Gene families and evolution of trehalose metabolism in plants

John E. Lunn

Max Planck Institute of Molecular Physiology, Am Mühlenberg 1, 14424 Potsdam, Germany.

Email: lunn@mpimp-golm.mpg.de

This paper originates from a presentation at the 8th International Congress of Plant Molecular Biology, Adelaide, Australia, August 2006.

Abstract. The genomes of Arabidopsis thaliana L., rice (Oryza sativa L.) and poplar (Populus trichocarpa Torr. & A.Gray) contain large families of genes encoding trehalose-phosphate synthase (TPS) and trehalose-phosphatase (TPP). The class I subfamily of TPS genes encodes catalytically active TPS enzymes, and is represented by only one or two genes in most species. A. thaliana is atypical in having four class I TPS genes, three of which (AtTPS1–4) encode unusual short isoforms of TPS that appear to be found only in members of the Brassicaceae family. The class II TPS genes encode TPS-like proteins with a C-terminal TPP-like domain, but there is no experimental evidence that they have any enzymatic activity and their function is unknown. Both classes of TPS gene are represented in the genomes of chlorophyte algae (Ostreococcus species) and non-flowering plants [Physcomitrella patens (Hedw.) Bruch & Schimp. (B.S.G.) and Selaginella moellendorffii (Hieron. in Engl. & Prantl)]. This survey shows that the gene families encoding the enzymes of trehalose metabolism are very ancient, pre-dating the divergence of the streptophyte and chlorophyte lineages. It also provides a frame of reference for future studies to elucidate the function of trehalose metabolism in plants.

Additional keywords: trehalose-phosphate synthase, trehalose-phosphatase, trehalase.

Introduction

α,α'-Trehalose (α-D-glucopyranosyl-1,1-α-D-glucopyranoside) is a non-reducing disaccharide that is commonly found in bacteria, fungi and some invertebrates (Elbein 1974). Within the past 10 years it has become clear that many, perhaps all, green plants can also synthesise trehalose, but the function of trehalose metabolism in plants is still something of a mystery. Trehalose has been found in several groups of non-vascular plants, including green algae and liverworts, as well as in lower vascular plants, such as spike-mosses (Selaginella species) and ferns (Anselmino and Gilg 1913; Elbein 1974; Kandler and Hopf 1980). Selaginella lepidophylla (Hook. and Grev.) Spring is a resurrection plant that can recover from almost complete desiccation, and it accumulates high levels of trehalose during dehydration. Trehalose is able to stabilise proteins and membranes in vitro by displacing water molecules and forming a glass-like structure. On the basis of these properties, it has been proposed that the accumulation of trehalose in S. lepidophylla, and some other resurrection plants, helps to protect intracellular structures from damage during anhydrobiosis. Yeast (Saccharomyces cerevisiae) is another desiccation-tolerant organism that accumulates large amounts of trehalose. However, a study of yeast mutants that cannot synthesise trehalose found that the cells were still able to survive desiccation, whereas induction of trehalose accumulation by osmotic stress did not improve desiccation tolerance (Ratnakumar and Tunacliffe 2000). These results, and some previous observations on rotifers (Tunacliffe and Lapinsku 2003), suggest that trehalose accumulation is neither necessary nor sufficient on its own for desiccation tolerance in these organisms, and call into question its role in trehalose-accumulating resurrection plants. As an alternative function, trehalose accumulation has been shown to protect yeast cells from damage by reactive oxygen species (Benaroudj et al. 2001). With the exception of a few resurrection plants, e.g. Mynythhuemum flabellifolium Weib. and Sporobolus antunesiens Kunth (Drennan et al. 1993; Iturriaga et al. 2000), most angiosperms do not accumulate trehalose, and the trace amounts that have been found in some species were sometimes attributable to bacterial or fungal contamination (Kandler and Hopf 1980). In early studies on several desiccation-intolerant flowering plants – tobacco (Nicotiana tabacum L.), potato (Solanum tuberosum L.) and soybean (Glycine max (L.) Merr.) – trehalose levels in untreated plants were below the limits of detection of the assays used (Müller et al. 1995; Goddijn et al. 1997). Therefore, the discovery of genes encoding trehalose-phosphate synthase (TPS; EC 2.4.1.15) and trehalose-phosphatase (TPP; EC 3.1.3.12) in A. thaliana L. – another desiccation-intolerant species – was rather unexpected (Blázquez et al. 1998; Vogel et al. 1998). The complete sequencing of the A. thaliana genome (The Arabidopsis Initiative 2000) subsequently revealed a family of 11 genes (AtTPS1–11) encoding TPS or TPP-like proteins, and a family of 10 genes (AtTPH1–10) encoding TPP, whereas trehalase
(EC 3.2.1.28) appears to be encoded by a single gene (AdTRE) (Leyman et al. 2001).

Phylogenetic analysis of the A. thaliana TPS genes showed that they cluster into two distinct subfamilies, designated class I (AdTPS1–4) and class II (AdTPS5–11) (Leyman et al. 2001).

