Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar

Benjamin Petre, Ian Major, Nicolas Rouhier, Sébastien Duplessis

HAL Id: hal-02548481

https://hal.science/hal-02548481

Submitted on 29 May 2020

BMC
Plant Biology

RESEARCH ARTICLE   Open Access

Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar

Benjamin Petre[1], Ian Major[2], Nicolas Rouhier[1], Sébastien Duplessis[1*]

Abstract

Background: Plant inducible immunity includes the accumulation of a set of defense proteins during infection called pathogenesis-related (PR) proteins, which are grouped into families termed PR-1 to PR-17. The PR-5 family is composed of thaumatin-like proteins (TLPs), which are responsive to biotic and abiotic stress and are widely studied in plants. TLPs were also recently discovered in fungi and animals. In the poplar genome, TLPs are over-represented compared with annual species and their transcripts strongly accumulate during stress conditions.

Results: Our analysis of the poplar TLP family suggests that the expansion of this gene family was followed by diversification, as differences in expression patterns and predicted properties correlate with phylogeny. In particular, we identified a clade of poplar TLPs that cluster to a single 350 kb locus of chromosome I and that are up-regulated by poplar leaf rust infection. A wider phylogenetic analysis of eukaryote TLPs - including plant, animal and fungi sequences - shows that *TLP* gene content and diversity increased markedly during land plant evolution. Mapping the reported functions of characterized TLPs to the eukaryote phylogenetic tree showed that antifungal or glycan-lytic properties are widespread across eukaryote phylogeny, suggesting that these properties are shared by most TLPs and are likely associated with the presence of a conserved acidic cleft in their 3D structure. Also, we established an exhaustive catalog of TLPs with atypical architectures such as small-TLPs, TLP-kinases and small-TLP-kinases, which have potentially developed alternative functions (such as putative receptor kinases for pathogen sensing and signaling).

Conclusion: Our study, based on the most recent plant genome sequences, provides evidence for *TLP* gene family diversification during land plant evolution. We have shown that the diverse functions described for TLPs are not restricted to specific clades but seem to be universal among eukaryotes, with some exceptions likely attributable to atypical protein structures. In the perennial plant model *Populus*, we unravelled the TLPs likely involved in leaf rust resistance, which will provide the foundation for further functional investigations.

Background

Plants respond to challenge from pathogens by activating an inducible protein-based defense system that includes 17 families of pathogenesis-related (PR) proteins termed PR-1 to PR-17 [1,2]. Proteins of the PR-5 family have high sequence identity with thaumatins, which are sweet-tasting proteins isolated from the West African shrub *Thaumatococcus daniellii* and are thus referred to as thaumatin-like proteins (TLPs) [3]. For decades, TLPs have been studied extensively in plants

for their antifungal properties. The recent identification of TLPs in animals [4] and fungi [5] indicates that these proteins are more widely distributed and not only restricted to plants [6].

Molecular studies of TLP expression, localisation and activity support a role for TLPs in host defense during pathogen infection. TLP up-regulation has been described in many higher plants infected by pathogens such as bacteria, oomycetes and fungi [7,8]. Localisation studies revealed that plant pathogen-inducible TLPs are secreted into the apoplast [9,10]. More than 20 TLPs from animals, fungi and plants have been shown to exhibit an antifungal activity [7], although the mechanisms by which TLPs exert this activity remain unclear. Several antifungal modes of action have been described

* Correspondence: duplessi@nancy.inra.fr
[1]INRA†/Nancy Université, Unité Mixte de Recherche 1136 'Interactions Arbres/Micro-organismes', Centre INRA de Nancy, F-54280 Champenoux, France
Full list of author information is available at the end of the article

such as membrane permeabilization [11], β-glucan binding and degradation [5], inhibition of enzymes such as xylanases [12], α-amylase, or trypsin [13], as well as an apoptosis-inducing mechanism reported in yeast [14]. Other functional properties have been reported for TLPs, including antifreeze activity [15], protection from abiotic stress [16] and binding to proteins such as actin, viral CMV-1 protein, yeast glycoproteins and G-Protein Coupled Receptor (GPCR) or to hormones such as cytokinins [7].

Most typical TLPs described to date have a molecular weight ranging from 20 to 26 kDa, and generally possess 16 conserved cysteine residues that form eight disulfide bonds [17]. Recently, small TLPs (sTLPs) have been identified in monocots and conifers. These are characterized by a smaller molecular weight (around 17 kDa) and only 10 conserved cysteine residues that form five disulfide bonds [18-20]. Seven TLP structures have been solved so far, revealing a strongly conserved 3D organisation with a characteristic acidic cleft domain that comprises the five highly conserved amino acids REDDD that are dispersed in the primary sequence [21]. Despite good conservation of these amino acids in sTLP primary sequences, they do not organize into an acidic cleft at the 3D level [22]. Unusual TLP and protein kinase fusion proteins referred to as PR5-kinase or TLP-kinase (TLP-K) have also been reported in a few plant species [23,7].

The analysis of the *Populus trichocarpa* 'Nisqually-1' genome revealed a substantial over-representation of genes encoding disease resistance proteins compared with annual species such as *Arabidopsis thaliana*, and this increase is not solely attributable to the genome expansion in *Populus* [24]. In particular, 55 putative *TLP* genes were initially identified in *P. trichocarpa* versus 24 for *A. thaliana* [24]. *Populus* spp. are economically important and hybrid poplars in particular are used extensively worldwide for wood production. Breeding programs particularly target resistance to *Melampsora* spp. fungi, which are responsible for leaf rust, a major disease of poplars that severely impacts tree growth and wood production [25]. With the availability of both *P. trichocarpa* and *M. larici-populina* genome sequences, the biotrophic poplar-rust interaction is emerging as a model pathosystem in forest biology [26]. Several transcriptome-based studies revealed transcriptional reprogramming in poplar leaves infected by *Melampsora* spp., including the up-regulation of many PR proteins [26]. In particular, transcript profiling of poplar leaves during an incompatible interaction (i.e. host-specific resistance) with *M. larici-populina* established a set of host-defense marker genes, including several TLPs [27].

The present study describes the annotation of 42 *TLP* gene models in the *P. trichocarpa* 'Nisqually-1' genome

version 2.0. In addition, comparison of expression studies conducted on poplar subjected to biotic (i.e. *Melampsora* spp. infection) and abiotic stresses identified stress-responsive clades. The comparison of 598 complete eukaryote TLP amino acid sequences, of which 410 come from the 18 plant genome sequences currently available, allowed us to establish a link between function and phylogeny by systematically mapping functional data mined from the literature to the phylogenetic tree. *In silico* structural analysis confirmed that, with the exception of sTLPs, the acidic cleft domain is strongly conserved among eukaryote TLPs.
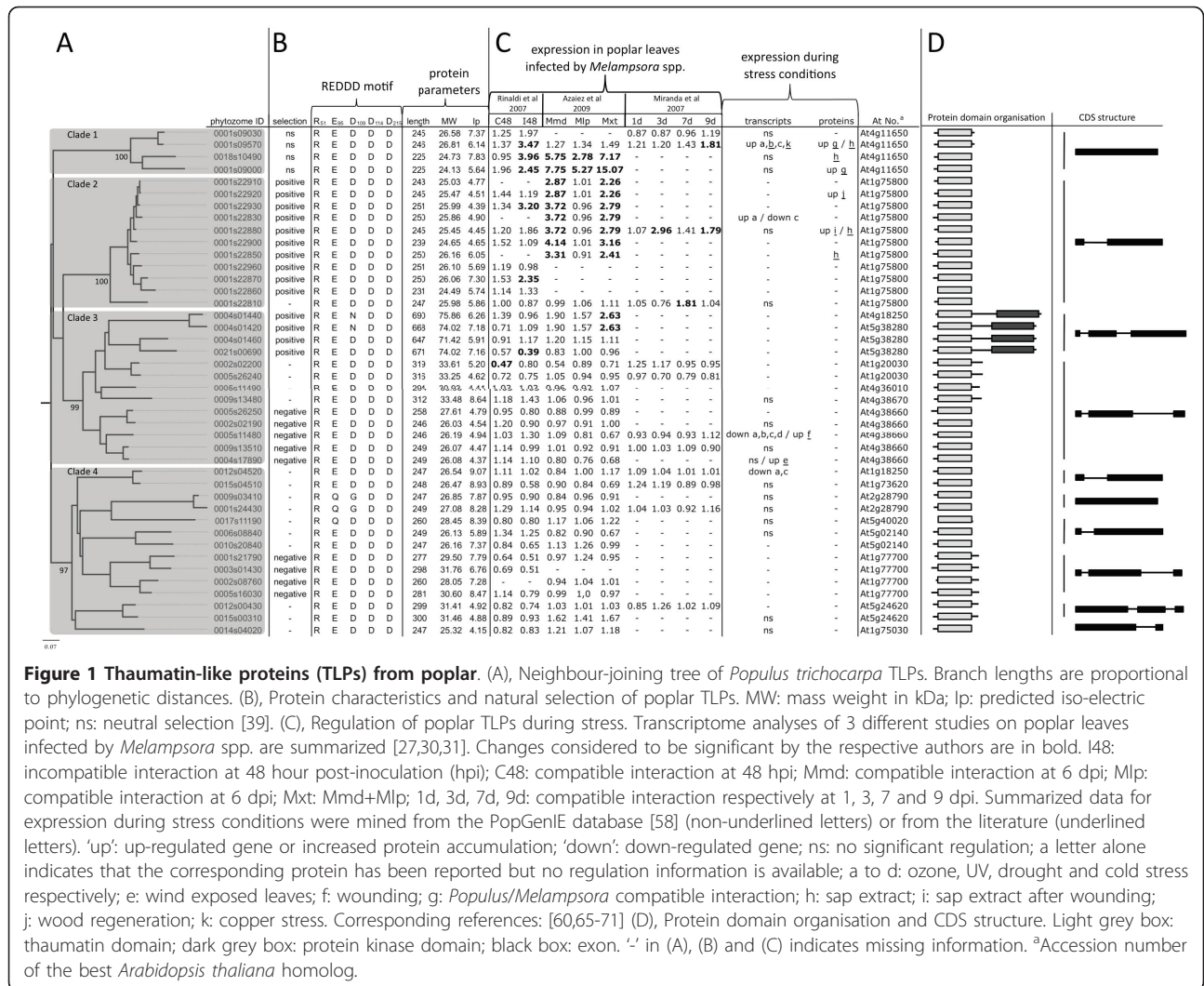
Results
Annotation, phylogeny, genomic distribution and gene expression of poplar TLPs

In contrast to Tuskan and collaborators [24], we identified a total of 59 putative *TLP* genes in the *P. trichocarpa* 'Nisqually-1' genome version 1.1. In version 2.0 of the genome, now integrated in the Phytozome portal [28,29], 17 of these *TLP* gene models are not validated. These 17 invalidated models include 11 predicted alleles that were previously considered to be independent genes and six probable pseudogenes that are interrupted by stop codons (Additional file 1). The remaining 42 *TLP* genes that are validated in version 2.0 of the genome comprise 38 typical TLPs and four genes with strong homology to TLP-K from *A. thaliana*, including fusion to a putative protein kinase (Pfam: PF00069) ([23], Additional file 2).

A phylogenetic tree constructed with the validated poplar TLPs reveals four well-defined clades, numbered here from 1 to 4. Among these clades, the REDDD residues are highly conserved with only small variations for five TLPs (Figure 1). The size of the proteins varies from 225 to 319 amino acids (~24 to 34 kDa) for the 38 typical TLPs and is approximately 650 amino acids (~73 kDa) for the four TLP-Ks. The predicted isoelectric points vary from 4.15 to 9.07 and correspond well with the TLP phylogeny (Figure 1). Analysis of the protein domain organisation showed that the thaumatin domain (Pfam: PF00314) covers almost 95% of the entire mature TLPs, except 10 TLPs in clades 3 and 4 that have approximately 40 additional amino acids in their C-terminal region. The four TLP-Ks are grouped in a specific branch of clade 3, suggesting that they are monophyletic in poplar. The gene structure of poplar *TLPs* is well conserved within clades 1-3, with genes belonging to clade 1 formed by a single exon, *TLPs* from clade 2 by two exons and *TLPs* of clade 3 by three exons (Figure 1); clade 4 is an exception with genes composed of one, two or three exons.

