenie
				PITA_000066768	238,42	7,18E-78	235	1	Congenie
				PITA_000010739	238,23	2,76E-73	510	3	Congenie
				PITA_000087912	238,04	7,06E-74	567	6	Congenie
				PITA_000070827	197,02	4,53E-47	379	2	Congenie
				PITA_000069801	232,65	1,03E-72	475	5	Congenie
				PITA_000058482	236,11	1,50E-72	644	7	Congenie
				PITA_000069097	219,55	1,48E-70	234	1	Congenie
				PITA_000000597	218,78	1,55E-70	208	1	Congenie
				PITA_000041533	135,45	3,90E-10	444	4	Congenie
				PITA_000002550	206,45	2,20E-65	240	1	Congenie
				PITA_000000202	197,98	1,89E-62	214	2	Congenie
				PITA_000091230	199,9	7,81E-62	324	2	Congenie
				PITA_000089669*	178,33	6,70E-55	202	1	Congenie
				PITA_000078522	178,33	2,91E-54	252	1	Congenie
				PITA_000091324	180,84	2,06E-51	838	6	Congenie
				PITA_000010734*	169,47	3,37E-52	151	2	Congenie
				PITA_000037146	178,72	2,07E-51	667	8	Congenie
				PITA_000033863	167,55	9,34E-50	286	1	Congenie
				PITA_000081579*	163,7	1,97E-49	184	1	Congenie
				PITA_000051863	164,47	1,27E-48	277	1	Congenie
				PITA_000087079	164,47	3,38E-48	322	2	Congenie
				PITA_000093937	166,2	2,84E-47	510	2	Congenie
				PITA_000080209	159,07	4,85E-47	244	1	Congenie
				PITA_000012478	158,3	1,20E-46	246	1	Congenie
				PITA_000041401	159,46	1,53E-46	302	2	Congenie

				PITA_000018832	104,93	3,23E-40	308	2	Conifers
				PITA_000042353	155,61	9,64E-46	240	1	Conifers
				PITA_000038163	154,45	2,63E-45	237	1	Conifers
				PITA_000093129	152,91	1,44E-44	251	1	Conifers
				PITA_000021487*	149,06	6,72E-44	184	1	Conifers
				PITA_000074950	150,98	7,37E-44	252	1	Conifers
				PITA_000066987*	147,52	3,15E-43	184	1	Conifers
				PITA_000008671	146,75	3,10E-42	248	1	Conifers
				PITA_000042205	148,87	5,33E-41	511	5	Conifers
				PITA_000039064	144,82	1,84E-41	247	1	Conifers
				PITA_000033081*	142,51	2,15E-41	184	1	Conifers
				PITA_000018739*	141,74	4,57E-41	184	1	Conifers
				PITA_000043543	147,52	1,93E-40	630	9	Conifers
				PITA_000044167*	139,04	1,94E-40	148	2	Conifers
				PITA_000016471*	139,04	3,71E-40	177	1	Conifers
				PITA_000004949	145,59	1,68E-39	698	8	Conifers
				PITA_000038798	137,12	1,24E-35	508	3	Conifers
				PITA_000058143*	131,34	3,29E-37	169	1	Conifers
				PITA_000013861	119,01	8,57E-32	252	1	Conifers
Angiosperm (dicotyledons; most	Amborellaceae	*Amborella trichopoda*	Atr	evm_27.model.AmTr_v1.0_scaffold00002.295	108,6	2,60E-28	209	1	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00010.64	159,1	3,30E-47	239	1	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00012.258	142,9	6,30E-41	248	1	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00012.94	124,4	5,70E-34	246	0	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00017.121	170,2	1,20E-51	224	1	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00017.123	178,7	1,40E-53	340	2	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00022.297	160,6	8,40E-48	236	3	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00025.195	159,5	4,70E-47	259	2	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00032.17	171,4	3,80E-52	217	0	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00032.18	263,5	7,60E-88	231	0	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00032.19	178,7	3,00E-55	199	1	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00041.116	157,5	1,10E-45	319	2	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00041.117	133,7	3,30E-37	272	1	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00051.103*	72,8	2,20E-16	112	0	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00059.132	155,6	6,60E-46	234	2	Phytozome
				evm_27.model.AmTr_v1.0_scaffold00124.20	101,7	1,70E-25	248	0	Phytozome
Angiosperm (mocot)	Araceae	*Spirodela polyrhiza*	Spo	Spipo11G0042300	143,3	6,50E-41	258	1	Phytozome
				Spipo1G0019200	142,1	1,20E-40	245	1	Phytozome
				Spipo1G0065500	175,6	6,30E-53	296	3	Phytozome
				Spipo1G0065600	157,9	2,00E-46	266	0	Phytozome
				Spipo20G0017700	170,6	1,40E-51	249	1	Phytozome
				Spipo23G0021200	140,2	7,10E-40	248	0	Phytozome
				Spipo28G0022500	159,5	1,00E-46	300	2	Phytozome
				Spipo29G0025500	132,5	5,30E-37	247	1	Phytozome
				Spipo2G0061400	106,7	2,50E-27	266	2	Phytozome
				Spipo2G0113900	136,7	3,60E-38	290	0	Phytozome
				Spipo32G0003200	282	2,80E-95	223	0	Phytozome
				Spipo9G0041100	158,7	1,40E-46	277	2	Phytozome
Angiosperm (mocot)	Poaceae	*Brachypodium distachyon*	Bdi	Bradi1g13060.1	261,2	2,20E-86	239	0	Phytozome
				Bradi1g13070.1	257,3	6,30E-85	234	0	Phytozome
				Bradi1g30117.1	141	5,50E-39	319	2	Phytozome
				Bradi1g33540.1	172,2	1,20E-51	248	0	Phytozome
				Bradi1g68330.1	141	5,50E-39	332	2	Phytozome
				Bradi1g68340.1	155,2	2,70E-44	329	2	Phytozome
				Bradi1g69277.2	154,1	1,60E-44	267	1	Phytozome

				Bradi2g01146.1	210,7	5,80E-63	612	0	Phytozome
				Bradi2g01200.2	213,4	8,80E-64	634	2	Phytozome
				Bradi2g01217.1	207,6	5,30E-62	585	0	Phytozome
				Bradi2g01228.1	189,1	2,60E-56	421	1	Phytozome
				Bradi2g54560.1	113,6	1,80E-29	254	0	Phytozome
				Bradi3g04330.2	156,8	8,80E-45	337	2	Phytozome
				Bradi3g07960.1*	143,7	1,60E-41	174	0	Phytozome
				Bradi3g21100.1	129	3,40E-34	378	2	Phytozome
				Bradi3g26630.2	111,3	3,20E-28	298	2	Phytozome
				Bradi3g40596.2	145,2	1,10E-39	466	3	Phytozome
				Bradi3g42380.2	143,7	6,00E-40	329	2	Phytozome
				Bradi4g03280.1*	94,7	4,00E-23	186	0	Phytozome
				Bradi4g03285.1*	62	1,70E-11	201	0	Phytozome
				Bradi4g03290.1	245,7	1,30E-80	226	0	Phytozome
				Bradi4g04150.1*	65,9	9,40E-13	218	0	Phytozome
				Bradi4g04160.1*	71,2	1,20E-14	219	0	Phytozome
				Bradi4g04180.1*	67,4	3,20E-13	229	0	Phytozome
				Bradi4g05430.2	231,9	8,90E-75	256	1	Phytozome
				Bradi4g05440.1	220,3	1,20E-70	227	0	Phytozome
				Bradi4g09130.1	132,1	2,00E-36	243	0	Phytozome
				Bradi4g09220.1*	139,8	4,70E-40	177	0	Phytozome
				Bradi4g09370.1*	71,2	1,60E-14	258	0	Phytozome
				Bradi4g34180.1	156,8	2,20E-44	383	2	Phytozome
				Bradi4g36400.1	147,5	1,90E-41	325	1	Phytozome
				Bradi4g36410.1	135,2	6,90E-37	322	2	Phytozome
				Bradi5g00550.1	137,5	3,90E-38	270	0	Phytozome
				Bradi5g27280.3	147,5	1,20E-41	301	2	Phytozome
Angiosperm (mocot)	Poaceae	*Oryza sativa*	Osa	LOC_Os01g02310.1	206,8	2,10E-61	627	2	Phytozome
				LOC_Os01g12150.1*	53,1	9,70E-09	143	1	Phytozome
				LOC_Os01g62260.1	115,9	2,40E-30	247	0	Phytozome
				LOC_Os02g02200.1*	75,5	3,70E-16	244	1	Phytozome
				LOC_Os03g13070.1	159,8	9,50E-47	265	1	Phytozome
				LOC_Os03g14030.1	145,2	1,40E-40	336	2	Phytozome
				LOC_Os03g14050.1*	154,8	3,30E-44	332	2	Phytozome
				LOC_Os03g45960.1	253,1	2,20E-83	232	0	Phytozome
				LOC_Os03g46060.1	189,9	5,90E-59	222	0	Phytozome
				LOC_Os03g46070.1	262,7	3,10E-87	229	0	Phytozome
				LOC_Os04g24130.1	47,8	4,10E-06	433	4	Phytozome
				LOC_Os04g59370.1	161,4	2,60E-47	278	1	Phytozome
				LOC_Os06g16990.1*	34,3	3,00E-02	122	1	Phytozome
				LOC_Os06g19250.1	60,5	7,60E-11	223	4	Phytozome
				LOC_Os06g47600.1	171,4	2,40E-51	251	0	Phytozome
				LOC_Os06g50240.1	156,8	5,30E-45	318	2	Phytozome
				LOC_Os07g04730.1	51,2	3,00E-07	695	4	Phytozome
				LOC_Os07g23470.1	142,5	2,40E-40	249	1	Phytozome
				LOC_Os07g23730.1	137,1	5,00E-38	276	1	Phytozome
				LOC_Os08g29320.1*	55,5	9,20E-10	113	0	Phytozome
				LOC_Os08g40600.1	129,8	1,60E-34	383	2	Phytozome
				LOC_Os08g43510.1	153,7	8,50E-44	323	2	Phytozome
				LOC_Os09g32280.1	158,3	1,40E-45	331	2	Phytozome
				LOC_Os09g36560.1	159,1	7,70E-46	330	2	Phytozome
				LOC_Os09g36580.1	142,5	8,80E-40	312	2	Phytozome
				LOC_Os10g05600.1	133,7	6,60E-36	389	2	Phytozome

				LOC_Os10g05660.1	186,4	2,40E-56	325	1	Phytozome
				LOC_Os10g27280.1	117,1	1,20E-30	275	1	Phytozome
				LOC_Os10g38100.1*	57,4	6,60E-10	196	0	Phytozome
				LOC_Os11g47670.1*	64,7	2,10E-12	212	0	Phytozome
				LOC_Os11g47680.1*	82,8	9,20E-19	206	0	Phytozome
				LOC_Os11g47944.1	138,7	6,50E-39	253	1	Phytozome
				LOC_Os12g38120.1*	82	2,30E-18	230	0	Phytozome
				LOC_Os12g38150.1*	55,8	4,60E-09	233	0	Phytozome
				LOC_Os12g38170.1	77,4	7,10E-17	233	0	Phytozome
				LOC_Os12g43380.1*	147,9	3,10E-43	177	0	Phytozome
				LOC_Os12g43390.1*	135,2	2,50E-38	174	0	Phytozome
				LOC_Os12g43410.1*	95,1	1,20E-23	150	1	Phytozome
				LOC_Os12g43430.1*	147,9	3,20E-43	178	1	Phytozome
				LOC_Os12g43440.1*	149,1	1,50E-43	182	0	Phytozome
				LOC_Os12g43450.1	204,1	3,60E-64	238	0	Phytozome
				LOC_Os12g43490.1	214,5	4,50E-68	258	0	Phytozome
Angiosperm (mocot)	Poaceae	*Setaria italica*	Sit	Seita.1G148000.1	164,5	5,20E-48	316	2	Phytozome
				Seita.1G148300.1	172,9	2,90E-51	313	2	Phytozome
				Seita.1G148900.1	167,9	1,70E-49	309	2	Phytozome
				Seita.1G157500.1	166,8	5,30E-49	316	2	Phytozome
				Seita.2G113200.1	137,9	1,20E-38	247	1	Phytozome
				Seita.2G113300.1	140,6	5,80E-40	223	0	Phytozome
				Seita.2G114300.1	148,7	1,60E-42	277	1	Phytozome
				Seita.2G260400.1	165,6	2,50E-48	336	2	Phytozome
				Seita.2G289000.1	169,5	8,50E-50	331	2	Phytozome
				Seita.2G289100.1	151	5,60E-43	310	2	Phytozome
				Seita.2G365800.1	263,1	2,60E-87	231	0	Phytozome
				Seita.3G003900.1	142,1	7,50E-40	281	1	Phytozome
				Seita.3G328000.1*	48,9	1,10E-06	240	1	Phytozome
				Seita.3G328100.1*	63,5	5,70E-12	241	0	Phytozome
				Seita.3G328300.1*	72	5,40E-15	226	0	Phytozome
				Seita.3G398800.1*	47	4,70E-06	225	3	Phytozome
				Seita.3G398900.1*	99	1,30E-24	197	0	Phytozome
				Seita.3G399100.1*	115,2	8,50E-31	183	0	Phytozome
				Seita.3G399800.1*	120,9	4,60E-33	172	0	Phytozome
				Seita.3G399900.1*	145,2	2,40E-42	173	0	Phytozome
				Seita.3G400000.1*	131,3	5,70E-37	173	0	Phytozome
				Seita.4G252200.1	166,4	1,70E-49	254	0	Phytozome
				Seita.4G281400.1	157,5	5,60E-46	256	1	Phytozome
				Seita.5G078100.1	218,4	2,60E-67	458	2	Phytozome
				Seita.5G078200.1	194,9	3,40E-57	623	1	Phytozome
				Seita.5G078300.1	195,3	7,10E-59	409	1	Phytozome
				Seita.5G355700.1	214,5	1,40E-68	222	0	Phytozome
				Seita.5G375800.1	106,3	8,50E-27	249	0	Phytozome
				Seita.6G194300.1	148,7	4,70E-42	325	2	Phytozome
				Seita.6G214000.1	163,7	3,90E-47	392	2	Phytozome
				Seita.6G239100.1	164,9	3,40E-48	316	2	Phytozome
				Seita.7G019500.1	141,4	9,20E-40	273	1	Phytozome
				Seita.8G248100.1*	65,9	9,10E-13	245	0	Phytozome
				Seita.8G251000.1	134,8	2,30E-37	256	0	Phytozome
				Seita.9G148300.1	263,5	1,70E-87	227	0	Phytozome
				Seita.9G148400.1*	140,6	4,60E-40	215	1	Phytozome
				Seita.9G261300.1	110,9	2,40E-28	274	1	Phytozome
				Seita.9G288400.1	167,5	2,90E-49	317	1	Phytozome

				Seita.9G288500.1	120,9	2,50E-31	390	2	Phytozome
				Seita.9G471800.1	162,2	3,80E-47	326	2	Phytozome
				Seita.9G471900.1	139,8	1,50E-38	340	2	Phytozome
				Seita.9G483200.1	142,9	3,00E-40	275	1	Phytozome
Angiosperm (mocot)	Poaceae	*Zea mays*	Zma	AC207628.4_FGP003	80,1	2,20E-16	1822	16	Phytozome
				AC210168.4_FGP003*	62,4	1,20E-11	186	1	Phytozome
				GRMZM2G002555_T01	185,3	2,00E-53	645	0	Phytozome
				GRMZM2G006853_T01*	142,1	7,40E-41	177	0	Phytozome
				GRMZM2G010048_T01	146	9,90E-42	233	1	Phytozome
				GRMZM2G023655_T01	169,9	4,70E-50	311	2	Phytozome
				GRMZM2G026948_T01*	42,7	3,50E-05	105	0	Phytozome
				GRMZM2G036826_T01	161	1,20E-46	317	2	Phytozome
				GRMZM2G038490_T01	167,2	1,10E-49	242	0	Phytozome
				GRMZM2G038846_T01	144,4	1,20E-40	277	1	Phytozome
				GRMZM2G039639_T01	260,4	3,80E-86	230	0	Phytozome
				GRMZM2G049057_T02	157,5	2,80E-45	319	2	Phytozome
				GRMZM2G050867_T01	63,5	3,50E-11	682	9	Phytozome
				GRMZM2G053890_T01*	98,2	3,90E-24	209	1	Phytozome
				GRMZM2G058394_T01	132,1	3,30E-36	256	1	Phytozome
				GRMZM2G064605_T01	121,7	3,90E-32	288	0	Phytozome
				GRMZM2G066602_T01	67	3,50E-13	226	0	Phytozome
				GRMZM2G086410_T01	147,9	1,10E-41	318	2	Phytozome
				GRMZM2G092474_T01	219,6	2,70E-70	220	0	Phytozome
				GRMZM2G100747_T02	140,6	3,90E-39	281	1	Phytozome
				GRMZM2G108396_T01	63,5	1,00E-11	240	0	Phytozome
				GRMZM2G125918_T02	166,4	7,30E-49	295	1	Phytozome
				GRMZM2G136372_T01*	142,1	7,40E-41	177	0	Phytozome
				GRMZM2G138896_T03	157,9	2,20E-45	319	2	Phytozome
				GRMZM2G148536_T03	138,7	1,50E-37	398	2	Phytozome
				GRMZM2G149798_T01	134	2,50E-36	333	2	Phytozome
				GRMZM2G149809_T01	160,6	2,50E-46	326	2	Phytozome
				GRMZM2G151589_T01	130,6	8,00E-35	371	2	Phytozome
				GRMZM2G154449_T01	140,6	4,50E-39	301	2	Phytozome
				GRMZM2G159110_T01	156,8	6,30E-45	330	2	Phytozome
				GRMZM2G340534_T01	159,8	5,30E-46	329	2	Phytozome
				GRMZM2G346861_T02	149,1	6,50E-42	335	2	Phytozome
				GRMZM2G374971_T01	268,5	2,10E-89	227	0	Phytozome
				GRMZM2G377143_T01*	113,2	5,90E-30	170	0	Phytozome
				GRMZM2G393507_T01	151,8	1,40E-43	263	1	Phytozome
				GRMZM2G402631_T01*	131,3	6,80E-37	174	0	Phytozome
				GRMZM2G429982_T01*	51,2	5,20E-08	117	0	Phytozome
				GRMZM2G435592_T02	188,7	1,20E-54	653	0	Phytozome
				GRMZM2G475505_T01*	45,1	3,00E-05	254	0	Phytozome
				GRMZM2G476523_T01	181,8	2,40E-55	249	0	Phytozome
				GRMZM2G477139_T01	183	4,90E-54	407	1	Phytozome
				GRMZM2G541730_T01	126,3	1,40E-33	324	2	Phytozome
				GRMZM5G861959_T02	67	2,40E-12	641	7	Phytozome
				GRMZM5G882302_T01*	62,8	5,90E-12	168	1	Phytozome
				GRMZM6G397851_T01*	135,6	1,60E-38	171	0	Phytozome
Angiosperm (eudicot)	Ranunculaceae	*Aquilegia coerulea*	Aco	Aqcoe1G104300.1	137,9	1,30E-38	247	1	Phytozome
				Aqcoe1G124700.1	98,6	2,40E-24	216	0	Phytozome
				Aqcoe1G125000.1	132,1	1,50E-36	246	1	Phytozome
				Aqcoe1G125100.1	157,1	2,50E-46	221	0	Phytozome
				Aqcoe1G125200.1	60,8	6,30E-11	252	1	Phytozome
				Aqcoe1G125400.1	140,6	9,80E-40	235	1	Phytozome
				Aqcoe1G133800.1	150,6	1,40E-43	234	1	Phytozome
				Aqcoe1G133900.1	141,7	3,10E-40	234	1	Phytozome
				Aqcoe1G134200.1	132,1	1,10E-36	231	1	Phytozome

				Aqcoe1G274300.1*	139,8	5,00E-40	188	0	Phytozome
				Aqcoe1G274400.1	179,1	2,20E-54	257	0	Phytozome
				Aqcoe1G274500.1	174,9	1,40E-52	268	0	Phytozome
				Aqcoe1G303100.1*	80,9	7,70E-19	112	1	Phytozome
				Aqcoe1G465700.1	116,7	1,50E-30	256	0	Phytozome
				Aqcoe1G490700.1	117,5	4,90E-31	248	1	Phytozome
				Aqcoe1G490800.1	146	1,70E-41	266	2	Phytozome
				Aqcoe2G226300.1	133,7	1,00E-34	868	8	Phytozome
				Aqcoe3G011200.1	153,7	9,40E-45	246	1	Phytozome
				Aqcoe3G114100.1	223,8	1,50E-71	263	0	Phytozome
				Aqcoe3G114200.1	253,4	7,90E-83	292	1	Phytozome
				Aqcoe3G317400.1	120,9	3,70E-32	258	1	Phytozome
				Aqcoe5G259000.1	105,1	9,20E-27	206	1	Phytozome
				Aqcoe5G278000.1	171,4	4,80E-51	285	2	Phytozome
				Aqcoe6G108900.1	151,4	1,00E-43	250	1	Phytozome
				Aqcoe6G211700.1	162,2	2,20E-47	303	2	Phytozome
				Aqcoe6G212100.1	162,2	8,20E-47	353	2	Phytozome
				Aqcoe7G040400.1	236,9	4,20E-77	237	0	Phytozome
				Aqcoe7G040600.1	300,4	3,60E-102	224	0	Phytozome
				Aqcoe7G040700.1	299,7	3,80E-102	201	0	Phytozome
				Aqcoe7G040800.1*	196,4	4,90E-62	182	1	Phytozome
				Aqcoe7G295800.1	158,3	1,20E-45	322	2	Phytozome
Angiosperm (eudicot)	Phrymaceae	*Mimulus guttatus*	Mgu	Migut.A00799.1	257,3	2,90E-85	227	0	Phytozome
				Migut.A01135.1	139,8	3,50E-39	282	0	Phytozome
				Migut.B00098.1	161,4	8,50E-47	342	2	Phytozome
				Migut.B00099.1	141,4	1,00E-39	286	3	Phytozome
				Migut.B00845.1	151	1,80E-43	282	2	Phytozome
				Migut.B01335.1	153,7	8,40E-45	247	0	Phytozome
				Migut.B01336.1	99,4	7,10E-25	194	2	Phytozome
				Migut.C00688.1	161	1,40E-47	249	1	Phytozome
				Migut.C00849.1	166,8	4,20E-49	316	2	Phytozome
				Migut.C00850.1	136,3	1,20E-37	307	2	Phytozome
				Migut.D01556.1	150,6	1,30E-43	248	2	Phytozome
				Migut.E00851.1*	162,5	2,00E-49	143	2	Phytozome
				Migut.E01123.1	268,9	1,00E-89	230	0	Phytozome
				Migut.E01125.1	287,3	4,40E-97	229	0	Phytozome
				Migut.E01128.1	203,4	3,10E-64	223	0	Phytozome
				Migut.F00124.1	167,9	9,70E-50	295	2	Phytozome
				Migut.H00882.1	143,3	1,60E-40	278	1	Phytozome
				Migut.H02013.1*	53,5	3,30E-09	103	1	Phytozome
				Migut.H02015.1	139,8	1,60E-39	241	1	Phytozome
				Migut.J00460.1	145,2	1,10E-41	239	1	Phytozome
				Migut.J00461.1	146	6,30E-42	239	1	Phytozome
				Migut.J00462.1	131	3,00E-36	239	1	Phytozome
				Migut.J00463.1	114,8	3,60E-30	241	1	Phytozome
				Migut.K00892.1	120,2	4,90E-32	248	0	Phytozome
				Migut.L00322.1	123,3	4,80E-33	276	0	Phytozome
				Migut.L00502.1	134,4	2,40E-35	689	4	Phytozome
				Migut.L00673.1	142,5	6,40E-40	313	2	Phytozome
				Migut.L00700.1	143,7	2,30E-40	315	2	Phytozome
				Migut.L00701.1	134	4,10E-37	268	0	Phytozome
				Migut.L00960.1	137,5	1,20E-38	244	2	Phytozome
				Migut.L01961.1	146,4	5,90E-42	247	1	Phytozome
				Migut.N00519.1	150,2	9,10E-44	218	1	Phytozome
				Migut.N03166.1	125,2	2,10E-33	320	2	Phytozome
				Migut.N03301.1	167,9	7,10E-50	281	2	Phytozome

Angiosperm (eudicot)	Solanaceae	*Solanum lycopersicum*	Sly	Solyc01g086840.1.1	165,2	2,20E-49	238	2	Phytozome
				Solyc01g104290.1.1	154,8	1,30E-44	316	3	Phytozome
				Solyc01g111320.2.1	157,5	5,90E-46	276	2	Phytozome
				Solyc01g111330.2.1	141,4	3,70E-39	356	2	Phytozome
				Solyc02g083760.2.1	148,7	3,20E-42	326	2	Phytozome
				Solyc02g083790.2.1	168,7	5,80E-50	304	2	Phytozome
				Solyc02g087520.2.1	120,9	3,10E-32	270	2	Phytozome
				Solyc03g033490.1.1	155,2	8,00E-45	311	2	Phytozome
				Solyc03g079960.2.1	166,8	2,60E-49	297	2	Phytozome
				Solyc03g118780.2.1	154,8	2,50E-45	245	2	Phytozome
				Solyc04g007310.1.1	106,7	4,50E-27	255	1	Phytozome
				Solyc04g079890.2.1	144,8	1,50E-41	247	1	Phytozome
				Solyc04g081550.2.1	156,8	4,50E-45	335	2	Phytozome
				Solyc04g081560.2.1	150,2	8,30E-43	325	3	Phytozome
				Solyc05g053020.2.1	147,9	1,20E-42	244	1	Phytozome
				Solyc06g073000.2.1	181	1,00E-54	306	2	Phytozome
				Solyc07g017970.1.1	132,9	5,50E-37	240	1	Phytozome
				Solyc08g080590.2.1*	154,1	4,40E-46	157	1	Phytozome
				Solyc08g080600.1.1*	134,4	1,50E-39	75	0	Phytozome
				Solyc08g080610.1.1*	313,9	2,00E-108	164	0	Phytozome
				Solyc08g080620.1.1	399,8	2,60E-141	226	0	Phytozome
				Solyc08g080640.1.1	434,1	1,60E-154	247	0	Phytozome
				Solyc08g080650.1.1	443,7	2,50E-158	246	0	Phytozome
				Solyc08g080660.1.1	381,7	6,10E-134	250	0	Phytozome
				Solyc08g080670.1.1	385,6	2,30E-135	250	0	Phytozome
				Solyc10g084840.1.1	147,5	1,90E-42	245	1	Phytozome
				Solyc11g013300.1.1	172,9	9,10E-52	287	2	Phytozome
				Solyc11g044390.1.1*	109,8	5,20E-30	78	0	Phytozome
				Solyc11g044400.1.1*	124,4	1,30E-35	87	0	Phytozome
				Solyc11g066130.1.1	134	2,70E-37	252	0	Phytozome
				Solyc12g056360.1.1	258,5	8,50E-86	229	0	Phytozome
				Solyc12g056370.1.1*	165,6	1,60E-49	234	2	Phytozome
				Solyc12g056380.1.1	241,9	3,20E-79	229	0	Phytozome
				Solyc12g056390.1.1	285,4	2,30E-96	227	0	Phytozome
Angiosperm (eudicot)	Solanaceae	*Solanum tuberosum*	Stu	PGSC0003DMT400009196	170,2	2,20E-50	303	2	Phytozome
				PGSC0003DMT400009202	143,7	3,20E-40	314	4	Phytozome
				PGSC0003DMT400003284	124	2,30E-33	262	2	Phytozome
				PGSC0003DMT400034395	364,8	1,30E-127	216	0	Phytozome
				PGSC0003DMT400061140	151,4	2,20E-43	288	2	Phytozome
				PGSC0003DMT400015871	131,3	3,30E-35	366	2	Phytozome
				PGSC0003DMT400015872	159,1	6,10E-46	323	3	Phytozome
				PGSC0003DMT400065004	173,7	2,00E-52	238	2	Phytozome
				PGSC0003DMT400069795	144,8	2,60E-41	244	1	Phytozome
				PGSC0003DMT400010890	289,7	3,60E-95	426	1	Phytozome
				PGSC0003DMT400010888*	136,7	1,80E-39	141	0	Phytozome
				PGSC0003DMT400010887*	206,1	6,90E-66	170	0	Phytozome
				PGSC0003DMT400010892*	162,5	1,30E-49	129	0	Phytozome
				PGSC0003DMT400010886	286,6	1,00E-96	227	0	Phytozome
				PGSC0003DMT400010891	226,9	1,90E-73	216	0	Phytozome
				PGSC0003DMT400069224	181,4	1,10E-54	306	2	Phytozome
				PGSC0003DMT400022986*	133,3	1,90E-37	199	0	Phytozome
				PGSC0003DMT400014747	151	1,10E-43	247	2	Phytozome
				PGSC0003DMT400058117*	67	6,10E-14	128	1	Phytozome
				PGSC0003DMT400001544	166,4	5,70E-49	297	2	Phytozome
				PGSC0003DMT400001066	134	3,70E-37	251	0	Phytozome
				PGSC0003DMT400016741	172,2	5,10E-52	233	0	Phytozome

