Molecular Genetics and Genomics (2019) 294:1137–1157
https://doi.org/10.1007/s00438-019-01554-y

ORIGINAL ARTICLE

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[2] · Luisa Abruzzi de Oliveira-Busatto[1] · Maria Helena Bodanese-Zanettini[1]

Received: 14 November 2018 / Accepted: 25 March 2019 / Published online: 27 April 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

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

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00438-019-01554-y) contains supplementary material, which is available to authorized users.

✉ Maria Helena Bodanese-Zanettini
maria.zanettini@ufrgs.br

Giulia Ramos Faillace
giuliafaillace@yahoo.com.br

Andreia Carina Turchetto-Zolet
aturchetto@gmail.com

Frank Lino Guzman
frank.guzman.e@gmail.com

Luisa Abruzzi de Oliveira-Busatto
luisaabruzzi@gmail.com

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

[2] Programa de Pós-Graduação em Biologia Celular e Molecular, Centro de Biotecnologia (CBiot), Porto Alegre, Brazil

Abbreviations

TLP	Thaumatin-like proteins
OLPs	Osmotin-like proteins
PR-5	Protein family 5
AA	Amino acids
CDS	Coding sequences
WGD	Whole-genome duplication

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 daniellii* (Cao et al. 2016). Most typical proteins from the PR-5 family—also called thaumatin-like proteins (TLPs)—have

molecular masses ranging from 21 to 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), 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).

Thaumatin-like proteins 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 ten 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 TLP 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 query 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 *A. 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* was 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 sequence 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. 2018) 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 into 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.

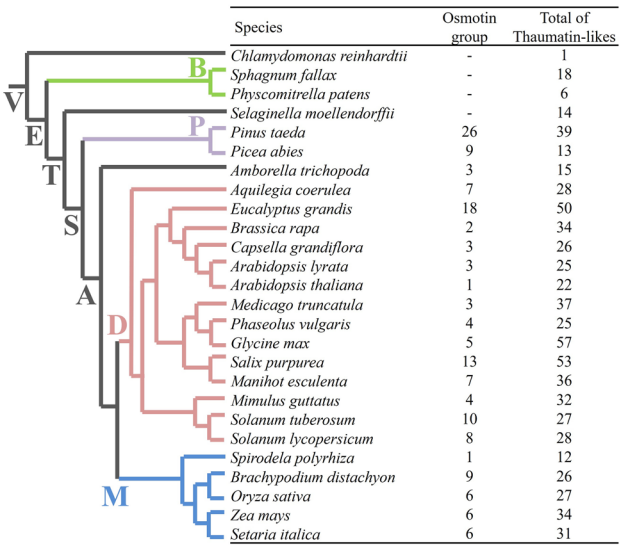
Results

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 *P. 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 (Fig. 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 (Fig. 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 (Fig. 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 26 (*P. taeda*) (Figs. 1, 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 (Fig. 3). The majority of *P. abies* and *P. taeda* sequences grouped in the osmotin group (Fig. 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



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 (color figure online)

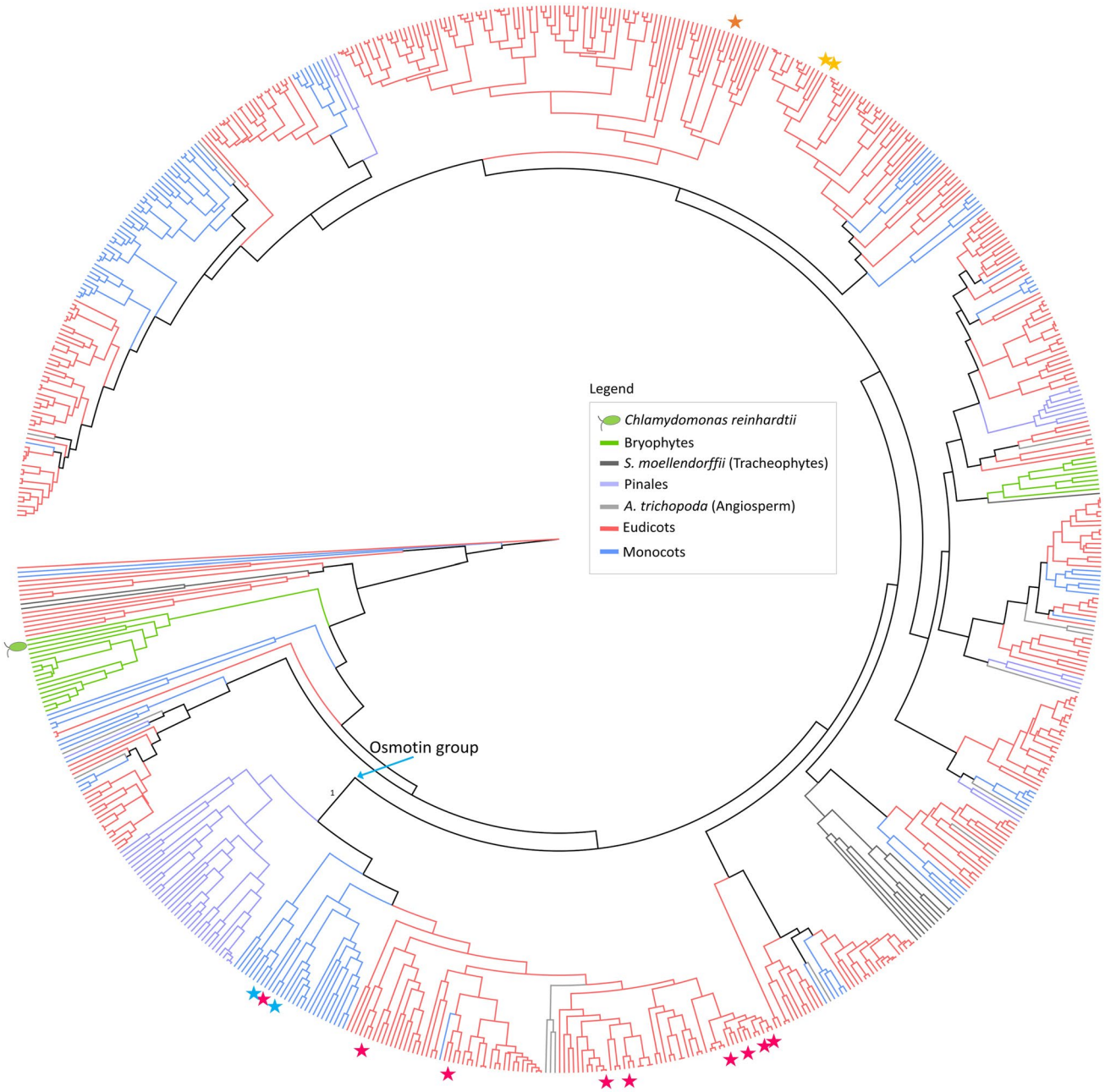


Fig. 2 Phylogenetic tree comprising 722 thaumatin domains from 32 plant species. Pink, blue, yellow, and orange stars indicate the previously characterized permatin, osmotin, allergenic, and kinase proteins, respectively. Blue arrow indicates the osmotin group. For clarity, protein names and posterior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

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 (Figs. 1 and S1).