The version 2.0 of the *P. trichocarpa* genome incorporates a greatly improved physical map compared with

Figure 1 Thaumatin-like proteins (TLPs) from poplar. (A), Neighbour-joining tree of *Populus trichocarpa* TLPs. Branch lengths are proportional to phylogenetic distances. (B), Protein characteristics and natural selection of poplar TLPs. MW: mass weight in kDa; Ip: predicted iso-electric point; ns: neutral selection [39]. (C), Regulation of poplar TLPs during stress. Transcriptome analyses of 3 different studies on poplar leaves infected by *Melampsora* spp. are summarized [27,30,31]. Changes considered to be significant by the respective authors are in bold. I48: incompatible interaction at 48 hour post-inoculation (hpi); C48: compatible interaction at 48 hpi; Mmd: compatible interaction at 6 dpi; Mlp: compatible interaction at 6 dpi; Mxt: Mmd+Mlp; 1d, 3d, 7d, 9d: compatible interaction respectively at 1, 3, 7 and 9 dpi. Summarized data for expression during stress conditions were mined from the PopGenIE database [58] (non-underlined letters) or from the literature (underlined letters). 'up': up-regulated gene or increased protein accumulation; 'down': down-regulated gene; ns: no significant regulation; a letter alone indicates that the corresponding protein has been reported but no regulation information is available; a to d: ozone, UV, drought and cold stress respectively; e: wind exposed leaves; f: wounding; g: *Populus/Melampsora* compatible interaction; h: sap extract; i: sap extract after wounding; j: wood regeneration; k: copper stress. Corresponding references: [60,65-71] (D), Protein domain organisation and CDS structure. Light grey box: thaumatin domain; dark grey box: protein kinase domain; black box: exon. '-' in (A), (B) and (C) indicates missing information. [a]Accession number of the best *Arabidopsis thaliana* homolog.
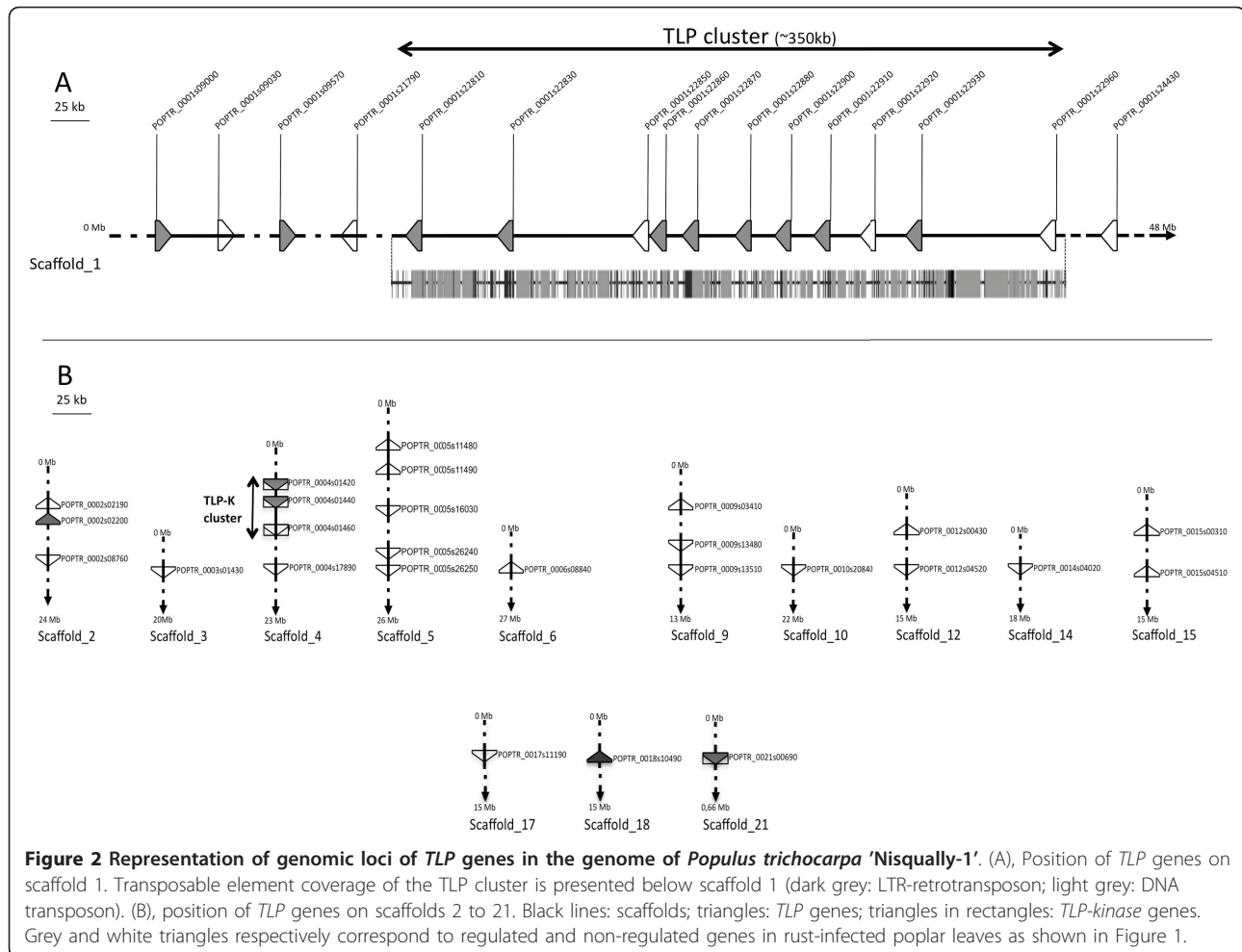
version 1.1. This helped localise 41 of the 42 annotated *TLP* genes on 13 of the 19 chromosomes (i.e. scaffolds 1 to 19 on the Phytozome portal [29]) (Figure 2). Scaffold 1 contains 16 *TLP* genes, including all 11 *TLP* genes from clade 2 which are located within a 350 kb segment that encodes *TLPs* exclusively. We named this region the TLP cluster. Transposable elements (TE) cover 52% of this 350 kb region, with a particular over-representation of long terminal repeat (LTR) Gypsy elements that cover 37% of the cluster (Figure 2 and Additional file 3).

Results compiled from three different previously published transcriptome analyses of poplar leaves infected by *Melampsora* spp. fungi [27,30,31] indicate that, of the 42 *TLP* genes, 14 are significantly up-regulated and two are significantly down-regulated (Figure 1). Among the 14 up-regulated TLP transcripts, 12 belong to clades 1 and 2 and 11 of these are located on scaffold 1 (Figure 1 and 2). Interestingly, five *TLP* genes are up-regulated during an incompatible poplar/rust interaction, of which

three are grouped in clade 1. Under abiotic stress conditions, five poplar TLP transcripts showed differential accumulation. In addition, six TLPs were identified by different proteomic studies, of which four were shown to accumulate during biotic or abiotic stress (Figure 1). More specifically, the *PopTLP1* gene (*P. trichocarpa* geneID Poptr_0001s09570) from clade 1 is associated with several biotic and abiotic stresses and we confirmed with a detailed time-course analysis by RT-qPCR that *PopTLP1* expression increases in poplar leaves challenged by *M. larici-populina* (Additional file 4).

TLPs in green plant genome sequences
We performed an exhaustive genomic analysis of plant *TLPs* by collecting *TLP* gene models from 18 sequenced plants available at the Phytozome portal [29]. Models encoding proteins with an incomplete thaumatin domain were ignored (Table 1). A single but incomplete *TLP* gene was identified in the unicellular green algae

Figure 2 Representation of genomic loci of *TLP* genes in the genome of *Populus trichocarpa* 'Nisqually-1'. (A), Position of *TLP* genes on scaffold 1. Transposable element coverage of the TLP cluster is presented below scaffold 1 (dark grey: LTR-retrotransposon; light grey: DNA transposon). (B), position of *TLP* genes on scaffolds 2 to 21. Black lines: scaffolds; triangles: *TLP* genes; triangles in rectangles: *TLP-kinase* genes. Grey and white triangles respectively correspond to regulated and non-regulated genes in rust-infected poplar leaves as shown in Figure 1.

Chlamydomonas reinhardtii, which represents the evolutionary starting point of viridiplantae, and thus makes the origin of complete TLPs in the green lineage unclear (Table 1). Three complete *TLP* genes were identified in the moss *Physcomitrella patens* and 12 were found in the vascular plant *Selaginella moellendorffii*, indicating that an important gene expansion occurred in the transition from bryophytes to tracheophytes. Among the 15 angiosperm genomes, the *TLP* gene number varies from 16 in the barrel clover *Medicago truncatula* to 42 in the black cottonwood *P. trichocarpa*, whereas *A. thaliana* has 22 *TLP* genes. An average of 26 *TLP* genes are present in angiosperms, with similar numbers of *TLPs* in dicots or monocots (Table 1). sTLP-encoding genes were identified exclusively in monocots (from 2 in *Zea mays* to 9 in *Sorghum bicolor*), whereas TLP-Ks have been identified in both monocots and dicots, although dicot TLP-Ks were restricted to the *A. thaliana* and *P. trichocarpa* genomes. To identify the genes that are most similar to TLP-Ks in the remaining dicots, we performed homology searches with the kinase domain of

TLP-Ks and retrieved only lectin-kinase genes, confirming the absence of *TLP-Ks* in these dicot genomes (data not shown). In *S. bicolor*, a small-TLP-kinase (here termed sTLP-K) composed of a N-terminal sTLP domain and a C-terminal protein kinase domain, separated by a predicted transmembrane (TM) domain, was identified (Additional file 5). The origin of this arrangement is puzzling and has apparently evolved independently of TLP-Ks. To our knowledge, this is the first report of such a domain organisation.

Eukaryote TLPs: linking phylogeny with protein structure and function

To achieve an accurate and complete phylogeny of eukaryote TLPs, we retrieved an additional 188 sequences with a complete thaumatin domain from the NCBI protein database [32] and combined them with the 410 plant sequences that we identified earlier (Additional file 6). These include several sequences from fungi (basidiomycetes and ascomycetes) and invertebrate animals (nematods and arthropods), as well as other

Table 1 *TLP* gene content in sequenced plant species

organism	code	common organism name	phylum	class	order	TLP blast result[a]	complete TLP domain[c]	small-TLP/TLP-K[d]
Chlamydomonas reinhardtii	Chlre	Green algae	Chlorophyte	Chlorophyceae	Volvocales	1	0	0/0
Physcomitrella patens	Phypa	Moss	Bryophyte	Bryopsides	Funariales	5	3	0/0
Selaginella moelledorffii	Selmo	Lycophyte	Tracheophyte	Sellaginellopsides	Selaginellales	18	12	0/0
Oryza sativa	Orysa	Rice	Angiosperm	Monocotyledon	Cyperales	37	26	4/1
Brachypodium distachyon	Bradi	Purple false brome	Angiosperm	Monocotyledon	Poales	32	24	3/2
Sorghum bicolor	Sorbi	Sorghum	Angiosperm	Monocotyledon	Poales	45	36	9/1(1[e])
Zea mays	Zeama	Maize	Angiosperm	Monocotyledon	Poales	38	29	2/2
Mimulus guttatus	Mimgu	Common monkey-flower	Angiosperm	Dicotyledon	Lamiales	33	23	0/0
Vitis vinifera	Vitvi	Grapevine	Angiosperm	Dicotyledon	Rosales	27	18	0/0
Carica papaya	Carpa	Papaya tree	Angiosperm	Dicotyledon	Brassicales	18	16	0/0
Arabidopsis thaliana	Arath	Thale cress	Angiosperm	Dicotyledon	Brassicales	30	22	0/3
Cucumis sativus	Cucsa	Cucumber	Angiosperm	Dicotyledon	Cucurbitales	29	28	0/0
Glycine max	Glyma	Soya	Angiosperm	Dicotyledon	Fabales	58	38	0/0
Medicago truncatula	Medtr	Barrel clover	Angiosperm	Dicotyledon	Fabales	21	16	0/0
Prunus persica	Prupe	Peach tree	Angiosperm	Dicotyledon	Rosales	37	28	0/0
Manihot esculenta	Manes	Manioc	Angiosperm	Dicotyledon	Malpighiales	34	27	0/0
Ricin communis	Ricco	Castor oil plant	Angiosperm	Dicotyledon	Malpighiales	24	22	0/0
Populus trichocarpa	Poptr	Poplar	Angiosperm	Dicotyledon	Malpighiales	59[b]	42	0/4

[a]Number of putative *TLP* genes identified by amino acid homology searches of plant genome sequences on the Phytozome portal [29].