				PGSC0003DMT400007865	408,7	9,60E-145	226	0	Phytozome
				PGSC0003DMT400007905	380,2	4,40E-133	250	0	Phytozome
				PGSC0003DMT400007869	435,6	5,60E-155	247	0	Phytozome
				PGSC0003DMT400007866*	109,8	8,70E-29	182	2	Phytozome
				PGSC0003DMT400007906	389	1,30E-136	250	0	Phytozome
				PGSC0003DMT400007870	441,8	2,00E-157	246	0	Phytozome
				PGSC0003DMT400007868	434,1	1,90E-154	247	0	Phytozome
				PGSC0003DMT400043496	148,3	1,20E-42	248	0	Phytozome
				PGSC0003DMT400025547	156,8	5,00E-45	335	2	Phytozome
				PGSC0003DMT400015398	102,1	2,20E-25	249	1	Phytozome
				PGSC0003DMT400028761	144,8	2,70E-41	245	1	Phytozome
Angiosperm (eudicot)	Myrtaceae	*Eucalyptus grandis*	Egr	Eucgr.A00487.1	152,1	4,00E-44	248	1	Phytozome
				Eucgr.A01288.1	154,1	9,40E-45	249	1	Phytozome
				Eucgr.A01612.1	127,1	9,80E-35	237	1	Phytozome
				Eucgr.B00944.1	150,2	3,00E-43	249	2	Phytozome
				Eucgr.D01887.1	283,9	1,50E-95	227	0	Phytozome
				Eucgr.D01888.1	344,4	3,80E-119	241	0	Phytozome
				Eucgr.D01890.1*	172,2	7,70E-53	163	1	Phytozome
				Eucgr.D01891.1*	74,7	1,80E-16	146	3	Phytozome
				Eucgr.D01892.1	298,1	3,30E-101	225	0	Phytozome
				Eucgr.D01893.1	228	1,00E-73	223	0	Phytozome
				Eucgr.D01894.1	255,8	3,00E-84	243	1	Phytozome
				Eucgr.D01897.1*	111,7	1,30E-29	162	2	Phytozome
				Eucgr.D01898.1	287,3	5,90E-97	225	0	Phytozome
				Eucgr.D01899.1	230,3	8,90E-75	211	0	Phytozome
				Eucgr.D01900.1	285,4	4,40E-96	231	1	Phytozome
				Eucgr.D01904.1	296,2	2,20E-100	225	0	Phytozome
				Eucgr.E00557.1*	225,3	4,40E-73	196	1	Phytozome
				Eucgr.E00560.1	280,4	4,30E-94	235	0	Phytozome
				Eucgr.E00561.1	205,3	4,10E-65	208	0	Phytozome
				Eucgr.E01381.1	132,9	1,00E-36	248	1	Phytozome
				Eucgr.E01382.1	143,3	1,10E-40	249	1	Phytozome
				Eucgr.E01384.1	143,3	1,10E-40	249	1	Phytozome
				Eucgr.E01385.1	135,2	1,10E-37	249	1	Phytozome
				Eucgr.E01389.1	143,3	1,10E-40	249	1	Phytozome
				Eucgr.E01392.1	146,4	6,60E-42	249	1	Phytozome
				Eucgr.F00728.1*	101,7	1,70E-25	211	3	Phytozome
				Eucgr.F01670.1	150,2	2,60E-43	255	1	Phytozome
				Eucgr.F03757.1	166	6,70E-49	287	2	Phytozome
				Eucgr.F03936.1	122,5	9,70E-33	265	1	Phytozome
				Eucgr.F03937.1	121,3	5,90E-32	311	2	Phytozome
				Eucgr.F03940.1	125,6	2,00E-33	311	2	Phytozome
				Eucgr.G00607.1	147,1	1,10E-41	289	2	Phytozome
				Eucgr.G01772.1	157,5	4,20E-45	348	3	Phytozome
				Eucgr.H00974.1	146	7,30E-41	329	2	Phytozome
				Eucgr.H00975.1	162,2	4,80E-47	322	2	Phytozome
				Eucgr.H03863.1	254,2	4,10E-84	212	2	Phytozome
				Eucgr.H03864.1	303,9	2,10E-103	225	0	Phytozome
				Eucgr.H03865.1	302,4	6,80E-103	225	0	Phytozome
				Eucgr.I01396.1	162,9	1,50E-47	312	2	Phytozome
				Eucgr.I01898.1	122,9	9,40E-33	268	1	Phytozome
				Eucgr.I02339.1	162,5	2,20E-47	306	2	Phytozome
				Eucgr.J02061.1	99,4	1,80E-24	224	0	Phytozome
				Eucgr.J02439.1	158,7	6,30E-46	302	2	Phytozome
				Eucgr.J02440.1	156	6,50E-45	302	2	Phytozome
				Eucgr.J02442.1	149,4	9,20E-43	268	1	Phytozome
				Eucgr.J02443.1	127,1	1,10E-34	234	1	Phytozome

				Eucgr.J02445.1	158,3	5,10E-46	281	1	Phytozome
				Eucgr.J02446.1	159,1	1,90E-45	373	2	Phytozome
				Eucgr.K00478.1	153,3	2,20E-44	256	1	Phytozome
				Eucgr.K03107.1	138,3	6,50E-39	234	1	Phytozome
				Eucgr.L01962.1	303,9	2,10E-103	225	0	Phytozome
				Eucgr.L02491.1	157,5	1,70E-45	310	2	Phytozome
				Eucgr.L02566.1	279,3	7,50E-94	218	1	Phytozome
				Eucgr.L02568.1	282	7,50E-95	223	0	Phytozome
				Eucgr.L03623.1	285,8	2,50E-96	227	0	Phytozome
Angiosperm (eudicot)	Euphorbiaceae	*Manihot esculenta*	Mes	Manes.01G050400.1	159,1	3,80E-46	305	2	Phytozome
				Manes.01G064200.1	322	3,00E-110	247	0	Phytozome
				Manes.01G064300.1	321,6	4,00E-110	247	0	Phytozome
				Manes.01G064400.1	264,2	5,70E-88	217	1	Phytozome
				Manes.01G146800.1	159,8	1,60E-46	298	2	Phytozome
				Manes.01G146900.1	151,4	4,70E-43	320	2	Phytozome
				Manes.02G025100.1	290,8	2,70E-98	226	0	Phytozome
				Manes.02G025200.1	265,8	1,60E-88	223	0	Phytozome
				Manes.02G028300.1	293,5	2,00E-99	225	0	Phytozome
				Manes.02G028400.1	218	4,30E-70	213	0	Phytozome
				Manes.02G105300.1	154,5	1,90E-44	302	2	Phytozome
				Manes.02G105400.1	158,3	9,30E-46	309	2	Phytozome
				Manes.02G160200.1	160,2	7,60E-47	284	2	Phytozome
				Manes.02G160300.1	152,1	1,10E-43	288	2	Phytozome
				Manes.02G199600.1	158,3	1,60E-46	240	1	Phytozome
				Manes.04G072700.1	171,4	2,20E-51	254	1	Phytozome
				Manes.05G146000.1	140,6	1,40E-38	365	1	Phytozome
				Manes.05G146200.1	168,3	1,60E-49	321	2	Phytozome
				Manes.06G007400.1	155,2	3,10E-45	249	1	Phytozome
				Manes.07G014800.1	140,2	6,50E-39	321	2	Phytozome
				Manes.07G118500.1	138,7	6,00E-39	251	1	Phytozome
				Manes.08G013800.1	93,6	2,20E-22	238	1	Phytozome
				Manes.09G094100.1	142,5	2,40E-40	249	1	Phytozome
				Manes.10G130700.1	139,8	6,60E-39	300	2	Phytozome
				Manes.11G095900.1	159,8	5,10E-46	340	2	Phytozome
				Manes.11G132100.1	118,6	2,60E-31	254	0	Phytozome
				Manes.12G003700.1	102,1	2,50E-25	250	0	Phytozome
				Manes.12G003800.1	162,5	3,20E-48	235	1	Phytozome
				Manes.13G003800.1	93,2	3,70E-22	252	1	Phytozome
				Manes.13G003900.1	102,8	1,50E-25	256	1	Phytozome
				Manes.14G170800.1	149,4	4,50E-43	248	1	Phytozome
				Manes.15G175600.1	122,1	1,20E-32	259	1	Phytozome
				Manes.18G012600.1	151	6,10E-43	316	1	Phytozome
				Manes.18G012700.1	167,5	3,90E-49	321	2	Phytozome
				Manes.18G030000.1	147,5	2,30E-42	245	1	Phytozome
				Manes.S049600.1	149,1	3,60E-43	232	0	Phytozome
Angiosperm (eudicot)	Salicaceae	*Salix purpurea*	Spu	SapurV1A.0001s1550.1	146,7	7,10E-42	248	1	Phytozome
				SapurV1A.0008s1120.1	118,6	3,60E-31	249	0	Phytozome
				SapurV1A.0010s0780.1	167,2	2,60E-49	281	2	Phytozome
				SapurV1A.0010s0790.1	167,2	2,60E-49	281	2	Phytozome
				SapurV1A.0010s0800.1	167,2	2,60E-49	281	2	Phytozome
				SapurV1A.0025s0190.1	149,1	5,10E-42	316	2	Phytozome
				SapurV1A.0025s0200.1	165,2	3,70E-48	317	2	Phytozome
				SapurV1A.0040s0150.1	118,2	4,00E-31	247	0	Phytozome
				SapurV1A.0067s0640.1	49,3	1,80E-06	575	3	Phytozome
				SapurV1A.0098s0100.1	325,1	2,30E-111	246	0	Phytozome
				SapurV1A.0105s0010.1	164,1	3,60E-48	282	2	Phytozome
				SapurV1A.0117s0110.1	152,5	5,10E-43	354	2	Phytozome
				SapurV1A.0130s0410.1	157,1	4,60E-45	328	2	Phytozome
				SapurV1A.0130s0420.1	141	7,60E-39	342	2	Phytozome

				SapurV1A.0183s0020.1*	195,7	1,30E-61	172	1	Phytozome
				SapurV1A.0183s0030.1	298,1	4,80E-101	226	0	Phytozome
				SapurV1A.0183s0040.1	303,9	2,30E-103	225	0	Phytozome
				SapurV1A.0183s0050.1	305,4	5,60E-104	225	0	Phytozome
				SapurV1A.0271s0330.1	297,4	1,00E-100	226	0	Phytozome
				SapurV1A.0271s0340.1	301,6	2,40E-102	225	0	Phytozome
				SapurV1A.0271s0350.1	308,5	3,80E-105	225	0	Phytozome
				SapurV1A.0281s0160.1	127,1	1,80E-34	247	1	Phytozome
				SapurV1A.0357s0050.1*	104,8	1,50E-26	193	1	Phytozome
				SapurV1A.0384s0040.1	169,1	1,10E-49	313	2	Phytozome
				SapurV1A.0384s0050.1	147,9	1,50E-41	329	2	Phytozome
				SapurV1A.0507s0090.1*	271,9	9,60E-91	224	1	Phytozome
				SapurV1A.0507s0100.1	321,2	7,40E-110	246	0	Phytozome
				SapurV1A.0507s0110.1	277,3	6,80E-93	225	0	Phytozome
				SapurV1A.0514s0180.1	160,6	2,10E-46	313	2	Phytozome
				SapurV1A.0514s0200.1	157,1	7,40E-45	342	2	Phytozome
				SapurV1A.0519s0110.1	155,6	2,70E-45	248	1	Phytozome
				SapurV1A.0519s0120.1*	46,2	2,00E-06	97	2	Phytozome
				SapurV1A.0610s0070.1*	94,4	2,40E-21	583	1	Phytozome
				SapurV1A.0715s0040.1	150,6	7,30E-43	298	1	Phytozome
				SapurV1A.0761s0080.1	280,4	4,00E-94	227	0	Phytozome
				SapurV1A.0977s0010.1	162,2	8,30E-48	249	1	Phytozome
				SapurV1A.0977s0020.1	146,7	8,40E-42	247	2	Phytozome
				SapurV1A.0977s0040.1	152,5	4,50E-44	247	1	Phytozome
				SapurV1A.1042s0010.1	116,3	2,50E-30	247	0	Phytozome
				SapurV1A.1146s0050.1	154,5	9,40E-45	250	1	Phytozome
				SapurV1A.1265s0060.1	164,9	2,10E-48	282	2	Phytozome
				SapurV1A.1323s0010.1	47,4	8,60E-06	637	3	Phytozome
				SapurV1A.1323s0020.1	152,9	6,40E-42	649	3	Phytozome
				SapurV1A.1323s0030.1*	138,7	5,40E-37	599	3	Phytozome
				SapurV1A.1323s0060.1	149,1	1,60E-40	668	2	Phytozome
				SapurV1A.1365s0040.1*	135,6	2,40E-37	273	3	Phytozome
				SapurV1A.1469s0040.1	157,5	6,00E-46	249	1	Phytozome
				SapurV1A.1469s0050.1	129,4	2,10E-35	237	1	Phytozome
				SapurV1A.1469s0060.1*	98,6	1,80E-24	180	2	Phytozome
				SapurV1A.1469s0070.1	157,1	9,50E-46	249	1	Phytozome
				SapurV1A.1469s0080.1*	96,3	4,70E-24	137	2	Phytozome
				SapurV1A.1518s0060.1	159,5	3,30E-46	300	2	Phytozome
				SapurV1A.1658s0010.1	146,7	6,80E-42	249	2	Phytozome
				SapurV1A.1658s0020.1	163,3	3,80E-48	249	1	Phytozome
				SapurV1A.1658s0030.1*	84,3	8,40E-20	129	1	Phytozome
				SapurV1A.1668s0040.1	162,9	1,70E-47	302	2	Phytozome
				SapurV1A.2357s0010.1	277,3	6,80E-93	225	0	Phytozome
				SapurV1A.2357s0020.1	319,7	3,10E-109	246	0	Phytozome
				SapurV1A.2569s0010.1	121,7	2,80E-32	260	1	Phytozome
				SapurV1A.2698s0020.1	152,9	3,30E-44	250	1	Phytozome
				SapurV1A.3598s0010.1	327,8	2,20E-112	246	0	Phytozome
				SapurV1A.3604s0020.1	159,5	3,30E-46	300	2	Phytozome
				SapurV1A.4851s0010.1	149,1	5,10E-42	316	2	Phytozome
Angiosperm (eudicot)	Brassicaceae	*Arabidopsis lyrata*	Aly	AL1G30620.t1	152,9	1,30E-44	243	2	Phytozome
				AL1G31910.t1	162,5	3,80E-48	248	1	Phytozome
				AL1G32810.t1	160,6	6,70E-47	303	1	Phytozome
				AL2G29540.t2	122,5	3,40E-31	667	6	Phytozome
				AL2G33400.t1	153,7	8,20E-45	244	2	Phytozome
				AL2G34910.t1	170,6	2,50E-51	246	1	Phytozome
				AL2G34930.t1	170,6	1,90E-51	239	1	Phytozome
				AL2G34940.t1	177,6	4,30E-54	245	1	Phytozome
				AL2G35790.t1	167,2	5,00E-49	330	2	Phytozome

				AL2G37920.t1	171,8	2,60E-51	282	2	Phytozome
				AL3G50900.t1	162,9	8,70E-48	298	1	Phytozome
				AL4G23320.t1	123,6	2,10E-33	249	0	Phytozome
				AL4G42380.t1	124,4	1,50E-32	489	4	Phytozome
				AL6G11370.t1	129,8	1,10E-35	248	1	Phytozome
				AL6G36230.t1*	112,1	9,50E-29	312	1	Phytozome
				AL6G46220.t1*	132,1	1,90E-37	172	1	Phytozome
				AL6G46230.t1	298,9	9,60E-102	223	1	Phytozome
				AL6G46240.t1	311,6	2,50E-106	244	1	Phytozome
				AL7G10550.t1	155,6	1,20E-44	345	2	Phytozome
				AL7G10560.t1	158,7	2,50E-46	281	2	Phytozome
				AL7G15040.t1	159,5	1,90E-46	302	1	Phytozome
				AL7G15050.t1*	94,7	2,70E-23	190	2	Phytozome
				AL7G29100.t2	138,7	1,90E-38	314	2	Phytozome
				AL7G30420.t1	266,5	7,70E-89	231	1	Phytozome
				AL7G36160.t1	96,7	3,50E-22	852	4	Phytozome
				AL7G49980.t1	127,9	6,30E-35	256	3	Phytozome
				AL7G52530.t1	153,3	2,40E-42	663	4	Phytozome
				AL7G52540.t1	164,9	1,90E-46	665	4	Phytozome
Angiosperm (eudicot)	Brassicaceae	*Arabidopsis thaliana*	Ath	AT1G18250.2	153,3	1,10E-44	244	2	Phytozome
				AT1G19320.1	152,1	3,20E-44	247	1	Phytozome
				AT1G20030.2	164,9	2,30E-48	316	2	Phytozome
				AT1G70250.1	128,3	5,90E-33	799	5	Phytozome
				AT1G73620.1	156	2,10E-45	264	2	Phytozome
				AT1G75030.1	169,1	9,80E-51	246	1	Phytozome
				AT1G75040.1	169,9	5,20E-51	239	1	Phytozome
				AT1G75050.1	177,6	5,20E-54	246	1	Phytozome
				AT1G75800.1	167,2	4,30E-49	330	2	Phytozome
				AT1G77700.1	173,7	3,30E-51	356	2	Phytozome
				AT2G17860.1	159,5	6,70E-47	253	0	Phytozome
				AT2G24810.1*	55,8	1,70E-09	193	3	Phytozome
				AT2G28790.1	120,9	2,50E-32	249	0	Phytozome
				AT4G11650.1	313,5	4,70E-107	244	1	Phytozome
				AT4G18250.1	100,1	2,30E-23	853	4	Phytozome
				AT4G24180.1	140,6	1,30E-39	260	1	Phytozome
				AT4G36000.1*	107,5	1,10E-27	208	1	Phytozome
				AT4G36010.1	160,2	1,00E-46	301	1	Phytozome
				AT4G38660.1	155,2	2,00E-44	345	2	Phytozome
				AT4G38670.1	156,4	1,90E-45	281	2	Phytozome
				AT5G02140.1	128,6	7,20E-35	294	2	Phytozome
				AT5G24620.1	164,1	4,20E-47	420	1	Phytozome
				AT5G40020.1	127,9	7,80E-35	256	3	Phytozome
				AT5G38280.1	156	3,10E-43	665	4	Phytozome
Angiosperm (eudicot)	Brassicaceae	*Brassica rapa*	Bra	Brara.A00011.1	155,6	3,60E-45	274	2	Phytozome
				Brara.A00247.1*	113,6	8,50E-30	227	2	Phytozome
				Brara.A00940.1	102,1	2,10E-25	249	1	Phytozome
				Brara.A01323.1	263,5	1,70E-87	231	1	Phytozome
				Brara.A01423.1	142,5	9,20E-40	313	2	Phytozome
				Brara.B02149.1	172,9	3,70E-52	247	1	Phytozome
				Brara.B02213.1	167,9	2,10E-49	321	2	Phytozome
				Brara.B02282.1	176,8	3,50E-53	286	2	Phytozome
				Brara.B02758.1	305,8	4,80E-104	244	1	Phytozome
				Brara.B03671.1	160,6	1,10E-45	434	1	Phytozome
				Brara.C00059.1	129,4	1,70E-35	247	1	Phytozome
				Brara.D01016.1	132,5	1,70E-36	256	3	Phytozome
				Brara.D01724.1	108,6	1,10E-27	250	0	Phytozome
				Brara.D02646.1	126,3	1,40E-32	662	4	Phytozome
				Brara.F01282.1	157,5	3,10E-46	243	2	Phytozome
				Brara.F01364.1	166,4	1,20E-49	246	1	Phytozome
				Brara.F01366.1	155,2	2,70E-45	248	1	Phytozome

				Brara.F02676.1	162,2	2,00E-46	403	1	Phytozome
				Brara.G00232.1	157,1	5,90E-46	249	0	Phytozome
				Brara.G01401.1	113,6	1,40E-29	251	0	Phytozome
				Brara.G02134.1	175,6	1,10E-52	284	2	Phytozome
				Brara.G02246.1*	154,5	3,30E-45	228	2	Phytozome
				Brara.G03254.1	155,2	4,30E-45	265	2	Phytozome
				Brara.G03321.1	169,9	1,60E-49	373	2	Phytozome
				Brara.G03322.1	148,3	1,20E-42	250	1	Phytozome
				Brara.G03323.1	170,6	3,40E-51	246	1	Phytozome
				Brara.G03379.1	170,2	2,90E-50	325	1	Phytozome
				Brara.G03527.1	173,7	2,00E-52	248	1	Phytozome
				Brara.H01620.1	151,8	1,30E-43	288	2	Phytozome
				Brara.H01849.1	154,1	8,80E-44	355	2	Phytozome
				Brara.H02374.1	167,9	3,30E-50	247	1	Phytozome
				Brara.I04743.1	172,6	5,90E-52	246	1	Phytozome
				Brara.J01946.1	100,1	1,80E-23	661	4	Phytozome
				Brara.J02899.1	128,6	3,90E-35	248	1	Phytozome
				Brara.K00455.1	92,4	6,10E-21	622	3	Phytozome
				Brara.K00744.1	144,4	1,30E-40	313	2	Phytozome
Angiosperm (eudicot)	Brassicaceae	*Capsella grandiflora*	Cgr	Cagra.0096s0064.1	172,6	9,30E-52	280	2	Phytozome
				Cagra.0402s0024.1	177,6	4,30E-54	249	1	Phytozome
				Cagra.0402s0025.1	172,9	2,20E-52	240	1	Phytozome
				Cagra.0402s0026.1	172,9	2,70E-52	246	1	Phytozome
				Cagra.0909s0003.1	154,5	3,20E-45	243	2	Phytozome
				Cagra.1226s0084.1	144,1	1,40E-40	312	2	Phytozome
				Cagra.1383s0050.1	152,1	2,00E-43	340	2	Phytozome
				Cagra.1383s0051.1	155,6	3,20E-45	281	2	Phytozome
				Cagra.1515&s0009.1	262,7	2,10E-87	231	1	Phytozome
				Cagra.1595s0024.1	96,3	2,80E-22	657	4	Phytozome
				Cagra.1595s0025.1	90,9	1,80E-20	652	4	Phytozome
				Cagra.1595s0026.1*	56,6	3,10E-09	598	5	Phytozome
				Cagra.1785&s0001.1	108,2	2,10E-26	679	5	Phytozome
				Cagra.1813&s0003.1	122,9	3,40E-33	248	3	Phytozome
				Cagra.1912s0002.1	295	3,50E-100	223	1	Phytozome
				Cagra.1912s0003.1	311,6	2,00E-106	244	1	Phytozome
				Cagra.1961s0004.1	163,3	1,10E-47	323	2	Phytozome
				Cagra.1961s0069.1	155,2	2,10E-45	250	1	Phytozome
				Cagra.2236s0010.1	155,2	2,20E-45	255	0	Phytozome
				Cagra.2637s0020.1	129,8	1,30E-35	266	2	Phytozome
				Cagra.2913s0003.1	152,9	2,20E-44	269	2	Phytozome
				Cagra.3126s0018.1	159,5	5,40E-47	254	0	Phytozome
				Cagra.3138s0010.1	142,1	1,60E-38	622	6	Phytozome
				Cagra.3138s0011.1	154,1	6,30E-45	257	1	Phytozome
				Cagra.3706s0043.1	119	4,10E-30	663	3	Phytozome
				Cagra.4538s0005.1	100,5	2,80E-25	212	1	Phytozome
				Cagra.9017s0003.1	164,9	1,60E-47	413	1	Phytozome
Angiosperm (eudicot)	Fabaceae	*Glycine max*	Gma	Glyma.01G165400.1	160,6	2,60E-46	318	2	Phytozome
				Glyma.01G165600.1	161	1,70E-46	304	2	Phytozome
				Glyma.01G165800.1*	97,4	2,30E-23	256	1	Phytozome
				Glyma.01G217600.1	263,1	3,60E-87	223	0	Phytozome
				Glyma.01G217700.1	305,4	2,40E-103	249	2	Phytozome
				Glyma.02G047400.1	160,6	2,80E-46	309	2	Phytozome
				Glyma.02G220900.1	168,7	2,00E-49	313	2	Phytozome
				Glyma.02G249500.1	141,4	1,20E-39	246	2	Phytozome
				Glyma.04G023700.1	167,9	6,30E-49	329	2	Phytozome
				Glyma.04G034300.1	129,4	2,00E-35	223	2	Phytozome
				Glyma.04G176200.1	114,4	3,00E-29	277	2	Phytozome

				Glyma.04G225800.1	112,1	1,30E-28	251	4	Phytozome
				Glyma.05G169700.1	142,9	8,20E-40	287	2	Phytozome
				Glyma.05G204600.1	288,9	2,50E-97	224	0	Phytozome
				Glyma.05G204800.1	289,3	2,10E-97	222	0	Phytozome
				Glyma.05G245800.1	151	2,80E-43	245	0	Phytozome
				Glyma.06G023900.1	164,5	1,40E-47	328	2	Phytozome
				Glyma.06G139000.1	164,5	2,90E-48	273	3	Phytozome
				Glyma.06G189100.1	120,2	3,60E-31	321	2	Phytozome
				Glyma.07G135100.1	136,7	6,30E-38	245	1	Phytozome
				Glyma.08G053600.1	153,3	4,40E-44	253	0	Phytozome
				Glyma.08G128000.1	141,4	3,30E-39	289	2	Phytozome
				Glyma.08G172700.1	120,2	6,80E-32	216	1	Phytozome
				Glyma.09G003600.1*	133,7	1,40E-37	166	0	Phytozome
				Glyma.10G060300.1	131	1,10E-35	243	1	Phytozome
				Glyma.10G060800.1	132,9	3,20E-36	263	1	Phytozome
				Glyma.10G060900.1	65,9	8,80E-13	192	0	Phytozome
				Glyma.10G061000.1	132,5	2,50E-36	242	1	Phytozome
				Glyma.10G061100.1	65,5	1,20E-12	192	0	Phytozome
				Glyma.10G061700.1	131	1,10E-35	243	1	Phytozome
				Glyma.10G061800.1	132,5	3,40E-36	261	1	Phytozome
				Glyma.10G062100.1	92	1,70E-21	240	1	Phytozome
				Glyma.11G025600.1	263,1	3,70E-87	224	0	Phytozome
				Glyma.11G077700.1	164,1	9,10E-48	304	2	Phytozome
				Glyma.11G077800.1	158,3	2,40E-45	316	2	Phytozome
				Glyma.11G106100.1	173,7	2,80E-51	312	2	Phytozome
				Glyma.11G106200.1*	42,7	8,30E-05	132	1	Phytozome
				Glyma.11G140800.1	114,8	1,10E-29	248	0	Phytozome
				Glyma.11G220700.1	134,8	3,90E-37	251	1	Phytozome
				Glyma.12G031000.1	171	2,60E-50	312	2	Phytozome
				Glyma.12G031200.1	161,8	1,90E-46	346	2	Phytozome
				Glyma.12G064300.1	117,5	1,10E-30	248	0	Phytozome
				Glyma.12G238900.1	151,4	1,50E-43	236	1	Phytozome
				Glyma.13G082700.1	159,5	4,40E-46	287	2	Phytozome
				Glyma.14G077300.1	154,5	1,20E-44	250	1	Phytozome
				Glyma.14G077400.1	169,1	2,90E-50	251	1	Phytozome
				Glyma.14G163700.1	157,1	3,60E-45	287	2	Phytozome
				Glyma.14G188400.1	171	3,40E-50	313	2	Phytozome
				Glyma.14G219600.1	168,3	3,40E-49	318	2	Phytozome
				Glyma.15G254700.1	120,9	3,90E-32	232	0	Phytozome
				Glyma.15G258400.1	65,9	9,40E-12	885	1	Phytozome
				Glyma.15G275200.1	117,1	6,10E-31	206	1	Phytozome
				Glyma.16G126900.1	156,4	8,10E-45	306	2	Phytozome
				Glyma.16G127100.1	156,8	5,60E-45	299	2	Phytozome
				Glyma.17G248300.1	161	4,00E-47	248	1	Phytozome
				Glyma.17G258500.1	164,5	1,00E-47	315	2	Phytozome
				Glyma.17G258600.1	131,3	3,00E-35	314	2	Phytozome
				Glyma.18G036700.1	109,4	8,80E-28	235	2	Phytozome
				Glyma.18G185100.1	113,6	9,80E-30	193	1	Phytozome
				Glyma.19G018400.1	139	1,40E-38	261	1	Phytozome
				Glyma.19G087400.1*	34,3	2,80E-02	73	0	Phytozome
Angiosperm (eudicot)	Fabaceae	*Medicago truncatula*	Mtr	Medtr1g021945.1	164,9	6,90E-49	249	1	Phytozome
				Medtr1g025420.1	162,5	4,10E-47	323	2	Phytozome
				Medtr1g062340.1	130,6	8,30E-36	244	1	Phytozome
				Medtr1g062350.1	129,8	1,70E-35	244	1	Phytozome
				Medtr1g062370.1	127,9	7,80E-35	244	1	Phytozome
				Medtr1g062390.1	89	1,20E-20	244	1	Phytozome
				Medtr1g062590.1	134,4	2,80E-37	242	1	Phytozome
				Medtr1g062630.1	139,4	4,80E-39	262	1	Phytozome