All of the proteins encoded by these genes contain a glucosyltransferase-like domain similar to the TPS enzymes from yeast (S. cerevisiae) and Escherichia coli (E. coli), but only the A. thaliana proteins appear to encode a TPS activity. Expression of the AdTPS1 or S. lepidophylla TPS1 genes in the yeast tps1Δ (TPS-) mutant showed not only that they can restore trehalose synthesis and growth of the mutant on glucose, but also that they encode active TPS enzymes (Blázquez et al. 1998; Zentella et al. 1999; van Dijck et al. 2002). In contrast, AtTPS7 or AtTPS8 were unable to complement the yeast tps1Δ mutant (Vogel et al. 2001), and AtTPS5 shows no TPS activity (Harthill et al. 2006). The TPS1 proteins from both A. thaliana and S. lepidophylla contain an N-terminal extension compared to the yeast enzyme, and this extension appears to be specific to plant TPSs. The N-terminal extension of the S. lepidophylla TPS1 did not act as an intracellular targeting signal when expressed as a fusion protein with green fluorescent protein in tobacco cells, indicating that the S. lepidophylla TPS1 is likely to be cytosolic (van Dijck et al. 2002). Removal of the extension from the S. lepidophylla TPS1 and A. thaliana TPS1 dramatically increased the catalytic activity of the enzymes when expressed in yeast (van Dijck et al. 2002), suggesting that it acts as an autoinhibitory domain. The class I TPS2-4 isoforms from A. thaliana do not appear to possess this domain.

In addition to the TPS-like domain, the A. thaliana TPS genes encode two isoforms that resemble the phosphate domain of the yeast TPS enzyme encoded by the $\delta$ Trps2 gene, which also contains a non-catalytic, TPS-like domain at the N-terminus (Bell et al. 1998; Leyman et al. 2001). The phosphate domain of all these proteins includes three amino acid motifs that are characteristic of family II phosphatases: motif I $\rightarrow$ DX(D/T)Y(V/I)L/V/V, motif II $\rightarrow$ a Ser or Thr, generally in a hydrophobic context, and motif III $\rightarrow$ KX$_{i-1}$, (G)SD/SXX(D/G)DX$_{i-2}$ (Bell et al. 1998). The most highly conserved, and the invariant first Asp is the functional nucleophile that forms a phospho-acyl intermediate during catalysis (Freudig et al. 2005). The HAD superfamily embraces a wide range of phosphatases and hydrolases, including the analogous enzyme of sucrose synthesis, sucrose-phosphate phosphatase (Lunn et al. 2000), and crystal structures show that the conserved motifs form the active site of the superfamily enzymes (Freudig et al. 2005; Burroughs et al. 2006). The presence of these motifs in the class II isoforms of A. thaliana TPS has prompted speculation that they might have TPS activity (Leyman et al. 2001), but no such activity has been found. Neither AtTPS7 nor AtTPS8 was able to restore growth at 28.6 °C, or trehalose synthesis after heat shock, when expressed in the yeast tps2Δ (TPP+) mutant (Vogel et al. 2001) and AtTPS5 showed no TPS activity (Harthill et al. 2006). Thus, the function of the class II isoforms of TPSs in A. thaliana remains unresolved, and is one of the main puzzles in the enigma of trehalose metabolism in plants.

These proteins are encoded by the 10 A. thaliana TPS genes also contain the three HAD motifs, and two members of the gene family (AtTPP4 and AtTPP9) encode active TPP enzymes that complement the yeast tps2Δ mutant (Vogel et al. 1998), as do two homologues from rice and maize (Pramanik and Imai 2005; Satoh-Nagasawa et al. 2006). The maize TPP (RA3) is of particular interest, because a lesion in the gene encoding this enzyme in the maize (Zea mays) inbred line shows much increased florescence branching (Satoh-Nagasawa et al. 2006). It was proposed that the RA3 gene product could either interfere with a sugar signal that modifies the development of seedlings, or that it acts directly as a transcriptional regulator (Satoh-Nagasawa et al. 2006).

This finding adds to the growing evidence that trehalose metabolism has an important function in plants, even in species that do not accumulate large amounts of this sugar. Loss of TPS1 gene function in A. thaliana is embryo lethal, and leads to growth arrest at the torpedo stage of embryo development (Eastmond et al. 2002). The tps1Δ mutant can be rescued through embryogenesis by inductive expression of TPSs, but the resulting plants show retarded vegetative growth and are unable to flower when TPS expression is no longer induced (van Dijken et al. 2004). Expression of yeast or bacterial TPS or TPP genes in plants also gives rise to striking morphological and biochemical phenotypes (Goddijn et al. 1997; Romero et al. 1997; Piltun-Smittle et al. 1998; Schluempmann et al. 2003). The contrasting phenotypic effects of TPS v. TPP or phosphotrehalase overexpression in A. thaliana suggested that these were caused by changes in the level of trehalose-6-phosphate (Tre6P) – the intermediate of trehalose synthesis – rather than trehalose itself (Schluempmann et al. 2003). The level of Tre6P is correlated with the amount of sucrose in A. thaliana rosette leaves, and addition of 15 mM sucrose to sugar-starved seedlings induced a 26-fold increase in Tre6P (Lunn et al. 2006), consistent with a role for Tre6P in sugar signalling. Increases in Tre6P were also correlated with redox activation of ADP-glucose pyrophosphorylase (Lunn et al. 2006), supporting the hypothesis that one of the functions of Tre6P is to mediate sugar-dependent regulation of starch synthesis (Kolbe et al. 2006). However, this is unlikely to be the only function of Tre6P in plants, because embryos of the A. thaliana tps1Δ mutant show relatively normal synthesis of starch before their growth is arrested, despite the lack of Tre6P (Gomez et al. 2006). The complex morphological phenotypes of TPS and TPP overexpressing plants suggest that changes in Tre6P levels have a far-reaching influence on growth and development, which would be difficult to explain by effects on starch metabolism alone.