[b]Number of putative *TLP* genes identified from version 1.1 of the *Populus trichocarpa* 'Nisqually-1' genome on the JGI website [55].

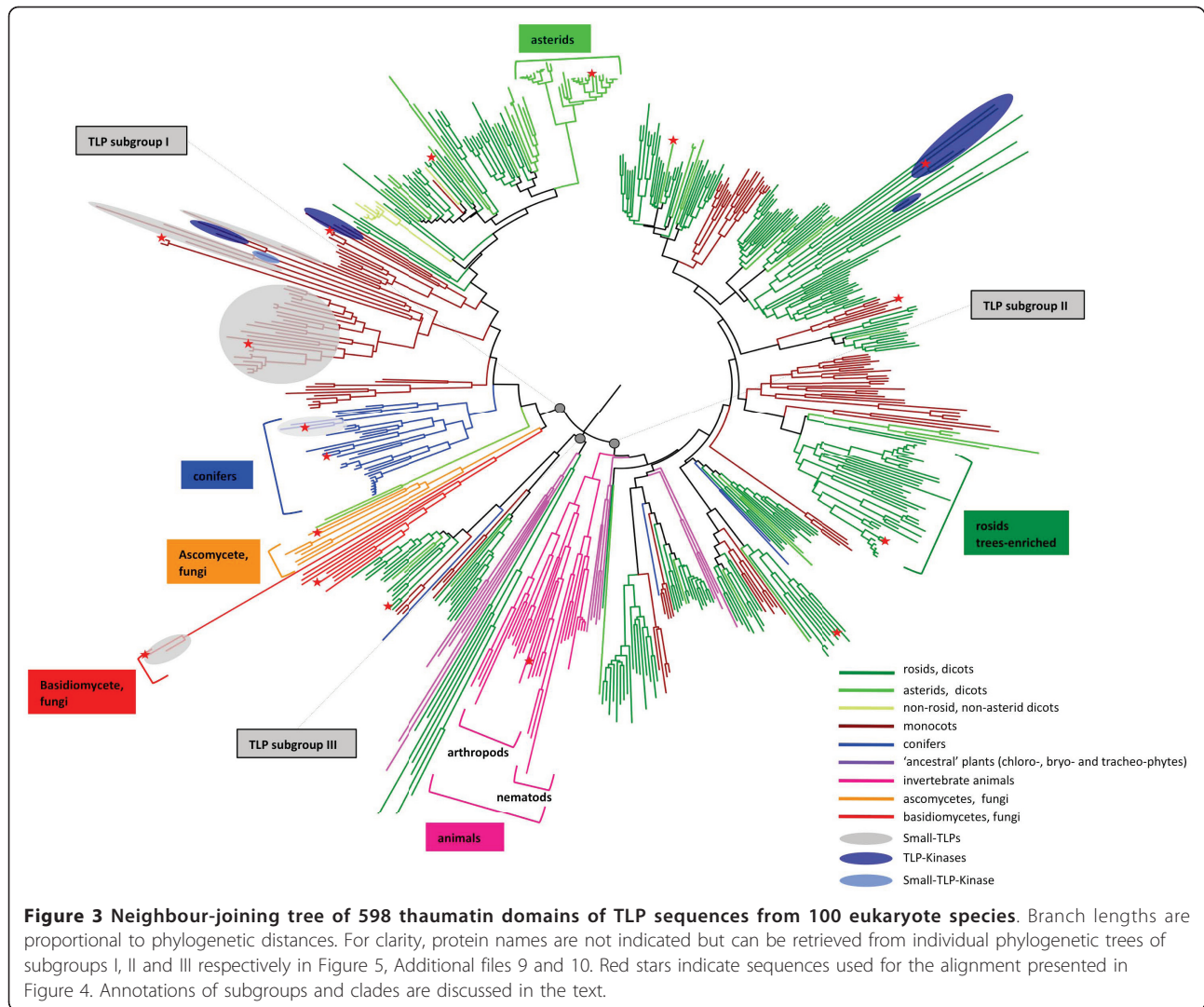[c]TLP sequences with a complete thaumatin domain.

[d]Proportion of sTLP and TLP-K with a complete thaumatin domain.

[e]small-TLP/kinase domain fusion (sTLP-K).

plants from mainly the asterid and conifer divisions. We report for the first time the identification of *sTLP* genes in basidiomycetes, precisely in the pucciniales *M. larici-populina* and *Puccinia graminis* f.sp. *tritici*. Fungal sTLPs appear to be monophyletic, suggesting that sTLPs evolved independently in pucciniales, monocots and conifers or that sTLPs were lost during evolution from other phyla such as dicots and animals (Additional file 7). Overall, a total of 598 sequences were retrieved from 100 different species (12 animals, 12 fungi and 76 green plants) and were used for comparative genomic analyses. The phylogeny of these eukaryote TLPs reveals three major monophyletic groups (Figure 3). TLP subgroup I consists of 211 sequences and includes highly specific clades, such as a fungal clade containing TLPs from both ascomycetes and basidiomycetes, as well as plant clades that are specific to conifers, monocots, monocot sTLPs, monocot TLP-Ks, dicots or asterids. TLP subgroup II is composed of 341 sequences and includes an animal-specific clade with distinct sub-clades for nematodes and arthropods. Because of their over-representation, a large clade of plant sequences constitutes the vast majority of TLP subgroup II, with several subclades

composed of relatively balanced numbers of monocot and dicot sequences (Figure 3). TLP subgroup II notably includes a clade enriched in rosid and tree TLPs that in particular contains the poplar TLP cluster. Dicot TLP-Ks also belong to TLP subgroup II. TLP subgroup III contains only 46 sequences from 20 different plant species, with a large number of sequences from the vascular plant S. *moellendorffii* (Figure 3).
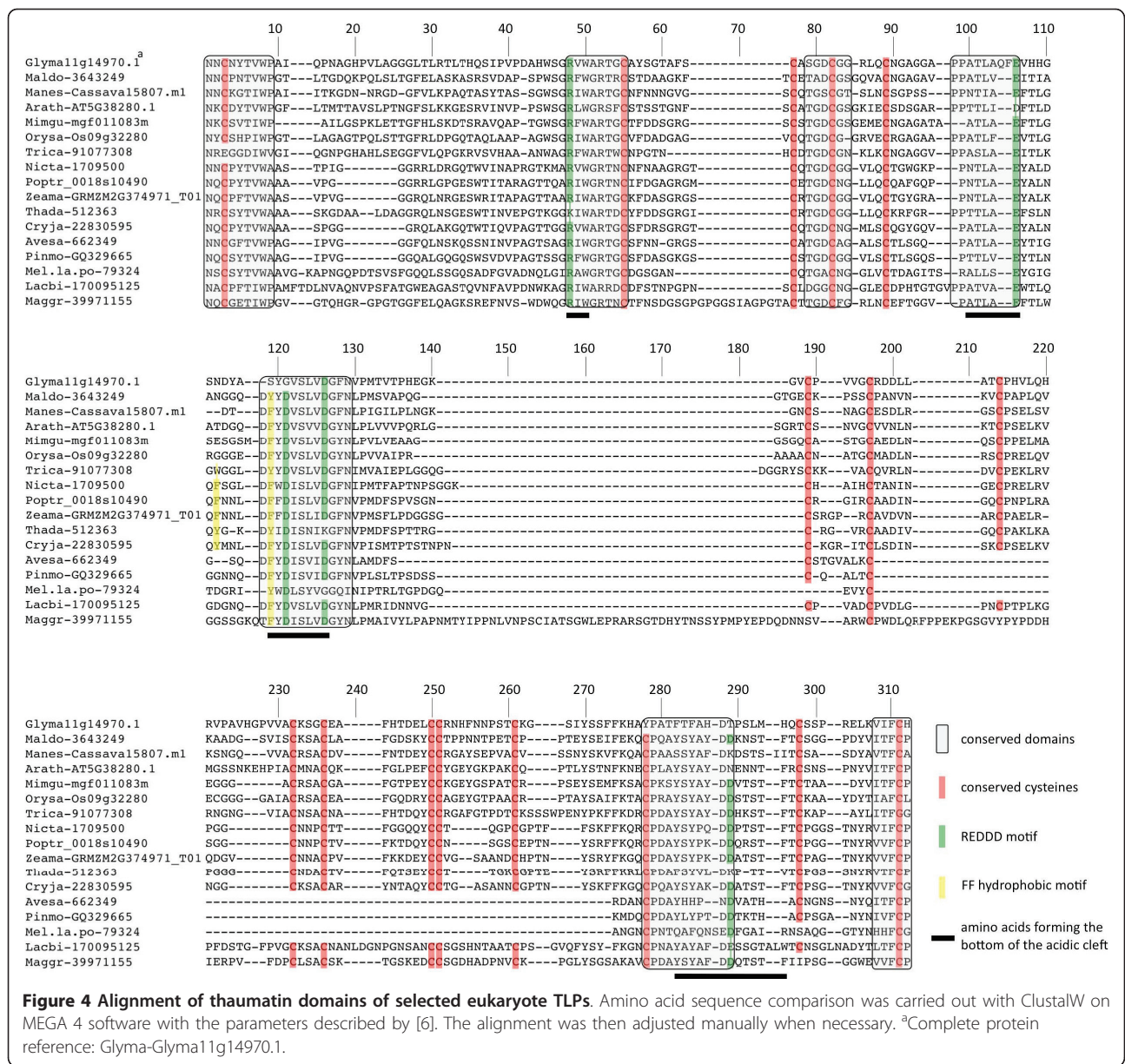
An alignment with 18 representative TLP sequences from the major sub-clades shows the diversity of eukaryote TLPs (Figure 4). The thaumatin domain of ascomycetes is almost 30% longer than that of typical TLPs (~280 versus ~215 amino acids), mainly due to three insertions in less-conserved regions of the domain. By contrast, sTLPs are almost 30% smaller than typical TLPs (~150 versus ~215 amino acids) due to a large deletion. The 16 cysteine residues (10 for sTLPs) are extremely well conserved, except for 1-2 residues in ascomycete and basidiomycete sTLPs and in some animal sequences (Figure 4). The REDDD motif or its equivalent (i.e. amino acids with similar biochemical properties) is fully conserved in 13 of the 18 representative sequences. Similarly, the amino acids forming the

Figure 3 Neighbour-joining tree of 598 thaumatin domains of TLP sequences from 100 eukaryote species. Branch lengths are proportional to phylogenetic distances. For clarity, protein names are not indicated but can be retrieved from individual phylogenetic trees of subgroups I, II and III respectively in Figure 5, Additional files 9 and 10. Red stars indicate sequences used for the alignment presented in Figure 4. Annotations of subgroups and clades are discussed in the text.

bottom of the acidic cleft and those at each extremity of the thaumatin domain are generally well conserved.