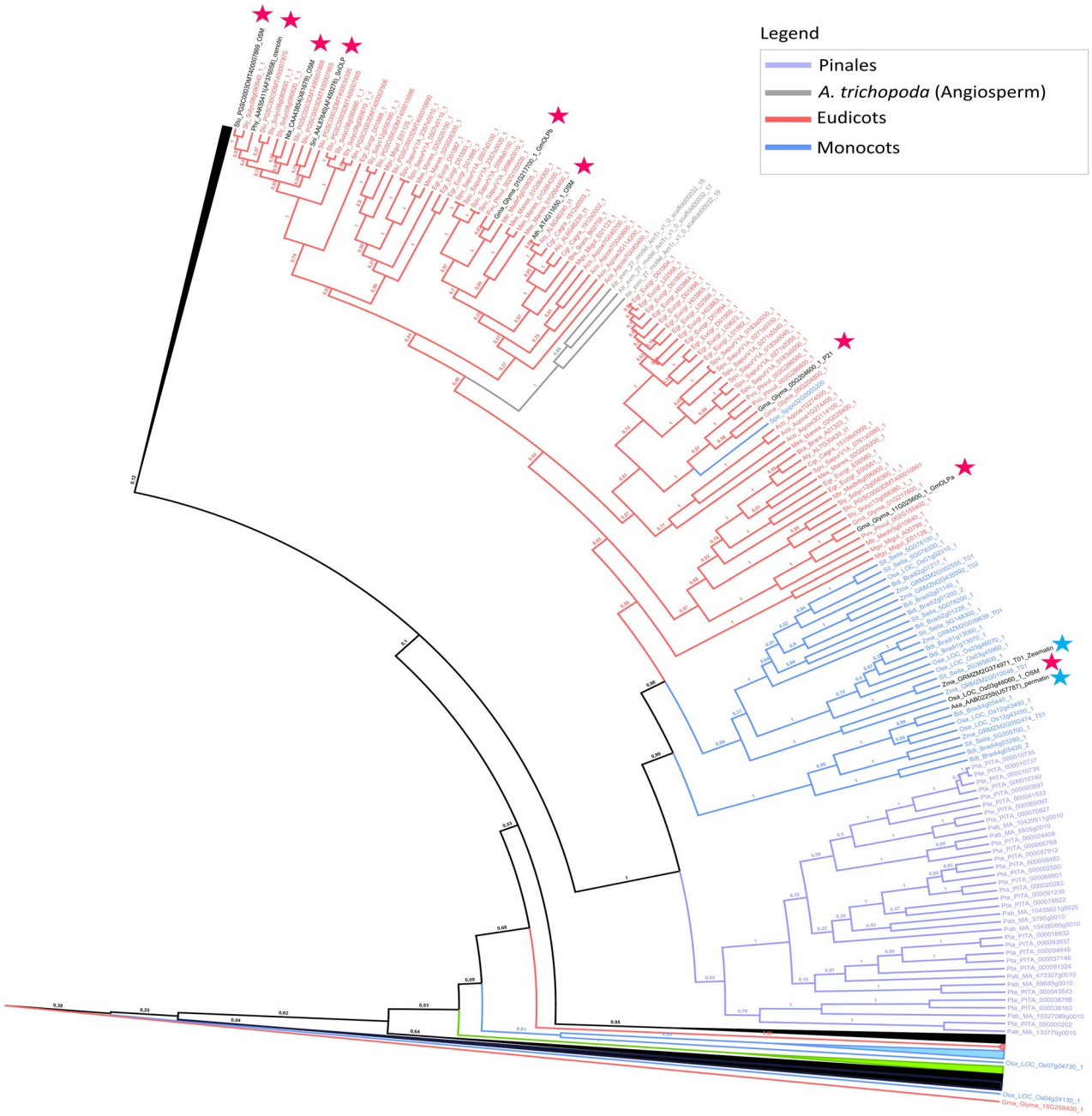


Fig. 3 Detailed representation of the phylogenetic relationship among thaumatin domains within the osmotin group. Branch lengths are proportional to phylogenetic distances. Previously characterized perматins and osmotins are indicated by blue and pink stars, respectively (color figure online)

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 (Fig. 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