Although much current research on trehalose metabolism in plants is focused on discovering the role of Tre6P, we cannot dismiss the possibility that trehalose itself also has a function in plants. Inhibition of trehalose activity in vivo by treatment with valinomycin A led to accumulation of trehalose, and reduced levels of sucrose and starch in flowers, leaves and stems of A. thaliana (Muller et al. 2001), and in root nodules of soybean (G. max) and cowpea (Vigna unguiculata (L.) Walp.) (Muller et al. 1995). Exogenous application of trehalose to plants can induce both abiotic and biotic stress responses, including transcription of genes involved in protection against oxidative stress and pathogen defence (Reignault et al. 2001; Bae et al. 2005a, 2005b). Exogenous trehalose derived from pathogenic and symbiotic microbes also appears to modify the metabolism of their host plants (Muller et al. 1998; Brodmann et al. 2002).
To understand the role of trehalose metabolism in plants, we need to answer two questions: (1) what factors influence the synthesis and breakdown of Tre6P and trehalose in plants? and (2) what are the downstream effects of changes in Tre6P and trehalose? The large diversity of TPS and TPP isoforms in A. thaliana complicates efforts to understand how trehalose synthesis is controlled in this species. In particular, we do not know the significance of the unusual classification of TPS isoforms (TPS2–4), or the function of class II isoforms (TPS5–11) and the need for so many isoforms of TPP. There is also no clear picture of how far this diversity is present in plants other than A. thaliana. The recent completion of genome sequencing for rice (International Rice Genome Sequencing Project 2005) and poplar (Tsukam et al. 2006), and the release of genomic sequences from other species, offer a new opportunity to address some of these questions. This report presents a survey of genome sequences from angiosperms and non-flowering plants to trace the origins and evolution of trehalose metabolism in the plant kingdom, with the aim of helping us to understand its function in plants.

Materials and methods

Materials

The tomato (Solanum lycopersicum L.) TPS1 cDNA clone (cTOD3K4) was a gift from Dr. Alisdair Fernie (Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany). The Physcomitrella patens (Hedw.) Bruch & Schimp. (B.S.G.) cDNA clones (PyTPS1, phppg7108; PyTPS2, phppg1122; PyTPS4, phppg1521 and PyTPR1, phppg1616) (Nishiyama et al. 2003) were obtained from the RIKEN BioResource Center, Japan (http://www.brc.riken.jp/inf/en/, accessed 18 February 2007).

DNA sequencing

Sequencing was carried out on both strands by the dideoxy chain terminal method by Big Dye terminator chemistry (Applied Biosystems, Foster City, CA, USA). The sequences of cDNA clones reported in this paper are available in the NCBI GenBank database with the following accession numbers: EF151131 – S. lycopersicum TPS1, EF151132 – P. patens TPS1, EF151133 – P. patens TPS2, EF151134 – P. patens TPS4 and EF151135 – P. patens TPS6.

Genomic sequence assembly and gene annotation

The rice (Oryza sativa L. subsp. japonica) and poplar (black cottonwood, Populus trichocarpa Torr. & A.Gray) genome sequences were searched for genes with similarity to the TPS, TPP and TRE genes of A. thaliana with the BLASTN and TBLASTN algorithms (Altschul et al. 1990). Each matching sequence was manually curated by checking for consistency with available cDNA and EST sequences, and by comparison of the translated sequence with related protein sequences from A. thaliana and other species. The gene identifiers of sequences that agree with the current annotations of the rice (RAP1; Ohyanagi et al. 2006) and poplar (Release 1.1; http://www.jgi.doe.gov/), accessed 18 February 2007) genomes are shown in Table S1 (see supplementary material). All protein sequences that are not available from public databases are provided in the supplementary material.
Evolution of trehalose metabolism in plants

Fig. 1. Phylogenetic tree of trehalose-phosphate synthase (TPS) sequences from plants. A maximum likelihood tree was constructed with the PHYML program (Guindon et al. 2005), implementing the evolutionary model (Jones et al. 1992) of amino acid substitution, from an alignment of full length plant TPS protein sequences (without gaps) with the Escherichia coli otsA as an outgroup. Bootstrap sampling was carried out with 100 replicates, and internal nodes supported by >70 bootstraps are shown. The tree is displayed as a phylogram rooted with the E. coli otsA. Sequences were obtained from the NCBI GenBank and DOE Joint Genome Institute databases - see supplementary material available on the Functional Plant Biology website for details.
To investigate the origins of the short isoforms of TPS in *A. thaliana*, the search for class I TPS genes was extended to non-flowering plants and algae. The genomes of the prasinophyte algae, *O. lucimarinus* and *O. tauri*, contain single class I TPS genes (OITPS1 and OITPS3) that encode proteins with long N-terminal extensions (136 and 141 amino acids, respectively). The moss *P. patens* has two class I isoforms (PtTPS1 and PtTPS2), which also have N-terminal extensions, although these are only 46 amino acids long. The very low similarity of the N-terminal extensions from these lower plants with the angiosperm sequences makes it difficult to identify any conserved residues in this region with certainty. In contrast, the TPS1 from the spike-moss *S. lepidophylla* is known to have an N-terminal extension that inhibits catalytic activity, and the two closely related sequences from *S. moellendorfii* (SmTPS2 and SmWT4P52) also have N-terminal extensions containing the putative autoinhibitory domain. These observations show that the plant-specific, N-terminal extension of class I TPSs appeared very early in the evolution of plants, but it may only have acquired its autoinhibitory function after the divergence of vascular (Tracheophyta) and non-vascular plants. It can also be inferred that the lack of an N-terminal extension in the *A. thaliana* TPS-4 isoforms is a derived feature resulting from loss of this region.