Information about the biological and/or biochemical properties were compiled for 26 TLPs with a complete amino acid sequence from an exhaustive survey of the literature (Additional file 8). These data were added systematically to the phylogenetic sub-trees of TLP subgroups I (Figure 5) and II (Additional file 9). Among these 26 TLPs, 21 have antifungal activity and nine have endo-β-1,3-glucanase activity. Surprisingly, antifungal TLPs are widespread among eukaryotes, as 13 are present in TLP subgroup I and 8 are in TLP subgroup II. A similar widespread assortment across TLP subgroups I and II was obtained for TLPs that exhibit endo-β-1,3-glucanase or antifreeze activities. Compared with the large amount of information available concerning asterid TLPs (many functions have been described for two TLPs of subgroup I: tobacco osmotin, Nicta-1709500, and maize

zeamatin, Zeama-grmzm2g394771), there is almost no functional characterization of conifer and fungal TLPs or sTLPs. One exception is TLX1, a sTLP from wheat (Triae-110836639), which is the only sTLP characterized to date and the only TLP shown to have xylanase inhibitor activity (Additional files 7 and 8). Among poplar TLPs, only the four TLPs from the poplar clade 1 (Figure 1) belong to the eukaryote TLP subgroup I (Figure 5). Proteins from TLP subgroup II have been poorly characterized, except for the rosid-specific and tree-enriched clade, which contains several proteins with described antifungal or endo-β-1,3-glucanase activities (Additional file 9). Thirty-one poplar TLPs are distributed in TLP subgroup II, including the 11 TLPs that form the poplar TLP cluster and which are assembled in the tree-enriched clade. To our knowledge, none of the proteins from subgroup III have been characterized at the functional level so far (Additional file 10).

Figure 4 Alignment of thaumatin domains of selected eukaryote TLPs. Amino acid sequence comparison was carried out with ClustalW on MEGA 4 software with the parameters described by [6]. The alignment was then adjusted manually when necessary. [a]Complete protein reference: Glyma-Glyma11g14970.1.

To estimate how TLP structural diversity influences biological and biochemical functions, a 3D structure alignment (3D-SA) was performed with the phylogenetically most distinct TLP structures available among the seven solved to date: the tobacco PR-5d (Nicta-1709500; PDB: 1AUN) from TLP subgroup I and the cherry Pru Av 2 (Pruav-1729981; PDB: 2AHN) from TLP subgroup II (Figure 6). In general, the 3D structures of these TLPs superimpose well, especially the region forming the acidic cleft. Indeed, this region, as well as two hydrophobic or aromatic residues (generally Phe or Tyr), are important for the antifungal or lytic activities of TLPs (Figure 6, [21]). However, although well conserved, some residues of the REDDD and FF motifs adopt

slightly different positions in these two TLPs. For example in the Pru Av 2 structure, the side chain of the aspartate at position 289 (D_{289}) is oriented outside the acidic cleft and the phenylalanine residue F_{119} is replaced by a small non-aromatic residue (Gly) that is positioned differently. It is not clear whether these small differences have a significant impact on the substrate selectivity or protein function. Primary sequence alignment mapping on 3D structures (AM-3D) of PR-5d and Pru Av 2 with sequences from subgroups I and II, respectively, confirmed that the acidic cleft is the most conserved region among eukaryote TLPs (Figure 6). By contrast, although the REDDD amino acids are conserved in most sTLPs, AM-3D of several sTLP
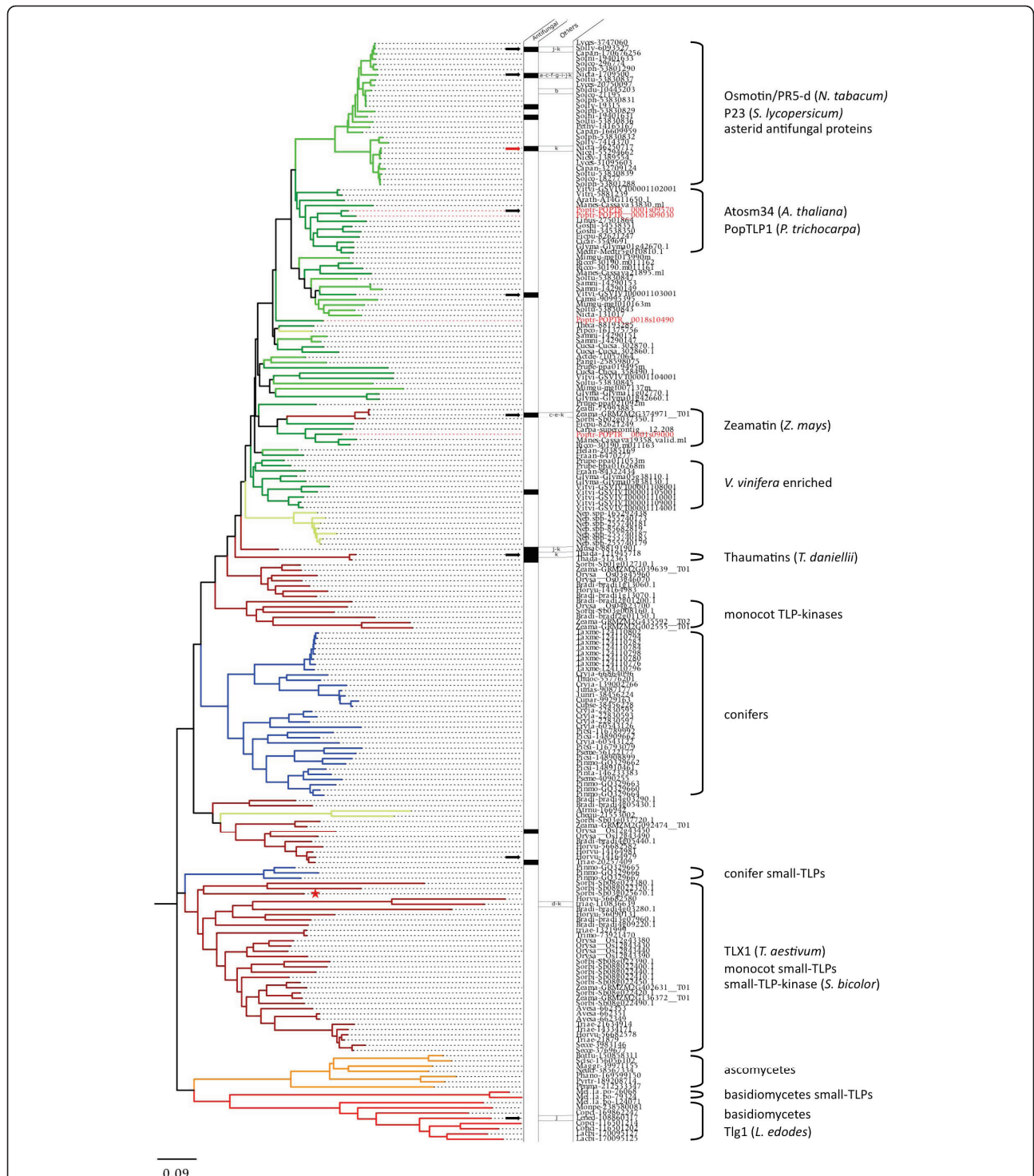
Figure 5 Neighbour-joining tree of the 211 thaumatin domains of TLP subgroup I. Functionally characterized TLPs and corresponding functions are indicated. Poplar sequence names are in red. The 5 letter code before each protein ID corresponds to the 3 first letters of the genus name followed by the 2 first letters of the species name. The red arrow indicates PR-5d used for 3D structure alignment and black arrows indicate sequences used for alignment mapping on 3D Structure (see Figure 6). The red star indicates the Small-TLP-Kinase from *Sorghum bicolor* (Sb03g025670). The two columns successively indicate proteins with demonstrated antifungal activity and other functions. a: protection against abiotic stress; b: antifreeze activity; c: membrane permeabilization activity; d: xylanase inhibitor; e: α-amylase/trypsin inhibition; f: apoptosis-inducing in yeast; g: GPCR binding; h: CMV1-a binding; i: glycoprotein binding; j: endo-β-1,3-glucanase activity; k: solved 3D structures. References corresponding to these data are summarized in Additional file 8. Branch lengths are proportional to phylogenetic distances.
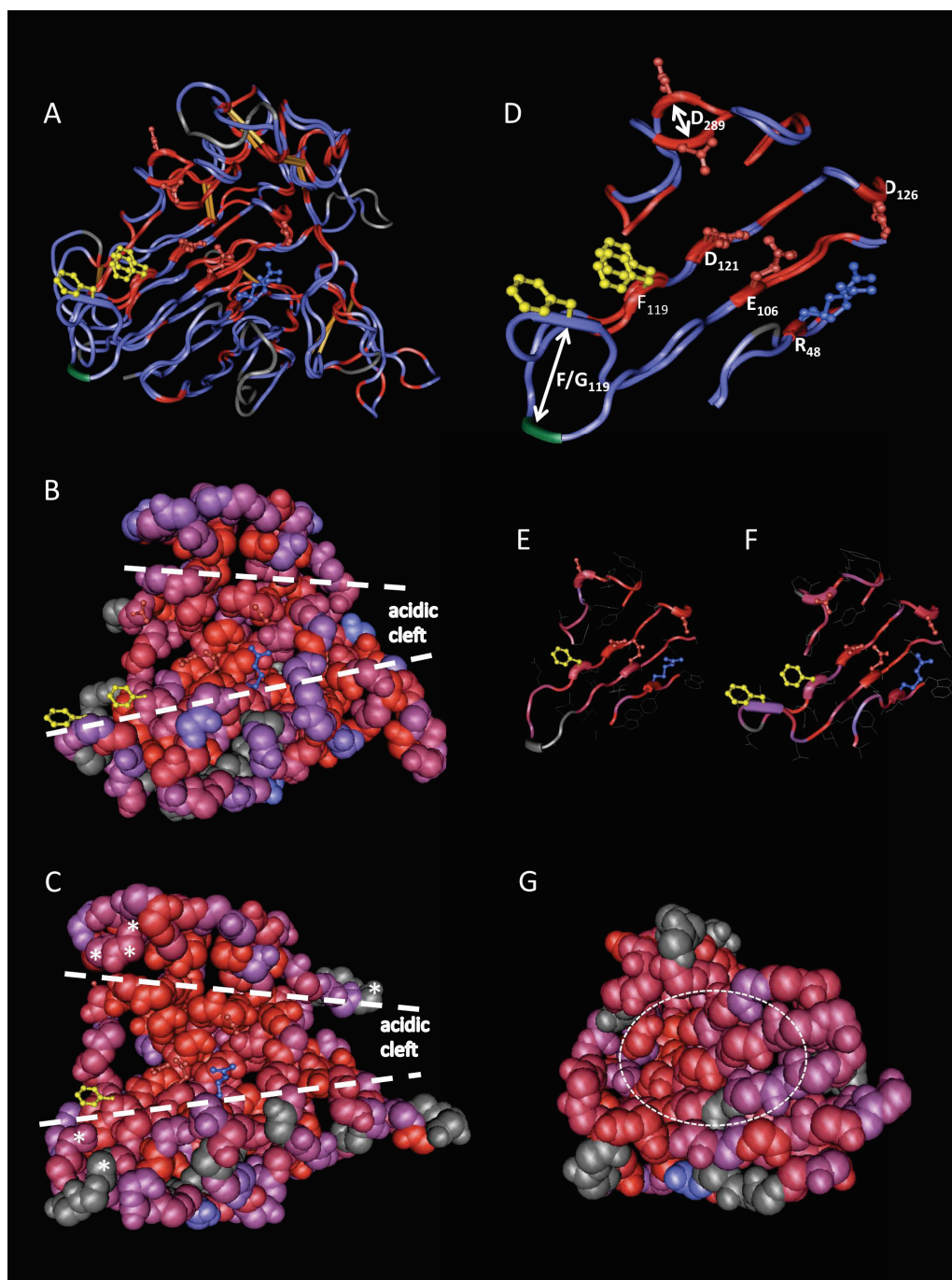
Figure 6 3D structure alignment (3D-SA) and alignment mapping on 3D structure (AM-3D) of eukaryote TLPs. Amino acids of the REDDD and FF motifs are represented with side-chains in balls and sticks. Color code of side-chains, red: negatively charged; blue: positively charged; yellow: hydrophobic. White dashed-lines indicate acidic cleft limits. (A), 3D-SA of tobacco PR-5d and cherry Pru av 2. Protein backbone color code, red: identical amino acids; blue: different amino acids; grey: unaligned residues, green: glycine/phenylalanine residues discussed in the text. Disulfide bonds are in orange. (B), AM-3D of 9 subgroup I TLPs using the PR-5d structure as template. The four-color code of the protein backbone (from red to blue) corresponds to a decrease in amino acid conservation. (C), AM-3D of 15 subgroup II TLPs using the Pru Av 2 structure as template. Color code and annotations are as in B. Amino acids under diversifying selection [39] are indicated by white asterisks. (D, E and F), Highlights of β-sheets forming the acidic cleft in A, B and C respectively. Color code is similar to that in A, B and C. In D, the residues forming the REDDD and FF motifs are numbered as in Figure 4. White arrows indicate motif differences discussed in the text. (G), AM-3D of the 9 small-TLPs indicated in Additional file 7 using the TLX1 structure as template. Color code is similar to that in B. A white dashed ellipse marks the missing acidic cleft.

sequences with the recently solved structure of wheat TLX1 (PDB: 1KWN) revealed neither an acidic cleft nor any particular conserved region which could be linked to the reported xylanase inhibitor function (Figuer 6, [12]).