				Medtr1g062640.1*	45,8	5,00E-06	154	1	Phytozome
				Medtr1g062660.1	125,6	7,10E-34	244	1	Phytozome
				Medtr1g062690.1*	105,1	1,30E-26	220	2	Phytozome
				Medtr2g063150.1	63,9	7,20E-12	256	1	Phytozome
				Medtr2g067970.1*	107,1	1,90E-25	895	1	Phytozome
				Medtr2g067980.1	162,2	2,40E-44	1033	2	Phytozome
				Medtr2g067990.1	129,8	4,00E-33	1006	2	Phytozome
				Medtr2g068030.1	148,7	1,20E-42	254	1	Phytozome
				Medtr2g068655.1	152,5	1,90E-43	309	1	Phytozome
				Medtr2g069660.1	136,3	5,30E-38	243	1	Phytozome
				Medtr3g068015.1	138,7	7,80E-39	249	1	Phytozome
				Medtr3g081550.1*	54,3	3,10E-09	115	1	Phytozome
				Medtr3g111620.1	156,8	9,60E-46	250	1	Phytozome
				Medtr3g114030.1	155,6	2,20E-45	237	1	Phytozome
				Medtr4g063630.1	126,3	3,70E-34	252	0	Phytozome
				Medtr4g073720.1	158,7	1,60E-45	344	2	Phytozome
				Medtr4g073730.3	164,9	6,00E-48	326	3	Phytozome
				Medtr5g010635.1	304,3	2,60E-103	241	0	Phytozome
				Medtr5g010640.1	246,5	1,00E-80	236	0	Phytozome
				Medtr5g022310.1	153,7	4,20E-44	300	2	Phytozome
				Medtr5g022350.1	159,1	7,20E-46	316	3	Phytozome
				Medtr5g023850.1*	92,4	6,30E-22	224	2	Phytozome
				Medtr5g059200.1	160,6	8,10E-47	282	2	Phytozome
				Medtr5g068670.1*	58,5	1,10E-10	132	0	Phytozome
				Medtr6g009480.1	125,2	1,40E-33	272	1	Phytozome
				Medtr6g079580.1	162,9	2,90E-48	240	1	Phytozome
				Medtr7g062610.1	134,4	2,90E-37	245	1	Phytozome
				Medtr7g076360.1*	76,3	9,90E-17	172	2	Phytozome
				Medtr7g102380.1	137,1	2,30E-38	243	1	Phytozome
				Medtr8g036215.1	155,6	1,80E-45	237	1	Phytozome
				Medtr8g037890.1	164,9	4,30E-48	317	0	Phytozome
				Medtr8g056820.1	159,5	1,20E-46	257	0	Phytozome
				Medtr8g075510.1	153,7	5,70E-44	308	2	Phytozome
				Medtr8g075550.1	152,9	1,10E-43	306	2	Phytozome
				Medtr8g088960.1*	40,4	1,70E-04	104	2	Phytozome
				Medtr8g089020.1*	82,8	4,60E-19	159	2	Phytozome
				Medtr8g096900.1	268,1	6,30E-89	248	0	Phytozome
				Medtr8g096910.1*	169,1	1,90E-51	176	0	Phytozome
				Medtr8g096920.1*	165,6	4,90E-50	173	0	Phytozome
				Medtr8g107140.1	152,9	2,70E-44	249	1	Phytozome
Angiosperm (eudicot)	Fabaceae	*Phaseolus vulgaris*	Pvu	Phvul.001G005000.1	163,7	8,10E-48	315	2	Phytozome
				Phvul.001G005100.1	130,6	2,00E-35	309	2	Phytozome
				Phvul.001G016600.1	151,4	6,60E-44	244	1	Phytozome
				Phvul.001G016700.1	157,1	5,80E-46	248	1	Phytozome
				Phvul.001G234000.1	139,8	1,80E-39	244	1	Phytozome
				Phvul.002G107800.1	163,7	1,10E-47	328	2	Phytozome
				Phvul.002G107900.1	164,1	8,90E-48	341	2	Phytozome
				Phvul.002G155400.1	258,8	6,70E-86	225	0	Phytozome
				Phvul.002G155500.1	299,3	1,80E-101	242	0	Phytozome
				Phvul.002G250400.1	150,2	6,50E-43	286	2	Phytozome
				Phvul.002G286500.1	258,5	9,40E-86	222	0	Phytozome
				Phvul.002G286600.1	276,9	4,70E-93	221	0	Phytozome
				Phvul.002G329400.1	160,2	3,20E-47	248	0	Phytozome
				Phvul.003G263400.1	157,5	1,40E-45	308	2	Phytozome
				Phvul.004G024600.1	124,4	1,40E-33	253	1	Phytozome
				Phvul.004G151550.1	167,5	5,70E-50	252	1	Phytozome
				Phvul.006G033200.1*	36,6	6,90E-03	169	1	Phytozome
				Phvul.008G118200.1	138,7	5,20E-39	245	1	Phytozome
				Phvul.008G166500.1	158,3	4,00E-46	286	2	Phytozome

				Phvul.009G017700.1	172,9	3,70E-51	330	2	Phytozome
				Phvul.009G040100.1	160,2	3,00E-47	251	1	Phytozome
				Phvul.011G034100.1	164,1	6,00E-48	310	2	Phytozome
				Phvul.011G034200.1	157,1	6,10E-45	348	2	Phytozome
				Phvul.011G065300.1	116,7	1,10E-30	248	0	Phytozome
Angiosperm (eudicot)	Solanaceae	*Nicotiana tabacum*	Nta	CAA43854	438	1,00E-157	245	0	NCBI
Angiosperm (eudicot)	Solanaceae	*Solanum nigrum*	Sni	AAL87640	-	-	247	0	NCBI
Angiosperm (eudicot)	Solanaceae	*Petunia x hybrida*	Phy	AAK55411	446	4,00E-163	246	0	NCBI
Angiosperm (mocot)	Poaceae	*Avena sativa*	Asa	AAB02259.1	220	2,00E-74	228	0	NCBI
Angiosperm (eudicot)	Rosaceae	*Prunus avium*	Pav	P50694	140	6,00E-43	245	-	NCBI
Angiosperm (eudicot)	Rosaceae	*Malus domestica*	Mdo	AAX19849	152	4,00E-45	246	1	NCBI

*protein ID sequences excluded in phylogenetic analyses

Phytozome 12.0 accessed 8 May 2017

Table S2 TLP structural information.

Sequences according tree position	Signal peptide	Transmembrane portion	Domains[1]	pI acid cleft
Spu_SapurV1A_0514s0200_1	YES	YES	THN	3.45
Spu_SapurV1A_0117s0110_1	YES	YES	THN	3.45
Mes_Manes_11G095900_1	YES	NO	THN	3.45
Gma_Glyma_12G031200_1	YES	NO	THN	3.45
Pvu_Phvul_011G034200_1	YES	NO	THN	3.45
Mtr_Medtr4g073720_1	YES	NO	THN	3.45
Egr_Eucgr_G01772_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400009202	YES	NO	THN	3.45
Sly_Solyc02g083760_2_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400015871	YES	NO	THN	3.45
Sly_Solyc01g111330_2_1	YES	NO	THN	3.45
Mgu_Migut_B00098_1	NO	YES (2)	THN	3.45
Aly_AL7G10550_t1	YES	NO	THN	3.45
Bra_Brara_H01849_1	YES	YES	THN	3.45
Cgr_Cagra_1383s0050_1	YES	NO	THN	3.45
Ath_AT4G38660_1	YES	NO	THN	3.45
Spo_Spipo1G0065600	YES	NO	THN	3.45
Egr_Eucgr_I02339_1	NO	NO	THN	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00041_116	YES	YES	THN	3.45
Aco_Aqcoe6G212100_1	YES	NO	THN	3.45
Gma_Glyma_16G126900_1	YES	NO	THN	3.45
Gma_Glyma_02G047400_1	YES	YES	THN	3.45
Pvu_Phvul_003G263400_1	YES	NO	THN	3.45
Mtr_Medtr8g075510_1	NO	YES	THN	3.45
Gma_Glyma_01G165400_1	YES	YES	THN	3.45
Gma_Glyma_11G077800_1	YES	NO	THN	3.45
Pvu_Phvul_002G107800_1	NO	YES	THN	3.45
Mtr_Medtr5g022350_1	NO	YES	THN	3.83
Mgu_Migut_L00673_1	YES	NO	THN	3.45
Mgu_Migut_L00700_1	YES	NO	THN	3.45
Mes_Manes_02G105400_1	YES	YES	THN	3.45
Mes_Manes_01G146900_1	YES	YES	THN	3.45
Spu_SapurV1A_0130s0410_1	YES	NO	THN	3.45
Egr_Eucgr_J02446_1	NO	YES (2)	THN	3.45
Mes_Manes_18G012600_1	NO	YES	THN	3.45
Mes_Manes_05G146000_1	NO	YES	THN	3.45
Egr_Eucgr_H00974_1	YES	NO	THN	3.83
Spu_SapurV1A_0025s0190_1	YES	YES	THN	3.45
Spu_SapurV1A_4851s0010_1	YES	YES	THN	3.45
Spu_SapurV1A_0384s0050_1	NO	NO	THN	3.45
Mgu_Migut_N03166_1	NO	NO	THN	3.45
Mgu_Migut_C00850_1	NO	YES	THN	3.45
Gma_Glyma_17G258600_1	YES	NO	THN	3.45
Pvu_Phvul_001G005100_1	YES	NO	THN	3.45
Ath_AT4G24180_1	YES	NO	THN	3.45
Aly_AL7G29100_t2	NO	YES	THN	3.45
Cgr_Cagra_1226s0084_1	YES	NO	THN	3.45
Bra_Brara_K00744_1	NO	NO	THN	3.45
Bra_Brara_A01423_1	NO	NO	THN	3.45

Sly_Solyc04g081560_2_1	YES	YES	THN	3.45
Zma_GRMZM2G138896_T03	YES	NO	THN	3.45
Zma_GRMZM2G049057_T02	YES	NO	THN	3.45
Sit_Seita_6G239100_1	YES	YES	THN	3.45
Osa_LOC_Os08g43510_1	YES	NO	THN	3.45
Bdi_Bradi3g42380_2	YES	NO	THN	3.45
Sit_Seita_2G289100_1	YES	NO	THN	3.45
Zma_GRMZM2G154449_T01	YES	NO	THN	3.45
Osa_LOC_Os09g36580_1	NO	NO	THN	3.45
Bdi_Bradi4g36410_1	YES	NO	THN	3.45
Sit_Seita_9G288500_1	YES	NO	THN	3.45
Zma_GRMZM2G148536_T03	YES	NO	THN	3.45
Zma_GRMZM2G151589_T01	YES	NO	THN	3.45
Osa_LOC_Os10g05600_1	YES	NO	THN	3.45
Bdi_Bradi3g21100_1	YES	NO	THN	3.45
Sit_Seita_9G471900_1	YES	NO	THN	3.45
Zma_GRMZM2G346861_T02	YES	NO	THN	3.45
Zma_GRMZM2G149798_T01	YES	NO	THN	3.45
Bdi_Bradi1g68330_1	YES	NO	THN	3.45
Osa_LOC_Os03g14030_1	YES	NO	THN	3.45
Zma_GRMZM2G038490_T01	NO	NO	THN	3.45
Sit_Seita_2G289000_1	YES	NO	THN	3.45
Bdi_Bradi4g36400_1	YES	NO	THN	3.45
Osa_LOC_Os09g36560_1	YES	NO	THN	3.45
Bdi_Bradi1g68340_1	YES	NO	THN	3.45
Sit_Seita_9G471800_1	YES	NO	THN	3.45
Zma_GRMZM2G149809_T01	YES	NO	THN	3.45
Zma_GRMZM2G159110_T01	YES	NO	THN	3.45
Sit_Seita_9G288400_1	YES	YES	THN	3.45
Zma_GRMZM2G477139_T01	NO	NO	THN	3.45
Osa_LOC_Os10g05660_1	YES	NO	THN	3.45
Zma_GRMZM2G541730_T01	YES	NO	THN	3.45
Sit_Seita_4G281400_1	YES	NO	THN	3.45
Osa_LOC_Os06g50240_1	YES	NO	THN	3.45
Bdi_Bradi1g30117_1	YES	NO	THN	3.45
Sit_Seita_1G148000_1	YES	NO	THN	3.45
Zma_GRMZM2G036826_T01	YES	NO	THN	3.45
Sit_Seita_1G148300_1	YES	NO	THN	3.45
Zma_GRMZM2G086410_T01	YES	YES	THN	3.66
Sit_Seita_1G157500_1	YES	NO	THN	3.45
Sit_Seita_1G148900_1	YES	NO	THN	3.45
Zma_GRMZM2G023655_T01	YES	NO	THN	3.45
Spo_Spipo1G0065500	YES	NO	THN	3.45
Spo_Spipo11G0042300	YES	NO	THN	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00041_117	NO	NO	THN	3.45
Mgu_Migut_L00502_1	NO	YES (2)	THN; S_TKc	3.02
Mgu_Migut_L00322_1	NO	NO	THN	3.63
Pvu_Phvul_009G040100_1	YES	NO	THN	3.45
Gma_Glyma_06G139000_1	NO	NO	THN	3.45
Gma_Glyma_04G225800_1	NO	NO	THN	2.97
Pvu_Phvul_008G188000_1	YES	YES	THN	3.45

Gma_Glyma_02G220900_1	NO	YES	THN	3.45
Gma_Glyma_14G188400_1	YES	NO	THN	3.45
Egr_Eucgr_K00478_1	YES	NO	THN	3.45
Aco_Aqcoe7G295800_1	YES	NO	THN	3.45
Spu_SapurV1A_3604s0020_1	YES	NO	THN	3.45
Spu_SapurV1A_1518s0060_1	YES	NO	THN	3.45
Spu_SapurV1A_1668s0040_1	YES	NO	THN	3.45
Mes_Manes_01G050400_1	YES	NO	THN	3.45
Mgu_Migut_F00124_1	YES	YES	THN	3.45
Sly_Solyc03g079960_2_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400001544	YES	NO	THN	3.45
Sly_Solyc06g073000_2_1	NO	YES	THN	3.45
Stu_PGSC0003DMT400069224	NO	NO	THN	3.45
Bra_Brara_B03671_1	YES	NO	THN; 2 RPT1	3.45
Bra_Brara_F02676_1	YES	NO	THN; 2 RPT1	3.45
Cgr_Cagra_9017s0003_1	YES	NO	THN; 2 RPT1	3.45
Ath_AT5G24620_1	YES	NO	THN; 2 RPT1	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00025_195	YES	NO	THN	3.45
Zma_GRMZM2G340534_T01	YES	NO	THN	3.45
Sit_Seita_2G260400_1	YES	NO	THN	3.45
Osa_LOC_Os09g32280_1	YES	YES	THN	3.45
Bdi_Bradi4g34180_1	NO	NO	THN	3.45
Sit_Seita_6G214000_1	YES	NO	THN	3.45
Osa_LOC_Os08g40600_1	YES	NO	THN; 2 RPT1	3.45
Bdi_Bradi3g40596_2	NO	NO	THN; 2 RPT1	3.45
Spo_Spipo28G0022500	YES	NO	THN	3.45
Pta_PITA_000087079	YES	YES	THN	3.45
Pab_MA_132199g0010	YES	YES	THN	3.45
Pab_MA_133864g0010	YES	YES	THN	3.36
Egr_Eucgr_J02440_1	YES	NO	THN	3.45
Egr_Eucgr_J02439_1	YES	NO	THN	3.45
Egr_Eucgr_J02442_1	NO	NO	THN	3.45
Egr_Eucgr_J02445_1	NO	NO	THN	3.45
Egr_Eucgr_J02443_1	NO	NO	THN	3.5
Egr_Eucgr_L02491_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400009196	YES	YES	THN	3.45
Sly_Solyc02g083790_2_1	YES	YES	THN	3.45
Sly_Solyc03g033490_1_1	YES	NO	THN	3.45
Mgu_Migut_L00701_1	YES	NO	THN	3.45
Mgu_Migut_L00960_1	NO	NO	THN	3.45
Gma_Glyma_11G077700_1	YES	YES	THN	3.45
Gma_Glyma_01G165600_1	YES	YES	THN	3.45
Pvu_Phvul_002G107900_1	NO	YES	THN	3.45
Mtr_Medtr8g037890_1	YES	NO	THN	3.42
Mtr_Medtr8g056820_1	YES	NO	THN	3.45
Mtr_Medtr5g022310_1	YES	NO	THN	3.45
Gma_Glyma_16G127100_1	YES	NO	THN	3.45
Mtr_Medtr8g075550_1	YES	NO	THN	3.45
Mes_Manes_01G146800_1	YES	NO	THN	3.45
Mes_Manes_02G105300_1	YES	YES	THN	3.45

Aco_Aqcoe6G211700_1	YES	YES	THN	3.45
Spu_SapurV1A_0130s0420_1	NO	NO	THN	3.45
Aly_AL3G50900_t1	YES	NO	THN	3.45
Cgr_Cagra_3126s0018_1	YES	NO	THN	3.45
Ath_AT2G17860_1	YES	NO	THN	3.45
Bra_Brara_G00232_1	YES	NO	THN	3.45
Ath_AT4G36010_1	YES	NO	THN	3.45
Aly_AL7G15040_t1	YES	NO	THN	3.45
Cgr_Cagra_2236s0010_1	YES	NO	THN	3.45
Bra_Brara_H01620_1	YES	NO	THN	3.45
Gma_Glyma_11G106100_1	YES	YES	THN	3.83
Gma_Glyma_12G031000_1	YES	YES	THN	3.83
Pvu_Phvul_011G034100_1	YES	NO	THN	3.83
Mtr_Medtr4g073730_3	YES	YES	THN	3.83
Aly_AL7G10560_t1	YES	NO	THN	3.83
Ath_AT4G38670_1	YES	NO	THN	3.83
Cgr_Cagra_1383s0051_1	YES	NO	THN	3.83
Bra_Brara_A00011_1	YES	NO	THN	3.83
Spu_SapurV1A_0514s0180_1	YES	YES	THN	3.83
Mes_Manes_04G072700_1	YES	NO	THN	3.83
Egr_Eucgr_I01396_1	YES	YES	THN	3.83
Sly_Solyc01g111320_2_1	YES	NO	THN	3.83
Stu_PGSC0003DMT400015872	YES	NO	THN	3.83
Mgu_Migut_B00099_1	YES	NO	THN	3.83
Gma_Glyma_04G023700_1	YES	NO	THN	3.83
Gma_Glyma_06G023900_1	YES	NO	THN	3.83
Pvu_Phvul_009G017700_1	YES	NO	THN	3.83
Mtr_Medtr1g025420_1	YES	NO	THN	3.83
Mtr_Medtr3g114030_1	NO	NO	THN	3.83
Pvu_Phvul_001G005000_1	NO	NO	THN	3.83
Gma_Glyma_17G258500_1	YES	NO	THN	3.83
Gma_Glyma_14G219600_1	YES	NO	THN	3.83
Gma_Glyma_15G275200_1	NO	NO	THN	3.8
Sly_Solyc04g081550_2_1	YES	NO	THN	3.83
Stu_PGSC0003DMT400025547	YES	NO	THN	3.83
Egr_Eucgr_F01670_1	YES	NO	THN	3.45
Egr_Eucgr_H00975_1	YES	NO	THN	3.45
Mes_Manes_05G146200_1	YES	NO	THN	3.83
Mes_Manes_18G012700_1	NO	YES (2)	THN	3.83
Spu_SapurV1A_0384s0040_1	YES	NO	THN	3.83
Spu_SapurV1A_0025s0200_1	NO	NO	THN	3.83
Mgu_Migut_C00849_1	YES	NO	THN	3.83
Bra_Brara_G03379_1	YES	NO	THN	3.45
Bra_Brara_B02213_1	YES	NO	THN	3.45
Aly_AL2G35790_t1	YES	NO	THN	3.45
Ath_AT1G75800_1	YES	NO	THN	3.45
Ath_AT1G20030_2	YES	NO	THN	3.83
Aly_AL1G32810_t1	NO	NO	THN	3.83
Cgr_Cagra_1961s0004_1	YES	NO	THN	3.83
Mtr_Medtr2g068030_1	YES	NO	THN	4.03
Mtr_Medtr2g067980_1	YES	YES	THN; B lectin; S_TKc	3.83

Mtr_Medtr2g067990_1	YES	YES	THN; B lectin; S_TKc	4.2
Mtr_Medtr2g068655_1	YES	YES	THN	6.47
Gma_Glyma_04G176200_1	NO	NO	THN	4.36
Gma_Glyma_06G189100_1	YES	YES	THN	8.75
Spu_SapurV1A_1323s0020_1	YES	YES	THN; S_TKc	3.82
Spu_SapurV1A_1323s0060_1	YES	YES	THN; S_TKc	4.03
Aly_AL7G36160_t1	NO	YES	2 THN; Pkinase	5.03
Ath_AT4G18250_1	NO	YES	2 THN; Pkinase	3.64
Cgr_Cagra_1595s0025_1	YES	YES	THN; S_TKc	3.29
Bra_Brara_A00940_1	YES	NO	THN	4.93
Cgr_Cagra_1595s0024_1	YES	YES	THN; S_TKc	3.79
Bra_Brara_K00455_1	NO	YES	THN; Pkinase	3.95
Bra_Brara_J01946_1	YES	YES	THN; S_TKc	3.66
Aly_AL7G52530_t1	YES	YES	THN; S_TKc	3.45
Ath_AT5G38280_1_PR5K	YES	YES	THN; Pkinase	3.45
Cgr_Cagra_3138s0010_1	YES	YES	THN; S_TKc	3.5
Cgr_Cagra_3138s0011_1	YES	NO	THN	3.36
Aly_AL7G52540_t1	YES	YES	THN; S_TKc	3.45
Bra_Brara_D02646_1	YES	YES	THN; S_TKc	3.13
Cgr_Cagra_17854s0001_1	YES	YES	THN; S_TKc	3.8
Cgr_Cagra_3706s0043_1	YES	YES	THN; S_TKc	3.8
Aly_AL4G42380_t1	YES	YES	THN; S_TKc	3.13
Ath_AT1G70250_1	NO	NO	LPT_2; THN; Pkinase	3.8
Aly_AL2G29540_t2	YES	YES	AAI; THN; S_TKc	3.5
Gma_Glyma_10G060800_1	NO	NO	THN	3.83
Gma_Glyma_10G061000_1	YES	NO	THN	3.83
Gma_Glyma_10G061800_1	NO	NO	THN	3.83
Gma_Glyma_10G061700_1	YES	NO	THN	3.83
Gma_Glyma_10G060300_1	YES	NO	THN	3.83
Mtr_Medtr1g062590_1	YES	NO	THN	3.83
Mtr_Medtr1g062350_1	YES	NO	THN	3.83
Mtr_Medtr1g062370_1	YES	NO	THN	3.83
Mtr_Medtr1g062340_1	YES	NO	THN	3.83
Mtr_Medtr1g062660_1	YES	NO	THN	3.83
Mtr_Medtr1g062630_1	YES	NO	THN	3.59
Mdo_AAX19849_MalD2	YES	NO	THN	3.83
Mtr_Medtr7g102380_1	YES	NO	THN	3.83
Pav_P50694_PruAv2	YES	NO	THN	4.18
Egr_Eucgr_E01384_1	YES	NO	THN	3.45
Egr_Eucgr_E01382_1	YES	NO	THN	3.45
Egr_Eucgr_E01389_1	YES	NO	THN	3.45
Egr_Eucgr_E01385_1	YES	NO	THN	3.45
Egr_Eucgr_E01381_1	YES	NO	THN	3.45
Egr_Eucgr_E01392_1	YES	NO	THN	3.83
Egr_Eucgr_A01612_1	YES	NO	THN	3.83
Mes_Manes_02G199600_1	YES	NO	THN	3.83
Spu_SapurV1A_0977s0010_1	YES	NO	THN	3.83
Spu_SapurV1A_1658s0020_1	YES	NO	THN	3.83
Spu_SapurV1A_1469s0070_1	YES	NO	THN	3.83
Spu_SapurV1A_0519s0110_1	YES	NO	THN	3.45
Spu_SapurV1A_1658s0010_1	YES	NO	THN	3.83
Spu_SapurV1A_0977s0020_1	YES	NO	THN	3.83

Spu_SapurV1A_1469s0040_1	YES	NO	THN	3.69
Spu_SapurV1A_1469s0050_1	NO	NO	THN	3.45
Spu_SapurV1A_0977s0040_1	YES	NO	THN	3.83
Gma_Glyma_15G254700_1	NO	NO	THN	3.83
Gma_Glyma_08G172700_1	NO	NO	THN	4.03
Mtr_Medtr2g069660_1	YES	NO	THN	3.83
Mes_Manes_S049600_1	NO	NO	THN	3.45
Mgu_Migut_J00460_1	YES	NO	THN	3.83
Mgu_Migut_J00461_1	YES	NO	THN	3.83
Mgu_Migut_J00463_1	YES	NO	THN	3.92
Mgu_Migut_J00462_1	YES	NO	THN	3.83
Sly_Solyc07g017970_1_1	NO	NO	THN	3.42
Osa_LOC_Os07g23470_1	YES	NO	THN	3.45
Sit_Seita_2G113200_1	YES	NO	THN	3.83
Sit_Seita_2G113300_1	NO	NO	THN	3.45
Sit_Seita_2G114300_1	YES	NO	THN	3.83
Osa_LOC_Os07g23730_1	YES	NO	THN	3.83
Aco_Aqcoe1G490800_1	NO	NO	THN	3.83
Aco_Aqcoe1G125000_1	YES	NO	THN	3.83
Aco_Aqcoe1G490700_1	NO	NO	THN	6.56
Aco_Aqcoe1G134200_1	YES	NO	THN	3.83
Aco_Aqcoe1G133900_1	YES	NO	THN	3.83
Aco_Aqcoe1G133800_1	YES	NO	THN	3.83
Aco_Aqcoe1G125400_1	YES	NO	THN	3.83
Aco_Aqcoe5G259000_1	NO	NO	THN	3.83
Aco_Aqcoe1G125100_1	NO	NO	THN	3.83
Zma_GRMZM2G100747_T02	YES	NO	THN	2.95
Zma_GRMZM2G038846_T01	YES	NO	THN	2.95
Sit_Seita_7G019500_1	NO	NO	THN	2.95
Bdi_Bradi5g00550_1	YES	NO	THN	2.95
Spo_Spipo1G0019200	YES	NO	THN	3.45
Gma_Glyma_14G077400_1	YES	NO	THN	3.45
Gma_Glyma_17G248300_1	YES	NO	THN	3.45
Mtr_Medtr1g021945_1	YES	NO	THN	3.45
Gma_Glyma_14G077300_1	YES	NO	THN	3.45
Pvu_Phvul_001G016600_1	YES	NO	THN	3.83
Gma_Glyma_04G034300_1	YES	NO	THN	3.45
Mtr_Medtr3g111620_1	YES	NO	THN	3.45
Pvu_Phvul_001G016700_1	YES	NO	THN	3.45
Spu_SapurV1A_0715s0040_1	NO	YES	THN	3.45
Mes_Manes_18G030000_1	YES	NO	THN	3.45
Mgu_Migut_C00688_1	YES	NO	THN	3.45
Mgu_Migut_N00519_1	NO	NO	THN	3.45
Aco_Aqcoe6G108900_1	YES	NO	THN	3.45
Spo_Spipo2G0113900	NO	NO	THN	3.45
Sly_Solyc04g079890_2_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400043496	NO	NO	THN	3.45
Egr_Eucgr_A00487_1	YES	NO	THN	3.45
Sit_Seita_9G483200_1	YES	NO	THN	3.45
Zma_GRMZM2G393507_T01	YES	NO	THN	3.45
Bdi_Bradi1g69277_2	YES	NO	THN	3.45
Osa_LOC_Os03g13070_1	YES	NO	THN	3.45
Mgu_Migut_B01335_1	YES	NO	THN	3.83
Mgu_Migut_B01336_1	NO	NO	THN	3.83