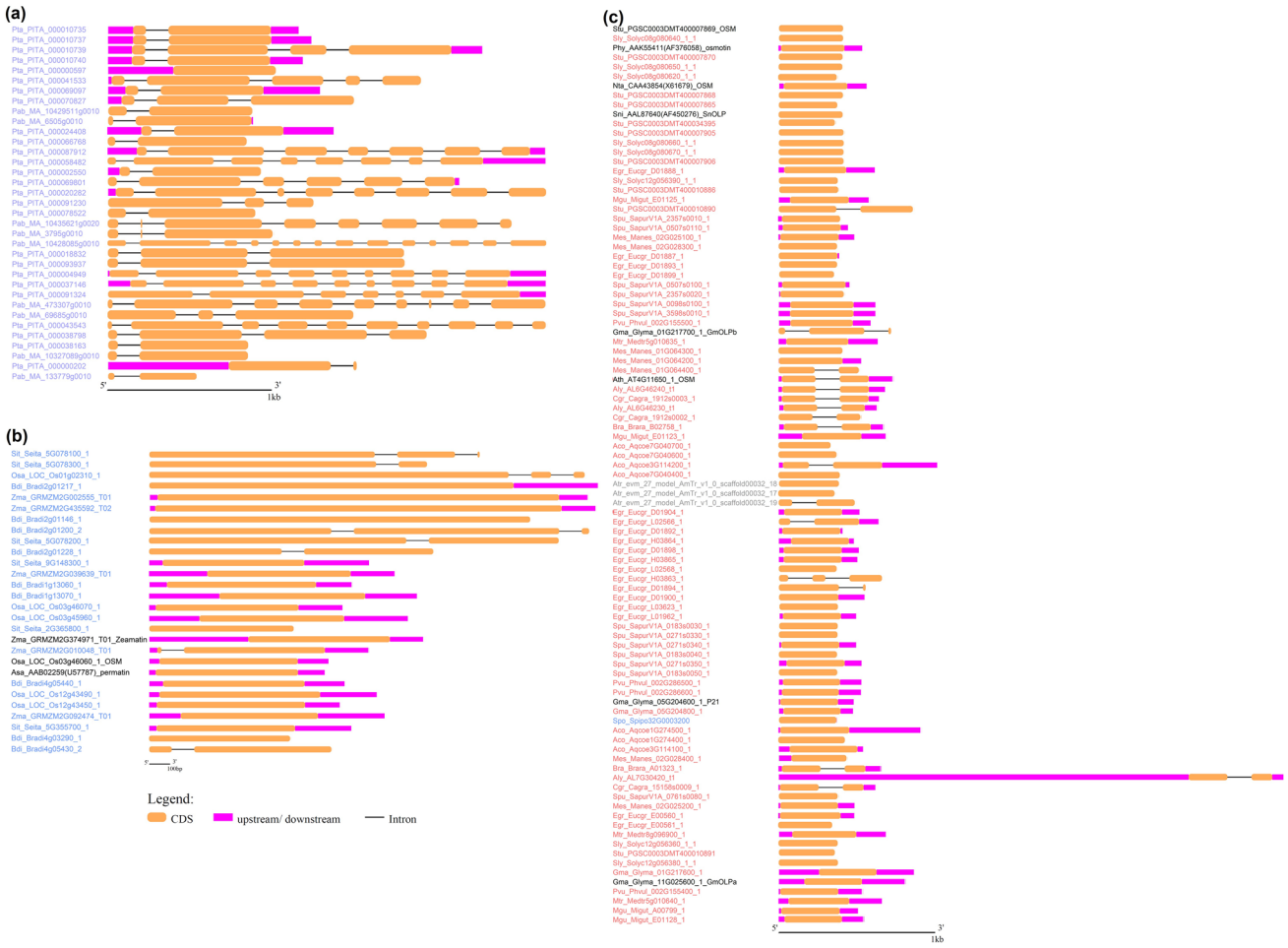


Fig. 4 Gene structure of putative osmotins according to the phylogenetic tree organization. **a** Pinales, **b** monocots, and **c** eudicots, and *Spirodela polyrhiza*

present introns. Five and three monocot sequences presented one and two introns, respectively (Fig. 4b). Fifteen and two eudicot sequences presented one and two introns, respectively (Fig. 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

Table 1 Gene duplication pattern and protein sequence structures of osmotin group

Sequences according tree position	Duplication pattern	Lenght	MW (kDa)	pI	Signal peptide	Subcelular location[a]	Trans-membrane portion	Domains[b]	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

Table 1 (continued)

Sequences according tree position	Duplication pattern	Lenght	MW (kDa)	pI	Signal peptide	Subcelular location[a]	Transmembrane portion	Domains[b]	REDDD motif					pI acid cleft
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

Table 1 (continued)

Sequences according tree position	Duplication pattern	Lenght	MW (kDa)	pI	Signal peptide	Subcelular location[a]	Transmembrane portion	Domains[b]	REDDD motif					pI acid cleft
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

Table 1 (continued)

Sequences according tree position	Duplication pattern	Lenght	MW (kDa)	pI	Signal peptide	Subcelular location[a]	Transmembrane portion	Domains[b]	REDDD motif					pI acid cleft
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

Table 1 (continued)

Sequences according tree position	Duplication pattern	Lenght	MW (kDa)	pI	Signal peptide	Subcelular location[a]	Transmembrane portion	Domains[b]	REDDD motif					pI acid cleft
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

[a]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

[b]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

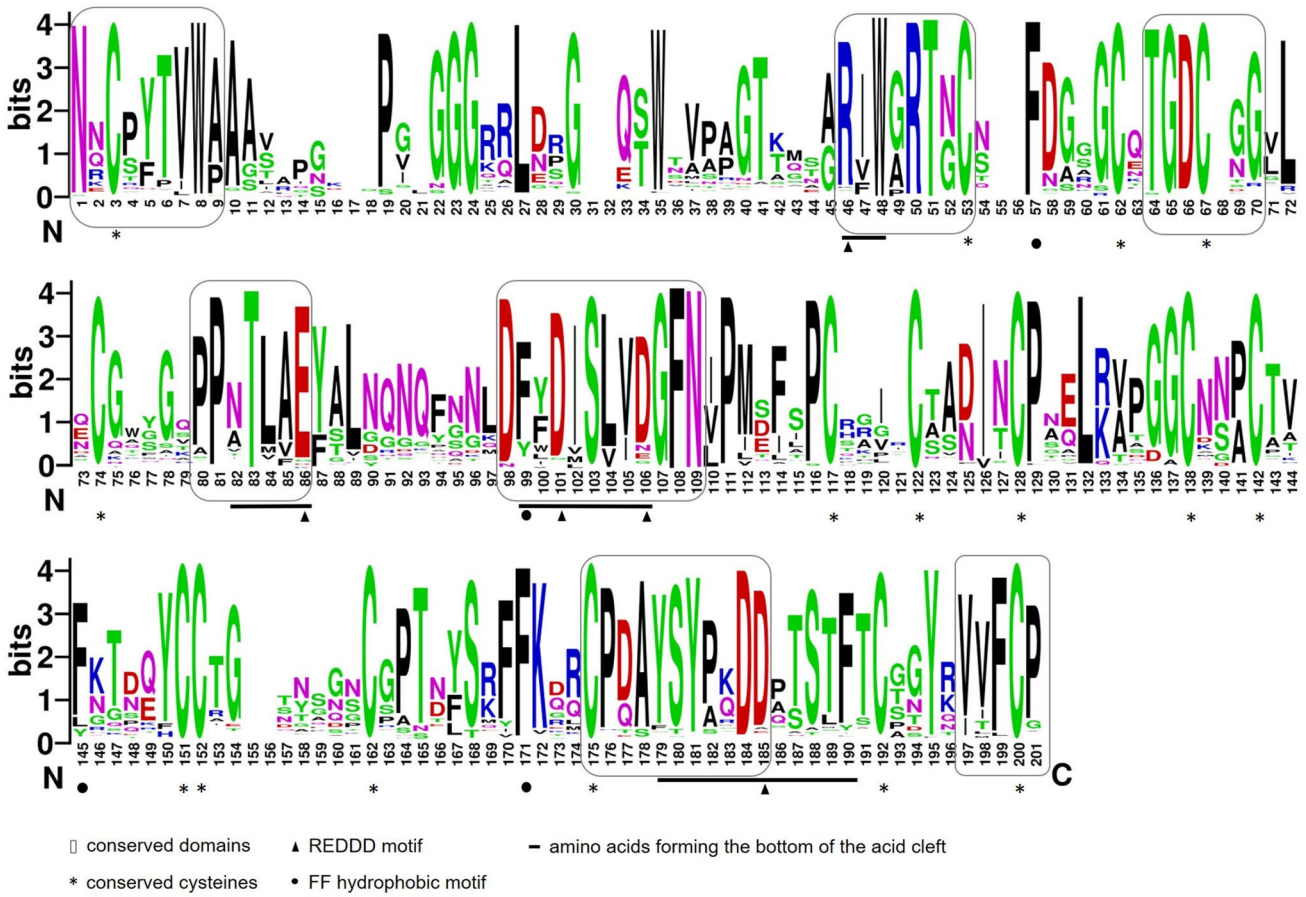


Fig. 5 Logo of thaumatin domain sequence alignments of the osmotin group. The logo alignment residue numeration does not represent the canonical domain sequence numbering. Conserved residues and protein secondary structure are indicated according to Petre et al. (2011). Amino acid frequency was analyzed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