Alignment of the 14 TLP-Ks identified from six different plant species (two dicots and four monocots), including the four poplar TLP-Ks, revealed that the thaumatin domain of TLP-Ks is similar to that of typical TLPs, possessing both the conserved residues involved in the acidic cleft and the cysteine residues (Figure 7). The protein kinase domain of TLP-Ks is extremely well conserved, even among monocots and dicots, and contains two fully conserved residues D_{740} and D_{758} known to be part of the catalytic motif [33]. A predicted TM domain is present between the thaumatin and the

protein kinase domains in all TLP-K sequences (Figure 7, Additional file 5), except Bradi-2g01200, which might be due to an erroneous interdomain annotation in the *Brachypodium distachyon* genome.

Discussion

The recent release of the *P. trichocarpa* genome, the first tree genome available, paved the way for high-throughput genomic and computational analyses of multigene families, and has defined *Populus* as a model organism in forest biology [34]. Considering that leaf rust fungi are responsible for considerable damage in poplar plantations, the *Populus/Melampsora* interaction has emerged as a model pathosystem in forest pathology [26]. In order to decipher the molecular basis of poplar resistance against this biotrophic fungus, in-
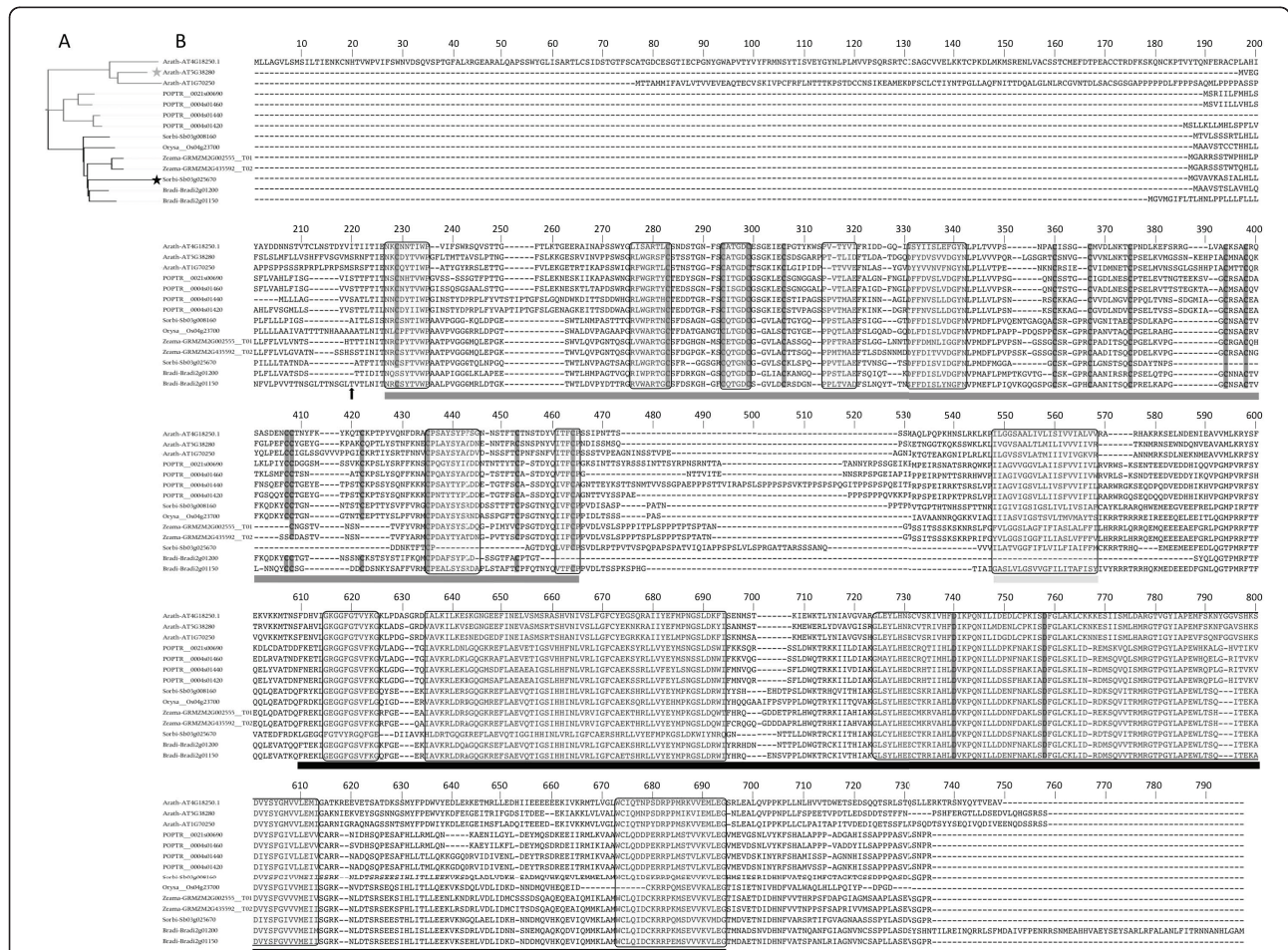

Figure 7 Amino acid sequence comparison of plant TLP-kinases (TLP-Ks). (A), Neighbour-joining tree of the 14 TLP-Ks identified in plants. Branch lengths are proportional to phylogenetic distances. Black star: sTLP-K from *Brachypodium distachyon*; grey star: TLP-K from *Arabidopsis thaliana* described in [51]. (B), ClustalW amino acid alignment using the parameters described by [6] and manually adjusted. Thaumatin and protein kinase domains are respectively underlined in dark grey and black. Phobius [72] predicted transmembrane domain is underlined in light grey. Shaded boxes indicate highly conserved sequences. The arrow indicates the end of the predicted signal peptide. Vertical bars indicate cysteine residues in the thaumatin domain and aspartate residues forming the catalytic site of the kinase domain.

depth and exhaustive studies of defense-related functions require a reliable annotation of gene families before we can understand their structural and functional diversity. We have therefore performed a genome-wide analysis of the TLP multigene family, which comprises many stress-inducible proteins in *P. trichocarpa*. Extension of the phylogenetic analysis to include TLPs from other eukaryotes extends our knowledge about the evolution of TLPs.

TLPs in plant genomes: an evolutionary diversification

The history of TLP multigene families retraced from 18 plant genome sequences shows a strong evolutionary diversification from the green alga *C. reinhardtii* to the black cottonwood tree *P. trichocarpa* (Table 1). In the co-evolution of host-microbe interactions, plants and fungi acquire new weapons that promote their resistance or virulence, respectively [35]. As a consequence of this arms race, the size of some multigene families involved in resistance (such as NB-LRRs) has greatly increased in poplar and other higher plants [36]. It is accepted that this increase, sometimes specific to certain organisms, represents an important means for generating functional diversity via sub- or neo-functionalization of paralogs [37]. Analysis of natural selection is increasingly used in plant pathology to estimate how evolutionary forces impact genes and corresponding proteins, from the scale of amino acid sites to gene families [38]. Poplar TLP genes have recently been investigated for natural selection. The four TLP-Ks and 10 of the 11 TLPs that belong to clade 2 (i.e. the poplar TLP cluster) were shown to be driven by diversifying (positive) selection (Figure 1, [39]). More precisely, several exposed amino acids of TLPs are under diversifying selection, whereas amino acids forming the acidic cleft are under purifying (negative) selection and thus well-conserved (Figure 6). Conservation of the acidic cleft could be necessary to maintain antifungal activity, whereas diversification of exposed amino acids could avoid recognition by pathogen enzyme-inhibitors or proteases [40].

Is the antifungal activity of TLPs a universal property?

Historically, TLPs have been described as biotic and abiotic stress-responsive proteins and were called TLPs/PR5 or osmotin/osmotin-like proteins (OLPs), depending on the stress condition (i.e. biotic or abiotic stress, respectively) in which these proteins or their closest homologs were first described. As already suggested by Shatters and collaborators [6], phylogenetic analyses do not support this separate nomenclature that generates semantic confusion in the literature [41]. Our broad sequence analysis of eukaryote TLPs confirms that there is no clear difference among TLPs and OLPs, since different TLP functions are not separated by distinct

phylogenetic clades. In fact, the major biochemical properties of TLPs such as antifungal or endo-β-1,3-glucanase activity are widespread among eukaryotes (Figure 5 and Additional file 9). At the structural level, most TLPs are predicted to share a conserved acidic cleft, which is usually associated with an antifungal property, suggesting that this property is universal among eukaryote TLPs (Figure 6). Although more subtle conformational differences might explain the large variety of properties described so far for TLPs, the phylogenetic and functional data do not justify adoption of a distinct nomenclature between biotic- and abiotic-responsive TLPs. An exception to this statement might be considered for TLPs with different domain organisations such as sTLPs or TLP-Ks. Indeed, sTLPs are assumed to act as xylanase inhibitors and no antifungal activity has yet been reported [12]. This functional divergence is consistent with the important structural differences observed and in particular with the absence of a well-defined acidic cleft (Figure 6).

Poplar TLPs: stress-responsive proteins, but not only

The release of the *P. trichocarpa* genome version 2.0 and its integration into the Phytozome portal enabled a drastic improvement of TLP gene annotation and the validation of more than 70% of the 59 *TLP* gene models from the *Populus* genome version 1.1 [Additional file 1, [24]). The expression analysis of TLPs during biotic and abiotic stresses supports the idea that TLPs, like other PR proteins, belong to a general plant stress response pathway rather than being specific to distinct stresses, as often hypothesized [1]. This is exemplified by *PopTLP1*, whose expression is induced by diverse environmental constraints such as high ozone, UV-B, drought, copper and infection by rust fungi (Figure 1 and Additional file 4). In addition, *PopTLP1* is the closest homolog of *A. thaliana Atosm34* (At4g11650), which also accumulates during both biotic and abiotic stress conditions [8,42]. The RT-qPCR expression profile of *PopTLP1* in rust-infected poplar leaves confirmed transient transcript accumulation during host-specific resistance (Additional file 4). This profile is in accordance with results obtained by similar approaches for several PR proteins in this pathosystem [27] and confirms the involvement of TLPs in poplar defense.

However, the fact that only 19 of the 42 poplar *TLPs* are transcriptionally regulated in the stress conditions investigated suggests that their role in poplar might not be restricted to stress response but that they could have other roles, such as during development. Indeed, some TLPs have been reported to accumulate during plant developmental stages or conditions such as in ovular secretions or during leaf aging [43,44]. In addition, TLPs have been extensively described as ripening-associated

proteins that accumulate strongly in fruit during maturation [45,46]. It has been shown recently in hybrid poplars that two TLPs belonging to the tree-specific and stress-responsive TLP cluster (Figure 1 and 2) are present in the phloem of healthy non-stressed plants [47]. Taken together, these results suggest that the expansion of this multigene family in poplar could also be related to tree-specific developmental stages.