The *A. thaliana* TPSs (AtTPS1–4) are adjacent to each other on chromosome 1 of *A. thaliana* (Fig. 2) (n.b. there is no AtTPS2 gene), and at least two nearby genes (AtTg17010 and AtTg17040) downstream of these have homologues downstream of the AtTPS1 gene (Atg78580) (Fig. 2). This co-linearity suggests that the AtTPS2 and AtTPS3 genes could have arisen from a segmental duplication of the AtTPS1 gene region or a whole genome duplication, followed by a tandem duplication giving rise to the AtTPS2–AtTPS3 pair. Close inspection of the AtTPS1 gene suggests that it is unlikely to encode a functional TPS enzyme because two regions corresponding to exons 2 and 3 of the AtTPS2 coding sequence appear to be corrupted. There is no evidence from available ESTs that AtTPS3 is expressed, whereas the AtTPS2 gene is represented by an EST (BE523335) from developing seeds, and microarray data from the AtGeneExpress study (Schmid et al. 2005) confirm that this gene is expressed in developing seeds/siliquecs. The region of chromosome 4 in the immediate vicinity of the AtTPS4 gene does not show any obvious collinearity with the AtTPS1 or AtTPS2/3 regions on chromosome 1 (Fig. 2), suggesting that the AtTPS4 gene might have arisen from a more ancient duplication, or a transposition event that only conserved microsynteny within the TPS gene itself. The AtTPS4 gene is represented by a single EST (BG462642) from a mixed tissue sample, and the AtGeneExpress data indicate that expression is essentially restricted to developing seeds/siliquecs, as with AtTPS2 (Schmid et al. 2005).

The only close homologues of the AtTPS2–4 genes in the NCBI EST and genomic sequence databases were found in genomic sequences from Chinese cabbage [*Brassica rapa L. subsp. pekinensis* (Lour.) Kitam.] and *Boechera stricta* (Graham) Al-Shahbaz (GenBank accession number EU693992), both of which, like *A. thaliana*, belong to the Brassicaceae family. One of the genomic sequences from Chinese cabbage (AC172862) also includes a tandem pair of TPS genes, and shows substantial collinearity with the chromosomal region containing the AtTPS2.
suggest that the genes encode short isoforms of TPS without Chinese cabbage, but analysis of potential open reading frames to have occurred ∼ of the lineage from other flowering plants, but predated the divergence of the Brassicaceae. Genes may have occurred after the divergence of the Brassicaceae and AtTPS3 transposition events that gave rise to the region. The second gene region, and was presumably lost after the duplication event that gave rise to this region. The second genomic sequence from Chinese cabbage (AC189237) shows co-linearity with the genomic sequence from Chinese cabbage (AC189237). Homologous genes are indicated by the same colours: red, F-box family protein; orange, Ku70-like protein (DNA-binding protein); yellow, TPS; green, RHM1 protein involved in rhamnogalacturonan I synthesis; blue, iron ion-binding/isopenicillin-N synthase/oxidoreductase; mauve, unknown protein; purple, unknown protein (contains Src homology 2 domain); black, proton pump interactor 1 kinase/ribokinase; brown, unknown protein.

On current evidence it appears that the duplication or loss of the AtTPS1 gene in heterozygous tps1 mutants causes arrest of embryo development at the torpedo stage (Eastmond et al. 2002), showing that AtTPS2 and AtTPS4 cannot compensate for the loss of AtTPS1 gene function. In contrast, expression of the E. coli otsA gene in heterozygous TPS1/tps1 plants could rescue tps1/tps1 embryos (van Dijken et al. 2004), showing that restoration of Tre6P and/or trehalose synthesis is sufficient to complement the mutation. Further work is needed to establish if the AtTPS2 and AtTPS4 genes are expressed in the same cells of developing embryos (van Dijken et al. 2004), showing that restoration of Tre6P and/or trehalose synthesis is sufficient to complement the mutation. Further work is needed to establish if the AtTPS2 and AtTPS4 genes are expressed in the same cells of developing seeds/siliques as AtTPS1, and if so, whether the transcripts are translated and the resulting proteins have TPS activity. Phenotypic analysis of knockout mutants, particularly tps2/tps4...
double knockouts, could also help to reveal the functions of these unusual isoforms of TPS.