Ath_AT1G75030_1	YES	NO	THN	3.45
Aly_AL2G34910_t1	YES	NO	THN	3.45
Cgr_Cagra_0402s0026_1	YES	NO	THN	3.45
Bra_Brara_B02149_1	YES	NO	THN	3.45
Bra_Brara_G03322_1	NO	NO	THN	3.59
Aly_AL2G34940_t1	YES	NO	THN	3.45
Ath_AT1G75050_1	YES	NO	THN	3.45
Cgr_Cagra_0402s0024_1	YES	NO	THN	3.45
Bra_Brara_G03323_1	YES	NO	THN	3.45
Aly_AL2G34930_t1	YES	NO	THN	3.42
Ath_AT1G75040_1	YES	NO	THN	3.42
Cgr_Cagra_0402s0025_1	YES	NO	THN	3.42
Bra_Brara_G03321_1	YES	NO	2 THN	3.45
Bra_Brara_F01364_1	YES	NO	THN	3.45
Bra_Brara_F01366_1	NO	NO	THN	3.83
Bra_Brara_H02374_1	NO	NO	THN	3.45
Bra_Brara_I04743_1	YES	NO	THN	3.45
Ath_AT1G19320_1	YES	NO	THN	3.45
Aly_AL1G31910_t1	YES	NO	THN	3.45
Cgr_Cagra_1961s0069_1	YES	NO	THN	3.45
Pta_PITA_000042205	NO	NO	2 THN	3.45
Pta_PITA_000074950	NO	NO	THN	3.45
Pta_PITA_000008671	YES	NO	THN	3.45
Pab_MA_19953g0020	NO	NO	THN	3.45
Pta_PITA_000012478	YES	NO	THN	3.45
Pta_PITA_000039064	YES	NO	THN	3.45
Pta_PITA_000093129	NO	NO	THN	3.45
Pta_PITA_000042353	YES	NO	THN	3.83
Gma_Glyma_12G238900_1	YES	NO	THN	3.59
Mtr_Medtr8g036215_1	YES	NO	THN	3.42
Mes_Manes_12G003800_1	YES	NO	THN	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00022_297	YES	NO	THN	3.45
Stu_PGSC0003DMT400065004	NO	NO	THN	3.45
Sly_Solyc01g086840_1_1	YES	NO	THN	3.45
Mgu_Migut_H02015_1	YES	NO	THN	3.45
Mgu_Migut_K00892_1	YES	NO	THN	4.49
Sfa_Sphfalx0073s0078_1	NO	NO	THN	3.45
Sfa_Sphfalx0092s0064_1	YES	NO	THN	3.45
Sfa_Sphfalx0075s0069_1	YES	YES	THN	3.45
Sfa_Sphfalx0259s0005_1	NO	NO	THN	3.45
Ppa_Pp3c16_17280V3_1	YES	YES	THN	3.45
Ppa_Pp3c25_4860V3_1	YES	NO	THN	3.45
Ppa_Pp3c6_11450V3_1	YES	NO	THN	3.45
Sfa_Sphfalx0006s0097_1	NO	NO	THN	3.94
Smo_419455	NO	NO	THN	3.45
Spu_SapurV1A_0010s0800_1	YES	NO	THN	3.45
Spu_SapurV1A_0010s0790_1	YES	NO	THN	3.45
Spu_SapurV1A_0010s0780_1	YES	NO	THN	3.45
Spu_SapurV1A_0105s0010_1	YES	NO	THN	3.45
Spu_SapurV1A_1265s0060_1	YES	NO	THN	3.45
Egr_Eucgr_F03757_1	YES	YES	THN	3.45
Mes_Manes_02G160200_1	YES	NO	THN	3.45

Aly_AL2G37920_t1	YES	NO	THN	3.45
Ath_AT1G77700_1	NO	YES	THN	3.45
Cgr_Cagra_0096s0064_1	YES	NO	THN	3.45
Bra_Brara_G03527_1	YES	NO	THN	3.45
Bra_Brara_B02282_1	YES	NO	THN	3.45
Bra_Brara_G02134_1	YES	NO	THN	3.45
Sly_Solyc11g013300_1_1	YES	YES	THN	3.45
Stu_PGSC0003DMT400016741	NO	NO	THN	3.45
Aco_Aqcoe5G278000_1	YES	NO	THN	3.45
Bdi_Bradi5g27280_3	YES	NO	THN	3.45
Osa_LOC_Os04g59370_1	YES	NO	THN	3.45
Sit_Seita_3G003900_1	YES	NO	THN	3.45
Zma_GRMZM2G125918_T02	YES	NO	THN	3.45
Bdi_Bradi3g04330_2	NO	YES	THN	3.45
Sit_Seita_6G194300_1	NO	NO	THN	3.83
Mtr_Medtr5g059200_1	YES	NO	THN	3.45
Gma_Glyma_13G082700_1	YES	YES	THN	3.45
Pvu_Phvul_008G166500_1	YES	NO	THN	3.45
Gma_Glyma_14G163700_1	YES	YES	THN	3.45
Mgu_Migut_N03301_1	YES	NO	THN	3.45
Spo_Spipo9G0041100	YES	NO	THN	3.45
Mes_Manes_02G160300_1	YES	NO	THN	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00017_121	NO	NO	THN	3.59
Atr_evm_27_model_AmTr_v1_0_scaffold00017_123	YES	NO	THN	3.45
Pvu_Phvul_002G250400_1	YES	NO	THN	3.5
Gma_Glyma_05G169700_1	YES	NO	THN	3.39
Gma_Glyma_08G128000_1	YES	NO	THN	3.39
Sly_Solyc01g104290_1_1	NO	NO	THN	3.45
Stu_PGSC0003DMT400061140	YES	NO	THN	3.45
Mgu_Migut_B00845_1	YES	NO	THN	3.45
Mes_Manes_10G130700_1	YES	YES	THN	3.83
Mes_Manes_07G014800_1	NO	NO	THN	3.45
Egr_Eucgr_G00607_1	YES	YES	THN	4.53
Pta_PITA_000033863	YES	NO	THN	3.83
Pta_PITA_000041401	YES	NO	THN	3.45
Pab_MA_10432704g0010	YES	NO	THN	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00059_132	YES	NO	THN	3.76
Aly_AL1G30620_t1	YES	NO	THN	3.45
Ath_AT1G18250_2	YES	NO	THN	3.45
Cgr_Cagra_0909s0003_1	YES	NO	THN	3.45
Bra_Brara_F01282_1	YES	NO	THN	3.45
Aly_AL2G33400_t1	YES	NO	THN	3.45
Ath_AT1G73620_1	YES	NO	THN	3.45
Bra_Brara_G03254_1	YES	NO	THN	3.45
Cgr_Cagra_2913s0003_1	NO	NO	THN	3.45
Sly_Solyc03g118780_2_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400014747	YES	NO	THN	3.45
Mgu_Migut_D01556_1	YES	NO	THN	3.45
Aco_Aqcoe3G011200_1	YES	NO	THN	3.45
Mes_Manes_14G170800_1	YES	NO	THN	3.45
Mes_Manes_06G007400_1	YES	NO	THN	3.45

Spu_SapurV1A_1146s0050_1	YES	NO	THN	3.45
Spu_SapurV1A_2698s0020_1	YES	NO	THN	3.45
Egr_Eucgr_B00944_1	YES	NO	THN	3.45
Gma_Glyma_05G245800_1	YES	NO	THN	3.45
Pvu_Phvul_002G329400_1	YES	NO	THN	3.45
Gma_Glyma_08G053600_1	YES	NO	THN	3.45
Mtr_Medtr8g107140_1	YES	NO	THN	3.45
Pvu_Phvul_004G151550_1	YES	NO	THN	3.45
Mtr_Medtr6g079580_1	YES	NO	THN	3.45
Sit_Seita_4G252200_1	YES	NO	THN	3.45
Zma_GRMZM2G476523_T01	YES	NO	THN	3.45
Osa_LOC_Os06g47600_1	YES	NO	THN	3.45
Bdi_Bradi1g33540_1	YES	NO	THN	3.45
Spo_Spipo20G0017700	YES	NO	THN	3.45
Atr_evm_27_model_AmTr_v1_0_scaffold00010_64	YES	NO	THN	3.45
Pta_PITA_000051863	NO	NO	THN	3.45
Pta_PITA_000080209	YES	NO	THN	3.45
Cgr_Cagra_18132s0003_1	NO	NO	THN	3.66
Bra_Brara_D01016_1	YES	NO	THN	3.66
Aly_AL7G49980_t1	YES	NO	THN	3.66
Ath_AT5G40020_1	YES	NO	THN	3.8
Mtr_Medtr6g009480_1	NO	YES	THN	3.59
Aco_Aqcoe3G317400_1	YES	NO	THN	3.66
Mes_Manes_15G175600_1	YES	NO	THN	3.66
Spu_SapurV1A_2569s0010_1	YES	NO	THN	3.66
Atr_evm_27_model_AmTr_v1_0_scaffold00002_295	NO	NO	THN	4.02
Mgu_Migut_H00882_1	NO	YES	THN	3.59
Stu_PGSC0003DMT400003284	YES	NO	THN	3.59
Sly_Solyc02g087520_2_1	YES	NO	THN	3.59
Pvu_Phvul_004G024600_1	YES	NO	THN	3.59
Gma_Glyma_19G018400_1	YES	NO	THN	3.59
Egr_Eucgr_F03937_1	NO	NO	THN	5.03
Egr_Eucgr_F03940_1	NO	NO	THN	7.36
Egr_Eucgr_F03936_1	YES	NO	THN	5.03
Egr_Eucgr_I01898_1	YES	NO	THN	3.66
Sit_Seita_9G261300_1	YES	NO	THN	3.59
Zma_GRMZM2G064605_T01	YES	NO	THN	3.59
Osa_LOC_Os10g27280_1	YES	NO	THN	4.03
Bdi_Bradi3g26630_2	YES	NO	THN	3.59
Spo_Spipo2G0061400	NO	YES	THN	4.03
Smo_404034	YES	NO	THN	3.45
Smo_404385	YES	NO	THN	3.45
Smo_271835	YES	NO	THN	3.45
Smo_426298	YES	NO	THN	3.45
Smo_402999	YES	NO	THN	3.45
Smo_236128	YES	NO	THN	3.45
Smo_422821	YES	NO	THN	3.85
Smo_81648	NO	NO	THN	8.76
Smo_230073	NO	NO	THN	6.13
Smo_410984	NO	NO	THN	4.29
Smo_105934	NO	NO	THN	3.82

Stu_PGSC0003DMT400007869_OSM	YES	NO	THN	3.45
Sly_Solyc08g080640_1_1	YES	NO	THN	3.45
Phy_AAK55411(AF376058)_osmotin	YES	NO	THN	3.45
Stu_PGSC0003DMT400007870	YES	NO	THN	3.45
Sly_Solyc08g080650_1_1	YES	NO	THN	3.45
Sly_Solyc08g080620_1_1	YES	NO	THN	3.45
Nta_CAA43854(X61679)_OSM	YES	NO	THN	3.45
Stu_PGSC0003DMT400007868	YES	NO	THN	3.45
Stu_PGSC0003DMT400007865	YES	NO	THN	3.45
Sni_AAL87640(AF450276)_SnOLP	YES	NO	THN	2.95
Stu_PGSC0003DMT400034395	YES	NO	THN	3.53
Stu_PGSC0003DMT400007905	YES	NO	THN	3.45
Sly_Solyc08g080660_1_1	YES	NO	THN	3.45
Sly_Solyc08g080670_1_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400007906	YES	NO	THN	3.45
Egr_Eucgr_D01888_1	YES	NO	THN	3.45
Sly_Solyc12g056390_1_1	NO	NO	THN	3.45
Stu_PGSC0003DMT400010886	NO	NO	THN	3.45
Mgu_Migut_E01125_1	YES	NO	THN	3.45
Stu_PGSC0003DMT400010890	YES	NO	THN;THN	3.45
Spu_SapurV1A_2357s0010_1	YES	NO	THN	3.45
Spu_SapurV1A_0507s0110_1	YES	NO	THN	3.45
Mes_Manes_02G025100_1	YES	NO	THN	3.83
Mes_Manes_02G028300_1	YES	NO	THN	3.83
Egr_Eucgr_D01887_1	YES	NO	THN	3.45
Egr_Eucgr_D01893_1	YES	NO	THN	4.03
Egr_Eucgr_D01899_1	NO	NO	THN	4.03
Spu_SapurV1A_0507s0100_1	YES	NO	THN	3.45
Spu_SapurV1A_2357s0020_1	YES	NO	THN	3.45
Spu_SapurV1A_0098s0100_1	YES	NO	THN	3.45
Spu_SapurV1A_3598s0010_1	YES	NO	THN	3.45
Pvu_Phvul_002G155500_1	YES	NO	THN	3.45
Gma_Glyma_01G217700_1_GmOLPb	NO	NO	THN	3.45
Mtr_Medtr5g010635_1	YES	NO	THN	3.45
Mes_Manes_01G064300_1	YES	NO	THN	3.45
Mes_Manes_01G064200_1	YES	NO	THN	3.45
Mes_Manes_01G064400_1	YES	NO	THN	3.45
Ath_AT4G11650_1_OSM	YES	NO	THN	3.45
Aly_AL6G46240_t1	YES	NO	THN	3.45
Cgr_Cagra_1912s0003_1	YES	NO	THN	3.45
Aly_AL6G46230_t1	YES	NO	THN	3.45
Cgr_Cagra_1912s0002_1	YES	NO	THN	3.45
Bra_Brara_B02758_1	YES	NO	THN	3.45
Mgu_Migut_E01123_1	YES	NO	THN	3.45
Aco_Aqcoe7G040700_1	NO	NO	THN	3.45
Aco_Aqcoe7G040600_1	YES	NO	THN	3.45
Aco_Aqcoe3G114200_1	YES	NO	THN;RPT1	3.45
Aco_Aqcoe7G040400_1	YES	NO	THN	3.76
Atr_evm_27_model_AmTr_v1_0_scaffold00032_	YES	NO	THN	3.83
Atr_evm_27_model_AmTr_v1_0_scaffold00032_	NO	NO	THN	3.45

Atr_evm_27_model_AmTr_v1_0_scaffold00032_	YES	NO	THN	4.29
Egr_Eucgr_D01904_1	YES	NO	THN	3.83
Egr_Eucgr_L02566_1	YES	NO	THN	3.83
Egr_Eucgr_D01892_1	YES	NO	THN	3.83
Egr_Eucgr_H03864_1	YES	NO	THN	3.83
Egr_Eucgr_D01898_1	YES	NO	THN	4.03
Egr_Eucgr_H03865_1	YES	NO	THN	3.83
Egr_Eucgr_L02568_1	NO	NO	THN	3.83
Egr_Eucgr_H03863_1	YES	NO	THN	3.83
Egr_Eucgr_D01894_1	YES	NO	THN	3.45
Egr_Eucgr_D01900_1	YES	NO	THN	3.83
Egr_Eucgr_L03623_1	YES	NO	THN	3.83
Egr_Eucgr_L01962_1	YES	NO	THN	3.83
Spu_SapurV1A_0183s0030_1	YES	NO	THN	3.76
Spu_SapurV1A_0271s0330_1	YES	NO	THN	3.76
Spu_SapurV1A_0271s0340_1	YES	NO	THN	3.45
Spu_SapurV1A_0183s0040_1	YES	NO	THN	3.83
Spu_SapurV1A_0271s0350_1	YES	NO	THN	3.83
Spu_SapurV1A_0183s0050_1	YES	NO	THN	3.83
Pvu_Phvul_002G286500_1	YES	NO	THN	3.83
Pvu_Phvul_002G286600_1	YES	NO	THN	3.83
Gma_Glyma_05G204600_1_P21	YES	NO	THN	3.83
Gma_Glyma_05G204800_1	YES	NO	THN	3.83
Spo_Spipo32G0003200	YES	NO	THN	3.83
Aco_Aqcoe1G274500_1	NO	NO	THN	4.93
Aco_Aqcoe1G274400_1	YES	NO	THN	4.93
Aco_Aqcoe3G114100_1	NO	YES	THN	3.45
Mes_Manes_02G028400_1	YES	NO	THN	3.42
Bra_Brara_A01323_1	NO	NO	THN	4.18
Aly_AL7G30420_t1	YES	NO	THN	3.83
Cgr_Cagra_15158s0009_1	YES	NO	THN	3.83
Spu_SapurV1A_0761s0080_1	YES	NO	THN	3.83
Mes_Manes_02G025200_1	YES	NO	THN	3.83
Egr_Eucgr_E00560_1	YES	NO	THN	3.83
Egr_Eucgr_E00561_1	NO	NO	THN	4.29
Mtr_Medtr8g096900_1	NO	YES	THN	3.83
Sly_Solyc12g056360_1_1	YES	NO	THN	3.83
Stu_PGSC0003DMT400010891	NO	NO	THN	3.53
Sly_Solyc12g056380_1_1	NO	NO	THN	3.8
Gma_Glyma_01G217600_1	YES	NO	THN	3.39
Gma_Glyma_11G025600_1_GmOLPa	YES	NO	THN	3.45
Pvu_Phvul_002G155400_1	YES	NO	THN	3.45
Mtr_Medtr5g010640_1	YES	NO	THN	4.03
Mgu_Migut_A00799_1	YES	NO	THN	3.93
Mgu_Migut_E01128_1	YES	NO	THN	9.91
Sit_Seita_5G078100_1	YES	YES	THN;STYKc	3.83
Sit_Seita_5G078300_1	YES	YES	THN;SCOP d1qpca	3.83
Osa_LOC_Os01g02310_1	YES	YES	THN;S_TKc	4.03
Bdi_Bradi2g01217_1	YES	YES	THN;Pkinase	3.8
Zma_GRMZM2G002555_T01	YES	YES	THN;Pkinase	3.8
Zma_GRMZM2G435592_T02	YES	YES	THN;S_TKc	3.59

Bdi_Bradi2g01146_1	YES	YES	THN;S_TKc	4.2
Bdi_Bradi2g01200_2	YES	NO	THN;S_TKc	3.59
Sit_Seita_5G078200_1	YES	YES	THN;S_TKc	2.95
Bdi_Bradi2g01228_1	YES	NO	THN;S_TKc	3.59
Sit_Seita_9G148300_1	YES	NO	THN	3.83
Zma_GRMZM2G039639_T01	YES	NO	THN	3.83
Bdi_Bradi1g13060_1	YES	NO	THN	3.83
Bdi_Bradi1g13070_1	YES	NO	THN	3.83
Osa_LOC_Os03g46070_1	YES	NO	THN	3.83
Osa_LOC_Os03g45960_1	YES	NO	THN	3.83
Sit_Seita_2G365800_1	YES	NO	THN	3.83
Zma_GRMZM2G374971_T01_Zeamatin	YES	NO	THN	3.83
Zma_GRMZM2G010048_T01	YES	NO	THN	7.38
Osa_LOC_Os03g46060_1_OSM	YES	NO	THN	4.03
Asa_AAB02259(U57787)_permatin	YES	NO	THN	3.83
Bdi_Bradi4g05440_1	YES	NO	THN	3.83
Osa_LOC_Os12g43490_1	YES	NO	THN	3.83
Osa_LOC_Os12g43450_1	YES	NO	THN	4.18
Zma_GRMZM2G092474_T01	YES	NO	THN	3.83
Sit_Seita_5G355700_1	YES	NO	THN	4.18
Bdi_Bradi4g03290_1	YES	NO	THN	3.83
Bdi_Bradi4g05430_2	NO	NO	THN	3.83
Pta_PITA_000010735	YES	NO	THN	3.83
Pta_PITA_000010737	YES	NO	THN	3.83
Pta_PITA_000010739	YES	NO	THN;THN	3.83
Pta_PITA_000010740	YES	NO	THN	3.83
Pta_PITA_000000597	NO	NO	THN	3.92
Pta_PITA_000041533	YES	NO	THN;THN	3.45
Pta_PITA_000069097	YES	NO	THN	3.83
Pta_PITA_000070827	YES	NO	THN;THN	3.83
Pab_MA_10429511g0010	NO	NO	THN	3.83
Pab_MA_6505g0010	NO	NO	THN	3.83
Pta_PITA_000024408	YES	NO	THN	3.83
Pta_PITA_000066768	NO	NO	THN	3.83
Pta_PITA_000087912	YES	YES	THN;S_TKc	3.45
Pta_PITA_000058482	YES	YES	THN;S_TKc	3.45
Pta_PITA_000002550	YES	NO	THN	3.8
Pta_PITA_000069801	NO	NO	THN;S_TKc	3.45
Pta_PITA_000020282	YES	YES	THN;S_TKc	3.83
Pta_PITA_000091230	YES	NO	THN	3.45
Pta_PITA_000078522	NO	NO	THN	3.09
Pab_MA_10435621g0020	NO	YES	THN;S_TKc	3.45
Pab_MA_3795g0010	NO	NO	THN	3.53
Pab_MA_10428085g0010	NO	YES	THN;RPT 1; LRR_8; S_TKc	6.56
Pta_PITA_000018832	YES	NO	THN;THN	4.49
Pta_PITA_000093937	YES	NO	THN;THN	4.03
Pta_PITA_000004949	NO	YES	THN;S_TKc	3.7
Pta_PITA_000037146	NO	YES	THN;S_TKc	3.8
Pta_PITA_000091324	YES	YES	THN;THN;S_TKc	4.4
Pab_MA_473307g0010	NO	YES	THN;STYKc	4.4

Pab_MA_69685g0010	NO	YES	THN	6.64
Pta_PITA_000043543	NO	YES	THN;STYKc	8.58
Pta_PITA_000038798	YES	NO	THN;THN	3.53
Pta_PITA_000038163	YES	NO	THN	4.19
Pab_MA_10327089g0010	YES	NO	THN	3.96
Pta_PITA_000000202	NO	NO	THN	3.83
Pab_MA_133779g0010	YES	NO	THN	3.83
Aly_AL4G23320_t1	YES	NO	THN	5.67
Ath_AT2G28790_1	YES	NO	THN	5.67
Cgr_Cagra_4538s0005_1	NO	NO	THN	5.67
Bra_Brara_D01724_1	YES	NO	THN	5.67
Bra_Brara_G01401_1	YES	NO	THN	5.67
Spu_SapurV1A_0040s0150_1	YES	NO	THN	5.67
Spu_SapurV1A_1042s0010_1	YES	NO	THN	5.67
Spu_SapurV1A_0008s1120_1	YES	NO	THN	5.67
Mes_Manes_11G132100_1	YES	NO	THN	5.67
Sly_Solyc11g066130_1_1	YES	NO	THN	5.67
Stu_PGSC0003DMT400001066	YES	NO	THN	5.67
Aco_Aqcoe1G465700_1	YES	NO	THN	5.67
Gma_Glyma_12G064300_1	YES	NO	THN	5.67
Gma_Glyma_11G140800_1	YES	NO	THN	5.67
Pvu_Phvul_011G065300_1	YES	NO	THN	5.67
Osa_LOC_Os01g62260_1	YES	NO	THN	5.67
Bdi_Bradi2g54560_1	YES	NO	THN	5.67
Sit_Seita_5G375800_1	YES	NO	THN	5.67
Atr_evm_27_model_AmTr_v1_0_scaffold00012_94	YES	NO	THN	5.67
Mgu_Migut_A01135_1	NO	NO	THN	5.67
Egr_Eucgr_J02061_1	NO	NO	THN	5.67
Mtr_Medtr4g063630_1	YES	NO	THN	5.67
Spo_Spipo23G0021200	YES	NO	THN	7.35
Pta_PITA_000013861	YES	NO	THN	3.93
Atr_evm_27_model_AmTr_v1_0_scaffold00124_20	YES	NO	THN	4.2
Osa_LOC_Os12g38170_1	YES	NO	THN	4.79
Zma_GRMZM2G066602_T01	YES	NO	THN	4.4
Zma_GRMZM2G108396_T01	YES	NO	THN	4.17
Gma_Glyma_10G061100_1	NO	NO	THN	8.58
Gma_Glyma_10G060900_1	NO	NO	THN	8.58
Zma_GRMZM5G861959_T02	NO	NO	THN	4.15
Zma_GRMZM2G050867_T01	NO	NO	THN; Nup96	4.22
Zma_AC207628_4_FGP003	NO	NO	'HN; 5 PPR; PPR_2; DYW_deaminas	4.52
Osa_LOC_Os07g04730_1	NO	NO	2 RPT1; THN; 2 RPT2	5.03
Sfa_Sphfalx0188s0003_1	YES	NO	THN	3.45
Sfa_Sphfalx0431s0001_1	YES	NO	THN	3.45
Sfa_Sphfalx0261s0001_1	YES	NO	THN	3.59
Sfa_Sphfalx0261s0004_1	YES	NO	THN	3.59
Sfa_Sphfalx0057s0089_1	YES	NO	THN	3.92
Sfa_Sphfalx0046s0111_1	YES	NO	THN	3.59
Sfa_Sphfalx0163s0028_1	YES	NO	THN	6.39
Sfa_Sphfalx0140s0007_1	YES	NO	THN	3.82
Sfa_Sphfalx0078s0097_1	YES	NO	THN	3.95
Sfa_Sphfalx0140s0006_1	YES	NO	THN	3.59

Sfa_Sphfalx0113s0056_1	YES	NO	THN	3.45
Sfa_Sphfalx0016s0013_1	YES	NO	THN	3.59
Sfa_Sphfalx0287s0019_1	YES	NO	THN	3.59
Ppa_Pp3c9_14830V3_1	YES	NO	THN	3.02
Ppa_Pp3c17_5160V3_1	YES	NO	THN	3.45
Cre_Cre02_g102300_t1_2	NO	NO	THN	3.59
Ppa_Pp3c9_21030V3_1	YES	NO	THN	5.17
Mes_Manes_08G013800_1	YES	NO	THN	4.49
Gma_Glyma_10G062100_1	YES	NO	THN	5.69
Mtr_Medtr1g062390_1	YES	NO	THN	5.69
Aco_Aqcoe1G124700_1	NO	NO	THN	5.11
Mtr_Medtr2g063150_1	YES	NO	THN	5.21
Aco_Aqcoe1G125200_1	YES	NO	THN	8.76
Smo_118249	NO	NO	THN	9.04
Smo_26188	NO	NO	THN	5.7
Mes_Manes_13G003800_1	YES	NO	THN	4.19
Mes_Manes_13G003900_1	YES	NO	THN	4.36
Mes_Manes_12G003700_1	YES	NO	THN	7.21
Spu_SapurV1A_1323s0010_1	YES	YES	THN; S_TKc	4.69
Spu_SapurV1A_0067s0640_1	YES	YES	THN; S_TKc	4.84
Osa_LOC_Os06g19250_1	NO	NO	THN	9.02
Osa_LOC_Os04g24130_1	YES	NO	THN	7.81
Gma_Glyma_15G258400_1	NO	YES	THN; B lectin; Pkinase	7.89

[1]Domains predicted by SMART: THN, Thaumatin; RPT1, Internal Repeat 1; STYKc, phosphotransferases, possible dual-specificity Ser/Thr/Tyr kinase; SCOP d1qpca, protein-kinase like; S_TKc, phosphotransferases, serine or threonine-specific kinase; Pkinase, protein kinase domain; LRR_8, leucine-rich repeat domain.

Osmotin sequences are in bold

Table S3 Transposable elements that surround the *Arabidopsis thaliana* osmotin gene. Data accessed in TAIR10 (Wang et al., 2012).

Transposon_Name	orientation_is_5prime	Transposon_min_Start	Transposon_max_End	Transposon_Family	Transposon_Super_Family
AT4TE30225	FALSE	6992629	6994363	ATCOPIA57	LTR/Copia
AT4TE30230	TRUE	6993930	6994045	ARNOLD2	DNA/MuDR
AT4TE30235	FALSE	6995629	6995903	ATLINEIII	LINE/L1
AT4TE30270	FALSE	7003461	7003936	ATREP3	RC/Helitron
AT4TE30310	FALSE	7011697	7012856	VANDALNX2	DNA/MuDR
AT4TE30315	FALSE	7012857	7012910	ATDNA2T9C	DNA/MuDR
AT4G11650 (osmotin)		7024856	7026146		
AT4TE30390	FALSE	7027939	7028334	ATREP10A	RC/Helitron
AT4TE30395	FALSE	7028335	7028429	BRODYAGA1A	DNA/MuDR
AT4TE30400	TRUE	7028430	7028539	ATREP4	RC/Helitron
AT4TE30405	FALSE	7028540	7028708	ATDNAI27T9B	DNA/MuDR
AT4TE30410	TRUE	7029423	7029476	ATDNAI26T9	DNA/MuDR
AT4TE30415	FALSE	7029509	7029939	ATDNAI27T9C	DNA/MuDR
AT4TE30420	FALSE	7030209	7030311	VANDAL5A	DNA/MuDR
AT4TE30425	FALSE	7030312	7030464	VANDAL5	DNA/MuDR
AT4TE30430	FALSE	7030897	7031032	VANDAL4	DNA/MuDR
AT4TE30435	FALSE	7031033	7031138	VANDAL5A	DNA/MuDR
AT4TE30440	FALSE	7031139	7032281	VANDAL5A	DNA/MuDR
AT4TE30445	FALSE	7032282	7032552	VANDAL5	DNA/MuDR
AT4TE30450	TRUE	7032553	7034348	ATREP4	RC/Helitron
AT4TE30455	FALSE	7034349	7034984	ATREP2	RC/Helitron
AT4TE30460	TRUE	7034985	7035399	ATREP4	RC/Helitron
AT4TE30470	FALSE	7037328	7037987	ATREP15	RC/Helitron
AT4TE30475	FALSE	7037988	7038080	ATHILA8A	LTR/Gypsy
AT4TE30480	FALSE	7038890	7039272	ATREP19	DNA
AT4TE30485	TRUE	7040259	7040335	RathE2_cons	RathE2_cons

Table S4 Ontology annotation for putative *Arabidopsis thaliana* and *Oryza sativa* osmotin genes.