The poplar TLP cluster contains tree-specific and stress-responsive proteins

The poplar TLP cluster is an assembly of 11 successive genes on scaffold 1 and is considerably enriched in TE for a gene-containing genomic region. Indeed, TE account for 52% of the TLP cluster region compared with an average coverage of 42% in the whole genome [24]. Moreover, LTR TE from the Gypsy class are specifically over-represented, covering 37% of the TLP cluster compared with 5% in the whole *P. trichocarpa* genome sequence (Figure 2, [24]). This class of TE might be a source of genome plasticity in plants [48]. The very well-conserved exon-intron structure of the genes in the TLP cluster supports a mechanism of tandem duplication from a unique ancestral gene (Figure 1). Taken together, these results strongly suggest that this cluster likely resulted from recent tandem duplications driven by TE activity. Futhermore, the TLP cluster appears to be highly responsive to fungal infection in poplar and belongs to a rosid-specific and tree-enriched clade in our complete phylogeny of eukaryote TLPs (Figure 1, Figure 3 and Additional file 9). TLPs from cherry, chestnut, apple and peach trees that exhibit antifungal and/or endo-β-1,3-glucanase activities (Additional file 8 [49,50]) also belong to this clade. Thus, the TLP cluster appears to be a tree-specific and rust-responsive group of TLPs that are of outstanding interest for further analyses focusing on tree and TLP specificities in defense against pathogens. More precisely, two TLPs recently identified at the protein level in the phloem of hybrid poplar constitute excellent candidates for future investigations ([47], Figure 1).

TLP-Ks: defense proteins recruited for signaling?

TLP-Ks result from the fusion between two genes coding for a TLP and a protein kinase. They have been hypothesized to act as receptor-like kinases (RLKs), where the extracellular TLP could sense pathogens and the cytoplasmic kinase could relay downstream signaling [23]. This assumption was strengthened by the demonstration that plants overexpressing an *A. thaliana* TLP-K showed a delay in the appearance of disease symptoms [51]. The ability of plants to recruit defense proteins to form a RLK involved in pathogen sensing has already been suggested for PR1 and PR3 [1]. The

strong homology of the kinase domain between TLP-K and some lectin-kinases reinforces the speculation about the potential role of TLP-K in the induction of the defense system, since a rice lectin-kinase has been shown to confer resistance to the rice blast [52]. In the *P. trichocarpa* genome, three TLP-Ks are organized in tandem on scaffold 4 and are interspersed by other protein kinase-encoding genes. This genomic region is referred to as the 'TLP-K cluster' (Figure 2). Genetic and physical mapping of *Melampsora* rust resistance genes in natural populations of *P. trichocarpa* identified a locus encoding two *TLP-K* genes on chromosome 4, which possibly corresponds to the TLP-K cluster in scaffold 4 of the *P. trichocarpa* genome sequence [53]. Hence, although evidence is still needed to clarify the exact role of these TLP-Ks in poplar, this opens interesting perspectives concerning new RLK types related to poplar defense against *Melampsora* spp. rust pathogens.

Conclusion

TLPs are eukaryote proteins that constitute small and monophyletic families in invertebrate animals and fungi whereas they are more diverse and are organized in large multigene families in plants. Regardless of their origin, it appears that many typical TLPs possess an antifungal activity, which is probably linked to a conserved acidic cleft in their 3D structure. In plants, TLPs have undergone a drastic evolutionary diversification including the evolution of tree-enriched clades and of TLPs fused to protein kinase domains. The poplar genome encodes 42 validated *TLP* gene models, including four TLP-kinases. Some poplar TLP transcripts accumulate specifically under abiotic or biotic stress conditions, which can be strongly correlated with their phylogeny. In the poplar genome, a tree-specific and stress-responsive cluster of tandemly-duplicated *TLP* genes should be of interest for understanding the unique attributes of defense against pathogen attacks that have evolved in trees.

Methods

Identification and annotation of TLP genes in P. trichocarpa

TLP genes were identified in the *P. trichocarpa* 'Nisqually-1' genome using thaumatin and osmotin keywords and amino acid sequence homology searches. Manual gene annotation was performed by finding missing start/stop codons, by defining correct exon/intron borders, by analyzing perfectly matching ESTs and by taking into account the amino acid conservation of the thaumatin domain (Pfam: PF00314) and in particular the position of conserved cysteines. Alignments with closest homologs in the Phytozome portal were used to reconstruct gene structure and corresponding amino acid sequences.

Allelic versions detected by the Phytozome annotation (two adjacent genes in the genome assembly v1.1 that correspond to a single locus in the v2.0 assembly) or incomplete genes were discarded and not considered for further analysis. Transposable element analysis was carried out with the CENSOR software available on the Giri database [54].

Search for TLP in public genomic databases and sequence analyses

The 18 plant genome sequences available on the Phytozome portal [29] (June 2010) were mined using sequence homology searches. Only *TLP* gene models encoding a complete TLP domain were reserved for sequence comparison. The NCBI protein database [32] was mined using sequence homology and keyword searches (i.e. thaumatin and osmotin). TLP sequences from non-sequenced plants, fungi and animals were individually examined and amino acid sequences with a complete thaumatin family domain were retained for further analyses. Sequences from *M. larici-populina* and *P. graminis* f.sp. *tritici* were retrieved from the Joint Genome Institue (JGI, [55]) and the Broad Institute [56] websites, respectively.

Sequence alignment and construction of phylogenetic tree

For all amino acid sequence comparisons, the thaumatin domain covering almost 95% of the mature TLP protein was considered. Limits of the TLP domain were defined as N-x-C-x(3)-V/I-W and Y-x-I/V-x-F-C-x in the N- and C-terminal ends, respectively. Amino acid sequence alignments were performed using ClustalW as described by [6], DIALIGN (http://dialign-tx.gobics.de/) as well as MAFFT (http://mafft.cbrc.jp/alignment/server/) methods. In all cases, the raw output alignments required deep manual re-adjustment to proceed further with phylogenetic reconstruction. Raw alignments were thus imported into the Molecular Evolutionary Genetics Analysis (MEGA) package 4.1 [57] and manually adjusted. Phylogenetic analyses were conducted using the Neighbour-Joining method with the pairwise deletion option for handling alignment gaps and the Poisson correction model for distance computation. Bootstrap tests were conducted using 1,000 replicates. Branch lengths are proportional to phylogenetic distances.

Expression of poplar TLPs

Transcriptional data for *Populus*/*Melampsora* interaction were extracted from published studies with significant fold-changes as described by the respective authors (Figure 1, [27,30,31]). Other information pertaining to the poplar transcriptome during stress-related situations were extracted from the PopGenIE portal [58,59] and from the literature. Proteomic data for *Populus/Melampsora* interaction were retrieved from the PROTICdb database [60,61] and from the literature for other stress-related situations.

3D structure analyses

3D-SA and MA-3D were carried out with Cn3D software [62]. Sequences were manually aligned using the integrated sequence viewer in editor mode and reference structures were retrieved from NCBI structure database [32]. For MA-3D, sequence conservation has been estimated with a four level color code from red to blue, reporting a respectively strong to weak amino acid variety.

RT-qPCR analyses

Isolates 98AG31 (virulent, pathotype 3-4-7) and 93ID6 (avirulent, pathotype 3-4) of *M. larici-populina* were used in this study. Rust urediniospore multiplication and plant inoculation procedures were performed as previously described [27], using the same inoculum doses (100,000 urediniospores/ml), leaf plastochron indexes for detached *P. trichocarpa* X *Populus deltoides* 'Beaupré' leaves and identical culture conditions. For time-course infection analyses, leaves were harvested at the following time-points: 0, 12, 15, 18, 21, 24, 36, 48, 72, 96, 120 and 168 hpi. RNA extraction, quality control and cDNA synthesis were performed as previously described in [27]. In order to assess transcript levels by RT-qPCR, specific primers for the *PopTLP1* gene (Poptr_0001s09570 in *P. trichocarpa* genome; 5' CCAGACTTGGTATCTTAATG; 3' GTTAC-CAAACTGATTTAACG) were designed and quantitative PCR was carried out as previously described [63], with technical and biological duplicates. Transcript expression was normalized to a reference ubiquitin transcript (Poptr_0015s01600 in *P. trichocarpa* genome; 5' GCAGG-GAAACAGTGAGGAAGG; 3' TGGACTCACGAGGA-CAG) using ratio calculation as described in [64].

Additional material

Additional file 1: Annotation of *TLP* genes in the *Populus trichocarpa* 'Nisqually-1' genome sequence. [a]*TLP* gene models retrieved in the *Populus trichocarpa* 'Nisqually-1' genome version 1.1 from the JGI website [55]. [b]*TLP* gene models retrieved in the *P. trichocarpa* 'Nisqually-1' genome version 2.0 from the Phytozome portal [29].

Additional file 2: List of amino acid sequences deduced from the 42 *P. trichocarpa* TLP gene models. Stop codons are represented by asterisks.

Additional file 3: Transposable element (TE) features of the TLP cluster. [a]Percentage of the 350 kb total length of the TLP cluster covered by TE.

Additional file 4: *PopTLP1* RTqPCR expression profile. Total RNA was isolated from mock-inoculated or inoculated leaves of *Populus trichocarpa* X *Populus deltoides* 'Beaupré' with either compatible (white diamonds, strain 98AG31) or incompatible (black diamonds, strain 93ID6) strains of *Melampsora larici-populina* between 12 and 168 hours post-

inoculation (hpi). RT-qPCR results are presented as expression ratios. *Populus* ubiquitin transcripts were as a reference gene for normalization. n = 2 (except for I168, n = 1), error bar: standard deviation.

Additional file 5: Small-TLP-kinase domains and features. The signal peptide and the transmembrane domain of the small-TLP-kinase of *Sorghum bicolor* (Sb03g025670) are predicted by the Phobius program [72].

Additional file 6: Protein accession numbers and phylogenetic classification of corresponding species used in this study. [a]Organism code used in the study, corresponding to the 3 first letters of the genus name followed by the 2 first letters of the species name (ex.: *Arabidopsis thaliana* = Arath), except for *Nepenthes species pluralis* (Nep.spp) and *Melampsora larici-populina* (Mel.la.po). [b]Number of TLP sequences used in this study. [c]Accession numbers are preceded by the organism code. Sequences retrieved from the NCBI protein database [32] are in black, those retrieved from the Phytozome portal [29] are in red. The 3 sequences of *M. larici-populina* were retrieved from the JGI genome website [55] and are labelled with JGI protein IDs.

Additional file 7: Neighbour-joining tree of eukaryote small-TLPs. Branch lengths are proportional to phylogenetic distances. Branch color and protein ID codes correspond to those in Figures 3 and 5, respectively. Supplemental sequences from the *Puccinia graminis* f.sp. *tritici* genome sequence were retrieved from the Broad Institute website [56] (gene IDs PGTG_00965.2; PGTG_00963.2; PGTG_19613.2; PGTG_19646.2). Black star: small-TLP-Kinase from *Sorghum bicolor*; grey star: TLX 1 from *Triticum aestivum*; black arrows: sequences used for the structural analysis in Figure 6.

Additional file 8: List of functionally characterized TLPs described in this study. In some cases, several studies participated to the characterization of a given TLP function. For clarity, only one reference is given per function and per protein (most relevant, otherwise first published). [a]GPCRs: G-Protein-Coupled Receptors. [b]Numbers refer to the complete reference in the text.

Additional file 9: Neighbour-joining tree of the 341 thaumatin domains of TLP Subgroup II. Functionally characterized TLPs and corresponding functions are indicated. Poplar sequence names are in red. The five-letter code before proteins IDs indicate genus and species. Red arrows indicate protein structures used for 3D structure alignment while black arrows indicate sequences used for alignment mapping on 3D structure in Figure 6. Antifungal column includes both *in vitro*- and transgenic-based antifungal demonstrations. In the other column, a: transgenic abiotic stress protection; b: antifreeze activity; c: membrane permeabilization activity; d: xylanase inhibitor; e: α-amylase/trypsin inhibition; f: apoptosis-inducing in yeast; g: GPCR binding; h: CMV1-a binding; i: glycoprotein binding; j: endo-β-1,3-glucanase activity; k: 3D structure solved. References corresponding to these data are summarized in Additional file 8. Branch lengths are proportional to phylogenetic distances.