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 (Fig. 5). Notably, some putative osmotin sequences did not present all REDDD residues (Fig. 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 (Figs. 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

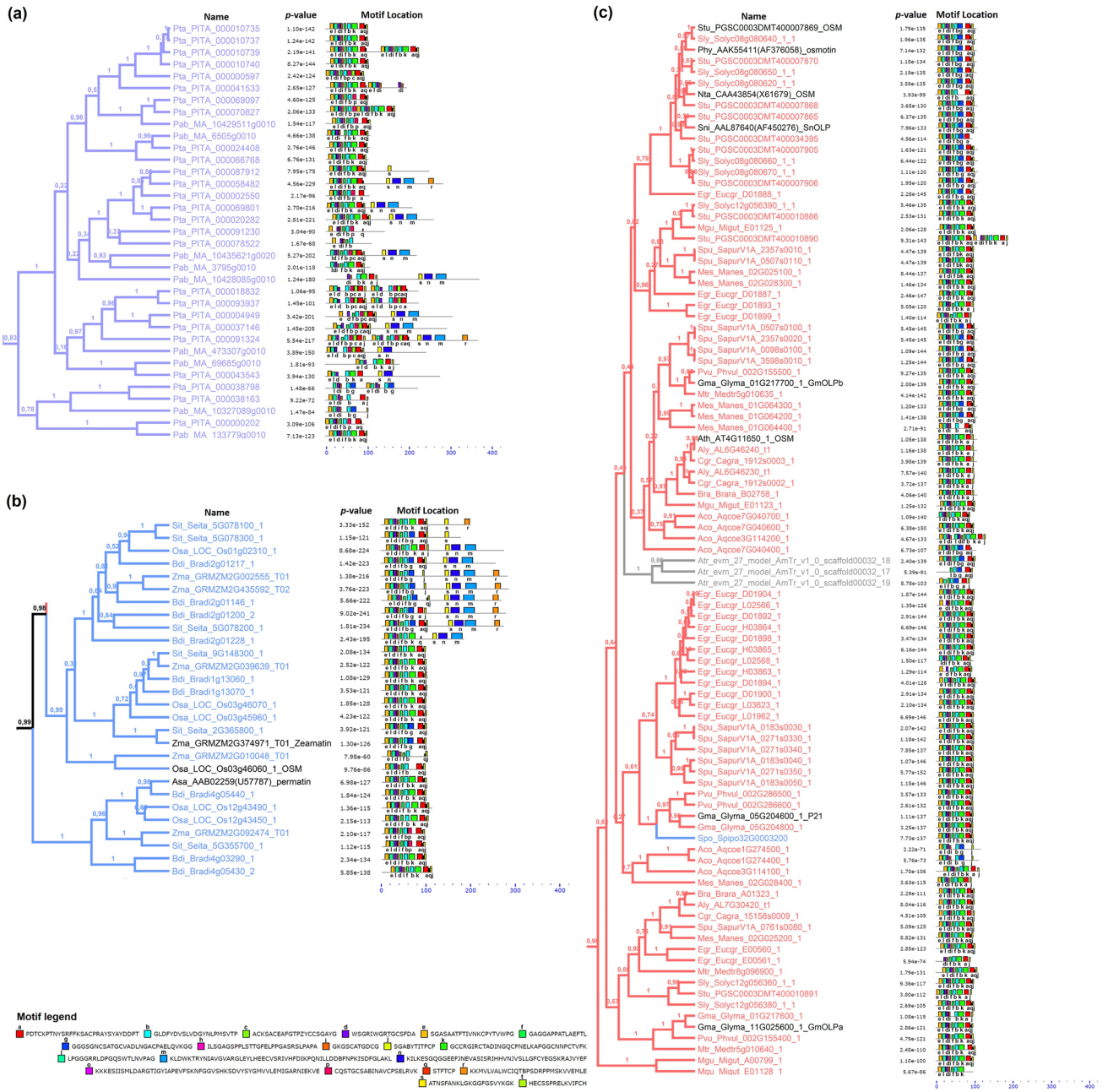


Fig. 6 Architecture of conserved protein motifs in putative osmotin sequences identified using the MEME tool. Each motif is represented by a colored block with a letter below. The lengths and positions of the blocks correspond to the lengths and positions of motifs in the individual protein sequences. 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*

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 (Figs. 6 and S3b), with the "k" motif being