ID	GO accession	Type	Name	With
AT4G11650	GO:0009651	biological_process	response to salt stress	
	GO:0009816	biological_process	defense response to bacterium, incompatible interaction	
	GO:0009817	biological_process	defense response to fungus, incompatible interaction	
LOC_Os01g02310.1	GO:0008152	biological_process	metabolic process	TAIR:AT5G38280
	GO:0009987	biological_process	cellular process	TAIR:AT5G38280
	GO:0007165	biological_process	signal transduction	TAIR:AT5G38280
	GO:0009607	biological_process	response to biotic stimulus	TAIR:AT5G38280
	GO:0005575	cellular_component	cellular_component	TAIR:AT5G38280
	GO:0016301	molecular_function	kinase activity	TAIR:AT5G38280
	GO:0006464	biological_process	protein modification process	TAIR:AT5G38280
	GO:0004872	molecular_function	receptor activity	TAIR:AT5G38280
	GO:0004871	molecular_function	signal transducer activity	TAIR:AT5G38280
LOC_Os03g46070.1	GO:0005623	cellular_component	cell	TAIR:AT4G11650
	GO:0006950	biological_process	response to stress	TAIR:AT4G11650
	GO:0009607	biological_process	response to biotic stimulus	TAIR:AT4G11650
	GO:0009628	biological_process	response to abiotic stimulus	TAIR:AT4G11650
LOC_Os03g45960.1	GO:0005623	cellular_component	cell	TAIR:AT4G11650
	GO:0006950	biological_process	response to stress	TAIR:AT4G11650
	GO:0009607	biological_process	response to biotic stimulus	TAIR:AT4G11650
	GO:0009628	biological_process	response to abiotic stimulus	TAIR:AT4G11650
LOC_Os03g46060.1	GO:0005623	cellular_component	cell	TAIR:AT4G11650
	GO:0006950	biological_process	response to stress	TAIR:AT4G11650
	GO:0009607	biological_process	response to biotic stimulus	TAIR:AT4G11650
	GO:0009628	biological_process	response to abiotic stimulus	TAIR:AT4G11650
LOC_Os12g43490.1	GO:0005623	cellular_component	cell	TAIR:AT4G11650
	GO:0006950	biological_process	response to stress	TAIR:AT4G11650
	GO:0009607	biological_process	response to biotic stimulus	TAIR:AT4G11650
	GO:0009628	biological_process	response to abiotic stimulus	TAIR:AT4G11650
LOC_Os12g43450.1	GO:0005623	cellular_component	cell	TAIR:AT4G11650
	GO:0006950	biological_process	response to stress	TAIR:AT4G11650
	GO:0009607	biological_process	response to biotic stimulus	TAIR:AT4G11650
	GO:0009628	biological_process	response to abiotic stimulus	TAIR:AT4G11650

File S1 Amino acid sequence alignment in FASTA format.

File S2 Phylogenetic tree in Newick format.

Capítulo III
ARTIGO CIENTÍFICO 2

Molecular characterization of soybean osmotins and their involvement in the drought stress response

---

Manuscrito submetido para publicação na Physiologia Plantarum

Molecular characterisation of soybean osmotins and their involvement in drought stress response

Giulia Ramos Faillace[1], Paula Bacaicoa Caruso[2], Luis Fernando Saraiva Macedo Timmers[2,3,‡], Débora Favero[4], Frank Lino Guzman Escudero[5], Ciliana Rechenmacher[1], Luisa Abruzzi de Oliveira-Busatto[1], Osmar Norberto de Souza[2,3], Christian Bredemeier[4], Maria Helena Bodanese-Zanettini[1]*

[1] *Programa de Pós-Graduação em Genética e Biologia Molecular and Instituto Nacional de Ciência e Tecnologia: Biotec Seca-Pragas, Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil*

[2] *Laboratório de Bioinformática, Modelagem e Simulação de Biossistemas (LABIO), Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, RS, Brazil*

[3] *Programa de Pós-Graduação em Biologia Celular e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)*

[4] *Programa de Pós-Graduação em Fitotecnia, Departamento de Plantas de Lavoura, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil*

[5] *Programa de Pós-Graduação em Biologia Celular e Molecular, Centro de Biotecnologia (CBiot), Universidade Federal do Rio Grande do Sul (UFRGS), 91501-970, Porto Alegre, RS, Brazil*

[‡] *Present Address: Programa de Pós-Graduação em Biotecnologia, Universidade do Vale do Taquari (UNIVATES), Lajeado, RS, Brazil*

*Author for correspondence:

Maria Helena Bodanese-Zanettini

Tel: +55 51 33086725

Email: maria.zanettini@ufrgs.br

Abstract

Osmotins are multifunctional proteins belonging to the thaumatin-like family related to plant stress responses. To better understand the functions of soybean osmotins in drought stress response, the current study presents the characterisation of four previously described proteins and a novel putative soybean osmotin (GmOLPc). Gene and protein structure as well as gene expression analyses were conducted on different tissues and developmental stages of two soybean cultivars with varying dehydration sensitivities (BR16 and EMB48 are highly and slightly sensitive, respectively). The analysed osmotin sequences share the conserved amino acid signature and 3D structure of the thaumatin-like family. Some differences were observed in the conserved regions of protein sequences and in the electrostatic surface potential. P21-like present the most similar electrostatic potential to osmotins previously characterised as promoters of drought tolerance in *Nicotiana tabacum* and *Solanum nigrum*. Gene expression analysis indicated that soybean osmotins were differentially expressed in different organs (leaves and roots), developmental stages (R1 and V3), and cultivars in response to dehydration. In addition, under dehydration conditions, the highest level of gene expression was detected for GmOLPc and P21-like osmotins in the leaves and roots, respectively, of the less drought sensitive cultivar. Altogether, the results suggest an involvement of these genes in drought stress tolerance.

Abbreviations

ABA, abscisic acid; ΔT, air temperature; aa, amino acids; bp, base pairs; EMB48, EMBRAPA48; GSDS, gene structure display server; PGDD, genome duplication database; IRR, irrigated; MCMC, markov chain monte carlo; MEGA7, molecular evolutionary genetics analysis; MS medium, Murashige and Skoog basal salt mixture; NIRR, non-irrigated; PR, pathogenesis-related; PEG, polyethylene glycol; RWC, relative leaf water content; R2, reproductive stage; SMART, simple modular architecture research tool; TLPs, thaumatin-like proteins; 3D, three-dimensional; TFs, transcript factors; TSS, transcription start site; $\Psi_{MIN,}$ water potential; WGD, whole genome duplication.

Introduction

Due to plants being sessile organisms, they are often exposed to various abiotic stresses such as drought, cold, and soil salinity. These environmental stresses result in osmotic changes that can lead to the disruption of normal cellular activities, thus affecting plant growth and development. In crop plants, these effects could also hamper productivity (Husaini and Rafiqi, 2012). To avoid the consequences of multiple stresses, plants have evolved highly complex and sophisticated defence mechanisms such as the production of various defence-related proteins known as pathogenesis-related (PR) proteins. PRs are classified into 17 families (PR1 to PR17) based on their amino acid composition, structure, and biochemical function (Misra et al., 2016).

Osmotins are stress responsive multifunctional proteins belonging to the PR5 family (also known as the thaumatin-like family) due to their high sequence similarity to thaumatin, a sweet-tasting protein from the West African shrub *Thaumatococcus danielli* (Abdin et al., 2011). Osmotins were first isolated and characterised in tobacco cells adapted to a low osmotic potential, but also have been induced in several plant species - including *Glycine max* (soybean) - in response to various abiotic and biotic stresses (Parkhi et al., 2009; Weber et al., 2014). In view of their involvement in stress responses, many studies have aimed to overexpress osmotin in agronomically important crops (Barthakur et al., 2001; Parkhi et al., 2009; Goel et al., 2010; Annon et al., 2014; Weber et al., 2014). Transgenic plants overexpressing osmotin-like proteins have been shown to confer tolerance to salt, drought, freezing, as well as fungal and bacterial infection (Kumar et al., 2016).

As a primary contributor to global food production, soybean is one of the most important commodities. Current and predicted climate change, which suggests increased frequency, duration, and severity of drought periods, or intense heat, represent a serious challenge for agricultural production in Brazil and worldwide (Cunha et al., 2014; Chen et al., 2016). The evaluation of the relative effect of climate and agricultural technology on soybean productivity in Brazil indicated that some regions can be more heavily affected by climate change. Although the environmental suitability of some areas would increase, an overall decrease in environmental suitability was observed, indicating that soybean cultivation in Brazil could be highly threatened in the future (Caetano et al., 2018).

Drought is one of the most relevant environmental factors that dramatically limits soybean grain yield. Over the years, the development of drought-tolerant cultivars has served as a solution to yield losses (Shin et al., 2015). Moreover, a promising strategy to develop tolerance against abiotic stress is based on the overexpression of PR proteins - such as osmotins - in transgenic plants (Ahmed et al., 2013). Notably, Weber et al. (2014) demonstrated that the expression of *Solanum nigrum* osmotin in soybeans improved the physiological responses and yield components of transgenic plants subjected to water deficit. However, the molecular basis of osmotin action in response to drought remains unclear. In this sense, any attempt to improve stress tolerance first requires an improved understanding of the underlying physiological, biochemical, and molecular events (Abdin et al., 2011).

To date, soybean presents four identified osmotins: P21, GmOLPa, GmOLPb, and P21e (P21-like). P21 protein was the first osmotin identified in soybean, and has been purified from mature leaves without any stress treatments (Graham et al., 1992). GmOLPa, GmOLPb, and P21e isoforms were further characterised as being involved in high-salt stress and hormonal responses (Onishi et al., 2006; Tachi et al., 2009). Despite their high sequence similarities, some differences in three-dimensional (3D) structure, electrostatic potential, subcellular location, and gene expression suggested that each soybean osmotin could play a distinctive role in defence against high salt stress (Tachi et al., 2009). Although three soybean osmotins have been suggested as being related to salinity response, the roles of soybean osmotins in drought stress and tolerance remain unclear.

To better understand the functions of soybean osmotins in drought stress response, we present the characterisation of four previously described proteins and a novel putative soybean osmotin. The analysis included the protein sequence, subcellular location, individual 3D protein structure, electrostatic potential, gene structure, chromosomal position, gene duplication, putative *cis*-elements, and expression pattern of these osmotins in different tissues and developmental stages of two soybean cultivars with different sensitivities to dehydration.

Materials and Methods

Data mining and phylogenetic analyses

Data mining was performed to identify osmotins that have been expressed in transgenic plants and have demonstrated a relationship to drought tolerance. Previously identified osmotins from *Nicotiana tabacum* (X61679) (Das et al., 2011) and *Solanum nigrum* (AF450276) (Weber et al., 2014) were used as queries for a blastp search against the *Glycine max* full genome available at Phytozome v.012. Two other previously characterised *S. nigrum* osmotins (AF473702 and KC292261) were also included in the analysis. All soybean sequences retrieved from blastp share the thaumatin domain, thus suggesting that they belong to the thaumatin-like family. Protein isoforms were excluded to refine the analysis. The full-length coding sequences (cds) were translated into amino acid sequences and aligned using the Muscle algorithm from MEGA7 (Molecular Evolutionary Genetics Analysis) software (Kumar et al., 2016). The thaumatin domain sequences of *N. tabacum* and *S. nigrum* were identified by the Simple Modular Architecture Research Tool (SMART) (Letunic and Bork, 2017) and used as a reference to determine the thaumatin domain region of the other aligned sequences. The thaumatin domain sequence was then used for phylogenetic analysis. The sequences were manually edited, and the Prottest 3.4 program (Abascal et al., 2005) was used to identify the optimal protein evolution model for Bayesian analysis. The phylogenetic analysis was reconstructed using the Bayesian method in the Beast 1.8 package (Drummond et al., 2012). WAG+G was the best model for protein sequences dataset according Prottest (Abascal et al., 2005). The Birth-Death process was selected as a tree prior to Bayesian analysis, and 50,000,000 generations were performed with the Markov Chain Monte Carlo (MCMC) algorithms. Tracer 1.6 (Rambaut et al., 2014) was used to verify the effectivity of obtained data by the convergence of Markov chains and adequate effective sample sizes (>200) after the first 10% of generations had been deleted as burn-in. The TreeAnnotator (Beast 1.8 package) was used to access the maximum clade credibility of the consensus tree. The tree was visualised and edited using FigTree v1.4.3 (http:// tree.bio.ed.ac.uk/software/figtree/). Statistical support for the clades was determined by accessing the Bayesian posterior probability. Further editions were performed by visual analysis of sequence organisation in the reconstructed tree.

Gene and protein structure analysis

Osmotin sequences were analysed for gene and protein structures as well as signal peptide information. Intron/exon structures and organisation were determined from the Gene Structure Display Server (GSDS) program, developed by the Centre of Bioinformatics (CBI), Peking University (Hu et al., 2015). The protein structures, signal peptides, and subcellular locations were predicted using SMART (Letunic and Bork, 2017), SignalP 4.1 Server (http://www.cbs.dtu.dk/services/SignalP/), and TargetP 1.1 Server (http://www.cbs.dtu.dk/services/TargetP/), respectively. To verify the conserved residues, protein sequences were aligned using MEGA 7 software, and the alignments were further visually inspected using GeneDoc (Nicholas et al., 1997). The conserved residues were identified according to Petre et al. (2011) and Ahmed et al. (2013).

Chromosomal localisation and duplication pattern

The chromosomal localisation and gene duplication patterns were determined for soybean osmotin sequences using the Genome Duplication Database (PGDD), considering a 100kb syntenic region (http://chibba.agtec.uga.edu/duplication/) (Lee et al., 2013) and MCScanX software (http://chibba.pgml.uga.edu/mcscan2/) (Wang et al., 2012).

Bioinformatic sequences analysis and comparative modelling protocol

The SignalP 4.1 sequence analysis (Petersen et al., 2011) revealed that the first 20 amino acids were identified as a signal peptide. Hence, these residues were excluded from the molecular modelling procedure. We used the comparative modelling approach, and implemented in the MODELLER 9v19 program (Sali and Blundell, 1993) to construct 3D models of osmotins Gma_OLPb, Gma_P21like, Gma_P21, Gma_GmOLPc, and Gma_OLPa based on the 3D structure of grape thaumatin-like protein (PDB ID 4L5H) (Marangon et al., 2014). For Sni_SnOLP, Sni_SindOLP and Sni_Jami models were based on the 3D structure of NP24-I from *Solanum lycopersicum* (PDB ID: 2I0W) (Ghosh and Chakrabarti, 2008).

The protocol used to perform molecular modelling experiments was: generation of 10 models, from which one model for each osmotin sequence was selected. All models were submitted to the DOPE energy scoring function (Shen and Sali, 2006) implemented in

MODELLER 9v19 to select the best structures. The MOLPROBITY webserver (Chen et al., 2010) and PROCHECK (Laskowski et al., 1993) were used to verify and validate the stereochemical quality of the models. Electrostatic surface potential were calculated using the program APBS and displayed with the PyMOL program (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC). All images were generated using the PyMOL program. Multiple sequence alignment comparisons were performed using ClustalW (Combet et al., 2000) using the Blosum matrix for amino acid substitutions and the default parameters to infer possible structural similarities.

Gene expression data mining

In order to gain insights regarding gene expression, soybean osmotin RNA-seq expression data were searched in the Soybean eFP Browser (http://bar.utoronto.ca/efpsoybean/cgi-bin/efpWeb.cgi) (Libault et al., 2010), RNA-Seq Atlas of *Glycine max* (Severin et al., 2010), and data from drought stress experiments (Le et al., 2012; Belamkar et al., 2014; Shin et al., 2015; Chen et al.,2016).

Soybean dehydration assay for gene expression analysis

To improve the investigation on soybean osmotin expression in response to drought stress, two experiments were performed using the BR16 and EMBRAPA48 (EMB48) soybean cultivars, which are highly and slightly sensitive to dehydration stress (Oya et al., 2004), respectively.

The first experiment was performed under greenhouse conditions at an air temperature of 28±5ºC using natural illumination. Plants were grown in 5L-plastic pots filled with a substrate/soil mixture. Eight seeds were sown per pot, and four plants remained after thinning following plant emergence. Prior to sowing, seeds were inoculated with two *Bradyrhizobium* strains (*B. elkanii* and *B. japonicum*). Once a week, Hoagland solution was applied to each pot. Treatments consisted of two water regimes (control/irrigated and drought stress/non-irrigated), and were imposed for 38 days after sowing, when plants were at R1 growth stage (onset of flowering) (Fehr and Caviness, 1977). From plant emergence until the R1 growth stage, pots were weighed daily and irrigated with water (if necessary) to maintain soil moisture at approximately 90% of field capacity. The experiment was performed using a randomised block design with four biological replicates.

Each pot was considered an experimental unit. In order to characterise drought stress and physiological responses, relative leaf water content (RWC), minimum leaf water potential, leaf temperature, and the quantum yield of photosystem PSII were determined at 8 (moderate stress) and 10 (severe stress) days after watering suspension. The uppermost fully expanded leaves were used for data collection. Leaves from both cultivars with four biological replicates were also collected and frozen in liquid nitrogen for gene expression analysis. Samples were stored at -80°C until the analyses were performed.

In the second experiment, plants from the two cultivars (BR16 and EMB48) were grown in pots containing vermiculite supplemented with half strength MS medium (Murashige and Skoog Basal Salt Mixture) once per week. The pots were maintained under growth chamber conditions (26 ± 1°C with a 16/8 h light/dark cycle at a light intensity of 22.5 μEm^{-2s-1}) and irrigated once per day. At the V3 stage, plants were removed from pots, roots were washed, and entire plants were exposed to air. Leaves and roots were collected from both cultivars with five biological replicates at three time points (0, 6, and 12 hours). Samples were frozen in liquid nitrogen and stored at -80°C.

Physiological analysis

Relative leaf water content (RWC) estimates the current water content of sampled leaf tissue relative to the maximum water content it can hold at full turgidity, which was determined by the following relationship: RWC(%)=[(fresh weight – dry weight)/(turgid weight – dry weight)]*100. Firstly, the fresh weight of two leaves per plant and per treatment was determined. Thereafter, leaves were left floating in distilled water in Petri dishes for 24 hours, and the turgid weight was then recorded. Finally, leaves were dried in an oven at 65ºC for 48 hours, and their dry weight was measured (Salvador et al., 2012). Minimum leaf water potential was measured at 1:00pm using a Scholander-type pressure chamber (Model 3000, Soil Moisture Co., EUA) (Scholander et al., 1965; Boyer, 1967).

Leaf temperature was measured remotely by an infrared thermometer (Incoterm Co., São Paulo, Brazil) with temperature range of -10 to 60ºC, emissivity of 0.98 W m^{-2}, and a field of view of 2.80°. Readings were performed at the same time (11:00am) and at same distance (15 cm) between the leaf and thermometer on the adaxial surface of the uppermost fully expanded leaf in four plants per replicate. At this distance, the measured area on leaf

surface is approximately 9 mm^2. During measurements, air temperature was recorded, and the difference between both variables (ΔT) was calculated by the following relationship: $\Delta T_{leaf - air}$.

The quantum yield of photosystem II was determined under natural light conditions with a portable pulse modulation fluorometer (Model OS1-FL, Opti-Sciences, Hudson, USA). Measurements were performed at 11:00am on the adaxial surface of the uppermost fully expanded leaf in four leaves per replicate.

Gene expression analysis

Total RNA was extracted with a Trizol reagent (ThermoScientific) and treated with DNase I (ThermoScientific) according to the manufacturer's instructions. First-strand cDNAs were obtained using 1 μg of DNA-free RNA, M-MLV Reverse Transcriptase System™ (ThermoScientific) and oligo(dT) primers. To evaluate relative gene expression, the first-strand cDNA reaction product was diluted at 1:100.

Gene expression was analysed using real-time quantitative polymerase chain reaction (RT-qPCR). Primers were designed using Primer3 software (http://frodo.wi.mit.edu/) (Koressaar and Remm, 2007; Untergasser et al., 2012) according to the gene sequences (Table S1). The reactions were performed in a 25 μL volume containing 10 μM of each primer, 12.5 μL of diluted cDNA sample (1:100), 1 × PCR buffer, 50 mM $MgCl_2$, 10 mM of each dNTP, 2.5 μL SYBR-Green solution (1:100,000, Molecular Probes Inc., Eugene, OR), and 0.06 U Platinum Taq DNA Polymerase (ThermoScientific). The RT-qPCR was performed using a StepOne Applied Real-Time Cycler in a 96-well plate. Cycling conditions were implemented as follows: 5 min at 94 °C for an initial denaturation, 40 cycles of a 10 s denaturation step at 94 °C, a 15 s annealing step at 60 °C, and a 15 s extension step at 72 °C ending with 2 min at 72 °C for a final extension. Melting curve analysis was performed at the end of the PCR run over a range of 55–99 °C, increasing the temperature stepwise by 0.1 °C/s. Technical triplicate reactions were performed for each sample. The CYP2 and ELF1A genes were used as references for expression normalisation. Relative expression fold changes were determined using the $2^{-\Delta\Delta Ct}$ method described by Livak and Schmittgen (2001).

Statistical analysis

Physiological data from the two soybean cultivars, which were submitted or not to drought stress at the R1 reproductive stage, were subjected to analysis of variance (ANOVA). When the F-test was significant (p<0.05), a comparison of means was performed by Duncan's test (p=0.05) using the software SPSS Statistics version 17.0. Gene expression data were subjected to a $\log_2$ transformation prior to analysis in order to make data distribution more symmetrical (less skewed), since log-transformed data have less extreme values compared to untransformed data (Willems et al., 2008; Zwiener et al., 2014).

***In silico* promoter analysis**

The putative promoter region - located 2,000 base pairs (bp) upstream of the transcription start site (TSS) of each soybean osmotin gene - was used to search for putative *cis*-elements. The analysis was performed using the Plant Pan Database (http://plantpan2.itps.ncku.edu.tw/) (Chang et al., 2008). In addition, the transcript factors (TFs) identified by Plant Pan Database were searched in the RNA-seq data available from the literature on soybean submitted to drought stress. Only upregulated TFs genes were included.

Results

Data mining and phylogenetic analysis

Transgenic plants of different species expressing *N. tabacum* and *S. nigrum* osmotin genes have been shown to exhibit enhanced drought stress tolerance (Table 1). Two previously characterised osmotins (Nta_X61679_OSM and Sni_AF450276_SnOLP) were used as query sequences for blastp in soybean genome. A total of 61 soybean sequences were retrieved. The four *N. tabacum* and *S. nigrum* osmotin sequences (Nta_X61679_OSM, Sni_AF450276_SnOLP, Sni_AF473702_Jami, and Sni_KC292261_SindOLP) (Table 1) were included in further analyses. Following protein alignment, six sequences showed missing data, and therefore were excluded.

A phylogenetic tree reconstructed with the remaining 59 thaumatin-like sequences allowed the identification of an osmotin monophyletic clade including the eight previously identified osmotins from *N. tabacum*, *S. nigrum*, *G. max* (Nta_OSM, Sni_SnOLP,

Sni_SindOLP, Sni_Jami, Gma_01G217700_GmOLPb, Gma_05G204600_P21, Gma_05G204800_P21-like, and Gma_11G025600_GmOLPa), and a fifth soybean osmotin (Gma_Glyma_01G217600, GmOLPc) that has yet to be characterised (Fig. S1). GmOLPc was identified as a homologous sequence of GmOLPa (Fig. 1a). A subclade for the Solanaceae sequences (*N. tabacum* and *S. nigrum*) was formed inside the osmotin group. Moreover, a common ancestor was shared by the Solanaceae subclade and the GmOLPb sequence. Furthermore, the homologous sequences P21 and P21-like share a common ancestor with GmOLPb and the Solanaceae subclade, indicating more similarities among these sequences than GmOLPa and GmOLPc (Fig. 1a).

Gene and protein structure

In order to identify similarities between the osmotins, gene and protein structures were analysed for exon-intron organisation pattern, protein length, domain, signal peptide, and conserved residues.

Gene and protein structures were very similar among osmotins, which generally have no introns along the gene sequence and present a signal peptide followed by the thaumatin domain (Fig. 1a). GmOLPb was the unique exception, presenting two introns in its sequence and no signal peptide. Osmotin protein length varied from 222 (P21like) to 249 (GmOLPb) amino acids (aa). All osmotin proteins presented a thaumatin domain, and are predicted to be targeted to the secretory pathway. The three *S. nigrum* osmotins presented the same gene and protein structures, THN domain, and signal peptide length (Fig. 1a).

In general, the conserved amino acids described for the thaumatin domain - 16 cysteine residues, REDDD, and FF hydrophobic motifs - were identified in the osmotin sequences in similar positions to those described in previous publications (Petre et al., 2011; Ahmed et al., 2013), with the exception of Nta_OSM - which lost a cysteine residue - and for GmOLPc, which presents the aspartate^{109} residue (D) instead of the glutamate^{109} (E) in the REDDD motif (Fig. 1b). Despite this change, the two amino acids (D and E) have the same characteristics (polar, hydrophilic, and negatively charged), implying a substitution without different properties. Other amino acid substitutions were observed in the acid cleft position, some of which between amino acids with similar properties and others with different properties. For example, the substitution of methionine^{210} (M) in GmOLPa and

GmOLPc by lysine^{210} (K) in P21 and P21-like, and the substitution of glutamine^{210} (Q) in Nta_OSM, Sni_SnOLP, Sni_SindOLP, Sni_Jami, and GmOLPb. Methionine residue is hydrophobic and neutral, while lysine is hydrophilic, polar, and negatively charged, and glutamine is hydrophilic, polar, and neutral. Other substitutions between amino acids with different properties were observed in the methionine^{216} (M) of GmOLPc and GmOLPa sequences. This residue was replaced by a threonine^{216} (T) in the other osmotins, which is a polar, hydrophilic, and neutral amino acid (Fig. 1b).

Chromosomal localisation and duplication pattern

The chromosomal localisation of soybean osmotin-encoding genes and the putative mechanism of their duplication are illustrated in Figure 1c. The soybean osmotin-encoding genes are distributed on chromosomes 1, 5, and 11. On chromosome 1, two osmotin genes are identified (GmOLPc and GmOLPb), which are classified as WGD (whole genome duplication) and tandem duplication. P21 and P21-like, WGD and proximally duplicated, respectively, are localised on chromosome 5. In addition, the osmotin gene GmOLPa was identified on chromosome 11 and classified as WGD. All soybean osmotin genes were localised at the end of chromosomes (Fig. 1c). A synteny analysis confirmed the WGD classification of the paralogous soybean osmotin genes GmOLPc, P21, and GmOLPa (Fig. 2).

Protein sequence analysis and comparative modelling

Although significant differences have not been observed in amino acid content, sequence analysis of osmotins revealed differences in net charge. The net charges of *S. nigrum* (Sni_SnOLP, Sni_SindOLP, and Sni_Jami) and *N. tabacum* (Nta_OSM) proteins ranged from 0 to +5, whereas those of *G. max* (GmOLPa, Gma_P21like, Gma_P21, Gma_GmOLPc, and Gma_OLPb) varied from -5 to +4 (Table 2). Notably, only the Gma_P21-like osmotin has a positive net charge (+4), which is similar to the cationic osmotins of *S. nigrum* and *N. tabacum*.

In order to characterise differences among the osmotins of *N. tabacum*, *S. nigrum*, and *G. max*, three-dimensional (3D) structures for all proteins were built. The analysis revealed a conserved overall architecture of the osmotin proteins Nta_OSM, Sni_SnOLP, Sni_SindOLP, Sni_Jami, Gma_OLPa, Gma_GmOLPc, Gma_P21, Gma_P21-like, and

Gma_OLPb, which is also preserved among the thaumatin protein family (Anzlovar et al., 2003). The modelled structures comprise three domains: (i) domain I, containing 11 stranded β-sheets organised as a β-barrel, forming the protein core; (ii) domain II, containing an α-helix and a set of disulphide rich-loops; and (iii) domain III, containing a β-hairpin and a coil motif, both maintained by a disulphide bond (Fig. 3).

Since the fold is conserved, the distribution of charges in the protein surface was evaluated based on the electrostatic potential surface, which aimed to identify different patterns among proteins. It was observed that all osmotins displayed a negatively charged cavity (Fig. 4). However, the *N. tabacum* and *S. nigrum* proteins presented a predominantly positively charged posterior region, whereas for *G. max* that region is predominantly negatively charged, with the exception of Gma_P21-like osmotin (Fig. 4c).

Gene expression data mining

In order to gain insights regarding soybean osmotins expression, RNA-seq data available from different sources were investigated. Among soybean osmotins, only the P21, P21-like, and GmOLPa relative expression was observed in the investigated sources (Fig. 5 and Table 3). Using the Soybean eFP Browser, an up regulation of P21like and GmOLPa was detected in roots, while P21 up-regulation was detected in leaves. A slightly increased expression of the three genes was also observed in soybean flowers (Fig. 5a). Using the RNA-Seq Atlas of *Glycine max*, a greater number of read counts in flowers was detected in the three genes, especially P21. P21 also exhibited a high number of read counts in young leaves and pods, while GmOLPa showed higher counts in roots. GmOLPa and P21-like were also expressed in nodules (Fig. 5b).

Considering RNA-seq data, the differential expression of osmotin genes in response to dehydration and developmental stage has been verified (Table 3). P21 has shown to be upregulated in response to drought stress in leaves, during reproductive stage (R2), and 6 hours after treatment in the PI 416937 genotype (slightly sensitive to dehydration stress) (Shin et al., 2015) and 6 days after treatment in the Williams 82 cultivar (highly sensitive to drought) (Ha et al., 2015). Data from Belamkar et al. (2014) and Shin et al. (2015), showed that P21-like and GmOLPa osmotins were upregulated 6 and 12 hours after drought treatment in the roots of Williams 82 during the vegetative stage, and 6, 12, and 24

hours after treatment in leaves of the PI 416937 and Benning cultivars (highly sensitive to dehydration stress) during the reproductive stage. According to Le et al. (2012) data, P21-like osmotin was upregulated and downregulated 6 days after drought treatment in the leaves of Williams 82 during the reproductive and vegetative stages, respectively. However, the response was the reverse for GmOLPa (Table 3).

Gene expression analyses

Plants of the two soybean cultivars BR16 and EMB48 have been described as highly and slightly sensitive to dehydration stress (Oya et al., 2004), and were used in two different experiments to investigate soybean osmotin expression in response to drought stress.

In the first experiment physiological variables were determined and gene expression was evaluated (Table 4 and Fig. 6, respectively).