Additional file 10: Neighbour-joining tree of uncharacterized eukaryote TLPs from TLP subgroup III. Branch lengths are proportional to phylogenetic distances. Branch color and protein IDs codes correspond to those in Figures 3 and 5, respectively. Poplar sequence names are in red.

Abbreviations
TLP: thaumatin-like protein; PR: pathogenesis-related; GPCR: G protein-coupled receptor; sTLP: small-TLP; TM: transmembrane domain; TLP-K: TLP-kinase; CDS: coding DNA sequence; sTLP-K: small-TLP-kinase; TE: transposable element; 3D-SA: 3D structure alignment; AM-3D: alignment mapping on 3D structure; OLP: osmotin-like protein; JGI: joint genome institute; RLK; receptor-like kinase; EST: expressed sequence tag; hpi: hour-post inoculation; NB-LRR: nucleotide binding-leucine rich repeat; LTR: long terminal repeat.

Author details
[1]INRA†/Nancy Université, Unité Mixte de Recherche 1136 'Interactions Arbres/Micro-organismes', Centre INRA de Nancy, F-54280 Champenoux, France. [2]Plant Research Laboratory, 122 Plant Biology Laboratory, Michigan State University, East Lansing, Michigan, 48864, USA.

Authors' contributions
BP and SD performed conceptual and experimental designs. BP carried out experimental procedures, *in silico* analyses and drafted the manuscript. IM compiled transcriptional data concerning *Populus-Melampsora* interactions from the literature. SD and NR supervised the work and helped with conceptual design and data analysis. All authors participated in depth reading and revising the manuscript. All authors read and approved the final manuscript.

References
1. Van Loon LC, Rep M, Pieterse CMJ: **Significance of Inducible Defense-related Proteins in Infected Plants.** *Annu Rev Phytopathol* 2006, **44**:135-62.
2. Dodds PN, Rathjen JP: **Plant immunity: towards an integrated view of plant-pathogen interactions.** *Nat Rev Genet* 2010, **11**:538-548.
3. Van Der Wel H, Loeve K: **Isolation and Characterization of Thaumatin I and II, the Sweet-Tasting Proteins from *Thaumatococcus daniellii* Benth.** *Eur J Biochem* 1972, **31**:221-225.
4. Brandazza A, Angeli S, Tegoni M, Cambillau C, Pelosi P: **Plant stress proteins of the thaumatin-like family discovered in animals.** *FEBS Letters* 2004, **572**:3-7.
5. Sakamoto Y, Watnabe H, Nagai M, Nakade K, Takahashi M, Sato T: ***Lentinula edodes tlg1* Encodes a Thaumatin-Like Protein That Is Involved in Lentinan Degradation and Fruiting Body Senescence.** *Plant Physiol* 2006, **141**:793-801.
6. Shatters RG Jr, Boykin LM, Lapointe SL, Hunter WB, Weathersbee AA: **Phylogenetic and Structural Relationships of the PR5 Gene Family Reveal an Ancient Multigene Family Conserved in Plants and Select Animal Taxa.** *J Mol Evol* 2006, **63**:12-29.
7. Liu JJ, Sturrock R, Ekramoddoullah AKM: **The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function.** *Plant Cell Rep* 2010, **29**:419-436.
8. Mukherjee AK, Carp MJ, Zuchman R, Ziv T, Horwitz BA, Gepstein S: **Proteomics of the response of *Arabidopsis thaliana* to infection with *Alternaria brassicicola*.** *Journal of Proteomics* 2010, **73**:709-720.
9. Islam MA, Sturrock RN, Holmes A, Ekramoddoullah AKM: **Ultrastructural studies of *Phellinus sulphurascens* infection of Douglas-fir roots and immunolocalization of host pathogenesis-related proteins.** *Mycol Res* 2009, **113**:700-712.
10. Wang X, Tang C, Deng L, Cai G, Liu X, Han Q, Buchenauer H, Wei G, Han D, Huang L, Kang Z: **Characterization of a pathogenesis-related thaumatin-like protein gene *TaPR5* from wheat induced by stripe rust fungus.** *Physiol Plant* 2010, **139**:27-38.
11. Vigers AJ, Roberts WK, Selitrennikoff CP: **A new family of plant antifungal proteins.** *Mol Plant Microbe Interact* 1991, **4**:315-23.
12. Fierens E, Rombouts S, Gebruers K, Goesaert H, Brijs K, Beaugrand J, Volckaert G, Van Campenhout S, Proost P, Courtin CM, Delcour JA: **TLX1, a novel type of xylanase inhibitor from wheat (*Triticum aestivum*) belonging to the thaumatin family.** *Biochem J* 2007, **403**:583-591.
13. Schimoler-O'Rourke R, Richardson M, Selitrennikoff CP: **Zeamatin Inhibits Trypsin and α-Amylase Activities.** *Appl Env Microbiol* 2001, **67**:2365-2366.
14. Narasimhan ML, Coca MA, Jin J, Yamauchi T, Ito Y, Kadowaki T, Kim KK, Pardo JM, Damsz B, Hasegawa PM, Yun DJ, Bressan RA: **Osmotin Is a**

Homolog of Mammalian Adiponectin and Controls Apoptosis in Yeast through a homolog of Mammalian Adiponectin Receptor. *Mol Cell* 2005, **17**:171-180.

15. Hon WC, Griffith M, Mlynarz A, Kwok YC, Yang DSC: **Antifreeze Proteins in Winter Rye Similar to Pathogenesis-Related Proteins.** *Plant Physiol* 1995, **109**:879-889.

16. Rajam MV, Chandola N, Saiprasad Goud P, Singh D, Kashyap V, Choudhary ML, Sihachakr D: **Thaumatin gene confers resistance to fungal pathogen as well as tolerance to abiotic stresses in transgenic tobacco plants.** *Biol Plant* 2007, **51**:135-141.

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

18. Fierens E, Gebruers K, Courtin CM, Delcour JA: **Xylanase Inhibitors Bind to Nonstarch Polysaccharides.** *J Agri Food Chem* 2008, **56**:564-570.

19. Fierens E, Gebruers K, Voet ARD, De Maeyer M, Courtin CM, Delcour JA: **Biochemical and structural characterization of TLXI, the *Triticum aestivum* L. thaumatin-like xylanase inhibitor.** *J Enz Inhibit Med Chem* 2009, **24**:646-654.

20. Liu JJ, Zamani A, Ekramoddoullah AKM: **Expression profiling of a complex thaumatin-like protein family in western white pine.** *Planta* 2010, **231**:637-651.

21. Koiwa H, Kato H, Nakatsu T, Oda J, Yamada Y, Sato F: **Crystal Structure of Tobacco PR-5d Protein at 1,8 Å Resolution Reveals a Conserved Acidic Cleft Structure in Antifungal Thaumatin-Like Proteins.** *J Mol Biol* 1999, **286**:1137-1145.

22. Vandermarliere E, Lammens W, Schoepe J, Rombouts S, Fierens E, Gebruers K, Volckaert G, Rabijns A, Delcours JA, Strelkov SV, Courtin CM: **Crystal structure of the noncompetitive xylanase inhibitor TLX1, member of the small thaumatin-like protein family.** *Proteins* 2010, **78**:2391-2394.

23. Wang X, Zafian P, Choudhary M, Lawton M: **The PR5K receptor protein kinase from *Arabidopsis thaliana* is structurally related to a family of plant defense proteins.** *Proc Natl Acad Sci* 1996, **93**:2598-2602.

24. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, *et al*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**:1596-1604.

25. Pinon J, Frey P: **Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance.** In *Rust Diseases of Willow and Poplar. Edited by: MH Pei and AR McCracken.* CAB International, Wallingford; 2005:139-154.

26. Duplessis S, Major I, Martin F, Séguin A: **Poplar and Pathogen Interactions: Insights from *Populus* Genome-Wide Analyses of Resistance and Defense Gene Families and Gene Expression Profiling.** *Crit Rev Plant Sci* 2009, **28**:309-334.

27. Rinaldi C, Kohler A, Frey P, Duchaussoy F, Ningre N, Couloux A, Wincker P, Le Thiec D, Fluch S, Martin F, Duplessis S: **Transcript Profiling of Poplar Leaves upon Infection with Compatible and Incompatible Strains of the Foliar Rust *Melampsora larici-populina*.** *Plant Physiol* 2007, **144**:347-366.

28. Street NR, Tsai CJ: ***Populus* Resources and Bioinformatics.** In *Genetics and Genomics of Populus. Edited by: S Jansson. Plant Genetics and Genomics: Crops and Models;* .

29. **Phytozome portal.** [http://www.phytozome.net/].

30. Miranda M, Ralph SG, Mellway R, White R, Heath MC, Bohlmann J, Constabel CP: **The Transcriptionnal Response of Hybrid Poplar (*Populus trichocarpa* x *P. deltoides*) to infection by *Melampsora medusae* Leaf Rust Involves Induction of Flavonoid Pathway Genes Leading to the Accumulation of Proanthocyanidins.** *Mol Plant Microbe Interact* 2007, **20**:816-83147.

31. Azaiez A, Boyle B, Levée V, Séguin A: **Transcriptome Profiling in Hybrid Poplar Following Interactions with *Melampsora* Rust Fungi.** *Mol Plant Microbe Interact* 2009, **22**:190-200.

32. **NCBI database.** [http://www.ncbi.nlm.nih.gov/].

33. Hanks SK, Quinn AM: **Protein Kinase Catalytic Domain Sequence Database: Identification of Conserved Features of Primary Structure and Classification of Family Members.** *Methods Enzymol* 1991, **200**:38-62.

34. Yang X, Kalluri UC, DiFazio SP, Wullschleger SD, Tschaplinski TJ, Cheng ZM, Tuskan GA: **Poplar Genomics: State of the Science.** *Crit Rev Plant Sci* 2009, **28**:285-308.

35. Burdon JJ, Thrall PH: **Coevolution of Plants and Their Pathogens in Natural Habitats.** *Science* 2009, **324**:755-756.

36. Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F: **Genome-wide identification of NBS resistance genes in *Populus trichocarpa*.** *Plant Mol Biol* 2008, **66**:619-636.

37. Paterson AH, Freeling M, Tang H, Wang X: **Insights from the Comparison of Plant Genome Sequences.** *Annu Rev Plant Biol* 2010, **61**:349-372.

38. Ellegren H: **Comparative genomics and the study of evolution by natural selection.** *Mol ecol* 2008, **17**:4586-4596.

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

40. Misas-Villamil JC, van der Hoorn RAL: **Enzyme-inhibitor interactions at the plant-pathogen interface.** *Curr Opin Plant Biol* 2008, **11**:380-388.

41. Monteiro S, Barakat M, Piçarra-Pereira MA, Teixeira AR, Ferreira RB: **Osmotin and Thaumatin from Grape: A Putative General Defense Mechanism Against Pathogenic Fungi.** *Biochem Cell Biol* 2003, **93**:1505-1512.

42. **Arabidopsis eFP Browser.** [http://www.bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi].

43. O'Leary SJB, Poulis BAD, Von Aderkas P: **Identification of two thaumatin-like proteins (TLPs) in the pollination drop of hybrid yew that may play a role in pathogen defence during pollen collection.** *Tree Physiol* 2007, **27**:1649-1659.

44. Capelli N, Diogon T, Greppin H, Simon P: **Isolation and characterization of a cDNA clone encoding an osmotin-like protein from *Arabidopsis thaliana*.** *Gene* 1997, **191**:51-56.

45. Kim YS, Park JY, Kim KS, Ko MK, Cheong SJ, Oh BJ: **A thaumatin-like gene in nonclimateric pepper fruit used as molecular marker in probing resistance, ripening, and sugar accumulation.** *Plant Mol Biol* 2002, **49**:125-135.

46. Tattersall DB, van Heeswijck R, Hoj PB: **Identification and Characterization of a Fruit-Specific, Thaumatin-Like Protein That Accumulates at Very high Levels in Conjunction with the Onset of Sugar Accumulation and Berry Softening in Grapes.** *Plant Physiol* 1997, **114**:759-769.