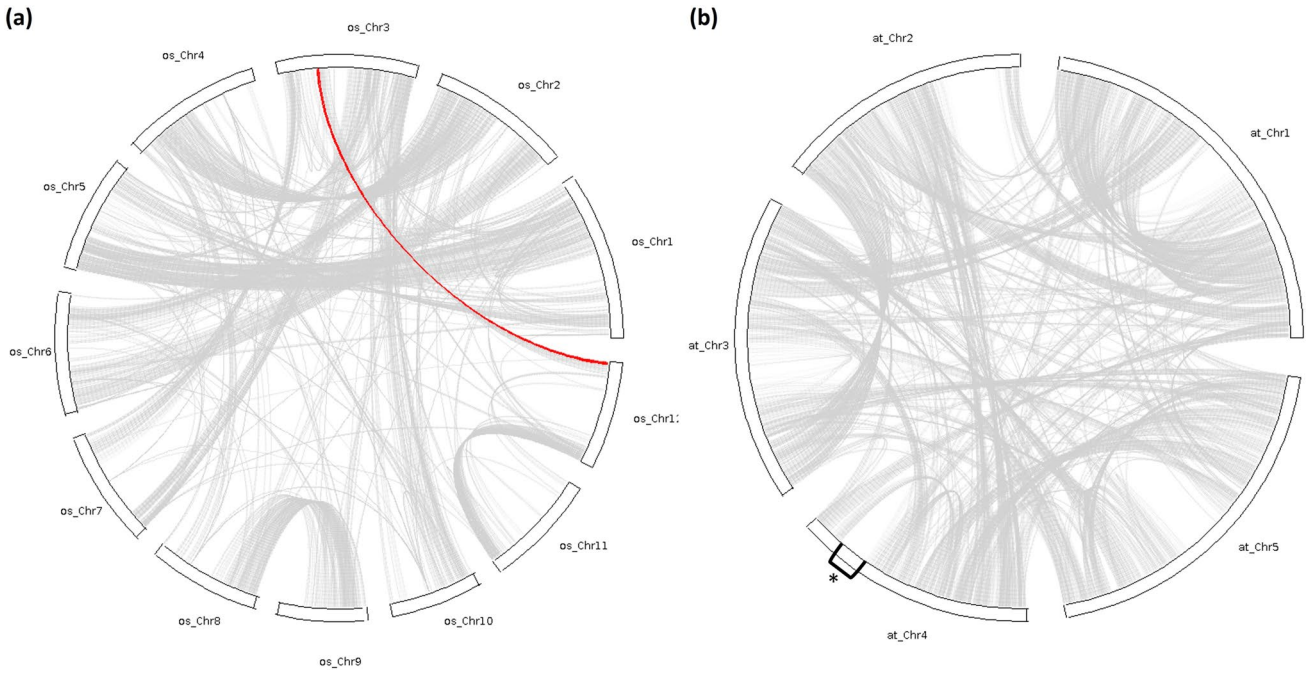


Fig. 7 Collinear gene pairs for osmotin genes on **a** 12 *O. sativa* chromosomes and **b** 5 *A. thaliana* chromosomes. Gray lines indicate the collinear gene pairs in the whole genome, while red lines indicate the collinear gene pairs for osmotin genes. An asterisk (*) indicates the non-duplicated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

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 patters of putative osmotin genes. In the whole genome of *O. sativa*, 6567 (15.56%) and 3706 (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 (Fig. 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*, 10367 (37.81%) and 8039 (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 into 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 (Figs. 8, 9). The putative rice osmotins LOC_Os01g02310, LOC_03g46060, and LOC_Os12g43450 were highly expressed during drought stress (Fig. 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* (Fig. 8b). The *A. thaliana* osmotin was predicted to be highly expressed at different time points during drought, salt, and osmotic stresses (Fig. 9a). This osmotin was also observed to be highly expressed during challenge with three types of fungi (Fig. 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 (Figs. 1, 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) (Figs. 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 with the domain I of these proteins (Mani et al. 2012; 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 phosphotransferase/kinase 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

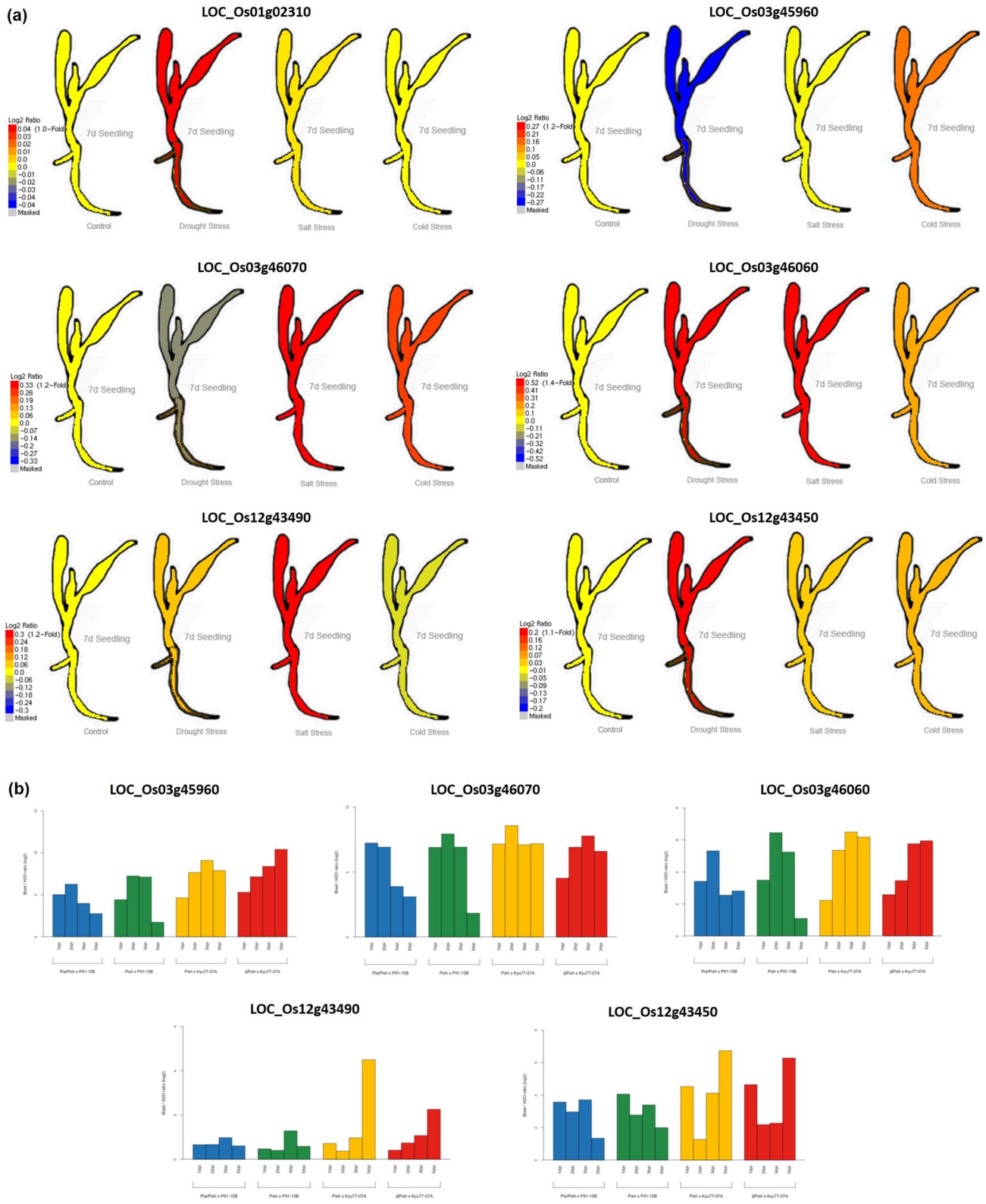


Fig. 8 Relative expression profile of *O. sativa* putative osmotin genes. **(a)** As plotted in the bar scale, red and blue colors indicate up- and downregulated genes (Log 2 ratio) under abiotic stresses. **b** Signal intensity data based on 75 percentile normalization and log2 transformation for the 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) (color figure online)