Trehalose-phosphate synthase gene families – class II

The rice and poplar genomes both contain 10 class II TPS genes, slightly more than in A. thaliana, which has seven (AtTPS5-11). All of these were clearly separated from the class I sequences in every phylogenetic tree (Fig. 1), but no consistent tree topology was found within the class II group (Fig. 1), possibly due to the low bootstrap values for several of the branch points within this subfamily (Fig. 1). Separate analyses of the class II sequences alone also resulted in multiple trees with different topologies, depending on which region of the protein sequences was used for the analysis, and the method used for selecting the optimal tree – parsimony, distance, maximum likelihood or Bayesian inference (data not shown). Nevertheless, the following clusters were observed in nearly all trees: (1) AtTPS5/poplar3/popl5, (2) AtTPS6/popl4/popl6/rice3/rice9, (3) AtTPS7/popl7/popl3/popl11/rice10/rice7a/rice29a, (4) AtTPS8-10/popl9/popl10/rice6a/rice9a, (5) AtTPS11/popl12/ricc12/ricc16b, (6) ricc2/ricc8b/ricc9f (Fig. 1, and data not shown). In two out of five equally parsimonious trees, constructed with the C-terminal phosphatase domain, the sequences within the AtTPS8-10 group (cluster 4) were split along species lines. However, this grouping was supported in the other three parsimony trees, and in the distance and maximum likelihood trees from the same region, as well as in all trees constructed using the full-length sequences or the N-terminal glucosyltransferase-like region. The AtTPS5 (1) and AtTPS6 (2) groups were always joined together, and this joint AtTPS5/AtTPS6 group was also clustered with the AtTPS7 (3) group in all trees except those constructed with the C-terminal phosphatase domain only. The AtTPS5/1/popl12 (5) and ricc2/ricc8b/ricc9f (6) groups were clustered together along the respective species trees with the phosphatase domain (data not shown).

The strong support for groups 1-5 in the phylogenetic analyses suggests that the genes within each group are orthologous. Each group includes a pair of closely related genes from P. trichocarpa, which may represent homologues from the two ancestral genomes of this ancient tetraploid species. The presence of at least one rice gene as sister to the eudicot sequences in groups 2, 3 and 4 suggests that their common ancestor diverged from the other TPS genes before the monocot-eudicot split, which is estimated to have occurred ~200 million years ago (Mitchell-Olks and Clausen 2002). Group 1 appears to be eudicot-specific, resulting from a duplication within group 2 after the divergence of the monocots from the eudicots. The prasinophyte algae are thought to represent one of the most basal lineages among the green plants (Derecic et al. 2006), so the presence of class I and class II TPS genes in the genomes of O. tauri and O. lucimarinus (Fig. 1) indicates that they were already present in the streptophyte lineage, which includes land plants, before it diverged from the chlorophyte algal lineage. BLAST searches of the NCBI GenBank database indicate that the plant TPS sequences are most closely related to those from fungi and other eukaryotes, e.g. red algae (Rhodophyta) and Dictyostelium sp. (Myxozoa) (data not shown). This suggests a eukaryotic rather than prokaryotic origin for the plant TPS gene families. Further analyses including more green algae and other groups of photosynthetic eukaryotes (e.g. Rhodophyta or red algae) might reveal whether the class I and class II genes arose by divergence of a common ancestor within the plant lineage, or were acquired independently.

Within the streptophyte plant lineage, the moss P. patens has four closely related class II TPS genes, which account for all of the class II ESTs from this species (Fig. 1). Two of these class II genes were found in the genomic sequences available from S. moellendorfii, and both of these clusters with those from P. patens. A single class II TPS sequence from the gymnosperm Ginkgo biloba L. usually clustered with this group as well. The relationship between these genes and those from flowering plants is unclear, as there was disagreement between the distance-based trees on the one hand, and the parsimony and maximum likelihood trees on the other. The former tended to place the sequences from non-flowering plants as a sister group to the AtTPS5/AtTPS6 cluster (data not shown), whereas the latter placed them with the AtTPS7 cluster (Fig. 1).

As noted previously, the function of the class II isoforms of TPS in plants is still unresolved. Comparison with crystal structures of the E. coli tnaA (Gibson et al. 2002, 2004) shows that several of the TPS active site residues are less well conserved in the TPS class II proteins than in the class I proteins, which are known to have TPS activity (Table 1). In particular, residues involved in binding the ribosyl and distal phosphate moieties of UDP-glucose differ in the class II proteins, as well as one of the residues involved in Glc6P-binding. The imperfect conservation of active site residues could explain the inability of these proteins to complement the yeast ΔtpS (TPP-) mutant, and their lack of detectable TPS activity (Vogel et al. 2001; Harthill et al. 2006). In contrast, their lack of TPP activity and inability to complement the yeast ΔtpS (TPP-) mutant are more surprising (Vogel et al. 2001; Harthill et al. 2006). Comparison with a crystal structure of the Thermoplasma acidititum TPP, with Tre6P modelled in the active site (Bao et al. 2006), shows that most of the active site residues are highly conserved in the C-terminal phosphatase-like domain of the class II TPS proteins (Table 2). This includes the critical Asp residue (Asp7) that forms a phospho-acyl intermediate during the dephosphorylation reaction. All of the class II TPSs have a Ser in place of Thr45, which is involved in orientating the phosphate group of Tre6P in the correct position for nucleophilic attack by Asp7. However, a similarly conservative substitution is found in the E. coli tnaB (TPP) and most of the plant TPP proteins, suggesting that this difference is unlikely to explain the lack of TPP activity. One of the residues involved in binding the trehalose moiety of Tre6P (His118) is conserved in some TPS class II sequences, but another residue (Lys149) is not conserved at all. For comparison, in plant TPSs, His118 is perfectly conserved and Lys149 is either conserved or, in most cases, conservatively substituted by Arg. This imperfect conservation of residues involved in binding the trehalose moiety of Tre6P might account for the lack of TPP activity in the class II TPS proteins, despite the conservation of the active site HAD motifs. A substitution of Asn for the first Asp (Asp179) residue in HAD motif III, as found in the rice O/S/TPS9a and O/S/TPS9b proteins, is associated with the absence...
of a metal ion cofactor in some members of the HAD superfamily (Burroughs et al. 2006).