47. Dafoe NJ, Gowen BE, Constabel P: **Thaumatin-like proteins are differentially expressed and localized in phloem tissues of hybrid poplar.** *BMC Plant Biology* 2010, **10**:191.

48. Zedek F, Smerda J, Smarda P, Bures P: **Correlated evolution of LTR retrotransposons and genome size in the genus *Eleocharis*.** *BMC Plant Biology* 2010, **10**:265.

49. Menu-Bouaouiche L, Vriet C, Peumans WJ, Barre A, Van Damme EJM, Rougé P: **A molecular basis for the endo-β-1,3-glucanase activity of the thaumatin-like proteins from edible fruits.** *Biochimie* 2003, **85**:123-131.

50. Palacin A, Tordesillas L, Gamboa P, Sanchez-Monge R, Cueast-Herranz J, Sanz ML, Barber D, Salcedo G, Diaz-Perales A: **Characterization of peach thaumatin-like proteins and their identification as major peach allergens.** *Clin Exp Allergy* 2010, **40**:1422-1430.

51. Guo Z, Bonos S, Meyer WA, Day PR, Belanger FC: **Transgenic creeping bentgrass with delayed dollar spot symptoms.** *Mol Breeding* 2003, **11**:95-101.

52. Chen X, Shang J, Chen D, Lei C, Zou Y, Zhai W, Liu G, Xu J, Ling Z, Ma B, Wang Y, Zhao X, Zhu L: **A B-lectin receptor kinase gene conferring rice blast resistance.** *Plant J* 2006, **46**:794-804.

53. Yin TM, Difazio SP, Gunter LE, Jawdy SS, Boerjan W, Tuskan GA: **Genetic and physical mapping of *Melampsora* rust resistance in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence.** *New Phytol* 2004, **164**:95-105.

54. **Giri website.** [http://www.girinst.org/censor/].

55. **Joint Genome Institute.** [http://www.jgi.doe.gov/].

56. **Broad Institute.** [http://www.broadinstitute.org/].

57. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.** *Mol Biol Evol* 2007, **24**:1596-1599.

58. **PopGenIE: The Populus Genome Integrative Explorer.** [http://www.popgenie.org/].

59. Sjödin A, Street NR, Sandberg G, Gustafsson P, Jansson S: **PopGenIE: The *Populus* Genome Integrative Explorer. A new tool for exploring the *Populus* genome.** *New Phytol* 2009, **182**:1013-1025.

60. **PROTICdb.** [http://www.cbib.u-bordeaux2.fr/outils/Protic/home/].

61. Ferry-Dumazet H, Houel G, Montalent P, Moreau L, Langella O, Negroni L, Vincent D, Lalanne C, de Daruvar A, Plomion C, Zivy M, Joets J: **PROTICdb:**

A web-based application to store, track, query, and compare plant proteome data. *Proteomics* 2005, **5**:2069-2081.

62. Hogue CW: **Cn3D: a new generation of three-dimensional molecular structure viewer.** *Trends Biochem Sci* 1997, **8**:314-316.

63. Hacquard S, Delaruelle C, Legué V, Kohler A, Frey P, Martin F, Duplessis S: **Laser capture microdissection of uredinia formed by *Melampsora larici-populina* revealed a transcriptionnal switch between biotrophic and sporulation phase.** *Mol Plant Microbe Interact* 2010, **23**:1275-1286.

64. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29**:2002-2007.

65. Dafoe NJ, Zamani A, Ekramoddoullah AKM, Lippert D, Bohlmann J, Constabel CP: **Analysis of the Poplar Phloem Proteome and Its Response to Leaf Wounding.** *J Proteome Res* 2008, **8**:2341-2350.

66. Dafoe NJ, Constabel CP: **Proteomic analysis of hybrid poplar xylem sap.** *Phytochemistry* 2009, **70**:856-853.

67. Du J, Xie HL, Zhang DQ, He XQ, Wang MJ, Li YZ, Cui KM, Lu MZ: **Regeneration of the secondary vascular system in poplar as a novel system to investigate gene expression by a proteomic approach.** *Proteomics* 2006, **6**:881-895.

68. Christopher ME, Miranda M, Major IT, Constabel CP: **Gene expression profiling of systematically wound-induced defenses in hybrid poplar.** *Planta* 2004, **219**:936-947.

69. Gupta P, Duplessis S, White H, Karnosky DF, Martin F, Podila GK: **Gene expression patterns of trembling aspen trees following long-term exposure to interacting elevated CO_2 and tropospheric O_3.** *New Phytol* 2003, **167**:129-142.

70. Bogeat-triboulot MB, Brosché M, Renaut J, Jouve L, Le Thiec D, Fayyaz P, Vinocur B, Witters E, Laukens K, Teichmann T, Altman A, Hausman JF, Polle A, Kangasjärvi J, Dreyer E: **Gradual Soil Water Depletion Results in Reversible Changes of Gene Expression, Protein Profiles, Ecophysiology, and Growth Performance in *Populus euphratica*, a poplar growing in Arid Regions.** *Plant Physiol* 2007, **143**:876-892.

71. Guerra F, Duplessis S, Kohler A, Martin F, Tapia J, Lebed P, Zamudio F, Gonzàlez E: **Gene expression analysis of *Populus deltoides* roots subjected to copper stress.** *Env Exp Bot* 2009, **67**:335-344.

72. Phobius: **A combined transmembrane topology and signal peptide predictor.** [http://phobius.sbc.su.se/].

73. Hu X, Reddy ASN: **Cloning and expression of a PR5-like protein from *Arabidopsis*: inhibition of fungal growth by bacterially expressed protein.** *Plant Mol Biol* 1997, **34**:949-959.

74. Garcia-Casado G, Collada C, Allona I, Soto A, Casado R, Rodriguez-Cerezo E, Gomez L, Aragoncillo C: **Characterization of an apoplastic basic thaumatin-like protein from recalcitrant chestnut seeds.** *Physiol Plant* 2000, **110**:172-180.

75. Wang L, Duman JG: **A Thaumatin-like Protein from Larvae of the Beetle *Dendroides canadensis* Enhances the Activity of Antifreeze Proteins.** *Biochemistry* 2006, **45**:1278-1284.

76. Krebitz M, Wagner B, Ferreira F, Peterbauer C, Campillo N, Witty M, Kolarich D, Steinkellner H, Scheiner O, Breiteneder H: **Plant-based Heterologous Expression of Mal d 2, a Thaumatin-like Protein and Allergen of Apple (*Malus domestica*), and its Characterization as an Antifungal Protein.** *J Mol Biol* 2003, **329**:721-730.

77. Barre A, Peumans WJ, Menu-Bouaouiche L, Van Damme EJM, May GD, Fernandez Herrera A, Van Leuven F, Rougé P: **Purification and structural analysis of an abundant thaumatin-like protein from ripe banana fruit.** *Planta* 2000, **211**:791-799.

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

79. Grenier J, Potvin C, Trudel J, Asselin A: **Some thaumatin-like proteins hydrolyse polymeric β-1,3-glucans.** *Plant J* 1999, **19**:473-480.

80. Woloshuk CP, Meulenhoff JS, Sela-Buurlage M, van den Elzen PJM, Cornelissen BJC: **Pathogen-Induced Proteins with Inhibitory Activity toward *Phytophtora infestans*.** *Plant Cell* 1991, **3**:619-628.

81. Narasimhan ML, Damsz B, Coca MA, Ibeas JI, Yun DJ, Pardo JM, Hasegawa PM, Bressan RA: **A plant Defense Response Effector Induces Microbial apoptosis.** *Mol Cell* 2001, **8**:921-930.

82. D'Angeli S, Altamura MM: **Osmotin induces cold protection in olive trees by affecting programmed cell death and cytoskeleton organisation.** *Planta* 2007, **225**:1147-1163.

83. Abad LR, D'Urzo MP, Liu D, Narasimhan ML, Reuveni M, Zhu JK, Niu X, Singh NK, Hasegawa PM, Bressan RA: **Antifungal activity of tobacco osmotin has specificity and involves plasma membrane permeabilization.** *Plant Sci* 1996, **118**:11-23.

84. Ibeas JI, Yun DJ, Damsz B, Narasimhan ML, Uesono Y, Ribas JC, Lee H, Hasegawa PM, Bressan RA, Pardo JM: **Resistance to the plant PR-5 protein osmotin in the model fungus *Saccharomyces cerevisiae* is mediated by the regulatory effects of SSD1 on cell wall composition.** *Plant J* 2001, **25**:271-280.

85. Min K, Ha SC, Hasegawa PM, Bressan RA, Yun DJ, Kim KK: **Crystal Structure of Osmotin, a Plant Antifungal Protein.** *Proteins* 2004, **54**:170-173.

86. Koiwa H, Kato H, Nakatsu T, Oda J, Yamada Y, Sato F: **Purification and Characterization of Tobacco Pathogenesis-Related Protein PR-5d, an Antifungal Thaumatin-like Protein.** *Plant Cell Physiol* 1997, **38**:783-791.

87. Velazhahan R, Muthukrishnan S: **Transgenic tobacco plants constitutively overexpressing a rice thaumatin-like protein (PR-5) show enhanced resistance to *Alternaria alternata*.** *Biol Plant* 2003, **47**:347-354.

88. Dall'Antonia Y, Pavkov T, Fuchs H, Breiteneder H, Keller W: **Crystallization and preliminary structure determination of the plant food allergen Pru av 2.** *Acta Cryst* 2005, **61**:186-188.

89. Newton SS, Duman JG: **An osmotin-like cryoprotective protein from the bittersweet nightshade *Solanum dulcamara*.** *Plant Mol Biol* 2000, **44**:581-589.

90. Fagoaga C, Rodrigo I, Conejero V, Hinarejos C, Tuset JJ, Arnau J, Pina JA, Navarro L, Peña L: **Increased tolerance to *Phytophthora citrophthora* in transgenic orange plants constitutively expressing a tomato pathogenesis related protein PR-5.** *Mol Breeding* 2001, **7**:175-185.

91. Pressey R: **Two isoforms of NP24: A Thaumatin-Like Protein in Tomato Fruit.** *Phytochemistry* 1997, **44**:1241-1245.

92. Campos M de A, Silva MS, Magalhaes CP, Ribeiro SG, Sarto RPD, Vieira EA, Grossi de Sa MF: **Expression in *Escherichia coli*, purification, refolding and antifungal activity of an osmotin from *Solanum nigrum*.** *Microbe Cell Fact* 2008, **7**:7.

93. Ogata CM, Gordon PF, de Vos AM, Kim SH: **Crystal structure of a sweet tasting protein thaumatin I, at 1.65 Å Resolution.** *J Mol Biol* 1992, **228**:893-908.

94. Schestibratov KA, Dolgov SV: **Transgenic strawberry plants expressing a thaumatin II gene desmonstrate enhanced resistance to *Botrytis cinerea*.** *Sci Hortic* 2005, **106**:177-189.

95. Kuwabara C, Takezawa D, Shimada T, Hamasa T, Fujikawa S, Arakawa K: **Abscisic acid- and cold-induced thaumatin-like protein in winter wheat has an antifungal activity against snow mould, *Microdochium nivale*.** *Physiol Plant* 2002, **115**:101-110.

96. Altincicek B, Knorr E, Vilcinskas A: **Beetle immunity: Identification of immune-inducible genes from the model insect Tribolium castaneum.** *Dev Comp Immunol* 2008, **32**:585-595, .

97. Roberts WK, Selitrennikoff CP: **Zeamatin, an antifungal protein from maize with membrane-permeabilizing activity.** *J Gen Microbiol* 1990, **136**:1771-1778.

98. Huynh QK, Borgmeyer JR, Zobel JF: **Isolation and Characterization of a 22 kDa Protein with Antifungal Properties from Maize Seeds.** *Bioch Biophys Res Com* 1992, **182**:1.

99. Batalia MA, Monzingo AF, Ernst S, Roberts W, Robertus JD: **The crystal structure of the antifungal protein Zeamatin, a member of the thaumatin-like, PR-5 protein family.** *Nat Struct Biol* 1996, **3**:19-23.