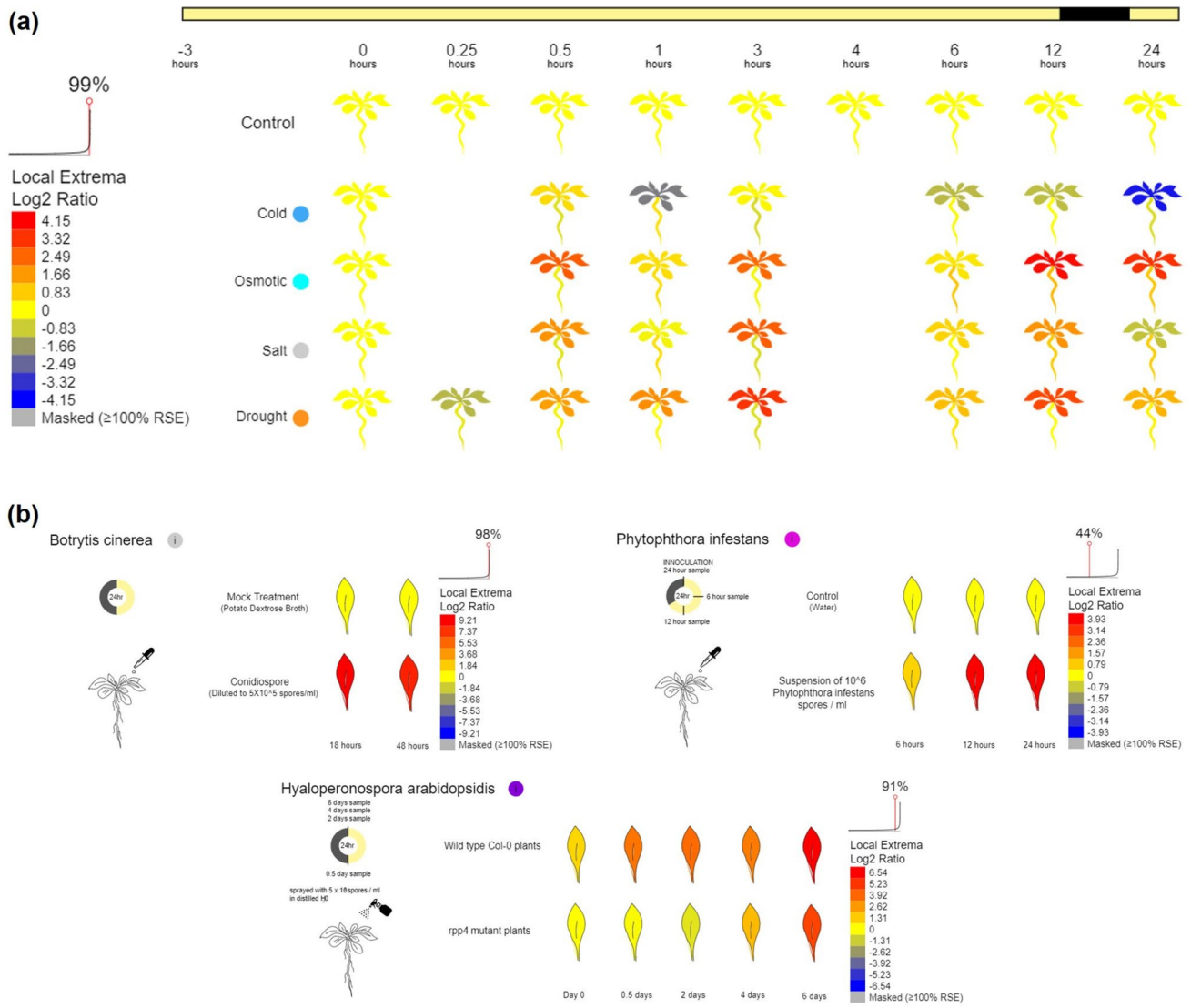


Fig. 9 Relative expression profile of a putative *A. thaliana* osmotin gene. As plotted in the bar scale, red and blue color indicate up- and down-regulated genes (Log 2 ratio) under **a** abiotic stresses and **b** biotic stresses (color figure online)

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 et 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 (Fig. 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 indicate an upregulation of putative osmotins during *M. oryzae* challenge (Fig. 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 (Figs. 8, 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 provide 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 conserved defense mechanisms against pathogen infection. 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 coevolution, 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 with 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 (Fig. 4). According to Koonin (2006), introns are important components of eukaryotic genes, and their gain or loss affects 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 duplication 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.

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: FLG. 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.

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.

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 is 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/.

Compliance with ethical standards

Conflict of interest GRF, ACT-Z, FLG, LAO-B and MHB-Z declares that they have 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 191:51–56

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 Genom 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 Biol Evol 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 42:105–112

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 Physiol Biochem 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. J Exp Bot 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. Biol 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. Front Plant Sci 6:163

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 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. Front Plant Sci 8:366

Letunic I, Bork P (2017) 20 years of the SMART protein domain annotation resource. Nucl Acids Res 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. Mol Biotechnol 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*. Funct Integr Genom 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*. J Proteom 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 Biol 11:33

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67(5):901–904. https://doi.org/10.1093/sysbio/syy032

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. Mol Plant Microbe Interact 17:780–788

Shatters RG, Boykin LM, Lapointe SL, Hunter WB, Weathersbee AA III (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 4:315–323

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 93:2598–2602

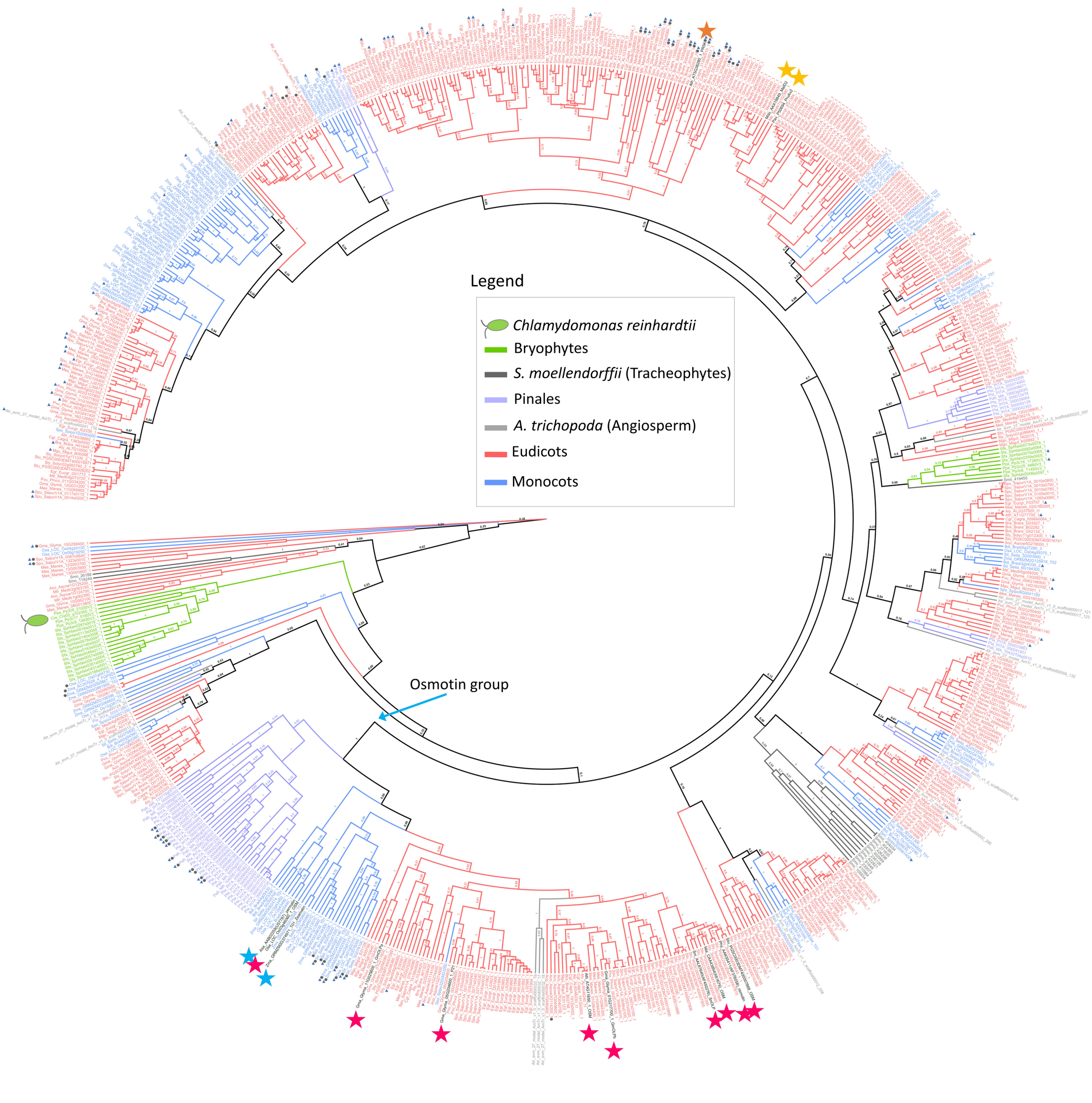
Wang Y, Wang X, Tang H, Tan X, Ficklin SP, Feltus FA, Paterson AH (2011) Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. PLoS One 6:e28150