Relative water content (RWC) and leaf water potential can be used to determine plant water status, and integrates the effects of several drought adaptive traits (Mir et al., 2012). In the present study, the greatest difference in RWC was observed in the severe water stress regime (Table 4). At this point, the irrigated (IRR) plants of both cultivars exhibited higher RWC compared to the non-irrigated (NIRR) plants. Moreover, EMB48 NIRR plants presented higher RWC than BR16 NIRR. Differences among treatments in the same cultivar under moderate and severe stress were detected for minimum leaf water potential (Ψ_{MIN}). NIRR plants of both cultivars exhibited a decrease in leaf water potential, indicating their efforts to cope with water deficit.

Leaf temperature could also be indicative of plant stress (Martynenko et al., 2016). Plants of both cultivars exhibited increased leaf temperature in relation to air temperature (ΔT) under drought stress conditions. However, under severe stress, EMB48 NIRR plants exhibited a lower ΔT compared to BR16 NIRR plants (Table 4).

The quantum yield of photosystem II reinforces the differences between the two cultivars under drought stress. Under both stress conditions (moderate and severe), EMB48 plants responded better than BR16 plants, and did not present significant differences between IRR and NIRR plants.

The gene expression of soybean osmotins and ABA response markers genes were also accessed in plants involved in this first experiment (Fig. 6). All transcript levels were calibrated in relation to the expression level of BR16 IRR plants under moderate stress. The expression of three ABA response markers (GmABI1, GmBZIP1, and GmDREB2) increased in BR16 NIRR plants under moderate stress. Under this stress condition, no differences were observed between cultivars in the expression patterns of GmAB1 and GmBZIP1. However, in BR16 NIRR plants, the expression of GmDREB2 was greater than in EMB48 NIRR plants. An increment in the expression of GmAB1 and GmDREB2 was observed in EMB48 NIRR plants. Under severe stress, no differences were detected in GmAB1 and GmBZIP1 expression. Under this condition, the induction of GmDREB2 was observed for both cultivars in NIRR plants. BR16 NIRR plants presented higher expression of GmDREB2 than EMB48 NIRR plants under both moderate and severe stress levels.

In addition, the expression profiles of osmotin genes were also evaluated (Fig. 6). No differences were observed for GmOLPb or P21-like osmotins under moderate stress. BR16 NIRR plants exhibited incremental expression of GmOLPa and GmOLPc under both stress conditions, while EMB48 NIRR exhibited an increment only in GmOLPc expression under severe stress. An increment of P21-like expression was observed in NIRR plants for both cultivars under severe stress. When differences were detected between cultivars, BR16 plants generally presented higher expression of osmotin encoding-genes than EMB48 plants.

In the second experiment, BR16 and EMB48 plants were grown in a growth chamber until the V3 stage. At this stage, plants were removed from vermiculite and exposed to air at 0, 6, and 12 hours. Leaves and roots were collected, and osmotin gene expression was evaluated. All transcript levels were calibrated in relation to the expression level of BR16 roots at 0h. At 0h, no expression was detected for osmotins in the leaves of both cultivars, except slight expression of GmOLPb in BR16. At the same time point, all genes presented expression in the roots of both cultivars. GmOLPa expression was regularly observed in roots of both cultivars. Its expression increased until 12h in leaves and until 6h in roots. The highest expression of GmOLPa was observed in EMB48 roots at 6h. The expression of GmOLPc also increased until 12h in leaves and until 6h in the roots of both cultivars. The highest level of GmOLPc expression was observed in EMB48 leaves at 12h, while

EMB48 leaves and roots presented higher GmOLPc expression than BR16 at 6h. GmOLPb also presented higher expression in BR16 leaves at 6 and 12h. In EMB48, slight expression was detected for this gene in leaves at 12h and roots at 0 and 12h. P21-like osmotin expression was detected only in roots at 0 and 6h. The highest expression level for this gene was observed in the roots of EMB48 at 6h.

In silico promoter analysis

Temporal and spatial gene expression is influenced by the presence of different *cis*-regulatory elements in the promoter region, where transcription factors can bind. The *cis*-elements analysis revealed a strong presence of binding motifs for AT-Hook and homeodomain transcriptional factors family (Table 5). Myb c*is*-elements were the third-most present among all soybean osmotin promoter regions. NAC and bZIP c*is*-elements were also well represented. Analysis of the available RNA-seq data (Le et al., 2012; Belamkar et al., 2014; Chen et al., 2016) from drought stressed soybean did not identify the upregulation of AT-Hook genes. However, homeodomain and Myb genes were among the most abundant upregulated transcription factors recorded under drought conditions.

Discussion

Various osmotin proteins have been identified from a variety of plants and characterised based on their potential subcellular location, p*I* value, and gene expression in response to biotic and abiotic stresses (Tachi et al., 2009; Chowdhury et al., 2015; Ullah et al., 2017). Furthermore, studies on transgenic plants overexpressing osmotins have demonstrated the potential of this overexpression to protect plants against different stresses (Weber et al., 2014; Kumar et al., 2016). In the present study, we aim to elucidate the roles of soybean osmotin in drought stress response, and the structural and transcriptional characteristics of four previously described proteins and a novel putative soybean osmotin. *N. tabacum* and *S. nigrum* osmotins, previously characterised as providers of drought tolerance in plants, were used as references (Table 1). A phylogenetic tree was reconstructed and an osmotin clade was formed, thus allowing the identification of a novel soybean osmotin sequence (GmOLPc) (Figs 1 and S1). The GmOLPb soybean sequence was shown to be the most similar sequence to the Solanaceae osmotins, although it is a unique osmotin sequence that has two introns and no signal peptide (Fig. 1a). In this context, Xu et al. (2012) highlighted

that exon-intron structure alterations are prevalent in duplicated genes and, in many cases, have led to the generation of functionally distinct paralogs.

Duplication of genomic content can occur by many independent mechanisms, such as tandem duplication (consecutive duplications involving one or two genes), proximal duplications (duplications near one another but separated by a few genes), and whole-genome duplications (WGD; originate by polyploidy events) (Flagel and Wendel, 2009; Wang et al., 2012). GmOLPb previously characterised as a neutral osmotin (Tachi et al., 2009) was here classified as a tandem duplication found in chromosome 1 proximal to GmOLPc (Fig 1c), the novel putative osmotin sequence. GmOLPc is homologous to the acidic GmOLPa previously characterised by Onishi et al. (2006). The duplication pattern of GmOLPc, GmOLPa, and P21 was classified as WGD (Fig. 1c). P21-like osmotin was located near P21 (its homologous sequence) on chromosome 5 and was classified as a proximal duplication. Duplicated genes can undergo neofunctionalisation (when one copy acquires a novel function) or subfunctionalisation (when both copies are mutated and adopt complementary functions) (Lynch and Conery, 2000; Lynch and Force, 2000).

Responses to stress generally involve integrated circuits involving multiple pathways and cellular compartments. The subcellular localisation of a protein can provide important information regarding its function within the cell (Abdin et al., 2011). The score location assignment program suggests that all osmotin sequences are secretory proteins (Fig. 1a). Osmotins could also be synthesised as precursors presenting a C-terminal elongation that mediates their transport to the vacuole. This vacuole targeting was already identified for the *N. tabacum* and *S. nigrum* osmotins (Fig. 1b) (Campos et al., 2002). A C-terminal elongation was also identified in GmOLPb and proposed as vacuole targeting (Tachi et al., 2009). The acidic isoforms GmOLPa and P21 as well as P21-like and GmOLPc osmotins lack C-terminal elongation (Fig. 1b). Osmotin proteins that lack C-terminal elongation have been predicted to be released into an extracellular space by the direct effects of the N-terminal signal peptide (Onishi et al., 2006). The secretory nature and multiple location targeting of osmotins are in agreement with their multifunctional role in plant responses to biotic and abiotic stresses (Abdin et al., 2011). Furthermore, it has been also demonstrated that both intracellular and extracellular osmotins may be involved in plant drought tolerance (Onishi et al., 2006; Parkhi et al. 2009). According to Chowdhury et al. (2017),

higher concentrations of glycine, proline, and threonine in the interior of the cell might play an important role in the control of abiotic stress. However, in the present study, no significant differences in these amino acid contents among osmotin sequences were observed (Table 2).

The nine protein sequences here analysed (Nta_OSM, Sni_SnOLP, Sni_SindOLP, Sni_Jami, Gma_OLPa, Gma_GmOLPc, Gma_P21, Gma_P21-like, and Gma_OLPb) share the conserved amino acid signature and 3D structure of the thaumatin family, indicating their inclusion in this protein family (Figs 1b and 3). Despite their similarities, some differences were observed in the conserved regions of protein sequences and in electrostatic surface potential. The GmOLPc and GmOLPa sequences have two important substitutions involving amino acids with different properties in the acidic cleft region. These substitutions change a hydrophilic amino acid by a hydrophobic methionine$^{(210,216)}$ in both sequences. GmOLPc also has an aspartate^{109} residue (D) instead of the glutamate^{109} residue (E) in the REDDD motif, which is an important sequence in the acidic cleft related to PR5 antifungal activity (Koiwa et al., 1999). According to the authors, in addition to its acidic nature, the cleft region is rich in hydrophilic residues, which is a characteristic fairly typical of carbohydrate-binding sites. Therefore, the acidic REDDD motif and the hydrophilic residues of the acidic cleft region are important in determining protein antifungal activity (Jami et al., 2007). In addition, a study conducted on the basic osmotin of tobacco suggests that the acidic cleft of this protein forms a hollow for Ca^{+2} electrostatic binding that facilitates the interaction to glycans on the surface of fungal cells, thereby leading to plasma membrane permeabilisation and damage (Salzman et al., 2004). Notably, as shown for some osmotins, Ca^{+2} is also related to enhanced plant drought tolerance by protecting the structure and stability of cellular plasma membranes against lipid peroxidation, elevating proline content, and maintaining normal photosynthesis (Song et al., 2008; Kumar et al., 2015). In spite of all nine analysed osmotins presenting a negative cavity, some differences in electrostatic potential were observed in the posterior region of the proteins and in their net charge. All soybean osmotins are negatively charged, except for P21-like, which has a positive net charge and a posterior region that is predominantly positively charged. The charge characteristics of P21-like are similar to those of *N. tabacum* and *S. nigrum* osmotins (Table 2 and Fig. 4). Differences in the topology and

surface electrostatic potential surrounding the cleft are thought to determine the specificity of TLPs to their target proteins and ligands (Min et al., 2004).

The two cultivars BR16 and EMB48 - highly and slightly sensitive to dehydration, respectively - were evaluated for physiological variables to determine plant stress status. Differences between the two cultivars were observed for RWC, $\Delta T_{leaf - air}$, and quantum yield of photosystem II (Table 4). EMB48 retained more water in its leaves than BR16, maintaining a lower leaf temperature and photosystem II integrity, thereby reinforcing its characterisation as slightly sensitive to dehydration. A previous study showed that drought tolerant soybean genotypes were able to maintain RWC values and chlorophyll content at steady-state levels, even under stress conditions (Hossain et al., 2014). These physiological adaptive traits are frequently associated with abscisic acid (ABA) phytohormone signalling (Mir et al., 2012; Hossain et al., 2014; Martynenko et al., 2016). The results presented in the current study indicated that GmDREB2 ABA marker gene expression generally increased in both BR16 and EMB48 cultivars under moderate and severe water stress (Fig. 6). GmDREB2 was previously described as being responsive to ABA signalling and being involved in ABA-dependent signal pathways in soybean (Chen et al., 2007). According to the authors, GmDREB2 acts as an important transcriptional activator and may be useful in improving plant tolerance to abiotic stresses. It has also been revealed that tobacco osmotin is induced in cultured cells and roots in response to ABA treatment and under polyethylene glycol (PEG)-mediated water or salt stresses (Ullah et al., 2017). According to Onishi et al. (2006), the expression of GmOLPa osmotin in response to ABA and dehydration may be primarily induced via an ABA-independent transcriptional pathway. The *in silico* promoter analysis of osmotins performed in the present study revealed a strong presence of AT-Hook, homeodomain, and Myb *cis*-elements (Table 5). The presence of Myb *cis*-elements (involved in dehydration and abscisic acid (ABA) response) upstream to the GmOLPa coding sequence was also reported by Onishi et al. (2006). Although NAC and bZIP *cis*-elements have not been as numerous in the promoter region of soybean osmotins, their encoding genes have been frequently identified as being upregulated under drought stress conditions. Notably, the GmOLPc and GmOLPa promoter region also have a small number of WRKY *cis*-elements. Moreover, the upregulation of WRKY-encoding genes has been related to drought stress response (Dias et al., 2016).

As previously mentioned, the overexpression of osmotins could also promote abiotic stress tolerance in transgenic plants (Ahmed et al., 2013). The present study demonstrated that soybean osmotins (GmOLPa, GmOLPb, P21-like, and GmOLPc) were differentially expressed in different organs (leaves and roots), developmental stages (V3 and R1), cultivars (BR16 and EMB48), and in response to dehydration (Figs 6 and 7). In the first experiment, in which the leaves of soybean cultivars were collected at the R1 developmental stage, BR16 osmotins were more induced compared to EMB48 osmotins (Fig. 6). However, in the second experiment, in which soybean plants were sampled at V3 stage, the GmOLPa, GmOLPc, and P21-like osmotins of EMB48 exhibited higher expression than those of BR16 at certain time points and organs (Fig. 7). The expression of these three osmotins at 0h (control) was exclusive to roots. GmOLPa and GmOLPc exhibited expression in leaves following dehydration treatment, while P21-like osmotin showed expression only in roots (Fig. 7). These results are congruent with *in silico* data that supports the expression of P21-like and GmOLPa osmotins being observed only in the roots of non-stressed soybean plants (Fig. 5a). Onishi et al. (2006) also reported that dehydration for 24 h markedly increased expression of the GmOLPa gene in roots, and also induced low levels of expression in stems and leaves.

In conclusion, in the present work we characterized a new soybean osmotin-encoding gene (GmOLPc) and its expression pattern and putative product were compared to the already known osmotin isoforms (P21, GmOLPb, GmOLPa and P21-like). Our results show that the soybean osmotins expression pattern is organ and developmental stage dependent. The highest level of gene expression was detected for GmOLPc and P21-like osmotins in leaves and roots, respectively, of the less drought sensitive cultivar. We have also demonstrated that P21-like osmotin presents the most similar net charge to those osmotins previously characterised as promoters of drought tolerance in *N. tabacum* and *S. nigrum*. Overall, the results suggest the involvement of GmOLPc and P21-like osmotins in drought stress tolerance.

Author contributions

Study design: GRF and MHB-Z. *In silico* analysis: GRF. Duplication pattern analysis: FLGE. Bioinformatic sequences analysis and comparative modelling: PBC, LFSMT, and ONS. Soybean dehydration assays: GRF, DF, and CB. Physiological analysis: DF and CB.

Gene expression analysis: GRF and CR. Statistical analysis: CB. Manuscript: GRF, LFSMT, and CB. Manuscript revision: LAO-B and MHB-Z. Study supervision and coordination: MHB-Z. All authors read and approved the final manuscript.

Acknowledgements

We thank Dr. Rogério Margis for the help with gene expression analysis.

Funding

This work was supported by grants from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Chamada INCT – MCTI/CNPq/CAPES/FAPs nº 16/2014, Ativos Biotecnológicos Aplicados a Seca e Pragas em Culturas Relevantes para o Agronegócio (INCT Biotec Seca-Pragas) [88887.136360/2017-00 - 465480/2014-4].

References

Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

Abdin MZ, Kiran U, Alam A (2011) Analysis of osmotin, a PR protein as metabolic modulator in plants. Bioinformation 5:336

Ahmed NU, Park JI, Jung HJ, Chung MY, Cho YG, Nou IS (2013) Characterization of Thaumatin-like gene family and identification of *Pectobacterium carotovorum* subsp. *carotovorum* inducible genes in *Brassica oleracea*. Plant Breed Biotech 1:111-121

Annon A, Rathore K, Crosby K (2014) Overexpression of a tobacco osmotin gene in carrot (*Daucus carota* L.) enhances drought tolerance. In Vitro Cell Dev Biol—Plant 50:299–306

Anzlovar S, Dermastia M (2003) The comparative analysis of osmotins and osmotic-like PR-5 proteins. Plant Biol 5:116-124

Barthakur S, Babu V, Bansal KC (2001) Over-expression of Osmotin induces proline accumulation and confers tolerance to osmotic stress in transgenic tobacco. J Plant Biochemistry & Biotechnology 10:31-37

Belamkar V, Weeks NT, Bharti AK, Farmer AD, Graham MA (2014) Cannon S.B., Comprehensive characterization and RNA-seq profiling of the HD-Zip transcription factor family in soybean (*Glycine max*) during dehydration and salt stress. BMC Genomics 15:950

Bhattacharya A, Saini U, Joshi R, Kaur D, Pal AK, Kumar N, Gulati A, Mohanpuria P, Yadav SK, Kumar S, Ahuja PS (2014) Osmotin-expressing transgenic tea plants have improved stress tolerance and are of higher quality. Transgenic Res 23:211–223

Boyer JS (1967) Leaf water potencial measured with a pressure chamber. Plant Physiology 42:1056-1066

Caetano JM, Tessarolo G, de Oliveira G, e Souza KDS, Diniz-Filho JAF, Nabout JC (2018) Geographical patterns in climate and agricultural technology drive soybean productivity in Brazil. PloS one 13:e0191273

Campos MA, Ribeiro SG, Rigden DJ, Monte DC, Grossi de Sa MF (2002) Putative pathogenesis-related genes within *Solanum nigrum* L. var americium genome: isolation of two genes coding for PR5-like proteins, phylogenetic and sequence analysis. Physiological and Molecular Plant Pathology 61:205-216

Chang WC, Lee TY, Huang HD, Huang HY, Pan RL (2008) PlantPAN: plant promoter analysis navigator, for identifying combinatorial cisregulatory elements with distance constraint in plant gene group. BMC Genom 9:561

Chen M, Wang QY, Cheng XG, Xu ZS, Li LC, Ye XG, Xia LQ, Ma YZ (2007) GmDREB2, a soybean DRE-binding transcription factor, conferred drought and high-salt tolerance in transgenic plants. Biochem Biophys Res Commun 353:299-305

Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr D66:12-21

Chen W, Yao Q, Patil GB, Agarwal G, Deshmukh RK, Wang LLB, Wang Y, Prince SJ, Song L, Xu D, An YC, Valliyodan B, Varshney RK, Nguyen HT (2016) Identification and

comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-Seq. Front Plant Sci 7:1044

Chowdhury S, Basu A, Kundu S (2015) Cloning, characterization, and bacterial over-expression of an osmotin-like protein gene from *Solanum nigrum* L. with antifungal activity against three necrotrophic fungi. Mol Biotechnol 57:371-81

Chowdhury S, Basu A, Kundu S (2017) Overexpression of a new osmotin-like protein gene (SindOLP) confers tolerance against biotic and abiotic stresses in sesame. Frontiers in Plant Science 8:410

Combet C, Blanchet C, Geourjon C, Deléage G (2000) NPS@: Network Protein Sequence Analysis. Trends Biochem Sci 25:147-150

Cunha DA, Coelho AB, Féres JG (2014) Irrigation as an adaptive strategy to climate change: an economic perspective on Brazilian agriculture. Envir and Develop Econ 20:57–79

Das M, Chauhan H, Chhibbar A, Haq QMR, Khurana P (2011) High-efficiency transformation and selective tolerance against biotic and abiotic stress in mulberry, *Morus indica* cv. K2, by constitutive and inducible expression of tobacco osmotin. Transgenic Res 20:231–246

Dias LP, de Oliveira-Busatto LA, Bodanese-Zanettini MH (2016) The differential expression of soybean [*Glycine max* (L.) Merrill] WRKY genes in response to water deficit. Plant Physiology and Biochemistry 107:288-300

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973

Fehr W, Caviness RH (1977) Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill. Plant Science for a Better World 11:929-931

Flagel LE, Wendel JF (2009) Gene duplication and evolutionary novelty in plants. New Phytol 183:557–564

Gao SQ, Chen M, Xu ZS, Zhao CP, Li L, Xu HJ, Tang Y, Zhao X, Ma YZ (2011) The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants. Plant molecular biology 75:537-553

Ghosh R, Chakrabarti C (2008) Crystal structure analysis of NP24-I: a thaumatin-like protein. Planta 228:883-890

Goel D, Singh AK, Yadav V, Babbar SB, Bansal KC (2010) Overexpression of osmotin gene confers tolerance to salt and drought stresses in transgenic tomato (*Solanum lycopersicum* L.). Protoplasma 245:133–141

Graham JS, Burkhart W, Xiong J, Gillikin JW (1992) Complete amino acid sequence of soybean leaf P21. Plant Physiol 98:163-165

Ha CV, Watanabe Y, Tran UT, Le DT, Tanaka M, Nguyen KH, Tran LS (2015) Comparative analysis of root transcriptomes from two contrasting drought-responsive Williams 82 and DT2008 soybean cultivars under normal and dehydration conditions. Frontiers in plant science 6:551

Hossain MM, Liu X, Qi X, Lam HM, Zhang J (2014) Differences between soybean genotypes in physiological response to sequential soil drying and rewetting. The Crop Journal 2:366-380

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31:1296-1297

Husaini AM, Rafiqi AM (2012) Role of osmotin in strawberry improvement. Plant molecular biology reporter 30:1055-1064

Jami SK, Anuradha TS, Guruprasad L, Kirti PB (2007) Molecular, biochemical and structural characterization of osmotin-like protein from black nightshade (*Solanum nigrum*). Journal of plant physiology 164:238-252

Koiwa H, Kato H, Nakatsu T, Oda JI, Yamada Y, Sato F (1999) Crystal structure of tobacco PR-5d protein at 1.8 Å resolution reveals a conserved acidic cleft structure in antifungal thaumatin-like proteins1. Journal of molecular biology 286:1137-1145

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. Bioinformatics 23:1289-1291

Kumar SA, Kumari PH, Kumar GS, Mohanalatha C, Kishor PBK (2015) Osmotin: a plant sentinel and a possible agonist of mammalian adiponectin. Front Plant Sci 6:163

Kumar SA, Kumari PH, Jawahar G, Prashanth S, Suravajhala P, Katam R, Sivan P, Rao KS, Kirti PB, Kishor PBK (2016) Beyond just being foot soldiers – osmotin like protein (OLP) and chitinase (Chi11) genes act as sentinels to confront salt, drought, and fungal stress tolerance in tomato. Environmental and Experimental Botany 132:53-65

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33:1870-1874

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283-291

Le DT, Nishiyama R, Watanabe Y, Tanaka M, Seki M, Ham LH, Yamaguchi-Shinozaki K, Shinozaki K, Tran LP (2012) Differential Gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. Plos One 7:e49522

Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res 41:D1152–D1158

Letunic I, Bork P (2017) 20 years of the SMART protein domain annotation resource. Nucleic Acids Research gkx922.

Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, He J, Xu D, May G, Stacey G (2010) An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant J 63:86–99

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25:402-408

Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459-473

Marangon M, Van Sluyter SC, Waters EJ, Menz RI (2014) Structure of haze forming proteins in white wines: *Vitis vinifera* thaumatin-like proteins. PLoS One 9:e113757

Martynenko A, Shotton K, Astatkie T, Petrash G, Fowler C, Neily W, Critchley AT (2016) Thermal imaging of soybean response to drought stress: the effect of *Ascophyllum nodosum* seaweed extract. Springerplus 5:1393

Min K, Ha SC, Hasegawa PM, Bressan RA, Yun DJ, Kim KK (2004) Crystal structure of osmotin, a plant antifungal protein. Proteins: Structure, Function, and Bioinformatics 54:170-173

Mir RR, Zaman-Allah M, Sreenivasulu N, Trethowan R, Varshney RK (2012) Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. Theor Appl Genet 125:625-45

Misra RC, Sandeep, Kamthan M, Kumar S, Ghosh S (2016) A thaumatin-like protein of *Ocimum basilicum* confers tolerance to fungal pathogen and abiotic stress in transgenic Arabidopsis. Scientific Reports 6:25340

Nicholas KB, Nicholas HB, Deerfield DW (1997) Genedoc: analysis and visualization of genetic variation. Embnew News 4:14

Onishi M, Tachi H, Kojima T, Shiraiwa M, Takahara H (2006) Molecular cloning and characterization of a novel salt-inducible gene encoding an acidic isoform of PR-5 protein in soybean. Plant Physiol Biochem 44:574–580

Oya T, Nepomuceno AL, Neumaier N, Farias JRB, Tobita S, Ito O (2004) Drought tolerance characteristics of Brazilian soybean cultivars—evaluation and characterization of drought tolerance of various Brazilian soybean cultivars in the field. Plant Prod Sci 7:129-137

Parkhi V, Kumar V, Sunilkumar G, Campbell LM, Singh NK, Rathore KS (2009) Expression of apoplastically secreted tobacco osmotin in cotton confers drought tolerance. Mol Breeding 23:625–639

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785-786

Petre B, Major I, Rouhier N, Duplessis S (2011) Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar. BMC Plant Biology 11:33

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6. Available from http://tree.bio.ed.ac.uk/software/tracer/

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779-815

Salvador V, Pagan M, Cooper M, Kantartzi SK, Lightfoot DA, Meksem K, Kassem MA (2012) Genetic analysis of relative water content (RWC) in two re¬combinant inbred line populations of soybean [*Glycine max* (l.) merr.]. Journal of Plant Genome Sciences 1:46–53

Salzman RA, Koiwa H, Ibeas JI, Pardo JM, Hasegawa PM, Bressan RA (2004) Inorganic cations mediate plant PR-5 protein antifungal activity through fungal Mnn1-nad Mnn4-regulated cell surface glycans. Mol Plant Microbe Interact 17:780-788

Scholander PF, Bradstreet ED, Hemmingsen EA, Hammel HT (1965) Sap pressure in vascular plants: negative hydrostatic pressure can be measured in plants. Science 148:339-46

Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC (2010) RNA-Seq Atlas of *Glycine max*: A guide to the soybean transcriptome. BMC Plant Biology 10:160

Shen MY, Sali A (2006) Statistical potential for assessment and prediction of protein structures. Prot Science 15:2507-2524

Shin JH, Vaughn JN, Abdel-Haleem H, Chavarro C, Abernathy B, Kim KD, Jackson SA, Li Z (2015) Transcriptomic changes due to water deficit define a general soybean response and accession-specific pathways for drought avoidance. BMC Plant Biology 15:26

Silvestri C, Celletti S, Cristofori V, Astolfi S, Ruggiero B, Rugini E (2017) Olive (*Olea europaea* L.) plants transgenic for tobacco osmotin gene are less sensitive to in vitro-induced drought stress. Acta Physiol Plant 39:229

Song WY, Zhang ZB, Shao HB, Guo XL, Cao HX, Zhao HB, Fu ZY, Hu XJ (2008) Relationship between calcium decoding elements and plant abiotic-stress resistance. International Journal of Biological Sciences 4:116–125

Tachi H, Yamada KF, Kojima T, Shiraiwa M, Takahara H (2009) Molecular characterization of a novel soybean gene encoding a neutral PR-5 protein induced by high-salt stress. Plant Physiol Biochem 47:73–79

Ullah A, Hussain A, Shaban M, Khan AH, Alariqi M, Gul S, Jun Z, Lin S, Li J, Jin S, Munis MFH (2017) Osmotin: a plant defense tool against biotic and abiotic stresses. Plant Physiology and Biochemistry 123:149-159

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:115

Wang Y, Suo H, Zheng Y, Liu K, Zhuang C, Kahle KT, Ma H, Yan X (2010) The soybean root specific protein kinase GmWNK1 regulates stress responsive ABA signaling on the root system architecture. The Plant Journal 64:230-242

Wang Y, Tang H, DeBarry JD, Xu T, Li J, Wang X, Lee T, Jin H, Marler B, Kissinger HGJC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research 40:e49

Weber RLM, Wiebke-Strohm B, Bredemeier C, Margis-Pinheiro M, Brito GG, Rechenmacher C, Bertagnolli PF, Sá MEL, Campos MA, Amorim RMS, Beneventi MA, Margis R, Grossi-de-Sa MF, Bodanese-Zanettini MH (2014) Expression of an osmotin-like protein from *Solanum nigrum* confers drought tolerance in transgenic soybean. BMC Plant Biology 14:343

Willems E, Leyns L, Vandesompele J (2008) Standardization of real-time PCR gene expression data from independent biological replicates. Analytical biochemistry 379:127-129

Xu G, Guo C, Shan H, Kong H (2012) Divergence of duplicate genes in exon–intron structure. Proceedings of the National Academy of Sciences 109:1187-1192

Zwiener I, Frisch B, Binder H (2014) Transforming RNA-Seq data to improve the performance of prognostic gene signatures. PloS one 9:e85150

Tables

Table 1 Transgenic plants overexpressing osmotins that confer drought tolerance.

Gene name	ID	Donor	Transgenic plant	Reference
	-		*Nicotiana tabacum*	Barthakur et al. (2001)
	-		*Gossypium hirsutum L.*	Parkhi et al. (2009)
	-		*Solanum lycopersicum*	Goel et al. (2010)
Osmotin	X61679	*Nicotiana tabacum*	*Morus indica*	Das et al. (2011)
	-		*Daucus carota L.*	Annon et al. (2014)
	M29279		*Camellia sinensis L.*	Bhattacharya et al. (2014)
	-		*Olea europaea L.*	Silvestri et al. (2017)
SnOLP	AF450276		*Glycine max*	Weber et al. (2014)
OLP	AF473702	*Solanum nigrum*	*Solanum lycopersicum*	Kumar et al. (2016)
SindOLP	KC292261		*Sesamum indicum*	Chowdhury et al. (2017)

"-" indicates that the sequence ID was not available in the manuscript.