With no experimental evidence of any enzymatic activity, we can only speculate on the function of the class II TPS isoforms. One possibility is that they act as regulatory subunits in a heterooligomeric complex with the class I TPS, analogous to the non-catalytic TPS3 and TSLI subunits in the yeast trehalose-synthesising complex (Bell et al. 1998). Another possibility is that they are somehow involved in Tre6P signalling, perhaps as Tre6P-binding proteins. Although we do not know the function(s) of the class II TPS proteins, it seems clear that they are subject to a high degree of regulation. In *A. thaliana* leaves, the *AtTPS*-11 genes all show strong diurnal cycles in transcript abundance, with *AtTPS* peaking at the end of the day, while the others peak at the end of the night (Bläsin et al. 2005).

Expression of several members of the class II subfamily also responds dramatically to changes in sugar levels in seedlings; *AtTPS* is repressed during sugar starvation and induced by sucrose and glucose, whereas *AtTPS*-10 are strongly induced by sucrose starvation and repressed on sugar re-addition (Price et al. 2004; Osuna et al. 2007). These changes in transcript abundance suggest a high degree of regulation at the level of transcription. In addition, several of the class II TPS proteins are the targets of multi-site phosphorylation by sucrose-non-fermenting-1-related protein kinases and calcium-dependent protein kinases (Glimski and Weckwerth 2005; Harthill et al. 2000), that are conserved in the plant proteins are shown in bold. Residues that are present in all of the plant TPS-II or TPS-IIs are shown in three-letter code, whereas variable residues are shown in single-letter code in order of frequency.

### Table 1. Conservation of active site residues in plant trehalose-phosphate synthase (TPS) proteins

Plant TPS class I and TPS class II protein sequences were aligned with the *Escherichia coli* TPS (otsA). Active site residues, identified from otsA crystal structures (Gibson et al. 2002, 2004), that are conserved in the plant proteins are shown in bold. Residues that are present in all of the plant TPS-IIs or TPS-IIs are shown in three-letter code, whereas variable residues are shown in single-letter code in order of frequency.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Atom(s)</th>
<th>otsA</th>
<th>Residue(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDPglucose</td>
<td>N3, O4</td>
<td>Phe340</td>
<td>L,M,I/V, L,LM</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H, O3H</td>
<td>Asp122</td>
<td></td>
</tr>
<tr>
<td>Phosphate</td>
<td>O1</td>
<td>Val367</td>
<td></td>
</tr>
<tr>
<td>Phosphate</td>
<td>O2</td>
<td>Asn365</td>
<td>Asn</td>
</tr>
<tr>
<td>Phosphate</td>
<td>O3</td>
<td>Lys268</td>
<td>Lys</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>Asp362</td>
<td>Asp</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Asn365</td>
<td>Asn</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>His155</td>
<td>His</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Gln186</td>
<td>His, His</td>
</tr>
<tr>
<td>Glucose</td>
<td>O6H</td>
<td>Phe340</td>
<td>Phe, L, M, V, I</td>
</tr>
<tr>
<td>Glucose</td>
<td>O6H</td>
<td>His155</td>
<td>His, His</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Tyr77</td>
<td>Tyr</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>Arg301</td>
<td>Arg, Arg</td>
</tr>
<tr>
<td>Glucose</td>
<td></td>
<td>His133</td>
<td>His</td>
</tr>
</tbody>
</table>

### Table 2. Conservation of trehalose-phosphatase (TPP) active site residues in plant trehalose-phosphate synthase (TPS) class II and TPP proteins

Plant TPS class II and TPP protein sequences were aligned with the *Thermoplasma acidophilum* TA1209 TPP. Active site residues, identified from the *T. acidophilum* TPP crystal structure (Rao et al. 2006), and conserved in the plant proteins are shown in bold. Residues that are present in all of the plant TPS-II or TPP sequences are shown in three-letter code, whereas variable residues are shown in single-letter code in order of frequency.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Atom(s)</th>
<th>otsA</th>
<th>Residue(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>Asp340</td>
<td>Asp, Asp</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>Asp301</td>
<td>Asp</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Val367</td>
<td>Val</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Asn365</td>
<td>Asn</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>Lys268</td>
<td>Lys</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Asp362</td>
<td>Asp</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>Asn365</td>
<td>Asn</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>His155</td>
<td>His</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Gln186</td>
<td>Gln, His</td>
</tr>
<tr>
<td>Glucose</td>
<td>O6H</td>
<td>Phe340</td>
<td>Phe, L, M, V, I</td>
</tr>
<tr>
<td>Glucose</td>
<td>O6H</td>
<td>His155</td>
<td>His, His</td>
</tr>
<tr>
<td>Glucose</td>
<td>O1H</td>
<td>Tyr77</td>
<td>Tyr</td>
</tr>
<tr>
<td>Glucose</td>
<td>O2H</td>
<td>Arg301</td>
<td>Arg, Arg</td>
</tr>
<tr>
<td>Glucose</td>
<td></td>
<td>His133</td>
<td>His</td>
</tr>
</tbody>
</table>

### Notes

1. Interaction with main chain amide and carbonyl.
2. Interaction with main chain amide.
3. Arg in rice TPS2, 1b, 3a, only.
4. Asn in Arabidopsis thaliana TPS3 only.
5. Water-mediated interaction with main chain amide.
6. Water-mediated interaction between phosphate moiety of Glc6P and His133.
that the plant TPP sequences are most closely related to those from bacteria, especially proteobacteria (data not shown). This suggests that the plant TPP genes may have originated from the endosymbiotic ancestor of mitochondria, which is thought to have been similar to present day α-proteobacteria (Brown et al. 2003).