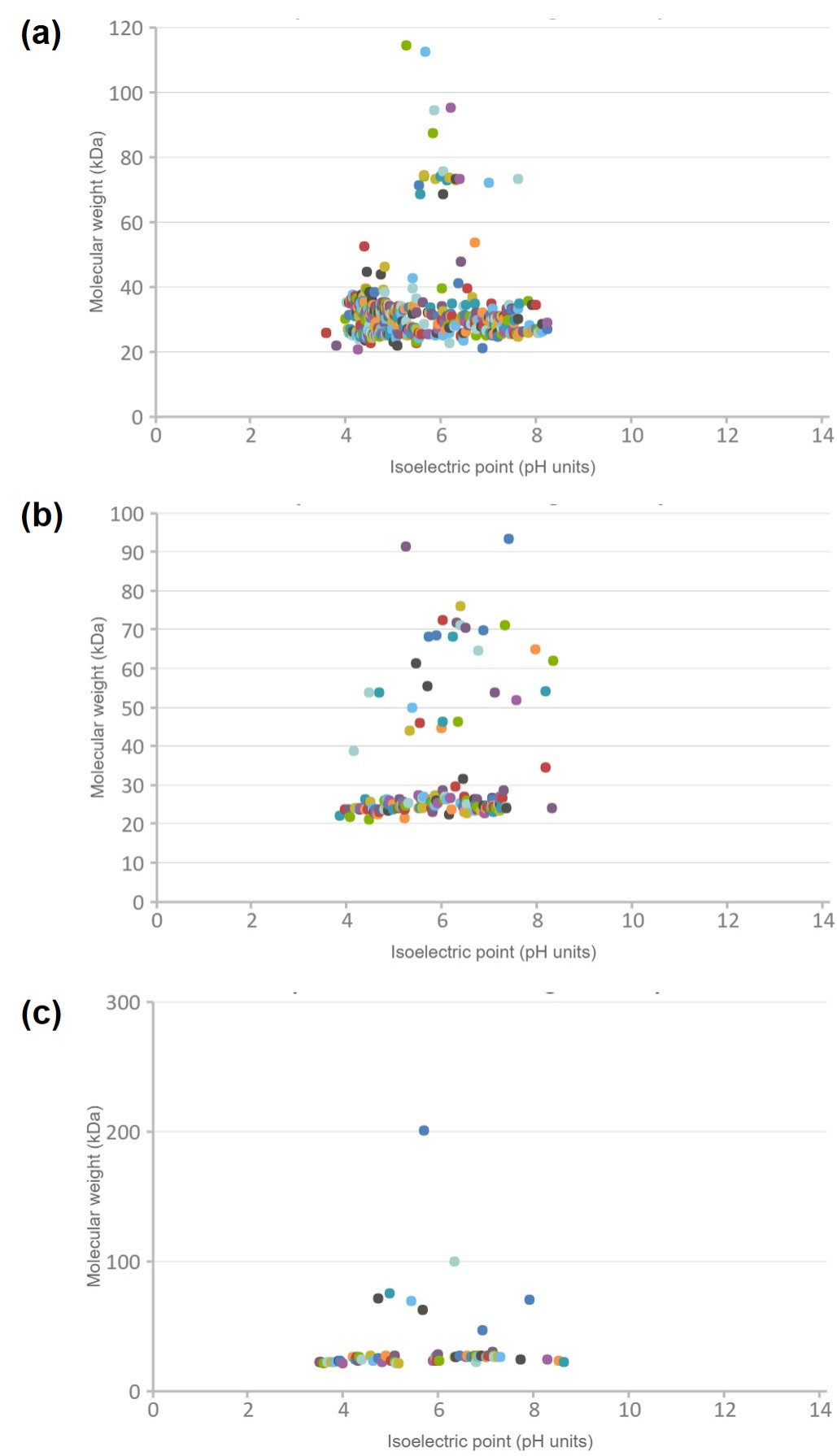
Wang Y, Tang H, DeBarry JD, Xu T, Li J, Wang X, Lee T, Jin H, Marler B, Kissinger HGJC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucl Acids Res 40:e49–e49


Wang H, Devos KM, Bennetzen JL (2014) Recurrent loss of specific introns during angiosperm evolution. PLoS Genet 10:e1004843