Table 2 Glycine, proline, and threonine amino acids content in *Nicotiana tabacum*, *Solanum nigrum*, and *Glycine max* osmotin sequences.

	Glycine	Proline	Threonine	Net charge
Nta_OSM	11.4 %	8.2 %	9.4 %	5
Sni_SnOLP	10.5 %	8.5 %	10.1 %	0
Sni_SindOLP	10.9 %	8.5 %	9.7 %	2
Sni_SnJami	10.5 %	8.5 %	9.7 %	0
Gma_GmOLPb	12.0 %	7.6 %	7.2 %	-1
Gma_GmP21like	10.8%	6.3%	9.5%	4
Gma_GmP21	12.1 %	6.2 %	8.5 %	-1
Gma_GmOLPc	9%	5.8%	10.3%	-4
Gma_GmOLPa	9.8 %	5.8 %	9.4 %	-5

Table 3 Microarray and RNA-seq data from soybean submitted to drought stress experiments.

Reference	Genotype	Time	Developmental stage	Organ	P21-like	P21	GmOLPa
Le et al., 2012[1]	Williams 82	6 days	R2	Leaves			
			V6	Leaves			
Belamkar et al., 2014[2]	Williams 82	1h	V1	Roots			
		6h					
		12h					
Shin et al., 2015[2]	PI	6h	R2	Leaves			
		12h					
		24h					
	Benning	6h	R2	Leaves			
		12h					
		24h					
Chen et al., 2016[2]	Williams 82	7 days	V4	Leaves			

[1] Microarray analysis

[2] RNA-seq analysis

Red and *green* rectangles indicate *upregulation* and *downregulation*, respectively.

Table 4 Physiological variables of two soybean cultivars (BR16 and EMB48) at the R1 developmental stage under moderate and severe water stress.

Physiological variable	Stress	Treatment	BR16		EMB48	
RWC (%)	Moderate	IRR	86.3 (± 4.5)	Ab	90.7 (± 0.8)	Aa
		NIRR	71.7 (± 5.0)	Ab	74.1 (± 5.2)	Ab
	Severe	IRR	90.8 (± 0.4)	Aa	91.1 (± 0.9)	Aa
		NIRR	54.9 (± 0.8)	Bb	70.6 (± 3.6)	Ab
Ψ_{MIN} (MPa)	Moderate	IRR	-0.67 (± 0.05)	Aa	-0.77 (± 0.05)	Aa
		NIRR	-1.31 (± 0.18)	Ab	-1.24 (± 0.28)	Ab
	Severe	IRR	-0.73 (± 0.03)	Aa	-0.70 (± 0.05)	Aa
		NIRR	-1.95 (± 0.27)	Ab	-1.59 (± 0.23)	Ab
$\Delta T_{leaf-air}$ (ºC)	Moderate	IRR	-2.57 (± 0.79)	Ab	-1.83 (± 0.97)	Ab
		NIRR	0.77 (± 0.90)	Aa	0.67 (± 1.07)	Aa
	Severe	IRR	-2.25 (± 0.28)	Ab	-3.13 (± 0.65)	Bb
		NIRR	0.74 (± 0.64)	Aa	-0.66 (± 0.35)	Ba
PSII quantum yield	Moderate	IRR	0.60 (± 0.04)	Aa	0.53 (± 0.05)	Aa
		NIRR	0.44 (± 0.05)	Bb	0.64 (± 0.03)	Aa
	Severe	IRR	0.66 (± 0.01)	Aa	0.68 (± 0.003)	Aa
		NIRR	0.50 (± 0.06)	Bb	0.67 (± 0.02)	Aa

IRR, irrigated.
NIRR, non-irrigated.
RWC, relative water content.
Ψ_{MIN}, minimum leaf water potential.
$\Delta T_{leaf-air,}$ leaf temperature in relation to air temperature.
PSII, photosystem II
Capital letters correspond to the comparison among means of cultivars within the same treatment.
Small letters correspond to the comparison among means of irrigation treatments within the same cultivar.
Means followed by the same letter do not differ significantly (Duncan's test, $p < 0.05$). Values

between parentheses represent the standard error of the mean.

Table 5 Transcription factor binding sites (TFBS) identified in the promoter sequences of soybean osmotin genes and the number of TFs upregulated in expression data.

	Number of *cis*-elements					Upregulated TFs in drought stressed soybean[1]		
TF family	**GmOLPa**	**GmOLPc**	**GmOLPb**	**P21**	**P21-like**	**Le et al., 2012**	**Chen et al., 2016**	**Belamkar et al., 2014**
AP2	8	3				17	19	35
AP2; B3	7	12	3	8	10			
AT-Hook	135	93	143	67	74	1		
bHLH	9	9	3	5	5	9	50	24
bZIP	13	27	23	33	21	9	32	9
C2H2	8	2	13	8	17	6	20	10
bZIP; homeodomain; HD-ZIP	2		4	2	2	3		
EIN3	6	5	3	8	11	2	2	
GATA	22	6	2	7	5		11	2
Sox	2		2	2	1			
Homeodomain	59	58	79	70	54	2	20	7
MADF	6	2	13	12	14		1	
MADS box	9	6	4	12	8	1	4	3
Myb/SANT	53	40	48	55	37	27	89	28
NAC; NAM	20	22	14	10	14	30	54	10
SBP	35	52	1	14	1	1		
TBP	25	47	83	40	25		4	
TCP	1	1	1	6		2	3	1
TCR	12	7	40	12	10			
WRKY	21	34	1	2	2	39	56	6
CG-1	2	6			2			
GRAS		2					13	5
B3			1		2		6	
Motif sequence only	82	96	75	99	68	-	-	-
Others	1	1		1	1	-	-	-

[1] Available expression data with gene function annotation. Number of TFs upregulated genes according to each study.

Figures

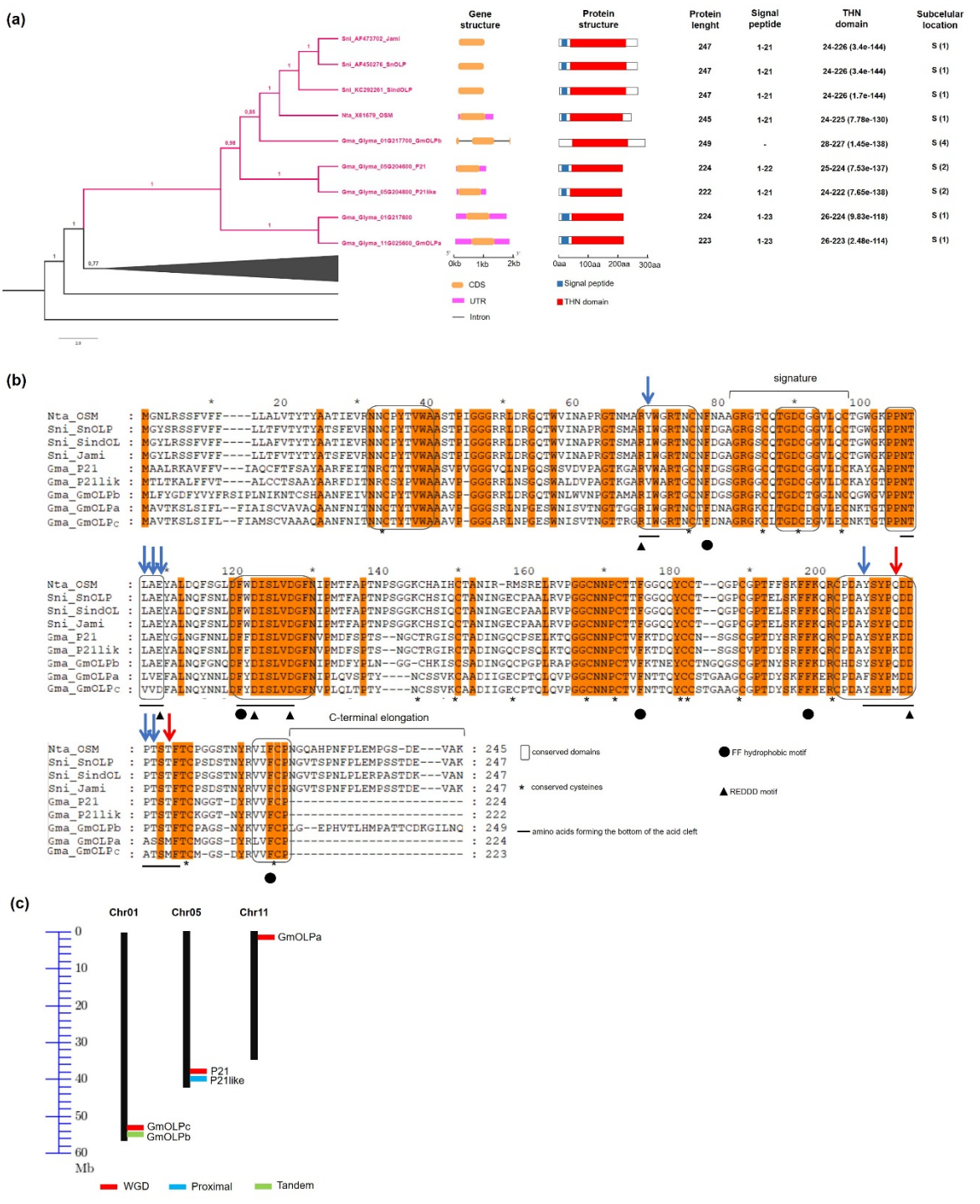


Fig. 1. Osmotin *in silico* analysis. (a) Detailed representation of the phylogenetic relationship among the thaumatin domain within the osmotin monophyletic clade with corresponding gene and protein structure information. Subcellular locations are numbered

according to the TargetP server from one to five, where one indicates the strongest prediction. S, secretory pathway. **(b)** Alignment of osmotin protein sequences. Conserved residues and protein secondary structure are indicated according to Petre et al. [23]. *Orange* letter background corresponds to conserved amino acids (aa) among the nine analysed sequences. *Blue* and *red* arrows indicate aa substitutions with similar and different properties, respectively. **(c)** Chromosomal location and duplication of soybean osmotins. *Black* vertical lines represent the chromosomes with their numbers at the top. WGD, whole-genome duplication.

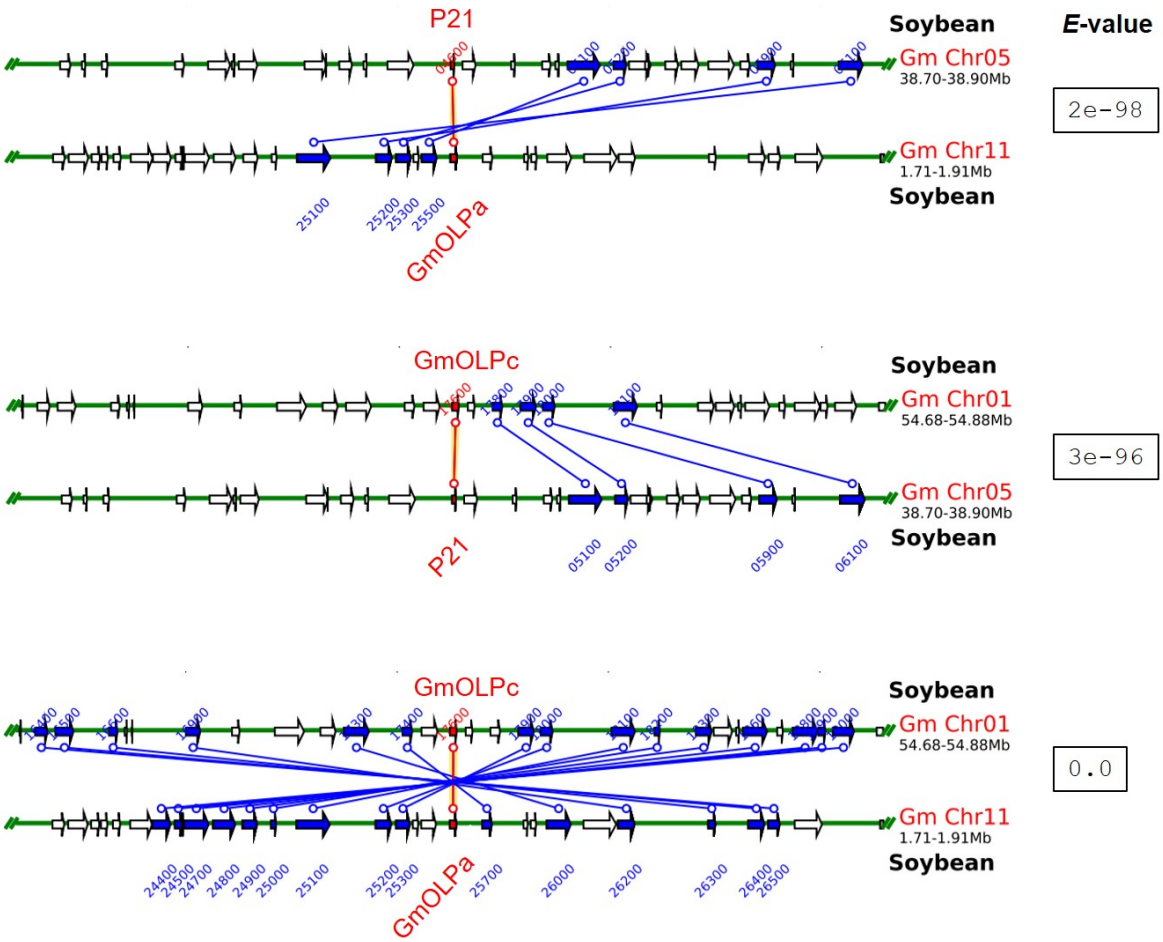


Fig. 2. Syntenic regions between paralogous soybean osmotin genes. *Green* horizontal lines represent the chromosomes. *Blue* and *red* vertical lines and arrows represent the

duplicated paralogous genes. The *red* lines represent the sequence used as the search query (indicated at the top of each syntenic region). The *white* arrow indicates no duplicated genes. The *e*-value for each syntenic region is shown.

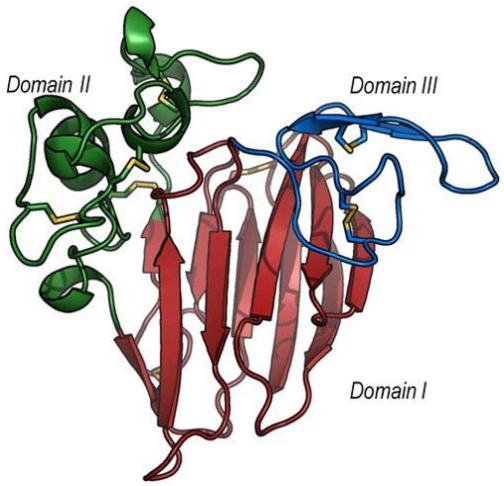


Fig. 3. Three-dimensional structure of Gma_P21. The protein structure is coloured based on its domains: **(i)** domain I (*red*), consisting of a 11 stranded β-sheet organised as a β-barrel that forms the protein core; **(ii)** domain II (*green*), consisting of an α-helix and a set of disulphide-rich loops; and **(iii)** domain III (*blue*), presenting a β-hairpin and a coil region, which are both maintained by one disulphide bond each. The main chain is represented as an illustration and the disulphide bonds are presented as sticks. The Sni_SnOLP, Sni_SindOLP, Sni_Jami, Gma_OLPb, Gma_P21like, Gma_P21, Gma_OLPc, and Gma_OLPb structures share the same topology. The image was generated with PyMOL Molecular Graphics System version 1.5.0.4 (Schrödinger, LLC).

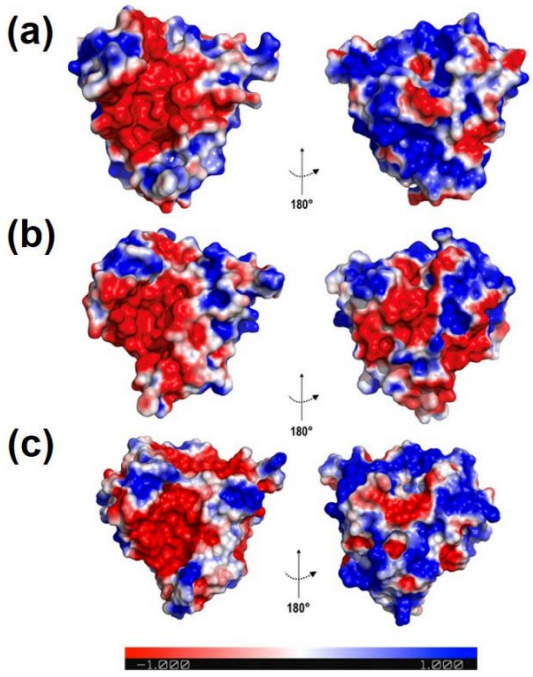
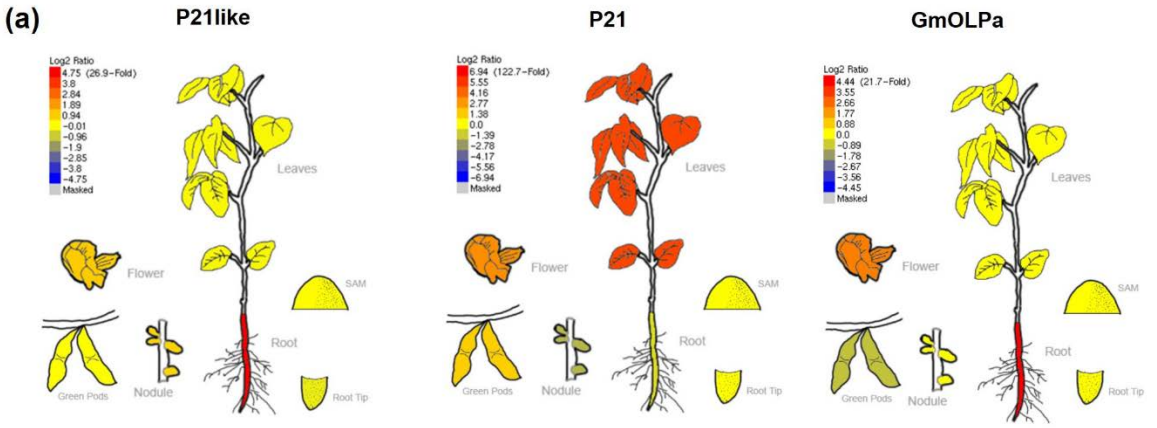


Fig. 4. Electrostatic surface potential of osmotin proteins. The colours indicate charges in the electrostatic surface potential; where *red, blue*, and *white* represent negative, positive, and neutral regions, respectively. **(a)** Sni_SnOLP image representing the electrostatic surface potential of *Solanum nigrum* and *Nicotiana tabacum* osmotins. **(b)** Gma_OLPb image representing the electrostatic surface potential of *Glycine max* osmotins, except Gm_P21-like. **(c)** Gma_P21-like. Images generated with PyMOL Molecular Graphics System version 1.5.0.4 (Schrödinger, LLC).

(a)



(b)

	Young leaf	Flower	One cm pod	Pod shell 10DAF	Pod shell 14DAF	Seed 10DAF	Seed 14DAF	Seed 21DAF	Seed 25DAF	Seed 28DAF	Seed 35DAF	Seed 42DAF	Root	Nodule
P21like	0	19	7	6	1	1	0	0	0	0	1	12	11	47
P21	288	123	129	82	348	12	10	0	0	1	3	9	13	1
GmOLPa	0	57	9	1	0	0	0	0	0	1	0	2	379	46

Fig. 5. *In silico* gene expression analysis. (a) Relative expression profile of soybean osmotin genes from the RNA-seq BAR database. *Red* and *blue* colours indicate upregulated and downregulated genes, respectively (Log ^{2} ratio) in different organs/tissues as plotted in the *bar* scale. **(b)** Digital gene expression counts of the uniquely mappable reads from RNA-Seq Atlas of *Glycine max* raw data. RNA-seq reads were only mapped to the initial genome assembly (i.e. Wm82.a1.v1).

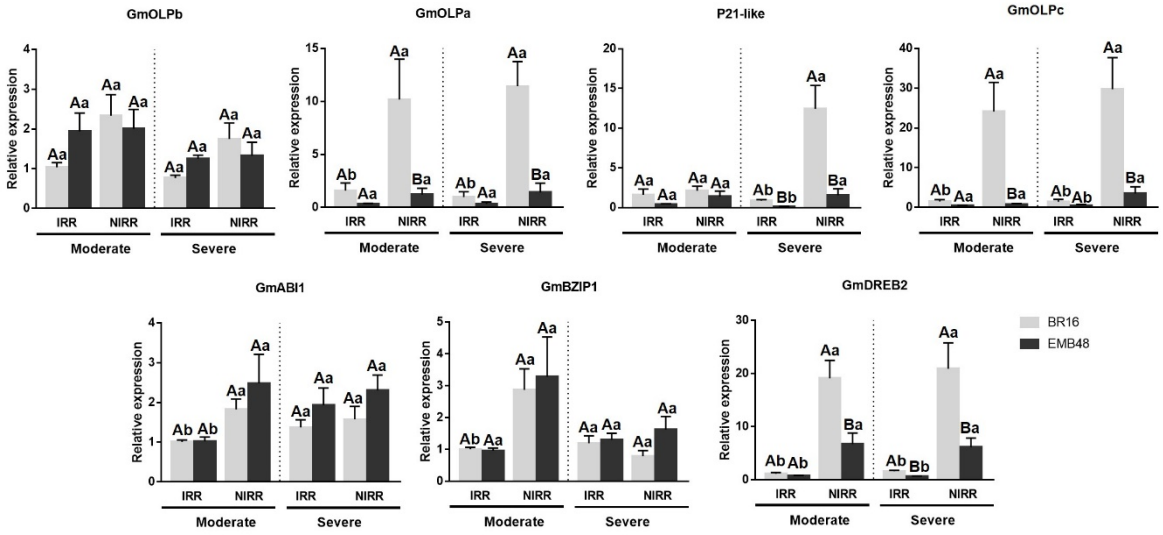


Fig. 6. Gene expression analyses in a greenhouse experiment involving plants at the R1 development stage submitted to moderate and severe stress. The relative expression levels of genes in leaves were measured by RT-qPCR. NIRR, non-irrigated plants. Irrigated (IRR) plants of each cultivar were used as control. CYP2 and ELF1A reference genes were used as internal controls to normalise the amount of mRNA present in each sample. All transcript levels were calibrated in relation to the expression level of BR16 IRR plants under moderate water stress. Data represent the means of four biological replicates with three technical replicates each. Capital letters correspond to the comparison among means of cultivars within the same treatment. Small letters correspond to the comparison among means of treatments within the same cultivar. Means labeled with the same letter do not differ significantly (Duncan's test, $p < 0.05$). Error bars represent the standard error of the mean.

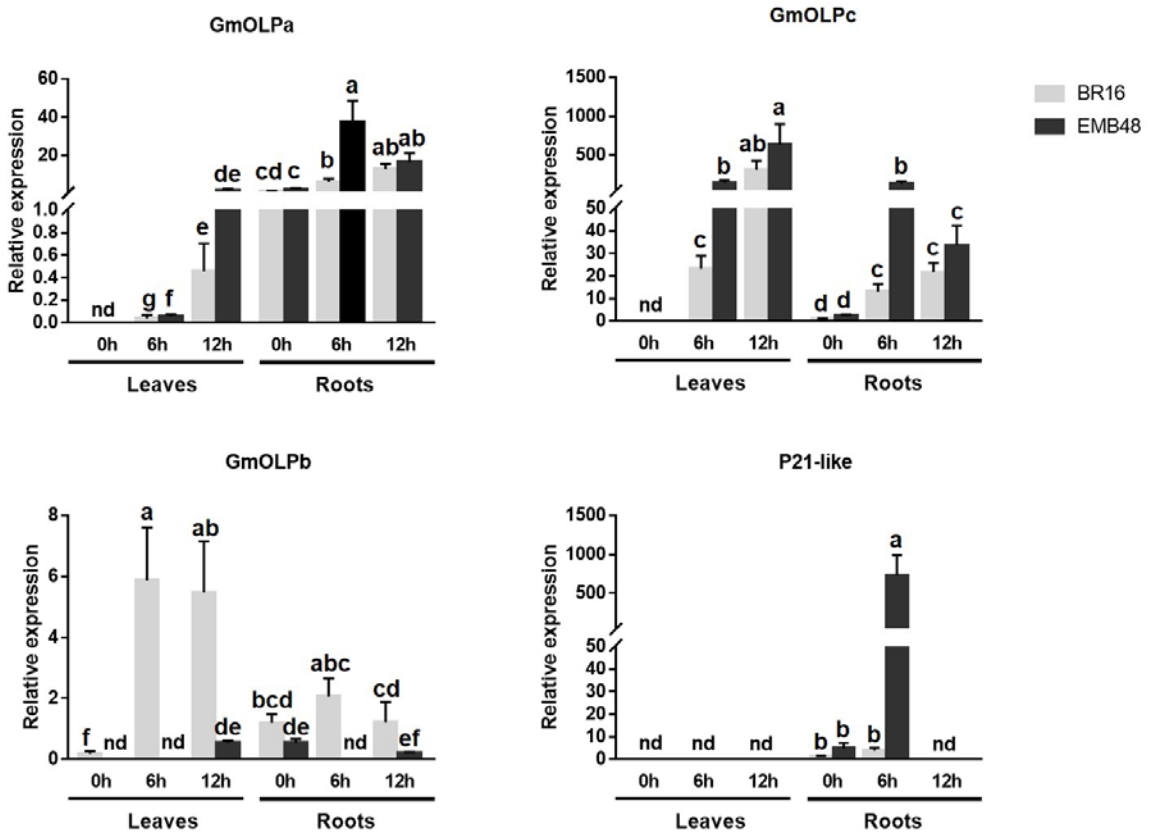


Fig. 7. Gene expression analyses in a growth chamber experiment of plants at the V3 development stage submitted to drought stress. The relative expression levels of genes in leaves and roots were measured by RT-qPCR at 0 (control), 6, and 12 hours after dehydration treatment. CYP2 and ELF1A reference genes were used as internal controls to normalise the amount of mRNA present in each sample. All transcript levels were calibrated in relation to the expression level of BR16 roots at 0h. Data represent the means of five biological replicates with three technical replicates each. Means labeled with the same letter do not differ significantly (Duncan's test, $p < 0.05$). Error bars represent the standard error of the mean. Nd, not detected.

Supporting information

Table S1 Primer set designed for RT-qPCR.

Genes	Prímers (5′→3′)		References
GmOLPb	Forward	CCAATTTGGCAACCAGGATTT	This study
	Reverse	TTGTGACACCCACCGTTTAG	
GmOLPc	Forward	GTGCGGTCCCACAGATTATT	
	Reverse	TTGCATCGTCCATAGGGTAAC	
GmOLPa	Forward	GTTGCGGTCCCACTGATTAT	
	Reverse	GCGTCATCCATAGGGTAACTAAA	
P21like	Forward	TTGTTCTTTGTGACCGCTCTAT	
	Reverse	GGGTATGAGCATCGGTTTGT	
P21	Forward	AAAGCCGTGTTCTTTGTAATCG	
	Reverse	GTGTATGTGCATCGGTTTGTG	
GmDREB2	Forward	GCTGACGTGGCTGGAACTAA	Chen et al. (2007)
	Reverse	TTCCGCTCGCCTTAACTTCC	
GmABI1	Forward	TGTGCCAGAGAAGTAAGCGC	Wang et al. (2010)
	Reverse	GCAAGAATTGAGGCTGCTGG	
GmbZIP1	Forward	ATTGCCACCACTTCCACCAT	Gao et al. (2011)
	Reverse	GCAGGAGGAGTAGAAGGCCA	

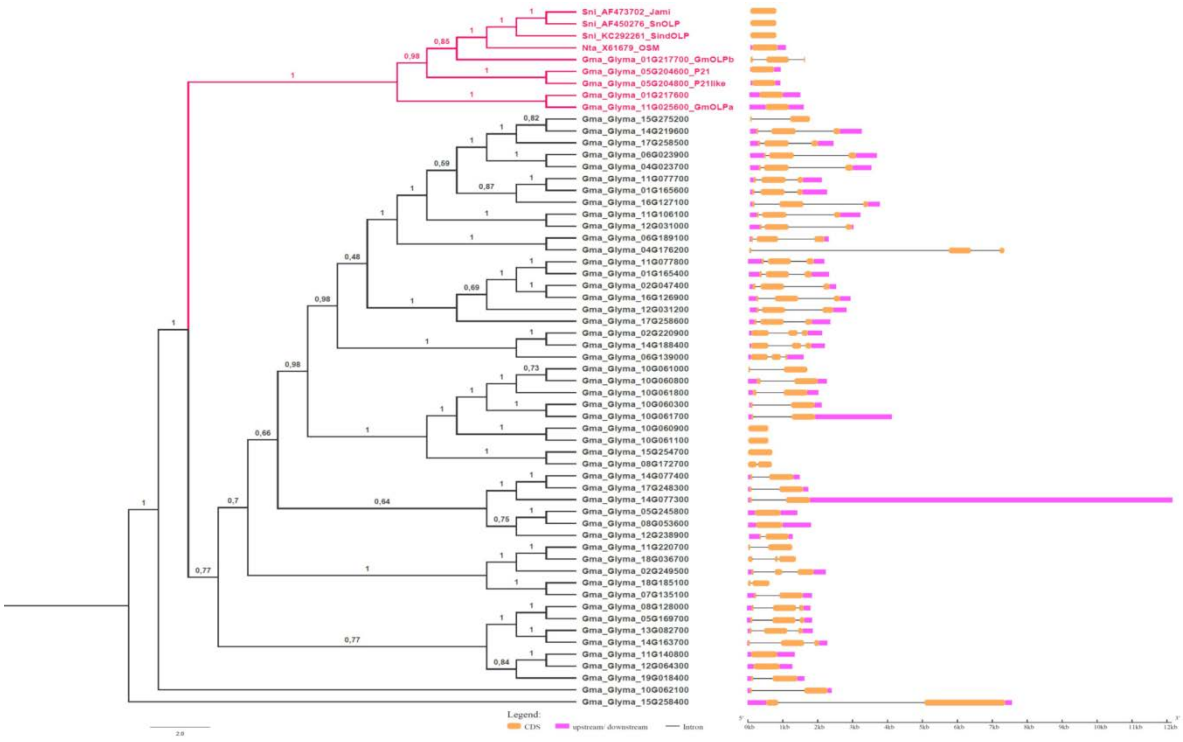


Fig. S1 Phylogenetic tree of 59 thaumatin domain sequences and gene structures from *G. max* and previously characterised *N. tabacum* and *S. nigrum* osmotins. Osmotin group is *pink* coloured according to posterior probability values.