In phylogenetic analyses of the plant TPP protein sequences, no consensus tree topology was discovered when the sequences from non-flowering plants were included, although the angiosperm sequences were consistently divided into two major groups, except for poplar TPP9 and TPP10, which always clustered separately as outliers (Fig. 3). Separate analyses were carried out with the sequences from angiosperms only, and these provided robust support for a fundamental dichotomy within the angiosperm TPP family, with poplar TPP9 and TPP10 as outliers. In contrast to the clear differences in gene structure between the class I and class II TPS gene families, the position of introns is generally conserved in all of the TPP genes, although there appear to be several examples where introns have been lost. The broad conservation of gene structure, and the presence of introns is generally conserved in all of the TPP genes, although there appear to be several examples where introns have been lost. The broad conservation of gene structure, and the presence of A. thaliana, poplar and rice sequences within both of the major groups, indicate that the two subfamilies of TPP genes arose by duplication of a common ancestor before the separation of the monocot and eudicot lineages. There is also evidence of more recent duplications of TPP genes within each of the three species, for example, the A. thaliana TTPB/TPPD, TPE/TPPG, TPF/TPPH and TPI/TPPJ isoforms were consistently paired together. Several pairs of A. thaliana and poplar sequences were always clustered together, e.g. AtTPPI/AtTPPII and poplar TPP3/TPP4, suggesting that these represent orthologous groups, but it was more difficult to identify putative orthologues from rice (Fig. 3). The maize RA3 protein, an active TPP enzyme that is involved in control of inflorescence branching (Satoh-Nagasawa et al. 2006), was also included in the analyses, and belongs to a distinct clade of monocot sequences that includes the rice TPP3 and TPP7a (Fig. 3; see also Satoh-Nagasawa et al. 2006).

The AtTPPII and AtTPPIIIB genes can complement the yeast tps2Δ (TPP−) mutant and have been shown to encode active TPP enzymes (Vogel et al. 1998), as have the rice TPP2α and maize RA3 proteins (Pramanik and Imai 2005; Satoh-Nagasawa et al. 2006). All of the plant TPP proteins show perfect conservation of the three HAD motifs, and good conservation of the other residues involved in binding the trehalose moiety of Tre6P in the T. acidophilum TPP (Kao et al. 2006, Table 2). There is a conservative substitution of Lys149 by Arg in many of the plant TPPs, and Lys111 is substituted by Asn or His in the plant TPPs, and Lys111 is substituted by Asn or His in the plant TPPs, except those from Ostreococcus species which have thr111 by arg119 in the plant TPPs, except those from Ostreococcus species (Met), S. moellendorffii TPS2 (Thr) and poplar TPP9–10 (Arg), might be expected to replace the function of Lys111 in the T. acidophilum TPP. The high degree of conservation of active site residues indicates that all of the plant TPP genes are likely to encode active TPP enzymes. In addition to the HAD phosphate domain, the plant TPPs also have a highly variable N-terminal region of unknown function, although one possibility is that this region could be involved in intracellular targeting of the proteins.

The fact that so many isoforms of TPP is not immediately obvious. However, analysis of the Arabidopsis genome data indicates that the TPP genes in A. thaliana have different spatial and temporal patterns of expression. For example, the transcript abundance of AtTPPII shows a strong diurnal rhythm in rosette leaves, peaking at the end of the night, whereas AtTPPIII peaks in the middle of the day (Blasing et al. 2005), and AtTPPIIIB is highly expressed in flowers and developing seeds/silicates but hardly at all in rosette leaves (Schmid et al. 2005). The RA3 gene in maize shows very specific expression in localised domains subterfacing the ovules is located in the inflorescence (Satoh-Nagasawa et al. 2006). In A. thaliana, salt stress induces expression of several TPP genes in roots (e.g. AtTPPI, AtTPPII, AtTPPIII, AtTPPIV and AtTPPIVB), and low temperature induces expression of other TPP genes in both shoots and roots (e.g. AtTPPIVA). Chilling stress in rice has been shown to induce expression of the rice TPP2α gene, together with transient increases in both TPP activity and the level of trehalose in the roots (Pramanik and Imai 2005). However, at present we do not know if the stress-induced changes in the expression of TPP genes in A. thaliana lead to any change in TPP activity, or altered levels of Tre6P and trehalose. To understand the functions of the many isoforms of TPP will require more detailed expression analyses at the transcript and protein levels, coupled with measurements of TPP activity and the amounts of Tre6P and trehalose. Phenotypic characterisation of A. thaliana knockout mutants could also be informative, although the presence of several pairs of closely related TPP genes in this species suggests there may be considerable redundancy of function, and so double knockout mutants could be needed to see any phenotype.