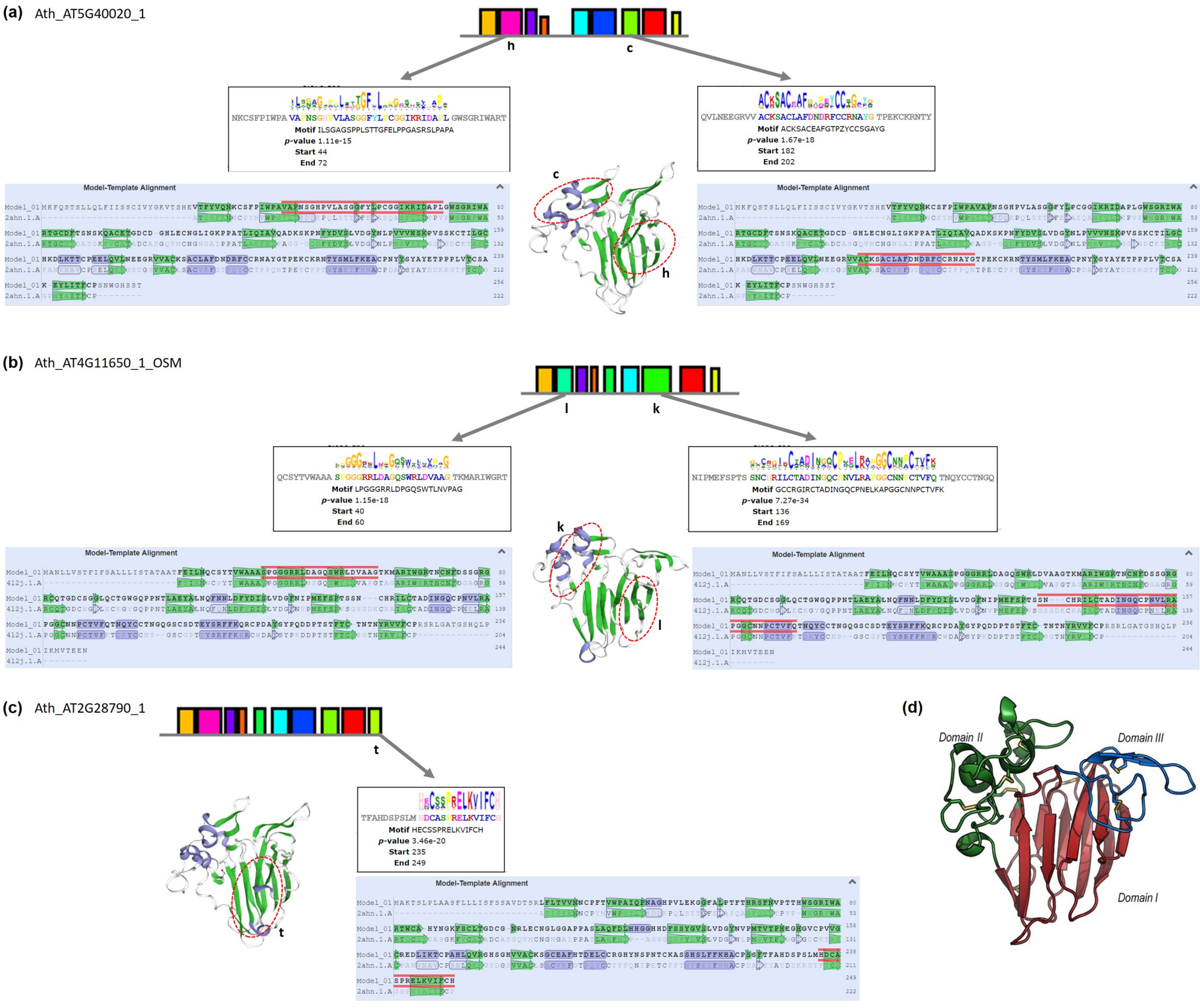
Zong W, Zhong X, You J, Xiong L (2013) Genome-wide profiling of histone H3K4-tri-methylation and gene expression in rice under drought stress. Plant Mol Biol 81:175–188

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.


Legend
Chlamydomonas reinhardtii
Bryophytes
S. moellendorffii (Tracheophytes)
Pinales
A. trichopoda (Angiosperm)
Eudicots
Monocots
Osmotin group


(a)
Molecular weight (kDa)
120
100
80
60
40
20
0
0
2
4
6
8
10
12
14
Isoelectric point (pH units)
(b)
Molecular weight (kDa)
100
90
80
70
60
50
40
30
20
10
0
0
2
4
6
8
10
12
14
Isoelectric point (pH units)
(c)
Molecular weight (kDa)
300
200
100
0
0
2
4
6
8
10
12
14
Isoelectric point (pH units)


(a)
(b)
(c)
Name
p-value
Motif Location
Motif legend
PDTCKPTNYSRFFKSACPRAYSYAYDDPT
GLDFYDVSLVDGYNLPMSVTP
ACKSACEAFGTPZYCCSGAYG
WSGRIWGRTGCSFDA
SGASAATFTIVNKCPYTVWPG
GAGGAPPATLAEFTL
GGGSGNCSATGCVADLNGACPAELQVKGG
ILSGAGSPPLSTTGFELPPGASRSLPAPA
GKGSCATGDCG
SGABYTITFCP
GCCRGIRCTADINGQCPNELKAPGGCNNPCTVFK
LPGGGRRLDPGQSWTLNVPAG
KLDWKTRYNIAVGVARGLEYLHEECVSRIVHFDIKPQNILLDDBFNPKISDFGLAKL
KILKESGQGGEEFJNEVASISRIHHVNJVSLLGFCYEGSKRAJVYEF
KKKESIISMLDARGTIGYIAPEVFSKNFGGVSHKSDVYSYGMVVLEMIGARNIEKVE
CQSTGCSABINAVCPSELRVK
STFTCP
KKMVLVALWCIQTBPSDRPPMSKVVEMLE
ATNSFANKLGKGGFGSVYKGK
HECSSPRELKVIFCH


(a) Ath_AT5G40020_1
h
c
NKCSFPIWPA VAPNSGHPVLASGGFYLPCGGIKRIDAPLG WSGRIWART
Motif ILSGAGSPPLSTTGFELPPGASRSLPAPA
p-value 1.11e-15
Start 44
End 72
QVLNEEGRVV ACKSACLAFDNDRFCCRNAYG TPEKCKRNTY
Motif ACKSACEAFGTPZYCCSGAYG
p-value 1.67e-18
Start 182
End 202
Model-Template Alignment
(b) Ath_AT4G11650_1_OSM
l
k
QCSYTVWAAA SPGGGRRLDAGQSWRLDVAAG TKMARIWGRT
Motif LPGGGRRLDPGQSWTLNVPAG
p-value 1.15e-18
Start 40
End 60
NIPMEFSPTS SNCHRILCTADINGQCPNVLRAPGGCNNPCTVFQ TNQYCCTNGQ
Motif GCCRGIRCTADINGQCPNELKAPGGCNNPCTVFK
p-value 7.27e-34
Start 136
End 169
Model-Template Alignment
(c) Ath_AT2G28790_1
t
TFAHDSPSLM HDCASPRELKVIFCH
Motif HECSSPRELKVIFCH
p-value 3.46e-20
Start 235
End 249
Model-Template Alignment
(d)
Domain II
Domain III
Domain I