Capítulo IV
CONSIDERAÇÕES FINAIS

CONSIDERAÇÕES FINAIS

Desde sua identificação em 1987, as osmotinas vem sendo estudadas devido ao seu grande potencial biotecnológico para o desenvolvimento de cultivares mais tolerantes a estresses bióticos e abióticos. Ao longo desses anos, muito se descobriu sobre sua distribuição em diferentes organismos e espécies, sobre sua estrutura e função, seu mecanismo de expressão e ação, e sua utilidade biotecnológica. Porém, até então, nenhum estudo havia focado na origem evolutiva das osmotinas e na sua diversificação dentro da família *Thaumatins-like proteins* (TLPs).

No presente estudo, reconstruímos a história evolutiva das TLPs e identificamos a origem do grupo das osmotinas. A partir da árvore filogenética gerada, observamos um complexo padrão de evolução molecular, no qual eventos de duplicação distintos e mecanismos de ganhos e perdas de genes dirigiram a evolução e diversificação da família gênica. A partir das análises, identificamos que as duplicações em tandem e em bloco tiveram um importante papel na expansão das TLPs a partir das embriófitas. Organizações e motivos específicos para alguns subgrupos de proteínas também foram encontrados. Os resultados obtidos permitem hipotetizar que a adaptação a novos ambientes levou à fixação de diferentes aminoácidos e a consequente formação de diferentes subgrupos de proteínas dentro desta família. Para confirmar esta hipótese análises futuras de resíduos de aminoácidos sob seleção positiva devem ser realizadas, no intuito de melhor compreender a evolução e possível diversificação funcional dessas proteínas.

Dentre os subgrupos identificados na árvore filogenética encontra-se o subgrupo das osmotinas, que tiveram sua origem a partir das espermatófitas. Neste subgrupo, encontram-se todas as sequências de proteínas já caracterizadas como osmotinas e outras identificadas como possíveis osmotinas. O acesso a bancos de dados de ontologia e expressão gênica das possíveis osmotinas de Arabidopsis e arroz contribuiu para a caracterização do subgrupo das osmotinas. Apesar disso, ainda se fazem necessárias novas análises de expressão e estudos funcionais das osmotinas de diferentes espécies em

resposta a estresses, a fim de melhor caracterizar funcionalmente este subgrupo de proteínas.

Osmotinas de espécies vegetais distintas já foram relacionadas com repostas a estresses. Dentre elas encontram-se as osmotinas de tabaco e *Solanum nigrum*. Estudos de superexpressão de osmotinas de tabaco demonstraram que as mesmas podem conferir tolerância à seca em plantas transgênicas de variadas espécies vegetais: *Nicotiana tabacum* (Barthakur et al., 2001), *Gossypium hirsutum* L. (Parkhi et al., 2009), *Solanum lycopersicum* (Goel et al., 2010), *Morus indica* (Das et al., 2011), *Daucus carota* L. (Annon et al., 2014), *Camellia sinensis* L. (Bhattacharya et al., 2014), *Olea europaea* L. (Silvestri et al., 2017). Da mesma forma, osmotinas de *Solanum nigrum* tornaram as plantas de *Glycine max* (Weber et al., 2014), *Solanum lycopersicum* (Kumar et al,. 2016) e *Sesamum indicum* (Chowdhury et al., 2017) menos sensíveis à falta de água.

A soja apresenta quatro osmotinas já caracterizadas e denominadas P21, GmOLPa, GmOLPb e P21-like. A partir de nossa análise filogenética confirmamos a presença destas sequências no subgrupo das osmotinas e identificar uma nova sequência até então desconhecida, a qual denominamos GmOLPc.

Estudos prévios com as osmotinas de soja descritas até então, demonstraram seu envolvimento nas respostas ao estresse salino (Onishi et al., 2006; Tachi et al., 2009). Sabe-se que as respostas das plantas ao sal e à seca estão intimamente relacionadas e os mecanismos se sobrepõem. Apesar disso, o papel das osmotinas de soja na resposta ao estresse hídrico e sua relevância na tolerância de plantas à seca continuou desconhecido. No presente estudo, demonstramos que as osmotinas de soja respondem diferentemente ao estresse hídrico, em diferentes órgãos (folha e raiz), estádios de desenvolvimento (V3 e R1) e cultivares com sensibilidades distintas à seca (EMB48, tolerante; BR16, sensível). A partir deste estudo verificamos que os genes GmOLPc e P21-like foram os mais expressos nas folhas e nas raízes, respectivamente, na cultivar tolerante EMB48. Além disso, mostramos que a proteína P21-like apresenta o potencial eletrostático e a carga total mais semelhante às cargas e potenciais das

osmotinas de *Nicotiana tabacum* e *Solanum nigrum*, já caracterizadas como promotoras de tolerância à seca em plantas transgênicas de diferentes espécies. Esses resultados sugerem o envolvimento das osmotinas GmOLPc e P21-like na tolerância à seca em soja. Porém para confirmar seus reais efeitos se faz necessário um estudo funcional dos genes que codificam essas proteínas.

A superexpressão de genes relacionados à resposta a estresses tem se mostrado muito útil não só para a análise funcional desses genes, como também para obter o aumento de tolerância ao estresse em plantas transgênicas (Shinozaki and Yamaguchi-Shinozaki, 2007). Nosso grupo de pesquisa ao longo dos anos vem obtendo sucesso na técnica de transformação pelo método de bombardeamento de partículas. Weber et al. (2014), por exemplo, utilizaram esta técnica para superexpressão do gene da osmotina de *Solanum nigrum* (SnOLP), sob o controle do promotor constitutivo da Ubiquitina 3 de Arabidopsis (UBQ3-P), que resultou na obtenção de plantas de soja menos sensíveis ao estresse hídrico. Outro método implementado no nosso laboratório consiste na obtenção de raízes transformadas, mediada por *Agrobacterium rhizogenes*. Esta técnica se torna interessante uma vez que a percepção e a resposta da planta ao estresse hídrico se dão primeiramente pela raiz. Além disso, o método de obtenção de raízes transformadas por *A. rhizogenes* é uma ferramenta bastante útil para testar o efeito de genes de interesse em um curto espaço de tempo (Weber and Bodanese-Zanettini 2011).

No contexto acima, com o intuito de melhor compreender o papel das osmotinas GmOLPc e P21-like na tolerância à seca em soja foi iniciado o estudo funcional desses genes. As sequências codificadoras foram clonadas a partir do DNA extraído da cultivar tolerante EMB48. Primers específicos foram projetados utilizando o programa Primer3plus adicionando-se a sequência CACC necessária para a clonagem via sistema GateWay (Invitrogen) (Tabela 1).

Tabela 1. Primers para clonagem via sistema GateWay.

Genes	**Primers (5′ 3′)**	
GmOLPc	Forward	CACCATGGCCGTCACGAAAAGC
	Reverse	TTAAGGGCAAAACACAACCCT
P21-like	Forward	CACCATGACTCTCACAAAGGCCT
	Reverse	TCAAGGGCAAAAGACAACCCT

Os vetores pENTR::GmOLPc/P21-like foram enviados para sequenciamento e os resultados alinhados com as sequências dos genes GmOLPc e P21-like obtidas no Phytozome utilizando o programa MEGA 7 para confirmação da correta amplificação. De acordo com as análises realizadas, apenas o gene P21-like foi clonado e recombinado corretamente no vetor pENTR. Este foi recombinado para o vetor de destino pmdc43 resultando no vetor final pmdc43::P21-like que contêm o gene da osmotina P21-like, gene repórter GFP (*green fluorescent protein*), dirigidos pelo promotor de expressão constitutiva 35S do vírus mosaico da couve-flor (P35S CaMV), e gene HygR (que confere resistência a higromicina) para seleção das plantas (Figura 1).

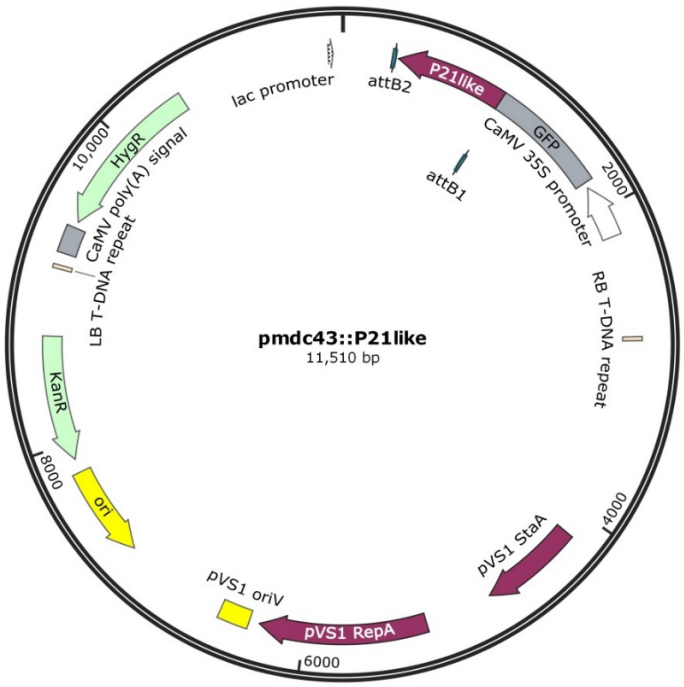


Figura 1 - Vetor pmdc43::P21like. HygR, gene que confere resistência a higromicina para seleção das plantas; KanR, gene que confere resistência a canamicina para seleção das bactérias; CaMV 35S, promotor de expressão constitutiva 35S do vírus mosaico de couve flor; GFP, *green fluorescent protein*.

Na próxima etapa está prevista a obtenção de plantas compostas com raízes transgênicas da cultivar sensível BR16 via transformação por *A. rhizogenes*. As plantas compostas serão submetidas ao estresse hídrico e a avaliadas quanto ao fenótipo. Espera-se que as plantas com raízes transformadas superexpressando o gene da osmotina P21-like sejam menos sensíveis à falta de água do que as plantas com raízes transformadas com o vetor vazio (sem o gene de interesse P21-like). Novos experimentos de localização subcelular em protoplastos de *A. thaliana*, duplo híbrido e transformação estável de plantas de soja via bombardeamento visando a superexpressão e/ou silenciamento do gene P21-like deverão ser realizados. Além disso, a clonagem do gene GmOLPc deverá ser repetida para a realização de experimentos de transformação, a fim de confirmar a possível relação do gene com a tolerância ao estresse hídrico em plantas de soja.

REFERÊNCIAS

Abdin MZ, Kiran U, Alam A. 2011. Analysis of osmotin, a PR protein as metabolic modulator in plants. Bioinformation 5(8), 336.

Adams J, Mansfield MJ, Richard DJ, Doxey AC. 2017. Lineage-specific mutational clustering in protein structures predicts evolutionary shifts in function. Bioinformatics 33(9), 1338-1345.

Ahmed NU, Park JI, Jung HJ, Chung MY, Cho YG, Nou IS. 2013. Characterization of Thaumatin-like Gene Family and Identification of *Pectobacterium carotovorum* subsp. *carotovorum* Inducible Genes in *Brassica oleracea*. Plant Breed Biotech. 1(2), 111-121.

Annon A, Rathore K, Crosby K. 2014. Overexpression of a tobacco osmotin gene in carrot (*Daucus carota* L.) enhances drought tolerance. In Vitro Cellular & Developmental Biology-Plant 50(3), 299-306.

Anzlovar and Dermastia. 2003. The comparative analysis of osmotins and osmotin-like PR-5 proteins. Plant biol. 5, 116-124.

Barthakur S, Babu V, Bansa K. 2001. Over-expression of osmotin induces proline accumulation and confers tolerance to osmotic stress in transgenic tobacco. J. Plant Biochem. Biotechnol. 10, 31–37.

Batalia MA, Monzingo AF, Ernst S, Roberts W, Robertus JD. 1996. The crystal structure of the antifungal protein zeamatin, a member of the thaumatin-like, PR-5 protein family. Nature Structural and Molecular Biology 3(1), 19.

Bencke-Malato M, Cabreira C, Wiebke-Strohm B, Bücker-Neto L, Mancini E, Osorio MB, Homrich MS, Turchetto-Zolet AC, De Carvalho MCCG, Stolf R, Weber RLM, Westergaard G, Castagnaro AP, Abdelnoor RV, Marcelino-Guimarães FC, Margis-Pinheiro M, Bodanese-Zanettini MH. 2014. Genome-wide annotation of the soybean WRKY family and functional characterization of genes involved in response to Phakopsora pachyrhizi infection. BMC plant biology 14(1), 236.

Bhattacharya A, Saini U, Joshi R, Kaur D, Pal AK, Kumar N, Gulati A, Mohanpuria P, Yadav SK, Kumar S, Ahuja PS. 2014. Osmotin-expressing transgenic tea plants have improved stress tolerance and are of higher quality. Transgenic Res. 23, 211–223.

Brando PM, Goetz SJ, Baccini A, Nepstad DC, Beck PSA, Christman MC: 2010. Seasonal and interannual variability of climate and vegetation indices cross the Amazon. Proc Natl Acad Sci USA 107, 14685–14690.

Breiteneder H. 2004. Thaumatin-like proteins–a new family of pollen and fruit allergens. Allergy 59(5), 479-481.

Câmara GMS, Heiffig LS. 2000. Fisiologia, ambiente e rendimento da cultura da soja. In: Câmara G M S. Soja: tecnologia da produção II. Piracicaba: ESALQ/LPV, p.450.

Cao J, Lv Y, Hou Z, Li X, Ding L. 2016. Expansion and evolution of thaumatin-like protein (TLP) gene family in six plants. Plant Growth Regul 79, 299–307.

Chen WP, Chen PD, Liu DJ, Kynast R, Friebe B, Velazhahan R, Muthukrishnan S, Gill BS. 1999. Development of wheat scab symptoms is delayed in transgenic wheat plants that constitutively express a rice thaumatin-like protein gene. Theor Appl Genet 99, 755–760.

Chen W, Yao Q, Patil GB, Agarwal G, Deshmukh RK, Wang LLB, Wang Y, Prince SJ, Song L, Xu D, An YC, Valliyodan B, Varshney RK, Nguyen HT. 2016. Identification and comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-Seq. Front Plant Sci., 7, 1044.

Chowdhury S, Basu A, Kundu S. 2017. Overexpression of a new osmotin-like protein gene (SindOLP) confers tolerance against biotic and abiotic stresses in sesame. Frontiers in Plant Science 8, 410.

Conab - Compania Nacional de Abastecimento, acompanhamento da safra brasileira de grãos, safra 2017/18, disponível em <www.conab.gov.br>

Das M, Chauhan H, Chhibbar A, Haq QMR, Khurana P. 2011. High-efficiency transformation and selective tolerance against biotic and abiotic stress in mulberry, *Morus indica* cv. K2, by constitutive and inducible expression of tobacco osmotin. Transgenic research 20(2), 231-246.

Goel D, Singh AK, Yadav V, Babbar SB, Bansal KC. 2010. Overexpression of osmotin gene confers tolerance to salt and drought stresses in transgenic tomato (*Solanum lycopersicum* L.). Protoplasma 245, 133–141.

Graham JS, Burkhart W, Xiong J, Gillikin JW. 1992. Complete amino acid sequence of soybean leaf P21. Plant Physiol. 98, 163-165.

Guimarães-Dias F, Neves-Borges AC, Viana AA, Mesquita RO, Romano E, de Fátima Grossi-de-Sá M, Nepomuceno AL, Loureiro ME, Alves-Ferreira M. 2012. Expression analysis in response to drought stress in soybean: Shedding light on the regulation of metabolic pathway genes. Genet Mol Biol 35, 222–232.

Gutha LR, Reddy AR. 2008. Rice DREB1B promoter shows distinct stress-specific responses, and the overexpression of cDNA in tobacco confers improved abiotic and biotic stress tolerance. Plant molecular biology 68(6), 533.

Hakim, Ullah A, Hussain A, Shaban M, Khan AH, Alariqi M, Gul S, Jun Z, Lin S, Li J, Jin S, Munis MFH. 2018. Osmotin: A plant defense tool against biotic and abiotic stresses. Plant Physiology & Biochemistry 123, 149-159.

Hoffmann-Sommergruber K. 2002. Pathogenesis-related (PR)-proteins identified as allergens. Biochem Soc Trans 30, 930–935.

Hua L, Challa GS, Subramanian S, Gu X, Li W. 2018. Genome-Wide Identification of Drought Response Genes in Soybean Seedlings and Development of Biomarkers for Early Diagnoses. Plant Molecular Biology Reporter 36(2), 350-362.

Ibeas JI, Lee H, Damsz B, Prasad DT, Pardo JM, Hasegawa PM, Bressan RA. Narasimhan ML. 2000. Fungal cell wall phosphomannans facilitate the toxic activity of a plant PR-5 protein. The Plant Journal 23(3), 375-383.

Jami SK, Anuradha TS, Guruprasad L, Kirti PB. 2007. Molecular, biochemical and structural characterization of osmotin-like protein from black nightshade (*Solanum nigrum*). Journal of plant physiology 164(3), 238-252.

Ku YS, Au-Yeung WK, Yung YL, Li MW, Wen CQ, Liu X, Lam HM. 2013. Drought stress and tolerance in soybean. In A comprehensive survey of international soybean research-genetics, physiology, agronomy and nitrogen relationships. InTech.

Kumar AS, Kumari HP, Kumar SG, Mohanalatha C, Kishor KPB. 2015. Osmotin: a plant sentinel and a possible agonist of mammalian adiponectin. Frontiers in plant science 6, 163.

Kumar SA, Kumari PH, Jawahar G, Prashanth S, Suravajhala P, Katam R, Sivan P, Rao KS, Kirti PB, Kishor PBK. 2016. Beyond just being foot soldiers – osmotin like protein (OLP) and chitinase (Chi11) genes act as sentinels to confront salt, drought, and fungal stress tolerance in tomato. Environmental and Experimental Botany 132, 53-65.

Le TT, Williams B, Mundree SG. 2018. An osmotin from the resurrection plant *Tripogon loliiformis* (TlOsm) confers tolerance to multiple abiotic stresses in transgenic rice. Physiologia plantarum 162(1), 13-34.

Leone P, Menu-Bouaouiche L, Peumans WJ, Payan F, Barre A, Roussel A, Van Damme EJM, Rougé P. 2006. Resolution of the structure of the allergenic and antifungal banana fruit thaumatin-like protein at 1.7-Å. Biochimie 88, 45–52.

Liu JJ, Sturrock R, Ekramoddoullah AK. 2010. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant cell reports 29(5), 419-436.

Manavalan LP, Guttikonda SK, Tran LSP, Nguyen HT. 2009. Physiological and molecular approaches to improve drought resistance in soybean. Plant Cell Physiol 2009 50, 1260–1276.

Mani T, Sivakumar KC, Manjula S. 2012. Expression and functional analysis of two osmotin (PR5) isoforms with differential antifungal activity from Piper colubrinum: prediction of structure–function relationship by bioinformatics approach. Molecular biotechnology 52(3), 251-261.

Menu-Bouaouiche L, Vriet C, Peumans WJ, Barre A, van Damme EJM, Rouge´ P (2003) A molecular basis for the endo-beta 1,3-glucanase activity of the thaumatin-like proteins from edible fruits. Biochimie 85,123–131.

Min K, Ha SC, Hasegawa PM, Bressan RA, Yun DJ, Kim KK. 2004. Crystal structure of osmotin, a plant antifungal protein. Proteins: Structure, Function, and Bioinformatics 54(1), 170-173.

Misra RC, Kamthan M, Kumar S, Ghosh S. 2016. A thaumatin-like protein of *Ocimum basilicum* confers tolerance to fungal pathogen and abiotic stress in transgenic Arabidopsis. Scientific Reports 6, 25340.

Moon SJ, Han SY, Kim DY, Kim BG, Yoon IS, Shin D, Kwon HB, Byun MO. 2014. Ectopic expression of CaWRKY1, a pepper transcription factor, enhances drought tolerance in transgenic potato plants. Journal of Plant Biology 57(3), 198-207.

Muoki RC, Paul A, Kumar S. 2012. A shared response of thaumatin like protein, chitinase, and late embryogenesis abundant protein3 to environmental stresses in tea [*Camellia sinensis* (L.) O. Kuntze]. Functional & integrative genomics 12(3), 565-571.

Narasimhan ML, Damsz B, Coca MA, Ibeas JI, Yun DJ, Pardo JM, Hasegawa PM, Bressan RA. 2001. A plant defense response effector induces microbial apoptosis. Mol Cell 8, 921–930.

Narasimhan M, Coca M, Jin J, Yamauchi T, Ito Y, Kadowaki T, Kim K, Pardo J, Damsz B, Hasegawa P. 2005. Osmotin is a homolog of mammalian adiponectin and controls apoptosis in yeast through a homolog of mammalian adiponectin receptor. Mol Cell 17, 171–180.

Neale AD, Wahleithner JA, Lund M, Bonnett HT, Kelly A, Meeks-Wagner DR, Peacock WJ, Dennisa ES. 1990. Chitinase, 6-1,3-Glucanase, Osmotin, and Extensin Are Expressed in Tobacco Explants during Flower Formation. The Plant Cell 2, 673-684.

Onishi M, Tachi H, Kojima T, Shiraiwa M, Takahara H. 2006. Molecular cloning and characterization of a novel salt-inducible gene encoding an acidic isoform of PR-5 protein in soybean. Plant Physiol Biochem. 44, 574–580.

Pathan MS, Lee JD, Shannon JG, Nguyen HT. 2007. Recent advances in breeding for drought and salt stress tolerance in soybean. In Advances in molecular breeding toward drought and salt tolerant crops. Edited by Jenks MA, Hasegawa PM, Jain SM. Springer: New York, 739 – 773.

Parkhi V, Kumar V, Sunilkumar G, Campbell LM, Singh NK, Rathore KS. 2009. Expression of apoplastically secreted tobacco osmotin in cotton confers drought tolerance. Mol Breeding 23,625–639.

Petre B, Major I, Rouhier N, Duplessis S. 2011. Genome-wide analysis of eukaryote thaumatin like proteins (TLPs) with an emphasis on poplar. BMC Plant Biology 11, 33.

Ramos MV, Oliveira RSB, Pereira HM, Moreno FBMB, Lobo MDP, Rebelo LM, Brandão-Neto J, Sousa JS, Monteiro-Moreira ACO, Freitas CDT, Grangeiro TB. 2015. Crystal structure of an antifungal osmotin-like protein from *Calotropis procera* and its effects on *Fusarium solani* spores, as revealed by atomic force microscopy: Insights into the mechanism of action. Phytochemistry 119, 5–18.

Reetz ER. 2012. Anuário brasileiro da soja 2012. Editora Gazeta Santa Cruz, p 156.

Salzman RA, Koiwa H, Ibeas JI, Pardo JM, Hasegawa PM, Bressan RA. 2004. Inorganic cations mediate plant PR5 protein antifungal activity through fungal Mnn1-and Mnn4-regulated cell surface glycans. Molecular plant-microbe interactions 17(7), 780-788.

Sels J, Mathys J, De Coninck BMA, Cammue BPA, De Bolle MFC. 2008. Plant pathogenesis-related (PR) proteins: A focus on PR peptides. Plant Physiology and Biochemistry 46, 941–950.

Shatters RG, Boykin LM, Lapointe SL, Hunter WB, Weathersbee AA. 2006. Phylogenetic and structural relationships of the PR5 gene family reveal an ancient multigene family conserved in plants and select animal taxa. Journal of molecular evolution 63(1), 12-29.

Shin JH, Vaughn JN, Abdel-Haleem H, Chavarro C, Abernathy B, Kim KD, Jackson SA, Li Z. 2015. Transcriptomic changes due to water deficit define a general soybean response and accession-specific pathways for drought avoidance. BMC Plant Biology 15, 26.

Shinozaki K, Yamaguchi-Shinozaki K. 2007. Gene networks involved in drought stress response and tolerance. Journal of experimental botany, 58(2), 221-227.

Silvestri C, Celletti S, Cristofori V, Astolfi S, Ruggiero B, Rugini E. 2017. Olive (*Olea europaea* L.) plants transgenic for tobacco osmotin gene are less sensitive to in vitro-induced drought stress. Acta Physiol. Plant. 39, 229.

Singh NK, Bracker CA, Hasegawa PM, Handa AK, Buckel S, Hermodson MA, Pfankoch E, Regnier FE, Bressan RA. 1987. Characterization of Osmotin: A Thaumatin-like protein associated with osmotic adaptation in plant cells. Plant Physiol. 85, 529-536.

Slootstra JW, De Geus P, Haas H, Verrips CT, Meloen RH. 1995. Possible active site of the sweet-tasting protein thaumatin. Chem. Senses. 20, 535–543.

Stintzi A, Heitz T, Kauffmann S, Legrand M, Fritig B. 1991. Identification of a basic pathogenesis-related, thaumatin-like protein of virus-infected tobacco as osmotin. Physiological and Molecular Plant Pathology 38 (2), 137-146.

Tachi H, Yamada KF, Kojima T, Shiraiwa M, Takahara H. 2009. Molecular characterization of a novel soybean gene encoding a neutral PR-5 protein induced by high-salt stress. Plant Physiol Biochem., 47, 73–79.

Trudel J, Grenier J, Potvin C, Asselin A. 1998. Several thaumatin-like proteins bind to b-1, 3-glucans. Plant Physiology,118, 1431–1438.

Van Loon LC, Pierpoint WS, Boller T, Conejero V. 1994. Recommendations for naming plant pathogenesis-related proteins. Plant Mol. Biol. Reporter, 12, 245-264.

Velazhahan R, Datta SK, Muthukrishnan S. 1999. The PR-5 family: thaumatin-like proteins. Pathogenesis-related proteins in plants, 107-129.

Vitali A, Pacini L, Bordi E, De Mori P, Pucillo L, Maras B, Botta B, Brancaccio A, Giardina B. 2006. Purification and characterization of an antifungal thaumatin-like protein from Cassia didymobotrya cell culture. Plant Physiology and Biochemistry 44(10), 604-610.

Wang X, Zafian P, Choudhary M, Lawton M. 1996. The PR5K receptor protein kinase from Arabidopsis thaliana is structurally related to a family of plant defense proteins. Proceedings of the National Academy of Sciences 93(6), 2598-2602.

Weber RLM, Bodanese-Zanettini MH. 2011. Induction of transgenic hairy roots in soybean genotypes by *Agrobacterium rhizogenes*-mediated transformation. Pesquisa Agropecuária Brasileira 46, 1070-1075.

Weber RLM, Wiebke-Strohm B, Bredemeier C, Margis-Pinheiro M, Brito GG, Rechenmacher C, Bertagnolli PF, Sá MEL, Campos MA, Amorim RMS, Beneventi MA, Margis R, Grossi-de-Sa MF, Bodanese-Zanettini MH. 2014. Expression of an osmotin-like protein from Solanum nigrum confers drought tolerance in transgenic soybean. BMC Plant Biology 14, 343.

Wiebke-Strohm B, Droste A, Pasquali G, Osório MB, Bücker-Neto L, Passaglia LMP, Bencke M, Homrich MS, Margis-Pinheiro M, Bodanese-Zanettini MH. 2011. Transgenic fertile soybean plants derived from somatic embryos transformed via the combined DNA-free particle bombardment and Agrobacterium system. Euphytica (Wageningen) 177, 343-354.

Yun DJ, Ibeas JI, Lee H, Coca MA, Narasimhan ML, Uesono Y, Hasegawa PM, Pardo JM, Bressan RA. 1998. Osmotin, a plant antifungal protein, subverts signal transduction to enhance fungal cell susceptibility. Molecular cell 1(6), 807-817.

Zhang H, Huang Z, Xie B, Chen Q, Tian X, Zhang X, Zhang H, Lu X, Huang D, Huang R. 2004. The ethylene-, jasmonate-, abscisic acid-and NaCl-responsive tomato transcription factor JERF1 modulates expression of GCC box-containing genes and salt tolerance in tobacco. Planta 220, 262–270.

Zhao JP, Su XH. 2010. Patterns of molecular evolution and predicted function in thaumatin-like proteins of *Populus trichocarpa*. Planta 232(4), 949-962.

Zhu JK. 2002. Salt and drought stress signal transduction in plants. Annual review of plant biology 53(1), 247-273.