Trehalase

In contrast to the large TPS and TPP gene families, trehalase is encoded by a single gene in all of the plant and green algal species examined except poplar and P. patens, which both have three TRE genes. Two of the three closely related (>92% identity) poplar, PtTRE1 and PtTRE2, are adjacent to each other on chromosome 1, whereas PtTRE3 is located on chromosome 9. It seems likely that the PtTRE1/PtTRE2 pair resulted from a relatively recent duplication, and that the ancestral gene was homeologous with PtTRE3. The trehalases from several plants cluster together in a separate clade from the flowering plant sequences, with the algal (Ostreococcus species) trehalases in yet another group (Fig. 4). There is a clear monocot/eudicot divide among the trehalases from flowering plants. The plant trehalases are most closely related to those from animals, indicating a eukaryotic origin for the plant TRE genes.

Trehalase activity has been detected in several plant species, including A. thaliana, and is particularly high in legume root nodules (Müller et al. 1995, 2001). Trehalase activity was also found to be strongly induced in roots and hypocotyls of A. thaliana plants infected with the clubroot pathogen Plasmodyctichium brassicae (Brodmann et al. 2002). An apoplastic trehalase from soybean was reported to have a broad pH optimum and to hydrolyse only trehalose and maltose (Müller et al. 1992; Aeschbach et al. 1999), whereas an acid trehalase from Phascolus vulgaris L. root nodules was less specific, hydrolysing sucrose, melibiose, cellobiose and raffinose, in addition to trehalose and maltose (Garcia et al. 2005). Inhibition of endogenous trehalase activity by
Evolution of trehalose metabolism in plants

Fig. 3. Phylogenetic tree of trehalose-phosphatase (TPP) in plants. A maximum likelihood tree was constructed with the PHYML program (Guindon et al. 2005), implementing the JTT evolutionary model of amino acid substitution, from an alignment of full-length TPP protein sequences (without gaps). Bootstrap sampling was carried out with 100 replicates, and internal nodes supported by >70 bootstraps are shown. The tree is displayed as a phylogram rooted with the Ostreococcus species TPPs. Sequences were obtained from the NCBI GenBank and DOE Joint Genome Institute databases – see supplementary material available on the Functional Plant Biology website for details.
validamycin A was reported to increase levels of trehalose in several plant species, indicating that the enzyme normally prevents accumulation of trehalose that could otherwise interfere with its metabolism (Brodmann et al. 2002). Trehalose produced by symbiotic nitrogen-fixing bacteria in legume root nodules appears to induce changes in the carbohydrate metabolism of the plant cells that favour the bacteroids, and the concomitant induction of trehalase in the plant cells may help to maintain a balance between the needs of the bacteroids and those of the host cell (Xie et al. 2003). Further work is needed to fully understand the importance of trehalase in these microbial-plant interactions, and whether it has any role in other plant partnerships such as those with mycorrhizal fungi or endophytic organisms (Secks et al. 1999).

**Concluding comments**

The discovery of *TPS* and *TPP* genes in *A. thaliana* less than 10 years ago (Blázquez et al. 1998; Vogel et al. 1998) has led to a complete reappraisal of the importance of trehalose metabolism in plants. Previously it was thought to be restricted to just a few specialised resurrection plants, but is now known to be widespread, perhaps universal, in plants. However, the functions of trehalose metabolism in plants are less well understood. There is growing evidence that Tre6P, the intermediate of trehalose synthesis, is a signalling molecule that influences both metabolic and developmental processes, but the details are still sketchy. Trehalase itself may also be a signal metabolite, especially in plant-microbe interactions.

Analysis of the *A. thaliana* genome sequence first showed the great diversity of genes encoding the enzymes of trehalose metabolism in plants, in particular the presence of two distinct subfamilies of *TPS* genes (Leyman et al. 2001). In the present study, the phylogenetic analysis has been extended to a much broader range of species, including monocots, non-flowering plants and green algae. This has revealed that the four main gene families related to trehalose metabolism in flowering plants – *TPS* class I, *TPS* class II, *TPP* and *TRE* – have very ancient origins, dating back to before the divergence of the streptophyte and chlorophyte lineages. Both classes of *TPS* genes appear to have a eukaryotic origin, whereas the *TPP* genes may be derived from the endosymbiotic bacterial ancestor of mitochondria. The domain structure of the plant class II TPSs closely resembles that of the yeast TTP enzyme (TPS2), and it is tempting to speculate that the plant class II TPSs might once have had TTP activity, but this became redundant after the acquisition of the prokaryotic type TTP, allowing the plant class II TPSs to evolve a new function.

Proliferation of *TPS* class II and *TPP* genes has occurred independently in several plant lineages, whereas the *TPS* class I and *TRE* gene families are usually much smaller, and often represented by only a single gene. In this respect, *A. thaliana* is atypical in having four *TPS* class I genes, three of which encode the unusual short isoforms of *TPS* that appear to be
restricted to the Brassicaceae. Analysis of the TPF gene families from Arabidopsis thaliana together with rice and poplar showed for the first time that the genes are divided into two subfamilies (Fig. 3), although the differences between the two groups are less distinctive than between the class I and class II TPS genes. This division suggests that there are fundamental differences in the properties and/or functions of the isoforms encoded by the two subfamilies of TPF genes, but at present we do not know what these might be.

In conclusion, this survey of trehalose metabolism-related genes in the genomes of rice, poplar and other more primitive plants has provided new insights into the origins and evolution of trehalose metabolism in plants, and although it raises several new questions, it should provide a framework on which to base future studies into the function(s) of trehalose metabolism in plants.

References


Evolution of trehalose metabolism in plants


Manuscript received 29 November 2006, accepted 11 January 2007